The use of Kohonen self-organizing maps to study meteorological parameters in Meknès city (Morocco)

Mustapha BEN EL HOUARI
Research Teams "Analytical Chemistry and Environment" ; "Applied Materials and Catalysis"
University Moulay Ismail
Faculty of Sciences
bel_houari@hotmail.fr

Omar ZEGAOUI
Research Team "Applied Materials and Catalysis"
University Moulay Ismail
Faculty of Sciences
ozegaoui@gmail.com

Abdelaziz ABDALLAOUI
Research Team "Analytical Chemistry and Environment"
University Moulay Ismail
Faculty of Sciences
a.abdallaoui@gmail.com

Abstract— In order to classify the meteorological parameters in Meknes city (Morocco), we have proposed in this study a linear analysis based on the Principal Component Analysis (PCA) and unsupervised classification (clustering). This allowed to identify groups of similar objects from monthly weather parameters recorded between 1996 and 2013. These parameters are the air temperature (Tair), the atmospheric pressure (Pr), the humidity (H), the precipitation (P), the visibility (Vis), the wind speed (V), and the dew point (Tr). This study is based on the use of the Kohonen self-organizing map (SOM). The obtained results showed good correlations between meteorological parameters and the existence of three classes depending on the period of the study, and directly related to typical meteorological conditions in Meknes city.

Index Terms—PCA, classification, clustering,meteorological parameters, Air temperature, Kohonen SOM, Meknes city.

1 Introduction

Artificial neural networks are improved techniques of data processing able to model relationships between particularly complex functions. Recently, several studies have confirmed that artificial neural networks are also well suited to the study of meteorological parameters **[1; 2; 3; 4]**. One of the main tasks to accomplish in this regard is the use of knowledge's extraction systems from data in order to study and analyze meteorological parameters by using automatic classification that allows to the extraction of a set of prototypes (clusters) representative weather conditions **[5; 6; 7]**.

The objective of this work was to analyze the meteorological data in Meknes city (Morocco) which are the air temperature (Tair), the atmospheric pressure (Pr), the humidity (H), the precipitation (P), the visibility (Vis), the wind speed (V), and the dew point (Tr). To do this, tools are necessary to extract useful features, providing simpler and more manageable information. The knowledge extraction techniques are advanced statistical methods developed for this task. The Kohonen self-organizing map (SOM) is one of the most popular data mining techniques that proved effective for the grouping of multidimensional data.

2 Materials and Methods

2.1 Database, geographic and climatic profiles

The database used in this study includes seven meteorological parameters measured during 216 months (1996-2013). These parameters are: the air temperature (Tair), the atmospheric pressure (Pr), the humidity (H), the precipitation (P), the visibility (Vis), the wind speed (V) and the dew point (Tr).

Meknes city is located in the central part of Morocco on plain of the Saïs with 33° 57′ Northern latitude, 05° 33′ Western longitude and 500 meter elevation. She has borders with the Middle Atlas moutains in the south and with the Pre-Rif hills in the north. It is characterized by a semi-continental climate of Mediterranean countries, whose winters are cold and rainy, and summers are hot and dry.

The mean values of monthly meteorological parameters recorded in Meknes city between 1996 and 2013 are shown Table 1. These parameters show that it is possible to extract several meteorological situations that have similar parameters (meteorological class). According to Table 1, we note that the meteorological parameters for the months of June, July and August are almost identical with air temperature that exceeds an average of 23 °C, and low humidity. Several other classes can be easily identified as the class of December, January, February, and the weather class representing the months of October and May.

TABLE 1
MEAN VALUES OF MONTHLY METEOROLOGICAL PARAMETERS RECORDED IN MEKNES CITY (MOROCCO) (1996-2013).

