A Comprehensive Survey of Data Mining Techniques on Time Series Data for Rainfall Prediction

Neelam Mishra[1], Hemant Kumar Soni[2,*], Sanjiv Sharma[3] & A.K. Upadhyay[2]

[1]Department of Computer Science and Engineering, NRI College of Engineering and Management, Gwalior, Madhya Pradesh, India
[2]Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University, Madhya Pradesh, Gwalior, India
[3]Department of Computer Science and Engineering, Madhav Institute of Technology and Science, Gwalior, Madhya Pradesh, India
*E-mail: soni_hemant@rediffmail.com

Abstract. Time series data available in huge amounts can be used in decision-making. Such time series data can be converted into information to be used for forecasting. Various techniques are available for prediction and forecasting on the basis of time series data. Presently, the use of data mining techniques for this purpose is increasing day by day. In the present study, a comprehensive survey of data mining approaches and statistical techniques for rainfall prediction on time series data was conducted. A detailed comparison of different relevant techniques was also conducted and some plausible solutions are suggested for efficient time series data mining techniques for future algorithms.

1   Introduction

Data mining is used in various areas where ample amounts of data are available. With the help of data mining techniques, the user can extract valuable hidden information that can be helpful in decision-making or used for making predictions. Time series data collected over a constant period of time, viz. daily, weekly, monthly, quarterly or yearly, can be used by managers to inform appropriate decisions and identify suitable plans for the future, based on the assumption that past patterns will be repeated in the future. With the help of time series data analysis, long-term forecasting over years may be possible. This can lead to better assessment of future requirements and also planning for potential development [1]. Data mining is used in various domains. With the help of time series data, data mining is used among others for weather prediction.

In a country like India, where most of the farmers are dependent on rain for their crops and the growth and GDP of the country are based on agriculture, rainfall prediction is a sensitive and important issue. Rainfall prediction can be considered a significant and hot issue [2,3]. Intelligent forecasting models have achieved better results than traditional statistical methods. Although intelligent forecasting methods perform better, we can still improve their results in terms of accuracy in addition to other factors [4].

The main contribution of this paper is to present a comprehensive survey of traditional statistical methods along with the latest approaches to data mining for time series data analysis and rainfall forecasting. Further, a comparison of different approaches for rainfall prediction is given. Some plausible solutions for efficient weather prediction techniques are also suggested.

This paper is organized as follows. A detailed review methodology is given in Section 2. A review of the literature on statistical techniques for time series forecasting is presented in Section 3. A comprehensive survey on the use of data mining with time series data is presented in Section 4. In Section 5, some plausible solutions for efficient time series data mining techniques are given. The paper is closed in Section 6 with the conclusion and future research directions.

2   Review Methodology

In this study, we critically examined 51 papers published in various journals and conference proceedings. The paper selection strategy and evaluation criteria considered in this paper were as follows (see also Figure 1):

1. *Selection of papers*

The digital libraries of IEEE, ACM, Springer, Science Direct, World Scientific, Taylor and France, Research Gate, Delnet and Google Scholar were explored by using the keywords 'time series data mining', 'rainfall prediction', 'weather prediction', 'time series data analysis', 'statistical techniques', 'neural network', 'weather prediction', 'rainfall forecasting, 'intelligent forecasting model', 'rainfall and runoff patterns', 'genetic algorithm for time series data mining', 'evolutionary algorithm for time series data mining', 'neural network', etc.

In this study, we also downloaded papers published in the proceedings of international and national conferences and seminars available online. Papers published as hard copy of proceedings were not considered due to unavailability. We found 103 papers and selected 51 papers for review and filtered out 52 papers that were not useful or not relevant.

2. *Parameters*

The following parameters were considered to evaluate the proposals made by various researchers and academicians:

1. Techniques used for time series data analysis
2. Length of temporal database
3. Parameters used for prediction
4. Duration of forecasting
5. Size of the database
6. Performance of the techniques

We classified the papers in following categories:

1. Conventional approach by using statistical methods
2. Data mining techniques for time series data analysis
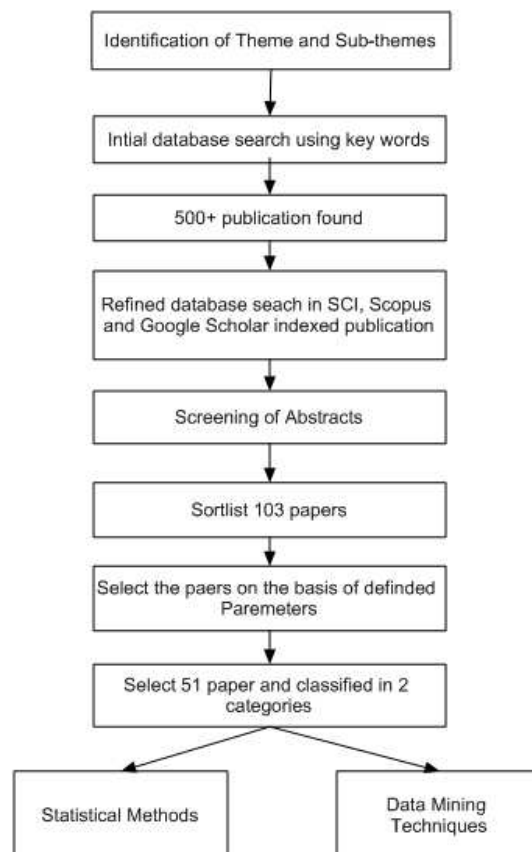
Figure 1 Review methodology for paper selection.

