A Survey on Clustering based Meteorological Data Mining

Wei Tian[1], Yuhui Zheng[1], Runzhi Yang[2], Sai Ji[1] and Jin Wang[1]

[1] College of Computer and Software, Nanjing University of Information Science &
Technology, Nanjing 210044, China
[2] National Meteorological Information Center, China Meteorological Administration,
Beijing 100080, China

Abstract

Data mining is an important tool in meteorological problems solved. Cluster analysis techniques in data mining play an important role in the study of meteorological applications. The research progress of the clustering algorithms in meteorology in recent years is summarized in this paper. First, we give a brief introduction of the principles and characteristics of the clustering algorithms that are commonly used in meteorology. On the other hand, the applications of clustering algorithms in meteorology are analyzed, and the relationship between the various clustering algorithms and meteorological applications are summarized. Then we interpret the relationship from the perspectives of algorithms' characteristics and practical applications. Finally, some main research issues and directions of the clustering algorithms in meteorological applications are pointed out.

Keywords: data mining; clustering; meteorological applications

1. Introduction

Climate change and its consequences are increasingly being recognized as one of the most significant challenges which faced by humanity. Managing and taking advantage of large mass of meteorological data is the basis for studying climate change as well as the key to improve the accuracy of prediction of disastrous weather. A growing number of technologies have been or will be introduced into the field of meteorology, in which the cluster analysis in data mining has occupied an important position in the study of meteorology. K-means and hierarchical methods are classical algorithms in cluster analysis and they are widely used in meteorology. Some other clustering algorithms, such as DBSCAN (Density-based Special Clustering of Applications with Noise), EM (Expectation Maximization), SOM (Self-organizing Feature Maps), FCM (Fuzzy C-means Clustering), SA (Simulated Annealing), WaveCluster and graph clustering, also perform well in some special meteorological applications.

In order to make better use of clustering algorithms in meteorology, we analyze the applications of commonly used clustering algorithms in meteorology, and summarize the distribution and characteristics of clustering algorithms in meteorological applications. Then we interpret the results from the perspectives of characteristics of the algorithms and practical applications.

2. Problem Definition

The cluster analysis of meteorological data, both observed and model-generated, poses a number of unique challenges: (i) the data is spatially and temporally correlated,

(ii) the data is potentially noisy, (iii) massive quantities of data are available for mining, *etc.* Common data types in cluster analysis mainly contain interval-scale variables, binary variables, categorical variables, ordinal variables, ratio-scaled variables, variables of mixed types and vector objects. There are some commonly used distance measures, for example, Euclidean distance, Manhattan distance, Minkowski distance, Chebyshev distance, Mahalanobis distance, Hamming distance and correlation coefficient [1]. Different measure methods have different features and advantages. Thus, in a specific meteorological application, combined with the data characteristics, choosing the appropriate clustering algorithm is the premise of successful experiments.

3. The Commonly used Clustering Algorithms in Meteorology

At present, many clustering algorithms are discussed [2]. But few clustering algorithms can be used in meteorology. In the rest of this section, clustering algorithms that are commonly used in meteorological applications are summarized briefly from the perspectives of principles and characteristics.

3.1 Partitioning clustering

Given a dataset with n objects, a partitioning method classifies the data into k groups. Then an iterative relocation technique is used to improve the partitioning. The process stops until the criterion function converges.

The k-means algorithm [3] is a prominent partitioning clustering. It chooses k initial centroids where k is a user-specified parameter. The k-means proceeds as follows:

(1) Select k points as initial centroids.

(2) Repeat.

(3) Form k clusters by assigning each point to its closest centroid.

(4) Recomputed the centroid of each cluster.

(5) Until Centroids do not change.

The k-means algorithm is very simple and can be easily implemented in many practical problems. It can work very well for compact and hyperspherical clusters. But there is no efficient and universal method for identifying the initial partitions and the number of clusters k. The iteratively optimal procedure cannot guarantee convergence to a global optimum. It is also sensitive to outliers and noise.

