

Soft Computing techniques applied to a case study of air quality in industrial areas in the Czech Republic

Ángel Arroyo¹, Emilio Corchado^{2,5}, Verónica Tricio³, Laura García-Hernández⁴ and Václav Snášel⁵

Abstract. This multidisciplinary research analyzes the atmospheric pollution conditions of two different places in Czech Republic. The case study is based on real data provided by the Czech Hydrometeorological Institute along the period between 2006 and 2010. Seven variables with atmospheric pollution information are considered. Different Soft Computing models are applied to reduce the dimensionality of this data set and show the variability of the atmospheric pollution conditions among the two places selected, as well as the significant variability of the air quality along the time.

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

1. Introduction

Soft computing [1] [2] [3] consists of various techniques 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.

There are many studies based on the application of different soft computing paradigms [4] [5] [6] to the field of air quality and environmental conditions. Some of them are based in the Czech Republic [7] where an air quality model is approached. Other studies visualize high dimensionality data sets with environ-

¹Department of Civil Engineering, University of Burgos, Burgos, Spain. aarroyop@ubu.es.

²Departamento de Informática y Automática, University of Salamanca, Salamanca, Spain. escorchado@usal.es.

³Department of Physics, University of Burgos, Burgos, Spain. vtricio@ubu.es.

⁴Area of Project Engineering, University of Cordoba, Spain. ir1gahel@uco.es.

⁵Department of Computer Science, VSB-Technical University of Ostrava, Czech Republic IT4Innovations, Ostrava, Czech Republic. vaclav.snasel@vsb.cz.

mental information in order to find patterns of behavior in the climatology and pollution in local areas [8] or global areas [9].

In this study it is tested the validity of soft computing models to analyze the atmospheric pollution in two different places in Czech Republic. The data are provided by the ISKO (Czech Hydrometeorological Institute) [10].

The Czech Republic is the object of this study for its important tradition in the field of environmental conditions analysis [11]. This is a contribution to the study of environmental pollution in areas of high industrial activity.

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

2. Soft Computing techniques

In order to analyze data sets with atmospheric pollution information, several dimensionality reduction techniques are applied, although the results are only shown for those that achieve the best performance.

Principal Components Analysis (PCA). PCA [12] gives the best linear data compression in terms of least mean square error and can be implemented by several artificial neural networks [13] [14].

Isometric Featured Mapping (ISOMAP). ISOMAP [15] is a nonlinear dimensionality reduction method which preserves the global properties of the data. Methods for nonlinear dimensionality reduction have proven successful in many applications, although the weakness of a method such as Multidimensional Scaling (MDS) [16] is that they are based on Euclidean distances and do not take the distribution of the neighboring data points into account. ISOMAP nonlinear dimensionality reduction [17] resolves this problem by attempting to preserve pairwise geodesic (or curvilinear) distance between data points. Geodesic distance is the distance between two points measured over the manifold. ISOMAP defines the geodesic distance as the sum of edge weights along the shortest path between two nodes (computed using Dijkstra's algorithm [18], for example). The doubly-centered geodesic distance matrix K in ISOMAP is presented by "Eq. 1":

$$K = \frac{1}{2} HD^2 H \quad (1)$$

Where $D^2=D_{ij}^2$ means the element wise square of the geodesic distance matrix $D=[D_{ij}]$, and H is the centring matrix, given by "Eq. 2":

$$H = I_n - \frac{1}{N} e_N e_N^T \quad (2)$$

In which $e_N=[1 \dots 1]^T \in \mathbb{R}^N$.

The top N eigenvectors of the geodesic distance matrix represent the coordinates in the new n -dimensional Euclidean space.

Local Linear Embedding (LLE). LLE [19] is an unsupervised learning algorithm that computes low-dimensional, neighborhood-preserving embeddings of high-dimensional inputs [20]. LLE attempts to discover nonlinear structure in high dimensional data by exploiting the local symmetries of linear reconstructions. Notably, LLE maps its inputs into a single global coordinate system of lower dimensionality, and its optimizations — though capable of generating highly nonlinear embeddings — do not involve local minima.

