UNIVERSIDAD DE SALAMANCA
DEPARTAMENTO DE ESTADÍSTICA
Doctorado en Estadística MultivarianteAplicada

TESIS DOCTORAL

CONTRIBUCIONES AL ANÁLISIS MULTIVARIANTE DE DATOS PONDERADOS GEOGRÁFICAMENTE

Autor:
JAVIER ANTONIO DE LA HOZ MAESTRE

Directoras:
MARÍA JOSÉ FERNÁNDEZ GÓMEZ
SUSANA LUÍSA DA CUSTÓDIA MACHADO MENDES

Salamanca, España
2022

CONTRIBUCIONES AL ANÁLISIS MULTIVARIANTE DE DATOS PONDERADOS GEOGRÁFICAMENTE

Departamento de Estadística

Universidad de Salamanca

Memoria que, para optar algrado de Doctor en Estadística Multivariante Aplicada por el Departamento de Estadística de la Universidad de Salamanca, presenta:

Javier Antonio De la Hoz Maestre

Salamanca

2022

Departamento de Estadística
Universidad de Salamanca

MARÍA JOSÉ FERNÁNDEZ GÓMEZ

Profesora Titular de Universidad del Departamento de Estadística de laUniversidad de Salamanca

Y

SUSANA LUÍSA DA CUSTÓDIA MACHADO MENDES

Profesora Adjunta del Instituto Politécnico de Leiria (Portugal)

CERTIFICAN:

Que Don Javier Antonio De la Hoz Maestre, Ingeniero Pesquero, ha realizado en el Departamento de Estadística de la Universidad de Salamanca, bajo su dirección, el trabajo para optar el grado de Doctor, cuyo título es: *Contribuciones al Análisis Multivariante de Datos Ponderados Geográficamente*; y para que conste, firman el presente certificado en Salamanca, a 19 de junio 2022.

Fdo.: Mª José Fernández Gómez.

Fdo.: Susana Luísa da Custódia Machado Mendes

"Somos lo que hacemos repetidamente.
La excelencia, entonces no es un acto, sino
un hábito" (Will Durant)

A mi esposa, Iraida Isabel
A mi madre, Magarita q.e.p.d
A mi padre, Luis

"Lo escuché y lo olvidé, lo vi y lo entendí, lo hice y lo aprendí" (Confucio)

AGRADECIMIENTOS

Agradezco a mis tutoras, Doctoras María José Fernández Gómez y Susana Machado Mendes, por el apoyo, dedicación y aporte de conocimientos en el desarrollo de este trabajo.

A la vez es justo reconocer y agradecer la contribución de los profesores del Departamento de Estadística de la Universidad de Salamanca en mi formación doctoral.

A la Universidad del Magdalena porque tuve su respaldo y apoyo durante todo el proceso académico.

A todos mis compañeros de doctorado, en especial a aquellos con los que fraternicé bajo el mismo techo, compartiendo vinos, café, comidas y tiempo de calidad, aún más en tiempos de pandemia.

¡ Muchas Gracias!

RESUMEN

Dentro de la estadística espacial hay una subárea particular denominada modelos ponderados geográficamente. Estos modelos se utilizan en situaciones donde la dependencia y la heterogeneidad espacial se convierte en el mayor foco de investigación. El paradigma de los modelos ponderados geográficamente es amplio y ha incluido una variedad de modelos entre los cuales tenemos la Regresión Ponderada Geográficamente, el Análisis de Componentes Principales Ponderados Geográficamente, el Análisis Discriminante Ponderado Geográficamente y el Análisis de Cluster Ponderado Geográficamente. En este trabajo se, ha realizado una exhaustiva revisión bibliográfica tanto de las técnicas estadísticas que se pueden utilizar para analizar datos ponderados geográficamente como de sus aplicaciones a las diferentes áreas científicas. También se ha revisado el software existente en la actualidad para llevar a cabo la aplicación de estos métodos y se ha desarrollado una herramienta en un entorno informático que permite la utilización de esas técnicas de una manera fácil, amigable y flexible. Para la revisión bibliográfica se propuso una novedosa metodología que fue implementada en una aplicación de código abierto llamada LDAShiny, que utiliza herramientas de aprendizaje automático y modelado de una manera interactiva y fácil de usar. Las matrices resultantes de la modelización de tópicos fueron analizadas mediante técnicas de Análisis Multivariante, en concreto mediante Escalamiento Multidimensional no métrico y HJ-Biplot. Tras la revisión de los programas informáticos que implementan los modelos ponderados geográficamente, se propuso una nueva herramienta de análisis, denominada GeoWeightedModel. Esta se presenta como una interfaz simple e intuitiva en donde los análisis se pueden realizar de forma interactiva ("apuntar y hacer clic") en un navegador web. La aplicación GeoWeightedModel se utilizó para el análisis de datos reales que recogen información para explorar y visualizar la heterogeneidad espacial de las relaciones entre varias variables (a saber, datos sobre la mortalidad por cáncer de pulmón y bronquios y factores de riesgo a nivel de condados de EE.UU, y datos sobre resultados electorales en EE.UU). A partir de los resultados obtenidos, concluimos que la Regresión Ponderada Geográficamente fue la técnica con el mayor número de extensiones y publicaciones (a saber, 3183). Además, el uso de la metodología de revisión propuesta a través del programa LDAShiny, permitió identificar con éxito 22 tópicos de investigación que definen el estado actual de la investigación en el área de los modelos ponderados geográficamente. Los resultados del

escalamiento multidimensional no métrico permitieron validar el etiquetado de los tópicos, al mostrar agrupaciones coherentes y superposición de nodos, lo que indica distribuciones de palabras similares. El HJ-Biplot permitió analizar y visualizar de manera sencilla la distribución por países de los tópicos encontrados. El análisis de datos reales mediante el programa propuesto GeoWeightedModel, puso de manifiesto que es una interesante herramienta para el análisis de datos ponderados geográficamente que no exige que los investigadores aplicados, como usuarios, tengan grandes conocimientos de programación y/o manejo de software. Con la interfaz gráfica desarrollada para el programa GeoWeightedModel se pudo demostrar que todas las acciones necesarias para el proceso de análisis de datos pueden ser accesibles para cualquier usuario, así como extensible a cualquier área de interés.

Palabras claves: Regresión Ponderada Geográficamente, Análisis de Componentes Principales Ponderados Geográficamente, Análisis Discriminante Ponderado Geográficamente, Análisis de Cluster Ponderado Geográficamente, LDAShiny, GeoWeightedModel, HJ-Biplot, NMDS.

ÍNDICE DE CONTENIDO

LISTA DE FIGURAS

LISTA DE TABLAS

CAPÍTULO I.

1. INTRODUCCIÓN Y OBJETIVOS

1.1. INTRODUCCIÓN

Gran parte de la teoría y práctica de la estadística clásica supone que las observaciones son independientes e idénticamente distribuidas (iid), a menudo con distribución normal. Esta suposición permite que los términos de covariación se establezcan en 0, lo que simplifica drásticamente la teoría estadística matemática (Chun & Griffith, 2013).

Los datos espaciales no cumplen la hipótesis de independencia, debido a que normalmente están autocorrelacionados. La autocorrelación espacial se produce cuando el valor observado de una variable, en un lugar determinado depende de los valores de la misma variable observados en lugares vecinos, por ello, con frecuencia, también es llamada dependencia espacial (Getis & Ord, 2010).

Los modelos estadísticos clásicos pueden mostrar problemas de especificación cuando no consideran los efectos espaciales, es decir, cuando no son consideradas ni la dependencia ni la heterogeneidad espacial (Páez, 2006).

La heterogeneidad espacial surge cuando se trabaja con unidades espaciales por ejemplo, continentes, países, municipios, entre otros, en las que un fenómeno se distribuye de manera distinta sobre el espacio. Por eso, este efecto espacial suele estar directamente relacionado con la localización geográfica (Yrigoyen, 2004).

Considerar los efectos espaciales tiene una importante relevancia. Por un lado, la dependencia espacial puede expresar un proceso de influencia entre las unidades de observación o bien, puede ser producto de variables que tienden a agrupar a poblaciones con rasgos comunes en ciertas áreas (Voss et al., 2006).

Por otro lado la violación del supuesto de la independencia de las observaciones genera problemas en los modelos de regresión, puesto que los estimadores de mínimos cuadrados serán sesgados e ineficientes (Anselin, 1988), lo que trae consigo problemas en la estimación y significación de los coeficientes, sus intervalos de confianza, así como también problemas para la bondad de ajuste del modelo (Wooldridge, 2010).

La presencia de heterogeneidad espacial implica variaciones en la media y varianza dependiendo de la ubicación de las observaciones (Kopczewska, 2020), es decir, viola el supuesto de que los parámetros estimados son los mismos a lo largo y ancho de la región analizada. Esta variación puede, sin embargo, intentar modelarse a fin de comprender las razones que la causan.

En resumidas palabras, nuestro interés por los efectos espaciales puede ser metodológico en tanto buscamos evaluar la estabilidad y confiabilidad de nuestras estimaciones, o bien, podemos hacer de esa distribución espacial parte de nuestra pregunta de investigación (Sánchez-Peña, 2012).

Dado que este efecto de heterogeneidad espacial suele producirse conjuntamente con el efecto de autocorrelación espacial (Anselin, 1988), las herramientas utilizadas por la estadística clásica no resultan adecuadas, es por ello que el tratamiento de datos espaciales debe ser manejado bajo el enfoque de la estadística espacial.

El término estadística espacial hace referencia a las técnicas estadísticas en donde se analizan y cuantifican datos espaciales en los que, unidades espaciales, la dependencia espacial y la heterogeneidad espacial se convierte en el mayor foco de investigación (Kopczewska, 2020).

La estadística espacial tiene origen en la decada de 1950 con trabajos dedicados al análisis de la autocorrrelación espacial (Moran, 1948; Geary, 1954), mientras que, desde la Biología y la Ecología, en esa misma década aparecen las formulaciones sobre la variación espacial estacionaria (Whittle, 1954).

La aparición de los sistemas de información geográfica (SIG) ha impulsado una expansión masiva del uso de la estadística espacial en muchas disciplinas (Kitchin & Thrift, 2009). La estadística espacial se utiliza en las Ciencias Económicas, para el estudio de

patrones espaciales de tendencia, ubicación, concentración económica, comercio internacional, procesos de valoración de bienes inmuebles y terrenos, finanzas públicas locales Medio Ambiente, en Sociología (comportamiento en el espacio, patrones étnicos, crimen, tendencias demográficas), Ciencias Políticas (patrones de votación), planificación urbana y Geografía (patrones de relaciones interpersonales, interacciones, reducción de distancia, uso del suelo), transporte (sistemas de transporte, análisis de accidentes y tráfico), historia (cambios socioeconómicos en el tiempo y el espacio), cuidado de la salud y epidemiología (propagación de enfermedades), entre otros campos.

Dentro de la estadística espacial hay una subárea particular denominada modelos ponderados geográficamente (GW, por sus iniciales en inglés). Estos modelos están diseñados para situaciones en las que los datos espaciales no están bien descritos por un modelo estacionario en el espacio (Gollini et al., 2015).

Los modelos GW utilizan una técnica de ponderación de ventana móvil (esta ventana indica el número de observaciones adyacentes que han de tenerse en cuenta para realizar las estimación), en donde para un modelo individual en alguna ubicación de destino, se ponderan todas las observaciones vecinas de acuerdo con alguna función Kernel, por ejemplo, bicuadrada, gausiana o tricubo y luego se aplica localmente el modelo a estos datos ponderados.

El tamaño de la ventana sobre la que se podría aplicar el modelo localizado está controlado por el ancho de banda. Este se puede encontrar de manera óptima, utilizando diferentes métodos como el de la validación cruzada o enfoques relacionados (Gollini et al., 2015).

El paradigma de los modelos ponderados geográficamente es amplio y ha incluido una variedad de estrategias que se pueden utilizar cuando existe la sospecha de heterogeneidad o no estacionariedad en el proceso espacial del estudio. Es así como dentro de los modelos ponderados geográficamente encontramos:

- la Regresión Ponderada Geográficamente (GWR, por sus iniciales en inglés) es una técnica de análisis espacial introducida en la década de 1990 en el campo de la Geografía (Brunsdon et al., 1996),

- el Análisis de Componentes Principales Ponderados Geográficamente (GWPCA, por sus iniciales en inglés), término acuñado por Fotheringham et al. (2002), considerado como una extensión del Análisis de Componentes Principales cuando se tiene en cuenta que la estructura de covariación de los datos no es constante a través del espacio (Harris et al., 2011b),
- el Análisis Discriminante Ponderado Geográficamente (GWDA, por sus iniciales en inglés) (Brunsdon et al., 2007) donde esencialmente, la idea que se asume es que la matriz de varianzas-covarianzas, el vector de medias y las probabilidades a priori no son fijas, sino que dependen de la ubicación espacial, y
- el Análisis de Cluster Ponderado Geográficamente (GWCA, por sus iniciales en inglés) propuesto por Mason & Jacobson en 2007, los cuales incorporan un modelo básico de interacción espacial en la ponderación de pertenencia a un clúster proponiendo el Algoritmo de Agrupamiento Ponderado Geográficamente Difuso (FGWC, por sus iniciales en inglés).

En esta investigación se pretende realizar una exhaustiva revisión bibliográfica de las técnicas estadísticas que se pueden utilizar para analizar datos ponderados geográficamente y sus aplicaciones, así como del software existente para la puesta en práctica de las mismas. Así mismo se pretende desarrollar alguna herramienta en un entorno informático que permita la revisión y puesta en práctica de esas técnicas de una manera fácil y flexible.

Este trabajo de investigación se estructura de la siguiente manera:

En el Capítulo II se realiza una revisión exhaustiva de los modelos ponderados geográficamente utilizando una combinación entre la búsqueda de palabras clave y la búsqueda de citas, con la finalidad de presentar los desarrollos iniciales y extensiones recientes de estos modelos.

En el Capítulo III se presenta una nueva propuesta metodológica, basada en el aprendizaje automático, para la identificación y modelización de tópicos, con el fin de analizar la vasta y creciente literatura científica relativa a la aplicación de los métodos descritos en el Capítulo II. Esta nueva herramienta nos permite no sólo identificar los tópicos sino también las tendencias de las investigaciones aplicadas donde se utilizan estos modelos ponderados geográficamente. Adémas, en este capítulo, proponemos una aplicación de

código abierto llamada LDAShiny, que proporciona una interfaz gráfica de usuario para realizar una revisión bibliográfica utilizando herramientas de aprendizaje automático de una manera interactiva y fácil de usar.

En el Capítulo IV se presenta una revisión de los programas (software) estadísticos conocidos que permiten aplicar los modelos ponderados geográficamente. Como producto de esta revisión, en el Capítulo V se propone un nuevo paquete de análisis, que hemos denominado GeoWeightedModel. El objetivo principal de dicho paquete, consiste en facilitar, a cualquier investigador que no posea un amplio conocimiento de programación, el análisis de datos con métodos ponderados geográficamente. Se presenta con una interfaz simple e intuitiva en donde los análisis se puedan realizar de forma interactiva ("apuntar y hacer clic") en un navegador web.

Por último, se presentan las conclusiones de nuestro trabajo.

1.2. OBJETIVOS

1.2.1 Objetivos Generales

- Realizar una revisión bibliográfica sobre los métodos estadísticos que permiten analizar datos ponderados geográficamente.
- Desarrollar una herramienta informática de uso libre, flexible y de fácil manejo, en un entorno web, que permita al investigador aplicado utilizar de forma sencilla las técnicas de análisis de datos ponderados geográficamente.

1.2.2 Objetivos Específicos

- Realizar una revisión de las diferentes técnicas, tanto univariantes como multivariantes, de análisis de datos ponderados geográficamente, desarrolladas hasta la actualidad.
- Realizar una revisión de cuál es el estado actual y la evolución cronológica de su utilización en las diferentes áreas científicas.

- Realizar una revisión del software estadístico disponible en la actualidad para llevar a cabo el análisis de datos ponderados geográficamente.

CAPÍTULO II.

2. REVISIÓN DE LOS MÉTODOS ESTADÍSTICOS PARA ANALIZAR DATOS PONDERADOS GEOGRÁFICAMENTE

La Ciencia es un proceso de esfuerzo acumulativo donde se crean nuevos conocimientos a partir de la combinación e interpretación de conocimientos existentes, es por esto que la revisión de la literatura se considera parte integral del proceso investigativo en cualquier área científica (Brocke et al., 2009).

El propósito de una revisión es resumido según Paré et al. (2015) en cinco categorías, donde se reflejan los motivos por los cuales se efectúa dicha revisión de la literatura:

(i) identificar lo que se ha escrito sobre el tema;
(ii) determinar en qué medida un área de investigación específica revela tendencias o patrones interpretables;
(iii) agregar hallazgos empíricos relacionados con una pregunta de investigación para apoyar la práctica basada en evidencia;
(iv) generar nuevos marcos y teorías e;
(v) identificar temas o preguntas que requieren más investigación.

Los métodos más utilizados para la realización de búsquedas bibliográficas son la búsqueda de citas hacia delante/hacia atrás y búsquedas de palabras clave (Adams et al., 2007). Ambos tipos de búsquedas tienen limitaciones (de Wildt et al., 2018) e implican un trabajo considerable.

Para la revisión de métodos presentada en este capítulo se utilizó una combinación entre búsqueda de palabras clave y búsqueda de citas.

2.1 REGRESIÓN PONDERADA GEOGRÁFICAMENTE

La primera ley de la Geografía indica que "Todo se relaciona con todo, pero las cosas cercanas se relacionan más que las lejanas" (Tobler, 1970), lo que implica que en procesos espaciales puede no cumplirse la suposición de independencia entre las observaciones asumidas en los modelos de regresión clásicos.

La GWR extiende el método de regresión clásico, permitiendo estimar parámetros locales en vez de globales (Charlton et al., 2006). De este modo, el modelo de regresión lineal múltiple es un caso particular de la GWR cuando se asume que los parámetros son constantes.

La base metodológica de la GWR es la técnica no paramétrica denominada Regresión Local Ponderada, introducida por Cleveland (1979) para el ajuste y suavizado univariado de curvas que luego fue ampliado por Cleveland y Devlin (1988) hacia la forma multivariante.

2.1.1. Modelo de Regresión Ponderada Geográficamente.

Un modelo de regresión básico está compuesto por un vector **Y** correspondiente a la variable dependiente o variable respuesta y una matriz **X** que contiene un conjunto de variables predictores o independientes. La relación entre estos se modeliza como sigue:

$$y_i = \sum_j \boldsymbol{\beta_j} \boldsymbol{X}_{ij} + \boldsymbol{\varepsilon_i} \tag{1}$$

donde los errores $\boldsymbol{\varepsilon_i} \overset{i.i.d}{\sim} N(0, \sigma^2)$

Los subíndices indican la selección de elementos individuales de vectores o matrices, $\boldsymbol{\beta}$ es un vector de coeficientes de regresión y $\boldsymbol{\varepsilon}$ es un vector aleatorio de los errores del modelo. Los coeficientes de la regresión se obtienen, como estimación de máxima verosimilitud,

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{y} \tag{2}$$

Brunsdon et al. (1996), apoyados en la idea de los métodos de regresión no paramétricos, proponen la técnica GWR cuyo modelo de regresión de coeficientes que varían espacialmente es el siguiente:

$$y_i = \sum_j^p \beta_j(u_i, v_i)x_{ij} + \varepsilon_i\,,\ i = 1,2,3,\ldots,n \qquad (3)$$

donde:

y_i son observaciones de la variable respuesta y,

$x_1, x_2 \ldots, x_p$ las variables explicativas,

(u_i, v_i) representan las coordenadas espaciales de la ubicación *i*, y

$\beta_j(u_i, v_i)$ (*j*=1,2,…,*p*) son *p* funciones desconocidas de las ubicaciones geográficas.

De esta forma los parámetros βj constantes se reemplazan por las funciones βj (.,.) de la ubicación (u, v). En términos más simples, estamos explorando si la relación entre las variables predictoras y la variable respuesta está cambiando geográficamente.

2.1.2. Ajuste del modelo de Regresión Ponderada Geográficamente.

Los parámetros $\beta_j(u_i, v_i)$ en el modelo de GWR se estiman localmente mediante el método de los mínimos cuadrados ponderados. Los pesos (**W**) en cada ubicación (u_i, v_i) se toman como una función de la distancia desde (u_i, v_i) a otras ubicaciones donde se recogen las observaciones. Dicha función de ponderación está relacionada con la atenuación de un patrón o proceso con la distancia, al cual se le denomina a menudo "decadencia de la distancia" y se basa en la primera ley de la Geografía formulada por Tobler (1970). Tal se relaciona con la idea de que es más probable que los datos de ubicaciones cercanas sean similares que los datos de ubicaciones más distantes. Otros términos utilizados para la misma función es la de impedancia de inclusión (Koenig, 1980) y la tasa de lapso de distancia (Johnston et al., 2000).

Siguiendo a Fotheringham et al. (2002), el ajuste de la GWR se realiza suponiendo que las ponderaciones en la localización (u_i, v_i) son $w_j(u_i, v_i)$ (*j*=1,2,…,*n*) y los parámetros en la ubicación (u_i, v_i) son estimados minimizando:

$$\sum_{j=1}^{n} w_j\,(u_i, v_i)\left[y_j - \beta_1(u_i, v_i)x_{j1} - \beta_2(u_i, v_i)x_{j2} - \ldots - \beta_p(u_i, v_i)x_{jp}\right]^2 \quad (4)$$

Siendo

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1p} \\ 1 & x_{21} & x_{22} & \ldots & x_{2p} \\ \vdots & & & & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{np} \end{bmatrix}, \qquad \boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\boldsymbol{W}(u_i, v_i) = \begin{bmatrix} w_{i1} & 0 & \ldots & 0 \\ 0 & w_{i2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & w_{in} \end{bmatrix}$$

Luego, de acuerdo con la teoría de los mínimos cuadrados ponderados, los parámetros estimados en (u_i, v_i) son:

$$\hat{\beta}\,(u_i, v_i) = [\boldsymbol{X^T W}(u_i, v_i)\boldsymbol{X}]^{-1}\boldsymbol{X^T W}(u_i, v_i)\boldsymbol{Y} \quad (5)$$

donde:

$\boldsymbol{X}_i^T = \left(x_{i1}, x_{i2}, \ldots x_{ip}\right)$ es la i-ésima fila de $\boldsymbol{X}$, matriz de diseño de covariables dejando la columna inicial de 1 para el intercepto,

$\boldsymbol{Y}$ es el vector de la variable respuesta y

$\boldsymbol{W}$ es la matriz diagonal $n \times n$ de ponderaciones en donde w_{in} corresponde a la ponderación o peso del dato en el punto n en la calibración del modelo alrededor del punto i.

El valor ajustado de y en la localización (u_i, v_i) se obtiene como:

$$\hat{y} = \boldsymbol{x}_i^T[\boldsymbol{X}^T\boldsymbol{W}(u_i, v_i)\boldsymbol{X}]^{-1}\boldsymbol{X}^T\boldsymbol{W}(u_i, v_i)\boldsymbol{Y} \quad (6)$$

Si denotamos respectivamente por $\hat{\boldsymbol{Y}} = \left(\hat{y_1}, \hat{y_2}, \ldots, \hat{y_n}\right)^T$ y $\hat{\boldsymbol{\varepsilon}} = \left(\hat{\varepsilon_1}, \hat{\varepsilon_2}, \ldots, \hat{\varepsilon_n}\right)^T$ a los vectores de valores ajustados de $\boldsymbol{y}$ y de residuos en las n localizaciones(u_i, v_i).

Reescribiendo tenemos,

$$\begin{cases} \hat{\boldsymbol{Y}} = \boldsymbol{LY}; \\ \hat{\varepsilon} = \boldsymbol{Y} - \hat{\boldsymbol{Y}} = (\boldsymbol{I} - \boldsymbol{L})\boldsymbol{Y} \end{cases}$$

donde,

$$\boldsymbol{L} = \begin{bmatrix} \boldsymbol{x}_1^T[\boldsymbol{X}^T\boldsymbol{W}(u_1, v_1)\boldsymbol{X}]^{-1}\boldsymbol{X}^T\boldsymbol{W}(u_1, v_1) \\ \boldsymbol{x}_2^T[\boldsymbol{X}^T\boldsymbol{W}(u_2, v_2)\boldsymbol{X}]^{-1}\boldsymbol{X}^T\boldsymbol{W}(u_2, v_2) \\ \vdots \\ \boldsymbol{x}_n^T[\boldsymbol{X}^T\boldsymbol{W}(u_n, v_n)\boldsymbol{X}]^{-1}\boldsymbol{X}^T\boldsymbol{W}(u_n, v_n) \end{bmatrix}$$

donde, $\boldsymbol{I}$ es una matriz de identidad de orden $\boldsymbol{n}$, y

$\boldsymbol{L}$ es conocida como la matriz de sombrero (*hat matrix*) (Hoaglin & Welsch, 1978).

Los errores estándar locales de las estimaciones de los parámetros de la GWR se obtienen, considerando al reescribir el estimador de los parámetros locales dados en la Ecuación (6), como:

$$\hat{\boldsymbol{\beta}}(u_i, v_i) = \boldsymbol{C}\boldsymbol{y} \tag{7}$$

donde

$$\boldsymbol{C} = [\boldsymbol{X}^T\boldsymbol{W}(u_i, v_i)\boldsymbol{X}]^{-1}\boldsymbol{X}^T\boldsymbol{W}(u_i, v_i) \tag{8}$$

Las varianzas de las estimaciones de los parámetros está dada por:

$$Var\left[\hat{\boldsymbol{\beta}}(u_i, v_i)\right] = \boldsymbol{C}\boldsymbol{C}^T\,\hat{\sigma}^2 \tag{9}$$

donde $\hat{\sigma}^2$ es la suma de cuadrados residual normalizada de la regresión local y que se define como:

$$\hat{\sigma}^2 = \sum_i \frac{(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i)^2}{[n - 2tr(\boldsymbol{L}) - tr(\boldsymbol{L}^T\boldsymbol{L})]} \tag{10}$$

Al denominador de la expresion anterior se le denomina grados efectivos de libertad, mientras que el término $2tr(\boldsymbol{L}) - tr(\boldsymbol{L}^T\boldsymbol{L})$ (tr es la traza) es equivalente al número de parámetros en un modelo de regresión lineal global y puede denominarse el número efectivo

de parámetros en el modelo de GWR local. Debido a que $tr(\boldsymbol{L})$ y $tr(\boldsymbol{L}^T\boldsymbol{L})$ son generalmente muy similares, el número efectivo de parámetros en la regresión local generalmente puede ser aproximado por $tr(\boldsymbol{L})$, lo que ahorra tener que calcular la traza $tr(\boldsymbol{L}^T\boldsymbol{L})$. Una vez que la varianza de cada parámetro estimado se obtiene de la Ecuación (10), los errores estándar se obtienen de:

$$SE\left[\hat{\boldsymbol{\beta}}\,(u_i, v_i)\right] = \sqrt{Var\left[\hat{\boldsymbol{\beta}}\,(u_i, v_i)\right]} \tag{11}$$

Las matrices de ponderación se obtienen a través de una función de distancia denominada función de ponderación o Kernel y se aplican a la variable respuesta y a las covariables. Los requisitos generales para una función de ponderación, según lo establecido por Brunsdon et al. (1998) son:

- tiene que ser una función monótona decreciente para números reales positivos, para garantizar que el peso disminuya a medida que aumenta la distancia;
- $\mathbf{W}(0) = 1$, de modo que los datos observados en el punto de regresión contribuyen completamente a la regresión;
- $\lim_{d\to\infty} \boldsymbol{W}(d) = 0$, para que los datos obtenidos en distancias muy alejadas no contribuyan a la regresión.

La introducción de los pesos en el modelo se deduce del supuesto de autocorrelación espacial, donde se establece que las observaciones más próximas en el espacio son más similares. La autocorrelación espacial, cuando existe, si es ignorada en el modelo de regresión lineal da como resultado errores correlacionados espacialmente, y es, por tanto, una violación de la suposición del modelo de que los errores deben ser independientes e idénticamente distribuidos. Se puede elegir modelizar la autocorrelación espacial ya sea a través del término de error o a través de los coeficientes de regresión, que es el enfoque que utiliza el modelo de GWR (Wheeler, 2019).

Los parámetros de la función Kernel que se consideran en el cálculo de las ponderaciones en la GWR son: la forma de la función, el tipo de Kernel (fijo versus adaptativo) y el tamaño del ancho de banda (γ).

Forma de la función

Podemos distinguir la función Kernel continua, que pondera todas las observaciones de la función Kernel con soporte compacto para el cual el peso de las observaciones es cero más allá de cierta distancia. Sin embargo, la forma de la función solo cambia los resultados ligeramente (Brunsdon et al., 1998).

Dentro de las funciones Kernel continuas tenemos la uniforme, en la que en el modelo global cada observación tiene como peso la unidad. Elegir una función Kernel uniforme significa hacer una regresión de mínimos cuadrados ordinarios en cada punto.

$$w_j(u_i, v_i) = 1 \tag{12}$$

Otras funciones Kernel continuas y quizas las más utilizadas aplicadas dentro del contexto de la GWR son las funciones continuas que producen pesos que disminuyen monótonamente con la distancia, como la función Kernel gaussiana (Ecuación 13) y la función Kernel exponencial (Ecuación 14):

$$w_j(u_i, v_i) = exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{\gamma}\right)^2\right] \tag{13}$$

$$w_j(u_i, v_i) = exp\left(-\frac{d_{ij}}{\gamma}\right) \tag{14}$$

donde:

d_{ij} es la distancia desde (u_i, v_i) a (u_j, v_j) y

γ es el denominado ancho de banda del núcleo que controla la disminución y el rango de correlación espacial para los pesos.

Las alternativas de funciones Kernel con soporte compacto son la función bicuadrada (Ecuación 15), la tricubo (Ecuación 16) y la función Box-Car (Ecuación 17). Esta última maneja un fenómeno continuo de forma discontinua.

$$w_j(u_i, v_i) = \begin{cases} \left[1 - \left(\frac{d_{ij}}{\gamma}\right)^2\right]^2 & \text{si } d_{ij} \leq \gamma \\ 0 & \text{si } d_{ij} > \gamma \end{cases} \tag{15}$$

$$w_j(u_i, v_i) = \begin{cases} \left[1 - \left(\frac{d_{ij}}{\gamma}\right)^3\right]^3 & \text{si } d_{ij} \leq \gamma \\ 0 & \text{si } d_{ij} > \gamma \end{cases} \quad (16)$$

$$w_j(u_i, v_i) = \begin{cases} 1 & \text{si } d_{ij} \leq \gamma \\ 0 & \text{de otra forma} \end{cases} \quad (17)$$

Las funciones Kernel bicuadrada y tricubo proporcionan una función de ponderación continua casi gaussiana hasta la distancia γ y luego pone en cero cualquier punto de datos más allá de γ. Por lo tanto, esto significa que la función es continua hasta que se alcanza un umbral de distancia y luego es constante (cero) más allá de dicho umbral (Brunsdon et al., 1998).

Kernel fijo versus Kernel adaptativo

Según Wheeler (2019) hay básicamente dos tipos generales de funciones Kernel: las fijas y las adaptativas. Las funciones Kernel fijas tienen un γ expresado como una distancia constante. En este sentido, la función Kernel tiene el mismo tamaño espacial independientemente de la densidad de los puntos de datos. Esto podría dar lugar a un número variable de observaciones ponderadas en toda el área de estudio si la función tiene ponderaciones que son cero más allá de una cierta distancia, mientras que las funciones Kernel adaptativas se ajustan a la densidad de los puntos de datos utilizando un γ expresado en función del número de observaciones.

Un Kernel fijo es adecuado para una distribución espacial uniforme de datos, pero no es muy efectivo en el caso de una distribución no homogénea. En áreas de baja densidad, un Kernel fijo demasiado pequeño incluirá muy pocos puntos en la regresión y por tanto la varianza será mayor, mientras que en áreas muy densas, un Kernel fijo pasará por alto las variaciones en una escala más fina y por consiguiente el sesgo será mayor (Wheeler, 2019).

Una de las funciones Kernel adaptativas más populares es la bicuadrada. La función es:

$$w_j(u_i, v_i) = \begin{cases} \left[1 - \left(\frac{d_{ij}}{d_{iN}}\right)^2\right]^2 & \text{si } j \text{ es uno de los N} - \text{ésimos vecinos más cercanos a } i \\ 0 & \text{de otra forma} \end{cases} \quad (18)$$

donde:

d_{iN} es la distancia al N-ésimo vecino más cercano de la ubicación *i* y

N es el número de vecinos espacialmente más cercanos.

Esta función asigna un peso distinto de cero, que disminuye con la distancia, a los puntos dentro del número umbral de vecinos y asigna un peso de cero a los puntos que están más allá de la distancia al N-ésimo vecino más cercano.

Tamaño del ancho de banda $(\boldsymbol{\gamma})$

Tanto la función como el tipo del Kernel seleccionado para el modelo de GWR pueden afectar a los parámetros estimados. Sin embargo, el efecto de la selección del ancho de banda es mucho mayor en dichas estimaciones (Brunsdon et al., 1996, Fotheringham et al., 1998).

Cuando el ancho de banda es grande, más puntos de datos estarán involucrados en la regresión, de modo que la varianza será pequeña, mientras que el sesgo será grande. Si el ancho de banda tiende a infinito, el modelo tenderá a un modelo de regresión global, donde la superficie de los parámetros tenderan a ser planos y se enmascararán las anomalías locales. El empleo de un ancho de banda pequeño, conduce a grandes variaciones en las estimaciones de los coeficientes del modelo, ya que solo se considera una pequeña cantidad de observaciones en cada regresión local (Beenstock & Felsenstein, 2018).

En la GWR se han desarrollado varios métodos para estimar el γ óptimo: el de la Validación Cruzada (VC) sugerido para la regresión local por Cleveland (1979) y para la estimación de la densidad de Kernel (una forma no paramétrica de estimar la función de densidad de probabilidad de una variable aleatoria) (Bowman, 1984) para minimizar el error

de predicción. Además de este, también se desarrollaron métodos de Validación Cruzada Generalizada (Craven & Wahba, 1978) para suavizar splines, el Criterio de Información de Akaike modificado (AIC, Akaike 1973) para equilibrar la bondad de ajuste y la complejidad del modelo (Fotheringham et al., 2002), y el Criterio de Información Bayesiano (BIC) (Schwarz, 1978), utilizado por Nakaya (2001) en la GWR. Sin embargo, de todos los enfoques, el de la Validación Cruzada es el más utilizado, probablemente debido a su uso intensivo en otras áreas relacionadas de la Estadística y a su simplicidad conceptual.

El γ se estima en la VC buscando el γ que minimiza la puntuación de la VC. Tanto la suma de los errores de la VC como la raíz cuadrada del promedio de los errores en términos porcentuales al cuadrado (RMSPE) han sido utilizados como puntuación de la VC.

El ancho de banda del núcleo que minimiza la suma de errores de la VC, se define como:

$$\hat{\gamma} = argmin \sum_{i=1}^{n} \left[y_i - \hat{y}_i (\gamma) \right]^2 \tag{19}$$

donde $\hat{y}_i$ es el valor predicho de la observación *i* con la ubicación de calibración que dejó fuera del conjunto de datos de estimación.

El ancho de banda del Kernel que minimiza el RMSPE es

$$\hat{\gamma} = argmin \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left[y_i - \hat{y}_i (\gamma) \right]^2} \tag{20}$$

La validación cruzada generalizada (VCG) se define como (Loader, 1999):

$$VCG(\gamma) = n \sum_i \frac{\left(y_i - \hat{y}_i(\gamma) \right)^2}{[(n - tr(\boldsymbol{L}))^2]} \tag{21}$$

A diferencia de la VC, el enfoque AIC corregido para estimar γ se basa en minimizar el error de estimación de la variable resultado, no en la predicción de la variable de resultado.

El AIC corregido en la GWR se adopta de la regresión ponderada localmente. En el AIC corregido existe una penalización por la cantidad efectiva de parámetros en el modelo:

$$AIC_c = 2nlog_e(\hat{\sigma}) + nlog_e(2\pi) + n\left[\frac{n+tr(\boldsymbol{L})}{n-2-tr(\boldsymbol{L})}\right] \quad (22)$$

El BIC, a veces denominado el Criterio de Información de Schwartz (SIC) (Schwartz, 1978) se define como:

$$BIC = -2log_e(l) + klog_e(n) \quad (23)$$

donde:

$log_e(l)$ es la función del log-verosimilitud para el modelo y

k es el número de parámetros.

El proceso de selección de ancho de banda es, por lo tanto, un proceso de optimización.

2.1.3. Extensiones de la GWR

En este apartado, se describen las extensiones desarrolladas tras la introducción de la GWR. En algunos casos, estas extensiones provienen del propio modelo, como por ejemplo en el caso del modelo de GWR mixto. Otras buscan remediar ciertos problemas inherentes al modelo básico de GWR como el caso de las que abordan el tema de la colinealidad. Algunas extensiones se preocupan por la naturaleza del término de error.

Otras consideran la heterocedasticidad espacial y las extensiones más recientes se centran en estimar diferentes anchos de banda para los términos de regresión.

Regresión Ponderada Geográficamente Mixta

El modelo Mixto de Regresión Pondearda Geográficamente (MGWR, por sus iniciales en inglés) fue propuesto por Brunsdon et al. (1999) y considera situaciones en las

que algunas variables explicativas pueden influir en la variable respuesta de forma global, mientras que otras lo pueden hacer localmente. En dicho modelo, algunos coeficientes del modelo de GWR en la Ecuación (24) se asume que son fijos, pero otros pueden variar en la región. Después de ajustar el orden de las variables explicativas, un modelo de MGWR tiene la forma:

$$y_i = \sum_{j=1}^{q} \beta_j \, x_{ij} + \sum_{j=q+1}^{p} \beta_j \, (u_i, v_i) x_{ij} + \varepsilon_i, \quad i=1,2,\ldots,n. \tag{24}$$

Al tomar $x_{i1} = 1$ ó igual a $x_{i,q+1}$ para todo *i*, el modelo puede involucrar un intercepto constante o espacialmente variable.

Brunsdon et al (1999) propusieron un método iterativo para estimar los coeficientes del modelo de MGWR. Sin embargo, es computacionalmente costoso, motivo por el cual Mei et al. (2006) propusieron un método de estimación basado en el enfoque de Speckman (1988) que reduce significativamente la sobrecarga computacional. Mei et al. (2006), sugieren además, que cuando se aplica un modelo de MGWR para analizar un conjunto de datos reales, se debería, inicialmente, determinar qué coeficientes se pueden mantener fijos y cuáles no. Consecuentemente, una vez identificados los coeficientes constantes en el modelo MGWR, se pueden estimar tanto los coeficientes constantes como los coeficientes de variación espacial que son importantes para reflejar la no estacionariedad espacial de la relación de regresión.

Regresión Pondearda Geográficamente Heterocedástica

A diferencia del modelo básico de GWR en el que la varianza del término de error se supone fija, en la Regresión Heterocedastica Ponderada Geográficamente (HGWR, por sus iniciales en inglés), propuesta por Fotheringham et al. (2002), se asume que $\varepsilon_i \sim N(0, \sigma^2(u_i, v_i))$, es decir, que la varianza es función de la posición (u_i, v_i) reemplazando la estimación global de la varianza residual en GWR básico con una estimación local. Harris et al. (2011a) extienden aún más modelo de GWR heteroscedástico a una forma predictiva que puede aumentar la precisión de la predicción, así como la precisión de la incertidumbre de la predicción sobre el modelo básico de GWR.

Así como los Modelos Lineales Generalizados (GLM) (Nelder & Wedderburn, 1972) extienden el modelo de regresión lineal básica, Fotheringham et al. (2002) extendieron el modelo GWR a los denominados Modelos Lineales Generalizados Ponderados Geográficamente (GWGLM, por sus iniciales en inglés) que permiten que la variable de respuesta siga una distribución de la familia exponencial tales como, Gausiana, Logística, Poisson, Multinomial y Binomial Negativa.

El modelo GLM tiene tres componentes en concreto, el componente estocástico de la predicción, el predictor lineal que usa variables explicativas y la función de enlace que transforma de forma monótona y continua la esperanza de la respuesta al predictor lineal. Aunque todos los GLM utilizan predictores lineales, se diferencian en los componentes estocásticos y las funciones de enlace (Nelder & Wedderburn, 1972). De forma general el modelo GWGLM se puede escribir como:

$$\eta_i(u_i, v_i) = \sum_j \boldsymbol{\beta}_j (u_i, v_i)\boldsymbol{x}_{ij} \tag{25}$$

El modelo de regresión gaussiana es equivalente al modelo GWR convencional. Aquí, en la Euación 5 se reemplaza $\mu_i = \eta_i(u_i, v_i)$ (función de enlace); el componente estocástico (η_i) viene dado por $y_i \sim N(\mu_i, \sigma^2(u_i, v_i))$.

En el modelo de Regresión Logística Ponderada Geográficamente (GWLR, por sus iniciales en inglés) (Atkinson et al., 2003) la respuesta es dicotómica, esto es, 0 o 1, por lo que μ_i debe estar entre 0 y 1. Esta condición se obtiene utilizando una función logit:

$$y_i \sim Bernoulli[\mu_i]$$

$$log\left(\frac{\mu_i}{1-\mu_i}\right) = \sum_j \beta_j (u_i, v_i)x_{ij} \quad o \quad \mu_i = \frac{1}{1+exp[-\sum_j \beta_j(u_i,v_i)x_{ij}]} \tag{26}$$

Bernoulli [μ_i] denota la distribución de Bernoulli con el parámetro μ_i. Esto simplemente significa que se espera que y_i sea 1 con una probabilidad de μ_i y 0 con una probabilidad de 1 - μ_i.

Nakaya et al. (2005) propusieron el modelo de regresión de Poisson ponderado geográficamente (GWPR, por sus iniciales en inglés) para tratar una variable de respuesta de recuentos, en donde tenemos el componente estocástico y la función de enlace como sigue:

$$y_i \sim Poisson[\mu_i] = Poisson[\exp(\eta_i(u_i, v_i))] \quad (27)$$

$$\log(\mu_i) = \eta_i(u_i, v_i) \text{ o } \mu_i = \exp(\eta_i(u_i, v_i)) \quad (28)$$

El modelo de regresión de Poisson a menudo incluye un término adicional, denominado desplazamiento, ρ_i

$$\log(\mu_i) = \log(\rho_i) + \sum_j \beta_j (u_i, v_i) x_{ij} \quad \text{o} \quad \mu_i = \rho_i \exp[\sum_j \beta_j (u_i, v_i) x_{ij}] \quad (29)$$

La regresión logística multinomial convencional genera estimaciones para el área de estudio en su conjunto, de esta forma las variaciones locales quedan enmascaradas. A diferencia de la regresión logística binaria, que maneja variables dependientes dicotómicas, el modelo de regresión logística multinomial puede analizar más de dos categorías de variables dependientes.

La Regresión Logística Multinomial Geográficamente Ponderada (GWMLR, por sus iniciales en inglés) (Luo & Kanala, 2008) estima los coeficientes para cada ubicación de la muestra, utilizando todas las otras muestras con ponderaciones asignadas, siguiendo el siguiente modelo:

$$log\left[\frac{Pr(y=j)}{Pr(y=r)}\right] = log\left(\frac{P_{ij}}{P_{ir}}\right) = \sum_{k=1}^{K} \beta_{ijk}\, x_{ijk} \quad (30)$$

donde:

r es la r-ésima categoría de referencia y

j una categoría diferente a la de referencia y el subindice *i* indica la ubicación de una muestra.

Da Silva & Rodrigues (2014) desarrollaron una metodología para la GWR para datos que siguen una distribución binomial negativa y lo denotaron como Regresión Binomial Negativa Ponderada Geográficamente (GWNBR, por sus iniciales en inglés). La ventaja de una distribución binomial negativa es la capacidad de modelizar datos con sobredispersión

(Hilbe, 2011) porque este tipo de datos tiene un parámetro adicional, α. Además, esta distribución es una generalización de las distribuciones geométricas y de Poisson, cuando α = 1 y α → 0, respectivamente.

Dada la dificultad asociada con la búsqueda del número efectivo de parámetros, Da Silva & Rodrigues (2014) también desarrollaron el enfoque GWNBRg (GWNBR con metodología α global). La diferencia entre estos dos métodos está en la estimación del parámetro de sobredispersión α. En GWNBRg, α se estima de forma global, lo que genera una estimación sesgada para dicho parámetro. Sin embargo, la simplicidad de este enfoque permite el cálculo del número efectivo de parámetros del modelo. En GWNBR, aunque esta cantidad es desconocida, GWNBR permite que el parámetro α varíe espacialmente.

La regresión binomial negativa global considera una función de enlace logarítmico y el componente estocástico como sigue:

$$y_i \sim BN\left[\rho_i exp\left(\sum_j \boldsymbol{\beta}_j (u_i, v_i) x_{ij}\right), \alpha(u_i, v_i)\right] \quad (31)$$

donde:

ρ_i es el parámetro de desplazamiento y

α es el parámetro de sobredispersión.

Por otro lado, en el modelo de GWNBRg, la estimación del parámetro α se hace globalmente, es decir, se asume que todos los parámetros en el modelo son estacionarios, y se estima una sobredispersión global $\hat{\alpha}$ para ser utilizado en las estimaciones locales $\beta_j(u_i, v_i)$.

Regresión Ponderada Geográficamente y Funciones de Varianza Anisotrópica

El modelo de regresión lineal básica supone, además de la independencia, un proceso de generación que es estacionario (es decir, homogéneo o invariante espacialmente) e isótropo (es decir, regular o invariante en todas las direcciones). Páez (2004) cuestionó el supuesto de isotropía y menciona que hay razones para creer que la anisotropía puede ser una situación bastante común en muchos conjuntos de datos espaciales. Motivado por lo anterior propuso una extensión que permite la investigación de superficies paramétricas no estacionarias que pueden variar a diferentes velocidades en diferentes direcciones con

respecto a una ubicación dada. En este sentido, Páez concluye que la estimación de modelos con funciones de Kernel anisotrópicas requiere solo modificaciones menores de los métodos existentes utilizados en el caso de las funciones de varianza isotrópica (o Kernel).

Regresión Ponderada Geográficamente Bayesiana

Fotheringham et al. (2002) mencionan que, entre los problemas más importantes que surgen en la modelación de la GWR, el que la presencia de valores atípicos distorsionaría los resultados de los puntos de predicción. Estos autores recomendaron eliminar el conjunto de datos atípicos y luego volver a ajustar el modelo. Si bien tal detección y remoción puede ser útil para tratar los valores atípicos comunes, no puede manejar las observaciones o influencias locales causadas por los efectos de la comunidad, como lo son los valores atípicos de la regresión local y la función de ponderación de descenso de la distancia (LeSage, 2004).

Además del problema de los valores atípicos, LeSage (2004) también identificó otros problemas relacionados con la GWR. Uno de ellos estaba relacionado con la validez de las inferencias para los parámetros de regresión con los enfoques de mínimos cuadrados clásicos y otro es el problema de los "datos débiles" (es decir, que el número efectivo de observaciones para cada estimación puede ser demasiado pequeño porque tenga insuficientes vecinos dentro de su ancho de banda). Por lo anterior propuso el enfoque bayesiano del modelo de GWR al que denominó Regresión Ponderada Geográficamente Bayesiana (BGWR por sus iniciales en inglés) (Subedi et al., 2018).

El modelo subyacente de BGWR es el mismo que el del modelo de GWR en la Ecuación 3. No obstante, el término de error del modelo asume que $\varepsilon_i \overset{i.i.d}{\sim} N(0, \sigma^2, V_i)$, donde V_i es una matriz diagonal n × n que representa un conjunto de variaciones no constantes a lo largo del espacio y debe estimarse a partir de los datos (Shing, 2008).

El modelo de BGWR requiere el uso de Método de Monte Carlo basado en cadenas de Markov, denominado muestreo de Gibbs.

Extensiones Centradas en la Estimación de Anchos de Banda

Dentro los estudios para seleccionar el ancho de banda en el modelo de GWR se incluyen Páez (2004) para ancho de banda anisotrópico, Farber & Páez (2007) para una selección robusta de ancho de banda y Brunsdon et al. (1999) y Nakaya et al. (2005) en el modelo de GWR mixto.

Yang (2014) propuso un modelo de GWR con banda flexible (FB-GWR) o GWR multiescala que reemplaza el ancho de banda común con anchos de banda para cada término. Este modelo FB-GWR extiende el modelo de GWR mediante la sustitución de $\beta_j(u_i, v_i)$ en la Ecuación (4) por $\beta_{jbw}(u_i, v_i)$. Esta nueva notación indica que un ancho de banda diferente es permisible para cada parámetro local estimado asociado a una ubicación, lo que indica que no solo los coeficientes de regresión varían en diferentes ubicaciones, sino que la extensión implica que también varían en diferentes escalas espaciales para cada variable independiente.

Fotheringham et al. (2015) propusieron una extensión del modelo de GWR involucrando un ancho de banda espaciotemporal a la que denominaron Regresión Ponderada Geográfica y Temporalmente (GTWR, por sus iniciales en inglés) a través del desarrollo del concepto de ancho de banda en el modelo de GWR al incluir una dimensión temporal. Específicamente, propusieron una función espatio-temporal Kernel y desarrollaron un procedimiento para elegir el ancho de banda espatio-temporal óptimo. El modelo GTWR se distingue del básico en la forma en la que se construye la matriz de pesos **W**, pues el modelo GTWR captura los efectos espaciales y temporales de las observaciones cercanas tanto en el espacio como en el tiempo.

Lu et al. (2017) propusieron un modelo de GWR con métricas de distancia específicas para cada parámetro (Métricas de Parámetros de Distancia Específica PSDM GWR, por sus iniciales en inglés), que especifica anchos de banda específicos para cada parámetro. Estos autores extienden el modelo de GWR a varias métricas de distancia no euclidianas en lugar de la distancia euclídea predeterminada que se suele emplear para el esquema de ponderación. Según Lu et al. (2017) el modelo puede considerarse como una fusión de los modelos de

GWR métricos con distancia no euclídea de Lu et al. (2011, 2014a) y el modelo FB-GWR de Yang (2014). El modelo PSDM GWR puede ser expresado generalmente de la siguiente forma:

$$y_i = \beta_{0i}^{(DM_0, bw_0)} + \sum_{k=1}^{m} \beta_{ki}^{(DM_k, bw_k)} x_{ik} + \varepsilon_i \quad (32)$$

donde DM_k y bw_k ($k = 0, 1, \cdots, m$) representa la métrica de distancia específica y el ancho de banda para cada estimación del parámetro de cada variable independiente (e intercepto).

Modelos de Regresión Penalizados Ponderados Geográficamente

La regresión Ridge (Hoerl & Kennard, 1970a, 1970b) y la regresión Lasso (Tibshirani, 1996) son métodos de regresión penalizados, que ponen una restricción en los coeficientes de regresión para reducir la multicolinealidad. La integración de los términos de regresión regularizados en el marco de la GWR puede mejorar el modelo al minimizar la multicolinealidad. Sin embargo, ambas regresiones tienen sus propias desventajas (Zou & Hastie, 2005). La regresión Ridge puede reducir continuamente los valores de los coeficientes, pero no puede eliminar ninguna variable mediante el establecimiento de su coeficiente en cero. Como resultado, la regresión de Ridge siempre mantiene todas las variables explicativas en el modelo y no puede ofrecer un modelo parsimonioso. En contraste, la regresión Lasso se puede usar para seleccionar variables mientras minimiza la pérdida de precisión de la predicción.

Teniendo en cuenta lo anterior, Zou & Hastie (2005) desarrollaron un método denominado Elastic Net (EN, por sus iniciales en inglés) que puede considerarse como un híbrido de las regresiones Rigde y Lasso. Por lo tanto, el método EN sirve tanto para seleccionar automáticamente las variables (como se hace con el método Lasso), así como mantener un alto rendimiento de predicción cuando existe colinealidad (al igual que hace el método Ridge).

Las versiones penalizadas de la GWR, Regresión Ridge Ponderadas Geográficamente (GWRR, por sus iniciales en inglés; Wheeler, 2007), la Regresión Lasso Ponderada

Geográficamente (GWL, por sus iniciales en inglés; Wheeler, 2009) y la Regresión Elástica Ponderada Geográficamente (GWEN, por sus iniciales en inglés; Li & Lam, 2018), pueden estabilizar los coeficientes de regresión y, por tanto, mejorar la interpretación de los resultados de una GWR (Wheeler, 2007, 2009).

Los coeficientes de la GWRR, minimizan la suma de una penalización sobre la suma residual de cuadrados y el tamaño de los coeficientes al cuadrado:

$$\hat{\beta}^{R} = argmin_{\beta} \left\{ \sum_{j=1}^{n} \left(y_j - \beta_0 - \sum_{i=1}^{p} \beta_i \, x_{ij} \right)^2 + \lambda \sum_{i=1}^{p} \beta_i^2 \right\} \quad (33)$$

Los coeficientes de la GWL minimizan la suma de cuadrados residual y el valor absoluto de los coeficientes:

$$\hat{\beta}^{L} = argmin_{\beta} \left\{ \sum_{j=1}^{n} \left(y_j - \beta_0 - \sum_{i=1}^{p} \beta_i \, x_{ij} \right)^2 + \lambda \sum_{i=1}^{p} |\beta_i| \right\} \quad (34)$$

donde λ es un parámetro de regularización que determina el número de variables explicativas seleccionadas (con coeficientes distintos de cero).

El parámetro λ es 0 cuando todas las variables están seleccionadas. A medida que aumenta λ, aumenta el número de coeficientes β_i que son cero, lo que significa que el número de variables seleccionadas disminuye hasta alcanzar un mínimo cuando no se seleccionan variables.

El término de penalización del modelo GWEN es una compensación entre la penalización del modelo GWL (la penalización Lasso a menudo denominada penalización L1) y la penalización del modelo de GWRR (a menudo denominada penalización L2). Li & Lam (2018) muestran que el modelo GWEN reune las ventajas tanto de la alta exactitud de predicción de la GWR como de la baja multicolinearidad propia de las variables explicativas del EN:

$$\hat{\beta}^{EN} = argmin_{\beta} \left\{ \sum_{j=1}^{n} \left(y_j - \beta_0 - \sum_{i=1}^{p} \beta_i \, x_{ij} \right)^2 + \lambda \sum_{i=1}^{p} \left((1-\alpha)\beta_i^2 + \alpha|\beta_i| \right) \right\} \quad (35)$$

donde α, que varía entre 0 a 1, es el parámetro que determina si el modelo de regresión se parece más a una regresión Lasso o a una Ridge.

Comber & Harris (2018) desarrollaron una Regresión Logística Elastic Net Ponderada Geográficamente (GW-ENLR por sus iniciales en inglés). La implementación de este modelo GWENLR simplemente requiere la asignación de una ENLR a cada subconjunto de datos ponderados geográficamente específicos de la ubicación. La expresión del modelo es como sigue:

$$\min_{\beta_{i0},\beta_i} -\left[\frac{1}{N}\sum_{i=1}^{N} y_{i1}\left(\beta_{i0}+\beta_i x_i^T\right) - log\left(1+exp^{(\beta_{i0}+\beta_i x_i^T)}\right)\right]+\lambda\left[\frac{1}{2}(1-\alpha)\parallel\beta_i\parallel_2^2+\parallel\beta_i\parallel_1\right] \quad (36)$$

Regresión Cuantílica y M-Cuantílica Ponderada Geográficamente

El objetivo de los métodos de regresión clásicos consiste en minimizar la suma de los cuadrados de los residuos y utilizar la media como estimador. La regresión cuantílica (QR por sus iniciales en inglés) (Koenker & Bassett, 1978) busca minimizar una suma de errores absolutos ponderados, con pesos asimétricos, y utiliza los cuantiles como estimadores. Breckling & Chambers (1988) proporcionan una generalización adicional de la regresión por cuantiles mediante la regresión por M-cuantiles.

Salvati et al. (2009, 2010) exploraron el uso de la GWR en la estimación de áreas pequeñas con enfoque de un modelo local para los cuantiles M de la distribución condicional de la variable respuesta dadas las covariables. El modelo de GWR de M-cuantil se centra en áreas pequeñas e integra los conceptos de estimación de área pequeña robusta y atípica en el espacio dentro de un marco de modelado unificado.

Chen et al. (2012) extendieron el modelo de GWR a un ajuste de QR con la especificación del siguiente modelo:

$$y_i = \boldsymbol{\beta}_0^{\tau}(u_i,v_i) + \sum_{k=1}^{p}\boldsymbol{\beta}_k^{\tau}(u_i,v_i)x_{ik} + \varepsilon_i^{\tau} \quad (37)$$

donde:

el cuantil τ-ésimo del término de error ε_i dado $\mathbf{X}_i$ es cero, y

el vector $\boldsymbol{\beta}^{\tau}(u_i, v_i)$ es el vector de los coeficientes de regresión cuantil local ($0<\tau<1$) en la ubicación (u_i, v_i).

Los coeficientes QR $\boldsymbol{\beta}_k^{\tau}(u_i, v_i)$ miden el cambio en un cuantil τ específico de la variable de respuesta **Y** correspondiente a un cambio de unidad en la variable independiente $\mathbf{X}_k$. Chen et al. (2012) afirman que esta conceptualización permite una comparación de cómo algunos percentiles de una variable respuesta pueden verse más (o menos) afectados por ciertas variables independientes.

Wang et al. (2018) extienden la GWQR a una versión penalizada a la que denominaron Regresión Lasso Cuantílica Ponderada Geográficamente (GWQLR, por sus iniciales en inglés). El método propuesto combina la estimación local-lineal del modelo GWQR y el grupo Lasso adaptativo, que puede identificar simultáneamente coeficientes que varían espacialmente, coeficientes constantes no-cero y coeficientes cero.

Regresión Ponderada Geográficamente Kriging

La regresión Kriging, acuñada por primera vez por Odeha et al. (1994), es una técnica de predicción espacial, que puede combinar diferentes modelos de regresión para generar muchos métodos combinados (Li & Heap, 2014) entre los cuales se encuentra el de regresión ponderada geográficamente Kriging (GWRK, por sus iniciales en inglés). Este modelo de GWRK combina dos componentes, uno determinista y otro estocástico, donde ambos se modelizan por separado (Kumar et al., 2012a). El componente determinista se modeliza mediante una GWR y utiliza la información de covariables disponible para predecir la tendencia de una variable objetivo. El componente estocástico (residuos) se puede interpolar con Kriging y se suma a la tendencia estimada. Los residuos se consideran los errores, y es posible que los errores tengan alguna estructura de correlación espacial que se pueda modelizar. Se podría considerar que los errores son el componente del modelo que no puede ser explicado por el componente determinista, pero es importante agregarlo ya que ayuda a explicar la variación de la variable objetivo a través del espacio. El enfoque del modelo de GWRK se puede utilizar para estimar los residuos y para predecir la tendencia de manera más eficiente (Harris et al., 2011a).

Regresión Beta Ponderada Geográficamente

El uso de una distribución beta en un modelo de regresión se remonta a Falls (1974). Sin embargo, la estructura de datos restringidos al intervalo (0, 1) fue definido por Ferrari & Cribari-Neto (2004). Aunque la distribución beta no pertenece a los modelos lineales generalizados (GLM), la idea de Ferrari & Cribari-Neto era crear una estructura similar a lo que denominaron regresión beta. Da Silva & de Oliveira (2017) extendieron los conceptos de regresión beta al modelo de GWR, y propusieron la Regresión Beta Ponderada Geográficamente (GWBR por sus iniciales en inglés) y la Regresión Beta Ponderada Geográficamente global (GWBRg, por sus iniciales en inglés), para modelizar datos de tasa o proporción restringidos al intervalo (0, 1) en un contexto espacial.

Para el modelo de GWBR, Da Silva & de Oliveira (2017) propusieron que el promedio de la variable respuesta en la ubicación *j* se puede modelar como:

$$g(\mu_j) = \sum_k^p \boldsymbol{\beta}_k (u_j, v_j) x_{jk} + \varepsilon_j \quad (38)$$

donde,

g (.) es una función de enlace que asocia el intervalo (0, 1) a $\mathbb{R}$;

(u_j, v_j) representa las coordenadas geográficas de la observación j-ésima, j = 1, ..., n;

$\beta_k (u_j, v_j)$ es el parámetro para la k-ésima variable explicativa en función de la ubicación de la j-ésima observación;

x_{jk} es el valor de la variable explicativa k-ésima para la ubicación *j*;

ε_j es el error.

Tal como señalaron Da Silva & Rodrigues (2014), es posible estimar los parámetros en cualquier punto de regresión *i*. Sin embargo, se calculan las medias estimadas sólo para los puntos observados j donde se conoce la información necesaria x_{jk}.

Debido a la dificultad asociada a la búsqueda del número efectivo de parámetros, Da Silva & De Oliveira (2017) desarrollaron el modelo de GWBRg que se diferencia del GWBR en la

estimación del parámetro de precisión φ en donde este parámetro se estima globalmente en lugar de variar en el espacio.

Regresión Inversa Ponderada Geográficamente Tridimensional

Chybicki (2017), presenta la técnica de GWR tridimensional (3GWR) que une el modelo de regresión local ponderada geográficamente (LGWR, por sus iniciales en inglés), con una optimización inversa dependiente de la profundidad, es decir, se aplica la regresión local no solo en el plano geográfico, sino también en el plano vertical (profundidad) de la región analizada.

Regresión Ordinal Ponderada Geográficamente

Dong et al. (2018) propusieron el modelo de Regresión Ordinal Ponderada Geográficamente (GWOR, por sus iniciales en inglés) ofreciendo una herramienta estadística adecuada para analizar datos espaciales con respuestas categóricas ordinales. El modelo, siguiendo la notación del modelo de GWR (Fotheringham et al., 2002) es:

$$Y_i^* = x_i\beta(u_i) + \varepsilon_i; i = 1\ldots N \quad (39)$$

$$Y_i = 1\, si\ -\infty < Y_i^* \leq a_1(u_i) \quad (40)$$

$$Y_i = j\, si\, a_{j-1} < Y_i^* \leq a_j(u_i),\, j = 2\ldots J-1 \quad (41)$$

$$Y_i = J\, si\, a_{J-1} < Y_i^* \leq -\infty \quad (42)$$

donde:

Y_i^* denota una variable continua latente de la respuesta categórica observada Yi que depende de un conjunto de puntos de corte o valores umbral a, y

J es el número de categorias de respuesta y x_i es un vector fila del conjunto de variables predictoras.

Tanto los puntos de corte J − 1 como el vector de coeficiente coeficientes β a estimar, están asociados con un indicador de ubicación u_i. Siguiendo extensiones previas, Dong et al.

(2018) emplean un proceso de verosimilitud local ponderado localmente para estimar los parámetros en el modelo de GWOR.

2.2. ANÁLISIS DE COMPONENTES PRINCIPALES PONDERADAS GEOGRÁFICAMENTE

El análisis de componentes principales (PCA, por sus iniciales en inglés) es quizás el más antiguo y el mejor conocido de los métodos multivariantes. Fue introducido por Pearson (1901) y posteriormente desarrollado por Hotelling (1933). Este método es usado comúnmente para explicar la estructura de covariación de un conjunto de datos multivariantes (de alta dimensionalidad) utilizando sólo unas pocas componentes (es decir, proporciona una alternativa de baja dimensión). Su idea es simple, es decir, reducir la dimensión de un conjunto de datos, mientras se conserva la mayor variabilidad como sea posible (Jolliffe & Cadima, 2016).

El PCA convencional ha sido criticado al no considerar las variaciones e ignorar los efectos geográficos y/o espaciales que a menudo son esenciales para una comprensión más completa de datos con estas características. El Análisis de Componentes Principales Ponderadas Geográficamente (GWPCA, por sus iniciales en inglés), término acuñado por Fotheringham et al. (2002), incorpora al PCA convencional los efectos espaciales. En otras palabras, el GWPCA es una extensión del PCA cuando se tiene en cuenta que la estructura de covariación de los datos no es constante a través del espacio (Harris et al., 2011b).

2.2.1. Modelo del Análisis de Componentes Principales Ponderados Geográficamente

En el modelo GWPCA, las componentes se pueden calcular mediante la descomposición de la covarianza local. Primero, asumiendo que cada variable X_i tiene un par de coordenadas en la ubicación *i*, que se representa como $\mathbf{X}\ (u_i, v_i)$ (matriz con n filas para observaciones y m columnas para variables) que se supone que tiene una distribución normal multivariante.

$$x_i|(u,v)N(\mu(u,v),\sum(u,v)) \tag{43}$$

Luego, la matriz de varianzas-covarianzas geográficas se expresa de la siguiente manera:

$$\sum(u_i, v_i) = \boldsymbol{X}^T \boldsymbol{W}(u_i, v_i)\boldsymbol{X} \tag{44}$$

donde:

$\boldsymbol{W}(u_i, v_i)$ es una matriz diagonal de pesos geográficos y, como en cualquier método ponderado, la elección de la función de ponderación Kernel es un tema fundamental (Harris et al., 2014).

Para encontrar las componentes principales locales en la ubicación (u_i, v_i), la descomposición de la matriz de varianzas-covarianzas local proporciona los valores propios locales y los vectores propios locales (o vectores de carga). El producto de la fila *i* de la matriz de datos por los valores propios locales para la ubicación *i* proporciona la fila *i* de las puntuaciones de las componentes locales (Gollini et al., 2015):

$$\boldsymbol{L}_{(u_i,v_i)}\boldsymbol{V}_{(u_i,v_i)}\boldsymbol{L}^T_{(u_i,v_i)} = \boldsymbol{\Sigma}_{(u_i,v_i,)} \tag{45}$$

donde:

$\boldsymbol{L}(u_i, v_i)$ es la matriz local de vectores propios, y

$\boldsymbol{V}(u_i, v_i)$ es la matriz diagonal local de valores propios.

Una matriz de puntuaciones ($\boldsymbol{T}$) se puede calcular como sigue (Lu et al., 2014b):

$$\boldsymbol{T}(u_i, v_i) = \boldsymbol{X}\boldsymbol{L}(u_i, v_i) \tag{46}$$

donde el producto de la fila *i* de la matriz de datos por los vectores propios locales para la ubicación *i* da como resultado la fila *i* de las puntuaciones de las componentes locales.

Si dividimos cada valor propio local por la traza de $\boldsymbol{V}(u_i, v_i)$ se encuentran las versiones local de la proporción de la varianza total en los datos originales por cada componente. Por lo tanto, en cada ubicación observada para un GWPCA con m variables, hay m componentes, m valores propios, m conjuntos de cargas de componentes (cada uno de tamaño m×m) y m conjuntos de puntuaciones de componentes (cada uno de tamaño n×m). También se pueden obtener valores propios y sus vectores propios asociados en ubicaciones no observadas, aunque como no existen datos para estas ubicaciones, no se

pueden obtener puntuaciones de las componentes para ellas. En ecología numérica, cargas y puntuaciones de las componentes se conocen comúnmente como puntuaciones variables (o especies) y puntuaciones del sitio (lugares), respectivamente. Las cargas de componentes son los coeficientes de correlación entre las puntuaciones de las componentes y los datos en bruto (Harris et al. , 2014).

Para describir cómo se selecciona el ancho de banda en un GWPCA, Harris et al. (2011b) utilizaron el método de la validación cruzada basado en las aproximaciones obtenidas con las primeras q componentes principales. En un PCA clásico el subespacio q dimensional está abarcado por las primeras q cargas (también vectores m-dimensionales), y es el subespacio que maximiza la varianza de las proyecciones de datos proyectados en ese subespacio. Para la reducción de la dimensionalidad, q, se elige para que este subespacio contenga una proporción de varianza total alta. Por lo tanto, los componentes q + 1 a m que representan las distancias euclídeas a lo largo de los ejes de los vectores ortogonales correspondientes a un subespacio lineal q-dimensional que representan la desviación de este subespacio. En términos de puntuaciones de componentes, las primeros q componentes se describen mediante $\mathbf{XL}_q$, y las componentes restantes por $\mathbf{XL}_{(-q)}$ (donde $\mathbf{L}_q$ denota la matriz de carga **L** con todas menos las primeras q columnas eliminadas y $\mathbf{L}_{(-q)}$ denota **L** con las primeras q columnas eliminadas).

La mejor aproximación q (según el criterio de los mínimos cuadrados) para **X** es $\mathbf{XL}_q\mathbf{L}_q^T$ y la matriz residual **S**, dada por $\mathbf{S} = \mathbf{X} - \mathbf{XL}_q\mathbf{L}_q^T$, también se puede escribir como $\mathbf{S} = \mathbf{XL}_{(-q)}\ \mathbf{L}_{(-q)}^T$ (Jolliffe, 2002). Así, a través de un análisis de componentes principales, el objetivo es encontrar el mínimo de la expresión:

$$\sum_{ik}([\boldsymbol{X}]_{ik} - [\boldsymbol{S}]_{ik})^2 \tag{47}$$

Para el GWPCA, las componentes principales locales para la ubicación *i* representan una proyección similar, pero con las correspondientes cargas definidas localmente. Es decir, ahora encontramos **S** para minimizar (Harris et al., 2011b):

$$\sum_{ik} w_i\,([\boldsymbol{X}]_{ik} - [\boldsymbol{S}]_{ik})^2. \tag{48}$$

2.2.2. Extensiones de Análisis de Componentes Principales Ponderadas Geográficamente

Análisis de componentes principales ponderadas geográficamente no negativo

Teniendo en cuenta que los vectores propios de un GWPCA pueden variar espacialmente (pero pueden ser negativos, lo cual dificulta la construcción de un índice compuesto multidimensional (MCI), Tsutsumida et al. (2019)) propusieron el GWPCA no negativo (GWnnegPCA, por sus iniciales en inglés) que combina un PCA no negativo y un GWPCA para hacer que los vectores propios locales no sean negativos en cualquier ubicación teniendo en cuenta una heterogeneidad espacial.

Dada una matriz $\mathbf{X}_{n\times m}$ que consta de m variables objetivas medidas en n sitios de observación, el GWPCA en la i-ésima ubicación con coordenadas u_i, v_i en el espacio geográfico, descompone la matriz de varianzas-covarianzas GW de **X** y utilizando enfoques de optimización como dejar un residuo (LOOR, por sus iniciales en inglés) que se encuentra descrito en Harris et al. (2011a). Para el análisis GWPCA, el primer vector propio l_i^1 para GWPCA en la ubicación i se calcula de modo que:

$$l_i^1 = argmax\, \boldsymbol{l}_i^T \Sigma i_i \boldsymbol{l}_{i,}, \quad sujeto\ a\ \|\boldsymbol{l}_i\| = 1 \tag{49}$$

donde:

$\boldsymbol{l}_i$ es la primera columna de longitud

m de la matriz de vectores propios de GW

$\boldsymbol{L}_i$ en la ubicación i.

El GWnnegPCA utiliza una restricción adicional:

$$sujeto\ a\ l_i \geq 0$$

de modo que todos los valores propios en cualquier ubicación no sean negativos.

Análisis de componentes principales ponderadas geográfica y temporalmente

En esta extensión, Han et al. (2022) incorporaron el efecto temporal al análisis GWPCA y desarrollaron un modelo extendido, el Análisis de Componentes Principales

Ponderados Geográfica y Temporalmente (GTWPCA, por sus iniciales eni nglès), para explorar simultáneamente la no estacionariedad espacial y temporal.

Para el modelo GTWPCA, se requiere una suposición de normalidad multivariante, y su escala operativa se localiza en un entorno espaciotemporal (Han et al., 2022). La matriz de varianzas-covarianzas ponderada geográfica y temporalmente es la siguiente:

$$\sum(u_i, v_i) = \boldsymbol{X}^T \boldsymbol{W}_{(u_i, v_i, t_i)} \boldsymbol{X} \tag{50}$$

donde:

$\boldsymbol{X}$ es la matriz de datos y

$\boldsymbol{W}_{(u_i, v_i, t_i)}$ es una matriz diagonal de pesos espaciotemporales.

Se supone que los datos observados están ubicados en un sistema tridimensional de coordenadas espacio-temporales con efectos de escala relativamente equilibrados en el tiempo y el espacio. La distancia espaciotemporal d_{ij}^{ST} entre el punto (u_i, v_i, t_i) y su punto circundante (u_j, v_j, t_j) se puede describir mediante la combinación lineal de la distancia espacial d_{ij}^{S} y la distancia temporal d_{ij}^{T} de la forma siguiente:

$$d_{ij}^{ST} = \alpha d_{ij}^{S} + \beta d_{ij}^{T} = \alpha \left[\left(\mu_i - \mu_j\right)^2 + \left(v_i - v_j\right)^2 \right] + \beta \left(t_i - t_j\right)^2 \tag{51}$$

Los parámetros α y β se utilizan para equilibrar los efectos de las distancias espaciales y temporales. La construcción de la matriz de ponderación espaciotemporal se realiza a través de la función Kernel bicuadrada:

$$\boldsymbol{W}_{ij}^{ST} = \begin{cases} \left[1 - \left(\frac{d_{ij}^{ST}}{r^{ST}}\right)^2\right]^2 & \text{si } d_{ij}^{ST} < r^{ST} \\ 0 & \text{dsi } d_{ij}^{ST} \geq r^{ST} \end{cases} \tag{52}$$

donde:

r^{ST} es el ancho de banda, es decir, es la distancia espaciotemporal,

d_{ij}^{ST} se refiere a la distancia entre las ubicaciones espaciotemporales desde la *i*-ésima hasta la *j*-ésima filas en la matriz de datos

$\boldsymbol{W}_{ij}^{ST}$ es el peso espaciotemporal en la ubicación *j* al calcular la componente principal local en la ubicación *i*.

Para encontrar las componentes principales locales en la ubicación (u_j, v_j, t_j) la descomposición de la matriz de varianzas-covarianzas proporciona la matriz de valores propios y vectores propios ponderados geográfica y temporalmente

$$\boldsymbol{L}_{(u_i,v_it_i)}\boldsymbol{V}_{(u_i,v_it_i)}\boldsymbol{L}^{T}_{(u_i,v_i,t_i)} = \boldsymbol{\Sigma}_{(u_i,v_i,t_i)} \tag{53}$$

La componente ponderada geográfica y temporalmente que puntúa en la ubicación espaciotemporal (u_j, v_j, t_j) se encuentra post-multiplicando los datos originales por matrices de vectores propios ponderados geográfica y temporalmente $\boldsymbol{L}_{(u_i,v_it_i)}$, como se muestra en:

$$\boldsymbol{T}_{(u_i,v_it_i)} = \boldsymbol{X}\boldsymbol{L}_{(u_i,v_it_i)} \tag{54}$$

2.3. ANÁLISIS DISCRIMINANTE PONDERADO GEOGRÁFICAMENTE

El Análisis Discriminante (DA, por sus iniciales en inglés) abarca una serie de métodos o modelos de clasificación supervisados que incluye el Análisis Discriminante Líneal (LDA, por sus iniciales en inglés), el Análisis Discriminante Cuadrático (QDA, por sus iniciales en inglés) y el Análisis Discriminante Kernel (KDA, por sus iniciales en inglés), entre otros. Se utilizan para estudiar las características típicas de objetos que pertenecen a dos o más grupos conocidos a priori y las nuevas observaciones se clasifican en uno de dichos grupos en función de sus características o de una regla discriminante. El modelo puede usarse en un contexto exploratorio para determinar la importancia relativa de las variables predictoras para construir a la clasificación o de manera inferencial para realiar la predicción (Manly & Navarro, 2016).

Brunsdon et al. (2007) proponen una extensión de LDA que permite la predicción y el análisis de datos espaciales categóricos donde la relación entre las variables predictoras y las categorías varía espacialmente y la denominaron Análisis Discriminante Ponderados Geográficamente (GWDA, por sus iniciales en inglés). Si la relación entre las variables predictoras y las categorías no es estacionaria, siempre que la distribución espacial de las categorías sea bastante uniforme, se esperaría que el GWDA modele con mayor precisión la relación entre las variables predictoras, proporcionando así una clasificación de mayor calidad que el LDA (Foley & Demšar , 2013).

2.3.1. Modelo del Análisis Discriminante Ponderado Geográficamente.

Brunsdon et al. (2007)**,** asume que la matriz de varianzas, de varianzas-covarianzas, el vector de medias y las probabilidades a priori no son fijas, sino que dependen de la ubicación espacial, es decir, las probabiidades usadas para derivar las reglas de decisión están condicionadas a la ubicación

Brunsdon et al. (2007) propuso el siguiente modelo:

$$f_p\left(\frac{\boldsymbol{x}}{\boldsymbol{u}}\right) = \frac{1}{(2\pi|\sum_j(\boldsymbol{u})|)^{\frac{q}{2}}} exp\left[-\frac{1}{2}\left(\boldsymbol{x}-\mu_j(\boldsymbol{u})\right)^T \sum_j^{-1}(\boldsymbol{u})(\boldsymbol{x}-\mu_j(\boldsymbol{u}))\right] \quad (55)$$

donde:

$\boldsymbol{\Sigma_j}$ la matriz de varianzas-covarianzas;

μ_j el promedio para la población j;

$\boldsymbol{u}$ la ubicación espacial, y

q es el número de variables predictoras en el vector de observaciones **x.**

2.3.2. Estimación del Modelo del Análisis Discriminante Ponderado Geográficamente.

Un enfoque para estimar $C_j(\mathbf{u})$ (colección de funciones asociadas con la ubicación) en cualquier punto **u** es a través de la estimación por máxima verosimilitud local, minimizando la siguiente función:

$$L\left(\frac{C_j(\boldsymbol{u})}{\boldsymbol{x}}\right) = \sum_{i=1}^{n} w_i\,(\boldsymbol{u})\left[log\left(\left|\sum_{\boldsymbol{j}}(\boldsymbol{u})\right|\right) - -\frac{1}{2}\left(\boldsymbol{x}-\mu_j\right)^T \sum_{\boldsymbol{j}}^{-\mathbf{1}}(\boldsymbol{x}-\mu_j)\right] \quad (56)$$

donde $w_i(\boldsymbol{u})$ es el peso aplicado a la observación i cuando se realiza la calibración en el punto u.

El enfoque propuesto conduce a los siguientes estimadores para Cj (u):

$$\hat{\mu}_j = \frac{\sum_{i=1}^{n} w_i(\boldsymbol{u})\boldsymbol{x}_i}{\sum_{i=1}^{n} w_i(\boldsymbol{u})} \quad (57)$$

y

$$\sum_j(\boldsymbol{u}) = \frac{\sum_{i=1}^{n} w_i(\boldsymbol{x}_i - \mu_j(\boldsymbol{u}))(\boldsymbol{x}_i - \mu_j(\boldsymbol{u}))^T}{\sum_{i=1}^{n} w_i(\boldsymbol{u})} \quad (58)$$

donde p_j (**u**), la probabilidad a priori de que una observación provenga de la población j en la localización (**u**)

La probabilidad se obtiene de manera sencillla si los datos de entrenamiento en la vecindad de **u** son una muestra aleatoria de la población completa obtenida al fusionar cada una de las poblaciones asociadas con los valores de la ***y*** individuales.

En este caso, los p_j se estiman como sigue (Brunsdon et al., 2007):

$$p_j = \frac{n_j}{n_1 + n_2 + \ldots + n_m} \quad (59)$$

Si además se asume que Σ es igual para cada población, como en el LDA, se obtiene una estimación de Σ de la siguiente forma (Brunsdon et al., 2007):

$$\sum = \frac{n_1 \sum_1 + n_2 \sum_2 + \ldots + n_m \Sigma_m}{n_1 + n_2 + \ldots + n_m} \quad (60)$$

Estos mismos autores suponen de esta forma que las proporciones de diferentes poblaciones en la muestra reflejan las proporciones globales. Sin embargo, cuando ese no sea el caso, entonces suponen que cada p_j es igual a 1/m. Además, si las proporciones en la población fusionada asociadas con cada uno de los valores *y* pueden variar geográficamente en su composición, deben usarse las estimaciones de probabilidad local de p_j como sigue:

$$p_j(\boldsymbol{u}) = \frac{\sum_{x_i \in j} w_i(\boldsymbol{u})}{\sum_j \sum_{x_i \in j} w_i(\boldsymbol{u})} \quad (61)$$

Un tema clave en el enfoque anterior es la decisión sobre qué partes del modelo deben variar geográficamente. Brunsdon et al. (2007) permitieron que los tres componentes,

la media, la matriz de varianzas-covarianzas y las probabilidades a priori varien geográficamente. Sin embargo, manifiestan que hay multiples formas en que se podría especificar el GWDA, pues es probable que las diferentes opciones de combinación de componentes con valores que varien local o globalmente sean mejores en diferentes situaciones. Por tanto, se debería proponer alguna metodología para seleccionar el mejor enfoque para un conjunto de datos determinado.

Al igual que en la GWR, la elección de ancho de banda y de si se utiliza un enfoque fijo o adaptativo para la ponderación geográfica requiere una metodología de selección, pues se puede esperar que el grado de variabilidad espacial en las funciones discriminantes varíe geográficamente. Brunsdon et al. (2007) sugieren una función de ponderación de Kernel bicuadrada para asignar pesos geográficos, $\mathbf{w}_i$, a los objetos que rodean al objeto x que se está clasificando. En dicho trabajo, propusieron la validación cruzada, que implica eliminar cada observación de los datos de entrenamiento, aplicar el GWDA al resto del conjunto de datos y luego asignar la observación eliminada a una población utilizando la regla discriminante que se ha derivado y anotando la proporción de asignaciones correctas. De esta manera se obtiene una puntuación (valoración) de desempeño para un modelo dado que trabaja con un ancho de banda determinado. Esta puntuación obtenida mediante la validación cruzada se puede usar para identificar la "mejor" combinación de modelo y ancho de banda para un conjunto de datos determinado.

Otro enfoque sería el de seleccionar aleatoriamente una gran proporción de los datos de entrenamiento (por ejemplo, 90%), calibrar el GWDA con éstos y luego aplicar el procedimiento de validación cruzada al 10% restante. Sin embargo, un enfoque más sofisticado para la puntuación de validación cruzada puede ser, usar la probabilidad de validación cruzada, en lugar de la proporción de asignaciones correctas, teniendo en cuenta que el DA asigna observaciones a las poblaciones sobre la base de probabilidades posteriores. Aplicando el Teorema de Bayes, la probabilidad de que la observación x esté en la población j viene dada por:

$$Pr(j|\boldsymbol{x},\boldsymbol{u}) = \frac{\Pr(\boldsymbol{x}|j,\boldsymbol{u})}{\Pr(\boldsymbol{x}|1,\boldsymbol{u})+\cdots+\Pr(\boldsymbol{x}|m,\boldsymbol{u})} \tag{62}$$

Para cada observación en el conjunto de datos de validación cruzada retenida, conocemos el valor real de j, y al usar un modelo y un ancho de banda determinados podemos estimar la probabilidad posterior de que la observación se encuentre en su población real (Brunsdon et al., 2007). Al tomar registros y sumar estas cantidades, se obtiene la probabilidad del conjunto de datos de validación para una elección dada de ancho de banda y modelo:

$$Puntuación = \sum_{\substack{Conjunto \\ de\ validación}} \log\left(Pr(j|\boldsymbol{x}, \boldsymbol{u})\right) \tag{63}$$

Esta puntuación se puede utilizar para seleccionar la mejor combinación de modelo y ancho de banda, aunque es computacionalmente más costosa, tiene la ventaja de que al ser una función continua con respecto al ancho de banda es ventajoso en los algoritmos de optimización, como el descenso de gradiente o el método simplex (Brunsdon et al., 2007).

2.4. ANÁLISIS DE CLUSTER PONDERADO GEOGRÁFICAMENTE

Mason & Jacobson (2007), basándose en los principios de la interacción espacial geográfica (Birkin & Clarke, 1991), incorporan un modelo básico de interacción espacial en la ponderación de pertenencia a un clúster proponiendo el Algoritmo de Agrupamiento Ponderado Geográficamente Difuso (FGWC, por sus iniciales en inglés), el cúal, según los autores, ofrece una alternativa al agrupamiento regular que proporciona la capacidad de aplicar los efectos geográficos de población y de distancia en un análisis de agrupamiento geodemográfico (GDA, por sus iniciales en inglés).

La autocorrelación espacial de los valores de pertenencia al clúster aumenta a medida que aumenta la ponderación de la interacción espacial, y los valores de pertenencia indican una homogeneización creciente de los agrupamientos, que viene medida por la desviación estándar de los valores de pertenencia (Mason & Jacobson, 2007).

La pertenencia ajustada para el algoritmo de agrupamiento difuso ponderado geográficamente, que se calcula en cada iteración del algoritmo de agrupamiento difuso, se muestra en la Ecuación 64:

$$\mu_i^* = \alpha\mu_i + \beta\frac{1}{A}\sum_j^n w_{ij}\,\mu_j \tag{64}$$

donde:

μ_i^* la nueva pertenencia a cluster;

μ_i la antigua pertenencia al cluster del área i;

α y β ($\alpha + \beta$=1) son variables de escala que afectan la proporción de la membresía original versus la membresía ponderada (calculada)

$$\boldsymbol{w}_{ij} = \frac{(m_i m_j)^b}{d_{ij}^a} \tag{65}$$

donde:

m_i, es la población de las áreas i;

m_j es la población de las áreas j;

d_{ij} es la distancia entre i y j, y

A es un factor de escala para el sumatorio y se calcula en todos los grupos, asegurando que la suma de las pertenencias para un área determinada para todos los grupos es igual a uno.

Hasta la fecha sólo se han publicado en revistas revisadas por pares, cuatro artículos con la temática FGWC, todos ellos hacen referencia a mejoras del algoritmo. A continuación se hará una breve descripción de cada uno de ellos.

Son et al. (2012) proponen un algoritmo de agrupación para aplicar un GDA, que se extiende a FGWC y lo denominan Agrupación Ponderada Geográficamente Difusa Posibilista Intuicionista (IPFGWC, por sus iniciales en inglés). Prueban dicho algoritmo con un conjunto de datos de variables socioeconómicas y demográficas de la Organización de las Naciones Unidas (ONU) que contienen estadísticas anuales sobre el tamaño y composición de la población, nacimientos, defunciones, matrimonio y divorcio, actividad económica, nivel educativo, hogar, características, vivienda, etnia y lengua, etc. Según los autores, los resultados del algoritmo IPFGWC fueron más estables para diferentes valores de α. En el caso de un α<0,5, encontaron que el número de agrupaciones del algoritmo IPFGWC es más óptimo que el de FGWC.

Wijayanto & Purwarianti (2016) propusieron la integración entre el algoritmo de colonia de abeja artificial (ABC, por sus iniciales en inglés) y el algoritmo FGWC, cuyo objetivo era alcanzar una mejor precisión de agrupamiento geodemográfico. El nuevo algoritmo utiliza el algoritmo ABC para seleccionar, de forma automática, los centros de los

clústers (centroides) o la matriz de pertenencia en la fase de inicialización del procedimiento. Los autores, utilizando simulación en varios conjuntos de datos geodemográficos y mostraron que la calidad de agrupamiento del algoritmo propuesto, denominado FGWC-ABC, es mejor que la del FCM, Efectos de vecindario (NE, por sus iniciales en inglés) y FGWC.

Son (2015) introdujo un nuevo algoritmo de agrupamiento difuso intuitivo basado en funciones de núcleo o Kernel llamado Algoritmo de Agrupamiento Difuso del Núcleo para el Análisis Geodemográfico (KFGC, por sus iniciales en inglés) con el fin de abordar problemas de análisis geodemográfico Una de las características del algoritmo empleado es que la función objetivo del KFGC emplea una función Kernel gaussiana en lugar de la euclídea tradicional.

Nurmala & Purwarianti (2017) proponen una mejora del algoritmo FGWC mediante la integración de un algoritmo metaheurístico, denominado Optimización de Colonias de Hormigas (ACO, por sus iniciales en inglés), como una herramienta de optimización global para aumentar la precisión de agrupación en la etapa inicial del algoritmo FGWC.

CAPÍTULO III.

3. REVISIÓN DE APLICACIONES DE LOS MÉTODOS PONDERADOS GEOGRÁFICAMENTE

Aunque para la revisión de las aplicaciones del modelo de GWR realizamos un estudio preliminar hasta el año 2019, que se encuentra publicado en la revista *Area* (Journal Citation Report™: factor de impacto 2,280; cuartil Q1) que lleva por título "Trends and topics in geographically weighted regression research from 1996 to 2019" (https://doi.org/10.1111/area.12757) (Anexo 1). Cabe resaltar que en ese artículo se utilizaron paquetes de R de otros autores, tales como topicmodels (Grun, B & Hornik, 2011) textmineR (Jones, 2019) y ldatunning (Nikita, 2019). Sin embargo, en el presente documento de tesis, en el que aparece una revisión actualizada de GWR a diciembre de 2021 se utilizó el paquete LDAShiny, creado en R "ad hoc" precisamente para realizar esta revisión actualizada, y que por tanto, es producto de la presente investigación y será explicado en el apartado 3.1.2 de la memoria. Aunque el paquete LDAShiny se encuentra alojado en el Comprehensive Archive Network (CRAN) de R, la presentación de este paquete a la comunidad científica se ha realizado mediante una publicación en la revista *Mathematics* (Journal Citation Report™: factor de impacto 2,258, cuartil Q1) que lleva como título "LDAShiny: An R Package for Exploratory Review of Scientific Literature Based on a Bayesian Probabilistic Model and Machine Learning Tools" (https://doi.org/10.3390/math9141671) (Anexo 2).

Aunque el LDAShiny se creó para resolver el problema de revisión bibliográfica de las aplicaciones de los métodos de GWR en el transcurso de este trabajo de investigación, como hemos apuntado, la publicación de *Mathematics* pone en evidencia su facilidad de uso en otros contextos como puede ser la Medicina y análisis textual en redes sociales. Hay que resaltar también que el paquete LDAShiny ya ha sido utilizado con éxito en otro trabajo, "Frequency of Neuroendocrine Tumor Studies: Using Latent Dirichlet Allocation and HJ-Biplot

Statistical Methods" (Escobar et al., 2021) y en el trabajo titulado como "Lending ears to unheard voices: An empirical analysis of user-generated content on social media" (Gour et al., 2021). Además, el paquete también se utilizó en el trabajo "Capturing the Complexity of Covid-19 Research: Trend Analy-sis in the First Two Years of the Pandemic Using a Bayesian Probabilistic Model and Machine Learning Tools" sometido actualmente también a la revista *Mathematics*.

3.1. PROPUESTA DE UNA APLICACIÓN EN R PARA LA REVISIÓN SISTEMÁTICA DE LITERATURA

Como se mencionó anteriormente, al comenzar con la revisión de artículos sobre aplicaciones o usos de los métodos de análisis de datos ponderados geográficamente, el número de artículos científicos encontrados en diferentes bases de datos fue tan alta que se tornó inviable hacer una clasificación sencilla de todos ellos.

Para llevar a cabo una revisión de las aplicaciones de los métodos clásicos ponderados geográficamente, en esta tesis se presenta una nueva propuesta metodológica en la que se utiliza el modelado de tópicos haciendo uso del aprendizaje automático para explorar artículos que permitan identificar en qué direcciones se han movido las investigaciones sobre estos temas en los últimos años.

En la actualidad existe una cantidad cada vez mayor de literatura científica publicada en forma digital en bases de datos como Scopus o Web of Science, por mencionar dos de las más utilizadas por los investigadores (Harzing & Alakangas, 2016). Por lo tanto, se puede inferir que existe una brecha entre la disponibilidad y el uso de la información. Una revisión bibliográfica de forma convencional es restringida, tiene un alto costo en términos de tiempo y presenta un poder de procesamiento limitado, lo que lleva a los investigadores a restringir la cantidad de documentos a revisar.

Asmussen & Møller (2019) mencionan que la revisión bibliogáfica de literatura, de forma convencional, pronto quedará obsoleta, ya que como hemos apuntado se trata de un procedimiento en el cual el investigador debe emplear una gran cantidad de tiempo. De hecho, esto es un problema en la fase exploratoria inicial de una investigación ya que lo que se necesita es una visión general del estado del arte de la investigación. La gran cantidad de

información disponible hace que buscar, recuperar y resumir la información sea engorroso y desafiante, por lo que se demanda el uso de herramientas capaces de buscar, organizar y resumir una gran colección de documentos de texto en el campo científico.

La solución puede encontrarse utilizando métodos de aprendizaje automático (machine learning) que hacen posible procesar grandes cantidades de datos, y hacen que los investigadores dediquen menos tiempo a examinar sus hallazgos. Cuando el procesamiento de información asistido por humanos, como el cifrado, se reemplaza por el procesamiento asistido por computador, mejora la fiabilidad y los costos disminuyen (DiMaggio et al., 2013).

3.1.1. Modelado de tópicos

El modelado de tópicos es un problema clásico en procesamiento del lenguaje natural. Se trata de una estrategia de aprendizaje automático no supervisado, que gracias a un conjunto de algoritmos y métodos estadísticos, tiene la capacidad de verificar y analizar la estructura oculta de una colección de documentos, que permite identificar palabras y descubrir los patrones dentro de ellos y en consecuencia, encontrar los grupos de palabras y expresiones comparativas que mejor caracterizan un conjunto de corpus. En resumen, los algoritmos de modelado de temas generan colecciones de frases y palabras que creen que están relacionadas, lo que permite al lector analizar y comprender qué significan esas relaciones, mientras clasifica los temas. Tiene la ventaja de no requerir anotaciones previas ni rotulación del documento ya que los temas surgen a partir del análisis de los textos originales (Blei, 2012).

El modelado de tópicos ha demostrado ser una herramienta muy útil para el análisis exploratorio de un gran número de artículos. Sin embargo, rara vez se ha aplicado en el contexto de una revisión bibliográfica exploratoria (Asmussen & Møller, 2019).

Su uso abarca prácticamente todos los aspectos de la extracción de textos y el procesamiento de información, incluido el resumen del texto, la recuperación de la información y la clasificación de texto. El modelado de temas nos permite organizar y resumir archivos electrónicos en varios formatos (páginas web, artículos científicos, libros, imágenes, sonido, videos y redes sociales) a una escala que sería imposible para la anotación humana (Asmussen & Møller, 2019).

En la actualidad, dentro los algoritmos utilizados para el modelado de tópicos se encuentran el Análisis Semántico Latente (LSA, por sus iniciales en inglés) (Deerwester et al., 1990), el Análisis Semántico Latente Probabilístico (PLSA, por sus iniciales en inglés) (Hofmann, 1999) y la Asignación Latente de Dirichlet (LDA, por sus iniciales en inglés) propuesto por Blei et al. (2003).

Para la nueva propuesta metodológica, en este trabajo nos decidimos por este último algoritmo, dado que es considerado como uno de los más populares (Jacobi et al., 2016; Blei, 2012; DiMaggio et al., 2013; Grimmer, 2010).

El LSA representa el texto como una matriz de términos de documentos y utiliza la descomposición en valores singulares (SVD, por sus iniciales en inglés) para reducir su dimensionalidad y codificarlas usando características latentes (es decir, temas) (Figura 1), mientras que el PLSA expande el LSA con una base probabilística.

En lugar de realizar una SVD, el PLSA se basa en el principio de verosimilitud y define un modelo generativo adecuado de los datos (Hofmann, 1999) (Figura 2).

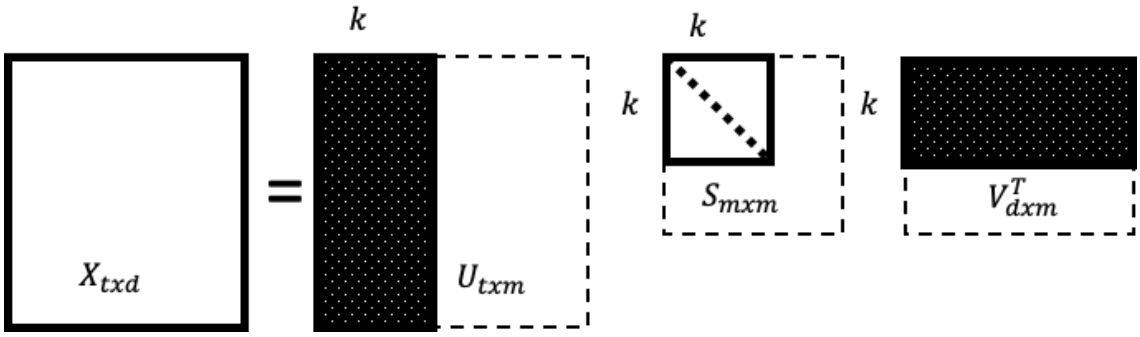


Figura 1. Notación, descomposición y truncamiento de la matriz términos-documento (adoptado de Berry et al., 1995). La SVD descompone la matriz términos x documentos $\boldsymbol{X}_{txd}$, en el producto de otras tres matrices: una matriz $\boldsymbol{U}_{txm}$ columna-ortogonal donde m representa la dimensionalidad, una matriz diagonal $\boldsymbol{S}_{mxm}$ con valores singulares dispuestos en orden decreciente y una matriz transpuesta $\boldsymbol{V}_{dxm}$ columna-ortogonal

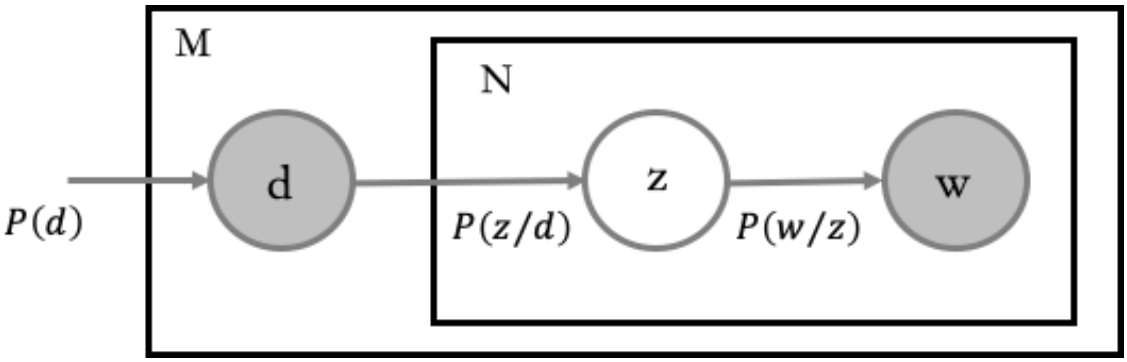


Figura 2. "Notación de placa" que representa el modelo asimétrico del Análisis Semántico Latente Probabilístico (adoptado Blei et al., 2003). Dado un documento d, el tema z está presente en ese documento con probabilidad $P(z/d)$; dado un tema z, la palabra w se extrae de z con probabilidad $P(w/z)$

Como un paso más, el algoritmo LDA desarrolla un modelo probabilístico generativo de un corpus mediante la representación de documentos como mezclas aleatorias sobre temas latentes y temas como distribuciones probabilísticas sobre palabras (Blei et al., 2003). El objetivo del LDA es inferir o estimar las variables latentes, es decir, calcular su distribución condicionada a los documentos. Una representación gráfica en "notación de placas" del modelo LDA se encuentra en la Figura 3, en la que se puede deducir la siguiente distribución conjunta:

$$P(w, z, \theta, \phi/\alpha, \beta) = P(\theta/\alpha)\, P(z/\theta)\, P(\phi/\beta)\, P(w/z, \phi) \tag{66}$$

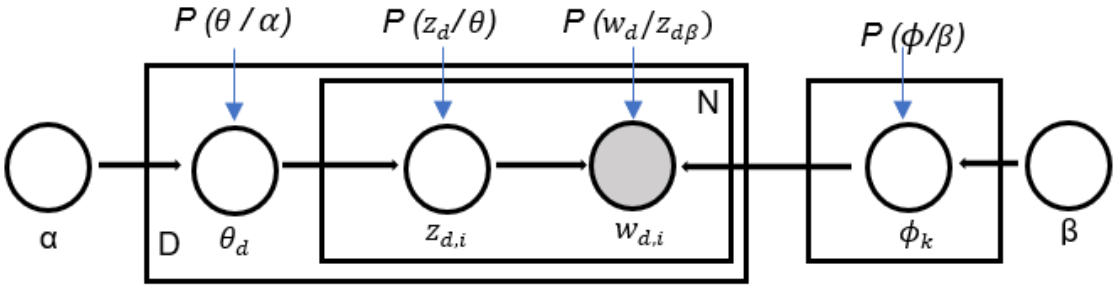


Figura 3. Notación de placa del modelo de Asignación Latente de Dirichlet (adaptado Blei et al., 2003). Los nodos sin sombrear representan las variables aleatorias ocultas, los nodos sombreados las variables aleatorias observadas y los bordes las dependencias condicionales entre ellas. Los rectángulos se llaman placas que representan la replicación

A la derecha de la Ecuación 63 se encuentra:

- $P\left(\frac{\theta}{\alpha}\right)$: representa la distribución de tópicos por documento, teniendo en cuenta el parámetro Dirichlet, el cual es un vector K–dimencional con componentes $\alpha_k > 0$

(el operador punto $(\alpha .)$ que se introduce en la Ecuación, es la abreviatura en el índice de las variables para la suma de todos los valores).

$$P\left(\frac{\theta}{\alpha}\right) = \frac{\Gamma(\alpha .)}{\prod_{k=1}^{K}\Gamma(\alpha_k)}\prod_{k=1}^{K}\theta_k^{\alpha k^{-1}} \quad (67)$$

- $P\left(\frac{z}{\theta}\right)$ corresponde a la distribución del tópico z en el corpus, este depende de la distribución anteriormente mencionada. Consecuentemente, a cada palabra w_i en el documento de N palabras, se le da un valor de 1, …, K.

  Esta distribución $P\left(\frac{z}{\theta}\right)$ expresa la probabilidad del tópico z para todos los documentos y también los tópicos en términos de número de palabras $n_{d,k}$. Esta es la cantidad de veces que se ha asignado un tema k a cualquier palabra que se encuentra en el documento *d*.

$$P\left(\frac{z}{\theta}\right) = \prod_{d=1}^{D}\prod_{k=1}^{K}\theta_{d,k}^{n_{d,k}} \quad (68)$$

- $P(\phi/\beta)$: representa las distribuciones de los tópicos según el término, teniendo en cuenta todo el corpus ϕ_k. Estas se obtienen de una distribución de Dirichlet con parámetro β. $\phi_{k,v}$ proporciona la probabilidad en donde el término *v* se adquiere de acuerdo con el tópico elegido, este se enuncia para todos los tópicos y cada una de las palabras del vocabulario así:

$$P(\phi/\beta) = \prod_{K=1}^{K}\frac{\Gamma(\beta_{k,\cdot})}{\prod_{v=1}^{V}\Gamma(\beta_{k,v})}\prod_{v=1}^{V}\phi_{k,v}^{\beta_{k,v}-1} \quad (69)$$

Finalmente, la probabilidad de un corpus *w* teniendo en cuenta z y ϕ en el modelo gráfico es:

$$P\left(\frac{w}{z},\phi\right) = \coprod_{k=1}^{K}\prod_{v=1}^{V}\phi_{k,v}^{n_{\cdot,\cdot,k,v}} \quad (70)$$

La Ecuación 66 se podría reescribir eliminando las variables latentes, con el fin de considerar un modelo de probabilidad dado, un corpus *w* y los hiperparámetros (α, β). Para poder realizar una "estimación máximo-verosímil" de los parámetros del modelo y para inferir la distribución de las variables latentes (Blei et al., 2003) es necesaria la probabilidad anteriormente presentada.

$$P\left(\frac{w}{z}, \phi\right) =$$

$$\int_{\phi}\int_{\theta}\sum_{z}\left(\prod_{d=1}^{D}\frac{\Gamma(\alpha\cdot)}{\prod_{k=1}^{K}\Gamma(\alpha_k)}\prod_{k=1}^{K}\theta_{d,k}^{\alpha_k+n_{\cdot,k,v}-1}\right)\left(\prod_{k=1}^{K}\frac{\Gamma(\beta_{k,\cdot})}{\prod_{v=1}^{V}\Gamma(\beta_{k,v})}\prod_{v=1}^{V}\phi_{k,v}^{\beta_{k,v}+n_{\cdot,k,v}-1}\right)d\theta d\phi \quad (71)$$

La adición sobre todas las composiciones posibles de asignaciones de tópicos hace que esta probabilidad sea computacionalmente intratable (Blei et al., 2003) y es por ello que se debe hacer uso de algoritmos de aprendizaje automático (Machine Learning) para hallar aproximaciones de la probabilidad marginal. Sin embargo, esta puede hallarse mediante una aproximación lo suficientemente cerca al verdadero valor teniendo en cuenta inferencia estadística, que puede tratarse aplicando el algoritmo de maximización de la expectativa variacional (Blei et al., 2003) o mediante el muestreo de Gibbs (Griffiths & Steyvers, 2004).

Estas técnicas pueden proporcionar una aproximación cercana al verdadero valor de probabilidad posterior. Con este fin, los tópicos identificados, y cómo se distribuyen en cada documento, se combinan con la información del corpus evaluado (por ejemplo, nombre de la revista, año de publicación y país de afiliación del autor de correspondencia) para descubrir las tendencias temporales y regionales en la investigación.

3.1.2. Paquete LDAShiny

Justificación del Paquete

Ni la teoría del modelado de tópicos ni cualquier otra tendría trascendencia sin la existencia de una herramienta que facilite su uso. En la actualidad, en un mundo cada vez más globalizado, los investigadores tienen prisa y cuentan con un número considerable de opciones a su disposición. Un usuario interesado en el uso de cualquier aplicación dejará de utilizarla si al intertar obtener resultados no lo consigue en poco tiempo.

En el entorno de código abierto R (R Core Team, 2021), en el Comprehensive Archive Network (CRAN), podemos encontrar una lista de 59 paquetes relacionados con el procesamiento del lenguaje natural (NLP, por sus iniciales en inglés), ocho de los cuales implementan el modelado de temas a través de la LDA:

- **lda**: Collapsed Gibbs Sampling Methods for Topic Models (Chang, 2015);

- **lda.svi**: Fit Latent Dirichlet Allocation Models using Stochastic Variational Inference (Erskine, 2015);
- **ldaPrototype**: Prototype of Multiple Latent Dirichlet Allocation Runs (Rieger, 2015);
- **ldatuning**: Tuning of the Latent Dirichlet Allocation Models Parameters (Nikita, 2019);
- **LDAvis**: Interactive Visualization of Topic Models (Sievert & Shirley, 2016);
- **Topicdoc**: Topic-Specific Diagnostics for LDA and CTM Topic Models (Friedman, 2019);
- **topicmodels**: Topic Models (Grun, B & Hornik, 2011);
- **textmineR** (Jones, 2019).

No obstante, la utilización de los paquetes mencionados requieren algunas habilidades estadísticas y de aprendizaje automático que no todos los investigadores poseen.

Con lo anterior en mente el objetivo principal que nos ha llevado a la creación del paquete LDAShiny, es conseguir que el flujo de trabajo típico de la LDA sea más fácil de usar, especialmente para aquéllos que no están familiarizados con el lenguaje R. Mediante la utilización del LDAShiny el análisis se puede realizar de forma interactiva en un navegador web, lo que facilita que muchos más investigadores puedan aplicar la metodología y la técnica para revisar la literatura científica de su área de estudio.

Con LDAShiny se logra reducir o reemplazar el tiempo frente a la computadora al generar automáticamente temas de revisión en función de las cualidades estadísticas de los documentos utilizados, sin necesidad de clasificación, categorización o etiquetado previos. Por lo tanto, el posible sesgo debido a las elecciones subjetivas de los investigadores podría evitarse o minimizarse. Además, la investigación y las tendencias históricas y actuales en el campo de estudio se podrían sintetizar más fácilmente.

Especificación de requisitos

Se propuso que la aplicación o paquete propuesto cumpliera con los siguientes requisitos:

- Accesible a la mayor cantidad posible de usuarios;
- De fácil distribución e instalación;

- Entorno gráfico y funcional que fuese amigable con el usuario;
- Con soporte en idioma inglés y castellano.

El entorno del programa R cumple con algunos de los requisitos planteados, pues a través de la infraestructura CRAN puede ser distribuido y ser accesible a una gran cantidad de usuarios. Dado que el software está disponible gratuitamente desde el repositorio CRAN, es posible monitorear las descargas. LDAShiny y a la fecha de cierre de este documneto ha registrado 14 mil descargas (https://cranlogs.r-pkg.org/badges/grand-total/LDAShiny) desde su publicación oficial en marzo de 2021.

Como valor añadido, R es un entorno multiplataforma, es decir, el programa construido puede ser ejecutado en los sistemas operativos utilizados por la mayoría de los computadores personales como son Windows, Mac o Linux.

Otro de los requisitos es que el paquete tiene que ser fácil de instalar, dado que si utilizamos elementos externos al software R que requieran ser instalados previamente o compilados para el funcionamiento del módulo, es probable que el usuario desista de utilizarlo. En nuestro caso la instalación no requiere de ningún elemento externo.

En cuanto al entorno gráfico, el software R no contiene comandos que permitan generar una interfaz gráfica para el usuario, es decir, no presenta un conjunto de ventanas, marcos, botones, listas y demás elementos gráficos para generar un entorno amigable y facilitar el uso. Para generar el entorno gráfico se recurrió al paquete Shiny de R que facilita la creación de aplicaciones web interactivas (apps) directamente desde R. El enlace "reactivo" automático entre entradas y salidas y los amplios widgets preconstruidos hacen posible crear aplicaciones atractivas, potentes y con gran capacidad de respuesta.

Con respecto al soporte, la aplicación desarrollada presenta tutoriales en inglés y castellano con videos explicativos del uso de la aplicación. Estos tutoriales estan alojados en la CRAN de R y pueden ser consultados desde los siguientes enlaces:

- https://cran.rproject.org/web/packages/LDAShiny/vignettes/A_brief_introduction_to_LDAShiny.html
- https://cran.r-project.org/web/packages/LDAShiny/vignettes/Una_breve_introducci-n_a_LDAShiny.html

Descripción de la aplicación

Dado que el programa LDAShiny se basa en enfoques de modelado de tópicos, la principal contribución de este parte del trabajo no es introducir nuevas formas de procesar datos, sino proponer cómo se combinan los métodos y cómo pueden ser utilizados fácilmente por los investigadores mediante el uso de esta aplicación. Se consideró que el proceso de revisión de la literatura constaba de cuatro pasos: (i) preprocesamiento, (ii) inferencia, (iii) modelado de tópicos y (iv) postprocesamiento (Figura 4).

(i) *Preprocesamiento*

El preprocesamiento consiste en cargar y preparar los documentos para procesos posteriores. Los datos no estructurados contienen generalmente mucha información irrelevante que debe eliminarse antes de pasar los datos a la fase de entrenamiento y de análisis. Esta etapa implica muchos pasos, se siguen técnicas de tokenización (reconocimiento de unidades sintácticas), eliminación de palabras vacías y lematización para el preprocesamiento de los resúmenes de los artículos.

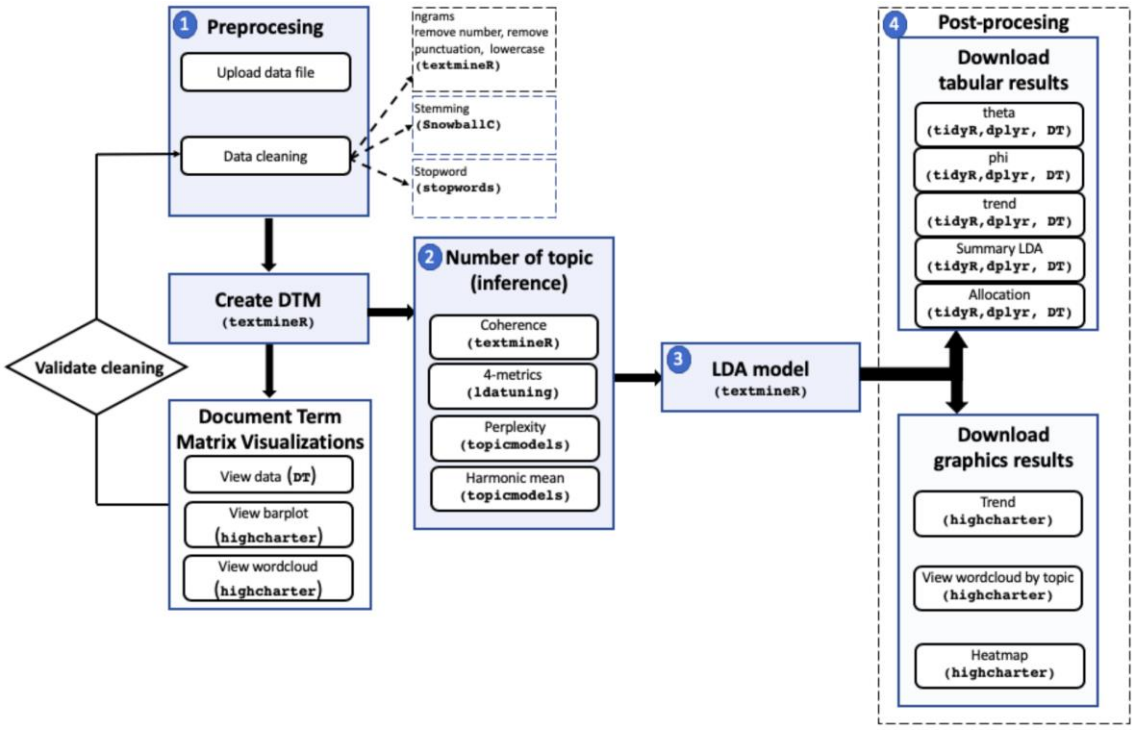


Figura 4. Esquema del paquete LDAShiny. Entre paréntesis se encuentran los principales paquetes utilizados

Esta fase juega un papel muy importante en todo el procedimiento, siendo generalmente el primer paso en las técnicas y aplicaciones de minería de textos (Vijayarani et al., 2015). El preprocesamiento busca normalizar o convertir el conjunto de un texto a una forma estándar más conveniente que permita reducir la dimensionalidad de la matriz de datos al eliminar el ruido o los términos sin sentido. Dentro del preprocesamiento tenemos la "limpieza" en la que se realizan las siguientes tareas:

- Eliminar espacios en blanco y los signos de puntuación, así como poner en minúsculas todo el texto;
- Tokenización, que es el procedimiento de separación de morfemas (palabras). Según Jurafsky & Martin (2008) es beneficioso tanto en lingüística como en informática;
- Inclusión de n-gramas: un n-grama es una secuencia contigua de n palabras (Manning & Schutze, 1999). Aunque es más habitual analizar palabras individuales, en algunos casos, como en las ciencias de la vida, sería ventajoso incorporar bigramas porque los nombres científicos de las especies se componen de dos palabras. En LDAShiny podemos trabajar con unigramas, bigramas o trigramas (tres palabras consecutivas de ocurrencia frecuente);
- Eliminar números: a pesar de que con frecuencia se piensa que los números no son informativos, hay algunas áreas de conocimiento donde los números pueden proporcionar información valiosa, por ejemplo, en materia legislativa, los proyectos de ley o decretos pueden ser significativos con respecto al contenido de la legislación. Es por ello que en la aplicación desarrollada el investigador puede decidir si elimina o no los números;
- Eliminar StopWord (un término acuñado por Luhn en 1958). El procedimiento consiste en descartar palabras que no tienen significado léxico y que aparecen en los textos con mucha frecuencia (como artículos y pronombres). Hay muchas listas potenciales de StopWord, sin embargo, nos limitamos a una lista precompilada de palabras proporcionada por R StopWord (Porter, 1980);
- LDAShiny permite realizar este procedimiento en 14 idiomas: danés, holandés, inglés, finlandés, francés, alemán, húngaro, italiano, noruego, portugués, rumano, ruso, español y sueco;
- Stemming, que es la versión más simple de lematización. Consiste en reducir las palabras a formas básicas (Porter, 1980). Aunque a menudo se usa como una técnica de reducción, debe usarse con cuidado, ya que podría combinar palabras con

diferentes significados, por ejemplo en las frases "college students partying", and "political parties", la derivación reduciría *partying* y los *parties* a la misma forma básica;

- Eliminar los términos que se usan con poca frecuencia (sparcity). Este procedimiento es muy útil porque permite eliminar los términos que aparecen en muy pocos documentos antes de continuar con las sucesivas fases. Entre las razones de este paso está la factibilidad computacional, ya que este proceso reduce drásticamente el tamaño de la matriz sin perder información significativa y también puede eliminar errores en los datos, como palabras mal escritas. Esto solo se aplica a los términos que cumplen con la condición:

$$df(t) > N(1 - sparce) \tag{72}$$

donde:

df(t) es la frecuencia de documentos del término t y

N es el número de términos .

