



## An Alternative Approach to Structure Specification Based on Fuzzy Multidimensional Membership Function Using Forward Selection Rule

Sreyasi Ghosh, Assistant Professor, Department of Commerce (Mathematics), The Bhawanipur Education Society College ghosh.sreyasi1@gmail.com,+919433785429

Sarbari Ghosh, Department of Mathematics, Vidyasagar Evening College and Guest Faculty Department of Atmospheric Science, C.U..

Pradip Kumar Sen, Department of Mathematics, J.U.

Subhra Chatterjee, Assistant Professor, Academy of Technology, Kolkata.

*Published: 4 December 2015*

---

**Abstract:** Fuzzy logic first established in July,1964 by Lofti A. Zadeh, is usually used to develop cost-effective approximate solutions to complex real-world problems exploiting the tolerance of imprecision. The present study attempts to develop a general computational technique based on fuzzy multi-dimensional membership function using forward selection rule for discriminating two different situations which is basically non-linear. Incidentally the technique suggested here is validated with atmospheric data. Earlier the fundamental principal component analysis (PCA) technique was applied to identify the significant parameters for the occurrence of pre-monsoon thunderstorms (TS) in Kolkata. They showed how the linear discriminant analysis (LDA) technique alone as well as in conjunction with PCA can be successfully applied for the purpose (Ghosh et al. 1999, 2004; Chatterjee et al., 2009). Also a comparative study was performed between the existing multivariate technique, the linear discriminant analysis and a technique based on fuzzy membership roster method (Chatterjee et al., 2011). Recently a fuzzy –neuro based algorithm for weather prediction has been developed (T. Rahman et al. 2014). The main objective of the study is to address the numerical imprecision of some quantified physical variables. In this rule, a product form is taken to construct the multivariate membership function where the univariate membership function is Gaussian in nature as well as continuously differentiable. Since the parameters may have different units so they are made dimensionless before taking the product. To develop the technique no software package or fuzzy toolbox is used. The program for the study is developed by the authors themselves. This rule is applied to two datasets of different categories consisting of the parameters of the days with convective development and fair weather respectively during pre-monsoon season of Kolkata (22.53°N, 88.33°E), India. Basic parameters for discriminating the

situation (convective development and fair weather in pre-monsoon season of Kolkata) is constructed from the known data set of 12 years covering the period 1985-1996. The results are validated for the period 1997-1999 using the dataset consisting of variables of unknown nature. The study reveals that the technique can classify the two different situation to give the best possible combination of parameters with atmost 88% success rate. Moreover, the study indicates that the two datasets are structurally different. The technique suggested here is expected to work in any other domain too. It is found that the method works with better accuracy than the existing ones so far the atmospheric parameters are concerned.

---

## INTRODUCTION

Pre monsoon thunder squalls/ Nor'westers are one of the most important events that occupy a major portion in the pre-monsoon weather system over Eastern India. Contribution of different meteorological parameter like temperature, pressure, humidity, etc are the most important variables which play a significant role for the development of the Pre monsoon thunder squalls/ Nor'westers. Although other factors can also be taken in to consideration for the ideal situation and actual time and place of occurrence of all these thunder squalls during the hot weather period of Summer season( mainly within the optimum period of March to May), every year, in most cases, over the region of Eastern and North-Eastern states. Normally these Pre monsoon Thunder squalls/ Nor'westers are very much violent and destructive in nature and appear suddenly in form of dark big clouds with sudden rise in wind speed associated with frequent lightning and thunder. Meteorological Scientist on previous occasions have contributed their valuable ideas and thoughts involving the phenomena by their contributions at different times. However, for forecasting the occurrence of Pre monsoon thunder squalls/ Nor'weaster, some times it is necessary to know the favorable conditions and mechanism of these phenomena.

Expert forecasters use well-developed subjective techniques for weather prediction. They improve their accuracy and skill over time by learning through experiences. Over the years forecasters have a huge collection of dataset and products. So, they can use, intelligent system approaches for data analysis, interpretation, verification etc. Fuzzy logic is one of such intelligent or expert systems, the goal of which is to perform at the level of a human expert by leveraging knowledge and experience gained over time. Ghosh *et al.* have applied Fuzzy multivariate membership function using forward selection rule to the data set of three years to identify significant parameters for the occurrence of pre-monsoon thunderstorms (TS) at Kolkata (22.53°N, 88.33°E). The present work aims at the reduction of parameters using some objective method as well as to predict the convective development during pre-monsoon season at Kolkata utilizing radiosonde data of 15 years. In fact, many researchers in different situations have used these multivariate techniques. The principal component analysis (PCA) technique was applied by previous workers to identify the significant parameters for the occurrence of pre-monsoon thunderstorms (TS) in Kolkata. They showed how the linear discriminant analysis