Suppose the data consist of N real-valued vectors x_i , each of dimensionality D , sampled from some smooth underlying manifold. Provided there is sufficient data (such that the manifold is well-sampled), it is expected that each data point and its respective neighbors will lie on or close to a locally linear patch of the manifold. The method can be defined as follows:

1. Compute the neighbors of each vector, x_i .
2. Compute the weights W_{ij} that best reconstruct each vector x_i from its neighbors minimizing the cost in by constrained linear fits, “Eq. 3”.

$$\varepsilon (W) = \sum_i \left| x_i - \sum_j W_{ij} x_j \right|^2 \quad (3)$$

3. Finally, find point y_i in a lower dimensional space to minimize:, “Eq. 4”:

$$\Phi (Y) = \sum_i \left| y_i - \sum_j W_{ij} y_j \right|^2 \quad (4)$$

This cost function in “Eq. 4” like the previous one in “Eq. 3” is based on locally linear reconstruction errors, but here the weights W_{ij} are fixed while optimizing the coordinates y_i . The embedding cost in “Eq. 4” defines a quadratic form in the vectors y_i .

3. An Atmospheric Pollution real Case Study

This multidisciplinary study is focused on the analysis of real pollution data recorded in the Czech Republic [21]. It is based on the information collected by the national network stations in the Czech Republic. The National air pollution network (SIS) is an open network [22], member of the International Air Quality network AIRNOW [23]. This network classifies the stations as traffic, urban, suburban, rural and industrial stations. In this study two industrial data acquisition stations are selected in order to analyze its air pollution.

1. Station coded as TOPRA, located in Ostrava-Privoz, Region Moravskoslezsky. Coordinates: 49° 51' North latitude 18° 16' East longitude, 207 masl (meters above sea level), the official estimated population of Ostrava was 310,464 inhabitants in 2011. The main industrial activity is focused on steel works, concentrates on metallurgy and machine engineering.

2. Station coded as SVELA, located in Veltrusy, Region Stredocesky. Coordinates: 50° 16' North latitude 14° 19' East longitude, 174 masl, 1,600 inhabitants. The main industrial activity is focused on engineering, chemical industry, food industry, glass industry, ceramics manufacture and printing industry.

The years under study are the following: 2006, 2008, 2009, 2010. Year 2007 is omitted because there are multiple data corrupt or missing, (the monthly summarized database is used). The study is focused on the twelve months of 2006, 2008, 2009 and 2010 and the 2 stations analyzed, 12 samples for each station and year, one sample per month. There is a total of 96 samples. In this research, the following seven variables were analyzed:

- Nitrogen Oxide (NO) - $\mu\text{g}/\text{m}^3$, Nitrogen Dioxide (NO_2)- $\mu\text{g}/\text{m}^3$, Nitrogen Oxides (NO_x) - $\mu\text{g}/\text{m}^3$, Particulate Matter (PM10) - $\mu\text{g}/\text{m}^3$, Sulphur Dioxide (SO_2) - $\mu\text{g}/\text{m}^3$. The choice of these contaminants is because they represent the majority of air emissions in areas of industrial activity.
- Benzene (BZN) - $\mu\text{g}/\text{m}^3$, Toluene (TLN) - $\mu\text{g}/\text{m}^3$. These contaminants belong to the Volatile Organic Compounds (VOCs). VOCs are linked to automobile emissions as a result of incomplete gas combustion. VOCs are also used in the production of paints, varnishes, lacquers, thinners, adhesives, plastics, resins and synthetic fibers.

This study examines the performance of several statistical and soft computing methods, in order to analyze the behaviour of the pollutants cited above to compare the air quality of an important industrial area in Czech Republic as Ostrava with other industrial point located in a very different location in Czech Republic as Veltrusy, along a wide period of time.

4. Results and Discussions

The results shown below correspond to the mean monthly values of the parameters described in the previous section, after applying the soft computing methods described above for clustering and dimensionality reduction:

Table 1. Label interpretation for (Figure 1, Figure 2 and Figure 3).