3   Statistical Techniques for Time Series Data Analysis for Rainfall Prediction

Various organizations in India and abroad have done modeling using time series data. Different methodologies have been applied, viz. Statistic Decomposition Models (SDM), Exponential Smoothing Models (ESM), ARIMA models and their variations like Seasonal ARIMA Models (SAM), Vector ARIMA Models (VAM) using variable time series, ARMAX models, i.e. ARIMA with instructive variables, etc. Numerous studies have been conducted on the analysis of patterns and distribution of rainfall in different regions of the globe. Various time series methods with different objectives have been used to investigate rain information in numerous literatures. Different parameters have been used by various researchers to predict rainfall, runoff, heavy rain, monthly and annual rainfall. A detailed comparison of the various statistical techniques for time series data analysis is given in Table 1.

Stringer [5] reports that a minimum of thirty-five quasi-periods of over one year long are revealed in records of pressure, temperature, precipitation, and extreme climatic conditions over the earth of the world surface. A very common quasi-periodic oscillation is the quasi-biennial oscillation (QBO), during which the environmental condition events recur each 2 to 2.5 years. Winstanley [6,7] predicts that from the years 1957 to 1970, the monsoon rain decreased over 500 times from Africa to India. It is also expected that by the year 2030, the long monsoon rain will have decreased to a minimum.

Harvey, *et al*. [8] have investigated how patterns of rainfall correlate with general climatic conditions and the frequency of rain cycles. They used rain information from Brazil for a selected region that frequently suffers from drought to assess the alternate behavior of rain. They used the Stochastic Cycles Model, which permits alternate parts to be modeled explicitly. They found that cyclical components are random instead of deterministic and also that the gains achieved from a forecast by taking account of the cyclic element are tiny in the case of Brazil.

Kuo and Sun [9] exploited an average 10-day stream flow forecast by using an intervention model. They synthesized and investigated the factors that affect the extraordinary phenomena caused by typhoons and different severe irregularities in the weather of the Tanshui river basin in Taiwan. Chiew, *et al*. [10] carried out an evaluation of six rainfall runoff modeling approaches to simulate daily, monthly and annual flows in eight unregulated catchments. They found that a time series approach offers sufficient estimates of monthly and annual yields within the water resources of the catchments.

Langu [11] applied time series analysis to identify changes in rainfall and runoff patterns. These patterns help in finding significant changes in rainfall series. The author used statistic analysis to scrutinize changes in rainfall and runoff patterns to identify important alterations in rainfall statistics. In the early 1970s, Box and Jenkins [1] led in developing methodologies for statistic modeling within univariate cases, often referred to as Univariate Box-Jenkins (UBJ) ARIMA modeling. Based on this, many researchers have developed different approached, viz. time series decomposition models, exponential smoothing model, vector ARIMA, ARNAx, etc.

Carter and Elsner [12] followed the outcome from a factor analysis regionalization of non- tropical storm convective rainfall over the island of Puerto Rico. They used a statistical technique to explore its potential to predict rainfall over limited areas. Island regionalization was carried out on a 15-year dataset. A set covering 3 years of surface and rainfall data was used in this predictive model. Surface data from two first-order stations were adopted as input to a partially adaptive classification tree in order to forecast the incidence of heavy rain.

Al-Ansari and Baban [13] have proposed an applied math analysis of rainfall measurements for 3 meteorological stations in Jordan: Amman aerodrome (central Jordan), Irbid (northern Jordan) and Mafraq (eastern Jordan). Traditional applied math and power spectrum analyses as well as an ARIMA model were applied to semi-permanent annual rainfall measurements from the 3 stations. The result showed that potential periodicities in the order of 2.3 - 3.45, 2.5 - 3.4 and 2.44-4.1 years for Amman, Irbid and Mafraq stations, respectively, were obtained. A statistic model for every station was adjusted, processed, diagnostically checked and finally an ARIMA model for every station was established with a 95% confidence interval and also the model was used to forecast annual rainfall values over five years for Amman, Irbid and Mafraq meteorological stations.