3.2 Hierarchical clustering

Hierarchical methods organize data into a hierarchical structure according to the proximity matrix. They can be further classified as agglomerative methods or divisive methods, based on how the hierarchical decomposition is formed.

Agglomerative clustering [4] starts with N clusters and each of them includes exactly one object. Based on different definitions of distances between two clusters, there are mainly five agglomerative methods including single linkage, complete linkage, average linkage, median linkage and ward's method. The different distances for the five methods are defined as follows.

Single linkage method $\quad D_{\min}(C_i, C_j) = MIN_{x \in C_i, y \in C_j} |x - y|$ $\qquad$ (1)

Complete linkage method $D_{max}(C_i, C_j) = MAX_{x \in C_i, y \in C_j} |x - y|$　　　　(2)

Average linkage method $D_{avg}(C_i, C_j) = \sum_{x \in C_i} \sum_{y \in C_j} \frac{|x - y|}{n_i n_j}$　　　　(3)

Median linkage method $D_{mean}(C_i, C_j) = (1/C(n_i + n_j, 2)) \sum_{x, y \in (C_i, C_j)} |x - y|$　(4)

Ward's method　　　$D_{ward}(C_i, C_j) = (1/(n_i + n_j)) \sum_{x \in (C_i, C_j)} |x - n|^2$　　　　(5)

where n is the merged cluster center. $|x\text{-}y|$ is the distance between two objects x and y. n_i is the number of objects in cluster C_i. And $C(n_i + n_j, 2)$ is the total number of the methods for different combinations of two elements extracted from the $n_i + n_j$ elements.

3.3 Model-based clustering

Model-based clustering methods attempt to optimize the fit between the given data and some mathematical model. Such methods are often based on the assumption that the data are generated by a mixture of underlying probability distribution. EM and SOM are two popular methods of model-based clustering.

(1) EM method: EM [5] is a popular iterative refinement algorithm for finding maximum likelihood estimate of parameters in statistical models, where the model depends on unobserved latent variables.

EM can be viewed as an extension of the k-means paradigm. However, it assigns each object to a cluster according to a weight representing the probability of membership rather than cluster mean. EM iteration alternates between performing an expectation (E) step, which creates a function for the expectation of the log-likelihood evaluated using the current estimate for the parameters, and a maximization (M) step, which computes parameters maximizing the expected log-likelihood found on the E step. These parameter-estimates are then used to determine the distribution of the latent variables in the next E step.

EM is simple and easy to implement. In practice, it converges fast but may not reach the global optima. It is sensitive to the initial parameters too.

(2)SOM method: SOM [6] is a type of the artificial neural network that is trained using unsupervised learning to produce a low-dimensional, discretized representation of the input space of the training samples, called a map. It is one of the most popular neural network methods for clustering analysis.

The method is particularly useful when a nonlinear mapping is inherent in the problem itself. Just like the k-means, SOM also needs to predefine the number of the clusters, which is unknown for most circumstances.

3.4 Fuzzy clustering

In fuzzy clustering, data elements can belong to more than one cluster, and associated with each element is a set of membership levels. Fuzzy clustering is a process of assigning these membership levels, and then using them to assign data elements to one or more clusters.

FCM [7] is one of the most widely used fuzzy clustering algorithms. Like the k-means algorithm, the FCM aims to minimize the following objective function:

$$J(U,M) = \sum_{i=1}^{c} \sum_{j=1}^{N} (u_{i,j})^m D_{ij}$$

(6)

where $U = [u_{i,j}]_{c \times N}$ is a fuzzy partition matrix. $u_{i,j} \in [0,1]$ is the membership value of element xj in the cluster j. $M = [m_1, ..., m_c]$ is a cluster centers matrix. D_{ij} is the distance between element xj and cluster center mi. The function differs from the k-means objective function by the addition of the membership value $u_{i,j}$ and the fuzzy coefficient m. The fuzzy coefficient m determines the level of cluster fuzziness. A large m results in smaller membership values, hence, fuzzier clusters. In the absence of experimentation or domain knowledge, m is commonly set to 2.

