

Chapter 16

CLIMAGE: A New Software for the Prediction of Short-Term Weather with the Help of Satellite Data and Neuro-Fuzzy Clustering

Mrinmoy Majumder and Tilottama Chackraborty

Abstract Weather is an essential part of decision making in large-scale manufacturing, agroforestry, electrical, and various other rainfall-dependent industries. The generation and distribution of electricity, water supply, and related essential amenities of any region also depend on daily weather variations. That is why short-range weather forecasting has a direct impact on the day-to-day operations of these industries. But due to the amount of manpower, equipment, and computational capacity involved in the prediction of short-term weather, many scientists are now opting to utilize remote sensing and image processing for the estimation of next-day or 3-day weather. In short-range weather forecasting, due to the complex and non-linear interrelationship between rainfall and weather parameters, quantification of the forecast can be error-prone. Probabilistic forecasting has proven to be more accurate. For such a level of nonlinearity as one finds in short-range forecasting, normal regression methods fail to achieve the level of accuracy required. That is why nature-based algorithms, which are well known for their ability to map the nonlinearity between dependent and independent variables, are more suitable than statistical algorithms. The present investigation tries to estimate the probability of occurrence of rainfall 1 day in advance based on the processing of satellite images and neuro-fuzzy clustering algorithms. A software module is also developed so that the model can be utilized to achieve given objectives.

Keywords Short-range weather forecasting • Neuro-fuzzy clustering • Image processing

M. Majumder (✉) • T. Chackraborty
School of Hydro-Informatics Engineering, National Institute of Technology Agartala,
Barjala, Jirania 799055, Tripura, India
e-mail: mmajumder15@gmail.com; tilottama86@gmail.com

16.1 Introduction

There are four types of weather forecast: short-range, extended-range, medium-range, and long-range. Short-range weather forecasts are forecasts of weather 48–72 h in advance. Weather forecasting involving 3–5, 5–7, and more than 10 days in advance is referred to as extended-range, mid-range, and long-range weather forecasting, respectively. The range for short-range weather forecasting (SRWF) is mainly concerned with weather systems observed in the latest weather charts and by considering the generation of new systems within a certain time period.

16.1.1 Importance of Short-Term Weather Forecasting

In recent days, because of abnormalities in climate patterns perceptible in many places of the world, SRWF has become very complex but very important. What follows are some of the reasons why SRWF is essential and its accuracy of utmost importance.

- Very SRWF and now casting is included in day-to-day decision making in the aviation and shipping industries.
- The impact of small-scale, short-lived, extreme phenomena like thunderstorms, tornadoes, flash floods, dense fog, and freezing rain on the socioeconomic lifestyle of vulnerable populations and agricultural productivity is high. That is why predictions of such short-duration events is important for the prevention of large-scale damage.
- Tropical storms or hurricanes are phenomena of special importance in short-range forecasting.
- SRWF is complex, and the amount of spatial and temporal detail required for forecasting decreases as the period increases.

Predictions from SRWF are used for the optimization of different managerial decisions. Weather can hamper the normal efficiency of any industry that is directly or indirectly connected to weather. In the distribution of electricity an extreme event lasting for a few minutes can wreck the entire supply framework. The demand for electricity also varies with weather patterns (Bolzern and Fronza 1986). A rise in temperature, along with humidity, often decreases the comfortability index (CI) of a region. The reduction in CI will also increase the demand for artificial wind sources like fans or air-conditioning systems, which in turn will increase the demand for electricity. But a nor'wester or westerlies, which are common after a few days of hot and humid weather, can reduce the temperature for a few days and increase the CI, which will again decrease the demand for electricity. But if a power plant is conditioned for excess output due to rising demand from the previous day, then the excess energy will be wasted because of this sudden reduction of demand.

The harvesting pattern for cash crops also requires knowledge of weather patterns and often use SRWF to decide about harvesting time for which maximum production and profit can be realized (Wilks et al. 1993).

