International Journal of Computer Applications (0975 – 8887)
Volume 85 – No 15, January 2014

A Fast Learning Algorithm for Rainfall Prediction

Seema Mahajan
L. J. Institute of Engineering & Technology, S.G.Road
Ahmedabad-382210, Gujarat, India

Himanshu Mazumdar
Professor E.C and Head R&D Center
D.D.University, Nadiad-387001, Gujarat, India

ABSTRACT
A PC based application is developed using 51 years of Indian rainfall data for long range prediction of average rain fall. This learning algorithm iteratively estimates 96 coefficients of a 5th order polynomial in few minutes. Proposed prediction model is based on modelling of time series rainfall data using 5th order non-linear predicting code. Steepest descent algorithm is used for extraction of appropriate coefficient of rainfall time series data. These coefficients are reinforced in model learning process. The rainfall data of 1960 to 2010 is used for the development of the model. Model has been tested on rainfall data for different training sets. The proposed model is capable of forecasting yearly rainfall 1 year in advance. Rainfall estimation accuracy is above 85%.

Keywords
Rainfall forecasting, non linear predicting code

1. INTRODUCTION
Rainfall data is needed for many systems like hydrological systems, an irrigation system and mainly agriculture system for crop cultivations. The variation of the all India rainfall is not consistent with all the Indian regions. Locations, distribution of land and water, distance from sea, Himalayan Mountain are the causative factor of the spatial variation in the climate of India. In past years many researchers have introduced several new concepts and ideas in rainfall forecasting. Abraham and Philip (2004) attempted 5 soft computing models based on 40 years rainfall of kerla[1]. Their simulation result for monthly prediction shows that a soft computing technique gives RMSE as 0.090, 0.094, 0.093, 0.092 and 0.093%. Munot and Krishna Kumar developed prediction model using temperature, zonal and meridional data[2]. They used stepwise multiple regressions technique. The model shows 87% of the variance and CC (correlation coefficient) is 0.934 with observed and predicted rainfall. Iyengar and Raghukanth developed rainfall-forecasting model using intrinsic mode function. They used Artificial Neural Network and regression for model development [3]. The rainfall prediction efficiency is 75% to 80% of the inter-annual variability. Karmakar, Kowar and Guhathakurta attempts to predict long range monsoon rainfall over the subdivision of Chhattisgarth using feed forward back propagation neural net[7]. Their model results 0.8 correlations between actual and predicted rainfall. Surajit and Goutami Chattopadhyay (2008) identify a non linear methodology to forecast the time series of average summer monsoon rainfall over India[8]. The RMSE for four NN based techniques is 0.15, 0.26, 0.42 and 0.47. Krisha Kumar discussed about the predictors and their relationship of these predictors and methods of long range forecasting [6].The relationship of the predictors with rainfall is not coherent. Many researchers worked on many different methods such as regression techniques, artificial neural network, discrete wavelet transform, self organizing map, fuzzy rule systems, support vector machine, Gray Markov model, learning vector quantization etc, but still rainfall prediction remains a challenge for researchers. Rainfall is a complex phenomenon and is a function of various atmospheric parameters. We assume this function is continues and its higher order derivatives exists in all points of the sample and prediction period. This will in turn assumes that a corresponding Taylor's series exists for bounded period of the sample space. In this paper we use non linear predicting code for rainfall prediction. The whole India rainfall is obtained as the area-weighted mean of sub divisional rainfall and is available from the website Indian Institute of Tropical Meteorology, Pune. The rainfall data from 1960 to 2010 is used for model implementation.

2. PAGE SIZE
Researchers have used various approaches for forecasting such as statistical method, dynamical method (numerical weather prediction) and dynamical cum statistical method.

- Atmosphere structure is very complex. 51 samples of yearly averaged data of 51 years are not sufficient to build such complex model. To increase the number of samples within the function and sample space we use interpolation to get desired size of data set. Figure 1 shows the yearly average rainfall data for the period 1960 to 2010 of India. This data is collected from the website of IITM, Pune,

- 2001 data points containing the original set of 51 are generated using interpolation within the sample space using dot Net frame work. Interpolation is achieved using C# function DrawCurve() with tension parameter set to 0.5. DrawCurve() method draws cardinal spline through given set of points. Interpolated rainfall data points are at the interval of $((51\times 365)/2001) \approx 9$ days. The model is implemented using interactive user friendly application written in C# as shown in figure 2.

