

# A Climatological Analysis by means of Soft Computing Models

Ángel Arroyo<sup>1</sup>, Emilio Corchado<sup>2</sup>, Verónica Tricio<sup>3</sup>

**Abstract.** This research analyzes the meteorological conditions of four different places in Spain. The case study is based on real data provided by the AEMET (Meteorological Spanish Agency) in 2009. Thirteen variables with atmospheric conditions are considered. Different Statistical and Soft Computing Models are applied to shown the great variability of the environmental conditions in the four well selected places. The results are confirmed by the Annual Environmental Summarized 2009 provided by the AEMET.

**Keywords.** Artificial neural networks, soft computing, meteorology, statistical models, environmental conditions.

## 1. Introduction

The meteorology is defined as the science that deals with the phenomena of the atmosphere, especially weather and weather conditions. It is necessary to distinguish between the current conditions and its evolution so called atmospheric weather, and the average conditions during a long period that is known as local climate or region.

Meteorology includes atmospheric chemistry and atmospheric physics with a major focus on weather forecasting. Climatology is the study of atmospheric changes (both long and short-term) that define average climates and their change over time, due to both natural and anthropogenic climate variability.

In this study it is proved the validity of soft computing models to analyze the atmospheric conditions in four different places in Spain. The data are provided by the AEMET (Meteorological Spanish Agency) [1]. The results obtained by these

---

<sup>1</sup> Department of Civil Engineering, University of Burgos, c\Francisco de Vitoria s/n 09006 Burgos, Spain. aarroyop@ubu.es

<sup>2</sup> Department of Computer Science and Automatic, University of Salamanca, Plaza de La Merced s/n, 37008, Salamanca, Spain. escorchado@usal.es.

<sup>3</sup> Department of Physics, University of Burgos, Burgos, Spain. vtricio@ubu.es

studies are confirmed by the 2009 Annual Climatology Summarized provided by the AEMET [2].

This study is part of an ambitious project about the application of Soft Computing Models to Atmospheric Pollution [3, 4] and Climate Conditions [5, 6].

Soft computing [7, 8] consists of various technologies which are used to solve inexact and complex problems. It is used to investigate, simulate, and analyze complex issues and phenomena in an attempt to solve real-world problems.

The rest of this study is organized as follows. Section 2 presents the statistical and soft computing methods applied throughout this research. Section 3 details the case studies and Section 4 describes the experiments and results. Finally, Section 5 sets out the conclusions.

## 2. Soft Computing and Statistical Models

Several statistical and soft computing models are used in this study, although the results are only shown of those that offer the best performance.

**Principal Components Analysis (PCA).** PCA [9] gives the best linear compression of the data in terms of least mean square error and can be implemented by several artificial neural networks [10, 11, 12]. The basic PCA network [11] applied in this study is described by the next three equations (Eq.(1) to Eq.(3)): an  $N$ -dimensional input vector at time  $t$ ,  $x(t)$ , and an  $M$ -dimensional output vector,  $y$ , with  $W_{ij}$  being the weight linking input  $j$  to output  $i$ , and  $\eta$  being the learning rate. Its activation and learning may be described as follows:

Feedforward step, "Eq. (1)":

$$y_i = \sum_{j=1}^N W_{ij} x_j, \forall i \quad (1)$$

Feedback step, "Eq. (2)":

$$e_j = x_j - \sum_{i=1}^M W_{ij} y_i \quad (2)$$

Change weights, "Eq. (3)":

$$\Delta W_{ij} = \eta e_j y_i \quad (3)$$

**A Cooperative Exploratory Projection Pursuit Method.** Cooperative Maximum Likelihood Hebbian Learning (CMLHL) is an extended version of Maximum Likelihood Hebbian Learning (MLHL) [13] adding lateral connections which have been derived from the Rectified Gaussian Distribution [13]. The resultant net can find the independent factors of a data set but does so in a way that captures some type of global ordering in the data set.