(LDA) technique alone as well as in conjunction with PCA can be successfully applied for the purpose (Ghosh et al. 1999, 2004; Chatterjee et al., 2009)<sup>1</sup>. Also a comparative study was performed between the existing multivariate technique, the linear discriminant analysis and a technique based on fuzzy membership roster method (Chatterjee et al., 2011)<sup>2</sup>. Eigenvector methods have been applied to study the principal anomaly patterns of winter temperature<sup>3</sup>. The principal components derived from a 500 hPa height data set had been linearly transformed to interpret spatial patterns<sup>4-5</sup>. The principal components based on covariance matrix and correlation matrix for a given data set of cyclone frequencies have been compared<sup>6</sup>. Cluster analysis and linear discriminant analysis (LDA) have been utilized to describe a multivariate statistical model for forecasting anomalies of surface pressure present over Europe and North Atlantic<sup>7</sup>. A comparative study of rotated and unrotated PCA has been performed<sup>8</sup>. A composite empirical orthogonal function (EOF) analysis of the monthly sea surface temperature variations and those of precipitation in the tropical Pacific Ocean region was performed<sup>9</sup>. Multiple linear regressions was compared with LDA for making hind casts and real time forecasts of north-east Brazil wet season rainfall using sea surface temperature<sup>10</sup>. Though a number of attempts were made to establish empirical models for the prediction of atmospheric stability<sup>11-12</sup>, the work done on Kano is perhaps the first successful attempt for tropical region<sup>13</sup>. Some statistical forecast models, based on logistic regression, have been reported in the literature. While, in one case, six variables were selected from 27 variables by constructing correlation matrix<sup>14</sup>, in the other case, variable reduction was done using forward and

backward selection procedures<sup>15</sup>. In India, a number of attempts have been made to describe occurrence of rainfall by two states Markov-chain<sup>16-19</sup>. Another attempt has been made to predict the occurrence of CD at Dhaka (Bangladesh) in terms of stability indices<sup>20</sup>. Complex empirical orthogonal function was used to determine vertical wind profiles over the Indian Ocean<sup>21</sup>. Another study has produced a computational algorithm for one-dimensional cyclostationary empirical orthogonal functions and examined their properties<sup>22</sup>. A low order barotropic EOF model has also been reformulated<sup>23</sup>.

Convective developments are strongly favoured by convective instability, abundant moisture at lower levels, strong wind shear, and a dynamical lifting mechanism that can release the instability<sup>24</sup>. Not only that, the vertical shear of the environmental winds has to match the value of the convective instability for proper development of a large convective cloud<sup>25</sup>. It has been emphasized that the presence of conditional instability is an essential criterion for supporting electrification and lightning<sup>26</sup>. Apart from the parameters mentioned, two more parameters, viz.  $(\theta_{es} - \theta_e)$  and  $(P - PLCL)$  have been included in the present study, where,  $\theta_{es}$  and  $\theta_e$ , denote the saturated equivalent potential temperature and equivalent potential temperature, respectively; P, is a level pressure; and PLCL, is the pressure at the corresponding lifting condensation level.

The thermodynamic parameter  $(\theta_{es} - \theta_e)$  was originally introduced by Betts as a measure of the unsaturation of the atmosphere<sup>27</sup>. PLCL for the surface parcel was considered as the cloud base and hence,  $(P - PLCL)$  has been taken as a forcing factor for the saturation of a parcel<sup>28</sup>.

The number of convective development) and fair weather (FW) days linked with the morning and evening radio sonde / rawin sonde (RS/RW) observations are presented in Tables 1 and 2. These data are used for the construction of discriminant functions. Any convective development occurring within the next 12 hrs of the morning RS/RW observation taken at 0000 hrs GMT (0530 hrs IST) is considered as CD related with morning RS/RW, otherwise it is FW related with the same RS/RW. A similar consideration for evening RS/RW observation taken at 1200 hrs GMT is utilized for the classification of CD or FW linked with evening

RS/RW. On many occasions, the data either at one or more of the significant levels, i.e. 1000, 850, 700, 600 and 500 hPa were not available. Naturally, those occasions could not be taken into consideration.

These limitations have greatly reduced the data size. The linear discriminant functions (morning and evening) for forecasting the convective development at Kolkata have been constructed utilizing all the available radiosonde data of 12 years (1985-1996) and for the validation of these functions, the radiosonde data of 3 years (1997-1999) have been used.

