



Global Prediction algorithms and predictability of anomalous points in a time series

K C Tripathi, Rashi Agarwal, P N Hrisheeksha

Department of Computer Science and Engineering,
Inderprastha Engineering College, Ghaziabad, India,
kctripathi@rediffmail.com

Department of IT, Maharaja Agrasen College of Management Studies, New Delhi, India
rashii27@gmail.com

Director, Inderprastha Engineering College, Ghaziabad, India
director@ipec.org.in

ABSTRACT

The Indian Ocean Dipole (IOD) event of 1997 had significant impacts on regional climate variability as it produced large Sea Surface Temperature (SST) anomalies in the Indian Ocean (IO). This dipole mode accounted for about 12% of the sea surface temperature variability in the Indian Ocean—and, in its active years, also causes severe rainfall in eastern Africa and droughts in Indonesia. There have been speculations that the El-nino and Intra-Seasonal Oscillations (ISOs) act as stochastic forcing to reinvigorate the natural damped mode and hence contribute to the development of the IOD. In the present paper we have statistically investigated whether the formation of the dipole was an anomalous phenomenon with respect to the time series generated by the deterministic laws of the system governing the IO SST or was this event a consequence of the dynamics of the system itself. For this, we have used the a global prediction algorithm i.e. linear regression. Prediction errors of global prediction algorithms such as the regression and Artificial Neural Network (ANN) models at various points in time contain important information regarding the statistical nature of the data. On the basis of the error analysis it is found that the occurrence of the IOD is a consequence of the state of the SST system as a whole together with the evolution laws. The role of ISO in providing external forcing to the dipole is investigated by analyzing the prediction errors of the global prediction algorithms for processes at intra-seasonal time scales. It is concluded that the Intra-Seasonal Oscillations may provide the stochastic forcing to the IOD.

Indexing terms/Keywords

Linear Regression, Sea Surface Temperature, Indian Ocean Dipole, Predictability

Academic Discipline And Sub-Disciplines

Predictability, Indian Ocean dipole, algorithms, dynamic al systems

SUBJECT CLASSIFICATION

Computer Science and Mathematics

TYPE (METHOD/APPROACH)

Experimental and analytical

Council for Innovative Research

Peer Review Research Publishing System

Journal: INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY

Vol 10, No 9.

editor@cirworld.com

www.cirworld.com. member.cirworld.com



1 INTRODUCTION

Study of sea surface temperature (SST) variations is important for accurate weather prediction as it is the main variable of interest for atmospheric forcing. A successful prediction of the seasonal mean rainfall and circulation is based on the premise that the monsoon is dynamically a stable system and its seasonal mean rainfall and circulation and their interannual variability are largely governed by slowly varying boundary conditions such as SST, snow cover, soil moisture, etc. [1], [2]. Studies have shown that it affects the local as well as global monsoon [3] including Indian monsoon rainfall [4]. Prediction of SST over the Indian Ocean (IO) region contributes to the forecast of the monsoon in the Indian subcontinent [5] and the El-Nino and la-Nina onset [6]. Since comparison of various statistical methods has always been a curiosity among the researchers [7], therefore apart from dynamical modeling, ANN and linear regression [8] have been recently used for SST forecast in the IO.

Despite an investigative history of more than a century, the role of the IO in regional monsoon variability is yet to be fully understood and remains an active area of diagnostic and modeling research. Hence, state of the art forecasting systems have attained only a limited success. Anomalous events in the IO during 1994 and 1997 have generated much renewed interest in all aspects of the coupled climate system. Several studies have been done to investigate the origin of the Indian Ocean Dipole (IOD) and its impact on the Asian monsoon [9]. The role of Intraseasonal oscillations and El-nino has been recently investigated and it has been found that the ISO may act as stochastic forcing to reinvigorate the natural damped mode and that the El Niño influences the IODM mainly through the change of intensity of the Indian monsoon [10].