That is why managers in related industries decide on their daily productivity after analyzing the weather forecast of a given day.

In this study a software was developed for the prediction of short-range weather (24 h) with the help of fuzzy logic and a neurogenetic algorithm. The input data were retrieved from cloud composition as observed from the satellite images of the target region. This method of weather forecasting can eliminate the need for expensive instruments and supercomputers or skilled workers for collecting relevant data from the target region.

Simple processing of satellite images showing the shape, pattern, and concentration of clouds of 16 square-shaped equal-area grids (eight grids of the first level and eight of the second) having a length or width of 105 km.

Predicting the probability of rainfall 24 h in advance requires satellite images showing the shape, size, thickness, texture, and type of cloud. The information thus collected must be entered into the software as input where the weight of the input and that of each grid adjacent to the target grid are estimated with the help of fuzzy logic in accordance with their influence on the occurrence of rainfall. Once the input data are submitted, the model will automatically estimate the probability of rainfall in the next 24 h.

The advantage of fuzzy logic (e.g., Niros and Tsekouras 2012; Maharaj and D'Urso 2011) and neural network models (e.g., Mitrofanov 2006; Kim et al. 2001) have been discussed in various scientific studies, and the relevancy of the fuzzy logic in application to the present problem domain is also available through various scientific articles published in related journals (e.g., Awan and Awais 2011; Cobaner 2011; Moustris et al. 2011; Chang et al. 2005; Cao and Chen 1983).

16.1.2 Satellite Imagery

Satellite imagery consists of photographs of Earth or other planets made by means of artificial satellites. Satellites equipped with high-resolution sensors revolve around Earth. During rotation the sensors try to capture images from the Earth's surface. These images are then sent to a receiving station on Earth for processing. Once the images are received by the Earth stations, they are converted into a computer readable format. The processed images are then used for the retrieval of relevant and extractable information. Various types of satellites are deployed to perform multiple jobs. The main apparatus of the satellites are the sensors, which are normally high-resolution cameras, or signal generators, which are sent to observatories on the Earth's surface for further processing.

16.1.2.1 Importance of Satellite Imagery

Images captured by satellites are used in meteorology, agriculture, geology, forestry, landscape, biodiversity conservation, regional planning, education, intelligence, and warfare. Satellite imagery is also used in seismology and oceanography in deducing changes to land formation, water depth and sea beds, by color caused by earthquakes,

volcanoes, and tsunamis; the potential of prescription and precision farming can also be monitored from the rich source of data retrievable from satellite images.

The images captured by satellites are further processed by transferring them from some accessible source. For example, one of the earliest two-dimensional Fourier transforms applied to digital image processing of NASA photos as well as national security applications by ESL Incorporated (Atlas and Hoffman 2000).

Due to the real-time observation and direct capture of images, satellite images can also produce real-time weather forecasting with greater accuracy and reliability. The images show the concentration, speed, and direction of storms, which enable weather forecasters to estimate more precise information for the areas in the path of a storm. By merely observing a satellite image, the forecaster can see a storm escalating or breaking apart, changing directions, or speeding up or slowing down.

Satellite images have made weather forecasting much more factual and functional than it was even just a few years ago because of the details and quantity of information obtainable. "Forecasters are now able to give the public a much better image of what to anticipate in the next week, day or hour because of information provided by these satellite images (Anton 2010)."

16.1.3 Clustering

Clustering is an unsupervised learning process to classify a collection of unlabeled data into different groups whose members are similar in some way to other members of the group but unlike members of other groups. A loose definition of clustering could be the process of organizing objects into groups whose members are similar in some way.

Thus cluster analysis can be defined as a collection of statistical methods to identify groups of samples that show similar characteristics.