- The proposed predictor function is defined in equation 1. The expression contains M previous rainfall samples P(t-m) to predict the next sample P(t+1). P(t+1) represents rainfall prediction for succeeding one year. Each of M samples P(t-m) contains Nth order polynomial with N×M coefficients knm.

$$P(t+1)_{pred} = \sum_{n=0}^{N} \sum_{m=0}^{M} P^{n}(t-m)k_{nm} \;\ldots\ldots 1$$

The values of N and M are experimentally optimized.

The coefficient array knm is computed iteratively using steepest descent algorithm to minimize the prediction error.

- The iterative method is developed by defining error cost function and knm is estimated by minimizing the error cost function. The error cost function is defined as given in eq. (2)

$$\mathbf{E} = \frac{\mathbf{1}}{\mathbf{2}}(\mathbf{P_{pred}}(\mathbf{t}+\mathbf{1}) - \mathbf{P}(\mathbf{t}+\mathbf{1}))^{\mathbf{2}} \ldots\ldots\ldots\ldots\ldots .2$$

- The error sensitivity of knm is shown in equation 3. Equation 4 shows the reinforcement of knm with learning rate ή with value ranging from 0.001 to 0.01 where (knm)new the corrected value of the coefficient is (knm)old

$$\frac{\partial \mathrm{E}}{\partial \mathrm{k_n}} = \mathrm{S_{nm}} \ldots\ldots\ldots\ldots\ldots\ldots .2.1$$

$$\boldsymbol{S_{nm}} = [\boldsymbol{P_{pred}}(\boldsymbol{t}+\mathbf{1})\boldsymbol{P}(\boldsymbol{t}+\mathbf{1})]\frac{\boldsymbol{\partial P_{pred}}(\boldsymbol{t}+\mathbf{1})}{\boldsymbol{\partial P}(\boldsymbol{t}+\mathbf{1})} ..\mathbf{3}$$

$$(\boldsymbol{k_{nm}})_{\boldsymbol{new}} = (\boldsymbol{k_{nm}})_{\boldsymbol{old}} - \eta\, \boldsymbol{S_{nm}} \ldots\ldots\ldots\ldots\ldots\ldots .\mathbf{4}$$

To make the algorithm robust and noise tolerant we have added a small random noise to the final predicted value during training.

3. 3. RESULT AND DISCUSSION

The above algorithm is employed to predict the rainfall. The choice of the order of the polynomial (N) and the number of samples (M) in eq.-1 is made by carrying out exhaustive experiments for different values of N and M. The experiments are carried out by varying sample size from 10 to 30 and also varying the order of polynomial N from 5 to 17. Training period for all these experiments is selected as 1960 to 2002. All the samples are taken at an interval of 365 days. Fig. 3 shows rainfall prediction error for year 2003 for all above experiments. Sample size 16 shows minimum rainfall prediction error (below 10 %) for all the order of polynomial. So sample size 16 is optimized for the proposed study. The year to year variation in rainfall is traced by M = 16 samples i.e. 16 years. Previous 16 points are used in the model to predict next 17th point of rainfall data. The rainfall prediction period is one year ahead of time. Fig.4 shows rainfall prediction error for years 1988 to 2011 with sample size 16 with order of polynomial N =5, 9 and 11. An average prediction error for year 1988 to 2011 for the order of polynomial N = 5, 9 and 11 is 11.8%, 12.9% and 13.4% respectively. So the order of polynomial, N = 5 is chosen for further study. A pruning algorithm is used to test the sensitivity of all 96 coefficients knm on predicted value.

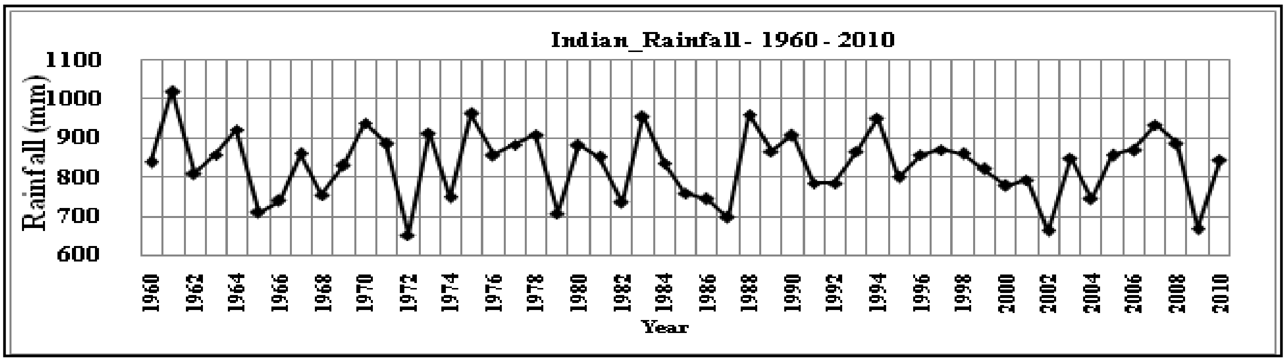


Fig 1: Plot of yearly average Indian rainfall data from 1960 to 2010. The data is taken from the website of IITM, Pune.

