Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 13, Number 9 (2017), pp. 6167-6174

http://www.ripublication.com

Prediction of Seasonal Rainfall Data in India using Fuzzy Stochastic Modelling

P. Arumugam and S.M. Karthik
Department of Statistics, Manonmaniam Sundaranar University
Tirunelveli – 627 012, Tamil Nadu, India.

Abstract

In this paper we present the existence and uniqueness of solutions to the fuzzy linear regression for the prediction of seasonal rainfall in India. In our daily human life, forecasting techniques are often used to predict the rainfall, population growth, economy and stock prices. In recent years many researchers used fuzzy time series to handle prediction and forecasting problems. However the computational costs associated with traditional neuro fuzzy solutions can be prohibitive. Also neuro fuzzy solutions of higher order required to get accurate results suffer from over fitting. So a simple linear regression with fuzzy variables and fuzzy coefficients is proposed in this work. Triangular membership function is used with pre-computed fuzzy intervals based on data. Experimental results indicate better performance with both training and testing data. The proposed method outperforms AR (1) models, ARIMA, SARIMA and neuro fuzzy solutions. The computational efficiency are also demonstrated through experiments.

Keywords: Seasonal rainfall, Fuzzy linear regression, Neuro Fuzzy, Stochastic Modelling, Prediction, Time Series.

1. INTRODUCTION

Time series analysis is a well developed research area with variety of mathematical techniques [1]. A univariate time series is a series of measurements taken on a dynamic system at different points in time. The problem of forecasting is to predict the values assumed by the variable at a future time instance. The fundamental assumption of time series prediction is that the past history can be used to construct a mathematical model of the system generating the time series. This model is then used to predict the future values of the time series. Time series models have been widely used in several areas like climate science, econometrics, life sciences, control theory, astronomy, business analytics etc. The variation in time series consists of three components. The longer term gradual change is called trend, the cyclical variation pattern is called seasonality and then the short term fluctuations that are not predictable. Mathematical models have been highly successful in accurately predicting seasonality and trends. Stochastic models have been used to predict the bounds for the short term fluctuations.

The climate system is a chaotic system subject to the famous butterfly effect. A flap of a butterfly wing on one side of the world can eventually cause a hurricane on the other. In technical terms, a small change in initial conditions can cause large variation in the system in the future [2]. However the basic patterns of weather are cyclical in nature influenced by the seasonal variations such as the revolution of the earth, wind patterns over the ocean etc. The Indian weather is influenced by the monsoon system. Every year, during July through September, the winds over the Bay of Bengal bring much needed rainfall to the Indian peninsula. The Northwest monsoon operate out of the Arabian Sea. The Indian agriculture, the main economic activity of the large Indian population is dependent on this rainfall [3]. Proper forecasting of Indian rainfall is essential for agricultural planning, urban water management and flood control measures. The Indian Meteorological department releases its predictions of the rainfall at different time scales including high resolution forecast for ten days. The rural economy of India that provides the livelihood of a billion people depends on this weather system. The accurate and reliable prediction of rainfall is a challenging and rewarding task.

2. RAINFALL DATA

The monthly seasonal and annual rainfall data is provided by the data.gov.in website contributed by the Ministry of Earth Sciences [4]. It contains the monthly rainfall in mm from 1901 till 2014. The data is described briefly in this section. Figure 1 shows the total annual rainfall in India from 1901 to 2014. The highest rainfall of 1463.9 mm occurred in 1917 and the lowest of 947.1 mm in 1972.