First Parameter - Month	Second Parameter - Year	Third Parameter - Station
(1 to 12)	2006, 2008, 2009, 2010	O – Ostrava - Privoz V – Veltrusy

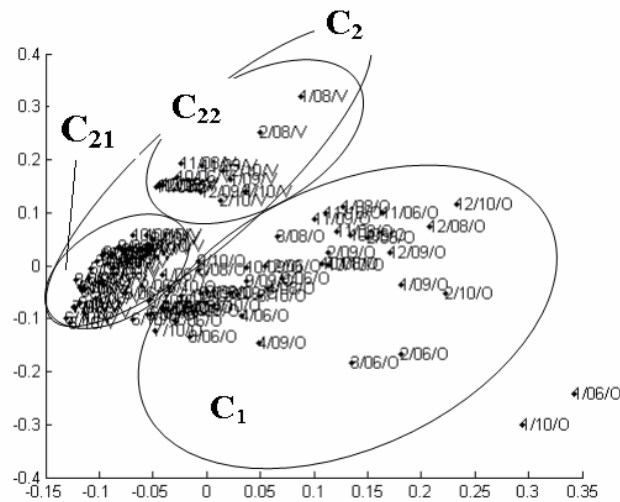


Fig. 1. PCA projections.

Fig. 1. PCA identifies two main clusters, C_1 corresponds with the samples related to Ostrava station and C_2 corresponds with the samples of Veltrusy station. In cluster C_1 there is an important concentration of samples corresponding to the seasons of spring and summer, (see Table 2). The samples in C_1 corresponding to the seasons of winter and autumn are not so concentrated and present high levels of air pollution. Table 2 indicates the high values of PM_{10} and NO_x . The points labeled with '1/10/O' and '1/06/O' represents the days when the levels of SO_2 and specially PM_{10} reach values which exceed the health protection fixed in $40 \mu g/m^3$ [24], in more than $80 \mu g/m^3$ in both cases. In [24], the report about stations with annual average concentration corroborates this fact. Cluster C_2 (see Fig. 1) contains the samples corresponding to the seasons of spring, summer and autumn in Veltrusy, where the levels of air pollution are lower than winter in Veltrusy and lower than in Ostrava. In the season of winter, the concentration of NO_x is higher than in the rest of the year in both stations.

Table 2. Samples of each cluster identified by PCA and the range of values of each variable and units.

Cluster #	Samples						
C_1	2/10/O 3/10/O 4/10/O 5/10/O 6/10/O 7/10/O 8/10/O 9/10/O 10/10/O 11/10/O 12/10/O 1/09/O 2/09/O 3/09/O 4/09/O 5/09/O 6/09/O 7/09/O 8/09/O 9/09/O 10/09/O 11/09/O 12/09/O 1/08/O 2/08/O 3/08/O 4/08/O 5/08/O 6/08/O 7/08/O 8/08/O 9/08/O 10/08/O 11/08/O 12/08/O 2/06/O 3/06/O 4/06/O 5/06/O 6/06/O 7/06/O 8/06/O 9/06/O 10/06/O 11/06/O 12/06/O. Total number of samples: 46						
Var.	BZN ($\mu g/m^3$)	TLN ($\mu g/m^3$)	SO_2 ($\mu g/m^3$)	PM_{10} ($\mu g/m^3$)	NO_2 ($\mu g/m^3$)	NO ($\mu g/m^3$)	NO_x ($\mu g/m^3$)
Range of values	2.2-21.9	4.3 -26.4	4.3 – 20.3	30.7 – 87.2	3.0-28.9	3.0-26.1	25.0-76.4

Description	Groups all the samples from the station based in Ostrava, except '1/10/O' and '1/06/O'. High values of PM10 and NO _x in the season of winter						
C₂₁	3/10/V 4/10/V 5/10/V 6/10/V 7/10/V 8/10/V 9/10/V 10/10/V 3/09/V 4/09/V 5/09/V 6/09/V 7/09/V 8/09/V 9/09/V 10/09/V 11/09/V 3/08/V 4/08/V 5/08/V 6/08/V 7/08/V 8/08/V 9/08/V 10/08/V 11/08/V 12/08/V 1/06/V 11/06/V 12/06/V. Total number of samples: 30						
Range of Values	0.4 – 10.3	0.2 – 22.9	2.6 – 6.4	9.7 – 24.0	11.4 – 27.8	0.9 – 15.8	12.0 – 46.5
Description	Groups most of the samples from the station based in Veltrusy, samples corresponding to the seasons of spring, summer and autumn. High values only in NO _x .						
C₂₂	1/10/V 2/10/V 11/10/V 12/10/V 1/09/V 2/09/V 12/09/V 1/08/V 2/08/V 11/08/V 12/08/V 1/06/V 11/06/V 12/06/V. Total number of samples: 16						
Range of Values	1.3 – 10.5	1.2 – 14.6	2.6 – 11.8	11.5 – 28.1	23.9 – 40	3.4 – 17.9	30.0 – 75.0
Description	Groups most of the samples from the station based in Veltrusy corresponding to the season of winter. The samples with very high levels of NO _x are located out of the cluster.						