Month	Tair	Pr	H	P	Vis	V	Tr
Jan	9.25	1021.88	76.14	55.07	8.75	11.24	4.79
Feb	11.15	1022.13	73.34	33.10	8.79	11.49	5.76
Mar	14.27	1017.90	70.52	34.24	8.86	12.45	8.16
Apr	15.03	1015.40	70.42	41.99	8.55	11.36	8.88
May	18.22	1014.97	66.01	32.59	8.56	11.64	11.96
Jun	23.22	1014.80	58.67	13.72	8.78	11.03	15.13
Jul	24.90	1013.95	57.00	0.80	8.74	9.41	15.98
Aug	25.16	1013.93	56.48	2.59	8.72	10.14	16.22
Sep	22.26	1015.02	63.35	18.13	8.61	10.85	15.74
Oct	18.98	1016.77	66.90	42.32	8.95	10.94	13.26
Nov	14.12	1018.75	69.92	62.18	9.05	10.96	9.32
Dec	11.23	1020.26	76.20	79.16	8.79	12.93	6.61

2.2 Standardization of the database

A data preprocessing phase is needed to prepare the data for the study phase. It consists to a normalization of the data in the interval of [0, 1] according to the equation (1)

$$X_n = \frac{(X - X_{\min})}{(X_{\max} - X_{\min})} \qquad (1)$$

Xn : normalized values,
X : original values,
Xmin: minimum values,
Xmax: maximum values.

2.3 Principal component analysis (PCA)

The principal component analysis (PCA) is method of data analysis used in statistics and in the size reduction of multidimensional data projection **[8]**. It is widely used in meteorological applications **[9; 10; 11; 12]**.

The objective of the PCA is the transformation of the data space to feature space (feature space) or factor space. The transformation matrix is constructed around of the eigenvectors of the input correlation matrix. These vectors are ordered according to their eigenvalues.

The data set is compiled in the matrix A defined by the equation (2):

$$A = X_N^P \qquad (2)$$

Where, N individuals are described by the p variables.

If we consider $\overrightarrow{x_i}$ the means variables P in the matrix A, the covariance matrix is given by the equation (3):

$$\phi_{jk} = \frac{1}{n} \sum_{i=1} (x_{ji} - \overline{x_j})(x_{ik} - \overline{x_k}),\ j\neq k \qquad (3)$$

with j=1, 2, …, n and k=1, 2, …, p.

2.4 Kohonen networks and hierarchical classification

Kohonen maps are an important group of artificial neural networks. They have received special importance since the work of Von der Malsburg **[13]** and Kohonen **[14; 15]**.

They are classified as methods of vector quantization and data projection algorithms. Quantification of N training samples to the M prototypes reduced the original set of data to a smaller set, while maintaining the initial properties of data. There are no explicit rules for selecting the number of nodes of a Kohonen network, but the principle is that the size should allow easy detection of the structure of a Kohonen map **[16]**. The total number of units of the Kohonen map (M) is estimated by using the heuristic formula $M=5\sqrt{N}$ with N is number of samples. Once the number of units of the card is determined, the size of the map is then determined.
To determine the optimal number of units of the Kohonen map, several experiments must be carried out by changing the number of nodes and checking the performance of each solution.

Automatic classification can build an entire hierarchy of objects as a tree in ascending order. This method considers the singletons (classes formed only of a single individual) and proceeds by merging classes according to a similarity measure to form a new class. The process is iterated until only one cluster containing all individuals is obtained. This classification generates a tree that can be cut at different levels according to the distance measurement for a selected number of classes more or less large. Different interclass distance measurements can be used: Euclidean distance, the less distance (which encourages the creation of low inertia classes) or greater distance.

The construction principle of bottom-up methods is to develop, step by step, a series of nested partitions from the finest score (composed of n singletons {x}, i = 1. 2... n) up the coarsest partition ({X}). We begin by aggregating the two closest peoples; it remains only n - 1 items (the two individuals firstly grouped are considered as a new element). This process is iterated until all elements have been processed. The classification is based on a measure of the distance between individuals that quantifies the heterogeneity of a part, based on the distance between individuals who are in as well as on a dissimilarity measure between two parts based on the distance between an individual in a class and another in another class. Euclidean distance is generally used for individual points. There are no known criterions able to distinguish the most suitable distance type to use. After grouping the two closest individuals, we can assemble individuals, an individual and a class, or two classes **[17; 18; 19]**.

3 RESULTS AND DISCUSSION

3.1 Principal component analysis (2)

The principal component analysis (PCA) is a data analysis

method widely used for the extraction of meteorological characteristics **[11]**. The obtained results of the meteorological projection data using PCA are presented in Figure 1.
The goal of the total inertia representing bar chart is to get the maximum inertia kept with the minimum factors.