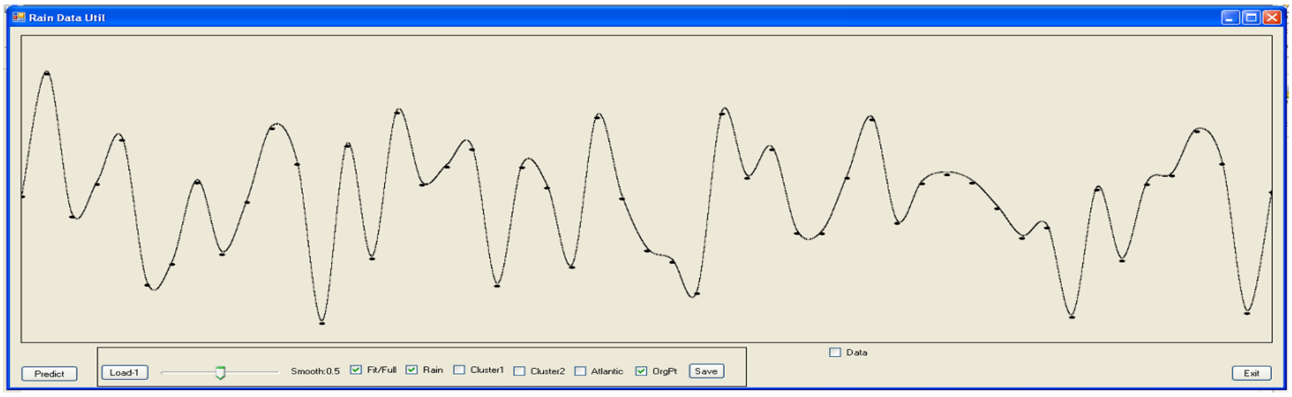


Fig 2: A screenshot of the application shows interpolated rain fall plot from 1960 to 2010 overlapping with 51 yearly Indian rainfall average value points