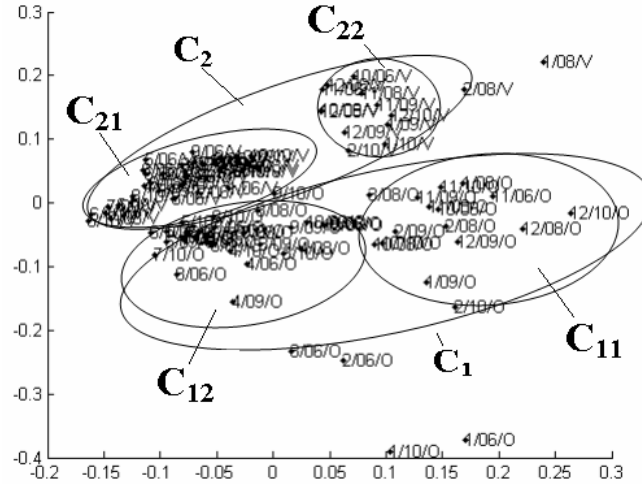


Fig. 2. LLE projections. Number of dimensions: 2, Neighbors: 20. Applies Euclidian distance function.

Fig. 2. LLE identifies the same two clusters of data than PCA (see Fig. 1) but in a more clear way. LLE is also capable of differentiating two subclusters of data (C_{11} and C_{12}), corresponding to the samples of data in C_1 (see Fig. 1). C_{11} corresponds to the most samples of winter in Ostrava (see Table 3). These samples offer the highest values of air pollution presenting in the data set, and C_{12} corresponds to the samples of the rest of the year. LLE in this case is more sensitive with samples with very high value in any of its parameters; i.e. ('2/06/O', '3/06/O'), where an important level of PM10 is detected and '1/08/V' and '2/08/V' where the concentrations of NO, NO₂, NO_x and PM10 are higher than in the rest of the year.

Table 3. Samples of each cluster identified by LLE and the range of values of each variable and units.

Cluster #	Samples						
C₁₁	11/10/O 12/10/O 1/09/O 2/09/O 11/09/O 12/09/O 1/08/O 2/08/O 11/08/O 12/08/O 11/06/O 12/06/O. Total number of samples: 12						
Var.	BZN (ug/m ³)	TLN(ug/m ³)	SO ₂ (ug/m ³)	PM10(ug/m ³)	NO ₂ (ug/m ³)	NO (ug/m ³)	NO _x (ug/m ³)
Range of Values	5.0 – 10.9	1.6 – 6.9	6.7 – 20.3	48.2 – 76.5	30.2 – 44.2	16.4 - 31	49.9 – 76.4
Description	Groups most of the samples of winter based in Ostrava. These samples offer the highest values of air pollution presenting in the data set. The samples with the maximum levels of NO _x and PM10 are located out of the cluster.						
C₁₂	3/10/O 4/10/O 5/10/O 6/10/O 7/10/O 8/10/O 9/10/O 10/10/O 3/09/O 4/09/O 5/09/O 6/09/O 7/09/O 8/09/O 9/09/O 10/09/O 3/08/O 4/08/O 5/08/O 6/08/O 7/08/O 8/08/O 9/08/O 10/08/O 3/06/O 4/06/O 5/06/O 6/06/O 7/06/O 8/06/O 9/06/O 10/06/O. Total number of samples: 32						
Range of Values	3.1 – 21.9	1.6 – 8.0	4.1 – 16.7	26.4 – 75.0	21.0 – 42.3	2.8 – 21.9	25.0 – 68.1
Description	Groups most of the samples of summer, spring and autumn based in Ostrava. These samples offer high values of pollution, but not so high as in C ₁₁ .						
C₂₁	3/10/V 4/10/V 5/10/V 6/10/V 7/10/V 8/10/V 9/10/V 10/10/V 3/09/V 4/09/V 5/09/V 6/09/V 7/09/V 8/09/V 9/09/V 10/09/V 11/09/V 3/08/V 4/08/V 5/08/V 6/08/V 7/08/V 8/08/V 9/08/V 10/08/V 11/08/V 12/08/V 1/06/V 11/06/V 12/06/V. Total number of samples: 30						
Range of Values	0.4 – 10.3	0.2 – 22.9	2.6 – 6.4	9.7 - 24	11.4 – 27.8	0.9 – 15.8	12.0 – 46.5
Description	Groups most of the samples from the station located in Veltrusy, samples corresponding to the seasons of spring, summer and autumn. High values only of NO _x .						
C₂₂	1/10/V 2/10/V 11/10/V 12/10/V 1/09/V 2/09/V 2/09/V 11/08/V 12/08/V 1/06/V 12/06/V. Total number of samples: 11						
Range of Values	1.3 – 10.5	1.2 – 14.6	2.6 – 11.8	11.0 – 28.1	22.3 – 38.4	3.4 – 12.1	30.0 – 54.4
Description	Groups most of the samples from the station located in Veltrusy corresponding to the season of winter. Very high values of NO _x .						