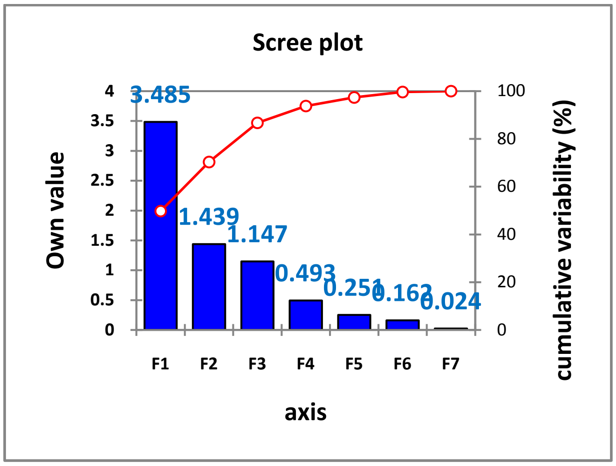

Fig. 1. The first seven principal components and their variability.

A significant drop in 4th axis (F4) is observed (from 16.39% of inertia). However the 2nd axis (F2) representing about 70.33% of the total inertia that is that we can explain 70.33% of the overall information.

The first principal component CP1 expresses 49.78% of the inertia. This component is mainly characterized by air temperature and a high dew point, and a moderate degree of moisture and pressure. CP2 which expresses 20.55% of the variability is characterized by a high visibility and high wind speed. CP3 is characterized by large measures of precipitation and pressure. CP4 is characterized by high visibility and high wind speed. CP5 reflects the influence of large measures of humidity. The remaining components, CP6 and CP7, are positively correlated with the air temperature and the dew point.

The PC1 and PC2 express a cumulative variability of 70.33% only of all principal axes. This low percentage (<80%) indicates that the PCA is not inappropriate to explain the distribution of weather clusters, mainly due to the nonlinear behavior of meteorological data.

3.2 Classification with SIM card

The topology of the Kohonen map used to regroup meteorological data is shown in Figure 2, where the cards are connected to the adjacent hexagonal nodes (size = 10x10) by adapting the meteorological situations in Meknes city.

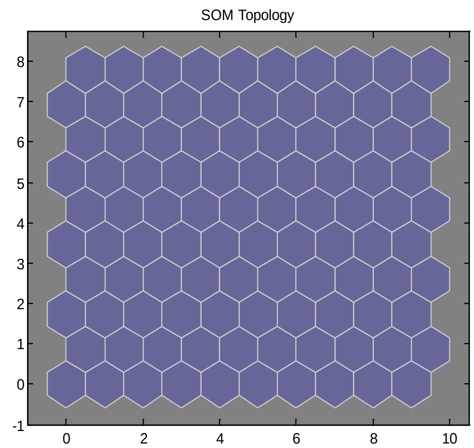

Fig. 2. Topology of the Kohonen map used to regroup meteorological data in Meknes city.

Plans of SOM components of the data set are shown in Figure 3. Patterns ~~of~~ with the same corresponding to a positive correlation can be considered among the Tair variables and variables Tr, Pr and H. The negative correlation can be observed between Tair, and H and Pr, as welle as between Tr, and H and Pr. The remaining variables are neither positive nor negative correlations, in particular those related to P and Vis.

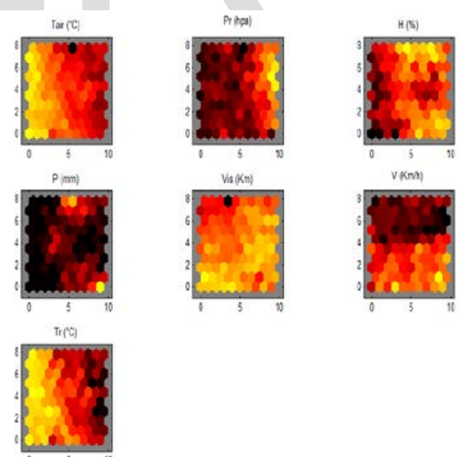

Fig. 3. Plans for the seven components of the input variables

3.3 Hierarchical classification of the SIM card

The dendrogram obtained by the hierarchical classification leads to classify the 216 observations corresponding to the studied period into three classes following the meteorological parameter values in Meknes city (Figure 4).