Al-Ansari, *et al.* [14] used statistical analysis of rain records at 3 major meteorological stations in Jordan. The authors performed normal statistical, harmonic and power spectrum analysis and time series analysis. An ARIMA model for each station was established with a 95% confidence interval. The results showed a decreasing trend for forecasted rainfall results in all stations. Ingsrisawang [15] implemented three statistical techniques: First-order Markov Chain, Logistic model, and Generalized Estimating Equation (GEE) in modeling the rainfall prediction over the eastern part of Thailand. Two daily datasets were used, called Meteor and GPCM, collected during 2004-2008. By the combination of the GPCM dataset and the Meteor data, the GPCM+Meteor

dataset was generated. With the help of the Meteor dataset, the First-order Markov Chain model was implemented.

Seyed, *et al*. [16] modeled weather parameters using random methods (ARIMA Model). The authors used time series methods to model weather parameters in Iran at Abadeh Station and recommended ARIMA(0,0,1)(1,1,1) as the best fit for monthly rainfall data and ARIMA(2,1,0)(2,1,0) for monthly average temperature for Abadeh station. Mahsin, *et al*. [17] used the Box-Jenkins technique to create a seasonal ARIMA model for monthly rainfall information taken from Dhaka Station, Bangladesh, covering 30 years, from 1981-2010. In their paper, the ARIMA (0,0,1)(0,1,1) model was found adequate. This model was used for forecasting monthly rainfall.

Table 1   Comparison of various statistical techniques for Time Series data analysis.

S.No.	Author(s)	Technique used	Area/ country	Dataset used	Data used (for no. of years)	Parameter used	Prediction
1.	Stringer [5]	Quasi-biennial Oscillation (QBO)	Africa	Thirty-five quasi-periods	1 Year	Pressure, temperature, precipitation, and extreme climatic conditions	2-2.5 years
2.	Winstanley [6,7]	Rainfall Observation	Africa - India	Monsoon rain	13 years	Rain	Up to 2030
3.	Harvey, *et al.* [8]	Modelling Stochastic Cycles	Brazil	131 annual rainfall	130 years	Drought	1 year
4.	Kuo & Sun [9]	Intervention Model	Taiwan	Rainfall	1 year	Typhoons	Average 10 days
5.	Chiew, *et al.*, (1993)[10]	Rainfall-Runoff Model	Eight unregulated catchments	Rainfall	Not given	Catchments	Monthly and annual
6.	Langu [11]	Statistic analysis	Not specified	Rainfall series	Not given	Changes in rainfall	Rainfall and runoff patterns
7.	Carter & Elsner [12]	Statistical technique	Puerto Rico	Surface and rainfall data	3 years	Non-tropical storm convective	Heavy rain
8.	Al-Ansari & Baban [13]	Applied Math Analysis, Power Spectrum Analyses, ARIMA	Jordan	Rainfall data	More than 5 years	95% confidence interval	5 years

Table 1 *Continued*. Comparison of various statistical techniques for Time Series data analysis.

S.No.	Author(s)	Technique used	Area/ country	Dataset used	Data used (for no. of years)	Parameter used	Prediction
9.	Ingsrisawang, *et al.* [15]	First-order Markov Chain, Logistic model, and Generalized Estimating Equation	Thailand	GPCM + Meteor datasets	4 years	Climate of previous day	Prediction of rainfall estimates on wet days
10.	Seyed, *et al.* [16]	ARIMA	Iran	Weather data	More than 2 years	Weather parameter	Monthly rainfall information and monthly average temperature
11.	Mahsin, *et al.* [17]	Box-Jenkins Technique + ARIMA	Bangladesh	Rainfall data	30 years	Rainfall	Monthly rain

4 Data Mining Techniques for Rainfall Forecasting

Data mining is now used in various domains, including time series data. Time series data analysis is used for weather forecasting or rainfall prediction with the help of data mining techniques. For time series data analysis, intelligent forecasting models perform better than methods that are traditionally used in forecasting. Neural network (NN) and genetic algorithm (GA) are two of the most popular techniques based on computational intelligence. In the literature, hybrid methods, which consist of combining more than one technique, are also commonly found. There are two significant categories for time series forecasting, i.e. neural network based methods and evolutionary computation based methods.

Neural networks are extensively used to model a number of nonlinear hydrological processes such as weather forecasting. The ASCE Task committee has presented some ideas about the application of artificial neural networks (ANN) in the geophysical science domain [18]. Hu [19] uses the concept of ANN for weather forecasting. This was one of the very first attempts to implement a soft computing technique in this domain, which opened up a new dimension in environment-related research.

French, *et al*. [20] suggested a two-dimensional rainfall-forecasting model, which predicts 1 hour prior to occurrence. This ANN model is basically a mathematical rainfall simulation model. The results of this model can be taken as input for further forecasting. However, in this model interaction and training time are not balanced. Another issue was that in the comparison of input and output nodes, the quantity of hidden layers and hidden nodes was insufficient. This was required to reserve the upper-order relationship for adequately abstracting the method. Although there were many other issues with this scheme, it was the first attempt to use ANN on geophysical processes. Michaelides, *et al*. [21] evaluated the performance of ANN and judged it against multiple regression. They worked on the estimation of missing rainfall information for the Cyprus region. Kalogirou, *et al*. [22] also attempted the application of ANN, using time series data to reconstruct rainfall information for Cyprus.

