Advances in Geosciences

Investigation of trends in synoptic patterns over Europe with artificial neural networks

S. Michaelides[1]**, F. Tymvios**[1]**, and D. Charalambous**[1,2]

[1]Meteorological Service, 1418 Nicosia, Cyprus
[2]Department of Physics, University of Lancaster, Lancaster, LA1 4YB, UK

Abstract. The present study is a comprehensive application of a methodology developed for the classification of synoptic situations using artificial neural networks. In this respect, the 500 hPa geopotential height patterns at 12:00 UTC (Universal Time Coordinated) determined from the reanalysis data (ERA-40 dataset) of the European Centre for Medium range Weather Forecasts (ECMWF) over Europe were used. The dataset covers a period of 45 years (1957–2002) and the neural network methodology applied is the SOM architecture (Self Organizing Maps). The classification of the synoptic scale systems was conducted by considering 9, 18, 27 and 36 synoptic patterns. The statistical analysis of the frequency distribution of the classification results for the 36 clusters over the entire 44-year period revealed significant tendencies in the frequency distribution of certain clusters, thus substantiating a possible climatic change. In the following, the database was split into two periods, the "reference" period that includes the first 30 years and the "test" period comprising the remaining 14 years.

1 Introduction

There is a strong relationship between large scale circulation patterns and regional surface variables, such as surface pressure, relative humidity, evaporation, precipitation, wind and temperature. As a consequence, the analysis of synoptic upper air charts at certain levels, and especially the 500 hPa level that represents the middle troposphere, provides a valuable tool for the operational weather forecaster to diagnose and even predict surface parameters in a qualitative manner. Considering a longer time scale, these upper-air patterns could be representative of the climate that characterizes

the domain under study. An investigation of the time evolution of the geopotential height pattern could produce supporting evidence for any climatic trends. In order to simplify the statistical processing of the height pattern distribution, stochastic downscaling methods are often applied to the input weather patterns in order to create clusters of synoptic cases with similar characteristics. There is a variety of methods to classify synoptic patterns. For a review of various classifications, including their applications, refer to Key and Crane (1986), El-Kadi and Smithoson (1992), Hewitson and Crane (1996). The method explored in this work is the Kohonen' Self Organizing Maps (SOM) architecture which is an Artificial Neural Networks (ANN) method with unsupervised learning (Kohonen, 1990). A detailed description of the method and procedures proposed is provided elsewhere (Michaelides et al., 2007).

The aim of this paper is to show that the ANN methodology developed to classify synoptic patterns into distinctive clusters (see Michaelides et al., 2007; Philipp et al., 2010) can be applied in an effort to investigate trends in synoptic patterns which can be attributed to climatic change. Also, a (climatic) "reference" period can be compared against a "test" period, thus revealing differences between the frequencies of synoptic patterns that are associated with specific weather conditions over an area.

In the following, the data and methodology used are described in Sect. 2; the results and discussion are presented in Sect. 3 and concluding remarks in Sect. 4.

2 Data and methodology

2.1 Database

For the present study, gridded data from the ERA-40 reanalysis dataset were used. ERA-40 is a re-analysis of meteorological fields from September 1957 to August 2002,

Correspondence to: S. Michaelides
(silas@ucy.ac.cy)

produced by the European Centre for Medium-Range Weather Forecasts (ECMWF; see Uppala et al., 2005). The ERA-40 project started in 2000, aiming at producing global re-analyses for the 1957–2002 period with a 6-h temporal resolution (Simmons and Gibson, 2000). All of the available observations including both in-situ and remote-sensing measurements were taken into account. The basic analysed variables included not only the conventional meteorological wind, temperature and humidity fields, but also stratospheric ozone and ocean-wave and soil conditions. The re-analyses were produced on a T159 horizontal spectral resolution (nearly 125 km) and 60 vertical levels (up to 65 km height), using the 3-D-FGAT data assimilation technique with 6 hourly frequency throughout the period, supplemented by intermediate 3-h forecasts. The results are available through the ECMWF MARS database. Model biases are mostly responsible for discrepancies among re-analyses and real observations. There are two major problems reported through the years, regarding systematic errors in the database, namely, excessive precipitation over tropical oceans and too strong Brewer-Dobson circulation, both of which are pronounced in later years. These are considered negligible for our pressure level of interest, namely, the 500 hPa and the area of interest in this study.