FCM is particularly useful when the boundaries among the clusters are not well separated or ambiguous. Moreover, the membership values may help us discover more sophisticated relations between a given object and the disclosed clusters. However, FCM suffers from the presence of noise and outliers. It is also difficult to identify the initial partitions.

3.5 Combinatorial search techniques-based clustering

The basic objective of search techniques is to find the global or approximate global optimum for combinatorial optimization problems, which usually has NP-hard complexity and need to search an exponentially large solution space. Clustering can be regarded as a category of optimization problem. Give a set of data points, clustering algorithms aim to organize them into K subsets that optimize some criterion function.

4. The Applications of Clustering Algorithms in Meteorology

4.1 Climate change

Climate change and its impact is one of the greatest challenges facing humanity and the earth. Investigating the climate change and exploring the law for the benefit of mankind is everyone's responsibility. Cluster analysis techniques study the climate change from three aspects of meteorological elements classification, weather type classification and atmospheric circulation.

4.1.1 Meteorological elements classification: The classification is a task of classifying regions based on similar basic characteristics and changes of meteorological elements. Cluster analysis techniques classify regions to study regional weather, climate change, prediction and prevention of extreme weather from aspects of wind, precipitation, temperature, clouds, pressure etc. At present the study of classification based on meteorological elements focus on k-means method, hierarchical method, DBSCAN method, SOM method and FCM method.

(1) K-means method: In 2009, Jiménez, P. A. *et al.* [8] classified the daily surface wind fields into wind pattern types with the combination of the complete linkage algorithm and an algorithm similar to the k-means according to their spatial similarity over the Comunidad Foral de Navarra region. In the same year, Christos J. Lolis [9] applied the k-means to the factor scores time series classifying the 56 years into six distinct clusters that describe the main modes of spatial distribution of cloudiness. Then in 2011, İbrahim Sönmez *et al.* [10]

applied the k-means to reclassify rainfall regions of Turkey [11] and investigated their temporal variability in relation to North Atlantic Oscillation.

(2) Hierarchical method: Robeson, S.M. *et al.* and J. Kysely *et al.* [12] classified homogeneous regions according to the temperature and precipitation by different hierarchical clustering methods respectively. In 2008, M. Burlando *et al.* [13] applied 15 different clustering techniques resulting from the combination of three distance measures and five agglomerative methods to study the wind climate of Corsica.

(3) DBSCAN method: In 2005, a data mining application based on DBSCAN was carried out on air temperature database and obtained clusters that have similar temperature trends [14].

(4) SOM method: In 2007, Reusch, D.B. *et al.* [15] analyzed the mean sea level pressure data from 1957 to 2002 with SOM method in order to study North Atlantic climate variability in general, especially the North Atlantic oscillation.

4.1.2 Atmospheric circulation: Atmospheric circulation has been topics of interest to climatologists for years [16]. It is the dominant factor in global climate and a wide range of weather situation as well as the background of the weather system activity on a variety of scales. Cluster analysis techniques analyze the characteristics and long-term changes of the atmospheric circulation as well as its relationship with the climate change. The clustering algorithms used in this field contain k-means method, hierarchical method, SA method and graph theory-based method.

(1) K-means method: In 2006, Esteban, P. *et al.* [17] applied PCA and the k-means to characterize the daily surface synoptic circulation patterns. The results are consistent with the subjective knowledge of the daily atmospheric circulation over the area.

(2) Hierarchical method: In 2007, EM and a traditional hierarchical agglomerative clustering method (HAC) were applied and generated distinctly different atmospheric patterns characterizing seasonal circulation over eastern North America.