### 3 METHODOLOGY

The study is performed separately for morning and evening, as two radio soundings are available in a day. The atmospheric layer (1000-500)h Pa is subdivided into the following four layers : - (1000-850)h Pa, (850-700)h Pa, (700-600)h Pa and (600-500)h Pa. A day associated with a sub layer is considered as a vector containing at most five components, among which the first two components represent two thermodynamic parameters and the

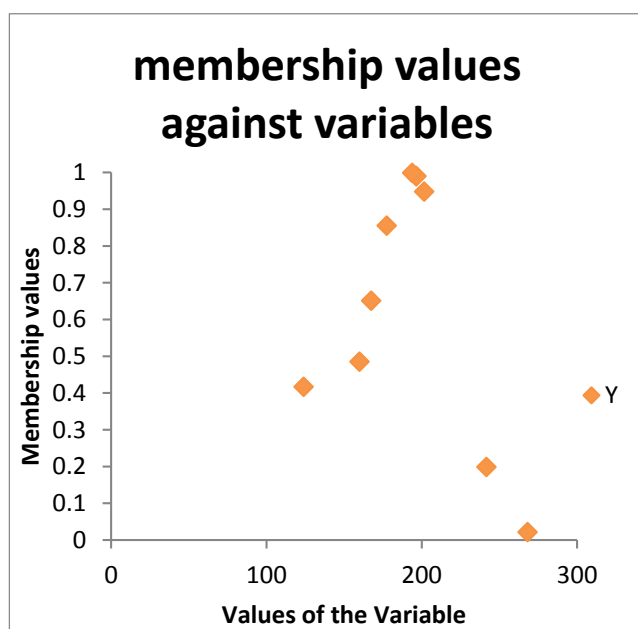
remaining three are the dynamic parameters. The category or pattern of an unknown day is predicted for next 12 hours from the time of observations depending on its degree of compatibility with the sets of days of two known categories or standard patterns ( i.e. convective development and fair-weather). The study is performed separately for morning and evening. The sets of days of known categories are termed as fuzzy sets as the two sets have overlapping area so far the quantified values of the dynamic parameters like conditional instability, convective instability and vertical shear are concerned.

#### Forward selection

The simplest data-driven model building approach is called forward selection. In this approach, one adds variables to the model one at a time. At each step, each variable that is not already in the model is tested for inclusion in the model. The most significant of these variables is added to the model, so long as it's P-value is below some pre-set level. Thus we begin with a model including the variable that is most significant in the initial analysis, and continue adding variables until none of remaining variables are "significant" when added to the model. We have only verified the result for the layer (1000-850)h Pa. There are five parameters. A product form is taken to construct the multivariate membership function. Since the parameters have different units so they are made dimensionless before taking the product. To develop the technique no software package or fuzzy toolbox is used. The program for the study is developed by the authors themselves. Since some of the parameters are found to follow Gaussian distribution and usually the physical parameters are assumed to be Gaussian or quasi Gaussian in nature, for each parameter, the

Gaussian membership function has been chosen to construct the one dimensional or univariate degree of compatibility. Gaussian membership functions are continuously differentiable as well as parameterizable. Gaussian membership functions are factorizable. Hence, we may synthesize a multi dimensional or multivariate degree of compatibility as the product of one dimensional or univariate degree of compatibility.

The graph of the membership grade values against the variables also suggests Gaussian nature of the membership function. The graph given below is an example of the relation between the values of the variables and their corresponding membership grade values.



Let us consider two groups X and Y, where X consists of the parameters of FW situations and the elements of Y are the parameters representing the situations of CD. Let us suppose that there are k parameters,  $U_i$  ( $i = 1$  to k) on which we have the following two sets of observations:

$$X = [X_{ij}] \quad (i = 1 \text{ to } k, j = 1 \text{ to } m) \text{ and}$$

$$Y = [Y_{ij}] \quad (i = 1 \text{ to } k, j = 1 \text{ to } n).$$

In the present study,  $U_i$  ( $i = 1$  to 5) denote the above mentioned 5 parameters,  $X_{ij}$  denotes the value of the  $i$ th parameter on  $j$ th FW day and  $Y_{ij}$  gives the value of the  $i$ th parameter on  $j$ th CD day.