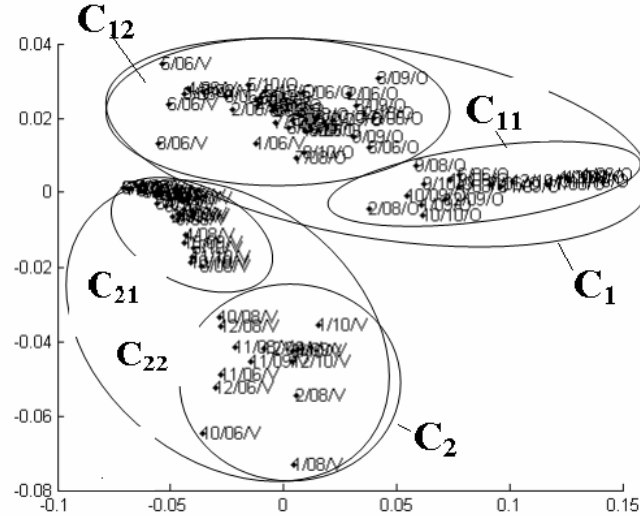


Fig. 3. ISOMAP projections. Number of dimensions: 2, Number of neighbors: 12. Applies Geodesic distance function, using Dijkstra's algorithm.

Fig. 3. ISOMAP offers the most clear results. Identifies four clusters as LLE does (LLE), but in a more sparse way. The clusters contain the same samples of data than applying LLE, except C_{11} (see Table 4). The drawback is that ISOMAP can not identify isolate samples with unusually high values in any of its parameters, as LLE does (see Fig. 2), which is very useful to identify exceptional situations. This property of ISOMAP is due to the fact that this method tries to preserve the global properties of the data, which may be an advantage or a disadvantage depending on the situation.

Table 4. Samples of each cluster identified by ISOMAP and the range of values of each variable and units.

Cluster #	Samples							
C_{11}	11/10/O 12/10/O 1/09/O 2/09/O 11/09/O 12/09/O 1/08/O 2/08/O 11/08/O 12/08/O 11/06/O 12/06/O 9/08/O 9/10/O 10/10/O 10/09/O 6/06/O. Total number of samples: 17							
Var.	BZN (ug/m ³)	TLN (ug/m ³)	SO ₂ (ug/m ³)	PM10(ug/m ³)	NO ₂ (ug/m ³)	NO (ug/m ³)	NO _x (ug/m ³)	
Range of Values	2.2 – 16.6	1.6 – 6.9	6.2 – 33.2	28.9 – 123.7	21.1 – 48.5	2.8 – 27.7	25.0 – 76.4	
Description	Groups most of the samples of winter based in Ostrava. These samples offer the highest values of air pollution presenting in the data set. ISOMAP is able to group samples than in LLE are located out of the cluster (see Fig. 2).							
C_{12}	3/10/O 4/10/O 5/10/O 6/10/O 7/10/O 8/10/O 9/10/O 10/10/O 3/09/O 4/09/O 5/09/O 6/09/O 7/09/O 8/09/O 9/09/O 10/09/O 3/08/O 4/08/O 5/08/O 6/08/O 7/08/O 8/08/O 9/08/O 10/08/O 1/06/O 2/06/O 3/06/O 4/06/O 5/06/O 7/06/O 8/06/O 9/06/O 10/06/O 1/06/V 5/06/V 6/06/V 7/06/V 8/06/V. Total number of samples: 38							
Range of Values	3.1 – 21.9	1.6 – 29.9	2.8 – 26.4	10.1 – 130.1	12.8 – 51.5	3.0 – 21.9	17.7 – 88.6	