(3) SA method: In 2007, Philipp, A *et al.* [18] proposed a new clustering scheme combining the concepts of simulated annealing and diversified randomization (ASNDRA) when exploring long-term variability of daily North Atlantic-European pressure patterns since 1850. The method is able to reduce substantially the influence of chance in the cluster assignment, leading to partitions that are noticeably nearer to the global optimum and more stable. Based on the former method, a representative set of patterns that fully characterize the different circulation types appearing in each season was generated [19], and in 2009 [20], it was used to classify daily pressure-pattern sequences and evaluate whether sequence classification is more suitable to describe surface air-temperature conditions in Europe.

4.2. Urban meteorology

Urban meteorology, including the city meteorological observation, urban atmospheric pollution, urban fine weather forecasting, is an important part of human life. High-impact weather research, air quality prediction and weather forecasting are the main applications of clustering algorithms in the field of urban meteorology.

4.2.1 Air quality: Air pollutants, such as SO2, CO, O3, PM2.5, PM10, acid rain, acid fog, have a definite impact on human living conditions. At present, cluster analysis techniques are mainly used to study the impact of air pollution as well as the relationship between weather conditions and air pollution. The main clustering techniques in this field are the k-means, hierarchical method, the SOM and the FCM.

(1) **K-means method:** Cervone, G. *et al.* [21] and Li, L. *et al.* [22] investigated the impact of air pollution by k-means. In 2011, Alex Mace *et al.* [23] proposed adaptive k-means clustering algorithm that used both the trajectory variables and the associated chemical value. The air mass trajectories were clustered to identify source regions of certain chemical species.

(2) **Hierarchical method:** Joseph H. Casola *et al.* [24] identified weather regimes through different hierarchical clustering methods. S. Yonemura *et al.* [25] and Charbel Afif *et al.* [26] analyzed the characterization of pollutant gas concentrations by different hierarchical clustering methods. In 2011, Seungmin Lee *et al.* [27] applied an agglomerative hierarchical clustering algorithm based on the Ward's minimum variance clustering criterion to the back trajectories to examine the origin of and favorable meteorological conditions for high concentrations of PM10 in Seoul, Korea.

(3) **SOM method:** Considering the influence of synoptic-scale meteorology on air quality, Ignacio J. Turias *et al.* [28] and John L. Pearce *et al.* [29] have done something by SOM method respectively. Although the SOM has proven its efficiency in meteorological parameters and air pollution concentration clustering, it is difficult to clearly identify cluster's borders if the SOM is very populated. Moreover, using only k-means generates high computational time when clustering a large multidimensional time series data. In 2012, Soufiane Khedairia *et al.* [30] proposed a two stages clustering approach based on the SOM and the k-means algorithm to the characterization of meteorological conditions.

4.2.2 Weather forecasting: Weather forecasting is made based on analysis of satellite cloud images and weather maps, combined with relevant meteorological data, topography, seasonal characteristics and the mass experiences. At present, forecasting is still not accurate because of the uncertain data of nature. On the other hand, it is time-consuming to process large amount of data. All these deficiencies make cluster algorithms more applicable in weather forecasting field. The main clustering algorithms in this field are hierarchical method and FCM method.

4.3 Hydrometeorology

Hydrometeorology is a subject that applies the principles and methods of meteorology to study the issues related to precipitation and evaporation in hydrological cycles and water balance. Clustering algorithms are applied in hydrometeorology to tackle hydrological forecasting and water resources management.

4.3.1 Hydrological forecasting: Hydrological forecasting is an important aspect of hydrological services on economy and society, especially making prediction of disastrous hydrological phenomena. Currently, the main applications of cluster algorithms in hydrological forecasting include rainfall monitoring, flood forecasting,

water quality monitoring, inflow and outflow calculation. The main clustering methods contain k-means method, hierarchical method, SOM method and FCM method.

(1) K-means method: In 2010, Wang Min *et al.* [31] applied k-means clustering algorithm to segment the SAR water image. In the same year, the k-means was used to categorize the data points of the different flooding characteristics in the study area and identify the control point(s) from individual flooding cluster(s) when forecasting flood inundation depths [32].