Adyal and Collopy [23] present 11 guidelines to assess ANN. They implemented their theory of NNs to business forecasting and prediction. During 1988 to 1994 they conducted 48 studies. For each study, they assessed the effectiveness of the proposed technique in comparison to alternatives like the effectiveness of validation. They also worked on the effectiveness of implementation. In their research, they found that only eleven studies out of the total number of studies were effectively validated and implemented, whereas another eleven studies were effectively validated and generated positive results. Within these 22 studies, they found better results using a neural network in 18 studies. Lee, *et al*. [2] used ANN for rainfall forecasting by grouping the available data into subpopulations. Wong, *et al*. [3] applied fuzzy rules based on the Kyrgyzstani monetary unit and back-propagation neural networks. This model predicts rainfall over Switzerland using spatial interpolation.

Pucheta, *et al*. [27] designed a feed-forward NN based NAR model for forecasting time series. The Levenberg-Marquardt method was adopted for learning rules to adjust the NN weights. The technique examined 5 time series obtained from Mackey-Glass delay differential equations and from monthly cumulative rainfall. Three sets of parameters for the MG solution were used. Herein, the monthly cumulative rainfall belongs to two different sites and time periods, i.e. La Perla during the years 1962-1971 and Santa Francisca during the years 2000-2010, both located in Córdoba, Argentina. This technique predicts 18 future values of each time series simulated by 500 Monte Carlo trials to specify the variance using fractional Gaussian noise.

Adhikari and Agarwal [28] comprehensively explored the outstanding ability of artificial neural networks in recognizing and forecasting strong seasonal patterns without removing them from the raw data. Six real-world time series

data with dominant seasonal fluctuations were used in this work. The empirical results showed that properly designed ANNs are remarkably efficient in forecasting strong seasonal variation and outperform each of the three statistical models for all six-time series.

Nanda, *et al.* [29] worked on various artificial neural network models, including Multi Layer Perceptron (MLP), Functional Link Artificial Neural Network (FLANN) and Legendre Polynomial Equation (LPE). They observed that MLP, FLANN and LPE performed better for time series data prediction. In their work, the authors proposed an ARIMA-based approach with ANN. A simulation study was carried out using MATLAB and was validated using data collected from India Meteorological Department covering June to September 2012. The authors claim that the FLANN predictions were better and closer in comparison to ARIMA with less Absolute Average Percentage Error (AAPE). Sethi, *et al.* [30] introduced a multiple linear regression (MLR) technique for rainfall prediction. They followed an empirical statistical technique and used 30 years of climate data from Udaipur City, Rajasthan India. The climate data included average temperature, rainfall precipitation, cloud cover over the city, and vapor pressure. The authors performed an experiment to evaluate the rainfall prediction accuracy. To identify the quality of the MLR they compared the prediction with actual data. With the help of graphs the authors showed that their method generates values that are close to the actual results. Prasad and Neeraj [31] conducted a study on weather prediction using data covering 9 years for Basra City. They used data mining techniques such as association rule mining, aggregation, classification and outlier analysis for weather prediction.

Apart from the abovementioned literature, some other work has been done by various researchers [24-27, 32-35]. Many authors, like Wang and Sheng [36], Htike and Khalifa [39], Phusakulkajorn [42], Charaniya, *et al.* [45] developed neural network based techniques for rainfall prediction. The CART and C4.5 technique proposed by Ji, *et al.* [43] was developed for hourly rainfall prediction, while Phusakulkajorn's method [42] forecasts daily rainfall based on previous rainfall data. Techniques proposed by Jesada, *et al.* [38], Htike and Khalifa [39], Charaniya, *et al.* [45], Jin, *et al.* [49], and Suhartono, *et al.* [50] predict monthly rainfall. Only Wang and Sheng [36], Kannan, *et al.* [37], Awan, *et al.* [47] proposed methodologies that predict the rainfall for a year or more. The NNARX and ANFIS technique proposed by Ramesan, *et al.* [40] is the only technique that predicts rainfall runoff. As far as the accuracy is concerned, it is between 72.3%, acquired by Decision Tree using the SLIQ technique developed by Narasimha, *et al.* [51], to more than 99%, acquired by the CART and C4.5 techniques proposed by Ji, *et al.* [43]. A detailed comparison of different data mining techniques used for time series data analysis is given in Table 2.

Table 2 Comparison of different data mining techniques used for Time Series data analysis.