IOD being an anomalous phenomenon because of the large deviation from the trend, may be a state of the dynamical system governing the SST time series of the IO or it may be a consequence of some local process which need not be governed by the deterministic system in question. Hence, it is not a priori obvious that any model will or will not be able to simulate this event.

Is IOD a part of the dynamical system governing the SST time series of the IO? Or is it a consequence of some local process which need not be governed by the deterministic system in question? What is the role of Intra-Seasonal Oscillations (ISOs) (frequencies from 15 days to 60 days)? Do the global prediction algorithms give a conclusive answer to these questions? These are the basic questions that are explored in this study.

These questions can be addressed by analyzing the errors in prediction of SST time series using a global prediction algorithm. Global Statistical models used in meteorology and oceanography are the regression model [11], principal component analysis [12] and the canonical correlation analysis [13]. Recent studies have shown that ANN has better prediction and classification capabilities than statistical models [14].

It is well settled that ANN models can better tolerate chaotic components than the traditional models such as the regression. It is thus pertinent to explore how various global prediction algorithms are suited for forecasting anomalous phenomena. It is not possible to answer the above question in all entirety without actually exploring the models. However, as a first step we have investigated the case using the linear regression model. Regarding prediction by the ANN model, if the phenomenon under question is a random phenomenon and bears no connection with the underlying dynamics of the system then there will be a sudden rise in the prediction error at the point of occurrence of the rare phenomenon. This is because global prediction algorithm such as the ANN essentially model the underlying deterministic function and ignores the noise factors (i.e. the impact of local processes on the state of the dynamical system in question). Any anomalous event is a major deviation from the deterministic function and as such the prediction errors at these points will increase sharply. The linear regression model under discussion is not an ideal model for modeling chaotic time series. Hence the prediction errors may or may not rise abruptly during such anomalous occurrence. This point needs investigation.

On the other hand, if the phenomenon in question is governed by the deterministic dynamical laws of the system, the prediction error at this instance will not differ much with the prediction error at other temporal instances. The reason being the anomalous point is generated from the dynamical law governing the system and hence will be identified by the global algorithm. Again, the question that remains to be explored is whether predictions by ANN and linear regression essentially give the same results? If the results by most of the global prediction algorithms are similar, it may safely be concluded that the anomalous point under question is a point on the trajectory of the dynamical system.

In the present study linear regression model is used to analyze the predictability of the SST over the two regions of the IOD: positive pole [(0o-5o) N and (60o - 65o) E] and negative pole [(5o -10o) S and (95o - 100o) E] formed in the year 1997 [15]. The prediction errors are analysed at different time scales of prediction to study the statistical nature of the event.

2 METHODOLOGY

2.1 Data

The extended reconstructed sea surface temperature called ERSST version 2 (ERSST.v2) (Smith and Reynolds, 2003; Smith and Reynolds, 2004) from January 1871 to May 2004 for the IO region (in the neighborhood of IOD) has been used for the present analysis. The ERSST was constructed using the most recently available International Comprehensive Ocean -Atmosphere Data Set (ICOADS) SST data and improved statistical methods that allow stable reconstruction using sparse data. This monthly analysis begins from January 1854, but because of sparse data the analyzed signal is heavily damped before 1880. Afterwards the strength of the signal is more consistent over time. The ERSST analysis is updated continuously as new data become available. ERSST.v2 is an improved extended reconstruction. In the reconstruction the

high-frequency SST anomalies are reconstructed by fitting to a set of spatial modes. Compared to the earlier reconstruction, version 1 (v1), the improved reconstruction better resolves variations in weak-variance regions. It also uses sea-ice concentrations to improve the high-latitude SST analysis, a modified historical bias correction for the 1939 -1941 period, and it includes an improved error estimate.