Description	Groups most of the samples of summer, spring and autumn based in Ostrava. These samples offer high values of air pollution but not so high as in C ₁₁ . This time some samples based in Veltrusy are presented, those with a similar pollution to Ostrava, belong to the season of summer of 2006.						
C₂₁	3/10/V 4/10/V 5/10/V 6/10/V 7/10/V 8/10/V 9/10/V 10/10/V 3/09/V 4/09/V 5/09/V 6/09/V 7/09/V 8/09/V 9/09/V 10/09/V 11/09/V 3/08/V 4/08/V 5/08/V 6/08/V 7/08/V 8/08/V 9/08/V 10/08/V 11/06/V 12/06/V. Total number of samples: 27						
Range of Values	0.5 – 3.1	0.2 - 3.5	0.6 – 4.5	9.7 – 19.9	10.3 – 31.5	0.9 – 7.6	12.0 – 29.6
C₂₂	1/10/V 2/10/V 11/10/V 12/10/V 1/09/V 2/09/V 1/08/V 2/08/V 11/08/V 12/08/V 10/06/V 11/06/V 12/06/V. Total number of samples: 13						
Range of Values	0.4 – 10.5	0.2 – 14.6	2.6 – 11.8	11.0 – 28.1	22.3 – 40.0	3.4 – 25.0	30.0 – 75.0
Description	Groups most of the samples from the station located in Veltrusy corresponding to the season of winter. Very high values of NO _x .						

5. Conclusions

This study shows the different values of air pollution over several years in two Czech Republic localities with high industrial activity, Ostrava - Privoz and Veltrusy, appreciating higher levels in Ostrava. The study also reflects a similar behavior in both of the localities along the year, showing a significant increase of air pollution in the season of winter, especially in January. The pollutant with a higher increase of level in winter is PM₁₀ and in a less important way NO_x. Ostrava is an important core of industrial production, especially production of steel and metal industries. These metal industries emit large amounts of PM₁₀ and NO_x, which justifies these results, also the heavy traffic, specially the diesel engine type, represent another major source of PM₁₀ emissions to the atmosphere. The high industrial production in winter and the low levels of precipitations in the first months of the year lead to more air pollution in winter. In Veltrusy the type of industry is different and more diversified than in Ostrava, resulting in a slight increase in all contaminants in winter when the weather is drier, but not so marked an increase of PM₁₀ as in Ostrava.

Finally, the study demonstrates the behavior of the methods applied. PCA is the first method used in the data analysis process. It identifies the internal structure of the data. The rest of the techniques applied reaffirms and improves these graphical results. ISOMAP is able to group the data in compacted clusters. LLE is able to identify these same clusters and also detect unusual or interesting situations, that could be due to abnormal climatological situations in those days of high pollution, identifying these significant data points.

Acknowledgments. This research is partially supported through a projects of the Spanish Ministry of Economy and Competitiveness [ref: TIN2010-21272-C02-01] (funded by the European Regional Development Fund). This work was also supported in the framework of the IT4Innovations Centre of Excellence project, reg. no. CZ.1.05/1.1.00/02.0070 by operational

programme \Research and Development for Innovations\ funded by the Structural Funds of the European Union and state budget of the Czech Republic.