16.1.3.1 Importance of Clustering

Cluster analysis is used to perform the following functions:

- To describe and to make spatial and temporal comparisons of communities (assemblages) of organisms in heterogeneous environments;
- To generate artificial phylogenies or clusters of organisms (individuals) at the species, genus, or higher level that share a number of attributes;
- To build groups of genes with related expression patterns (also known as coexpressed genes);
- To automatically assign genotypes;
- To infer population structures;
- To classify multivariate data from surveys and test panels in market research;
- To create a more relevant set of search results compared with normal search engines like Google; there currently exist a number of Web-based clustering tools such as Clusty;

- To divide a digital image into distinct regions;
- To locate and characterize extrema in a target distribution;
- To classify chemical properties in different sample locations.

The use of clustering is widespread in different decision-making problems where clusters are developed to identify problematic regions and accordingly solutions are searched to solve the problems. K-means, K-medoid, or hierarchical or principal component analysis are some of the methods for clustering data into groups of similar objects. Neuro-fuzzy clustering is one of the widely used hybrid clusterization methods where the classification simplicity of fuzzy logic and mapping capacity of neural networks are utilized to cluster a sample population based on the importance of input variables.

16.1.4 Neuro-Fuzzy Clustering

When a sample population obtained from experimental analysis or mathematical models is grouped according to expert-specified thresholds, uncertainty may creep in and can be removed only if the relevant information to identify the system is present and clustering is done by the application of fuzzy logic.

For example, a major problem in bioinformatics analysis or medical science is obtaining the correct diagnosis from certain important information. Nowadays computers are used to gather, store, analyze, and integrate patterns and biological data, which can then be applied to find a new, useful diagnosis or information with the help of neuro-fuzzy clustering techniques.

16.2 Description of the Software: CLIMAGE

CLIMAGE, or Climate from Image, is a software that allows users to estimate the probability of rainfall on the next day based on cloud information retrieved from the target grid and nine adjacent grids. Each of the variables is encoded with a weight representing its importance for the objective. All the weights are deduced based on their importance as decided by the implementation of the fuzzy logic.

The determination of weights was performed with the help of fuzzy logic. All the input variables were arranged in rows and columns. Then each input was compared with the other inputs with respect to its importance in estimation of the output. The rule utilized to rate the input variable with respect to every other variable is given below (Rule 1):

*If Input N is much more important than Input N-1, Then
give Input N a rating of 1.*

*If Input N is more important than Input N-1, Then
give Input N a rating of 2.*

If Input N is as important as Input N-1, Then

Table 16.1 Rating received by input variables with respect to other variables

	Cloud area	Cloud type	Cloud thickness	Cloud shape	Cloud texture	Rainfall
Cloud area	0	5	5	4	4	5
Cloud type	1	0	3	2	3	3
Cloud thickness	1	3	0	2	3	4
Cloud shape	2	3	4	0	3	4
Cloud texture	2	3	3	3	0	4
Rainfall	1	3	2	2	2	0

give Input N a rating of 3.

If Input N is less important than Input N-1, Then

give Input N a rating of 4.

If Input N is much less important than Input N-1, Then

give Input N a rating of 5.

Applying the same rule to rate each of the input variables with respect to every other input variable will produce a data matrix as shown in Table 16.1.

After rating each of the input variables, each row of the matrix is divided by the maximum rating achieved by the row. The dividend will be minimum for those variables that has the highest rating (Table 16.2).

That is why the complement of the dividend was taken as the weight of the variable. Similarly, the locations of adjacent grids were also rated, and their weight was also derived using the same method. Once the weight was decided, the same weight as for the variable and the grid was multiplied by the magnitude of the input variable for an image. The weighted averages of all the input variables were taken as the objective function. The value of the variables is also rated with the help of their influence on the occurrence of rainfall.

All the values of the input and output variables are then encoded into nine groups, each representing the intensity of the variables. The encoded category and the ratings given to them are presented in Table 16.3.

Similar to the Input variables the output variable or objective function is also grouped into nine categories based on magnitude. The category of the objective function represents the occurrence of rainfall in the target grid.