2.2 Preprocessing (Fourier analysis for the removal of dominant modes)

Figure 1 shows the difference in anomalies of the two regions. The ERSST time series for the plus and minus regions have been analyzed using discrete Fourier analysis for identifying the dominant cycles. The power spectrum of the ERSST series after discrete Fourier transform (dft) analysis is shown in Figure 2(a). The dft for the series is same for both the Plus and Minus regions. It can be seen that there are three dominant harmonics viz. the harmonic numbers 133, 266 and 399. These correspond to the period of 1596/133, 1596/266 and 1596/399 months, which approximately corresponds to 12, 6 and 4 months respectively.

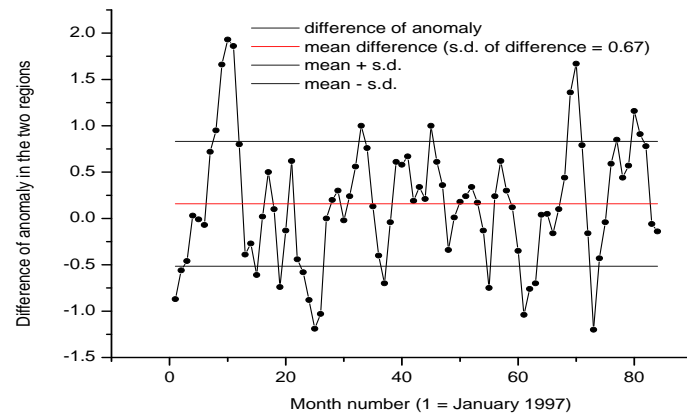


Figure 1: Difference in anomalies of the plus and minus

Importance of the ISO is now widely recognised. Both during summer and winter seasons, convection associated with ISO (ISOs), typically with a time scale of 30-50 days, originate over the equatorial IO [16]. Although the ISOs are inherent to the atmosphere, both observational and coupled modeling studies demonstrate that IO SST plays a crucial role in the organization, intensification, and propagation of the convection and circulation associated with the ISOs [17]. In the context of the IOD, it is established that there is high negative correlation between the IOD and ISO activity over the southeastern IO [18].

From the frequency domain representation of the time series we can look into the strength of the ISOs. The relative strength of the ISO will be visible if we remove the three dominant cycles, which are essentially the variations at the seasonal times-cales. This is shown in Figure 2(b). It can be seen that after the removal of the three very strong components the 35th harmonic (corresponding to a cycle of 45 days or Madden-Julian Oscillation) is the most dominating. Other dominating harmonics also belong to the intraseasonal timescale and are covered by the circle in the figure. Thus, any simulation of the SST variations without the annual, semi-annual and quad-annual cycles is, to a significant acceptance, the simulations of the ISOs.

To be able to look into the aspect of the IOD being sustained by the ISOs, we need to observe the prediction errors of global prediction algorithms (in this case linear regression) after removing the three dominant cycles so that the residues are essentially the observed and simulated ISOs. In the present study we have not looked into the aspects of sustainability of the IOD by the ISOs. The errors are analyzed for the following observations:

- (1) If the prediction errors rise abruptly during the formation of the dipole, then dipole formation is not a consequence of the underlying deterministic mechanism which governs the SST time series in the respective regions. The 'abrupt rise' is quantified as 'the error outside the mean+2*sd range' where sd is the standard deviation of the prediction errors while predicting the anomalies with the dominant cycles present.
- (2) If the prediction errors do not rise abruptly during the formation of the dipole but does so during the step (ii) of the experiments (i.e. the error goes beyond $2 * sd$) then formation of the dipole is related to the processes going on at longer time scales and ISO has no important role in sustaining the state.