S. No.	Author(s)	Technique used	Comparison with	Performance of the technique	Characteristics of the technique	Prediction
1	Wang & Sheng [36]	Generalized Regression Neural Network	BP Neural Network	Accuracy better than BP	Simple and stable network structure is	Annual rainfall
2	Kannan, *et al.* [37]	Pearson Coefficient	Regression Approach	The predicted values lie below the computed values.	Shows an approximate value	Five years
3	Jesada, *et al.* [38]	Fuzzy Inference System	Box-Jenkins and artificial neural networks model	Good alternative method to predict accurately.	Accuracy and human-understand-able prediction mechanism	Monthly rainfall
4	Htike & Khalifa [39]	Focused Time Delay Neural Network	Conventional techniques	Yearly dataset gave most accurate results	Suitable for time series prediction	Monthly, quarterly and annually
5	Ramesan, *et al.* [40]	NNARX and ANFIS	Conventional techniques	Work efficiently in rainfall – runoff model	Accurate and reliable in run-off prediction	Rainfall-runoff
6	Castro, *et al.* [41]	Neuro – Fuzzy Neuron Technique	Dynamic downscaling model	Improved results in compare to the dynamic model, using RMSE	Low computational cost	Seasonal rainfall forecast
7	Phusakulka-jorn [42]	Artificial Neural Network and Wavelet Decompo-sition	ANN model without the transformation of time series by using wavelet decomposition	Satisfactory Accuracy in one-day daily rainfall prediction with R2=0.9948 & RMSE=0.9852m m.	Identify ANN based wavelet transform as a practical tool	Daily rainfall based on previous of rainfall data
8	Ji, *et al.* [43]	CART and C4.5hourly rainfall	Conventional model	99.2% accuracy predicted C4.5 predicted accurately 99.3%.	High accuracy	Hourly rainfall
9	Cheng, *et al.* [44]	Gray theory with Markov chain	Conventional method	Less fit for random and volatile data sequence.	A new approach to forecasting the volatile random objects	Rainfall prediction
10	Charaniya, *et al.* [45]	Artificial Neural Network based Model with Wavelet Decompo-sition	Conventional method	Reasonably Accurate	Reliable rainfall prediction	Monthly rainfall

Table 2 *Continued.* Comparison of different data mining techniques used for Time Series data analysis.

S. No.	Author(s)	Technique used	Comparison with	Performance of the technique	Characteris-tics of the technique	Prediction
11	Liu, *et al.* [46]	Improved Naïve Bayes Classifier (INBC) technique	Genetic algorithm with average classification	Accuracy rate 90% on the rain/no-rain classification problems.	Rainfall prediction with (Depth3) & (Depth5), which are around 65%-70%	Rainfall prediction
12	Awan, *et al.* [47]	BP and learning vector Quantization	Multiple linear regressions and statistical downscaling models	Better performance in accuracy, better lead time and fewer resources required.	LVQ takes less training time than BP	One year
13	Du, *et al.* [48]	Immune Evolutionary Algorithm based on back propagation network	BP network algorithm model	Higher accuracy and better stability.	Solving complicated problems of optimization	Rainfall prediction
14	Jin, *et al.* [49]	PSO-NN Ensemble Prediction (PNNEP) model	Traditional linear statistical forecast method	Superior predicting capability	Enhanced generalization capacity	Monthly mean rainfall
15	Suhartono, *et al.* [50]	Ensemble method based on ANFIS and ARIMA	Ensemble method	Individual method is more accurate	Results in line with M3 competition results	Monthly rainfall
16	Narasimha, *et al.* [51]	Decision Tree Method using SLIQ	Fuzzy Logic, NN	Gives accuracy of 72.3%	Classification rule is generated	Rainfall prediction

5   Some Plausible Solutions for Efficient Time series Data Mining Techniques

Over the last two decades, in the area of time series data mining, a number of algorithms have been proposed. Time series data mining is applied in various areas, such economy, climate forecasting, medical surveillance, hydrology, etc. Hence there is a requirement and scope of developing new algorithms to deal with problem complexity and to achieve accurate results [32,33]. On the basis of this deep study, critical review, identification of shortcomings in existing techniques and development of new methodologies, some characteristics were identified.

The following are some characteristics that should be incorporated in future algorithms for better results:

1. **Data representation**: Proper representation of time series data is important because it is very difficult to manipulate the original structure of time series data. Since time series data are highly dimensional, implementing available data mining techniques is very difficult and hence it is very much required to identify a proper way to represent time series data.
2. **Stream analysis**: New algorithms must incorporate the handling capability of the data since in the development of hardware and networking technology and advancement in bandwidth capabilities, massive streams are generated with fluctuating data, hence analysis in mining such data flows is an important issue. There is an urgent requirement to design new techniques that can cope with ever flowing data streams.
3. **Pattern matching**: Time series data are collected at fixed intervals of time. The collected data have different lengths. Therefore, it is required that future algorithms have the capability of conducting pattern-matching effectively and efficiently for complete-sequence and subsequence pattern matching.
4. **Multi-parameter handling**: Accurate or close prediction on the basis of time series data mining is challenging. This can be done only when algorithms have the capability to handle multiple parameters simultaneously.
5. **User interaction**: The ultimate objective of time series data mining is to provide higher-order knowledge with an effective and efficient solution to the user. It is suggested that user interaction should be incorporated in future algorithms for dynamic exploration and refinement of solutions.
6. **Incremental data**: Time series datasets are huge and grow continuously in a rapid manner. The existing time series data mining algorithms support static data. As and when new data are included, the existing algorithms start again from scratch. This is simply a waste of previously mined results, time, resources and human efforts. It is strongly recommended that the handling of incremental data should be part of future algorithms.