After determination of the weight for the input variables and the weight for the grids a combinatorial data matrix considering every possible combination among the input variables is prepared and the corresponding objective function is determined and encoded into relevant groups. This categorized data matrix is fed to the feedforward, fully connected neural network for training the network according to the conjugate gradient descent training algorithm, logistic activation function, and a topology as determined by a heuristic search method: 7-13-1.

The rule for encoding the objective function is prescribed as follows (Rule 2):

$$\begin{aligned}
 \text{OBJECTIVE FUNCTION}(Q) = & \left(\begin{aligned} & w_n \times A + w_n \times T_c + w_n \times d_c + w_n \times S_c \\ & + w_n \times T_x + w_n \times P / \end{aligned} \right) \\
 & \sum w_n \times (A + T_c + d_c + S_c + T_x + P)
 \end{aligned}
 \tag{16.1}$$

Table 16.2 Rating received by grids with respect to other grids

		1st level								2nd level							
		North	South	East	West	Northwest	Northeast	Southwest	Southeast	North	South	East	West	Northwest	Northeast	Southwest	Southeast
1st level	North																
	South	2				1	4	1	1	2	1	1	2	3	1	1	1
	East	4	3			3	5	3	3	3	3	3	4	4	2	2	3
	West	3	2	2		2	3	2	2	1	1	1	2	2	1	1	3
	Northeast	4	3	3	4	3	5	3	3	4	3	3	4	3	2	2	3
	Northwest	2	1	1	3	1	1	2	1	2	1	1	2	3	1	1	2
	Southeast	5	3	3	4	3	5	3	3	4	3	3	4	4	3	3	3
	Southwest	5	2	2	4	2	4	3	3	4	3	3	4	4	3	3	3
2nd Level	North	4	3	3	5	2	4	2	2	4	2	2	3	3	2	2	2
	South	5	3	3	5	3	5	3	3	4	4	2	4	4	3	3	3
	East	5	3	4	5	3	5	3	3	4	4	4	4	4	3	3	3
	West	4	2	3	4	2	4	2	2	3	2	2	4	3	2	2	2
	Northeast	5	4	4	5	4	5	3	3	4	3	3	4	4	3	3	3
	Northwest	3	2	2	4	3	3	2	2	3	2	2	3	2	2	2	2
	Southeast	5	4	4	5	4	5	3	3	4	3	3	4	4	3	4	3
	Southwest	5	3	3	5	3	4	3	3	4	3	3	4	4	3	4	3

Table 16.3 Groups utilized for encoding input variables

Variable name	Groups encoded	Rating
Cloud area (A)	EH, VH, SH, H, M, L, SL, VL, EL	9–1
Cloud type (T_c)	Cumulonimbus	11
	Nimbostratus	10
	Cirrocumulus	9
	Alto cumulus	8
	Stratocumulus	7
	Fog	6
	Cumulus	5
	Cirrus	4
	Cirrostratus	3
	Altostratus	2
	Stratus	1
Cloud thickness (d_c)	EH, VH, SH, H, M, L, SL, VL, EL	9–1
Cloud shape (S_c)	ELa, VL a, SL a, La, M, S, SS, VS, ES	9–1
Cloud texture (Tx_c)	Stormy	18
	Mixed with rain	17
	Light rain	16
	Overcast	15
	Jumpy	14
	Dark rain	13
	Incoming storm	12
	Hazy	11
	Stippled	10
	Fuzzy	9
	Fluffy	8
	Crepuscular	7
	Wispy	6
	Distant	5
	Diffuse	4
	Lumpy	3
	Line	2
	Vapor	1
Rainfall (P)	EH, VH, SH, H, M, L, SL, VL, EL	1–9
Objective function (Q)	EH, VH, SH, H, M, L, SL, VL, EL	9–1

If ($Q < 10$,
Then “EL”,
Else If ($Q < 25$,
Then “VL”
Else If ($Q < 35$,
Then “SL”,
Else If ($Q < 45$,
Then “L”,

Else If ($Q < 55$,
Then "M",
Else If ($Q < 65$,
Then "H",
Else If ($Q < 75$,
Then "SH",
Else If ($Q < 90$,
Then "VH",
Or Else "EH"

The training and testing correct classification rate was 99.99%, but the testing CCR was 77.04%. The reduction in testing accuracy can be attributed to the small amount of data available for testing purposes.