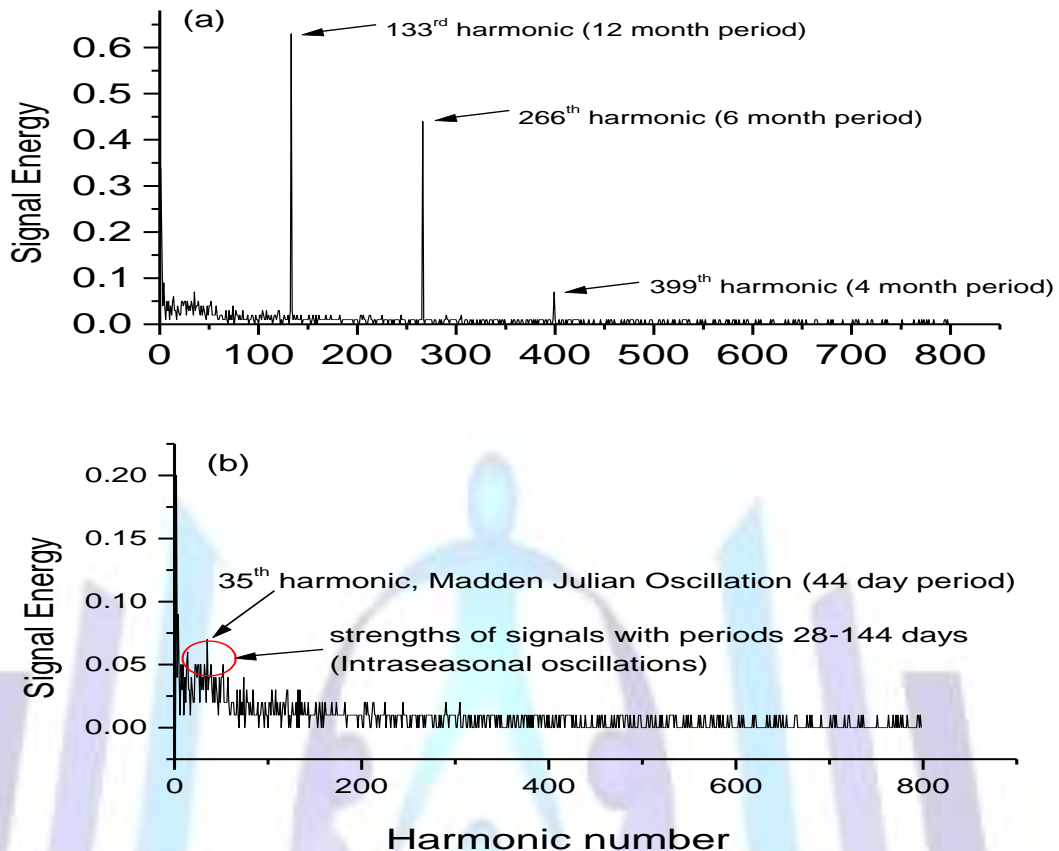


Figure 2 (a): Power spectra of the ERSST data showing all the cycles in the frequency domain representation of the time series (b) Same as in (a) except that the three strong harmonics have not been shown

2.3 Input Data (Predictors)

For deciding the predictors, the lag correlations were calculated for lags starting from 12 months to 24 months. The best correlated lag series having a correlation coefficient of 0.5 or more were taken as predictors. The series' with 12, 13 and 14 months lag was found to give the best correlation coefficient for the analysis of the Plus region. For the analysis of the Minus region, the series with 12, 14 and 21 months lag had the best correlation coefficient. Hence, we used the following as the inputs to the regression model: Plus region - three inputs corresponding to the series' with 12, 13 and 14 months' lag and Minus region - three inputs corresponding to the series' with 12, 14, 21 months' lag.

2.4 Partitioning

We have partitioned the data in two sets viz. training and test. The entire data set of 133 years (1596 months) is partitioned with the last 7 years (1997 to 2003) forming the 'test set' and the remaining forming the 'training set'.

3 RESULTS AND DISCUSSION

The training was done with 1294 points for the plus region and 1287 points for the minus region. The multiple regression line fitted the training data for the plus region with a correlation coefficient of 0.81 and a root mean square error (rmse) of 0.36. The multiple regression line fitted the training data for the minus region with a correlation coefficient of 0.61 and a root mean square error (rmse) of 0.59. This is within acceptable limits.