6   Conclusion

In the present paper, the authors studied various statistical techniques for time series data analysis for rainfall prediction. During the survey, various models were studied, for example the stochastic cycles model, intervention model, rainfall and runoff patterns model, Univariate Box-Jenkins (UBJ) ARIMA modeling, first-order Markov chain, logistic model, generalized estimating equation (GEE), etc. Further, statistical data mining approaches for rainfall forecasting were also taken into consideration. Artificial neural networks were

applied on datasets collected from various sources, such as the geophysical science domain, environmental related research, microwave radar, satellite, weather station information, etc. The authors also discussed the use of Generalized, Fuzzy Inference System, Regression Neural Network, Focused Time Delay Neural Network, etc. More than 50 papers were studied thoroughly to find plausible solutions for efficient time series data mining techniques. In Tables 1 and 2, the authors present various aspects of feasible and suitable techniques for rainfall prediction.

On the basis of this study, the authors conclude that data mining is a technique that can be used not only for finding hidden information or patterns but also in forecasting. Time series data analysis is an example of weather prediction by using data mining. It was clearly observed that the use of soft computing techniques and evolutionary algorithms can also be applied successfully in time series data analysis.

This paper presents a comprehensive study of statistical techniques and data mining approaches for time series data analysis. It was found that statistical techniques and neural network based techniques can be used in parallel in time series data analysis. However with the help of soft computing techniques, better results may be found.

On the basis of the above study, the following issues are identified as unaddressed:

1. Most of the abovementioned researches used artificial neural networks, whereas other techniques such as genetic algorithms, swarm optimization etc. can also work and may give better results.
2. Accuracy is a major concern. There is solid evidence of accurate predictions.
3. All research is done for specific areas. The environment, atmosphere and weather parameters vary for each and every area/location. The results of one area could be used as a standard for other locations.
4. All methods give predictions for different periods of time. Hence, it is difficult to identify which method/technique is best for rainfall prediction.
5. The studies only represent rainfall prediction. Other phenomena than this, such as rainfall-runoff patterns, effects of climate change, tsunamis etc., are not covered.

From this study, it is observed that time series data analysis has the attention of many researchers. However, a number of research issues remain unexplored. There is a strong need to develop new models for time series data analysis that provide more accurate and region-specific weather forecasting. Nature-inspired algorithms may be helpful in this.

References

[1]   Box, G.E.P., Jenkins, G.M. & Reinsel, G.C., *Time Series Analysis: Forecasting and Control*, Pearson Education, 1994.

[2]   Lee, S., Cho, S. & Wong, P.M., *Rainfall Prediction Using Artificial Neural Network*, J. Geog. Inf. Decision Anal., **2**(2), pp. 233-242, 1998.

[3]   Wong, K.W., Wong, P.M., Gedeon, T. & Fung, C.C., *Rainfall Prediction Model Using Soft Computing Technique*, Soft. Computing, **7**(6) pp. 434-438, 2003.

[4]   Azevedo, J.M., Almeida, R. & Almeida, P., *Using Data Mining with Time Series Data in Short-Term Stocks Prediction: A Literature Review*, International Journal of Intelligence Science, **2**, pp. 76-180, 2012.

[5]   Stringer, E.T., *Techniques of Climatology*, San Francisco: WH Freeman & Co., 1972.

[6]   Winstanley, D., *Recent Rainfall Trends in Africa, The Middle East and India*, Nature, **243**, pp. 464-465, 1973.

[7]   Winstanley, D., *Rain Patterns and General Atmospherical Circulation*, Nature, **245**, pp. 190-194, 1973.

[8]   Harvey, R., Andrew, C. & Souza, R.C., *Assessing and Modeling The Alternate Behavior of Rainfall in North-East Brazil*, Journal of Climate and Applied Meteorology, **26**(10), pp. 1339-1344, 1987.

[9]   Kuo, J.T. & Sun, Y.H., *An Intervention Model for Average 10 Day Stream Flow Forecast and Synthesis,* Journal of Hydrology, **151**(1), pp. 35-56, 1993.

[10]   Chiew, F.H.S., Stewardson, M.J. & McMahon, T.A., *Comparison of Six Rainfall-Runoff Modeling Approaches*, Journal of Hydrology, **147**(1), pp. 1-36, 1993.

[11]   Langu, E.M., *Detection of Changes in Rainfall l and Runoff Patterns*, Journal of Hydrology, **147**, pp. 153-167,1993.