In the present study, the dataset consists of the 12:00 UTC 500 hPa isobaric heights, measured in geopotential meters, over an area defined by the geographical latitude circles 30° N and 76° N and meridians 37° W and 56° E; a time period of 44 years (1958 to 2001) was considered, as calendar years which were not complete (i.e., 1957 and 2002) were excluded. The dataset is originally available at high resolution (i.e., 1° × 1°) but for computational reasons the density of the grid was coarser, namely, 2° × 3°.

2.2 Artificial Neural Networks

An Artificial Neural Network (ANN) is an interconnected structure of simple processing units whose functionality resembles that of the biological processing elements, the neurons, organized in such a way that the network structure adapts itself to the problem being considered. The processing capabilities of this artificial network assembly are determined by the strength of the connections between the processing units, the specific architecture pattern followed during the construction of the network and a special set of parameters adopted during the training of the network. During the last two decades, ANN proved to be excellent tools for research, as they are able to handle non-linear interrelations (non-linear function approximation), separate data (data classification), locate hidden relations in data groups (clustering) or model natural systems (simulation).

The building block of any neural computing system is an artificial representation of the fundamental cell of the brain: the neuron. Artificial neurons (or processing elements) are designed to respond to the applied inputs and to behave

consistently. The original artificial neuron is considered to be the Threshold Logic Unit, proposed by McCulloc and Pitts (1943). For a comprehensive introduction to artificial neurons the reader can refer to Tymvios et al. (2008).

The artificial implementation of a biological neuron is, in reality, an algorithm or an electronic circuit whose operation can be summarized in a few simple steps:

– All the input values are multiplied by a predetermined weight and summed.

– A bias is aggregated to the result.

– The sum is introduced to the activation function and is altered accordingly. The activation function's role is to control the amplitude of the output according to the chosen ANN architecture.

– The signal flows to the next neuron(s).

ANN are a collective sets of neural units, in which the individual neurons are connected through complex synaptic joints characterized by weight coefficients; every single neuron makes its contribution towards the computational properties of the whole system. As models of specific biological computational structures, ANN consist of distributed information processing elements, possessing an inherent potential for parallel computation. In fact, parallel processing operates in the brain but not as yet in ANN: in most personal computers the processes are performed sequentially.

The neural network architecture used in this work is the Kohonen' Self Organizing Maps (SOM). Kohonen networks provide a way of representing multidimensional data in much lower dimensional spaces – usually one or two dimensions. In order to implement this methodology, the matrix of height patterns is transformed to a one dimensional vector. This merging was achieved by placing the 24 rows in each day sequentially, thus transforming each of the two-dimensional arrays corresponding to each day (24 rows and 32 columns) into a single column. The array resulting from the merging of the daily vectors has a dimension of 16 436 columns and 768 rows.

The 16 436 vectors containing the data to be processed were randomly fed as input to a neural network implemented in Matlab 7.5. Using Matlab's own annotation, the SOM implemented for the European domain as defined above is shown in Fig. 1.

The training procedure is summarized as follows:

1. The initial value of the weights was set to small random numbers around the mean height value of the first day while the learning rate and the neighbourhood were assigned to relatively large values. Steps 2 to 5 (below) are repeated until the weights of the network are stabilized.

2. The input vectors were randomized. One vector is chosen from the dataset as an input to the network.
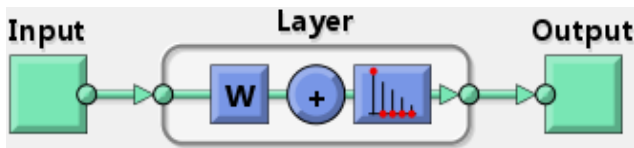
Fig. 1. A Kohonen Network using Matlab's notation, build to classify a vector of 768 members to 36 classes. The number of the ouput nodes of the SOM equals the number of classes.