The network thus trained is utilized for prediction of rainfall in the target grid.

That is why if the necessary information of the input variables is given for some day, the next day occurrence probability can be predicted from this model.

After the probability of occurrence of rainfall is predicted for the target grid with respect to the other 16 grids surrounding the target grid the category of the objective variable as predicted by the software is rated using the same method as with the ratings shown in Table 16.3.

A weighted average is calculated based on the weight of the grids and the respective ratings of the objective function. The value of this weighted average is again converted into a category of the objective function to find the actual occurrence of rainfall in the target grid with respect to the surrounding 16 grids.

16.3 Benefits of CLIMAGE

The neural network model thus trained and tested was embedded into the software framework and can be distributed to different users for testing purposes. All the user needs to collect is the satellite images of nine grids representing the cloud variation. The necessary information about the cloud and the rainfall magnitude can be collected from the meteorologic stations nearest to the grid.

The relevant data when entered into the software will predict the occurrence probability of rainfall in the center grid. The prediction can be verified by waiting till the next day or by comparing with an already available set of data.

The benefits of the model are manifold. The model can be utilized by practitioners and experts in many fields like farmers, power generation companies, energy efficiency managers, business owners, disaster management groups, town planners, construction engineers, material managers, factory supervisors, etc.

The farmers can schedule their harvesting work according to the outcome from the model. Based on the requirements of the crop and the model prediction farmers can plan their harvesting schedule.

As rainfall directly impacts the demand for electricity by decreasing CI operation managers of generating units can adjust their production according to the outcome predicted by the software. The adjustment can save a lot of energy.

Energy efficiency managers can create an optimal plan for the usage of energy in running High Energy Consuming Appliances. The controlling of such high-energy devices can conserve a noticeable amount of energy. The conservation of energy will also increase income from selling carbon credits.

Rainfall or abnormal weather patterns can also impact the level of demand for a specific product. For example, when the temperature is high, sales of ice-cream are also high. But after a sudden shower, sales of ice-cream declines. Thus, a business owner whose profit mainly depends on the efficiency of his or her inventory will modify the management plan to counter a sudden reduction in demand.

The occurrence of extreme events implies that disaster management groups in the region will be under immense pressure to mitigate the sudden rise in the level of uncertainty. That is why if those groups can know in advance about an upcoming surge in rainfall, then they can prepare for the impact 1 day earlier.

Town planners can also benefit from the software. If they analyze the number of extreme events based on different variations of the input variables that are common in the targeted region, they can propose a plan that will already be optimized to prevent the more dire consequences of extreme events. The specification and maintenance required will also be preplanned. Thus a significant amount of money can be saved by saving energy and additional expenses.

Construction engineers can adjust their construction schedule according to the outcome of the model. They can plan in advance by utilizing the software for different combinations of those input variables that are most common in and around the construction site.

The quality of materials utilized at construction sites or manufacturing plants varies with the occurrence of monsoons. A sudden decrease in water content in the air can alter the quality of the materials themselves. Wide but frequent variation in temperature can either wear or tear materials affected by thermal conduction. That is why material managers can preplan their use of specific materials on a specific day when the chance of extreme events will be at a minimum.

Manufacturing and production units are controlled by weather patterns. Water can contaminate and reduce the quality of cement. If seepage of water remains undetected, then an entire inventory can be wasted. That is why managers in many companies are extremely concerned about the occurrence of rainfall on a given day. Accordingly, they prepare their production schedule and manage output. Thus, CLIMAGE can help them in planning at the postproduction stage for the level and time of output they desire on any given day.