[12]   Carter, M.M. & Elsner, D.J.B., *A Statistical Method for Forecasting Rainfall over Puerto Rico*, American Meteorological Society, **12**(3), pp. 515-525, 1997.

[13]   Al-Ansari, N.A. & Baban, S., *Rainfall Trends in the Badia Region of Jordan*, Surveying and Land Information Science, **65**(4), pp. 233-243, 2005.

[14]   Al-Ansari, N.A., Al-Shamali, B. & Shatnawi, A., *Statistical Analysis at Three Major Meteorological Stations*, Jordan, Al Manara Journal for Scientific Studies and Research, **12**, pp. 93-120, 2006.

[15]   Ingsrisawang, L., Ingsriswang, S., Luenam, P., Trisaranuwatana, P., Klinpratoom, S., Aungsuratana, P. & Khantiyanan, W., *Applications of Statistical Methods for Rainfall Prediction over The Eastern Thailand*, Proceedings of the Multi Conference of Engineers and Computer Scientists, March 17-19, Hong Kong, IMECS 2010, III: 2024-2027, 2010.

[16] Seyed, A., Shamsnia, M., Naeem, S. & Ali, L., *Modelling weather parameter using stochastic methods (ARIMA Model)(Case Study:Abadeh station, Iran)*, 2011 International conference on Environment and Industrial innovation IPCBEE, 12, 2011.

[17] Mahsin, M.D., Yesmin, A. & Monira, B., *Modeling Rain in Dacca (National Capital) Division of Bangladesh Using Time Series Analysis*, Journal of Mathematical Modelling and Application, **1**(5), pp. 67-73, 2012.

[18] ASCE: Task Committee on Application of Artificial Neural Networks in Hydrology, *I: Preliminary Concepts*, J. Hydrol. Eng., **5**(2), pp. 115-123, 2000.

[19] Hu, M.J.C., *Application of ADALINE System to Weather Forecasting, Technical Report,* Master Thesis, Technical Report 6775-1, Stanford Electronic Laboratories, Stanford, United States, 1964.

[20] French, M.N., Krajewski, W.F. & Cuykendall, R.R., *Rainfall Forecasting in Space and Time Using Neural Networ*k, Journal of Hydrology, **137**(1), pp. 1-31, 1992.

[21] Michaelides, S.C., Neocleous, C.C. & Schizas, C.N., *Artificial Neural Networks and Multiple Linear Regression in Estimating Missing Rainfall Data,* Proceedings of the DSP95 International Conference on Digital Signal Processing, X. Limassol, Cyprus, pp. 668-673, 1992.

[22] Kalogirou, S.A., Neocleous, C.C. & Schizas, C.N., *Artificial Neural Networks for the Estimation of the Performance of a Parabolic Trough Collector Steam Generation System*, Proceedings of the International Conference EANN'97, Stockholm, Sweden, pp. 227-232, 1997.

[23] Adya, M. & Collopy, F., *How Effective are Neural Networks at Forecasting and Prediction*? *A Review and Evaluation*, J. Forecast., **17**(5-6), pp. 481-495, 1998.

[24] Koizumi, K., *An Objective Method to Modify Numerical Model Forecasts with Newly Given Weather Data Using an Artificial Neural Network*, Weather Forecast, **14**(1), pp. 109-118, 1999.

[25] Toth, E., Brath, A. & Montanari, A., *Comparison of Short-Term Rainfall Prediction Models for Real-Time Flood Forecasting*, Journal of Hydrology, **239**(1), pp. 132-147, 2000.

[26] Abraham, A., Steinberg, D. & Philip, N.S., *Rainfall Forecasting Using Soft Computing Models and Multivariate Adaptive Regression. Splines*, IEEE SMC Transactions, Special Issue on Fusion of Soft Computing and Hard Computing in Industrial Applications, **1**, pp. 1-6, 2001.

[27] Pucheta, J.A, Rivero, C.M.R., Herrera, M.R., Patiño, H.D. & Kuchen, B.R., *A Feed-Forward Neural Networks-Based Nonlinear Autoregressive Model for Forecasting Time Series*, Computación y Sistemas, **14**(4), pp. 423-435, 2011.

[28] Adhikari, R. & Agrawal, R.K., *Forecasting Strong Seasonal Time Series with Artificial Neural Network*, Journal of Scientific and Industrial Research, **71**(10), pp. 657-666, 2012.

[29] Nanda, S.K., Tripathy, D.P. Nayak, S.K., & Mohapatra, S., *Prediction of Rainfall in India Using Artificial Intelligent Systems and Applications*, **12**, pp. 1-22, 2013.

[30] Sethi, N. & Garg, K., *Exploiting Data Mining Technique for Rainfall Prediction*, International Journal of Computer Science and Information Technologies, **5**(3), pp. 3982-3984, 2014.