Por ejemplo, si el valor sparce es 0,99, se toman los términos que aparecen en más del 1% de los documentos. Como regla general, los términos que aparecen en menos del 0,5 % -1 % de los artículos deben descartarse (Grimmer, 2010; Yano, et al., 2012; Grimmer & Stewart, 2013). Sin embargo, no existe en la literatura un examen sistemático de las implicaciones que esta decisión de procesamiento previo tiene en la fase final de los análisis.

El preprocesamiento debe ser validado. Sin embargo, hasta la fecha no ha existido una forma científica de establecer cuándo finaliza, por lo que debe ser iterativo, ya que no es posible garantizar un procedimiento de "limpieza" idéntico al realizar una revisión exploratoria (Asmussen & Møller, 2019). Una vez completada la fase de preprocesamiento, se obtiene la matriz documento-término (DTM, por sus iniciales en inglés) como datos de entrada para los procesos posteriores.

(ii) *Inferencia (Número de tópicos)*

LDA es un modelo de variables latentes mediante correlaciones entre palabras y temas semánticos latentes en una colección de documentos (Blei & Lafferty, 2007). Esto implica

que el parámetro k del algoritmo (número de temas) es crucial y debe establecerse a priori, ya que la validez de los resultados obtenidos depende en gran medida del proceso de inferencia del modelo. En términos teóricos, un número muy grande de temas producirá temas demasiado específicos, mientras que, por el contrario, un número muy pequeño tratará temas amplios y heterogéneos (Sbalchiero & Eder, 2020). Existen una variedad de métricas que se pueden utilizar para determinar el número óptimo de temas. En nuestro paquete implementamos las siguientes:

- **Perplexity**: definida por Blei et al. (2003) para un conjunto de textos de documentos M como:

$$perplexity(D_{test}) = exp\left\{-\frac{\sum_{d=1}^{M} \log P(w_d)}{\sum_{d=1}^{M} N_d}\right\} \quad (73)$$

  donde:

  N_d es el número de palabras en el d-ésimo documento del corpus de texto D_{text} , y w_d , es el d-esímo documento en el corpus.

  Es monótonamente decreciente. Al comparar varios modelos, el que tiene el menor valor de perplejidad se considera el mejor (Blei et al., 2003);

- **Verosimilitud marginal**, que se puede aproximar por la media armónica. Este método fue aplicado por primera vez por Griffiths & Steyvers (2004) en su enfoque bayesiano, con el fin de encontrar el número óptimo de temas;
- **Coherencia** (Röder et al., 2015). Se basa en la hipótesis de distribución conjunta (Cao et al., 2009) que establece que las palabras con significados similares tienden a coexistir en contextos similares. El procedimiento se encuentra implementado en el paquete TextmineR (Jones, 2019) y consiste en ajustar varios modelos y calcular la coherencia de cada uno de ellos. El mejor modelo será el que ofrezca la mayor medida de coherencia;
- **Otras métricas:** están implementadas en el paquete ldatuning y son: las métricas Arun 2010 (Arun et al., 2010), CaoJuan 2009 (Cao et al., 2009), Deveaud 2014 (Deveaud et al., 2014), Griffiths 2004 (Griffiths & Steyvers, 2004). El planteamiento de estas métricas es sencillo ya que se basan en la búsqueda de valores extremos (minimización: Arun 2010 y CaoJuan 2009; maximización: Deveaud 2014 y Griffiths 2004).

*(*iii) *Modelado de tópicos*

Una vez determinado el número de tópicos, se procede a ejecutar el modelo LDA. Algunos parámetros como el número de iteraciones pueden ser modificados por un número de iteraciones mayor al utilizado para hacer la inferencia. Como resultado, la matriz DTM de modelado se reduce a dos matrices. En la primera, $\boldsymbol{\theta}$, sus filas indican la distribución de tópicos en los documentos $P(topic_k/document_d)$. En la segunda, φ, sus filas que indican la distribución de palabras sobre los temas $P(token_v/topic_k)$.

(iv) *Postprocesamiento*

Este paso implica procesar los resultados y obtener una descripción de los tópicos. La distribución de términos temáticos no viene con una interpretación semántica. Sin embargo, dependiendo de la frecuencia de las palabras, los temas se pueden etiquetar correctamente en la mayoría de los casos. Lewis et al. (2013) mencionan que los análisis algorítmicos tienen una capacidad muy limitada para comprender significados latentes en el lenguaje humano, por lo que el etiquetado manual se considera un estándar (Lau et al., 2011). Sin embargo, en este último caso, el etiquetado puede proporcionar diferentes etiquetas de temas según el investigador que lo realice. En la aplicación LDAShiny se implementó una función del paquete textmineR (Jones, 2019) que proporciona un etiquetado de tópicos basado en un algoritmo de etiquetado naive basado en bigramas. Sin embargo, como se mencionó, estos algoritmos tienen capacidades limitadas, pero pueden servir como guía.

Una vez etiquetados todos los temas, con la ayuda de la matriz $\boldsymbol{\theta}$ (distribución de los tópicos en los documentos), se continúa con el procedimiento, asignando documentos a cada tema, clasificándolos Dependiendo de cuál es la mayor probabilidad de que un documento trate sobre un tema. De esta forma también se agruparán los documentos. El etiquetado requiere la validación por parte de un experto en el campo de investigación sobre el que se realiza la revisión, de lo contrario, se podrían obtener temas mal etiquetados y un resultado no válido (Asmussen & Møller, 2019).

Para facilitar la caracterización de los temas en términos de sus tendencias a lo largo del período de revisión considerado (años, meses, etc.), se utilizan las pendientes de regresión

simple para cada tema. En cada regresión para cada tema, el año es la variable independiente y las proporciones de los temas en cada año la variable respuesta (Griffiths & Steyvers, 2004):

$$\theta_k^y = \frac{\sum_{m \epsilon y} \theta_{mk}}{n^y} \quad (74)$$

donde:

$m \epsilon y$ representan los artículos publicados en un año determinado y

θ_{mk} la proporción del tópico k-ésimo y

n^y el número total de artículos publicados en el año y (Xiong et al., 2019).

Los tópicos cuyas pendientes de regresión son positivas (negativas) a un nivel de significación estadística dado, se interpretan como que han presentado un interés creciente (decreciente, respectivamente). Si las pendientes no son significativas, los temas se clasificarán como que han presentado tendencias fluctuantes.

Interfaz gráfica de usuario (GUI) LDAShiny

LDAShiny está basado en la web y como hemos apuntado con anterioridad ha sido desarrollado en R utilizando el paquete de aplicaciones web Shiny (Chang et al., 2009). Este paquete proporciona una plataforma integrada para la revisión exploratoria de información científica, que ofrece una serie de opciones para administrar, explorar, analizar y visualizar datos. Como ya hemos dicho, es particularmente beneficioso para los investigadores que no están tan familiarizados con el entorno R o con la programación en general, pero que desean utilizar los métodos que aquí se han descrito. En el Anexo 3 se encuentra el código fuente de la aplicación desarrollada.

Se accede al paquete LDAShiny desde Comprehensive R Archive Network (CRAN) en http://CRAN.R-project.org/package=LDAShiny.

Para instalarlo, cargarlo y ejecutarlo, se debe escribir lo siguiente en la consola de R:

```
R > install.packages("LDAShiny")
R > library("LDAShiny")
R > LDAShiny::runLDAShiny()
```

La GUI propuesta en este trabajo proporciona un menú que, de arriba a abajo, guía al usuario a través del análisis (Figura 5):

- ***About LDAShiny***: este panel sirve como página de introducción del software. Muestra la información general de la aplicación, así como el objetivo del software, (tanto en inglés como en español);
- ***Preprocessing***: proporciona una interfaz para que los usuarios carguen los datos a analizar. Además, también existen diferentes opciones para realizar el preprocesamiento (Figuras 6 y 7);
- ***Document Term Matrix Visualizations***: en este menú se puede elegir cómo visualizar la matriz de términos y documentos, tanto en forma tabular como en forma gráfica. La tabla de datos se puede descargar en formato ".csv", ".xlxs" o ".pdf" o se puede copiar al portapapeles. Los gráficos (barplot o wordcloud) se pueden descargar en formato ".png", ".pdf", ."jpeg" y ".svg". (Figura 8);
- ***Number of topics (inference)***: En este menú están disponibles las opciones para configurar los parámetros de entrada de cada una de las métricas que pueden utilizarse para encontrar el número de tópicos. Todas las pestañas tienen en común una secuencia numérica de ingreso ("desde", "hasta" y "por"), "Candidato número de temas k" y además se incluye un campo "Otros K" donde se puede escribir cualquier número(s) diferente del rango que desea probar. En todos los casos se muestra un gráfico con los resultados de la métrica utilizada, como resultado de su ejecución (Figura 9);
- ***LDA Model***: Una vez definido el número de temas, se ajusta el modelo LDA. Los parámetros de inferencia deben utilizarse como guía. Sin embargo, algunos pueden modificarse, como el número de iteraciones, que puede ser mayor. También se podría utilizar la recomendación de Griffiths & Steyvers (2004), estableciendo un valor de α de 50/k (Figura 10);
- ***Tutorial***: este menú ofrece una viñeta (en inglés y español) con vídeos que sirven como guía rápida donde se explican los pasos básicos para utilizar el software.

Figura 5. Interfaz gráfica de usuario (GUI) de la aplicación LDAShiny

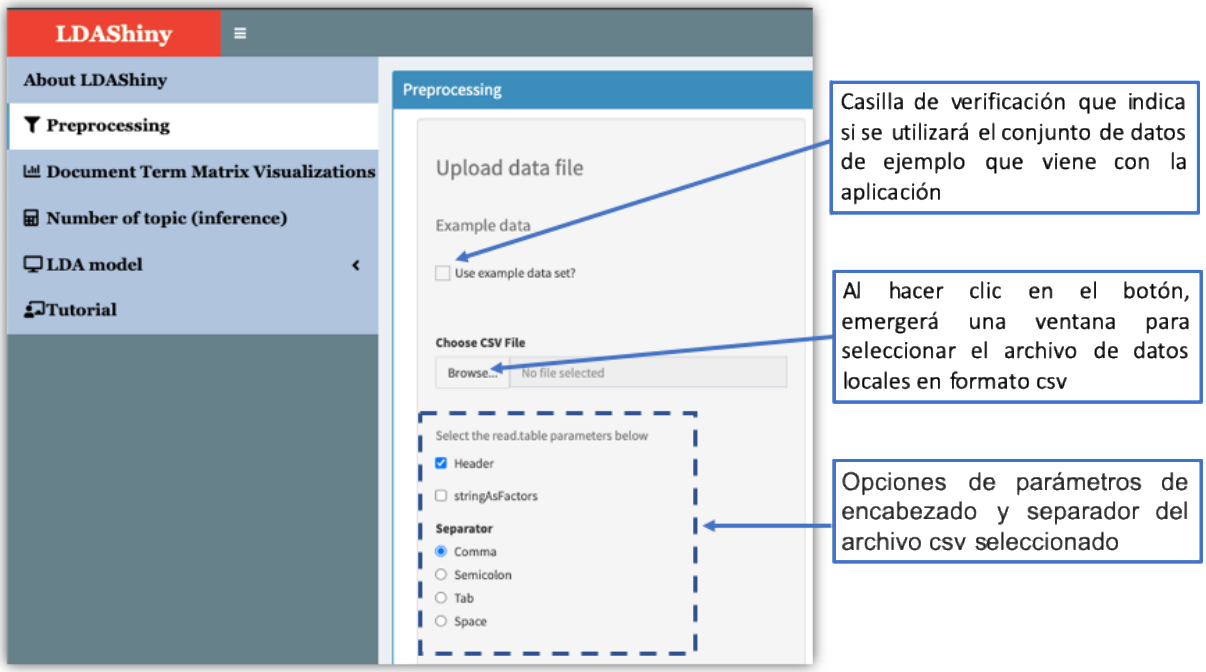


Figura 6. Opciones de carga de datos del menú *preprocessing*

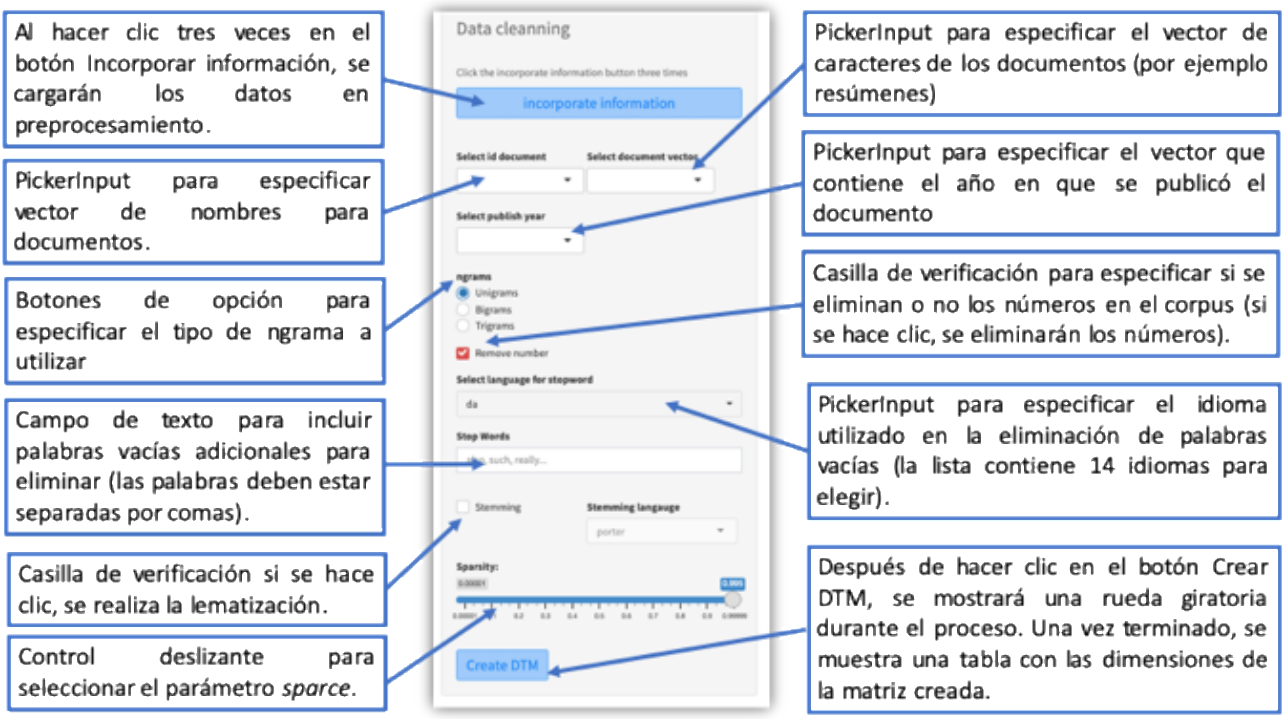


Figura 7. Opciones de "limpieza" de datos del menú *preprocessing*

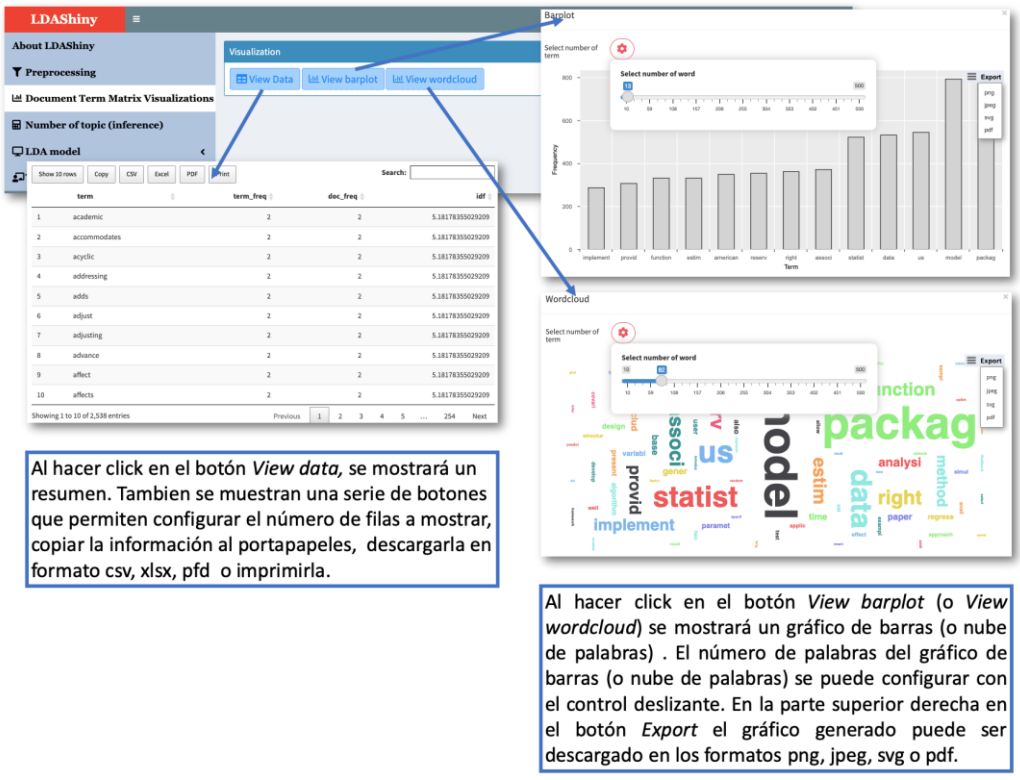


Figura 8. Opciones del menú *Document Term Matrix Visualizations*

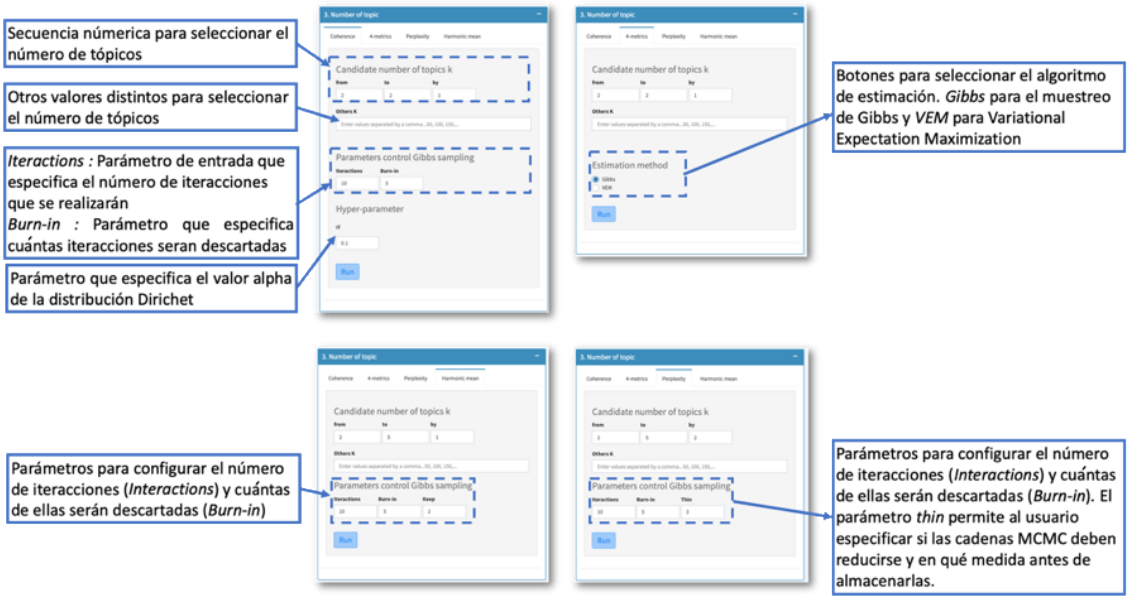


Figura 9. Opciones del menú *Inference* para las métricas implementadas en LDAShiny

3.2. UTILIZACIÓN DEL PAQUETE DESARROLLADO PARA LA REVISIÓN BIBLIOGRÁFICA DE LAS APLICACIONES DE LOS MÉTODOS PONDERADOS GEOGRÁFICAMENTE.

En este apartado del trabajo se presenta la revisión de las aplicaciones de los métodos de análisis de datos ponderados geográficamente descritos en el capítulo II. Sin embargo, es importante destacar que la metodología aplicada es muy útil cuando la cantidad de documentos es relativamente grande como es el caso de las aplicaciones de los GWR.

Además de las extensiones mencionadas, la GWR también ha expandido su aplicación a una amplia gama de áreas de investigación, que incluyen no solo la Geografía y otras Ciencias de la Tierra, sino también Estadística, Economía, Física, Ingeniería y Ciencias de la Salud, entre otras. Esto pone de manifiesto la problemática de que cada vez existe un mayor número de publicaciones y que, aunque hoy día se pueden explorar realizando búsquedas por palabras clave, esto tiene a menudo el efecto de limitar la posibilidad de identificar toda la literatura relevante. Además, podría ocurrir que un artículo en particular (que puede compartirse con artículos similares) no siempre se pueda descubrir mediante este tipo de búsqueda (Srivastava & Sahami, 2009). Esto dificulta la capacidad para usar la información

de manera efectiva y para realizar un seguimiento de las nuevas investigaciones (Larsen & von Ins, 2010).

Ante la creciente diversidad de aplicaciones de la técnica GWR y sus extensiones, surge la necesidad de realizar estudios cuantitativos que ayuden a comprender mejor los siguientes problemas:

P1: ¿Cuáles fueron las principales revistas y cuáles los principales autores en la investigación de la GWR?

P2: ¿Cómo son las colaboraciones científicas entre países en relación a la aplicación de la GWR?

P3: ¿Cuáles son los principales temas de investigación en este campo?

P4: ¿Cómo han evolucionado estos temas de investigación con el tiempo?

P5: ¿Cómo se distribuyen estos temas entre países?

Las respuestas a estas preguntas nos ayudan a proporcionar una descripción completa y una comprensión de vanguardia de las aplicaciones de la GWR. Para abordar estos problemas, se utilizó la combinación de algoritmos avanzados de modelado de tópico LDA (Blei et al., 2003) y análisis bibliométrico. Esto permite una revisión más objetiva, robusta, estructurada y completa de este dominio de investigación -en rápida expansión- de lo que pueden proporcionar las revisiones y un análisis de la literatura tradicional (Vanhala et al., 2020).

3.2.1. Metodología

El proceso metodológico empleado se dividió en tres etapas:

(i) Búsqueda y recopilación de artículos científicos;

(ii) Estudio bibliométrico;

(iii) Identificación de tópicos de investigación.

(iv) Identifiación de las tendencia de lostópicos

(i) *Búsqueda y recopilación de artículos*

La búsqueda de artículos se realizó a través de las bases de datos Scopus y la Web of Science. Decidimos seleccionar estas bases de datos porque son de las más utilizadas por los

investigadores en diferentes áreas (Harzing & Alakangas, 2016). La búsqueda se restringió a artículos en donde se menciona la técnica GWR o sus extensiones, publicados en idioma inglés en el periodo comprendido desde octubre de 1996 (fecha de la primera publicación) hasta diciembre de 2021.

La base de datos preliminar, con los documentos obtenidos tras ejecutar la consulta de búsqueda, contenía 2572 artículos para la Web of Science y 2622 para Scopus. Esta muestra inicial se sometió a un proceso de filtrado donde se eliminaron los artículos repetidos, artículos que no contaban con resúmenes, año de publicación o afiliación de los autores. La muestra final obtenida consistió de 3183 artículos (Figura 10), de los cuales se obtuvieron las metavariables título, resumen, año de publicación, autores y afiliación de autores.

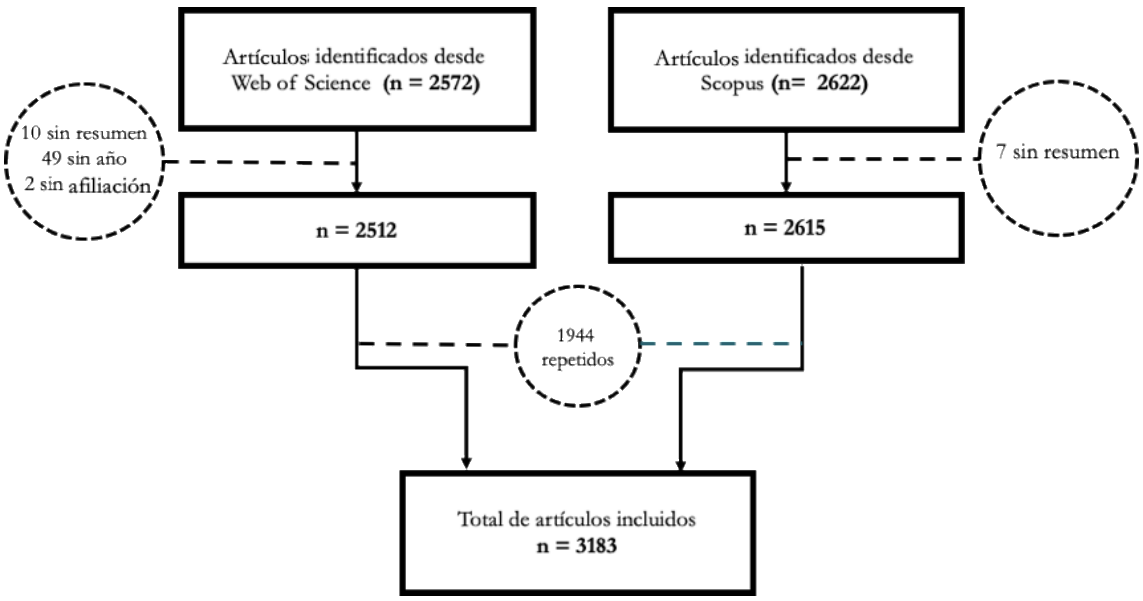


Figura 10. Diagrama de la selección de artículos sobre GWR entre octubre 1996 y diciembre de 2021

(ii) *Estudio bibliométrico*

Para responder a las preguntas planteadas, P1 y P2, utilizamos el análisis bibliométrico. Esto permite determinar diversos aspectos de la producción científica a partir de estructuras intelectuales y sociales de la muestra de publicaciones (Aria & Cuccurullo, 2017). El procesamiento de datos en esta parte del estudio se llevó a cabo utilizando el

paquete bibliometrix (Aria & Cuccurullo, 2017), un software de código abierto en lenguaje de programación R (R Core Team 2021).

(iii) *Identificación de tópicos de investigación*

Para responder a las preguntas P3, P4 y P5 se utilizó el método del modelo de tópicos LDA, descrito en el item 3.1.1. El procedimiento para la identificación de temas a través de LDA se dividió en tres etapas: (a) preprocesamiento, (b) creación del modelo LDA y (c) etiquetado de temas. Esta parte del estudio se llevó a cabo utilizando el paquete LDAshiny que ha sido programado como parte de esta tesis.

(a) Preprocesamiento

Para aumentar la coherencia de los tópicos, cada resumen fue tokenizado usando bigramas y se eliminó una lista estándar de palabras denominadas StopWord. Esta lista de palabras pueden ser descargadas de https://github.com/JavierDeLaHoz/stopword. Para la eliminación de números se utilizó el stemming y se fijó el parámetro sparce en 0,997, es decir, se toman los términos que aparecen en más del 0,3% de los documentos.

(b) Creación del modelo LDA

Para determinar el número de tópicos se realizaron simulaciones variando dicho valor, k, de 4 a 30, utilizando el muestreo de Gibbs, con 500 iteraciones. Se utilizó la coherencia como métrica para determinar el número de tópicos óptimo. Una vez determinado el número de tópicos se procedio a ajustar el módelo LDA utilizando 1000 iteraciones, y como parámetro α de la distribución Dirichlet se tomó el valor de 50/k, según recomendación de Griffiths & Steyvers (2004). El β fue calculado con información del corpus.

(c) Etiquetado de temas

En primer lugar, utilizamos un algoritmo naive de etiquetado basado en bigramas probables (estas etiquetas se basan en $P(\text{bi-gram} \mid \text{topic}) - P(\text{bi-gram})$) proporcionado por la aplicación. Sin embargo, dado que los algoritmos tienen una capacidad muy limitada para comprender los significados latentes del lenguaje humano (Lau et al., 2011), también se decidió utilizar un

etiquetado manual. Este último etiquetado se considera un estándar en el modelado de temas, y utiliza dos fuentes de información: una lista de 15 de palabras frecuentes de cada tópico (las más probables) y una muestra de los títulos de los artículos clasificados en el tópico correspondiente. Además, para mejorar el etiquetado de los tópicos, los visualizamos en dos dimensiones mediante un análisis de escalamiento multidimensional no métrico (NMDS, por sus iniciales en inglés) (Siever & Shirley, 2014). Este análisis muestra la similitud entre tópicos con respecto a su distribución de probabilidad sobre las palabras.

(iv)Tendencia de tópicos

Teniendo en cuenta que el número de documentos y palabras es bastante amplio y no es posible comprender la tendencia de los tópicos de forma intuitiva, se utilizaron algunos índices cuantitativos propuestos por Xiong et al. (2019), los cuales se describen a continuación:

- Distribución de tópicos a lo largo del tiempo (θ_k^y):

$$\theta_k^y = \frac{\sum_{m \epsilon y} \theta_{mk}}{n^y} \tag{75}$$

donde:

$m \epsilon y$ representa los artículos publicados en un año determinado

θ_{mk} es la proporción del k-ésimo tópico en cada artículo, y

n^y es el número total de artículos publicados en el año *y* (Xiong et al., 2019).

Con el fin de facilitar la caracterización de los tópicos en cuanto a su tendencia se emplearon pendientes de regresión simple para cada tema, siendo el año la variable independiente y las proporciones de los temas en los años correspondientes la variable de respuesta (Griffiths & Steyvers, 2004). Como ya explicamos anteriormente, aquellos tópicos cuyas pendientes de regresión resultaron positivas (o negativas) a un nivel de significación estadística de 0,01 se clasificaron como de tendencia positiva (o negativa). Del mismo modo aquellos tópicos cuyas pendientes no fueron significativas se clasificaron como de tendencia fluctuantes o sin tendencia.

- Popularidad de tópicos (P^i):

$$P^i = S_{NP}^i + S_{Tr}^i \tag{76}$$

$$S_{NP}^{i} = P_A^i / P_A^{max} \quad (77)$$

donde:

P^i es la popularidad de tópico *i,*

S_{NP}^{i} es la probabilidad normalizada (se divide cada probabilidad por el valor mayor) y

S_{Tr}^{i} toma valores de 1 si el tópico muestra tendencia positiva, 0,67 si no se muestra tendencia y 0,33 si hay tendencia a disminuir (Xiong et al., 2019).

- Distribución de tópicos por revista (θ_k^r):

$$\theta_k^r = \frac{\sum_{m \epsilon r} \theta_{mk}}{n^r} \quad (78)$$

donde

$m\epsilon r$ representa los documentos publicados en una determinada revista

θ_{mk} la proporción del k-ésimo tópico en cada documento, y

n^r el número de documentos publicados en la revista *r* .

Con el objetivo de describir las relaciones entre países y tópicos se utilizó el siguiente índice

- Distribución de tópicos por país (θ_k^p):

$$\theta_k^r = \frac{\sum_{m \epsilon p} \theta_{mk}}{n^p} \quad (79)$$

donde:

$m\epsilon p$ representa los documentos cuyos autores trabajan en un determinado país

θ_{mk} la proporción del k-ésimo tópico en cada documento, y

n^p el número de documentos del país p.

Para el análisis del índice anterior se utilizó la técnica HJ-Biplot (Galindo, 1986), una extensión de los biplots clásicos introducidos por Gabriel (1971).

La diferencia fundamental entre los métodos clásicos y el HJ-Biplot es que los primeros son capaces de reconstruir los datos de la matriz a través del producto interno de los marcadores fila y columna (Gabriel los nombró como GH-Biplot, con alta calidad de

representación para las columnas, y JK-Biplot, con alta calidad de representación para las filas), mientras que el HJ-Biplot consigue obtener alta calidad de representación para las filas y para las columnas de forma simultánea (Galindo, 1986).

Existen excelentes libros donde se explican en detalle la fundamentación teórica de los métodos biplot, como Biplots (Gower & Hand 1995), Biplots in practice (Greenacre, 2010), Understanding biplots (Gower et al. 2011), y logicamente los "seminal papers" de Galindo (1986) y Gabriel (1971). A ellos remitimos al lector.

Para realizar una interpretación correcta de un gráfico biplot debemos evaluar inicialmente la absorción de inercia recogida por los ejes factoriales, posteriormente debemos evaluar que los individuos y variables, estén bien representados en dicho plano.

Una característica importante de la representación gráfica producida por los métodos Biplot (en la que las variables se representan mediante vectores y los individuos mediante puntos) es que su interpretación se basa en conceptos geométricos simples, como es la proyección de las filas y columnas de la matriz de datos sobre una representación plana (Galindo, 1986):

- Las dimenciones o ejes factoriales representan gradientes. La interpretación se hace evaluando las contribuciones del factor al elemento. Aquellas variables que presentan altas contribuciones relativas de un eje y baja de los demás son las más importantes en la interpretación de los gradientes latentes;
- La longitud del vector que representa a una variable indica la variabilidad de la misma. Cuanto mayor es longitud, más variabilidad presenta la variable en el estudio y por tanto mayor información aporta a la hora de realizar la interpretación;
- El ángulo que forman los vectores se interpretan como la relación existente entre las variables que representan. Un ángulo agudo indica fuerte asociación positiva. Un ángulo recto indica independencia entre las variables, mientras que un ángulo llano indica fuerte relación inversa. Un ángulo aproximadamente recto sería indicativo de una posible independencia entre dichas variables;
- La distancia entre los puntos que representan a los individuos se interpreta como disimilaridad. La proximidad entre puntos indica similitud entre ellos, en relación a las distintas variables;

- La relación de un individuo con las variables se analiza a través de la proyección perpendicular del punto individuo sobre los vectores variables, aproximando el correspondiente valor de la variable, pudiendo determinar qué variables son las que caracterizan más al individuo. Los individuos que se proyectan en la dirección de la punta del vector (variable) son los que presentan valores más altos en esa variable; los que se proyectan en la dirección contraria son los que toman los valores más bajos.

Los cálculos y figuras para el análisis HJ-biplot, se realizaron utilizando una modificación (de elaboración propia) de la función HJ.Biplot del paquete MultBiplotR (Vicente-Villardon, 2021). Dicha modificación permite incluir las bandera de los países en el biplot generado.

3.2.2. Resultados

Estudio bibliométrico

Las estadísticas básicas sobre el conjunto de datos analizados se presentan en la Tabla 1. Los 3183 artículos considerados en el estudio procedían de 900 revistas diferentes.

Los autores totales fueron 5732, de los cuales sólo 139 lo eran de artículos de un único autor (4,3%). La media de autores por documento es de 2,19. El promedio de citas por documento es de 21,81 lo que indica que los artículos son muy valorados en la academia. Estas estadísticas sugieren que la colaboración entre especialistas de diferentes disciplinas es necesaria debido a la naturaleza interdisciplinaria de la GWR.

Los resultados obtenidos durante el período de estudio considerado (1996-2021), nos permiten afirmar, en general, que el número de publicaciones sigue una tendencia de tipo exponencial (Figura 11). Además, el número de publicaciones aumentó a una tasa anual promedio de 36,37%. Por otro lado, se observa que, durante los 15 primeros años, el número de publicaciones no era superior a 100 artículos (Figura 11). Sin embargo, se observa un fuerte incremento del número de artículos a partir del año 2015 (Figura 11).

Tabla 1. Principales estadísticas sobre la colección de artículos de GWR entre 1996 y 2021

Descripción		Resultados
Información principal sobre los datos	Espacio de tiempo	1996:2021
	Fuentes (Revistas)	900
	Documentos	3183
	Promedio de años desde la publicación	4,8
	Citas promedio por documentos	21,81
	Promedio de citas por año por documento	3
Contenidos de documentos	Keywords Plus (ID)	8503
	Palabras clave del autor	6048
Autores	Autores	5732
	Apariciones del autor	10517
	Autores de documentos de un solo autor	139
	Autores de documentos de varios autores	5593
Colaboración de autores	Documentos de un solo autor	163
	Documentos por autor	0,456
	Autores por documento	2,19
	Coautores por documentos	4,02
	Índice de colaboración	2,28

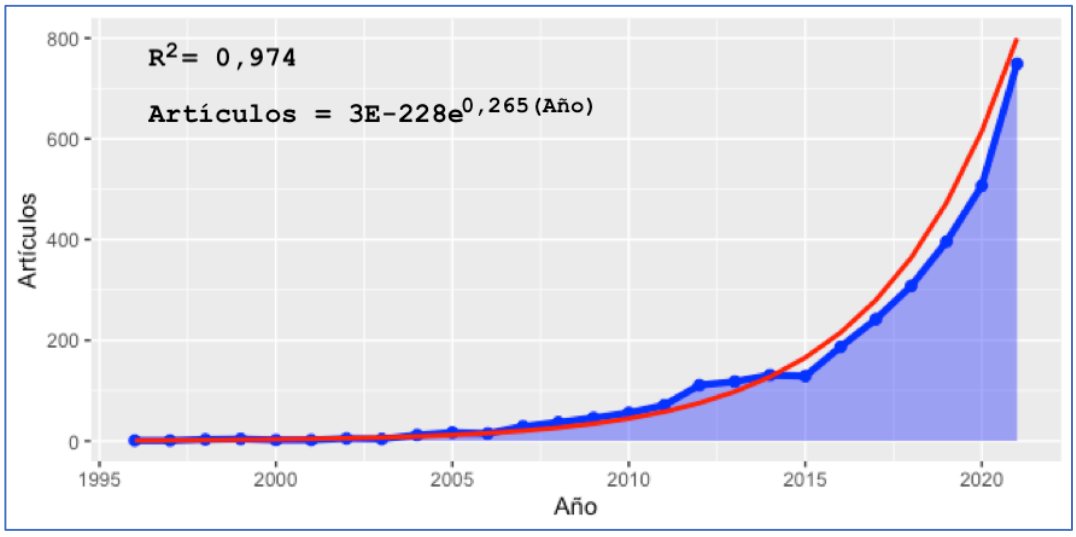


Figura 11. Número de publicaciones sobre GWR entre octubre 1996 y diciembre de 2021. La línea roja muestra la tendencia de tipo exponencial

En lo que respecta a las revistas, observamos que según el número de publicaciones, el *International Journal of Environmental Research and Public Health*, con 107 artículos, ocupa el primer lugar, seguido de *Sustainability* (Switzerland), con 100 publicaciones (Figura 12). El resto de revistas registradas no supera los 100 artículos cada una (Figura 12).

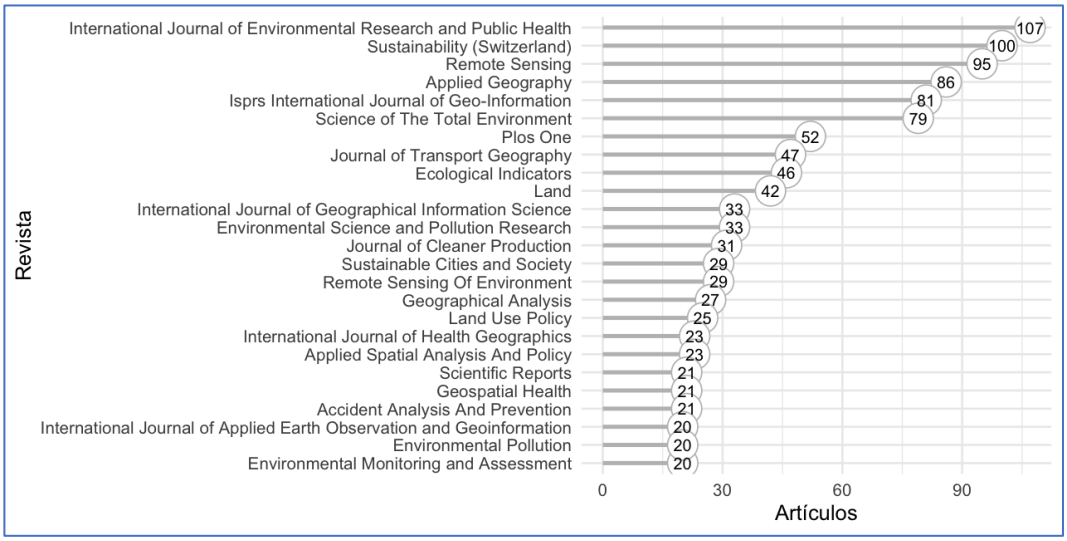


Figura 12. Número de artículos sobre GWR por revista, entre octubre 1996 y diciembre de 2021

El país al que se asignó a un documento fue el de afiliación del primer autor. Según esta asignación, los artículos analizados provienen de instituciones ligadas a 90 países, entre los que destacan China (1106; 34,75%), Estados Unidos de América (715; 22,43%), Reino Unido (104; 3,27%), Canada (77; 2,42%), Irán (75; 2,36%), Brasil (71; 2,23%), España (64; 2,01%), Italia (58; 1,82%), Australia (56; 1,76%), India (51; 1,6%) y Corea del Sur (51; 1,6%) (Figura 13).

La red de colaboración de los autores es crucial para comprender la dirección de la investigación en numerosos dominios de estudio (Mumu et al., 2021). Esta colaboración es a menudo el resultado de la formación de centros de investigación que impulsan el crecimiento y la expansión futura de un tema particular en estudio. La red de coautores ilustra las conexiones intelectuales entre investigadores según el país (Figura 14). De los 90 países encontrados se tuvieron en cuenta para la construcción de la red aquellos cuyo número de artículos fuera igual o superior a 10, por lo que solo se muestran 50 países. En la figura 14, el tamaño del círculo indica el número de publicaciones de un país. La intensidad de la colaboración se mide por el grosor de las líneas y el espaciado de los círculos.

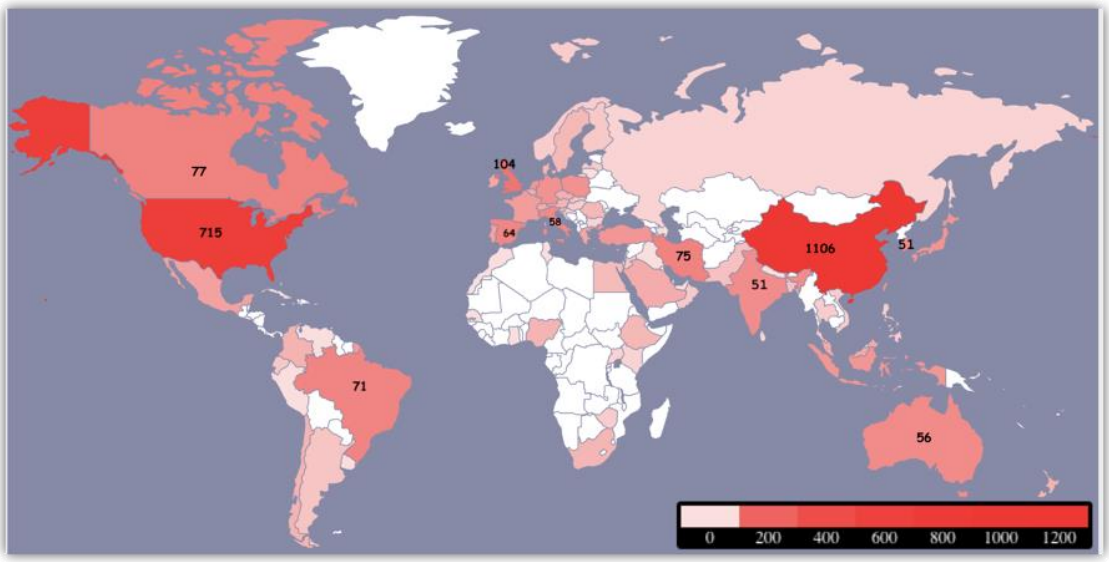


Figura 13. Origen geográfico de los 3183 artículos sobre GWR entre octubre 1996 y diciembre de 2021

En la investigación sobre las aplicaciones de la GWR, se observa en la Figura 14 la red de colaboración científica en donde los nodos representan los países y los enlaces son coautorías. Surgieron cinco redes principales, lideradas por China (clúster rojo), España (clúster verde), Reino Unido (clúster azul), Alemania (clúster naranja) y Australia (cluster morado), mientras que Eslovenia, Ecuador y Finlandia no se encuentran en ninguna de las redes. Se observa que existe una relación de cooperación más fuerte entre países del mismo grupo que entre países en diferentes agrupaciones (Figura 14). Por supuesto, esto no significa que no haya cooperación entre países situados en diferentes clústeres, más bien indica que puede haber algunos temas de investigación comunes entre los países del mismo grupo, haciendo que su cooperación sea más intensa. La colaboración internacional está muy concentrada en EE. UU y China, en términos de coautoría internacional de publicaciones relacionadas con la GWR. Esto es consistente con la tendencia general en ciencia (Wagner et al., 2015), en la que China se destacó como un importante colaborador estadounidense, superando las históricas relaciones de colaboración entre Europa y la nación norteamericana en otros campos.

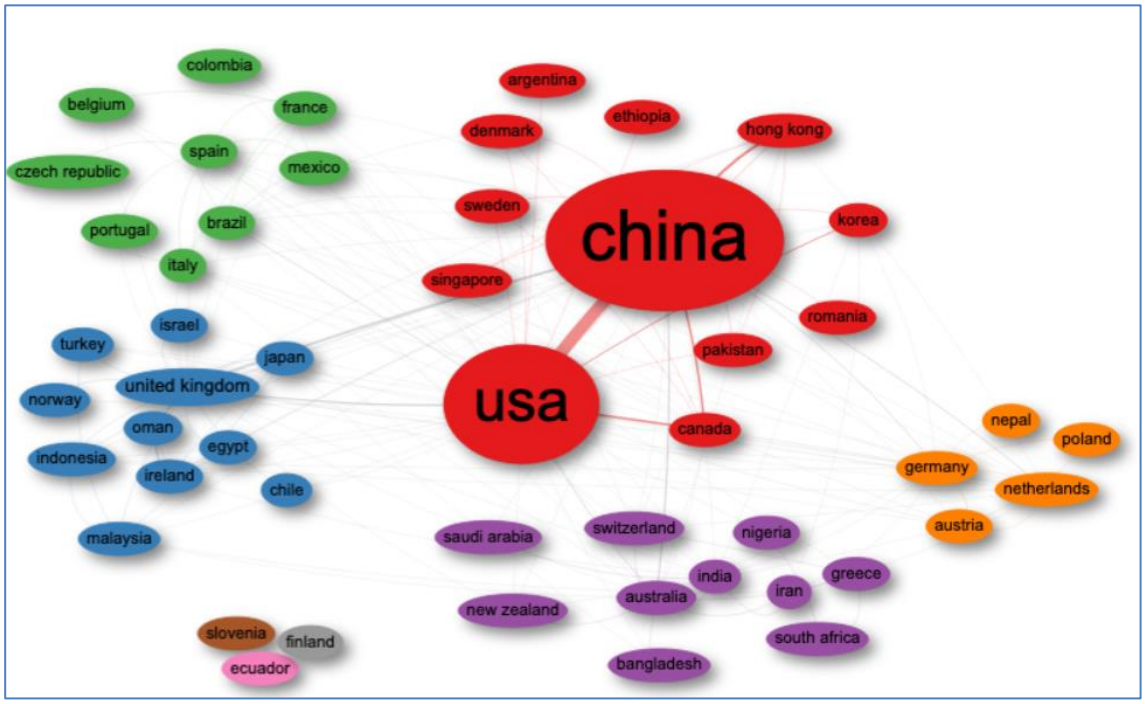


Figura 14. Red de coautorias basada en países de los artículos sobre GWR

Análisis de modelado de temas

Los resultados sugieren que el modelo LDA, con la puntuación de coherencia óptima, contiene 22 tópicos (k = 22).

En la Tabla 2 se pueden consultar los 22 temas latentes extraídos para el conjunto de artículos, mientras que en la Figura 15 se muestran las 10 palabras más frecuentes para cada tópico.

Cada tópico se considera como una colección de palabras que son semánticamente relevantes. Las palabras con mayores probabilidades en un determinado tópico reflejan el tema que se trata en ellos. Por lo tanto, podemos asignar un determinado subcampo de investigación a cada tema después de analizar cuáles son dichas palabras principales.

Por ejemplo, el tópico t_16 "COVID" está asociado con los términos *covid*, *vulner*, *social*, *case*, *popul*, *pandem*, *communiti*, *resili*, *transmiss*, …, todas ellas relacionadas con la COVID-19. Este tópico aparece en 70 documentos a pesar de que sólo se registran artículos desde el año 2020.

Tabla 2. 22 tópicos latentes en los 3183 artículos sobre GWR publicados en el período 1996-2021. Se muestra para cada tópico el etiquetado manual (Etiqueta) y el etiquetado proporcionado por la aplicación (Etiqueta 2), el número de documentos y la prevalencia estadística sobre la colección de artículos

Tópico	Etiqueta	Etiqueta 2	Número de documentos	Prevalencia (%)
t_1	"Factores de riesgo en enfermedades"	risk_factor	170	4,452
t_2	"Precios de vivienda"	hous_price	182	3,477
t_3	"Calidad del agua"	water_qualiti	91	2,602
t_4	"China"	influenc_factor	246	7,819
t_5	"Cambio climático"	climat_chang	88	4,436
t_6	"Riqueza de especies"	speci_rich	94	2,466
t_7	"Factores socioeconómicos"	socio_econom	133	6,014
t_8	"Heterogeneidad espacial"	spatial_heterogen	210	9,834
t_9	"Contaminación del aire"	pm_concentr	125	3,086
t_10	"Transporte"	built_environ	180	4,276
t_11	"Modelos GWR"	vari_coeffici	344	10,722
t_12	"Precipitación"	spatial_resolut	71	2,045
t_13	"Carbono orgánico"	soil_organ	135	3,557
t_14	"Uso del suelo"	land_cover	103	3,296
t_15	"Clima"	surfac_temperatur	137	4,104
t_16	"COVID"	covid_case	70	1,709
t_17	"Sensores remotos"	remot_sens	85	2,954
t_18	"Emisión de carbono"	spatial_tempor	90	3,508
t_19	"Modelos GWR"	ordinari_squar	129	6,045
t_20	"Área urbana"	urban_area	153	4,257
t_21	"Incendios forestales"	forest_fire	64	2,258
t_22	"Salud pública"	risk_factor	283	7,084

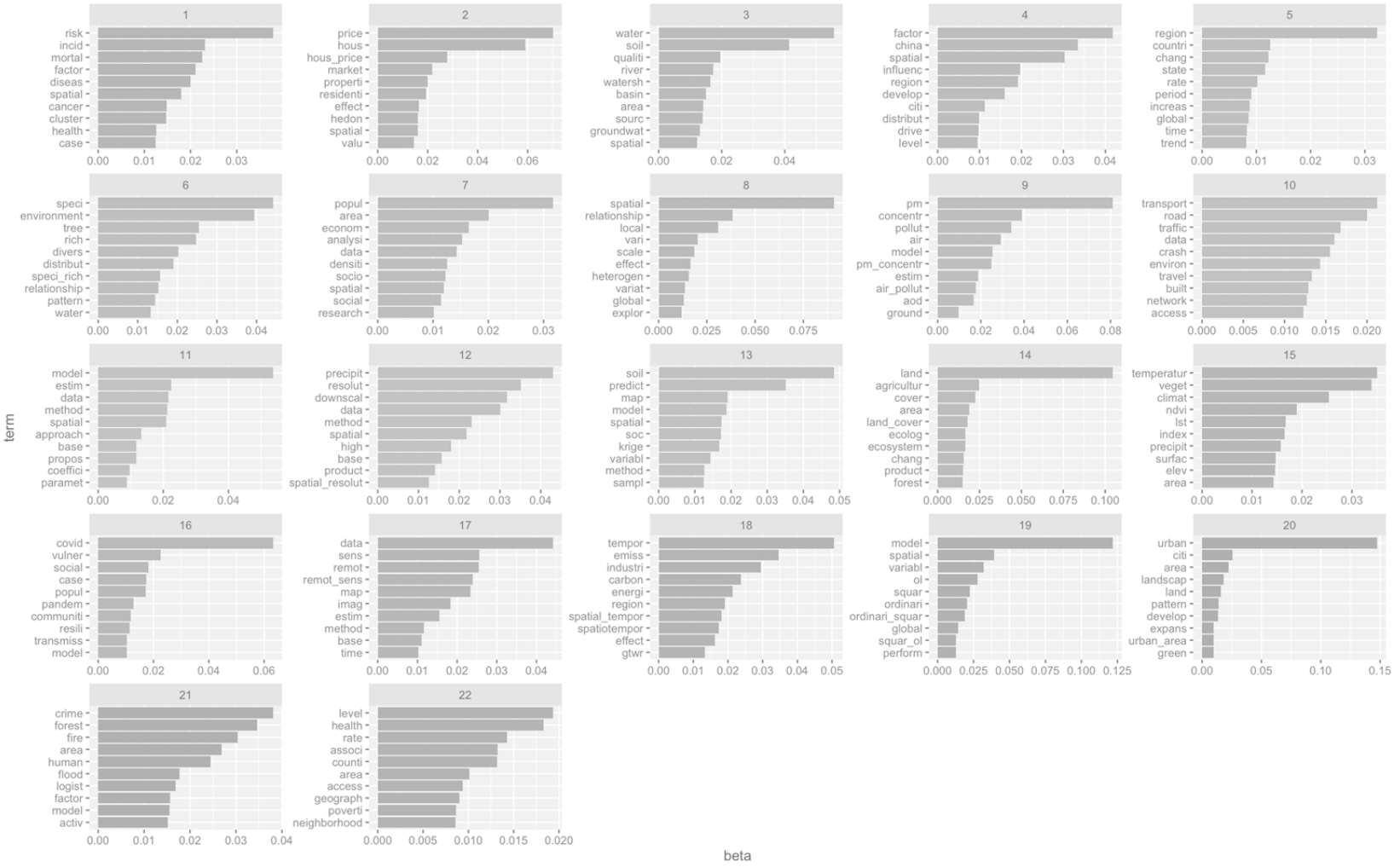


Figura 15. Top-10 términos para cada uno de los 22 tópicos

En la Figura 16 se puede visualizar los resultados del análisis NMDS, los tópicos como círculos en el plano bidimensional cuyos centros se determinan calculando la divergencia de Jensen-Shannon entre los temas (las proporciones proporcionadas por la matriz θ) y luego usando una escala multidimensional para proyectar las distancias entre temas en dos dimensiones.

La distancia entre los nodos en el plano representa la similitud entre nodos con respecto a la distribución de palabras. Por tanto, cuánto más cercanos se presenten dos nodos más similares serán, en cuanto a dicha distribución de palabras y cuanto más alejados más disimilares serán. El área de los nodos indican la prevalencia del tema dentro del corpus, con nodos más grandes aquéllos temas más prevalentes dentro del corpus.

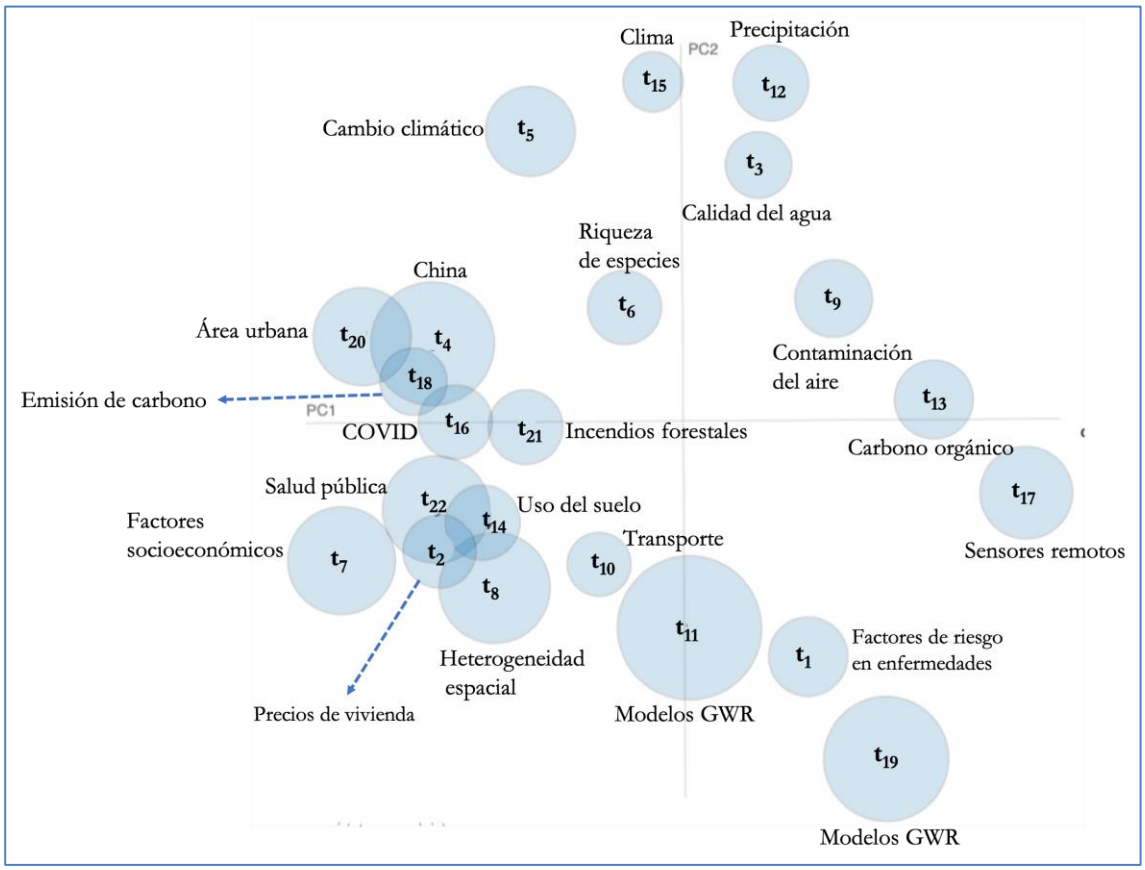


Figura 16. Representación bidimensional de la distancia inter-tópico mediante escalamiento multidimensional no métrico

Tendencia de tópicos

El análisis de la dinámica de la proporción de los tópicos a lo largo del tiempo, nos permite comprender la tendencia general de la investigación sobre GWR.

Identificamos la tendencia de cada uno de los 22 tópicos a lo largo del tiempo de modo que encontramos que las probabilidades de 14 de ellos, a saber: "Factores de riesgo en enfermedades"(t_1), "Precios de vivienda"(t_2), "Calidad del agua"(t_3), "China"(t_4), "Contaminación del aire"(t_9), "Transporte"(t_10), "Precipitacion"(t_12), "Carbono orgánico"(t_13), "Uso del suelo"(t_14), "COVID"(t_16), "Sensores remotos"(t_17), "Emisión de carbono"(t_18), "Área urbana"(t_20), "Incendios forestales"(t_21), aumentaron progresivamente con el tiempo. En dos de ellos, "Heterogeneidad espacial"(t_8) y "Modelos GWR"(t_11) disminuyeron, mientras que en los tópicos restantes "Cambio climático"(t_5), "Riqueza de especies"(t_6), "Factores socioeconómicos"(t_7), "Clima"(t_15), "Modelos GWR"(t_19) y "Salud pública"(t_22) su proporción fluctuó en el tiempo, sin presentar tendencias marcadas (Figura 17).

Los resultados encontrados son consistentes con la idea de que la investigación muestra fuertes tendencias con temas que surgen en un momento dado y cuya prevalencica disminuye regularmente con el tiempo (Griffiths & Steyvers 2004).

El aumento en la proporción de algunos temas en tiempos más recientes, indica que se trata de campos de investigación emergentes, mientras que la disminución de otros muestra una disminución del interés por parte de los investigadores en los últimos años. Además en algún tema el alto patrón de frecuencia encontrado fue seguido por una tendencia negativa durante el período de estudio, lo que indica una posible disminución de su interés dentro de la comunidad científica.

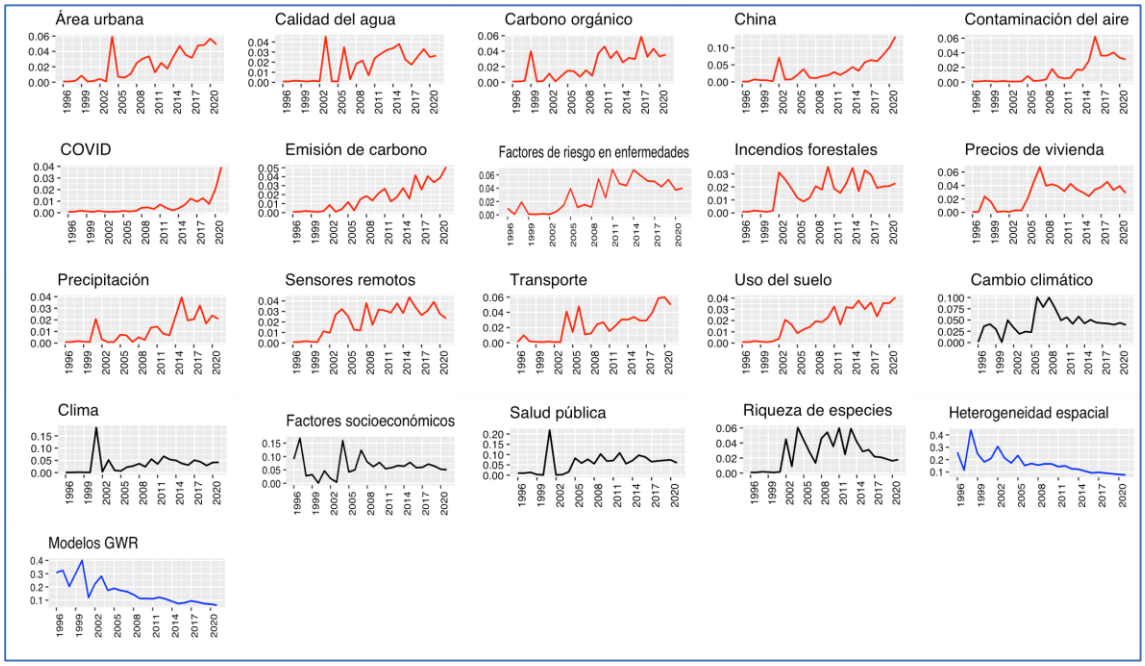


Figura 17. Tendencias de tópicos latentes en artículos sobre GWR publicados en el período 1996-2021. En color rojo temas con tendencia creciente, azul tendencia decreciente y negro tendencia fluctuante

En la Figura 18 se presenta un mapa de calor en donde se puede observar la distribución de tópicos por año de forma comparativa. Se forman cuatro grupos de años, sin embargo, los tópicos que más predominan de forma general son: "Modelos GWR" (t_11), "Modelos GWR" (t_19). Estos dos temas, aunque recbieron la misma etiqueta, hacen referencia a dos cuestiones distintas sobre el modelo GWR (el primero hace referencia a la estimacón del modelo mientras el segundo a las extensiones del mismo) y "Heterogeneidad espacial" (t_8). Se observa que el año 2001 no se agrupa y esto es debido a que adémas de los tópicos "Modelos GWR" (t_11), "Heterogeneidad espacial" (t_8) y Modelos GWR"(t_19) aparece el t_15 (“Clima”) como otro de los tópicos principales.

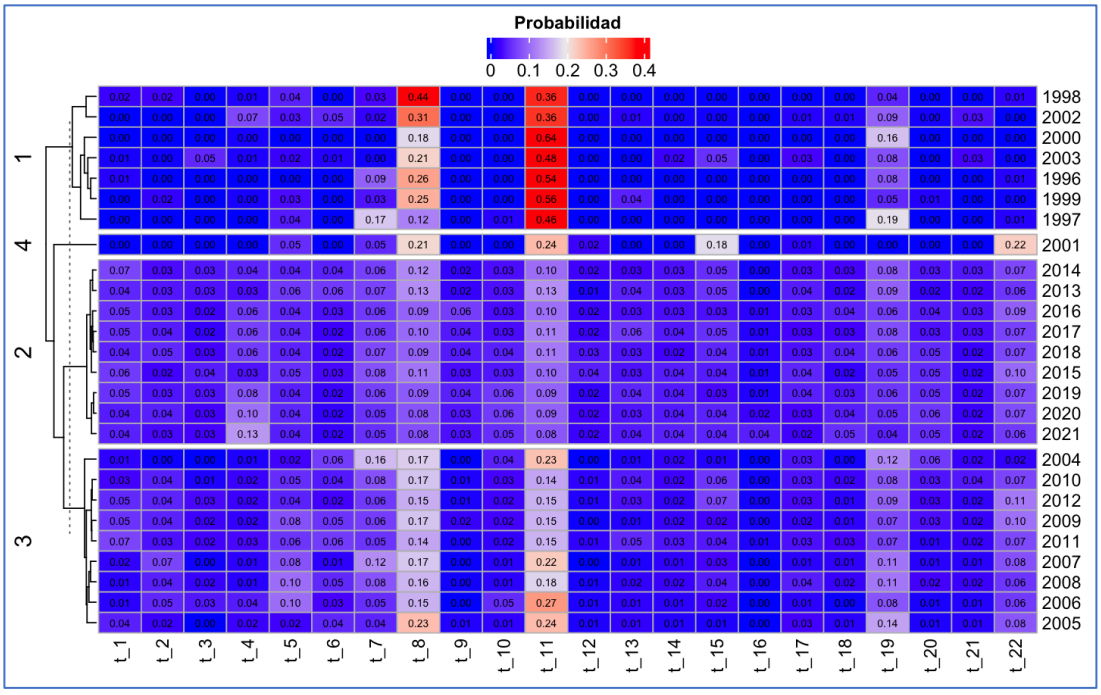


Figura 18. Mapa de calor de tópicos entre los años 1996 y 2021. Se muestra la distribución de las proporciones de los tópicos por año en donde la suma de filas es igual a 1

A pesar de los resultados anteriores, identificar los tópicos más representativos utilizando el análisis de las tendencias de forma aislada, es complejo. Por lo tanto, el interés por algunos tópicos puede haber disminuido (o aumentado) con el tiempo, pero aún tienen probabilidades relativamente grandes (o bajas). La popularidad de los tópicos, teniendo en cuenta la tendencia y la probabilidad del tema, se presentan en la Tabla 3. Se observa que los tres tópicos más populares en orden descendente son "China" (t_4), "Factores de riesgo en enfermedades" (t_1) y "Transporte"(t_10). En general, los tópicos con tendencias crecientes con probabilidades altas o moderadas tienen puntuaciones de popularidad más altos. Los tópicos con los valores de popularidad más bajos fueron "Cambio climático" (t_5), "Riqueza de especies" (t_6) y "Clima" (t_15).

La existencia de colores rojos (que indican probabilidades mayores) en el mapa de calor de la Figura 19 denota una escasa distribución de tópicos de investigación en una revista, en otras palabras, estas revistas se encuentran especializadas en tópicos particulares, como por ejemplo las revistas del cluster 6 (*Geographical Analysis*, *Environment and Planning A, Spatial Statistics, International Journal of Geographical Information Science y Journal of Geographical Systems*) tienen al tópico t_11 ("Modelos GWR") como su principal tema. De la misma forma en el cluster 1 las revistas se encuentran especializadas en el tópico t_10 ("Transporte"). En

el cluster 3 se observa que los temas de investigación tienen una distribución más amplia en sus revistas.

Tabla 3. Popularidad de los tópicos organizados en orden descendente. El tópico más popular será aquel con el valor de popularidad mayor. La popularidad normalizada se obtiene dividiendo cada probabilidad entre el valor máximo encontrado

Tópico	Etiqueta	Probabilidad	Probabilidad normalizada	Tendencia	Popularidad
t_4	"China"	0,078	0,729	1,000	1,729
t_1	"Factores de riesgo en enfermedades"	0,045	0,415	1,000	1,415
t_10	"Transporte"	0,043	0,399	1,000	1,399
t_20	"Área urbana"	0,043	0,397	1,000	1,397
t_13	"Carbono orgánico"	0,036	0,332	1,000	1,332
t_22	"Salud pública"	0,071	0,661	0,670	1,331
t_11	"Modelos GWR"	0,107	1,000	0,330	1,330
t_18	"Emisión de carbono"	0,035	0,327	1,000	1,327
t_2	"Precios de vivienda"	0,035	0,324	1,000	1,324
t_14	"Uso del suelo"	0,033	0,307	1,000	1,307
t_9	"Contaminación del aire"	0,031	0,288	1,000	1,288
t_17	"Sensores remotos"	0,030	0,276	1,000	1,276
t_8	"Heterogeneidad espacial"	0,098	0,917	0,330	1,247
t_3	"Calidad del agua"	0,026	0,243	1,000	1,243
t_19	"Modelos GWR"	0,060	0,564	0,670	1,234
t_7	"Factores socioeconómicos"	0,060	0,561	0,670	1,231
t_21	"Incendios forestales"	0,023	0,211	1,000	1,211
t_12	"Precipitación"	0,020	0,191	1,000	1,191
t_16	"COVID"	0,017	0,159	1,000	1,159
t_5	"Cambio climático"	0,044	0,414	0,670	1,084
t_15	"Clima"	0,041	0,383	0,670	1,053
t_6	"Riqueza de especies"	0,025	0,230	0,670	0,900

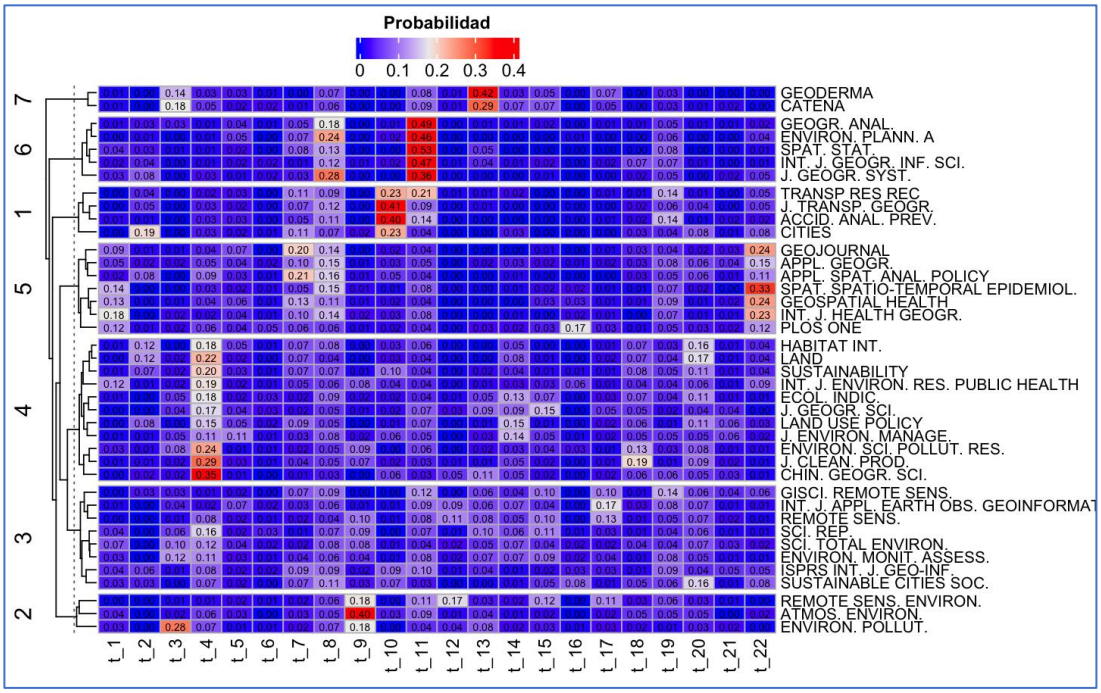


Figura 19. Mapa de calor de la distribución de tópicos por revista. La suma de filas es igual a 1

Relación entre tópicos y países

De manera similar al análisis de distribución de tópicos por revista o año, en la Figura 20 se muestra el mapa de calor de la distribución de tópicos en cada país. En este caso sólo se muestran los países que tienen más de 10 artículos publicados, teniendo en cuenta que se le asignó a un país la procedencia de un artículo considerando la afilición del primer autor.

El análisis de cluster identifica con éxito algunos países que muestran una gran similitud, como por ejemplo, Indonesia, Japón, México, Colombia, Brasil y República Checa. Como se puede observar, las ubicaciones geográficas, no juegan necesariamnete un papel crucial en la determinación de las aplicaciones de la técnica GWR.

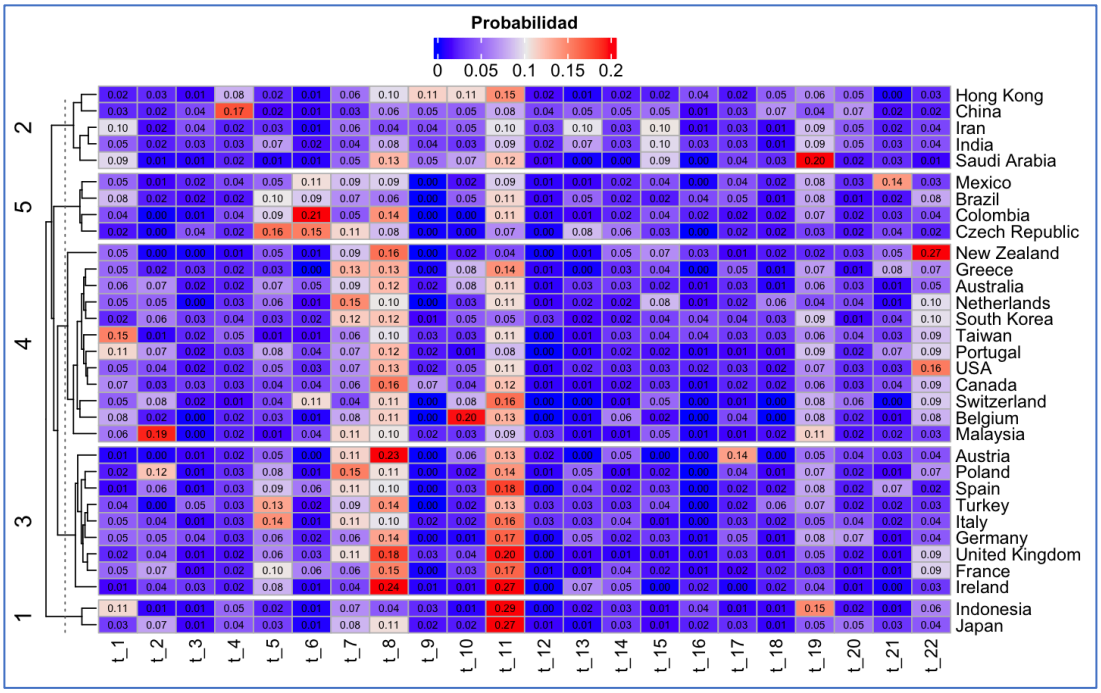


Figura 20. Mapa de calor de la distribución de tópicos por país. La suma de filas es igual a 1

En lo que respecta al análisis mediante el HJ-Biplot, de la matriz de distribución de tópicos por país (Ecuación 78), el porcentaje de varianza explicada de las dos primeras dimensiones o componentes, recoge una inercia del 29,7 % de la que el 15,9 % se corresponde al primer eje factorial (Figura 21).

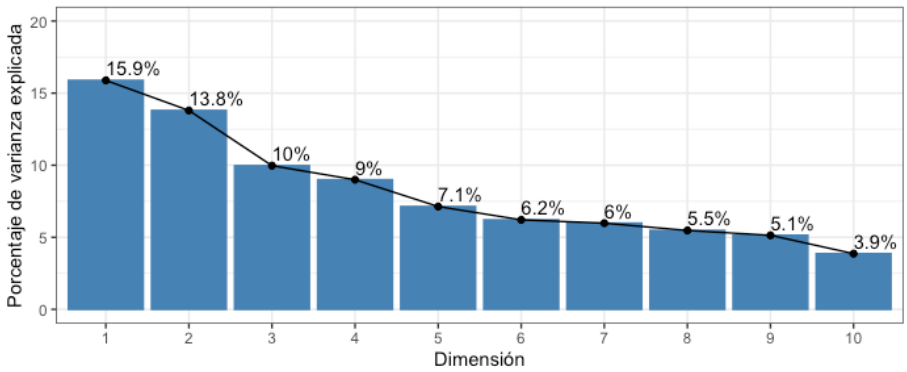


Figura 21. Porcentaje de varianza explicada por cada eje del análisis HJ-Biplot

Atendiendo a las contribuciones del factor al elemento para las variables (Tópicos), se observa que no todas las variables pueden ser interpretadas en el primer plano factorial. La línea discontinua roja en la Figura 22 indica la contribución promedio esperada. Si la contribución de las variables fueran uniformes, el valor esperado sería (Kassambara, 2017)

$$\frac{1}{Longitud(tópicos)} \tag{80}$$

Por lo que para una componente dada, una variable con una contribución mayor que este límite podría considerarse importante en la contribución a la dimensión o componente (Kassambara, 2017).

En el caso del plano i-j el valor esperado de la contribución fue calculado como (Kassambara, 2017):

$$Contrib = \frac{(Dim_i * Eig_i) + (Dim_j * Eig_j)}{(Eig_i * Eig_j)} \quad (81)$$

donde:

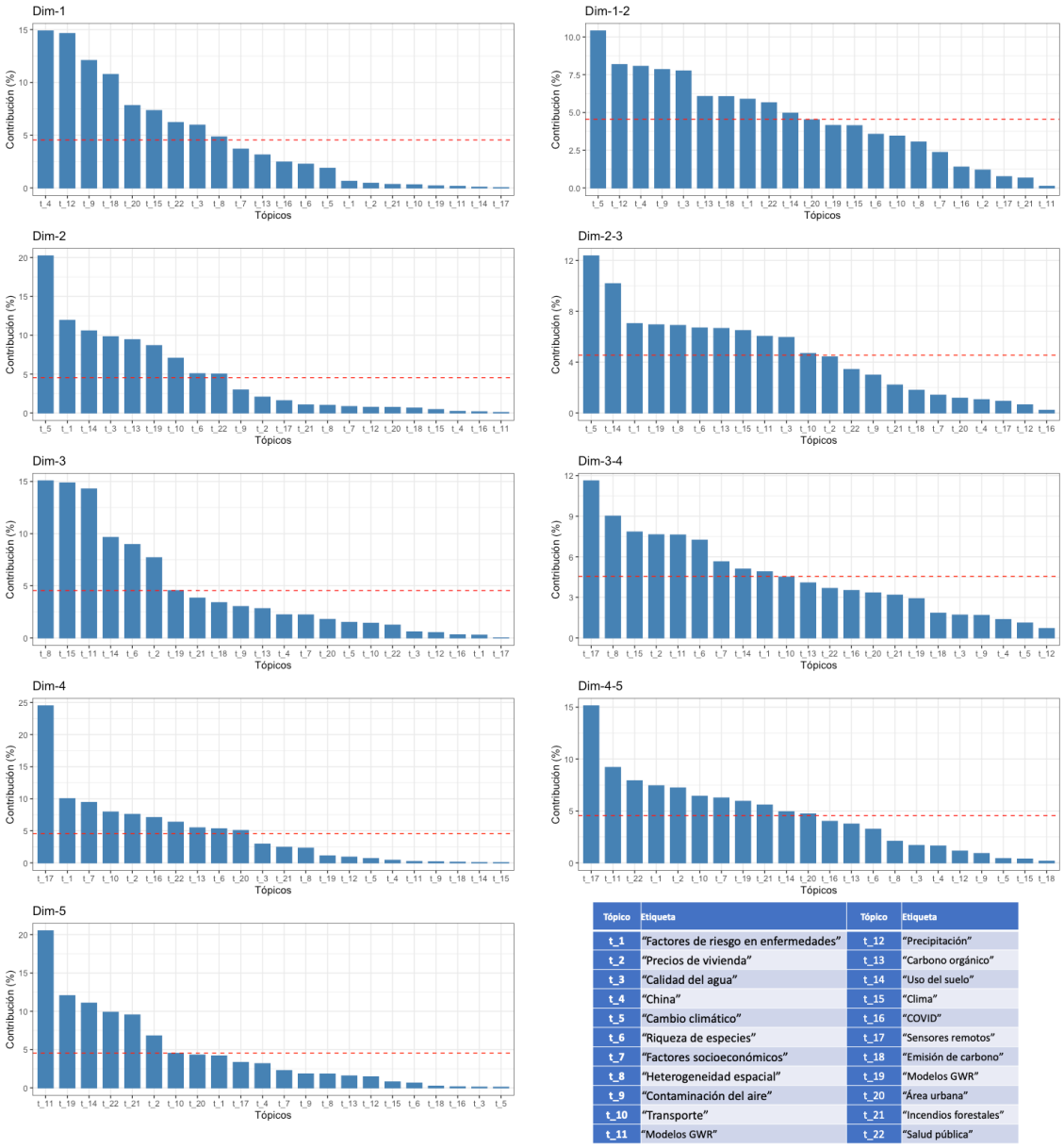
Dim_i y Dim_j son las coordenadas de la variable sobre el plano i-j,

Eig_i y Eig_j son los valores propios correspondientes a las dimensiones i y j, respectivamente.

En cuanto a los países (Figura 23) se presentan sus contribuciones (los 15 que más contribuyen) para cada dimensión y plano.

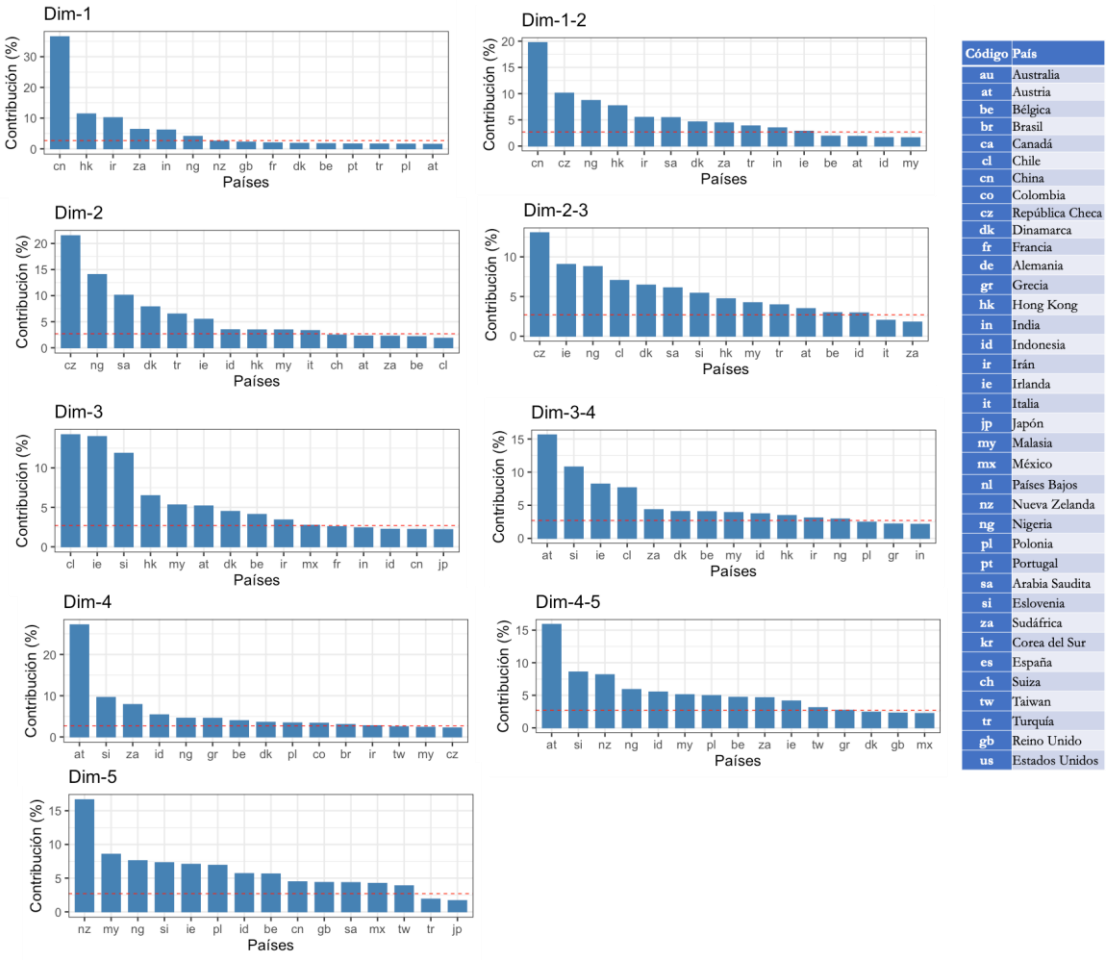
En la Figura 24a se muestra el gráfico factorial de los plano 1-2. Este gráfico incluye todos los tópicos y países. El color de los tópicos varía en tonos de azul (siendo los más oscuros los que presentan una contribución más alta). Los países se identifican mediante banderas y etiquetas del nombre abreviado utilizando el código ISO 3166-1 alfa-2.

En la Figura 24b se grafican los tópicos con la mayor calidad de representación ($cos^2 \geq 0,5$; calculado como la suma de los cuadrados de las coordenadas de las variables sobre los ejes) y los 10 países con mayor contribución del plano al elemento. Observamos que los tópicos que caracterizan el plano 1-2 estan relacionados con temas ambientales ("Cambio climático", "Calidad del agua", "Precipitación" y "Contaminación del aire").



Tópico	Etiqueta	Tópico	Etiqueta
t_1	"Factores de riesgo en enfermedades"	t_12	"Precipitación"
t_2	"Precios de vivienda"	t_13	"Carbono orgánico"
t_3	"Calidad del agua"	t_14	"Uso del suelo"
t_4	"China"	t_15	"Clima"
t_5	"Cambio climático"	t_16	"COVID"
t_6	"Riqueza de especies"	t_17	"Sensores remotos"
t_7	"Factores socioeconómicos"	t_18	"Emisión de carbono"
t_8	"Heterogeneidad espacial"	t_19	"Modelos GWR"
t_9	"Contaminación del aire"	t_20	"Área urbana"
t_10	"Transporte"	t_21	"Incendios forestales"
t_11	"Modelos GWR"	t_22	"Salud pública"

Figura 22. Contribuciones del factor al elemento para los tópicos, en las cinco primeras dimensiones (Dim-i) y para los planos Dim1-2, Dim 2-3, Dim 3-4 y Dim4-5. La línea discontinua roja indica la contribución promedio esperada



Código	País
au	Australia
at	Austria
be	Bélgica
br	Brasil
ca	Canadá
cl	Chile
cn	China
co	Colombia
cz	República Checa
dk	Dinamarca
fr	Francia
de	Alemania
gr	Grecia
hk	Hong Kong
in	India
id	Indonesia
ir	Irán
ie	Irlanda
it	Italia
jp	Japón
my	Malasia
mx	México
nl	Países Bajos
nz	Nueva Zelanda
ng	Nigeria
pl	Polonia
pt	Portugal
sa	Arabia Saudita
si	Eslovenia
za	Sudáfrica
kr	Corea del Sur
es	España
ch	Suiza
tw	Taiwan
tr	Turquía
gb	Reino Unido
us	Estados Unidos

Figura 23. Contribuciones del factor al elemento de los países en las cinco primeras dimensiones (Dim-i) y para los planos Dim1-2, Dim 2-3, Dim 3-4 y Dim 4-5. La línea discontinua roja indica la contribución promedio esperada

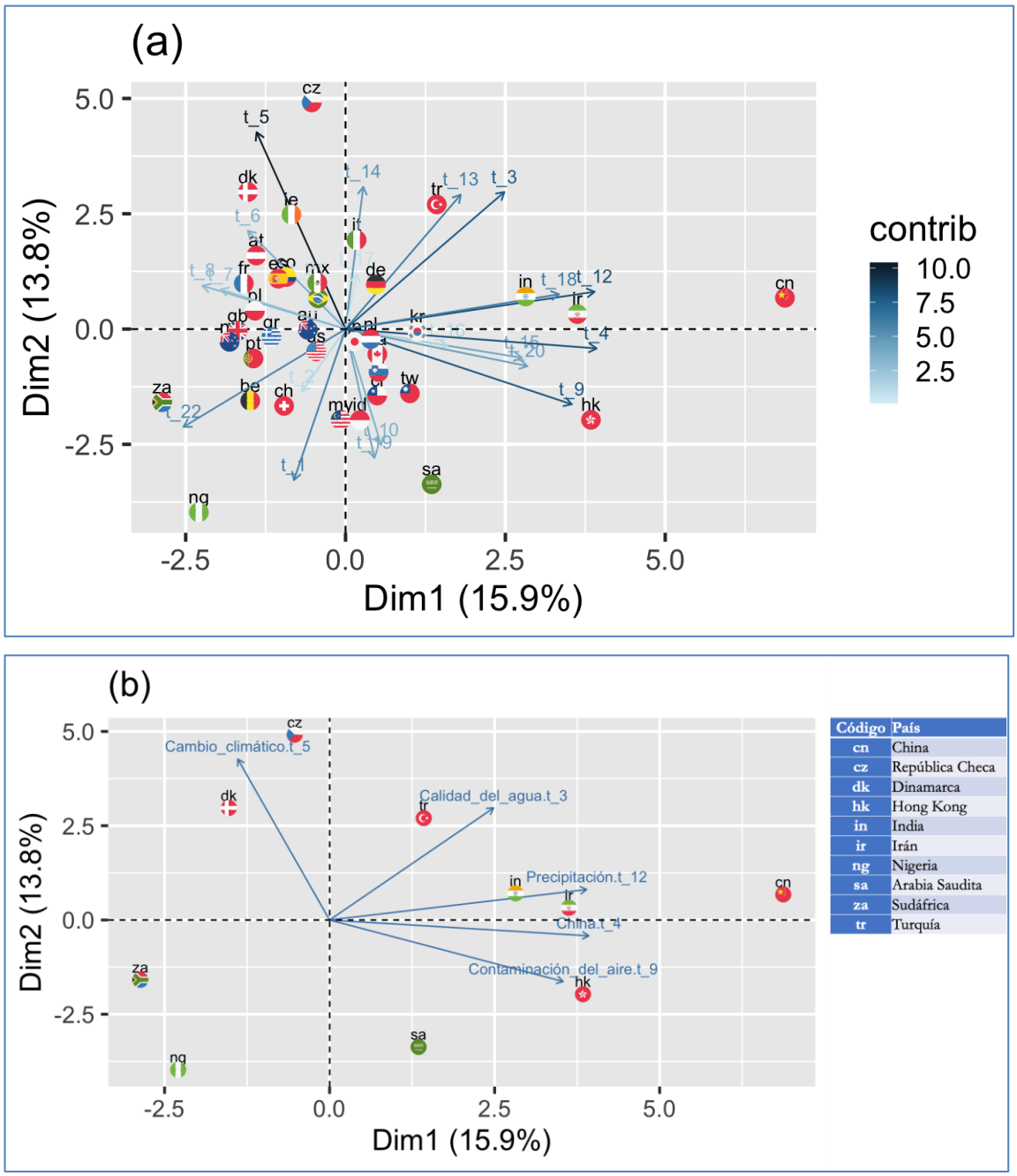


Figura 24. Representación factorial resultante del HJ-Biplot para el plano 1-2. (a) Todos los tópicos y países. (b) Tópicos con la mayor calidad de representación ($cos^2 \geq 0{,}5$) y los 10 países con la mayor contribución

La información más relevante que podríamos extraer de la información recogida por el plano 1-2 sería que en China, India, Irán, y Hong Kong fundamentalmente se utilizan estas técnicas para investigar temas sobre "Precipitación" (t_12), "China" (t_4) "Contaminación del aire" (t_9) y "Emisión de Carbono" (t_18), mientras que los países situados a la izquierda del eje 1 lo hacen en menor medida, siendo Sudáfrica y Nigeria los que menos investigan en ellos.

La República Checa, Dinamarca e Irlanda son los más preocupados por el "Cambio climático" (t_5) y "Uso del suelo" (t_14) que caracterizan el extremo superior del eje 2, siendo Arabia Saudita es el país al que menos investiga sobre estos temas, pero es, sin embargo, un país preocupado por temas como: "Factores de riesgo en enfermedades" (t_1) "Modelos GWR"(t_19) Transporte (t_10).

Los países situados en el tercer cuadrante, sobre todo Sudáfica, Nigeria, Bélgica y Suiza, investigan mediante el uso de estas técnicas en temas relacionados con la Salud pública (t_22) y "Factores de riesgo en enfermedades" (t_1).Turquía es el país que más utiliza estas técnicas para investigar sobre Carbono orgánico (t_13) y "Calidad del agua" (t_3) y "Uso del suelo" (t_14).

En la Figura 25a se representan los resultados de las proyecciones en el plano 2-3. Los tópicos con la mayor calidad de representación son "Cambio climático" (t_5) y "Uso del suelo" (t_14), mientras que los países con contribuciones más altas son Chile (cl), República Checa (cz), Dinamarca (dk), Hong Kong (hk), Irlanda (ie), Malasia (my) Nigeria(ni), Arabia Saudita (sa), Eslovenia (si) y Turquía (tr) (Figura 25b).

Si tenemos en cuenta el plano 2-3, la información más relevante es que el eje 1 está separando aquéllos países (situados a la derecha del mismo, Dinamarca, Turquía y República Checa) que más utilizan estas técnicas para el estudio del cambio climático y el uso del suelo, de aquéllos (situados a la izquierda del mismo: Nigeria, Hong Kong, Arabia Saudí, Malasia, Chile y Eslovenia) que menos los utilizan para analizar estos temas.

El eje 3 está caracterizado en su parte inferior por los tópicos t_11 y t_8 ("Modelos GWR" y "Heterogeneidad espacial" respectivamente). Los países que investigan esos temas estan situados en el semiplano inferior. La parte superior del eje 3 viene caraterizada por el tópico t_15 ("Clima"). Por tanto, los países que más utilizan estas técnicas para investigaciones sobre dicho tópico serían Chile y Eslovenia, mientras que los situados en el semiplano inferior presentan menos investigaciones que traten sobre ello.

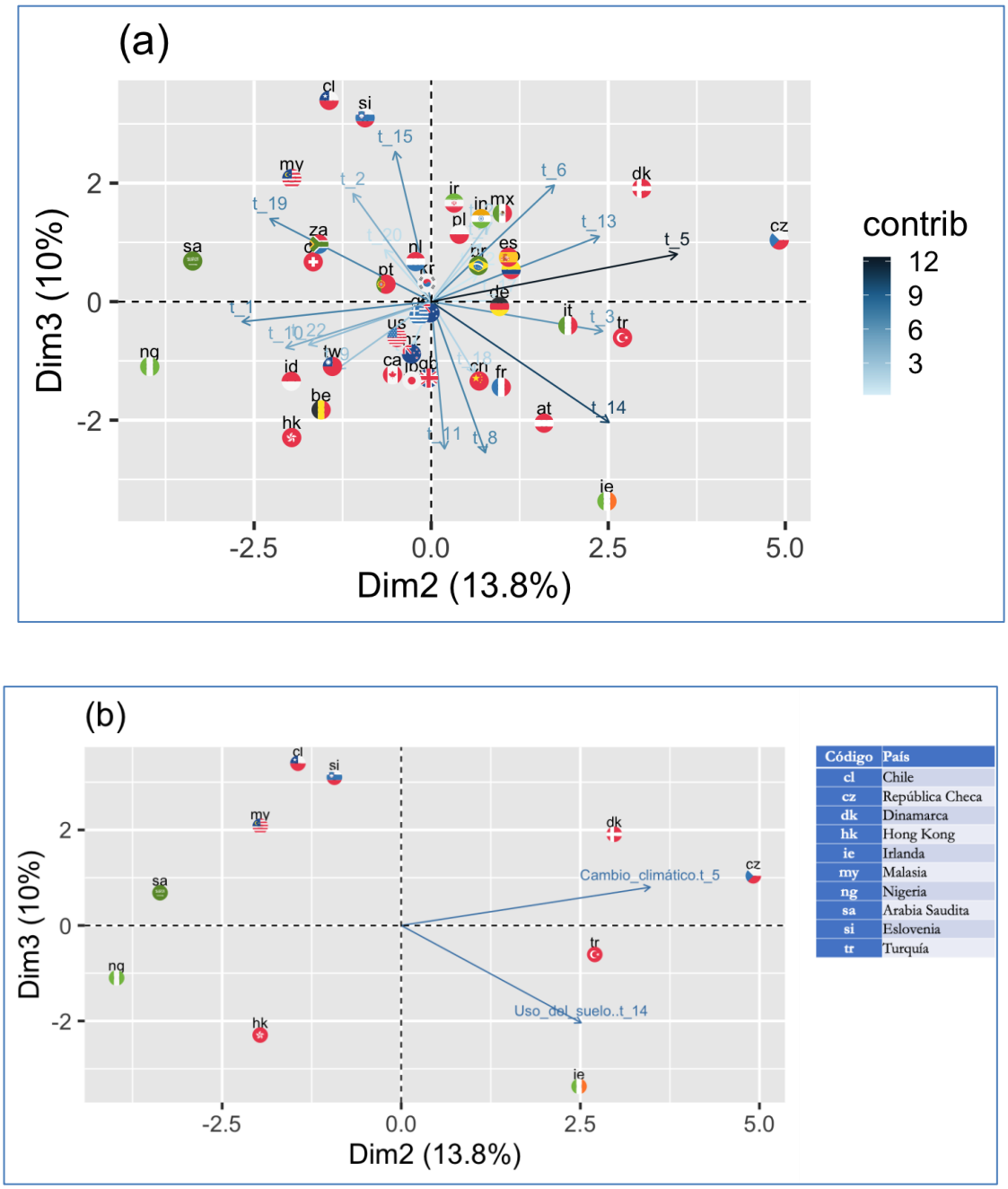


Figura 25. Representación factorial resultante del HJ-Biplot para el plano 2-3. (a) Todos los tópicos y países. (b) Tópicos con la mayor calidad de representación ($cos^2 \geq 0{,}5$) y los 10 países con la mayor contribución

En la Figura 26a se encuentran todos los tópicos y países representados en el plano 3-4, mientras que en la Figura 26b observamos que el tópico "Sensores remotos" es el que mejor representado que se encuentra en él. Además, los 10 países con contribuciones más altas son Austria (at), Eslovenia (si), Irlanda (ie), Chile (Cl), Sudáfrica (za), Dinamarca (dk), Bélgica (be), Malasia (my), Indonesia (id) y Hong Kong (hk).

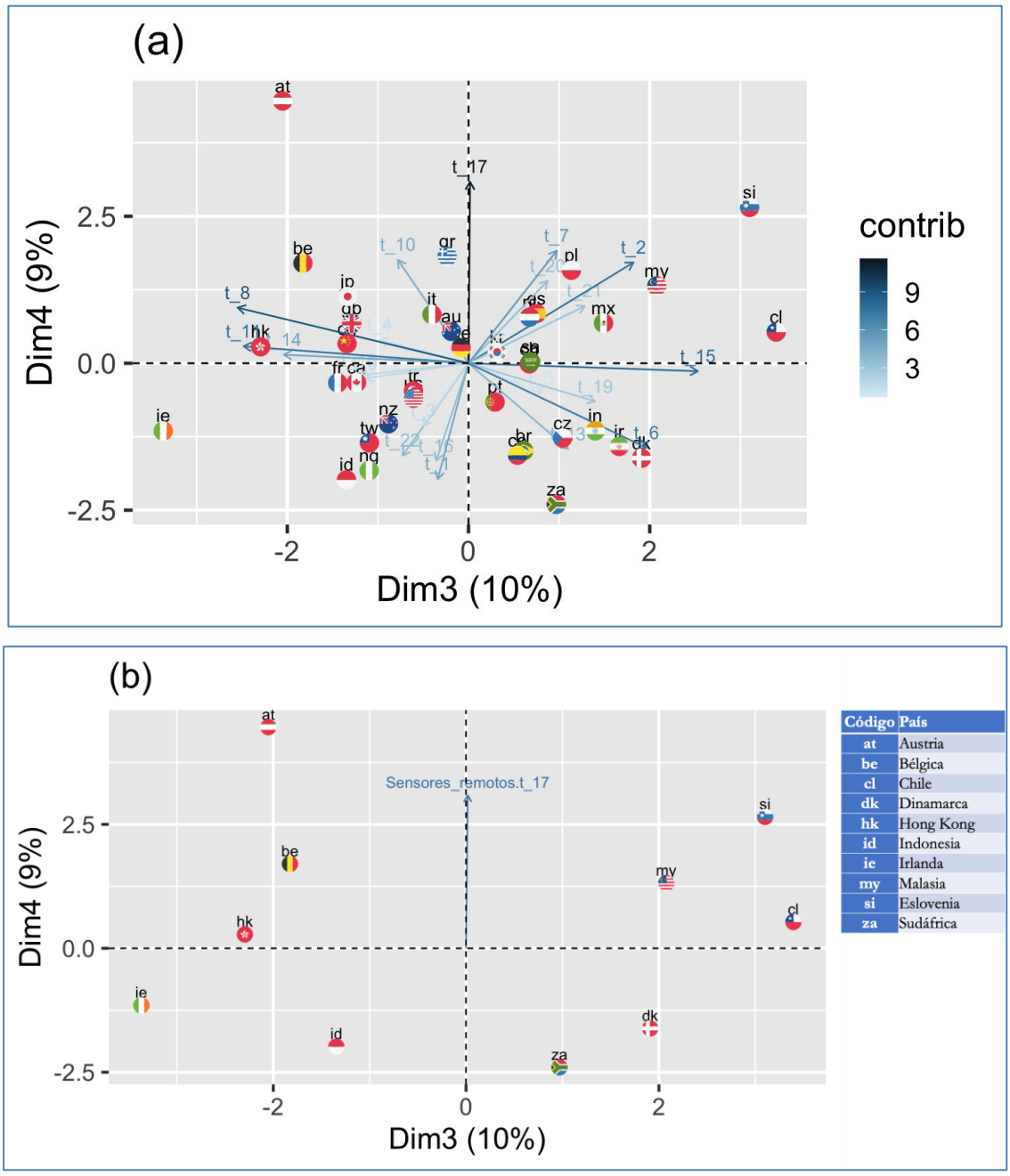


Figura 26. Representación factorial resultante del HJ-Biplot para el plano 3-4. (a) Todos los tópicos y países. (b) Tópicos con la mayor calidad de representación ($cos^2 \geq 0{,}45$) y los 10 países con la mayor contribución

En relación al tópico de "Sensores remotos" (t_17), los países situados en la parte superior del plano se caracterizan porque utilizan este tipo de técnicas para analizar esta temática (sobre todo Austria, pero también Bélgica, Eslovenia, y Malasia), mientras que los situados en el semiplano inferior, son los que menos utilizan estás técnicas para el estudio de sensores remotos (Sudáfrica, Indonesia, Dinamarca e Irlanda).

La calidad de representación de los tópicos en el plano 4-5 no alcanzan a superar el 0,3 y por esta razón no se ha representado.

Como conclusiones más importantes del análisis de los artículos de investigación donde se aplica la GWR, podemos establecer que mediante la aplicación LDAShiny:

- Fue posible analizar 3183 resúmenes de artículos sobre las invesigaciones realizadas utilizando los modelos GWR en el período de publicación comprendido entre 1996 a 2021.

- Se pudo identificar con éxito 22 temas de investigación que definen el estado actual de la investigación en esa área. Las probabilidades posteriores de documento-tópico y tópico-palabra se combinan además con el año de publicación, la revista y la afiliación del primer autor que, mediante una serie de análisis, nos permiten explorar el contexto de cada tópico, las tendencias de investigación asociadas y la similitud/disimilitud de tópicos entre revistas y paises. Los resultados de estos análisis nos brindan una estrategia viable para investigar el contenido central de la investigación en GWR que se espera beneficie a todas las partes interesadas de la comunidad científia (por ejemplo, estudiantes, investigadores, organizadores de conferencias, editores de revistas, agencias de financiación) en múltiples formas.

- Se analizó la distribución de los temas a nivel de revista. Los hallazgos sobre la similitud de los tópicos a nivel de revista pueden ayudar a los investigadores a identificar aquellas revistas objetivo para el envío de sus manuscritos. Los editores podrían utilizar la información de distribución de tópicos por revistas para identificar y priorizar temas que podrían mejorar el impacto, y además, podrían considerar ajustar o reevaluar el alcance y el enfoque de la revista.

- Se pudo analizar la distribución de tópicos por país. En general, la distribución no se ve necesariamente afectada por factores como la ubicación geográfica. La distribución de temas por país que se ha identificado puede ayudar a las agencias de financiación a comprender mejor las necesidades de investigación de sus regiones.

3.3. REVISIÓN DE LA APLICACIONES DE GWPCA

Al igual que en el caso de las aplicaciones de los modelos de GWR la búsqueda de artículos se realizó a través de las bases de datos de Scopus y la Web of Science. La búsqueda no tuvo restrincción de fechas, idioma o tipo de documento. La búsqueda se realizó el 12 de abril de 2022 y se encontraron 48 documentos en la base de Scopus y 11 en la Web of Scence. Estos últimos se encontraban en ambas bases por lo que el número total de artículos evaluados fueron 48. Los artículos procedian de 41 revistas, con 151 autores de los cuales los autores de documentos de un solo firmante fueron tres. La media de autores por documento es de 3,08. El promedio de citas por documento es de 17,94. En la Tabla 4 se muestra un resumen (en el Anexo 4 se encuentra la información de cada uno de ellos).

Tabla 4. Principales estadísticas sobre la colección de artículos sobre GWPCA

Descripción		Resultados
Información principal sobre los datos	Espacio de tiempo	2010:2022
	Fuentes (Revistas)	41
	Documentos	49
	Promedio de años desde la publicación	3,65
	Citas promedio por documentos	17,94
	Promedio de citas por año por documento	3,143
Tipos de documento	Artículos	46
	Documento de conferencia	2
Contenidos de documentos	Keywords Plus (ID)	451
	Palabras clave del autor	173
Autores	Autores	151
	Apariciones del autor	196
	Autores de documentos de un solo autor	3
	Autores de documentos de varios autores	148
Colaboración de autores	Documentos de un solo autor	3
	Documentos por autor	0,325
	Autores por Documento	3,08
	Coautores por Documentos	4
	Índice de colaboración	3,22

Como resultado de la búsqueda se encontraron documentos entre los años 2010 y 2022. El número de publicaciones aumentó a una tasa anual promedio de 10,5%. En la Figura 27 se puede observar la variación anual del número de artículos publicados.

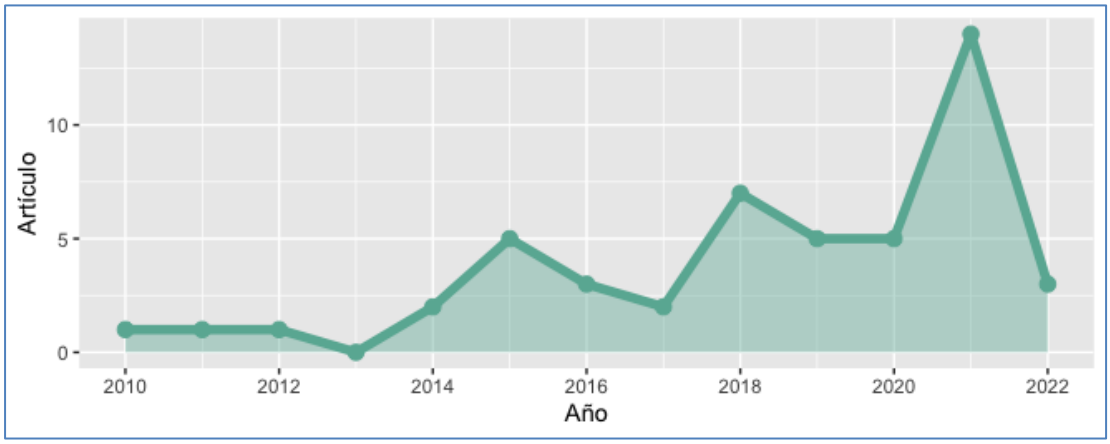


Figura 27. Número de publicaciones sobre GWPCA entre 2010 y 2022

En lo que respecta a las revistas, observamos que según el número de publicaciones, *Sustainability* (Switzerland), con tres públicaciones, es la revista con más artículos sobre GWPCA, seguído de seis revistas con dos artículos publicados y el resto de revistas (34) con sólo una publicación cada una (Figura 28).

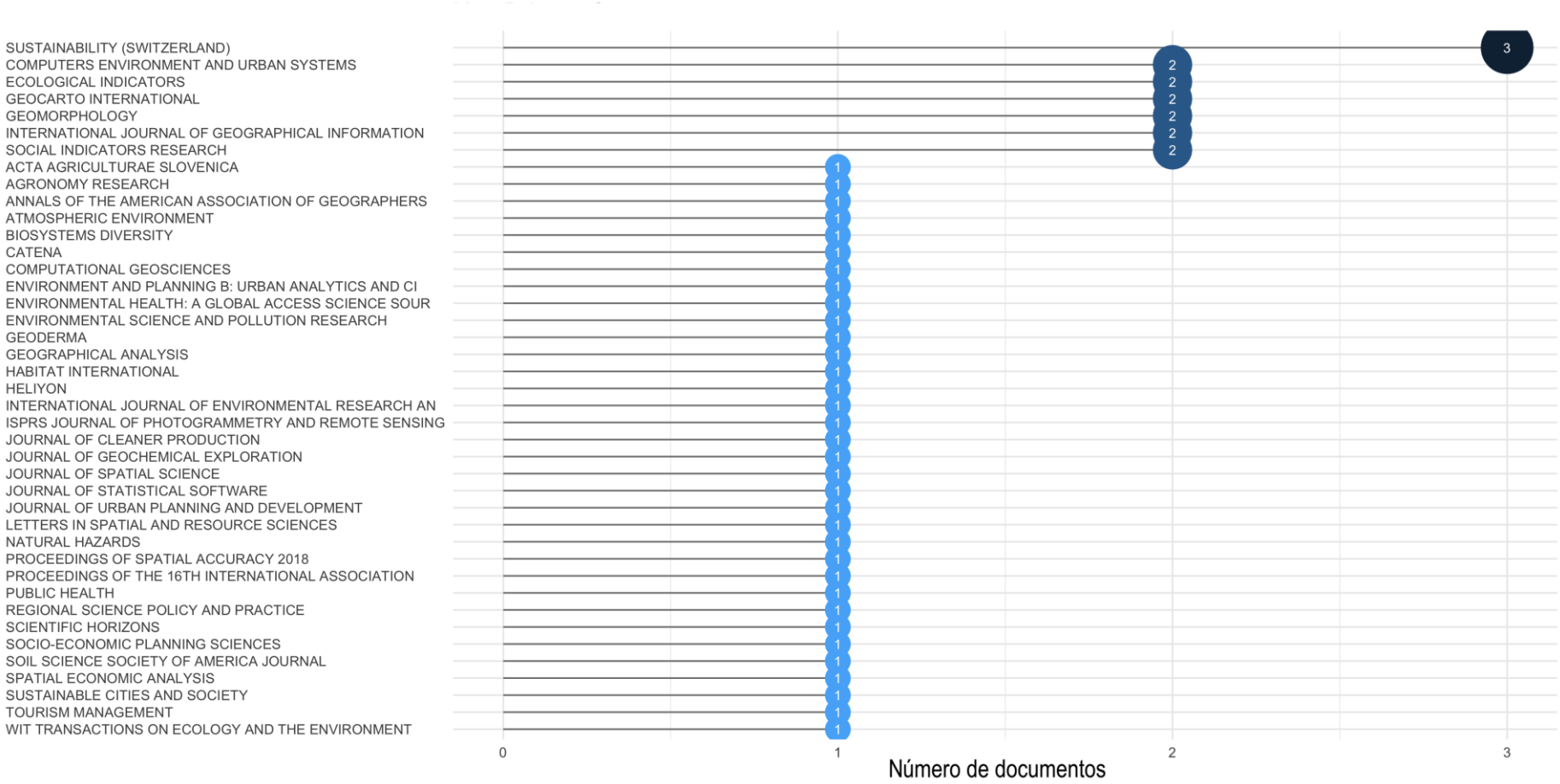


Figura 28. Número de artículos por revista sobre GWPCA entre 2010 y 2022

Según la afiliación del primer autor, los artículos analizados provienen de China, Reino Unido, India, Italia, Ucrania, Estados Unidos de América, Irán, España, Ghana, Portugal, Irlanda, Austria, Francia, Colombia, Panamá y México (Figura 29).

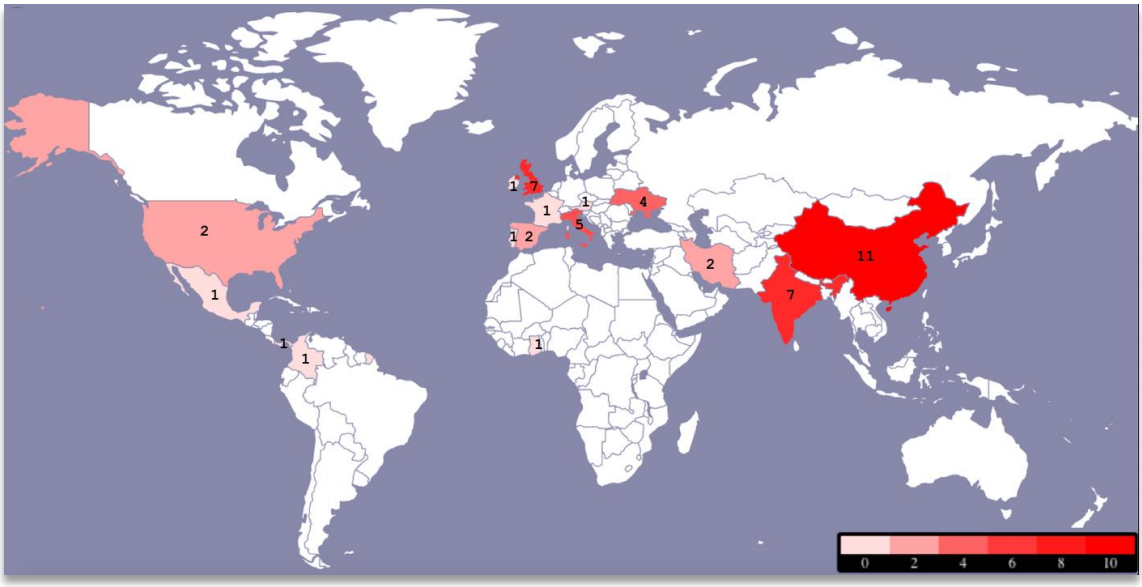


Figura 29. Origen geográfico de los 49 artículos sobre GWPCA entre 2010 y 2022

La red de coautores (Figura 30), ilustra las conexiones intelectuales entre investigadores que utilizan un GWPCA. El país con mayores conexiones es el Reino Unido.