3. A winning neuron is determined by examining the Euclidian distance between input vector and all individual vectors represented by neurons in the network.

4. The winning neuron and (to a far less extent) its neighbourhood are altered towards the input vector.

5. The neighbourhood and the learning rate of the parameters are decreased.

When all vectors in the training set are presented once at the input, an "epoch" is finished. The procedure is then repeated as many times as necessary until no further improvement is observed.

2.3 Experimental setup

Kohonen networks are trained with no supervision for a fixed number of epochs. Small scale problems may require a training period of just 50 epochs while for large problems this number may be as large as 10 000 or even more. These networks are extremely demanding as far as computer processing power and memory are concerned. For the classification purpose of this work, the input data were randomized and the networks were trained for 2000 epochs. The learning rate and neighbourhood size are automatically reduced as training progresses (Charalambous et al., 2001). The neighbourhood size should begin with a relatively high number, such as 90% of the number of neurons in the output layer. The termination of the learning process (presentation of learning events) occurs after the completion of the preset number of epochs or when the neighbourhood becomes zero.

2.4 Selection of cluster classification

A series of experimentations was conducted in which the 500 hPa patterns in the 44-year period were classified into 9, 18, 27 and 36 clusters. The 36-cluster classification appeared to be the most suitable to utilize over a synoptic size domain such as the one used in this work, because it provides sufficient discretization of patterns and also because it can be connected with surface observed parameters such as rainfall or dust events. The justification is provided by Tymvios et al. (2007, 2010) and will not be discussed further here. As a result of these arguments, the 36-cluster classification was chosen for further study.
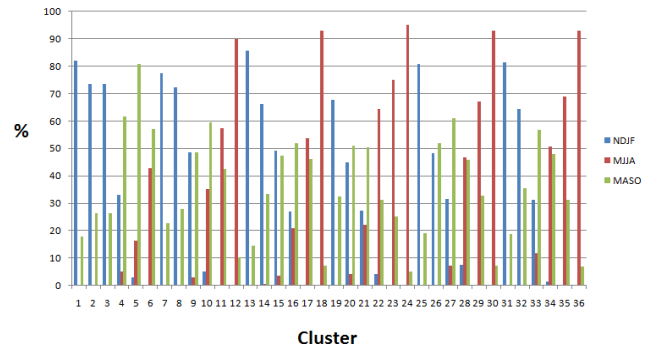

Fig. 2. Seasonal appearance (percentage) of the 36 clusters (NDJF = November, December, January and February; MJJA = May, June, July and August; MASO = March, April, September, October).

3 Results and discussion

3.1 Seasonality of clusters

The ability of ANN to group synoptic patterns into seasonally dependent clusters was noted by Michaelides et al. (2007). This seasonal discretization of clusters should be an essential attribute of a classification technique. Figure 2 shows the percentage of occurrence of the 36 clusters in each of the three seasonal periods adopted for demonstrating the seasonality discretization of the ANN methodology. The three periods were chosen by considering the Mediterranean climate as a reference: November, December, January and February were considered to belong to the "cool and wet" season (i.e., they belong to the Mediterranean Winter season); May, June, July and August were considered to belong to the "warm and dry" season (i.e., they belong to the Mediterranean Summer season); March, April, September and October are considered as "transitional" months (i.e., the transition from "cool and wet" to "warm and dry" and vice versa).

It is worth noting that the "transitional" months include patterns from all of the 36 clusters. Certain clusters are present in the "warm and dry" season but not in the "cool and dry" season and vice versa.

The frequency of occurrence of each of the 36 clusters in the entire 44-year period studied is shown in Fig. 3. The most frequent is Cluster 24 that is a "warm and dry" cluster; this is followed by Cluster 1that is a "cold and wet" cluster (see Fig. 2).