The work has been performed with  $k = 5$ ,  $m =$  the number of FW days, which is 280 for morning and 201 for evening and  $n =$  the number of CD days, which is 123 for morning and 165 for evening. X and Y are termed as fuzzy sets since it is difficult to identify sharp boundaries between these two sets so far the parameters, viz., convective instability, conditional instability and vertical shear are concerned. Then the degrees of compatibility of a parameter,  $U_i$  ( $i=1$  to 5) with the standard pattern classes, Y and X are computed by  $AY(U_i)$  and  $AX(U_i)$ , again  $AY(U)$  and  $AX(U)$  are the product of the uni variates respectively. If, now, an unknown pattern or a day, say  $U = (U_1, U_2, \dots, U_5)$  is given, where  $U_i$  is the measurement associated with the  $i$ th parameter of the pattern, then the degrees of compatibility of U with the standard patterns, Y and X, denoted by  $AY(U)$  and  $AX(U)$  respectively, are computed. Next, an unknown pattern or a day, U is classified by the larger value of  $AY(U)$  or  $AX(U)$ , i.e. if  $AY(U) > AX(U)$ , then there is a possibility for U to be more of the pattern Y than of the pattern X for next 12 hours. Hence it may be predicted that U is possibly a day with convective development for next 12 hours (Klir and Yuan, 2002)<sup>27</sup>. Here, the number of days of unknown category involved in the validation is 44 for MCD, 84 for MFW, 53 for ECD and 65 for EFW. It is worth mentioning in this context that there is no sound principle yet for guiding the choice of membership function or degree of compatibility.

## RESULT

No. of Combinations	Nature of Day	No. of Days	Combination	No. of Correct Prediction	% of Correct Result
1	EC D	53	1	29	54.72
	EF W	65	4	40	61.54
	MF W	84	3	45	53.57
	MC D	44	3	24	54.55
2	EC D	53	1,3	45	84.91
	EF W	65	4,3	55	84.62
	MF W	84	3,4	64	76.19
	MC D	44	3,5	34	77.27
3	EC D	53	1,3,2	42	79.25
	EF W	65	4,3,1	56	87.69
	MF W	84	3,4,1	73	86.91
	MC D	44	3,5,4	36	81.82
4	EC D	53	1,3,2,4	47	88.68

	EF W	65	4,3,1,2	57	86.15
	MF W	84	3,4,1,2	70	83.33
	MC D	44	3,5,4,2	32	72.73
5	EC D	53	1,3,2,4,5	44	83.02
	EF W	65	4,3,1,2,5	57	87.69
	MF W	84	3,4,1,2,5	74	86.91
	MC D	44	3,5,4,2,1	39	88.64

Thus we can see from the table, for the layer we have discussed here combination of all the parameters give the best result for Morning convective development, with a whopping 88.64% correct result. Whereas for Evening convective development combination of 4 parameters 1,2,3,4 gives a better result compared to the one with all 5 parameters, with again 88.68% correct result.

#### CONCLUSION

As any natural phenomenon is inherently complex and multivariate, its exact prediction is difficult. Not only that any natural phenomenon is obviously nonlinear in nature. Many attempts for pre-monsoon weather forecasting of Kolkata(22.53°N,88.33°E),India, have been tried by previous workers. But it is worth mentioning that this method works with better accuracy than the existing ones.

We have applied this rule to two datasets (morning and evening) of different situation (convective development and fair weather). The



study reveals that the rule can classify the situation with almost 88% success rate. Moreover, the study indicates that the two datasets are structurally different. The study reveals that the technique used in the present analysis, are efficient for forecasting pre-monsoon weather of Kolkata, India. It might be preferable to the field forecasters as it will reduce the number of parameters.

The technique suggested here is expected to work in any other domain too.

### Acknowledgements

The authors wish to express their gratitude to India Meteorological Department, Kolkata for supplying necessary data for the study.