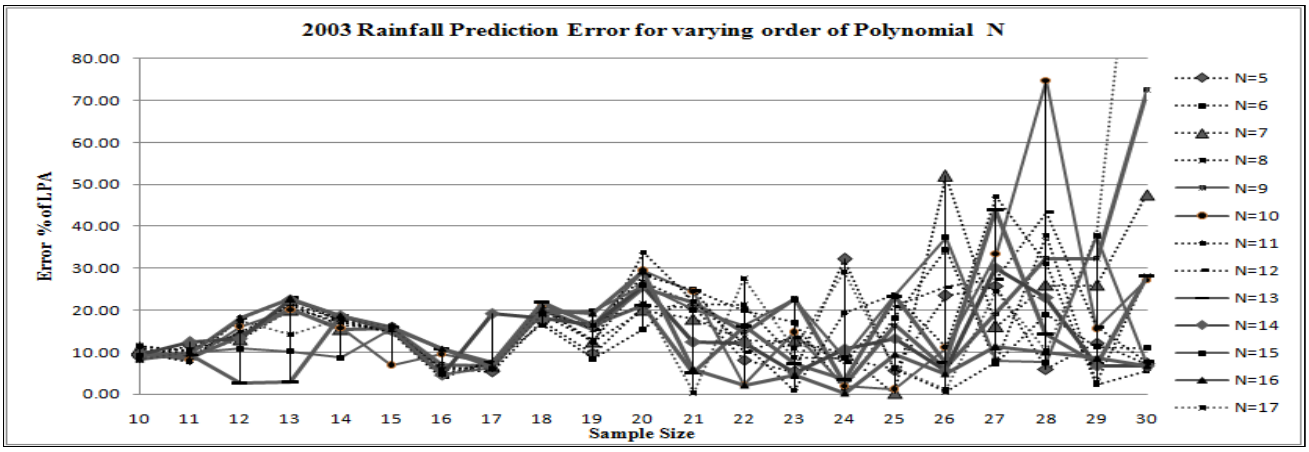
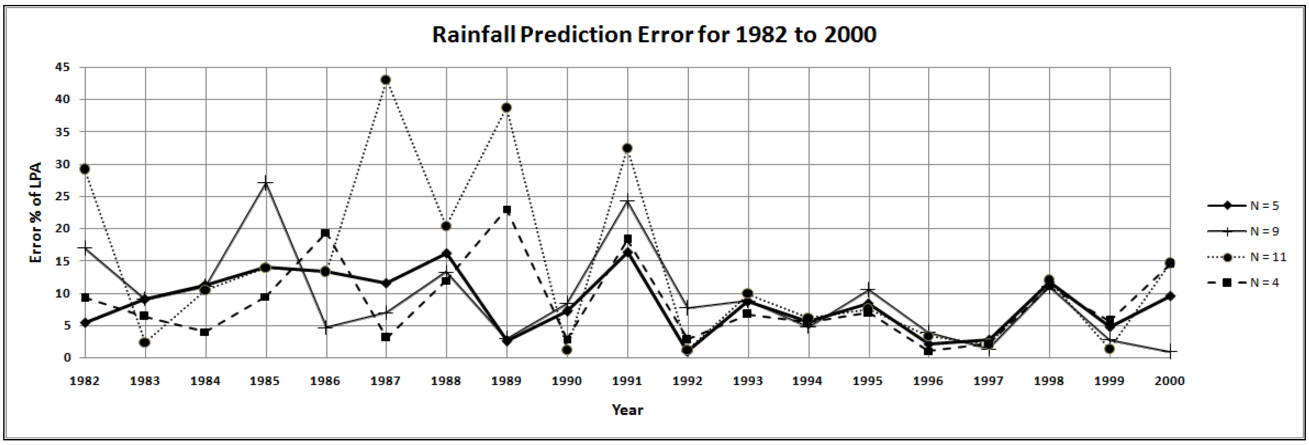


Fig 3: Rainfall prediction error for year 2003 for varying order of polynomial N from 5 to 17



4: Rainfall prediction error for 1988 to 2011for order of polynomial N = 5, 9 and 11, sample size 16

The application is implemented with Intel i3 CPU, @ 2.40 GHZ, with installed memory 3 GB laptop to evaluate the prediction algorithm. The .Net C# interactive application takes few minutes for each experiment to converge cost function error to desired minima. Fig. 5 shows the Inter annual variability of India rainfall, expressed as a percentage departure from Long Period Average (LPA) for the period of 1960-2010. Long period average is average of rainfall for entire data set. Rainfall for the season is counted with respective long period average. LPA for the period of 1960 – 2010 is 833.3 mm. Figure 5 shows Indian rainfall variability with respect to LPA, which lies between ± 25%. Table 1 shows prediction error of the proposed model from 1988 to 2011. In all the cases error is calculated as % of LPA. In the year 2002, rainfall was below normal during first half of the season so model is more error in this case16.All other models are also having more error for 2002 compare to other years. With sufficient training period, prediction accuracy is above 90%. Table-1 shows that error lies between ± 15%. Table 2 shows the predicted values for 5 different cases from year 2008 to 2011.Table 3 shows prediction error of the models proposed by other researchers.

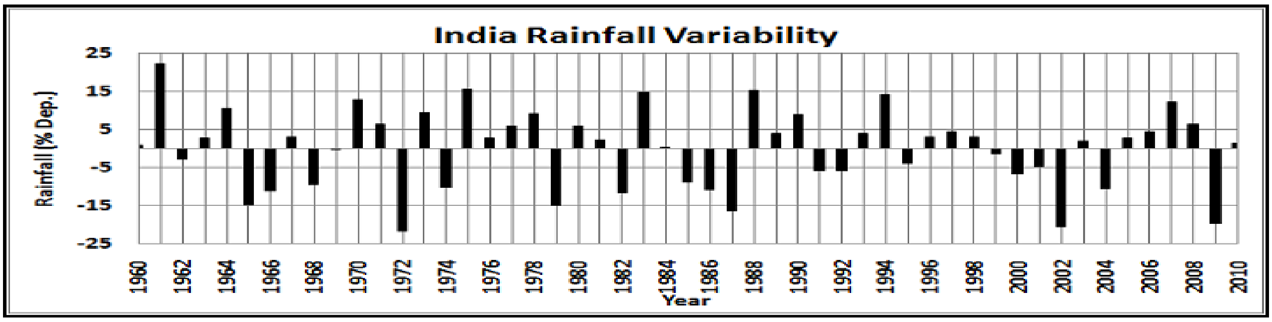


Fig 5: Variability of average Indian rainfall expressed as % departure from LPA

4. FUTURE WORK

Research is in progress to increase the prediction range by progressively using inputs from predicted values. The main difficulty using such method is that the system becomes unstable and output soon goes out of bound because of cumulating error in feedback loop.

