A Markov Chain approach in the prediction of severe pre-monsoon thunderstorms through artificial neural network with daily total ozone as predictor
XXXth URSI General Assembly and Scientific Symposium to be Held in Istanbul, Turkey, August 13-20, 2011

Goutami Chattopadhyay and S. S. De

Centre of Advanced Study in Radio Physics and Electronics, University of Calcutta, 1, Girish Vidyaratna Lane, Kolkata 700 009, India
*E-mail: goutami15@yahoo.co.in (corresponding author)

Abstract

Purpose of the present paper is to examine the predictability of the occurrence of the severe pre-monsoon thunderstorm over Gangetic West Bengal. Instead of considering various meteorological predictors, the daily total ozone concentration is chosen as the predictor because of the influence of tropospheric as well as stratospheric ozone on the genesis of meteorological phenomena. Considering the occurrence/non-occurrence of thunderstorm in the pre-monsoon season (March-May) of the year 2005 as the dichotomous time series $\{X_t\}$ that realizes 0 and 1 for non-occurrence and occurrence of TS respectively, a first order two state (FOTS) Markov dependence is revealed within this time series.

1. Introduction

The climate of India is dominated by the summer monsoon (June to September). The entire year is, however, divided into four seasons, namely, winter (January and February), pre-monsoon (March-May), southwest or summer monsoon (June - September) and post monsoon (October-December). Pre-monsoon season is characterized by severe thunderstorms (TS), hailstorms and duststorms [1]. The hazardous consequences of pre-monsoon thunderstorms over the regions encompassing Kolkata are documented in works like [1] and [2]. The TS is a cumulus scale weather phenomenon and is generated due to instabilities created in the meso-scale [3, 4]. These weather systems play dual role in the atmosphere. The TSs are important because they take heat and moisture near the earth's surface and transport it to the upper levels to maintain the general atmospheric circulation [4].

Since 1940s, association between ozone and TS has been an area of interest [5, 6]. Ozone is a secondary pollutant formed in the atmosphere through photochemical reactions involving nitrogen oxides (NOx) and volatile organic compounds (VOC) in the presence of sunlight. The temporal evolution of ozone concentrations at the ground level is controlled strongly by the diurnal variation of the atmospheric boundary layer. Under certain meteorological conditions, ozone can be formed and accumulated in the daytime boundary layer [7]. In a study by Dickerson *et al* (1987) [8] it was explained how the occurrence of TS influences the transport of air pollutants. Pickering *et al* (1990) [9] determined that ozone production in the upper troposphere can be enhanced by convective transport of precursor species. Cloud scale chemical transport models can provide valuable insight into the impact of convection and lightning on tropospheric chemistry [10,11]. Using simulations of the photochemistry, DeCaria *et al* (2005) [12] have shown that the occurrence of TS leads to increase in the ozone production in the upper atmosphere. Orville (1967) [6] has shown that the ozone increase is associated with thunderstorm phenomenon such as lightning or induced corona. Transport of air parcels from the stratosphere into the cloud anvil has been identified as a source of high concentrations of ozone measured in cloud top regions by several authors [6,13].

In the present work, the procedure of hypothesis testing is adopted with the null hypothesis proposing serial independence in the data series and alternative hypothesis proposing First-order Two-state (FOTS) Markovian dependence in the pre-monsoon thunderstorm occurrence/non-occurrence time series over Gangetic West Bengal (GWB), which includes the broad urban area of Kolkata.

2. Methodology

2.1 Formulation of Markov chain model

The present work basically aims at modeling the occurrence/non-occurrence of severe pre-monsoon TS over GWB. For this purpose, the thunderstorm occurrence/ non-occurrence is observed over GWB during March-May in 2005. On a

978-1-4244-6051-9/11/$26.00 ©2011 IEEE

given day, a TS either occurs or does not occur over a specified geographical zone. Thus, the TS event would generate a dichotomous time series $\{X_t\}$ where the random variable would be defined as [14]

X_t = 0 if TS does not occur on day '*t*'

= 1 if TS occurs on day '*t*' (1)

The transition probabilities would be the conditional probabilities

$p_{00} = \{X_{t+1} = 0 / X_t = 0\}$, $p_{01} = \{X_{t+1} = 1 / X_t = 0\}$, $p_{10} = \{X_{t+1} = 0 / X_t = 1\}$, $p_{11} = \{X_{t+1} = 1 / X_t = 1\}$.