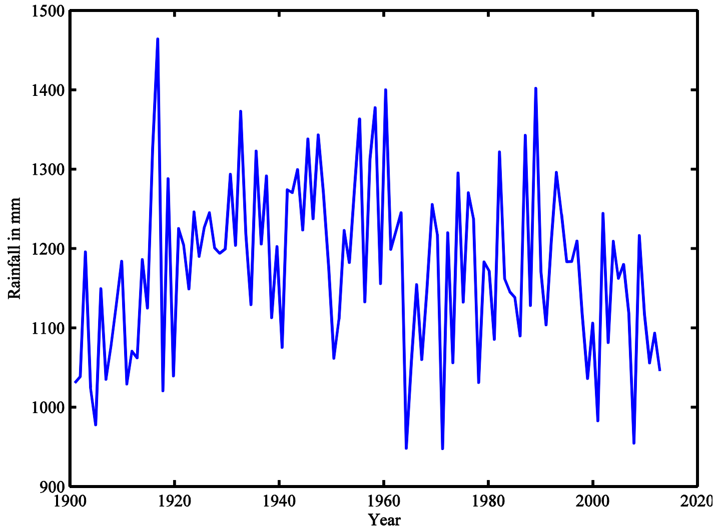

Figure 1. Average Annual Rainfall in mm from 1901 to 2014

Figure 2 shows the box plot of the monthly variation in rainfall. The months from June to October forms the monsoon period marked by both high amount and variability in rainfall. July brings the highest rainfall and is also the most variable. The other months bring less than 60 mm in rainfall and are less erratic.

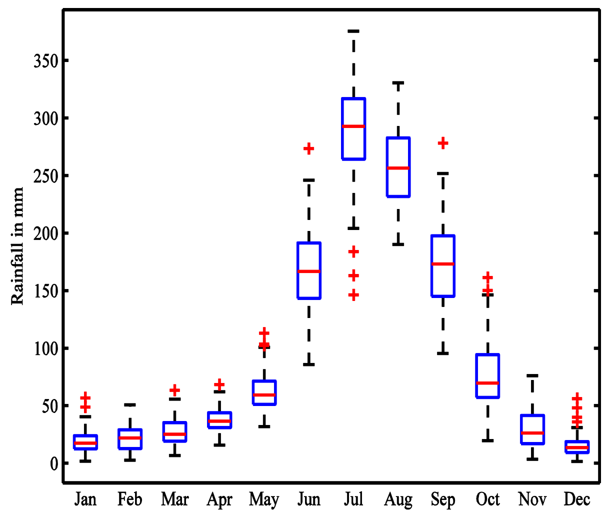

Figure 2. Box Plot of Monthly Variation in Rainfall

3. PROPOSED METHOD

3.1 FUZZY TIME SERIES

FTS methods [5] divide the universe of discourse $\mathrm{U} = \{\mathrm{u_1}, \mathrm{u_2}, \dots, \mathrm{u_b}\}$ into several fuzzy sets $\mathrm{A_i}$ defined as

$$A_i = \frac{f_{A_1}(u_1)}{u_1} + \frac{f_{A_2}(u_2)}{u_2} + \cdots + \frac{f_{A_b}(u_b)}{u_b} \qquad \text{... (3)}$$

Where $\mathrm{f_{A_i}}: \mathrm{U} \to [0,1]$ is the membership function of the fuzzy set $\mathrm{A_i}$ that maps each element to a real number in the unit interval representing its degree of belongingness in the set. A Fuzzy Time Series on real numbers $\mathrm{Y(t)}$ is defined as the collection $\mathrm{F(t)}$ of fuzzy sets $\mathrm{f_i(t)}, (\mathrm{i} = 1,2, \dots)$ that are defined using $\mathrm{Y(t)}$ as the universe of discourse.

A Fuzzy Relation between $\mathrm{F(t)}$ and $\mathrm{F(t-1)}$ is denoted by $\mathrm{R(t-1,t)}$ and written as

$$\mathrm{F(t)} = \mathrm{F(t-1)} \odot \mathrm{R(t-1,t)} \qquad \text{... (4)}$$

If such a relation exist then $\mathrm{F(t)}$ is said to be caused by $\mathrm{F(t-1)}$. The variable t denotes time. In short a fuzzy relation is expressed as $\mathrm{F(t-1)} \to \mathrm{F(t)}$. This allows the expression of rules involving linguistic quantities. This enables FTS models to capture human like intelligence. The right hand side is the fuzzy forecast and the left hand side can involve more than one fuzzy sets. If there are N fuzzy sets in the left hand side, it is referred to as a N - order relation. High order relations were introduced by Chen. A group of fuzzy relations is a fuzzy relationship group (FRG).