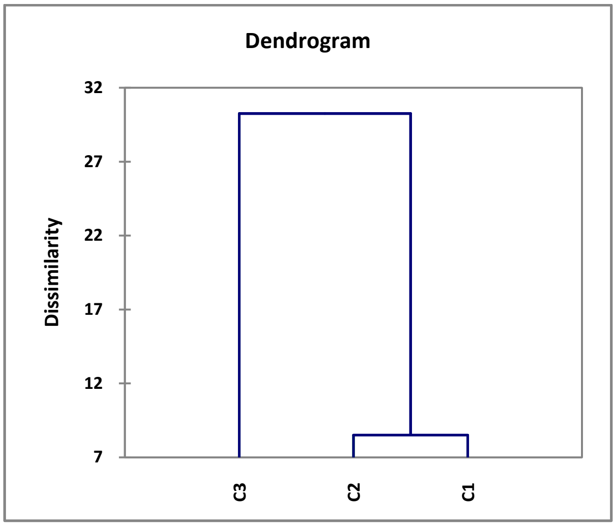

Fig. 4. Dendrogram with hierarchical classification of the SOM map

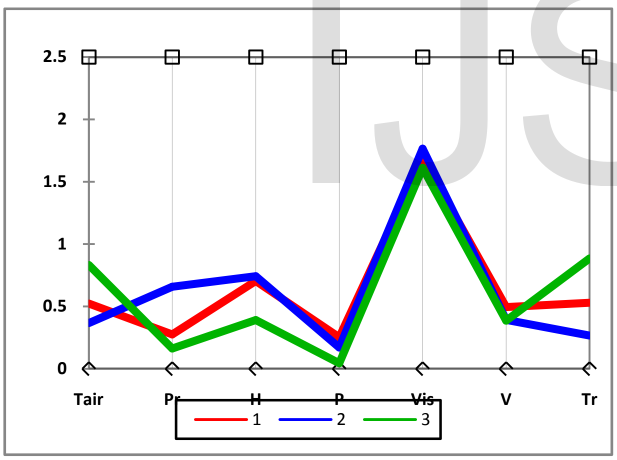

Fig. 5. Profile classes obtained with the hierarchical classification of the SOM map

With: 1: Class I, 2: Class II, 3: Class III.

Plans for the seven components of the input parameters (Fig. 5) were a great help in interpreting classes obtained.

In table 2, we presente the basic statistical parameters (minimum, average, maximum) of all data as well as the different classes related to the meteorological parameters.

Classe I: This class contains 76 months. It is characterized by the average air temperature and the dew point with high measures of precipitation, humidity and wind speed.

TABLE 2
BASIC STATISTICAL QUANTITIES (MINIMUM, AVERAGE, MAXIMUM) RELATED TO METEOROLOGICAL PARAMETERS IN MEKNES CITY (1996-2013)

		Tair (°C)	Pr (hpa)	H (%)	P (mm)	Vis (km)	V (km/h)	Tr (°C)
Total of observations	Min	1.80	1012.10	43.40	0.00	6.80	3.00	2.00
	Moy	17.15	1020.00	67.07	33.13	8.76	11.18	11.07
	Max	27.30	1027.10	83.00	233.71	10.20	17.60	17.00
Classe I	Min	1.80	1012.10	48.30	0.00	7.30	6.00	5.80
	Moy	15.51	1016.37	70.68	58.36	8.65	11.90	9.93
	Max	23.40	1021.40	83.00	233.71	9.70	16.90	16.30
	Min	8.40	1018.10	60.80	0.00	7.70	3.00	2.00

Classe II: This class is composed of 56 months and represents mainly meteorological parameters recorded at December, January and February. It is characterized by very low values of air temperature that does not exceed an average of 11.15 ° C, low measures of humidity and wind speed, and high measures of pressure and precipitation.

Classe III: This class contains 84 months and is characterized by low values of humidity, pressure and wind speed as well as highvalues of air temperature and dew point.

It is noteworthy that Visibility values remain without remarkable change for the three classes.

4 CONCLUSION

The obtained results of the principal component analysis indicated the existence of fundamental relationships between the variables of the database. However, classes that are extracted by the PCA are not appropriate to represent the meteorological clusters because of this linear analysis method is not a discriminating processing.