Above all, laypeople can also avoid the hassles they may face due to a sudden influx of rainfall. The software can predict the weather 24 h in advance, making it possible to plan for the next day's activities.

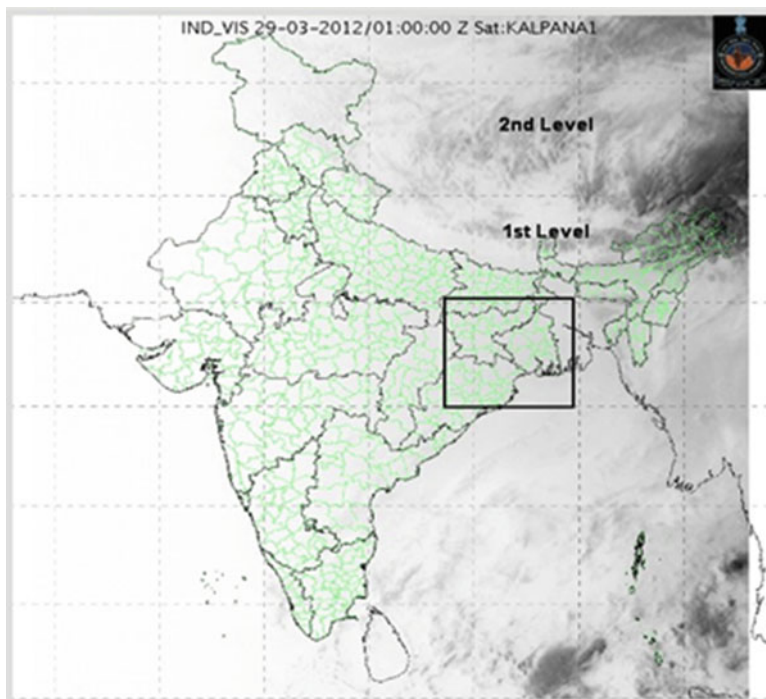


Fig. 16.1 Satellite picture of India as captured by INSAT IIIA

16.4 Case Study

Satellite imagery of India was captured by INSAT satellite Kalpana, where the presence of a cloud with its shape, size, and texture was clearly visible. An image from eastern India was captured and the already gridded image was processed to estimate the type, thickness, area, shape, and texture of the cloud. The data were fed to the software along with the data on present-day rainfall of that region. The software successfully predict rainfall on the following day at the same time of day.

The present study aims to estimate the probability of rainfall in the entire state of Orissa and southern part of West Bengal. The region is highlighted by the white box in Fig. 16.1.

According to the picture, the northeastern part of the target grid in the second level has a sufficient amount of rain clouds engulfing most of the regions of those grids.

But in other areas of the targeted grids no such clouds are observed. Accordingly, the categories of the input variables are configured and shown in Table 16.4.

Table 16.4 Categories assigned to input variables with respect to target grid depicted in Fig. 16.1

Grid	Cloud area	Cloud type	Cloud thickness	Cloud shape	Cloud texture	Rainfall	OBJ
First level							
Southwest	EL	St	EL	ES	Distant	EL	EL
Southeast	EL	St	EL	ES	Overcast	EL	VL
West	EL	St	EL	ES	Distant	EL	EL
South	EL	St	VL	VS	Line	EL	SH
Northwest	EL	St	VL	VS	Fluffy	EL	M
Northeast	VH	Ns	VH	L	Dark rain	EL	VL
North	M	Ns	M	ES	Light rain	EL	M
East	VL	Sc	M	ES	Light rain	EL	VH
Second level							
Southwest	EL	St	EL	ES	Diffuse	EL	EL
Southeast	M	Ns	M	M	Incoming storm	EL	EL
West	EL	St	EL	EL	Wispy	EL	EL
South	VL	St	L	SS	Light rain	EL	EL
Northwest	VL	St	L	SS	Distant	EL	EL
Northeast	EH	Cs	EH	EL	Stormy	EL	VH
North	VH	Ns	H	VL	Dark rain	EL	EH
East	H	Ns	SH	L	Mixed with rain	EL	VH
Next day (24 h in advance) probability of rainfall (according to Rule 2)							EL