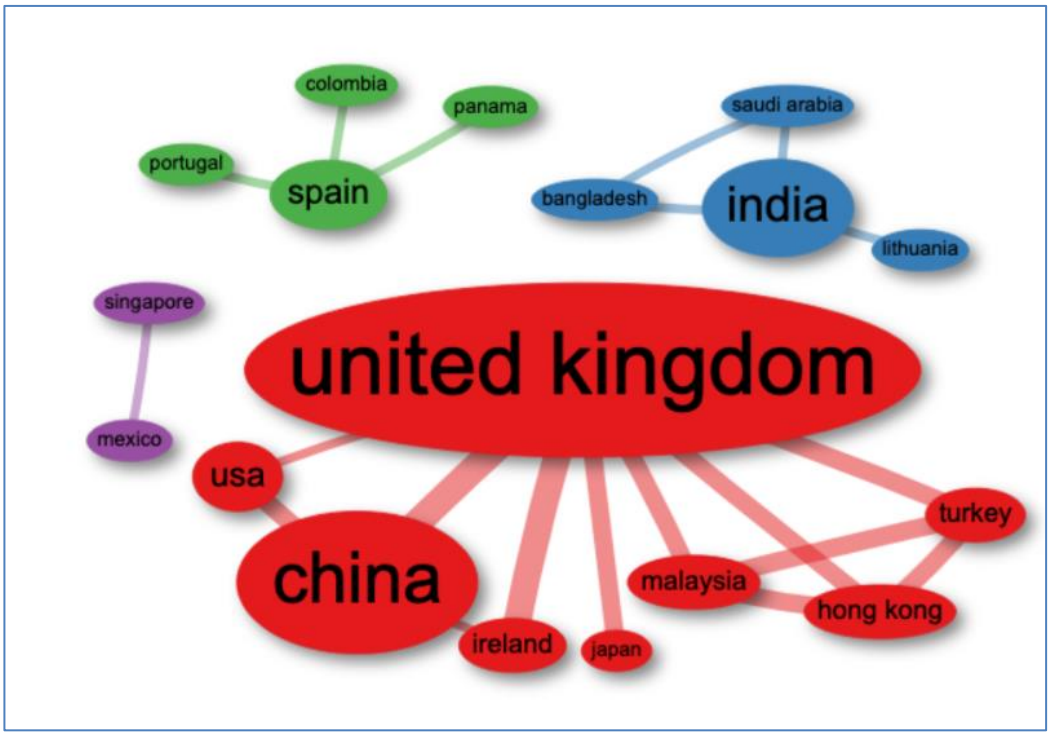


Figura 30. Red de coautorías basada en los países de los artículos sobre GWPCA

El GWPCA se ha aplicado en varios contextos, como por ejemplo, para analizar características multivariantes de la población (Lloyd, 2010), estructuras sociales (Harris et al., 2011b), características del suelo (Kumar et al., 2012b; Roca-Pardiñas et al., 2016; Fernández et al., 2018), estimación de requerimiento de agua (Wang et al., 2013), datos de composicón química del agua (Harris et al., 2015), patrones de desarrollo económico regional (Li et al., 2016), análisis de contaminación (Shrestha & Luo, 2017), comportamiento de turistas (Losada et al., 2019), entre otros.

3.4. REVISIÓN DE LA APLICACIONES DE GWDA

Al realizar la revisión en las bases de datos de la Web of Science y Scopus observamos que después de la primera publicación en el año 2007 sólo se han publicado dos artículos y una nota relacionados con la técnica GWDA (Tabla 5).

Tabla 5. Publicaciones sobre GWDA, entre los años 2007 y 2022

Autores	Título	Año	Revista	DOI	Tipo de documento
Brunsdon C., Fotheringham S., Charlton M.	Geographically weighted discriminant analysis	2007	Geographical Analysis	10.1111/j.1538-4632.2007.00709.x	Artículo
Johnston R., Pattie C.	Comment: Geographically weighted discriminant analysis and the 2005 british general election	2009	Geographical Analysis	10.1111/j.1538-4632.2009.00756.x	Nota
Foley P., Demšar U.	Using geovisual analytics to compare the performance of geographically weighted discriminant analysis versus its global counterpart, linear discriminant analysis	2013	International Journal of Geographical Information Science	10.1080/13658816.2012.722638	Artículo
Nicholas D.E., Delamater P.L., Waters N.M., Jacobsen K.H.	Geographically weighted discriminant analysis of environmental conditions associated with Rift Valley fever outbreaks in South Africa	2016	Spatial and Spatio-temporal Epidemiology	10.1016/j.sste.2016.04.005	Artículo

La primera nota es un comentario escrito por Johnston & Pattie (2009) que identifica un error relacionado con el análisis de patrones de votación en las elecciones generales de 2005 en el Reino Unido, que fueron los datos utilizados por Brunsdon et al. (2007) para

ilustrar el GWDA. La nota sugiere que la aplicación de GWDA no ofrece ninguna mejora con respecto al análisis discriminante lineal estándar para ese problema en particular. Además, cuestionan las razones para trabajar con componentes principales en lugar de datos sin procesar. Sin embargo, no niegan el uso potencial de GWDA en otros contextos, en donde las variables discriminantes varíen en el espacio.

Foley & Demšar (2013), en su publicación, sugieren y desarrollan técnicas analíticas, que combinan herramientas visuales interactivas y métodos computacionales para explorar los resultados del GWDA y del LDA, que permiten examinar el rendimiento de los métodos de clasificación globales y locales en términos de calidad e incertidumbre de clasificación y no estacionariedad espacial. Los autores aplicaron su enfoque en un estudio de caso utilizando un conjunto de datos que vincula los resultados de las elecciones de Estados Unidos de América de 2004 con una selección de cinco variables socioeconómicas.

El último artículo registrado con la técnica GWDA es una aplicación a la investigación en epidemiología de enfermedades infecciosas. Data del año 2016 y fue publicado por Nicholas et al. (2016). En dicho estudio se comparó el rendimiento de la clasificación obtenida por un GWDA y la del DA clásico para distinguir entre los lugares donde las reses están en riesgo o no de adquirir brotes de fiebre en el Valle del Rift en Sudáfrica. Según los autores, el GWDA produjo mejores resultados que el DA clásico para todas las combinaciones de ancho de banda y Kernel. El mejor modelo de GWDA fue obtenido con ancho de banda fijo y núcleo exponencial y gaussiano, habiendo clasificado correctamente el 96,6% de los datos originales (frente al 84,5% obtenido con el DA clásico). Además, con el GWDA, los falsos positivos disminuyeron de un 10,9% a un 3,7%, y los falsos negativos disminuyeron del 19,9% al 3,2% (Nicholas et al., 2016).

3.5. REVISIÓN DE LAS APLICACIONES DE CLUSTER PONDERADO GEOGRÁFICAMENTE.

Se encontrarón ocho documentos en la Web of Science y 14 en la base de datos de Scopus, pero al eliminar los duplicados quedaron sólo 14 documentos. Estos documentos procedian de 14 fuentes, con 26 autores de los cuales, los autores de documentos de un solo autor, son únicamente tres, adémas, los documentos encontrados estaban en el lapso de

tiempo comprendido entre 2012 y 2022 . La media de autores por documento es de 1,62, mientras que el promedio de citas por documento es de 10,75 (Tabla 6).

Tabla 6. Principales estadísticas de la colección de artículos sobre Cluster Ponderado Geográficamente

Descripción		Resultados
Información principal sobre los datos	Espacio de tiempo	2012:2022
	Fuentes (Revistas)	14
	Documentos	14
	Promedio de años desde la publicación	4,5
	Citas promedio por documentos	10,75
	Promedio de citas por año por documento	1,64
Tipos de documento	Artículo	8
	Documento de conferencia	5
	Documento de datos	1
Contenidos de documentos	Keywords Plus (ID)	88
	Palabras clave del autor	33
Autores	Autores	26
	Apariciones del autor	43
	Autores de documentos de un solo autor	3
	Autores de documentos de varios autores	23
Colaboración de autores	Documentos de un solo autor	4
	Documentos por autor	0,615
	Autores por Documento	1,62
	Coautores por Documentos	2,69
	Índice de colaboración	1,92

La mayoria de artículos publicados hacen referencia a mejoras de algoritmos (ocho documentos). Por otro lado, se encontró un artículo que presenta el conjunto de datos sobre vulnerabilidad social en Indonesia publicado en la revista *Data in Brief*. Dentro de las aplicaciones encontramos dos artículo relacionados con la COVID-19 (a saber, "Mapping the geodemographics of racial, economic, health, and COVID-19 deaths inequalities in the conterminous US" publicado en *Applied Geography* y "Micro, small, and medium enterprises' business vulnerability cluster in Indonesia: An analysis using optimized fuzzy geodemographic clustering" publicado en *Sustainability*) (Tabla 7).

Tabla 7. Publicaciones sobre cluster ponderado geográficamente, entre los años 2012 y 2022

Autores	Título	Año	Revista	DOI	Tipo de documento
Son L.H., Cuong B.C., Lanzi P.L., Thong N.T.	A novel intuitionistic fuzzy clustering method for geo-demographic analysis	2012	Expert Systems with Applications	10.1016/j.eswa.2012.02.167	Artículo
Wijayanto A.W., Purwarianti A.	Improvement of fuzzy geographically weighted clustering using particle swarm optimization	2014	2014 International Conference on Information Technology Systems and Innovation, ICITSI 2014 - Proceedings	10.1109/ICITSI.2014.7048229	Conference Paper
Wijayanto A.W., Purwarianti A.	Improvement design of fuzzy geo-demographic clustering using Artificial Bee Colony optimization	2014	2014 International Conference on Cyber and IT Service Management, CITSM 2014	10.1109/CITSM.2014.7042178	Conference Paper
Son L.H.	A novel kernel fuzzy clustering algorithm for Geo-Demographic Analysis	2015	Information Sciences	10.1016/j.ins.2015.04.050	Artículo
Nurmala N., Purwarianti A.	Improvement of Fuzzy Geographically Weighted Clustering-Ant Colony Optimization using context-based clustering	2016	2015 International Conference on Information Technology Systems and Innovation, ICITSI 2015 - Proceedings	10.1109/ICITSI.2015.7437726	Conference Paper
Wijayanto A.W., Purwarianti A., Son L.H.	Fuzzy geographically weighted clustering using artificial bee colony: An efficient geo-demographic analysis algorithm and applications to the analysis of crime behavior in population	2016	Applied Intelligence	10.1007/s10489-015-0705-7	Artículo
Nurmala N., Purwarianti A.	Improvement of fuzzy geographically weighted clustering-ant colony optimization performance using context-based clustering and CUDA parallel programming	2017	Journal of ICT Research and Applications	10.5614/itbj.ict.res.appl.2017.11.1.2	Artículo
Wijayanto A.W., Mariyah S., Purwarianti A.	Enhancing clustering quality of fuzzy geographically weighted clustering using Ant Colony optimization	2018	Proceedings of 2017 International Conference on	10.1109/ICODSE.2017.8285858	Conference Paper

Autores	Título	Año	Revista	DOI	Tipo de documento
			Data and Software Engineering, ICoDSE 2017		
Nasution B.I., Kurniawan R., Siagian T.H., Fudholi A.	Revisiting social vulnerability analysis in Indonesia: An optimized spatial fuzzy clustering approach	2020	International Journal of Disaster Risk Reduction	10.1016/j.ijdrr.2020.101801	Artículo
Grekousis G., Wang R., Liu Y.	Mapping the geodemographics of racial, economic, health, and COVID-19 deaths inequalities in the conterminous US	2021	Applied Geography	10.1016/j.apgeog.2021.102558	Artículo
Caraka R.E., Kurniawan R., Nasution B.I., Jamilatuzzahro J., Gio P.U., Basyuni M., Pardamean B.	Micro, small, and medium enterprises' business vulnerability cluster in indonesia: An analysis using optimized fuzzy geodemographic clustering	2021	Sustainability (Switzerland)	10.3390/su13147807	Artículo
Abdussamad S.N., Astutik S., Efendi A.	Fuzziness Evaluation on Hybrid Context Based Clustering Methods with Fuzzy Geographically Weighted Clustering-Particle Swarm Optimization Algorithm	2021	Journal of Physics: Conference Series	10.1088/1742-6596/1811/1/012087	Conference Paper
Grekousis G.	Local fuzzy geographically weighted clustering: a new method for geodemographic segmentation	2021	International Journal of Geographical Information Science	10.1080/13658816.2020.1808221	Artículo
Kurniawan R., Nasution B.I., Agustina N., Yuniarto B.	Revisiting social vulnerability analysis in Indonesia data	2022	Data in Brief	10.1016/j.dib.2021.107743	Documento de datos

CAPÍTULO IV

4. SOFTWARE PARA LOS MÉTODOS PONDERADOS GEOGRÁFICAMENTE

4.1. SOFTWARE PARA LA GWR

Como se ha comentado en los apartados anteriores son numerosas las extensiones que se han desarrollado a partir del modelo básico de GWR tras su introducción por Brunsdon et al. (1996). Sin embargo, para que esos desarrollos se utilicen por los investigadores en situaciones prácticas se deben implementar herramientas de software confiables. No obstante, el software disponible siempre va por detrás del desarrollo de nuevos métodos o extensiones. Por otras palabras, siempre habrá un retraso entre el desarrollo de nuevas técnicas y la provisión de software que sea apto para el investigador en el sentido de que pueda manejar grandes conjuntos de datos, sea fácil de usar y tenga una buena documentación (Fotheringham et al., 2002).

La implementación de la GWR en el software, como la mayoría de los métodos estadísticos, presenta en la actualidad dos tendencias en su desarrollo. Una es la implementación de técnicas de GWR más avanzadas en el software que ya está desarrollado inicialmente y otra es la implementación de la GWR en paquetes informáticos más generales y flexibles.

El software más antiguo conocido por el público que permite realizar una GWR se denomina GWR, cuya ultima versión es la 4.0 (Nakaya et al., 2009) y que lógicamente es una actualización de sus versiones predecesoras 3.0 (Charlton et al., 2007) y GWR 2.0 introducida por Fotheringham et al. (2002).

El software GWR 4.0 se ejecuta en entornos Windows Vista, Windows 7, 8 y 10 con .NET Framework 4. Permite especificar de la función del núcleo (tiene cuatro opciones Kernel disponibles, Gausiana y Bicuadrada en sus versiones fijas y adaptativas), la selección

del ancho de banda y permite elegir entre los criterios AICc y BIC para la selección del modelo. Tiene implementado el modelo de GWR mixto, donde los usuarios pueden especificar cada coeficiente como geográficamente fijo o variando, ya sea manualmente o mediante una rutina de selección automática de variables basada en comparaciones de modelos recursivos. También permite ajustar modelos GWR semiparamétricos, que le permiten combinar términos globalmente fijos y términos explicativos variables de manera simultánea. Además, se puede aplicar a los modelos GWGLM que incluyen regresiones de Gauss, Poisson y Logística.

Las salidas del software GWR 4.0 incluyen los resultados de la estimación de los parámetros junto con elementos de diagnósticos esenciales, incluidos los residuos estandarizados, las R^2 locales, las predicciones en puntos de no regresión, etc. Las estimaciones de los parámetros se pueden exportar posteriormente a una hoja de cálculo o a las plataformas de Sistemas de Información Geográficas (GIS) para su visualización.

El modelo GWR también está disponible en softwares GIS convencionales como ArcGIS en su Spatial Analyst Toolbox (Environmental Systems Research Institute (ESRI), 2018) y en software de análisis espacial, como por ejemplo, SAM (Análisis Espacial en Macroecología, (Rangel et al., 2010)). Asimismo, el modelo también está disponible en otros softwares como STATA (Pearce, 1999), MATLAB en su Econometrics toolbox (LeSage, 2001) y en SAS en una macro desarrollada por Chen & Yang (2012).

En cuanto al entorno R (Core Team, 2021) existe un conjunto de paquetes o librerías con funcionalidades específicas aplicadas a la GWR. Al hacer la revisión se encontraron 10 paquetes publicados en el CRAN que abordaban la temática GWR o sus extensiones. La Tabla 8 muestra una comparación de las funciones que se encuentran en dichos paquetes y se establecen las similitudes y divergencias de los mismos en relación a seis aspectos (especificaciones de la matriz de ponderación, estadísticas de resumen de GW de funciones, otras funciones de regresión de GW, funciones para abordar la colinealidad, función de regresión de GW básica y funciones de regresión de GW generalizadas).

Li et al. (2019) proponen una implementación de código abierto altamente escalable basada en Python denominada por FastGWR. Este código permite resolver la limitante de la cantidad o número máximo de datos que se pueden manejar a través de software GWR de

código abierto, que es de aproximadamente 15.000 observaciones, lo cual es un problema en esta era del Big Data. Los autores, con el algoritmo propuesto, permiten aplicaciones con conjuntos de datos muy grandes, del orden de millones de observaciones, optimizando el uso de la memoria junto con la paralelización para aumentar significativamente el rendimiento.

Oshan et al. (2019) presentan una implementación basada en Python, denominada mgwr, que se enfoca explícitamente en el análisis multiescala de la heterogeneidad espacial. Además, según los autores, la implementación proporciona una funcionalidad novedosa para análisis inferencial y exploratorio de procesos espaciales locales, nuevos diagnósticos exclusivos de escala múltiple, modelos locales y mejoras en la eficiencia de las rutinas de estimación.

Tabla 8. Comparación de los diferentes paquetes de R utilizados en GWR (modificado de Gollini et al., 2015)

		Gwmodel [1]	**spgwr [2]**	**gwrr [3]**	**McSpatial [4]**	**gwer [5]**	**gwfa [6]**	**GWLelast [7]**	**lctools [8]**	**mgwrsar [9]**	**scgwr [10]**
Especificaciones de la matriz de ponderación	Función Kernel	Bi-square Box-car Exponential Gaussian Tri-cube	Bi-square Tri-cube Gaussian	Exponential Gaussian	Bi-square Epanechnikov Gaussian Rectangular Tri-cube Tri-weight Triangular	Bi-square Gaussian	Bi-square	Bi-square Gaussian	Bi-square Binary	Bi-square Binary Gaussian	Exponencial Gaussian
	¿Ancho de banda adaptativo?	Si	Si	No	No	SI	No	Si	Si	Si	Si
	¿Ancho de banda fijo?	Si	Si	Si	Si	SI	No	Si	Si	Si	Si
	Métricas de distancia espacial	Euclidea Gran circulo Minkowski	Euclidea Gran circulo	Euclidea	Euclidea Gran circulo Mahalanobis	Euclidea	Euclidea	Euclidea	Euclidea	Euclidea	Euclidea
	¿Funciones para el cálculo matricial de ponderaciones?	Si	Si	No	Si	Si	No	Si	Si	Si	Si
Funciones estadísticas de resumen GW	¿Estadísticos básicos?	Si	Si	No	No	No	No	No	No	No	No
	¿Estadísticos robustos?	Si	No	No	No	No	No	No	No	No	No
	¿Test de Monte-Carlo?	Si	No	No	No	No	No	No	Si	No	No
Otras funciones de regresión GW	¿Regresión GW robusta?	Si	No	No	No	No	No	No	No	No	No
	¿Regresión GW Heterocedástica?	Si	No	No	No	No	No	No	No	No	No
	¿Regresión GW Mixed (semiparamétrica)?	Si	No	No	Si	No	No	No	No	Si	Si

		Gwmodel [1]	spgwr [2]	gwrr [3]	McSpatial [4]	gwer [5]	gwfa [6]	GWLelast [7]	lctools [8]	mgwrsar [9]	scgwr [10]
	¿Regresión GW Elastic net?	No	No	No	No	No	No	Si	No	No	No
	¿Regresión GW Quantile?	No	No	No	Si	No	No	No	No	No	No
Funciones para abordar los problemas de colinealidad en la regresión GW	¿Correlación local?	Si	Si	No	No	No	No	No	Si	No	No
	¿Local VIFs?	Si	No	No	No	No	No	No	No	No	Si
	¿Local VDPs?	Si	No	Si	No	No	No	No	No	No	No
	¿Números de condición local?	Si	No	Si	No	No	No	No	Si	No	No
	¿Regresión GW con un término de ridge global (especificado por el usuario)?	Si	No	Si	No	Si	No	No	No	No	No
	¿Regresión GW con un término de ridge global (estimado)?	No	No	Si	No	Si	No	No	No	No	No
	¿Regresión GW con un término de ridge local ?	Si	No	No	No	Si	No	No	No	No	No
	¿GW lasso?	No	No	Si	No	Si	No	No	No	No	No
Basic GW regression functions	¿Regresión gw básica?	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si
	¿Ancho de banda CV?	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si
	¿Ancho de banda CV generalizado?	No	No	No	Si	No	No	No	No	No	No
	¿Ancho de banda AICc?	Si	Si	No	No	Si	No	Si	Si	Si	Si
	¿Herramientas de selección del modelo?	Si	No	No	No	No	No	No	No	No	No
	¿Test de Regresión local vs global?	Si	Si	No	No	No	No	No	No	No	No
	¿Moran's I test?	No	Si	No	No	No	No	No	No	No	No
	¿Opciones de computación en paralelo?	No	Si	No	No	Si	No	Si	No	Si	Si
Funciones de Regresión GW generalizadas	¿Regresión GW Generalizada?	Si	Si	No	Si	No	No	Si	Si	No	No
	Familia/link:	Binomial,Poisson Poisson	Como para glm	No	Logit, Multinomial, Probit probit			Logit	Poisson		

[1] Gollini I, Lu B, Charlton M, Brunsdon C, Harris P. GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models. J. Stat. Softw. 2015; 63(17): 1–50. URL https://doi.org/10.18637/jss.v063.i17

[2] Roger Bivand and Danlin. spgwr: Geographically Weighted Regression. R package version 0.6-35, 2022, URL, https://CRAN.R-project.org/package=spgwr.

[3] Wheeler D. gwrr: Fits Geographically Weighted Regression Models with Diagnostic Tools. R package version 0.2-2, 2022; URL, https://CRAN.R-project.org/package=gwrr

[4] McMillen . McSpatial: Nonparametric Spatial Data Analysis. R package version 2.0, 2013, URL, http://CRAN.R-project.org/package=McSpatial.

[5] Cysneiros, F. J. A., Paula, G. A., and Galea, M. gwer: Geographically Weighted Elliptical Regression. R package version 3.0, 2021, URL, https://CRAN.R-project.org/package=gwer

[6] Semecurbe F, Roux SG, and Tannier C. gwfa: Geographically Weighted Fractal Analysis. R package version 0.04, 2016, URL, https://CRAN.R-project.org/package=gwer

[7] Yoneoka D, Saito E. GWLelast: Geographically Weighted Logistic Elastic Net Regression. R package version 1.2.2, 2019, URL, https://CRAN.R-project.org/package=GWLelast

[8] Kalogirou S. lctools: Local Correlation, Spatial Inequalities, Geographically Weighted Regression and Other Tools. R package version 0.2.8, 2020, URL, https://CRAN.R-project.org/package=lctools

[9] Geniaux G and Martinetti D. mgwrsar: GWR and MGWR with Spatial Autocorrelation. R package version 1.0, 2022, URL, https://CRAN.R-project.org/package=mgwrsar

[10] Murakami D, Tsutsumida N, Yoshida T, Nakaya T and Lu B. scgwr: Scalable Geographically Weighted Regression R package version 0.1.2-21, 2021, URL, https://CRAN.R-project.org/package=scgwr

4.2. SOFTWARE PARA EL GWPCA

Actualmente sólo encontramos dos paquetes en R para realizar un GWPCA, el primero es el paquete GWmodel (Gollini et al., 2015) que proporciona funciones para realizar un GWPCA básico y robusto. Los dos ajustes del GWPCA se encuentran utilizando la función gwpca y sólo hay que cambiar uno de los argumentos de la función para seleccionar uno u otro ajuste. El segundo paquete es el GWnnegPCA (Tsutsumida, 2020), que consiste en la fusión de un GWPCA y un PCA sparse no negativo.

4.3. SOFTWARE PARA EL GWDA

A la fecha, sólo se encuentra implementado el Análisis Discriminante Ponderado Geográficamente en el paquete de R denominado por GWmodel (Gollini et al, 2015). Este paquete ofrece una colección de funciones para el análisis de técnicas geográficamente ponderadas dentro de las que se incluye la función gwda que implementa el análisis GWDA y la bw.gwda que es una función para la selección automática del ancho de banda para el análisis GWDA utilizando el enfoque de la validación cruzada.

4.4. SOFTWARE PARA ANÁLISIS CLUSTER PONDERADO GEOGRÁFICAMENTE

En septiembre de 2016 se publicó el paquete de R spatialClust (Pramana & Pamungkas, 2016) que era el único paquete que realizaba un análisis de agrupación espacial utilizando agrupación ponderada geográficamente difusa. En la actualidad este paquete ya no se encuentra disponible en la CRAN. Sin embargo, una versión de desarrollo se puede encontrar en el repositorio de Github https://github.com/imamhabib/spatialClust.

CAPÍTULO V

5. EL PAQUETE *GEOWEIGHTEDMODEL*

Los modelos ponderados geográficamente son técnicas que forman parte de la estadística espacial. Son útiles para una amplia gama de áreas, incluidas las ciencias sociales, ambientales y de la salud (entre otras), donde la información georreferenciada se recopila regularmente.

A la fecha, aunque existen varios paquetes para modelos GW en el entorno R, como hemos visto en el capítulo anterior, se requieren algunas habilidades computacionales que no todos los investigadores poseen.

En el transcurso de esta tesis se ha elaborado el paquete GeoWeightedModel que presentamos en este capítulo. La aplicación tiene como objetivo principal el facilitar, a cualquier investigador que no posea un amplio conocimiento de programación en el análisis de datos de métodos ponderados geográficamente, una interfaz simple e intuitiva en donde los análisis se puedan realizar de forma interactiva ("apuntar" y "hacer clic") en un navegador web.

Los análisis que podemos realizar con el paquete GeoWeightedModel son: análisis de autocorrelación espacial, las estadísticas de resumen locales ponderadas geográficamente, GWR y sus extenciones, GWPCA (básico y robusto) y GWDA.

Se pretende que la aplicación o paquete propuesto sea accesible para la mayor cantidad de usuarios posible, fácil de distribuir e instalar y tener un entorno gráfico funcional y sencillo de usar. El entorno del programa R cumple con los requisitos mencionados. Efectivamente, a través de la infraestructura de Comprehensive R Archive Network (CRAN), la aplicación se puede distribuir y puede ser accesible para una gran cantidad de usuarios. Como valor agregado, R es un entorno multiplataforma, es decir, el programa construido puede ejecutarse en los sistemas operativos utilizados por la mayoría de las computadoras personales como Windows, Mac o Linux. Otra característica es que el paquete es fácil de instalar, ya que si utilizamos elementos externos al software de R que necesitan ser instalados o compilados

previamente para que el módulo funcione, es probable que el usuario desista de utilizarlo. En nuestro caso la instalación no requiere ningún elemento externo.

GeoWeightedModel fue creado en Shiny el cual es un paquete de extensión de R creado por Joe Cheng de Rstudio (Allaire, 2012) que facilita la creación de aplicaciones web interactivas. Esta es la razón por la que Shiny es ideal para hacer prototipos rápidamente, pero también es una buena opción para construir aplicaciones listas para producción .

5.1. ARQUITECTURA Y FUNCIONALIDADES DEL PAQUETE

La estructura de cualquier aplicación desarrollada en Shiny es relativamente sencilla. Este se compone de dos partes que pueden localizarse en dos archivos separados, en concreto una secuencia de comandos de interfaz de usuario (ui.R) y una secuencia de comandos de servidor (server.R). Alternativamente, también se puede ubicar en un solo archivo donde se encuentran integrados los dos anteriormente mencionados (app.R). Estos archivos se pueden escribir en R "puro", pero también se puede incorporar código JavaScript, CSS o HTML. En el Anexo 5 se encuentra el código fuente de la aplicación GeoWeightedModel que se ha desarrollado.

La interfaz de usuario (ui.R) controla el diseño y la apariencia de la aplicación Shiny y proporciona interactividad a la aplicación tomando la entrada del usuario y mostrando dinámicamente la salida generada en la pantalla. Generalmente consiste en un esquema de diseño, con controles de usuario o widgets (como controles deslizantes, cuadros de selección, cuadros de radio, etc.). En la Tabla 9 encontramos una lista de los widgets estándar de Shiny utilizados en esta aplicación.

La función del servidor (server.R) contiene las instrucciones para la reactividad necesaria para la aplicación Shiny. Es decir, contiene la serie de pasos o instrucciones para convertir la entrada dada por el usuario en la salida que se desea mostrar.

La aplicación se sirve al cliente (usuario de la aplicación) a través de un host (Protocolo de Internet o dirección IP) y un número de puerto. Luego, el servidor mantiene abierta una conexión websocket para recibir solicitudes (cuando los ejecuta en RStudio, su servidor es su computadora). La sesión de R detrás de la aplicación se asegurará de que esta solicitud se

traduzca en la interactividad deseada y envíe la respuesta, generalmente un objeto actualizado, como un gráfico o una tabla (Figura 31).

Tabla 9. Widgets estándar de Shiny

Widget	Función
`actionButton`	Botón de acción
`checkboxGroupInput`	Un grupo de casillas de verificación
`checkboxInput`	Una sola casilla de verificación
`fileInput`	Un asistente de control de carga de archivos
`helpText`	Texto de ayuda que se puede agregar a un formulario de entrada
`numericInput`	Un campo para ingresar números
`radioButtons`	Un conjunto de botones de radio
`selectInput`	Un cuadro con opciones para seleccionar
`sliderInput`	Una barra deslizante
`submitButton`	Un botón de enviar
`textInput`	Un campo para ingresar texto

El paquete GeoWeightedModel (actualmente en la versión 1.0.2) está alojado en el repositorio de Comprehensive R Archive Network (CRAN) y se puede instalar con el siguiente comando:

```
R> install.packages("GeoWeightedModel")
R> library(GeoWeightedModel)
```

Posteriormente en la consola de R o RStudio, la aplicación o interfaz gráfica de usuario (GUI) se inicia en una pestaña del navegador web con la siguiente instrucción:

```
R> runGeoWeightedModel()
```

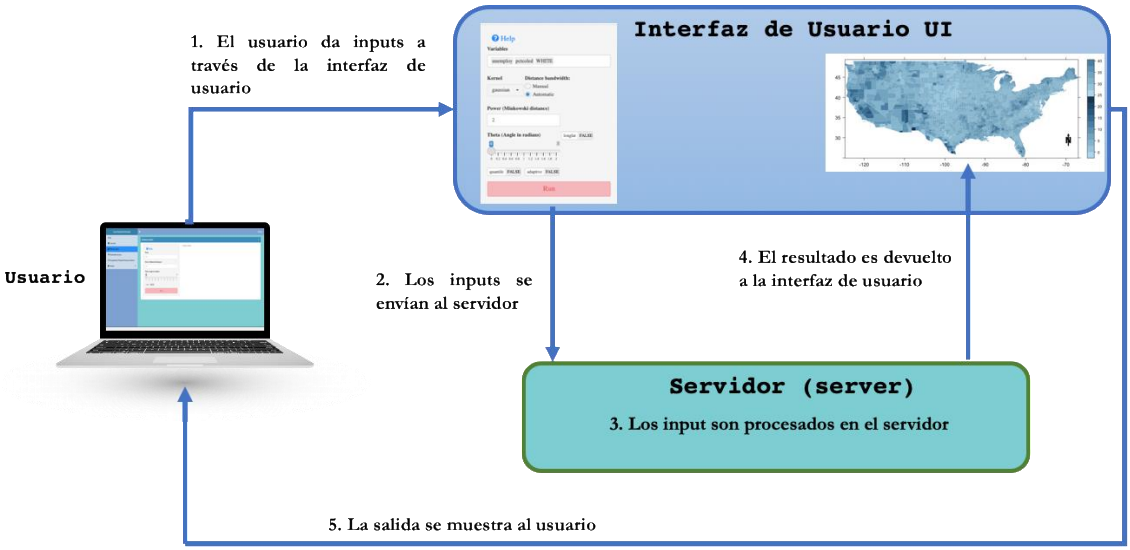


Figura 31. Arquitectura básica de una aplicación desarrollada en Shiny

5.2. INTERFAZ GRÁFICA DE USUARIO (GUI)

La GUI implementada proporciona un menú que, de arriba a abajo, guía al usuario a través del análisis. Consta de seis paneles de menús principales localizados en la parte izquierda de la página (Figura 32).

Figura 32. Interfaz gráfica de usuario de la aplicación GeoWeighthedModel

5.2.1. Menú load data

Este menú consta de dos cuadros. El primero nos permite cargar el archivo de datos, el cual debe estar en formato ".xls" o ".xlsx". Este procedimiento es realizado haciendo clic en el botón *Browse*, luego se debe seleccionar la hoja de trabajo y posteriormente se debe hacer clic en el botón para obtener datos. Este archivo debe tener una columna de identificación de área que debe coincidir con el identificador de los *shapefiles*. El segundo recuadro nos permite cargar el archivo del mapa que contiene las áreas de la región de estudio. El archivo del mapa debe estar en formato *shapefile* (".shp", ".dbf", ".prj", ".shx", ".xml") y se puede cargar haciendo clic en el botón *Browse* y seleccionando todos los archivos correspondientes al mapa. También encontramos en este menú una casilla de verificación donde podemos cargar un archivo de ejemplo. Adicionalmente, en la parte superior, existe un botón de ayuda *Help,* en donde se despliega la descripción del módulo y un vídeo corto a modo de tutorial donde se explica cómo funciona el menú (Figura 33). Vale mencionar que este botón de ayuda también se encuentra disponible en todos los menús con su vídeo correspondiente.

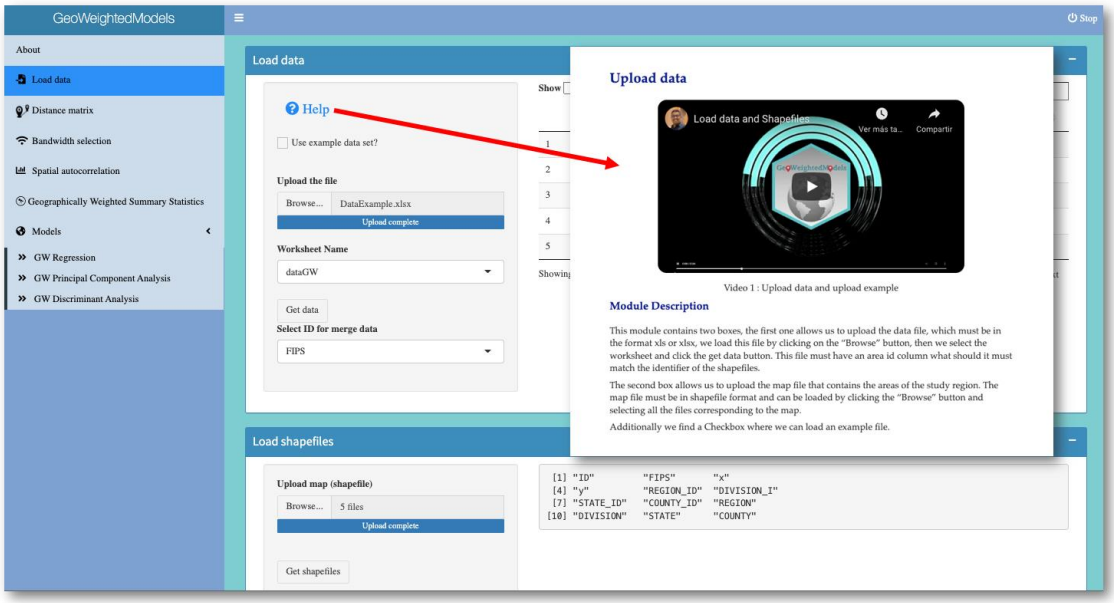


Figura 33. Menú *load data* de la aplicación GeoWeighthedModel

5.2.2. Menú Distance Matrix

El menú *Distance matrix* (Figura 34) permite calcular una matriz de distancias entre cualquier punto de calibración del modelo GW y los puntos de datos. Los datos de salida se pueden descargar en formato ".csv", ".xls" o ".pdf" o se pueden copiar al portapapeles. En

la parte superior de la caja encontramos un botón de ayuda donde se especifican los parámetros de entrada y también se muestra un vídeo explicativo del uso del menú.

Dentro de los parámetros de entrada (*inputs*) de este menú tenemos:

- ***Focus***: es un número entero, indexado al punto del modelo GW actual, si focus = 0, se calcularán todas las distancias entre todos los puntos de calibración del modelo GW y los puntos de datos y se devolverá una matriz de distancias; si 0<focus<longitud (número de puntos de datos), entonces se calcularán las distancias entre los puntos del modelo GW del 'foco' y los puntos de datos y se devolverá un vector de distancias.
- ***Power(Minkowski distance)***: es la potencia de la distancia de Minkowski, por defecto es p=2, es decir, la distancia Euclídea.
- ***Theta***: Un ángulo en radianes para rotar el sistema de coordenadas, el valor predeterminado es 0.
- ***longlat***: si es VERDADERO, se calcularán las distancias del círculo máximo.

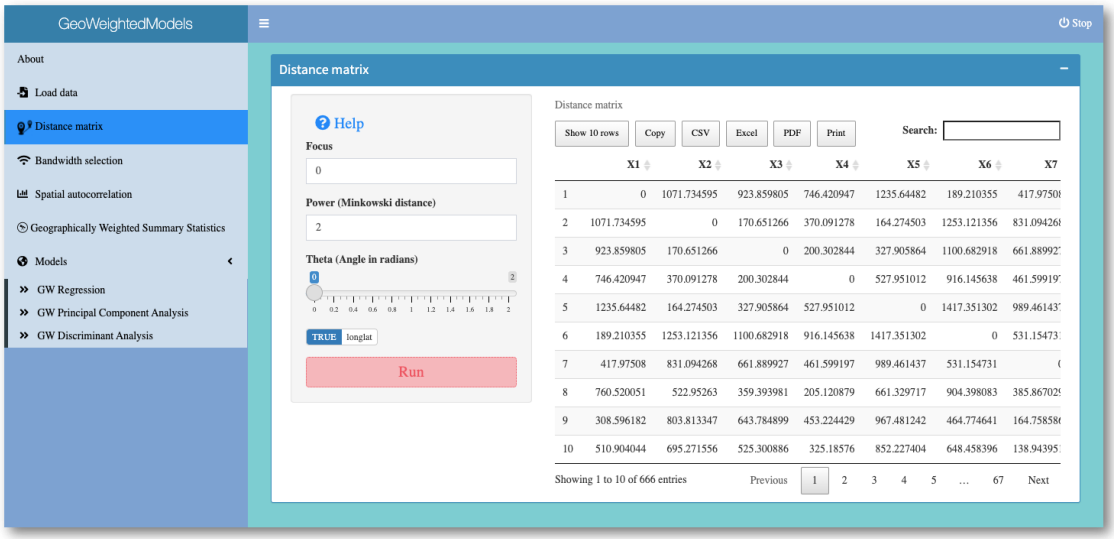


Figura 34. Menú *Distance matrix* de la aplicación GeoWeighthedModel

5.2.3. Menú Bandwidth

Este menú contiene los inputs para la selección automática del ancho de banda para calibrar la regresión GW básica (*bw.gwr*), el modelo GWR generalizado (*bw.ggwr*), el análisis

de componentes principales GW (*bw.gwpca*) y el análisis discriminante GW (*bw.gwda*) (Figura 35).

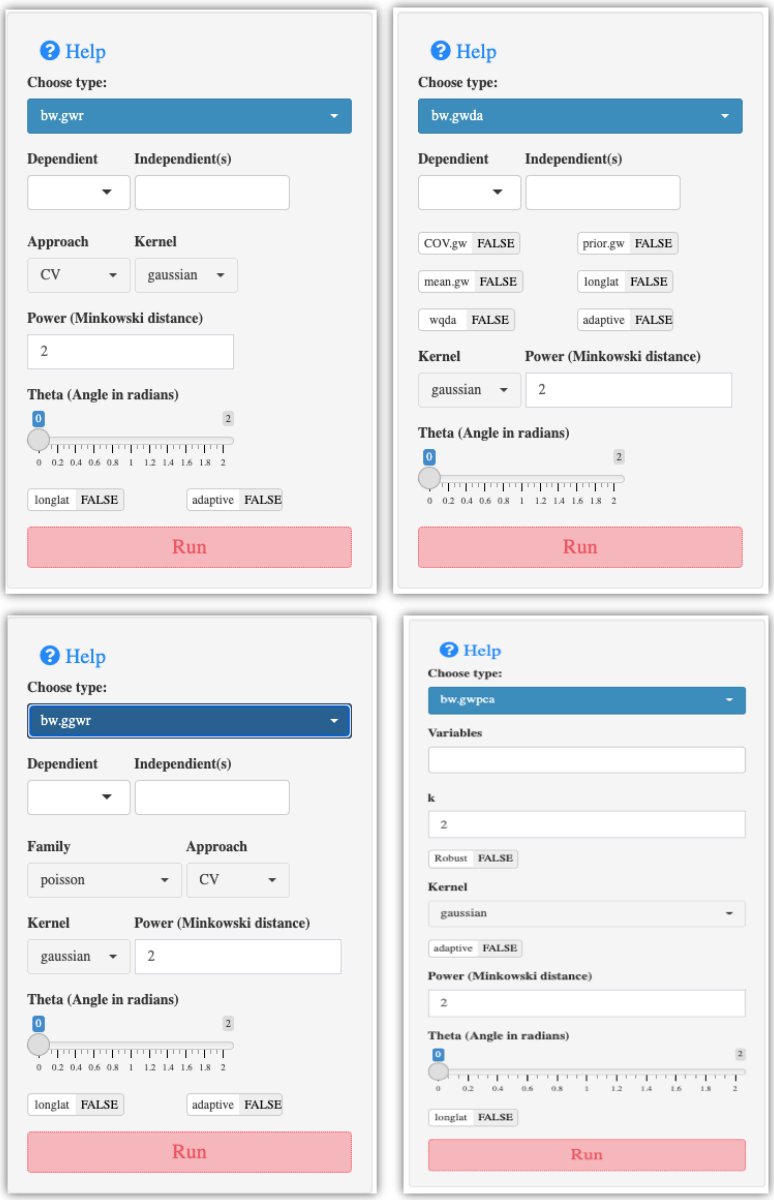


Figura 35. Opciones del Menú *Bandwidth* de la aplicación GeoWeighthedModel

En la Tabla 10 se enumeran los argumentos del menú *Bandwidth* que se describen a continuación:

- ***Dependent***: Variable dependiente del modelo de regresión.
- ***Independent(s)***: Variable(s) independiente(s) del modelo de regresión.
- ***Family***: una descripción de la distribución de errores del modelo y la función de enlace, que puede ser Poisson o Binomial.

- ***Approach***: tiene dos opciones: CV para el enfoque de validación cruzada o por el enfoque corregido del criterio de información de Akaike (AIC).
- ***Kernel***: un conjunto de cinco funciones de kernel de uso común entre las cuales se puede elegir: gaussiana, exponencial, box-car, bicuadrada y tricubo.
- ***Power***: la potencia de la distancia de Minkowski (p=1 es la distancia de Manhattan, p=2 es la distancia Euclídea).
- ***Theta***: un ángulo en radianes para rotar el sistema de coordenadas, el valor predeterminado es 0 (sin rotación).
- ***longlat***: si es VERDADERO, se calcularán las distancias del círculo máximo.
- ***Adaptative***: si es VERDADERO, encuentra un núcleo adaptable con un ancho de banda proporcional al número de vecinos más cercanos (es decir, distancia adaptable); de lo contrario, encuentra un Kernel fijo (el ancho de banda es una distancia fija).
- ***Cov.gw***: si es VERDADERO, se usa la matriz de varianzas-covarianzas local para el análisis discriminante del modelo GW; de lo contrario, se utiliza la matriz de varianzas-covarianzas global.
- ***Prior.gw***: si es VERDADERO, se utiliza la probabilidad previa local para el análisis discriminante de GW; de lo contrario, se utiliza una probabilidad previa fija.
- ***Mean.gw***: si es VERDADERO, se usa la media local para el análisis discriminante de GW; de lo contrario, se utiliza la media global.
- ***wqda***: si es VERDADERO, se aplicará un análisis discriminante cuadrático ponderado; de lo contrario, se aplicará un análisis discriminante lineal ponderado.
- ***Variables***: un vector de nombres de variables a evaluar en el GWPCA.
- ***Robust***: si es VERDADERO se utilizará un GWPCA robusto; de lo contrario, se utilizará el GWPCA clásico.

Tabla 10. Argumentos *input* del Menú *Bandwidth* de la aplicación GeoWeighthedModel

Argumentos	***bw.gwr***	***bw.ggwr***	***bw.gwda***	***bw.gwpca***
Dependient	✓	✓	✓	
Independient	✓	✓	✓	
Family	x	✓		
Approach	✓	✓		
Kernel	✓	✓	✓	✓
Power	✓	✓	✓	✓
Theta	✓	✓	✓	✓
Longlat	✓	✓	✓	✓
Adaptative	✓	✓	✓	✓
Cov.gw			✓	
Prior.gw			✓	
Mean.gw			✓	
wqda			✓	
Variables				✓
Robust				✓

5.2.4. Menú Spatial Autocorrelation.

Una pregunta frecuente es si las características con valores similares están agrupadas, distribuidas aleatoriamente o dispersas. La autocorrelación espacial mide el grado de correlación en el espacio (Cliff & Ord, 1981). Las pruebas de autocorrelación espacial examinan la independencia del valor observado en relación con los valores de esa variable en ubicaciones vecinas. En este módulo, se calcula el valor del índice I de Moran y tanto la puntuación z como el valor p para evaluar la importancia de ese índice. Los valores de probabilidad son aproximaciones numéricas del área bajo la curva para una distribución conocida, limitada por el estadístico de prueba.

Los argumentos o *inputs* de entrada en este menú son los siguientes:

- ***Variable***: un vector numérico de la misma longitud que la lista de vecinos.

- ***zero.policy***: predeterminado NULL, usa el valor de opción global; si se escoge VERDADERO asigna cero al valor perdido de zonas sin vecinos; si escoge FALSO asigna NA.
- ***style***: a partir de una lista de vecinos binarios, en la que las regiones se enumeran como vecinos o están ausentes, la función agrega una lista de pesos con valores dados por el estilo de esquema de codificación elegido. *B* es la codificación binaria básica, *W* está estandarizado por filas (sumas de todos los enlaces a n), *C* está estandarizado globalmente (sumas de todos los enlaces a n), *U* es igual a *C* dividido por el número de vecinos (sumas de todos los enlaces a n), mientras que *S* es el esquema de codificación estabilizador de varianza propuesto por Tiefelsdorf et al. (1999) (sumas sobre todos los enlaces a n).
- ***nsim***: número de permutaciones.
- ***alternative:*** una cadena de caracteres que especifica la hipótesis alternativa "*greater*" (por defecto), "l*ess*" or "*two.sided*".

Como resultado se encuentran los siguientes componentes: prueba del índice I de Moran bajo aleatorización, simulación de Monte-Carlo del índice I de Moran, resumen de estadísticas del índice I de Moran local y simulación de Monte-Carlo de resumen de estadísticas de I de Moran local.

El índice I de Moran, así como el p-valor se pueden representar gráficamente y personalizar utilizando las opciones de configuración que se encuentran en la pestaña *Plot*. Entre estas opciones de configuración tenemos *Select* (la cual muestra las opciones *Imoran_i* y *Imoran_p*), *Main title*, *Select pallet* (16 paletas de colores disponibles), *North arrow* (posición de la flecha que señala el Norte, latitud y longitud**),** *Select the Option* (".png" or ".pdf" format), y un botón de descarga (Figura 36).

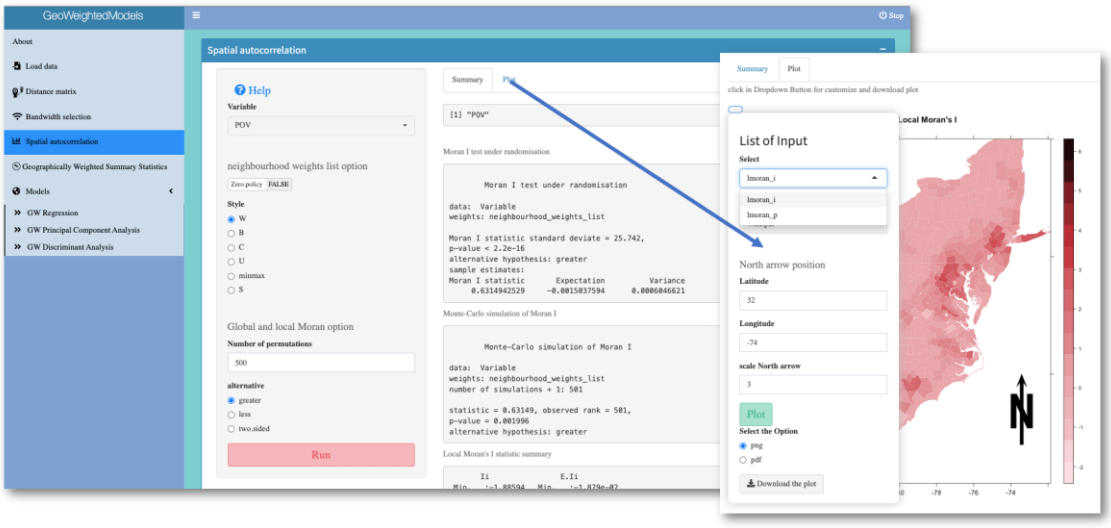


Figura 36. Opciones del Menú *Spatial autocorrelation* de la aplicación GeoWeightedModel. Se muestra las ventanas *summary* y *plot*

5.2.5. Menú Geographically Weighted Summary Statistics

Brunsdon et al. (2002), describieron los estadísticos de resumen locales ponderados geográficamente (media, desviación estándar y correlación de Pearson) para visualizar la variación geográfica en la distribución estadística. Sin embrago, Gollini et al. (2015), en el entorno de la computación estadística R en el paquete Gwmodel, describen cómo calcular medias GW, desviaciones estándar de GW y medidas de sesgo de GW, que constituyen un conjunto de estadísticos básicos de resumen GW. Además, para mitigar cualquier efecto adverso de valores atípicos en los estadísticos locales, describen también un conjunto de alternativas robustas como medianas GW y rangos intercuartílicos GW y desequilibrio cuantil GW. Además, también describen las correlaciones GW tanto en sus formas básicas y robustas (de Pearson y de Spearman, respectivamente), proporcionando un conjunto de estadísticos resumidos bivariados locales. En la Figura 37 podemos observar las opciones del Menú Geographically Weighted Summary Statistics de la aplicación GeoWeighthedModel.

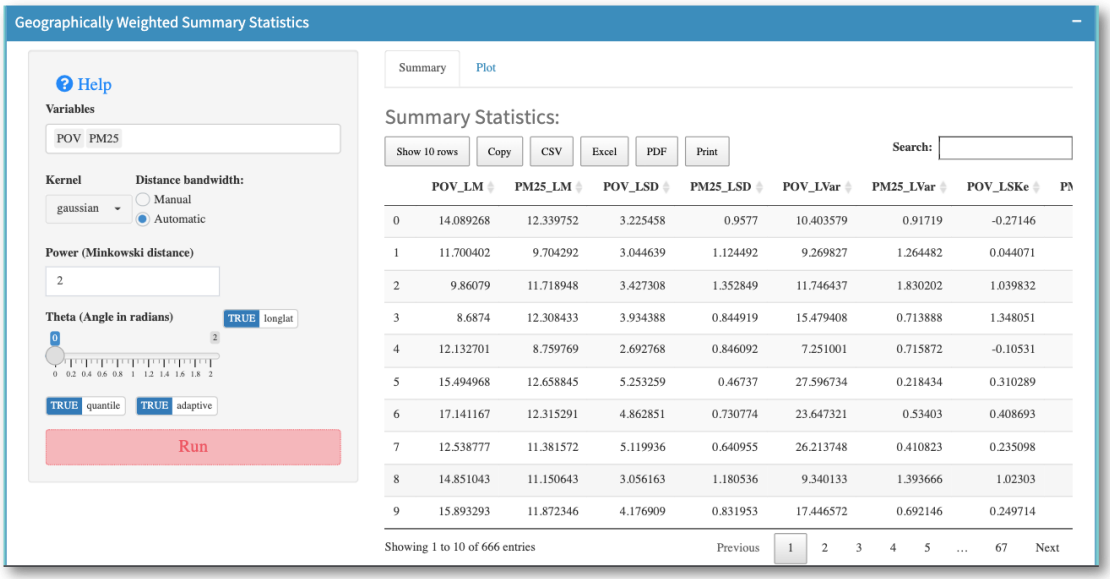


Figura 37. Opciones del Menú *Geographically Weighted Summary Statistics* de la aplicación GeoWeighthedModel. Se muestra la ventana *summary*

Los argumentos o *inputs* de entrada en este menú son los siguientes:

- ***Variables***: un vector de nombres de variables a ser evaluadas.
- ***Kernel***: un conjunto de cinco funciones de Kernel de uso común entre las cuales se puede elegir: gaussiana, exponencial, box-car, bicuadrada y tricubo.
- ***Distance bandwidth***: ancho de banda utilizado en la función de ponderación. Tiene dos opciones, automática que se calcula en el módulo de selección de *bandwidth* y manual en la que el usuario introduce el valor.
- ***Power***: la potencia de la distancia de Minkowski (p=1 es la distancia de Manhattan, p=2 es la distancia Euclídea).
- ***Theta***: un ángulo en radianes para rotar el sistema de coordenadas, el valor predeterminado es 0 (sin rotación).
- ***longlat***: si es VERDADERO, se calcularán las distancias del círculo máximo.
- ***quantile***: si es VERDADERO, se calcularán la mediana y el rango intercuartílico GW.

Este menú tiene dos tipos de salida: una tabular en donde se presenta una tabla que se puede descargar en diversos formatos (".csv", ".xlxs" y ".pdf"), copiar en el portapapeles o imprimir. En la Tabla 11 se encuentran todos los estadísticos de resumen, medias locales, desviaciones estándar locales, varianza local, sesgo local, coeficientes de variación locales,

covarianzas locales, correlaciones locales (de Pearson), correlaciones locales (de Spearman), medianas locales, rangos intercuartílicos locales, desequilibrio cuantil locales y coordenadas.

Tabla 11. Lista de estadísticos de resumen ponderados geográficamente

Nombre de la variable	Estadístico referido	¿Robusto ?
X_LM	Media GW	No
X_LSD	Desviación estándar GW	No
X_Lvar	Varianza GW	No
X_LSKe	Asimetría GW	No
X_LCV	Coeficiente de variación GW	Si
X_IQR	Rango intercuartílico GW	Si
X_QI	Desequilibrio cuantil GW	Si
Cov_X.Y	Covarianza GW	No
Corr_X.Y	Correlación de GW Pearson	No
Spearman_rho_X.Y	Correlación de GW Spearman	Si

Nota: Hay que tener en cuenta que X y Y se reemplazan por los nombres de las variables que se están investigando y que los estadísticos bivariados se calculan solo si se proporcionan dos nombres de variables en el input.

El segundo tipo de salida, en la pestaña *Plot*, se pueden personalizar y descargar gráficos utilizando las opciones disponibles para tal fin. Entre estas opciones tenemos *Select variable*, *Main title*, *Select pallet* (16 paletas de colores disponibles), *North arrow* (posición de la flecha que señala el Norte, latitud y longitud), *Select the Option* (".png" or ".pdf" format), y un botón de descarga (Figura 38).

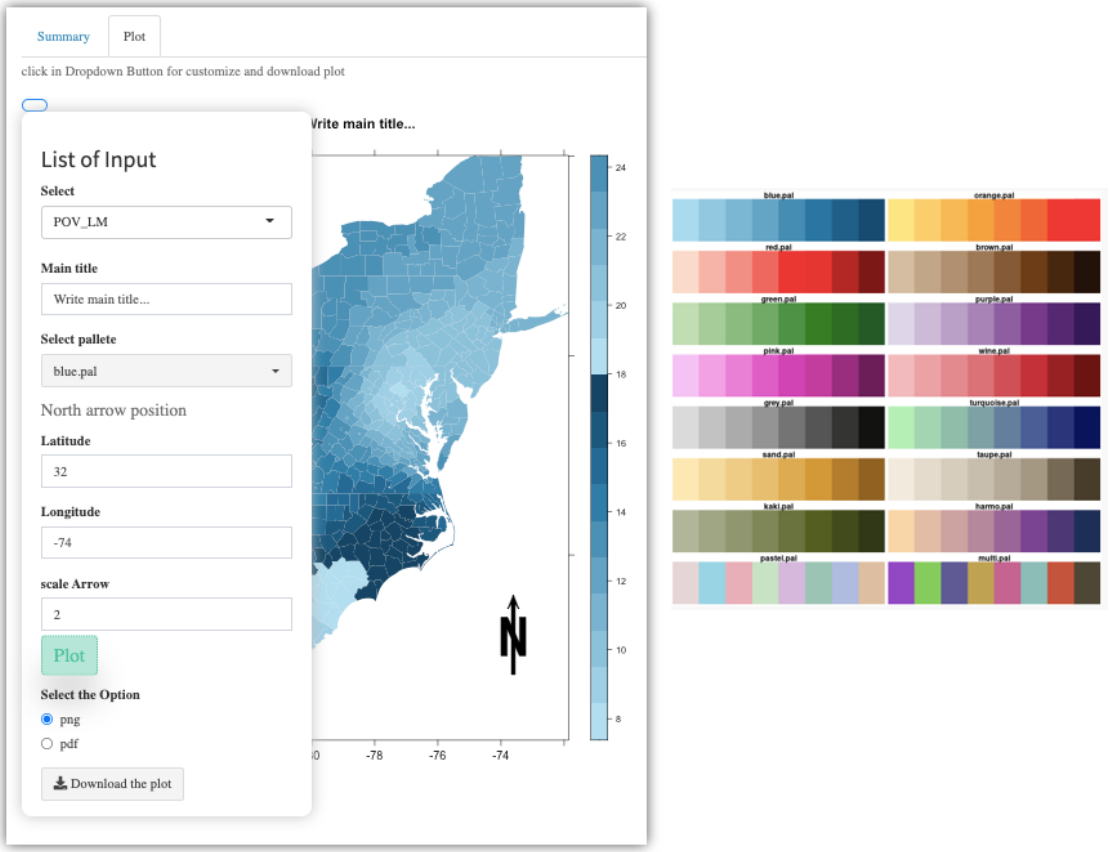


Figura 38. Opciones de configuración de la salida gráfica del Menú *Geographically Weighted Summary Statistics* de la aplicación GeoWeighthedModel. A la derecha la paleta de colores disponible

5.2.6. Menú GW Regression

En este menú se pueden seleccionar varias opciones entre ellas *Local collinearity diagnostics*, y además del modelo básico, *Basic GWR model*, otros cinco: *Robust GWR*, *Generalised GWR*, *Heteroskedastic GWR*, *Mixed GWR* y *Scalabe GWR*. En la Tabla 12 y Figura 39 se pueden observar los parámetros de entradas de cada modelo.

Tabla 12. Argumentos *input* del Menú GW Regression de la aplicación GeoWeighthedModel

Argumento		GWR					
	Local collinearity diagnostics	*Basic*	*Robust*	*Generalized*	*Heterocedastic*	*Mixed*	*Scalable*
Dependent	✓	✓	✓	✓	✓	✓	✓
Independient	✓	✓	✓	✓	✓	✓	✓
Family				✓			
Cv		✓		✓			
Kernel	✓	✓	✓	✓	✓	✓	✓
Power	✓	✓	✓	✓	✓	✓	✓
Theta	✓	✓	✓	✓	✓	✓	✓
Longlat	✓	✓	✓	✓	✓	✓	✓
Adaptive	✓	✓	✓	✓	✓	✓	
Distance bandwidth	✓	✓	✓	✓	✓	✓	
Max iter			✓		✓		
Fixed						✓	
Intercep fixed						✓	
Diagnostic						✓	
F123			✓				
Filtered			✓				
bw.adapt							✓
Polynomial							✓

- ***Dependent***: Variable dependiente del modelo de regresión.
- ***Independient(s)***: Variable(s) independiente(s) del modelo de regresión.
- ***Family***: una descripción de la distribución de errores del modelo y la función de enlace, que puede ser "Poisson" o "Binomial".
- ***Cv***: si es VERDADERO, se calcularán los datos de validación cruzada.
- ***Kernel***: Un conjunto de cinco funciones del Kernel de uso común, entre las cuales se puede elegir: gaussiana, exponencial, box-car, bicuadrada y tricubo.
- ***Power***: la potencia de la distancia de Minkowski (p=1 es la distancia de Manhattan, p=2 es la distancia Euclídea).
- ***Theta***: un ángulo en radianes para rotar el sistema de coordenadas, el valor predeterminado es 0 (sin rotación).

- ***Adaptative***: si es VERDADERO, encuentra un núcleo adaptable con un ancho de banda proporcional al número de vecinos más cercanos (es decir, distancia adaptable); de lo contrario, encuentra un Kernel fijo (el ancho de banda es una distancia fija).
- ***Distance bandwidth***: ancho de banda utilizado en la función de ponderación. Tiene dos opciones: automática, que se calcula en el módulo de selección de "Ancho de Banda" y manual, en la cual el usuario ingresa el valor.
- ***Max iter***: número máximo de iteraciones para el enfoque automático.
- ***Fixed***: variables independientes que aparecieron en la fórmula que deben ser tratadas como globales.
- ***Intercep fixed***: si es VERDADERO, la intersección se tratará como global.
- ***Diagnostic***: si es VERDADERO se calcularán los diagnósticos.
- ***F123***: predeterminado FALSO de lo contrario calcula los resultados de la prueba F
- ***Filtered***: predeterminado a FALSO se usa el enfoque automático; si es VERDADERO, se emplea el enfoque de datos filtrados, como se describe en Fotheringham et al. (2002).
- ***bw.adapt***: ancho de banda adaptativo (es decir, número de vecinos más cercanos) utilizado para la ponderación geográfica.
- ***Polynomial***: grado del polinomio para aproximar la función Kernel.

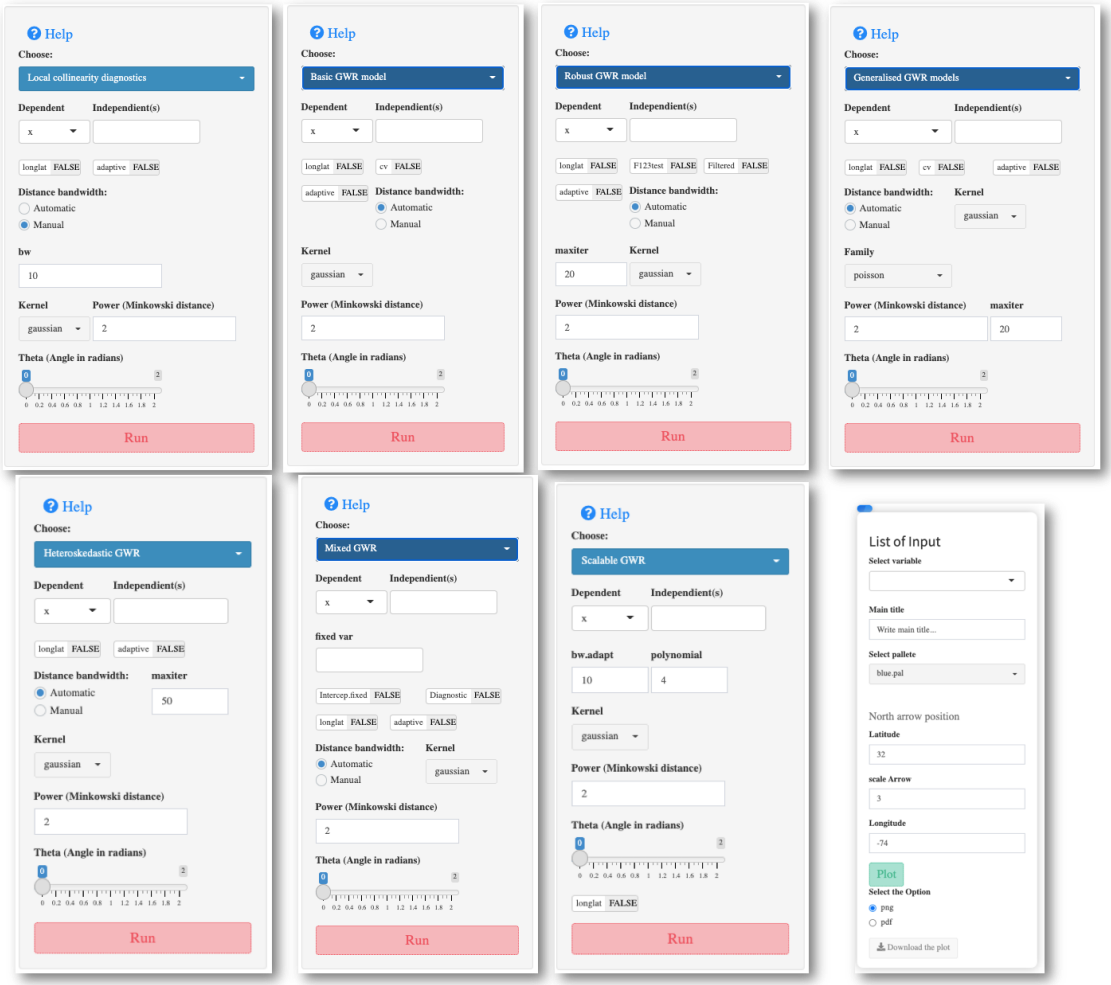


Figura 39. Opciones del Menú *GW Regression* de la aplicación GeoWeighthedModel

Al igual que en el caso del ítem anterior, las salidas son tanto tabulares como gráficas. Entre las salidas podemos encontrar los fit.points, estimaciones de los coeficientes GWR, valores pronosticados, errores estándar de los coeficientes y valores t. Todos esos valores pueden ser representados gráficamente y las opciones de configuración de los gráficos son las mismas que las descritas anteriormente.

5.2.7. Menú GW Principal Component Analysis

Este módulo implementa un Análisis de Componentes Principales Ponderado Geográficamente (GWPCA, por sus iniciales en inglés) básico o robusto.

Los inputs de este menú son los siguientes:

- ***Variables***: un vector de nombres de variables a evaluar.
- ***k***: el número de componentes a retener; k debe ser menor que el número de variables.
- ***Kernel***: Un conjunto de cinco funciones del Kernel de uso común; entre las cuales se puede elegir: gaussiana, exponencial, box-car, bicuadrada y tricubo.
- ***Power***: la potencia de la distancia de Minkowski (p=1 es la distancia de Manhattan, p=2 es la distancia Euclídea).
- ***Theta***: un ángulo en radianes para rotar el sistema de coordenadas, el valor predeterminado es 0 (sin rotación).
- ***Adaptative***: si es VERDADERO, encuentra un núcleo adaptable con un ancho de banda proporcional al número de vecinos más cercanos (es decir, distancia adaptable); de lo contrario, encuentra un Kernel fijo (el ancho de banda es una distancia fija).
- ***Distance bandwidth***: ancho de banda utilizado en la función de ponderación. Tiene dos opciones: automática, que se calcula en el módulo de selección de "Ancho de Banda" y manual en la cual el usuario ingresa el valor.
- ***longlat***: si es VERDADERO, se calcularán las distancias del círculo máximo.
- Robust: si es VERDADERO, se aplicará GWPCA robusto; de lo contrario, se aplicará la GWPCA básico.

La ventana que se despliega en este menú consta además de tres pestañas, *Summary, Plot winning variable* y *Percent total variation* (Figura 40).

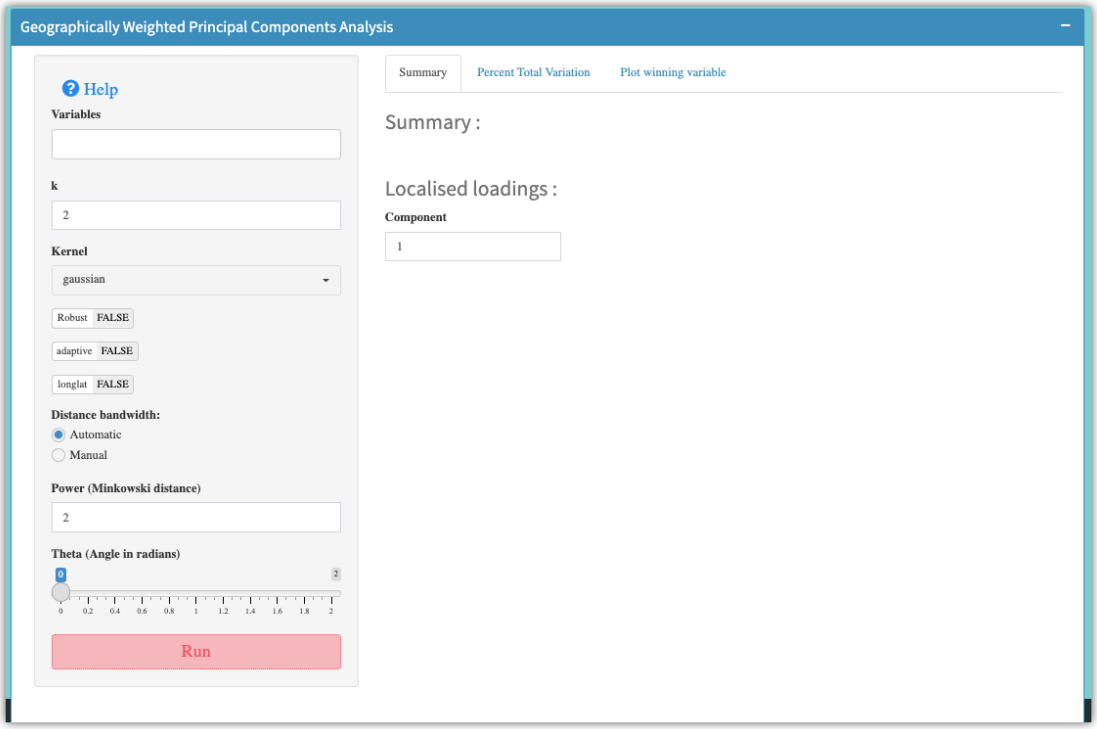


Figura 40. Opciones del Menú *GW Principal Component Analysis* de la aplicación GeoWeighthedModel

En la pestaña *Summary* es en donde se despliega información tabular que incluye los parámetros de ajuste del modelo para generar el archivo de informe, las cargas locals, proporciones locales de varianza para cada componente principal, proporción acumulada y variable ganadora para cada componente principal (es decir, la que tiene la carga local absoluta más alta).

Podemos visualizar cómo cada una de las variables influye localmente en una componente dada, en la pestaña *Plot winning variable* al mapear la variable ganadora con la carga absoluta más alta.

En la pestaña *Percent total variation* encontramos el porcentaje de total de variación explicada. Al igual que en los otros menús, los gráficos generados se pueden personalizar y descargar utilizando las opciones disponibles para tal fin.

5.2.8. Menú GW Discriminat Analysis

Este menú permite realizar un Análisis Discriminante Ponderado Geográficamente (GWDA, por sus iniciales en inglés), incluido el cálculo de las probabilidades y la entropía de la ubicación. Los parámetros de entrada son :

- ***Grouping factor***: Variable utilizada para agrupar.
- ***Discriminators***: Variables utilizadas como discriminadores.
- ***Mean.gw***: si es VERDADERO, se utiliza la media local para el análisis discriminante de GW; de lo contrario, se utiliza la media global.
- ***Cov.gw***: si es VERDADERO, se usa la matriz de varianzas-covarianzas local para el análisis discriminante de GW; de lo contrario, se utiliza la matriz de varianzas-covarianzas global.
- ***Prior.gw***: si es VERDADERO, se utiliza la probabilidad previa local para el análisis discriminante de GW; de lo contrario, se utiliza una probabilidad previa fija.
- ***longlat***: si es VERDADERO se calcularán las distancias del gran círculo.
- ***wqda***: si es VERDADERO, se aplicará un análisis discriminante cuadrático ponderado; de lo contrario, se aplicará un análisis discriminante lineal ponderado.
- ***Adaptive***: Si es VERDADERO encuentre un Kernel adaptativo con un ancho de banda proporcional al número de vecinos más cercanos (es decir, distancia adaptativa); de lo contrario, busque un núcleo fijo (el ancho de banda es una distancia fija).
- ***Distance bandwidth***: ancho de banda utilizado en la función de ponderación. Tiene dos opciones: automática que se calcula en el módulo de selección de "Ancho de Banda" y manual en la cual el usuario ingresa el valor.
- ***Power (Minkowski distance)***: la potencia de la distancia de Minkowski (p=1 es la distancia de Manhattan, p=2 es la distancia Euclídea).
- ***Kernel***: Un conjunto de cinco funciones del Kernel de uso común; entre las cuales se puede elegir: gaussiana, exponencial, box-car, bicuadrada y tricubo.
- ***Theta:*** un ángulo en radianes para rotar el sistema de coordenadas, el valor predeterminado es 0 (sin rotación).

Dentro de las salidas, se incluye un objeto con las probabilidades para cada nivel, la probabilidad más alta y la entropía de las probabilidades. Por otro lado, en la pestaña *Plot*, se

pueden personalizar y descargar gráficos utilizando las opciones disponibles para tal fin (Figura 41).

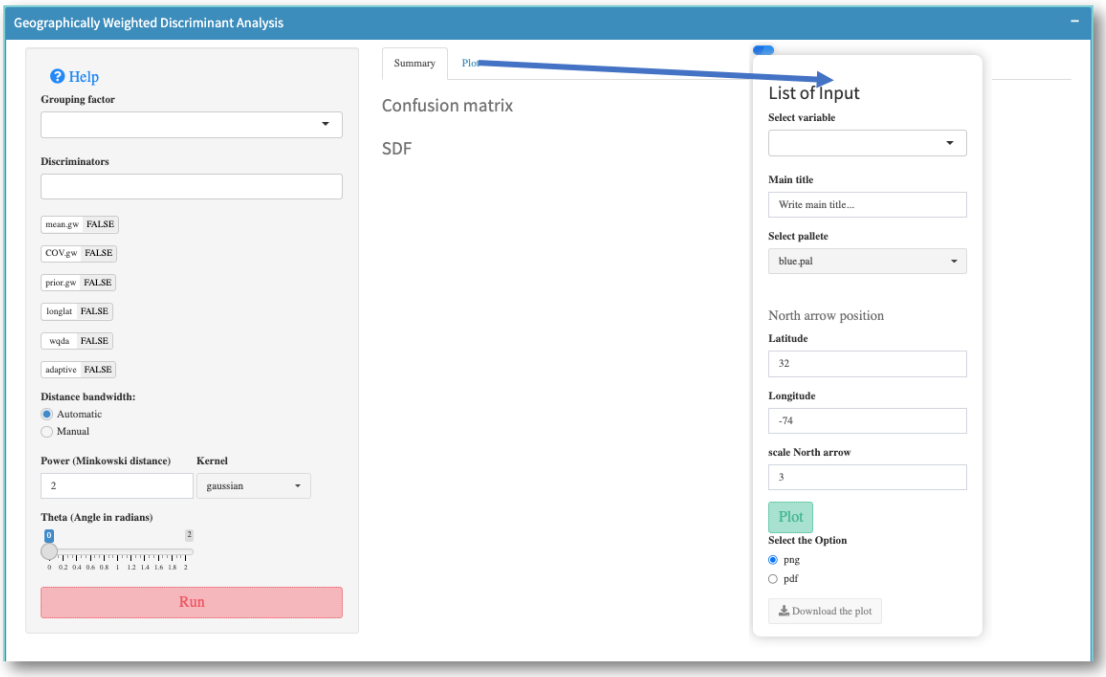


Figura 41. Opciones del Menú *GW Principal Component Analysis* de la aplicación GeoWeighthedModel

5.3. Ejemplo de Aplicación

Para demostrar cómo se utiliza la GUI, se utilizará una base de datos que recoge información para explorar y visualizar la heterogeneidad espacial de las relaciones entre la mortalidad por cáncer de pulmón y bronquios y factores de riesgo a nivel de condados de EE. UU. Esta base de datos resulta propicia para aplicar todos los métodos implementados en GeoWeightedModel, con exepción de GWDA.

El archivo de las variables disponible, así como los *shapefiles,* se encuentran alojados en el enlace https://www.dropbox.com/s/lrz6og0ld2m64df/Data_GWR.7z?dl=00. En este enlace se encuentra parte de la información utilizada del artículo "Explainable artificial intelligence (XAI) for exploring spatial variability of lung and bronchus cancer (LBC) mortality rates in the contiguous USA" (Ahmed, et al. 2021).

Para facilidad de lectura de los datos en la aplicación, el archivo fue convertido a formato de Excel. Aunque en el artículo mencionado hacen referencia a información de 3107 condados, en el repositorio sólo hay información disponible para 666, ubicados en la zona del Atlántico de EE.UU..

El archivo de datos de Excel con las 666 observaciones contiene seis variables (Anexo 6), la posición geográfica (x = latitud, y = longitud) y una columna (FIPS) para la identificación del área para cada condado. Las variables se describen a continuación:

- **Tasas de mortalidad por cáncer de pulmón y bronquios (Rate).**

Tasas de mortalidad por condado, ajustadas por edad, debidas a cáncer de pulmón y bronquios, desde 1980 hasta 2014; estas se obtuvieron del Instituto para la Métrica y Evaluación de la Salud (Institute for Health Metrics and Evaluation, IHME). La tasa se estimó utilizando modelos de regresión de efectos mixtos bayesianos espacialmente explícitos para hombres y mujeres (Mokdad et al., 2017). La tasa está dada en número de muertes por 100.000 habitantes.

- **Prevalencia del tabaquismo estandarizado por edad (%) (SMOK)**

Prevalencia de tabaquismo estandarizada por edad (%) y por condado, estimada a partir de los datos de la encuesta del Sistema de Vigilancia de Factores de Riesgo del Comportamiento (BRFSS) (Dwyer-Lindgren et al., 2014).

- **Tasa de pobreza estandarizada por edad por condado (POV)**

Los datos de pobreza a nivel de condado (% de población por debajo del nivel de pobreza), provienen del Programa de Estimaciones de Pobreza e Ingresos en Áreas Pequeñas del Censo de EE. UU. (SAIPE, Small Area Income and Poverty Estimates).

- **Partículas PM2.5 (PM2.5)**

El promedio anual de material de partículas respirable PM2.5. Se modelizó a partir de la profundidad óptica de aerosoles de múltiples instrumentos satelitales y se validó mediante sitios de monitoreo de PM2.5 en tierra (Van Donkelaar et al., 2016).

- **Dióxido de nitrógeno (NO2)**

Concentraciones medias anuales de NO2 en el ambiente, estimadas a partir de tres instrumentos satelitales, como el Experimento de Monitoreo Global de Ozono (GOME, Global Ozone Monitoring Experiment), el Espectrómetro de Absorción de Imágenes de Barrido para Cartografía Atmosférica (SCIAMACHY, SCanning Imaging Absorption spectroMeter for Atmospheric CartograpHY) y el satélite GOME-2 en combinación con el modelo de transporte químico (Geddes et al., 2016).

- **Dióxido de azufre (SO2)**

Concentraciones medias anuales de SO2 en el ambiente obtenidas de mediciones tanto satelitales como de superficie y estimadas a partir de series temporales (Fioletov et al., 2017)

5.3.1. Ejemplo: Análisis de la Autocorrelación Espacial.

El índice de autocorrelación espacial I de Moran es un estadístico inferencial, lo que significa que los resultados del análisis siempre se interpretan dentro del contexto de su hipótesis nula, es decir, que no existe autocorrelación. Consecuentemente::

- Cuando el p-valor no es estadísticamente significativo, no se puede rechazar la hipótesis nula. Es muy posible que la distribución espacial de los valores de las características sea el resultado de procesos espaciales aleatorios.
- Si el valor p-valor es estadísticamente significativo y la puntuación z es positiva. Hay evidencias en contra de la hipótesis nula. La distribución espacial de valores altos y/o valores bajos en el conjunto de datos está más agrupada espacialmente de lo que se esperaría si los procesos espaciales subyacentes fueran aleatorios.
- Si el valor p-valor es estadísticamente significativo y la puntuación z es negativa. Hay evidencias en contra de la hipótesis nula. La distribución espacial de valores altos y valores bajos en el conjunto de datos está más dispersa espacialmente de lo que se esperaría si los procesos espaciales subyacentes fueran aleatorios.

Resultados con los datos del ejemplo

En la Figura 42 se pueden observar las opciones de configuración utilizadas para los datos del ejemplo y parte de las salidas numéricas obtenidas.

Los valores tanto de la prueba del índice I de Moran como de la simulación de Monte-Carlo están, de forma global, más cerca de 1 (0,631), lo que se interpreta como la existencia de una autocorrelación espacial positiva. Conviene recordar que la autocorrelación espacial positiva ocurre cuando los valores altos (bajos) de la variable de estudio en una ubicación se asocian con valores altos (bajos) en las regiones vecinas. Por lo que la POV (tasa de pobreza) se presentará de manera similar en los diferentes puntos y esto confirma la posible existencia de clusters. Además se muestran agrupaciones estadísticamente significativas dado que el p-valor para cada método o criterio es menor a 0,05.

En lo que respecta a los índices locales de Moran, se puede ver en la Figura 43 que existe una autocorrelación espacial bien negativa bien positiva para la variable POV (tasa de pobreza) en los diferentes condados, siendo en muchos casos no significativa.

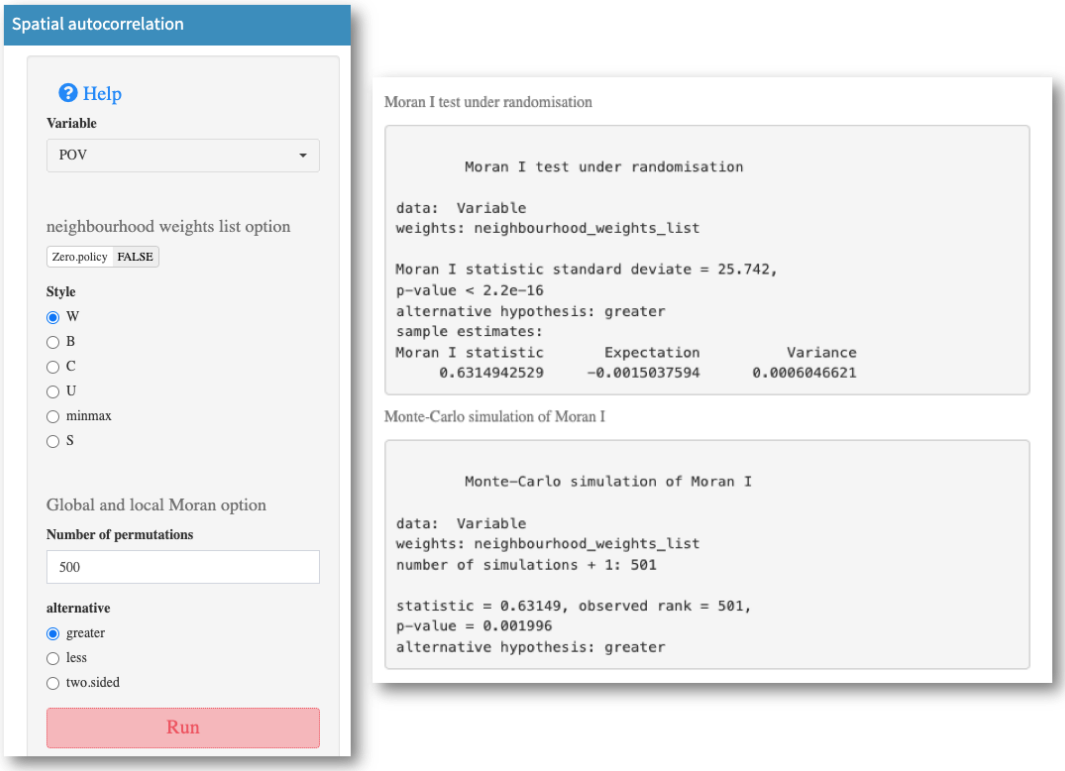


Figura 42. Opciones de configuración utilizadas en el ejemplo para *Spatial autocorrelation* y parte de las salidas numéricas

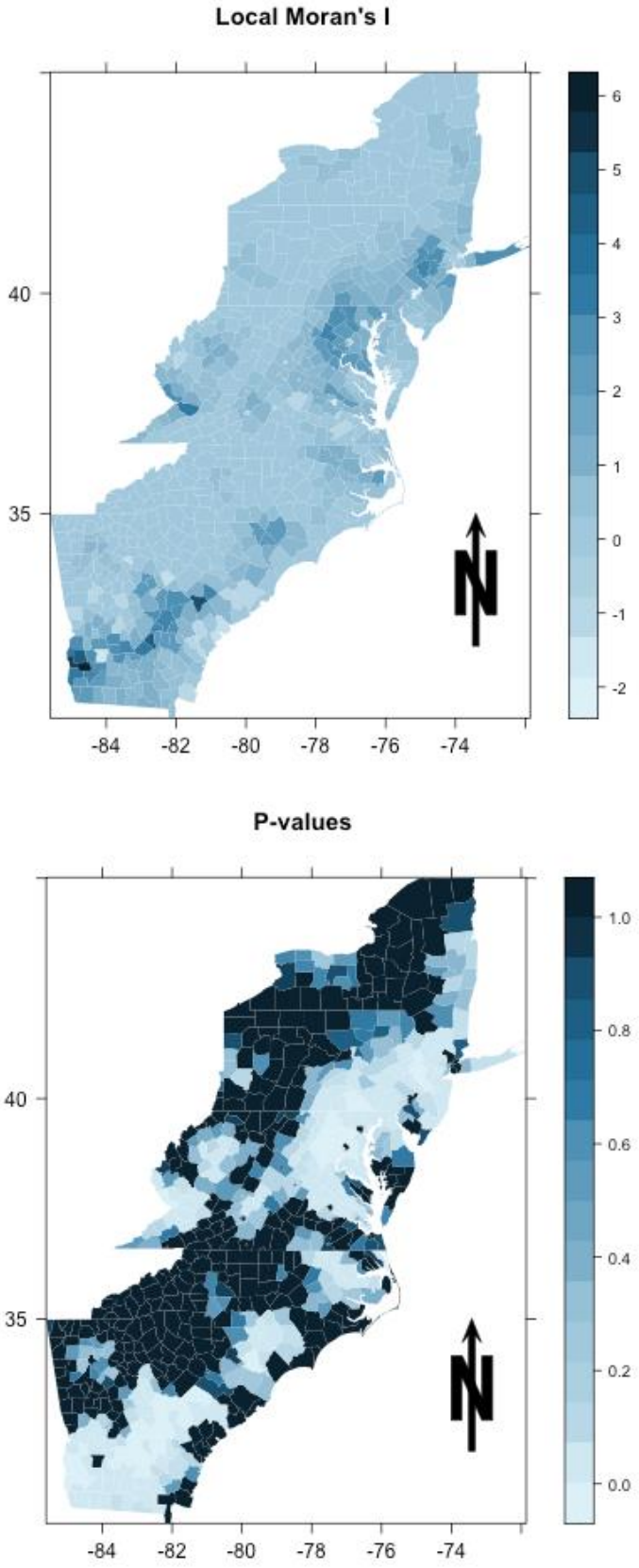


Figura 43. Índice I de Moran local y valores de probabilidad para la variable POV

5.3.2. Ejemplo: Estadísticas Resumidas Ponderadas Geográficamente

Para proporcionar una comprensión intuitiva de las técnicas GW, un comienzo útil es realizar un análisis exploratorio de los datos de las variables en estudio. En la Tabla 13 se observan algunos estadísticos descriptivos univariados, mientras que en la Figura 44 se presentan los histogramas y diagramas de caja de las variables consideradas en el ejemplo. Observamos como las variables NO2 (Dióxido de nitrógeno) y SO2 (Dióxido de azufre) presentan sesgo positivo con presencia de muchos valores atípicos, las variables SMOK (Prevalencia del tabaquismo estandarizado por edad) y PM25 (Partículas PM2.5) presentan asimetría negativa, mientras que Rate (Tasas de mortalidad por cáncer de pulmón y bronquios) y POV (Tasa de pobreza estandarizada por edad por condado) presentan asimetría positiva pero menos marcada.

Tabla 13. Estadísticos descriptivos de las variables consideradas en el ejemplo

Variable	Mínimo	Desviación	Media	Q1	Mediana	Q3	IQR	Máximo
NO2	0,5954	1,5892	2,2015	1,2608	1,7766	2,6475	1,3867	14,6895
PM25	7,3227	1,3247	11,3937	10,5077	11,6517	12,3843	1,8767	14,1333
POV	3,2600	5,6772	15,2085	11,3483	14,5367	18,9050	7,5567	35,2600
Rate	39	12,0742	69,9685	62	69	77	15	120
SMOK	13,5133	3,3275	26,0627	24,5717	26,5233	28,1850	3,6133	34,9600
SO2	0,0012	0,0711	0,0725	0,0279	0,0493	0,0920	0,0640	0,4208

El correlograma (Figura 45) muestra la relación de nuestro conjunto de variables. El área del círculo refleja la fuerza de la asociación recogida por el coeficiente de correlación de Pearson. Las cruces indican relaciones que no son estadísticamente significativas con un nivel de confianza del 95%. La dirección de la relación se indica con colores oscuros para la asociación negativa, mientras que, con color se indica una asociación positiva.

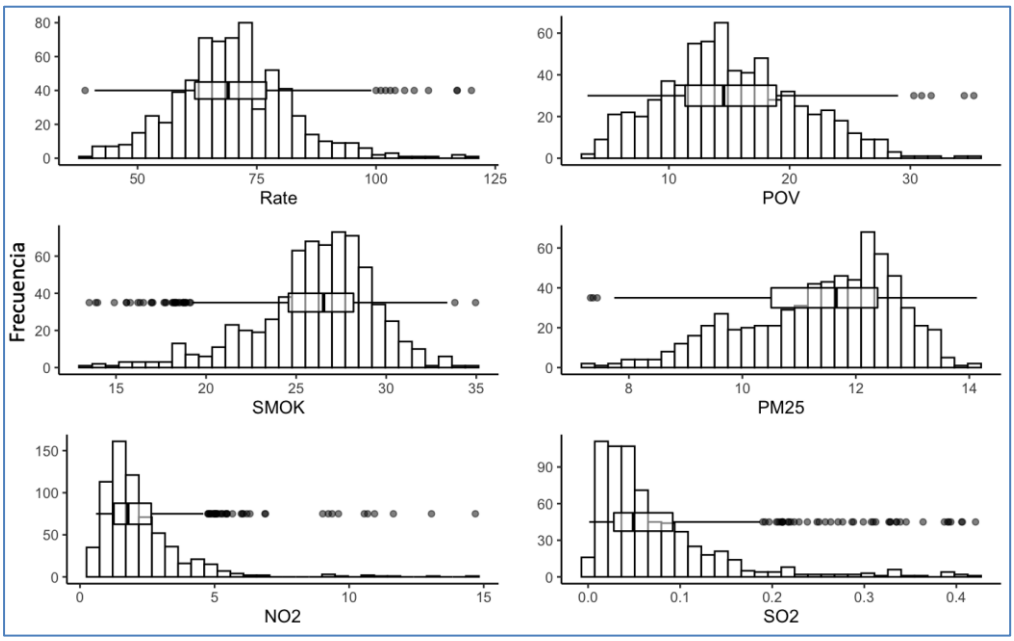


Figura 44. Histogramas y diagramas de caja de las variables utilizadas en el ejemplo para GWSS. Rate (Tasas de mortalidad por cáncer de pulmón y bronquios), SMOK (Prevalencia del tabaquismo estandarizado por edad), POV (Tasa de pobreza estandarizada por edad y por condado), NO2 (Dióxido de nitrógeno), PM25 (Partículas PM2.5) y SO2 (Dióxido de azufre)

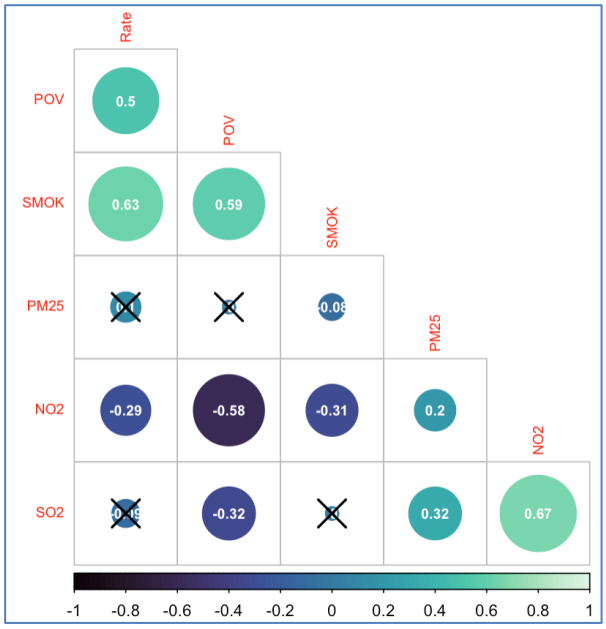


Figura 45. Correlograma de las variables utilizadas en el ejemplo. Rate (Tasas de mortalidad por cáncer de pulmón y bronquios), SMOK (Prevalencia del tabaquismo estandarizado por edad), POV (Tasa de pobreza estandarizada por edad y por condado), NO2 (Dióxido de nitrógeno), PM25 (Partículas PM2.5) y SO2 (Dióxido de azufre)

Los resultados del correlograma indican que la variable Rate (Tasas de mortalidad por cáncer de pulmón y bronquios) es significativa y está positivamente relacionada con las

variables SMOK (Prevalencia del tabaquismo estandarizado por edad) (0,63; p-valor<0,05) y POV (Tasa de pobreza estandarizada por edad y por condado) (0,50; p-valor<0,05). Sin embargo, la correlación con la variable NO2 (Dióxido de nitrógeno) parece ser negativa (-0,29; p-valor<0,05), mientras que no muestra una asociación significativa con PM25 (Partículas PM2.5) (0,1; p-valor>0,05) ni con SO2 (Dióxido de azufre) (-0,08; p-valor>0,05). Sin embargo, hay que tener en cuenta las limitaciones descriptivas de los coeficientes de correlación simple, ya que muestran la relación entre un par de variables, sin controlar las otras, por lo que, los coeficientes de correlación pueden producir relaciones ficticias en presencia de variables de confusión.

Los estadísticos GW de resumen (GWSS, por sus iniciales en inglés) son un elemento básico que debe preceder a la aplicación de cualquier modelo GW (Gollini et al., 2013). En la estimación de los GWSS intervienen varios aspectos, uno de ellos es la función de ponderación espacial (Fotheringham et al., 2002), es decir, la matriz de ponderación utilizada, en la cual intervienen tres aspectos fundamentales: la distancia, la función Kernel y el ancho de banda.

Aunque Gollini et al. (2013) recomiendan experimentar con diferentes funciones Kernel, en nuestro ejemplo utilizamos la información presentada en la Figura 46 para la estimación de los estadísticos de resumen GWSS.

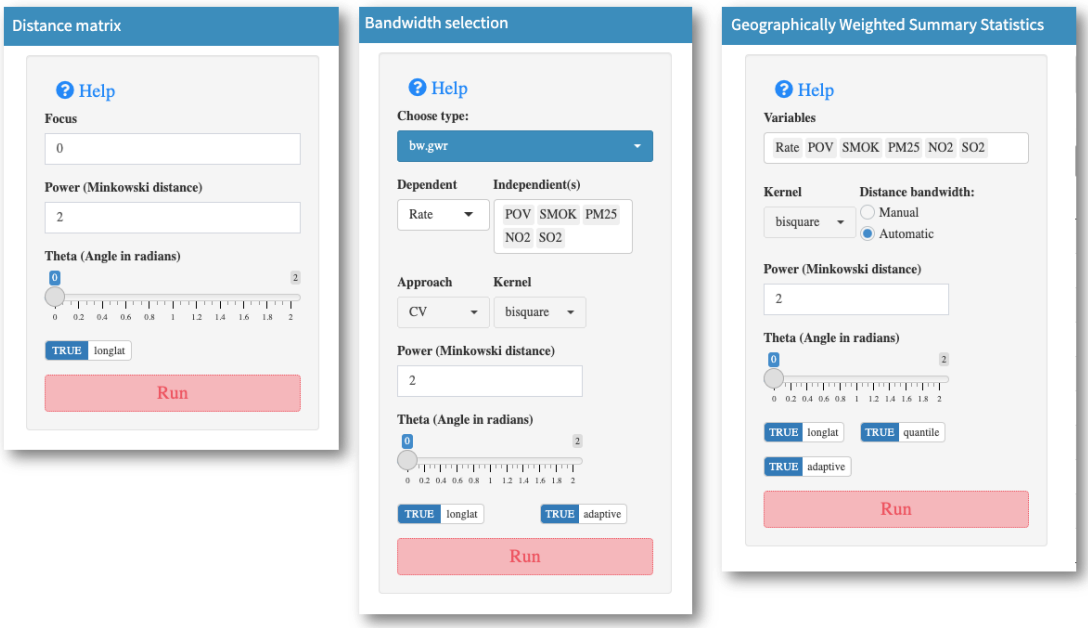


Figura 46. Opciones de configuración utilizadas en el ejemplo para GWSS

Teniendo en cuenta las opciones de calibración anterior, se presentan los resultados numéricos de los estadísticos mencionados en la Tabla 13 (todos estos estadísticos pueden mapearse). Las visualizaciones de los mapas con las medidas de tendencia central GW (básicas y robustas), medidas de variabilidad GW (básicas y robustas) y correlaciones GW (Pearson), para algunas de las variables consideradas (Figuras 47, 48 y 49 respectivamente).

Los estadísticas locales estimados se interpretan de la misma forma que los estadísticos en las estimaciones globales. En este sentido, por ejemplo, para la variable SO2, se observa que hubo mayor variabilidad en los condados ubicados en Noreste (Figura 48).

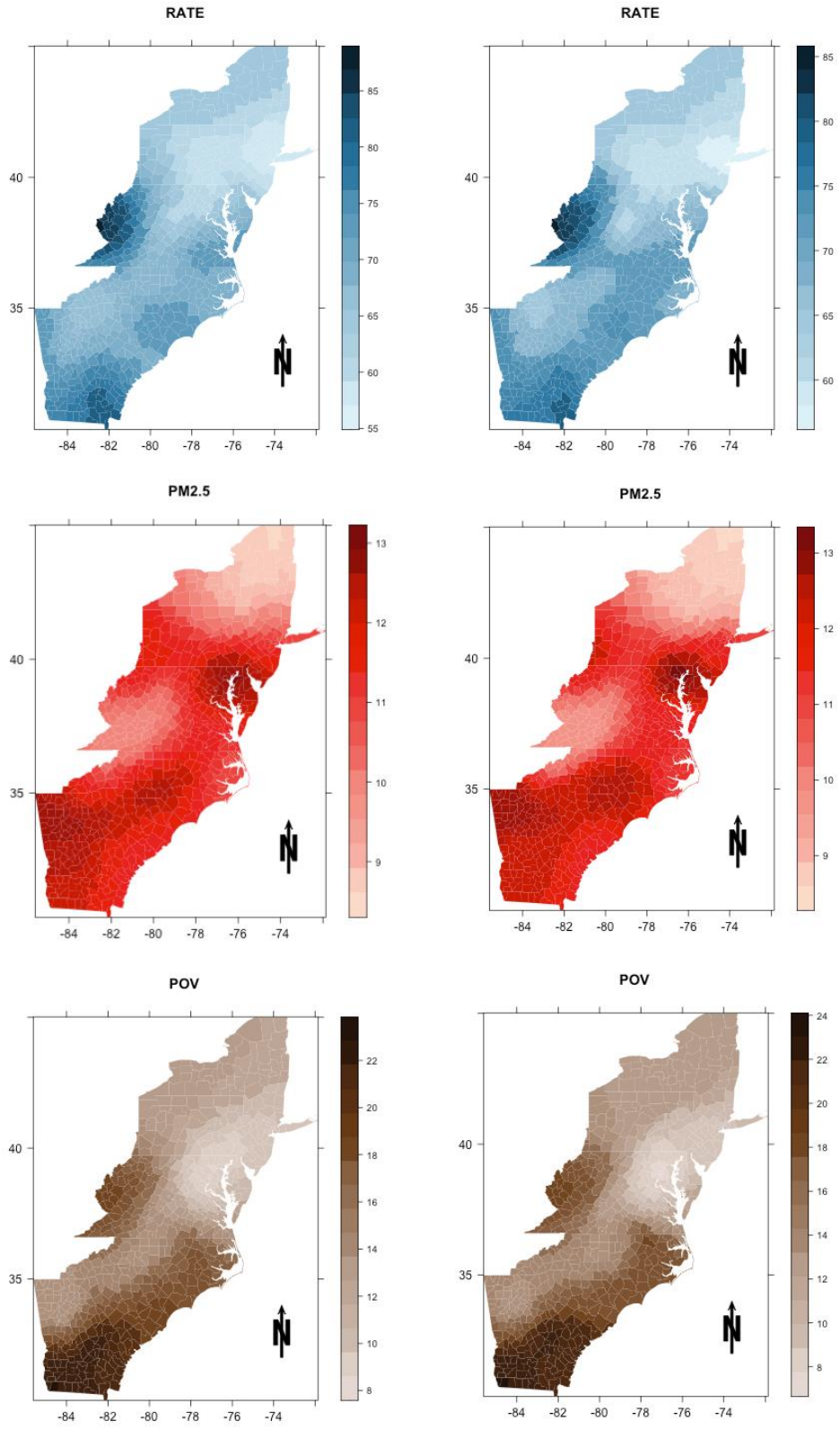


Figura 47. Mapas de las medidas de tendencia central GW locales. Para todas las variables, el mapa de la izquierda corresponde a la estimación de la media de GW y el mapa de la derecha a la estimación de la mediana de GW. Rate (Tasas de mortalidad por cáncer de pulmón y bronquios), POV (Tasa de pobreza estandarizada por edad y por condado) y PM25 (Partículas PM2.5).

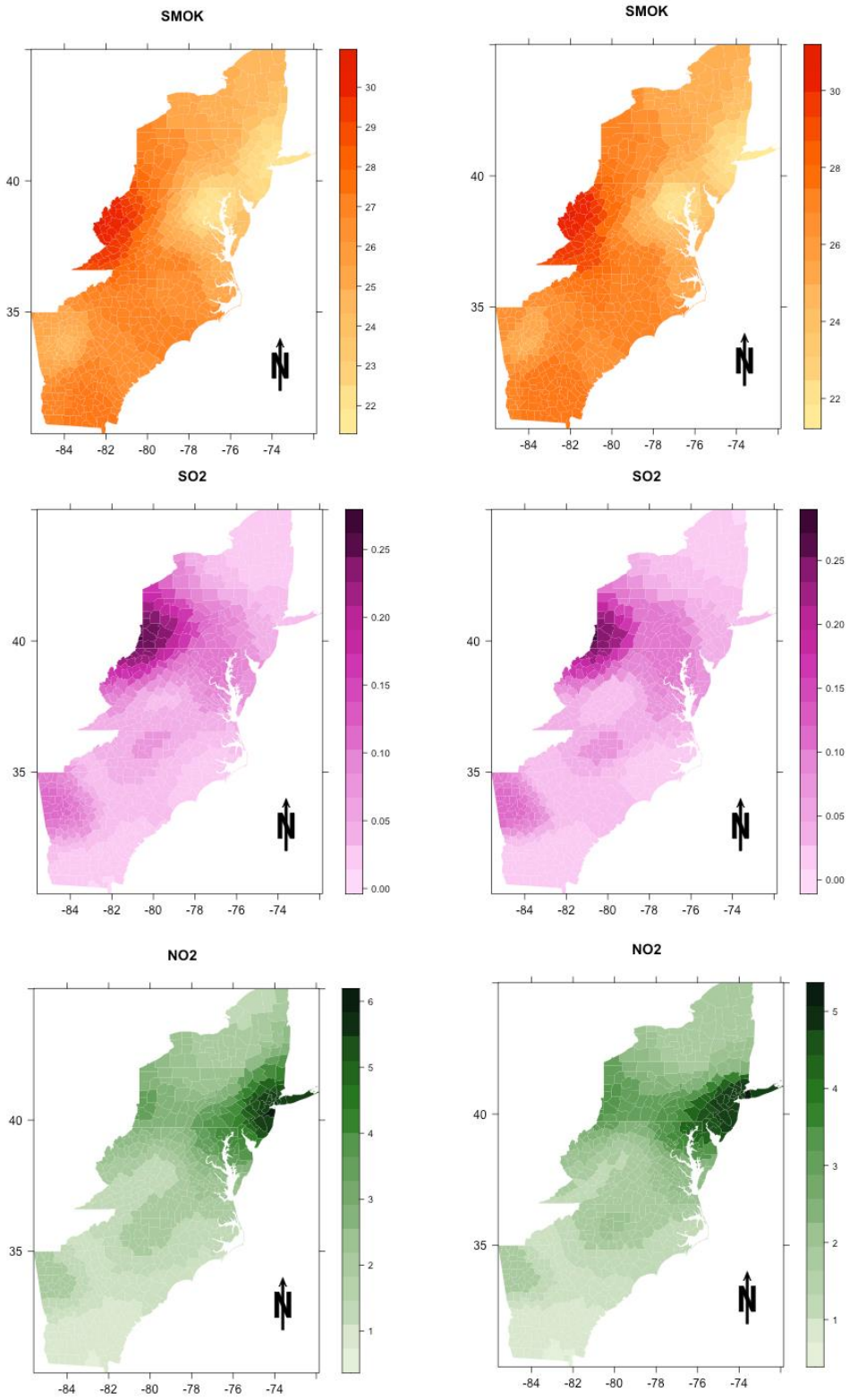


Figura 47 (continuación). Mapas de las medidas de tendencia central GW locales. Para todas las variables, el mapa de la izquierda corresponde a la estimación de la media de GW y el mapa de la derecha a la estimación de la mediana de GW. SMOK (Prevalencia del tabaquismo estandarizado por edad), NO2 (Dióxido de nitrógeno) y SO2 (Dióxido de azufre)

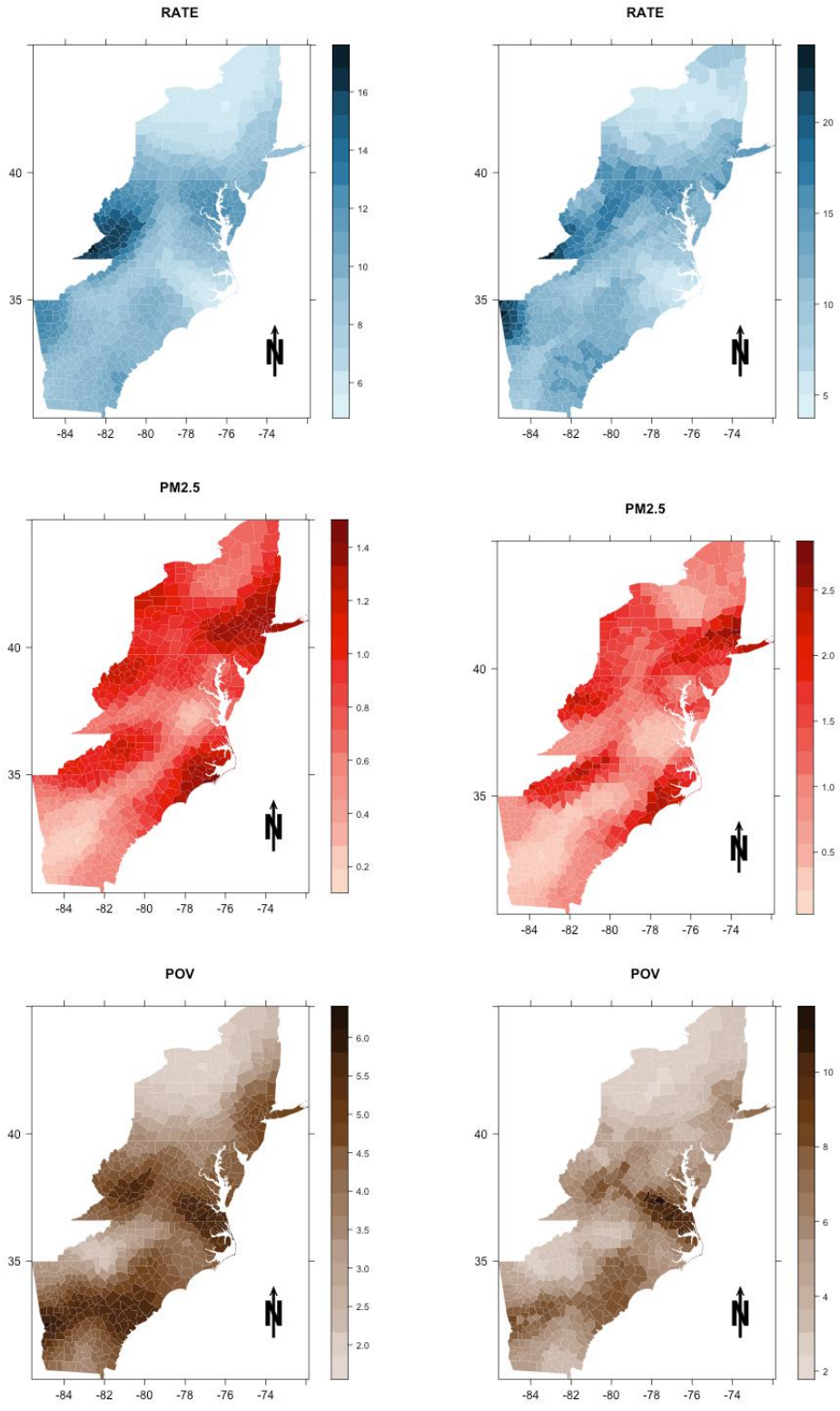


Figura 48. Mapas de las medidas de variabilidad GW locales. Para todas las variables, el mapa de la izquierda corresponde a la estimación de la desviación estándar GW y el mapa de la derecha a la estimación del rango intercuartílico GW. Rate (Tasas de mortalidad por cáncer de pulmón y bronquios), PM25 (Partículas PM2.5) y POV (Tasa de pobreza estandarizada por edad y por condado)

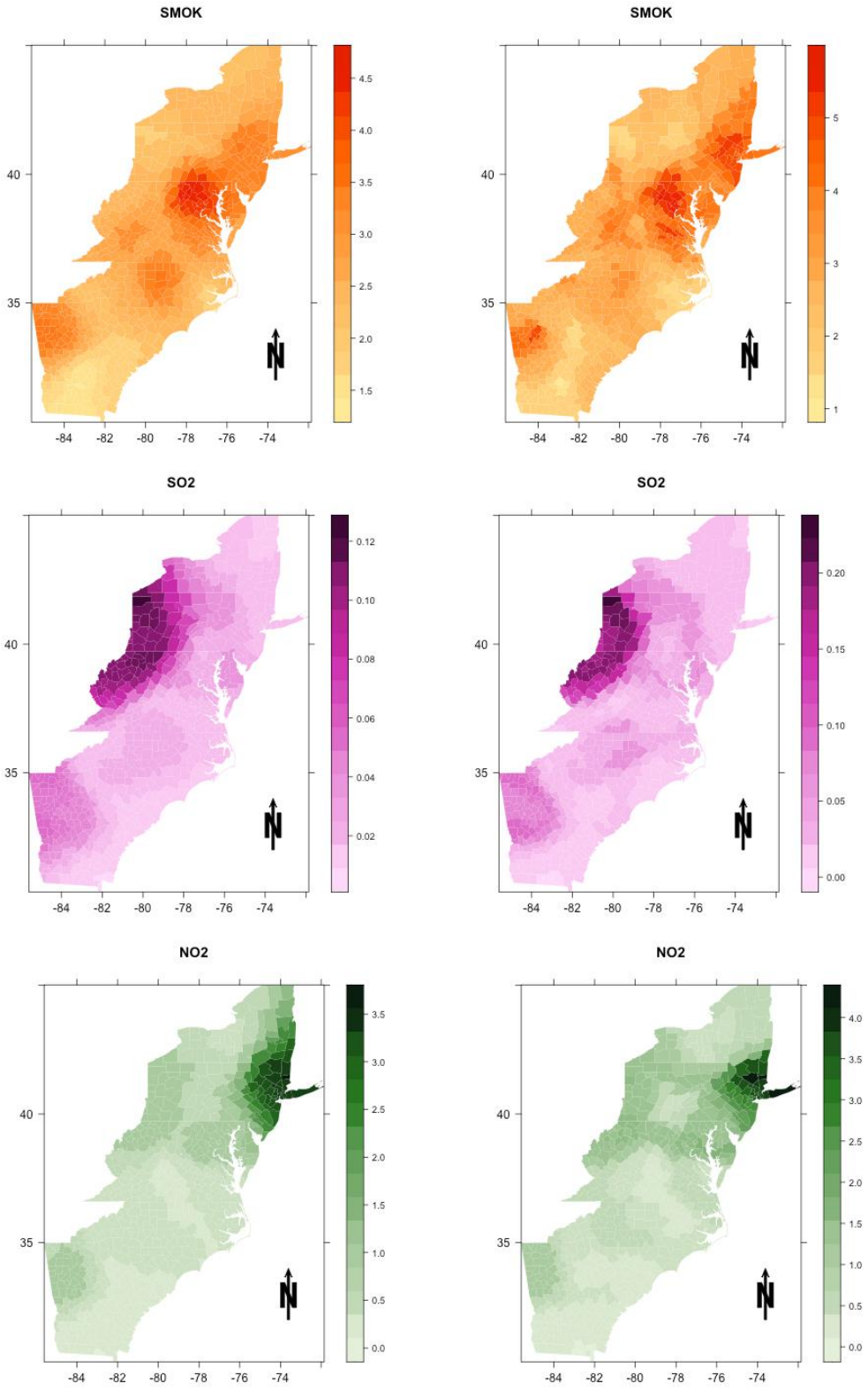


Figura 48 (continuación). Mapas de las medidas de variabilidad GW locales. Para todas las variables, el mapa de la izquierda corresponde a la estimación de la desviación estándar GW y el mapa de la derecha a la estimación del rango intercuartílico GW. SMOK (Prevalencia del tabaquismo estandarizado por edad), NO2 (Dióxido de nitrógeno) y SO2 (Dióxido de azufre)

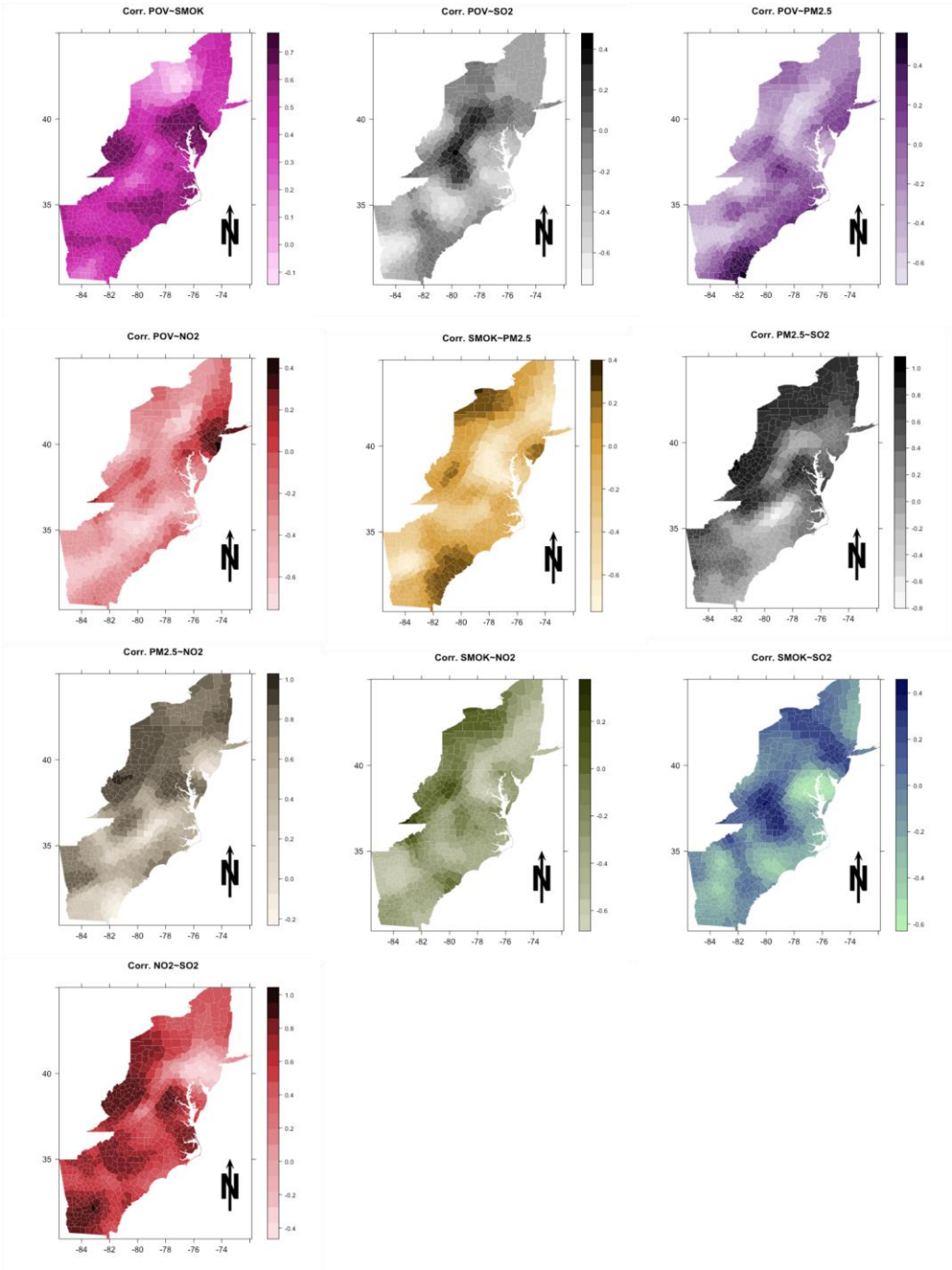


Figura 49. Coeficientes de correlación local de Pearson entre pares de variables. SMOK (Prevalencia del tabaquismo estandarizado por edad), POV (Tasa de pobreza estandarizada por edad y por condado), NO2 (Dióxido de nitrógeno), PM25 (Partículas PM2.5) y SO2 (Dióxido de azufre)

5.3.3. Ejemplo: Diagnóstico de Colinealidad Local

La colinealidad entre las variables independientes de cualquier modelo de regresión puede provocar una pérdida de potencia y precisión en las estimaciones de los coeficientes (Wheeler, 2007). Entre los posibles diagnósticos que se pueden utilizar con GeoWeightedModel encontramos cuatro que se consideran componentes esenciales de un conjunto de herramientas analíticas que deben usarse en cualquier análisis GWR, a saber: Coeficientes de Correlación (CC) entre pares de variables independientes, proporciones de descomposición de la varianza local (VDP), factores de inflación de varianza local (VIF) para cada variable independiente y números de condición local (CN, por sus iniciales en inglés) (Wheeler & Tiefelsdorf, 2005). Dado que estos criterios de diagnóstico son los habitualmente utilizados en cualquier estudio sobre colinealidad, los detalles sobre cómo usarlos, así como las posibles soluciones y discusiones críticas sobre el tema, se pueden consultar, por ejemplo, en Wheeler & Tiefelsdorf (2005) y Wheeler (2007).