5. CONCLUSION

The proposed method of Nth order polynomial can be useful for rainfall forecasting using previous years average rainfall data. The trend of rainfall prediction is up to 99% accuracy. Quantitative rainfall prediction error of the model is below 10%. Rainfall prediction achieved is for one year ahead of time which is most desired period for crop planning and other water management strategies.

6. ACKNOWLEDGMENTS

This work is carried out under a PLANEX project of Physical Research Laboratory, Ahmedabad with support of R&D Centre, D.D. University, Nadiad and L.J. Institute of Engineering & Technology, Ahmedabad. The authors would like to thank to Dr. Dilip Ahalpara for his valuable advice.

7. REFERENCES

[1] Ajith Ninan sajeeth Pilip, Soft Computing Models for Weather Forecasting (2004), Journal of Applied Science and Computations.

[2] A.A.Munot and K. Krishna Kumar, Long Range Prediction of Indian Summer Monsoon Rainfall – Indian Institute of Tropical Meteorology, Pune 411008, India (Feb-2007).

[3] Iyengar, R. N. and Raghukanth, S. T. G., Intrinsic mode functions and a strategy for forecasting Indian monsoon rainfall. Meteorol. Atmos. Phys., 2004.

[4] M. Rajeevan, Prediction of Indian summer monsoon: Status, problems and prospects Special Section: Climate, Monsoon and India's Water Current Science, VOL. 81, NO. 11, 10 December 2001 1451.

[5] M. Rajeevan, D.S. Pai, R.AnilKumar, New Statistical Models for Long RangeForecasting of SouthWest Monsoon Rainfall over India, NCC-Research Report, Indian Meteorological Department –Pune-411005, General of Meteorology (Research).

[6] K. Krishna Kumar , M. K. Soman and K. Rupa Kumar, Seasonal Forecasting of Indian Summer Monsoon Rainfall: A Review, IITM, Pune, India, Weather, 50, 12, pp.449-467May 26, 2010.

[7] S. Karmakar, M. K. Kowar, P. Guhathakurta, Long-Range Monsoon Rainfall Pattern Recognition and Prediction for the Subdivision 'EPMB' Chhattisgarh Using Deterministic and Probabilistic Neural Network, 7th International conference on Advance in Pattern Recognition.

[8] Surajit Chattopadhyay, Goutami Chattopadhyay, Comparative study among different neural net learning algorithms applied to rainfall time series, Meteorological Applications Volume 15, Issue 2, pages 273–280, June 2008.

[9] Sulochana Gadgil*, M. Rajeevan and Ravi Nanjundiah, Monsoon prediction – Why yet another failure?, CURRENT SCIENCE, VOL. 88, NO. 9, 10 MAY 2005.

[10] De Michele, C. and Bernardara, P. 2005, Spectral analysis and modeling of space-time. Atmospheric Research, pp. 124– 136.

[11] http://www.tropmet.res.in/ - Indian Institute of Tropical Meteorology, Pune

[12] Geoffrey E. Hinton Connectionist Learning Procedures, Computer Science Department, University of Toronto, 10 King's College Road, Toronto, Ontario, Canada.

[13] N. Krishnakutty, Long-range monsoon rainfall prediction of 2005 for the districts and sub-division Kerala with artificial neural network, Scientific Correspondence.

[14] Sarah Afshin, Hedayat Fahmi, Amin Alizadeh, Hussein Sedghi and Fereidoon Kaveh, Long term rainfall forecasting by integrated artificial neural network-fuzzy logic-wavelet model in Karoon basin, Scientific Research and Essays Vol. 6(6)

[15] Lorna R. Nayagam, Rajesh Janardanan and H. S. Ram Mohan, An empirical model for the seasonal prediction of southwest monsoon rainfall over Kerala, a meteorological subdivision of India, International Journal of Climatology

[16] Sulochana Gadgil, J Srinivasan, Ravi S Nanjundiah, On Forecasting the Indian Summer Monsoon: the Intringusing Season of 2002, Current Science