[31] Prasad, R.K. & Nejres, S.M., *Use of Data Mining Techniques for Weather Data in Basra City*, International Journal of Advanced Research in Computer Science and Software Engineering, **5**(12), pp. 135-139, 2015.

[32] Fu, T., *A Review on Time Series Data Mining*, Engineering Applications of Artificial Intelligence, **24**(1), pp. 164-181, 2011.

[33] Esling, P. & Agon, C., *Time series Data Mining*, ACM Comput. Surv. **45**(1), pp. 12:1-12:34, 2012.

[34] Shoba, G. & Shobha, G., *Rainfall Prediction Using Data Mining Techniques: A Survey*, International Journal of Engineering and Computer Science, **3**(5), pp. 6206-6211, 2014.

[35] Mishra, N. & Jain, A., *Time Series Data Analysis for Forecasting – A Literature Review*, International Journal of Modern Engineering Research, **4**(7), pp. 1-5, 2014.

[36] Wang, Z. & Sheng, H., *Rainfall Prediction Using Generalized Regression Neural Network: Case study Zhengzhou*, International Conference on Computational and Information Sciences, pp. 1265-1268, 2010.

[37] Kannan, M., Prabhakaran, S. & Ramachandran, P., *Rainfall Forecasting Using Data Mining Technique*, International Journal of Engineering and Technology, **2**(6), pp. 397-401, 2010.

[38] Kajornrit, J., Wong, K.W. & Fung, C.C., *Rainfall Prediction in the Northeast Region of Thailand Using Modular Fuzzy Inference System*, World Congress on Computational Intelligence, **10**(15), pp. 136-141, 2012.

[39] Htike, K.K. & Khalifa, O.O., *Rainfall Forecasting Models Using Focused Time-delay Neural Networks*, International Conference on Computer and Communication Engineering, **11**(13), pp. 1-6, 2010.

[40] Shamin, M.A., Han, D. & Mathew, J., *ANFIS and NNARX Based Rainfall-Runoff Modeling*, IEEE International Conference on Systems Man and Cybernetics, pp. 1454-1459, 2008.

[41] Castro, T.N., Souza, F., Alves, J.M.B., Pontes, R.S.T., Firmino, M.B.M. & Pereira, T.M., *Seasonal Rainfall Forest using A Neo-Fuzzy Neuron*

Model, IEEE International Conference on Industrial Informatics (INDIN), pp. 694-698, 2011.

[42] Phusakulkajorn, W., Lursinsap, C. & Asavanant, J., *Wavelet-Transform Based Artificial Neural Network for Daily Rainfall Prediction in Southern Thailand*, 9th International Symposium on Communications and Information Technology*, Icheon, pp. 432-437, 2009.

[43] Ji, S.-Y., Sharma, S., Yu, B. & Jeong, D. H., *Designing a Rule-Based Hourly Rainfall Prediction Model*, IEEE IRI, August 8-19, pp. 303-308, 2012.

[44] Cheng, L., Tian, Y-M. & Wang, X-H., *Study of Rainfall Prediction Model Based on GM (1,1) – Markov Chain*, National Water Pollution Control and Management of Science and Technology Project of China and the Science and Technology Innovation Special Foundation of Tianjin, pp. 744-747, 2011.

[45] Charaniya, N.A. & Dudul, S.V., *Committee of Artificial Neural Networks for Monthly Rainfall Prediction using Wavelet Transform*, International Conference on Business, Engineering and Industrial Applications, pp. 125-129, 2011.

[46] Liu, J.N.K., Li, B.N.L. & Dillon, T.S., *An Improved Naïve Bayesian Classifier Technique Coupled with a Novel Input Solution Method*, IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews, **31**(2), pp. 249-254, 2001.

[47] Awan, J.A. & Maqbool, O., *Application of Artificial Neural Networks for Monsoon Rainfall Prediction*, Sixth International Conference on Emerging Technologies, pp. 27-32, 2010.

[48] Du, J., Zhao, B. & Miao, S-H., *An Application on The Immune Evolutionary Algorithm Based on Back Propagation in The Rainfall Prediction*, International Conference on Computer Science and Electronics Engineering, pp. 313-317, 2012.

[49] Jin, L., Huang, Y. & Zhao, H., *Ensemble Prediction of Monthly Mean Rainfall with A Particle Swarm Optimization – Neural Network Model,* IEEE IRI, August 8-10, pp. 287-294, 2012.

[50] Suhartono, S., Faulina, R., Lusia, D.A., Otok, B.W., Sutikno, S. & Kuswanto, H., *Ensemble Method based on ANFIS-ARIMA for Rainfall Prediction*, International Conference on Statistics in Science, Business, and Engineering (ICSSBE)*, Langkawi, pp. 1-4, 2012.

[51] Prasad, N., Kumar, P. & Naidu, M.M., *An Approach to Prediction of Precipitation Using Gini Index in SLIQ Decision Tree*, 4th International Conference on Intelligent Systems, Modeling and Simulation, pp. 56-60, 2013.