En este ejemplo, nuestro objetivo es simplemente presentar cómo utilizar algunos de los diagnósticos de colinealidad implementados en el programa. Cada medida local tiene una regla general que determina si la colinealidad es un problema o no: la colinealidad se advierte mediante coeficientes de correlación > 0,8 o < -0,8, VIF superior a 10, CN superior a 30 y VDP superior a 0,5. Estos criterios no son absolutos, y obtener números más bajos no implica que la colinealidad no sea problemática, ni encontrar valores más grandes implica que la colinealidad lo sea (Golline et al., 2015).

Las entradas utilizadas en el ejemplo se muestran en la Figura 50.

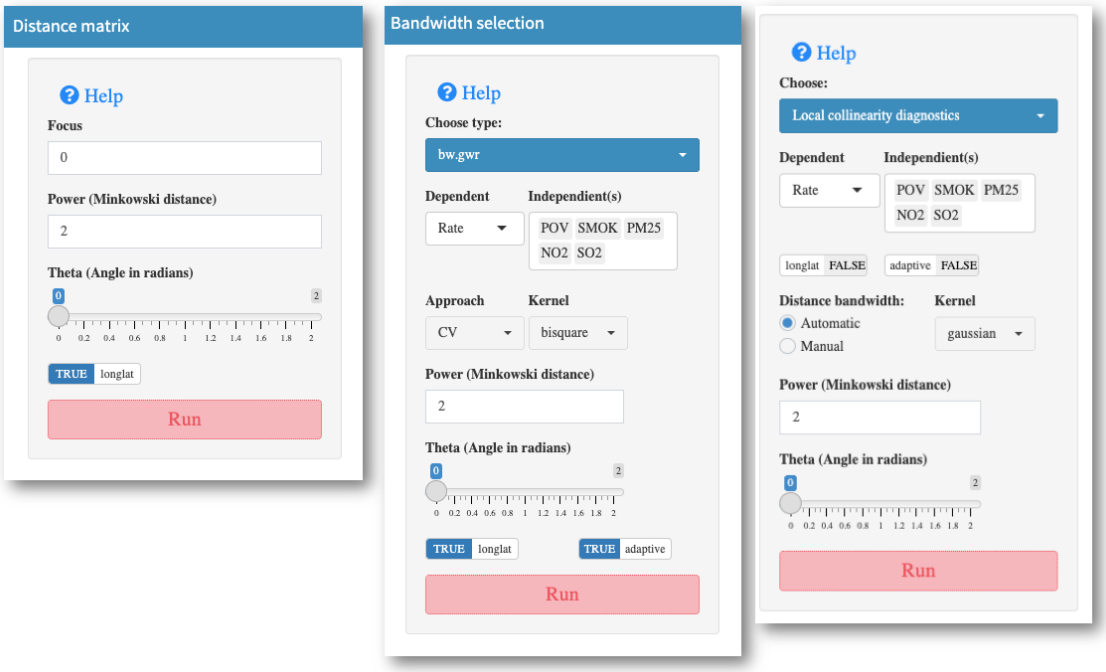


Figura 50. Configuración utilizada en el ejemplo para el diagnóstico de colinealidad local

Claramente, la colinealidad está presente en el modelo de regresión GW de nuestro ejemplo, donde encontramos, para los pares de variables NO2-SO2, PM2.5-NO2 y PM2.5-SO2, coeficientes de correlación locales mayores de 0,8 (en concreto, 197, 131, 84 valores de coeficientes de correlación locales > 0,8 para dichos pares de variables, respectivamente) (Figura 51).

La Figuras 52 y 53 muestran la distribución de los VIF para cada una de las variables del modelo GWR que nos sirve de ejemplo y muestran que hay algunas variables locales con VIF superiores al umbral 10. Se encontraron 39, 35 y 23 VIF mayores de 10 para las variables NO2, PM2.5 y SO2, respectivamente.

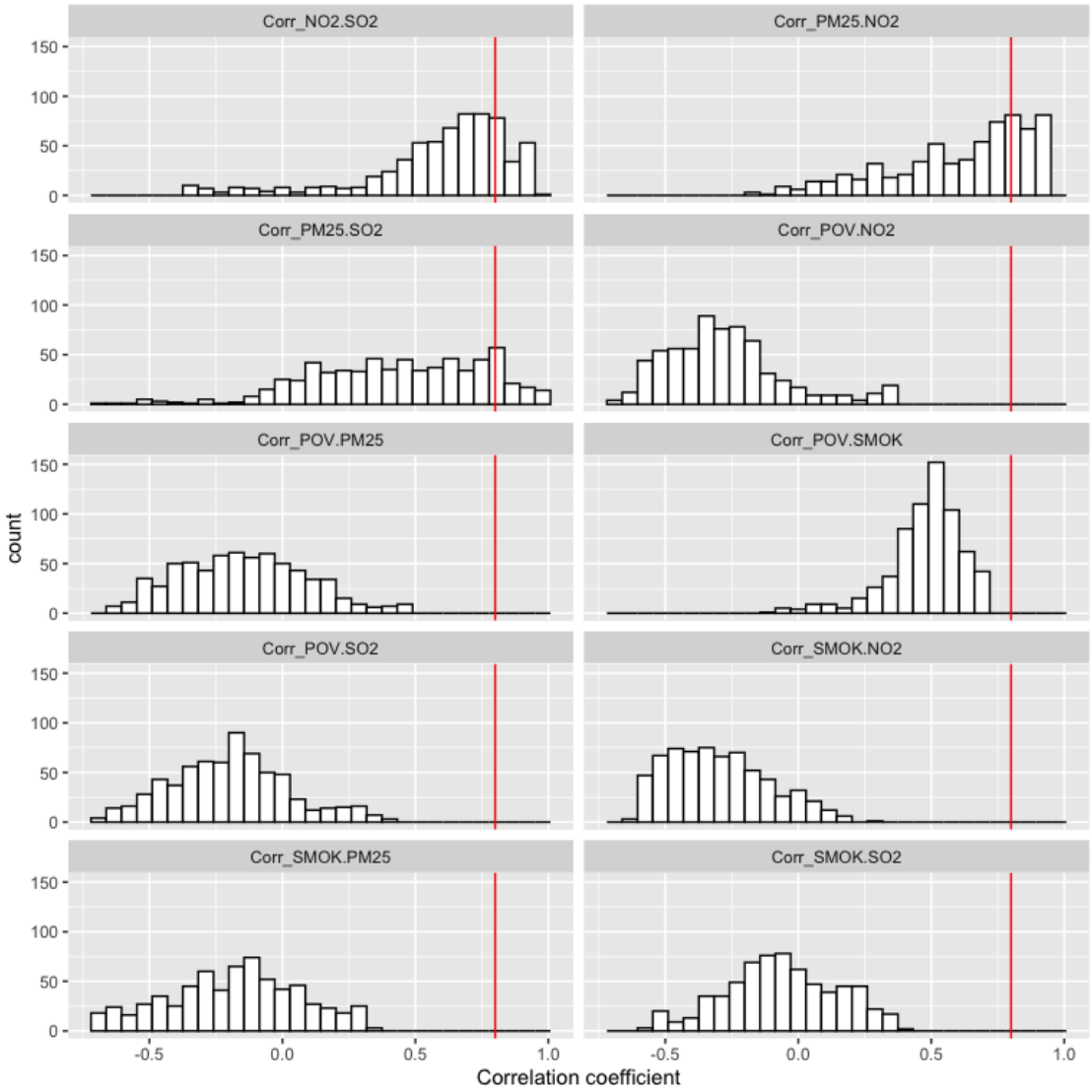


Figura 51. Distribución de los coeficientes de correlación locales entre pares de variables

La presencia de proporciones de descomposición de la varianza (VDP) superiores a 0,5 indica que existe colinealidad entre al menos dos términos de regresión, uno de los cuales puede ser el intercepto (Figuras 54 y 55).

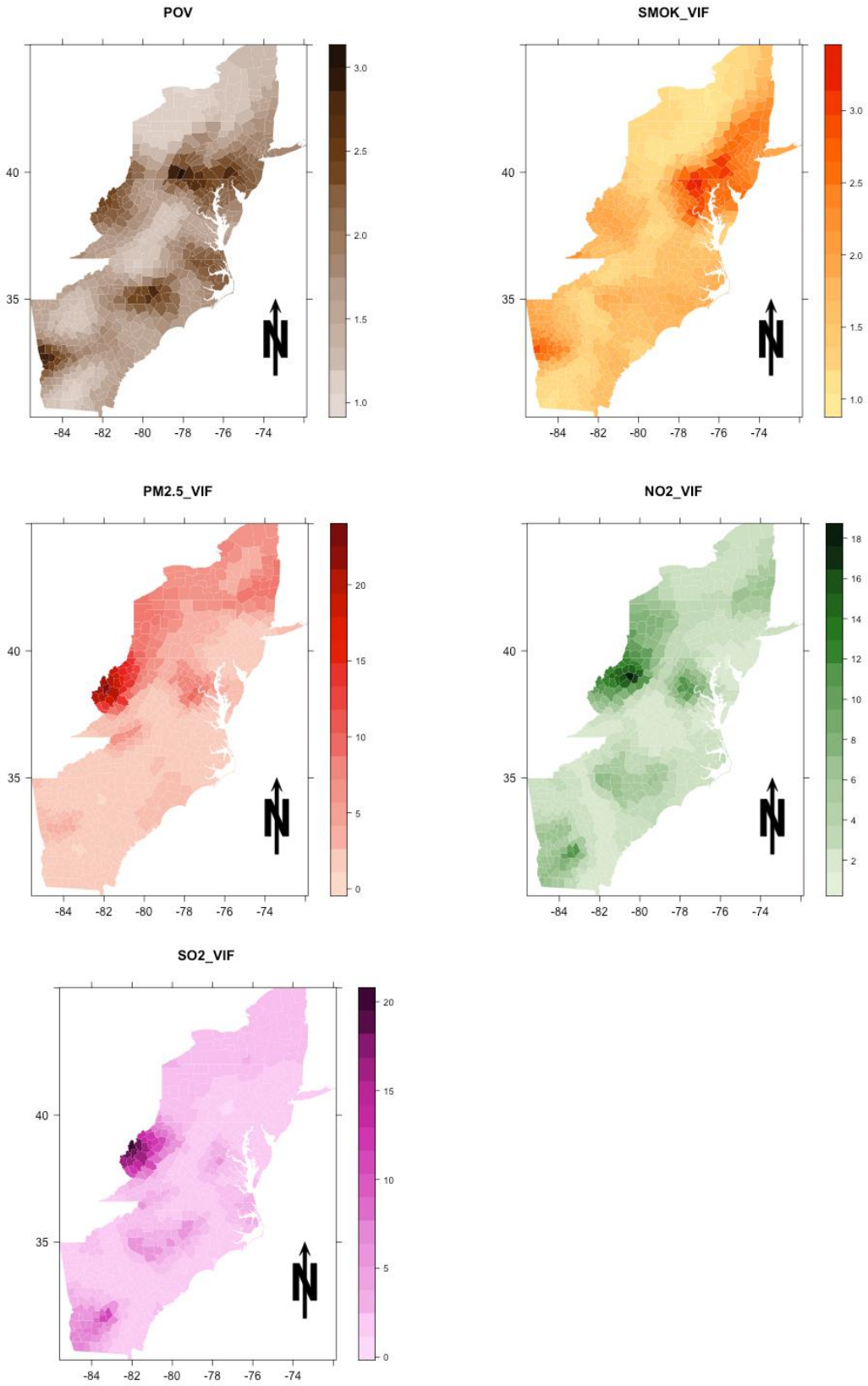


Figura 52. Factores de inflación de varianza locales (VIF). SMOK (Prevalencia del tabaquismo estandarizado por edad), POV (Tasa de pobreza estandarizada por edad y por condado), NO2 (Dióxido de nitrógeno), PM25 (Partículas PM2.5) y SO2 (Dióxido de azufre)

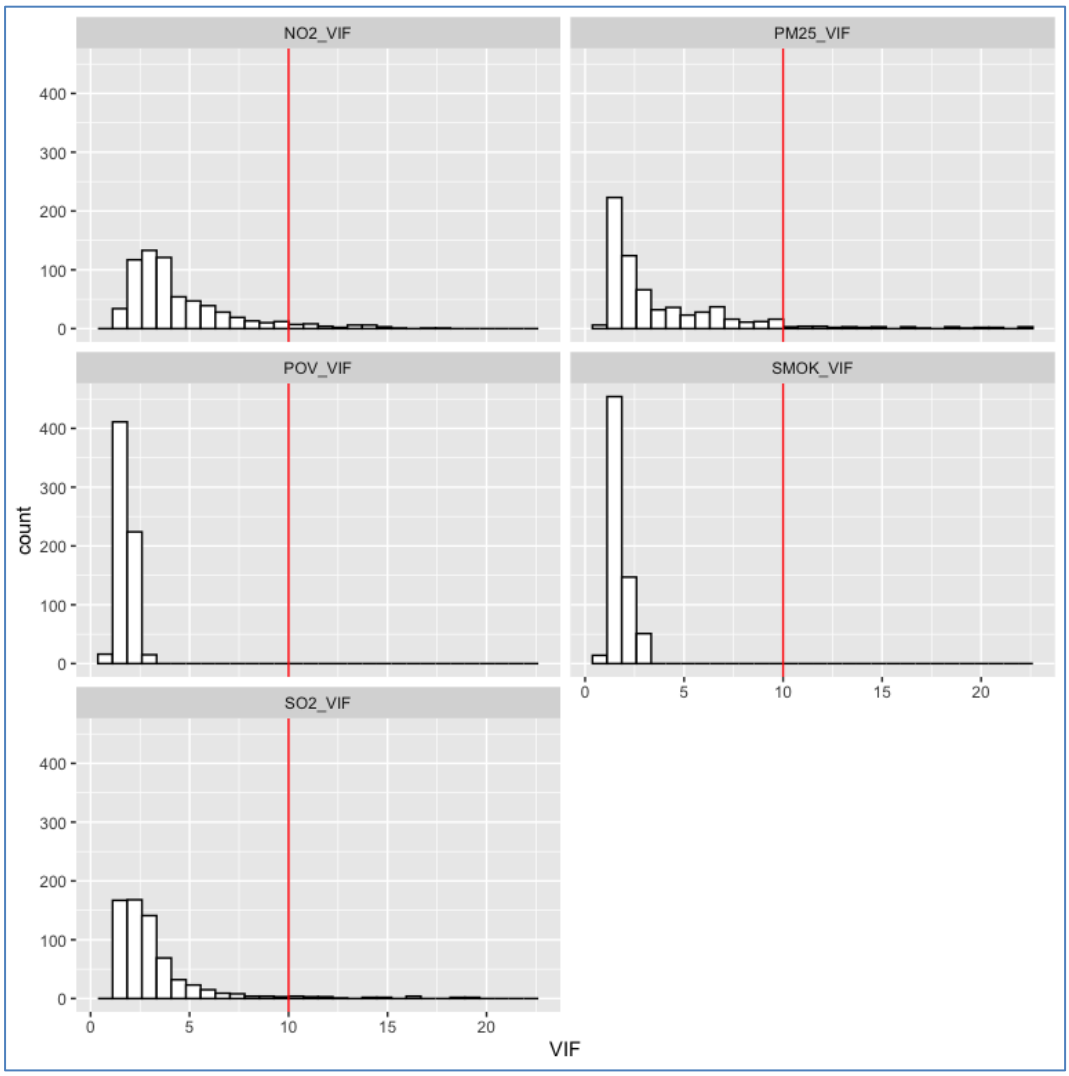


Figura 53. Distribución de los Factores de Inflación de Varianza local (VIF) para las diferentes variables. Rate (Tasas de mortalidad por cáncer de pulmón y bronquios), SMOK (Prevalencia del tabaquismo estandarizado por edad), POV (Tasa de pobreza estandarizada por edad y por condado), NO2 (Dióxido de nitrógeno), PM25 (Partículas PM2.5) y SO2 (Dióxido de azufre).

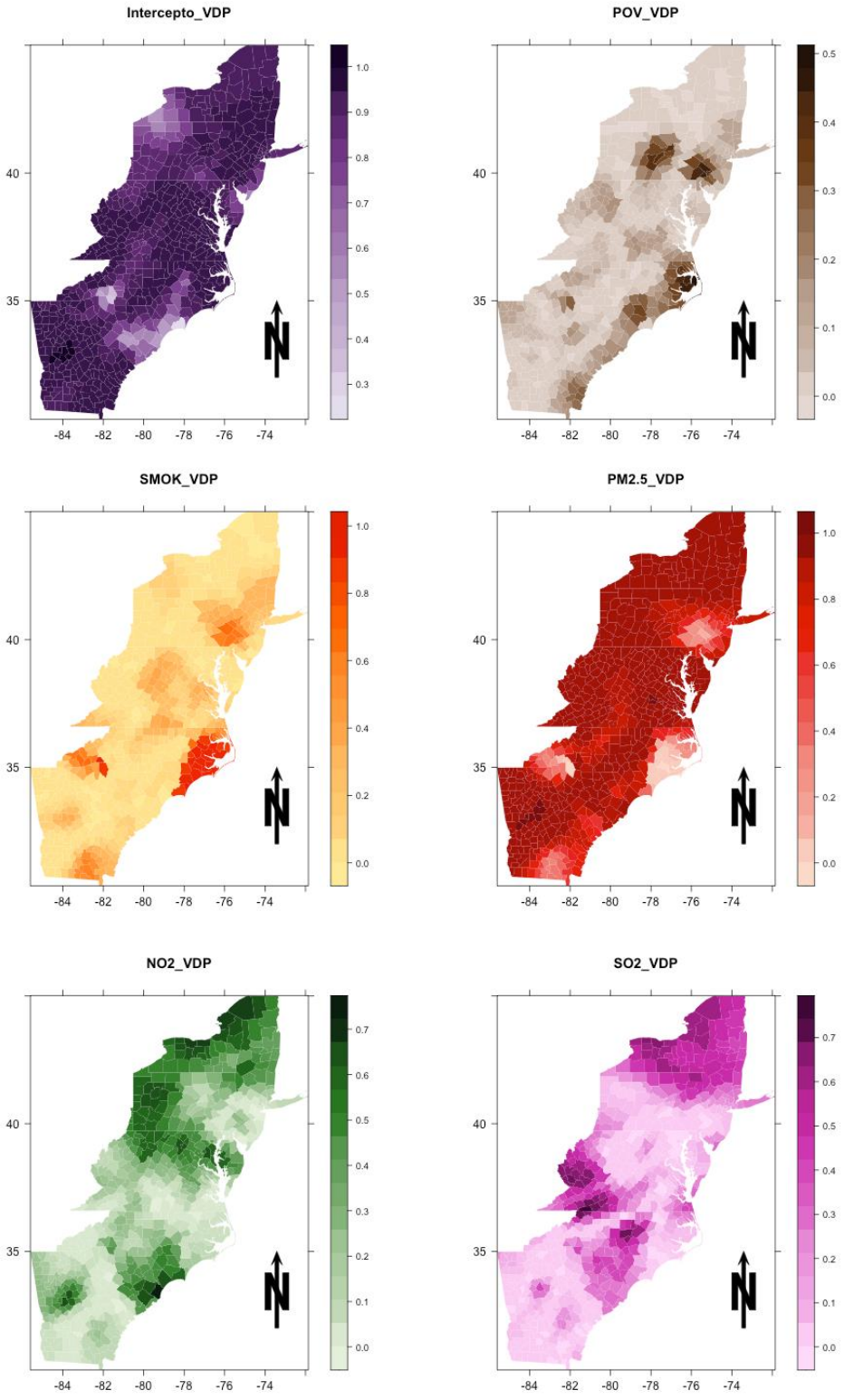


Figura 54. Proporciones de descomposición de la varianza local (VDP). Rate (Tasas de mortalidad por cáncer de pulmón y bronquios), SMOK (Prevalencia del tabaquismo estandarizado por edad), POV (Tasa de pobreza estandarizada por edad y por condado), NO2 (Dióxido de nitrógeno), PM25 (Partículas PM2.5) y SO2 (Dióxido de azufre)

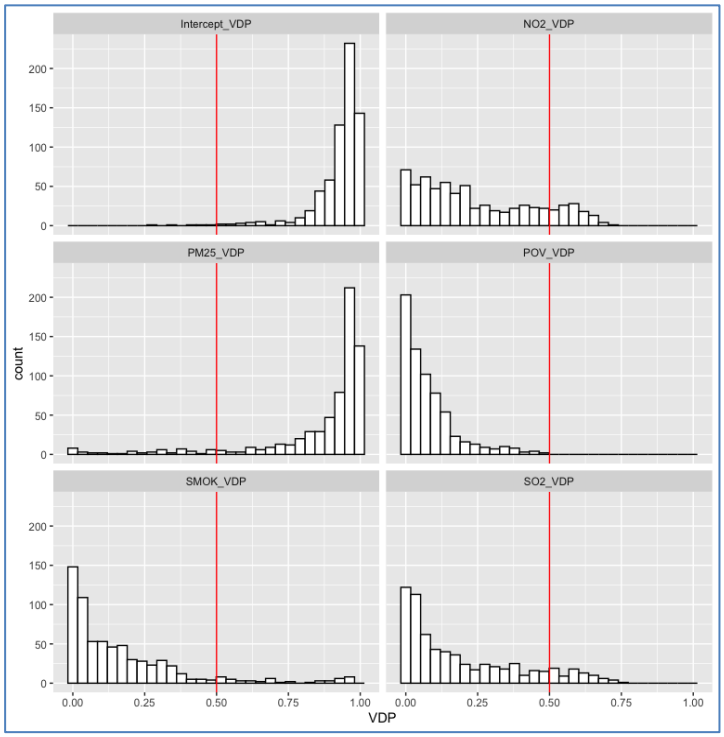


Figura 55. Distribución de las proporciones de descomposición de la varianza local (VDP). Rate (Tasas de mortalidad por cáncer de pulmón y bronquios), SMOK (Prevalencia del tabaquismo estandarizado por edad), POV (Tasa de pobreza estandarizada por edad y por condado), NO2 (Dióxido de nitrógeno), PM25 (Partículas PM2.5) y SO2 (Dióxido de azufre)

El CN local osciló entre 38 y 391, todos superiores al valor de referencia 30. Si bien las estimaciones locales son potencialmente más susceptibles a la colinealidad que el modelo global, la eliminación de uno o más predictores suele ser una solución sencilla. El desafío es determinar qué predictor(es) excluir, especialmente cuando se cree que todos son relevantes para caracterizar el proceso de estudio. Una regresión penalizada puede proporcionar una solución completa en este caso (Zou & Hastie, 2005). Los números de condición local para la estimación del modelo de ejemplo se muestran en la Figura 56.

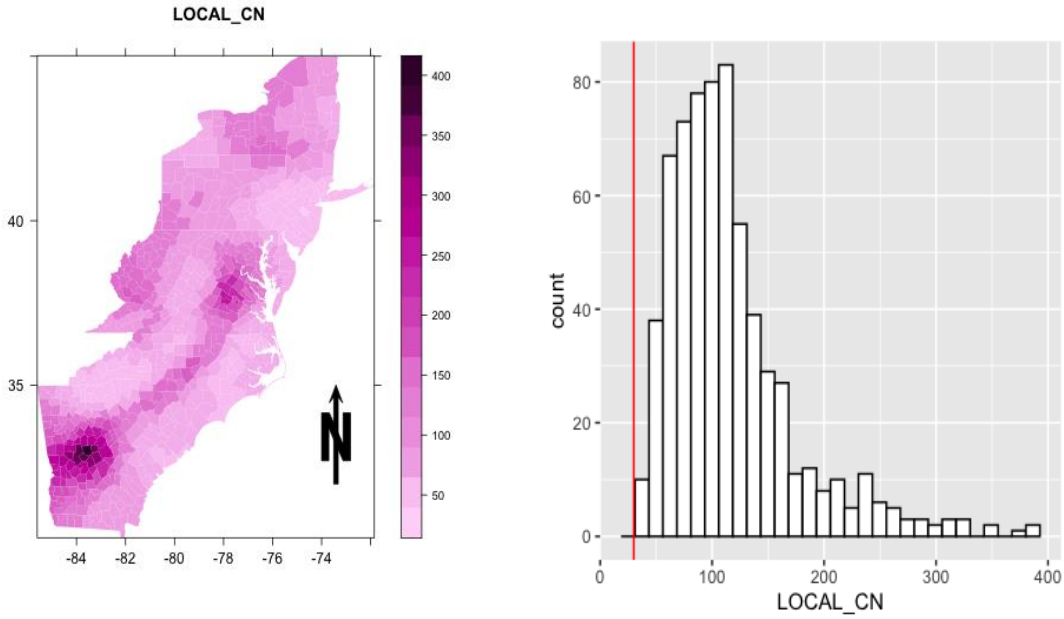


Figura 56. Números de condición locales (CN)

5.3.4. Ejemplo: GWR básica

Llevar a cabo un análisis de GWR debe ser justificado en términos de los objetivos del análisis y las características de los datos. Si los efectos espaciales son visibles en los datos, se puede considerar una GWR. Sin embargo, esto requiere demostrar que los modelos alternativos, específicamente aquellos con coeficientes fijos, son insuficientes. Comber et al. (2020) brindan orientación para la aplicación confiable de GWR.

Los pasos para ajustar un modelo de GWR son, primero, calcular una matriz de distancias que contenga las distancias entre puntos, seguido de la calibración del ancho de banda para el ajuste del modelo local (también puede especificar un ancho de banda el propio investigador), luego se ajusta el modelo y por último se busca la evidencia de variaciones espaciales.

Las entradas utilizadas para la estimación de la matriz de distancias con los datos de nuestro ejemplo, selección de Ancho de Banda y las opciones básicas de la GWR se muestran en la Figura 57.

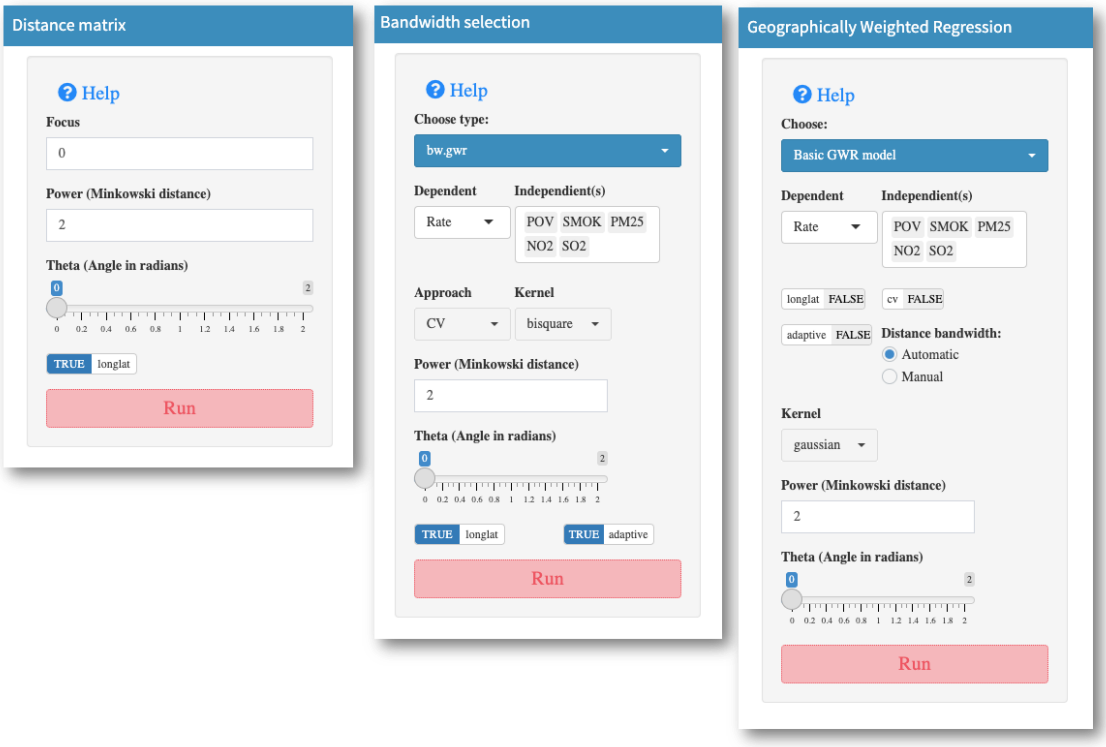


Figura 57. Configuración utilizada en el ejemplo para GWR básico

Para centrarnos en la descripción de la GWR, mantenemos nuestro análisis simple y estudiamos la Rate (Tasas de mortalidad por cáncer de bronquios y pulmón ajustadas por edad por condado) como una función de las variables SMOK (Prevalencia del tabaquismo estandarizado por edad), POV (Tasa de pobreza estandarizada por edad y por condado), NO2 (Dióxido de nitrógeno), PM25 (Partículas PM2.5) y SO2 (Dióxido de azufre).

Los resultados de la regresión global indican que existe una relación positiva entre Rate y las variables SMOK, PM2.5 y POV (esta última no es estadísticamente significativa), mientras que existe una relación inversamente proporcional con las variables NO2 y SO2 (aunque esta última tampoco es estadísticamente significativa). El valor del coeficiente de determinación (*Adjusted R-squared*) para la regresión es 0,4945, lo que indica que nuestro modelo explica el 49,45 % de la variabilidad de la variable Rate. Esto deja el 50,55% de la variabilidad sin explicación. Parte de esta varianza no explicada puede deberse a que el modelo de regresión asume que las relaciones en el modelo son constantes en el espacio, es decir, asume un proceso estacionario. En los resultados también encontramos: *Residual sum of squares*: 48636,74; *Sigma(hat):* 8,558511; *AIC:* 4761,729; *AICc*: 4761,899; *BIC*: 4172,747 que son valores que nos permiten comparar los modelos (Figura 58).

Evaluando el ajuste local del modelo se observan mejoras. El *Residual sum of squares* experimentó una reducción (39540,59), mientras que *R-square value* (0,592) aumentó.

Si bien los mapas de la Figura 59 brindan información útil sobre los patrones espaciales de las relaciones, no indican si estas relaciones son estadísticamente significativas o no. Es posible que algunas no lo sean. Aproximadamente, si la estimación de un coeficiente tiene un valor absoluto de t superior a 1,96 y la muestra es lo suficientemente grande, entonces es estadísticamente significativa. La Figura 60 muestra el valor t estimado para cada variable y el intercepto. Para SMOK, el 94,1% (627 condados) de todos los coeficientes locales son estadísticamente significativos, estos porcentajes calculados para PM2.5, NO2, POV y SO2 son respectivamente 38,9% (259 condados), 34,8% (232 condados), 31,8% (212 condados) y 25,2% (168 condados).

```
*********************************************************************
*                   Results of Global Regression                    *
*********************************************************************
## Call:
## lm(formula = Rate ~ POV + SMOK + PM25 + NO2 + SO2, data = DataExample)
## Residuals:
##     Min      1Q  Median      3Q     Max
## -25.237  -5.598  -0.400   4.992  33.543
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -6.58106    4.72460  -1.393  0.16411
## POV           0.15157    0.07771   1.950  0.05155 .
## SMOK          2.24165    0.13479  16.630  < 2e-16 ***
## PM25          1.60513    0.28176   5.697 1.84e-08 ***
## NO2          -0.75917    0.25448  -2.983  0.00296 **
## SO2         -10.97774    5.65825  -1.940  0.05279 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.584 on 660 degrees of freedom
## Multiple R-squared:  0.4983, Adjusted R-squared:  0.4945
## F-statistic: 131.1 on 5 and 660 DF,  p-value: < 2.2e-16
  ***Extra Diagnostic information
  Residual sum of squares: 48636.74
  Sigma(hat): 8.558511
  AIC:  4761.729
  AICc:  4761.899
  BIC:  4172.747
************************Diagnostic information*************************
  Number of data points: 666
  Effective number of parameters (2trace(S) - trace(S'S)): 390.0414
  Effective degrees of freedom (n-2trace(S) + trace(S'S)): 275.9586
  AICc (GWR book, Fotheringham, et al. 2002, p. 61, eq 2.33): 6055.75
  AIC (GWR book, Fotheringham, et al. 2002,GWR p. 96, eq. 4.22): 4954.968
  BIC (GWR book, Fotheringham, et al. 2002,GWR p. 61, eq. 2.34): 6187.656
  Residual sum of squares: 39540.59
  R-square value:  0.5921478
```

Figura 58. Resultados de la regresión global e información diagnóstica del ajuste local suministrados por el paquete GeoWeightedModel

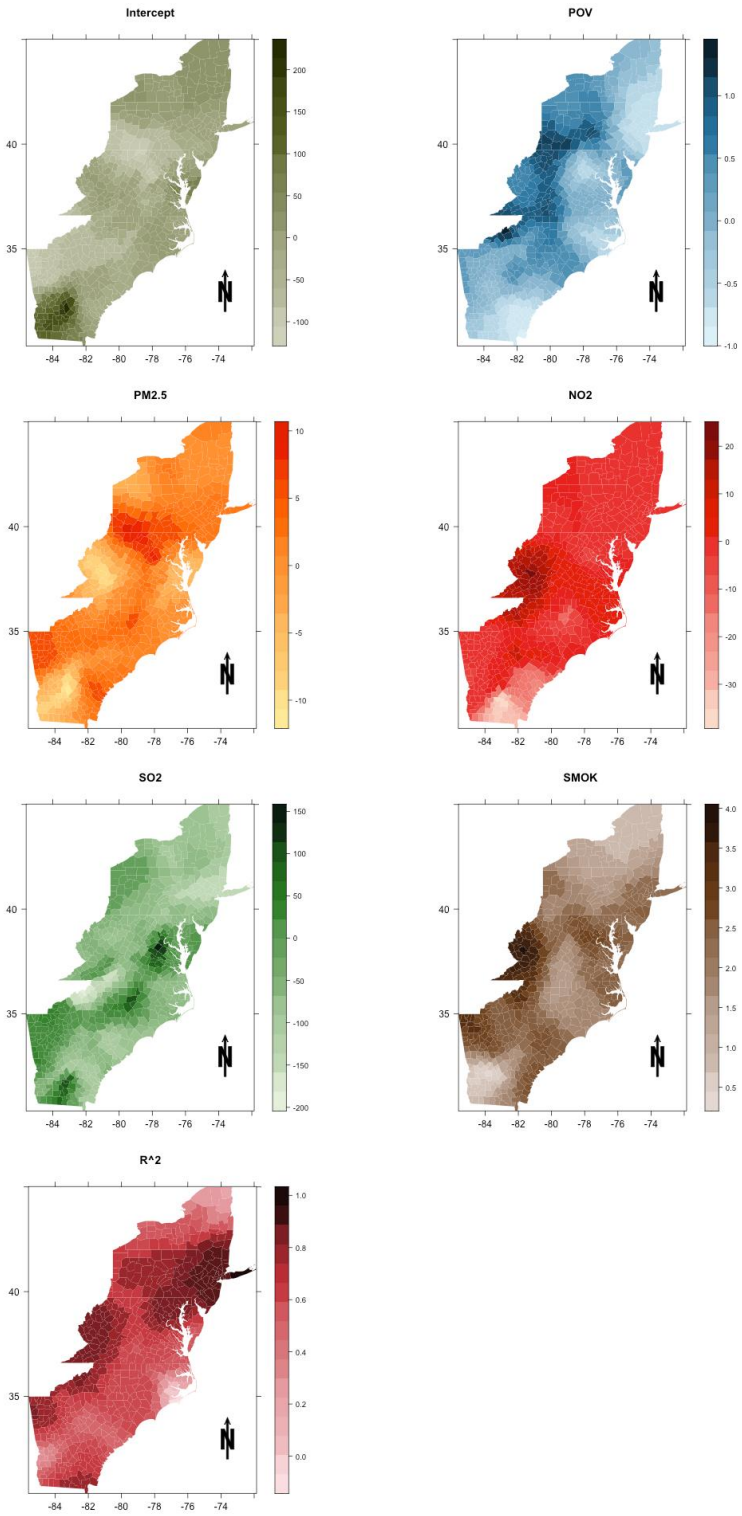


Figura 59. Estimaciones de los coeficientes de regresión GWR y coeficientes de determinación (R^2). SMOK (Prevalencia del tabaquismo estandarizado por edad), POV (Tasa de pobreza estandarizada por edad y por condado), NO2 (Dióxido de nitrógeno), PM25 (Partículas PM2.5) y SO2 (Dióxido de azufre)

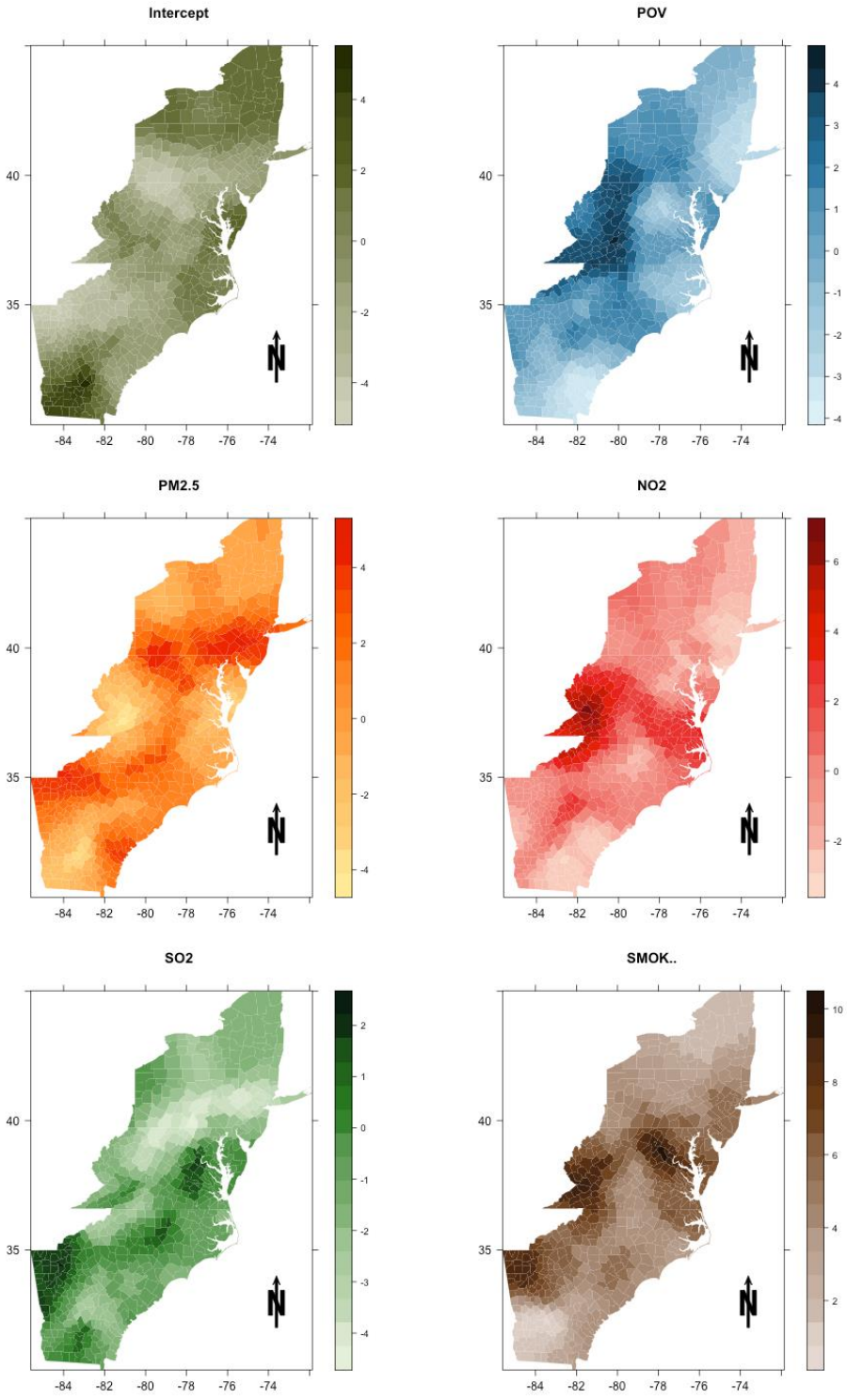


Figura 60. Valores t para las estimaciones de los coeficientes de GWR locales. SMOK (Prevalencia del tabaquismo estandarizado por edad), POV (Tasa de pobreza estandarizada por edad y por condado), NO2 (Dióxido de nitrógeno), PM25 (Partículas PM2.5) y SO2 (Dióxido de azufre)

5.3.5. Ejemplo: GWR robusta

Tanto las entradas como las salidas gráficas o numéricas son similares a las del modelo de la GWR básica. La diferencia entre este módulo y el anterior es que la GWR robusta permite identificar, y en algunos casos, reducir el efecto de valores atípicos en la GWR. En Gollini et al. (2015) se detalla y explica cómo se calculan los valores atípicos.

En el caso del ejemplo se identificaron tres outliers locales, los cuales se pueden encontrar en la salida SDF en la columna *E_wegts* (Figura 61). La eliminación de estos tres valores mejoró la estimación de R^2 y de AIC y BIC, en comparación con la GWR básica (Figura 62).

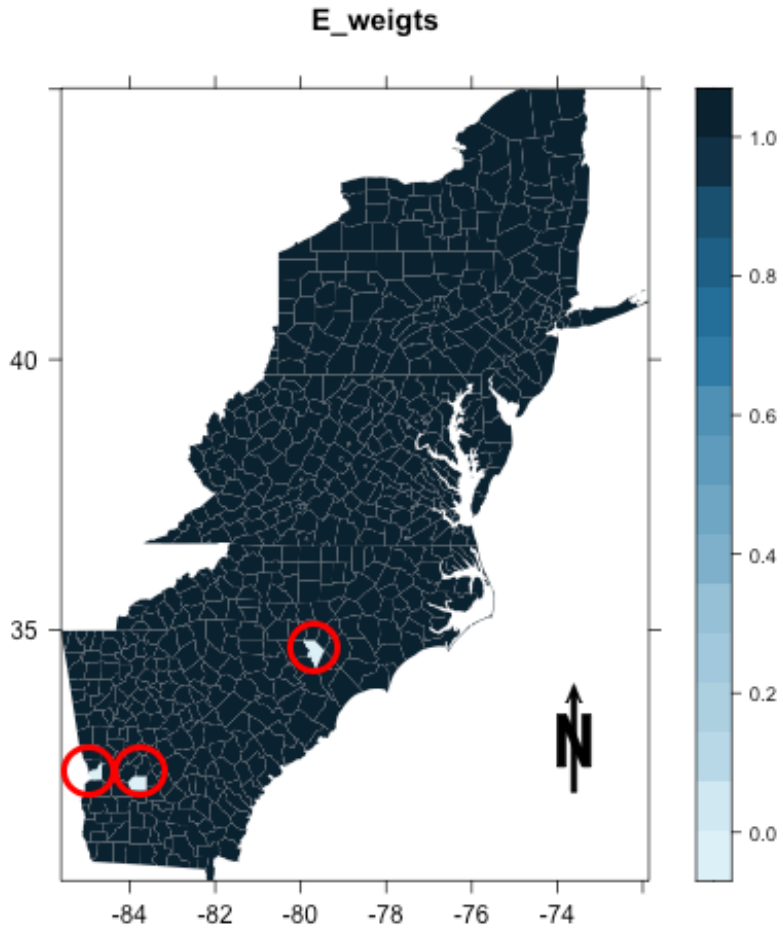


Figura 61. Valores atípicos identificados en GWR robusta

```
**********Diagnostic information**********
  Number of data points: 666
  Effective number of parameters (2trace(S) - trace(S'S)): 109.3008
  Effective degrees of freedom (n-2trace(S) + trace(S'S)): 556.6992
  AICc (GWR book, Fotheringham, et al. 2002, p. 61, eq 2.33): 4510.929
  AIC (GWR book, Fotheringham, et al. 2002,GWR p. 96, eq. 4.22): 4398.786
  BIC (GWR book, Fotheringham, et al. 2002,GWR p. 61, eq. 2.34): 4198.044
  Residual sum of squares: 25367.15
  R-square value:  0.7383436
  Adjusted R-square value:  0.6868782

  **********
```

Figura 62. Información diagnóstica de GWR robusta.

5.3.6. Ejemplo: Regresión Generalizada Ponderada Geográficamente (GGWR)

Así como los Modelos Lineales Generalizados (GLM, Nelder & Wedderburn, 1972) extienden el modelo de regresión lineal básico, Fotheringham et al. (2002) extendieron el modelo de GWR a los llamados Modelos Lineales Generalizados Ponderados Geográficamente (GGWR, por sus iniciales en inglés). Estos permiten que la variable de respuesta se distribuya según un miembro de la familia exponencial de distribuciones. En el caso del programa GeoWeightedModel podemos utilizar las distribuciones de Poisson y Binomial.

Las entradas utilizadas en esta sección se muestran en la Figura 63.

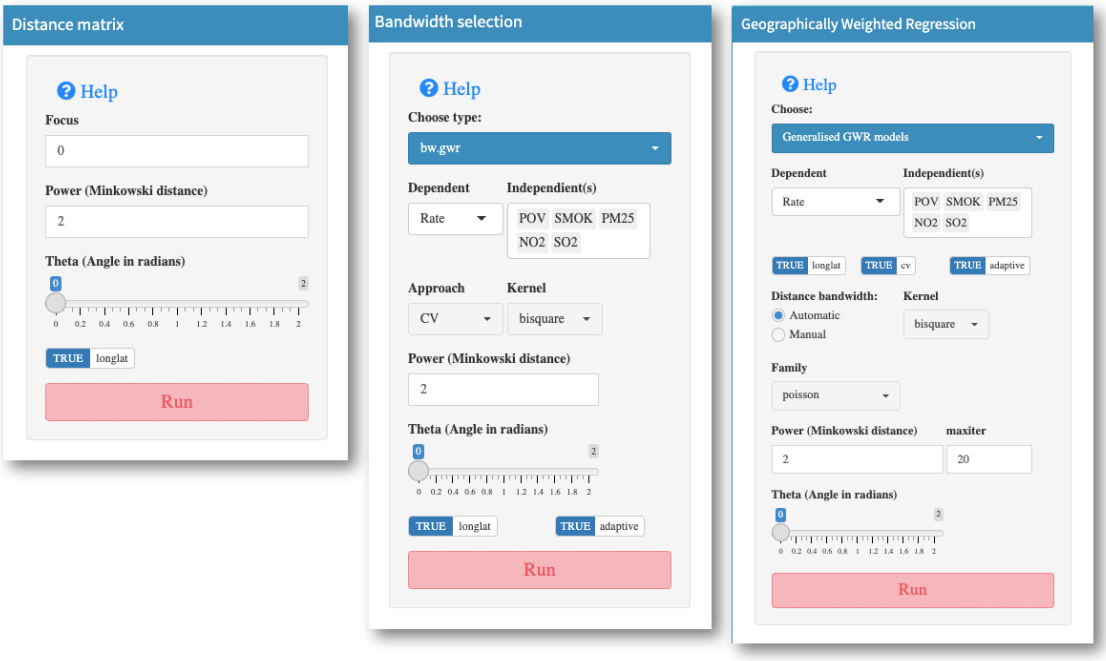


Figura 63. Configuración utilizada en el ejemplo para GGWR.

Al igual que en la GWR básica y robusta, tenemos una amplia salida numérica, que es muy útil. En la Figura 64 se proporciona el resumen del modelo de regresión global, mientras que en la Figura 65 el resumen de las estimaciones de los coeficiente de la GGWR.

Los valores diagnósticos clave como son los valores AIC y AICc fueron 670,89 y 671,02, respectivamente, para la regresión global, mientras que para el modelo GGWR estos valores son respectivamente 519,25 y 544,52. Esto indica que el GGWR resulta ser un mejor modelo. Hay que tener en cuenta que el GGWR casi siempre mejora el *Pseudo R-square* (de 0,52 a 0,74). Sin embargo, este valor nunca debe usarse como el único discriminador entre modelos. Siempre se debe informar sobre un diagnóstico de tipo AIC, ya que tiene en cuenta tanto la complejidad del modelo, como el ajuste del mismo. Otros puntos de interés son los resúmenes de los coeficientes GGWR estimados (mínimo, primer cuartíl, mediana, tercer cuartíl, máximo), que deben compararse con las estimaciones de coeficientes de la regresión global.

```
  ***********************************************************************
 *             Results of Generalized linear Regression                *
 ***********************************************************************

Call:
NULL
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.33778  -0.07926  -0.00795   0.07353   0.44683
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
Intercept  3.103325   0.068146  45.539  < 2e-16 ***
POV        0.001609   0.001092   1.474 0.140426
SMOK       0.034011   0.001954  17.402  < 2e-16 ***
PM25       0.023501   0.003966   5.926  3.1e-09 ***
NO2       -0.013945   0.003906  -3.570 0.000357 ***
SO2       -0.154462   0.079516  -1.943 0.052074 .
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1373.87  on 665  degrees of freedom
Residual deviance:  658.89  on 660  degrees of freedom
AIC: 670.89

Number of Fisher Scoring iterations: 4

 AICc:  671.0189
 Pseudo R-square value:  0.5204135
```

Figura 64. Resultados del modelo lineal generalizado global

```
**********Model calibration information**********
Kernel function: bisquare
Adaptive bandwidth: 91 (number of nearest neighbours)
Regression points: the same locations as observations are used.
Distance metric: A distance matrix is specified for this model calibration.

***********Summary of Generalized GWR coefficient estimates:**********
              Min.    1st Qu.     Median
Intercept  1.4658801  2.6351963  3.0646346
POV       -0.0126614 -0.0037604  0.0015772
SMOK       0.0059807  0.0281555  0.0339506
PM25      -0.1353828 -0.0156965  0.0196290
NO2       -0.4600854 -0.0198882  0.0023163
SO2       -2.4498812 -1.0201538 -0.5413040
           3rd Qu.   Max.
Intercept  3.4962785 6.0776
POV        0.0066516 0.0179
SMOK       0.0405401 0.0492
PM25       0.0539286 0.1400
NO2        0.0771013 0.2563
SO2        0.0869093 2.1680
**********Diagnostic information**********
Number of data points: 666
GW Deviance: 349.0459
AIC : 519.254
AICc : 544.5268
Pseudo R-square value:  0.7459404
```

Figura 65. Resumen de las estimaciones de los coeficientes GGWR e información diagnóstica del ajuste local suministrados por el paquete GeoWeightedModel.

Al mapear los coeficientes GGWR estimados podemos tener una idea más clara de cómo se desarrolla espacialmente la variación de las estimaciones (Figura 66).

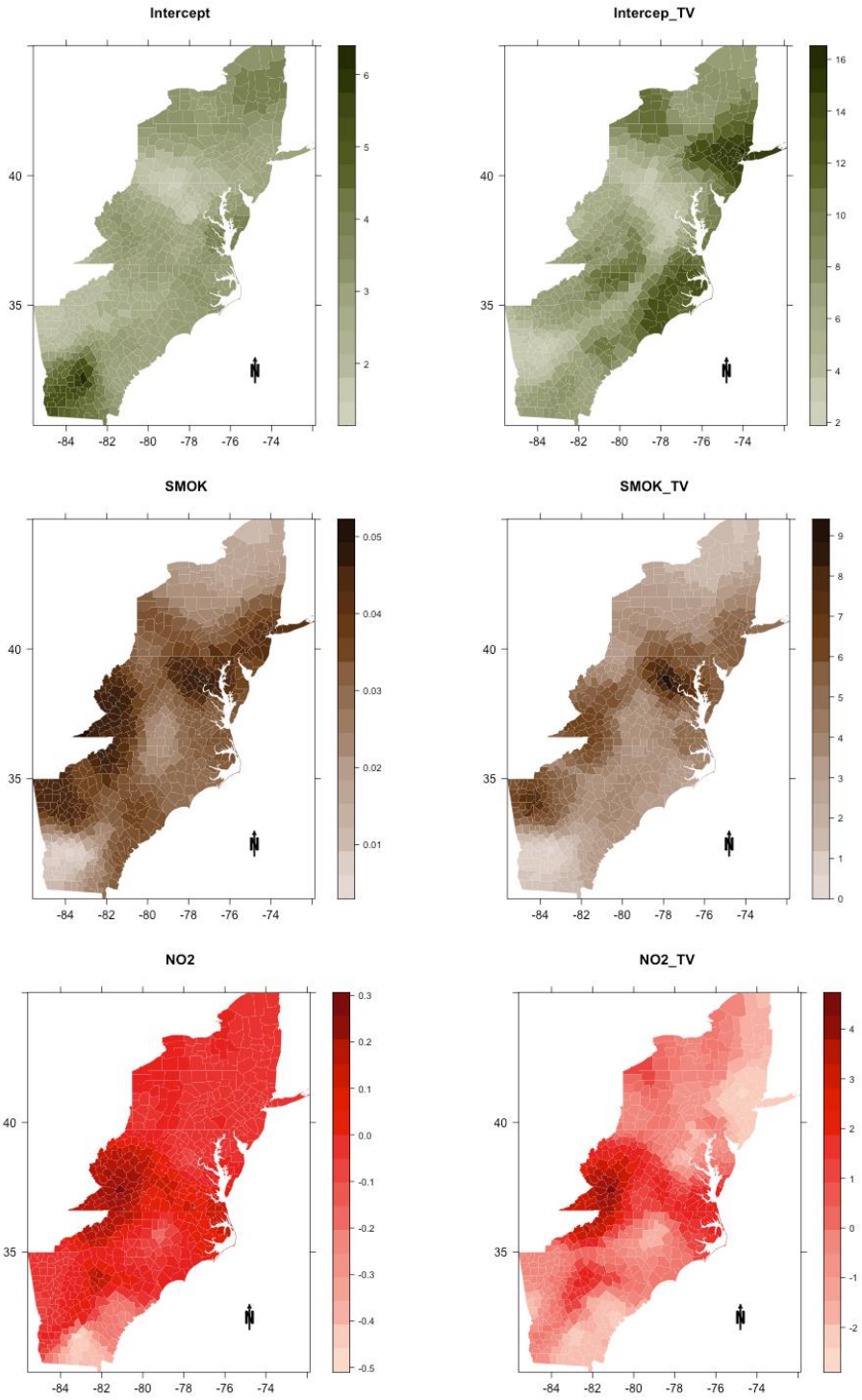


Figura 66. Estimaciones de coeficientes y valores t (TV) para la GGWR. SMOK (Prevalencia del tabaquismo estandarizado por edad), POV (Tasa de pobreza estandarizada por edad y por condado), NO2 (Dióxido de nitrógeno), PM25 (Partículas PM2.5) y SO2 (Dióxido de azufre)

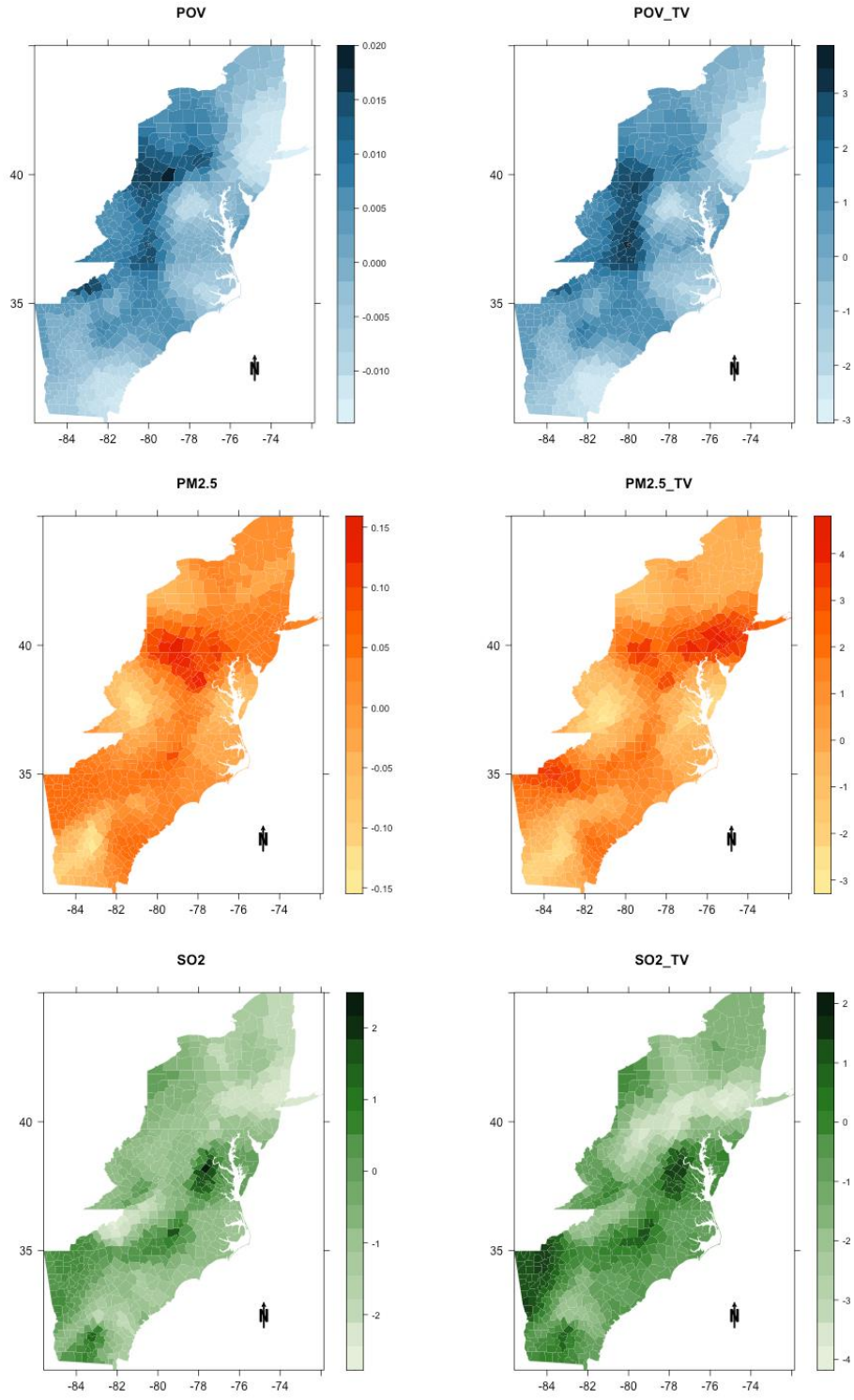


Figura 66 (continuación). Estimaciones de coeficientes y valores t (TV) para la GGWR. SMOK (Prevalencia del tabaquismo estandarizado por edad), POV (Tasa de pobreza estandarizada por edad y por condado), NO2 (Dióxido de nitrógeno), PM25 (Partículas PM2.5) y SO2 (Dióxido de azufre)

5.3.7. Ejemplo: Regresión Heterocedástica Ponderada Geográficamente

A diferencia del modelo de GWR básico en el que se supone que la varianza del término de error es fija, en la regresión heteroscedástica ponderada geográficamente (HGWR, por sus iniciales en inglés), se supone que la varianza es una función de la posición. Esto puede aumentar tanto la precisión del pronóstico como la precisión de la incertidumbre del pronóstico sobre el modelo de GWR básico.

Las entradas utilizadas para la estimación de la matriz de distancias son las mismas que en el ejemplo del ítem anterior. Las opciones utilizadas en el ejemplo y las salidas numéricas se muestran en la Figura 67, mientras que en la Figura 68 aparecen los coeficiente estimados.

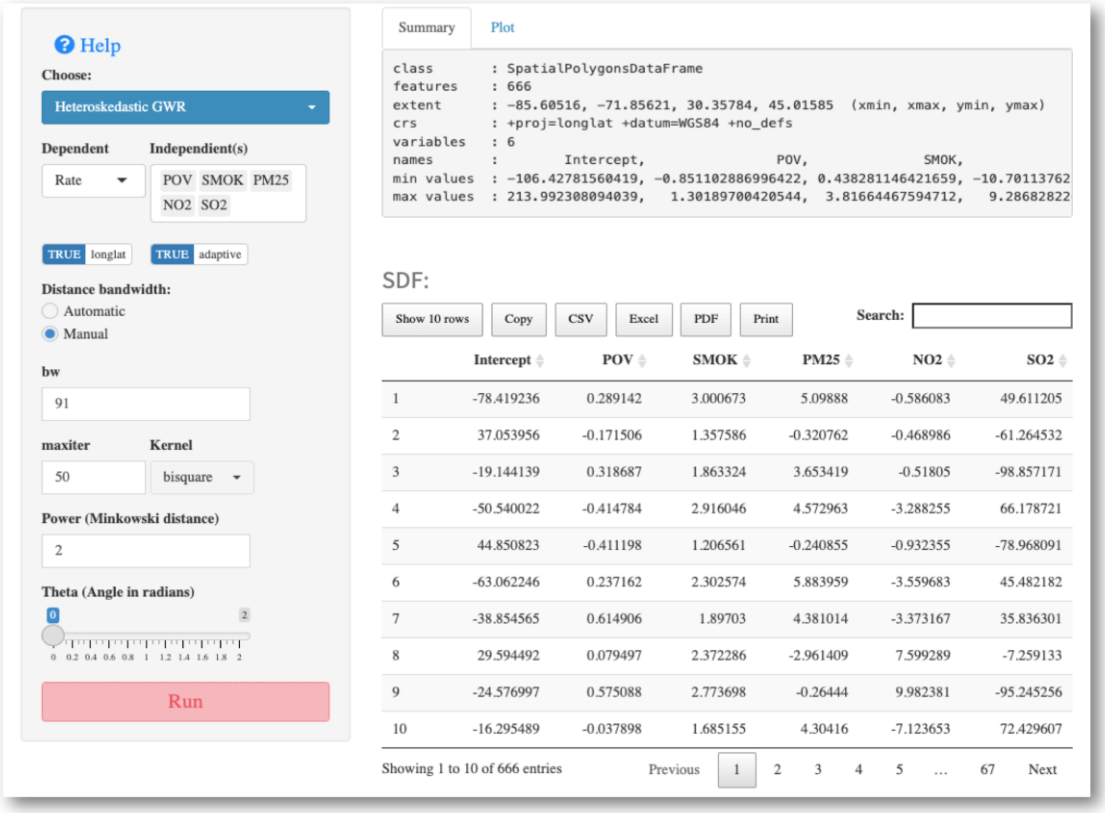


Figura 67. Entradas utilizadas para la estimación de HGWR

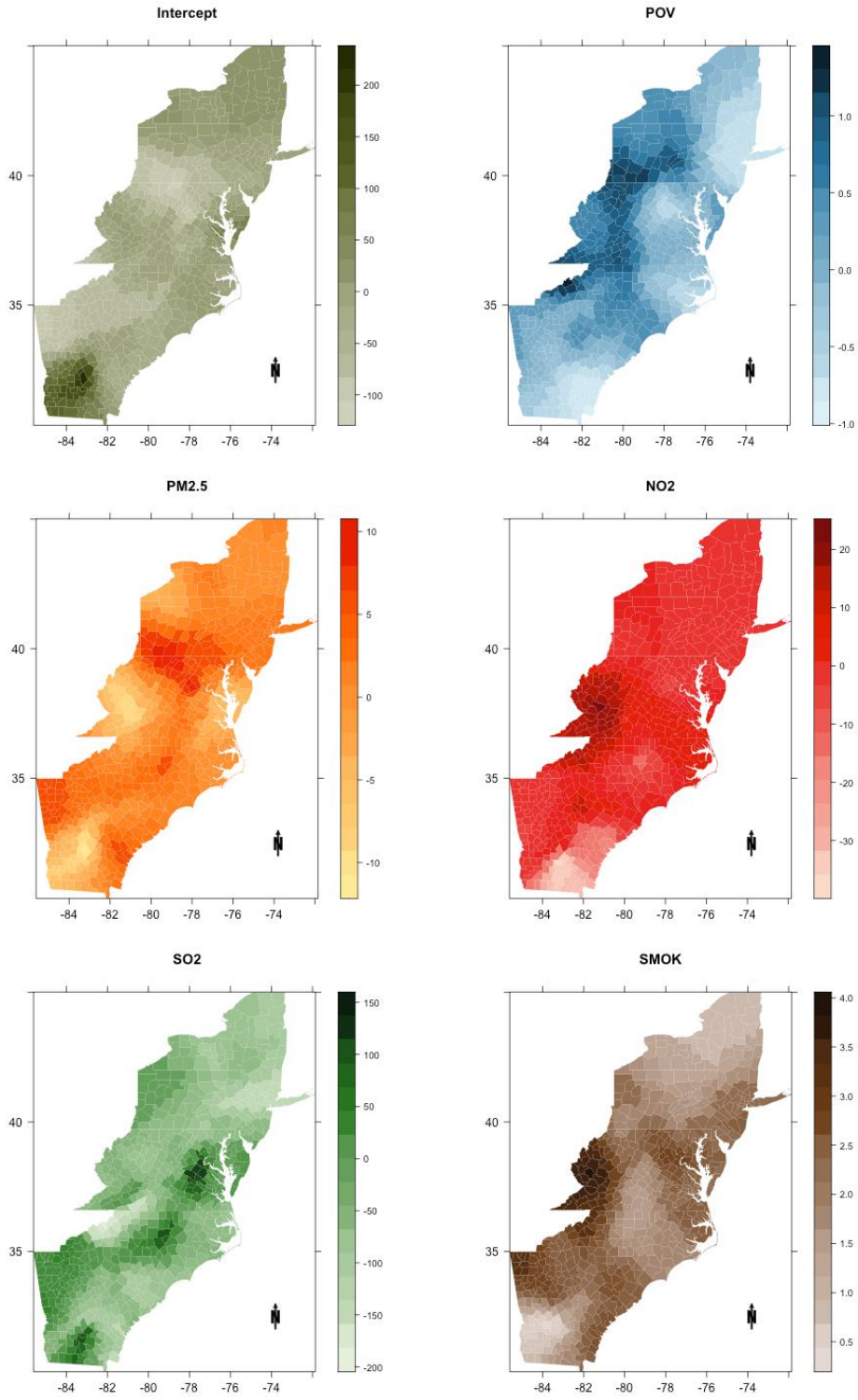


Figura 68. Estimaciones de coeficientes en el ejemplo de HGWR. SMOK (Prevalencia del tabaquismo estandarizado por edad), POV (Tasa de pobreza estandarizada por edad y por condado), NO2 (Dióxido de nitrógeno), PM25 (Partículas PM2.5) y SO2 (Dióxido de azufre)

5.3.8. Ejemplo: Regresión Mixta Ponderada Geográficamente (MGWR)

La GWR básica se calibra utilizando un único ancho de banda. Esto posiblemente no sea razonable porque implica que cada relación de respuesta vs predictor opera en la misma escala espacial. Algunas relaciones funcionan a mayor escala, mientras que otras funcionan a menor escala. Una GWR básica cancelará estas variaciones y producirá una relación de escala no estacionaria "mejor en promedio". Con la MGRW algunas variables pueden ser consideradas estacionarias y otras no estacionarias (Mei et al., 2006).

La Figura 69 muestra la entrada utilizada, mientras que la Figura 70 muestra la salida gráfica de los coeficientes estimados para cada variable. En este caso, la variable POV fue considerada como una variable fija.

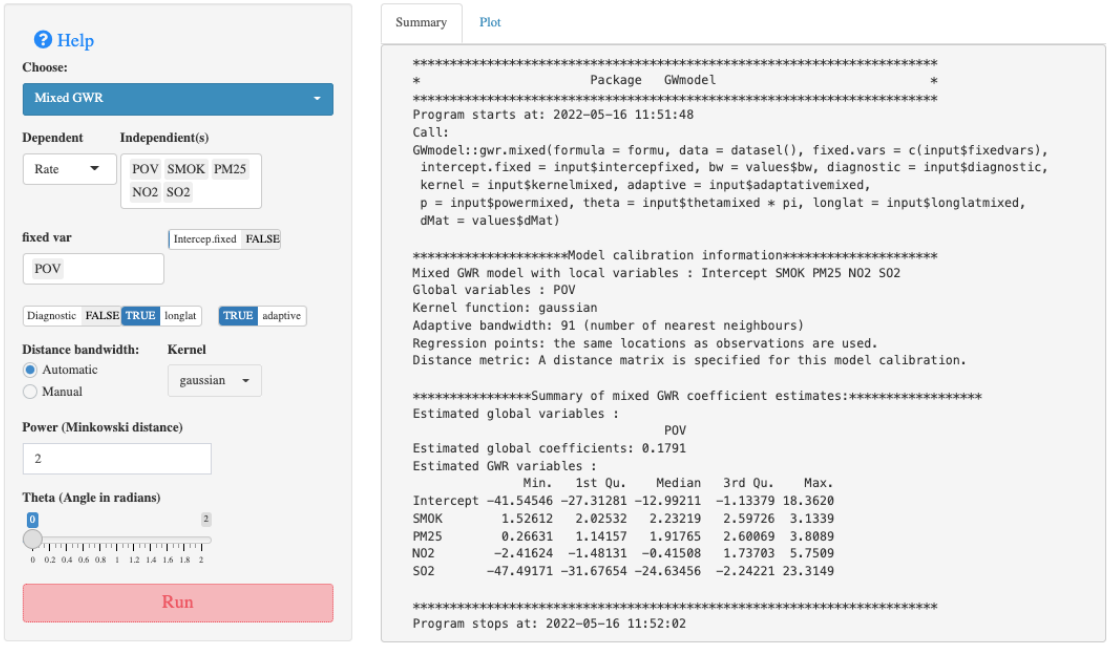


Figura 69. Configuración utilizada para la estimación de MGWR y resultado numérico

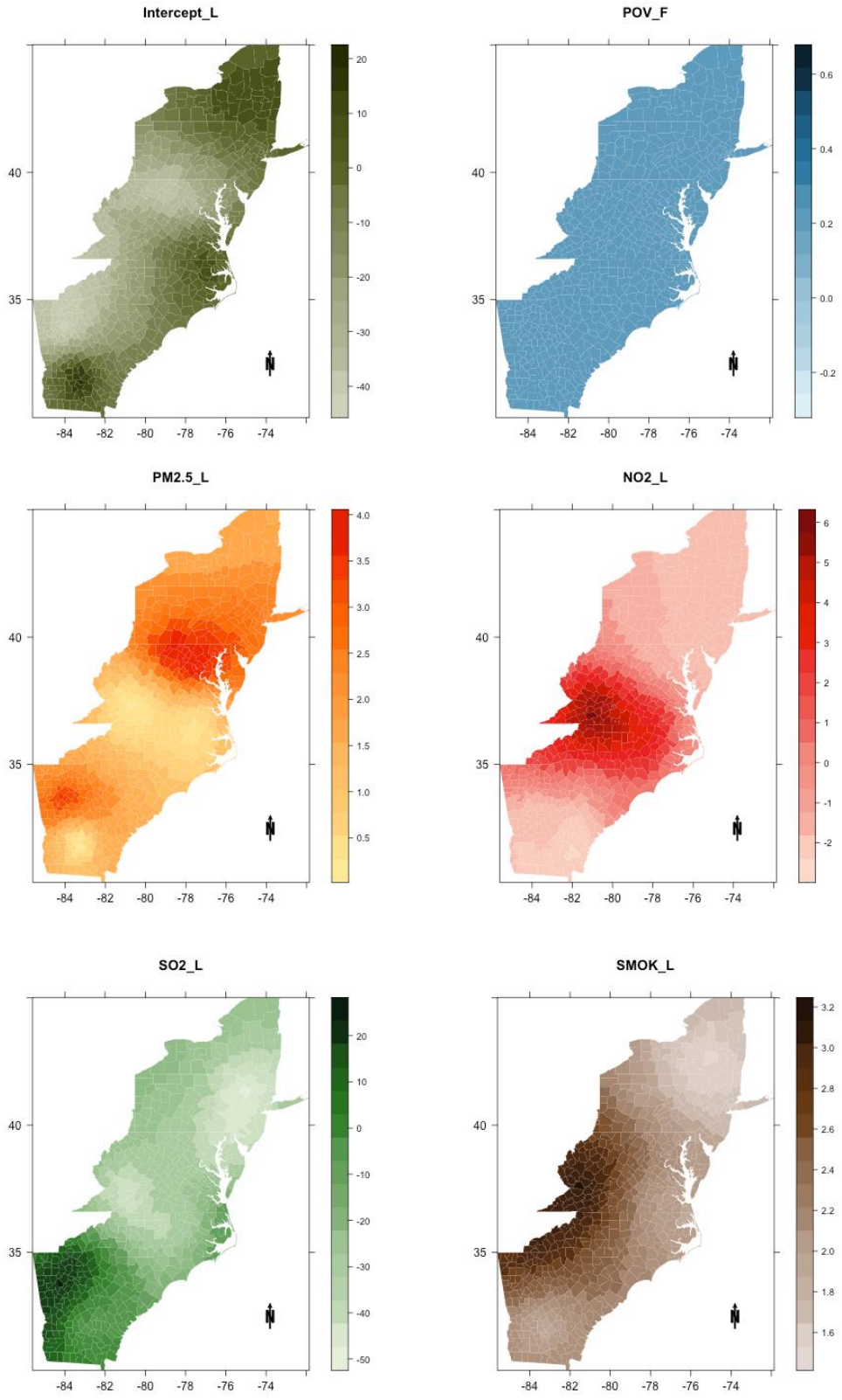


Figura 70. Estimaciones de coeficientes para MGWR. POV (Tasa de pobreza estandarizada por edad y por condado) fue considerada como variable fija. SMOK (Prevalencia del tabaquismo estandarizado por edad), NO2 (Dióxido de nitrógeno), PM25 (Partículas PM2.5) y SO2 (Dióxido de azufre)

5.3.9. Ejemplo: Regresión Mixta Ponderada Geográficamente Escalable

Murakami et al. (2020) propusieron un algoritmo rápido de GWR para grandes conjuntos de datos sin necesidad de paralelización. Las salidas de los resultados de este módulo son las misma que en la GWR básica. Se proporciona el resumen del modelo de regresión global y se presentan diagnósticos como los valores AIC para las regresiones globales y locales, además de los resúmenes de los coeficientes SGWR estimados (mínimo, primer cuartil, mediana, tercer cuartil, máximo). En cuanto a las opciones gráficas, se pueden realizar los mismos gráficos que estaban disponibles para Basic GWR.

5.3.10. Ejemplo: Análisis de Componentes Principales Ponderadas Geográficamente (GWPCA)

El PCA estándar puede reemplazarse con un análisis GWPCA (Harris et al., 2011b; Harris et al., 2014) en estudios geográficos para tener en cuenta la variabilidad espacial en la estructura multivariante de los datos. El GWPCA puede revelar regiones en las que suponer la misma estructura subyacente en todas las ubicaciones resulta inadecuado o demasiado simplista. El GWPCA puede determinar cómo varía regionalmente la dimensionalidad (efectiva) de los datos y cómo las variables originales influyen en cada componente que varía espacialmente (Harris et al., 2014).

Se analizaron los datos del mediante un GWPCA básico y un GWPCA robusto. En la Figura 71 podemos ver las entradas para el GWPCA básico.

El ancho de banda adaptativo se calculó utilizando un núcleo bicuadrado, tanto para un GWPCA básico como para el robusto. Los valores de ancho de banda óptimos encontrados fueron N = 599 y N = 137 para el ancho de banda básico y robusto, respectivamente. Además, hay que señalar que para ajustar el modelo debemos especificar todos los componentes posibles (k=5). Esta especificación asegura que la variación explicada localmente por cada componente se estime correctamente (Gollini et al., 2015).

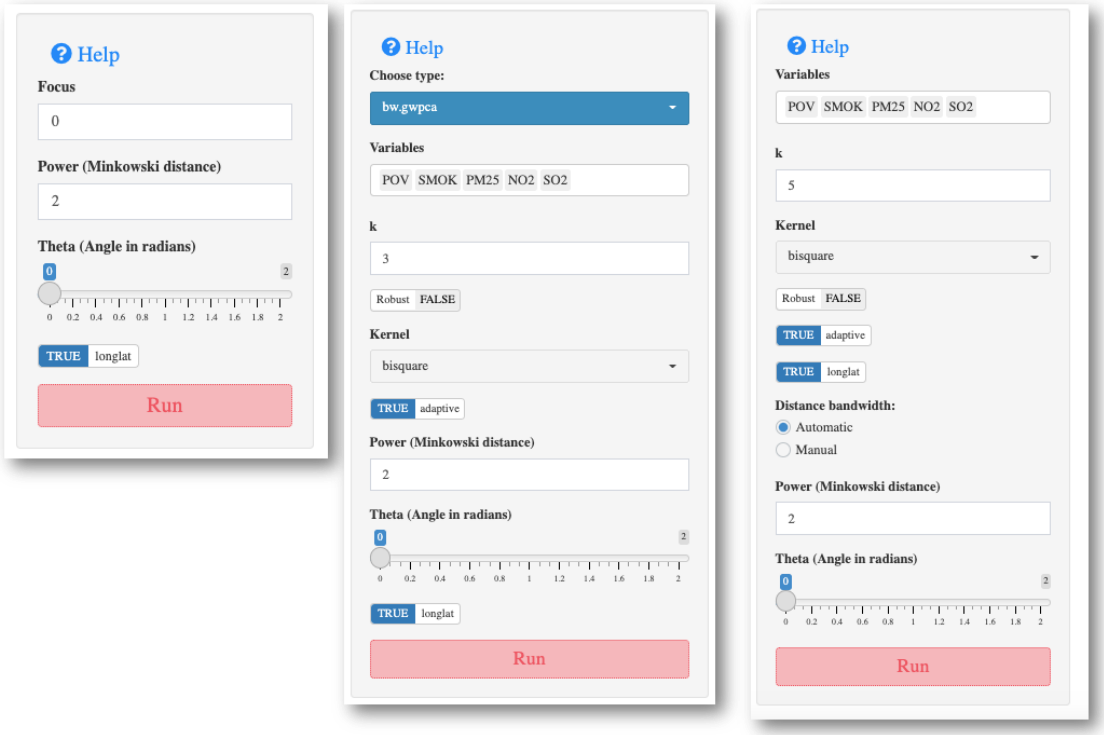


Figura 71. Configuración utilizada en el análisis GWPCA

Las salidas numéricas tanto del GWPCA básico como robusto se pueden consultar en el Anexo 7. En el PCA global, las tres primeras componentes, explican, en forma conjunta, el 83,35% de la variación de los datos.

La Figura 72 presenta los mapas del porcentaje de varianza total (PTV) locales para las tres primeras componentes de los dos GWPCA (básico y robusto). El patrón espacial en ambos mapas es diferente. En el GWPCA básico los mayores porcentajes se encuentran en los condados ubicados en el Centro-Oeste de la región, mientras que los menores porcentajes se ubican del Centro-Este al Norte. Para el GWPCA robusto no se observa un patrón claro, sin embargo, se puede observar que los PTV son superiores a los encontrados en el GWPCA básico. Las diferencias entre las salidas PTV básicas y robustas pueden ser tenidas en cuenta para advertir que pueden existir valores atípicos multivariados globales o posiblemente locales.

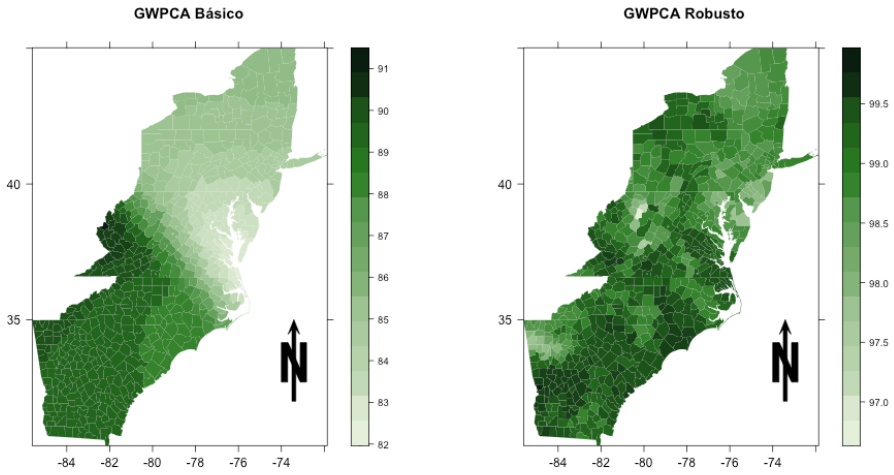


Figura 72. Porcentaje Total de Varianza (PVT) locales para las tres primeras componentes

Otro aspecto importante que se observa en la Figura 72 es que el porcentaje de varianza explicada por las primeras tres componentes locales excede el porcentaje del PCA global en a mayoría de condados tanto en el caso del GWPCA básico como en el robusto.

Un mayor porcentaje de la varianza total generalmente se explica en la primera componente en el caso GWPCA, que en el caso del PCA global (Harris et al., 2011a).

En la Figura 73 podemos visualizar cómo cada una de las cinco variables consideradas en el ejemplo influye localmente en una componente dada, mapeando la "variable ganadora", es decir, aquélla que presenta la carga absoluta más alta. Este tipo de visualización fue sugerida por Fotheringham et al. (2002)

Para el GWPCA básico en la primera componente la variable POV (Tasa de pobreza estandarizada por edad y por condado), domina en los condados ubicados en el Sur, la prevalencia del tabaquismo estandarizado por edad (SMOK) en los ubicados en la parte central, mientras que en el Norte predomina la variable dióxido de nitrógeno (NO2). El dióxido de nitrógeno (NO2) es un componente principal de la mezcla de contaminación del aire. Se han identificado fuertes asociaciones entre el NO2 y la mortalidad en estudios de varias ciudades de todo el mundo (Geddes et al., 2016)

En lo que respecta a la segunda componente del GWPCA básico, claramente la variable dióxido de azufre (SO2) domina la mayoria de condados. El dióxido de azufre (SO2) es un contaminante atmosférico que ingresa a la atmósfera a través de actividades antropogénicas,

como la combustión de combustibles que contienen azufre, procesos de refinación de petróleo, operaciones de fundición de minerales (Fioletov et al., 2017).

En la tercera componente el material particulado (PM2.5), las concentraciones de partículas finas en el ambiente contribuyen significativamente a la carga mundial de enfermedades, causando 3 millones de muertes prematuras en 2013 (Van Donkelaar et al., 2016).

La variación en los resultados del GWPCA básico es mucho mayor que la encontrada con el GWPCA robusto, se observa un predominio de la variable SMOK (Prevalencia del tabaquismo estandarizado por edad) en casi todo el territorio a excepción de los condados ubicados en el Sur que son dominados por la variable POV (Tasa de pobreza estandarizada por edad y por condado).

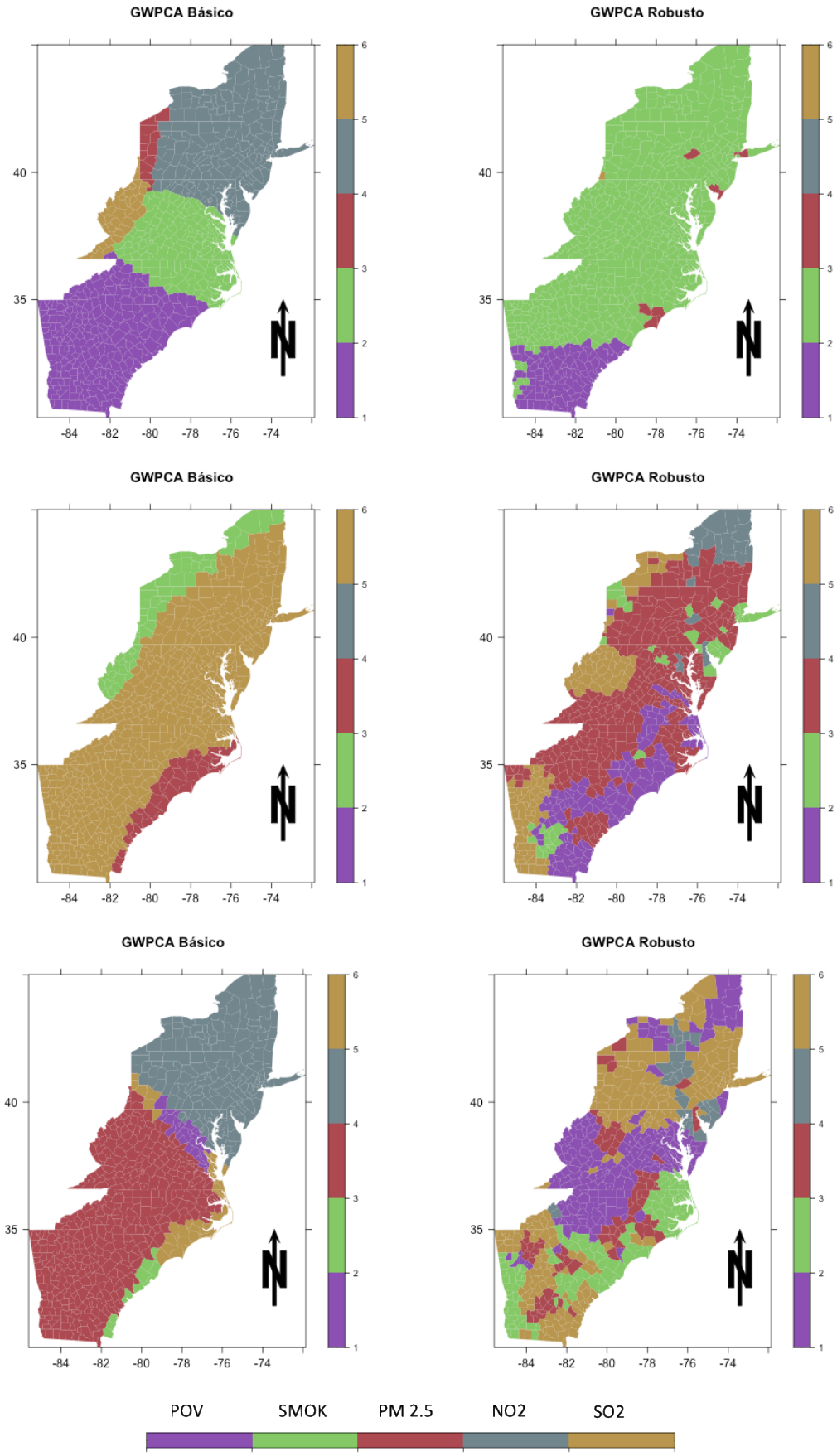


Figura 73. Resultados del GWPCA básico y robusto para la variable ganadora en las tres primeras componentes. Los componentes fueron gráficados de arriba hacia abajo (componentes 1, 2 y 3)

5.3.11. Ejemplo Análisis Discriminante Ponderado Geográficamente (GWDA)

Para mostrar cómo utilizar el programa y los resultados que se obtienen, utilizaremos los datos que se encuentran en el paquete GeoWeightedModel al seleccionar la opción *Use example data set?* y que corresponde a los resultados de las elecciones presidenciales de EE. UU. de 2004 a nivel de condado (N= 3111).

Como factor de agrupación se utilizaron tres niveles dependiendo de si ganó Bush o Kerry, niveles 1 y 2 respecivamente, y el nivel Borderline (nivel 3) en el caso de los condados en los que no hubo un claro ganador. Las variables utilizadas fueron: el porcentaje de desempleados (*unemployed*), el porcentaje de adultos mayores de 25 años con cuatro o más años de educación universitaria (*pctcoled*), el porcentaje de personas mayores de 65 años (*PEROVER65*) el porcentaje de población urbana en el condado (*pcturban*) y el porcentaje de blancos (WHITE).

En la Figura 74 se pueden observar los valores de entrada utilizados para realizar el GWDA, así como también la salida numérica.

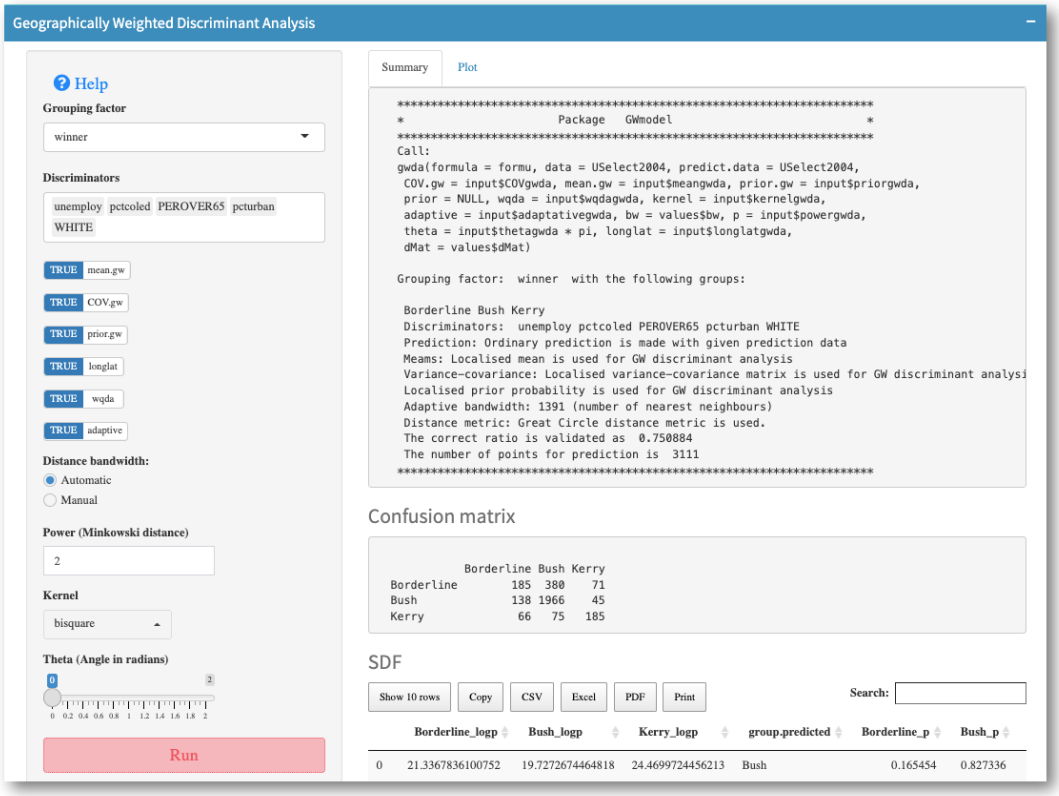


Figura 74. Configuración utilizada en la realización del GWDA. Se muestran también los resultados numéricos

Entre los resultados obtenidos encontramos la matriz de confusión resultante, además de la precisión de la clasificación (75,08%).

En la Figura 75 se presentan en un mapa, a nivel de condado, los resultados reales de la elección presidencial donde puede observarse que Bush fue un claro ganador en la mayoría de ellos (información original). En el mapa central se presentan los resultados de la clasificación usando el GWDA, donde podemos obsevar que el patrón espacial en las clasificaciones es muy cercano a los resultados reales.

En el mapa inferior se muestra la entropía (*shannon.entropy*), que mide la incertidumbre de una fuente de información y varía entre 0 y 1. El resultado se interpreta como 0, la clasificación es muy segura y 1 significa muy incierto.

Los resultados de las elecciones de asignados por GWDA revelan un interesante patrón de las clasificaciones erróneas, pues en el caso de los condados en los que no hubo un claro ganador, la mayoria fue asignado a Bush, sobre todo en los ubicados en la parte Norte-centro.

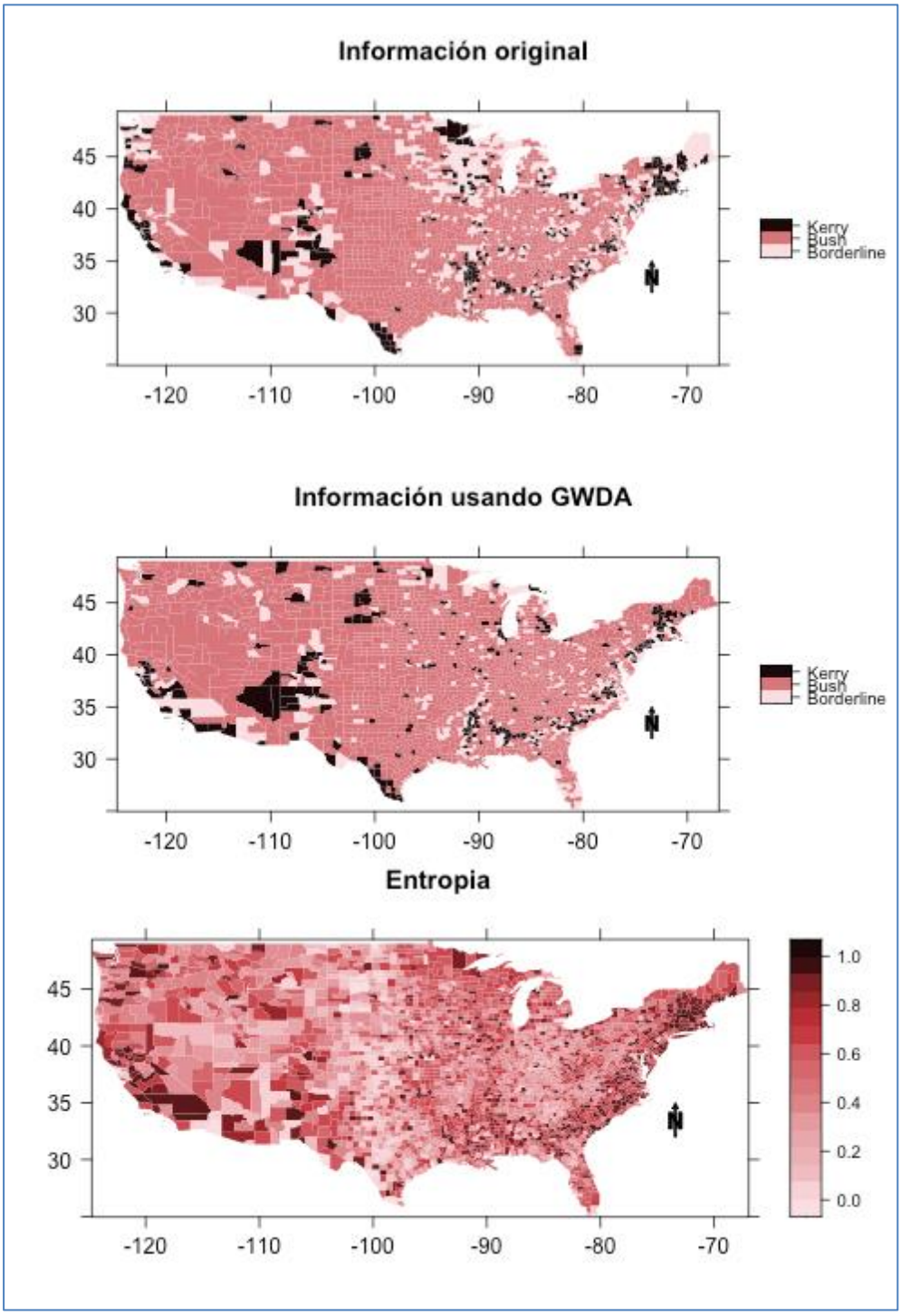


Figura 75. Resultados de las elecciones presidenciales de EE. UU. de 2004, resultados reales y los de la clasificación utilizando GWDA así como la entropía

CONCLUSIONES

1. Dentro de los modelos revisados, el de la Regresión Ponderada Geográficamente es el que despierta mayor interés dentro de la comunidad científica, dado el alto número de extensiones y publicaciones encontradas.

2. Se desarrolló el programa LDAShiny para dar soporte a la propuesta metodológica de revisión bibliográfica. Las características de esta aplicación permiten que esté a disposición de toda la comunidad científica a través de una conexión a internet, siendo fácil de instalar y fácil de utilizar, desde una interfaz gráfica.

3. La metodología propuesta para la revisión bibliográfica permitió analizar con éxito la información de 3183 artículos, que se procesó rápidamente con la aplicación LDAShiny. Se utilizaron menos de tres días para procesar los 3183 artículos y agruparlos en 22 tópicos de investigación, con el uso de una computadora portátil estándar.

4. El análisis de la tendencia de los tópicos identificados refleja el cambio de interéses de los investigadores que utilizan la Regresión Ponderada Geográficamente y puede ayudarles a comprender la tendencia de la investigación, advertirles, si es necesario, sobre la necesidad de realizar proyectos integrados y cambiar su enfoque de investigación.

5. Los resultados del escalamiento multidimensiona no métrico permitió validar el etiquetado previo de los tópicos, al mostrar agrupaciones coherentes y superposición de nodos, lo que indica distribuciones de palabras similares.

6. El análisis HJ-Biplot se configura como una herramienta efectiva en el análisis de la distribución por países de los tópicos encontrados en el análisis de la literatura sobre modelos geográficamente ponderados, y generalizable a otras áreas de interés.

7. La optimización de código realizada en el programa GeoWeightedModel le confiere gran potencia de cálculo.

8. La interfaz gráfica desarrollada para el programa GeoWeightedModel guía al usuario en todas las acciones que necesita para el proceso de análisis de datos sin exigirle que posea conocimientos de programación. La gran variedad de opciones gráficas implementadas permite que el usuario se centre en el objetivo de la interpretación de los resultados.

REFERENCIAS

Adams, J., Khan, H. T., Raeside, R., & White, D. I. (2007). *Research methods for graduate business and social science students*. SAGE publications India.

Ahmed, Z. U., Sun, K., Shelly, M., & Mu, L. (2021). Explainable artificial intelligence (XAI) for exploring spatial variability of lung and bronchus cancer (LBC) mortality rates in the contiguous USA. *Scientific reports*, *11*(1), 1-15. https://doi.org/10.1038/s41598-021-03198-8

Akaikei, H. (1973). Information theory and an extension of maximum likelihood principle. In *Proc. 2nd Int. Symp. on Information Theory* (pp. 267-281).

Allaire, J. (2012). RStudio: integrated development environment for R. Boston, MA, *770*(394), 165-171.

Anselin, L. (1988). *Spatial econometrics: methods and models* (Vol. 4). Springer Science & Business Media.

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of informetrics*, *11*(4), 959-975. https://doi.org/10.1016/j.joi.2017.08.007

Arun, R., Suresh, V., Veni Madhavan, C. E., & Murthy, N. (2010, June). On finding the natural number of topics with latent Dirichlet allocation: Some observations. In *Pacific-Asia conference on knowledge discovery and data mining* (pp. 391-402). Springer, Berlin, Heidelberg.

Atkinson, P. M., German, S. E., Sear, D. A., & Clark, M. J. (2003). Exploring the relations between riverbank erosion and geomorphological controls using geographically weighted logistic regression. *Geographical Analysis*, *35*(1),58-82. http://doi.org/10.1111/j.1538-4632.2003.tb01101.x

Asmussen, C. B., & Møller, C. (2019). Smart literature review: a practical topic modelling approach to exploratory literature review. *Journal of Big Data*, *6*(1), 1-18. https://doi.org/10.1186/s40537-019-0255-7

Beenstock, M., & Felsenstein, D. (2018). *Econometric analysis of nonstationary spatial panel data*. Heidelberg: Springer

Berry, M. W., Dumais, S. T., & O'Brien, G. W. (1995). Using linear algebra for intelligent information retrieval. *SIAM review*, *37*(4), 573-595. https://doi.org/10.1137/1037127

Birkin, M., & Clarke, G. P. (1991). Spatial interaction in geography. *Geography Review*, *4*(5), 16-24.

Bivand, R., Yu, D., Nakaya, T., Garcia-Lopez, M. A., & Bivand, M. R. (2017). Package 'spgwr'. *R software package*.

Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *The annals of applied statistics*, *1*(1), 17-35. https://doi.org/10.1214/07-AOAS114

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *the Journal of machine Learning research*, *3*, 993-1022.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77-84 https://doi.org/10.1145/2133806.2133826

Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, *71*(2), 353-360. http://doi.org/10.1093/biomet/71.2.353

Breckling, J., & Chambers, R. (1988). M-quantiles. *Biometrika*, *75*(4), 761-771.

Brocke, J. V., Simons, A., Niehaves, B., Niehaves, B., Reimer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the giant: On the importance of rigour in documenting the literature search process.

Brunsdon, C. (2009). Reply: GWDA and UK 2005 Election Results. *Geographical Analysis, 41*(3), 338–341. http://doi.org/10.1111/j.1538-4632.2009.00757.x

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis*, *28*(4), 281-298. http://doi.org/10.1111/j.1538-4632.1996.tb00936.x

Brunsdon, C., Fotheringham, A.S., & Charlton, M.E. (1998). Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *47*(3), 431-443. http://doi.org/:10.1111/1467-9884.00145

Brunsdon, C., Fotheringham, A. S., & Charlton, M. (1999). Some notes on parametric significance tests for geographically weighted regression. *Journal of Regional science*, *39*(3), 497-524. http://doi.org/10.1111/0022-4146.00146

Brunsdon C, Fotheringham A.S, Charlton M.E. (2002) Geographically weighted summary statistics -a framework for localised exploratory data analysis. *Computers, Environment and Urban Systems 26*:501-524. https://doi.org/10.1016/S0198-9715(01)00009-6

Brunsdon, C., Fotheringham, S., & Charlton, M. (2007). Geographically weighted discriminant analysis. *Geographical Analysis*, *39*(4), 376-396. http://doi.org/10.1111/j.1538-4632.2007.00709.

Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, *72*(7-9), 1775-1781. https://doi.org/10.1016/j.neucom.2008.06.011

Casetti, E. (1972). Generating models by the expansion method: applications to geographical research. *Geographical analysis*, *4*(1), 81-91. http://doi.org/10.1111/j.1538-4632.1972.tb00458.x

Chambers, R., & Tzavidis, N. (2006). M-quantile models for small area estimation. *Biometrika*, *93*(2), 255-268. http://doi.org/10.1093/biomet/93.2.255

Chang, J. *lda: Collapsed Gibbs Sampling Methods for Topic Models*, R package version 1.4.2; R Foundation for Statistical Computing: Vienna, Austria, 2015; Available online: https://CRAN.R-project.org/package=lda (accessed on 16 April 2021).

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J., & Blei, D. (2009). Reading tea leaves: How humans interpret topic models. In Advances in Neural Information Processing Systems 22 .

Chang, W.; Cheng, J.; Allaire, J.; Xie, Y.; McPherson, J. *Shiny: Web Application Framework for R*, R package version 1.4.0.2; R Foundation for Statistical Computing: Vienna, Austria. Available online: https://CRAN.R-project.org/package=shiny (accessed on 17 April 2021).

Charlton, M., Fotheringham, A. S., & Brunsdon, C. (2006). Geographically weighted regression: NCRM Methods Review Papers/NCRM/006. *Recuperado de http://eprints. ncrm. ac. uk/90*.

Charlton, M., Fotheringham, A. S., & Brunsdon, C. (2007). Geographically weighted regression: software for GWR. *National Centre for Geocomputation*.

Chen, V. Y. J., Deng, W. S., Yang, T. C., & Matthews, S. A. (2012). Geographically weighted quantile regression (GWQR): An application to US mortality data. *Geographical Analysis*, *44*(2), 134-150. http://doi.org/10.1111/j.1538-4632.2012.00841.x

Chen, V. Y. J., & Yang, T. C. (2012). SAS macro programs for geographically weighted generalized linear modeling with spatial point data: Applications to health research. *Computer methods and programs in biomedicine*, *107*(2), 262-273. https://doi.org/10.1016/j.cmpb.2011.10.006

Chybicki, A. (2017). Three-dimensional geographically weighted inverse regression (3GWR) model for satellite derived bathymetry using Sentinel-2 observations. *Marine Geodesy*, *41*(1), 1-23. https://doi.org/10.1080/01490419.2017.1373173

Chun, Y., & Griffith, D. A. (2013). *Spatial statistics and geostatistics: theory and applications for geographic information science and technology*. Sage

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American statistical association*, *74*(368), 829-836. http://doi.org/10.1080/01621459.1979.10481038

Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American statistical association*, *83*(403), 596-610. http://doi.org/10.1080/01621459.1988.10478639

Cliff, A. D., & Ord, J. K. (1981). Spatial processes: models & applications. Taylor & Francis.

Craven, P., & Wahba, G. (1978). Smoothing noisy data with spline functions. *Numerische mathematik*, *31*(4), 377-403. http://doi.org/10.1007/BF01404567

Comber, A., & Harris, P. (2018). Geographically weighted elastic net logistic regression. *Journal of Geographical Systems*, *20*(4), 317-341. https://doi.org/10.1007/s10109-018-0280-7

Comber, A., Brunsdon, C., Charlton, M., Dong, G., Harris, R., Lu, B., & Harris, P. (2020). The GWR route map: a guide to the informed application of Geographically Weighted Regression. *arXiv preprint arXiv:2004.06070*.

da Silva, A. R., & de Oliveira Lima, A. (2017). Geographically weighted beta regression. *Spatial Statistics*, *21*, 279-303. http://doi.org/10.1016/j.spasta.2017.07.011

da Silva, A. R., & Rodrigues, T. C. V. (2014). Geographically weighted negative binomial regression—incorporating overdispersion. *Statistics and Computing*, *24*(5), 769-783. http://doi.org/10.1007/s11222-013-9401-9

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American society for information science*, *41*(6), 391-407. https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9

Denny, M. J., & Spirling, A. (2018). Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it. *Political Analysis*, *26*(2), 168-189. https://doi.org/10.1017/pan.2017.44

Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. *Document numérique*, *17*(1), 61-84.

de Wildt, T. E., Chappin, E. J., van de Kaa, G., & Herder, P. M. (2018). A comprehensive approach to reviewing latent topics addressed by literature across multiple disciplines. Applied Energy, *228*, 2111-2128.

DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of US government arts funding. *Poetics*, *41*(6), 570-606. https://doi.org/10.1016/j.poetic.2013.08.004

Dong, G., Nakaya, T., & Brunsdon, C. (2018). Geographically weighted regression models for ordinal categorical response variables: An application to geo-referenced life satisfaction data. *Computers, Environment and Urban Systems*, *70*, 35-42. https://doi.org/10.1016/j.compenvurbsys.2018.01.012

Dwyer-Lindgren, L., Mokdad, A. H., Srebotnjak, T., Flaxman, A. D., Hansen, G. M., & Murray, C. J. (2014). Cigarette smoking prevalence in US counties: 1996-2012. *Population health metrics, 12*(1), 1-13. https://doi.org/10.1186/1478-7954-12-5

Environmental Systems Research Institute (ESRI) (2018). ArcMap 10.3 Spatial Analyst Toolbox. Redlands, CA. URL http://desktop.arcgis.com/en/arcmap/10.3/tools/ spatial-statistics-toolbox/geographically-weighted-regression.htm.

Erskine, N. *lda.svi: Fit Latent Dirichlet Allocation Models Using Stochastic Variational Inference*, R package version 0.1.0.; R Foundation for Statistical Computing: Vienna, Austria, 2015; Available online: https://CRAN.Rproject.org/package=lda.svi (accessed on 16 April 2021).

Escobar, K. M., Vicente-Villardon, J. L., de la Hoz-M, J., Useche-Castro, L. M., Alarcón Cano, D. F., & Siteneski, A. (2021). Frequency of Neuroendocrine Tumor Studies: Using Latent Dirichlet Allocation and HJ-Biplot Statistical Methods. *Mathematics*, *9*(18), 2281. https://doi.org/10.3390/math9182281

Falls, L. W. (1974). The beta distribution: a statistical model for world cloud cover. *Journal of Geophysical Research*, *79*(9), 1261-1264. http://doi.org/10.1029/JC079i009p01261

Farber, S., & Páez, A. (2007). A systematic investigation of cross-validation in GWR model estimation: empirical analysis and Monte Carlo simulations. *Journal of Geographical Systems*, *9*(4), 371-396. https://doi.org/10.1007/s10109-007-0051-3

Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of applied statistics*, *31*(7), 799-815. http://doi.org/10.1080/0266476042000214501

Fernández, S., Cotos-Yáñez, T., Roca-Pardiñas, J., & Ordóñez, C. (2018). Geographically weighted principal components analysis to assess diffuse pollution sources of soil heavy metal: application to rough mountain areas in Northwest Spain. *Geoderma*, *311*, 120-129. https://doi.org/10.1016/j.geoderma.2016.10.012

Fioletov, V., McLinden, C. A., Kharol, S. K., Krotkov, N. A., Li, C., Joiner, J., & Denier van der Gon, H. A. (2017). Multi-source SO2 emission retrievals and consistency of satellite and

surface measurements with reported emissions. *Atmospheric Chemistry and Physics*, *17*(20), 12597-12616. https://doi.org/10.5194/acp-17-12597-2017

Foley, P., & Demšar, U. (2013). Using geovisual analytics to compare the performance of Geographically Weighted Discriminant Analysis versus its global counterpart, Linear Discriminant Analysis. *International Journal of Geographical Information Science*, *27*(4), 633-661. http://doi.org/10.1080/13658816.2012.722638

Fotheringham, A. S., Charlton, M. E., & Brunsdon, C. (1998). Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and planning A*, *30*(11), 1905-1927.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons.

Fotheringham, A. S., Crespo, R., & Yao, J. (2015). Geographical and temporal weighted regression (GTWR). *Geographical Analysis*, *47*(4), 431-452. http://doi.org/10.1111/gean.12071

Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, *58*(3), 453-467. https://doi.org/10.1093/biomet/58.3.453

Galindo, M. P. (1986). An alternative for simultaneous representation: HJ-Biplot. *Questiio*, *10*, 12-23.

Geddes, J. A., Martin, R. V., Boys, B. L., & van Donkelaar, A. (2016). Long-term trends worldwide in ambient NO2 concentrations inferred from satellite observations. *Environmental health perspectives*, *124*(3), 281-289. https://doi.org/10.1289/ehp.1409567

Geary, R. C. (1954). The contiguity ratio and statistical mapping. *The incorporated statistician*, *5*(3), 115-146. https://doi.org/10.2307/2986645

Geniaux, G., & Martinetti, D. (2018). Package 'mgwrsar': GWR and MGWR with Spatial Autocorrelation. *R software package*

Getis, A., & Ord, J. K. (2010). The analysis of spatial association by use of distance statistics. In *Perspectives on spatial data analysis* (pp. 127-145). Springer, Berlin, Heidelberg.

Gollini, I., Lu, B., Charlton, M., Brunsdon, C., Harris, P. (2015).Gwmodel: An R package for exploring spatial heterogeneity using geographically weighted models. *Journal of Statistical Software*, *63* (17), 1-50. http://doi.org/10.18637/jss.v063.i17

Gour, A., Aggarwal, S., & Kumar, S. (2022)Lending ears to unheard voices: An empirical analysis of user generated content on social media. *Production and Operations Management*. https://doi.org/10.1111/poms.13732

Gower, J. C., & Hand, D. J. (1995). *Biplots* (Vol. 54). CRC Press.

Gower, J. C., Lubbe, S. G., & Le Roux, N. J. (2011). *Understanding biplots*. John Wiley & Sons.

Greenacre, M. J. (2010). *Biplots in practice*. Fundacion BBVA

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National academy of Sciences*, *101*(suppl 1), 5228-5235. https://doi.org/10.1073/pnas.0307752101

Grimmer, J. (2010). A Basiian hierarchical topic model for political texts: Measuring expressed agendas in Senate press releases. *Political Analysis*, *18*(1), 1-35. https://doi.org/10.1093/pan/mpp034

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, *21*(3), 267-297. https://doi.org/10.1093/pan/mps028

Grun, B.& Hornik, K. (2011). topicmodels: An R Package for Fitting Topic Models. *J. Stat. Softw. 40*, 1–30. http://doi.org/10.18637/jss.v040.i13

Han, J., Kang, X., Yang, Y., & Zhang, Y. (2022). Geographically and temporally weighted principal component analysis: a new approach for exploring air pollution non-stationarity in China, 2015–2019. *Journal of Spatial Science*, 1-18.
https://doi.org/10.1080/14498596.2022.2028270

Harris, P., Brunsdon, C., & Fotheringham, A. S. (2011a). Links, comparisons and extensions of the geographically weighted regression model when used as a spatial predictor. *Stochastic Environmental Research and Risk Assessment, 25(2), 123–138*. http://doi.org/10.1007/s00477-010-0444-6
Harris, P., Brunsdon, C., & Charlton, M. (2011b). Geographically weighted principal components analysis. *International Journal of Geographical Information Science*, *25*(10), 1717-1736. https://doi.org/10.1080/13658816.2011.554838

Harris, P., Clarke, A., Juggins, S., Brunsdon, C., & Charlton, M. (2014). Enhancements to a geographically weighted principal component analysis in the context of an application to an environmental data set. *Geographical Analysis*, 47(2), 146-172. https://doi.org/10.1111/gean.12048

Harzing, A. W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. *Scientometrics*, *106*(2), 787-804. https://doi.org/10.1007/s11192-015-1798-9

Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge University Press.

Hoaglin, D. C., & Welsch, R. E. (1978). The hat matrix in regression and ANOVA. *The American Statistician*, *32*(1), 17-22. http://doi.org/10.1080/00031305.1978.10479237

Hoerl, A. E., & Kennard, R. W. (1970a). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55-67. http://doi.org/10.1080/00401706.1970.10488634

Hoerl, A. E., & Kennard, R. W. (1970b). Ridge regression: applications to nonorthogonal problems. *Technometrics*, *12*(1), 69-82. http://doi.org/10.1080/00401706.1970.10488635

Hofmann, T. (1999, August). Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 50-57)

Hotelling, H. (1933). Analysis of a Complex of Statistical Variables Into Principal Components. *Journal of Educatonal Psychology*, *24*(6), 417–441. http://doi.org/10.1037/h0071325

Pramana, S., & Pamungkas, I. H. (2016). Improvement method of fuzzy geographically weighted clustering using gravitational search algorithm. *Jurnal Ilmu Komputer dan Informasi*, *11*(1), 10-16.