2.2 Formulation of Artificial Neural Network with daily total ozone as predictor

The TO datasets used in the present paper consist of the measurements made by EP/TOMS available at the website ftp://jwocky.gsfc.nasa.gov/pub/eptoms/data/overpass/OVP075 epc.txt, which provides measurements of Earth's total column ozone by measuring the backscattered Earth radiance in the six 1 nm bands. The Artificial Neural Network (ANN), in the form of multilayer perceptron (MLP) has been used with daily TO as the entries to the input matrix. Theoretical details of MLP are available in Haykin (2001) [15].

In general, a MLP consists of a lowermost input layer, any number of hidden layers, and an output layer at the top. In a network, the total input received by neuron 'j' in $(h+1)$ th layer is defined as;

$$X_j^{h+1} = \sum_i Y_i^h W_{ji}^h - \theta_j^{h+1} \quad (2)$$

For convenience, the threshold 'θ' is taken to be zero.

In the present problem, the input layer of the MLP has been constructed by taking elements from the time series $\{X_1, X_2, ..., X_n\}$, where X_i denotes the TO concentration of day i within the period of study. The models are denoted as ANN-k where 'k' denotes the number of days, for which the TO concentrations are being used as predictor for a particular model. In these MLPs, the target output is the dichotomous pre-monsoon TS time series. For example, the ANN-3 model is trained as the MLP, which has the input patterns as $[X_1, X_2, X_3], [X_2, X_3, X_4], [X_3, X_4, X_5]$ etc. Different ANN-k models are trained using equation (2) and during the training sigmoid non-linearity is adopted. Only one hidden layer is considered whose maximum size is taken as 30. In each model, 70% of the input data are takes as training cases and 30% of the input data are chosen for testing the model.

3. Examining Markov process for ANN generated time series

The transition probabilities are computed for the observed dichotomous TS occurrence/non occurrence time series spread over the pre-monsoon season of 2005 in the GWB. The daily TO time series is plotted in figure 1. Using backpropagation learning with sigmoid non-linearity, a time series is generated for TS by the ANN. For the observed as well as ANN generated time series, a hypothesis testing is executed as follows:

Null hypothesis H_0: the time series is serially independent i.e. it is generated by 0-order two state Markov Chain

Alternative hypothesis H_1: the time series is generated by a FOTS Markov Chain

To test the hypotheses, a test statistic chi-square (χ^2) is constructed as follows [14]

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (3)$$

with (2-1)×(2-1)=1 degree of freedom. The O_{ij} and E_{ij} are calculated from the contingency table.

From the standard χ^2 table, $\chi^2_{0.05}(1) = 3.841$. From the observed TS time series, it is found that $\chi^2_{0.05}(1) = 9.963$. Thus, the observed TS is generated by FOTS Markov Chain. Similarly, for each ANN-k model, the χ^2 value is calculated. All of the χ^2 values obtained in this manner are presented in figure 2 and the corresponding transition probabilities are presented in table 1. It is observed that for ANN-3, χ^2 = 3.9475, which exceeds tabular value marginally. However, for ANN-1, χ^2 = 5.9831 and for ANN-5, χ^2 = 7.1796, which dominate the tabular value significantly. Since the χ^2 obtained for ANN-1, ANN-3, and ANN-5 exceed the tabular value, the null hypothesis of

serial independence can be rejected. This implies that the TS occurrence/non-occurrence time series generated by these three ANN models are characterized by FOTS Markov process. From the above discussion, it is understood that the FOTS Markov chains generated by ANN-1, ANN-3 and ANN-5 are based on TO concentration of 1, 3 and 5 consecutive days respectively as the required predictors.