(2) Hierarchical method: In 2006, Sabu Paul *et al.* [33] grouped the water bodies into clusters having similar watershed characteristics with Ward's minimum variance method. This approach is helpful to identify possible sources and determine appropriate models. Later PCA and ward's method were used to determine surface water origin and its interactions with groundwater in Mediterranean streams in 2008 [34].

(3) SOM method: In 2007, Gwo-Fong Lin *et al.* [35] estimated design hyetographs of ungauged sites through a SOM-based approach and the number of clusters can be objectively decided by visual inspection. Later, they proposed a hybrid neural network model combined the SOM and the multilayer perceptron network to forecast the typhoon rainfall [36]. In 2011, they used the SOM to analyze the input data and reveal the topological relationships among input data when developing a reservoir inflow forecasting model [37].

4.3.2 Water resources management: Climate change and human activities have great impacts on water resources. It mainly includes precipitation change, runoff decrease, flood disaster and so on. Therefore, reasonable configuration and integrated planning of water resources are important parts of the water resources management. Clustering algorithms classify different watersheds with similar hydrological characteristics. And under certain climatic conditions, changes of precipitation, basin evaporation and snow density have been studied. At present, the clustering methods in this field contain the k-means, hierarchical method, SOM method and FCM method.

(1) K-means method: In 2005, k-means was performed aiming to isolate the weather regimes responsible for the inter-annual variability of the winter precipitation over Portugal [38]. In 2008, Mizukami, N. *et al.* [39] applied k-means clustering to obtain qualitative information on spatial patterns of snowpack density and densification rates. Density is one of the fundamental properties of snowpack because it directly affects many physical properties of snow.

(2) Hierarchical method: In 2008, V. Guldal *et al.* [40] applied a hierarchical clustering algorithm to cluster the monthly evaporation losses with the monthly winds speed and wind blow number of Eğirdir Lake in Turkey. Later, Daniel G. Kingston *et al.* [41] identified hydrologic regions for the northern North Atlantic periphery, based on the ward's method of inter-annual variability in 112 basins across this region. The primary climatic drivers of hydrological variation between regional classes have also been explored.

(3) SOM method: In 2008, Fi-John Chang *et al.* [42] applied k-means and SOM algorithm to group watersheds with similar values for hydrological statistics. In 2010, S. R. Fassnacht *et al.* [43] used SOM to define regions of homogeneity in the Colorado

River Basin using snow telemetry snow water equivalent data. In the same year, Chang, F.-J. *et al.* [44] accessed the effort of meteorological variables for evaporation estimation by the SOM. The results demonstrated that the topological structures of SOM could give a meaningful map to present the clusters of meteorological variables and the networks could well estimate the daily evaporation.

4.4 Energy meteorology

Wind and solar energy are renewable resources, clean and environmentally friendly, with a broad development space and huge value. Rational use of meteorological resources, strengthening the development and utilization of wind and solar energy are some of the important ways to solve the energy problem as well as for energy saving. Cluster analysis techniques may be used for short-term wind power prediction. They can also assess the quality of energy so as to identify the locations which are optimal for energy power plant. The main clustering techniques in this field are k-means method and hierarchical method.