Jacobi, C., Van Atteveldt, W., & Welbers, K. (2016). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital journalism*, *4*(1), 89-106. https://doi.org/10.1080/21670811.2015.1093271

Johnston, R. J., Gregory, D., Pratt, G., & Watts, M. (2000). *The Dictionary of Human Geography. ed.* Oxford, UK: Blackwell Publishers.

Johnston, R., & Pattie, C. (2009). Comment: geographically weighted discriminant analysis and the 2005 British general election. *Geographical Analysis*, *41*(3), 333-338. http://doi.org/10.1111/j.1538-4632.2009.00756.x

Jolliffe, I. T. (2002). *Principal component analysis for special types of data* (pp. 338-372). Springer New York.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065). http://doi.org/10.1098/rsta.2015.0202

Jones, T. *textmineR: Functions for Text Mining and Topic Modeling*, R package version 3.0.4.; R Foundation for Statistical Computing: Vienna, Austria, 2019; Available online: **https://CRAN.R-project.org/package=textmineR** (accessed on 16 April 2021)

Jurafsky, D., & Martin, J. H. (2008). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Pearson: Hoboken, NJ, USA.

Kalogirou, S., & Kalogirou, M. S. (2020). Package 'lctools'. Local Correlation, Spatial Inequalities and Other Tools. *R software package*

Kassambara, A. (2017). *Practical guide to principal component methods in R: PCA, M (CA), FAMD, MFA, HCPC, factoextra* (Vol. 2). Sthda.

Kitchin, R., & Thrift, N. (2009). *International encyclopedia of human geography*. Elsevier.

Koenig, J. G. (1980). Indicators of urban accessibility: Theory and application. *Transportation*, 9(2), 145–172. https://doi.org/10.1007/BF00167128

Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, 33-50. http://doi.org/10.2307/1913643.

Kopczewska, K. (2020). *Applied spatial statistics and econometrics: data analysis in R*. Routledge.

Kumar, S., Lal, R., & Liu, D. (2012a). A geographically weighted regression kriging approach for mapping soil organic carbon stock. *Geoderma*, *189*, 627-634. http://doi.org/10.1016/j.geoderma.2012.05.022

Kumar, S., Lal, R., & Lloyd, C. D. (2012b). Assessing spatial variability in soil characteristics with geographically weighted principal components analysis. *Computational Geosciences*, 16(3), 827-835.

https://doi.org/10.1007/s10596-012-9290-6

Lancichinetti, A., Sirer, M. I., Wang, J. X., Acuna, D., Körding, K., & Amaral, L. A. N. (2015). High-reproducibility and high-accuracy method for automated topic classification. *Physical Review X*, *5*(1), 011007. https://doi.org/10.1103/PhysRevX.5.011007

Larsen, P. & Von Ins, M. (2010). The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics*, *84*(3), 575-603. https://doi.org/10.1007/s11192-010-0202-z

Lau, J.H., Grieser, K., Newman, D. & Baldwin, T. (2011, June). Automatic labelling of topic models. In Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies (pp. 1536-1545).
LeSage, J.P. (2001). Econometrics: Matlab toolbox of econometrics functions.

LeSage, J. P. (2004). A family of geographically weighted regression models. In : Anselin L., Florax R.J.G.M., Rey S.J. (eds). *Advances in spatial econometrics*. *Advances in Spatial Science*. Springer, Berlin, Heidelberg. http://doi.org/10.1007/978-3-662-05617-2_11

Lewis, S. C., Zamith, R., & Hermida, A. (2013). Content analysis in an era of big data: A hybrid approach to computational and manual methods. *Journal of broadcasting & electronic media*, *57*(1), 34-52. https://doi.org/10.1080/08838151.2012.761702

Li, J., & Heap, A. D. (2014). Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software*, *53*, 173-189.

Li, Z., Cheng, J., & Wu, Q. (2016). Analyzing regional economic development patterns in a fast developing province of China through geographically weighted principal component analysis. *Letters in Spatial and Resource Sciences*, *9*(3), 233-245. https://doi.org/10.1007/s12076-015-0154-2

Li, K., & Lam, N. S. (2018). Geographically weighted elastic net: A variable-selection and modeling method under the spatially nonstationary condition. *Annals of the American Association of Geographers*, *108*(6), 1582-1600. http://doi.org/10.1080/24694452.2018.1425129

Li, Z., Fotheringham, A. S., Li, W., & Oshan, T. (2019). Fast Geographically Weighted Regression (FastGWR): a scalable algorithm to investigate spatial process heterogeneity in millions of observations. *International Journal of Geographical Information Science*, *33*(1), 155-175. http://doi.org/10.1080/13658816.2018.1521523

Lloyd, C. D. (2010). Analysing population characteristics using geographically weighted principal components analysis: a case study of Northern Ireland in 2001. *Computers, Environment and Urban Systems*, *34*(5), 389-399. https://doi.org/10.1016/j.compenvurbsys.2010.02.005

Loader, C. R. (1999). Bandwidth selection: classical or plug-in?. *The Annals of Statistics*, *27*(2), https://doi.org/415-438. 10.1214/aos/1018031201

Losada, N., Alen, E., Cotos-Yanez, T. R., & Dominguez, T. (2019). Spatial heterogeneity in Spain for senior travel behavior. *Tourism Management, 70,* 444-452. https://doi.org/10.1016/j.tourman.2018.09.011

Lu, B., Brunsdon, C., Charlton, M., & Harris, P. (2017). Geographically weighted regression with parameter-specific distance metrics. *International Journal of Geographical Information Science*, *31*(5), 982-998 https://doi.org/10.1080/13658816.2016.1263731

Lu, B., Charlton, M., & Fotheringhama, A. S. (2011). Geographically weighted regression using a non-Euclidean distance metric with a study on London house price data. *Procedia Environmental Sciences*, *7*, 92-97. https://doi.org/10.1016/j.proenv.2011.07.017

Lu, B., Charlton, M., Harris, P., & Fotheringham, A. S. (2014a). Geographically weighted regression with a non-Euclidean distance metric: a case study using hedonic house price data. *International Journal of Geographical Information Science*, *28*(4), 660-681. http://doi.org/10.1080/13658816.2013.865739

Lu, B., Harris, P., Charlton, M., & Brunsdon, C. (2014b). The GWmodel R package: further topics for exploring spatial heterogeneity using geographically weighted models. *Geo-spatial Information Science*, *17*(2), 85-101.
https://doi.org/10.1080/10095020.2014.917453

Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of research and development*, *2*(2), 159-165.

Luo, J. and N. Kanala (2008). Modelling urban growth with geographically weighted multinomial logistic regression. *Proceedings of SPIE, the International Society for Optical Engineering*, *7144* , 1–11

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., ... & Adam, S. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, *12*(2-3), 93-118. https://doi.org/10.1080/19312458.2018.1430754

Manly, B. F., & Navarro, J. A. (2016). *Multivariate statistical methods: a primer*. Chapman and Hall/CRC.

Manning, C., & Schutze, H. (1999). *Foundations of statistical natural language processing*. MIT press.

Mason, G. A., & Jacobson, R. D. (2007, September). Fuzzy geographically weighted clustering. In *Proceedings of the 9th international conference on geocomputation, Maynooth, Eire, Ireland* (pp. 3-5).

Mei, C. L., Wang, N., & Zhang, W. X. (2006). Testing the importance of the explanatory variables in a mixed geographically weighted regression model. *Environment and Planning A*, *38*(3), 587-59. http://doi.org/10.1068/a3768

McMillen, D., & McMillen, M. D. (2013). Package 'McSpatial'. *Nonparametric Spatial Data Analysis*. Package 's McSpatial. *R software package*.

Mokdad, A. H., Dwyer-Lindgren, L., Fitzmaurice, C., Stubbs, R. W., Bertozzi-Villa, A., Morozoff, C., ... & Murray, C. J. (2017). Trends and patterns of disparities in cancer mortality among US counties, 1980-2014. *Jama*, *317*(4), 388-406.

http://doi.org/10.1001/jama.2016.20324

Moran, P. A. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, *10*(2), 243-251.

Mumu, J. R., Tahmid, T., & Azad, M. A. K. (2021). Job satisfaction and intention to quit: A bibliometric review of work-family conflict and research agenda. *Applied Nursing Research*, *59*, 151334. http://doi.org/10.1016/j.apnr.2020.151334

Murakami, D., Tsutsumida, N., Yoshida, T., Nakaya, T., & Lu, B. (2020). Scalable GWR: A linear-time algorithm for large-scale geographically weighted regression with polynomial kernels. *Annals of the American Association of Geographers*, *111*(2), 459-480. https://doi.org/10.1080/24694452.2020.1774350

Nakaya, T. (2001). Local spatial interaction modelling based on the geographically weighted regression approach. *GeoJournal*, *53*(4), 347-358. http://doi.org/10.1023/A:1020149315435

Nakaya, T., Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2005). Geographically weighted Poisson regression for disease association mapping. *Statistics in medicine*, *24*(17), 2695-2717. http://doi.org/10.1002/sim.2129

Nakaya, T., Fotheringham, S., Charlton, M., & Brunsdon, C. (2009). Semiparametric geographically weighted generalised linear modelling in GWR 4.0. In Lees, B.G. & Laffan, S.W. (eds), *10th International Conference on GeoComputation*, UNSW, Sydney, November-December, 2009

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, *135*(3), 370-384. http://doi.org//10.2307/2344614

Nicholas, D. E., Delamater, P. L., Waters, N. M., & Jacobsen, K. H. (2016). Geographically weighted discriminant analysis of environmental conditions associated with Rift Valley fever outbreaks in South Africa. *Spatial and spatio-temporal epidemiology*, *17*, 75-83. http://doi.org/10.1016/j.sste.2016.04.005

Nikita, M. *ldatuning: Tuning of the Latent Dirichlet Allocation Models Parameters*, R package version 1.0.2.; R Foundation for Statistical Computing: Vienna, Austria; Available online: **https://CRAN.R-project.org/package=ldatuning** (accessed on 16 April 2021).

Nurmala, N., & Purwarianti, A. (2017). Improvement of Fuzzy Geographically Weighted Clustering-Ant Colony Optimization Performance using Context-Based Clustering and CUDA Parallel Programming. *Journal of ICT Research and Applications*, *11*(1), 21-37. http://doi.org/10.5614/itbj.ict.res.appl.2017.11.1.2

Odeha, I. O. A., McBratney, A. B., & Chittleborough, D. J. (1994). Spatial prediction of soil properties from landform attributes derived from a digital elevation model. *Geoderma*, *63*(3-4), 197-214. http://doi.org/10.1016/0016-7061(94)90063-9

Oshan, T., Li, Z., Kang, W., Wolf, L., & Fotheringham, A. S. (2019). mgwr: A Python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. *International Journal of Geo-Information* 8(6),269. http://doi.org/10.3390/ijgi8060269

Páez, A. (2004). Anisotropic variance functions in geographically weighted regression models. *Geographical Analysis*, *36*(4), 299-314. http://doi.org/10.1111/j.1538-4632.2004.tb01138.x

Páez, A. (2006). Exploring contextual variations in land use and transport analysis using a probit model with geographical weights. *Journal of Transport Geography*, *14*(3), 167-176. https://doi.org/10.1016/j.jtrangeo.2005.11.002

Paré, G., Trudel, M. C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management*, *52*(2), 183-199. https://doi.org/10.1016/j.im.2014.08.008

Pearce, M. S. (1999). Geographically weighted regression: A method for exploring spatial nonstationarity. *Stata Technical Bulletin*, *8*(46).

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6*, *2*(11), 559–572. http://doi.org/10.1080/14786440109462720

Porter, M.F. (1980). An algorithm for suffix stripping. *Programming* , *14*, 130–137. https://doi.org/10.1108/eb046814

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URLhttps://www.R-project.org/.

Rangel, T. F., Diniz-Filho, J. A. F., & Bini, L. M. (2010). SAM: a comprehensive application for spatial analysis in macroecology. *Ecography*, *33*(1), 46-50. http://doi.org/10.1111/j.1600-0587.2009.06299.x

Rieger, J. (2020). ldaPrototype: Prototype of Multiple Latent Dirichlet Allocation Runs. R package version 0.1. 1.

Roca-Pardiñas, J., Ordóñez, C., Cotos-Yáñez, T. R., & Pérez-Álvarez, R. (2016). Testing spatial heterogeneity in geographically weighted principal components analysis. *International Journal of Geographical Information Science,* 31(4), 676–693. http://doi.org/10.1080/13658816.2016.1224886

Röder, M., Both, A., & Hinneburg, A. (2015, February). Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining* (pp. 399-408).

Salvati, N., Tzavidis, N., Pratesi, M., & Chambers, R. (2009). Small area estimation via M-quantile geographically weighted regression. *Centre for Statistical and Survey Methodology, University of Wollongong, Working Paper 09-09, 2009, 30p.* http://ro.uow.edu.au/cssmwp/29

Salvati, N., Tzavidis, N., Pratesi, M., & Chambers, R. (2010). Small area estimation via M-quantile geographically weighted regression. *Test*, *21*(1), 1-28. https://doi.org/10.1007/s11749-010-0231-1

Sánchez-Peña, L. L. (2012). Alcances y límites de los métodos de análisis espacial para el estudio de la pobreza urbana. *Papeles de población*, *18*(72), 147-180.

Sbalchiero, S., & Eder, M. (2020). Topic modeling, long texts and the best number of topics. Some Problems and solutions. *Quality & Quantity*, *54*(4), 1095-1108.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461-464. http://doi.org/10.1214/aos/1176344136

Semecurbe, F., Roux, S. G., Tannier, C., & Semecurbe, M. F. (2016). Package 'gwfa'. *R software package*.

Shing, H. (2008). *Incorporating the Concept of'community'Into a Spatially-weighted Local Regression Analysis* (Doctoral dissertation, University of New Brunswick, Department of Geodesy and Geomatics Engineering).

Shrestha, A., & Luo, W. (2017). Analysis of groundwater nitrate contamination in the Central Valley: comparison of the geodetector method, principal component analysis and geographically weighted regression. ISPRS *International Journal of Geo-Information*, 6(10), 297.

Sievert, C.& Shirley, K. *LDAvis: Interactive Visualization of Topic Models*, R package version 0.3.2.; R Foundation for Statistical Computing: Vienna, Austria; Available online: https://CRAN.R-project.org/package=LDAvis (accessed on 16 April 2021).

Son, L. H., Cuong, B. C., Lanzi, P. L., & Thong, N. T. (2012). A novel intuitionistic fuzzy clustering method for geo-demographic analysis. *Expert Systems with Applications, 39(10),* 9848–9859. https://doi.org/10.1016/j.eswa.2012.02.167

Son, L. H. (2015). A novel kernel fuzzy clustering algorithm for geo-demographic analysis. *Information Sciences—Informatics and Computer Science, Intelligent Systems, Applications: An International Journal*, *317*(C), 202-223. https://doi.org/10.1016/j.ins.2015.04.050

Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society: Series B (Methodological)*, *50*(3), 413-436. http://doi.org/10.1111/j.2517-6161.1988.tb01738.x

Srivastava, A. N., & Sahami, M. (Eds.). (2009). *Text mining: Classification, clustering, and applications*. CRC press.

Subedi, N., Zhang, L., & Zhen, Z. (2018). Basiian geographically weighted regression and its application for local modeling of relationships between tree variables. *iForest-Biogeosciences and Forestry*, *11*(5), 542. https://doi.org/10.3832/ifor2574-011

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267-288. http://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tiefelsdorf, M., Griffith, D. A., Boots, B. (1999). A variance-stabilizing coding scheme for spatial link matrices, Environment and Planning A, 31, pp. 165–180. https://doi.org/10.1068/a310165

Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, *46*, 234. http://doi.org/10.2307/143141.

Tsutsumida, N., Murakami, D., Yoshida, T., Nakaya, T., Lu, B., & Harris, P. (2019). Geographically Weighted Non-negative Principal Components Analysis for Exploring Spatial Variation in Multidimensional Composite Index. *GeoComputation*

Van Donkelaar, A., Martin, R. V., Brauer, M., Hsu, N. C., Kahn, R. A., Levy, R. C. & Winker, D. M. (2016). Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. *Environmental science & technology*, *50*(7), 3762-3772.
https://doi.org/10.1021/acs.est.5b05833

Vanhala, M., Lu, C., Peltonen, J., Sundqvist, S., Nummenmaa, J. & Järvelin, K. (2020) The usage of large data sets in online consumer behaviour: A bibliometric and computational text-mining–driven analysis of previous research. *Journal of Business Research*, 106, 46–59. http://doi.org/10.1016/j.jbusres.2019.09.009

Vicente-Villardón, J. L. (2021). MultBiplotR: Mulivariate analysis using biplots. R package version 1.3.30.; Foundation for Statistical Computing: Vienna, Austria; Available online: https://CRAN.R-project.org/package=MultBiplotR.

Vijayarani, S., Ilamathi, M. J., & Nithya, M. (2015). Preprocessing techniques for text mining-an overview. *International Journal of Computer Science & Communication Networks*, *5*(1), 7-16.

Voss, P. R., Long, D. D., Hammer, R. B., & Friedman, S. (2006). County child poverty rates in the US: a spatial regression approach. *Population Research and Policy Review*, *25*(4), 369-391. https://doi.org/10.1007/s11113-006-9007-4

Wagner, C. S., Bornmann, L., & Leydesdorff, L. (2015). Recent developments in China–US cooperation in science. *Minerva*, *53*(3), 199-214. : https://doi.org/10.1007/s11024-015-9273-6

Wang, J., Kang, S., Sun, J., & Chen, Z. (2013). Estimation of crop water requirement based on principal component analysis and geographically weighted regression. *Chinese Science Bulletin*, *58*(27), 3371-3379. https://doi.org/10.1007/s11434-013-5750-1

Wang, W., Xu, S., & Yan, T. (2018). Structure identification and model selection in geographically weighted quantile regression models. *Spatial Statistics, 26*, 21–37. https://doi.org/10.1016/j.spasta.2018.05.003

Wheeler, D. C. (2007). Diagnostic Tools and a Remedial Method for Collinearity in Geographically Weighted Regression. *Environment and Planning A*, *39*(10), 2464–2481. http://doi.org/10.1068/a38325

Wheeler, D. C. (2009). Simultaneous coefficient penalization and model selection in geographically weighted regression: the geographically weighted lasso. *Environment and planning A*, *41*(3), 722-742. http://doi.org/10.1068/a40256.

Wheeler D.C. (2019) Geographically Weighted Regression. In: Fischer M., Nijkamp P. (eds) *Handbook of Regional Science*. Springer, Berlin, Heidelberg. http://doi.org/10.1007/978-3-642-36203-3_77-1

Wheeler, D., & Tiefelsdorf, M. (2005). Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems*, *7*(2), 161-187. https://doi.org/10.1007/s10109-005-0155-6

Wijayanto, A. W., & Purwarianti, A. (2016). Fuzzy geographically weighted clustering using artificial bee colony: An efficient geo-demographic analysis algorithm and applications to the analysis of crime behavior in population. *Applied Intelligence*, *44*(2), 377-398. https://doi.org/10.1007/s10489-015-0705-7

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, *41*, 434-449. https://doi.org/10.2307/2332724

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.
Xiong, H., Cheng, Y., Zhao, W., & Liu, J. (2019). Analyzing scientific research topics in manufacturing field using a topic model. *Computers & Industrial Engineering*, *135*, 333-347. https://doi.org/10.1016/j.cie.2019.06.010

Yang, W. (2014). *An extension of geographically weighted regression with flexible bandwidths* (Doctoral dissertation, University of St Andrews).

Yano, T., Smith, N. A., & Wilkerson, J. (2012, June). Textual predictors of bill survival in congressional committees. In *proceedings of the 2012 conference of the north American chapter of the Association for Computational Linguistics: human language technologies* (pp. 793-802).

Yoneoka, D., Saito, E., & Nakaoka, S. (2016). New algorithm for constructing area-based index with geographical heterogeneities and variable selection: An application to gastric cancer screening. *Scientific reports*, *6*(1), 1-7.

Yrigoyen, C. C. (2004). Modelos de heterogeneidad espacial. *University Library of Munich, Germany, Tech. Rep.*

Zhao, W., Chen, J. J., Perkins, R., Liu, Z., Ge, W., Ding, Y., & Zou, W. (2015, December). A heuristic approach to determine an appropriate number of topics in topic modeling. In *BMC bioinformatics* (Vol. 16, No. 13, pp. 1-10). BioMed Central.

.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, *67*(2), 301-320. http://doi.org/10.1111/j.1467-9868.2005.00503.x

ANEXOS

ANEXO 1. Artículo "Trends and topics in geographically weighted regression research from 1996 to 2019"

Received: 7 November 2020 | Revised: 6 July 2021 | Accepted: 8 October 2021

DOI: 10.1111/area.12757

AREA Royal Geographical Society with IBG
WILEY

REGULAR PAPER

Trends and topics in geographically weighted regression research from 1996 to 2019

Javier De La Hoz-M[1] | **María José Fernandez-Gómez**[2] | **Susana Mendes**[3]

[1]Universidad del Magdalena, Santa Marta, Colombia

[2]Department of Statistics, University of Salamanca, Salamanca, Spain

[3]MARE, School of Tourism and Maritime Technology, Polytechnic of Leiria, Peniche, Portugal

Correspondence
Javier De La Hoz, Universidad del Magdalena, Santa Marta, Colombia.
Email: jdelahozmaestre@gmail.com

Funding information
There are no funders to report for this submission.

Abstract

This research was conducted in order to improve the understanding of the structure, contents, and trend of topics within the existing literature in the field of geographically weighted regression. Additionally, it intended to determine and produce a mapping of scientific networks in the domain of geographically weighted regression. The proposed methodology implements a combination of bibliometric techniques and modelling of topics in order to extract the latent topics from the collected literature by utilising latent Dirichlet allocation and a machine learning tool. The results identified the most prolific authors, the most cited authors, the most representative articles and journals, and the countries which are responsible for the publications.

KEYWORDS

bibliometric analysis, collaboration pattern, geographically weighted regression, latent Dirichlet allocation, machine learning, topic modelling

1 | INTRODUCTION

Geographically weighted regression (GWR) is a technique of spatial analysis introduced in the in the 1990s in the field of geography (Brunsdon et al., 1996). Unlike the classical regression model, where the estimated coefficients are constant in space, GWR can model relationships that vary spatially between the independent variables and the dependent variable. In the field of analytical geography, GWR implements, within a locally linear regression framework, the distance-decay effect popularised by Tobler with the first law of geography (Paéz & Wheeler, 2009).

Due importance, complexity, and increasing demand for the application of spatial data models in different research, several models have been introduced to extend the concept of GWR (Table 1). In addition to these extensions, GWR has also expanded its application to a wide range of research areas, including social science, environmental science, ecology, and health sciences (among others). The extent of the models outlined, in combination with the growing worldwide increase in scientific literature, can cause researchers to become overwhelmed and thus their ability to carry out a review and follow-up of new research is effectively diminished (Larsen & von Ins, 2010).

In the face of the increasing diversity of research topics in this field, there is a need for quantitative studies that help better understand the following issues:

Q1: What were the primary publication sources, as well as major contributors in GWR research?
Q2: What were the scientific collaborations between countries in GWR research?

Q3: What are the major research topics in this field?
Q4: How do these research topics evolve with time?
Q5: What are the distributions of these topics across countries and journals?

Answers to these questions can provide a comprehensive depiction and a state-of-the-art understanding of GWR research. To address these issues, this study used the combination of advanced topic modelling algorithms – an unsupervised machine-learning algorithm called latent Dirichlet allocation (LDA; Blei et al., 2003) which has proven to be useful in the identification of scientific topics related to topics or existing categories in a set of articles (Griffiths & Steyvers, 2004) – and bibliometric analysis. This enables a more highly objective, robust, structured, and comprehensive review of this rapidly expanding research domain than traditional literature reviews and analyses can provide (Chen et al., 2019; Vanhala et al., 2020).

2 | METHOD

We performed a topic modelling-based bibliometric exploration of peer-reviewed literature reflecting the global status and trend of GWR research with 1588 target articles published from 1996 to 2019. The methods and general procedure used during the research are depicted in the methodological supplementary material (Figure S1). In the following subsections, we provide details of the methods and procedures.

2.1 | Dataset preparation

The search for articles was carried out through Scopus. We decided to select this database because it is one of the databases most used by researchers (Harzing & Alakangas, 2016). We systematically identified keywords for use in our study by performing searches using keyword combinations for research articles containing information discussing the technicalities of GWR and/or its extensions (Table 1), published in English in the period October 1996 (date of first publication) to December 2019. Only peer-reviewed articles were considered. Searching was conducted using the following query:

TABLE 1 Extensions of geographically weighted regression

Reference	Model/method
Brunsdon et al. (1999)	Mixed geographically weighted regression
Fotheringham et al. (2002)	Heteroscedastic geographically weighted regression
Fotheringham et al. (2002)	Geographically weighted generalised linear modelling
Atkinson et al. (2003)	Geographically weighted logistic regression
LeSage (2004)	Bayesian geographically weighted regression
Nakaya et al. (2005)	Geographically weighted Poisson regression
Wheeler (2007)	Geographically weighted ridge regression
Luo and Kanala (2008)	Geographically weighted multinomial logistic regression
Wheeler (2009)	Geographically weighted lasso
Huang et al. (2010); Fotheringham et al. (2002)	Geographically and temporally weighted regression
da Silva and Rodrigues (2014)	Geographically weighted negative binomial regression
Yang (2014)	Geographically weighted regression with flexible bandwidths
Lu et al. (2017)	Geographically weighted regression with parameter-specific distance metrics
da Silva and Oliveira (2017)	Geographically weighted beta regression
Li and Lam (2018)	Geographically weighted elastic net
Wang et al. (2018)	Geographically weighted quantile lasso
Chybicki (2018)	Three-dimensional geographically weighted regression
Dong et al. (2018)	Geographically weighted ordinal regression

TITLE-ABS-KEY (“Geographically Weighted regression” OR “Heteroscedastic GWR” OR “Geographically Weighted Generalized Linear Modeling” OR “Geographically Weighted Logistic Regression” OR “Geographically Weighted Poisson Regression” OR “Geographically Weighted Multinomial” OR “Geographically Weighted Negative Binomial Regression” OR “Geographically and Temporally Weighted Regression” OR “Geographically Weighted Ridge Regression” OR “Geographically Weighted Lasso” OR “Geographically Weighted Elastic Net” OR “Geographically Weighted Beta Regression” OR “Geographically Weighted Quantile Lasso” OR “Geographically Weighted Ordinal Regression”) AND (LIMIT-TO(DOCTYPE, “ar”)) AND (LIMIT-TO(LANGUAGE, “English”)) AND (EXCLUDE (PUBYEAR, 2021)) OR (EXCLUDE(PUBYEAR, 2020))

The preliminary database with the documents obtained after executing the search query contained 1,602 documents. This initial sample was subjected to a filtering process where repeated (one article) and misclassified articles (two articles with different languages) were eliminated. Also some documents could not be downloaded or were in image format, which prevented them from being processed (21 articles). The final sample obtained consisted of 1,588 articles. Subsequently, each of the 1,588 articles was downloaded in pdf format.

2.2 | Bibliometrics and research mapping

To answer Q1 and Q2, we adopted bibliometric analysis. This allows various aspects of scientific production to be determined using intellectual and social structures of the sample of publications (Aria & Cuccurullo, 2017; Cobo et al., 2011).

Data processing in this part of the study was carried out using bibliometrix (Aria & Cuccurullo, 2017), an opensource package in R programming language (R Core Team 2019), while for visualisation of the networks, the VOSviewer (Van Eck & Waltman, 2010) version 1.6.15 program was used.

2.3 | Identifying research topics

To answer Q3, Q4, and Q5, the topic model method LDA (Blei, 2012; Blei et al., 2003) was used. It is based on Bayesian models and is considered to be an extension of Probabilistic Latent Semantic Analysis (Blei et al., 2003; Hornik & Grün, 2011) (see methodological supplementary material, Figure S2).

We decided to use the texts of the complete articles since, in contrast to the exclusive use of abstracts, it has been shown that this increases the quality of the topics and provides greater detail of the latent themes that can be found in a document (Syed & Spruit, 2017).

The procedure for the identification of topics through LDA was divided into three stages: (i) preprocessing, (ii) creation of LDA model, and (iii) labelling topics (see methodological supplementary material, Figure S3).

Preprocessing texts: This phase plays a very important role and is generally the first procedure in text mining techniques and applications (Vijayarani et al., 2015). Each article was tokenised using bigrams, numbers, punctuation marks, and blank spaces, and words from a standard list called stop-words were eliminated. In addition, all the words were converted to lowercase and stemming was applied. These procedures resulted in a document terms matrix (dtm) with 5,870 unique terms. All stages of the preprocessing were made in the text mining package tm (Feinerer & Hornik, 2019), Software R (R Core Team, 2019).

*Creation model Latent Dirichl*et al*location*: Topic models are latent variable models of documents that use correlations between words and latent semantic themes in a collection of documents (Blei & Lafferty, 2007). This definition assumes that the expected number of topics k (i.e., latent variables) must be established a priori. Simulations were carried out varying k from 4 to 30 in incremental steps of two. An inference algorithm with 500 iterations was used, namely Gibbs sampling (Geman & Geman, 1984). Also, default values from the ‘textmineR’ (Jones, 2019) package were used for Dirichlet parameters α and ß was estimated based on the corpus. The quality of the LDA model was determined by utilising a topic coherence measure (Röder et al., 2015).

Labelling topics: First of all we use a naive labelling algorithm based on probable bigrams (these naive labels are based on P(bi-gram|topic)−P(bi-gram)) provided by the package textmineR (Jones, 2019). However, given that algorithms have a very limited ability to understand the latent meanings of human language (Lau et al., 2011), also it was decided to use manual labelling, which is considered a standard in topic modelling by using two sources of information: the 15 most

frequent word lists (most likely) and a sample of the titles, then summaries of the three most loaded articles. In addition, to improve the labelling of the topics, we visualised them in a two-dimensional area by computing the distance between topics (Chuang et al., 2012) by means of a multidimensional scaling analysis (Siever & Shirley, 2014). This analysis displays the similarity between topics with respect to their probability distribution over words.

2.4 | Quantitative indices used to analyse the trend of topics

Due to the large number of articles and, therefore, the quantity of words, it is difficult to understand the topics and their trends intuitively. Therefore, we used some quantitative indices proposed by Xiong et al. (2019), which are obtained by adding documents-topic and topic-words distributions in order to make clear the results and findings. The description of the indexes is as follows. The distribution of topics over time is obtained by:

$$\theta_k^y = \frac{\sum_{m \epsilon y} \theta_{mk}}{n^y} \tag{1}$$

where $m\epsilon y$ represents articles published in a given year, θ_{mk} the proportion of the k-th topic in each item, and n^y the total number of articles published in the year (Xiong et al., 2019).

Topic distribution across journals is defined as the ratio of the k-th topic in the journal j:

$$\theta_k^j = \frac{\sum_{m \epsilon j} \theta_{mk}}{n^j} \tag{2}$$

where $m\epsilon j$ represents the articles in a particular journal, θ_{mk} the proportion of the k-th topic on each item, and n^j the total number of articles published in the journal j.

With the purpose of facilitating the characterisation of the topics in terms of their tendency, we used simple regression slopes for each topic, where the year was a dependent variable and the proportion of the topics in the corresponding year was the response variable (Griffiths & Steyvers, 2004). The topics obtained by regression were positive or negative at a statistical significance level of 0.01 and were classified as positive or negative trends respectively.

In order to quantify the activity and influence of the topics, we calculated the popularity of topics. The popularity of a topic, taking into account trends and the probability of the topic, can be described quantitatively as follows:

$$P^i = S^i_{\mathrm{NP}} + S^i_{Tr} \tag{3}$$

$$S^i_{\mathrm{NP}} = P^i_A / P^{\max}_A \tag{4}$$

where P^i is the popularity of a topic i, S^i_{NP} is the normalised probability, and S^i_{Tr} takes values of 1 if a topic shows a positive trend, 0.67 fluctuating, and 0.33 negative trend (Xiong et al., 2019).

The final sample obtained from 1,588 articles with their metavariables, as well as the dtm and θ matrix can be found and downloaded in CSV format at: https://github.com/JavierDeLaHoz/Trends-and-topics-in-Geographically-Weighted-Regression-research-from-1996-to-2019

3 | RESULTS

3.1 | Bibliometric analysis

The summary generated includes basic statistics about the analysed dataset and is presented in the "Full results" supplementary material (Table S1). Documents stemmed from 594 different journals and were published over 24 years. A total of 3,791 authors were involved in the scientific production on GWR. Among the papers, 105 were created by a single author, whereas the overall Collaboration Index of the sample equals 2.5.

Table S2 (see Full results supplementary material) provides the list of top-20 most-cited authors. Of all the authors in the 1,588 articles, the three most cited are: Fotheringham AS, Brunsdon C, and Charlton M. All of them exceed 3,500 citations.

In addition to focusing on the number of citations, we also inspected the most important papers, in terms of global citations. The results are presented in Table S3 (see "Full results" supplementary material).

Table S4 (see "Full results" supplementary material) shows the most influential journals in terms of article count. These journals are distributed over different subject areas, such as environmental science, engineering, health, science, economics, and social sciences. The journals with the greatest number of published articles are *Applied Geography*, *Sustainability (Switzerland)*, and *International Journal of Environmental Research and Public Health*.

A world map of the scientific production (top of Figure 1) indicates research on GWR has been conducted on all continents (95 countries). The USA and China lead in this regard, with 597 and 502 articles, respectively. The United Kingdom (UK), Canada, and Australia appear in third, fourth, and fifth positions, with 140, 90, and 76 articles, respectively. The productivity of the countries is shown from the perspective of fractionalised production, indicating that a document was developed by several countries with each of them represented by a fraction of the authorship. Visual inspection of the map shows that from the African countries, only 11 countries contributed, among them South Africa, Egypt, Zimbabwe, and Ghana. Latin American participation is limited mainly to Brazil, Argentina, Chile, Colombia, Ecuador, Costa Rica, and Mexico.

Figure 1 (middle and lower part of the figure) shows the network of international co-author relationship among 90 countries, the largest set connected. Trinidad and Tobago, Latvia, Hungary, Cuba, and Tunisia do not present authorship connections with other countries. There are ten colours in Figure 1, indicating that these 90 countries are grouped into ten clusters. There is a stronger cooperative relationship between countries in the same cluster than between countries in different clusters. Of course, this does not mean that there is no cooperation between countries in different clusters, rather there may be some common research topics among the countries of the same cluster, making their cooperation closer. More frequent connections (represented by the thickness of the line) could be observed between the USA and China.

Table S5 (see Full results supplementary material) shows clusters and the affinity of the collaboration of countries. In this a link is a connection or a relation between two items (countries). Each link has a strength, represented by a positive numerical value. The higher this value, the stronger the link. For a given item, the Links and Total link strength attributes indicate, respectively, the number of links of an item with other items and the total strength of the links of an item with other items.

3.2 | Topic modelling analysis

The coherence score for all evaluated LDA models suggests that the LDA model with the optimal coherence score contains 20 subjects ($k = 20$).

Table S6 (see "Full results" supplementary material) provides an overview of the complete set of articles or corpus used in this study. The top-25 terms (unigram and bigram) with high occurrence frequency are listed in Table S7 (see "Full results" supplementary material). We observed that the words found in the articles considered represent the variety of topics investigated in the field of GWR.

Table S8 (see "Full results" supplementary material) presents the 20 topics estimated by our model and for each of them the 15 most common terms and topic label organised from most prevalent to least prevalent. Figure 2 provides an additional representation of the 20 uncovered GWR topics, alongside their proportions across the entire corpus. Here, the distance between the nodes represents the topic similarity with respect to the distribution of words, while the size of the nodes indicates the topic prevalence within the corpus, with larger nodes representing topics being more prominent within the corpus.

Trends of topics: We identified the tendency of each of the 20 topics over time. We found that the probabilities of 12 of them increased progressively over time, decreased in two of them, while the remainder fluctuated over time without prominent trends (Figure 3).

Although many journals included in our analysis overlap to some extent in their content, it is possible to identify journals that seem specialised in specific topics. The results for 25 of them are grouped according to the topics they addressed (Figure 4). There is some similar prevalence distribution of topics, for example *Accident Analysis and Prevention* (*Accid. Anal. Prev.*) and *Journal of Transport Geography* (*J. Transp. Geogr.*), which were focused on topics "Transport" and "Spatial variation".

Popularity topics: The popularity of the topics, taking into account the trend and the probability of the topic, are presented in Table 2. It is observed that the five most popular topics in descending order are: "GWR model," "Population

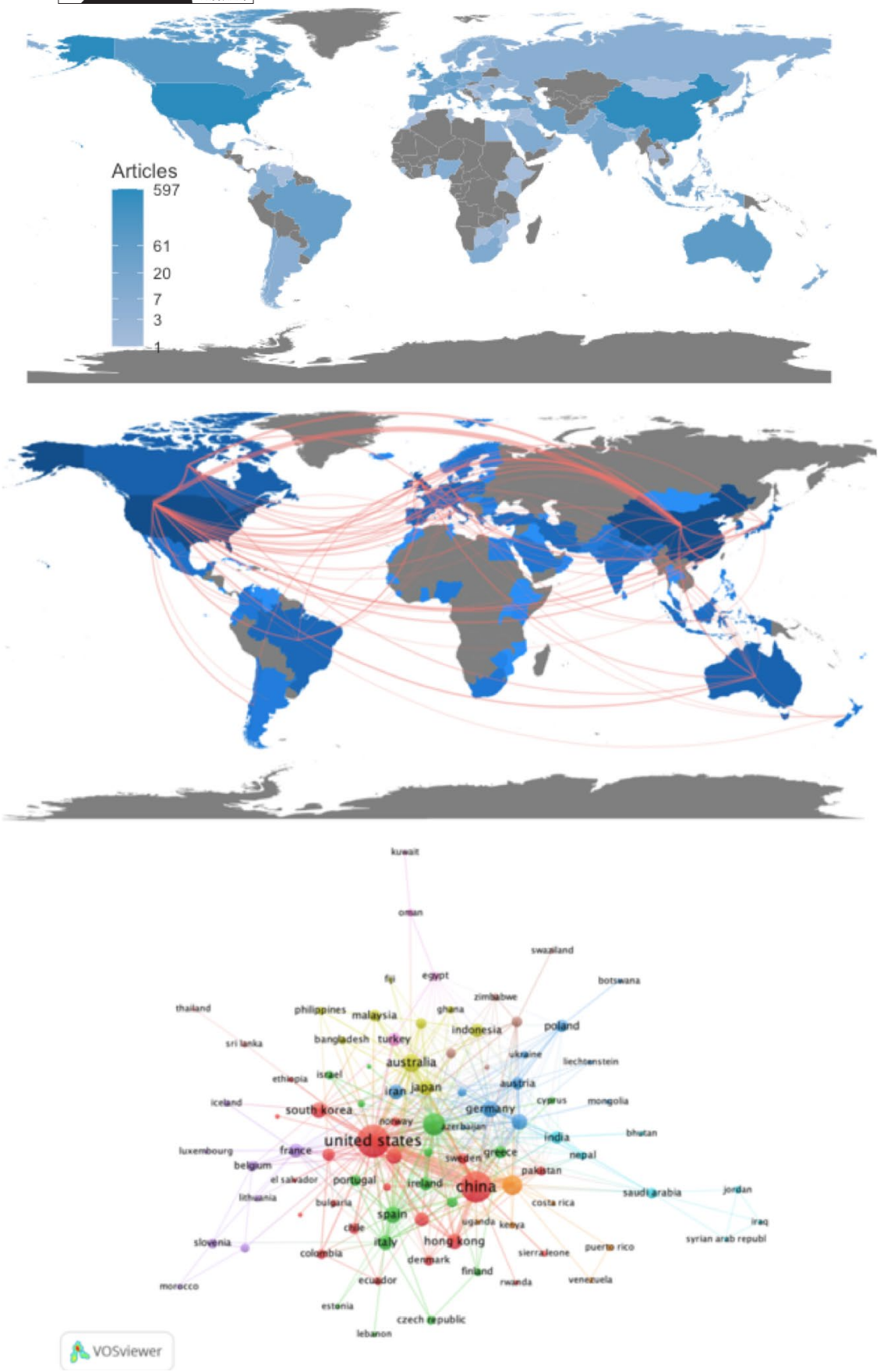


FIGURE 1 Worldwide distribution of articles and international co-authorship relationships on geographically weighted regression research between 1996 and 2019. Grey colour in countries or regions means that no publications were registered

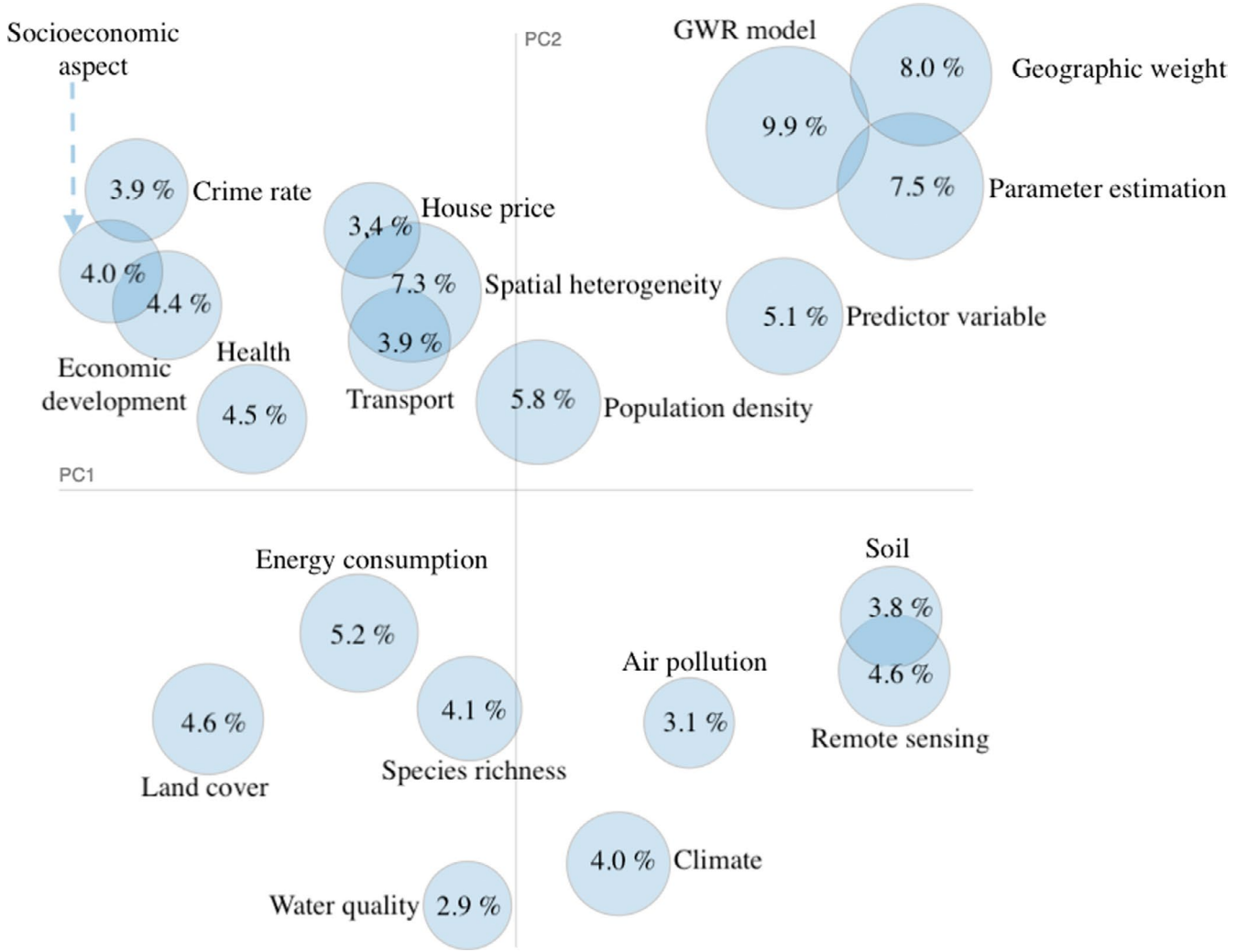


FIGURE 2 Inter-topic distance map that shows a two-dimensional representation via multidimensional scaling (all nodes add up to 100%)

density," "Health," "Energy consumption," "Specie richness," and "Remote sensing." In general, topics with increasing trends with high and moderate average probabilities have higher popularity scores, while some topic with moderate probabilities and increasing trends improved considerably.

Figure 5 details the topical foci of the top-25 countries. These were classified into five groups (based on the distribution of the probabilities of the topics). The largest group is made up of 13 countries, while the smallest is made up of Saudi Arabia.

4 | DISCUSSION

In this study, we have provided a comprehensive perspective on the evolution and development of research in the field of GWR. Based on the use of bibliometric methods and topic modelling techniques, we have analysed GWR publication patterns, geographic distribution, source journals, and international collaboration.

Similar to global bibliometric trends (Bornmann & Mutz, 2015), our results show that the GWR research landscape has evolved over the selected time period (1996–2019). The accelerated rate of growth of publications in countries with emerging economies (such as China, Brazil, and India) is due to the availability of funds for research and the increase in scientific and technological capacities (National Science Foundation (NSF), 2018).

In this study, we used the affiliation of the first author or corresponding author to define the country in which the studies were conducted. In the context of increased international collaboration, this assumption could have led to a

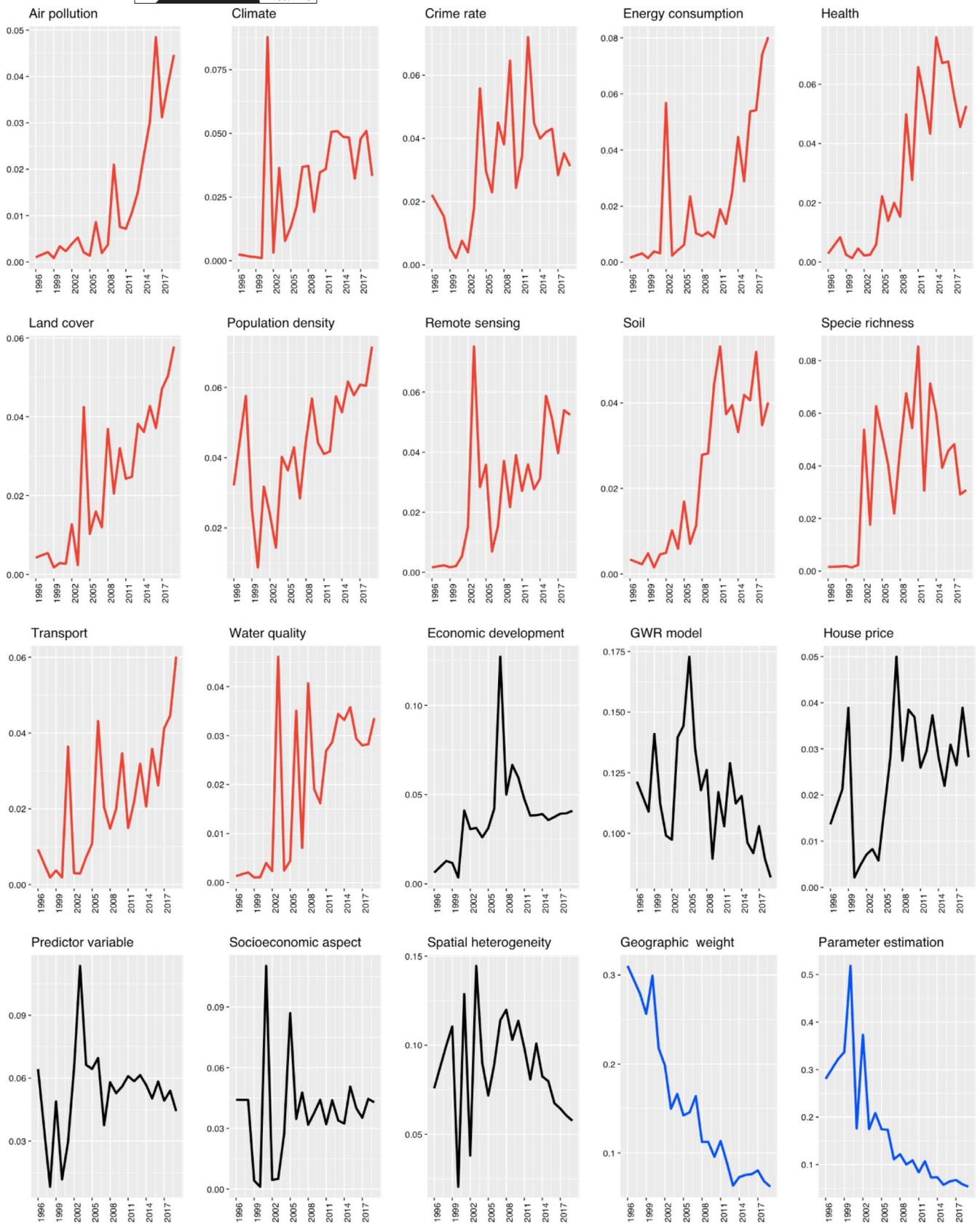


FIGURE 3 Topic trends research in geographically weighted regression during October 1996–December 2019. The red colour indicates topics with increasing tendency, blue with decreasing tendency, and black fluctuating

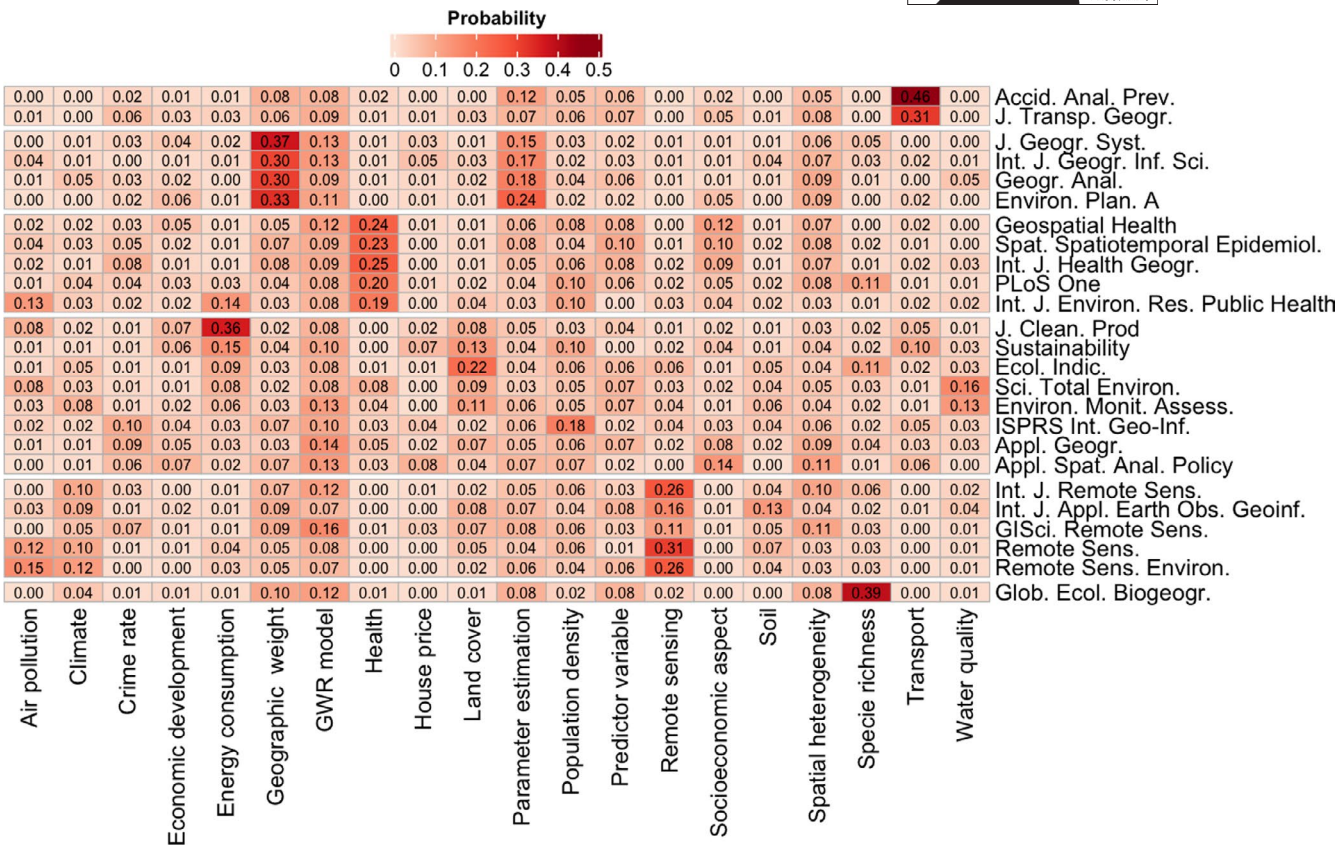


FIGURE 4 Heatmap overview of the proportional topic in the top-25 analysed journals. Values are in percentages and row totals sum up to 100%

geographic misrepresentation of the GWR's scientific output in favour of some countries that may carry out international research projects. We observed that, as in most other research fields, the USA and China were the main contributors in the field of GWR; in this regard, China has shown remarkable general growth in research production during the last decade (Aksnes et al., 2014).

International collaboration is highly concentrated in the USA, China, and the United Kingdom. However, China is the USA's top partner in terms of international co-authorship of GWR-related publications. This is consistent with the general trend in science (Wagner et al., 2015), in which China stood out as a major American contributor, surpassing the historical collaborative relationships between Europe and North American nations in other fields. Despite the fact that collaboration is widely accepted as a positive (Adams, 2013), some scholars have also argued that the way science is evolving follows earlier patterns of Western prejudice and domination (Peters, 2006). The homogenisation of a network can be aided by strong networks (Bodin & Crona, 2009). The more connected collaborators are with each other, the less likely new partnerships with "different others" will result, making introduction to new ideas equally unlikely (Lambiotte & Panzarasa, 2009). However, collaborations between academic communities in different geographic areas may have brought new perspectives and tools to established research approaches and contributed to shifting the interest of GWR research from model-related issues to technique application issues.

We quantitatively analysed and identified 20 latent themes and their trends. The data were presented in the form of topics distributions. These revealed how the scope of research in GWR has varied over time, increasing in the proportion of some topics (Population density, Health, Energy consumption, Specie richness, Remote sensing, Land cover, Climate, Soil, Crime rate, Transport, Air pollution, and Water quality), indicating that these are emerging research fields, while a decrease (Geographic weight, Parameter estimation) in a trend indicates a topic of less research interest. Furthermore, in some subjects the high frequency pattern found was followed by a negative trend during the study period (GWR model, Spatial heterogeneity, Predictor variable, Economic development, Socioeconomic aspect, and House price), which indicates a possible decrease in its popularity within the scientific community.

TABLE 2 Topic popularity

Topic	Probability	Normalised probability	Trend	Popularity
GWR model	0.1000	1.0000	0.67	1.6700
Population density	0.0578	0.5776	1	1.5776
Health	0.0527	0.5273	1	1.5273
Energy consumption	0.0489	0.4885	1	1.4885
Specie richness	0.0436	0.4360	1	1.4360
Remote sensing	0.0433	0.4331	1	1.4331
Land cover	0.0419	0.4194	1	1.4194
Spatial heterogeneity	0.0749	0.7484	0.67	1.4184
Climate	0.0412	0.4123	1	1.4123
Soil	0.0387	0.3871	1	1.3871
Crime rate	0.0382	0.3822	1	1.3822
Transport	0.0370	0.3704	1	1.3704
Air pollution	0.0300	0.2995	1	1.2995
Water quality	0.0294	0.2941	1	1.2941
Predictor variable	0.0529	0.5293	0.67	1.1993
Geographic weight	0.0808	0.8078	0.33	1.1378
Economic development	0.0421	0.4207	0.67	1.0907
Parameter estimation	0.0753	0.7529	0.33	1.0829
Socioeconomic aspect	0.0407	0.4064	0.67	1.0764
House price	0.0305	0.3048	0.67	0.9748

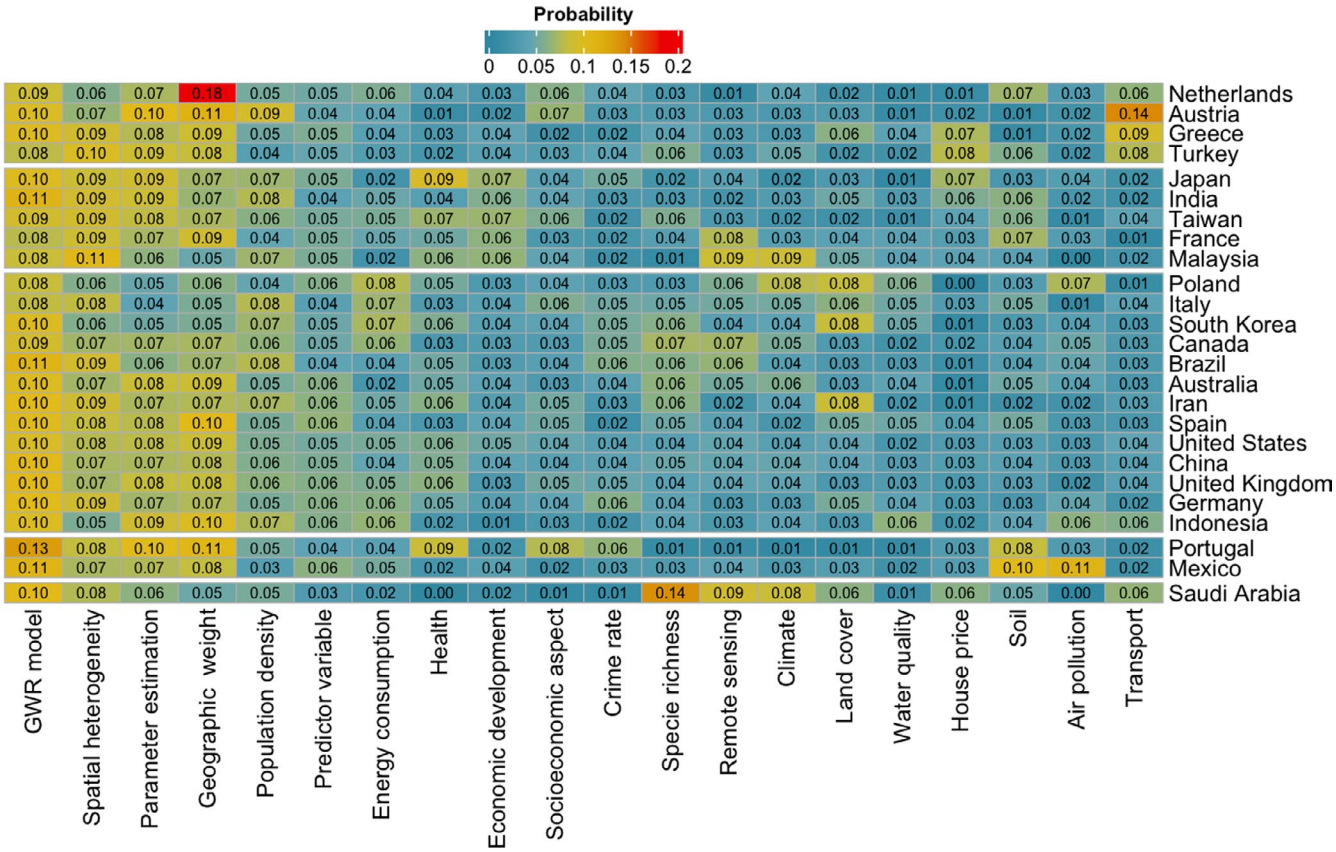


FIGURE 5 Heatmap overview of the proportional topic in the 25-top analysed countries. Values are in percentages and row totals sum up to 100%

The above results show that the GWR technique has been used by researchers who focus on current concerns such as climate, environment, and health. According to Cao et al. (2009) and Zhong et al. (2013), there are direct and indirect links between the listed topics. Furthermore, Beggs and Bambrick (2005) claim that there are direct links between climate change and disease.

Our results also show that the GWR theory and method are widely used and have attracted much attention from academics from academics from different disciplines. The results are consistent with the idea that research shows strong trends, with topics rising and falling regularly in the field prevalence (Griffiths & Steyvers, 2004).

It should be noted that this study is limited by the exclusion of books, reviews, book chapters, grey literature, and reports from the research. The data were collected only from the Scopus database and only articles were taken into account. For future research, academics might consider conducting an analysis using other databases as well, such as ISI Web of Science, which include non-indexed journals that are unavailable in Scopus. More articles could also be found on Google Scholar, for example. Related to this, future research could also compare the results obtained from other databases against the results of this study.

ACKNOWLEDGEMENTS

I would like to thank the anonymous reviewers for the productive comments that greatly improved the paper.

DATA AVAILABILITY STATEMENT

The final sample, obtained from 1588 articles with their metavariables, as well as the dtm and θ matrix, can be found and downloaded in CSV format at: https://github.com/JavierDeLaHoz/Trends-and-topics-in-Geographically-Weighted-Regression-research-from-1996-to-2019

ORCID

Javier De La Hoz-M https://orcid.org/0000-0001-7779-0803
María José Fernandez-Gómez https://orcid.org/0000-0002-5530-6416
Susana Mendes https://orcid.org/0000-0001-9681-3169

REFERENCES

Adams, J. (2013) The fourth age of research. *Nature*, 497(7451), 557–560. Available from: https://doi.org/10.1038/497557a

Aksnes, D.W., van Leeuwen, T.N. & Sivertsen, G. (2014) The effect of booming countries on changes in the relative specialization index (RSI) on country level. *Scientometrics*, 101(2), 1391–1401. Available from: https://doi.org/10.1007/s11192-014-1245-3

Aria, M. & Cuccurullo, C. (2017) Aibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975. Available from: https://doi.org/10.1016/j.joi.2017.08.007

Atkinson, P.M., German, S.E., Sear, D.A. & Clark, M.J. (2003) Exploring the relations between riverbank erosion and geomorphological controls using geographically weighted logistic regression. *Geographical Analysis*, 35(1), 58–82. Available from: https://doi.org/10.1111/j.1538-4632.2003.tb01101.x

Beggs, P.J. & Bambrick, H.J. (2005) Is the global rise of asthma an early impact of anthropogenic climate change? *Environmental Health Perspectives*, 113(8), 915–919. Available from: https://doi.org/10.1289/ehp.7724

Blei, D.M. (2012) Probabilistic topic models. *Communications of the ACM*, 55(4), 77–84. Available from: https://doi.org/10.1145/2133806.2133826

Blei, D.M. & Lafferty, J.D. (2007) A correlated topic model of science. *The Annals of Applied Statistics*, 1(1), 17–35. Available from: https://doi.org/10.1214/07-AOAS114

Blei, D.M., Ng, A.Y. & Jordan, M.I. (2003) Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(1), 993–1022.

Bodin, Ö. & Crona, B.I. (2009) The role of social networks in natural resource governance: What relational patterns make a difference? *Global Environmental Change*, 19(3), 366–374. Available from: https://doi.org/10.1016/j.gloenvcha.2009.05.002

Bornmann, L. & Mutz, R. (2015) Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, 66(11), 2215–2222.

Brunsdon, C., Fotheringham, A.S. & Charlton, M.E. (1996) Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis*, 28(4), 281–298. Available from: https://doi.org/10.1111/j.1538-4632.1996.tb00936.x

Brunsdon, C., Fotheringham, A.S. & Charlton, M. (1999) Some notes on parametric significance tests for geographically weighted regression. *Journal of Regional Science*, 39(3), 497–524. Available from: https://doi.org/10.1111/0022-4146.00146

Cao, S., Zhong, B., Yue, H., Zeng, H. & Zeng, J. (2009) Development and testing of a sustainable environmental restoration policy on eradicating the poverty trap in China's Changting County. *Proceedings of the National Academy of Sciences*, 106(26), 10712–10716. Available from: https://doi.org/10.1073/pnas.0900197106

Chen, X., Lun, Y., Yan, J., Hao, T. & Weng, H. (2019) Discovering thematic change and evolution of utilizing social media for healthcare research. *BMC Medical Informatics and Decision Making*, 19(2), 39–53. Available from: https://doi.org/10.1186/s12911-019-0757-4

Chuang, J., Ramage, D., Manning, C. & Heer, J. (2012, May). Interpretation and trust: Designing model-driven visualizations for text analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 443-452). Available from: https://doi.org/10.1145/2207676.2207738

Chybicki, A. (2018) Three-dimensional geographically weighted inverse regression (3GWR) model for satellite derived bathymetry using Sentinel-2 observations. *Marine Geodesy*, 41(1), 1–23. Available from: https://doi.org/10.1080/01490419.2017.1373173

Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E. & Herrera, F. (2011) Science mapping software tools: Review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology*, 62(7), 1382–1402.

da Silva, A.R. & de Oliveira, A. (2017) Geographically weighted beta regression. *Spatial Statistics*, 21, 279–303. Available from: https://doi.org/10.1016/j.spasta.2017.07.011

da Silva, A.R. & Rodrigues, T.C.V. (2014) Geographically weighted negative binomial regression-incorporating overdispersion. *Statistics and Computing*, 24(5), 769–783. Available from: https://doi.org/10.1007/s11222-013-9401-9

Dong, G., Nakaya, T. & Brunsdon, C. (2018) Geographically weighted regression models for ordinal categorical response variables: An application to geo-referenced life satisfaction data. *Computers, Environment and Urban Systems*, 70, 35–42. Available from: https://doi.org/10.1016/j.compenvurbsys.2018.01.012

Feinerer, I. & Hornik, K. (2019) *Text mining package*. Available from: https://cran.r-project.org/web/packages/tm Accessed: 22nd Jun 2020.

Fotheringham, A.S., Brunsdon, C. & Charlton, M. (2002) *Geographically weighted regression: The analysis of spatially varying relationships*. Chichester, UK: Wiley.

Geman, S. & Geman, D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741. Available from: https://doi.org/10.1109/TPAMI.1984.4767596

Griffiths, T.L. & Steyvers, M. (2004) Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, 101(suppl 1), 5228–5235. Available from: https://doi.org/10.1073/pnas.0307752101

Harzing, A.W. & Alakangas, S. (2016) Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison. *Scientometrics*, 106(2), 787–804. Available from: https://doi.org/10.1007/s11192-015-1798-9

Hornik, K. & Grün, B. (2011) Topicmodels: An R package for fitting topic models. *Journal of Statistical Software*, 40(13), 1–30. Available from: https://doi.org/10.18637/jss.v040.i13

Huang, B., Wu, B. & Barry, M. (2010) Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, 24(3), 383–401. Available from: https://doi.org/10.1080/13658810802672469

Jones, T. (2019) TextmineR: Functions for text mining and topic modeling. R Package Version, 3.0.4. Available from: https://CRAN.R-project.org/package=textmineR

Lambiotte, R. & Panzarasa, P. (2009) Communities, knowledge creation, and information diffusion. *Journal of Informetrics*, 3(3), 180–190. Available from: https://doi.org/10.1016/j.joi.2009.03.007

Larsen, P.O. & von Ins, M. (2010) The rate of growth in scientific publication and the decline in coverage provided by science citation index. *Scientometrics*, 84(3), 575–603. Available from: https://doi.org/10.1007/s11192-010-0202-z

Lau, J.H., Grieser, K., Newman, D. & Baldwin, T. (2011, June). Automatic labelling of topic models. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies* (pp. 1536-1545).

LeSage, J. P. (2004). A family of geographically weighted regression models. In L. Anselin, R. J. G. M. Florax & S. J. Rey, *Advances in spatial econometrics: Methodology, tools and applications* (pp. 241–264). Berlin, Germany: Springer-Verlag.

Li, K. & Lam, N.S.N. (2018) Geographically weighted elastic net: A variable-selection and modeling method under the spatially nonstationary condition. *Annals of the American Association of Geographers*, 108(6), 1582–1600. Available from: https://doi.org/10.1080/24694452.2018.1425129

Lu, B., Brunsdon, C., Charlton, M. & Harris, P. (2017) Geographically weighted regression with parameter-specific distance metrics. *International Journal of Geographical Information Science*, 31(5), 982–998. Available from: https://doi.org/10.1080/13658816.2016.1263731

Luo, J. & Kanala, N.K. (2008) Modeling urban growth with geographically weighted multinomial logistic regression. In L. Liu (Eds.), *Geoinformatics 2008 and Joint Conference on GIS and Built Environment: The built environment and its dynamics*, SPIE 71440M, SPIE Digital Library. Available from: https://doi.org/10.1117/12.812714

National Science Foundation (NSF). (2018) *National Science Foundation (NSF) Science and Engineering Indicators 2018 NSF*. NSF.

Nakaya, T., Fotheringham, A.S., Brunsdon, C. & Charlton, M. (2005) Geographically weighted poisson regression for disease association mapping. *Statistics in Medicine*, 24(17), 2695–2717. Available from: https://doi.org/10.1002/sim.2129

Páez, A. & Wheeler, D. C. (2009) Geographically Weighted Regression. In R. Kitchin, & N. Thrift (Eds.), *International Encyclopedia of Human Geography* (pp. 407-414). Amsterdam, Netherlands: Elsevier. Available from: https://doi.org/10.1016/B978-008044910-4.00447-8.

Peters, M.A. (2006) The rise of global science and the emerging political economy of international research collaborations. *European Journal of Education*, 41(2), 225–244. Available from: https://doi.org/10.1111/j.1465-3435.2006.00257.x

Team, R. C. (2019) *R: A language and environment for statistical computing*. Available from: https://www.r-project.org

Röder, M., Both, A. & Hinneburg, A. (2015) Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining* (pp. 399-408). Available from: https://doi.org/10.1145/2684822.2685324

Sievert, C. & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces* (pp. 63-70).

Syed, S. & Spruit, M. (2017) Full-text or abstract? Examining topic coherence scores using latent dirichlet allocation. In *2017 IEEE International conference on data science and advanced analytics (DSAA)* (pp. 165-174). IEEE. Available from: https://doi.org/10.1109/DSAA.2017.61

Van Eck, N.J. & Waltman, L. (2010) Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538. Available from: https://doi.org/10.1007/s11192-009-0146-3

Vanhala, M., Lu, C., Peltonen, J., Sundqvist, S., Nummenmaa, J. & Järvelin, K. (2020) The usage of large data sets in online consumer behaviour: A bibliometric andcomputational text-mining–driven analysis of previous research. *Journal of Business Research*, 106, 46–59.

Vijayarani, S., Ilamathi, M.J. & Nithya, M. (2015) Preprocessing techniques for text mining-an overview. *International Journal of Computer Science & Communication Networks*, 5(1), 7–16.

Wagner, C.S., Bornmann, L. & Leydesdorff, L. (2015) Recent developments in China–USA cooperation in science. *Minerva*, 53(3), 199–214. Available from: https://doi.org/10.1007/s11024-015-9273-6

Wang, W., Xu, S. & Yan, T. (2018) Structure identification and model selection in geographically weighted quantile regression models. *Spatial Statistics*, 26, 21–37. Available from: https://doi.org/10.1016/j.spasta.2018.05.003

Wheeler, D.C. (2007) Diagnostic tools and a remedial method for collinearity in geographically weighted regression. *Environment and Planning A: Economy and Space*, 39(10), 2464–2481. Available from: https://doi.org/10.1068/a38325

Wheeler, D.C. (2009) Simultaneous coefficient penalization and model selection in geographically weighted regression: The geographically weighted lasso. *Environment and Planning A: Economy and Space*, 41(3), 722–742. Available from: https://doi.org/10.1068/a40256

Xiong, H., Cheng, Y., Zhao, W. & Liu, J. (2019) Analyzing scientific research topics in manufacturing field using a topic model. *Computers & Industrial Engineering*, 135, 333–347. Available from: https://doi.org/10.1016/j.cie.2019.06.010

Yang, W. (2014) An extension Of geographically weighted regression with flexible bandwidths. St Andrews Research Repository, Available from: https://research-repository.st-andrews.ac.uk/ (Accessed 22nd May 2020).

Zhong, B., Peng, S., Zhang, Q., Ma, H. & Cao, S. (2013) Using an ecological economics approach to support the restoration of collapsing gullies in southern China. *Land Use Policy*, 32, 119–124. Available from: https://doi.org/10.1016/j.landusepol.2012.10.005

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

Supporting Information: Trends and topics in Geographically Weighted Regression research from 1996 to 2019.

Supporting Information: Trends and topics in Geographically Weighted Regression research from 1996 to 2019.

Figure S1: Methods and general procedure used in the performed study.

Figure S2: Graphic model for Latent Dirichlet Allocation LDA (Source: adapted from Blei et al. 2003). K, M, N and V denote the number of topics, number of article, words in article and words in the vocabulary respectively; α and η (Dirichlet hyper-parameters) are parameters of the prior distributions over θ and β respectively ; θm is the distribution of topics for article m (real vector of length K) ; zmn is the topic for the nth word in the mth article; wmn is the nth word of the mth document; βk is the distribution of words for topic k (real vector of length V).

Figure S3: Scheme of the methodological process used in the identification of research topics in Geographically Weighted Regression through latent Dirichlet allocation.

How to cite this article: De La Hoz, J., Fernandez-Gómez, M.J. & Mendes, S. (2021) Trends and topics in geographically weighted regression research from 1996 to 2019. *Area*, 00, 1–13. https://doi.org/10.1111/area.12757

ANEXO 2. LDAShiny: An R Package for exploratory Review of Scientific Literature Based on a Bayesian Probabilistic Model and Machine Learning Tools"

mathematics

MDPI

Article

LDAShiny: An R Package for Exploratory Review of Scientific Literature Based on a Bayesian Probabilistic Model and Machine Learning Tools

Javier De la Hoz-M [1,2,*], Mª José Fernández-Gómez [2,3] and Susana Mendes [4]

1 Facultad de Ingeniería, Universidad del Magdalena, Santa Marta 470004, Colombia
2 Department of Statistics, University of Salamanca, 37008 Salamanca, Spain; mjfg@usal.es
3 Institute of Biomedical Research of Salamanca, 37008 Salamanca, Spain
4 MARE, School of Tourism and Maritime Technology, Polytechnic of Leiria, 2520-614 Peniche, Portugal; susana.mendes@ipleiria.pt
* Correspondence: jdelahoz@unimagdalena.edu.co

Abstract: In this paper we propose an open source application called LDAShiny, which provides a graphical user interface to perform a review of scientific literature using the latent Dirichlet allocation algorithm and machine learning tools in an interactive and easy-to-use way. The procedures implemented are based on familiar approaches to modeling topics such as preprocessing, modeling, and postprocessing. The tool can be used by researchers or analysts who are not familiar with the R environment. We demonstrated the application by reviewing the literature published in the last three decades on the species *Oreochromis niloticus*. In total we reviewed 6196 abstracts of articles recorded in Scopus. LDAShiny allowed us to create the matrix of terms and documents. In the preprocessing phase it went from 530,143 unique terms to 3268. Thus, with the implemented options the number of unique terms was reduced, as well as the computational needs. The results showed that 14 topics were sufficient to describe the corpus of the example used in the demonstration. We also found that the general research topics on this species were related to growth performance, body weight, heavy metals, genetics and water quality, among others.

Keywords: text mining; topic modeling; latent dirichlet allocation; automatic literature review

Citation: De la Hoz-M, J.; Fernández-Gómez, M.J.; Mendes, S. LDAShiny: An R Package for Exploratory Review of Scientific Literature Based on a Bayesian Probabilistic Model and Machine Learning Tools. *Mathematics* **2021**, *9*, 1671. https://doi.org/10.3390/math9141671

Academic Editor: Maria Laura Manca

Received: 17 May 2021
Accepted: 24 June 2021
Published: 16 July 2021

1. Introduction

A literature review is considered an integral part of the research process in any scientific area, and seeks to discover the relevant sources of a particular subject of study. Thus, it plays a crucial role since wisdom is generated through the process of interpretation and integration of existing knowledge [1].

Nowadays there is an increasing amount of scientific literature published in digital form in databases such as Scopus or Web of Science, to mention two of the most used by researchers [2]. Therefore, it can be inferred that there is a gap between the availability and use of information. A literature review in a conventional way is restricted, has a high cost in terms of of time, and has limited processing power, which leads researchers to restrict the amount of documents to review. Nowadays, machine learning approaches make it feasible to process huge amounts of data, allowing researchers to spend less time examining their findings. When human-assisted information processing, such as encryption, is replaced with computer-assisted processing, dependability improves and costs fall [3].

Asmussen and Muller [4] mention that the exploratory review of literature in a conventional way will soon become outdated because it is a process that has a high cost in time, with limited processing power, which leads researchers to restrict the amount of documents to be reviewed, which is a problem in the initial exploratory phase of an investigation since what is needed is an overview of the state of the art of research. The large amount of information available makes searching, retrieving and summarizing information cumbersome

and challenging, so the use of tools capable of searching, organizing and summarizing a large collection of text documents in the scientific field is in demand.

In the open source environment R [5] in the Comprehensive Archive Network (CRAN) we can find a list of 59 packages related to natural language processing (NLP), eight of which implement the modeling of topics through latent Dirichlet assignment (LDA) [6]: lda collapsed Gibbs sampling methods for topic models [7]; lda.svi fit LDA models using stochastic variational inference [8]; ldaPrototype prototype of multiple LDA runs [9]; lda.svi LDA coupled with time series analyses [8]; ldatuning, tuning of the LDA models parameters [10]; LDAvis, interactive visualization of topic models [11]; topicdoc topic-specific diagnostics for LDA and Correlated Topic Models (CTM) topic models [12]; topicmodels [13] and textmineR [14] functions for text mining and topic modeling.

To date, there is no free statistical software package with a graphical user interface (GUI) where analysts and researchers can take advantage of the combined power of several packages to perform LDA-focused scientific literature reviews in an interactive (point-and-click) way. The LDAShiny application is primarily aimed at researchers who wish to use machine learning to explore a large number of documents (e.g., scientific articles) to identify research trends. This is beneficial for researchers who know little about the research field. The application allows a large number of documents to be grouped automatically in less time than if it were done manually, thus providing an overview of the directions of the investigation. Therefore, from the perspective of a literature review, this is valuable as the decision to include or exclude articles is made in a more informed way at a later stage.

This study presents the development of a computer tool for performing a literature review with a focus on topic modeling (a branch of unsupervised methods). It could help to reduce or to replace the time spent by the researcher at the computer by automatically generating review topics based on the statistical qualities of the documents utilized, without the need for prior classification, categorization, or labeling. Thus the possible bias due to subjective choices of the researchers could be avoided or minimized. Furthermore, historical and current research and trends in the field under study can be more easily synthesized.

There are several packages for modeling topics in the R environment. However, they require some statistical and machine learning skills that not all researchers possess [4]. Therefore, the main aim of LDAShiny was to make the typical LDA workflow easier to use, especially for those who are unfamiliar with R. With LDAShiny the analysis can be performed interactively in a web browser, which makes it easier for many more researchers to apply this technique to review the scientific literature.

Thus, and in order to facilitate the understanding of the work exposed here, the manuscript presents a section that introduce a quick overview of topic modeling with LDA. Then, the methods employed are presented (Section 3), followed by the detailed description of the LDAShiny GUI (Section 4). In Section 5, the use of the LDAShiny GUI using *Oreochromis niloticus* literature over the last three decades is explained. Finally, the conclusions are presented in Section 6.

2. Topic Modelling for Exploratory Literature

Topic modeling is a classic problem in NLP and machine learning. It refers to a set of algorithms and statistical methods of learning, recognition and extraction that aim to analyze the hidden structure of a collection of documents to discover the topics, how they are related to one another and how they have evolved over time. It has the advantage of not requiring any prior annotations or document labeling because the topics emerge from the analysis of the original texts [15].

It has the advantage that no previous annotations or labeling of the documents are required. Its use spans practically every aspect of text mining and information processing, including text summarization, information retrieval and text classification [16]. Topic modeling allows us to organize and summarize electronic files in various formats (web

pages, scientific articles, books, images, sound, videos and social networks) at a scale that would be impossible by human annotation [15].

Latent semantic analysis (LSA) [17] and probabilistic latent semantic analysis (PLSA) [18] are the predecessors of LDA However, considering that LDA is one of the most used methods [3,19,20], we decided on it due to its highly qualified ease of use, understanding and applicability [4].

LDA is a Bayesian variant of PLSA, based on a set of words assumption, which states that words in a text are interchangeable and that documents are represented as a series of individual words [6]. This algorithm was initially applied to text corpora but its use has been extended to images [21] and videos [22].

LDA is a generative model. In other words, it is a model that shows how data are produced, and once you have a model of how they are generated, you can know which target variable generated them. The Dirichlet distribution, which is a multivariate version of the beta distribution, is used by LDA to extract the features of the subjects and documents.

The generative process from which LDA assumes the documents come, is described as:

1. For every topic k:
   a. Draw a distribution over the words (i.e., vocabulary V) $\beta_k \sim Dir(\eta)$ [6]
2. For every document d:
   a. Draw a distribution over topics $\theta_d \sim Dir(\alpha)$ (i.e., per document topic proportion) [6]
   b. For each word w within document d:
      i. Draw a topic assignment, $z_{d,n} \sim Mult(\theta_d)$ (i.e., per-word topic assignment) [6]
      ii. Draw a word $w_{d,n} \sim Mult(\beta_{z,d,n})'$ [6,23,24].

Each topic k comes from a Dirichlet distribution $\beta_k \sim Dir(\eta)$, and is a multinomial distribution over the vocabulary. Furthermore, each document is represented as a topic distribution and originates from a $\theta_d \sim Dir(\alpha)$. The Dirichlet parameter η defines the smoothing of the words within topics, and α the smoothing of the topics within documents [6]. The joint distribution of all the hidden variables β_k, θ_D (document topic proportions within D), z_D (word topic assignments), and observed variables w_D (words in documents), is expressed by Equation (1):

$$P(\beta_k, \theta_D, z_D, w_D) = \prod_{k=1}^{K} P(\beta_k|\eta) \prod_{d=1}^{D} P(\theta_d|\alpha) \prod_{n=1}^{N} P(z_{d,n}|\theta_d)\, P(w_{d,n}|z_{d,n}, \beta_k) \quad (1)$$

This shows the statistical assumptions behind LDA's generative process. The per-word topic assignment $z_{d,n}$ depends on the previously drawn (step 2.a.) per-document topic proportion θ_d. Furthermore, the drawn word $w_{d,n}$ depends on the per-word topic assignment $z_{d,n}$ (step 2.b.i) and all the topics β_k (we retrieve the probability of $w_{d,n}$ (row) from $z_{d,n}$ (column) within the K $\times$ V topic matrix). The latent variables are the per-word topic assignment, the per-document topic distribution and the topics, which are not observed. To infer the hidden structure using statistical inference, we would have to condition on the single seen variable, i.e., the words within the documents. This might be thought of as a reversal of the generative process [6].

Equation (2) expresses the posterior or conditional probability. Unfortunately, due to the denominator, this probability cannot be computed [6]. Therefore, machine learning algorithms have to be used to find approximations of the marginal probability of the observations $P(w_D)$, This marginal probability of the observations is the chance of seeing the observed corpus under any topic model [15].

$$P(\beta_k, \theta_D, z_D|w_D) = \frac{P(\beta_k, \theta_D, z_D, w_D)}{P(w_D)} \quad (2)$$

Although it is impossible to accurately calculate the posterior probability, statistical posterior inference can be used to obtain an approximate value close enough to the true value. Two main types of reasoning technique can be identified: sampling-based algorithms [25,26] and variational-based algorithms [26–28]. Sampling-based algorithms sample from the posterior, usually taking one variable at a time, fixing the other variables. Repeating this process for several iterations makes the inference process converge, so the sample values have the same distribution as if they came from the true posterior value. An example of a sampling-based algorithm is the Gibbs sampler (a full explanation about Gibb sampling can be found in Griffths and Steyvers [23]), a Markov chain Monte Carlo (MCMC) algorithm. Variational-based algorithms create a family of distributions that are closest (distance is measured with Kullback–Leibler (KL) divergence) to the true posterior. It should be noted that both variational and sampling-based algorithms provide similar accurate results [29].

The latent variables θ and z are frequently used in inference to establish which subjects a document contains and from which subject a certain word in a document was derived. The variational posterior probability can be used to estimate latent variables on the premise that it is a reasonable approximation of the real posterior probability. If the variational expectation maximization (VEM) is employed for estimate, inference is always based on the variational posterior probabilities [13].

3. Materials and Methods

The methodology utilized to create the LDAShiny program is based on well-known topic modeling approaches to data cleansing and processing. The main contribution in this work is not to introduce new ways of processing data, but to learn how the methods are combined and how they can be easily used by researchers through the use of this application. The inspiration for the creation of LDAShiny can be found in Asmussen and Moller [4] who considered that the intelligent literature review process consists of three steps: preprocessing, topic modeling and post-processing.

In our proposal, the review process consists of four steps: preprocessing, inference, topic modeling and post-processing (Figure 1).

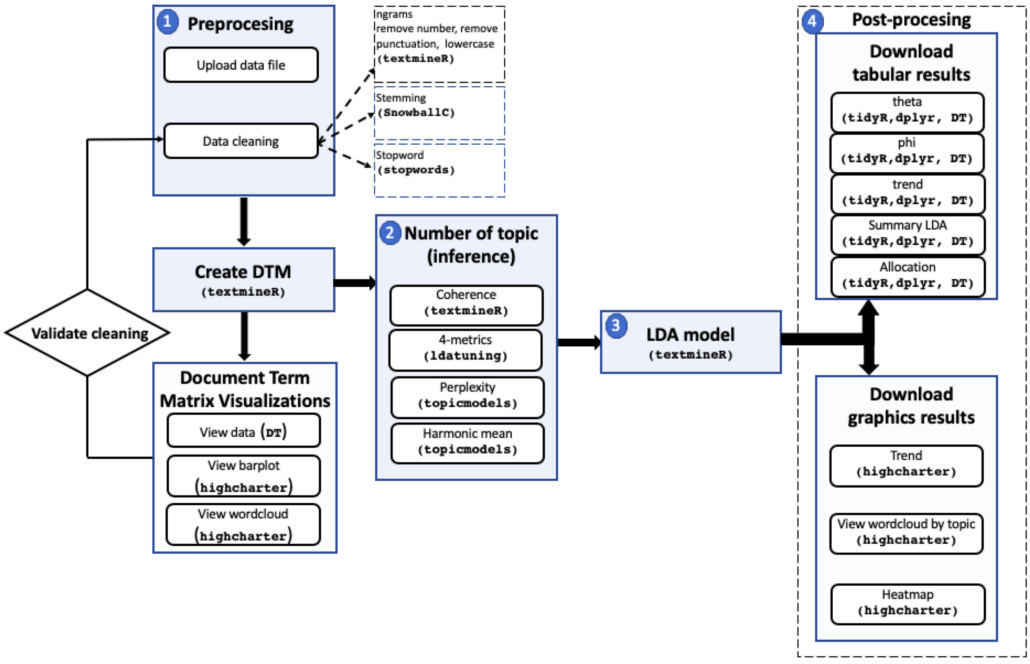


Figure 1. LDAShiny package outline. In parentheses are the main packages used.

3.1. Preprocessing

Preprocessing consists of loading and preparing the documents for subsequent processes. This phase plays a very important role, being generally the first step in text mining techniques and applications [30]. Pre-processing seeks to normalize or convert the set of text to a more convenient standard form that allows the reduction of the data dimensionality of the data matrix by eliminating noise or meaningless terms. Within the pre-processing we have the "cleaning" in which the following tasks are performed:

- Tokenization, which is the procedure of separating morphemes (words). According to Jurafsky and Martin [31] it is beneficial in both linguistics and computer science.
- n-gram inclusion: an n-gram is a contiguous sequence of n words [32]. Although it is more usual to analyze individual words, in some cases, such as in the life sciences, incorporating bigrams would be advantageous because scientific names of species are made up of two words. In LDAShiny we can work with unigrams, bigrams or trigrams (three words frequently occurring).
- Remove numbers, despite the fact that numbers are frequently thought to be uninformative, there are some areas of knowledge where numbers can provide valuable information, for instance, in legislative matters, bills or decrees can be significant with respect to content legislation. That is why in the developed application the researcher can decide whether or not to eliminate the numbers.
- Remove StopWord, a term coined by Luhn [33]. The procedure consists of discarding words that have no lexical meaning and that appear in texts very frequently (such as articles and pronouns). There are many potential StopWord lists, however, we restrict ourselves to a pre-compiled list of words provided by the R StopWord [34]. LDAShiny allows performing this procedure in 14 languages Danish, Dutch, English, Finnish, French, German, Hungarian, Italian, Norwegian, Portuguese, Romanian, Russian, Spanish, and Swedish.
- Stemming, which is the simplest version of lemmatization. It consists of reducing words to basic forms [35]. Although it is often used as a reduction technique, it must be used carefully, since it could combine words with different meanings, for example in the phrases "college students partying", and "political parties", stemming would reduce partying and parties as the same basic form.
- Remove infrequently used terms (sparsity). This procedure is very useful because it allows removing the terms that appear in very few documents before continuing with the successive phases. Among the reasons for this procedure is the computational feasibility, as this process drastically reduces the size of the matrix without losing significant information and can also eliminate errors in the data, such as misspelled words. This only applies to terms that comply with:

$$df(t) > N(1 - sparce) \tag{3}$$

  where df is the frequency of documents of the term t and N is the number of vectors. For example, if the sparse value is 0.99, the terms that appear in more than 1% of the documents are taken. As a general rule, terms that appear in less than 0.5–1 percent of the articles should be discarded [19,36,37]. However, there has been no systematic examination of the implications of this pre-processing decision on the analyses' final phase.
- Eliminating blank spaces and punctuation characters, as well as lowering the entire text, are other standard procedures used to prevent a word from being counted twice due to capitalization.

The cleaning process must be validated. However, to date there has been no scientific way to establish when this process ends, so the process must be iterative, since it is not possible to guarantee an identical cleaning procedure when conducting an exploratory review [4]. Once the pre-processing phase is completed, the document-term matrix (DTM) is obtained as input data for topic models.

3.2. Inference

LDA is a model for latent variables using correlations between words and latent semantic topics in a collection of documents [38]. This implies that the parameter k (number of topics) of the algorithm is crucial and must be established beforehand, since the validity of the results obtained depends largely on the inference process of the model. In theoretical terms, a very large number of topics will produce overly specific topics, while conversely, a very small number would handle broad and heterogeneous themes [39].

There are a variety of metrics that can be used to determine the optimal number of topics. In our package we implement the following:

- perplexity defined by [6] for a set of text of M documents as:

$$perplexity(D_{text}) = exp\left(\frac{\sum_{d=1}^{M}\log(w_d)}{\sum_{d=1}^{M}N_d}\right) \quad (4)$$

  where N_d is the number of words in the d-document of the text corpus D_{text} and w_d is the d^{th} document in the corpus. It is monotonically decreasing and algebraically equivalent to the inverse of the geometric mean probability per word. When comparing several models, the one with the lowest value of perplexity is considered the best [6].
- marginal likelihood that can be approximated by harmonic mean. This method has first been applied by Griffiths and Steyvers in their 2004 Bayesian approach, in order to find the optimal number of topics [23,40].
- coherence [41]. It is based on the distribution hypothesis [42] which states that words with similar meanings tend to coexist in similar contexts. The procedure used for this metric is based on the TextmineR package [14], which implements a thematic coherence measure based on probability theory and consists of fitting several models and calculating the coherence for each of them. The best model will be whichever offers the greatest measure of coherence.
- other metrics can be found in the ldatuning package Arun 2010 [43], CaoJuan 2009 [44], Deveaud 2014 [45], Griffiths 2004 [23]. The approach of these metrics is simple and they are based on finding extreme values (minimization Arun 2010 and CaoJuan 2009; maximization Deveaud 2014 and Griffiths 2004).

For a further description of each of the metrics used by the application, it is recommended to review the corresponding articles.

3.3. Latent Dirichlet Assignment (LDA) Model

Once the number of topics has been determined, LDAShiny proceeds to execute the LDA model. Some parameters such as the number of iterations can be modified by a number of iterations greater than that used to make the inference. As a result, the modeling DTM is reduced to two matrices. The first one, theta, has rows that indicate the distribution of topics on documents $P(topic_k|document_d)$. The second one, phi, has rows that indicate the distribution of words on topics $(token_v|topic_k)$.

3.4. Post-Processing

This step involves processing the results and obtaining a description of the topics. The distribution of topic terms does not come with a semantic interpretation. However, depending on the frequency of the words, the topics can be labeled correctly in most cases. Lewis, Zamith, and Hermida [46] mention that algorithmic analyses have a very limited capacity to understand latent meanings in human language, so manual labeling is considered a standard [47]. However, in the latter case, the labeling can provide different topic labels depending on the researcher. The textmineR [14] package provides a topic labeling based on a naive labeling algorithm built on bigrams. However, as mentioned, these algorithms have limited capabilities, but may well serve as a guide.

Once all the topics have been labeled, with the help of the theta matrix, the procedure continues assigning documents to each topic, classifying them according to the highest probability of each document for each topic. In this way the documents will also be grouped.

Labelling requires validation by an expert in the field of research, otherwise mislabeled topics and an invalid result could be obtained [4].

In order to facilitate the characterization of the topics in terms of their trends, the simple regression slopes for each theme are used. The year is the dependent variable and the proportions of the topics in each year the response variable [23]:

$$\theta_k^y = \frac{\sum_{m \in y} \theta_{mk}}{n^y} \tag{5}$$

where $m \in y$ represent the articles published in a certain year and, θ_{mk} the proportion of the k-topic and n^y the total number of articles published in the year y [48]. Topics whose regression slopes are positive (negative) at a statistical significance level are interpreted as increasing (declining) their interest respectively, and if the slopes are not significant, the topics will be classified as fluctuating trends.

4. LDAShiny Graphical User Interface (GUI)

The LDAShiny is web-based and has been developed in R using the shiny [49] web application framework. LDAShiny provides an integrated platform for exploratory review of scientific information, offering a number of options to manage, explore, analyze and visualize data. This is particularly beneficial to researchers who are not as familiar with R, or programming in general, but wish to use the methods described here.

The LDAShiny package is accessed from the Comprehensive R Archive Network (CRAN) at http://CRAN.R-project.org/package=LDAShiny. To install, load, and launch it, type the following in R:

- R > install.packages("LDAShiny")
- R > library("LDAShiny")
- R > LDAShiny::runLDAShiny()

The GUI proposed in this work provides a menu that, from top to bottom, guides the user through the analysis:

- About: this panel serves as the software's introduction page. The application's general information, as well as the software's goal, are displayed in English and Spanish.
- Data input and preprocessing: this provides an interface for users to load the data to be analyzed. In addition, there are also different options to perform preprocessing.
- Document term matrix visualizations: the matrix of terms and documents can be viewed both in tabular and graphical form in this menu. The tabular data can be downloaded in csv xlxs or pdf format or can be copied to the clipboard. The graphics (barplot or wordcloud) can be downloaded in .png, .pdf, .jpeg, .svg, and .pdf format.
- Number of topics (inference): The options to set the input parameters of each of the metrics used to find the number are available in this menu.
- LDA model tutorial: this menu offers a vignette (in English and Spanish) with videos that serve as a quick guide where the basic steps to use the software are explained.

Table 1 lists the details of each panel or menu.

Table 1. Details and description of LDAShiny panel.

Panel	Item	Menu	Description
Preprocessing	Upload data file	Use example data set?	Check box indicating whether a le that comes with the package.
		Choose csv file	Clicking the Browse button will load local data files in csv format
		Header	Checkbox indicating if the first line of the le contains the names of the columns
		stringAsFactors	String as factors.
		Separator	Field separating character.
		Select	PickerInput presents the loaded dataset and displays it in the Statistical summary table view.
	Data cleaning	Incorporate information	Clicking three times the Incorporate information button will load the data into preprocessing.
		Select id document	PickerInput for specifying vector of names for documents.
		Select document vector	PickerInput for specifying character vector of documents
		Select publish year	PickerInput for specify the vector containing the year the document was published
		ngrams	Radio buttons to specify the type of ngram to use (unigram, bigram or trigram).
		Remove number	Checkbox to specify whether or not to delete the numbers in thecorpus (if clicked it will remove the numbers).
		Select language for stopword	PickerInput to specify the language used in the stopword removal (the list contains 14 languages to choose from).
		Stop Words	Text field to include additional stop words to remove (words must be separated by commas).
		Stemming	Checkbox if clicked, stemming is performed
		Sparsity	Slider to select sparse parameter.
		Create Document-Term Matrix DTM	After clicking the Create DTM button, a spinner will be displayed during the process. Once finished, a table with the dimensions of the created matrix is displayed.
Document Term Matrix Visualizations		View Data	Clicking the View Data button will be display a summary. Also shown are a series of buttons that allow downloading in csv, xlxs or pdf formats, print the le Print, copy it Copy to the clipboard, and a button to configure the number of rows Show to be used in the summary.
		View barplot	Clicking the View barplot button will be display a barplot. The number of bars can be configured using the slider shown in the Dropdown button, Select number of term. In the upper right part of the graph (export button), clicking on it, you can download the graph in different formats (.png, .jpeg, .svg and .pdf)
		View wordcloud	Clicking the View wordcloud button will be display a wordcloud. The number of words can be configured by the slider shown in the Dropdown button Select number of term In the upper right part of the graph (export button), clicking on it, you can download the graph in different formats (.png, .jpeg, .svg and .pdf).
Number of topic (inference)	Tab Coherence	Iterations	Numeric input parameter that specifies how many iterations will be performed

Table 1. *Cont.*

Panel	Item	Menu	Description
		Burn-in	Numeric input parameter that specifies how many burn-in for posterior sampling will be performed
		Hyper-parameter	Numeric input parameter that specifies the alpha value of the Dirichlet distribution.
	Tab 4-metrics	Estimation method	There are two radio buttons to select the estimation algorithm, Gibb for Gibbs sampling and VEM for variational expectation maximization
	Tab Perplexity	Iteration, Burn-in, and Thin	These parameters control how many Gibbs sampling draws are made. The first burning iterations are discarded and then every thin iteration is returned for each iterations
	Tab Harmonic mean	Iteration, Burn-in, and Keep	If a keep parameter was given, the log-likelihood values of every keep iteration, are contained.
LDA model	Run model		The input parameters are the number of topics (K), number of iterations and the alpha parameter of the Dirichlet distribution. Clicking the Run LDA Model button, a spinner will be displayed. Once the process is complete, a table will be displayed that includes coherence score, prevalence, and 10 top-terms for each topic. Also shown are a series of buttons that allow downloading in csv, xlxs or pdf formats, print the file Print, copy it Copy to the clipboard and a button to configure the number of rows (Show rows).
	Download tabular results	theta	Clicking on the theta button, a table will be displayed that includes topic, document and theta
		phi	Clicking on the phi button, button, a table will be displayed that includes topic, term and phi
		trend	Clicking on the trend button, a table showing the results of a simple linear regression (intercept, slope, test statistic, standard error and *p*-value) where the year is the dependent variable and the proportions of the topics in the corresponding year is the response variable.
		Summary LDA	Clicking on the Summary LDA button, three sliders will be shown at the top, this allows the summary configuration: Select number of labels, Select number top terms, and Select assignments the latter is a documents by topics matrix similar to theta. This will work best if this matrix is sparse, with only a few non-zero topics per document
		Allocation	Clicking on the Allocation button, a table will be shown where the user can find the documents that can be organized by topic. Thanks to the slider located at the top one we can choose the number of documents per topic to be displayed.
	Download graphics	trend	Clicking on the trend button, a line graph will be shown (one line for each topic) where time trends can be visualized. The graphic is interactive, clicking on the lines they will be removed or displayed as the user decides.
		View wordcloud by topic	Clicking the View wordcloud by topic button will be display a wordcloud. In the drop-down button you can select the topic from which we want to generate the wordcloud, also, in the slider you can select the number of words to show
		heatmap	Clicking the heatmap button will display a heatmap. The years are shown on the *x*-axis, the *y*-axis shows the topics and the color variation represents the probabilities.

5. Demonstration of LDAShiny GUI

To demonstrate how the GUI is used, an exploratory review of scientific texts referring to the species *O. niloticus* was carried out. This species is used when considering that aquaculture research involves very diverse areas (engineering, ecology, biology, physiology, economics, environmental and political sciences, among others), which in most cases must be developed together to successfully produce a specific species at the industry level. It was assumed that an exploratory review of the literature on the species was necessary and that the number of documents to be reviewed was too large to carry out a manual review.

The inclusion criteria focused on selectin those research articles in which information about this species was discussed, using its scientific name as a keyword. Likewise, it was decided to take into account documents in which the name of the species was mentioned either in the title, in abstract or as keyword, ensuring that the largest number of potentially relevant documents was included.

The search for articles was carried out through Scopus database considering that it supports the downloading of metadata batches of the articles (which speeds up data collection). Furthermore, it is one of the databases most used by researchers [2]. A number of 6196 abstracts of articles were found (in the last three decades 1991–June 2020). This number of documents makes an individual exploratory review too time consuming, so the set of articles considered provides a good example to test the application. The file used for the demonstration can be downloaded at the link https://github.com/JavierDeLaHoz/o.niloticus/blob/main/O.niloticus.csv.

5.1. Preprocesing

The required dataset must be in a wide format (one article or abstract per row). Upload the *O. niloticus* data file to LDAShiny from the Upload Data panel. Next, on the Data cleaning panel, click the Incorporate information button, and then specify the columns for id document (Title in our case). Select document vector (Abstract), and select the year of publication (Year), Then click on the checkbox to select ngram (Bigrams). Remove the numbers, select the language for the stopwords and include the words you want to remove. In our example we use, in addition to the default list, a pre-compiled list called SMART (System for the Mechanical Analysis and Retrieval of Text) from the stopword package. In addition, some terms detected in the validation were also removed, such as all the terms with two letters and also the following words: article, articles, author, authors, blackwell, copyright, fish, francis, international, journal, licensee, nature, nile, niloticus, objective, oreochromis, present, press, published, publishing, reserved, result, resulted, results, rights, science, showed, significant, significantly, sons, springer, study, taylor, tilapia, total, verlag and wiley. The complete list of stopwords used in the example can be found at https://github.com/JavierDeLaHoz/stopword/blob/main/stopword.csv.

For this example, no stemming was performed and the Sparsity slider used was 99.5%, that is, the terms that appeared in more than 0.5%. Finally, the Create DTM button was clicked, after cleaning, 530,143 unique terms remained in the corpus, however the procedure reduced the number of unique terms to 3268, greatly reducing computational needs (Figure 2).

The resulting DTM matrix can be previewed in the Document Term Matrix Displays panel, in both tabular and graphical form (Figure 3). The information presented in tabular form contains the terms (term), their frequency of appearance (term_freq) and how many documents these terms appear (doc_freq). In addition, idf is the inverse frequency of the document, which measures if the term is common or not in the document collection. It is obtained by dividing the total number of documents by the number of documents that contain the term, and then the logarithm of that quotient is taken. We observe that words such as growth, levels, higher, protein, control weight, species, effects, days and observed, are the most frequent terms that appear the most in the evaluated documents (Figure 3).

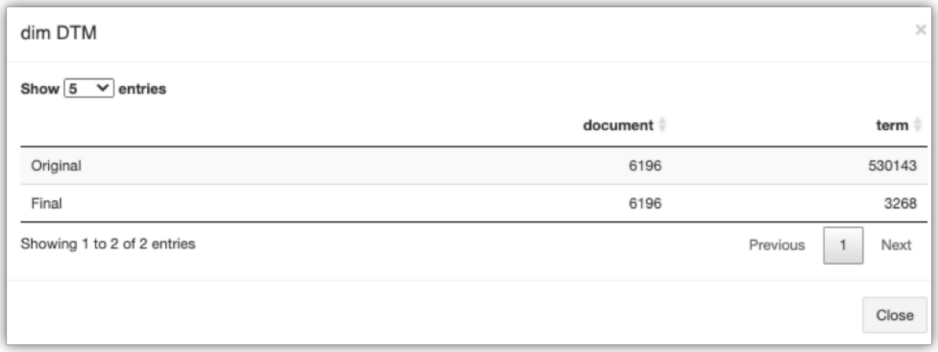


Figure 2. Document term matrix dimensions before (original) and after preprocessing (final).

The information on the frequency of terms can also be seen graphically in the form of a barplot or wordcloud. In both options the user can configure the number of words to display (Figure 3).

This statistical description, in the collection of articles, can provide a specific but limited overview of a particular field of research. As result, the words found in the evaluated articles represent the variety of topics investigated for *O. niloticus.*

5.2. Number of Topics (Inference)

Once the DTM matrix has been obtained, the next step is to determine the optimal number of topics. A very small number of topics can generate broad and heterogeneous topics. By contrast, a high number of k will produce themes that are too specific and in both cases the interpretation is complicated [39]. Therefore, the least number of topics was preferred as the intention is to provide an overview of the usefulness of the LDAShiny GUI. The highest quality LDA model can be determined using different metrics such as topic coherence [40]. This is a measure of the quality of a model topic from the point of view of human interpretability. Some authors consider it to be a more appropriate measure than computational metrics, such as perplexity [50] and likelihood of holdout data [24]. It should be noted that finding the number of topics is a computational expensive procedure and, although LDAShiny uses parallelism, the procedure may take anywhere from a few minutes to even a couple of days. It depends on the size of the DTM, the number of models (number of topics to evaluate), and the number of cores on the computer (LDAShiny works with the total number of cores).

In the left margin of Figures 4–7 the configuration options for each of the metrics used to calculate the number of topics are shown. The graphic outputs of each one appear on the right one. In every scenario, the amount of time it took to complete the inference is displayed. The time elapsed for estimating the number of topics in each of the metrics was 13,922, 2276, 5832 and 2755 s for "coherence", "4-metrics", "perplexity" and "Harmonic mean", respectively.

However, it should be noted that the times required are very dependent on the size of the DTM matrix, the number of iterations used (in all cases of the example there were 1000 iterations except for 4-metric, which uses 2000 by default), and the number of central processing unit (CPU) cores available (in our case a laptop with four cores was used).

Regarding the number of topics, the metrics Griffiths 2004, CaoJuan 2009, Arun 2010, Perplexity and Harmonic mean agreed to stablish that the number of suitable topics is between 45 and 50, while Deveaud 2014 showed 35 and Coherence 14 Topics. However, there are considerations that must be addressed when using LDAShiny. There is no common accepted way to choose the number of topics in a topic model. Thus, finding the right number of topics can be quite complex [4].

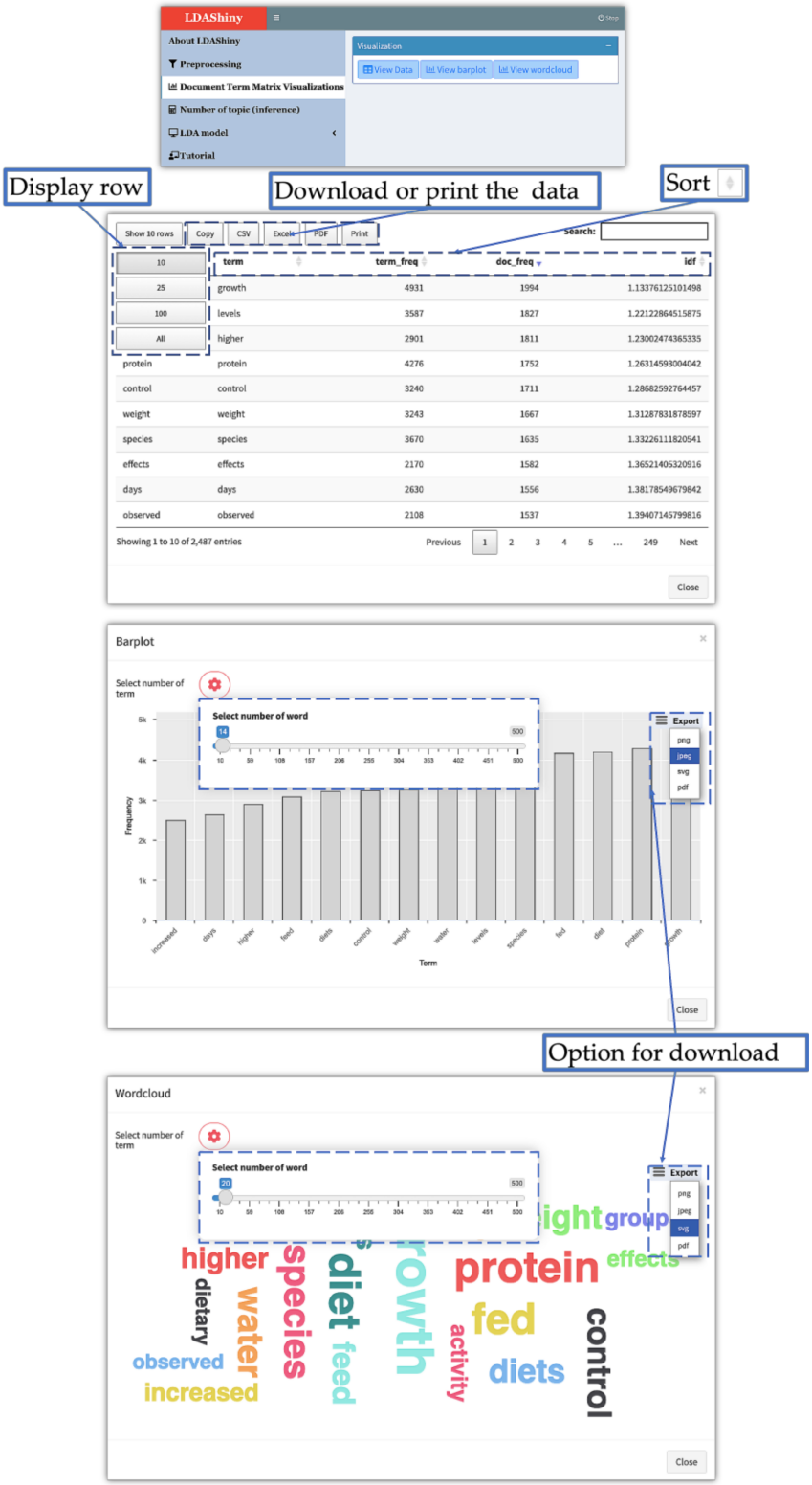


Figure 3. Document term matrix display options.

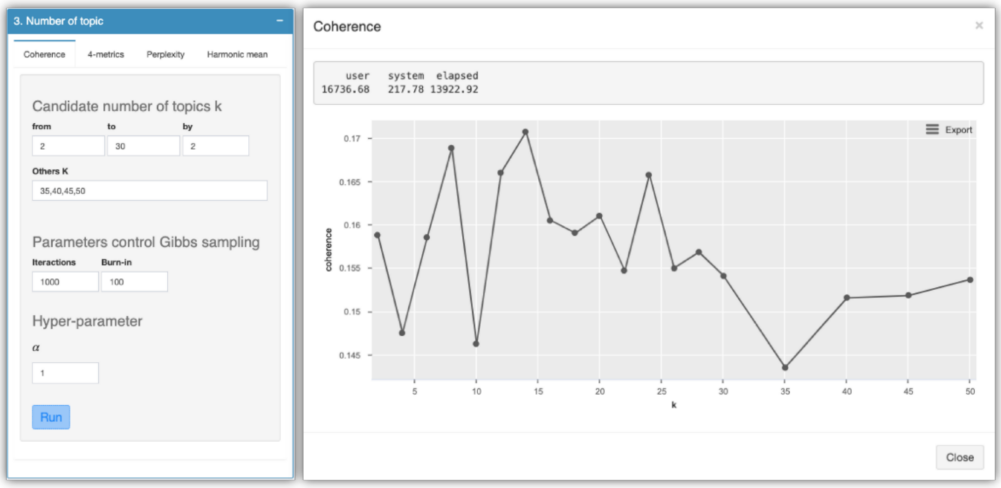


Figure 4. Configuration options used to calculate the number of topics (coherence method).

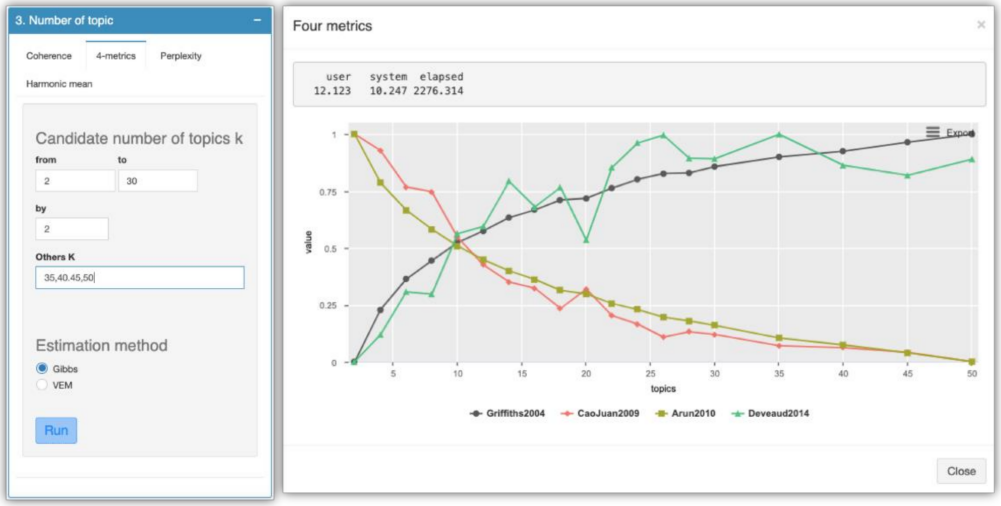


Figure 5. Configuration options used to calculate the number of topics (comparison of four metrics).

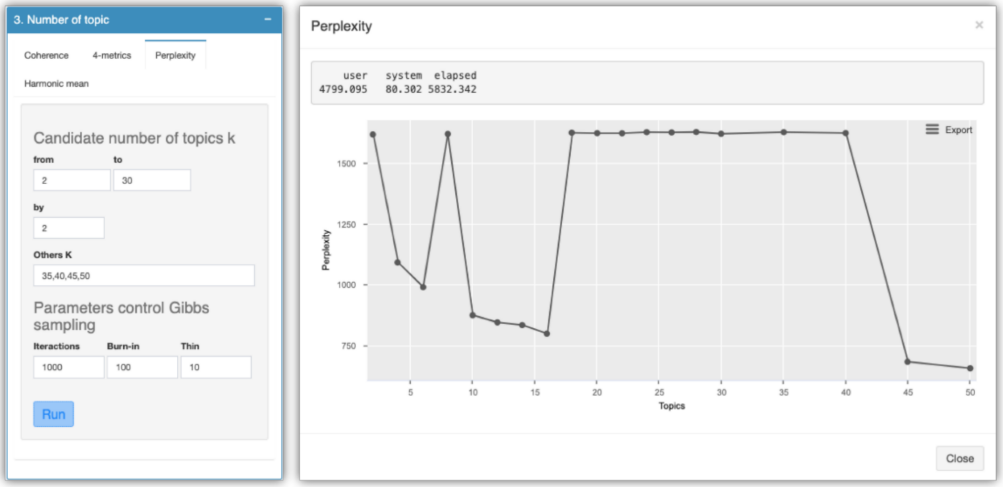


Figure 6. Configuration options used to calculate the number of topics (perplexity).