3.2 Annual variability of cluster membership

This sub-section deals with the annual variability in the number of synoptic patterns belonging to individual clusters. However, due to space limitations, it is not possible to discuss all of the 36 clusters. Instead, selected clusters are briefly presented, in order to illustrate the kind of analysis that can
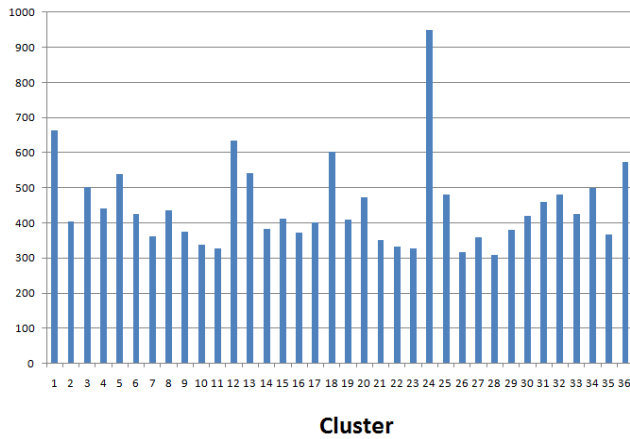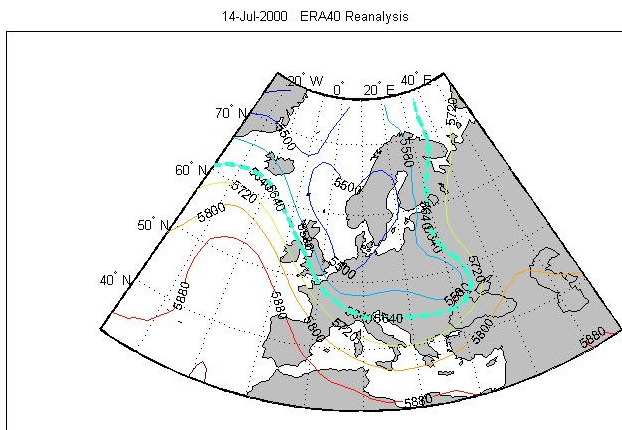
Fig. 3. Frequency of occurrence of the 36 clusters.

14-Jul-2000   ERA40 Reanalysis

Fig. 4. Typical pattern of Cluster 24: 14 July 2000 (units: gpm).

Fig. 5. Frequency of appearance of Cluster 24, a 5-year moving average and a linear fit.

24-Feb-2000   ERA40 Reanalysis
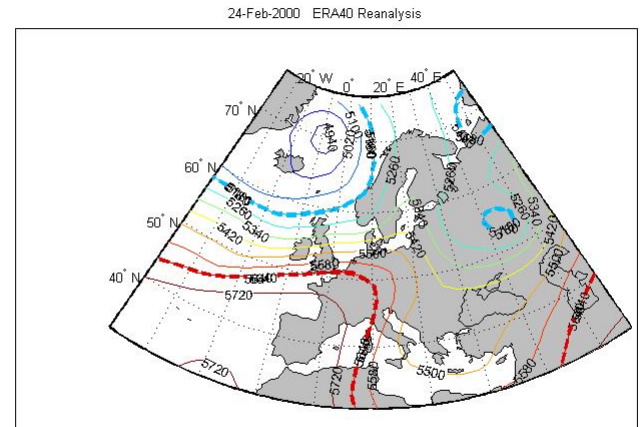
Fig. 6. Typical pattern of Cluster 1: 24 March 2000 (units: gpm).

be made. In this respect, one "warm and dry" and one "cool and wet" clusters are discussed; also, one cluster which is associated with extremely heavy rainfall events in the eastern Mediterranean is presented.

As mentioned above, the most frequent is Cluster 24. This is a "warm and dry" cluster, dominant in July and August. It corresponds to a persistent summer-time weather pattern. For instance, it persisted for 18 consecutive days in 1969 (from 25 July 1969 to 11 August 1969), for 21 consecutive days in 1972 (from 26 June 1972 to 16 July 1972) and for 22 consecutive days in 1981 (from 22 July 1981 to 13 August 1981). The frequency of the yearly occurrences of this cluster can be connected to the frequency of dry spells over eastern Mediterranean.