3.2 FUZZY LINEAR REGRESSION

Let the monthly rainfall data of consecutive years be represented by 12×1 vectors M_i and M_{i+1}. The rainfall data is expressed as fuzzy time series with three intervals. Trapezoidal membership functions are used to represent these three fuzzy classes corresponding to dry, normal and wet spells. The membership function of a fuzzy set in the range $\mathrm{A} = [\mathrm{a, b, c, d}]$ is defined as

$$\mu_A(x) = \begin{cases} 0, & (x < a) \; or \; (x > d) \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & b \le x \le c \\ \frac{d-x}{d-c}, & c \le x \le d \end{cases} \qquad \text{... (1)}$$

The parameters in A are determined from the data as the percentiles (P) of the range of values. The dry spell fuzzy class is characterized by A_{dry} = {minimum, P_{15}, P_{35}, P_{50}}. The normal class is centered around P_{50} i.e., A_{normal} = {P_{30}, P_{45}, P_{55}, P_{70}} and the wet class A_{wet} = {P_{60}, P_{80}, P_{90}, maximum}. The use of percentiles to define the fuzzy classes simplifies the interval calculations. The fuzzy regression equation applied to the problem of predicting M_{i+1} from M_i is

$$M_{i+1} = \sum_{k=1}^{3} C_k M_i \mu_k \quad \text{.... (2)}$$

where C_k is the 12×12 coefficient matrix, one for each of the three classes. The regression coefficients are solved from training data and tested on fresh testing data.

4. EXPERIMENTAL RESULTS AND ANALYSIS

The proposed method was tested and compared against well known time series forecasting methods. Figure 3 shows the prediction compared against the actual values. Table 1 shows a comparative analysis of the proposed method against Auto regression based methods [6], Artificial Neural networks [7] and Hidden Markov models [8]. The proposed method gives better accuracy on the testing set. It avoids the problem of overfitting by using the simpler linear model when compared against nonlinear neural networks based solutions. Let $R_{i,j}, i = 1,2,..,n; j = 1,2,...,12$ be the actual rainfall recorded in the test set for n years. Let $\hat{R}_{i,j}$ be the prediction for the same period from the SANN. Then the following performance measures are used to assess the accuracy of the prediction.

Mean Squared Error (MSE) is defined as

$$mse = \frac{1}{12n} \sum_{i=1}^{n} \sum_{j=1}^{12} \left(R_{i,j} - \hat{R}_{i,j}\right)^2 \quad \text{... (3)}$$

Root Mean Squared Error (RMSE) is defined as

$$rmse = \sqrt{mse} \quad \text{... (4)}$$

Mean Absolute Deviation (MAD) is defined as

$$mad = \frac{1}{12n} \sum_{i=1}^{n} \sum_{j=1}^{12} \left|R_{i,j} - \hat{R}_{i,j}\right| \quad \text{... (5)}$$

It is the average of all absolute deviations of the predicted from the actual values.
Mean Absolute Prediction Error (MAPE) [9] is also known as Mean Absolute Percentage Deviation (MAPD). It is defined as

$$mape = \frac{1}{12n} \sum_{i=1}^{n} \sum_{j=1}^{12} \left| \frac{R_{i,j} - \hat{R}_{i,j}}{R_{i,j}} \right| \qquad .... (6)$$

MAPE can be used in our current application since there are no zero values in the predicted variable and does not cause division by zero error. When MAPE is multiplied by 100, it is expressed as a percentage. MSE, RMSE, MAD and MAPE must be low for a good prediction. RMSE, MAD and MAPE are expressed in the same units as the predicted variable i.e., in mm in this case.

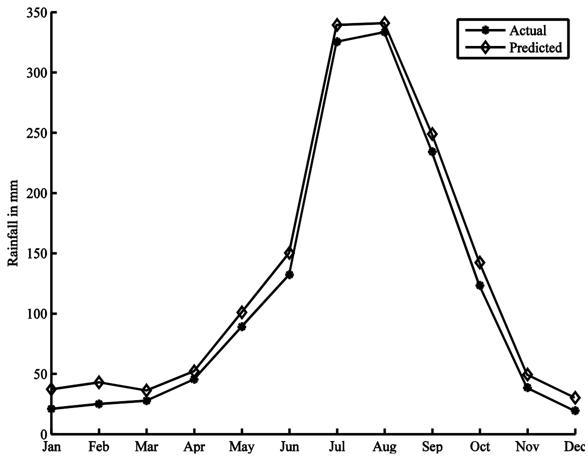

Figure 3. Actual vs. Predicted values for a representative year.