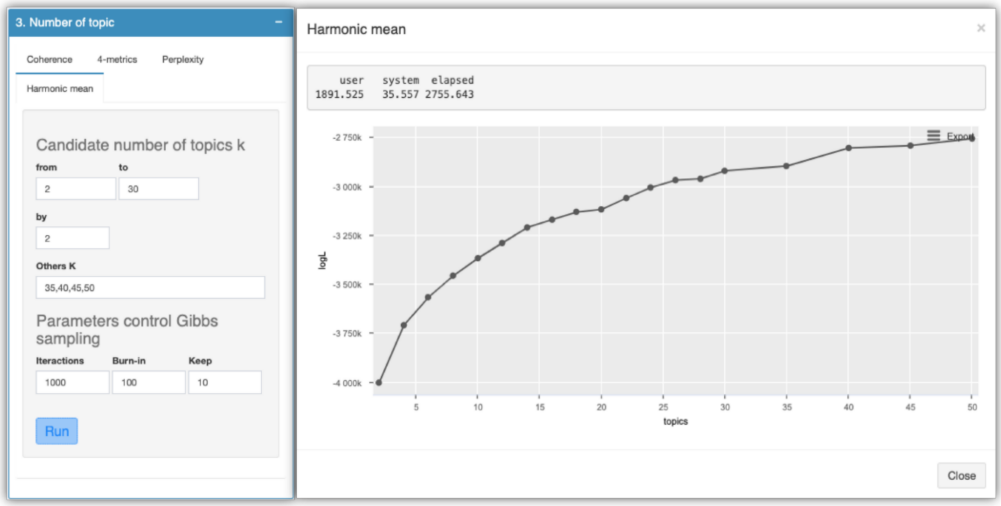


Figure 7. Configuration options used to calculate the number of topics (harmonic mean).

Because a general description of research on *O. niloticus* was required in our case, we preferred to use the smallest number of topics. However, determining what constitutes a small number of topics will differ from the model's input corpus. Nevertheless, visualizing the metric outputs can provide the appropriate guidance.

5.3. LDA Model

Once the number of topics has been defined, the LDA model is fitted. The parameters of inference should be used as a guide. However, some can be modified, such as the number of iterations, which may be higher. Also, the recommendation of Griffiths and Steyvers (2004) [23] could be used, setting a α value of 50/k. In this example, as input parameters, 1000 iterations and 100 burnin were used, and the α value was set to 3.57 (Figure 8).

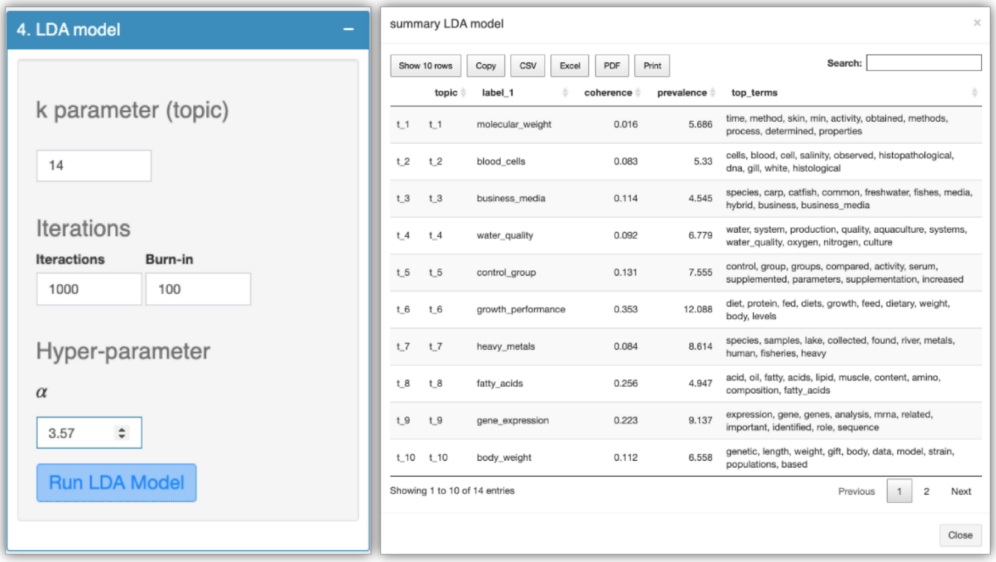


Figure 8. Configuration options used to calculate LDA model.

Within the tabular results of the model is the list of probabilities of each article for each topic (matrix theta) and the matrix that shows the most frequent words in each topic (phi) (Figure 9). The results of the estimations of the simple linear regression and their *p*-value (trends) (Figure 10, left). Also, they show the summary of the model where the

label, coherence score, prevalence and the top term for each topic are included (summary) (Figure 10 right) and finally a table with the allocation of topics (Figure 11).

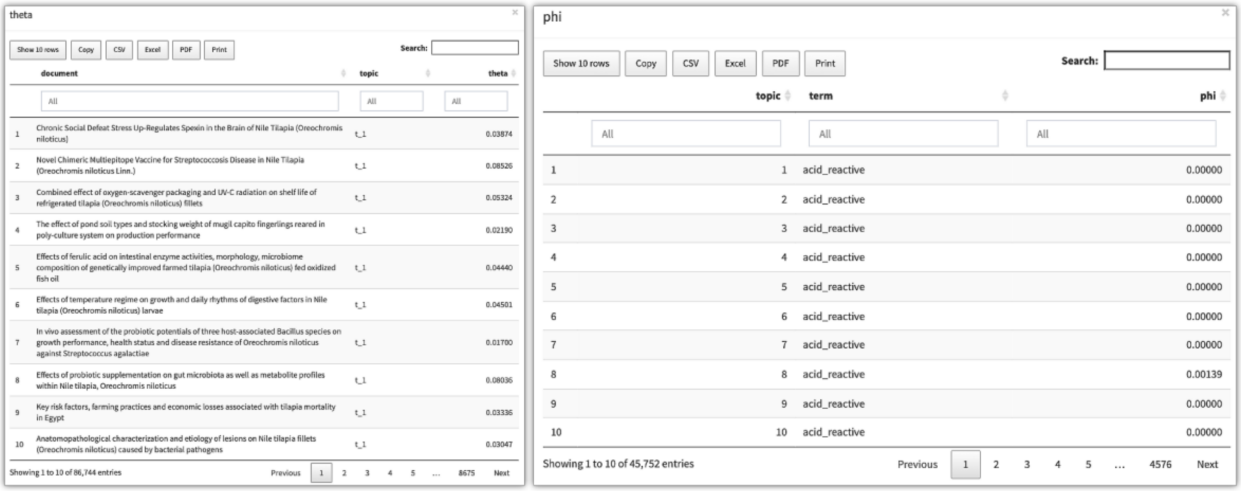


Figure 9. Output tabular of the model (theta and phi matrix).

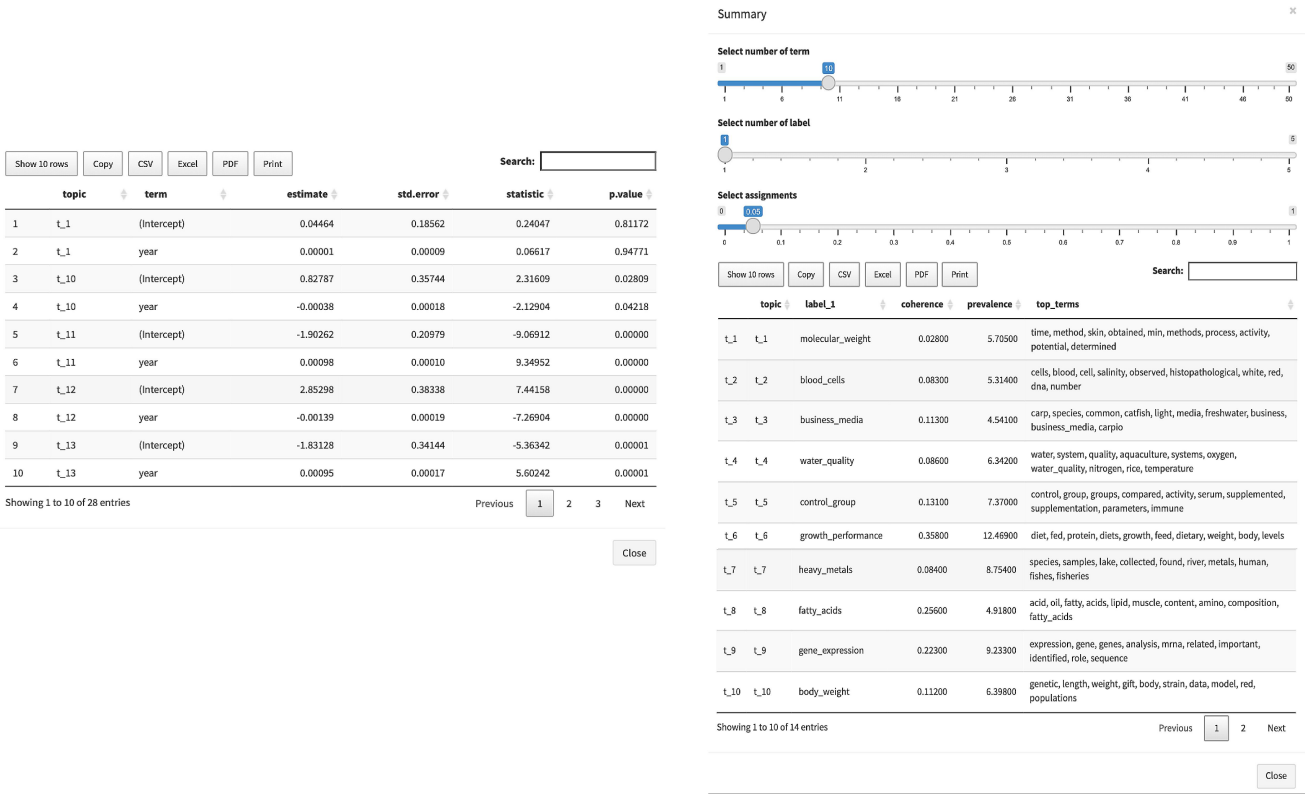


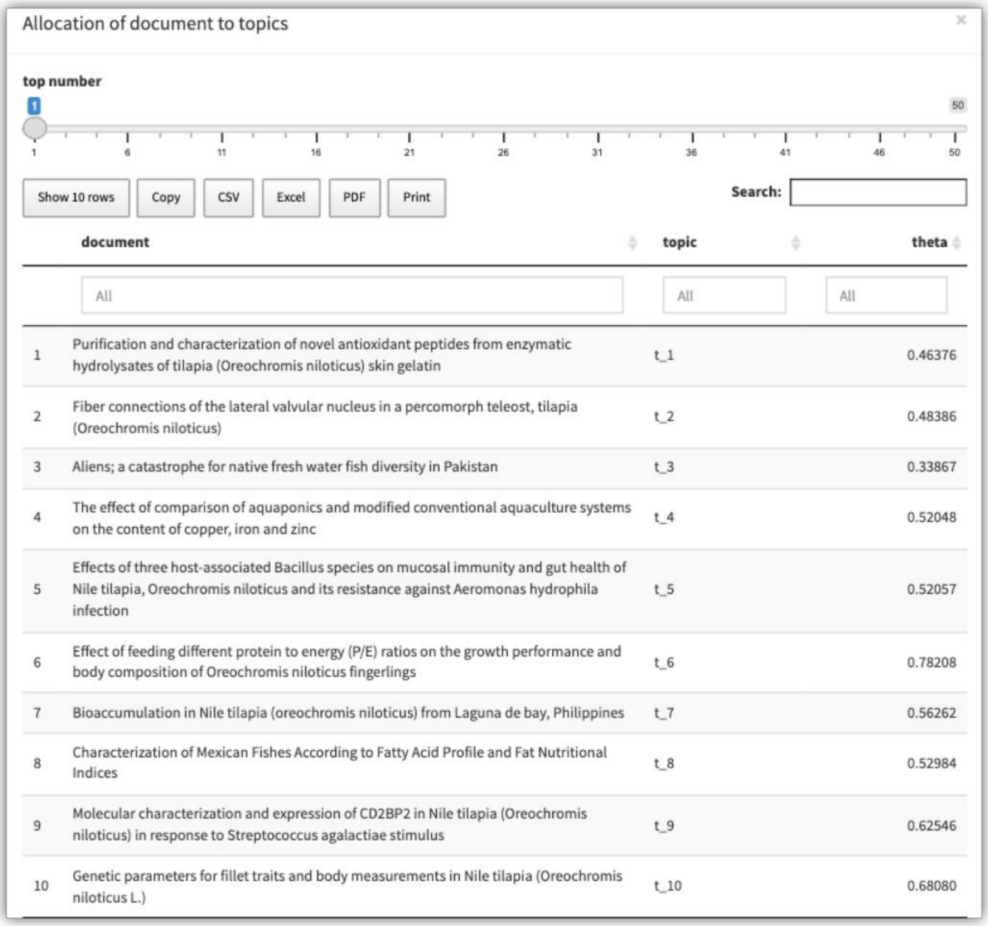
Figure 10. Output tabular of the model Trends (**left**) summary (**right**).

5.4. Postprocessing

Among the main outputs of the topic modelling algorithm are the collection of terms in relation to the frequencies of occurrence that characterize a topic and the composition, in percentage terms, for each document that has been analyzed. The distribution of topic terms does not come with a semantic interpretation. However, the topics can be properly labeled in most cases, inferring from the word frequency.

LDAShiny provides a topic labeling using a naive n-gram based topic algorithm from the textmineR [14] Package. However, as indicated above, these algorithms have limited capacity, so it is recommended that the labeling be validated by an expert in the research

area. If a domain expert is not available, it could generate incorrectly labeled topics and an invalid result [4].



	document	topic	theta
	All	All	All
1	Purification and characterization of novel antioxidant peptides from enzymatic hydrolysates of tilapia (Oreochromis niloticus) skin gelatin	t_1	0.46376
2	Fiber connections of the lateral valvular nucleus in a percomorph teleost, tilapia (Oreochromis niloticus)	t_2	0.48386
3	Aliens; a catastrophe for native fresh water fish diversity in Pakistan	t_3	0.33867
4	The effect of comparison of aquaponics and modified conventional aquaculture systems on the content of copper, iron and zinc	t_4	0.52048
5	Effects of three host-associated Bacillus species on mucosal immunity and gut health of Nile tilapia, Oreochromis niloticus and its resistance against Aeromonas hydrophila infection	t_5	0.52057
6	Effect of feeding different protein to energy (P/E) ratios on the growth performance and body composition of Oreochromis niloticus fingerlings	t_6	0.78208
7	Bioaccumulation in Nile tilapia (oreochromis niloticus) from Laguna de bay, Philippines	t_7	0.56262
8	Characterization of Mexican Fishes According to Fatty Acid Profile and Fat Nutritional Indices	t_8	0.52984
9	Molecular characterization and expression of CD2BP2 in Nile tilapia (Oreochromis niloticus) in response to Streptococcus agalactiae stimulus	t_9	0.62546
10	Genetic parameters for fillet traits and body measurements in Nile tilapia (Oreochromis niloticus L.)	t_10	0.68080

Figure 11. Output tabular allocations.

As a result, the 14 topics found reflect an overview of the research on the species *O. niloticus*. This shows one of the main benefits of the application, by providing information on a large collection of documents with relatively little effort on the part of the researcher.

After the label of the themes has been verified, the researcher can choose the articles that are relevant to the literature review. For example, if their main interest is in genetic expression, a specific number of articles on that topic can be selected by using the tabular output "allocation of document to topic."

LDAShiny allows the analysis of the dynamics of the topics over time in terms of their proportions, making it easier to understand the general trend of research. The increase in the proportion of some topics indicates that these are emerging fields of research, while their decrease shows a trend of less research interest. In addition, the high frequency pattern found at the beginning in some topics, which was followed by a negative trend during the period of study, has indicated a possible decrease in their popularity within the scientific community. This facilitates researchers not only to identify emerging research topics but also to visualize changes in the research focus.

The results obtained for the distribution of the topics by year, are also represented by a heatmap (Figure 12). In it, the color of the pixel represents the probability that a certain topic will be mentioned in a particular year.

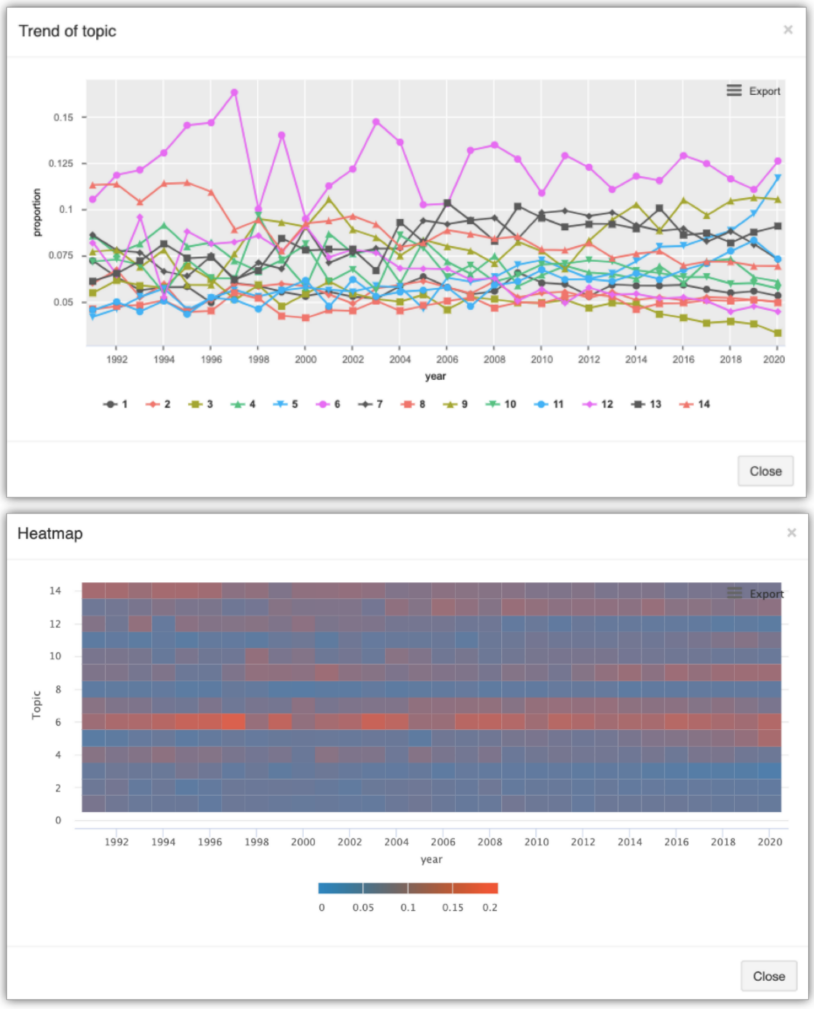


Figure 12. Graphical output of trend model.

6. Discussion

In 2004, Blei implemented the lda-c Software which was the first software that performs variational inference [6].

Commonly, available specialized open source software tools focus on stages or steps of text mining. Thus, they only focus for example on the preprocessing phase or in the inference phase. Some of these packages allow academics and researchers with a medium knowledge of a programming language (such as R) to follow the workflow required for an exploratory review of scientific literature. However, the available packages do not provide a GUI. In order to solve this problem, a R package with web-based GUI was developed in shiny, facilitating the execution of the exploratory review of scientific literature. Thus, LDAShiny facilitates the integral aspects of a review through LDA from preprocessing, inference (choosing between a set of models) and postprocessing (identifying trends in research). In addition, the information generated can be downloaded in various formats both in tabular and graph forms.

An additional benefit of LDAShiny is that it allows reproducibility, since all the steps of the exploratory review process can be reviewed and evaluated by other researchers in an agile and transparent way compared to a traditional review. In addition, the proposed application could be used to monitor the research trend. For instance, in the case of the example used, when more articles are published on the species under analysis, the review could be easily updated, since these new publications will be classified in related topics.

We found that the default parameters in the application example in the preprocessing steps offered a valid and usable result for the exploratory analysis of the literature on

O. niloticus. The execution time of the analysis did not take long, which is beneficial for the researcher. Usually, this time is mainly computer time and, although it is necessary to validate this verification, it requires less time than if a manual review were performed.

LDAShiny includes tools for undertaking an exploratory examination of scientific literature, as well as preprocessing features such as generating a corpus and removing stopwords, numbers and constructing ngrams. The tool also allows a document-term matrix to be created from a collection of documents, in a flexible manner, with a rudimentary understanding of the R programming language. Moreover, it facilitates researchers who are unfamiliar with R language to employ machine learning techniques. Users can point and click to generate a graphical or tabular of representation of the DTM matrix that can be downloaded in a variety of forms and saved and/or exported.

It is important to note that the preprocessing phase is an iterative process, as identifying stopwords, which might be difficult at initially [51,52] find that the preprocessing stages, in particular, can have a significant impact on the validity of the results, emphasizing the necessity of choosing the model parameters. However, for an exploratory study of the scientific literature, the default parameters and cleanup methods established in LDAShiny provide a legitimate and usable result.

In terms of inference, the app includes different metrics. While they are already available in R's CRAN in packages like topicmodel, ldatunning, and texmineR, the tool makes them easier to set by allowing them to be adjusted through easy-to-use interactive menus.

Although LDAShiny includes an algorithm for labeling, the identification of the topics is an important component of the post-processing phase. Because a mislabeled topic could lead to invalid results it is best if an expert reviews the labeling.

We might remark, for example, that one of the benefits of utilizing LDAShiny for a literature review is that the decision to include or delete articles can be postponed until a later stage when additional information is available, resulting in a decision-making process. Because all elements of the exploratory review process are reproducible, LDAShiny provides more reproducibility and transparency, allowing other researchers to analyze the entire review process in detail.

Although LDAShiny was evaluated in a study of academic scientific literature on the species *O. niloticus*, it is expected that researchers from various fields will put the tool to the test, as there is no technological reason why other types of documents cannot be included.

This is the first edition of the program. It is planned to add more features in future editions, such as the ability to read whole articles rather than just abstracts. This can improve the quality of the topics and provide more detail on latent themes [24].

7. Conclusions

In any scientific area, reviewing the scientific literature is a necessary step of the research process. As the number of publications increases over time, the task of acquiring knowledge becomes increasingly difficult.

This work aimed to present a tool, the LDAShiny package, that allow researchers to use topic modeling based on the use of the latent Dirichlet allocation. Thus, it is possible to perform an exploratory review of the literature, reducing the need to read articles manually and allowing the possibility to analyze a greater number of articles. The LDAShiny package was designed to be easily used by any researcher, as it requires less technical knowledge than using a normal topic model would imply.

LDAShiny development can also be addressed to the developer community, since the sources are published on GitHub (https://github.com/cran/LDAShiny), which allows the creation of shared development. The application can be run on a computer locally. Nonetheless, shiny can also be hosted on a server and deployed online.

There are options for preprocessing, inference, topic modeling and postprocessing in the application. The papers are loaded, cleaned, and authenticated during the preprocessing stage. The LDA approach is utilized in the inference step to estimate the number

of topics that were used in the topic modeling phase. The post-processing step generates topic model results.

LDAShiny was designed with a step-by-step approach, and with a friendly interface allowing accessibility. However, researchers from various fields are expected to test it and provide valuable evaluations to improve its use.

The application was tested with 6196 scientific publications on the species *O. niloticus*. This data was processed in a short amount of time, taking roughly three days on a five-core laptop. The data were divided into 14 categories.

We consider LDAShiny to be especially relevant for researchers in various areas, as the literature review is essential for gaining an overview of the different research fields, where a shiny-based graphical user interface can allow more documents to be reviewed, more frequently. The LDAShiny package provides an interface that allows users to use the features interactively and in a friendly way, which can be used not only by statisticians but also by analysts who are unfamiliar with the R environment.

Author Contributions: Conceptualization, J.D.l.H.-M. and M.J.F.-G.; methodology, J.D.l.H.-M.; software, J.D.l.H.-M.; validation, S.M.; formal analysis, M.J.F.-G.; investigation, J.D.l.H.-M.; resources, S.M.; data curation, J.D.l.H.-M.; writing—original draft preparation, J.D.l.H.-M.; writing—review and editing, S.M. and M.J.F.-G.; visualization, J.D.l.H.-M.; supervision, M.J.F.-G.; project administration, J.D.l.H.-M.; funding acquisition, S.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially funded by FCT (Fundação para a Ciência e a Tecnologia) for the financial support of this work through the project UID/MAR/04292/2020, attributed to MARE-Marine and Environmental Sciences Centre, Portugal, and by the Integrated Programme of SR&TD "SmartBioR" (reference Cen-tro-01-0145-FEDER-000018), cofounded by the Centro 2020 program, Portugal2020, European Union, through the European Regional Development Fund.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The software is distributed through CRAN https://cran.r-project.org, and data and code will be made available on Github at https://github.com/cran/LDAShiny.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Brocke, J.; Simons., A.; Niehaves, B.; Niehaves, B.; Reimer, K.; Plattfaut, R.; Cleven, A. Reconstructing the giant: On the importance of rigour in documenting the literature search process. In Proceedings of the 17th European Conference on Information Systems, Verona, Italy, 7–9 June 2009; p. 161. Available online: http://aisel.aisnet.org/ecis2009/161 (accessed on 16 April 2021).
2. Harzing, A.W.; Alakangas, S. Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison. *Scientometrics* **2016**, *106*, 787–804. [CrossRef]
3. DiMaggio, P.; Nag, M.; Blei, D. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of US government arts funding. *Poetics* **2013**, *41*, 570–606. [CrossRef]
4. Asmussen, C.B.; Muller, C. Smart literature review: A practical topic modelling approach to exploratory literature review. *J. Big Data.* **2019**, *6*, 93. [CrossRef]
5. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020. Available online: https://www.R-project.org/ (accessed on 16 April 2021).
6. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet allocation. *J. Mach. Learn Res.* **2003**, *3*, 993–1022.
7. Chang, J. *lda: Collapsed Gibbs Sampling Methods for Topic Models*, R package version 1.4.2; R Foundation for Statistical Computing: Vienna, Austria, 2015. Available online: https://CRAN.R-project.org/package=lda (accessed on 16 April 2021).
8. Erskine, N. *lda.svi: Fit Latent Dirichlet Allocation Models Using Stochastic Variational Inference*, R package version 0.1.0.; R Foundation for Statistical Computing: Vienna, Austria, 2015. Available online: https://CRAN.Rproject.org/package=lda.svi (accessed on 16 April 2021).
9. Rieger, J. *ldaPrototype: Prototype of Multiple Latent Dirichlet Allocation Runs*, R package version 0.1.1; R Foundation for Statistical Computing: Vienna, Austria, 2015. Available online: https://CRAN.R-project.org/package=ldaPrototype (accessed on 16 April 2021).
10. Nikita, M. *ldatuning: Tuning of the Latent Dirichlet Allocation Models Parameters*, R package version 1.0.2.; R Foundation for Statistical Computing: Vienna, Austria. Available online: https://CRAN.R-project.org/package=ldatuning (accessed on 16 April 2021).

11. Sievert, C.; Shirley, K. *LDAvis: Interactive Visualization of Topic Models*, R package version 0.3.2.; R Foundation for Statistical Computing: Vienna, Austria. Available online: https://CRAN.R-project.org/package=LDAvis (accessed on 16 April 2021).
12. Friedman, D. *topicdoc: Topic-Specific Diagnostics for LDA and CTM Topic Models*, R package version 0.1.0.; R Foundation for Statistical Computing: Vienna, Austria. Available online: https://CRAN.R-project.org/package=topicdoc (accessed on 16 April 2021).
13. Grun, B.; Hornik, K. topicmodels: An R Package for Fitting Topic Models. *J. Stat. Softw.* **2011**, *40*, 1–30. [CrossRef]
14. Jones, T. *textmineR: Functions for Text Mining and Topic Modeling*, R package version 3.0.4.; R Foundation for Statistical Computing: Vienna, Austria, 2019. Available online: https://CRAN.R-project.org/package=textmineR (accessed on 16 April 2021).
15. Blei, D.M. Probabilistic topic models. *Commun. ACM* **2012**, *55*, 77–84. [CrossRef]
16. Kao, A.; Poteet, S.R. *Natural Language Processing and Text Mining*, 1st ed.; Springer Science & Business Media: London, UK, 2007; p. 265.
17. Deerwester, S.; Dumais, S.T.; Furnas, G.W.; Landauer, T.K.; Harshman, R. Indexing by latent semantic analysis. *J. Assoc. Inf. Sci. Technol.* **1990**, *41*, 391–407. [CrossRef]
18. Hofmann, T. Probabilistic latent semantic indexing. In Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Berkeley, CA, USA, 15–19 August 1999; pp. 50–57. [CrossRef]
19. Grimmer, J. A Bayesian hierarchical topic model for political texts: Measuring expressed agendas in Senate press releases. *Political Anal.* **2010**, *18*, 1–35. [CrossRef]
20. Jacobi, C.; Van Atteveldt, W.; Welbers, K. Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digit. J.* **2016**, *4*, 89–106. [CrossRef]
21. Iwata, T.; Saito, K.; Ueda, N.; Stromsten, S.; Griffiths, T.L.; Tenenbaum, J.B. Parametric embedding for class visualization. *Neural. Comput.* **2007**, *19*, 2536–2556. [CrossRef] [PubMed]
22. Wang, Y.; Sabzmeydan, P.; Mori, G. Semi-latent Dirichlet allocation: A hierarchical model for human action recognition. In Proceedings of the Second Workshop, Human Motion—Understanding, Modeling, Capture and Animation, Rio de Janeiro, Brazil, 20 October 2007; Elgammal, A., Rosenhahn, B., Klette, R., Eds.; Springer: Berlin/Heidelberg, Germany, 2007. [CrossRef]
23. Griffths, T.L.; Steyvers, M. Finding scientific topics. *Proc. Natl. Acad. Sci. USA* **2004**, *101* (Suppl. 1), 5228–5235. [CrossRef]
24. Syed, S.; Spruit, M. Full-text or abstract? Examining topic coherence scores using latent dirichlet allocation. In Proceedings of the 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Tokyo, Japan, 19–21 October 2017; pp. 165–174. [CrossRef]
25. Newman, D.; Smyth, P.; Welling, M.; Asuncion, A.U. Distributed inference for latent Dirichlet allocation. In Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 3–6 December 2007; pp. 1081–1088.
26. Porteous, I.; Newman, D.; Ihler, A.; Asuncion, A.; Smyth, P.; Welling, M. Fast collapsed gibbs sampling for latent Dirichlet allocation. In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, NV, USA, 24–27 August 2008; pp. 569–577.
27. Blei, D.M.; Jordan, M.I. Variational inference for Dirichlet process mixtures. *Bayesian Anal.* **2006**, *1*, 121–143. [CrossRef]
28. Teh, Y.W.; Newman, D.; Welling, M. A collapsed variational Bayesian inference algorithm for latent Dirichlet allocation. In Proceedings of the 20th Annual Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 4–7 December 2007; pp. 1353–1360.
29. Wang, C.; Paisley, J.; Blei, D.M. Online variational inference for the hierarchical Dirichlet process. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, Machine Learning Research, Fort Lauderdale, FL, USA, 11–13 April 2011; pp. 752–760. Available online: http://proceedings.mlr.press/v15/wang11a.html (accessed on 17 April 2021).
30. Vijayarani, S.; Ilamathi, M.J.; Nithya, M. Preprocessing techniques for text mining-an overview. *Int. J. Comput. Sci. Commun. Netw.* **2015**, *5*, 7–16.
31. Jurafsky, D.; Martin, J.H. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*; Pearson: Hoboken, NJ, USA, 2008.
32. Manning, C.D.; Manning, C.D.; Schutze, H. *Foundations of Statistical Natural Language Processing*, 2nd ed.; MIT Press: Cambridge, MA, USA, 1999.
33. Luhn, H.P. The automatic creation of literature abstracts. *IBM J. Res. Dev.* **1958**, *2*, 159–165. [CrossRef]
34. Benoit, K.; Muhr, D.; Watanabe, K. *Stopwords: Multilingual Stopword Lists*, R package version 0.9.0; R Foundation for Statistical Computing: Vienna, Austria, 2017. Available online: https://CRAN.R-project.org/package=stopwords (accessed on 17 April 2021).
35. Porter, M.F. An algorithm for suffix stripping. *Programming* **1980**, *14*, 130–137. [CrossRef]
36. Yano, T.; Smith, N.A.; Wilkerson, J.D. Textual predictors of bill survival in congressional committees. In Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Montreal, QC, Canada, 3–8 June 2012; pp. 793–802.
37. Grimmer, J.; Stewart, B.M. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Anal.* **2013**, *21*, 267–297. [CrossRef]
38. Blei, D.M.; Lafferty, J.D. A correlated topic model of science. *Ann. Appl. Stat.* **2007**, *1*, 17–35. [CrossRef]
39. Sbalchiero, S.; Eder, M. Topic modeling, long texts and the best number of topics. Some problems and solutions. *Qual Quant.* **2020**, *54*, 1095–1108. [CrossRef]

40. Newton, M.A.; Raftery, A.E. Approximate Bayesian inference with the weighted likelihood bootstrap. *J. R. Stat. Soc. Series B. Stat. Methodol.* **1994**, *56*, 3–26. [CrossRef]
41. Roder, M.; Both, A.; Hinneburg, A. Exploring the space of topic coherence measures. In Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, 31 January–6 February 2015; pp. 399–408.
42. Harris, Z.S. Distributional structure. *Word* **1954**, *10*, 146–162. [CrossRef]
43. Arun, R.; Suresh, V.; Veni Madhavan, C.E.; Narasimha Murthy, M.N. On finding the natural number of topics with latent dirichlet allocation: Some observations. In *Advances in Knowledge Discovery and Data Mining. PAKDD 2010. Lecture Notes in Computer Science*; Zaki, M.J., Yu, J.X., Ravindran, B., Pudi, V., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; Volume 6118. [CrossRef]
44. Cao, J.; Xia, T.; Li, J.; Zhang, Y.; Tang, S. A density-based method for adaptive LDA model selection. *Neurocomputing* **2009**, *72*, 1775–1781. [CrossRef]
45. Deveaud, R.; SanJuan, E.; Bellot, P. Accurate and effective latent concept modeling for ad hoc information retrieval. *Doc. Numer.* **2014**, *17*, 61–84. [CrossRef]
46. Lewis, S.C.; Zamith, R.; Hermida, A. Content analysis in an era of big data: A hybrid approach to computational and manual methods. *J. Broadcast. Electron. Media.* **2013**, *57*, 34–52. [CrossRef]
47. Lau, J.H.; Grieser, K.; Newman, D.; Baldwin, T. Automatic labelling of topic models. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, OR, USA; 2011; pp. 1536–1545.
48. Xiong, H.; Cheng, Y.; Zhao, W.; Liu, J. Analyzing scientific research topics in manufacturing field using a topic model. *Comput. Ind. Eng.* **2019**, *135*, 333–347. [CrossRef]
49. Chang, W.; Cheng, J.; Allaire, J.; Xie, Y.; McPherson, J. *Shiny: Web Application Framework for R*, R package version 1.4.0.2; R Foundation for Statistical Computing: Vienna, Austria. Available online: https://CRAN.R-project.org/package=shiny (accessed on 17 April 2021).
50. Chang, J.; Gerrish, S.; Wang, C.; Boyd-Graber, J.L.; Blei, D.M. *Reading tea leaves: How humans interpret topic models. Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2009; pp. 288–296.
51. Denny, M.J.; Spirling, A. Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it. *Political Anal.* **2018**, *26*, 168–189. [CrossRef]
52. Maier, D.; Waldherr, A.; Miltner, P.; Wiedemann, G.; Niekler, A.; Keinert, A.; Adam, S. Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Commun. Methods Meas.* **2018**, *12*, 93–118. [CrossRef]

ANEXO 3. Código fuente Aplicación LDAShiny

ui.R (inicio)

```
require(shiny)
require(shinyWidgets)
require(shinycssloaders)

exp.stop <- c()

shinydashboard::dashboardPage( skin = "blue",
###header
  shinydashboard::dashboardHeader(title = "LDAShiny",tags$li(class = "dropdown", actionLink("stop_radiant", "Stop",
icon = icon("power-off"),
                                                            onclick = "setTimeout(function() {window.close();}, 100); ")
                                                )
                                ),
# dashboardSidebar############################
shinydashboard::dashboardSidebar(
  width = 400,
  shinydashboard::sidebarMenu(
    shinydashboard::menuItem("About LDAShiny",
                             tabName = "tab0"
    ),
    shinydashboard::menuItem("Preprocessing",
                             tabName = "tab1",
                             icon = icon("filter")
    ),
    shinydashboard::menuItem("Document Term Matrix Visualizations",
                             tabName = "tab2",
                             icon = icon("chart-bar")
    ),
    shinydashboard::menuItem("Number of topic (inference)",
                             tabName = "tab3",
                             icon = icon("calculator")
    ),
    shinydashboard::menuItem("LDA model",
                             tabName = "tab4",
                             icon = icon("desktop"),
                             shinydashboard::menuSubItem("Run model",
                                                         tabName = "tab41",
                                                         icon = icon("gear")
                             ),
                             shinydashboard::menuSubItem("Download tabular results",
                                                         tabName = "tab42",
                                                         icon = icon("download")
                             ),
                             shinydashboard::menuSubItem("Download graphics results",
                                                         tabName = "tab43",
                                                         icon = icon("download")
                             )
    ),

    shinydashboard::menuItem("Tutorial",
                             tabName = "tab5",
                             icon = icon("chalkboard-teacher")
    )
  )
),
## dashboardBody ############################
shinydashboard::dashboardBody(
  tags$head(
    tags$link(
      rel = "stylesheet",
      type = "text/css",
      href = "custom.css"
```

```
  )
),
shinyWidgets::useSweetAlert(),
shinydashboard::tabItems(
  shinydashboard::tabItem(
    tabName = "tab1",
    shinydashboard::box(
      width = 400,
      title = "Preprocessing",
      status = "primary",
      solidHeader = TRUE,
      collapsible = TRUE,
      shiny::sidebarPanel(
        shinyjs::useShinyjs(),
        helpText(h3("Upload data file")),
        br(),
        helpText(h4("Example data")),
        br(),
        shinyWidgets::prettyCheckbox(
          inputId = "example",
          label = "Use example data set?",
          value = FALSE,
          status = "warning"
        ),
        br(),
        br(),
        fileInput("file", "Choose CSV File",
                  multiple = T
        ),
        helpText("Select the read.table parameters below"),
        shiny::checkboxInput(
          inputId = "header",
          label = "Header",
          value = TRUE
        ),
        shiny::checkboxInput(
          inputId = "stringAsFactors",
          "stringAsFactors", FALSE
        ),
        shiny::radioButtons(
          inputId = "sep",
          label = "Separator",
          choices = c(
            "Comma" = ",",
            "Semicolon" = ";",
            "Tab" = "\t", "Space" = " "
          ),
          selected = ","
        ),
        uiOutput("selectfile"),
        br(),
        br(),
        helpText(h3("Data cleanning")),
        br(),
        helpText(" Click the incorporate information button three times"),
        shinyWidgets::actionBttn("choice", "incorporate information",
                                 style = "float",
                                 block = TRUE,
                                 color = "primary"
        ),
        br(),
        br(),
        div(
          style = "display: inline-block;vertical-align:top; width: 170px;",
          selectInput("column1", "Select id document",
                      choices = NULL
          )
        ), # no choices before uploading
        shinyBS::bsPopover("column1",
```

```
                "Select which column contains the id.",
                options = list(container = "body")
),
div(
  style = "display: inline-block;vertical-align:top; width: 170px;",
  selectInput("column2",
              "Select document vector",
              choices = NULL
  )
), # no choices before uploading

shinyBS::bsPopover("column2",
                "Select which column contains the column of text.",
                options = list(container = "body")
),

div(
  style = "display: inline-block;vertical-align:top; width: 170px;",
  selectInput(
    inputId = "column3",
    label = "Select publish year",
    choices = NULL
  )
), # no choices before uploading

shinyBS::bsPopover("column3",
                "Select which column contains years",
                options = list(container = "body")
),
shinyWidgets::awesomeRadio(
  inputId = "ngrams",
  label = "ngrams",
  choices = c("Unigrams" = 1L, "Bigrams" = 2L, "Trigrams" = 3L),
  selected = 1L,
  status = "primary"
),
shinyWidgets::awesomeCheckbox(
  inputId = "removenumber",
  label = "Remove number",
  value = TRUE,
  status = "danger"
),
shinyWidgets::pickerInput(
  inputId = "Language",
  label = "Select language for stopword",
  choices = c(
    "da", "nl", "en", "fi",
    "fr", "de", "hu", "it",
    "no", "pt", "ro",
    "ru", "es", "sv"
  ),
  choicesOpt = list(
    subtext = paste(Languages <- c(
      "danish", "dutch", "english", "finnish",
      "french", "german", "hungarian", "italian",
      "norwegian", "portuguese", "romanian",
      "russian", "spanish", "swedish"
    ))
  )
),
textInput("stopwords",
          label = "Stop Words",
          value = paste(exp.stop, collapse = ", "),
          placeholder = "also, such, really..."
),
shinyBS::bsPopover("stopwords", "Include additional stop words to remove(words
                           must be separated by commas)",
                options = list(container = "body")
),
```

```
      br(),
      div(
        style = "display: inline-block;vertical-align:top; width: 170px;",
        shinyWidgets::awesomeCheckbox(
          inputId = "checkStemming",
          label = "Stemming",
          value = FALSE,
          status = "danger"
        )
      ),
      shinyBS::bsPopover("checkStemming", "Click if you want to stemming",
                         options = list(container = "body")
      ),
      div(
        style = "display: inline-block;vertical-align:top; width: 200px;",
        selectInput("Stemm", " Stemming langauge",
                    choices = c(
                      "porter", "danish", "dutch", "english", "finnish",
                      "french", "german", "hungarian", "italian",
                      "norwegian", "portuguese", "romanian",
                      "russian", "spanish", "swedish"
                    )
        )
      ),
      br(),
      sliderInput(
        inputId = "sparce",
        label = "Sparsity:",
        min = 0.00001,
        max = 0.99999,
        value = 0.995,
        step = 0.0001
      ),
      shinyBS::bsPopover("minDoc", "Remove sparse terms:",
                         options = list(container = "body")
      ),
      br(),
      # shinyjs::hidden(
      div(
        id = "dTm",
        style = "display:inline-block",
        shinyWidgets::actionBttn(
          inputId = "dtm.update",
          label = "Create DTM",
          style = "float",
          block = TRUE,
          color = "primary"
        )
      ),
      bsModalNoClose(
        id = "DTMdim",
        title = "dim DTM",
        trigger = "dTm",
        size = "large",
        shinycssloaders::withSpinner(DT::DTOutput("Table_dim"))
      ),

      DT::DTOutput("table_display1"),
      DT::DTOutput("table_display2")
    ),
    mainPanel(
      uiOutput("tb"),
      uiOutput("tb2")
    ) # tab panel display
  )
), shinydashboard::tabItem(
  tabName = "tab2",
  collapsible = TRUE,
  shinydashboard::box(
```

```
width = 350,
title = "Visualization",
status = "primary",
solidHeader = TRUE,
collapsible = TRUE,
## View data action button hidden initially until the dataset is loaded
div(
  id = "data_b",
  style = "display:inline-block",
  shinyWidgets::actionBttn(
    inputId = "data",
    label = "View Data",
    icon = icon("table"),
    style = "float",
    block = TRUE,
    color = "primary"
  )
),
## View plot action button hidden initially until the dataset is loaded
div(
  id = "plot_b",
  style = "display:inline-block",
  shinyWidgets::actionBttn(
    inputId = "plot",
    label = "View barplot",
    icon = icon("bar-chart"),
    style = "float",
    block = TRUE,
    color = "primary"
  )
),

div(
  id = "plot_c",
  style = "display:inline-block",
  shinyWidgets::actionBttn(
    inputId = "plot2",
    label = "View wordcloud",
    icon = icon("bar-chart"),
    style = "float",
    block = TRUE,
    color = "primary"
  )
),

bsModalNoClose(
  id = "dataset",
  title = "Basic corpus statistics",
  trigger = "data",
  size = "large",
  shinycssloaders::withSpinner(DT::DTOutput("data_b"))
),

## Shiny BS Modal to display the plot inside a modal
## A spinner is also added
bsModalNoClose(
  id = "Plot", title = "Barplot", trigger = "plot", size = "large",
  div(
    style = "display: inline-block;vertical-align:top; width: 120px;",
    tags$h5("Select number of term")
  ),
  div(
    style = "display: inline-block;vertical-align:top; width: 120px;",
    shinyWidgets::dropdown(sliderInput(
      inputId = "b",
      label = "Select number of word",
      min = 10,
      max = 500,
      value = 10
```

```
        ),
        style = "unite",
        icon = icon("gear"),
        status = "danger",
        width = "500px",
        animate = shinyWidgets::animateOptions(
          enter = shinyWidgets::animations$fading_entrances$fadeInLeftBig,
          exit = shinyWidgets::animations$fading_exits$fadeOutRightBig
        )
        )
      ),
      shinycssloaders::withSpinner(highcharter::highchartOutput("plot_gg"))
    ),

    bsModalNoClose(
      id = "Plot2",
      title = "Wordcloud",
      trigger = "plot2",
      size = "large",
      div(
        style = "display: inline-block;vertical-align:top; width: 120px;",
        tags$h5("Select number of term")
      ),
      div(
        style = "display: inline-block;vertical-align:top; width: 120px;",
        shinyWidgets::dropdown(sliderInput(
          inputId = "c",
          label = "Select number of word",
          min = 10,
          max = 500,
          value = 10
        ),
        style = "unite",
        icon = icon("gear"),
        status = "danger",
        width = "500px",
        animate = shinyWidgets::animateOptions(
          enter = animations$fading_entrances$fadeInLeftBig,
          exit = animations$fading_exits$fadeOutRightBig
        )
        )
      ),
      shinycssloaders::withSpinner(highcharter::highchartOutput("plot_gf"))
    )
  )
),
shinydashboard::tabItem(
  tabName = "tab3",
  width = 300,
  shinydashboard::box(
    title = "3. Number of topic",
    status = "primary",
    solidHeader = TRUE,
    collapsible = TRUE,
    shinydashboard::tabBox(
      width = 300,
      title = "",
      # The id lets us use input$tabset1 on the server to find the current tab
      id = "tabset1",
      tabPanel(
        "Coherence",
        shiny::sidebarPanel(
          width = 300,
          helpText(h3("Candidate number of topics k")),
          div(
            style = "display: inline-block;vertical-align:top; width: 120px;",
            numericInput(
              inputId = "num1",
              label = "from",
```

```
          value = 2,
          min = 2
        )
      ),
      div(
        style = "display: inline-block;vertical-align:top; width: 120px;",
        numericInput(
          inputId = "num2",
          label = "to",
          value = 2,
          min = 2
        )
      ),
      div(
        style = "display: inline-block;vertical-align:top; width: 110px;",
        numericInput(
          inputId = "num3",
          label = "by",
          value = 1,
          min = 1
        )
      ),
      textInput(
        inputId = "OtherKCoherence",
        label = "Others K",
        value = paste(exp.stop, collapse = ", "),
        placeholder = "Enter values separated by a comma...50, 100, 150,..."
      ),
      textOutput("OthKcoh"),
      br(),
      helpText(h3("Parameters control Gibbs sampling")),
      div(
        style = "display: inline-block;vertical-align:top; width: 110px;",
        numericInput(
          inputId = "num4",
          label = "Iteractions",
          value = 10,
          min = 10
        )
      ),
      div(
        style = "display: inline-block;vertical-align:top; width: 110px;",
        numericInput(
          inputId = "num5",
          label = "Burn-in",
          value = 5,
          min = 5
        )
      ),
      shinyBS::bsPopover(
        id = "num5",
        "Specifies how many iterations are discarded If is too low,the Gibbs sample will be polluted from the wrong
distribution.If is too large, the only penalty is wasting computational effort",
        options = list(container = "body")
      ),
      br(),
      helpText(h3("Hyper-parameter")),
      withMathJax(),
      div(
        style = "display: inline-block;vertical-align:top; width: 110px;",
        numericInput(
          inputId = "num6",
          label = h4("\\( \\alpha \\)"),
          value = 0.1,
          min = 0.01,
          max = 1,
          step = 0.01
        )
      ),
```

```
          br(),
          shinyBS::bsPopover("num6",
                             "Assuming symmetric Dirichlet distributions (for simplicity), a low alpha value places more weight
on having each document composed of only a few dominant topics",
                             options = list(container = "body")
          ),
          br(),
          shinyalert::useShinyalert(),
          div(
            id = "Run1",
            style = "display:inline-block",
            shinyWidgets::actionBttn(
              inputId = "Run.model1",
              label = "Run",
              style = "float",
              block = TRUE,
              color = "primary"
            )
          ),
          bsModalNoClose(
            id = "Coherence",
            title = "Coherence",
            trigger = "Run1", size = "large",
            verbatimTextOutput("timeCoherence"),
            shinycssloaders::withSpinner(highcharter::highchartOutput("plot_gi"))
          )
        )
      ),
      tabPanel(
        "4-metrics",
        shiny::sidebarPanel(
          width = 300,
          helpText(h3("Candidate number of topics k")),
          div(
            style = "display: inline-block;vertical-align:top; width: 120px;",
            numericInput(
              inputId = "num7",
              label = "from",
              value = 2,
              min = 2
            )
          ),
          div(
            style = "display: inline-block;vertical-align:top; width: 120px;",
            numericInput(
              inputId = "num8",
              label = "to",
              value = 2,
              min = 2
            )
          ),
          div(
            style = "display: inline-block;vertical-align:top; width: 110px;",
            numericInput(
              inputId = "num9",
              label = "by",
              value = 1,
              min = 1
            )
          ),
          textInput(
            inputId = "OtherK4metric",
            label = "Others K",
            value = paste(exp.stop, collapse = ", "),
            placeholder = "Enter values separated by a comma...50, 100, 150,..."
          ),
          textOutput("OthK4metric"),
          br(),
          br(),
```

```
    awesomeRadio(
      inputId = "methods",
      label = helpText(h3("Estimation method")),
      choices = c("Gibbs", "VEM"),
      selected = "Gibbs"
    ),
    # checkboxGroupButtons(
    # inputId = "metric",
    # label = helpText(h3("Metrics")),
    # choices = c("Griffiths2004",
    # "CaoJuan2009",
    # "Arun2010",
    # "Deveaud2014"),
    # individual = TRUE,
    # checkIcon = list(
    # yes = tags$i(class = "fa fa-circle",
    #  style = "color: steelblue"),
    # no = tags$i(class = "fa fa-circle-o",
    # style = "color: steelblue"))
    # ),
    br(),
    shinyalert::useShinyalert(),
    div(
      id = "Run2",
      style = "display:inline-block",
      actionBttn(
        inputId = "Run.model2",
        label = "Run",
        style = "float",
        block = TRUE,
        color = "primary"
      )
    ),
    bsModalNoClose(
      id = "fourmetric",
      title = "Four metrics",
      trigger = "Run2",
      size = "large",
      verbatimTextOutput("timefourmetric"),
      withSpinner(highcharter::highchartOutput("plot_gj"))
    )
  )
),
tabPanel(
  "Perplexity",
  sidebarPanel(
    width = 300,
    helpText(h3("Candidate number of topics k")),
    div(
      style = "display: inline-block;vertical-align:top; width: 120px;",
      numericInput(
        inputId = "num13",
        label = "from",
        value = 2,
        min = 2
      )
    ),
    div(
      style = "display: inline-block;vertical-align:top; width: 120px;",
      numericInput(
        inputId = "num14",
        label = "to",
        value = 5,
        min = 2
      )
    ),
    div(
      style = "display: inline-block;vertical-align:top; width: 110px;",
      numericInput(
```

```
      inputId = "num15",
      label = "by",
      value = 2,
      min = 1
    )
   ),
   textInput(
    inputId = "OtherKLL",
    label = "Others K",
    value = paste(exp.stop,
                  collapse = ", "
    ),
    placeholder = "Enter values separated by a comma...50, 100, 150,..."
   ),
   textOutput("OthKLL"),
   helpText(h3("Parameters control Gibbs sampling")),
   div(
    style = "display: inline-block;vertical-align:top; width: 110px;",
    numericInput(
      inputId = "num16",
      label = "Iteractions",
      value = 10,
      min = 10
    )
   ),
   div(
    style = "display: inline-block;vertical-align:top; width: 110px;",
    numericInput(
      inputId = "num17",
      label = "Burn-in",
      value = 5,
      min = 5
    )
   ),
   div(
    style = "display: inline-block;vertical-align:top; width: 110px;",
    numericInput(
      inputId = "num18",
      label = "Thin",
      value = 3,
      min = 3
    )
   ),
   br(),
   br(),
   shinyalert::useShinyalert(),
   div(
    id = "Run3", style = "display:inline-block",
    actionBttn("Run.model3", "Run",
               style = "float",
               block = TRUE,
               color = "primary"
    )
   ),
   bsModalNoClose(
    id = "perplex",
    title = "Perplexity",
    trigger = "Run3",
    size = "large",
    verbatimTextOutput("timeloglike"),
    withSpinner(highcharter::highchartOutput("plot_gk"))
   )
  )
 ),
 tabPanel(
  "Harmonic mean",
  sidebarPanel(
   width = 300,
   helpText(h3("Candidate number of topics k")),
```

```
div(
  style = "display: inline-block;vertical-align:top; width: 120px;",
  numericInput("num19",
               label = "from",
               value = 2,
               min = 2
  )
),
div(
  style = "display: inline-block;vertical-align:top; width: 120px;",
  numericInput("num20",
               label = "to",
               value = 5,
               min = 5
  )
),
div(
  style = "display: inline-block;vertical-align:top; width: 110px;",
  numericInput("num21",
               label = "by",
               value = 1,
               min = 1
  )
),
textInput("OtherKHM",
          label = "Others K",
          value = paste(exp.stop, collapse = ", "),
          placeholder = "Enter values separated by a comma...50, 100, 150,..."
),
textOutput("Okhm"),
helpText(h3("Parameters control Gibbs sampling")),
div(
  style = "display: inline-block;vertical-align:top; width: 110px;",
  numericInput("num22",
               label = "Iteractions",
               value = 10,
               min = 10
  )
),
div(
  style = "display: inline-block;vertical-align:top; width: 110px;",
  numericInput("num23",
               label = "Burn-in",
               value = 5,
               min = 5
  )
),
div(
  style = "display: inline-block;vertical-align:top; width: 110px;",
  numericInput("num24",
               label = "Keep",
               value = 2,
               min = 2
  )
),
br(),
br(),
shinyalert::useShinyalert(),
div(
  id = "Run4", style = "display:inline-block",
  actionBttn("Run.model4",
             "Run",
             style = "float",
             block = TRUE,
             color = "primary"
  )
),
bsModalNoClose(
  id = "harmonic",
```

```
            title = "Harmonic mean",
            trigger = "Run4",
            size = "large", verbatimTextOutput("timeHmean"),
            withSpinner(highcharter::highchartOutput("plot_gl"))
          )
        )
      )
    )
  )
),

  shinydashboard::tabItem(
    tabName = "tab41",
    shinydashboard::box(
      width = 400,
      title = "4. LDA model",
      status = "primary",
      solidHeader = TRUE,
      collapsible = TRUE,
      sidebarPanel(
        width = 400,
        helpText(h3("k parameter (topic) ")),
        div(
          style = "display: inline-block;vertical-align:top; width: 120px;",
          numericInput("num25",
                       label = "",
                       value = 2,
                       min = 2
          )
        ),
        br(),
        helpText(h3("Iterations")),
        div(
          style = "display: inline-block;vertical-align:top; width: 110px;",
          numericInput("num26",
                       label = "Iteractions",
                       value = 10,
                       min = 10
          )
        ),
        div(
          style = "display: inline-block;vertical-align:top; width: 110px;",
          numericInput("num27",
                       label = "Burn-in",
                       value = 5,
                       min = 5
          )
        ),
        shinyBS::bsPopover("num9",
                           "Specifies how many iterations are discarded. If is too low,the Gibbs sample will be polluted from the
wrong distribution.If is too large, the only penalty is wasting computational effort",
                           options = list(container = "body")
        ),
        br(), helpText(h3("Hyper-parameter")),
        withMathJax(),
        div(
          style = "display: inline-block;vertical-align:top; width: 110px;",
          numericInput("num28",
                       label = h4("\\( \\alpha \\)"),
                       value = 0.1,
                       min = 0.01,
                       max = 1,
                       step = 0.01
          )
        ),
        br(),
        div(
          id = "Run5",
          style = "display:inline-block",
```

```
      actionBttn("Run.model5",
                 "Run LDA Model",
                 style = "float",
                 block = TRUE,
                 color = "primary"
      )
     ),
     bsModalNoClose(
      id = "ldasummary",
      title = "summary LDA model",
      trigger = "Run5",
      size = "large",
      # verbatimTextOutput("r2"),
      withSpinner(DT::DTOutput("sum"))
     )
    )
   )
),
shinydashboard::tabItem(
  tabName = "tab42",
  shinydashboard::box(
    width = 400,
    title = "Download tabular results",
    status = "primary",
    solidHeader = TRUE,
    collapsible = TRUE,
    div(
      id = "data_theta", style = "display:inline-block",
      shinyWidgets::actionBttn("datatheta",
                               "theta",
                               icon = icon("table"),
                               style = "float", block = TRUE, color = "primary"
      )
    ),
    bsModalNoClose(
      id = "datathet",
      title = "theta",
      trigger = "data_theta",
      size = "large",
      shinycssloaders::withSpinner(DT::DTOutput("theta"))
    ),
    div(
      id = "data_phi",
      style = "display:inline-block",
      shinyWidgets::actionBttn("dataphi",
                               "phi",
                               icon = icon("table"), style = "float",
                               block = TRUE, color = "primary"
      )
    ),
    bsModalNoClose(
      id = "datphi",
      title = "phi",
      trigger = "data_phi",
      size = "large",
      shinycssloaders::withSpinner(DT::DTOutput("phi"))
    ),
    div(
      id = "data_reg",
      style = "display:inline-block",
      shinyWidgets::actionBttn("reg",
                               "trend",
                               icon = icon("table"),
                               style = "float",
                               block = TRUE, color = "primary"
      )
    ),
    bsModalNoClose(
      id = "datareg",
```

```
        title = "Regression summary",
        trigger = "data_reg",
        size = "large",
        shinycssloaders::withSpinner(DT::DTOutput("reg"))
       ),
       div(
        id = "summ_LDA",
        style = "display:inline-block",
        shinyWidgets::actionBttn("sumLDA",
                                 "Summary LDA",
                                 icon = icon("table"),
                                 style = "float",
                                 block = TRUE,
                                 color = "primary"
        )
       ),
       bsModalNoClose(
        id = "SummLDA",
        title = "Summary",
        trigger = "summ_LDA",
        size = "large",
        sliderInput(
         inputId = "Topterm",
         label = "Select number of term",
         min = 1,
         max = 50,
         value = 10,
         step = 1
        ),
        sliderInput(
         inputId = "Labels",
         label = "Select number of label",
         min = 1,
         max = 5,
         value = 1,
         step = 1
        ),
        sliderInput(
         inputId = "assignments",
         label = "Select assignments",
         min = 0,
         max = 1,
         value = 0.05,
         step = 0.01
        ),
        shinycssloaders::withSpinner(DT::DTOutput("summLDA"))
       ),

       div(
        id = "Allocat", style = "display:inline-block",
        shinyWidgets::actionBttn("alloca", "Allocation", icon = icon("table"), style = "float", block = TRUE, color =
"primary")
       ),

       bsModalNoClose(
        id = "allocat", title = "Allocation of document to topics", trigger = "Allocat", size = "large",
        sliderInput(inputId = "topnumber", label = "top number", min = 1, max = 50, value = 1),
        shinycssloaders::withSpinner(DT::DTOutput("Alloca"))
       )
      )
     ),
     shinydashboard::tabItem(
      tabName = "tab43",
      shinydashboard::box(
       width = 400,
       title = "Download graphics results",
       status = "primary",
       solidHeader = TRUE,
       collapsible = TRUE,
```

```
div(
  id = "Trend",
  style = "display:inline-block",
  shinyWidgets::actionBttn("t_trend",
                           "trend",
                           style = "float",
                           block = TRUE,
                           color = "primary",
                           icon = icon("bar-chart")
  )
),
bsModalNoClose(
  id = "Trend_bs",
  title = "Trend of topic",
  trigger = "Trend",
  size = "large",
  shinycssloaders::withSpinner(highcharter::highchartOutput("plot_trend"))
),
div(
  id = "plot_cloud", style = "display:inline-block",
  actionBttn("plotcloud",
             "View wordcloud by topic",
             icon = icon("bar-chart"),
             style = "float",
             block = TRUE,
             color = "primary"
  )
),
bsModalNoClose(
  id = "plotworcloud",
  title = "Wordcloud",
  trigger = "plot_cloud",
  size = "large",
  div(style = "display: inline-block;vertical-align:top; width: 120px;", tags$h5("Select options")),
  div(
    style = "display: inline-block;vertical-align:top; width: 120px;",
    shinyWidgets::dropdown(numericInput("num29",
                                        label = "topic #",
                                        value = 1,
                                        min = 1
    ),
    sliderInput(
      inputId = "cloud",
      label = "Select number of word",
      min = 5,
      max = 100,
      value = 10
    ),
    style = "unite",
    icon = icon("gear"),
    status = "danger",
    width = "500px",
    animate = shinyWidgets::animateOptions(
      enter = animations$fading_entrances$fadeInLeftBig,
      exit = animations$fading_exits$fadeOutRightBig
    )
    )
  ),
  shinycssloaders::withSpinner(highcharter::highchartOutput("plot_worcloud"))
),
div(
  id = "heatmap",
  style = "display:inline-block",
  shinyWidgets::actionBttn("Heat",
                           "heat_map",
                           style = "float",
                           block = TRUE,
                           color = "primary",
                           icon = icon("bar-chart")
```

```
        )
      ),

      bsModalNoClose(
        id = "Heat_map",
        title = "Heatmap",
        trigger = "heatmap",
        size = "large",
        shinycssloaders::withSpinner(highcharter::highchartOutput("plot_heatmap"))
      )
    )
  ),

  shinydashboard::tabItem(
    tabName = "tab0",
    shiny::fluidPage(tags$iframe(
      src = "about.html",
      width = "100%",
      height = "1000px",
      frameborder = 0,
      scrolling = "auto"
    ))
  ),
  shinydashboard::tabItem(
    tabName = "tab5",
    shiny::fluidPage(shinydashboard::tabBox(
      title = "",
      # The id lets us use input$tabset1 on the server to find the current tab
      id = "tabset1", height = "1000px", width = "100%",
      tabPanel("English", tags$iframe(
        src = "A_brief_introduction_to_LDAShiny.html",
        width = "100%",
        height = "1000px",
        frameborder = 0,
        scrolling = "auto"
      )),
      tabPanel("Español", tags$iframe(
        src = "Una_breve_introducci-n_a_LDAShiny.html",
        width = "100%",
        height = "1000px",
        frameborder = 0,
        scrolling = "auto"
      ))
    ) )
  )
 )
)

)##final
```

Final ui.R

Server.UI

```
require("textmineR")
require("magrittr")
require("highcharter")
require("dplyr")
require("parallel")
require("ldatuning")
require("purrr")
require("topicmodels")
require("stringr")
require("broom")
require("DT")
```

```
shinyServer(function(input,output,session) {
  options(shiny.maxRequestSize=50000000*1024^2)
  output$selectfile <- renderUI({
       if(is.null(input$file)) {return()}
       list(hr(),
            helpText("Select the files for which you need
                     to see data and summary stats"),
            selectInput("Select", "Select",
                        choices=input$file$name)
            )
       })

## Summary Stats code ##
# this reactive output contains the summary of the dataset and display the summary in table format
 output$summexample <- renderPrint({
       if(input$example == FALSE){return()}
         dataexample <- system.file("extdata", "scopusJOSS.csv",
                                    package = "LDAShiny")
         data_example <- read.csv(dataexample)
       summary(data_example)
       })


output$summ <- renderPrint({
       if(is.null(input$example)){return()}
       summary(read.table (input$file$datapath[input$file$name==input$Select],
                           sep=input$sep,
                           header = input$header,
                           stringsAsFactors = input$stringAsFactors))
   })

observeEvent(input$example, {
       if(input$example == TRUE){
         shinyjs::disable("choice")
       } else {
         shinyjs::enable("choice")
       }
  })


    ## MainPanel tabset renderUI code ##
   # the following renderUI is used to dynamically g
   # enerate the tabsets when the file is loaded.
   # Until the file is loaded, app will not show the tabset.
   output$tb2 <- renderUI({
       if(input$example == FALSE){return()}
       else
          tabsetPanel(
             tabPanel("Statistical summary example",
                      verbatimTextOutput("summexample")
                      )
          )
     })

output$tb <- renderUI({
  if(is.null(input$file)){return()}
  else tabsetPanel(
       tabPanel("Statistical summary ",
                verbatimTextOutput("summ")
                )
     )
  })

info <- eventReactive(input$choice, {
       # Changes in read.table
       f <- read.table(file=input$file$datapath[input$file$name==input$Select],
                       sep=input$sep,
                       header = input$header,
```

```
                    stringsAsFactors = input$stringAsFactors)
      vars <- names(f)
      # Update select input immediately after clicking on the action button.
      updateSelectInput(session,
                        "column1",
                        "Select id document",
                        choices = vars)
      updateSelectInput(session,
                        "column2",
                        "Select document vector",
                        choices = vars)
      updateSelectInput(session,
                        "column3",
                        "Select publish year",
                        choices = vars)
      f
  })

   output$table_display1 <- renderTable({
      f <- info()
      f <- subset(f,
                  select = "input$column1",
                  drop = TRUE) #subsetting takes place here
  })

  output$table_display2 <- renderTable({
      f <- info()
      g <- subset(f,
                  select = "input$column2",
                  drop = TRUE) #subsetting takes place here
       })

observeEvent(input$checkStemming, {
     if(input$checkStemming == FALSE){
      shinyjs::disable("Stemm")}
 else {shinyjs::enable("Stemm")}
 })

observe({
 if (isTRUE(input$example == TRUE )) {
   shinyjs::disable("file")
 }
 else {shinyjs::enable("file")
 }
})

observe({
 if (!is.null(input$file)) {
   shinyjs::disable("example")
 }
 else {shinyjs::enable("example")
 }
})

observe({
 if (is.null(input$file)& input$example == FALSE) {
   shinyjs::disable("dtm.update")
 }
 else {shinyjs::enable("dtm.update")
 }
})


 z <- reactiveValues(odtm=NULL,
                     dtmt = NULL,
                     tf_mat = NULL,
                     dimen = NULL,
                     dtmF =NULL,
                     freq=NULL,
```

```
                wf=NULL,
                year=NULL,
                endtime=NULL)


 observeEvent(input$dtm.update, {
  if( input$example == TRUE){
    dataexample <- system.file("extdata",
                     "scopusJOSS.csv",
                     package = "LDAShiny")
    data_example <- read.csv(dataexample)
    filtro <-  data.frame(doc_names = data_example$Title,
                  doc_vec = data_example$Abstract,
                  year = data_example$Year)
     print(dataexample)
     }
   else {filtro <- tibble::tibble(read.table(file=input$file$datapath[input$file$name==input$Select],
                       sep=input$sep,
                       header = input$header,
                       stringsAsFactors = input$stringAsFactors))
      filtro <- dplyr::select(filtro,
                    doc_names=input$column1,
                    doc_vec=input$column2,
                    year=input$column3)
     }

    z$year <- filtro$year
    stp <- unlist(strsplit(input$stopwords,","))
    stp <- trimws(stp)

    if(input$example == TRUE){
      cpus <- 4}
    else {
    cpus <- parallel::detectCores()
    }
    ngram <- as.integer(input$ngrams)
    Stemm <- trimws(input$Stemm)
     odtm <- textmineR::CreateDtm(doc_vec = filtro$doc_vec,
                doc_names = filtro$doc_names,
                ngram_window = c(1,ngram),
                lower = FALSE,
                remove_punctuation = FALSE,
                remove_numbers = FALSE,
                #stem_lemma_function = function(x) SnowballC::wordStem(x, Stemm), ## primero se debe decidir si se
hace o no stemming y si se hace debe seleccionarse el idioma
                cpus = cpus)

    if(input$checkStemming)
    {
    dtm <- textmineR::CreateDtm(doc_vec = filtro$doc_vec,
                doc_names = filtro$doc_names,
                ngram_window = c(1,ngram),
                stopword_vec = c(stopwords::stopwords(input$Language),
                          letters,stp),
                lower = TRUE,
                remove_punctuation = TRUE,
                remove_numbers = input$removenumber,
                stem_lemma_function = function(x) SnowballC::wordStem(x, Stemm),
                cpus = cpus)
    } else
    {dtm <- textmineR::CreateDtm(doc_vec = filtro$doc_vec,
                 doc_names = filtro$doc_names,
                 lower = TRUE,
                 stopword_vec = c(stopwords::stopwords(input$Language),letters,stp),# Seleccionar el lenguaje
                 ngram_window = c(1,ngram),
                 remove_punctuation = TRUE,
                 remove_numbers = input$removenumber,
                 #stem_lemma_function = function(x) SnowballC::wordStem(x, Stemm), ## primero se debe decidir si se
hace o no stemming y si se hace debe seleccionarse el idioma
```

```
                cpus = cpus)    }

    z$dtm <- quanteda::as.dfm(dtm)
    CONVERT <- quanteda::convert(z$dtm,
                  to = "topicmodels")
    z$dtmt <- removeSparseTerms(CONVERT,
                        sparse= input$sparce)
    z$dtmF <-  chinese.misc::m3m(z$dtmt,
               to="dgCMatrix")
    Original <- dim(odtm)
    Without_Sparsity <- dim(z$dtm)
    Final  <- dim (z$dtmt)
    z$dimen <- rbind(Original,
                Final)
    colnames (z$dimen) <- c ("document", "term")
    z$tf_mat <- textmineR::TermDocFreq(dtm = z$dtmF)
    z$freq <- colSums(as.matrix(z$dtmF))  #
    z$wf <- tibble::tibble(word=names(z$freq), freq=z$freq)
   beepr::beep(2)
  })


  output$Table_dim <- DT::renderDT({
   DT::datatable(data = as.matrix(z$dimen),
            options = list(pageLength = 5,
                        searching = FALSE,
                         rownames = TRUE))
  })


  output$data_b <- DT::renderDT({
    DT::datatable(data = z$tf_mat, extensions = 'Buttons',
              options = list(dom = 'Bfrtip',
                          buttons = c('pageLength', 'copy', 'csv', 'excel', 'pdf', 'print'),
                          pagelength = 10,
                          lengthMenu = list(c(10, 25, 100, -1),
                                       c('10', '25', '100','All')
                                       )
                          )
              )
    })

output$plot_gg <-  highcharter::renderHighchart({
  export
    z$wf %>% top_n(input$b, freq) %>%
    hchart("column",
          hcaes(x = word, y = freq),
          color = "lightgray",
          borderColor = "black") %>%
     hc_add_theme(hc_theme_ggplot2()) %>%
     hc_xAxis(title = list(text = "Term")) %>%
     hc_yAxis(title = list(text = "Frequency")) %>%
     hc_exporting(
    enabled = TRUE,
     formAttributes = list(target = "_blank"),
     buttons = list(contextButton = list(
       text = "Export",
       theme = list(fill = "transparent"),
      menuItems = export)
      )
    )
    })

 output$plot_gf<-  highcharter::renderHighchart({
    export
    z$wf %>% top_n(input$c, freq) %>%
          hchart( "wordcloud",
                hcaes(name = word, weight = freq)) %>%
     hc_exporting(
```

```
        enabled = TRUE,
        formAttributes = list(target = "_blank"),
        buttons = list(contextButton = list(
          text = "Export",
         theme = list(fill = "transparent"),
          menuItems = export)))

  })

################################number topic################
observe({
  if (isTRUE(input$num1>=input$num2||input$num3>input$num2)) {
    shinyjs::disable("Run.model1")
     }
  else if(isTRUE(input$num5>=input$num4)) {
    shinyjs::disable("Run.model1")
     }
  else if(isTRUE(input$num1 < 2 )) {
    shinyjs::disable("Run.model1")
  }
  else if(is.na(as.numeric(input$num1))) {
    shinyjs::disable("Run.model1")
  }
  else if(is.na(as.numeric(input$num2))) {
    shinyjs::disable("Run.model1")
  }
  else if(is.na(as.numeric(input$num3))) {
    shinyjs::disable("Run.model1")
  }
  else if(is.na(as.numeric(input$num4))) {
    shinyjs::disable("Run.model1")
  }
  else if(is.na(as.numeric(input$num5))) {
    shinyjs::disable("Run.model1")
  }
  else if(is.na(as.numeric(input$num6))) {
    shinyjs::disable("Run.model1")
  }
   else if(is.null(z$dtmF)) {
     shinyjs::disable("Run.model1")
      }
  else {shinyjs::enable("Run.model1")
    }
  })
 output$OthKcoh <- renderText({
      stpCohe <- unlist(strsplit(input$OtherKCoherence,","))
      stpCohe <- as.numeric(trimws(stpCohe))
      if (anyNA(stpCohe)) {
        "Invalid input"
      }
  })

   alist <- reactiveValues(coherence_mat=NULL,
                           end_time=NULL)

   observeEvent(input$Run.model1,{
     set.seed(1234)
    ptm <- proc.time()
    stpCohe <- unlist(strsplit(input$OtherKCoherence,","))
    stpCohe <- as.numeric(trimws(stpCohe))
    seqk <- c(seq(from=input$num1,to=input$num2,by=input$num3),stpCohe)# Candidate number of topics k
    iterations <- input$num4     # Parameters control Gibbs sampling
    burnin <- input$num5  # Parameters control Gibbs sampling
    alpha <- input$num6 # Parameters control

    if(input$example == TRUE){
      cores <- 4}
    else {
      cores <- parallel::detectCores()
```

```
    }


    dtm <- z$dtmF
    coherence_list <- textmineR::TmParallelApply(X = seqk , FUN = function(k){
      m <- textmineR::FitLdaModel(dtm= dtm ,
                    k = k,
                    iterations =iterations ,
                    burnin = burnin,
                    alpha = alpha,
                    beta = colSums(dtm) / sum(dtm) * 100,
                    optimize_alpha = TRUE,
                    calc_likelihood = TRUE,
                    calc_coherence = TRUE,
                    calc_r2 = FALSE,
                    cpus = cores)
        m$k <- k

      m
    },export= ls(), # c("nih_sample_dtm"), # export only needed for Windows machines
    cpus = cores)

    alist$coherence_mat <- tibble::tibble(k = sapply(coherence_list, function(x) nrow(x$phi)),
                            coherence = sapply(coherence_list, function(x) mean(x$coherence)),
                            stringsAsFactors = FALSE)
    beepr::beep(2)
    alist$end_time <- proc.time() - ptm

})

output$timeCoherence <- renderPrint({
   print( alist$end_time)
  })

output$plot_gi <-  highcharter::renderHighchart({
    export
    alist$coherence_mat %>%
      hchart("line", hcaes(x = k, y = coherence)) %>%
      hc_add_theme(hc_theme_ggplot2())%>%
      hc_exporting(
        enabled = TRUE,
        formAttributes = list(target = "_blank"),
        buttons = list(contextButton = list(
          text = "Export",
          theme = list(fill = "transparent"),
          menuItems = export)))
  })

  observe({
    if (isTRUE(input$num7>=input$num8||input$num9>input$num8)) {
      shinyjs::disable("Run.model2")
    }

    else if(is.na(as.numeric(input$num7))) {
      shinyjs::disable("Run.model2")
    }
    else if(isTRUE(input$num7 < 2)) {
      shinyjs::disable("Run.model2")
    }
    else if(is.na(as.numeric(input$num8))) {
      shinyjs::disable("Run.model2")
    }
    else if(is.na(as.numeric(input$num9))) {
      shinyjs::disable("Run.model2")
    }
    else if(is.null(z$dtmt)) {
      shinyjs::disable("Run.model2")
     }
      else {
```

```
    shinyjs::enable("Run.model2")
  }
})

output$OthK4metric <- renderText({
 stpCohe <- unlist(strsplit(input$OtherK4metric,","))
 stpCohe <- as.numeric(trimws(stpCohe))
 if (anyNA(stpCohe)) {
  "Invalid input"
 }
})

blist <- reactiveValues(fourmetric_mat = NULL,
                        end_time2 = NULL)

 observeEvent(input$Run.model2, {

  set.seed(1234)
  ptm2 <- proc.time()
  #stp2 = unlist(strsplit(input$metric,","))
  #stp2 = trimws(stp2)
  method <- input$methods
  stpfourm <- unlist(strsplit(input$OtherK4metric,","))
  stpfourm <- as.numeric (trimws(stpfourm))
  seqk <- c(seq(from = input$num7,
                to = input$num8,
                by = input$num9),stpfourm)

  if(input$example == TRUE){
    cl <- makeCluster(4,
                      setup_strategy = "sequential")}
  else {
    cl <- makeCluster(parallel::detectCores(),
                      setup_strategy = "sequential")
  }

  fourmetric_mat <- ldatuning::FindTopicsNumber(
    z$dtmt,
    topics = seqk, # Select range number of topics
    metrics = c("Griffiths2004",
                "CaoJuan2009",
                "Arun2010",
                "Deveaud2014"),
    method = method,
    control = list(seed = 77),
    mc.cores = cl )
  blist$fourmetric_mat <- g4metric(fourmetric_mat)
   beepr::beep(2)
  blist$end_time2 <- proc.time() - ptm2
  stopCluster(cl)
  })

output$timefourmetric <- renderPrint({
  print(blist$end_time2)
})
output$plot_gj <-  highcharter::renderHighchart({
 export
 blist$fourmetric_mat %>%
    hchart("line", hcaes(x = topics, y = value, group =variable)) %>%
    hc_add_theme(hc_theme_ggplot2())%>%
    hc_exporting(
      enabled = TRUE,
      formAttributes = list(target = "_blank"),
      buttons = list(contextButton = list(
        text = "Export",
        theme = list(fill = "transparent"),
        menuItems = export)))
  })
```

```
observe({
  if (isTRUE(input$num13>=input$num14 | | input$num15>input$num14)) {
    shinyjs::disable("Run.model3")
  }
  else if(isTRUE(input$num17 > input$num16 | |input$num18 > input$num17)) {
    shinyjs::disable("Run.model3")
  }

 else if(isTRUE(input$num13 < 2 )) {
  shinyjs::disable("Run.model3")
 }
 else if(is.na(as.numeric(input$num13))) {
  shinyjs::disable("Run.model3")
 }
 else if(is.na(as.numeric(input$num14))) {
  shinyjs::disable("Run.model3")
 }
 else if(is.na(as.numeric(input$num15))) {
  shinyjs::disable("Run.model3")
 }
 else if(is.na(as.numeric(input$num16))) {
  shinyjs::disable("Run.model3")
 }
 else if(is.na(as.numeric(input$num17))) {
  shinyjs::disable("Run.model3")
 }

 else if(is.na(as.numeric(input$num17))) {
  shinyjs::disable("Run.model3")
 }
  else if(is.null(z$dtmt)) {
    shinyjs::disable("Run.model3")
  }
  else {
    shinyjs::enable("Run.model3")
  }
})

output$OthKLL <- renderText({
 stpCohe <- unlist(strsplit(input$OtherKLL,","))
 stpCohe <- as.numeric(trimws(stpCohe))
 if (anyNA(stpCohe)) {
   "Invalid input"
 }
})

clist <- reactiveValues(best.model = NULL,
                        end_time3 = NULL)

observeEvent(input$Run.model3,{

   set.seed(12345)
 ptm3 <- proc.time()
 stpLL <- unlist(strsplit(input$OtherKLL,","))
stpLL <- as.numeric (trimws(stpLL))
 seqk <- c(seq(from = input$num13,
               to = input$num14,
               by = input$num15),stpLL)
 iter <- input$num16
 burnin <- input$num17
 thin <- input$num18
# best.model <- lapply(seqk, function(k){LDA(z$dtmt, k, method = "Gibbs",iter =iter,burnin=burnin,thin=thin)})
  #best.model<- tibble(as.matrix(lapply(best.model, logLik)))
 #clist$best.model <- tibble(topics=seqk, logL=as.numeric(as.matrix(best.model)))

 perplex <- seqk %>%
   purrr::map(topicmodels::LDA, x =z$dtmt ,
              newdata = z$dtmt ,
```

```
                estimate_theta=FALSE,
                iter =iter,
                burnin=burnin,
                thin= thin)
    clist$best.model <- tibble::tibble(Topics = seqk,
                  Perplexity = map_dbl(perplex , perplexity))
    beepr::beep(2)
    clist$end_time3 <- proc.time() - ptm3
      })

  output$timeloglike <- renderPrint({
    print(clist$end_time3)
  })

   output$plot_gk <-  highcharter::renderHighchart({
    export
    Perplex <- clist$best.model
    Perplex %>%
      hchart("line", hcaes(x='Topics, y = Perplexity)) %>%
      hc_add_theme(hc_theme_ggplot2())%>%
      hc_exporting(
        enabled = TRUE,
        formAttributes = list(target = "_blank"),
        buttons = list(contextButton = list(
          text = "Export",
          theme = list(fill = "transparent"),
          menuItems = export)))
  })

 ################################
observe({
  if (isTRUE(input$num19>=input$num20 || input$num21>input$num20)) {
    shinyjs::disable("Run.model4")
      }
  else if(isTRUE(input$num23>input$num22 || input$num24>=input$num23)) {
    shinyjs::disable("Run.model4")
  }

  else if(isTRUE(input$num19 < 2 )) {
    shinyjs::disable("Run.model4")
  }
  else if(is.na(as.numeric(input$num19))) {
    shinyjs::disable("Run.model4")
  }
  else if(is.na(as.numeric(input$num20))) {
    shinyjs::disable("Run.model4")
  }
  else if(is.na(as.numeric(input$num21))) {
    shinyjs::disable("Run.model4")
  }
  else if(is.na(as.numeric(input$num22))) {
    shinyjs::disable("Run.model4")
  }
  else if(is.na(as.numeric(input$num23))) {
    shinyjs::disable("Run.model4")
  }
  else if(is.na(as.numeric(input$num24))) {
    shinyjs::disable("Run.model4")
  }
  else if(is.null(z$dtmt)) {
    shinyjs::disable("Run.model4")
      }
  else {
    shinyjs::enable("Run.model4")
      }
   })

    output$Okhm <- renderText({
      stpCohe <- unlist(strsplit(input$OtherKHM,","))
```

```
    stpCohe <- as.numeric(trimws(stpCohe))
    if (anyNA(stpCohe)) {
      "Invalid input"
    }
  })

dlist <- reactiveValues(hm_many = NULL)
  observeEvent(input$Run.model4,{
set.seed(12345)
    ptm4 <- proc.time()
    stpHM <- unlist(strsplit(input$OtherKHM,","))
    stpHM <- as.numeric (trimws(stpHM))
    seqk <- c(seq(from = input$num19,
                  to = input$num20,
                  by = input$num21),stpHM)
    iter <- input$num22
    burnin <- input$num23
    keep <- input$num24
    fitted_many <- lapply(seqk,
                          function(k)LDA(z$dtmt,
                                         k = k,
                                         method = "Gibbs",
                                         control = list(burnin = burnin,
                                                        iter = iter,
                                                        keep = keep)
                                         ))
    # extract logliks from each topic
    logLiks_many <- lapply(fitted_many, function(L)L@logLiks[-c(1:(burnin/keep))])
    # compute harmonic means
    hm_many <- tibble::tibble(as.matrix (sapply(logLiks_many,
                                        function(h) harmonicMean(h)
                                        )
                                  )
                          )
    # inspect
    dlist$hm_many <- tibble::tibble(topics=seqk,
                                    logL=as.numeric(as.matrix(hm_many)
                                                    )
                                    )
    beepr::beep(2)
    dlist$end_time4 <- proc.time() - ptm4
  })

  output$timeHmean<- renderPrint({
   print(dlist$end_time4)
  })

  output$plot_gl <-  highcharter::renderHighchart({
    export
    dlist$hm_many %>%
      hchart("line", hcaes(x=topics, y=logL)) %>%
      hc_add_theme(hc_theme_ggplot2())%>%
      hc_exporting(
        enabled = TRUE,
        formAttributes = list(target = "_blank"),
        buttons = list(contextButton = list(
          text = "Export",
          theme = list(fill = "transparent"),
          menuItems = export)))
  })

################################
  observe({
    if (isTRUE(input$num27 >= input$num26|| input$num25 < 2)) {
      shinyjs::disable("Run.model5")
    }
    else if(is.null(z$dtmF)) {
      shinyjs::disable("Run.model5")
    }
```

```
  else if(is.na(as.numeric(input$num25))) {
    shinyjs::disable("Run.model5")
  }
  else if(is.na(as.numeric(input$num26))) {
    shinyjs::disable("Run.model5")
  }
  else if(is.na(as.numeric(input$num27))) {
    shinyjs::disable("Run.model5")
  }
  else if(is.na(as.numeric(input$num28))) {
    shinyjs::disable("Run.model5")
  }
      else {
        shinyjs::enable("Run.model5")
   }
})
elist <- reactiveValues(summary= NULL,
                        tidy_thetha=NULL,
                        tidy_beta = NULL,
                        dfCoef=NULL,
                        model=NULL)

observeEvent(input$Run.model5,{
  set.seed(12345)
   k <- input$num25
  iter <- input$num26
  burnin <- input$num27
  alpha <- input$num28

 if(input$example == TRUE){
   cpus <- 4L}
else {
    cpus <- parallel::detectCores()
}
  elist$model <- textmineR::FitLdaModel(z$dtmF, # parameter
                      k = k ,# Number of topics k
                      iterations = iter, # parameter
                      burnin = burnin, #parameter
                      alpha = alpha,# parameter
                      beta = colSums(z$dtmF)/sum(z$dtmF)*100,
                      optimize_alpha = TRUE, # parameter
                      calc_likelihood = TRUE,
                      calc_coherence = TRUE,
                      calc_r2 = FALSE,
                      cpus = cpus)

top_terms <- GetTopTerms(phi = elist$model$phi,
                                M = 10)
prevalence <- colSums(elist$model$theta) / sum(elist$model$theta) * 100
 #textmineR has a naive topic labeling tool based on probable bigrams
 labels <- LabelTopics(assignments = elist$model$theta > 0.05,
                       dtm = z$dtmF,
                        M = input$Labels)

 elist$summary <- data.frame(topic = rownames(elist$model$phi),
                       label = labels,
                        coherence = round(elist$model$coherence, 3),
                        prevalence = round(prevalence,3),
                       top_terms = apply(top_terms, 2, function(x){
                         paste(x, collapse = ", ")
                        }),
                       stringsAsFactors = FALSE)

 elist$tidy_thetha <- data.frame(document = rownames(elist$model$theta),
                      round(elist$model$theta,5),
                      stringsAsFactors = FALSE) %>%
  tidyr::gather(topic, gamma, -document)
  elist$tidy_beta <- data.frame(topic = as.integer(str_replace_all(rownames(elist$model$phi),
```

```
                                                    "t_", "")
                                    ),
                          round(elist$model$phi,5),
                          stringsAsFactors = FALSE)%>%
      tidyr::gather(term, beta, -topic)

  elist$thetayear <- data.frame(elist$tidy_thetha,
                          year = rep(z$year))%>%
    group_by(topic,year) %>%
    summarise(proportion= mean(gamma))

  elist$dfreg  <- elist$thetayear %>% group_by(topic) %>%
   do(fitreg = lm(proportion ~ year, data = .))
  elist$thetayear <- data.frame(elist$thetayear)

  elist$dfCoef <- elist$thetayear %>%
   nest_by(topic) %>%
     #change do() to mutate(), then add list() before your model
     # make sure to change data = .  to data = data
    mutate(fitmodelreg = list(lm(proportion ~ year, data = data))) %>%
    summarise(tidy(fitmodelreg))

 classifications <- elist$tidy_thetha %>%
  dplyr::group_by(topic, document) %>%
   dplyr::top_n(1, gamma) %>%
   ungroup()
 beepr::beep(2)


    })

    output$sum <- DT::renderDT({
      DT::datatable(data = elist$summary, extensions = 'Buttons',
                options = list(dom = 'Bfrtip',
                               buttons = c('pageLength',
                                           'copy',
                                           'csv',
                                           'excel',
                                           'pdf',
                                           'print'),
                               pagelength = 10,
                               lengthMenu = list(c(10, 25, 100, -1),
                                                 c('10', '25', '100','All')
                                                 )
                               )
                )
      })

  output$summLDA <- DT::renderDT({
     model <- elist$model
      top_terms <- textmineR::GetTopTerms(phi = model$phi, M = input$Topterm)
      prevalence <- colSums(model$theta) / sum(model$theta) * 100
      # textmineR has a naive topic labeling tool based on probable bigrams
      labels <- LabelTopics(assignments = model$theta > input$assignments,
                            dtm = z$dtmF,
                            M = input$Labels)
      summary <- data.frame(topic = rownames(model$phi),
                               label = labels,
                               coherence = round(model$coherence, 3),
                               prevalence = round(prevalence,3),
                               top_terms = apply(top_terms, 2, function(x){
                                 paste(x, collapse = ", ")
                               }),
                               stringsAsFactors = FALSE)


      DT::datatable(data = summary, extensions = 'Buttons',
                options = list(dom = 'Bfrtip',
                               buttons = c('pageLength',
```

```
                                    'copy',
                                    'csv',
                                    'excel',
                                    'pdf',
                                    'print'),
                            pagelength = 10,
                            lengthMenu = list(c(10, 25, 100, -1),
                                        c('10', '25', '100','All'))))%>%
        formatRound( columns= c("coherence","prevalence"),
                     digits=5)
})


output$theta <- DT::renderDT({
     DT::datatable(data = elist$tidy_thetha ,
                 extensions = 'Buttons',
                 filter = 'top',
                 colnames=c("document","topic", "theta"),
                 options = list(dom = 'Bfrtip',
                                buttons = c('pageLength',
                                    'copy',
                                    'csv',
                                    'excel',
                                    'pdf',
                                    'print'),
                                pagelength = 10,
                                lengthMenu = list(c(10,100,20000,-1),
                                    c('10', '25', '10000','All')
                                    )
                            )
                 )%>% formatRound( columns= c("gamma"),digits=5)

  })

output$downloadData <- downloadHandler(
    filename = function() {
      paste('data-', Sys.Date(), '.csv', sep='')
    },
    content = function(con) {
      write.csv(elist$tidy_thetha, con)
    }
  )




  output$phi <- DT::renderDT({
    DT::datatable(data =elist$tidy_beta, extensions = 'Buttons',filter = 'top',
                 colnames=c("topic", "term", "phi"),
                 options = list(dom = 'Bfrtip',
                                buttons = c('pageLength',
                                    'copy', 'csv',
                                    'excel', 'pdf',
                                    'print'),
                                pagelength = 10,
                                lengthMenu = list(c(10, 25, 100, -1),
                                        c('10', '25', '100','All'))))%>%
     formatRound( columns= c("beta"),digits=5)

  })

output$Alloca <- DT::renderDT({
     classifications <- elist$tidy_thetha %>%
      dplyr::group_by(topic) %>%
      dplyr::top_n(input$topnumber, gamma) %>%
      ungroup()
    DT::datatable(data = classifications,
                 extensions = 'Buttons',
                 filter = 'top',
```

```
            colnames=c("document","topic", "theta"),
            options = list(dom = 'Bfrtip',
                           buttons = c('pageLength',
                                       'copy',
                                       'csv',
                                       'excel',
                                       'pdf',
                                       'print'),
                           pagelength = 10,
                           lengthMenu = list(c(10, 25, 100, -1),
                                             c('10', '25', '100','All'))))%>%
    formatRound( columns= c("gamma"),digits=5)
  })

output$reg <- DT::renderDT({
   datareg <- elist$dfCoef
   DT::datatable(data = datareg,
            extensions = 'Buttons',
            options = list(dom = 'Bfrtip',
                           buttons = c('pageLength',
                                       'copy',
                                       'csv',
                                       'excel',
                                       'pdf',
                                       'print'),
                           pagelength = 10,
                           lengthMenu = list(c(10, 25, 100, -1),
                                             c('10',
                                               '25', '100','All'))))%>%
     formatRound( columns= c("estimate",
                             "std.error",
                             "statistic",
                             "p.value"),digits=5)
    })

  output$plot_trend <-  highcharter::renderHighchart({
   export
   elist$thetayear %>%
    hchart("line", hcaes(x = year,
                         y = proportion,
                         group = as.integer(str_replace_all(topic,"t_", " ")))) %>%
    hc_add_theme(hc_theme_ggplot2())%>%
    hc_exporting(
     enabled = TRUE,
     formAttributes = list(target = "_blank"),
     buttons = list(contextButton = list(
      text = "Export",
      theme = list(fill = "transparent"),
      menuItems = export)))
  })

  output$plot_worcloud <-  highcharter::renderHighchart({
   export
   elist$tidy_beta %>% dplyr::filter(topic==input$num29)%>%
    dplyr::top_n(input$cloud, beta) %>%
    hchart( "wordcloud", hcaes(name = term,
                               weight = beta)) %>%
    hc_exporting(
     enabled = TRUE,
     formAttributes = list(target = "_blank"),
     buttons = list(contextButton = list(
      text = "Export",
      theme = list(fill = "transparent"),
      menuItems = export)))

  })

  output$plot_heatmap <-  highcharter::renderHighchart({
    colr <- list( list(0, '#2E86C1'),
```

```
                list(1, '#FF5733'))
  export
  elist$thetayear %>%
    hchart("heatmap", hcaes(x = year,
                            y = as.integer(str_replace_all(topic,"t_", " ")) ,
                            value =proportion)) %>%
    hc_colorAxis(  stops= colr,
                   min=min(elist$thetayear$proportion),
                   max= max(elist$thetayear$proportion)) %>%
    hc_yAxis(title = list(text = "Topic"))%>%
    hc_exporting(
      enabled = TRUE,
      formAttributes = list(target = "_blank"),
      buttons = list(contextButton = list(
        text = "Export",
        theme = list(fill = "transparent"),
        menuItems = export)))
 })

########################################end number topic
 #observe({
  #  if (input$Stop > 0) stopApp()  # stop shiny
  #})
  })
```

Final server.R

ANEXO 4. Publicaciones con aplicación y sobre GWPCA, entre los años 2010 y 2022

Autores	Título	Año	Revista	DOI	Tipo de documento
Lloyd CD	Analysing Population Characteristics Using Geographically Weighted Principal Components Analysis: A Case Study of Northern Ireland in 2001	2010	Comput. Environ. Urban Syst.	10.1016/j.compenvurbsys.2010.02.005	Artículo
Harris P;Brunsdon C;Charlton M	Geographically Weighted Principal Components Analysis	2011	Int. J. Geogr. Inf. Sci.	10.1080/13658816.2011.554838	Artículo
Kumar S;Lal R;Lloyd CD	Assessing Spatial Variability In Soil Characteristics With Geographically Weighted Principal Components Analysis	2012	Comput. Geosci.	10.1007/s10596-012-9290-6	Artículo
Faraji Sabokbar H;Shadman Roodposhti M;Tazik E	Landslide Susceptibility Mapping Using Geographically-Weighted Principal Component Analysis	2014	Geomorphology	10.1016/j.geomorph.2014.07.026	Artículo
Harris P;Howden Njk;Peukert S;Noacco V;Ramezani K;Tuominen E;Eludoyin B;Brazier R;Shepherd A;Griffith B;Orr R;Murray P	Contextualized Geographically Weighted Principal Components Analysis For Investigating Baseline Soils Data On The North Wyke Farm Platform	2014	Proc. Int. Assoc. Math. Geosci. - Geostat. Geospatial Approaches Charact. Nat. Resour. Environ.: Challenges, Process. Strateg., Iamg	ND	Conference Paper
Wang H;Cheng Q;Zuo R	Quantifying The Spatial Characteristics of Geochemical Patterns Via Gis-Based Geographically Weighted Statistics	2015	J. Geochem. Explor.	10.1016/j.gexplo.2015.06.004	Artículo
Saib M-S;Caudeville J;Beauchamp M;Carré F;Ganry O;Trugeon A;Cicolella A	Building Spatial Composite Indicators To Analyze Environmental Health Inequalities On A Regional Scale	2015	Environ. Health Global Access Sci. Sour.	10.1186/s12940-015-0054-3	Artículo
Harris P;Clarke A;Juggins S;Brunsdon C;Charlton M	Enhancements To A Geographically Weighted Principal Component Analysis In The Context of An Application To An Environmental Data Set	2015	Geogr. Anal.	10.1111/gean.12048	Artículo
Gollini I;Lu B;Charlton M;Brunsdon C;Harris P	Gwmodel: An R Package For Exploring Spatial Heterogeneity Using Geographically Weighted Models	2015	J. Stat. Software	10.18637/jss.v063.i17	Artículo
Wei C;Cabrera-Barona P;Blaschke T	Local Geographic Variation of Public Services Inequality: Does The Neighborhood Scale Matter?	2016	Int. J. Environ. Res. Public Health	10.3390/ijerph13100981	Artículo
Li Z;Cheng J;Wu Q	Analyzing Regional Economic Development Patterns In A Fast Developing Province of China Through Geographically Weighted Principal Component Analysis	2016	Lett. Spat. Resour. Sci.	10.1007/s12076-015-0154-2	Artículo
Comber Aj;Harris P;Tsutsumida N	Improving Land Cover Classification Using Input Variables Derived From A Geographically Weighted Principal Components Analysis	2016	Isprs J. Photogramm. Remote Sens.	10.1016/j.isprsjprs.2016.06.014	Artículo
Tsutsumida N;Harris P;Comber A	The Application of A Geographically Weighted Principal Component Analysis For Exploring Twenty-Three	2017	Ann. Am. Assoc. Geogr.	10.1080/24694452.2017.1309968	Artículo

Autores	Título	Año	Revista	DOI	Tipo de documento
	Years OfGoat Population Change Across Mongolia				
Roca-Pardiñas J;Ordóñez C;Cotos-Yáñez Tr;Pérez-Álvarez R	Testing Spatial Heterogeneity In Geographically Weighted Principal Components Analysis	2017	Int. J. Geogr. Inf. Sci.	10.1080/13658816.2016.1224886	Artículo
Wu C;Ye X;Ren F;Du Q	Modified Data-Driven Framework For Housing Market Segmentation	2018	J. Urban Plann. Dev.	10.1061/(ASCE)UP.1943-5444.0000473	Artículo
Kallio M;Guillaume Jha;Kummu M;Virrantaus K	Spatial Variation In Seasonal Water Poverty Index For Laos: An Application of Geographically Weighted Principal Component Analysis	2018	Soc. Indic. Res.	10.1007/s11205-017-1819-6	Artículo
Mishra Sv	Urban Deprivation In A Global South City-A Neighborhood Scale Study of Kolkata, India	2018	Habitat Int.	10.1016/j.habitatint.2018.08.006	Artículo
Trogu D;Campagna M	Towards Spatial Composite Indicators: A Case Study On Sardinian Landscape	2018	Sustainability	10.3390/su10051369	Artículo
Fernández S;Cotos-Yáñez T;Roca-Pardiñas J;Ordóñez C	Geographically Weighted Principal Components Analysis To Assess Diffuse Pollution Sources of Soil Heavy Metal: Application To Rough Mountain Areas In Northwest Spain	2018	Geoderma	10.1016/j.geoderma.2016.10.012	Artículo
Kunah Om;Pakhomov Oy;Zymaroieva Aa;Demchuk Ni;Skupskyi Rm;Bezuhla Ls;Vladyka Yp	Agroeconomic And Agroecological Aspects of Spatial Variation of Rye (Secale Cereale) Yields Within Polesia And The Forest-Steppe Zone of Ukraine: The Usage ff Geographically Weighted Principal Components Analysis	2018	Biosys. Diver.	10.15421/011842	Artículo
Harris P;Shao M;Lü Y;Wu L;Comber A	Alternative Interpretations of The PArtícle Size Distribution of Soils (0 -500cm) In The Loess Plateau, China	2018	Proc. Spat. Accuracy	ND	Conference Paper
Sepúlveda Murillo Fh;Chica Olmo J;Soto Builes Nm	Spatial Variability Analysis of Quality of Life And Its Determinants: A Case Study of Medellín, Colombia	2019	Soc. Indic. Res.	10.1007/s11205-019-02088-x	Artículo
Wu C;Hu W;Zhou M;Li S;Jia Y	Data-Driven Regionalization For Analyzing The Spatiotemporal Characteristics of Air Quality In China	2019	Atmos. Environ.	10.1016/j.atmosenv.2019.01.048	Artículo
Losada N;Alén E;Cotos-Yáñez Tr;Domínguez T	Spatial Heterogeneity In Spain For Senior Travel Behavior	2019	Tour. Manage.	10.1016/j.tourman.2018.09.011	Artículo
Zymaroieva A	The Perspectives of The Application of Geographically Weighted Principal Components Analysis For Estimation of Maize Yields Spatial Variability [Перспективи Використання Географічно Зваженого Аналізу Головних Компонент Для Оцінки Просторової Варіабельності Врожайності Кукурудзи]	2019	Sci. Horiz.	10.33249/2663-2144-2019-83-10-20-27	Artículo
Zymaroieva A;Zhukov O;Fedonyuk T;Pinkin A	Application of Geographically Weighted Principal Components Analysis Based On Soybean Yield Spatial Variation For Agro-Ecological Zoning of The Territory	2019	Agron. Res.	10.15159/AR.19.208	Artículo
Wu C;Peng N;Ma X;Li S;Rao J	Assessing Multiscale Visual Appearance Characteristics of Neighbourhoods Using Geographically Weighted Principal Component Analysis In Shenzhen, China	2020	Comput. Environ. Urban Syst.	10.1016/j.compenvurbsys.2020.101547	Artículo

Autores	Título	Año	Revista	DOI	Tipo de documento
Bellino A;Alfani A;De Riso L;Baldantoni D	Multivariate Spatial Analysis For The Identification of Criticalities And of The Subtended Causes In River Ecosystems	2020	Environ. Sci. Pollut. Res.	10.1007/s11356-019-07198-0	Artículo
Das A;Ghosh S;Das K;Basu T;Das M;Dutta I	Modeling The Effect of Area Deprivation On Covid-19 Incidences: A Study of Chennai Megacity, India	2020	Public Health	10.1016/j.puhe.2020.06.011	Artículo
Tejedor-Flores N;Vicente-Galindo P;Galindo-Villardón P	Geographically Weighted Principal Components Analysis Approach To Evaluate Electricity Consumption Behaviour	2020	Wit Trans. Ecol. Environ.	10.2495/SC200091	Artículo
Basu T;Das A;Pal S	Application of Geographically Weighted Principal Component Analysis And Fuzzy Approach For Unsupervised Landslide Susceptibility Mapping On Gish River Basin, India	2020	Geocarto Int.	10.1080/10106049.2020.1778105	Artículo
Liu Y;Yang S;Wang S	Heterogeneity Study of The Visual Features Based On Geographically Weighted Principal Components Analysis Applied To An Urban Community	2021	Sustainability	10.3390/su132313488	Artículo
Tomao A;Mattioli W;Fanfani D;Ferrara C;Quaranta G;Salvia R;Salvati L	Economic Downturns And Land-Use Change: A Spatial Analysis of Urban Transformations In Rome (Italy) Using A Geographically Weighted Principal Component Analysis	2021	Sustainability	10.3390/su132011293	Artículo
Fan Py;Chun Kp;Mijic A;Tan Ml;He Q;Yetemen O	Quantifying Land Use Heterogeneity On Drought Conditions For Mitigation Strategies Development In The Dongjiang River Basin, China	2021	Ecol. Indic.	10.1016/j.ecolind.2021.107945	Artículo
Aidoo En;Appiah Sk;Awashie Ge;Boateng A;Darko G	Geographically Weighted Principal Component Analysis For Characterising The Spatial Heterogeneity And Connectivity of Soil Heavy Metals In Kumasi, Ghana	2021	Heliyon	10.1016/j.heliyon.2021.e08039	Artículo
Benita F;Kalashnikov V;Tunçer B	A Spatial Livability Index For Dense Urban Centers	2021	Environ. Plann.	10.1177/2399808320960151	Artículo
Chen J;Qu M;Zhang J;Xie E;Huang B;Zhao Y	Soil Fertility Quality Assessment Based On Geographically Weighted Principal Component Analysis (Gwpca) In Large-Scale Areas	2021	Catena	10.1016/j.catena.2021.105197	Artículo
Chen J;Qu M;Zhang J;Xie E;Zhao Y;Huang B	Improving The Spatial Prediction Accuracy of Soil Alkaline Hydrolyzable Nitrogen Using Gwpca-Gwrk	2021	Soil Sci. Soc. Am. J.	10.1002/saj2.20189	Artículo
Basu T;Das A	Formulation of Deprivation Index For Identification of Regional Pattern of Deprivation In Rural India	2021	Socio-Econ. Plann. Sci.	10.1016/j.seps.2020.100924	Artículo
Das A;Ghosh S;Das K;Basu T;Dutta I;Das M	Living Environment Matters: Unravelling The Spatial Clustering of Covid-19 Hotspots In Kolkata Megacity, India	2021	Sustainable Cities Soc.	10.1016/j.scs.2020.102577	Artículo
Cartone A;Panzera D	Deprivation At Local Level: Practical Problems And Policy Implications For The Province of Milan	2021	Reg. Sci. Policy Pract.	10.1111/rsp3.12339	Artículo
Basu T;Das A;Pereira P	Urban Livability Index Assessment Based On Land-Use Changes In An Indian Medium-Sized City (Raiganj)	2021	Geocarto Int.	10.1080/10106049.2021.2002427	Artículo
Zymaroieva A;Zhukov O	Analyzing Cereal And Grain Legumes (Pulses) Yields Patterns In The Forest And Forest-	2021	Acta Agric. Slovenica	10.14720/aas.2020.116.2.873	Artículo

Autores	Título	Año	Revista	DOI	Tipo de documento
	Steppe Zones of Ukraine Using Geographically Weighted Principal Components Analysis [Analiza Vzorcev Pridelkov Žit In Zrnatih Stročnic Na Območju Gozda In Lesostepe Ukrajine Z Geografsko Tehtano Analizo Glavnih Komponent]				
Jin Y;Li A;Bian J;Nan X;Lei G;Muhammad K	Spatiotemporal Analysis of Ecological Vulnerability Along Bangladesh-China-India-Myanmar Economic Corridor Through A Grid Level Prototype Model	2021	Ecol. Indic.	10.1016/j.ecolind.2020.106933	Artículo
Cartone A;Postiglione P	Principal Component Analysis For Geographical Data: The Role of Spatial Effects In The Definition of Composite Indicators	2021	Spat. Econ. Anal.	10.1080/17421772.2020.1775876	Artículo
Ghosh S;Dinda S;Chatterjee Nd;Dutta S;Bera D	Spatial-Explicit Carbon Emission-Sequestration Balance Estimation And Evaluation of Emission Susceptible Zones In An Eastern Himalayan City Using Pressure-Sensitivity-Resilience Framework: An Approach Towards Achieving Low Carbon Cities	2022	J. Clean. Prod.	10.1016/j.jclepro.2022.130417	Artículo
Han J;Kang X;Yang Y;Zhang Y	Geographically And Temporally Weighted Principal Component Analysis: A New Approach For Exploring Air Pollution Non-Stationarity In China, 2015–2019	2022	J. Spat. Sci.	10.1080/14498596.2022.2028270	Artículo
Park G;Xu Z	The Constituent Components And Local Indicator Variables of Social Vulnerability Index	2022	Nat. Hazards	10.1007/s11069-021-04938-9	Artículo