The performances of the multiple regression models for the test cases are shown in Figure 3 (a) for the plus region and in Figure 3(b) for the minus region. The RMS error for the plus region is 0.42 and the correlation coefficient was 0.76. The standard deviation of the observed data for this region was 0.58. Hence the prediction is better than the mean prediction. The correlation coefficient between the observed and predicted anomaly for the test case of minus region is 0.56. The RMS error (0.62) is smaller than the standard deviation of the observed data (0.72) showing that the model prediction is better than the mean prediction. Further, since the training and test errors are of the same order it can be concluded that the sampling was proper.

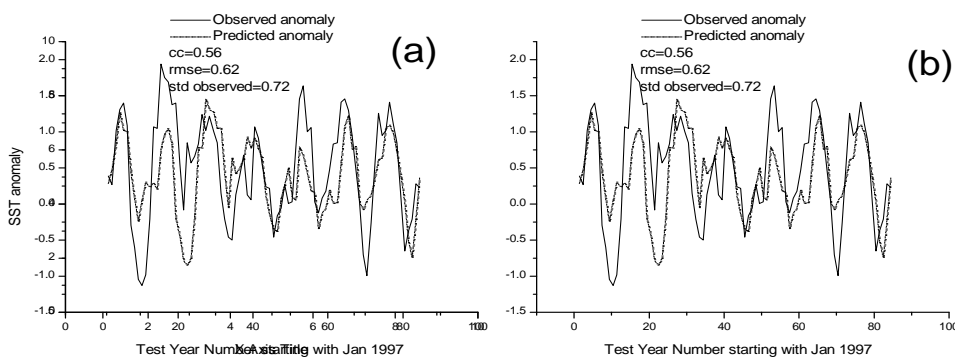


Figure 3 Comparison of observed and simulated anomalies (a): Test case for plus region and (b): Test case for minus region

We see that for both the test cases we get a significantly high correlation coefficient and small RMS errors (comparable to the corresponding training cases and smaller than the corresponding standard deviations). Thus the model gives acceptable performance on the test data sets.

Figure 4 shows the prediction errors for the test cases at various time steps for (a) the Plus region and (b) Minus region. It can be seen that there is an abrupt rise in the prediction errors for the plus region during the formation of the dipole (marked in the figure). The mean prediction error for the plus region is 0.18 and the standard deviation (SD) is 0.27. It can be seen that during the formation of the dipole, the error is well above the (Mean+2*SD). This shows a steep rise in error. For the minus region, the rise in error is not as steep as for the plus region during the formation of the dipole but a clear rise is visible. It is of the order of (Mean + 2*SD, = 0.38 + 0.55*2 ≈ 1.48). Hence it is observed for both the regions that the prediction errors rise abruptly during the formation of the dipole. We have calculated the correlation of the prediction errors with the difference in SST anomalies in the two regions values for both the regions. These were found to be 0.24 and 0.32 respectively.

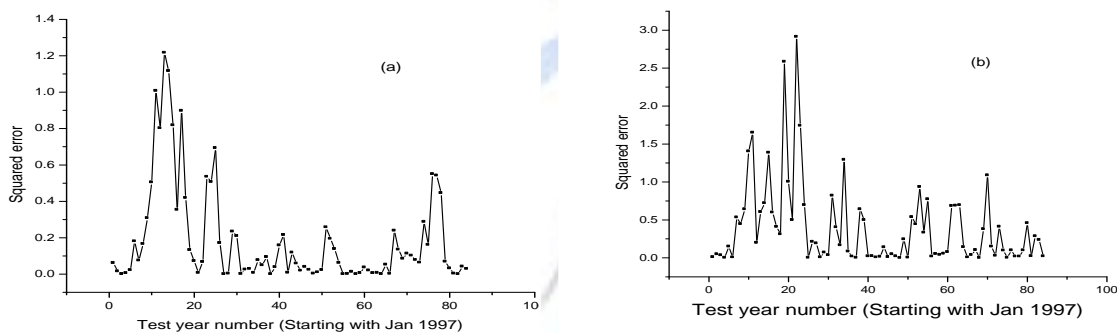


Figure 4: Prediction errors during the formation of the IOD (a) Plus region and (b) Minus region.