After the value of the objective function is calculated with respect to the predicted category of the grids, the value of the function is found to be equal to 5.97% and is thus encoded into the EL group. Thus, the probability of rainfall in the next 24 h for the target grid is extremely low. The weather report from the next day shows that no rainfall had occurred in the 24 h period from the time this picture was taken.

16.5 Conclusion

This research tried to predict the occurrence of rainfall within a grid of interest by the application of neuro-fuzzy clustering and image processing. The study developed a software that can predict the occurrence of rainfall with respect to the previous day’s rainfall and cloud type, shape, thickness, and texture. The data of these input variables must be entered for the adjacent 16 grids. The weight according to the importance as determined by fuzzy logic for the grids as well as for the input variables is predetermined. The prediction work is performed by the neurogenetic model developed with the help of a combinatorial data matrix considering every possible situation that might occur among the selected input variables. The study

will help a range of people from fields such as meteorology and agriculture as it will allow them to prepare for the onset of weather extremes based on the outcome of the software. Preplanned mitigation measures can save both money and manpower and increase the efficiency of any system. The drawback of the software lies in its inability to predict the quantity of rainfall. In the future, features such as prediction of rainfall 5 days in advance will be introduced, but such a prediction would also require more complexity in the model. Training of neural networks requires a sufficient amount of data, but because the model is fed with a combinatorial data matrix, no other data set can be included in the training data set of the model.

References

- Anton (2010) The advantages of satellite images, Retrieved from <http://www.stella2000.com/the-advantages-of-satellite-images/> on 25th Feb 2013
- Atlas R, Hoffman RN (2000) Chapter 4: The use of satellite surface wind data to improve weather analysis and forecasting at the NASA Data Assimilation Office. Elsevier Oceanogr Ser 63:57–78
- Awan MSK, Awais MM (2011) Predicting weather events using fuzzy rule based system. Appl Soft Comput 11(1):56–63
- Bolzern P, Fronza G (1986) Role of weather inputs in short-term forecasting of electric load. Int J Elec Power Energy Syst 8(1):42–46
- Cao H, Chen G (1983) Some applications of fuzzy sets to meteorological forecasting. Fuzzy Set Syst 9(1–3):1–12
- Chang CL, Lo SL, Yu SL (2005) Applying fuzzy theory and genetic algorithm to interpolate precipitation. J Hydrol 314(1–4):92–104
- Cobaner M (2011) Evapotranspiration estimation by two different neuro-fuzzy inference systems. J Hydrol 398(3–4):292–302
- Kim JH, Hwang SH, Shin HSA (2001) Neuro-genetic approach for daily water demand forecasting. KSCE J Civ Eng 5(3):281–288
- Maharaj EA, D'Urso P (2011) Fuzzy clustering of time series in the frequency domain. Inform Sci 181(7):1187–1211
- Mitrofanov DG (2006) Neural network identification of aerial objects from radar images. Meas Tech 49(2):178–186
- Moustris KP, Larissi IK, Nastos PT, Paliatsos AG (2011) Precipitation forecast using artificial neural networks in specific regions of Greece. Water Res Manag 25(8):1979–1993
- Niros AD, Tsekouras GE (2012) A novel training algorithm for RBF neural network using a hybrid fuzzy clustering approach. Fuzzy Set Syst 193:62–84
- Wilks DS, Pitt RE, Fick GW (1993) Modeling optimal alfalfa harvest scheduling using short-range weather forecasts. Agr Syst 42(3):277–305