A typical member of Cluster 24 on 14 July 2000 is illustrated in Fig. 4. It displays a combination of a well established ridge over Azores, 5880 gpm (geopotential metres) at the 500 hPa isobaric surface and a shallow low located over Scandinavia, 5500 gpm; this results in a meridional flow over the British Isles and further east. A weak upper trough is present over the eastern Mediterranean with heights around 5820 gpm, associated with a light westerly flow.
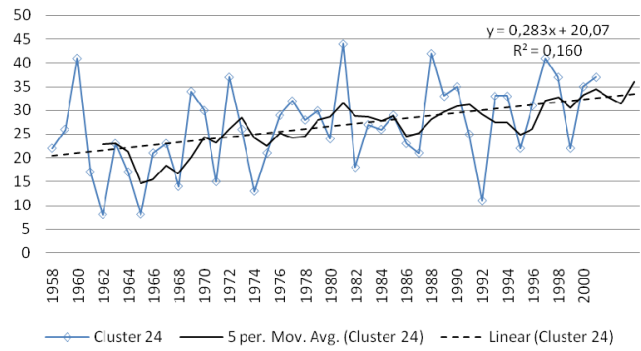
The annual variation of the membership of this cluster is presented in Fig. 5. In the same figure, the 5-year moving average and a linear projection in time is also presented. This figure exhibits a noticeable increase in the number of Cluster 24 occurrences over time. The increase of the occurrences can be connected with dry periods over he island of Cyprus, like the periods of 1957–1962, 1969–1974 and 1993–1997.

Cluster 1 is the most frequent "cold and wet" cluster and an example of a synoptic situation representative of this cluster is illustrated in Fig. 6, for 24 March 2000. The south-west to northeast orientation of the axis of the upper trough over the eastern Mediterranean frequently leads to surface cyclogenesis over the eastern part of the basin, often associated with heavy rainfall.

The dominant pattern for the heavy rainfall analysis is the one represented by Cluster 1 (Tymvios et al., 2010) and the frequency of appearance of this cluster over time is shown in Fig. 7, along with a 5-year moving average and a linear projection. Although there is a notable variation in the frequency of this cluster in the last 15 years, overall, there is a tendency for increase.
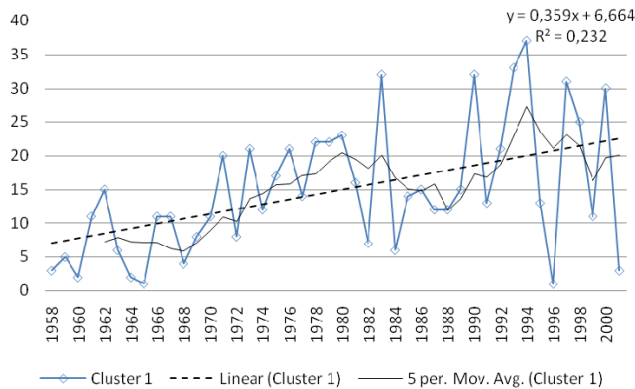
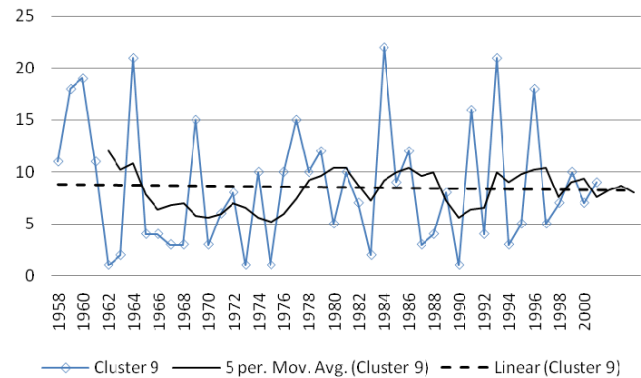**Fig. 7.** Frequency of appearance of Cluster 1, a 5-year moving average and a linear fit.