The abrupt rise in the prediction errors for both regions shows that the anomalous point was not on the trajectory of the dynamical system governing the SST anomalies. This observation pertains to the capability of linear multiple regression model to model the time series under discussion. The moderate correlation between the prediction errors and the difference in the anomalies of the two regions suggests that the model may be biased for certain regions of prediction. This result weakens the claim that the extreme weather phenomenon under discussion was not a part of the dynamical system and was an anomalous phenomenon caused by some local effects. For ideal models the errors must be independent of the region of prediction in temporal domain.

In order to conclusively establish or refute the above claim the study must be carried out using other global prediction models such as the ANN. Further, the role of ISOs for sustaining the IOD needs to be investigated in this light.

4 CONCLUSION

The question of whether or not the IOD of 1997 was part of the dynamical system governing the Indian ocean SST was attempted to be explored using the global prediction algorithm i.e. linear multiple regression. The applicability of global prediction algorithms as for such an investigation is important as it may lead to better predictability of anomalous phenomenon. A multiple linear regression model was constructed for the simulation of IO SST anomalies in the region of the 1997 dipole. The results show that the IOD event cannot be predicted to a reasonably good accuracy. The analyses of



regression prediction errors for both the regions reveal that the predictability declines during the formation of the dipole. It can thus be concluded that the formation of the dipole is not a consequence of the underlying deterministic system governing the evolution of SST in the IO. However, it may be too early to reach to a conclusion. For this, other global prediction algorithms such as the ANN, which have better modeling capability of modeling a chaotic time series, are required to be investigated.

ACKNOWLEDGMENTS

Our thanks to the experts who have contributed towards development of the template.