ANEXO 5. Código Fuente GeoWeightedModel

ui.R (inicio)

```
library(beepr)
library(cartography)
library(dplyr)
library(DT)
library(GWmodel)
library(raster)
library(readxl)
library(shiny)
library(shinyalert)
library(shinyBS)
library(shinybusy)
library(shinydashboard)
library(shinyjs)
library(shinyWidgets)
library(sp)
data(USelect)
################################################################################
###########
header <- shinydashboard:: dashboardHeader(
  title = "GeoWeightedModels",
  titleWidth = 320,
  tags$li(class = "dropdown",
          actionLink("stop_radiant",
                     "Stop",
                     icon = icon("power-off"),
                     onclick = "setTimeout(function(){window.close();}, 100); "
                     )
          )
 )
################################################################################
###########
sidebar <- shinydashboard::dashboardSidebar(
  width = 320,
 shinydashboard::sidebarMenu(
    shinydashboard::menuItem("About ",
             tabName = "tab0"
             ),
    shinydashboard::menuItem("Load data",
             tabName = "tab1",
             icon = icon("file-import")
             ),

    shinydashboard::menuItem(div(tags$img(src = "distance.png",
                         width="20px",
                         height="20px"),
                "Distance matrix"),
             tabName = "tab2"
    ),
    shinydashboard::menuItem("Bandwidth selection",
             tabName = "tab3",
             icon = icon("wifi")
    ),
    shinydashboard::menuItem("Spatial autocorrelation",
                          tabName = "tab6",
                          icon = icon("chart-bar")
    ),
    shinydashboard:: menuItem("Geographically Weighted Summary Statistics",
             tabName = "tab4",
             icon = icon("wpexplorer")
    ),
    shinydashboard::menuItem("Models",
```

```
            tabName = "tab5",
            icon = icon("globe-americas"),
            shinydashboard::menuSubItem("GW Regression",
                     tabName = "tab52"
                    # icon = tags$img(src = "normal1.png",
                              # width="30px")
                     ),
            shinydashboard:: menuSubItem("GW Principal Component Analysis",
                     tabName = "tab53"
                     #icon = icon("chart-bar")
                     ),
            shinydashboard::menuSubItem("GW Discriminant Analysis",
                     tabName = "tab54"
                     #icon = icon("chart-bar")
                     )
            )
    )
  )
##############################################################################
###########
body <- shinydashboard::dashboardBody(shinybusy::add_busy_spinner(spin = "pixel",
                                        height = "100px",
                                        width = "100px",
                                        color = "blue",
                                        position ="top-right"),
tags$head(tags$style(
HTML('/* logo when hovered */.skin-blue .main-header .logo:hover {
                         background-color: #7da2d1;
}
* { font-family: Garamond; }
/* navbar (rest of the header) */.skin-blue .main-header .navbar {
background-color: #7da2d1;
}
/* main sidebar */.skin-blue .main-sidebar { background-color:#7da2d1;}
/* active selected tab in the sidebarmenu
*/.skin-blue .main-sidebar .sidebar .sidebar-menu .active a{
background-color: #0091ff;
}
/* other links in the sidebarmenu */
.skin-blue .main-sidebar .sidebar .sidebar-menu a{ background-color: #ccdceb;
color: #000000;}
/* other links in the sidebarmenu when hovered */
.skin-blue .main-sidebar .sidebar .sidebar-menu a:hover{
background-color: #69c3ff;}
/* toggle button when hovered  */
.skin-blue .main-header .navbar .sidebar-toggle:hover{
                         background-color: #9bc2e8;
                         }

                         /* body */
                         .content-wrapper, .right-side {
                         background-color: #7dcdd1;
                         }'
)
                              )
                            ),
  shinydashboard::tabItems(
    shinydashboard::tabItem(
      tabName = "tab0",
      fluidRow(
        shinydashboard::tabBox(id = "tabset1",
              height = 400,
              width = 12,
              tabPanel(
                " ",
                h2(em(strong(
                  "GeoWeightedModels:
                  An R package for Geographically Weighted Models")
                ),
```

```
          align = "center"),
          br(),
          h4("Javier De La Hoz-M, María José Fernández Gómez
            & Susana Mendes",
            align="center"),
          br(),
          div(img(src = "icono.png",
                  height = 400,
                  width = 400),
              style="text-align: center;"),
          br(),
          tags$head(
            tags$style("h4 {font-family:Garamond}")
          ),
          tags$h4("GeoWeightedModels is an application developed in
                  Shiny to carry out some
                  models of a particular branch of spatial statistics,
                  named Geographically Weighted Models.
                  Includes functionsfor Exploratory Spatial Data
                  Analysis,various forms of Geographically
                  Weighted Regression,
                  Geographically Weighted Principal Component Analysis,
                  and Geographically Weighted Discriminant Analysis",
                  align="left"),


        )
    )
  )
  ),
shinydashboard::tabItem(tabName = "tab1",
                  fluidRow(column(width = 12,shinydashboard::box(width = 12,
                    title = "Load data ",
                    status = "primary",
                    solidHeader = TRUE,
                    collapsible = TRUE,
                    sidebarPanel(
                      shinyWidgets::actionBttn("helpload",
                                "Help",
                                icon = icon("question-circle"),
                                style = "stretch",
                                block = FALSE,
                                color = "primary"
                    ),
                    shinyBS::bsModal(id="helploaddata",
                                     title = "",
                                     trigger = "helpload",
                                     size="large",
                                     tags$iframe(
                                       src = "Upload-data.html",
                                       width = "100%",
                                       height = "1000px",
                                       frameborder = 0,
                                       scrolling = "auto"
                                     )
                    ),
                    br(), br(),
                    shinyWidgets::prettyCheckbox(
                      inputId = "example",
                      label = "Use example data set?",
                      value = FALSE,
                      status = "success"
                    ), br(),
                      fileInput("file1",
                                "Upload the file",
                                accept = c(".xlsx", ".xls"),
                                multiple = FALSE),
                      selectInput(inputId = "worksheet",
                                  label="Worksheet Name",
```

```
                        choices=NULL
                    ),actionButton(inputId = "getData",
                                   label="Get data"),
                    selectizeInput('colID',
                                   'Select ID for merge data',
                                   choices = NULL,
                                   multiple = FALSE)
                  ),mainPanel(DT::DTOutput("table1"),
                              DT::DTOutput("example")

                  ))),
                  column(width = 12,  shinydashboard::box(width = 12,
                   title = "Load shapefiles ",
                   status = "primary",
                   solidHeader = TRUE,
                   collapsible = TRUE,
                   sidebarPanel(fileInput("filemap",
                                          "Upload map (shapefile)",
                                          accept=c('.shp',
                                                   '.dbf','.prj',
                                                   '.shx','.xml'),
                                          multiple = TRUE),
                                br(),
                                actionButton(
                                  inputId = "getshape",
                                  label="Get shapefiles")
                                ),
                   mainPanel(verbatimTextOutput("shpi")
                   ))))

                ),

shinydashboard::tabItem(
  tabName = "tab2",
  shinydashboard::box(
    width = 12,
    title = "Distance matrix",
    status = "primary",
    solidHeader = TRUE,
    collapsible = TRUE,
    sidebarPanel(
      shinyWidgets::actionBttn("helpdMat",
                               "Help",
                               icon = icon("question-circle"),
                               style = "stretch",
                               block = FALSE,
                               color = "primary"
      ),
      shinyBS::bsModal(id="helpdistmat",
                       title = "",
                       trigger = "helpdMat",
                       size="large",
                       tags$iframe(
                         src = "Distance_Matrix.html",
                         width = "100%",
                         height = "1000px",
                         frameborder = 0,
                         scrolling = "auto"
                       )
      ),
      numericInput('focus',
                   'Focus',
                   0,
                   min = 1,
                   max = Inf),
      numericInput('power',
                   'Power (Minkowski distance)',
                   2,
                   min = 1,
```

```
              max = Inf),

        sliderInput("theta",
                    "Theta (Angle in radians)",
                    min = 0,
                    max = 2,
                    value = 0,
                    step = 0.05),
      shinyWidgets::switchInput("longlat",
                                inputId = "longlat",
                                onLabel = "TRUE",
                                offLabel = "FALSE",
                                size = "mini")
        ,
      shinyalert::useShinyalert(),
      shinyjs::useShinyjs(),
      shinyWidgets::actionBttn(
                  inputId = "Run1", # run dMat
                  label = "Run",
                  style = "float",
                  block = TRUE,
                  color = "danger"
                )
      ),
     mainPanel(helpText("Distance matrix"),

      DT::DTOutput("distmatrix")))
  ),
shinydashboard::tabItem(
  tabName = "tab3",fluidRow(column(width = 12,
  shinydashboard:: box(width = 12,
      title = "Bandwidth selection",
      status = "primary",
      solidHeader = TRUE,
      collapsible = TRUE,
      sidebarPanel(
        shinyWidgets::actionBttn("helpbw",
                                 "Help",
                                 icon = icon("question-circle"),
                                 style = "stretch",
                                 block = FALSE,
                                 color = "primary"
        ),
        shinyBS::bsModal(id="helpbandsel",
                         title = "",
                         trigger = "helpbw",
                         size="large",
                         tags$iframe(
                           src = "Help_bw.html",
                           width = "100%",
                           height = "1000px",
                           frameborder = 0,
                           scrolling = "auto"
                         )
        ),
        shinyWidgets::pickerInput(
          inputId = "bandSelect",
          label = "Choose type:",
          choices = c("Select type",
                      "bw.gwr",
                      "bw.ggwr",
                      #"bw.gtwr",
                      "bw.gwda",
                      "bw.gwpca"),
          options = list(style = "btn-primary")
        ),
        conditionalPanel(
          condition = "input.bandSelect == 'bw.ggwr'",
          div(style="display: inline-block; vertical-align:top;
```

```
    width: 100px;",
    selectizeInput('dependientggwr',
                   'Dependient',
                   choices = NULL,
                   multiple = TRUE,
                   options = list(maxItems = 1)
    )
  ),
  div(style="display: inline-block;vertical-align:top;
    width: 150px;",
    selectizeInput('independientggwr',
                   'Independient(s)',
                   choices = NULL,
                   multiple = TRUE)
  ),
  div(style="display: inline-block;vertical-align:top;
    width: 150px;",
    shinyWidgets::pickerInput(inputId = "familyggwr",
                label = "Family",
                choices = family
    )),
  div(style="display: inline-block;vertical-align:top;
    width: 100px;",
    shinyWidgets::pickerInput(
      inputId = "approachggwr",
      label = "Approach",
      choices = c("CV","AIC"))),
  div(style="display: inline-block;vertical-align:top;
    width: 100px;",
    shinyWidgets::pickerInput(
      inputId = "kernelggwr",
      label = "Kernel",choices = kernel)),

  div(style="display: inline-block;vertical-align:top;
    width: 200px;",
    numericInput(
      'powerggwr',
      'Power (Minkowski distance)',
      2,
      min = 1,
      max = Inf)),
  div(style="display: inline-block;vertical-align:top;
    width: 200px;",
    sliderInput("thetaggwr",
                "Theta (Angle in radians)",
                min = 0,
                max = 2,
                value = 0,
                step = 0.05 )),
  div(style="display: inline-block;vertical-align:top;
    width: 150px;",
    shinyWidgets::switchInput(
      "longlat",
      inputId = "longlatggwr",
      onLabel = "TRUE",
      offLabel = "FALSE",
      size = "mini")),
  div(style="display: inline-block;vertical-align:top;
    width: 100px;",
    shinyWidgets::switchInput("adaptive",
                inputId = "adaptativeggwr",
                onLabel = "TRUE",
                offLabel = "FALSE",
                size = "mini") )
  ),
conditionalPanel(
  condition = "input.bandSelect == 'bw.gwda'",
  div(style="display: inline-block;vertical-align:top;
    width: 100px;",
```

```
    selectizeInput(
      'depbwgda',
      'Dependient',
      choices = NULL,
      multiple = TRUE,
      options = list(maxItems = 1))),
  div(style="display: inline-block;vertical-align:top;
    width: 150px;",
    selectizeInput(
      'indepbwgda',
      'Independient(s)',
      choices = NULL,
      multiple = TRUE)),
  div(style="display: inline-block;vertical-align:top;
    width: 150px;",
    shinyWidgets::switchInput(
      "COV.gw",
      inputId = "COVbwgda",
      onLabel = "TRUE",
      offLabel = "FALSE",
      size = "mini")),
  div(style="display: inline-block;vertical-align:top;
    width: 150px;",
    shinyWidgets::switchInput(
      "prior.gw",
      inputId = "priorbwgda",
      onLabel = "TRUE",
      offLabel = "FALSE",
      size = "mini")),
  div(style="display: inline-block;vertical-align:top;
    width: 150px;",
    shinyWidgets::switchInput(
      "mean.gw",
      inputId = "meanbwgda",
      onLabel = "TRUE",
      offLabel = "FALSE",
      size = "mini")),
  div(style="display: inline-block;vertical-align:top;
    width: 150px;",
    shinyWidgets::switchInput(
      "longlat",
      inputId = "longlatbwgda",
      onLabel = "TRUE",
      offLabel = "FALSE",
      size = "mini")),
  div(style="display: inline-block;vertical-align:top;
    width: 150px;",
    shinyWidgets::switchInput(
      "wqda",
      inputId = "wqdabwgda",
      onLabel = "TRUE",
      offLabel = "FALSE",
      size = "mini")),
  div(style="display: inline-block;vertical-align:top;
    width: 100px;",
    shinyWidgets::switchInput(
      "adaptive",
      inputId = "adaptativebwgda",
      onLabel = "TRUE",
      offLabel = "FALSE",
      size = "mini")),
  div(style="display: inline-block;vertical-align:top;
    width: 100px;",
    shinyWidgets::pickerInput(
      inputId = "kernelbwgda",
      label = "Kernel",
      choices = kernel)),
  div(style="display: inline-block;vertical-align:top;
    width: 200px;",
```

```
      numericInput(
        'powerbwgda',
        'Power (Minkowski distance)',
        2,
        min = 1,max = Inf)),
  div(style="display: inline-block;vertical-align:top;
      width: 200px;",
      sliderInput(
        "thetabwgda",
        "Theta (Angle in radians)",
        min = 0,
        max = 2,
        value = 0,
        step = 0.05))),
conditionalPanel(
  condition = "input.bandSelect == 'bw.gwpca'",
  selectizeInput(
        'varpca',
        'Variables',
        choices = NULL,
        multiple = TRUE ),
  numericInput(
        'kpca',
        'Number of components k :',
        2,
        min = 1,
        max =Inf),
shinyWidgets::switchInput(
        "Robust",
        inputId = "robustpca",
        onLabel = "TRUE",
        offLabel = "FALSE",
        size = "mini"),
  shinyWidgets::pickerInput(
        inputId = "kernelpca",
        label = "Kernel",
        choices = kernel),
   shinyWidgets::switchInput("adaptive",
                inputId = "adaptativepca",
                onLabel = "TRUE",
                offLabel = "FALSE",
                size = "mini"),

      numericInput(
        'powerpca',
        'Power (Minkowski distance)',
        2,
        min = 1,max = Inf),
sliderInput(
        "thetapca","Theta (Angle in radians)",
        min = 0,
        max = 2,
        value = 0,
        step = 0.05),
 shinyWidgets::switchInput(
        "longlat",
        inputId = "longlatpca",
        onLabel = "TRUE",
        offLabel = "FALSE",
        size = "mini")),
conditionalPanel(
  condition = "input.bandSelect == 'bw.gwr'",
  div(style="display: inline-block;vertical-align:top;
      width: 100px;",
      selectizeInput(
        'dependientgwr',
        'Dependient',
        choices = NULL,
        multiple = TRUE,
```

```
            options = list(maxItems = 1))),
        div(style="display: inline-block;vertical-align:top;
            width: 150px;",
            selectizeInput(
              'independientgwr',
              'Independient(s)',
              choices = NULL,multiple = TRUE)),
        div(style="display: inline-block;vertical-align:top;
            width: 100px;",
            shinyWidgets::pickerInput(
              inputId = "approachgwr",
              label = "Approach",
              choices = c("CV","AIC"))),
        div(style="display: inline-block; vertical-align:top;
            width: 100px;",
            shinyWidgets::pickerInput(
              inputId = "kernelgwr",
              label = "Kernel",
              choices = kernel)),
        div(style="display: inline-block;vertical-align:top;
            width: 200px;",
            numericInput(
              'powergwr',
              'Power (Minkowski distance)',
              2,
              min = 1,max = Inf)),
        div(style="display: inline-block;vertical-align:top;
            width: 200px;",
            sliderInput(
              "thetagwr",
              "Theta (Angle in radians)",
              min = 0,
              max = 2,
              value = 0,
              step = 0.05)),
        div(style="display: inline-block;vertical-align:top;
            width: 150px;",
            shinyWidgets::switchInput(
              "longlat",
              inputId = "longlatgwr",
              onLabel = "TRUE",
              offLabel = "FALSE",
              size = "mini")),
        div(style="display: inline-block;vertical-align:top;
            width: 100px;",
            shinyWidgets::switchInput(
              "adaptive",
              inputId = "adaptativegwr",
              onLabel = "TRUE",
              offLabel = "FALSE",
              size = "mini")
        )),
       shinyalert::useShinyalert(),
       shinyjs::useShinyjs(),
       shinyWidgets::actionBttn(
        inputId = "Runbw",
        label = "Run",
        style = "float",
        block = TRUE,
        color = "danger"),
       ),
      mainPanel(verbatimTextOutput("selbw"))
  )))
),
shinydashboard::tabItem(
  tabName = "tab4",fluidRow(column(width = 12,
  shinydashboard::box(width = 16,
      title = "Geographically Weighted Summary Statistics",
      status = "primary",
```

```
solidHeader = TRUE,
collapsible = TRUE,
height = 66000,
sidebarPanel(shinyWidgets::actionBttn("helpgwss",
                                      "Help",
                                      icon = icon("question-circle"),
                                      style = "stretch",
                                      block = FALSE,
                                      color = "primary"
),
shinyBS::bsModal(id="helpgeowss",
                 title = "",
                 trigger = "helpgwss",
                 size="large",
                 tags$iframe(
                   src = "Help_gwss.html",
                   width = "100%",
                   height = "1000px",
                   frameborder = 0,
                   scrolling = "auto"
                 )
),
selectizeInput(
  'vargwss',
  'Variables',
  choices = NULL,
  multiple = TRUE),
div(style="display: inline-block;vertical-align:top; width: 100px;",
                 shinyWidgets::pickerInput(
                   inputId = "kernelgwss",
                   label = "Kernel",
                   choices = kernel)),
               div(style="display: inline-block;vertical-align:top;
                   width: 150px;",
                   shinyWidgets::prettyRadioButtons(
                     inputId = "selbw",
                     label = "Distance bandwidth:",
                     choices = c("Manual","Automatic"))),
               conditionalPanel(
                 condition = "input.selbw == 'Manual'",
                 div(style="display: inline-block;vertical-align:top;
                     width: 200px;",
                     numericInput('bwgwss',
                                  'bw',
                                  10,
                                  min = 1,
                                  max =Inf))),
               div(style="display: inline-block; vertical-align:top;
                   width: 200px;",
                   numericInput(
                     'powergwss',
                     'Power (Minkowski distance)',
                     2,
                     min = 1,
                     max = Inf)),
               div(style="display: inline-block;vertical-align:top;
                   width: 200px;",
                   sliderInput(
                     "thetagwss",
                     "Theta (Angle in radians)",
                     min = 0,
                     max = 2,
                     value = 0,
                     step = 0.05)),
               div(style="display: inline-block;vertical-align:top;
                   width: 100px;",
                   shinyWidgets::switchInput(
                     "longlat",
                     inputId = "longlatgwss",
```

```
                onLabel = "TRUE",
                offLabel = "FALSE",
                size = "mini")),
            div(style="display: inline-block;vertical-align:top;
                width: 100px;",
                shinyWidgets::switchInput(
                  "quantile",
                  inputId = "quantilegwss",
                  onLabel = "TRUE",
                  offLabel = "FALSE",
                  size = "mini")),
            div(style="display: inline-block;vertical-align:top;
                width: 100px;",
                shinyWidgets::switchInput(
                  "adaptive",
                  inputId = "adaptativegwss",
                  onLabel = "TRUE",
                  offLabel = "FALSE",
                  size = "mini")),
            shinyalert::useShinyalert(),
            shinyjs::useShinyjs(),
            shinyWidgets::actionBttn(
              inputId = "Rungwss",
              label = "Run",
              style = "float",
              block = TRUE,
              color = "danger")),
mainPanel(
  tabsetPanel(
    tabPanel(
      "Summary",
      helpText(h3("Summary Statistics:")),
      DT::DTOutput("DFgwss"),
      br(),
      verbatimTextOutput("gwss")),
    tabPanel("Plot",
             helpText("click in Dropdown Button for customize
                    and download plot"),
             shinyWidgets::dropdown(
               tags$h3("List of Input"),
               selectInput('plotgwss',
                           'Select variable',choices = NULL),
               textInput("maingwss",
                         label = "Main title",
                         value = "Write main title..."),

               shinyWidgets::pickerInput(
                 inputId = "colorgwss",
                 label = "Select pallete",
                 choices = Pallete_color),
               helpText(h4("North arrow position")),
               numericInput("latgwss", "Latitude",
                            min = -90, max = 90, value = 32
               ),
               numericInput("longwss", "Longitude",
                            min = -180, max = 180, value = -74
               ),
               numericInput("scalegwss", "scale Arrow",
                            min = 1, max = 10, value = 3
               ),
               shinyWidgets::actionBttn(
                 inputId = "Runplotgwss",
                 label = "Plot",
                 style = "float",
                 color = "success"
               ),
             radioButtons("butdowngwss", "Select the Option",
                          choices = list("png","pdf")),
               downloadButton(outputId = "downgwss",
```

```
                    label = "Download the plot"),
                  style = "unite",
                  #icon = icon("gear"),
                  status = "primary", width = "300px",
                  animate = shinyWidgets::animateOptions(enter = animations$fading_entrances$fadeInLeftBig,
                    exit = animations$fading_exits$fadeOutRightBig)
                ),
                plotOutput("mapgwss",width = "100%"))
        )
    )))),
shinydashboard::tabItem(
  tabName = "tab52",fluidRow(column(width = 12,height = 6000,
  shinydashboard::box(width = 12,
                      title = "Geographically Weighted Regression",
                      status = "primary",
                      solidHeader = TRUE,
                      collapsible = TRUE,
                      height = 66000,
                      sidebarPanel(
                        shinyWidgets::actionBttn(
                          "helpgwr",
                          "Help",
                          icon = icon("question-circle"),
                          style = "stretch",
                          block = FALSE,
                          color = "primary"),
                        shinyBS::bsModal(
                          id="helpgeowr",
                          title = "",
                          trigger = "helpgwr",
                          size="large",
                          tags$iframe(
                            src = "help_gwr.html",
                            width = "100%",
                            height = "1000px",
                            frameborder = 0,
                            scrolling = "auto")
                        ),
                        shinyWidgets::pickerInput(
                          inputId = "Selectgwr",
                          label = "Choose:",
                          choices = c(
                            "Select",
                            "Local collinearity diagnostics",
                            "Basic GWR model",
                            "Robust GWR model",
                            "Generalised GWR models",
                            "Heteroskedastic GWR",
                            "Mixed GWR",
                            "Scalable GWR"),
                          options = list(
                            style = "btn-primary")),
                        conditionalPanel(
                          condition = "input.Selectgwr == 'Local collinearity diagnostics'",
                          div(style="display: inline-block;vertical-align:top;
                                        width: 100px;",
                              selectizeInput('dependientbgwrd',
                                             'Dependient',
                                             choices = NULL,
                                             multiple = TRUE,
                                             options = list(maxItems = 1))),
                          div(style="display: inline-block;vertical-align:top;
                          width: 150px;",
                          selectizeInput(
                            'independientbgwrd',
                            'Independient(s)',
                            choices = NULL,
                            multiple = TRUE)),
                          div(style="display: inline-block;vertical-align:top;
```

```
    width: 100px;",
    shinyWidgets::switchInput(
      "longlat",
      inputId = "longlatbgwrd",
      onLabel = "TRUE",
      offLabel = "FALSE",
      size = "mini")),
    div(style="display: inline-block;vertical-align:top;
    width: 100px;",
    shinyWidgets::switchInput("adaptive",
                              inputId = "adaptativebgwrd",
                              onLabel = "TRUE",
                              offLabel = "FALSE",
                              size = "mini")),
    div(
      style="display: inline-block;vertical-align:top;
  width: 150px;",
  shinyWidgets::prettyRadioButtons(
    inputId = "selbgwrd",
    label = "Distance bandwidth:",
    choices = c("Automatic","Manual")
  )),
  conditionalPanel(condition = "input.selbgwrd == 'Manual'",
                   div(style="display: inline-block;
                   vertical-align:top; width: 200px;",
                   numericInput('bwbgwrd',
                                'bw',
                                10,
                                min = 1,
                                max =Inf))),
  div(style="display: inline-block;
  vertical-align:top; width: 100px;",
  shinyWidgets::pickerInput(
    inputId = "kernelbgwrd",
    label = "Kernel",
    choices = kernel
  )),
  div(style="display: inline-block;
              vertical-align:top; width: 200px;",
      numericInput('powerbgwrd',
                   'Power (Minkowski distance)',
                   2,
                   min = 1,
                   max = Inf)),
  div(style="display: inline-block;
              vertical-align:top; width: 200px;",
      sliderInput("thetabgwrd",
                  "Theta (Angle in radians)",
                  min = 0,
                  max = 2,
                  value = 0,
                  step = 0.05
      )),

  shinyalert::useShinyalert(),
  shinyjs::useShinyjs(),
  shinyWidgets::actionBttn(
    inputId = "Runbgwrd",
    label = "Run",
    style = "float",
    block = TRUE,
    color = "danger"
  )),
conditionalPanel(
  condition = "input.Selectgwr == 'Basic GWR model'",
  div(style="display: inline-block;vertical-align:top;
                        width: 100px;",
      selectizeInput('dependientbgwr',
                     'Dependient',
```

```
                    choices = NULL,
                    multiple = TRUE,
                    options = list(maxItems = 1))),
    div(style="display: inline-block;vertical-align:top;
        width: 150px;",
        selectizeInput(
          'independientbgwr',
          'Independient(s)',
          choices = NULL,
          multiple = TRUE)),
    div(style="display: inline-block;vertical-align:top;
        width: 100px;",
        shinyWidgets::switchInput(
          "longlat",
          inputId = "longlatbgwr",
          onLabel = "TRUE",
          offLabel = "FALSE",
          size = "mini")),
    div(style="display: inline-block;vertical-align:top;
        width: 100px;",
        shinyWidgets::switchInput(
          "cv",
          inputId = "cvbgwr",
          onLabel = "TRUE",
          offLabel = "FALSE",
          size = "mini")),
    div(style="display: inline-block;vertical-align:top;
        width: 100px;",
    shinyWidgets::switchInput("adaptive",
                inputId = "adaptativebgwr",
                onLabel = "TRUE",
                offLabel = "FALSE",
                size = "mini")),
    div(
      style="display: inline-block;vertical-align:top;
    width: 150px;",
shinyWidgets::prettyRadioButtons(
  inputId = "selbgwr",
  label = "Distance bandwidth:",
  choices = c("Automatic","Manual")
    )),
conditionalPanel(condition = "input.selbgwr == 'Manual'",
                 div(style="display: inline-block;
                     vertical-align:top; width: 200px;",
                     numericInput('bwbgwr',
                                  'bw',
                                  10,
                                  min = 1,
                                  max =Inf))),
div(style="display: inline-block;
    vertical-align:top; width: 100px;",
    shinyWidgets::pickerInput(
      inputId = "kernelbgwr",
      label = "Kernel",
      choices = kernel
    )),
div(style="display: inline-block;
                vertical-align:top; width: 200px;",
    numericInput('powerbgwr',
                'Power (Minkowski distance)',
                2,
                min = 1,
                max = Inf)),
div(style="display: inline-block;
                vertical-align:top; width: 200px;",
    sliderInput("thetabgwr",
                "Theta (Angle in radians)",
                min = 0,
                max = 2,
```

```
                value = 0,
                step = 0.05
      )),

    shinyalert::useShinyalert(),
    shinyjs::useShinyjs(),
    shinyWidgets::actionBttn(
     inputId = "Runbgwr",
     label = "Run",
     style = "float",
     block = TRUE,
     color = "danger"
    )),
  conditionalPanel(
   condition = "input.Selectgwr == 'Generalised GWR models'",
                div(
                  style="display: inline-block;
                  vertical-align:top; width: 150px;",
                  selectizeInput(
                   'dependientGGWR',
                   'Dependient',
                   choices = NULL,
                   multiple = TRUE,
                   options = list(maxItems = 1)
                   )),
   div(style="display: inline-block;vertical-align:top;
     width: 150px;",
     selectizeInput(
      'independientGGWR',
      'Independient(s)',
      choices = NULL,
      multiple = TRUE)),
   div(style="display: inline-block;vertical-align:top;
     width: 100px;",
     shinyWidgets::switchInput(
      "longlat",
      inputId = "longlatGGWR",
      onLabel = "TRUE",
      offLabel = "FALSE",
      size = "mini")),
   div(style="display: inline-block;vertical-align:top;
     width: 100px;",
     shinyWidgets::switchInput(
      "cv",
      inputId = "cvGGWR",
      onLabel = "TRUE",
      offLabel = "FALSE",
      size = "mini")),
   div(style="display: inline-block;vertical-align:top;
     width: 100px;",
     shinyWidgets::switchInput(
      "adaptive",
      inputId = "adaptativeGGWR",
      onLabel = "TRUE",
      offLabel = "FALSE",
      size = "mini")),
   div(style="display: inline-block;vertical-align:top;
     width: 150px;",
     shinyWidgets::prettyRadioButtons(
      inputId = "selGGWR",
      label = "Distance bandwidth:",
      choices = c("Automatic","Manual"))),
   conditionalPanel(
    condition = "input.selGGWR == 'Manual'",
    div(style="display: inline-block;vertical-align:top;
      width: 200px;",
      numericInput('bwGGWR',
                   'bw',
                   10,
```

```
                    min = 1,
                    max =Inf))),
      div(style="display: inline-block; vertical-align:top;
        width: 100px;",
        shinyWidgets::pickerInput(
          inputId = "kernelGGWR",
          label = "Kernel",
          choices = kernel)),
      div(style="display: inline-block;vertical-align:top;
        width: 150px;",
        shinyWidgets::pickerInput(
          inputId = "familyGGWR",
          label = "Family",
          choices = family)),
      div(style="display: inline-block; vertical-align:top;
        width: 200px;",
        numericInput(
          'powerGGWR',
          'Power (Minkowski distance)',
          2,
          min = 1,
          max = Inf)),
      div(style="display: inline-block; vertical-align:top;
        width: 100px;",
        numericInput('maxiterGGWR',
                     'maxiter',
                     20,
                     min = 2,
                     max = Inf)),
      div(style="display: inline-block;
        vertical-align:top; width: 200px;",
        sliderInput("thetaGGWR",
                    "Theta (Angle in radians)",
                    min = 0,
                    max = 2,
                    value = 0,
                    step = 0.05
                    )),
      shinyalert::useShinyalert(),
      shinyjs::useShinyjs(),
      shinyWidgets::actionBttn(
        inputId = "RunGGWR",
        label = "Run",
        style = "float",
        block = TRUE,
        color = "danger")),
  conditionalPanel(
    condition = "input.Selectgwr == 'Robust GWR model'",
                      div(style="display: inline-block;
        vertical-align:top; width: 100px;",
        selectizeInput('dependientRobust',
                     'Dependient',
                     choices = NULL,
                     multiple = TRUE,
                     options = list(maxItems = 1)
        )),
        div(style="display: inline-block;
        vertical-align:top; width: 150px;",
        selectizeInput('independientRobust',
                     'Independient(s)',
                     choices = NULL,
                     multiple = TRUE)),
        div(style="display: inline-block;vertical-align:top;
          width: 100px;",
          shinyWidgets::switchInput("longlat",
                      inputId = "longlatRobust",
                      onLabel = "TRUE",
                      offLabel = "FALSE",
                      size = "mini")),
```

```
div(style="display: inline-block;vertical-align:top;
        width: 100px;",
        shinyWidgets::switchInput("Filtered",
                    inputId = "filtered",
                    onLabel = "TRUE",
                    offLabel = "FALSE",
                    size = "mini")),
      div(style="display: inline-block;vertical-align:top;
        width: 100px;",
        shinyWidgets::switchInput("adaptive",
                    inputId = "adaptativeRobust",
                    onLabel = "TRUE",
                    offLabel = "FALSE",
                    size = "mini")

      ),

      div(style="display: inline-block;vertical-align:top;
        width: 150px;",
        shinyWidgets::prettyRadioButtons(
          inputId = "selRobust",
          label = "Distance bandwidth:",
          choices = c("Automatic","Manual"))),
      conditionalPanel(
        condition = "input.selRobust == 'Manual'",
        div(style="display: inline-block;vertical-align:top;
          width: 200px;",
          numericInput('bwRobust',
                       'bw',
                       10,
                       min = 1,
                       max =Inf))),
      div(style="display: inline-block; vertical-align:top;
        width: 100px;",
        numericInput('maxiterRobust',
                    'maxiter',
                    20,
                    min = 2,
                    max = Inf)),
      div(style="display: inline-block; vertical-align:top;
        width: 100px;",
        shinyWidgets::pickerInput(
          inputId = "kernelRobust",
          label = "Kernel",
          choices = kernel
        )),
      div(style="display: inline-block;
                    vertical-align:top; width: 200px;",
        numericInput('powerRobust',
                    'Power (Minkowski distance)',
                    2,
                    min = 1,
                    max = Inf)),
      div(style="display: inline-block;
                    vertical-align:top; width: 200px;",
        sliderInput("thetaRobust",
                    "Theta (Angle in radians)",
                    min = 0,
                    max = 2,
                    value = 0,
                    step = 0.05
        )),

      shinyalert::useShinyalert(),
      shinyjs::useShinyjs(),
      shinyWidgets::actionBttn(
        inputId = "RunRobust",
        label = "Run",
        style = "float",
```

```
          block = TRUE,
          color = "danger")),
conditionalPanel(
  condition = "input.Selectgwr == 'Heteroskedastic GWR'",
                    div(style="display: inline-block;
        vertical-align:top; width: 100px;",
        selectizeInput('dependienthetero',
                    'Dependient',
                    choices = NULL,
                    multiple = TRUE,
                    options = list(maxItems = 1)
        )),
        div(style="display: inline-block;
        vertical-align:top; width: 150px;",
        selectizeInput('independienthetero',
                    'Independient(s)',
                    choices = NULL,
                    multiple = TRUE)),
        div(style="display: inline-block;vertical-align:top;
            width: 100px;",
            shinyWidgets::switchInput("longlat",
                     inputId = "longlathetero",
                     onLabel = "TRUE",
                     offLabel = "FALSE",
                     size = "mini")),
        div(style="display: inline-block;vertical-align:top;
            width: 100px;",
            shinyWidgets::switchInput("adaptive",
                     inputId = "adaptativehetero",
                     onLabel = "TRUE",
                     offLabel = "FALSE",
                     size = "mini")

        ),

        div(style="display: inline-block;vertical-align:top;
            width: 150px;",
            shinyWidgets::prettyRadioButtons(
              inputId = "selhetero",
              label = "Distance bandwidth:",
              choices = c("Automatic","Manual"))),
  conditionalPanel(
    condition = "input.selhetero == 'Manual'",
    div(style="display: inline-block;vertical-align:top;
        width: 200px;",
        numericInput('bwhetero',
                     'bw',
                     10,
                     min = 1,
                     max =Inf))),
  div(style="display: inline-block; vertical-align:top;
      width: 100px;",
      numericInput(
        'maxiterhetero',
        'maxiter',
        50,
        min = 2,
        max = Inf)),
        div(style="display: inline-block; vertical-align:top;
            width: 100px;",
            shinyWidgets::pickerInput(
              inputId = "kernelhetero",
              label = "Kernel",
              choices = kernel
            )),
        div(style="display: inline-block;
                     vertical-align:top; width: 200px;",
            numericInput('powerhetero',
                     'Power (Minkowski distance)',
```

```
                    2,
                    min = 1,
                    max = Inf)),
        div(style="display: inline-block;
                    vertical-align:top; width: 200px;",
           sliderInput("thetahetero",
                    "Theta (Angle in radians)",
                    min = 0,
                    max = 2,
                    value = 0,
                    step = 0.05
           )),

        shinyalert::useShinyalert(),
        shinyjs::useShinyjs(),
        shinyWidgets::actionBttn(
          inputId = "Runhetero",
          label = "Run",
          style = "float",
          block = TRUE,
          color = "danger"
        )
  ),
conditionalPanel(condition = "input.Selectgwr == 'Mixed GWR'",
                    div(style="display: inline-block;
        vertical-align:top; width: 100px;",
        selectizeInput('dependientmixed',
                    'Dependient',
                    choices = NULL,
                    multiple = TRUE,
                    options = list(maxItems = 1)
        )),
        div(style="display: inline-block;
        vertical-align:top; width: 150px;",
        selectizeInput('independientmixed',
                    'Independient(s)',
                    choices = NULL,
                    multiple = TRUE)),
        div(style="display: inline-block;
        vertical-align:top; width: 150px;",
        selectizeInput('fixedvars',
                    ' fixed var',
                    choices = NULL,
                    multiple = TRUE)),
        div(style="display: inline-block;vertical-align:top;
           width: 150px;",
           shinyWidgets::switchInput("Intercep.fixed",
                    inputId = "intercepfixed",
                    onLabel = "TRUE",
                    offLabel = "FALSE",
                    size = "mini")),
        div(style="display: inline-block;vertical-align:top;
           width: 100px;",
           shinyWidgets::switchInput("Diagnostic",
                    inputId = "diagnostic",
                    onLabel = "TRUE",
                    offLabel = "FALSE",
                    size = "mini")),
        div(style="display: inline-block;vertical-align:top;
           width: 100px;",
           shinyWidgets::switchInput("longlat",
                    inputId = "longlatmixed",
                    onLabel = "TRUE",
                    offLabel = "FALSE",
                    size = "mini")),
        div(style="display: inline-block;vertical-align:top;
           width: 100px;",
           shinyWidgets::switchInput("adaptive",
                    inputId = "adaptativemixed",
```

```
                    onLabel = "TRUE",
                    offLabel = "FALSE",
                    size = "mini")
        ),

        div(style="display: inline-block;vertical-align:top;
            width: 150px;",
            shinyWidgets::prettyRadioButtons(
              inputId = "selmixed",
              label = "Distance bandwidth:",
              choices = c("Automatic","Manual")
              )),
        conditionalPanel(
          condition = "input.selmixed == 'Manual'",
          div(style="display: inline-block;vertical-align:top;
              width: 200px;",
              numericInput(
                'bwmixed',
                'bw',
                10,
                min = 1,
                max =Inf))),
        div(style="display: inline-block; vertical-align:top;
            width: 100px;",
            shinyWidgets::pickerInput(
              inputId = "kernelmixed",
              label = "Kernel",
              choices = kernel
            )),
        div(style="display: inline-block;
                    vertical-align:top; width: 200px;",
            numericInput('powermixed',
                    'Power (Minkowski distance)',
                    2,
                    min = 1,
                    max = Inf)),
        div(style="display: inline-block;
                    vertical-align:top; width: 200px;",
            sliderInput("thetamixed",
                    "Theta (Angle in radians)",
                    min = 0,
                    max = 2,
                    value = 0,
                    step = 0.05
            )),

        shinyalert::useShinyalert(),
        shinyjs::useShinyjs(),
        shinyWidgets::actionBttn(
          inputId = "Runmixed",
          label = "Run",
          style = "float",
          block = TRUE,
          color = "danger"
        )
        ),
conditionalPanel(
  condition = "input.Selectgwr == 'Scalable GWR'",
                    div(style="display: inline-block;
        vertical-align:top; width: 100px;",
        selectizeInput('dependientscal',
                    'Dependient',
                    choices = NULL,
                    multiple = TRUE,
                    options = list(maxItems = 1)
        )),
        div(style="display: inline-block;
        vertical-align:top; width: 150px;",
        selectizeInput('independientscal',
```

```
            'Independient(s)',
            choices = NULL,
            multiple = TRUE)),
div(style="display: inline-block;vertical-align:top;
    width: 100px;",
    numericInput('bwscal',
                 'bw.adapt',
                 10,
                 min = 1,
                 max =Inf)),
div(style="display: inline-block;vertical-align:top;
    width: 100px;",
    numericInput('polinom',
                 'polynomial',
                 4,
                 min = 1,
                 max = 10)),
div(style="display: inline-block; vertical-align:top;
    width: 100px;",
    shinyWidgets::pickerInput(
      inputId = "kernelscal",
      label = "Kernel",
      choices = c("gaussian",
                  "exponential")
    )),
div(style="display: inline-block;
            vertical-align:top; width: 200px;",
    numericInput('powerscal',
                 'Power (Minkowski distance)',
                 2,
                 min = 1,
                 max = Inf)),
div(style="display: inline-block;
            vertical-align:top; width: 200px;",
    sliderInput("thetascal",
                "Theta (Angle in radians)",
                min = 0,
                max = 2,
                value = 0,
                step = 0.05
    )),
div(style="display: inline-block;vertical-align:top;
    width: 100px;",
    shinyWidgets::switchInput("longlat",
                inputId = "longlatscal",
                onLabel = "TRUE",
                offLabel = "FALSE",
                size = "mini")),

shinyalert::useShinyalert(),
shinyjs::useShinyjs(),
shinyWidgets::actionBttn(
  inputId = "Runscal",
  label = "Run",
  style = "float",
  block = TRUE,
  color = "danger"
)
)
                    ),

mainPanel(
  tabsetPanel(
    tabPanel("Summary",
          conditionalPanel(condition = "input.Selectgwr == 'Local collinearity diagnostics'",
                       helpText(h3("SDF:")),
                        DT::DTOutput("gwrbasicdVIF"),
          ),
          conditionalPanel(condition = "input.Selectgwr == 'Basic GWR model'",
```

```
            verbatimTextOutput("gwrbasic")),
            conditionalPanel(condition ="input.Selectgwr == 'Generalised GWR models'",
            verbatimTextOutput("GGWRbasic")),
            conditionalPanel(condition ="input.Selectgwr == 'Robust GWR model'",
            verbatimTextOutput("GGWRobust")),
            conditionalPanel(condition ="input.Selectgwr == 'Heteroskedastic GWR'",
            verbatimTextOutput("GWRhetero")),
            conditionalPanel(condition ="input.Selectgwr == 'Mixed GWR'",
            verbatimTextOutput("GWRmixed")),
            conditionalPanel(condition ="input.Selectgwr == 'Scalable GWR'",
            verbatimTextOutput("GWRscalable")),
            br(),

            conditionalPanel(condition = "input.Selectgwr == 'Basic GWR model'",
                        helpText(h3("SDF:")),
                        DT::DTOutput("SDFgwrbasic")),
            conditionalPanel(condition ="input.Selectgwr == 'Generalised GWR models'",
                        helpText(h3("SDF:")),
                        DT::DTOutput("SDFGGWRbasic")),
            conditionalPanel(condition ="input.Selectgwr == 'Robust GWR model'",
                        helpText(h3("SDF:")),
                        DT::DTOutput("SDFGWRRobust")),
            conditionalPanel(condition ="input.Selectgwr == 'Heteroskedastic GWR'",
                        helpText(h3("SDF:")),
                        DT::DTOutput("SDFGWRhetero")),
            conditionalPanel(condition ="input.Selectgwr == 'Mixed GWR'",
                        helpText(h3("SDF:")),
                        DT::DTOutput("SDFGWRmixed")),
            conditionalPanel(condition ="input.Selectgwr == 'Scalable GWR'",
                        helpText(h3("SDF:")),
                        DT::DTOutput("SDFGWRscalable"))
),
tabPanel(
  "Plot",
  conditionalPanel(
    condition = "input.Selectgwr == 'Local collinearity diagnostics'",
    helpText("click in Dropdown Button for
             customize and download plot"),
    shinyWidgets::dropdown(
      tags$h3("List of Input"),
      selectInput('plotgwrd',
                  'Select',
                  choices = NULL),
      textInput("maingwrd",
                label = "Main title",
                value = "Write main title..."),
      shinyWidgets::pickerInput(
        inputId = "colorgwrd",
        label = "Select pallete",
        choices = Pallete_color),
      br(),
      helpText(h4("North arrow position")),
      numericInput("latgwrd",
                   "Latitude",
                   min = -90,
                   max = 90,
                   value = 32),
      numericInput("scalegwrd",
                   "scale Arrow",
                   min = 1,
                   max = 10,
                   value = 3),
      numericInput("longwrd",
                   "Longitude",
                   min = -180,
                   max = 180,
                   value = -74),
      shinyWidgets::actionBttn(
        inputId = "Runplotgwrd",
```

```
      label = "Plot",
      style = "float",
      color = "success"),
    radioButtons("butdowngwrd",
                 "Select the Option",
                 choices = list("png",
                                "pdf")),
    downloadButton(
      outputId = "downgwrd",
      label = "Download the plot"),
    style = "unite",
    #icon = icon("gear"),
    status = "primary",
    width = "300px",
    animate = shinyWidgets::animateOptions(
      enter = animations$fading_entrances$fadeInLeftBig,
      exit = animations$fading_exits$fadeOutRightBig))),
conditionalPanel(
  condition = "input.Selectgwr == 'Basic GWR model'",
  helpText("click in Dropdown Button for
           customize and download plot"),
  shinyWidgets::dropdown(
    tags$h3("List of Input"),
    selectInput('plotgwr',
                'Select variable',
                choices = NULL),
    textInput("maingwr",
              label = "Main title",
              value = "Write main title..."),
    shinyWidgets::pickerInput(
      inputId = "colorgwr",
      label = "Select pallete",
      choices = Pallete_color),
    br(),
    helpText(h4("North arrow position")),
    numericInput("latgwr",
                 "Latitude",
                 min = -90,
                 max = 90,
                 value = 32),
    numericInput("scalegwr",
                 "scale Arrow",
                 min = 1,
                 max = 10,
                 value = 3),
    numericInput("longwr",
                 "Longitude",
                 min = -180,
                 max = 180,
                 value = -74),
    shinyWidgets::actionBttn(
      inputId = "Runplotgwr",
      label = "Plot",
      style = "float",
      color = "success"),
    radioButtons("butdowngwr",
                 "Select the Option",
                 choices = list("png",
                                "pdf")),
    downloadButton(
      outputId = "downgwr",
      label = "Download the plot"),
    style = "unite",
    #icon = icon("gear"),
    status = "primary",
    width = "300px",
    animate = shinyWidgets::animateOptions(
      enter = animations$fading_entrances$fadeInLeftBig,
      exit = animations$fading_exits$fadeOutRightBig))),
```

```
        conditionalPanel(
          condition = "input.Selectgwr == 'Generalised GWR models'",
          helpText("click in Dropdown Button for customize
                   and download plot"),
          shinyWidgets::dropdown(
            tags$h3("List of Input"),
            selectInput('plotggwr',
                        'Select variable',
                        choices = NULL),
            textInput("mainggwr",
                      label = "Main title",
                      value = "Write main title..."),
            shinyWidgets::pickerInput(
              inputId = "colorggwr",
              label = "Select pallete",
              choices = Pallete_color),
            br(),
            helpText(h4("North arrow position")),
            numericInput("latggwr",
                         "Latitude",
                         min = -90,
                         max = 90,
                         value = 32),
            numericInput("scaleggwr",
                         "scale Arrow",
                         min = 1,
                         max = 10,
                         value = 3),
            numericInput("longgwr",
                         "Longitude",
                         min = -180,
                         max = 180,
                         value = -75),
            shinyWidgets::actionBttn(
              inputId = "Runplotggwr",
              label = "Plot",
              style = "float",
              color = "success"
              ),
            radioButtons("butdownggwr",
                         "Select the Option",
                         choices = list("png",
                                        "pdf")),
            downloadButton(outputId = "downggwr",
                           label = "Download the plot"),
            style = "unite",
            #icon = icon("gear"),
            status = "primary", width = "300px",
            animate = shinyWidgets::animateOptions(
              enter = animations$fading_entrances$fadeInLeftBig,
              exit = animations$fading_exits$fadeOutRightBig)
            )),
      conditionalPanel(
        condition = "input.Selectgwr == 'Robust GWR model'",
        helpText("click in Dropdown Button for customize and download plot"),
        shinyWidgets::dropdown(
          tags$h3("List of Input"),
          selectInput('plotRgwr',
                      'Select variable',
                      choices = NULL),
          textInput("mainRgwr",
                    label = "Main title",
                    value = "Write main title..."),
          shinyWidgets::pickerInput(
            inputId = "colorRgwr",
            label = "Select pallete",
            choices = Pallete_color),
          br(),
          helpText(h4("North arrow position")),
```

```
      numericInput("latRgwr", "Latitude",
                   min = -90,
                   max = 90,
                   value = 32),
      numericInput("scaleRgwr",
                   "scale Arrow",
                   min = 1,
                   max = 10,
                   value = 2 ),
      numericInput("lonRgwr",
                   "Longitude",
                   min = -180,
                   max = 180,
                   value = -74),
      shinyWidgets::actionBttn(
        inputId = "RunplotRgwr",
        label = "Plot",
        style = "float",
        color = "success"),
      radioButtons("butdownRgwr",
                   "Select the Option",
                   choices = list("png","pdf")),
      downloadButton(
        outputId = "downRgwr",
        label = "Download the plot"),
      style = "unite",
      #icon = icon("gear"),
      status = "primary", width = "300px",
      animate = shinyWidgets::animateOptions(
        enter = animations$fading_entrances$fadeInLeftBig,
        exit = animations$fading_exits$fadeOutRightBig))),
conditionalPanel(
  condition = "input.Selectgwr == 'Heteroskedastic GWR'",
  helpText("click in Dropdown Button for customize and download plot"),
  shinyWidgets::dropdown(
    tags$h3("List of Input"),
    selectInput('plotHgwr',
                'Select variable',
                choices = NULL),
    textInput("mainHgwr",
              label = "Main title",
              value = "Write main title..."),
    shinyWidgets::pickerInput(
      inputId = "colorHgwr",
      label = "Select pallete",
      choices = Pallete_color),
    br(),
    helpText(h4("North arrow position")),
    numericInput("latHgwr",
                 "Latitude",
                 min = -90,
                 max = 90,
                 value = 32),
    numericInput("scaleHgwr",
                 "scale Arrow",
                 min = 1,
                 max = 10,
                 value = 3 ),
    numericInput("lonHgwr",
                 "Longitude",
                 min = -180,
                 max = 180,
                 value = -74),
    shinyWidgets::actionBttn(
      inputId = "RunplotHgwr",
      label = "Plot",
      style = "float",
      color = "success"),
    radioButtons("butdownHgwr",
```

```
                "Select the Option",
                choices = list("png","pdf")),
    downloadButton(outputId = "downHgwr",
                   label = "Download the plot"),
    style = "unite",
    #icon = icon("gear"),
    status = "primary",
    width = "300px",
    animate = shinyWidgets::animateOptions(
      enter = animations$fading_entrances$fadeInLeftBig,
      exit = animations$fading_exits$fadeOutRightBig))),
  conditionalPanel(
    condition = "input.Selectgwr == 'Mixed GWR'",
    helpText("click in Dropdown Button for customize and download plot"),
    shinyWidgets::dropdown(
      tags$h3("List of Input"),
      selectInput('plotMgwr',
                  'Select variable',
                  choices = NULL),
      textInput("mainMgwr",
                label = "Main title",
                value = "Write main title..."),
      shinyWidgets::pickerInput(
        inputId = "colorMgwr",
        label = "Select pallete",
        choices = Pallete_color),
      br(),
      helpText(h4("North arrow position")),
      numericInput("latMgwr", "Latitude",
                   min = -90,
                   max = 90,
                   value = 32),
      numericInput("scaleMgwr",
                   "scale Arrow",
                   min = 1,
                   max = 10,
                   value = 2),
      numericInput("lonMgwr",
                   "Longitude",
                   min = -180,
                   max = 180,
                   value = -74),
      shinyWidgets::actionBttn(
        inputId = "RunplotMgwr",
        label = "Plot",
        style = "float",
        color = "success"),
      radioButtons("butdownMgwr",
                   "Select the Option",
                   choices = list("png","pdf")),
      downloadButton(outputId = "downMgwr",
                     label = "Download the plot"),
      style = "unite",
      #icon = icon("gear"),
      status = "primary", width = "300px",
      animate = shinyWidgets::animateOptions(
        enter = animations$fading_entrances$fadeInLeftBig,
        exit = animations$fading_exits$fadeOutRightBig)
      )),
  conditionalPanel(
    condition = "input.Selectgwr == 'Scalable GWR'",
    helpText("click in Dropdown Button for customize and download plot"),
    shinyWidgets::dropdown(
      tags$h3("List of Input"),
      selectInput('plotSgwr',
                  'Select variable',choices = NULL),
      textInput("mainSgwr",
                label = "Main title",
                value = "Write main title..."),
```

```
      shinyWidgets::pickerInput(
        inputId = "colorSgwr",
        label = "Select pallete",
        choices = Pallete_color),
      br(),
      helpText(h4("North arrow position")),
      numericInput("latSgwr",
                   "Latitude",
                   min = -90,
                   max = 90,
                   value = 32),
      numericInput("scaleSgwr",
                   "scale Arrow",
                   min = 1,
                   max = 10,
                   value = 3),
      numericInput("lonSgwr",
                   "Longitude",
                   min = -180,
                   max = 180,
                   value = -74),
      shinyWidgets::actionBttn(
        inputId = "RunplotSgwr",
        label = "Plot",
        style = "float",
        color = "success"
        ),
      radioButtons("butdownSgwr",
                   "Select the Option",
                   choices = list("png","pdf")),
      downloadButton(outputId = "downSgwr",
                     label = "Download the plot"),
      style = "unite",
      #icon = icon("gear"),
      status = "primary", width = "300px",
      animate = shinyWidgets::animateOptions(
        enter = animations$fading_entrances$fadeInLeftBig,
        exit = animations$fading_exits$fadeOutRightBig))),
    conditionalPanel(condition = "input.Selectgwr == 'Local collinearity diagnostics'",
                   plotOutput("mapgwrbasicd") ),
    conditionalPanel(condition = "input.Selectgwr == 'Basic GWR model'",
                   plotOutput("mapgwrbasic") ),
    conditionalPanel( condition = "input.Selectgwr == 'Scalable GWR'",
                    plotOutput("mapSgwr")),
    conditionalPanel( condition = "input.Selectgwr == 'Mixed GWR'",
                    plotOutput("mapMgwr")),
    conditionalPanel( condition = "input.Selectgwr == 'Heteroskedastic GWR'",
                    plotOutput("mapHgwr")),
    conditionalPanel( condition = "input.Selectgwr == 'Robust GWR model'",
                    plotOutput("mapRgwr")),
    conditionalPanel( condition = "input.Selectgwr == 'Generalised GWR models'",
                    plotOutput("mapggwr"))
)))))),
shinydashboard::tabItem(
  tabName = "tab53",fluidRow(column(width = 12,
  shinydashboard::box(
    width = 12,
    title = "Geographically Weighted Principal Components Analysis",
    status = "primary",
    solidHeader = TRUE,
    collapsible = TRUE,
    height = 66000,
    sidebarPanel(shinyWidgets::actionBttn(
      "helpgwpca",
      "Help",
      icon = icon("question-circle"),
      style = "stretch",
      block = FALSE,
      color = "primary"
```

```
        ),
shinyBS::bsModal(
  id="helpgeowpca",
  title = "",
  trigger = "helpgwpca",
  size="large",
  tags$iframe(
    src = "Help_gwpca.html",
    width = "100%",
    height = "1000px",
    frameborder = 0,
    scrolling = "auto")),
selectizeInput(
  'vargwpca',
  'Variables',
  choices = NULL,
  multiple = TRUE),
numericInput(
  'kgwpca',
  'k',
  2,
  min = 1,
  max =Inf),
pickerInput(
  inputId = "kernelgwpca",
  label = "Kernel",
  choices = kernel),
switchInput(
  "Robust",
  inputId = "robustgwpca",
  onLabel = "TRUE",
  offLabel = "FALSE",
  size = "mini"),
switchInput(
  "adaptive",
  inputId = "adaptativegwpca",
  onLabel = "TRUE",
  offLabel = "FALSE",
  size = "mini"),
shinyWidgets::switchInput(
  "longlat",
  inputId = "longlatgwpca",
  onLabel = "TRUE",
  offLabel = "FALSE",
  size = "mini"),
shinyWidgets::prettyRadioButtons(
  inputId = "selgwpca",
  label = "Distance bandwidth:",
  choices = c("Automatic",
              "Manual")),
conditionalPanel(condition = "input.selgwpca == 'Manual'",
                 div(style="display: inline-block;vertical-align:top;
                     width: 200px;",
                     numericInput(
                       'bwgwpca',
                       'bw',
                       10,
                       min = 1,
                       max =Inf))),
numericInput(
  'powergwpca',
  'Power (Minkowski distance)',
  2,
  min = 1,
  max = Inf),
sliderInput(
  "thetagwpca",
  "Theta (Angle in radians)",
  min = 0,
```

```
    max = 2,
    value = 0,
    step = 0.05),
  shinyalert::useShinyalert(),
  shinyjs::useShinyjs(),
  shinyWidgets::actionBttn(
    inputId = "Rungwpca",
    label = "Run",
    style = "float",
    block = TRUE,
    color = "danger")),
mainPanel(tabsetPanel(tabPanel(
  "Summary",
  helpText(h3("Summary :")),
  verbatimTextOutput("gwpca"),
  DT::DTOutput("DFgwpca"),
  br(),
  helpText(h3("Localised loadings :")),
  div(style="display: inline-block;vertical-align:top; width: 200px;",
      numericInput(
        'loading',
        'Component',
        1,
        min = 1,
        max = Inf)),
  DT::DTOutput("loadinggwpca")),
  tabPanel("Percent Total Variation",
           helpText("click in Dropdown Button
                    for customize and download plot"),
           shinyWidgets::dropdown(
             tags$h3("List of Input"),
             numericInput('loadingVgwpca',
                          'Number of components',
                          2,
                          min = 2,
                          max = Inf),
             textInput("mainVgwpca",
                       label = "Main title",
                       value = "Write main title..."),
             shinyWidgets::pickerInput(
               inputId = "colorVgwpca",
               label = "Select pallete",
               choices = Pallete_color),
             helpText(h4("North arrow position")),
             numericInput("latgwpca",
                          "Latitude",
                          min = -90,
                          max = 90,
                          value = 32
                          ),
             numericInput("scalegwpca",
                          "scale Arrow",
                          min = 1,
                          max = 10,
                          value = 3
                          ),
             numericInput("longwpca",
                          "Longitude",
                          min = -180,
                          max = 180,
                          value = -74),
             shinyWidgets::actionBttn(
               inputId = "RunplotVgwr",
               label = "Plot",
               style = "float",
               color = "success"),
             radioButtons("butdownVgwpca",
                          "Select the Option",
                          choices = list("png","pdf")),
```

```
            downloadButton("downVgwpca",
                           label = "Download the plot"),
            style = "unite",
            #icon = icon("gear"),
            status = "primary", width = "300px",
            animate = shinyWidgets::animateOptions(
              enter = animations$fading_entrances$fadeInLeftBig,
              exit = animations$fading_exits$fadeOutRightBig)),
          plotOutput("Vargwpca",
                     width = "100%")
        ),
    tabPanel(
      "Plot winning variable",
      helpText("click in Dropdown Button for customize and download plot"),
      shinyWidgets::dropdown(
        tags$h3("List of Input"),
        numericInput(
          'loadingwingwpca',
          'Components',
          1,
          min = 1,
          max = Inf),
        textInput(
          "mainwingwpca",
          label = "Main title",
          value = "Write main title..."),
        #shinyWidgets::pickerInput(
        # inputId = "colorwingwpca",
        # label = "Select pallete",
        #choices = Pallete_color),
        # helpText(h4("North arrow position")),
        numericInput("latwingwpca",
                     "Latitude",
                     min = -90,
                     max = 90,
                     value = 32),
        numericInput("scalewingwpca",
                     "scale Arrow",
                     min = 1,
                     max = 10,
                     value = 3),
        numericInput("lonwingwpca",
                     "Longitude",
                     min = -180,
                     max = 180,
                     value = -74),
        shinyWidgets::actionBttn(
          inputId = "Runplotwingwpca",
          label = "Plot",
          style = "float",
          color = "success"),
        radioButtons("butdownwingwpca",
                     "Select the Option",
                     choices = list("png","pdf")),
        downloadButton(
          "downwingwpca",
          label = "Download the plot"),
        style = "unite",
        #icon = icon("gear"),
        status = "primary", width = "300px",
        animate = shinyWidgets::animateOptions(
          enter = animations$fading_entrances$fadeInLeftBig,
          exit = animations$fading_exits$fadeOutRightBig)),
      plotOutput("wingwpca",width = "100%"))

                              )
            )))),
shinydashboard::tabItem(
  tabName = "tab54",fluidRow(column(width = 12,
```

```
shinydashboard::box(
  width = 12,
  title = "Geographically Weighted Discriminant Analysis",
  status = "primary",
  solidHeader = TRUE,
  collapsible = TRUE,
  height = 66000,
  sidebarPanel(shinyWidgets::actionBttn(
    "helpgwda",
    "Help",
    icon = icon("question-circle"),
    style = "stretch",
    block = FALSE,
    color = "primary"),
    shinyBS::bsModal(
      id="helpgeowda",
      title = "",
      trigger = "helpgwda",
      size="large",
      tags$iframe(
        src = "Help_gwda.html",
        width = "100%",
        height = "1000px",
        frameborder = 0,
        scrolling = "auto" )),
    selectizeInput('dependgwda',
                   'Grouping factor',
                   choices = NULL,
                   multiple = TRUE,
                   options = list(maxItems = 1)),
    selectizeInput('indepentgwda',
                   'Discriminators',
                   choices = NULL,
                   multiple = TRUE),
    shinyWidgets::switchInput("mean.gw",
                              inputId = "meangwda",
                              onLabel = "TRUE",
                              offLabel = "FALSE",
                              size = "mini"),
    shinyWidgets::switchInput(
      "COV.gw",
      inputId = "COVgwda",
      onLabel = "TRUE",
      offLabel = "FALSE",
      size = "mini"),
    shinyWidgets::switchInput(
      "prior.gw",
      inputId = "priorgwda",
      onLabel = "TRUE",
      offLabel = "FALSE",
      size = "mini"),
    shinyWidgets::switchInput(
      "longlat",
      inputId = "longlatgwda",
      onLabel = "TRUE",
      offLabel = "FALSE",
      size = "mini"),
    shinyWidgets::switchInput(
      "wqda",
      inputId = "wqdagwda",
      onLabel = "TRUE",
      offLabel = "FALSE",
      size = "mini"),
    shinyWidgets::switchInput(
      "adaptive",
      inputId = "adaptativegwda",
      onLabel = "TRUE",
      offLabel = "FALSE",
      size = "mini"),
```

```
  shinyWidgets::prettyRadioButtons(
    inputId ="selgwda",
    label = "Distance bandwidth:",
    choices = c("Automatic","Manual")
                 ),
  conditionalPanel(condition = "input.selgwda == 'Manual'",
                   div(style="display: inline-block;vertical-align:top;
                       width: 150px;",
                       numericInput('bwgwda',
                                    'bw',
                                    10,
                                    min = 1,
                                    max =Inf))),
  div(style="display: inline-block; vertical-align:top;
      width: 200px;",
      numericInput(
        'powergwda',
        'Power (Minkowski distance)',
        2,
        min = 1,
        max = Inf)),
  div(style="display: inline-block;vertical-align:top; width: 150px;",
      shinyWidgets:: pickerInput(
        inputId = "kernelgwda",
        label = "Kernel",
        choices = kernel)),
  div(style="display: inline-block; vertical-align:top; width: 200px;",
      sliderInput(
        "thetagwda",
        "Theta (Angle in radians)",
        min = 0,
        max = 2,
        value = 0,
        step = 0.05)),
  shinyalert::useShinyalert(),
  shinyjs::useShinyjs(),
  shinyWidgets::actionBttn(
    inputId = "Rungwda",
    label = "Run",
    style = "float",
    block = TRUE,
    color = "danger") ),# ffinsidebarpanel54
mainPanel(tabsetPanel(
  tabPanel(
    "Summary",
    verbatimTextOutput("gwda"),
    helpText(h3("Confusion matrix")),
    verbatimTextOutput("predict"),
    helpText(h3(" SDF")),
    DT::DTOutput("DFgwda")
    ),
  tabPanel(
    "Plot",
    helpText("click in Dropdown Button for customize and download plot"),
    shinyWidgets::dropdown(
      tags$h3("List of Input"),
      selectInput('plotgwda',
                  'Select variable',
                  choices = NULL),
      textInput("maingwda",
                label = "Main title",
                value = "Write main title..."),
      shinyWidgets::pickerInput(
        inputId = "colorgwda",
        label = "Select pallete",
        choices = Pallete_color),
      br(),
      helpText(h4("North arrow position")),
      numericInput("latgwda",
```

```
                "Latitude",
                min = -90,
                max = 90,
                value = 32),
          numericInput(
            "longwda",
            "Longitude",
            min = -180,
            max = 180,
            value = -74),
          numericInput(
            "scalegwda",
            "scale North arrow ",
            min = 1,
            max = 20,
            value = 3),
          shinyWidgets::actionBttn(
            inputId = "Runplotgwda",
            label = "Plot",
            style = "float",
            color = "success"),
          radioButtons("butdowngwda",
                       "Select the Option",
                       choices = list("png","pdf")),
          downloadButton(outputId = "downgwda",
                         label = "Download the plot"),
          style = "unite",
          #icon = icon("gear"),
          status = "primary", width = "300px",
          animate = shinyWidgets::animateOptions(
            enter = animations$fading_entrances$fadeInLeftBig,
            exit = animations$fading_exits$fadeOutRightBig)),
        plotOutput("mapgwda"))))
                )))),
shinydashboard::tabItem(tabName = "tab6",
                        fluidRow(column(width = 12,shinydashboard::box(
                          width = 12,
                          title = "Spatial autocorrelation",
                          status = "primary",
                          solidHeader = TRUE,
                          collapsible = TRUE,
                          height = 66000,
                          sidebarPanel(shinyWidgets::actionBttn(
                            "helpauto",
                            "Help",
                            icon = icon("question-circle"),
                            style = "stretch",
                            block = FALSE,
                            color = "primary"),
                            shinyBS::bsModal(id="helpautoc",
                                             title = "",
                                             trigger = "helpauto",
                                             size="large",
                                             tags$iframe(
                                               src = "Help_Autocorrelation.html",
                                               width = "100%",
                                               height = "1000px",
                                               frameborder = 0,
                                               scrolling = "auto" )),
                                           pickerInput('varauto',
                                            'Variable',
                                            choices = NULL
                                            ),
                          br(),
                          helpText(h4("neighbourhood weights list option")),
                          shinyWidgets::switchInput("Zero.policy",
                                                    inputId = "zeropolicy1",
                                                    onLabel = "TRUE",
                                                    offLabel = "FALSE",
```

```
                    size = "mini"),
  radioButtons("style",
               label = "Style",
               choices = list("W" = "W",
                              "B" = "B",
                              "C" = "C",
                              "U"= "U",
                              "minmax"= "minmax",
                              "S"="S"
               ), selected = "W"),
  br(),
  helpText(h4("Global and local Moran option")),
  numericInput("numsim",
               label = "Number of permutations",
               value = 500),
  radioButtons("alternative",
               label = "alternative",
               choices = list("greater"="greater",
                              "less"= "less",
                              "two.sided"= "two.sided"),
               selected = "greater"),
  shinyalert::useShinyalert(),
shinyjs::useShinyjs(),
shinyWidgets::actionBttn(
  inputId = "Runauto",
  label = "Run",
  style = "float",
  block = TRUE,
  color = "danger")),
mainPanel(tabsetPanel(
  tabPanel(
    "Summary",
    br(),
    verbatimTextOutput("variablename"),
    br(),
    helpText( "Moran I test under randomisation"),
    verbatimTextOutput("moran"),
    helpText( "Monte-Carlo simulation of Moran I"),
    verbatimTextOutput("moranmc"),
    helpText( "Local Moran's I statistic summary"),
    verbatimTextOutput("localmoran"),
    helpText( "Monte-Carlo simulation of Local Moran's I statistic summary"),
    verbatimTextOutput("localmoranperm"),
    br(),
    helpText( "Local Moran's I statistic"),
    DT::DTOutput("DFauto"),
  ),
  tabPanel(
    "Plot",
    helpText("click in Dropdown Button for customize and download plot"),
    shinyWidgets::dropdown(
      tags$h3("List of Input"),
      selectInput('plotauto',
                  'Select',
                  choices = c("lmoran_i" = "lmoran_i",
                              "lmoran_p" = "lmoran_p")),

      shinyWidgets::pickerInput(
        inputId = "colorauto",
        label = "Select pallete",
        choices = Pallete_color),
      br(),
      helpText(h4("North arrow position")),
      numericInput("latauto",
                   "Latitude",
                   min = -90,
                   max = 90,
                   value = 32),
      numericInput(
```

```
                    "longauto",
                    "Longitude",
                    min = -180,
                    max = 180,
                    value = -74),
                  numericInput(
                    "scaleauto",
                    "scale North arrow ",
                    min = 1,
                    max = 20,
                    value = 3),
                  shinyWidgets::actionBttn(
                    inputId = "Runplotauto",
                    label = "Plot",
                    style = "float",
                    color = "success"),
                  radioButtons("butdownauto",
                               "Select the Option",
                               choices = list("png","pdf")),
                  downloadButton(outputId = "downgauto",
                                 label = "Download the plot"),
                  style = "unite",
                  #icon = icon("gear"),
                  status = "primary", width = "300px",
                  animate = shinyWidgets::animateOptions(
                    enter = animations$fading_entrances$fadeInLeftBig,
                    exit = animations$fading_exits$fadeOutRightBig)),
                plotOutput("mapauto"))
              ))

                       )))

            )# fintab6
  )#fintabitemS
 )# findashboardbody

##########################################################################
###########
shinydashboard::dashboardPage(header, sidebar, body)

###ui.R (fin)
```

```
###server.R (inicio)
```

```
data(USelect)
shiny::shinyServer(function(input,output,session) {
  options(shiny.maxRequestSize=2000*1024^2)
  values <- reactiveValues()

  #--------------# Upload example #---------------#
  output$example <- DT::renderDT({
    if(input$example == FALSE){return()}
      data(USelect)
      data <- USelect2004@data
      DT::datatable(data, options = list(
      pageLength = 5,
      scrollCollapse = T,
      scrollX = TRUE,
      fixedColumns = TRUE,
      lengthMenu = c(10, 50, 100, 500)))
    })

  #--------------# Upload data #---------------#
```

```
observe({
 inFile <- input$file1
 if (is.null(inFile)) {
   return (NULL)
 }
 else {#Set up the selection for the worksheet names within the selected file
   filePath <<- inFile$datapath
   selectionWorksheet <- sort( unique(readxl::excel_sheets(inFile$datapath)))
   updateSelectInput(session, "worksheet", choices = selectionWorksheet)
 }
})

observe({
 if(is.null(input$file1)) {
   shinyjs::disable("getData")
 }
 else if(input$example == TRUE) {
   shinyjs::disable("getData")
 }
 else {shinyjs::enable("getData")
 }
})

z <- reactiveValues(dat = NULL,
                    map = NULL,
                    SPDF = NULL,
                    uploaddirectory = NULL,
                    shpdf=NULL )

observeEvent(input$getData, {# Get the data from the spreadsheet worksheets
 dat <- readxl::read_excel (filePath, sheet=input$worksheet)
 z$dat <- as.data.frame(unclass(dat), stringsAsFactors = TRUE)
 z$dat
})

output$table1 <- DT::renderDT({
  req(z$dat)
 DT::datatable(z$dat, options = list(
   pageLength = 5,
   scrollCollapse = T,
   scrollX = TRUE,
   fixedColumns = TRUE,
   lengthMenu = c(10, 50, 100, 500)))
 })
#--------------# Upload shapefile #---------------#
observe({
 if(is.null(input$filemap)) {
   shinyjs::disable("getshape")
 }
 else if(input$example == TRUE) {
   shinyjs::disable("getshape")
 }
 else {shinyjs::enable("getshape")
 }
})

observe({
 z$shpdf <- input$filemap
 if(is.null(z$shpdf)){
   return()
 }
 previouswd <- getwd()
 z$uploaddirectory <- dirname(z$shpdf$datapath[1])
 setwd(z$uploaddirectory)
 for(i in 1:nrow(z$shpdf)){
   file.rename(z$shpdf$datapath[i], z$shpdf$name[i])
 }
 setwd(previouswd)
 })
```

```
  observeEvent(input$getshape,{
     shinybusy::show_modal_spinner(
       spin = "atom",
       color = "#428ded",
       text = "Please wait...Work in progress. You'll have to be
     patient, because the result may take a while"
     )
   map <- raster::shapefile(paste(z$uploaddirectory,
                                  z$shpdf$name[grep(pattern="*.shp$",
                                                    z$shpdf$name)], sep="/"))
   map <- spTransform(map,
                      CRS("+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs"))
   z$map <- map
   shinybusy::remove_modal_spinner()
   beepr:: beep(2)
  })


  observe({
   req(z$dat)
   values$ovarsID <- names(z$dat)
   updateSelectizeInput(session,
                        'colID', 'Select ID for merge data',
                        choices = values$ovarsID,
                        #multiple = TRUE,
                        server = TRUE)
   })
  # select for merge data
  datasel <- reactive({
   req(z$dat)
   datasel <- merge(z$map,z$dat, by = input$colID)
   datasel
   })

  output$shpi <- renderPrint({
   if (is.null(z$map)) {
     return (NULL)
   }
   names(z$map)
  })

#--------------# Distance matrix #---------------#

    observe({
     if(input$example== TRUE | | !is.null(z$map)) {
       shinyjs::enable("Run1")
     }
     else {shinyjs::disable("Run1")}
    })


    observeEvent(input$Run1,{

     shinybusy::show_modal_spinner(
      spin = "atom",
      color = "blue",
      text = "Please wait...Work in progress. You'll have to be
      patient, because the result may take a while")

    if( input$example == TRUE){
      values$dMat <- gw.distGW(dp.locat = sp::coordinates(USelect2004),
                               focus = input$focus,
                               theta = input$theta*pi,
                               p = input$power ,
                               longlat = input$longlat)
    }
    else{
       values$dMat <- gw.distGW(dp.locat = sp::coordinates(datasel()),
```

```
                          focus = input$focus,
                          theta = input$theta*pi,
                          p = input$power ,
                          longlat = input$longlat)
    }
  shinybusy::remove_modal_spinner()
  beepr:: beep(2)

})

output$distmatrix <-  DT::renderDT({
  req(values$dMat)
  datmat <- data.frame(values$dMat)%>%
    dplyr::mutate_if(is.numeric, round, digits = 6)
  DT::datatable(data = datmat, extensions = 'Buttons',
                options = list(dom = 'Bfrtip',
                               scrollX = TRUE,
                               fixedColumns = TRUE,
                               buttons = c('pageLength',
                                           'copy',
                                           'csv',
                                           'excel',
                                           'pdf',
                                           'print'),
                               pagelength = 10,
                               lengthMenu = list(c(10, 25, 100, -1),
                                                 c('10', '25', '100','All'))))
})
############## Variable selection for bw (bandwidth)
#--------- bw.gwr------#
observe({
  if(input$example== TRUE){
  updateSelectizeInput(session,
                       'dependientgwr', 'Dependent',
                       choices =  USelect2004@data %>%
                         dplyr::select_if(is.numeric)%>% names,
                       #multiple = TRUE,
                       server = TRUE
  )
  updateSelectizeInput(session,
                       'independientgwr',
                       'Independient(s)',
                       choices = USelect2004@data %>%
                         dplyr::select_if(is.numeric)%>% names,
                       #multiple = TRUE,
                       server = TRUE)}
})

observe({
  req(z$dat)
  ovars <- z$dat %>% dplyr::select_if(is.numeric) %>% names
  updateSelectizeInput(session,
                       'dependientgwr', 'Dependent',
                       choices = ovars,
                       #multiple = TRUE,
                       server = TRUE
  )
  updateSelectizeInput(session,
                       'independientgwr',
                       'Independient(s)',
                       choices = ovars,
                       #multiple = TRUE,
                       server = TRUE)
})

observe({
  req(z$dat)
  ovars <- z$dat %>% dplyr::select_if(is.numeric) %>% names
  updateSelectizeInput(session,
```

```
                    'dependientggwr', 'Dependent',
                    choices = ovars,
                    #multiple = TRUE,
                    server = TRUE
 )

 updateSelectizeInput(session,
                    'independientggwr',
                    'Independient(s)',
                    choices = ovars,
                    #multiple = TRUE,
                    server = TRUE)
})
#--------- bw.gwda------#
observe({
 if(input$example == TRUE){
   updateSelectizeInput(session,
                    'depbwgda', 'Dependent',
                    choices =  USelect2004@data%>%
                      dplyr::select_if(is.factor) %>% names,
                    #multiple = TRUE,
                    server = TRUE)}
})
observe({
 req(z$dat)
 ovars <- z$dat %>% dplyr::select_if(is.factor) %>% names
 updateSelectizeInput(session,
                    'depbwgda', 'Dependent',
                    choices = ovars,
                    #multiple = TRUE,
                    server = TRUE)
})

observe({
 if(input$example == TRUE){
 updateSelectizeInput(session,
                    'indepbwgda', 'Independient(s)',
                    choices = USelect2004@data%>%
                      dplyr::select_if(is.numeric) %>% names,
                    #multiple = TRUE,
                    server = TRUE)}
})

observe({
 req(z$dat)
 ovars <- z$dat %>% dplyr::select_if(is.numeric) %>% names
 updateSelectizeInput(session,
                    'indepbwgda', 'Independient(s)',
                    choices = ovars,
                    #multiple = TRUE,
                    server = TRUE)
})
#---------bw.gwpca------#
observe({
 if(input$example == TRUE){
  updateSelectizeInput(session,
                   'varpca', 'Variables',
                    choices = USelect2004@data%>%
                     dplyr::select_if(is.numeric) %>% names,
                    #multiple = TRUE,
                    server = TRUE
)}
})
observe({
 if(input$example == FALSE){
   req(z$dat)
   ovarsbw <- z$dat %>% dplyr::select_if(is.numeric) %>% names
   updateSelectizeInput(session,
                      'varpca', 'Variables',
```

```
                    choices = ovarsbw,
                    #multiple = TRUE,
                    server = TRUE)}
  })

  observe({
    updateNumericInput(session,'kpca',
                'k',
                2,
                min = 2,
                max = length(input$varpca)-1)
 })

 datapca <- reactive({
  if(input$example ==FALSE){
    mf <- datasel()[, input$varpca]
    data.scaled <- scale(as.matrix(mf@data[, input$varpca]))
    coords <- sp::coordinates(datasel())
    datapca <- SpatialPointsDataFrame(coords, as.data.frame(data.scaled))
    datapca}
 })

# datapca <- reactive({
  # if(input$example == TRUE){
  # mf <- USelect2004[, input$varpca]
  # data.scaled <- scale(as.matrix(mf@data[, input$varpca]))
   #coords <- sp::coordinates(USelect2004)
  # datapca <- SpatialPointsDataFrame(coords, as.data.frame(data.scaled))
  # datapca}
 # })

 ## bw calculation

 observe({
  if (isTRUE(!is.null(input$independientgwr))) {
    shinyjs::enable("Runbw")
  }
  else if(isTRUE(!is.null(input$indepbwgda))) {
    shinyjs::enable("Runbw")
  }

  else if(isTRUE(!is.null(input$independientggwr))) {
    shinyjs::enable("Runbw")
  }
  else if(isTRUE(!is.null(input$varpca))) {
    shinyjs::enable("Runbw")
  }
  else {shinyjs::disable("Runbw")
  }
 })


observeEvent(input$Runbw, {
 req(values$dMat)
  shinybusy::show_modal_spinner(
    spin = "atom",
    color = "#428ded",
    text = "Please wait...Work in progress. You'll have to be
    patient, because the result may take a while"
  )
  if(input$example== FALSE){
  if(input$bandSelect == "bw.gwr"){
    formu <- as.formula(paste(input$dependientgwr," ~ ",
                        paste(input$independientgwr,collapse="+")))
  values$bw <-  GWmodel::bw.gwr(formula = formu,
                            data = datasel(),
                            approach = input$approachgwr,
                            kernel = input$kernelgwr,
                            adaptive = input$adaptativegwr,
```

```
                        p = input$powergwr,
                        theta = input$thetagwr*pi,
                        longlat= input$longlatgwr,
                        dMat = values$dMat)}

else if(input$bandSelect=="bw.ggwr"){
  formu <- as.formula(paste(input$dependientggwr," ~ ",
                      paste(input$independientggwr,collapse="+")))
values$bw <-  GWmodel::bw.ggwr(formula = formu,
                        data = datasel(),
                        family = input$familyggwr,
                        approach = input$approachggwr,
                        kernel = input$kernelggwr,
                        adaptive = input$adaptativeggwr,
                        p = input$powerggwr,
                        theta = input$thetaggwr*pi,
                        longlat= input$longlatggwr,
                        dMat = values$dMat)}
else if(input$bandSelect == "bw.gwda"){
  formu <- as.formula(paste(input$depbwgda," ~ ",
                      paste(input$indepbwgda,collapse="+")))
values$bw <-  GWmodel::bw.gwda(formula = formu,
                        data = datasel(),
                        COV.gw = input$COVbwgda,
                        prior.gw = input$priorbwgda,
                        mean.gw = input$meanbwgda,
                        prior = NULL,
                        wqda = input$wqdabwgda,
                        kernel = input$kernelbwgda,
                        adaptive = input$adaptativebwgda,
                        p = input$powerbwgda,
                        theta = input$thetabwgda*pi,
                        longlat= input$longlatbwgda,
                        dMat = values$dMat)}
else if(input$bandSelect == "bw.gwpca"){
  values$bw <- GWmodel::bw.gwpca(data = datapca(),
                          vars = colnames(datapca()@data),
                          k = input$kpca,
                          robust = input$robustpca,
                          kernel = input$kernelpca,
                          adaptive = input$adaptativepca,
                          p = input$powerpca,
                          theta = input$thetapca*pi,
                          longlat= input$longlatpca,
                          dMat = values$dMat)}
}
else if(input$example== TRUE){

  if(input$bandSelect == "bw.gwr"){
    formu <- as.formula(paste(input$dependientgwr," ~ ",
                        paste(input$independientgwr,collapse="+")))
  values$bw <-  GWmodel::bw.gwr(formula = formu,
                          data = USelect2004,
                          approach = input$approachgwr,
                          kernel = input$kernelgwr,
                          adaptive = input$adaptativegwr,
                          p = input$powergwr,
                          theta = input$thetagwr*pi,
                          longlat= input$longlatgwr,
                          dMat = values$dMat)}

  else if(input$bandSelect=="bw.ggwr"){
    formu <- as.formula(paste(input$dependientggwr," ~ ",
                        paste(input$independientggwr,collapse="+")))
    values$bw <-  GWmodel::bw.ggwr(formula = formu,
                            data = USelect2004,
                            family = input$familyggwr,
                            approach = input$approachggwr,
                            kernel = input$kernelggwr,
```

```
                              adaptive = input$adaptativeggwr,
                              p = input$powerggwr,
                              theta = input$thetaggwr*pi,
                              longlat= input$longlatggwr,
                              dMat = values$dMat)}
    else if(input$bandSelect == "bw.gwda"){
      formu <- as.formula(paste(input$depbwgda," ~ ",
                          paste(input$indepbwgda,collapse="+")))
    values$bw <-  GWmodel::bw.gwda(formula = formu,
                              data = USelect2004,
                              COV.gw = input$COVbwgda,
                              prior.gw = input$priorbwgda,
                              mean.gw = input$meanbwgda,
                              prior = NULL,
                              wqda = input$wqdabwgda,
                              kernel = input$kernelbwgda,
                              adaptive = input$adaptativebwgda,
                              p = input$powerbwgda,
                              theta = input$thetabwgda*pi,
                              longlat= input$longlatbwgda,
                              dMat = values$dMat)}
    else if(input$bandSelect == "bw.gwpca"){
      values$bw <- GWmodel::bw.gwpca(data = USelect2004,
                              vars = input$varpca,
                              k = input$kpca,
                              robust = input$robustpca,
                              kernel = input$kernelpca,
                              adaptive = input$adaptativepca,
                              p = input$powerpca,
                              theta = input$thetapca*pi,
                              longlat= input$longlatpca,
                              dMat = values$dMat)}
  }

  shinyWidgets::closeSweetAlert(session = session)
  shinyWidgets::sendSweetAlert(
    session = session,
    title =" Calculation completed !",
    type = "success"
  )
  beepr::beep(2)
  shinybusy::remove_modal_spinner()
})


output$selbw <- renderPrint({
  req(values$bw)
  values$bw
})
#--------------# gwss #---------------#
observe({
  if(input$example == TRUE){
  updateSelectizeInput(session,
                       'vargwss', 'Variables',
                       choices = USelect2004@data%>%
                         dplyr::select_if(is.numeric) %>% names,
                       #multiple = TRUE,
                       server = TRUE)}
  })

observe({
  req(z$dat)
  ovars <- names(z$dat)
  updateSelectizeInput(session,
                       'vargwss', 'Variables',
                       choices = ovars,
                       #multiple = TRUE,
                       server = TRUE)
})
```

```
  observe({
   if (isTRUE(is.null(input$vargwss))) {shinyjs::disable("Rungwss") }
   else {shinyjs::enable("Rungwss")}
  })

  observeEvent(input$Rungwss, {
   req(input$vargwss)
   req(values$dMat)
   shinybusy::show_modal_spinner(
    spin = "atom",
    color = "blue",
    text = "Please wait...Work in progress. You'll have to be
    patient, because the result may take a while"
   )
   if(input$example == FALSE){
   if (input$selbw == "Manual"){
    values$gwss <- gwss(data = datasel(),
                        vars = input$vargwss,
                        kernel = input$kernelgwss,
                        adaptive = input$adaptativegwss,
                        p = input$powergwss,
                        theta = input$thetagwss*pi,
                        longlat= input$longlatgwss,
                        bw = input$bwgwss,
                        quantile = input$quantilegwss,
                        dMat = values$dMat)}
   else {values$gwss <- gwss(data = datasel(),
                             vars = input$vargwss,
                             kernel = input$kernelgwss,
                             adaptive = input$adaptativegwss,
                             p = input$powergwss,
                             theta = input$thetagwss*pi,
                             longlat= input$longlatgwss,
                             bw = values$bw,
                             quantile = input$quantilegwss,
                             dMat = values$dMat)}
    }
else if (input$example == TRUE){
  if (input$selbw == "Manual"){
    values$gwss <- gwss(data = USelect2004,
                        vars = input$vargwss,
                        kernel = input$kernelgwss,
                        adaptive = input$adaptativegwss,
                        p = input$powergwss,
                        theta = input$thetagwss*pi,
                        longlat= input$longlatgwss,
                        bw = input$bwgwss,
                        quantile = input$quantilegwss,
                        dMat = values$dMat)}
  else {values$gwss <- gwss(data = USelect2004,
                            vars = input$vargwss,
                            kernel = input$kernelgwss,
                            adaptive = input$adaptativegwss,
                            p = input$powergwss,
                            theta = input$thetagwss*pi,
                            longlat= input$longlatgwss,
                            bw = values$bw,
                            quantile = input$quantilegwss,
                            dMat = values$dMat)}
  }

    shinyWidgets::closeSweetAlert(session = session)
    shinyWidgets::sendSweetAlert(
     session = session,
     title =" Calculation completed !",
     type = "success"
    )
    beepr::beep(2)
```

```
  shinybusy::remove_modal_spinner()
})

output$gwss <- renderPrint({
  req(values$gwss)
  values$gwss
})

output$DFgwss <-  DT::renderDT({
  req(values$gwss)
  datgwss <- data.frame(values$gwss$SDF)%>%
    dplyr::mutate_if(is.numeric, round, digits = 6)
  DT::datatable(data =  datgwss, extensions = 'Buttons',
                options = list(dom = 'Bfrtip',
                               scrollX = TRUE,
                               fixedColumns = TRUE,
                               buttons = c('pageLength',
                                           'copy',
                                           'csv',
                                           'excel',
                                           'pdf',
                                           'print'),
                               pagelength = 10,
                               lengthMenu = list(c(10, 25, 100, -1),
                                                 c('10', '25', '100','All'))))
})

observe({
  req(values$gwss)
  vargwss <- names(data.frame(values$gwss$SDF))
  updateSelectInput(session,'plotgwss',
                    'Select', choices = vargwss)
})

observeEvent(input$Runplotgwss, {
  req(values$gwss$SDF)
  if(input$example == FALSE){
    polys <- list("sp.lines", as(z$map, "SpatialLines"),
                  col="lightgrey", lwd=.5,lty=0.1)
    col.palette <- cartography::carto.pal(input$colorgwss, 20)
    map.na <- list("SpatialPolygonsRescale", layout.north.arrow(),
                   offset = c(input$longwss,input$latgwss),
                   scale = input$scalegwss, col=1)
    values$plotgwss <- sp::spplot(values$gwss$SDF,input$plotgwss,
                                  main = input$maingwss,
                                  sp.layout=list(polys,map.na),
                                  scales=list(cex = 1, col="black"),
                                  col="transparent",
                                  col.regions = col.palette)
    values$plotgwss}
  else if (input$example == TRUE){
    col.palette <- cartography::carto.pal(input$colorgwss,20)
    map.na <- list("SpatialPolygonsRescale", layout.north.arrow(),
                   offset = c(input$longwss,input$latgwss),
                   scale = input$scalegwss, col=1)
    values$plotgwss <- sp::spplot(values$gwss$SDF,input$plotgwss,
                                  main = input$maingwss,
                                  sp.layout=list(map.na),
                                  scales = list(cex = 1, col="black"),
                                  col = "transparent",
                                  col.regions = col.palette)
    values$plotgwss
  }

})


output$mapgwss <- renderPlot({
  values$plotgwss
```

```
  },height = 700, width = 700)

  ########download plot

  # output$down <-downloadHandler()
  observe({
    if (isTRUE(is.null(values$plotgwss))) {
      shinyjs::disable("downgwss")
    }
    else {shinyjs::enable("downgwss")
    }
  })

output$downgwss <- downloadHandler(
    filename =  function() {
      paste("GeoWeithedModel",input$butdowngwss, sep=".")
    },

    content = function(file) {
      if(input$butdowngwss == "png")
        png(file) # open the png device
      else
        pdf(file) # open the pdf device
      print(values$plotgwss) # draw the plot
      dev.off()  # turn the device off

    }
  )

#-----------------------------gwr Collin---------------------------------------
observe({
  if(input$example == TRUE){
    updateSelectizeInput(session,
                         'dependientbgwrd', 'Dependent',
                         choices = USelect2004@data %>%
                           dplyr::select_if(is.numeric)%>% names,
                         #multiple = TRUE,
                         server = TRUE
    )}
})
observe({
  if(input$example == TRUE){
    updateSelectizeInput(session,
                         'independientbgwrd', 'Independient(s)',
                         choices = USelect2004@data %>%
                           dplyr::select_if(is.numeric)%>% names,
                         #multiple = TRUE,
                         server = TRUE)}
})

observe({
  req(z$dat)
  ovars <- z$dat %>% dplyr::select_if(is.numeric) %>% names
  updateSelectizeInput(session,
                       'dependientbgwrd', 'Dependent',
                       choices = ovars,
                       #multiple = TRUE,
                       server = TRUE
  )
})

observe({
  req(z$dat)
  ovars <- z$dat %>% dplyr::select_if(is.numeric) %>% names
  updateSelectizeInput(session,
                       'independientbgwrd', 'Independient(s)',
                       choices = ovars,
                       #multiple = TRUE,
                       server = TRUE)
```

```
})


observe({
  if (isTRUE(is.null(input$independientbgwrd))) {
    shinyjs::disable("Runbgwrd")
  }
  else {shinyjs::enable("Runbgwrd")
  }
})

observeEvent(input$Runbgwrd, {
  req(values$dMat)
  req(input$independientbgwrd)
  shinybusy::show_modal_spinner(
    spin = "atom",
    color = "red",
    text = "Please wait...Work in progress. You'll have to be
      patient, because the result may take a while"
  )
  if(input$example == FALSE){
    formu <- as.formula(paste(input$dependientbgwrd,"~",
                              paste(input$independientbgwrd,collapse="+")))
    if( input$selbgwrd == "Automatic"){
      values$gwr.collin <- GWmodel::gwr.collin.diagno(formula = formu,
                                                      data = datasel(),
                                                      bw = values$bw,
                                                      kernel = input$kernelbgwrd,
                                                      adaptive = input$adaptativebgwrd,
                                                      p = input$powerbgwrd,
                                                      theta = input$thetabgwrd*pi,
                                                      longlat = input$longlatbgwrd,
                                                      dMat = values$dMat
                                                      # parallel.method = "cluster"
      )
    }
    else { values$gwr.collin <- GWmodel::gwr.collin.diagno(
      formula = formu,data = datasel(),
      bw = input$bwbgwrd,
      kernel = input$kernelbgwrd,
      adaptive = input$adaptativebgwrd,
      p = input$powerbgwrd,
      theta = input$thetabgwrd*pi,
      longlat = input$longlatbgwrd,
      dMat = values$dMat
      # parallel.method = "cluster"
    )
    }}
  else if(input$example == TRUE){
    formu <- as.formula(paste(input$dependientbgwrd,"~",
                              paste(input$independientbgwrd,collapse="+")))
    if( input$selbgwrd == "Automatic"){
      values$gwr.collin <- GWmodel::gwr.collin.diagno(formula = formu,
                                                      data = USelect2004,
                                                      bw = values$bw,
                                                      kernel = input$kernelbgwrd,
                                                      adaptive = input$adaptativebgwrd,
                                                      p = input$powerbgwrd,
                                                      theta = input$thetabgwrd*pi,
                                                      longlat = input$longlatbgwrd,
                                                      dMat = values$dMat
                                                      # parallel.method = "cluster"
      )
    }
    else if (input$selbgwrd == "Manual"){
      values$gwr.collin <- GWmodel::gwr.collin.diagno(
      formula = formu,
      data = USelect2004,
      bw = input$bwbgwrd,
```

```
      kernel = input$kernelbgwrd,
      adaptive = input$adaptativebgwrd,
      p = input$powerbgwrd,
      theta = input$thetabgwrd*pi,
      longlat = input$longlatbgwrd,
      dMat = values$dMat
      )}
    }

  shinyWidgets::closeSweetAlert(session = session)
  shinyWidgets::sendSweetAlert(
    session = session,
    title =" Calculation completed !",
    type = "success"
  )
  beepr::beep(2)
  shinybusy::remove_modal_spinner()
})


#DT::DTOutput("gwrbasicdlocalCN"),
#DT::DTOutput("gwrbasicdlocalVDP"),
#DT::DTOutput("gwrbasicdlocalcorr.mat")

output$gwrbasicdVIF <- DT::renderDT({
  req(values$gwr.collin)

  datgwr <- data.frame(values$gwr.collin$SDF)
  datgwr1 <- datgwr %>% dplyr:: select(!starts_with("Corr_Intercept"))
  #datgwr2 <- datgwr %>% dplyr:: select(ends_with("_VDP") | starts_with("SDF.Corr_"))
  #datgwrd <- merge(datgwr1,datgwr2)
  datgwrd <- datgwr1 %>% mutate_if(is.numeric,round, digits = 6)

  DT::datatable(data =  datgwrd, extensions = 'Buttons',
                options = list(dom = 'Bfrtip',
                               scrollX = TRUE,
                               fixedColumns = TRUE,
                               buttons = c('pageLength',
                                           'copy',
                                           'csv',
                                           'excel',
                                           'pdf',
                                           'print'),
                               pagelength = 10,
                               lengthMenu = list(c(10, 25, 100, -1),
                                                 c('10', '25', '100','All'))))
})


observe({
  req(values$gwr.collin)
  datgwr <- data.frame(values$gwr.collin$SDF)
  datgwr1 <- datgwr %>% dplyr:: select(!starts_with("Corr_Intercept"))
  #datgwr2 <- datgwr %>% dplyr:: select(ends_with("_VDP") | starts_with("SDF.Corr_"))
  #datgwrd <- merge(datgwr1,datgwr2)
  datgwrd <- datgwr1 %>% mutate_if(is.numeric,
            round,
            digits = 6)

  vargwsd <- names(datgwrd)

  updateSelectInput(session,'plotgwrd',
                    'Select', choices = vargwsd)
})

observeEvent(input$Runplotgwrd, {
  if(input$example == FALSE){
    req(values$gwr.collin$SDF)
    polys <- list("sp.lines", as(z$map, "SpatialLines"), col="lightgrey",
```

```
                lwd=.5,lty=0.1)
    map.na <- list("SpatialPolygonsRescale", layout.north.arrow(),
                offset = c(input$longwrd,input$latgwrd),
                scale = input$scalegwrd, col=1)
    col.palette <- cartography::carto.pal(input$colorgwrd, 20)
    values$plotgwrd <-sp::spplot(values$gwr.collin$SDF,input$plotgwrd,
                              main = input$maingwrd,
                              sp.layout=list(polys,map.na),
                              scales=list(cex = 1, col="black"),
                              col="transparent",
                              col.regions = col.palette)
    values$plotgwrd}
  else if(input$example == TRUE){
    map.na <- list("SpatialPolygonsRescale", layout.north.arrow(),
                offset = c(input$longwrd,input$latgwrd),
                scale = input$scalegwrd, col=1)
    col.palette <- cartography::carto.pal(input$colorgwrd, 20)
    values$plotgwrd <-sp::spplot(values$gwr.collin$SDF,input$plotgwrd,
                              main = input$maingwrd,
                              sp.layout=list(map.na),
                              scales=list(cex = 1, col="black"),
                              col="transparent",
                              col.regions = col.palette)
    values$plotgwrd
  }

})

observe({
  if (isTRUE(is.null(values$plotgwrd))) {
    shinyjs::disable("downgwrd")
  }
  else {shinyjs::enable("downgwrd")
  }
})

output$mapgwrbasicd <- renderPlot({
  values$plotgwrd
},height = 700, width = 700)


output$downgwrd <- downloadHandler(
  filename =  function() {
    paste("GeoWeithedModel",input$butdowngwrd, sep=".")
  },

  content = function(file) {
    if(input$butdowngwrd == "png")
      png(file) # open the png device
    else
      pdf(file) # open the pdf device
    print(values$plotgwrd) # draw the plot
    dev.off()  # turn the device off

  }
)



  #-----------------------------gwr.basic--------------------------------------

observe({
 if(input$example == TRUE){
  updateSelectizeInput(session,
                       'dependientbgwr', 'Dependent',
                       choices = USelect2004@data %>%
                         dplyr::select_if(is.numeric)%>% names,
                       #multiple = TRUE,
                       server = TRUE
```

```
  )}
  })
observe({
  if(input$example == TRUE){
  updateSelectizeInput(session,
                       'independientbgwr', 'Independient(s)',
                       choices = USelect2004@data %>%
                         dplyr::select_if(is.numeric)%>% names,
                       #multiple = TRUE,
                       server = TRUE)}
})

  observe({
    req(z$dat)
    ovars <- z$dat %>% dplyr::select_if(is.numeric) %>% names
    updateSelectizeInput(session,
                         'dependientbgwr', 'Dependent',
                         choices = ovars,
                         #multiple = TRUE,
                         server = TRUE
    )
  })

  observe({
    req(z$dat)
    ovars <- z$dat %>% dplyr::select_if(is.numeric) %>% names
    updateSelectizeInput(session,
                         'independientbgwr', 'Independient(s)',
                         choices = ovars,
                         #multiple = TRUE,
                         server = TRUE)
  })

  observe({
    if (isTRUE(is.null(input$independientbgwr))) {
      shinyjs::disable("Runbgwr")
    }
    else {shinyjs::enable("Runbgwr")
    }
  })
  observeEvent(input$Runbgwr, {
    req(values$dMat)
    req(input$independientbgwr)
    shinybusy::show_modal_spinner(
      spin = "atom",
      color = "red",
      text = "Please wait...Work in progress. You'll have to be
      patient, because the result may take a while"
    )
    if(input$example == FALSE){
      formu <- as.formula(paste(input$dependientbgwr,"~",
                                paste(input$independientbgwr,collapse="+")))
      if( input$selbgwr == "Automatic"){
      values$gwr.basic <- GWmodel::gwr.basic(formula = formu,
                                             data = datasel(),
                                             bw = values$bw,
                                             kernel = input$kernelbgwr,
                                             adaptive = input$adaptativebgwr,
                                             cv = input$cvbgwr,
                                             p = input$powerbgwr,
                                             theta = input$thetabgwr*pi,
                                             longlat = input$longlatbgwr,
                                             dMat = values$dMat
                                             # parallel.method = "cluster"
      )
    }
    else { values$gwr.basic <- GWmodel::gwr.basic(
      formula = formu,data = datasel(),
      bw = input$bwbgwr,
```

```
    kernel = input$kernelbgwr,
    adaptive = input$adaptativebgwr,
    cv = input$cvbgwr,
    p = input$powerbgwr,
    theta = input$thetabgwr*pi,
    longlat = input$longlatbgwr,
    dMat = values$dMat
    # parallel.method = "cluster"
  )
  }}
  else if(input$example == TRUE){
    formu <- as.formula(paste(input$dependientbgwr,"~",
                              paste(input$independientbgwr,collapse="+")))
    if( input$selbgwr == "Automatic"){
      values$gwr.basic <- GWmodel::gwr.basic(formula = formu,
                                             data = USelect2004,
                                             bw = values$bw,
                                             kernel = input$kernelbgwr,
                                             adaptive = input$adaptativebgwr,
                                             cv = input$cvbgwr,
                                             p = input$powerbgwr,
                                             theta = input$thetabgwr*pi,
                                             longlat = input$longlatbgwr,
                                             dMat = values$dMat
                                             # parallel.method = "cluster"
      )
    }
    else { values$gwr.basic <- GWmodel::gwr.basic(
      formula = formu,
      data = USelect2004,
      bw = input$bwbgwr,
      kernel = input$kernelbgwr,
      adaptive = input$adaptativebgwr,
      cv = input$cvbgwr,
      p = input$powerbgwr,
      theta = input$thetabgwr*pi,
      longlat = input$longlatbgwr,
      dMat = values$dMat
      # parallel.method = "cluster"
    )
    }
  }

  shinyWidgets::closeSweetAlert(session = session)
  shinyWidgets::sendSweetAlert(
    session = session,
    title =" Calculation completed !",
    type = "success"
  )
  beepr::beep(2)
  shinybusy::remove_modal_spinner()
})

output$gwrbasic <- renderPrint({
  req(values$gwr.basic)
  values$gwr.basic
})

output$SDFgwrbasic <- DT::renderDT({
  #req(values$gwr.basic)
  datgwr <- data.frame(values$gwr.basic$SDF)%>%
    mutate_if(is.numeric,
              round,
              digits = 6)
  DT::datatable(data =  datgwr, extensions = 'Buttons',
                options = list(dom = 'Bfrtip',
                               scrollX = TRUE,
                               fixedColumns = TRUE,
                               buttons = c('pageLength',
```

```
                                'copy',
                                'csv',
                                'excel',
                                'pdf',
                                'print'),
                    pagelength = 10,
                    lengthMenu = list(c(10, 25, 100, -1),
                                      c('10', '25', '100','All'))))
})

observe({
  req(values$gwr.basic)
  vargws <- names(data.frame(values$gwr.basic$SDF))

  updateSelectInput(session,'plotgwr',
                    'Select', choices = vargws)
})


observeEvent(input$Runplotgwr, {
  if(input$example == FALSE){
    req(values$gwr.basic$SDF)
    polys <- list("sp.lines", as(z$map, "SpatialLines"), col="lightgrey",
                  lwd=.5,lty=0.1)
    map.na <- list("SpatialPolygonsRescale", layout.north.arrow(),
                   offset = c(input$longwr,input$latgwr),
                   scale = input$scalegwr, col=1)
    col.palette <- cartography::carto.pal(input$colorgwr, 20)
    values$plotgwr <-sp::spplot(values$gwr.basic$SDF,input$plotgwr,
                        main = input$maingwr,
                        sp.layout=list(polys,map.na),
                        scales=list(cex = 1, col="black"),
                        col="transparent",
                        col.regions = col.palette)
    values$plotgwr}
  else if(input$example == TRUE){
    map.na <- list("SpatialPolygonsRescale", layout.north.arrow(),
                   offset = c(input$longwr,input$latgwr),
                   scale = input$scalegwr, col=1)
    col.palette <- cartography::carto.pal(input$colorgwr, 20)
    values$plotgwr <-sp::spplot(values$gwr.basic$SDF,input$plotgwr,
                        main = input$maingwr,
                        sp.layout=list(map.na),
                        scales=list(cex = 1, col="black"),
                        col="transparent",
                        col.regions = col.palette)
    values$plotgwr
  }

})

observe({
  if (isTRUE(is.null(values$plotgwr))) {
    shinyjs::disable("downgwr")
  }
  else {shinyjs::enable("downgwr")
  }
})

output$mapgwrbasic <- renderPlot({
  values$plotgwr
},height = 700, width = 700)


output$downgwr <- downloadHandler(
  filename =  function() {
    paste("GeoWeithedModel",input$butdowngwr, sep=".")
  },
```

```
  content = function(file) {
    if(input$butdowngwr == "png")
      png(file) # open the png device
    else
      pdf(file) # open the pdf device
    print(values$plotgwr) # draw the plot
    dev.off()  # turn the device off

  }
)
#-----------------------------generalized ggwr.basic-------------------------

observe({
  req(z$dat)
  ovars <- z$dat %>% dplyr::select_if(is.numeric) %>% names
  updateSelectizeInput(session,
                       'dependientGGWR', 'Dependent',
                       choices = ovars,
                       #multiple = TRUE,
                       server = TRUE
  )
})
observe({
  req(z$dat)
  ovars <- z$dat %>% dplyr::select_if(is.numeric) %>% names
  updateSelectizeInput(session,
                       'independientGGWR', 'Independient(s)',
                       choices = ovars,
                       #multiple = TRUE,
                       server = TRUE)
})

observe({
  if (isTRUE(is.null(input$independientGGWR))) {
    shinyjs::disable("RunGGWR")
  }
  else {shinyjs::enable("RunGGWR")
  }
})
observeEvent(input$RunGGWR, {
  req(values$dMat)
  shinybusy::show_modal_spinner(
    spin = "atom",
    color = "red",
    text = "Please wait...Work in progress. You'll have to be
    patient, because the result may take a while"
  )
  formu <- as.formula(paste(input$dependientGGWR," ~ ",
                            paste(input$independientGGWR,collapse="+")))

  if( input$selGGWR == "Automatic"){
    values$ggwr.basic <- GWmodel::ggwr.basic(formula = formu,
                                             data = datasel(),
                                             bw = values$bw,
                                             kernel = input$kernelGGWR,
                                             adaptive = input$adaptativeGGWR,
                                             cv = input$cvGGWR,
                                             tol = 1e-05,
                                             maxiter = input$maxiterGGWR,
                                             p = input$powerGGWR,
                                             theta = input$thetaGGWR*pi,
                                             longlat = input$longlatGGWR,
                                             dMat = values$dMat
                                             # parallel.method = "cluster"
    )
  }
  else { values$ggwr.basic <- GWmodel::ggwr.basic(
    formula = formu,
    data = datasel(),
```

```
    bw = input$bwGGWR,
    family = input$familyGGWR,
    kernel = input$kernelGGWR,
    adaptive = input$adaptativeGGWR,
    cv = input$cvGGWR,
    tol = 1e-05,
    maxiter = input$maxiterGGWR,
    p = input$powerGGWR,
    theta = input$thetaGGWR*pi,
    longlat = input$longlatGGWR,
    dMat = values$dMat
    # parallel.method = "cluster"
  )
  }
  shinyWidgets::closeSweetAlert(session = session)
  shinyWidgets::sendSweetAlert(
    session = session,
    title =" Calculation completed !",
    type = "success"
  )
  beepr::beep(2)
  shinybusy::remove_modal_spinner()
})

output$GGWRbasic <- renderPrint({
  req(values$ggwr.basic)
  values$ggwr.basic
})
output$SDFGGWRbasic <- DT::renderDT({
  req(values$ggwr.basic)
  datgwr <- data.frame(values$ggwr.basic$SDF)%>%
    dplyr::mutate_if(is.numeric, round, digits = 6)
  DT::datatable(data =  datgwr, extensions = 'Buttons',
                options = list(dom = 'Bfrtip',
                               scrollX = TRUE,
                               fixedColumns = TRUE,
                               buttons = c('pageLength',
                                           'copy',
                                           'csv',
                                           'excel',
                                           'pdf',
                                           'print'),
                               pagelength = 10,
                               lengthMenu = list(c(10, 25, 100, -1),
                                                 c('10', '25', '100','All'))))
})

#####plot ggwr.basic
observe({
  req(values$ggwr.basic)
  vargws <- names(data.frame(values$ggwr.basic$SDF))

  updateSelectInput(session,'plotggwr',
                    'Select', choices = vargws)
})


observeEvent(input$Runplotggwr, {
  req(values$ggwr.basic$SDF)
  polys <- list("sp.lines", as(z$map, "SpatialLines"), col="lightgrey",
                lwd=.5,lty=0.1)
  col.palette <- cartography::carto.pal(input$colorggwr, 20)
  map.na <- list("SpatialPolygonsRescale", layout.north.arrow(),
                 offset = c(input$longgwr,input$latggwr), scale = 1, col=1)
  values$plotggwr <-sp::spplot(values$ggwr.basic$SDF,input$plotggwr,
                               main = input$mainggwr,
                               sp.layout=list(polys,map.na),
                               scales=list(cex = 1, col="black"),
                               col="transparent",
```

```
                        col.regions = col.palette)
  values$plotggwr

})
output$mapggwr <- renderPlot({
  values$plotggwr
},height = 700, width = 700)

observe({
  if (isTRUE(is.null(values$plotggwr))) {
    shinyjs::disable("downggwr")
  }
  else {shinyjs::enable("downggwr")
  }
})


output$downggwr <- downloadHandler(
  filename =  function() {
    paste("GeoWeithedModel",input$butdownggwr, sep=".")
  },
  #
  content = function(file) {
    if(input$butdownggwr == "png")
      png(file) # open the png device
    else
      pdf(file) # open the pdf device
    print(values$plotggwr) # draw the plot
    dev.off()  # turn the device off

  }
)

#-----------------------------gwr.robust--------------------------------------
observe({
  if(input$example == TRUE){
    updateSelectizeInput(session,
                         'dependientRobust', 'Dependent',
                         choices = USelect2004@data %>%
                           dplyr::select_if(is.numeric)%>% names,
                         #multiple = TRUE,
                         server = TRUE)}
  })

observe({
  if(input$example == TRUE){
    updateSelectizeInput(session,
                          'independientRobust', 'Independient(s)',
                          choices = USelect2004@data %>%
                            dplyr::select_if(is.numeric)%>% names,
                          #multiple = TRUE,
                          server = TRUE)
    }
})
observe({
  req(z$dat)
  ovars <- z$dat %>% dplyr::select_if(is.numeric) %>% names
  updateSelectizeInput(session,
                         'dependientRobust', 'Dependent',
                         choices = ovars,
                         #multiple = TRUE,
                         server = TRUE
  )
})
observe({
  req(z$dat)
  ovars <- z$dat %>% dplyr::select_if(is.numeric) %>% names
  updateSelectizeInput(session,
                         'independientRobust', 'Independient(s)',
```

```
                choices = ovars,
                #multiple = TRUE,
                server = TRUE)
})

observe({
 if (isTRUE(is.null(input$independientRobust))) {
   shinyjs::disable("RunRobust")
 }
 else {shinyjs::enable("RunRobust")
 }
})

observeEvent(input$RunRobust, {
 req(values$dMat)
 shinybusy::show_modal_spinner(
   spin = "atom",
   color = "red",
   text = "Please wait...Work in progress. You'll have to be
   patient, because the result may take a while"
 )
 if (input$example == FALSE){formu <- as.formula(paste(input$dependientRobust," ~ ",
                    paste(input$independientRobust,collapse="+")))
 if( input$selRobust == "Automatic"){values$gwr.robust <- GWmodel::gwr.robust(formula = formu,
                                                    data = datasel(),
                          bw = values$bw,
                          kernel = input$kernelRobust,
                          adaptive = input$adaptativeRobust,
                          delta  = 1e-05,
                          filtered = input$filtered,
                          maxiter = input$maxiterRobust,
                          #F123.test = input$f123test,
                          p = input$powerRobust,
                          theta = input$thetaRobust*pi,
                          longlat = input$longlatRobust,
                          dMat = values$dMat)}
 else { values$gwr.robust <- GWmodel::gwr.robust(formula = formu,
                            data = datasel(),
                            bw = input$bwRobust,
                            kernel = input$kernelRobust,
                            adaptive = input$adaptativeRobust,
                            #F123.test = input$f123test,
                            delta  = 1e-05,
                            filtered = input$filtered,
                            maxiter = input$maxiterRobust,
                            p = input$powerRobust,
                            theta = input$thetaRobust*pi,
                            longlat = input$longlatRobust,
                            dMat = values$dMat )}
 }
 else if(input$example == TRUE){
  formu <- as.formula(paste(input$dependientRobust," ~ ",
                     paste(input$independientRobust,collapse="+")))
  if(input$selRobust == "Automatic"){
    values$gwr.robust <- GWmodel::gwr.robust(
      formula = formu,
      data = USelect2004,
      bw = values$bw,
      kernel = input$kernelRobust,
      adaptive = input$adaptativeRobust,
      delta  = 1e-05,
      #F123.test = input$f123test,
      filtered = input$filtered,
      maxiter = input$maxiterRobust,
      p = input$powerRobust,
      theta = input$thetaRobust*pi,
      longlat = input$longlatRobust,
      dMat = values$dMat
```

```
    )}
  else { values$gwr.robust <- GWmodel::gwr.robust(
    formula = formu,
    data = USelect2004,
    bw = input$bwRobust,
    kernel = input$kernelRobust,
    adaptive = input$adaptativeRobust,
    delta  = 1e-05,
    filtered = input$filtered,
    maxiter = input$maxiterRobust,
    #F123.test = input$f123test,
    p = input$powerRobust,
    theta = input$thetaRobust*pi,
    longlat = input$longlatRobust,
    dMat = values$dMat
   )}
 }

 shinyWidgets::closeSweetAlert(session = session)
 shinyWidgets::sendSweetAlert(
   session = session,
   title =" Calculation completed !",
   type = "success"
 )
 beepr::beep(2)
 shinybusy::remove_modal_spinner()
})

output$GGWRobust <- renderPrint({
 req(values$gwr.robust)
 values$gwr.robust
})
output$SDFGWRRobust <- DT::renderDT({
 req(values$gwr.robust)
 datgwr <- data.frame(values$gwr.robust$SDF)%>%
   mutate_if(is.numeric, round, digits = 6)
 DT::datatable(data =  datgwr, extensions = 'Buttons',
           options = list(dom = 'Bfrtip',
                          scrollX = TRUE,
                          fixedColumns = TRUE,
                          buttons = c('pageLength',
                                      'copy',
                                      'csv',
                                      'excel',
                                      'pdf',
                                      'print'),
                          pagelength = 10,
                          lengthMenu = list(c(10, 25, 100, -1),
                                            c('10', '25', '100','All'))))
})

## plot gwr robust
observe({
 req(values$gwr.robust)
 vargws <- names(data.frame(values$gwr.robust$SDF))

 updateSelectInput(session,'plotRgwr',
                   'Select', choices = vargws)
})


observeEvent(input$RunplotRgwr, {
 if(input$example == FALSE){
 req(values$gwr.robust$SDF)
 polys <- list("sp.lines", as(z$map, "SpatialLines"),
           col="lightgrey", lwd=.5,lty=0.1)
 col.palette <- cartography::carto.pal(input$colorRgwr, 20)
 map.na <- list("SpatialPolygonsRescale", layout.north.arrow(),
           offset = c(input$lonRgwr,input$latRgwr),
```

```
            scale = input$scaleRgwr, col=1)
  values$plotRgwr <-sp::spplot(values$gwr.robust$SDF,input$plotRgwr,
                       main = input$mainRgwr,
                       sp.layout=list(polys, map.na),
                       scales=list(cex = 1, col="black"),
                       col="transparent",
                       col.regions = col.palette)
  values$plotRgwr}
  else if(input$example == TRUE){
    col.palette <- cartography::carto.pal(input$colorRgwr, 20)
    map.na <- list("SpatialPolygonsRescale", layout.north.arrow(),
                offset = c(input$lonRgwr,input$latRgwr),
                scale = input$scaleRgwr, col=1)
    values$plotRgwr <-sp::spplot(values$gwr.robust$SDF,input$plotRgwr,
                         main = input$mainRgwr,
                         sp.layout=list(map.na),
                         scales=list(cex = 1, col="black"),
                         col="transparent",
                         col.regions = col.palette)
    values$plotRgwr
  }
  })

output$mapRgwr <- renderPlot({
  values$plotRgwr
},height = 700, width = 700)

observe({
  if (isTRUE(is.null(values$plotRgwr))) {
    shinyjs::disable("downRgwr")
  }
  else {shinyjs::enable("downRgwr")
  }
})

output$downRgwr <- downloadHandler(
  filename =  function() {
    paste("GeoWeithedModel",input$butdownRgwr, sep=".")
  },# content is a function with argument file. content writes the plot to
  #the device
  content = function(file) {
    if(input$butdownRgwr == "png")
      png(file) # open the png device
    else
      pdf(file) # open the pdf device
    print(values$plotRgwr) # draw the plot
    dev.off()  # turn the device off
    }
)
#------------------------------gwr.hetero--------------------------------------
observe({
  if(input$example == TRUE){
    updateSelectizeInput(session,
                         'dependienthetero', 'Dependent',
                         choices = USelect2004@data %>%
                           dplyr::select_if(is.numeric)%>% names,
                         #multiple = TRUE,
                         server = TRUE
  )}
})
observe({
  if(input$example == TRUE){
  updateSelectizeInput(session,
                       'independienthetero', 'Independient(s)',
                       choices = USelect2004@data %>%
                         dplyr::select_if(is.numeric)%>% names,
                       #multiple = TRUE,
                       server = TRUE)}
})
```

```
observe({
 req(z$dat)
 ovars <- z$dat %>% dplyr::select_if(is.numeric) %>% names
 updateSelectizeInput(session,
                      'dependienthetero', 'Dependent',
                      choices = ovars,
                      #multiple = TRUE,
                      server = TRUE
 )
})
observe({
 req(z$dat)
 ovars <- z$dat %>% dplyr::select_if(is.numeric) %>% names
 updateSelectizeInput(session,
                      'independienthetero', 'Independient(s)',
                      choices = ovars,
                      #multiple = TRUE,
                      server = TRUE)
})

observe({
 if (isTRUE(is.null(input$independienthetero))) {
   shinyjs::disable("Runhetero")
 }
 else {shinyjs::enable("Runhetero")
 }
})
observeEvent(input$Runhetero, {
 req(values$dMat)
 shinybusy::show_modal_spinner(
   spin = "atom",
   color = "red",
   text = "Please wait...Work in progress. You'll have to be
   patient, because the result may take a while"
 )
 if(input$example == FALSE){
 formu <- as.formula(paste(input$dependienthetero," ~ ",
                           paste(input$independienthetero,collapse="+")))

 if( input$selhetero == "Automatic"){
   values$gwr.hetero <- GWmodel::gwr.hetero(
     formula = formu,
     data = datasel(),
     bw = values$bw,
     kernel = input$kernelhetero,
     adaptive = input$adaptativehetero,
     tol=0.0001,
     maxiter = input$maxiterhetero,
     p = input$powerhetero,
     theta = input$thetahetero*pi,
     longlat = input$longlathetero,
     dMat = values$dMat

   )
 }
 else { values$gwr.hetero <- GWmodel::gwr.hetero(
   formula = formu,
   data = datasel(),
   bw = input$bwhetero,
   kernel = input$kernelhetero,
   adaptive = input$adaptativehetero,
   tol= 0.0001,
   maxiter = input$maxiterhetero,
   p = input$powerhetero,
   theta = input$thetahetero*pi,
   longlat = input$longlathetero,
   dMat = values$dMat
 )
```

```
  }
  }
  else if(input$example == TRUE){
    formu <- as.formula(paste(input$dependienthetero," ~ ",
                              paste(input$independienthetero,collapse="+")))
    if( input$selhetero == "Automatic"){
      values$gwr.hetero <- GWmodel::gwr.hetero(
        formula = formu,
        data = USelect2004,
        bw = values$bw,
        kernel = input$kernelhetero,
        adaptive = input$adaptativehetero,
        tol=0.0001,
        maxiter = input$maxiterhetero,
        p = input$powerhetero,
        theta = input$thetahetero*pi,
        longlat = input$longlathetero,
        dMat = values$dMat

      )
    }
    else { values$gwr.hetero <- GWmodel::gwr.hetero(
      formula = formu,
      data = USelect2004,
      bw = input$bwhetero,
      kernel = input$kernelhetero,
      adaptive = input$adaptativehetero,
      tol= 0.0001,
      maxiter = input$maxiterhetero,
      p = input$powerhetero,
      theta = input$thetahetero*pi,
      longlat = input$longlathetero,
      dMat = values$dMat
    )
    }


  }

  shinyWidgets::closeSweetAlert(session = session)
  shinyWidgets::sendSweetAlert(
    session = session,
    title =" Calculation completed !",
    type = "success"
  )
  beepr::beep(2)
  shinybusy::remove_modal_spinner()
})

output$GWRhetero <- renderPrint({
  req(values$gwr.hetero)
  values$gwr.hetero
})
output$SDFGWRhetero <- DT::renderDT({
  req(values$gwr.hetero)
  datgwr <- data.frame(values$gwr.hetero)%>%
    mutate_if(is.numeric, round, digits = 6)
  DT::datatable(data =  datgwr, extensions = 'Buttons',
                options = list(dom = 'Bfrtip',
                               scrollX = TRUE,
                               fixedColumns = TRUE,
                               buttons = c('pageLength',
                                           'copy',
                                           'csv',
                                           'excel',
                                           'pdf',
                                           'print'),
                               pagelength = 10,
                               lengthMenu = list(c(10, 25, 100, -1),
```

```
                                    c('10', '25', '100','All'))))
})

## plot gwr hetero
observe({
  req(values$gwr.hetero)
  vargws <- names(data.frame(values$gwr.hetero))

  updateSelectInput(session,'plotHgwr',
                    'Select', choices = vargws)
})

observeEvent(input$RunplotHgwr, {
  if(input$example== FALSE){
  req(values$gwr.hetero)
  polys <- list("sp.lines", as(z$map, "SpatialLines"), col="lightgrey",
                lwd=.5,lty=0.1)
  map.na <- list("SpatialPolygonsRescale", layout.north.arrow(),
                 offset = c(input$lonHgwr,input$latHgwr), scale = 1, col=1)
  col.palette <- cartography::carto.pal(input$colorHgwr, 20)
  values$plotHgwr <-sp::spplot(values$gwr.hetero,input$plotHgwr,
                           main = input$mainHgwr,
                           sp.layout=list(polys,map.na),
                           scales=list(cex = 1, col="black"),
                           col="transparent",
                           col.regions = col.palette)
  values$plotHgwr}
  else if(input$example == TRUE){
    map.na <- list("SpatialPolygonsRescale", layout.north.arrow(),
                   offset = c(input$lonHgwr,input$latHgwr),
                   scale = input$scaleHgwr, col=1)
    col.palette <- cartography::carto.pal(input$colorHgwr, 20)
    values$plotHgwr <-sp::spplot(values$gwr.hetero,input$plotHgwr,
                             main = input$mainHgwr,
                             sp.layout=list(map.na),
                             scales=list(cex = 1, col="black"),
                             col="transparent",
                             col.regions = col.palette)
    values$plotHgwr

  }


})

output$mapHgwr <- renderPlot({
  values$plotHgwr
},height = 700, width = 700)

observe({
  if (isTRUE(is.null(values$plotHgwr))) {
    shinyjs::disable("downHgwr")
  }
  else {shinyjs::enable("downHgwr")
  }
})
output$downHgwr <- downloadHandler(
  filename =  function() {
    paste("GeoWeithedModel",input$butdownHgwr, sep=".")
  },
  content = function(file) {
    if(input$butdownHgwr == "png")
      png(file) # open the png device
    else
      pdf(file) # open the pdf device
    print(values$plotHgwr) # draw the plot
    dev.off()  # turn the device off

  }
```

```
)

#----------------------------gwr.mixed--------------------------------------
observe({
 if(input$example == TRUE){
 updateSelectizeInput(session,
                      'dependientmixed', 'Dependent',
                      choices = USelect2004@data %>%
                        dplyr::select_if(is.numeric)%>% names,
                      #multiple = TRUE,
                      server = TRUE
 )}
})

observe({
 if(input$example == TRUE){
   updateSelectizeInput(session,
                      'independientmixed', 'Independient(s)',
                      choices = USelect2004@data %>%
                        dplyr::select_if(is.numeric)%>% names,
                      #multiple = TRUE,
                      server = TRUE)}
})
observe({
 if(input$example == TRUE){
   updateSelectizeInput(session,
                      'fixedvars', 'fixed var(s)',
                      choices = USelect2004@data %>%
                        dplyr::select_if(is.numeric)%>% names,
                      #multiple = TRUE,
                      server = TRUE)}
})

observe({
 req(z$dat)
 ovars <- z$dat %>% dplyr::select_if(is.numeric) %>% names
 updateSelectizeInput(session,
                      'dependientmixed', 'Dependent',
                      choices = ovars,
                      #multiple = TRUE,
                      server = TRUE

 )
})
observe({
 req(z$dat)
 ovars <- names(z$dat)
 updateSelectizeInput(session,
                      'independientmixed', 'Independient(s)',
                      choices = ovars,
                      #multiple = TRUE,
                      server = TRUE)
})

observe({
 req(z$dat)
 ovars <- c(input$independientmixed)
 updateSelectizeInput(session,
                      'fixedvars', 'fixed var',
                      choices = ovars,
                      #multiple = TRUE,
                      server = TRUE,
                      options = list(maxItems = 1))
})

observe({
 if (isTRUE(is.null(input$independientmixed))) {
   shinyjs::disable("Runmixed")
 }
```

```
 else {shinyjs::enable("Runmixed")
 }
})

observeEvent(input$Runmixed, {
 req(values$dMat)
 shinybusy::show_modal_spinner(
  spin = "atom",
  color = "red",
  text = "Please wait...Work in progress. You'll have to be
  patient, because the result may take a while"
 )
 if(input$example == FALSE){
 formu <- as.formula(paste(input$dependientmixed," ~ ",
                           paste(input$independientmixed,collapse="+")))

 if( input$selmixed == "Automatic"){
  values$gwr.mixed <- GWmodel::gwr.mixed(
   formula = formu,
   data = datasel(),
   bw = values$bw,
   fixed.vars = c(input$fixedvars),
   intercept.fixed= input$intercepfixed,
   diagnostic = input$diagnostic,
   kernel = input$kernelmixed,
   adaptive = input$adaptativemixed,
   p = input$powermixed,
   theta = input$thetamixed*pi,
   longlat = input$longlatmixed,
   dMat = values$dMat

  )
 }
 else { values$gwr.mixed <- GWmodel::gwr.mixed(
  formula = formu,
  data = datasel(),
  bw = input$bwmixed,
  fixed.vars = c(input$fixedvars),
  intercept.fixed= input$intercepfixed,
  diagnostic = input$diagnostic,
  kernel = input$kernelmixed,
  adaptive = input$adaptativemixed,
  p = input$powermixed,
  theta = input$thetamixed*pi,
  longlat = input$longlatmixed,
  dMat = values$dMat
 )
 }}
 else if(input$example == TRUE){
  formu <- as.formula(paste(input$dependientmixed," ~ ",
                            paste(input$independientmixed,collapse="+")))

  if( input$selmixed == "Automatic"){
   values$gwr.mixed <- GWmodel::gwr.mixed(
    formula = formu,
    data = USelect2004,
    bw = values$bw,
    fixed.vars = c(input$fixedvars),
    intercept.fixed= input$intercepfixed,
    diagnostic = input$diagnostic,
    kernel = input$kernelmixed,
    adaptive = input$adaptativemixed,
    p = input$powermixed,
    theta = input$thetamixed*pi,
    longlat = input$longlatmixed,
    dMat = values$dMat

   )
  }
```

```
    else { values$gwr.mixed <- GWmodel::gwr.mixed(
      formula = formu,
      data = USelect2004,
      bw = input$bwmixed,
      fixed.vars = c(input$fixedvars),
      intercept.fixed= input$intercepfixed,
      diagnostic = input$diagnostic,
      kernel = input$kernelmixed,
      adaptive = input$adaptativemixed,
      p = input$powermixed,
      theta = input$thetamixed*pi,
      longlat = input$longlatmixed,
      dMat = values$dMat
    )
    }}
  shinyWidgets::closeSweetAlert(session = session)
  shinyWidgets::sendSweetAlert(
    session = session,
    title =" Calculation completed !",
    type = "success"
  )
  beepr::beep(2)
  shinybusy::remove_modal_spinner()
})

output$GWRmixed <- renderPrint({
  req(values$gwr.mixed)
  values$gwr.mixed
})
output$SDFGWRmixed <- DT::renderDT({
  req(values$gwr.mixed)
  datgwr <- data.frame(values$gwr.mixed$SDF)%>%
    mutate_if(is.numeric, round, digits = 6)
  DT::datatable(data =  datgwr, extensions = 'Buttons',
                options = list(dom = 'Bfrtip',
                               scrollX = TRUE,
                               fixedColumns = TRUE,
                               buttons = c('pageLength',
                                           'copy',
                                           'csv',
                                           'excel',
                                           'pdf',
                                           'print'),
                               pagelength = 10,
                               lengthMenu = list(c(10, 25, 100, -1),
                                                 c('10', '25', '100','All'))))
})

## plot gwr mixed
observe({
  req(values$gwr.mixed)
  vargws <- names(data.frame(values$gwr.mixed$SDF))

  updateSelectInput(session,'plotMgwr',
                    'Select', choices = vargws)
})

observeEvent(input$RunplotMgwr, {
  if (input$example == FALSE){
    req(values$gwr.mixed$SDF)
    polys <- list("sp.lines", as(z$map, "SpatialLines"), col="lightgrey",
                  lwd=.5,lty=0.1)
    map.na <- list("SpatialPolygonsRescale", layout.north.arrow(),
                   offset = c(input$lonMgwr,input$latMgwr), scale = 1, col=1)
    col.palette <- cartography::carto.pal(input$colorMgwr, 20)
    values$plotMgwr <-sp::spplot(values$gwr.mixed$SDF,input$plotMgwr,
                                 main = input$mainMgwr,
                                 sp.layout=list(polys,map.na),
                                 scales=list(cex = 1, col="black"),
```

```
                        col="transparent",
                        col.regions = col.palette)
    values$plotMgwr}
  else if(input$example == TRUE){
    map.na <- list("SpatialPolygonsRescale", layout.north.arrow(),
                offset = c(input$lonMgwr,input$latMgwr),
                scale = input$scaleMgwr, col=1)
    col.palette <- cartography::carto.pal(input$colorMgwr, 20)
    values$plotMgwr <-sp::spplot(values$gwr.mixed$SDF,input$plotMgwr,
                        main = input$mainMgwr,
                        sp.layout=list(map.na),
                        scales=list(cex = 1, col="black"),
                        col="transparent",
                        col.regions = col.palette)
    values$plotMgwr
  }


})

output$mapMgwr <- renderPlot({
  values$plotMgwr
},height = 700, width = 700)

observe({
  if (isTRUE(is.null(values$plotMgwr))) {
    shinyjs::disable("downMgwr")
  }
  else {shinyjs::enable("downMgwr")
  }
})
output$downMgwr <- downloadHandler(
  filename =  function() {
    paste("GeoWeithedModel",input$butdownMgwr, sep=".")
  },

  content = function(file) {
    if(input$butdownMgwr == "png")
      png(file) # open the png device
    else
      pdf(file) # open the pdf device
    print(values$plotMgwr) # draw the plot
    dev.off()  # turn the device off
  }
)

#-----------------------------gwr.scalable-----------------------------------
observe({
  if(input$example == TRUE){
    updateSelectizeInput(session,
                         'dependientscal', 'Dependent',
                         choices = USelect2004@data %>%
                           dplyr::select_if(is.numeric)%>% names,
                         #multiple = TRUE,
                         server = TRUE
  )}
})

observe({
  if(input$example == TRUE){
  updateSelectizeInput(session,
                         'independientscal', 'Independient(s)',
                         choices = USelect2004@data %>%
                           dplyr::select_if(is.numeric)%>% names,
                         #multiple = TRUE,
                         server = TRUE)}
  })

observe({
```

```
 req(z$dat)
 ovars <- z$dat %>% dplyr::select_if(is.numeric) %>% names
 updateSelectizeInput(session,
                      'dependientscal', 'Dependent',
                      choices = ovars,
                      #multiple = TRUE,
                      server = TRUE
 )
})
observe({
 req(z$dat)
 ovars <- z$dat %>% dplyr::select_if(is.numeric) %>% names
 updateSelectizeInput(session,
                      'independientscal', 'Independient(s)',
                      choices = ovars,
                      #multiple = TRUE,
                      server = TRUE)
})


observe({
 if (isTRUE(is.null(input$independientscal))) {
   shinyjs::disable("Runscal")
 }
 else {shinyjs::enable("Runscal")
 }
})
observeEvent(input$Runscal, {
 req(values$dMat)
 shinybusy::show_modal_spinner(
   spin = "atom",
   color = "red",
   text = "Please wait...Work in progress. You'll have to be
   patient, because the result may take a while"
 )
 if(input$example == FALSE){
   formu <- as.formula(
     paste(input$dependientscal,
           " ~ ",paste(input$independientscal,collapse="+")))
   values$gwr.scalable <- GWmodel::gwr.scalable(formula = formu,
                                        data = datasel(),
                                        bw.adapt = input$bwscal,
                                        polynomial = input$polinom,
                                        kernel = input$kernelscal,
                                        p = input$powerscal,
                                        theta = input$thetascal*pi,
                                        longlat = input$longlatscal,
                                        dMat = values$dMat
                                        # parallel.method = "cluster"
   )}
 else if (input$example == TRUE){

   formu <- as.formula(paste(input$dependientscal," ~ ",
                             paste(input$independientscal,collapse="+")))
   values$gwr.scalable <- GWmodel::gwr.scalable(formula = formu,
                                        data = USelect2004,
                                        bw.adapt = input$bwscal,
                                        polynomial = input$polinom,
                                        kernel = input$kernelscal,
                                        p = input$powerscal,
                                        theta = input$thetascal*pi,
                                        longlat = input$longlatscal,
                                        dMat = values$dMat
                                        # parallel.method = "cluster"
   )
 }
 shinyWidgets::closeSweetAlert(session = session)
 shinyWidgets::sendSweetAlert(
   session = session,
```

```
    title =" Calculation completed !",
    type = "success"
  )
  beepr::beep(2)
  shinybusy::remove_modal_spinner()
})

output$GWRscalable <- renderPrint({
  req(values$gwr.scalable)
  values$gwr.scalable
})
output$SDFGWRscalable <- DT::renderDT({
  req(values$gwr.scalable)
  datgwr <- data.frame(values$gwr.scalable$SDF)%>%
    mutate_if(is.numeric, round, digits = 6)
  DT::datatable(data =  datgwr, extensions = 'Buttons',
                options = list(dom = 'Bfrtip',
                               scrollX = TRUE,
                               fixedColumns = TRUE,
                               buttons = c('pageLength',
                                           'copy',
                                           'csv',
                                           'excel',
                                           'pdf',
                                           'print'),
                               pagelength = 10,
                               lengthMenu = list(c(10, 25, 100, -1),
                                                 c('10', '25', '100','All'))))
})
## plot gwr scala
observe({
  req(values$gwr.scalable)
  vargws <- names(data.frame(values$gwr.scalable$SDF))

  updateSelectInput(session,'plotSgwr',
                    'Select', choices = vargws)
})

observeEvent(input$RunplotSgwr, {
  if(input$example == FALSE){
    req(values$gwr.scalable$SDF)
    polys <- list("sp.lines", as(z$map, "SpatialLines"), col="lightgrey",
                  lwd=.5,lty=0.1)
    map.na <- list("SpatialPolygonsRescale", layout.north.arrow(),
                   offset = c(input$lonSgwr,input$latSgwr),
                   scale = input$scaleSgwr, col=1)
    col.palette <- cartography::carto.pal(input$colorSgwr, 20)
    values$plotSgwr <-sp::spplot(values$gwr.scalable$SDF,
                                 input$plotSgwr, main = input$mainSgwr,
                                 sp.layout=list(polys,map.na),
                                 scales=list(cex = 1, col="black"),
                                 col="transparent",
                                 col.regions = col.palette)
    values$plotSgwr}
  else if(input$example == TRUE){
    map.na <- list("SpatialPolygonsRescale", layout.north.arrow(),
                   offset = c(input$lonSgwr,input$latSgwr),
                   scale = input$scaleSgwr, col=1)
    col.palette <- cartography::carto.pal(input$colorSgwr, 20)
    values$plotSgwr <-sp::spplot(values$gwr.scalable$SDF,
                                 input$plotSgwr, main = input$mainSgwr,
                                 sp.layout=list(map.na),
                                 scales=list(cex = 1, col="black"),
                                 col="transparent",
                                 col.regions = col.palette)
    values$plotSgwr
  }
```

```
  })
  output$mapSgwr <- renderPlot({
    values$plotSgwr
  },height = 700, width = 700)

  observe({
    if (isTRUE(is.null(values$plotSgwr))) {
      shinyjs::disable("downSgwr")
    }
    else {shinyjs::enable("downSgwr")
    }
  })
  output$downSgwr <- downloadHandler(
    filename =  function() {
      paste("GeoWeithedModel",input$butdownSgwr, sep=".")
    },
# is a function with argument file. content writes the plot to the device
    content = function(file) {
      if(input$butdownSgwr == "png")
        png(file) # open the png device
      else
        pdf(file) # open the pdf device
      print(values$plotSgwr) # draw the plot
      dev.off()  # turn the device off
    }
  )
  #-----------------------------------gwpca-----------------------------------
  observe({
    if(input$example == TRUE){
      updateSelectizeInput(session,
                           'vargwpca', 'Variables',
                           choices = USelect2004@data%>%
                             dplyr::select_if(is.numeric)
                           %>% names,
                           #multiple = TRUE,
                           server = TRUE)}
    })

  observe({
    req(z$dat)
    ovarspca <- z$dat %>% dplyr::select_if(is.numeric) %>% names
    updateSelectizeInput(session,
                         'vargwpca', 'Variables',
                         choices = ovarspca,
                         #multiple = TRUE,
                         server = TRUE
    )
  })
  observe({
  updateNumericInput(session,'kgwpca',
                     'k',
                     2,
                     min = 1,
                     max = length(input$vargwpca))
  })

  datagwpca <- reactive({
    if (input$example == FALSE){
    mf <- datasel()[, input$vargwpca]
    data.scaled <- scale(as.matrix(mf@data[, input$vargwpca]))
    coords <- sp::coordinates(datasel())
    datagwpca <- SpatialPointsDataFrame(coords, as.data.frame(data.scaled ))
    datagwpca}
  })
  #datagwpca <- reactive({
   # if(input$example== TRUE){
     # mf <- USelect2004[, input$vargwpca]
     # data.scaled <- scale(as.matrix(mf@data[, input$vargwpca]))
     # coords <- sp::coordinates(USelect2004)
```

```
  #datagwpca <- SpatialPointsDataFrame(coords, as.data.frame(data.scaled ))
  # datagwpca}
# })

 observe({
  if (isTRUE(is.null(input$vargwpca))) {
    shinyjs::disable("Rungwpca")
  }
  else {shinyjs::enable("Rungwpca")
  }
 })
 observeEvent(input$Rungwpca, {
  req(values$dMat)
  shinybusy::show_modal_spinner(
    spin = "atom",
    color = "#428ded",
    text = "Please wait...Work in progress. You'll have to be
    patient, because the result may take a while"
  )
  if(input$example == FALSE){

    if( input$selgwpca == "Automatic"){
    values$gwpca <- GWmodel::gwpca(data = datagwpca(),
                                vars = colnames(datagwpca()@data),
                                k = input$kgwpca,
                                robust = input$robustgwpca,
                                kernel = input$kernelgwpca,
                                adaptive = input$adaptativegwpca,
                                bw = values$bw,
                                p = input$powergwpca,
                                theta = input$thetagwpca*pi,
                                longlat= input$longlatgwpca,
                                dMat = values$dMat)
  }
  else { values$gwpca <- GWmodel::gwpca(data = datagwpca(),
                                  vars = colnames(datagwpca()@data),
                                  k = input$kgwpca,
                                  robust = input$robustgwpca,
                                  kernel = input$kernelgwpca,
                                  adaptive = input$adaptativegwpca,
                                  bw = input$bwgwpca,
                                  p = input$powergwpca,
                                  theta = input$thetagwpca*pi,
                                  longlat= input$longlatgwpca,
                                  dMat = values$dMat)
  }}
  else if(input$example == TRUE){

    if( input$selgwpca == "Automatic"){
      values$gwpca <- GWmodel::gwpca(data = USelect2004,
                                vars = input$vargwpca,
                                k = input$kgwpca,
                                robust = input$robustgwpca,
                                kernel = input$kernelgwpca,
                                adaptive = input$adaptativegwpca,
                                bw = values$bw,
                                p = input$powergwpca,
                                theta = input$thetagwpca*pi,
                                longlat= input$longlatgwpca,
                                dMat = values$dMat)
    }
    else { values$gwpca <- GWmodel::gwpca(data = USelect2004,
                                    vars = input$vargwpca,
                                    k = input$kgwpca,
                                    robust = input$robustgwpca,
                                    kernel = input$kernelgwpca,
                                    adaptive = input$adaptativegwpca,
                                    bw = input$bwgwpca,
                                    p = input$powergwpca,
```

```
                                theta = input$thetagwpca*pi,
                                longlat= input$longlatgwpca,
                                dMat = values$dMat)
    }

  }
  shinyWidgets::closeSweetAlert(session = session)
  shinyWidgets::sendSweetAlert(
    session = session,
    title =" Calculation completed !",
    type = "success"
  )
  shinybusy::remove_modal_spinner()
})
output$DFgwpca <- DT::renderDT({
  req(values$gwpca)
  values$gwpca$SDF
  datgwpca <- data.frame(values$gwpca$SDF)%>%
    dplyr::mutate_if(is.numeric, round, digits = 6)
  DT::datatable(data =  datgwpca, extensions = 'Buttons',
                options = list(dom = 'Bfrtip',
                               scrollX = TRUE,
                               fixedColumns = TRUE,
                               buttons = c('pageLength',
                                           'copy',
                                           'csv',
                                           'excel',
                                           'pdf',
                                           'print'),
                               pagelength = 10,
                               lengthMenu = list(c(10, 25, 100, -1),
                                                 c('10', '25', '100','All'))))
})

output$gwpca <- renderPrint({
  req(values$gwpca)
  values$gwpca
})
observe({
  updateNumericInput(session,'loading',
                     'Component',
                     1,
                     min = 1,
                     max = input$kgwpca)
})

output$loadinggwpca <- DT::renderDT({
  req(values$gwpca)
  loadinggwpca <- data.frame(values$gwpca$loadings[, , input$loading])%>%
    mutate_if(is.numeric, round, digits = 6)
  DT::datatable(data =  loadinggwpca, extensions = 'Buttons',
                options = list(dom = 'Bfrtip',
                               scrollX = TRUE,
                               fixedColumns = TRUE,
                               buttons = c('pageLength',
                                           'copy',
                                           'csv',
                                           'excel',
                                           'pdf',
                                           'print'),
                               pagelength = 10,
                               lengthMenu = list(c(10, 25, 100, -1),
                                                 c('10', '25', '100','All'))))
})

observe({
  updateNumericInput(session,'loadingVgwpca',
                     'Number of component',
                     2,
```

```
                min = 2,
                max = input$kgwpca)
})



observeEvent(input$RunplotVgwr, {
 if(input$example == FALSE){
  req(values$gwpca)
  var.gwpca <- prop.var(values$gwpca, input$loadingVgwpca)
  mf <- datasel()[, input$vargwpca]
  mf$var.gwpca <- var.gwpca
  polys <- list("sp.lines", as(z$map, "SpatialLines"), col="lightgrey",
                lwd=.5,lty=0.1)
  map.na <- list("SpatialPolygonsRescale", layout.north.arrow(),
                offset = c(input$longwpca,input$latgwpca),
                scale = input$scalegwpca, col=1)
  col.palette <- cartography::carto.pal(input$colorVgwpca, 20)
  values$plotVgwpca <- sp::spplot(mf,"var.gwpca", main = input$mainVgwpca,
                            sp.layout=list(polys,map.na),
                            scales=list(cex = 1, col="black"),
                            col="transparent",
                            col.regions = col.palette)
  values$plotVgwpca}
 else if(input$example == TRUE){
  req(values$gwpca)
  var.gwpca <- prop.var(values$gwpca, input$loadingVgwpca)
  mf <- USelect2004[, input$vargwpca]
  mf$var.gwpca <- var.gwpca
  map.na <- list("SpatialPolygonsRescale", layout.north.arrow(),
                offset = c(input$longwpca,input$latgwpca),
                scale = input$scalegwpca, col=1)
  col.palette <- cartography::carto.pal(input$colorVgwpca, 20)
  values$plotVgwpca <- sp::spplot(mf,"var.gwpca", main = input$mainVgwpca,
                            sp.layout=list(map.na),
                            scales=list(cex = 1, col="black"),
                            col="transparent",
                            col.regions = col.palette)
  values$plotVgwpca
 }




})
output$Vargwpca <- renderPlot({
 values$plotVgwpca
},height = 700, width = 700)

observe({
 if (isTRUE(is.null(values$plotVgwpca))) {
  shinyjs::disable("downVgwpca")
 }
 else {shinyjs::enable("downVgwpca")
 }
})

output$downVgwpca <- downloadHandler(
 filename =  function() {
  paste("GeoWeithedModel",input$butdownVgwpca, sep=".")
 },
 # is a function with argument file. content writes the plot to the device
 content = function(file) {
  if(input$butdownVgwpca == "png")
   png(file) # open the png device
  else
   pdf(file) # open the pdf device
  print(values$plotVgwpca) # draw the plot
  dev.off()  # turn the device off
```

```
  }
)

observe({
  updateNumericInput(session,'loadingwingwpca',
                     'Component',
                     1,
                     min = 1,
                     max = input$kgwpca)
})

observeEvent(input$Runplotwingwpca, {
  if(input$example== FALSE){
    mf <- datasel()[, input$vargwpca]
    loadings.pc1 <- values$gwpca$loadings[, , input$loadingwingwpca]
    win.item <- max.col(abs(loadings.pc1))
    mf$win.item <- win.item
    polys <- list("sp.lines", as(z$map, "SpatialLines"), col="lightgrey",
                  lwd=.5,lty=0.1)
    map.na <- list("SpatialPolygonsRescale", layout.north.arrow(),
                   offset = c(input$lonwingwpca,input$latwingwpca),
                   scale =input$scalewingwpca, col=1)
    mypalette.4 <- cartography::carto.pal("multi.pal",
                                          n1 = length(input$vargwpca))
    values$plotwingwpca <- sp::spplot(mf,"win.item",
                                      main = input$mainwingwpca,
                                sp.layout=list(polys,map.na),
                                scales=list(cex = 1, col="black"),
                                #legendEntries = c(input$vargwpca),
                                col="transparent",
                                col.regions = mypalette.4,
                                at = seq(from = 1,
                                         to = length(input$vargwpca)+1,
                                         by =1 ))
    values$plotwingwpca}
  else if(input$example == TRUE){
    mf <- USelect2004[, input$vargwpca]
    loadings.pc1 <- values$gwpca$loadings[, , input$loadingwingwpca]
    win.item <- max.col(abs(loadings.pc1))
    mf$win.item <- win.item
    map.na <- list("SpatialPolygonsRescale", layout.north.arrow(),
                   offset = c(input$lonwingwpca,input$latwingwpca),
                   scale = input$scalewingwpca, col=1)
    mypalette.4 <- cartography::carto.pal("multi.pal",
                                          n1 = length(input$vargwpca))
    values$plotwingwpca <- sp:: spplot(mf,"win.item", main = input$mainwingwpca,
                                sp.layout=list(map.na),
                                scales=list(cex = 1, col="black"),
                                #legendEntries = c(input$vargwpca),
                                col="transparent",
                                col.regions = mypalette.4,
                                at = seq(from = 1,
                                         to = length(input$vargwpca)+1,
                                         by =1 ))
    values$plotwingwpca
  }



})

output$wingwpca <- renderPlot({
  values$plotwingwpca
},height = 700, width = 700)

observe({
  if (isTRUE(is.null(values$plotwingwpca))) {
    shinyjs::disable("downwingwpca")
```

```
  }
  else {shinyjs::enable("downwingwpca")
  }
})
output$downwingwpca <- downloadHandler(
  filename =  function() {
    paste("GeoWeithedModel",input$butdownwingwpca, sep=".")
  },
  content = function(file) {
    if(input$butdownwingwpca == "png")
      png(file) # open the png device
    else
      pdf(file) # open the pdf device
    print(values$plotwingwpca) # draw the plot
    dev.off()  # turn the device off

  }
)

#-------------------------------------gwda-----------------------------------
  observe({
    #req(z$dat)
      #if (input$example == FALSE | is.null(z$dat)){return()}
     if (input$example == TRUE){
    updateSelectizeInput(session,
                         'dependgwda', 'Grouping factor',
                         choices = USelect2004@data %>% dplyr::select_if(is.factor)%>%
                          names,
                         #multiple = TRUE,
                         server = TRUE,
                         options = list(maxItems = 1)
                        )}
    })

observe({
  req(z$dat)
  #if (input$example == FALSE | is.null(z$dat)){return()}
    updateSelectizeInput(session,
                         'dependgwda', 'Grouping factor',
                         choices = z$dat %>% dplyr::select_if(is.factor)%>% names,
                         #multiple = TRUE,
                         server = TRUE,
                         options = list(maxItems = 1))
    })
observe({
  req(z$dat)
 ovars <- z$dat %>% dplyr::select_if(is.numeric)%>% names
 updateSelectizeInput(session,
                      'indepentgwda', 'Discriminators',
                      choices = ovars,
                      #multiple = TRUE,
                      server = TRUE)
 })
observe({
 #req(z$dat)
 if (input$example == TRUE ){
   updateSelectizeInput(session,
                        'indepentgwda', 'Discriminators',
                        choices =  USelect2004@data %>%
                          dplyr::select_if(is.numeric)%>% names,
                        #multiple = TRUE,
                        server = TRUE)
   }
  })
observe({
  if (isTRUE(is.null(input$indepentgwda))) {
    shinyjs::disable("Rungwda")
    }
  else {shinyjs::enable("Rungwda")
```

```
  }
  })

observeEvent(input$Rungwda,{
  req(values$dMat)
  shinybusy::show_modal_spinner(
    spin = "atom",
    color = "red",
    text = "Please wait...Work in progress. You'll have to be
    patient, because the result may take a while")

  if( input$example == TRUE ){
    formu <- as.formula(paste(input$dependgwda," ~ ",
                              paste(input$indepentgwda,collapse="+")))
     if(input$selgwda == "Automatic"){
      values$gwda <-  gwda(formula = formu,
                           data = USelect2004,
                           COV.gw = input$COVgwda,
                           predict.data = USelect2004,
                           prior.gw = input$priorgwda,
                           mean.gw = input$meangwda,
                           bw = values$bw,
                           prior = NULL,
                           wqda = input$wqdagwda,
                           kernel = input$kernelgwda,
                           adaptive = input$adaptativegwda,
                           p = input$powergwda,
                           theta = input$thetagwda*pi,
                           longlat= input$longlatgwda,
                           dMat = values$dMat)
    }
    else if(input$selgwda == "Manual"){
      values$gwda <-  gwda(formula = formu,
                           data = USelect2004,
                           COV.gw = input$COVgwda,
                           predict.data = USelect2004,
                           prior.gw = input$priorgwda,
                           mean.gw = input$meangwda,
                           bw = input$bwgwda,
                           prior = NULL,
                           wqda = input$wqdagwda,
                           kernel = input$kernelgwda,
                           adaptive = input$adaptativegwda,
                           p = input$powergwda,
                           theta = input$thetagwda*pi,
                           longlat= input$longlatgwda,
                           dMat = values$dMat)
    }

  }
  else if (input$example == FALSE) {
    formu <- as.formula(paste(input$dependgwda," ~ ",
                              paste(input$indepentgwda,collapse="+")))
    if(input$selgwda == "Manual"){
      values$gwda <-  gwda(formula = formu,
                           data = datasel(),
                           COV.gw = input$COVgwda,
                           predict.data = datasel(),
                           prior.gw = input$priorgwda,
                           mean.gw = input$meangwda,
                           bw = input$bwgwda,
                           prior = NULL,
                           wqda = input$wqdagwda,
                           kernel = input$kernelgwda,
                           adaptive = input$adaptativegwda,
                           p = input$powergwda,
                           theta = input$thetagwda*pi,
                           longlat= input$longlatgwda,
                           dMat = values$dMat)
```

```
    }
    else if(input$selgwda == "Automatic"){
      values$gwda <-  gwda(formula = formu,
                           data = datasel(),
                           COV.gw = input$COVgwda,
                           predict.data = datasel(),
                           prior.gw = input$priorgwda,
                           mean.gw = input$meangwda,
                           bw = values$bw,
                           prior = NULL,
                           wqda = input$wqdagwda,
                           kernel = input$kernelgwda,
                           adaptive = input$adaptativegwda,
                           p = input$powergwda,
                           theta = input$thetagwda*pi,
                           longlat= input$longlatgwda,
                           dMat = values$dMat)
    }

  }

  shinyWidgets::closeSweetAlert(session = session)
  shinyWidgets::sendSweetAlert(
    session = session,
    title =" Calculation completed !",
    type = "success")
  beepr::beep(2)
  shinybusy::remove_modal_spinner()
})


output$gwda <- renderPrint({
  req(values$gwda)
  values$gwda
  })

output$DFgwda <-  DT::renderDT({
  req(values$gwda$SDF)
  datgwda <- data.frame(values$gwda$SDF)%>%
    mutate_if(is.numeric, round, digits = 6)
  DT::datatable(data =  datgwda, extensions = 'Buttons',
                options = list(dom = 'Bfrtip',
                               scrollX = TRUE,
                               fixedColumns = TRUE,
                               buttons = c('pageLength',
                                           'copy',
                                           'csv',
                                           'excel',
                                           'pdf',
                                           'print'),
                               pagelength = 10,
                               lengthMenu = list(c(10, 25, 100, -1),
                                                 c('10', '25', '100','All')))))
})

output$predict <-  renderPrint({
  req(values$gwda$SDF)
  table(USelect2004$winner,values$gwda$SDF$group.predicted)
})

# plot gwda
  observe({
  req(values$gwda$SDF)
  vargws <- names(data.frame(values$gwda$SDF))

  updateSelectInput(session,'plotgwda',
                    'Select', choices = c("Original",
                                          "group.predicted",vargws))
  })
```

```
observeEvent(input$Runplotgwda, {

 req(values$gwda$SDF)
 if(input$example == TRUE){
  map.na <- list("SpatialPolygonsRescale", layout.north.arrow(),
              offset = c(input$longwda,input$latgwda),
              scale = input$scalegwda, col=1)
  col.palette <- cartography::carto.pal(input$colorgwda, 20)

  if(input$plotgwda == "Original"){
   USelect2004$winner <- factor(USelect2004$winner)
   values$plotgwda <- sp::spplot(USelect2004,"winner",
                      main = input$maingwda,
                      sp.layout=list(map.na),
                      scales=list(cex = 1, col="black"),
                      col="transparent",
                      col.regions = col.palette)
  values$plotgwda}
  else if(input$plotgwda == "group.predicted"){
   data <- values$gwda$SDF
   data$group.predicted <- factor(data$group.predicted)
   values$plotgwda <- sp::spplot(data,"group.predicted",
                        main = input$maingwda,
                        sp.layout=list(map.na),
                        scales=list(cex = 1, col="black"),
                        col="transparent",
                        col.regions = col.palette)
   values$plotgwda}
  else {
   values$plotgwda <- sp::spplot(values$gwda$SDF,input$plotgwda,
                      main = input$maingwda,
                      sp.layout=list(map.na),
                      scales=list(cex = 1, col="black"),
                      col="transparent",
                      col.regions = col.palette)
  values$plotgwda
  }

 }
 else if(input$example == FALSE){
  polys <- list("sp.lines", as(z$map, "SpatialLines"),
              col="lightgrey", lwd=.5,lty=0.1)
  map.na <- list("SpatialPolygonsRescale", layout.north.arrow(),
              offset = c(input$longwda,input$latgwda),
              scale = input$scalegwda, col=1)
  col.palette <- cartography::carto.pal(input$colorgwda, 20)
  values$plotgwda <- sp::spplot(values$gwda$SDF,input$plotgwda,
                      main = input$maingwda,
                      sp.layout=list(polys,map.na),
                      scales=list(cex = 1, col="black"),
                      col="transparent",
                      col.regions = col.palette)
  values$plotgwda}
})

output$mapgwda <- renderPlot({
 values$plotgwda
},height = 700, width = 700)

observe({
 if (isTRUE(is.null(values$plotgwda))) {
  shinyjs::disable("downgwda")
 }
 else {shinyjs::enable("downgwda")
 }
})
output$downgwda <- downloadHandler(
 filename =  function() {
```

```
    paste("GeoWeithedModel",input$butdowngwda, sep=".")
  },

  content = function(file) {
    if(input$butdowngwda == "png")
      png(file) # open the png device
    else
      pdf(file) # open the pdf device
    print(values$plotgwda) # draw the plot
    dev.off()  # turn the device off
  }
)

#------------------------Autocorrelation----------------------

 observe({
  req(z$dat)
  ovars <- names(z$dat)
  updatePickerInput(session,
                    'varauto',
                    choices = ovars)
  })

 observeEvent(input$Runauto, {
  req(datasel())
  datauto <- datasel()
  Variable <- pull(dplyr::select(z$dat,input$varauto))
  print(input$varauto)
  shinybusy::show_modal_spinner(
    spin = "atom",
    color = "blue",
    text = "Please wait...Work in progress. You'll have to be
    patient, because the result may take a while")
  neighbourhood <- spdep::poly2nb(datauto, queen= TRUE)
  neighbourhood_weights_list <- spdep::nb2listw(neighbourhood,
                                        style=input$style,
                                        zero.policy= TRUE)

  values$morantest <- spdep::moran.test(Variable,
                                neighbourhood_weights_list,
                                alternative = input$alternative)
  values$moranmc <-  spdep::moran.mc(Variable,
                               neighbourhood_weights_list,
                               alternative= input$alternative,
                               nsim= input$numsim)
  values$localmoran <- spdep::localmoran(Variable,
                                 neighbourhood_weights_list,
                                 p.adjust.method="bonferroni",
                                 alternative = input$alternative,
                                 na.action=na.exclude,
                                 zero.policy = TRUE)
  values$localmoranperm <- spdep::localmoran_perm(Variable,
                                          neighbourhood_weights_list,
                                          p.adjust.method="bonferroni",
                                          na.action=na.exclude,
                                          nsim= input$numsim,
                                          alternative = input$alternative,
                                          zero.policy = TRUE)
  shinybusy::remove_modal_spinner()
  beepr:: beep(2)
 })

 output$variablename<- renderPrint({
  req(input$varauto)
  print(input$varauto)
 })

 output$moran <- renderPrint({
  req(values$morantest)
```

```
  values$morantest
 })

 output$moranmc <- renderPrint({
  req(values$moranmc)
    values$moranmc
 })

 output$localmoran <- renderPrint({
  req(values$localmoran)
  summary(values$localmoran)
 })

 output$localmoranperm <- renderPrint({
  req(values$localmoranperm)
  summary(values$localmoranperm)
 })

 output$DFauto <-  DT::renderDT({
  req(values$localmoran)
  datmat <- data.frame(values$localmoran[,1:5])%>%
    dplyr::mutate_if(is.numeric, round, digits = 6)
  DT::datatable(data = datmat, extensions = 'Buttons',
                colnames = c("Ii","E.Ii","Var.Ii","Z.Ii","Pr.Z"),
              options = list(dom = 'Bfrtip', scrollX = TRUE,
                             fixedColumns = TRUE,
                             buttons = c('pageLength',
                                         'copy',
                                         'csv',
                                         'excel',
                                         'pdf',
                                         'print'),
                             pagelength = 10,
                             lengthMenu = list(c(10, 25, 100, -1),
                                               c('10', '25', '100','All'))))
 })

observeEvent(input$Runplotauto, {
  req(values$localmoran)
  SPDF <- datasel()
  SPDF@data$lmoran_i <- values$localmoran[,1]
  SPDF@data$lmoran_p <- values$localmoran[,5]
  polys <- list("sp.lines", as(z$map,
                               "SpatialLines"),
                col="lightgrey", lwd=.5,lty=0.1)
  col.palette <- cartography::carto.pal(input$colorauto, 20)
  map.na <- list("SpatialPolygonsRescale",
                 layout.north.arrow(),
                 offset = c(input$longauto,input$latauto),
                 scale = input$scaleauto, col=1)
  if(input$plotauto == "lmoran_i"){
    values$plotauto <- sp::spplot(SPDF, "lmoran_i",
                                  main = "Local Moran's I",
                                  sp.layout=list(polys,map.na),
                                  scales=list(cex = 1, col="black"),
                                  col="transparent",
                                  col.regions = col.palette)
    }
  else if(input$plotauto == "lmoran_p"){
    values$plotauto <- sp::spplot(SPDF, "lmoran_p",
                                  main = "P-values",
                                  sp.layout=list(polys,map.na),
                                  scales=list(cex = 1, col="black"),
                                  col="transparent",
                                  col.regions = col.palette)
    }
  })
```

```
  output$mapauto <- renderPlot({
    req(values$plotauto)
    values$plotauto
  },height = 700, width = 700)


  observe({
    if (isTRUE(is.null(values$plotauto))) {
      shinyjs::disable("downgauto")
    }
    else {shinyjs::enable("downgauto")
    }
  })

  output$downgauto <- downloadHandler(
    filename =  function() {
      paste("GeoWeithedModel",input$butdownauto, sep=".")
    },

    content = function(file) {
      if(input$butdownauto == "png")
        png(file) # open the png device
      else
        pdf(file) # open the pdf device
      print(values$plotauto) # draw the plot
      dev.off()  # turn the device off

    }
  )


  }) #FIN
```