In the next step, the climatological probabilities of occurrence of TS are computed for the time series generated by FOTS Markov chains. The climatological probability of occurrence of thunderstorm is computed as [14]

$$\pi_1 = \frac{p_{01}}{1 + p_{01} - p_{11}} \tag{4}$$

For ANN-1, ANN-3 and ANN-5 the values of π_1 are 0.2828, 0.4686 and 0.4489 respectively. To compare the performance of ANN-1, ANN-3 and ANN-5 in predicting TS from TO concentration, the lag-1 auto correlations are computed and are presented in figure 3. It is found that for ANN-5, the lag-1 autocorrelation is closest to the lag-1 auto correlation of the observed TS time series.

Thus, the FOTS Markov Chain processes implied by ANN-5 best represents the observed TS time series. Although both of TS time series generated by ANN-3 and ANN-5 are producing climatological probabilities of TS occurrence very close to the observed one (π_1=0.549), the higher value of lag-1 autocorrelation coefficient lead to the conclusion that ANN-5 is a better representative of the TS time series than the ANN-3. However, at 5% level of significance, the ANN-2 and ANN-4 do not produce FOTS Markov Chain i.e. these are serially independent at this level of significance.

The transition probabilities implied by ANN-5 are compared with those from the observed TS time series and are presented in figure 4. It is found that p_{11} and p_{01} are having values close to the observed TS time series. Thus, the ANN-5 is significantly successful in predicting a TS day when the previous day was a no TS or a TS day.

4. Conclusion

Occurrence/non-occurrence of pre-monsoon thunderstorms over Gangetic West Bengal is characterized by the first order two state Markov process. Artificial neural network in the form of a three layered multilayer perceptron is capable of generating a time series having similar Markov dependence to the observed occurrence/non-occurrence of thunderstorms over Gangetic West Bengal. Daily total ozone concentration can be a good predictor for the pre-monsoon thunderstorms over Gangetic West Bengal if artificial neural network is chosen as the predictive tool. If total ozone concentrations for five consecutive days are used as predictor, then the artificial neural network predicts the occurrence of thunderstorm on the next day with probability 0.727 if the given day has a thunderstorm.

5. Acknowledgments

This work is funded by Indian Space Research Organization (ISRO) through S. K. Mitra Centre for Research in Space Environment, University of Calcutta, Kolkata 700 009, India.

6. References

1. U. S. De, R. K. Dube and G. S. Prakasa Rao, "Extreme Weather Events over India in the last 100 years", *Journal of Indian Geophysical Union*, **9**, 2005, pp. 173-187.
2. U.S. De and M.M. Dandekar, "Natural Disasters in Urban Areas", *The Deccan Geographer*, **39**, 2001, pp. 1-12.
3. Y. L. Lin, "*Mesoscale Dynamics"*, Cambridge University Press, UK 2007.
4. S.M. Deshpandea and P.E. Raj, "UHF wind profiler observations during a tropical pre-monsoon thunderstorm — A case study", *Atmospheric Research*, **93**, 2009, pp. 179-187.
5. G.M.B. Dobson, A. W. Brewer and B.M. Cwilong, "Meteorology of the lower stratosphere", *Proceedings of Royal Society of London, A*, **185**, 1945, pp. 144-175.
6. R. E. Orville, "Ozone production during thunderstorms, measured by the absorption of ultraviolet radiation from lightning", *Journal of Geophysical Research*, **72**, 1967, pp. 3557-3561.
7. J. Zhang and S. T. Rao,"The Role of Vertical Mixing in the Temporal Evolution of Ground-Level Ozone Concentrations", *Journal of Applied Meteorology*, **38,** 1999, pp. 1674-1691.
8. R. R. Dickerson, *et al.*, "Thunderstorms: An important mechanism in the transport of air pollutants", *Science*, **235**, 1987, pp. 460– 465.