Consider an  $N$ -dimensional input vector  $x$ , an  $M$ -dimensional output vector  $y$  and a weight matrix  $W$ , where the element  $W_{ij}$  represents the relationship between

input  $x_j$  and output  $y_i$ , then as is shown in [14, 15], the CMLHL can be carried out as a four-step procedure:

Feed-forward step, outputs are calculated “Eq. (4)”:

$$y_i = \sum_{j=1}^N W_{ij} x_j, \forall i \quad (4)$$

Lateral activation passing step, “Eq. (5)”:

$$y_i(t+1) = [y_i(t) + \tau(b - Ay)]^+ \quad (5)$$

Feedback step, “Eq. (6)”:

$$e_j = x_j - \sum_{i=1}^M W_{ij} y_i, \forall j \quad (6)$$

Weights update step, learn the neural network, “Eq. (7)”:

$$\Delta W_{ij} = \eta y_i \text{sign}(e_j) |e_j|^{p-1} \quad (7)$$

Where  $t$  is the temperature,  $[\ ]^+$  is necessary to ensure that the  $y$ -values remain in the positive quadrant,  $\eta$  is the learning rate,  $\tau$  is the "strength" of the lateral connections,  $b$  the bias parameter,  $p$  a parameter related to the energy function, and  $A$  is a symmetric matrix used to modify the response to the data.

### 3. A Real Climatology Case Study

A climatology case study is presented in this research. This study analyses with data on the evolution of different meteorological parameters using the records of four meteorological data acquisition stations. These four stations are placed on:

1. Villafria (Burgos). Geographical coordinates 422122N; 033757W. Labelled as B on Figs. 1 to 3.
2. Santiago (A Coruña). Geographical coordinates 425316N; 082439W. Labelled as C on Figs. 1 to 3.
3. Almeria's Airport. Geographical coordinates 365047N; 022125W. Labelled as A on Figs. 1 to 3.
4. Mallorca. Geographical coordinates 393320N; 023735E. Labelled as M on Figs. 1 to 3.

As it was described in Section 1, the data were provided by AEMET. In this The study year 2009 is analyzed (the monthly summarized database is used [1]). In this research, the following variables were analyzed:

1. Maximum absolute temperature. Maximum temperature in the whole month, C°.
2. Minimum absolute temperature. Minimum temperature in the whole month, C°.
3. Frozen days. Number of days with lower temperature of below 0 C°.
4. Total amount of monthly rain in mm the month in mm.
5. Maximum rain in the whole month. Maximum daily rain in mm.

6. Number of rainy days over 0.1 mm. Days with a significant amount of rain (>0.1 mm).
7. Number of rainy days over 10 mm.
8. Rainy days. Number of days in the month with any amount of rain.
9. Snowy days. Number of days in the month with any amount of snow.
10. Hailey days. Number of days in the month with any amount of hail.
11. Average Sunstroke. Average number of sunny hours per day in the month., in hours.
12. Maximum absolute atmospheric press in tenths of hectopascal (hPa) in the whole month.
13. Minimum absolute atmospheric press in tenths of hectopascal (hPa) in the whole month.

This study examines the performance of several statistical and soft computing methods when analyzing the above-mentioned climatology variables, in order to study the behavior of the climatology in the four geographical points mentioned above.

#### 4. Experiments and Results

The study, which forms part of a more ambitious project in terms of Soft Computing Models applied to Meteorology [5, 6], is based on monthly summarized climatology data in 2009, 12 records per geographical location, thirteen variables, (see Section 3).

The information represented at each point is visually labeled in Fig. 1, Fig. 2 and Fig. 3, where the month is expressed in a number from 1 to 12 and the geographical place is represented by a letter (e.g., 6M means June in Mallorca). Letter B represents Burgos, letter A represents Almeria, letter M represents Mallorca and letter C represents Coruña. All data was were normalized for the study.

In the first step the projection of PCA (Fig.1) model and CMLHL (Fig.2) model to the described data are shown.

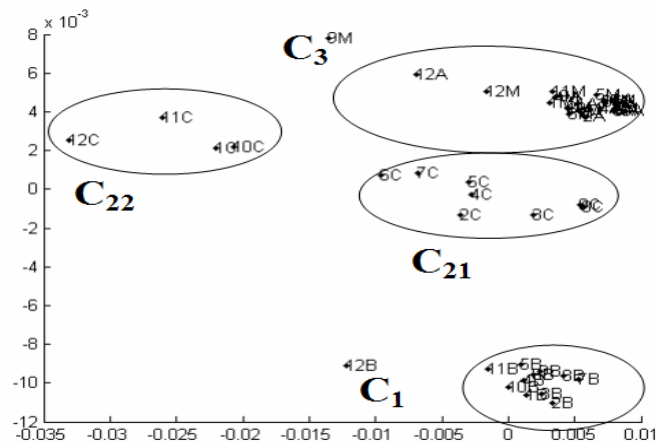


Fig. 1. PCA projections of the monthly meteorological information for the four geographical places in Spain, year 2009.