####Server.R (Fin)

ANEXO 6. Datos Utilizados en el Ejemplo de Aplicación del Paquete GeoWeightedModel.

x	y	Rate	POV	SMOK	PM25	NO2	SO2	FIPS
1056523,94	1376613,19	72	15,92	27,9333333	11,7553333	0,97766667	0,06418495	13111
1653442,06	2267301,48	59	12,22	26,9666667	9,026	1,49993333	0,03321098	42115
1633708,28	2096683,06	61	8,98666667	25,2733333	11,9633333	3,61646667	0,12028133	42075
1584048,85	1901443,4	62	7,86	22,9	12,7313333	3,62193333	0,11837113	51683
1735811,1	2409536,49	59	14,7466667	27,18	8,302	1,63273333	0,00640437	36057
1003646,75	1193901,75	86	17,5066667	30,2133333	12,2713333	1,6258	0,13878049	13149
1463896,2	1452940,21	79	21,9333333	29,48	12,7106667	1,48733333	0,0223045	37153
1716542,53	1743450,72	61	4,64666667	21,2333333	11,8433333	2,52026667	0,05727971	51735
1320295,85	1533039	64	12,6733333	27,7733333	11,5566667	1,7928	0,09070548	37003
1520852,15	1580156,88	65	14,8133333	19,58	12,3566667	1,93746667	0,06037865	37063
1768117,62	2407434,53	62	11,2466667	23,7866667	8,76733333	2,0944	0,00613167	36093
1463355,63	1632109,56	72	21,3733333	29,7133333	11,4726667	2,11326667	0,12612054	51590
1597510,11	1916180,73	41	5,57333333	15,58	13,132	4,147	0,13029202	51600
1445710,52	1759556,92	61	13,1666667	31,7	10,1246667	1,66666667	0,02803057	51530
1555147,6	1623917	79	22,22	29,66	11,7313333	1,73886667	0,06439823	37181
1240686,99	999295,872	93	24,16	28,2866667	11,87	0,636	0,00773525	13003
1622534,83	1672340,58	78	20,3533333	28,42	11,9273333	1,58486667	0,04330067	51081
1568845,62	2286060,4	73	14,36	26,1133333	9,17666667	1,64006667	0,04128695	36015
1462544,76	1811357,67	63	13,5533333	25,0533333	10,766	1,5146	0,03323845	51790
1434932,94	1605912,89	78	14,5266667	31,1	11,486	2,29733333	0,1052822	37157
1161855,58	1335144,38	72	16,5666667	27,5133333	12,7313333	1,39446667	0,03068753	13119
1581875,06	2029400,66	56	7,6	24,32	12,9446667	3,29246667	0,11301066	42001
1514326,82	1806985,04	53	19,7733333	23,1866667	11,3053333	1,98453333	0,0334955	51540
1644880,39	1977469,45	87	20,96	28,9933333	14,09	5,1886	0,14662918	24510
1485798,53	2224296,62	68	13,48	27	9,63	1,64146667	0,05687065	42105
1227378,73	968358,873	89	21,0066667	28,5133333	11,602	0,65173333	0,00759822	13173
1151573,41	1181611,12	62	11,6866667	24,82	12,2786667	1,46866667	0,13999563	13169
1584749,02	2324857,45	52	15,8133333	16,2133333	9,18133333	1,77753333	0,03277843	36109
1638111,94	2388843,96	67	10,9333333	24,8266667	8,768	1,84686667	0,01315113	36053
1432790,71	1443984,99	66	20,9333333	28,3733333	12,7226667	1,68553333	0,02962407	37007
1611858,91	1559545,42	71	19,1933333	25,6933333	13,0186667	1,69506667	0,01767137	37195
1385413,79	1946823,68	62	16,44	28,66	10,9553333	2,3982	0,25078285	54077
1420848,11	1636040,47	65	14,6933333	30,1533333	10,8953333	1,855	0,09123662	51089
1622210,6	1854839,18	74	6,79333333	21,5933333	12,0826667	2,94866667	0,12515072	51099
1431155,36	2177641,11	64	8,88	27,5866667	10,5346667	1,74226667	0,11571985	42047
1381551,31	2178527,87	74	17,6266667	30	10,9106667	2,31586667	0,1222416	42053
1099839,16	1245707,83	62	11,6866667	22,44	13,1333333	3,7826	0,13532285	13247
1653351,82	2193500,04	65	12,3266667	28,8466667	9,62266667	2,4324	0,07528687	42079
1834509,19	2196946,25	45	27,9733333	23,2533333	11,748	11,6445333	0,03987772	36005
1604588,65	2214031,26	63	12,98	26,9866667	9,344	1,61486667	0,08684919	42113
1444689,53	2074193,14	57	13,2733333	26,9466667	11,1873333	3,09953333	0,26352954	42021

x	y	Rate	POV	SMOK	PM25	NO2	SO2	FIPS
1285877,94	1284646,67	58	18,08	26,04	12,1866667	1,29026667	0,02473832	45037
1622394,82	2172315,89	56	12,4466667	25,78	10,06	2,702	0,11567753	42037
1508276,72	1804173,95	53	7,72666667	16,5266667	11,1873333	1,7494	0,03336686	51003
1445280,42	1529734,55	71	13,18	28,7466667	12,4046667	2,2968	0,05728067	37151
1816659,88	2353655,65	64	10,28	23,5133333	8,988	2,0608	0,0100048	36021
1800141,09	2141559,36	53	7,12666667	18,7133333	11,6366667	6,86953333	0,05096163	34023
1218893,13	1127443,53	70	20,5333333	25,9466667	12,1773333	0,84686667	0,0302641	13175
1179492,3	1247682,04	60	20,2933333	25,2333333	12,434	1,34466667	0,10151292	13133
1201807,39	1045952,93	96	23,0533333	28,6866667	12,4753333	0,64646667	0,01071437	13017
1177578,94	1008758,6	70	21,7133333	26,18	12,7546667	0,70926667	0,00898423	13277
1262954,26	959490,114	104	23,44	29,1	11,2953333	0,5954	0,01051375	13065
1623811,32	1739869,12	72	7,57333333	28,4	11,8306667	2,79906667	0,07902238	51570
1439791,53	1570775,95	61	14,4333333	23,2666667	12,1726667	2,71133333	0,08870687	37081
1330087,37	1619263,78	79	19,7466667	31,58	9,58733333	1,33013333	0,04283173	51640
1631867,74	1744490,81	101	16,5066667	32,3933333	11,836	3,2466	0,08019077	51670
1181765,03	1119115,17	85	17,5733333	27,9	12,2326667	0,88173333	0,03883442	13023
1363361,5	1097965,83	92	10,56	23,7466667	11,306	0,94953333	0,01031302	13029
1319776,97	1107445,31	75	23,98	28,5266667	11,85	0,82646667	0,01162779	13109
1097548,86	1180372,84	80	16,2866667	28,1666667	12,3206667	1,6782	0,1309916	13171
1167461,43	1307970,61	85	14,76	27,4733333	12,796	1,34393333	0,04357227	13195
1176211,4	1625932,37	87	19,5266667	31,2133333	9,72866667	2,209	0,05738237	51720
1683120,47	1752365,46	48	18,9066667	17,6666667	11,9473333	2,52346667	0,06950066	51830
1681804,14	2456217,56	60	13,16	25,12	7,75666667	1,20473333	0,00574632	36043
1643936,98	2304050,69	58	14,1133333	23,9466667	8,944	1,68733333	0,02485057	36007
1353126,64	1674729,09	78	23,04	22,7666667	9,702	1,91786667	0,0347425	51750
1253566,4	1687978,49	106	34,4666667	34,96	9,40866667	1,16826667	0,03828383	54047
1329015,14	1916161,46	84	17,3666667	28	11,6833333	2,94973333	0,21072637	54033
1310840,99	1846926,19	78	21,6133333	30,9266667	10,6306667	1,652	0,10218563	54007
1142177,55	1155786,94	75	21,44	25,3533333	12,6553333	1,3396	0,10016458	13021
1731568,73	1694759,54	72	8,19333333	21,5866667	10,29	2,3896	0,04355765	51550
1634724,24	2487703,53	61	13,64	26,7666667	7,93666667	1,52053333	0,00619001	36049
1746713,98	1664623,64	66	9,47333333	23,3933333	9,948	1,65993333	0,02423754	37029
1305889,3	2071193,13	65	10,9933333	27,3333333	13,158	5,27093333	0,31250119	42007
1179525,22	1190286,64	68	20,9466667	26,4733333	12,322	1,3204	0,1200595	13009
1682297,9	2170140,23	68	10,06	29,4466667	9,88733333	3,41206667	0,0727306	42025
1115419,38	1036903,96	77	9,92666667	24,68	12,4526667	0,88433333	0,01298697	13177
1720531,92	1943797,19	80	11,5933333	25,52	12,9286667	4,18906667	0,08052073	24011
949876,977	1363028,67	90	14,0266667	26,56	12,678	2,0276	0,09777243	13083
1119076,39	935895,365	77	22,2	27,0933333	11,876	0,64526667	0,00789093	13131
1317582,56	1729447,17	70	24,8333333	30,3333333	9,44866667	1,3772	0,03728186	54089
989780,592	1248202,62	89	17	30,2466667	12,6426667	1,5934	0,20456215	13143
1048750,48	1330599,25	74	11,5266667	26,4066667	12,8486667	1,5288	0,12275479	13227

x	y	Rate	POV	SMOK	PM25	NO2	SO2	FIPS
1574275,14	2130545,41	52	10,56	25,5733333	11,046	3,1288	0,14488425	42109
1421799,94	1636136,48	80	20,1266667	28,96	10,9446667	1,84813333	0,09080463	51690
1652705,23	2344047,23	63	14,2	26,0866667	8,73266667	1,44333333	0,01655284	36017
1149153,54	938437,028	78	20,18	25,8133333	11,886	0,66913333	0,00755534	13275
1717104,38	1631166,37	69	18,5466667	26,0066667	10,5593333	1,29626667	0,02000273	37041
1394188,12	1599626,41	77	11,9066667	30,9333333	11,198	1,99326667	0,08791065	37169
1294167,14	2002786,43	76	15,28	28,3666667	13,4333333	4,54506667	0,40655905	54069
1393922,75	1700093,24	71	8,75333333	26,5266667	10,322	2,20606667	0,0342634	51775
1593702,53	1852059,96	69	14,9	24,4933333	12,024	2,72193333	0,10601951	51630
1031382,34	1242709,45	73	10,76	24,7733333	13,2533333	2,90926667	0,20628079	13097
1571748,5	1594889,67	72	14,1	26,08	12,1573333	1,71193333	0,03852128	37069
1109248,87	1321624,54	62	13,34	24,8733333	13,088	1,55653333	0,05498553	13139
1242895,47	1448685,55	53	12,36	23,8333333	11,2586667	1,76313333	0,05993061	37149
1058282,46	1027981,07	69	27,06	27,0933333	11,9853333	0,72053333	0,01383411	13243
1252655,9	1125410,72	86	26,5866667	29,0533333	12,074	0,69973333	0,01963607	13283
1323677,53	1484314,99	66	11,6666667	26,3466667	12,2033333	2,45946667	0,1154378	37109
1731906,79	2432891,23	65	14,28	27,5866667	8,024	1,35086667	0,00524356	36035
1553949,12	2311832,58	67	12,3666667	25,8133333	9,09666667	1,772	0,03967289	36097
1535005,92	1613210,2	76	13,8333333	27,1466667	11,7113333	1,76473333	0,07535617	37077
1625527,45	1733268,28	97	20,62	29,5266667	11,636	2,49613333	0,07444644	51730
1248348,27	934440,329	73	23,7133333	28,5866667	11,1826667	0,6676	0,00970261	13101
1473184,12	2023798,4	52	11,9266667	26,8266667	11,0266667	2,84266667	0,15888642	42009
1243553,43	1245820,7	94	19,4466667	28,2933333	12,4213333	1,41006667	0,0454703	13189
1148323,71	1057203,73	78	27,5066667	27,8333333	12,6306667	0,80633333	0,01437081	13081
1730261,75	1719321,77	78	19,3933333	25,3866667	11,1406667	2,61733333	0,04962181	51710
1535867,51	1788758,04	55	6,97333333	23,1933333	11,2313333	1,84146667	0,04103907	51065
1474728,09	1572762,4	69	13,5	27	11,9	2,29846667	0,07881044	37001
1361780,11	1076726,92	76	17,1933333	25,4333333	11,1193333	0,8934	0,00977471	13179
1032302,68	1036603,89	93	23,9866667	28,8266667	12,018	0,76013333	0,01527326	13239
1549999,88	1837565,68	68	9,05333333	25,8266667	11,4973333	2,19993333	0,05785453	51137
1301653,5	1910050,19	76	19,32	31,5266667	11,968	3,15826667	0,22385305	54017
1068274,98	962213,133	78	21,9933333	26,7666667	12,5226667	0,66193333	0,01033905	13201
1768774,33	2011919,09	70	15,2333333	28,1133333	12,9146667	3,48773333	0,0456092	34011
1095759,99	1279796,44	53	9,27333333	19,2266667	13,4066667	3,24366667	0,10459696	13135
1337590,34	2102567,41	59	8,33333333	23,4	12,604	4,38326667	0,28662513	42019
1776076,01	2442465,77	63	6,59333333	21,1333333	8,52733333	2,01526667	0,00439067	36091
1646303	1513043,27	73	20,2333333	27,3533333	12,118	1,4862	0,01287791	37107
1037083,19	1109400,01	74	17,92	24,9533333	12,6886667	0,9664	0,03679928	13215
1699112,31	2138770,95	56	10,3666667	24,2866667	11,4393333	4,45166667	0,07234667	42077
1079948,56	1256389,65	52	15,0066667	18,88	13,422	4,4912	0,14099644	13089
1642946,12	1704445,42	84	19,4066667	27,9133333	11,4926667	1,90826667	0,05864335	51183
1517023,26	1936981,43	78	13,4533333	27,1066667	11,8393333	2,98966667	0,08177359	51840

x	y	Rate	POV	SMOK	PM25	NO2	SO2	FIPS
1509167,12	1462165,11	82	17,8733333	27,1666667	12,834	1,59966667	0,01735406	37093
1429885,05	2220922,96	73	14,3333333	28,5266667	10,05	1,46686667	0,07155714	42083
1434148,18	1352336,96	81	26,1933333	29,2866667	12,572	1,3402	0,03836309	45061
1552375,47	1345424,09	70	15,16	29,9866667	10,7166667	1,072	0,02784862	45051
1742993,49	1485548,21	75	12,7466667	26,84	8,12733333	0,82	0,00701613	37031
1354654,17	1206555,33	72	35,26	27,7466667	11,3453333	0,9478	0,02473639	45005
1075668,39	1409410,99	69	17,0866667	27,9733333	11,2953333	0,9582	0,04779432	37039
1820406,18	2298754,25	60	8,1	21,64	9,426	2,40266667	0,01841173	36027
1019493,25	1265907,64	74	7,59333333	25,84	13,292	2,62206667	0,22026504	13223
1082914,22	1307948,07	52	5,6	18,5533333	13,4626667	1,84426667	0,09054461	13117
1643088,78	2230410,23	64	10,86	26,0733333	9,13866667	2,1882	0,05527246	42131
1378205,38	1816237,45	74	17,68	28,2066667	9,21533333	0,9346	0,0446132	54075
1047514,93	1092319,58	88	18,64	20,6266667	12,2753333	0,78873333	0,02840256	13053
1291852,93	1767655,85	95	21,6133333	31,4866667	9,79533333	1,63333333	0,07084971	54019
1339894,38	1828331,82	89	27,1133333	31,5066667	9,648	1,1206	0,06210596	54101
1580830,18	1632431,43	61	22,88	27,2066667	11,7146667	1,61766667	0,04547771	37185
1375560,83	1041347,41	69	19,5866667	27,8133333	10,686	0,7158	0,01113753	13191
1111168,4	1358919,71	61	13,8133333	24,9066667	11,8946667	1,21653333	0,03645697	13311
1585461,89	1896563,54	62	5,58	20,5333333	12,4573333	3,24253333	0,10948789	51153
1534801,61	1506770,3	76	16,22	27,1333333	13,0206667	1,67686667	0,0188727	37085
1516215,83	1357000,88	79	24,6533333	28,66	11,6373333	1,21946667	0,02825202	45067
1530166,58	2334378,47	64	14,02	24,7933333	9,01533333	1,72446667	0,03952038	36123
1305712,57	1092823,63	92	25,6	28,5	11,97	0,7536	0,01050018	13267
1145798,96	1015128,4	71	19,6133333	27,9933333	12,538	0,7988	0,01023408	13321
1440375,92	1230614,93	64	11,2333333	24,4066667	10,7366667	1,24146667	0,04448493	45035
1509820,65	2355302,69	64	8,84666667	21,54	9,15	1,92586667	0,03878577	36069
1050761,13	1306115,37	61	6,9	22,2733333	13,4193333	2,4658	0,15243852	13057
1255596,04	1058523,51	99	20,2333333	27,54	12,1646667	0,67113333	0,0093006	13161
1239676,07	1027479,85	74	22,3066667	27,8466667	12,2626667	0,6426	0,00846247	13069
1427249,19	2384624,13	69	12,96	28,18	10,236	3,3304	0,04472398	36073
1515917,07	2289805,22	73	14,04	25,5533333	8,94733333	1,43426667	0,04050009	36101
1780222,65	2119053,23	54	9,12	19,9066667	11,8026667	5,28186667	0,05552765	34021
1081737,82	1199158,14	78	18,62	28,1	12,452	1,939	0,14970316	13255
1294707,94	1976028,34	74	15,62	30,7066667	13,0726667	4,18213333	0,38692167	54051
1290225,11	2050386,63	77	12,44	30,7	12,9853333	4,88753333	0,33491169	54029
1621663,15	2043017,97	59	8,04666667	24,2666667	12,976	3,75386667	0,12760804	42133
1186006,55	1505916,7	64	16,9666667	27,2133333	10,1466667	1,2564	0,04508436	37115
1676336,57	1460888,58	80	15,22	26,54	9,95	0,99353333	0,01375003	37133
1274478,73	1857390,21	87	22,74	32,9533333	11,5193333	2,13566667	0,18478184	54013
1484760,24	1773042,45	70	11,76	24,68	10,714	1,6068	0,02950524	51125
1466332,78	1725627,04	73	18,9066667	25,04	11,0206667	1,81846667	0,03556332	51680
1108099,22	1389069,29	61	13,9	24,7866667	11,2473333	0,8272	0,03193818	13281

x	y	Rate	POV	SMOK	PM25	NO2	SO2	FIPS
1133468,49	1127832,39	90	21,42	25,5733333	12,34	1,1298	0,0624172	13225
1373699	1656374,61	49	12,24	26,48	9,65	1,364	0,04934977	51063
1397300,33	2323592,89	67	13,28	24,3066667	10,3993333	2,8934	0,05711464	36029
1091082,44	1221936,75	68	7,78666667	22,5266667	12,828	2,972	0,15517424	13151
1222120,62	1597141,25	94	18,8	28,8133333	10,3626667	2,10313333	0,04576182	51520
1501884,1	1978349,93	80	12,14	29,8	11,454	3,22973333	0,09698803	54065
1693664,73	1807551,07	66	13,46	23,46	11,4133333	1,77133333	0,05694911	51103
1439265,33	2324433,93	60	10,36	27,2866667	9,48733333	1,90046667	0,0481577	36121
1561950,65	2193635,4	61	12,7533333	26,1866667	9,87866667	2,32446667	0,09709751	42081
1406203,34	1789944,57	69	9,01333333	27,0533333	9,51	1,36086667	0,03196677	51017
1752689,96	1775435,96	94	19,7266667	25,6533333	10,2273333	2,27726667	0,03581466	51131
1407800,14	1532761,68	73	13,0466667	28,9733333	12,4546667	2,68926667	0,07666492	37057
1528493,41	1422787,01	78	27,6733333	31,0333333	12,6273333	1,4948	0,01660844	37155
1602057,96	2145493,01	63	12,42	28,0866667	10,7353333	2,72366667	0,14771475	42097
1341070,72	2158752,11	72	14,4733333	29,2933333	11,614	3,26333333	0,14555582	42121
1311583,65	1644628,34	72	13,6066667	29,0333333	9,46266667	1,18866667	0,03080953	51197
1221701,6	1734554,27	120	22,0933333	33,8133333	10,2106667	2,0086	0,08861411	54045
1018911,29	1363530,39	97	15,28	31,0466667	12,724	1,82813333	0,1020938	13213
1709880,26	1554745,8	77	18,3866667	25,4466667	10,264	1,27113333	0,00934098	37013
1132960,35	1418734,13	58	14,9933333	26,4333333	10,3393333	0,74353333	0,02798745	37113
1412014,76	2267732,7	64	14,8333333	27,32	9,81533333	1,8274	0,06263911	36009
1471985,02	2381307,73	58	13,0066667	21,5	9,776	3,0206	0,04078241	36055
1757950,48	2048967,11	74	6,8	24,1733333	12,6826667	5,45653333	0,05371555	34015
1073561,98	1003649,31	81	30,94	26,66	12,1726667	0,75106667	0,01140711	13037
1248201,79	1795739,98	87	14,7133333	28,16	11,2873333	2,83853333	0,17350221	54039
1620145,77	1926595,62	62	17,8	22,08	13,512	5,44513333	0,14948016	11001
1427140,31	1323124,2	68	17,94	26,52	12,3466667	1,3676	0,04749971	45085
1533134,27	1933160,16	68	7,1	23,0666667	12,0073333	3,05213333	0,08444785	51043
1295165,51	1869902,77	71	23,9666667	29,1933333	11,466	2,0594	0,15492688	54021
1743771,15	2031005,45	73	9,77333333	25,0666667	13,3426667	4,57446667	0,05740113	34033
1116496,72	1102091,28	73	27,0066667	28,4466667	12,2533333	1,0106	0,03677861	13193
1066361,41	1134167,66	84	20,86	26,7	12,3566667	0,9238	0,06290726	13263
1142014,88	1082724,61	54	25,5	28,0533333	12,4386667	0,8602	0,02236816	13093
1693700,16	1825328,97	63	12,88	24,0866667	11,5133333	1,921	0,07030229	51133
1505940,86	1513658,57	73	14,4	26,2933333	12,8333333	1,705	0,02582737	37105
1446754,82	1298002,95	71	24,02	26,7733333	11,794	1,27426667	0,053876	45027
1600385,58	1587249	70	15,2533333	25,6	12,644	1,63566667	0,0246834	37127
1676303,2	1683927,06	73	18,78	26,3466667	11,3833333	1,79393333	0,04932482	51620
1545565,8	1752895,88	67	15,2733333	27,9733333	11,2813333	1,7732	0,04914474	51049
1519860,21	1691095,4	77	16,9733333	27,1	11,1346667	1,7392	0,06361169	51037
1710730,91	1777230,97	60	8,98666667	24,1333333	11,8593333	1,87993333	0,0540941	51115
1686849,32	1794380,82	75	12,44	24,02	11,5986667	2,02226667	0,07039741	51119

x	y	Rate	POV	SMOK	PM25	NO2	SO2	FIPS
1456078,92	1810843,5	63	8,22666667	25,9533333	10,1906667	1,48706667	0,0339999	51015
1523121,26	1971372,33	85	11,32	29,9733333	12,1886667	3,4102	0,0901126	54003
1586724,02	1473472,83	71	19,3733333	27,38	12,5106667	1,35513333	0,01644959	37163
1552612,15	1659453,18	74	16,1466667	28,2733333	11,5046667	1,67753333	0,07118343	51117
1321432,03	1503977,36	69	12,14	25,78	11,99	2,24733333	0,10893105	37035
1318756,01	952508,079	83	23,0266667	30,7533333	10,7353333	0,753	0,02336033	13049
1611551,67	1520555,88	76	17,3866667	26,9	12,8926667	1,73693333	0,01444243	37191
1613685,11	2347274,96	65	14,4533333	24,92	8,912	1,68666667	0,02249919	36023
1376002,5	1542989,92	60	10,0533333	25,9733333	12,1293333	2,32653333	0,09012095	37059
1086530,75	1376942,09	62	14,7933333	25,2133333	11,3473333	0,90093333	0,04341876	13291
1435433,15	2357644,15	67	10,1533333	25,44	10,012	2,55853333	0,0458211	36037
1666241,84	1825977,97	68	17,2666667	28,7933333	11,484	2,17586667	0,08299332	51159
983511,43	1300252,57	81	17,04	26,8266667	13,256	2,2278	0,18984728	13115
1378521,09	1235921,64	65	25,44	25,5	11,2906667	1,0986	0,03455298	45009
1467229,21	2185659,85	63	11,6933333	28,0533333	10,0973333	1,74046667	0,09180156	42023
1666038,77	1894678,83	68	5,46	22,8866667	12,5853333	2,8886	0,13336423	24009
1652922,9	1773161,12	80	5,62666667	24,48	11,3473333	2,57853333	0,08385762	51127
968986,589	1322457,97	87	17,4933333	30,5066667	13,202	1,92766667	0,15694681	13055
1138683,74	1389143,35	68	14,74	24,92	11,1026667	0,7984	0,02639801	13241
1537131,81	2237598,11	60	13,8066667	25,3933333	9,30866667	1,61266667	0,05207866	42117
1295852,48	1430438,16	79	17,1	30,2666667	12,254	1,971	0,07420892	45021
1640404,66	1734227,05	67	8,90666667	24,8866667	11,5513333	2,59386667	0,07433514	51149
1492001,5	1300584,53	73	28,4466667	26,0866667	10,8713333	1,21213333	0,05055695	45089
1319815,77	1263983,77	68	15,08	25,4866667	11,7193333	1,27133333	0,02671331	45003
1031170,74	1134100,57	60	9,39333333	21,38	12,3966667	0,95226667	0,057191	13145
1307535,89	2220156,84	69	14,0266667	27,2266667	11,4126667	2,5386	0,07851527	42049
1594290,54	2390104	66	12,7733333	23,46	9,342	2,21706667	0,01941565	36067
1642325,5	1646805,02	64	23,2466667	27,1333333	11,928	1,4642	0,0309088	37131
1650666,5	1755747,35	73	10,5333333	27,4	11,5233333	2,8706	0,08023074	51036
1414401,14	1670082,52	55	11,7466667	26,6533333	10,4333333	1,821	0,06122508	51067
1755884,16	2146097,44	46	3,41333333	16,98	10,9913333	4,99673333	0,06330873	34019
1135652,35	1212270,72	68	15,7533333	26,8133333	12,3926667	1,7292	0,16726126	13159
1121743,84	1455205,27	75	17,3066667	30,68	10,2206667	0,77993333	0,03815915	37173
1727246,94	1596012,32	69	22,26	27,06	9,89	1,51893333	0,0129759	37187
1017773,77	1301385,21	86	12,5	27,6333333	13,4873333	2,54006667	0,19660064	13015
1337018,76	1693779,24	83	11,3733333	27,88	9,33333333	1,86513333	0,03055773	51071
1723227,18	1712308,89	81	16,8066667	26,1466667	11,1486667	2,5748	0,0517809	51740
1639584,2	1994032,37	63	7,75333333	22,46	13,484	4,4376	0,13559614	24005
1694933,49	2018400,3	79	8,46	26,0733333	13,76	4,75593333	0,08946195	24015
1637921,41	1539543,09	68	20,4933333	26,16	12,7273333	1,61846667	0,01355609	37079
1713827,55	1896169,76	77	14,64	26,3733333	12,348	3,3664	0,07150207	24019
1247170,21	1494736,27	70	14,94	29,2266667	10,408	1,6632	0,06043457	37111

x	y	Rate	POV	SMOK	PM25	NO2	SO2	FIPS
1256171,79	1199257,3	77	23,84	27,06	12,1493333	1,0764	0,04054591	13163
1169114,23	1143614,23	73	19,92	28,86	12,1873333	1,0758	0,07071084	13289
1563959,5	1936625,55	50	3,26	14,9066667	12,896	3,70153333	0,10273129	51107
1742060,76	1653103,17	74	18,0866667	26,0933333	10,1253333	1,4822	0,024836	37139
1314686,84	1596855,37	63	17,64	28,0866667	9,84866667	1,2924	0,04406683	37005
1294492,84	2022929,84	76	12,4466667	30,5866667	13,258	4,1472	0,39190171	54009
1397694,61	1569407,77	67	14,1466667	25,4533333	12,1146667	2,74333333	0,09013944	37067
1346738,57	1308601,15	66	11,1266667	24,7866667	12,044	1,4006	0,03262457	45063
1737530,54	1972072,59	74	11,9333333	26,0733333	13,4226667	3,42153333	0,06929497	10001
1619605,46	1966057,21	42	4,44666667	13,88	13,384	4,7924	0,14096512	24027
1694052,15	1981950,22	65	11,78	22,96	13,5326667	4,58153333	0,1020788	24029
1599218,37	1948856,14	44	5,82	13,5133333	13,3806667	4,83206667	0,13019144	24031
1679783,72	1647579,89	70	23,3533333	26,52	11,362	1,4432	0,03043198	37091
1464342,96	2278548,82	67	16,5066667	24,6933333	9,23333333	1,30906667	0,04473419	36003
1626897,94	1474774,31	65	20,34	27,7333333	11,69	1,2014	0,01624083	37061
1057680,93	1255772,74	58	16,1133333	18,8866667	13,4646667	3,74773333	0,16628214	13121
1363554,58	1904822,79	72	20,64	29,0733333	10,5946667	2,03566667	0,16523128	54001
1557688,93	1690682,64	68	19,6866667	29,4733333	11,2266667	1,62513333	0,05725367	51111
1181659,32	1602179,41	84	16,82	29,6933333	10,5266667	2,51213333	0,05958188	51169
1668389,55	1867466,36	70	7,68666667	22,0666667	12,39	2,42186667	0,1010433	24037
1199434,59	1398170,84	64	14,14	24,5866667	11,932	1,52793333	0,02772757	45077
1301978,78	1390787,37	79	16,78	30,0866667	12,2813333	1,3432	0,04170773	45087
1264596,65	1412493,18	73	14,0533333	26,6866667	12,328	1,83606667	0,05122649	45083
1422969,88	1149284,84	50	11,6133333	21,5666667	10,86	0,9116	0,01545759	45013
1394211,9	1150416,93	60	23,1866667	27,96	10,896	0,96233333	0,01649858	45053
1545014,12	1472239,09	76	16,4	25,54	13,0273333	1,6628	0,01614699	37051
1719077,99	2024828,34	66	9,39333333	23,24	14,1333333	4,9514	0,06845867	10003
1036449	1204099,49	66	9,92	23,6266667	12,6246667	1,9452	0,16092483	13077
989224,302	1271527,44	100	17,1466667	29,36	12,866	1,94513333	0,21103309	13233
1652614,19	1944776,74	68	5,71333333	21,74	13,526	4,5748	0,1419863	24003
1166148,01	1468980,25	61	14,2	28,4666667	9,82666667	1,02466667	0,03662095	37087
1257108,31	1839934,89	82	21,4066667	33,2133333	11,6673333	2,57526667	0,21716079	54087
1361162,48	1566766,53	69	12,8666667	29,06	11,572	2,0944	0,08331289	37197
1762733,34	1928788,66	69	11,6266667	27,42	12,2926667	3,60133333	0,05763594	10005
1214308,51	1450981,49	56	12,1466667	24,2666667	10,6233333	1,5424	0,04508857	37089
1134916,52	1304291,6	83	13,32	26,22	13,0533333	1,7394	0,05700776	13157
1263152,82	1465512,04	69	16,9933333	28,6066667	11,2106667	1,90053333	0,07589796	37161
1385083,59	1629473,1	65	15,1266667	29,8733333	10,398	1,78353333	0,07027099	51141
1552118,76	2098795,28	51	9,49333333	26,2533333	11,2406667	3,05686667	0,10686305	42067
1658497	2064716,08	52	8,76	23,2266667	12,8113333	4,389	0,10805667	42071
1486274,33	1210584,56	62	15,8466667	23,3666667	9,59066667	1,02213333	0,02791982	45019
1316547,47	1564807,21	68	16,1533333	28,6733333	10,6973333	1,5656	0,06409672	37193

x	y	Rate	POV	SMOK	PM25	NO2	SO2	FIPS
1199327,05	1215604,88	72	28,94	27,4733333	12,2506667	1,3462	0,1015702	13141
1137907,67	1329501,64	60	14,3066667	27,9	12,8386667	1,4372	0,03761753	13011
1759744,96	1602276,31	70	26,16	29,2733333	8,868	1,40166667	0,01136837	37177
1302019,11	2140915,86	62	13	27,64	12,1906667	3,37513333	0,14867851	42085
1532755,46	2104065,46	63	13,76	28,1133333	11,0406667	2,89346667	0,10175133	42087
1509494,38	2134790,06	49	14,4866667	18,3533333	10,5413333	2,77833333	0,11580153	42027
1701412,13	2065940,82	54	5,72	18,2266667	13,038	5,0668	0,07530573	42029
1508707,79	1391005,58	94	25,0666667	30,48	12,386	1,3276	0,01920734	45033
1372665,67	2140579,48	58	14,4	25,1	11,7446667	3,21026667	0,21193319	42031
1351833,84	1365215,89	76	19,3066667	27,66	12,394	1,2808	0,03338413	45039
1123623,57	1286900,03	82	11,34	27,3133333	13,074	2,22286667	0,0797185	13013
1534615,89	1718046,14	70	20,8	25,88	11,206	1,9544	0,04759393	51147
1602149,32	2165666,68	54	9,88	27,0266667	10,764	3,07593333	0,14008771	42093
1718996,52	2159377,21	58	8,12	24,1	11,0746667	4,21806667	0,06679771	42095
1187967,65	1432054,99	50	13,48	24,6133333	10,3706667	1,1282	0,03185788	37175
1303279,47	1613812,51	74	15,9533333	29,8133333	9,492	1,10646667	0,03579787	51077
1636048,47	1921250,68	44	8,4	19,8866667	12,9493333	5,0034	0,15359939	24033
1770779,48	2061835,74	67	11,2066667	24,2133333	12,352	6,06933333	0,05132937	34007
1220636,09	1515441,14	59	17,2533333	27,62	9,68933333	1,14713333	0,04549115	37199
1282819,85	1237590,18	83	21,76	26,1333333	12,2393333	1,5142	0,02978399	13245
1205964,13	1249909,65	68	26,3333333	29,1	12,3473333	1,23113333	0,07274157	13265
1279101	1097466,73	96	24,2	26,5	12,1286667	0,72773333	0,01172462	13279
1356925,68	2255464,75	60	16,2	26,8133333	10,8026667	2,7216	0,07880145	36013
1354669,09	1594547,91	77	15,7533333	29,58	10,7953333	1,80166667	0,06846022	37171
1606103,32	2096485,22	61	10,52	23,4066667	11,7473333	3,30546667	0,1334082	42043
1499577,03	1728467,53	65	12,9266667	27,3	10,992	1,86953333	0,03846089	51011
1282966,39	1585422,59	60	16,1666667	27,18	9,61933333	1,26833333	0,0387374	37009
1611502,06	1824508,56	80	10,58	26,24	11,5673333	2,698	0,09964954	51033
1833913,47	2172024,91	42	22,8333333	21,98	12,294	9,60313333	0,03823125	36047
1272812,08	1630310,37	88	15,7933333	29,8	9,416	1,04353333	0,03280152	51173
1270662,38	1362776,66	75	17,3266667	28,74	12,4873333	1,29066667	0,02561883	45059
1640681,28	1793703,59	77	7,08666667	25,3933333	11,5766667	2,6488	0,09106466	51101
1397734,61	1264367,99	63	22,8266667	25,24	11,5013333	1,3372	0,04503425	45075
1599741,12	1914646,42	41	5,20666667	15,58	13,124	4,19726667	0,1326588	51059
1276514,13	1506834,9	71	15,0933333	29,3666667	10,76	1,63973333	0,07676277	37023
1581708,46	1836635,82	69	6,28666667	22,9866667	11,6686667	2,361	0,08411982	51177
1700070,46	1927775,07	52	8,58	21,46	13,02	3,82606667	0,09170723	24041
1506294,91	2020785,08	61	10,5933333	26,6266667	11,404	3,00693333	0,10741616	42057
1382960,57	1511317,32	72	14,2266667	27,7333333	12,7286667	2,4164	0,08593719	37159
1680413,53	1588520,68	79	20,8133333	25,6466667	11,318	1,58566667	0,01497305	37117
1085492,19	1156919,19	73	18,34	29,4333333	12,2106667	1,12966667	0,09451793	13293
998958,95	1363198,73	79	14,5533333	27,76	13,1386667	2,32813333	0,11112316	13313

x	y	Rate	POV	SMOK	PM25	NO2	SO2	FIPS
1593460,96	1865910,32	68	4,83333333	20,14	12,2086667	2,85	0,11226636	51179
1287060	1531518,53	80	14,5666667	29,48	10,8473333	1,59833333	0,06940971	37027
1570613,58	2151895,61	43	11,78	24,9466667	10,6926667	2,86586667	0,13960698	42119
1814336,67	2406093,9	74	10,66	23,66	8,70866667	2,2682	0,00598084	36083
1296774,67	1462918,92	72	17	27,8933333	12,0173333	2,03133333	0,09813909	37045
1750040,41	1892975,82	81	13,3133333	24,7	11,5853333	3,46746667	0,05660617	24045
1362697,35	1463374,17	56	12,08	19,4733333	13,0333333	3,1472	0,08653307	37119
1124773,13	1263220,21	68	11,9333333	26,2866667	12,866	2,13786667	0,10860491	13297
1372531,96	2211179,99	62	11,6533333	27,86	10,6773333	2,4872	0,08727414	42123
1803087,26	2167525,56	50	9,12	20,5	12,0273333	9,225	0,04970654	34039
1073563,46	1061689,58	66	19,3066667	26,8	12,0566667	0,7182	0,0179033	13307
1242618,54	1091702,72	67	31,7266667	28,0533333	12,1166667	0,74693333	0,01397589	13309
1340289,63	2052713,6	67	11,3933333	25,3	13,1493333	5,3952	0,40572267	42003
1385974,15	2042662,98	60	9,58666667	25,2933333	12,426	3,95446667	0,39051756	42129
1075715,83	1331660,45	74	10,5133333	25,7333333	12,7853333	1,25293333	0,08137945	13085
1071151,41	1229322,89	69	16,1333333	24,1133333	13,236	3,69293333	0,1628903	13063
1638010,36	1430021,55	69	15,14	27,3933333	10,2126667	0,89686667	0,02021901	37141
1646097,35	1398296,88	66	14,0666667	24,5066667	9,52	0,9528	0,0197437	37129
1744615,77	1862073,29	87	21,86	27,2933333	11,4486667	2,33686667	0,04860733	24039
1239454,73	1157124,13	68	26,9333333	28,34	12,0833333	0,85146667	0,03631817	13167
1524638,91	2392871,67	64	10,3733333	25,3466667	9,67733333	2,41786667	0,03207525	36117
1793128,14	2075366,68	60	5,54	21,64	11,64	4,75593333	0,04581557	34005
1763878,69	1670135,14	85	10,4466667	26,9333333	9,66066667	1,61126667	0,02206854	37053
1336110,31	1428342,56	71	11,68	25,6066667	12,576	1,9606	0,07064819	45091
1252619,99	1294113,47	57	18,9	26,2066667	12,4706667	1,14653333	0,02755893	45065
1303698,65	1313763,03	57	16,9066667	26,8666667	12,2193333	1,08146667	0,02317928	45081
1443588,97	1486516,4	63	18,6666667	27,22	12,5373333	1,7368	0,03594427	37123
1776833,62	2347254,63	75	13,5	26,4133333	8,798	2,17386667	0,01110178	36039
1170904,67	1378993,2	61	13,52	25,98	12,0166667	1,13353333	0,02453453	45073
1741874,53	2176288,67	61	6,24666667	22,5533333	10,152	4,03666667	0,06094802	34041
1376413,78	2098078,46	61	12,1466667	27,5866667	12,25	4,1228	0,34546719	42005
1517690,2	2172273,44	69	13,8933333	26,2533333	10,0826667	2,2874	0,09200365	42035
1111782,17	1008830,39	76	25,54	24,98	12,4866667	0,8886	0,01055159	13095
1804774,03	2472917,92	71	11,82	26,2666667	8,42533333	1,56386667	0,00332353	36115
1094443	1121507,84	72	24,2466667	28,86	12,34	0,96053333	0,05255585	13269
1045720,26	1271398,87	56	9,79333333	18,7666667	13,5653333	3,91293333	0,19218633	13067
1540247,38	1956911,5	79	9,60666667	25,5866667	12,4726667	3,5824	0,09080267	54037
1344969,2	1628917,5	66	15,0733333	29,3066667	9,62733333	1,3884	0,0476102	51035
1335318,35	1176789,81	81	21,1866667	26,5733333	11,5266667	0,74066667	0,01967182	13251
1703903,79	2595627,36	70	16,2133333	27,1666667	7,44133333	0,791	0,00172546	36033
1474589,85	2334175,81	67	11,5733333	23,9333333	9,17333333	1,77586667	0,04191675	36051
1174285,61	1763155,47	87	18,9866667	31,5933333	10,934	2,31533333	0,15609841	54099

x	y	Rate	POV	SMOK	PM25	NO2	SO2	FIPS
1222296,6	1670564,65	77	22,9066667	31,5133333	9,558	1,38693333	0,04609945	51027
1457757,86	1375458,33	80	20,5933333	27,84	12,722	1,50986667	0,02819108	45031
1378047,72	1327144,98	65	14,4666667	21,8066667	12,4713333	1,59313333	0,04011898	45079
1556868,48	1895059,37	64	5,87333333	21,6266667	11,9146667	2,56546667	0,0882615	51061
1447341,37	1951614	73	14,74	27,72	10,7173333	2,6718	0,1443067	54057
1206178,92	1720396,18	117	24,92	33,3933333	10,1746667	2,02386667	0,07471842	54059
1258731	1714945,02	96	22,8133333	33,22	9,56866667	1,48313333	0,04758733	54109
1778263,32	1880783,86	68	10,22	23,6666667	11,2526667	2,64373333	0,03958471	24047
1477747,46	1409462,87	111	25,0533333	31,22	12,71	1,44186667	0,01977629	45069
1569406,99	1781445,46	60	7,00666667	22,2	11,478	2,08486667	0,06482331	51075
1569996,51	1715378,67	79	18,9133333	30,1866667	11,32	1,9602	0,0585518	51135
1462694,7	1908032,11	59	13,5066667	29,9733333	10,206	1,93506667	0,09033027	54031
1730519,8	1640428,51	61	17,1133333	26,1933333	10,1446667	1,24346667	0,02090575	37143
1231184,55	1403328,2	63	12,8733333	23,2	12,3366667	1,69726667	0,03564866	45045
1086388,99	932261,374	87	24,1333333	26,8533333	11,9173333	0,6314	0,00905079	13087
1118028,34	975040,571	79	25,48	28,46	12,4626667	0,74593333	0,00880913	13205
1171003,42	1094923,93	66	19,76	28,5266667	12,4293333	0,81333333	0,02531628	13235
1476981,12	1945927,23	77	15,8066667	30,4466667	10,5686667	2,5572	0,10356995	54027
1300125,81	1698264,07	82	20,3466667	30,86	9,20866667	1,25933333	0,03237728	54055
1345865,98	1723197,08	72	15,9266667	28,7666667	9,28533333	1,54693333	0,03184743	54063
1311268,89	1800627,42	82	19,1533333	29,52	9,838	1,2936	0,0742697	54067
1212726,84	1809882,28	73	9,84666667	25,2333333	11,8366667	3,13433333	0,27146042	54079
1656055,09	1843192,98	81	14,44	27,42	11,804	2,3918	0,11307779	51193
1388452,8	1756142,05	76	11,4266667	26,7533333	9,616	1,57573333	0,03046243	51005
1804735,87	2182643,52	55	14,7333333	21,4733333	11,8693333	10,5624	0,04698671	34013
1294398,01	1205638,26	81	25,2133333	27,7666667	11,84	1,0434	0,02677244	13033
1763272,52	1826266,62	89	17,8133333	26,0066667	10,694	2,126	0,03360347	51001
1329138,94	1135899,88	67	23,3266667	22,4533333	11,7213333	0,88113333	0,01446308	13031
1302225,4	1176479,14	80	26,4333333	27,5533333	11,828	0,76833333	0,02143123	13165
1576397,47	1531400,34	75	13,5266667	25,8933333	13,1326667	2,07713333	0,01812385	37101
1135752,78	1594803,17	98	23,9133333	30	10,4706667	1,66606667	0,06322446	51105
1916558,97	2219152,49	60	6,49333333	21,8266667	10,4666667	5,66606667	0,0237588	36103
1439176,42	1711113,41	61	7,98	25,4	10,6473333	1,9292	0,04012971	51019
1549846,1	1864900,94	69	9,38	24,6466667	11,658	2,25926667	0,0689561	51047
1366794,28	1138324,1	81	10,22	25,72	11,286	0,98706667	0,01406846	13103
1381105,21	1867124,34	64	18,0066667	28,4733333	9,47333333	1,14366667	0,08417722	54083
1604974,7	1792754,05	69	4,87333333	21,92	11,7813333	2,72693333	0,08428972	51085
1508841,42	1939047,69	72	7,12	25,38	11,472	2,8908	0,0834059	51069
1175972,34	1630930,85	102	20,78	30,2933333	9,79466667	2,1212	0,05595153	51195
1631028,74	1883315,98	70	6,52666667	21,62	12,286	3,22733333	0,12646991	24017
1757119,24	2553752,17	63	12,5466667	22,38	7,366	1,0304	0,00189042	36031
1573084,28	1762556,11	68	5,98	24,4133333	11,4666667	2,05246667	0,06999977	51145

x	y	Rate	POV	SMOK	PM25	NO2	SO2	FIPS
1611344,22	1767745,31	84	21,6333333	24,86	12,282	3,31433333	0,08653221	51760
1549405,81	2355026,71	63	12,32	26,2866667	9,49666667	1,98093333	0,03413613	36099
1004456,19	1226100,57	77	15,3733333	26,6466667	12,5766667	1,8114	0,18685076	13045
1694092,15	1771191,17	84	9,06666667	25,0266667	11,782	1,87706667	0,06358423	51073
1155637,24	1286419,26	61	26,0733333	19,0866667	12,9166667	1,7166	0,06927095	13059
1500908,07	1659731,3	71	17,28	27,8933333	11,2626667	1,963	0,10573504	51083
1039034,9	1010413,01	66	30,2866667	26,92	11,9546667	0,6786	0,01287012	13061
1323351,94	2018108,58	70	10,1133333	26,4733333	13,104	4,34826667	0,42084654	42125
1400663,57	1366623,45	79	13,8	27,0266667	12,7333333	1,36513333	0,03622422	45055
1800201,82	1993222,58	71	9,82666667	25,06	12,2266667	2,24253333	0,03960326	34009
1281918,49	1736878,05	79	17,9733333	30,8066667	9,54666667	1,51093333	0,05355175	54081
1262487,03	1255436,35	63	7,14	20,9266667	12,438	1,55073333	0,03339277	13073
1200452,77	1092122,29	81	21,46	27,8866667	12,3053333	0,78686667	0,02021571	13091
1198413,32	1837359,46	93	18,16	31,9666667	12,1953333	3,34646667	0,32854737	54053
1093298,55	1436078,01	67	18,9133333	29,88	10,6	0,8758	0,04435937	37075
1378511,3	1996883,49	71	17,74	28,8	12,122	3,28953333	0,36387532	42051
1605034,7	1749188,26	64	6,05333333	21,0333333	11,846	2,79173333	0,08104605	51041
1502186,46	1871116,69	66	12,7666667	29,88	10,9313333	1,3806	0,0522102	51139
1393012,02	1697984,57	54	6,18666667	21,98	10,0226667	2,14	0,03561159	51161
1049875,39	977774,675	76	27,16	26,52	12,31	0,65046667	0,01143749	13099
1340728,1	1665907,14	74	13,7933333	29,42	9,49866667	1,44373333	0,03471435	51155
1261607,33	1915191,25	81	13,4133333	31,0933333	12,5973333	3,822	0,30969914	54073
1489940,52	1896026,7	68	9,49333333	26,1933333	10,9233333	1,608	0,06737963	51171
1413285,62	1479056,19	70	13,2933333	26,78	12,7193333	2,1168	0,04940727	37167
1265973,08	1661117,01	85	16,86	29,04	9,21666667	1,15013333	0,03358323	51185
1089047,82	1347799,34	72	14,4466667	25,3133333	12,2866667	1,1558	0,05479218	13187
1186214,8	1335649,72	64	17,2266667	26,8266667	12,7793333	1,24886667	0,02708039	13147
1234436,88	1892845,96	77	15,5	30,1	12,8646667	4,1584	0,33496652	54107
1323938,68	1882650,82	87	18,6533333	30,5666667	11,062	2,1718	0,13511538	54041
1660955,71	1685061,66	73	14,9666667	26,3466667	11,5893333	1,6544	0,0481272	51175
1181516,34	940200,05	76	23,5866667	26,68	11,7453333	0,6898	0,00745401	13027
1367612,98	1010378,06	75	16,02	24,2066667	10,648	0,67553333	0,01668144	13127
1150012,81	1117688,33	70	12,1533333	24,4533333	12,3853333	0,939	0,04782439	13153
1241432,54	1280516,51	76	18,1333333	27,7466667	12,4493333	1,26526667	0,03431777	13181
1237814,76	1758586,58	117	19,4066667	33,3333333	10,3966667	2,07786667	0,114511	54005
1227296,87	1849010,75	79	15,8466667	28,8666667	12,198	3,2112	0,30764743	54035
1417357,19	1959178,55	49	13,2733333	22,7866667	10,7326667	2,2226	0,20854219	24023
1118726,88	1176047,92	62	12,6066667	25,1133333	12,358	1,69286667	0,13645791	13207
1151027,26	1272737,38	50	7,32	17,7866667	12,7633333	1,8114	0,08811926	13219
1182710,86	1282146,61	68	14,1466667	26,1466667	12,5706667	1,29606667	0,06162576	13221
1190786,06	1161735,55	71	18,1533333	26,3533333	12,3466667	1,0366	0,0765142	13319
1432264,47	1913616,71	51	15,1866667	28,7066667	10,108	1,8462	0,12778116	54023

x	y	Rate	POV	SMOK	PM25	NO2	SO2	FIPS
1400892,18	1909096,14	64	16,08	27,8733333	9,78933333	1,54833333	0,15104295	54093
1859870,44	2189947,37	47	5,72666667	18,82	11,0513333	13,0592	0,03418871	36059
1588690,73	2436988,08	79	14,7066667	28,18	9,054	2,1862	0,01402763	36075
1046643,49	1062076,28	71	28,12	27,22	12,1146667	0,7876	0,01841575	13259
1702334,91	1662006,44	71	15,3866667	25,74	10,8386667	1,51246667	0,03577051	37073
1696239,43	1752034,18	68	4,8	18,16	12,0826667	2,2666	0,0648406	51199
1828676,41	2261806,23	61	5	19,14	9,526	3,03466667	0,02710716	36079
1477903,63	1803796,55	63	13,7466667	25,9533333	10,6226667	1,55673333	0,03125429	51820
1387460,16	2371359,14	72	12,0933333	27,2733333	10,9293333	3,77	0,05168716	36063
1093091,07	983304,178	63	24,6066667	26,1066667	12,4873333	0,76453333	0,00983018	13007
1575637,41	1982508,77	55	5,34666667	19,7266667	13,2526667	3,65773333	0,10854251	24021
1647136	2571722,28	77	16,6733333	25,84	7,97466667	0,98606667	0,00309673	36089
1225504,01	1068331,62	73	28,9866667	28,6466667	12,2313333	0,654	0,01161513	13271
1302309,01	1014963,54	92	18,4466667	28,42	11,5266667	0,6454	0,01058425	13229
1486637,48	1345492,05	77	18,0733333	26,1333333	11,896	1,3086	0,03560711	45041
1481683,85	1490671,68	61	12,48	23,5933333	12,6946667	1,64273333	0,02562493	37125
1408466,27	2139659,35	58	13,1333333	27,0666667	11,4146667	2,76473333	0,2108717	42065
1778629,86	2185127,61	49	4,00666667	16,96	10,4086667	6,31466667	0,05215598	34027
1617731,32	1374464,23	68	14,1333333	29,4533333	9,85266667	0,79106667	0,01984657	37019
1623563,73	1624783,65	71	24,4133333	27,96	12,1786667	1,65533333	0,02754656	37083
1191026,25	976480,722	81	21,28	28,88	12,4653333	0,68046667	0,00771704	13075
1420276,64	1202878,04	81	21,5066667	27,52	10,6753333	1,0124	0,03209786	45029
1531529,21	1286308,02	64	17,56	26,1	9,91133333	1,06613333	0,04238067	45043
1473661,73	1612333,89	73	16,92	29,2733333	11,4573333	2,2072	0,11489407	37033
1697494,86	1510019,83	72	14,3133333	24,64	9,94533333	1,35093333	0,00991089	37049
1149571,17	1245032,1	60	13,4866667	23,5266667	12,6286667	1,63273333	0,12866699	13211
1751441,69	2080937,92	80	22,82	29,42	13,31	6,893	0,05803297	42101
1817579,99	2161994,38	58	10,3466667	23,7066667	11,7066667	9,3702	0,03290444	36085
1814401,26	2204752,46	48	5,88666667	18,34	11,3753333	9,01833333	0,04112128	34003
1567737,52	2374439,53	61	12,3	27,7733333	9,52066667	2,13026667	0,02703454	36011
1505394,76	1617645,78	74	13,6066667	27,3733333	11,6253333	2,13246667	0,10469636	37145
1572483,12	1737932,54	82	10,1466667	28,5666667	11,296	1,99473333	0,06443133	51007
1472787,87	1845021,53	51	24,0333333	20,6666667	11,0546667	1,5598	0,04264868	51660
1348482,64	1875559,15	74	19,86	28,5266667	10,222	1,7296	0,10687214	54097
1711335,09	2323192,08	57	14,0466667	25,22	8,61	1,116	0,01506907	36025
1819905,97	2180131,94	53	15,0533333	22,0066667	12,1353333	10,9449333	0,04445196	34017
1326934,88	1040349,84	83	20,16	27,8466667	11,4626667	0,69566667	0,01044651	13305
1824854,7	2083480,23	65	8,42	24,9133333	11,718	3,2172	0,03761536	34029
1670078,51	1731435,22	66	11,72	25,02	11,464	2,4372	0,06855303	51181
1237850,25	1611510,93	72	12,7533333	26,56	9,882	1,6892	0,04137664	51191
1104919,49	1404573,35	57	15,26	26,1733333	10,8686667	0,78346667	0,03300759	37043
1695430,8	2369555,45	59	13,9133333	23,3466667	8,38466667	1,2666	0,01047439	36077

x	y	Rate	POV	SMOK	PM25	NO2	SO2	FIPS
1709056,97	2192003,06	69	9,77333333	24,9666667	9,652	2,6352	0,06026466	42089
1712275,97	2490987,62	73	10,1733333	22,3533333	7,32266667	0,88666667	0,00355774	36041
1662198,21	1556593,38	63	20,4333333	23,14	12,17	1,58353333	0,01312751	37147
1208146,79	992422,795	88	19,5933333	29,0666667	12,2473333	0,6612	0,00777428	13019
1784395,91	2248556,37	64	10,9133333	23,1066667	9,442	2,7786	0,03140666	36071
1780443,17	2385249,29	64	11,4266667	21,46	8,83533333	2,47533333	0,00757217	36001
1222799,47	1184983,22	73	22,82	27,5733333	12,2353333	1,01813333	0,06540251	13303
1523847,66	1851669,48	57	10,3466667	24,4533333	11,2933333	1,63686667	0,04760546	51113
1519532,58	1907327,62	80	9,50666667	26,8666667	11,3726667	1,96666667	0,07172545	51187
1635788,18	1587878,71	72	22,8266667	28,2733333	12,6373333	1,6888	0,0186096	37065
1392412,03	1101415,41	65	17,26	24,1666667	11,1093333	1,0518	0,00994462	13051
1607599,9	1715108,81	78	10,7533333	28,22	11,4773333	2,14313333	0,0645533	51053
1298399,5	1132052,35	77	23,82	27,38	12,0666667	0,7762	0,01507908	13043
1466190,93	1748958,49	74	11,9466667	27,1933333	10,7006667	1,6734	0,02986779	51009
1402346,68	1700671,94	82	17,5133333	28,96	10,36	2,14506667	0,03527919	51770
1394272,73	1439225,34	60	9,28666667	22,0866667	12,874	2,0826	0,04897361	37179
1269211,48	1560293,27	47	18,8466667	20,7933333	9,73133333	1,24493333	0,04243414	37189
1528372,04	1883439,69	56	8,28666667	21,96	11,3833333	1,71813333	0,06307779	51157
1273217,53	1895127,72	80	17,58	31,44	12,086	3,05326667	0,24868401	54085
1613596,31	1922262,73	43	6,76666667	13,98	13,568	5,00593333	0,14499371	51013
1016197,21	1165749,05	74	17,2733333	26,14	12,2626667	1,25866667	0,09715126	13285
1360362,39	975480,85	72	11,96	25,0533333	10,468	0,75213333	0,0272533	13039
1748150,05	2623675,29	69	13,7466667	23,7	7,742	1,1768	0,00120889	36019
1342753,03	1236098,03	73	21,4066667	27,2066667	11,5026667	1,09566667	0,02756709	45011
1384086,39	1403329,78	68	15,9466667	29,1333333	12,678	1,59146667	0,04155183	45057
1251796,43	1540204,59	66	17,58	26,4733333	9,70333333	1,1974	0,04133199	37011
1463675,83	1978141,47	64	15,16	26,38	11,0373333	2,70633333	0,14048924	24001
1490869,03	1438927,83	79	23,4933333	28,0866667	12,8546667	1,42613333	0,01773952	37165
1301449,84	2105533,83	62	13,28	27,2866667	12,9646667	4,10193333	0,21806551	42073
1307257,32	2185193,46	66	14,26	27,5666667	11,5146667	2,92306667	0,09747856	42039
1596515,27	2251514,69	65	12,86	26,88	9,10666667	1,47146667	0,0530943	42015
1111388,86	1205843,94	91	14,74	29,1133333	12,3813333	2,01446667	0,160238	13035
1288990,75	1057012,93	77	19,44	27,4333333	12,026	0,7092	0,00888677	13001
1604638,58	2296816,05	61	9,80666667	23,3133333	9,08	1,73453333	0,03447175	36107
1231540,19	1530025,88	72	15,86	27,7133333	9,962	1,08766667	0,04274761	37121
1233904,43	1327933,58	70	16,4	26,9266667	12,6666667	1,25253333	0,02427076	45001
1283161,73	1929358,59	86	16,5	30,6133333	12,554	3,47933333	0,29892331	54095
1701943,47	1690601,23	74	11,6333333	23,4666667	10,886	1,93346667	0,04945547	51800
1343252,82	1397436,61	81	18,5533333	30,4666667	12,3953333	1,4288	0,04506228	45023
1214993,69	1275513,36	76	19,9666667	28,1066667	12,4813333	1,18453333	0,04855345	13317
1198918,04	1027877,28	77	21,7733333	29,3	12,5566667	0,67886667	0,00960119	13155
1590775,7	2508226,52	72	14,8533333	26	8,586	1,80386667	0,0072077	36045

x	y	Rate	POV	SMOK	PM25	NO2	SO2	FIPS
1803710,6	2031834,04	69	10,9266667	25,9866667	11,7633333	3,11413333	0,04034153	34001
1745633,55	2117385,44	60	5,28666667	21,84	11,7513333	5,07733333	0,06283556	42017
1494044,09	1537413,37	51	10,5933333	21,0866667	12,5473333	2,254	0,04019597	37037
1349138,01	1524687,67	65	10,9866667	25,2133333	12,0873333	2,31533333	0,09785903	37097
1506866,67	2076919,91	61	12,5866667	26,34	10,9973333	2,75806667	0,11433182	42061
1577416,08	1428761,54	70	21,8133333	27,52	11,878	1,13306667	0,01923658	37017
1326661,91	1000420,53	99	18,2266667	30,06	11,0446667	0,67713333	0,01553918	13025
1178880,85	1066819,62	74	27,2266667	28,12	12,5873333	0,72913333	0,01505637	13315
1702968,97	1737731,4	68	13,8066667	24,9466667	12,1153333	2,6156	0,05076559	51700
1415121,01	1826079,89	56	12,7466667	25,2066667	9,14933333	1,11346667	0,04154865	51091
1752695,58	1705589,89	63	7,51333333	22,4533333	10,156	2,11973333	0,03788188	51810
1471173,59	1853131,42	51	8,92	23,8	10,7206667	1,48593333	0,04919018	51165
1339648,62	1064862,15	84	20,8866667	27,9533333	11,35	0,81473333	0,00952202	13183
1074360,62	1179283,59	85	11,2333333	25,3933333	12,3126667	1,46893333	0,12703505	13231
1060350,19	1213666,98	48	5,18	18,26	12,9646667	2,4054	0,16384844	13113
1276775,87	1033243,56	97	20,66	29,7733333	11,9253333	0,6524	0,00856262	13005
1545650,8	1556677,83	53	9,10666667	17,7266667	12,79	2,10593333	0,03194102	37183
1310463,89	1346837,15	69	16,2066667	27,22	12,446	1,22053333	0,02481782	45071
1158669,44	976083,731	80	22,5933333	27,98	12,4933333	0,695	0,00810521	13071
1047465,29	1170294,89	66	19,1533333	28,1866667	12,2406667	1,24706667	0,10973214	13199
1663852,67	2010730,91	64	6,06666667	21,7133333	13,4413333	4,54106667	0,11912654	24025
1197155,37	1651004,36	91	20,36	31,3266667	9,614	1,8648	0,05338612	51051
1780207,58	2151254,93	47	4,27333333	15,7933333	11,512	6,02833333	0,05694037	34035
1765770,9	2492083,5	63	10,66	24,2733333	7,88933333	1,2494	0,00286146	36113
1645709,5	2137970,77	67	11,4333333	28,6733333	10,458	3,33933333	0,1064991	42107
1571533,32	2060710,06	52	6,64666667	20,9933333	12,386	3,9012	0,10730805	42041
1427489,36	1864438,44	46	13,7933333	27,1266667	9,41133333	1,29006667	0,06815885	54071
1303204,7	1667921,16	71	12,9866667	28,0333333	9,20133333	1,154	0,02896316	51021
1538083,92	1990247,17	66	10,1733333	25,1133333	12,5146667	3,61593333	0,09678692	24043
1116766,6	1142102,13	75	16,6533333	27,5133333	12,2253333	1,2422	0,08211809	13079
1347254,65	1767353,56	80	17,3933333	28,5733333	9,438	1,29613333	0,04013834	54025
1731682,36	2066067,56	67	9,14666667	24,6066667	13,4273333	6,20053333	0,06136937	42045
1087754,1	1033206,98	85	27,5533333	27,72	12,1626667	0,79533333	0,01352705	13273
1697727,85	2257139,2	64	11,9133333	26,9066667	8,94066667	1,24606667	0,03140685	42127
1478293,44	1705612,14	68	11,7933333	26,74	10,8853333	1,83793333	0,0517628	51031
1011067,18	1330535,35	90	14,0133333	28,5666667	13,3093333	2,06213333	0,15785579	13129
1367109,72	1682899,99	55	17,4	19,98	9,554	1,96406667	0,03666562	51121
1519909,16	1754957,03	69	18,9	28,8133333	11,1173333	1,6644	0,03705136	51029
1798253,64	1604192,67	67	9,18	25,1066667	8,21666667	1,1462	0,00834759	37055
1130864,96	1359803,54	60	13,6533333	25,5866667	12,17	1,1376	0,03086892	13137
1204390,57	1770898,49	108	23,9066667	33,2	10,938	2,4316	0,17149628	54043
1454142,9	2133464,92	61	13,9666667	27,3933333	10,7753333	2,16806667	0,17547352	42033

x	y	Rate	POV	SMOK	PM25	NO2	SO2	FIPS
1739211,26	2274248,01	71	15,28	25,8066667	8,68066667	1,48413333	0,02333841	36105
1429094,61	2011809,21	49	12,9266667	27,22	11,344	2,63953333	0,25766174	42111
1409669,85	1733520,05	53	6,28	23,2066667	10,0653333	1,77733333	0,02879422	51023
1559975,42	1808855,19	76	10,14	26,84	11,4026667	2,0466	0,0579132	51109
1538800,03	2027269,53	55	8,53333333	24,3866667	12,3393333	3,27973333	0,09579465	42055
1058244,95	935725,169	81	23,4466667	27,06	12,0626667	0,6452	0,01091031	13253
1763204,22	1568816,33	69	22,0533333	28,3733333	8,82666667	1,18213333	0,00777405	37095
1727250,26	2227452,26	58	8,24666667	25,2466667	8,89733333	1,82566667	0,04171911	42103
1201057,91	1311274,26	74	18,6466667	28,7933333	12,7906667	1,11153333	0,03308846	13105
1213040,65	1358418,68	70	13,92	26,8	12,7266667	1,50146667	0,02415823	45007
1827575,68	2186232,24	43	17,9133333	18,74	12,272	10,6904	0,04245672	36061
1743539,37	2375273,68	67	12,1666667	25,1066667	8,22466667	1,49533333	0,00857669	36095
1226878,06	1235056,75	84	23,9	28,78	12,29	1,33246667	0,06279052	13301
1106655,28	1065071,71	70	25,0133333	25,1066667	12,2246667	0,89373333	0,01814369	13261
1376536,2	1719306,87	63	10,0466667	25,9733333	9,652	1,887	0,03033509	51045
1274318,47	1149501,5	84	25,76	27,9066667	11,9926667	0,74966667	0,02142951	13107
1284743,83	1815780,54	103	24,78	32,1066667	10,682	2,30173333	0,12296528	54015
1331673,94	1981403,64	74	16,4066667	27,3466667	12,764	4,15273333	0,39405763	42059
1841804,18	2182258,76	39	14,54	20	12,0106667	14,6895333	0,03737797	36081
1648771	2428756,76	63	14,28	26,1866667	8,44533333	1,834	0,00893315	36065
1510291,41	1835864,46	61	8,12666667	24,6733333	11,1786667	1,53306667	0,0387469	51079
1185759,53	1795938,61	83	18,7866667	28,8266667	11,5886667	2,78946667	0,23816696	54011
1350699,79	1959197,78	61	17,48	22,56	12,1993333	3,297	0,32675206	54061
1616258,35	1915795,15	48	8,16666667	16,3333333	13,338	4,85686667	0,14719117	51510
1410676,89	2085864,83	53	15,9266667	24,2133333	11,798	3,69353333	0,33912064	42063
1604981,21	1998805,92	61	4,86666667	21,1866667	13,1526667	3,2682	0,12737091	24013
1301405,28	1948184,98	83	17,3866667	30,0466667	12,572	3,59986667	0,32806033	54103
1287458,28	978514,547	78	21,82	28,4066667	11,3006667	0,63946667	0,01326026	13299
1356170,29	1926659,92	78	18,14	28,2266667	11,3806667	2,71493333	0,22915798	54091
1676368,49	1494220,47	70	17,8533333	27,88	10,4693333	1,34533333	0,01198182	37103
1378190,66	1186307,96	74	22,88	26,2266667	11,0173333	0,8428	0,02325097	45049
1717605,12	1735088,9	73	13,46	23,96	11,932	2,72933333	0,05644722	51650
1675785,76	1757564,59	48	6,69333333	18,48	11,7073333	2,33253333	0,07445928	51095
1678063,2	2228579,15	64	12,0333333	26,96	9,42066667	1,9836	0,04698714	42069
1475160,9	2078209,21	65	13,1533333	26,7666667	11,0193333	2,9708	0,17482969	42013
1153423,39	1354138,05	83	17,7733333	27,3266667	12,522	1,28673333	0,02793323	13257
1576210,28	1388925,42	80	22,56	29,2666667	11,0266667	0,96346667	0,02047931	37047
1330908,7	1463722,6	82	14,5333333	28,5933333	12,6306667	2,61266667	0,10438915	37071
1433973,06	1406304,16	81	20,3666667	29,2866667	12,806	1,38406667	0,02727738	45025
1775867,93	2302403,31	61	11,6	23,7533333	9,09	2,03246667	0,01797732	36111
1205489,37	1480787,54	62	14,0333333	25,36	10,22	1,41746667	0,04639695	37021
1686744,85	1615496,93	70	23,18	26,56	11,1793333	1,53386667	0,01956157	37015

x	y	Rate	POV	SMOK	PM25	NO2	SO2	FIPS
1794089,08	2208556,38	52	13,4933333	21,4733333	10,5013333	5,99206667	0,04291301	34031
1395339,66	1290613,6	53	16,7066667	24,6266667	11,8593333	1,47193333	0,047343	45017
1813564,82	2226668,67	44	10,7733333	18,28	10,4286667	5,0918	0,03671156	36087
1647038,65	1822048,17	74	13	25,6466667	11,3893333	2,438	0,10394836	51057
1498854,68	1579049,04	58	13,1933333	17,0466667	12,0893333	1,98126667	0,07261332	37135
1759786,53	2212712,26	63	4,68666667	21,7133333	9,34266667	3,157	0,04592587	34037
1046569,81	1355924,33	63	15,9666667	28,58	12,278	1,24706667	0,09288003	13123
969740,273	1351864,51	95	15,1466667	29,5	12,974	2,11746667	0,12057446	13295
1265105,22	1324705,81	65	16,1733333	26,5266667	12,3466667	1,12513333	0,02202201	45047
1656798,07	1798464,5	76	12,4866667	27,4266667	11,3326667	2,38926667	0,090404	51097
1566640,72	2086593,73	70	8,56	25,4066667	11,6453333	3,32426667	0,11305412	42099
982225,869	1372225,11	73	11,64	27,5466667	13,1	2,69766667	0,09758515	13047
1156077,94	1437274,96	58	17,2733333	25,8333333	10,082	0,866	0,02870949	37099
1259366,38	1100453,58	87	22,3533333	24,5666667	12,062	0,6928	0,01361859	13209
1164434,82	1216682,8	67	15,64	24,84	12,4566667	1,53333333	0,14887022	13237
1436054,55	1767464,31	61	10,6133333	24,9466667	9,96933333	1,71273333	0,02822363	51163
1729079,64	1519154,32	64	16,4266667	26,1733333	8,90466667	1,13546667	0,00776219	37137
1071815	1095905,65	64	22,0733333	29,1133333	12,1773333	0,76373333	0,03150351	13197
1820299,03	2125684,6	61	6,18	20,8133333	11,7353333	5,0214	0,04422624	34025
1117382,25	1236773,73	75	12,4666667	25,4733333	12,752	2,46813333	0,14437953	13217
1092478,71	1088370,43	77	17,7466667	27,5	12,1566667	0,82893333	0,02705802	13249
1459738,42	1658545,82	72	12,98	28,16	11,1073333	1,93373333	0,1020601	51143
1222483,59	1632905,46	83	18,1666667	29,4866667	9,67933333	1,76566667	0,04744314	51167
1164606,54	1036137,61	75	26,0666667	28,2866667	12,63	0,7356	0,01112687	13287
1249192,82	1873699,75	91	18,8533333	30,5266667	12,2553333	2,75933333	0,27650887	54105
1594645,86	1677031,06	74	20,12	28,4266667	11,4853333	1,5654	0,04961426	51025
1690124,34	1710628,9	70	9,01333333	22,8333333	11,3	2,22853333	0,05883096	51093
1699831,3	1961649,54	67	6,63333333	21,7	13,2706667	4,30586667	0,09655702	24035
1235765,27	1216094,35	70	16,1333333	28,7733333	12,2133333	1,25933333	0,05953033	13125
1727248,17	2098844,03	53	5,42666667	20,4333333	12,528	5,06093333	0,06394979	42091
1385195,63	1483013,98	66	10,2266667	24,38	12,9173333	2,67746667	0,07830983	37025
1676333,77	2111342,48	57	10,7733333	24,4	11,8073333	4,2978	0,08621708	42011
1479695,44	1250590,85	73	12,8933333	26,2266667	10,4233333	1,19866667	0,04941601	45015
1336346,91	1943122,15	80	16,3	26,5	12,2873333	3,2686	0,28828512	54049
1617200,64	1769762,46	64	8,18	20,6133333	12,0186667	3,25013333	0,08572237	51087
1835028,39	2232826,67	46	8,54666667	18,1266667	10,3533333	5,46933333	0,03241281	36119
1211163,79	943251,366	79	20,2	24,98	11,708	0,71066667	0,00749642	13185

ANEXO 7. Resultados númericos de GWPCA

GWPCA BÁSICO

```
   ***********************************************************************
   *                       Package   GWmodel                             *
   ***********************************************************************
   Program starts at: 2022-05-28 12:19:20
   Call:


   Variables concerned:  POV SMOK PM25 NO2 SO2
   The number of retained components:  5
   Number of data points: 666
   ***********************************************************************
   *                Results of Principal Components Analysis               *
   ***********************************************************************
Importance of components:
                          Comp.1    Comp.2    Comp.3    Comp.4
Standard deviation     1.4370788 1.1488786 0.8847254 0.7325775
Proportion of Variance 0.4130391 0.2639844 0.1565478 0.1073340
Cumulative Proportion  0.4130391 0.6770235 0.8335713 0.9409053
                           Comp.5
Standard deviation     0.54357493
Proportion of Variance 0.05909474
Cumulative Proportion  1.00000000


Loadings:
     Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
POV   0.523  0.353  0.290  0.467  0.548
SMOK  0.493  0.428 -0.415  0.114 -0.623
PM25 -0.262  0.584  0.674 -0.221 -0.295
NO2  -0.550  0.083 -0.094  0.808 -0.168
SO2  -0.336  0.586 -0.530 -0.258  0.443


   ***********************************************************************
   *   Results of Geographically Weighted Principal Components Analysis  *
   ***********************************************************************


   *********************Model calibration information*********************
   Kernel function for geographically weighting: bisquare
   Adaptive bandwidth for geographically and temporally  weighting: 599 (number of neare
st neighbours)
   Distance metric for geographically weighting: A distance matrix is specified for this
model calibration.
```

```
   ****************     Summary of GWPCA information:    *****************
   Local variance:
              Min.  1st Qu.   Median  3rd Qu.   Max.
   Comp.1 1.282622 1.683960 1.824477 2.395155 2.6841
   Comp.2 0.555828 0.916985 1.160684 1.643178 1.9032
   Comp.3 0.392443 0.610668 0.702313 0.766852 0.8433
   Comp.4 0.205417 0.293388 0.322072 0.532814 0.6496
   Comp.5 0.076753 0.094729 0.156598 0.320213 0.3640
   Local Proportion of Variance:
                  Min.  1st Qu.   Median  3rd Qu.     Max.
   Comp.1      41.4858  42.8474  44.3139  45.9204  50.0141
   Comp.2      21.1924  24.0847  27.0608  29.7649  35.1776
   Comp.3      10.3097  12.8811  13.8869  19.3736  19.8280
   Comp.4       6.2235   7.9618   8.1643   9.4830  11.1092
   Comp.5       2.2507   2.4889   3.9977   5.6506   7.0526
   Cumulative 100.0000 100.0000 100.0000 100.0000 100.0000

   ************************************************************************
   Program stops at: 2022-05-28 12:19:34
```

GWPCA ROBUSTO

```
************************************************************************
   *                       Package   GWmodel                            *
   ************************************************************************
   Program starts at: 2022-05-28 13:28:55
   Call:

   Variables concerned:  POV SMOK PM25 NO2 SO2
   The number of retained components:  5
   Number of data points: 666
   ************************************************************************
   *                Results of Principal Components Analysis              *
   ************************************************************************
Importance of components:
                          Comp.1    Comp.2    Comp.3    Comp.4
Standard deviation     1.4370788 1.1488786 0.8847254 0.7325775
Proportion of Variance 0.4130391 0.2639844 0.1565478 0.1073340
Cumulative Proportion  0.4130391 0.6770235 0.8335713 0.9409053
                           Comp.5
```

```
Standard deviation     0.54357493
Proportion of Variance 0.05909474
Cumulative Proportion  1.00000000


Loadings:
     Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
POV   0.523  0.353  0.290  0.467  0.548
SMOK  0.493  0.428 -0.415  0.114 -0.623
PM25 -0.262  0.584  0.674 -0.221 -0.295
NO2  -0.550  0.083 -0.094  0.808 -0.168
SO2  -0.336  0.586 -0.530 -0.258  0.443


   ***********************************************************************
   *   Results of Geographically Weighted Principal Components Analysis  *
   ***********************************************************************


   **********Model calibration information**********
   Kernel function for geographically weighting: bisquare
   Adaptive bandwidth for geographically and temporally  weighting: 137 (number of neare
st neighbours)
   Distance metric for geographically weighting: A distance matrix is specified for this
model calibration.


   ****************     Summary of GWPCA information:    *****************
   Local variance:
               Min.    1st Qu.     Median    3rd Qu.   Max.
   Comp.1 0.33908755 0.93240362 1.26205868 1.60942400 3.0537
   Comp.2 0.01931541 0.06094344 0.10148548 0.15231506 0.5796
   Comp.3 0.00577147 0.01827779 0.02684802 0.03512656 0.1185
   Comp.4 0.00172934 0.00608968 0.00839103 0.01226525 0.0325
   Comp.5 0.00027708 0.00102184 0.00208132 0.00372808 0.0161
   Local Proportion of Variance:
                    Min.    1st Qu.     Median    3rd Qu.     Max.
   Comp.1      68.269516  85.504341  89.287201  93.018350  96.8591
   Comp.2       1.400571   4.672402   7.515873  11.455892  27.8910
   Comp.3       0.521318   1.395047   1.759798   2.435368   7.3489
   Comp.4       0.218444   0.464270   0.627274   0.829090   2.7052
   Comp.5       0.012048   0.087005   0.159207   0.262842   0.7403
   Cumulative 100.000000 100.000000 100.000000 100.000000 100.0000


   ***********************************************************************
   Program stops at: 2022-05-28 13:29:52
```