9. K. E. Pickering,, A. M. Thompson, R. R. Dickerson, W. T. Luke, D. P. Mcnamara, J. Greenberg, and P. R. Zimmerman, "Model calculations of tropospheric ozone production potential following observed convective events," *J. Geophys. Res*., **95**, 1990, pp. 14049– 14062.
10. K. E. Pickering A. M. Thompson, J. R. Scala, W.-K. Tao, R. R. Dickerson, and J. Simpson, "Free tropospheric ozone production following entrainment of urban plumes into deep convection", *J. Geophys. Res*., **97**, 1992, pp. 17985– 18000.
11. K. E. Pickering *et al*., "Convective transport of biomass burning emissions over Brazil during TRACE-A", *J. Geophys. Res*., **101**, 1996, pp. 23993– 24012.
12. A. J. Decaria, K. E. Pickering, , G. L. Stenchikov and L. E. Ott, "Lightning-generated NOX and its impact on tropospheric ozone production: A three-dimensional modeling study of a Stratosphere-Troposphere Experiment: Radiation, Aerosols and Ozone (STERAO-A) thunderstorm", *Journal of Geophysical Research*, **110**, 2005, D14303, doi:10.1029/2004JD005556.
13. O. Poulida, et al., "Stratosphere-troposphere exchange in a midlatitude mesoscale convective complex 1. Observations", *J. Geophys. Res*., **101**, 1996, pp. 6823-6836.
14. D.S. Wilks, "*Statistical methods in atmospheric Science"s*, Second Edition, Elsevier Inc 2006.
15. S. Haykin, "Neural Networks: A Comprehensive Foundation", second ed. Pearson Education Inc., New Delhi, India 2001.

Table 1. Transition probabilities of order 1 for the observed and ANN-k generated TS time series.

	Occurrence/Non-occurrence of pre-monsoon TS					
Transition probability	*Observed*	*ANN-1*	*ANN-2*	*ANN-3*	*ANN-4*	*ANN-5*
p00	0.813	0.842	0.647	0.706	0.600	0.778
p01	0.188	0.158	0.353	0.294	0.400	0.222
p10	0.154	0.400	0.417	0.333	0.357	0.273
p11	0.846	0.600	0.583	0.667	0.643	0.727

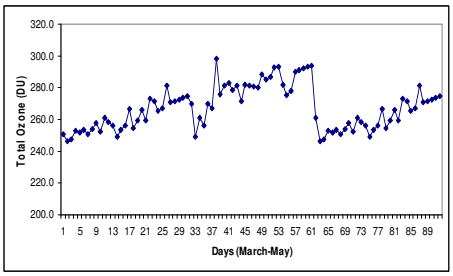

Figure 1. Line diagram showing the time series of daily total ozone in the pre-monsoon season of 2005

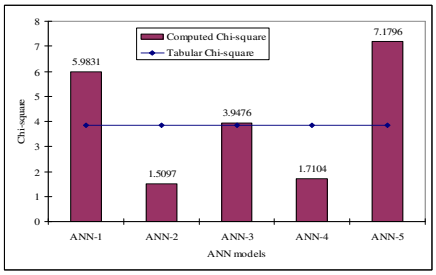

Figure 2. Bar diagram showing the Chi-square values for the ANN generated TS occurrence/non-occurrence time series.

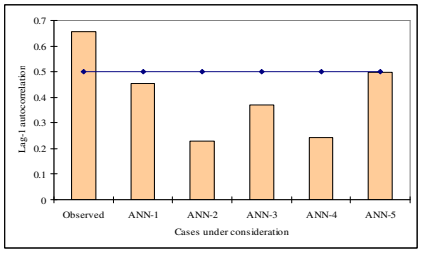

Figure 3. Bar diagram showing the lag-1 autocorrelation coefficients.

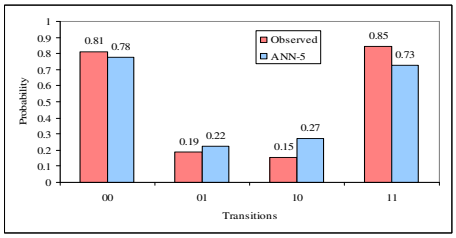

Figure 4. Schematic showing the transition probabilities pertaining to the FOTS Markov chains corresponding to the observed and ANN-5 generated pre-monsoon TS time series