5. Discussion

From the previous section, we can see the main distribution of clustering algorithms in meteorological applications. Then we give a simple summary of the distribution. The relationship between the clustering algorithms and its applications in meteorology are shown in Figure 1.
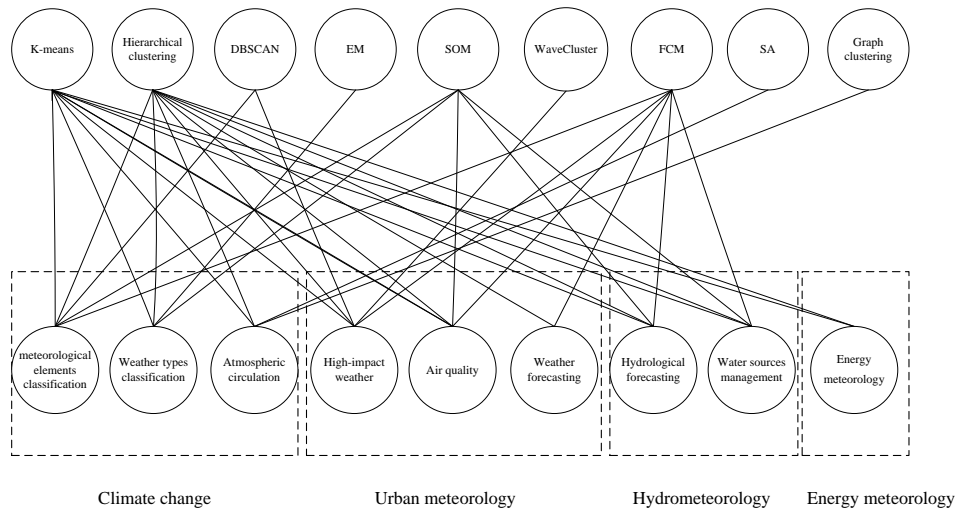
Figure 1. Algorithms/applications relationship map

Combined the research in section 3 and Figure 1, we see summary from two aspects of clustering algorithms and meteorological applications.

In the term of the clustering algorithms, we can see that k-means and hierarchical methods are widely used in almost all the meteorological applications mentioned above for their typicality and distinct advantages. The applications are mainly in climate change, urban meteorology, hydrometeorology and energy meteorology. FCM is a typical fuzzy clustering algorithm. Compared to the hard division, it fits many special meteorological applications based on fuzzy partition. These special applications are

mainly distributed in urban meteorology and hydrometeorology. It is also applied to classify Indian meteorological stations into homogeneous groups due to its advantage of assigning every station with partial membership in each group. As a method of neural network, SOM is widely applied in high-dimensional data visualization application for it can represent all points in a high-dimensional source space by points in a low-dimensional target space and try to maintain the topology of the input space. Its applications are mainly in climate change, air quality and hydrometeorology.

For some other algorithms, such as DBSCAN, EM, WaveCluster, SA and graph clustering method, their applications are limited in solving specific meteorology issues. DBSCAN is now mainly used for identifying and tracking storm cell and studying the light flash characteristics. EM is particularly used to determine the typical weather types, while WaveCluster identifies the outliers in meteorological data. SA and Graph clustering are mainly applied to investigate atmospheric circulation and its impact on climate change.

In terms of meteorological applications and the real world, the clustering algorithms are mostly used in applications which are closely related to human life, such as urban meteorology and climate change. With the rapid development of society and technology, people are increasingly concerned about the quality of life and scientific production, forecasting, monitoring, etc. This is also verified that people propose or create a variety of techniques to meet the requirements of the humans, which is in line with the law of human and social development.