Fig. 1 PCA. This method identifies four clusters. Cluster  $C_1$  groups most of the samples corresponding to location of Burgos. Clusters ( $C_{21}$  and  $C_{22}$ ) groups the samples corresponding to Coruña. Cluster  $C_3$  groups the samples corresponding to location of Mallorca and Almeria. These two geographical points had a similar weather in 2009 as we will be able to prove later on this section.

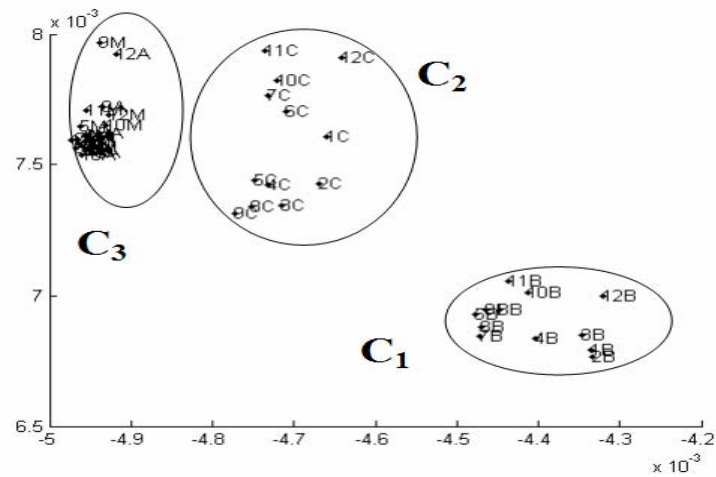
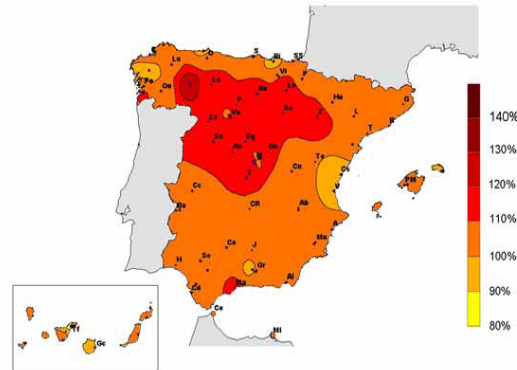


Fig. 2. CMLHL projections of the monthly meteorological information for the four geographical places in Spain, year 2009.

Fig. 2 shows the projection generated by the CMLHL method. Cluster  $C_1$  represents the samples of Burgos in the same way than in Fig. 1, but in this occasion all the samples are compacted inside the cluster. In the cluster  $C_2$ , all the samples belonging to Coruña are represented. In Fig.1, PCA was not able to group in a cluster all the samples of this city located at the north west of Spain. In the same way than in Fig. 1,  $C_3$  groups the samples corresponding to location of Mallorca and Almeria.

In the next step CMLHL is applied only to the records belonging to Almeria and Mallorca, in order to analyze cluster  $C_3$  of (Fig. 1 and Fig. 2), applying hierarchical CMLHL [16].





(c)

Fig. 4. 2009 Annual Climatology Summarized in terms of (a) Temperatures (b) Precipitations (c) Isolation.

Fig. 4 ((a), (b) and (c)) represents the climatology in Spain in 2009 in terms of temperature, precipitations and isolation. These three aspects summarize the thirteen variables applied in this study, (see Section 3). By analyzing these aspects (Fig. 4a, 4b, 4c) it would be easy to appreciate the difference between the weather in Burgos respected to the other three geographical places. The difference between the city of Coruña and the other three geographical places and the concordance between Mallorca and Almeria.

## 5. Conclusions

The PCA model and CMLHL model are able to perform properly with meteorological information, although the number of dimensions is high, the reduction to two variables is providing very interesting results. However, CMLHL offers a finer response, as the results with this model fix the reality in a better way than PCA.