6. References

- [1] Corchado, E. and Herrero, E.: Neural visualization of network traffic data for intrusion detection. *Appl. Soft Comput.* 11(2): 2042-2056 (2011).
- [2] Corchado, E., Arroyo, A. and Tricio, V.: Soft computing models to identify typical meteorological days. *Logic Journal of the IGPL* 19(2): 373-383 (2011).
- [3] Vaidya, V., Park, J. H., Arabnia, H. R., Pedrycz, W. and Peng, S.: Bio-inspired computing for hybrid information technology. *Soft Comput.* 16(3): 367-368 (2012).
- [4] Maqsood, I. and Abraham, A.: Weather analysis using ensemble of connectionist learning paradigms. *Applied Soft Computing*, vol. 7, no. 3, pp. 995–1004, jun. (2007).
- [5] Wang, W., Men, C. and Lu, W.: Online prediction model based on support vector machine. *Neurocomputing*, vol. 71, no. 4–6, pp. 550–558, ene. (2008).
- [6] Chattopadhyay, G., Chattopadhyay, S. and Chakraborty, P.: Principal component analysis and neurocomputing-based models for total ozone concentration over different urban regions of India. *Theoretical and Applied Climatology*, pp. 1–11, dec (2011).
- [7] Glezakos, T. J., Tsiligiridis, T. A., Iliadis, L. S., Yialouris, C. P., Maris, F. P. and Ferentinos, K. P.: Feature extraction for time-series data: An artificial neural network evolutionary training model for the management of mountainous watersheds. *Neurocomputing*, vol. 73, no. 1–3, pp. 49–59, dic. (2009).
- [8] Arroyo, A., Corchado, E. and Tricio, V.: Soft computing models to analyze atmospheric pollution issues. *Logic Jnl IGPL*, feb. (2011).
- [9] Arroyo, A., Corchado, E. and Tricio, V.: A climatological analysis by means of soft computing models. *Advances in Intelligent and Soft Computing*, vol. 87, pp. 551–559 (2011).
- [10] CHMI Portal : Home. [Online]. Available: http://www.chmi.cz/portal/dt?portal_lang=en&menu=JSPTabContainer/P1_0_Home. [Accessed: 19-apr-2012].
- [11] Dvorska, A., Lammel, G., Klanova, J. and Holoubek, I.: Kosetice, Czech Republic - ten years of air pollution monitoring and four years of evaluating the origin of persistent organic pollutants. *Environmental Pollution*, 156 (2), pp. 403-408. (2008).
- [12] Pearson, K.: On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, vol. 2, no. 2, pp. 559–572 (1901).
- [13] Oja, J., Ogawa, E. and Wangviwattana, H.: Principal component analysis by homogeneous neural networks, Part I: The weighted subspace criterion, *IEICE Trans. Inf. Syst.*, vol. E75-D, no. 3, pp. 366–375 (1992).
- [14] Oja, E.: Neural networks, principal components, and subspaces. *International Journal of Neural Systems*, vol. 1, no. 1, pp. 61–68 (1989).
- [15] Balasubramanian, M. and Schwartz, E. L.: The isomap algorithm and topological stability. *Science*, vol. 295, no. 5552, p. 7 (2002).
- [16] Cox, A. A. and Cox, T. F.: Multidimensional Scaling. *Handbook of data visualization. Springer Handbooks of Computational Statistics, Vol III*, 315-347 (2008).

- [17] Cheng Cai, S. H. and Yang, Z.: Dimensionality Reduction by Using Sparse Reconstruction Embedding. Lecture Notes in Computer Science, Vol 6298 ,167-178 (2011).
- [18] Shankar, N. R. and Sireesha, V.: Using Modified Dijkstra's Algorithm for Critical Path Method in a Project Network. International Journal of Computational and Applied Mathematics 5 (2), 217-225 (2010).
- [19] Roweis, S. T. and Saul, L. K.: Nonlinear dimensionality reduction by locally linear embedding. Saul. Science, vol. 290, no. 5500, pp. 2323 – 2326 (2000).
- [20] Smola, A. J. and Schölkopf, B.: A tutorial on support vector regression. Statistics and Computing, vol. 14, no. 3, pp. 199–222 (2004).
- [21] Annual Tabular Overview. [Online]. Available: http://portal.chmi.cz/files/portal/docs/uoco/isko/tab_roc/tab_roc_EN.html. [Accessed: 19-apr-2012].
- [22] Networks of Ambient Air Quality Monitoring Stations. [Online]. Available: <http://www.chmi.cz/files/portal/docs/uoco/isko/grafroc/groce/gr10e/akap21.html>. [Accessed: 08-may-2012].
- [23] International Air Quality. [Online]. Available: <http://airnow.gov/index.cfm?action=topics.world>. [Accessed: 09-may-2012].
- [24] Stations with annual average conc. [Online]. Available: http://portal.chmi.cz/files/portal/docs/uoco/isko/tab_roc/2010_enh/eng/pdf/MaximaH-RAP.pdf. [Accessed: 19-apr-2012].