Table 1. Comparative Analysis

Method	MSE	RMSE	MAD	MAPE
Fuzzy Linear Regression	33.19	5.93	5.06	0.1028
HMM	38.69	6.22	5.34	0.1131
ARMA	39.38	6.28	5.37	0.1134
ARIMA	40.36	6.35	5.55	0.1194
ANN	43.42	6.59	5.82	0.1200

The proposed method exhibits better performance than comparable methods in literature.

5. CONCLUSION

In this work presented a rainfall prediction model based on fuzzy linear regression. The rainfall data is classified into three fuzzy classes namely dry, normal and wet.

Trapezoidal functions with percentiles as the parameters were used to define the membership function. The simple linear relations among the variables in consecutive years brought out the different relations in each of the three different spells. It avoided overfitting and achieved better results for test data. The approach could be extended to prediction of other time series in the financial, econometric fields. Higher order methods could be explored to achieve better stability to forecast the values over a longer time frame.

ACKNOWLEDGEMENT

The Second author thanks the University Grants Commission, New Delhi for awarding fellowship under the scheme of UGC – Basic Science Research fellowship to carry out this work.

REFERENCES

[1] Arumugam, P., Senthamarai Kannan, K., Balamurugan. "Evaluation of the Fuzzy queries with Linguistic Hedges on Statistical Attributes." *Journal of Applied Mathematical Analysis and Applications*, Vol.4, No.1-2, (2008): 35-43.

[2] Arumugam, P. "Forecasting enrollments using Fuzzy stochastic model. " *Recent Trends in Statistical Research*, Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli, (2010): 21-28.

[3] Arumugam, P., Senthamarai Kannan, K. "Computational Algorithm of Fuzzy Stochastic Model For Forecasting. " *Journal of Algorithms and Computational Technology*, School of Computing and Mathematical Sciences, University of Greenwich, London, UK, Vol. 6, (3), (2012): 375-384.

[4] Arumugam, P. " Computational Algorithm for Clinical trial data based on Cox model. " *Recent trends in statistics and computer applications* (2011): 157-162.

[5] Arumugam, P., Karthik, S. M. "Stochastic Modelling in yearly Rainfall at Tirunelveli District, Tamil Nadu, India. " Published By Elsevier Materials today: Proceedings, 2214-7853(2016), *International Conference on Processing of Materials, Minerals and Energy,* Ongole, Andhra Pradesh, India, July 29 – 30, 2016: 1-7

[6] Arumugam, P., Karthik, S.M. "Rainfall prediction by seasonal neural networks enhanced with first order markov models. "*Bulletin of Mathematics and Statistics Research*, Vol.4. Issue.4 (2016): 69-76.

[7] Box, George EP, et al. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[8] Hilborn, Robert C. "Sea gulls, butterflies, and grasshoppers: A brief history of the butterfly effect in nonlinear dynamics." *American Journal of Physics* 72.4 (2004): 425-427.

[9] Hamilton, James Douglas. *Time series analysis*. Vol. 2. Princeton: Princeton university press, 1994.

[10] Khanal, Aditya R., and Ashok K. Mishra. "Enhancing food security: Food crop portfolio choice in response to climatic risk in India." *Global Food Security* 12 (2017): 22-30.

[11] Karthik. S.M., and Arumugam. P. "Stochastic modelling based monthly rainfall prediction using seasonal artificial neural networks. " *ICTACT Journal of Soft Computing*, Vol.2. Issue.2 (2017).

[12] MacDonald, Iain L., and Walter Zucchini. *Hidden Markov and other models for discrete-valued time series*. Vol. 110. CRC Press, 1997.

[13] Parthasarathy, B., K. Rupa Kumar, and A. A. Munot. "Forecast of rainy season foodgrain production based on monsoon rainfall." *Indian Journal of Agricultural Sciences* 62 (1992): 1-8.

[14] Song, Qiang, and Brad S. Chissom. "Fuzzy time series and its models." *Fuzzy sets and systems* 54.3 (1993): 269-277.

[15] Willmott, Cort J., and Kenji Matsuura. "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance." *Climate research* 30.1 (2005): 79-82.

[16] Zhang, G. Peter. "Time series forecasting using a hybrid ARIMA and neural network model." *Neurocomputing*: 50 (2003): 159-175.