Acknowledgements

References

[1] http://users.csc.calpoly.edu/~dekhtyar/466-Spring2009/lectures/lec09.466.pdf, Accessed on **(2013)** April 15.
[2] R. Xu and D. Wunsch, IEEE Transactions on Neural Networks, vol. 16, no. 3, **(2005).**
[3] J. MacQueen, L. M. L. Cam and J. Neyman, "Some methods for classification and analysis of multivariate observations", Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, **(1967)** June 21-July 18, **(1965)** December 27, **(1966)** January 7, Berkeley, California
[4] F. Murtagh, The Computer Journal, vol. 26, no. 4, **(1983).**
[5] G. McLachlan and T. Krishnan, Wiley, vol. 1, **(1997)**, pp. 997.
[6] T. Kohonen, H. J. Trussell and V. Damle, Proceedings of the IEEE, vol. 78, no. 9, **(1990)**.
[7] Y. H. Lu, T. H. Ma, C. H. Yin, X.Y. Xie, W. Tian and S. M. Zhong, International Journal of Database Theory and Application, vol. 6, no. 6, **(2013)**.
[8] P. A. Jiménez, J. F. González-Rouco, J. P. Montávez, E. García-Bustamante and J. Navarro, International Journal of Climatology, vol. 29, no. 4, **(2009).**
[9] C. Lolis, Theoretical and Applied Climatology, vol. 96, no. 3, **(2009).**
[10] İ. Sönmez and A. Ü. Kömüşcü, Theoretical and Applied Climatology, vol. 106, no. 3, **(2011).**
[11] S. M. Robeson and J. A. Doty, Journal of Climate, vol. 18, no. 8, **(2005).**
[12] J. Kysely, J. Picek and R. Huth, Studia Geophysica et Geodaetica, vol. 51, no. 2, **(2007).**
[13] M. Burlando, M. Antonelli and C. F. Ratto, International Journal of Climatology, vol. 28, no. 5, **(2008).**
[14] T. Bilgin and A. Çamurcu, (Eds.), Advances in Information Systems, Springer Berlin Heidelberg, New York, **(2005).**
[15] D. B. Reusch, R. B. Alley and B. C. Hewitson, Journal of Geophysical Research: Atmospheres, (1984-2012), vol. 112, no. D2, **(2007).**
[16] R. Huth, C. Beck, A. Philipp, M. Demuzere, Z. Ustrnul, M. Cahynov´a, J. Kysel´y and O. E. Tveito, Annals

of the New York Academy of Sciences, vol. 1146, no. 1, **(2008)**.

[17] P. Esteban, J. Martin-Vide and M. Mases, International Journal of Climatology, vol. 26, no. 11, **(2006)**.

[18] A. Philipp, P. M. Della-Marta, J. Jacobeit, D. R. Fereday, P. D. Jones, A. Moberg and H. Wanner, Journal of Climate, vol. 20, no. 16, **(2007)**.

[19] D. R. Fereday, J. R. Knight, A. A. Scaife, C. K. Folland and A. Philipp, Journal of Climate, vol. 21, no. 15, **(2008)**.

[20] A. Philipp, Theoretical and Applied Climatology, vol. 96, no. 1, **(2009)**.

[21] G. Cervone, P. Franzese, Y. Ezber, Z. Boybeyi, F. Bonchi, B. Berendt, F. Giannotti, D. Gunopulos, F. Turini, C. Zaniolo, N. Ramakrishnan and X. Wu, "Risk Assessment of Atmospheric Hazard Releases using K-means Clustering", Proceedings of IEEE International Conference on In Data Mining Workshops, **(2008)** December 15-19; Pisa, Italy.

[22] L. Li and S. Cheng, "A Calculated Methodology of Regional Contributions Based on MM5-CAMx in Typical City: A 2006 Case Study of SO2 and Sulfate", Proceedings of the 4th International Conference on Bioinformatics and Biomedical Engineering, **(2010)** June 18-20; Chengdu, China.

[23] A. Mace, R. Sommariva, Z. Fleming and W. Wang, in Adaptive K-means for clustering air mass trajectories, Edited H. Yin, W. Wang and V. Rayward-Smith, Springer Berlin, Heidelberg, vol. 6936, **(2011)**, pp. 1-8.

[24] J. H. Casola and J. M. Wallace, Journal of Applied Meteorology and Climatology, vol. 46, **(2007)**.

[25] S. Yonemura, S. Kawashima, H. Matsueda, Y. Sawa, S. Inoue and H. Tanimoto, Theoretical and Applied Climatology, vol. 92, no. 1, **(2008)**.

[26] C. Afif, A. L. Dutot, C. Jambert, M. Abboud, J. Adjizian-Gerard, W. Farah, P. E. Perros and T. Rizk, Air Quality, Atmosphere & Health, vol. 2, no. 2, **(2009)**.

[27] S. Lee, C. H. Ho and Y. S. Choi, Atmospheric Environment, vol. 45, no. 39, **(2011)**.