The self-organizing map of Kohonen has been used to study the classification of meteorological parameters in Meknes city (Morocco). The comparison with the hierarchical classification showed that the obtained results with the self-organizing map SOM were generally analogous.

REFERENCES

[1] M. Kolehmainen. H. Martikainen. J. Ruuskanen. "Neural networks

and periodic components used in air quality forecasting." *Atmospheric Environment*. vol. 35. pp. 815–825. 2001

[2] B. C. Hewitson. R. G. Crane. "Self-organizing maps: applications to synoptic climatology." *Climate Research*. vol. 22. pp. 13–26. 2002.

[3] Tereza Cavazos. "Using self-organizing maps to investigate extreme climate events: an application to wintertime precipitation in the Balkans." *Journal of climate*. vol. 13. pp. 1718–1732. 2000

[4] I. J. Turias. F. J. Gonzalez. M. L. Martin. P. L. Galindo. "A competitive neural network approach for meteorological situation clustering." *Atmospheric Environment*. vol. 40. pp. 532–541. 2006.

[5] C. Scott Sheridan. "The Redevelopment of a Weather-Type Classification Scheme for North America." *Int. J. Climatol*. vol. 22. pp. 51–68. 2002.

[6] L.S. Kalkstein. M.C. Nichols. C.D. Barthel. J.S. Greene. "A new spatial synoptic classification: application to air-mass analysis." *International Journal of Climatology*. vol. 16. pp. 983-1004. 1996.

[7] B. C. Hewitson. R. G. Crane. "Self-organizing maps: applications to synoptic climatology." *Climate Research*. vol. 22. pp. 13–26. 2002.

[8] A. Suwardi. K. Takenori. K. Shuhei. "Principal Component Analysis and Self Organizing Map for Visualizing and Classifying Fire Risks in Forest regions." *agricultural information research. Atmospheric Environment*. vol. 16(2). pp. 44-51. 2007.

[9] K.C. Lam. S. Cheng. "A synoptic climatological approach to forecast concentrations of sulfur dioxide and nitrogen oxides in Hong Kong." *Environmental Pollution*. vol. 101. pp. 183–19. 1998.

[10] J. Alex. P.H. Cannon. S. Whitfield. R.L. Edward. "Synoptic Map-Pattern Classification Using Recursive Partitioning and Principal Component Analysis. " *Monthly Weather Review*. vol. 130. pp. 1187-1206. 2002.

[11] A. Bjorn. A.W. Malmgren. "Climate Zonation in Puerto Rico Based on Principal Components Analysis and an Artificial Neural Network. " *Journal of Climate*. vol. 12. pp. 977-985. 1999.

[12] S.M. Shiva Nagendra. K. Mukesh. "Principal Component Analysis of Urban Traffic Characteristics and Meteorological Data." *Transportation Research*. Part D 8. pp. 285-297. 2003.

[13] V. Malsburg. "Self-organization of orientation sensitive cells in the striate cortex." *Kybernetics*. vol.14. 85-100. 1973.

[14] T. Kohonen. "Self-Organized Formation of topologically correct features maps." *Biological Cybernetics*. 509-521. 1988.

[15] T. Kohonen. "Self Organising Maps." *Edition Springer*. Second édition. p. 89. 1997.

[16] S.O. Hautaniemi. J. Yli-Harja. P. Astola. A. Kauraniemi. M. Kallioniemi. J. Wolf. S. Ruiz Mousses. O. kallioniemi. "Analysis and visualisation of gene expression microarray data in human cancer using self-organizing maps." *Machine Learning*. vol. 52. pp. 45-66. 2003.

[17] B. Mounzer. "contribution aux méthodes de classification non supervisée via des approches prétopologiques et d'agrégation d'opinions. " *Thèse de doctorat. Univ Claude Bernard-lyon 1*. pp. 19-35. 2006.

[18] D. Chessel. J. Thioulouse. A.B. Dufour. "Introduction à la classification hiérarchique." *Rapport technique*. pp. 27-38. 2004.

[19] L. Khodja. "Contribution à la Classification Floue non Supervisée." *Thèse de doctorat. Discip. Electronique - Electrotechnique – Automatique. Univ de Savoie*. pp. 143-158. 1997.