REFERENCES

- [1] Charney, J., Shukla, J. 1981 Predictability of monsoons, in *Monsoon Dynamics*, Cambridge University Press, 99-110.
- [2] Shukla, J., "Dynamical predictability of monthly means", *J. Atmos. Sci.*, 1981, 38, 2547-2572.
- [3] Yoo S.-H., S. Yang and C.-H Ho, "Variability of the Indian ocean sea surface temperature and its impacts on Asian-Australian monsoon climate", *J. Geophys. Res.*, 2006, 111, D03108, doi: 10.1029/2005JD006001.
- [4] Kucharski F., F. Molteni and J. H. Yoo, "SST forcing of decadal Indian Monsoon rainfall variability", *Geophys. Res. Lett.*, 2006, 33, L03709, doi:10.1029/2005GL025371.
- [5] Reddy P.R.C. and P.S. Salvekar, "Equatorial East Indian Ocean sea surface temperature: A new predictor for seasonal and annual rainfall", *Curr. sci.*, 2003, 85, 1600-1604.
- [6] Kug J.S., An Soon-II, J. Fie-Fie and In-Sik Kang, "Precondition for El Nino and La Nina onsets and their relation to Indian Ocean", *Geophys. Res. Lett.*, 2005, 32, L05706, doi:10.1029/2004GL021674.
- [7] Tang B., W. W. Hsieh, A. H. Monahan and F. T. Tangang, "Skill comparisons between neural networks and canonical correlation analysis in predicting the equatorial Pacific sea surface temperatures", *J. Clim.*, 2000, 13, 287 – 293.
- [8] Tripathi K. C., I. M. L. Das and A.K. Sahai, "Predictability of sea surface temperature anomalies in the Indian Ocean using artificial neural network", *Ind. J. Marine Sc.*, , 2006, 35, 210-220.
- [9] Yuan, Y, H. Yang, , W. Zhou, and Li. Chongyin,. "Influences of the Indian Ocean dipole on the Asian summer monsoon in the following year (p 1849-1859)", *Int. J. Clim.*, 2008, 28, 1835 – 1976.
- [10] Li, T., B. Wang, , C. P. Chang and Y. Zhang, "A Theory for the Indian Ocean Dipole–Zonal Mode", *Jour. Atm. Sci.* , 2003, 60, 2119–2135.
- [11] Smith T.M. and R.W. Reynolds, "A global merged land-air-sea surface temperature reconstruction based on historical observations (1880 – 1997)", *J. Climate.*, , 2005, 18, 2021-2036.
- [12] Lorenz E. N., "Deterministic nonperiodic flow", *J. Atmos. Sci.*, 1963, 20, 130–141.
- [13] Barnston A.G., and C.F. Ropelewski, "Prediction of ENSO episodes using canonical correlation analysis", *J. Climate.*, 1992, 5, 1316-1345.
- [14] Chau K.W., C.L. Wu and Y.S. Li, "Comparison of several flood forecasting models in Yangtze river", *J. Hydraul. Eng.* , 2005, – ASCE. 10, 485-491.
- [15] Saji N.H., B.N. Goswami, P.N. Vinayachandran and T. Yamagata, "A dipole in the tropical Indian Ocean", *Nature*, 1999, 401, 360 – 363.
- [16] H. Annamalai and J. M. Slingo, "Active/break cycles: Diagnosis of the intraseasonal variability of the Asian summer monsoon", *Clim. Dyn.*, 2001, 18, 85–102.
- [17] Waliser, D. E., K. M. Lau and W. Stern, "Potential predictability of the MJO", *Bull. Am. Meteorol. Soc.* , 2003, DOI: 10.1175/BAMS-84-1-33.
- [18] Ito, M. and, T. Satomura, "The Interannual Variation of Intraseasonal Oscillation Linked with the Indian Ocean Dipole Mode", *Scientific online letters on the Atmosphere*, 2009, 5, 69-72



Author' biography with Photo

K C Tripathi:



K C tripathi obtained his M.Sc degree in Computer Science from University of Allahabad, India, in 2003 and PhD degree from University of Allahabad, India in 2011. He has worked as a Scientist in the Department of Atmosphere and Ocean modeling in the University of Allahabad, India and is currently working as Assistant Professor in the department of Computer Science and Engineering in Inderprastha Engineering College, Ghaziabad, India. He has about 50 publications in various national and international journals and conferences. His research interestests include mathematical modeling of atmosphere and oceans, Artificial Neural Networks and pattern recognition.

Rashi Agarwal:



Rashi Agarwal has obtained her Masters degree in Computer Application from Indraprastha University, New Delhi, India in 2009. She is working as Assistant Professor in the department of IT, Maharaja Agrasen College of Management Studies, New Delhi, India. She is pursuing her PhD form Jamia Milia Islamia University, Delhi, India. Her research areas include Artificial Neural Networks, weather forecasting and statistical modeling.

P N Hrisheeksha:



Prof. P N Hrisheeksha is B.E. and M.Tech in electrical engineering. He obtained his PhD degree in Electrical Engineering from IIT Roorkee, India. He has over 25 years of teaching, research and consultancy experience. He is is currently Director of Inderprastha Engineering College, Ghaziabad, India and also consultant for few industries. He has several publications of national and international journals and is editor of several international journals of repute. He has over 20 years of teaching, research and consultancy experience and has been key note speakers in various national and international conferences in India and abroad. His research areas include power system operation and control, distributed generation, renewable energy resources, A. I. applications in the mentioned areas, distribution, automation, energy management, auditing and conservation and printed electronics.