[28] I. J. Turias, F. J. González, M. L. Martín and P. L. Galindo, Atmospheric Environment, vol. 40, no. 3, **(2006)**.

[29] J. L. Pearce, J. Beringer, N. Nicholls, R. J. Hyndman, P. Uotila and N. J. Tapper, Atmospheric Environment, vol. 45, no. 1, **(2011)**.

[30] S. Khedairia and M. T. Khadir, Atmospheric Research, vol. 113, **(2012)**.

[31] M. Wang, S. Zhou, B. Heng, M. Ning and Y. Song, "SAR Water Image Segmentation Based on GLCM and Wavelet Textures", Proceedings of the 6th International Conference on Wireless Communications, Networking and Mobile Computing, **(2010)** September 23-25; Chengdu, China.

[32] L. C. Chang, H. Y. Shen, Y. F. Wang, J. Y. Huang and Y. T. Lin, Journal of Hydrology, vol. 385, no. 1, **(2010)**.

[33] S. Paul, R. Srinivasan, J. Sanabria, P. K. Haan, S. Mukhtar and K. Neimann, Journal of The American Water Resources Association, **(2006)**.

[34] A. Menció and J. Mas-Pla, Journal of Hydrology, vol. 352, no. 3, **(2008)**.

[35] G. F. Lin and M. C. Wu, Journal of Hydrology, vol. 339, no. 3, **(2007)**.

[36] G. F. Lin and M. C. Wu, Journal of Hydrology, vol. 375, no. 3, **(2009)**.

[37] G. F. Lin and M. C. Wu, Journal of Hydrology, vol. 405, no. 3, **(2011)**.

[38] J. A. Santos, J. Corte-Real and S. M. Leite, International Journal of Climatology, vol. 25, no. 1, **(2005)**.

[39] N. Mizukami and S. Perica, Journal of Hydrometeorology, vol. 9, no. 6, **(2008)**.

[40] V. Guldal and H. Tongal, Applied Ecology and Environmental Research, vol. 6, no. 4, **(2008)**.

[41] D. G. Kingston, D. M. Hannah, D. M. Lawler and G. R. McGregor, Hydrological Processes, vol. 25, no. 7, **(2011)**.

[42] F. J. Chang, M. J. Tsai, W. P. Tsai and E. E. Herricks, Journal of Hydrology, vol. 354, no. 1, **(2008)**.

[43] S. R. Fassnacht and J. E. Derry, Water resources research, vol. 46, no. 4, **(2010)**.

[44] F. J. Chang, L. C. Chang, H. S. Kao and G. R. Wu, Journal of Hydrology, vol. 384, no. 1, **(2010)**.

Authors

Wei Tian , Lecturer at the School of Computer and Software, also a doctoral candidate of atmosphere science at School of Atmospheric Science, Nanjing University of Information Science and Technology. His research interest covers data mining , meteorological Data Processing, remote sensing image processing, typhoon track mutation mechanism.
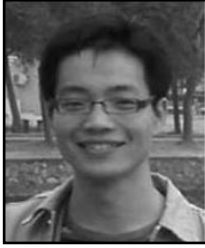
Yuhui Zheng, Assosiate Processor at the School of Computer and Software, Nanjing University of Information Science and Technology. His research interest covers image processing, pattern recognition, and remote sensing image restoration.

Runzhi Yang, Engineer at National Meteorological Information Center of China. She received her master degree from Beihang University,China, 2006. Her research interest covers meteorological data processing, cloud computation, and parallel computing.

Sai Ji, Professor at the School of Computer and Software, Nanjing University of Information Science and Technology. He received his Bachelor (NUIST, China, 1999), Master (NUAA, China, 2006). His research interests are in the areas of Data Mining, Computer Measurement and Control and Wireless sensor networks.

Jin Wang, Professor at the School of Computer and Software, Nanjing University of Information Science and Technology. His research interests mainly include routing protocol and algorithm design, performance evaluation and optimization for wireless ad hoc and sensor networks. He is a member of the IEEE and ACM.