Fig. 9. Frequency of appearance of Cluster 9, a 5-year moving average and a linear fit.
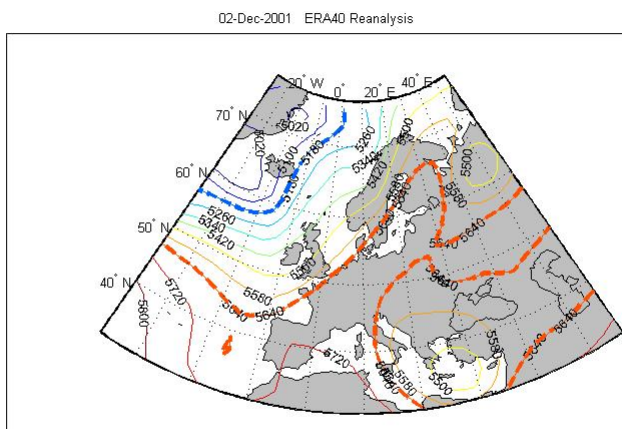

Fig. 8. Typical pattern of Cluster 9: 2 December 2001 (units: gpm).

Another important cluster is number 9. Cluster 9 is actually a collection of significant synoptic situations, in terms of extreme rainfall events. It is a cluster that is almost equally shared between the "cool and wet" and "transitional" months (see Fig. 2); it is connected with heavy rainfall events (e.g., with the most extreme rainfall day ever recorded on the island of Cyprus, as reported by Tymvios et al., 2010). A typical synoptic pattern of this cluster is shown in Fig. 8: a cold polar air mass gradually penetrates northern Europe, while warm humid air from the Mediterranean Sea is essentially stagnant over eastern Mediterranean. An upper low is centred near Crete, with a central height around 5500 gpm being associated with a surface low between Crete and Cyprus. This upper low has been cut-off from the circulation over central Europe and slowly progresses eastwards and filling. This system produced a mean accumulated rainfall of 62.3 mm on 2 December 2001 in the island of Cyprus.

The annual frequency of Cluster 9 is shown in Fig. 9. Overall, there is no significant change in the number occurrences of this cluster over time, as suggested by the year-to-year frequency of occurrences illustrated in this figure.

3.3 Comparing different time periods

In this sub-section, a procedure is put forward which can be followed in order to identify possible changes in the membership of clusters over different time periods. Although this procedure can be used in order to identify possible changes, it does not propose any reasoning behind them (e.g., climatic change).

In order to investigate whether the frequency of synoptic patterns (as represented by clusters) has changed with respect to a reference period, the database was split in two parts: in the first period, the first 30 years (1 January 1958 to 31 December 1987) and in the second period the subsequent 14 years are used. The first period is considered to represent a "reference" climatic period; the second represents the "test" period which was compared against the reference.

For each cluster, a Ko index is calculated as:

$$\mathrm{Ko} = \frac{\text{Relative frequency of cluster in the reference period}}{\text{Relative frequency of cluster in the test period}} \quad (1)$$

This index can be used to compare the number of occurrences of each cluster over the two periods. For Ko equal to unity, there is no change in the frequency of the respective cluster in the two periods; when Ko is greater (less) than one, then the respective cluster is less (more) frequent in the "test" period than in the "reference" period.

The Ko index for each of the 36 clusters is presented in Fig. 10. Comparing the relative frequencies in the two periods, a remarkable change is noted for some clusters, whereas, the change is not noteworthy for others. The most prominent change is noted for Cluster 1, whereby the Ko index is 0.64. As mentioned in Sect. 3.2, this cluster is the most frequent "cold and wet" cluster and it is often associated with heavy rainfall events. Regarding the most frequent "warm and dry" Cluster 24, the change between the two periods is very small. The same is true for the "transition" Cluster 9, discussed above.
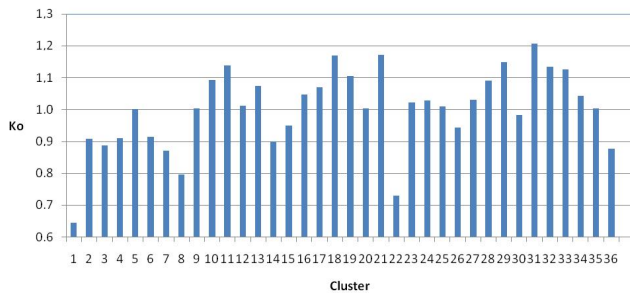
Fig. 10. The Ko index for the 36-cluster classification.