The big meteorological differences among well chosen geographical points are also proved. The results shown (in Fig.1, Fig.2 and Fig.3) are supported by the 2009 Annual Climatology Summarized (Fig. 4a, 4b, 4c).

**Acknowledgments.** This research has been partially supported through JCyL project BU035A08, and the Spanish Ministry of Science and Innovation project TIN2010-21272-C02-01. The authors would also like to thank the vehicle interior manufacturer, Grupo Antolin Ingeniería, S.A., within the framework of the project MAGNO2008 - 1028.- CENIT Project funded by the Spanish Ministry.

## 6. References

- 1 AEMET – Agencia Española de Meteorología, Public Access Database <http://www.aemet.es/es/servidor-datos/acceso-datos> Last Access: 03/01/2011.
- 2 AEMET - Agencia Española de Meteorología, 2009 Annual Climatology Summarized [http://www.aemet.es/documentos/es/elclima/datos\\_climat/resumenes\\_climat/anuales/res\\_a\\_nual\\_clim\\_2009.pdf](http://www.aemet.es/documentos/es/elclima/datos_climat/resumenes_climat/anuales/res_a_nual_clim_2009.pdf) Last Access: 03/01/2011.
- 3 Arroyo, A., Corchado, E., Tricio V. Atmospheric Pollution Analysis by Unsupervised Learning. Intelligent Data Engineering and Automated Learning - IDEAL 2009, pages: 767–772, 2009.
- 4 Arroyo, A., Corchado, E., Tricio V. Computational Methods for Immision Analysis of Urban Atmospheric Pollution. 9<sup>th</sup> International Conference Computational and Mathematical Methods in science and engineering, Gijón 09.
- 5 Arroyo, A., Corchado, E., Tricio V. Soft Computing Models for an Environmental Application. International Workshop on Soft Computing Models in Industrial Applications-SOCO 2010. Springer Advances in Soft Computing, ISSN1615-3871 (Print) 1860-0794 (Online).
- 6 Corchado, E., Arroyo, A., Tricio V. Soft Computing Models to Identify Typical Meteorological Days. Logic Journal of the IGPL, Oxford University Press Online ISSN 1368-9894 - Print ISSN 1367-0751, JULY 2010 \* Impact factor: 0.326.
- 7 Zadeh, Lofti A. Fuzzy logic, neural networks, and soft computing. Commun. ACM, Vol. 37, No. 3. (1994), pp. 77-84.
- 8 Subudhi, B., Morris, A. S. Soft computing methods applied to the control of a flexible robot manipulator. 149-158. Applied Soft Computing, Volume 9, Number 1, January 2009
- 9 Hyvärinen, A., Karhunen, J. and Oja, E. Independent Component Analysis, Wiley, 2002.
- 10 Oja, E., Ogawa, H. and Wangviwattana, J. Principal Components Analysis by Homogeneous Neural Networks, part 1. The Weighted Subspace Criterion, IEICE Transaction on Information and Systems E75D (1992), pp. 375-366.
- 11 Fyfe, C. and Baddeley, R. Non-linear data structure extraction using simple Hebbian networks, Biological Cybernetics 72(6) (1995), pp. 541-533.
- 12 Oja, E. Neural Networks, Principal Components and Subspaces, International Journal of Neural Systems, vol. 1 (1989), pp. 68-61.
- 13 Corchado, E., MacDonald, D., Fyfe, C. Maximum and Minimum Likelihood Hebbian Learning for Exploratory Projection Pursuit. Data Min. Knowl. Discov. 8(3) (2004), pp. 203-225.
- 14 Corchado, E., Fyfe, C. Connectionist Techniques for the Identification and Suppression of Interfering Underlying Factors. International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI), 17(8): 1447-1466 (2003).
- 15 Corchado, E., Han, Y., Fyfe, C. Structuring global responses of local filters using lateral connections. Journal of Experimental and Theoretical Artificial Intelligence 15(4): 473-487 (2003)
- 16 Corchado, E., Burgos, P., Rodriguez, M., Tricio, V. A Hierarchical Visualization Tool to Analyse the Thermal Evolution of Construction Materials. CDVE 2004, Lecture Notes in Computer Science 3190 Springer: 238-246.