### References

- 1 Ghosh S, et al. Reduction of number of parameters and forecasting convective developments at Kolkata (22.53°N, 88.33°E), India during pre-monsoon season: An application of multivariate techniques. (1999) pp 673-681.
- 2 Ghosh S, et al Comparison between LDA technique and fuzzy membership roster method for pre-monsoon weather forecasting (2011) pp 137-144.
- 3 Diaz H F & Fullbright D C, Eigenvector analysis of seasonal temperature, precipitation and synoptic-scale system frequency over contiguous United States: Part 1 (Winter), *Mon Weather Rev (USA)*, 109 (1981) pp 1267-1284.
- 4 Wallace J M & Gutzler D C, Telecommunications in the geopotential height field during the Northern Hemisphere winter, *Mon Weather Rev (USA)*, 109 (1981) pp 784-812.
- 5 Horel J D, A rotated principal component analysis of the interannual variability of the Northern Hemisphere 500 mb height field, *Mon Weather Rev (USA)*, 109 (1981) pp 2080-2092.
- 6 Overland J E & Preisendorfer R W, A significance test for principal components applied to a cyclone climatology, *Mon Weather Rev (USA)*, 110 (1982) pp 1-4.
- 7 Maryon R H & Storey A H, A multivariate statistical model for forecasting anomalies of half-monthly mean surface pressure, *Int J Climatol (UK)*, 5 (1985) pp 561-578.
- 8 Richman M B, Rotation of principal components: Review article, *Int J Climatol (UK)*, 6 (1986) pp 293-335.
- 9 Weare B C, Relationship between monthly precipitation and SST variation in the tropical Pacific region, *Mon Weather Rev (USA)*, 115 (1987) pp 687-698.
- 10 Ward M N & Folland C K, Prediction of seasonal rainfall in the North Nordeste of Brazil using eigenvector of sea-surface temperature, *Int J Climatol (UK)*, 11 (1991) pp 711-743.
- 11 Showalter A K, A stability index for thunderstorms forecasting, *Bull Am Meteorol Soc (USA)*, 34 (1953) pp 250-252.
- 12 Darkow G L, The total energy environment of storms, *J Appl Meteorol (USA)*, 7 (1968) pp 199-205
- 13 Oduro-Afriyie K & Adefolalu D O, Instability indices for severe weather forecasting in West Africa, *Atmos Res (Netherlands)*, 30 (1993) pp 51-68.
- 14 Reddy P J, Barbarick D E & Osterberg R D, Development of a statistical model for forecasting episodes of visibility degradation in the Denver Metropolitan area, *J Appl Meteorol (USA)*, 34 (1995) pp 616-625.
- 15 Dasgupta S & De U K, Binary regression models for short term prediction of pre-monsoon Convective developments over

Kolkata (India), *Int J Climatol (UK)*, 27 (2007) pp 831-836.

16 Dasgupta S & De U K, Markov chain models for pre- monsoon thunderstorms in Calcutta, India, *Indian J Radio Space Phys*, 31 (2001) pp 138-142.

17 Kulkarni M K, Kandalgaonkar S S, Tinmake M R & Nath A, Pre-monsoon season thunderstorms over Pune, *Int J Climatol (UK)*, 22 (2002) pp 1415-1420.

18 Pant B & Shivhare R P, Markov chain model for study of wet / dry spells at AF Stn Sarsawa during SW monsoon season, *Vatavaran (India)*, 22 (1998 ) pp 37-50.

19 Thiagarajan R, Ramadoss & Ramaraj, Markov chain model for daily rainfall occurrences at east Thanjavur district, *Mausam (India)*, 46 (1995) pp 383-388.

20 Chowdhury A M, Ghosh S & De U K, Analysis of pre- monsoon thunderstorm occurrence at Dhaka from 1983 to 1992 in terms of (  $\text{es- e}$ ) and convergence / divergence at surface, *Indian J Phys*, 70B (1996) pp 357-366.

21 Kishtawal C M, Basu S & Pandey P C, An algorithm for retrieving vertical wind profiles

from satellite-observed winds over the Indian Ocean using complex EOF analysis, *J Appl Meteorol (USA)*, 35 (1996) pp 532-540.

22 Kim K Y, North G R & Huang J, EOFs of one-dimensional cycle stationary time series, computation, examples and stochastic modeling, *J Atmos Sci (USA)*, 53 (1996) pp 1007-1017.

23 Selten F M, A statistical closure of a low-order barotropic model, *J Atmos Sci (USA)*, 54 (1997) pp 1085-1093.

24 Kessler E, *Thunderstorm morphology and dynamics* (US Department of Commerce, USA), 1982, pp 93-95, 146-149.

25 Asnani G C, *Tropical Meteorology Vol 2* (Pune, India), 1992, pp 829-833, 852.

26 Williams E & Renno N, An analysis of the conditional instability of the tropical atmosphere, *Mon Weather Rev (USA)*, 121 (1993) pp 23-26.

27 Klir G. J. and B. O. Yuan, 2002. *Fuzzy sets and fuzzy logic, theory and applications*. Prentice Hall of India Pvt. Ltd.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions defined by MDPI AG, the publisher of the Sciforum.net platform. Sciforum papers authors the copyright to their scholarly works. Hence, by submitting a paper to this conference, you retain the copyright, but you grant MDPI AG the non-exclusive and unrevocable license right to publish this paper online on the Sciforum.net platform. This means you can easily submit your paper to any scientific journal at a later stage and transfer the copyright to its publisher (if required by that publisher). (<http://sciforum.net/about>).