4 Concluding remarks

In the present study, a comprehensive application of a methodology developed for the classification of synoptic situations using artificial neural networks is presented. The neural network methodology applied is the SOM architecture (Self Organizing Maps) and the dataset used comprises the 500 hPa geopotential height patterns at 12:00 UTC covering a period of 44 years (1958–2001). Selected results of the classification with 36 clusters are presented. The statistical analysis of the frequency distribution of the classification results for the 36 clusters over the entire 44-year period revealed significant tendencies in the frequency distribution of certain clusters.

The pronounced trend for increase in the occurrences of Clusters 1 and 24 shows a distinctive tendency for change of the patterns' distribution. Cluster 1, a winter-time cluster, is associated with rainfall days and Cluster 24, a summer-time cluster, is the most frequent cluster representing dry conditions. The increase of the occurrences of both types can be used to explain the large variability of dry and above-average rainfall years observed during the last two decades in Cyprus.

Also, the database was split into two periods: the "reference" and the "test" periods. The classification with the 36 clusters features some interesting deviations over the two periods for some clusters, whereas, for some others no change is noted.

Edited by: A. Orphanou and K. Nicolaides
Reviewed by: two referees

References

Charalambous, C., Charitou, A., and Kaourou, F.: Comparative analysis of neural network models: Application in bankruptcy prediction, Ann. Oper. Res., 99, 403–425, 2001.

El-Kadi, A. K. A. and Smithoson, P. A.: Atmospheric classifications and synoptic climatology, Progr. Phys. Geogr., 16, 432–455, 1992.

Hewitson, B. C. and Crane, R. G.: Climate downscaling: Techniques and application, Clim. Res., 7, 85–95, 1996.

Key, J. and Crane, R. J.: A Comparison of Synoptic classification schemes based on "objective" procedures, J. Climatol., 6, 375–388, 1996.

Kohonen, T.: The Self-Organizing Map, Proceedings of IEEE, 78(9), 1464–1480, 1990.

Michaelides, S. C., Liassidou, F., and Schizas, C. N.: Synoptic classification and establishment of analogues with artificial neural networks, Pure Appl. Geophys., 164, 1347–1364, 2007.

Philipp, A., Bartholy, J., Beck, C., Erpicum, M., Esteban, P., Fettweis, X., Huth, R., James, P., Sylvie, J., Kreienkamp, F., Krennert, T., Lykoudis, S., Michalides, S. C., Pianko-Kluczynska, K., Post, P., Alvarez, D. R., Schiemann, R., Spekat, A., and Tymvios, F. S.: Cost733cat – A database of weather and circulation type classifications, J. Phys. Chem. Earth, 35, 360–373, doi:10.1016/j.pce.2009.12.010, 2010.

Simmons, A. J. and Gibson, J. K.: The ERA-40 Project Plan, ERA-40 Project Report, Series 1, European Centre for Medium range Forecasts, Reading, UK, 63 pp., 2000.

Tymvios, F. S., Constantinides, P., Retalis, A., Michaelides, S., Paronis, D., Evripidou, P., and Kleanthous, S.: The AERAS project – database implementation and Neural Network classification tests, Proceedings of the 6th International Conference on Urban Air Quality Limassol, Cyprus, 27–29 March 2007, 2007.

Tymvios, F. S., Savvidou, K., Michaelides, S. C., and Nicolaides, K. A.: Atmospheric circulation patterns associated with heavy precipitation over Cyprus, Geophys. Res. Abstr., 10, EGU2008-A-00000, 2008.

Tymvios, F. S., Michaelides, S. C., and Skouteli, C.: Estimation of surface solar radiation with Artificial Neural Networks, in: Recent Advances in Solar Radiation Modelling, edited by: Badescu, V., Springer-Verlag, Berlin, Heidelberg, Germany, 221–256, 2008.

Tymvios, F., Savvidou, K., and Michaelides, S: Association of geopotential height patterns with heavy rainfall events in Cyprus, Adv. Geosci., 23, 73–78, 2010,
http://www.adv-geosci.net/23/73/2010/.

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., da Costa Bechtold, V., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., McNally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Q. J. Roy. Meteor. Soc., 131, 2961–3012, 2005.