Asian J. Exp. Sci., Vol. 23, No. 1, 2009; 193-198

A Comparative Analysis of Forecasting Methods for Aerobiological Studies

Karuna S. Verma[1] and Apurva K. Pathak[2]*

[1] Aeroallergens and Immunology Laboratory, Department of Biological Sciences, Rani Durgavati University, Jabalpur-482001 (M.P.); India.

[2] Department of Pathology & Microbiology, Modern Dental College & Research Center, Indore-453112 (M.P.); India.

Abstract : Forecasting of air borne fungal spores are a problem for a long time to the many aero biologist. Time series analysis using ARIMA and Multiple Linear Regression Analysis are the methods generally in use. After introduction of artificial intelligent in forecasting the situation is changed, but the efficiency of these models are still under question. This study revealed that, for the short term forecasting all the three models are given significant level prediction of >95% accuracy. However, for long term due to the generalization made by the Seasonal Autoregressive Moving Average, Multiple Linear Regression Analysis and Neural Network Application can gave a significant prediction.

Key words : Forecasting, Multiple Linear Regression Analysis, SARIMA & Neural Network Application.

Introduction

Air is vital for life; the atmosphere contains many minutes' particles of organic and inorganic matters though the high light intensities, extreme temperature variations, low concentrations of organic matter, and a scarcity of water, make the atmosphere as unsuitable environment habitat for microbial growth; yet it contain a high varieties of organic substances (Atlas,1984) . Among the organic matters, a great parts of it comprising the fungal spores, which are discharged by various mechanisms. Fungal spores are ubiquitous on their occurrence but the quality and quantity of these spores are largely depends on the other factors.

It is important to note that geography and climate play an important role in determining the outdoor air microbial concentrations because the transport of bioaerosol is primarily governed by hydrodynamic and kinetic factors, while their fate is dependent on their specific chemical makeup and the meteorological factors to which they are exposed. The most significant environmental factors influencing the viability of microorganisms are temperature, relative humidity (RH), and wind velocity (Jones and Harrison, 2003). Thus, the forecasting of air borne fungal spores based on environmental factors is possible. Previous workers used Time and Series methods for forecasting but now a day's Neural Network Analysis is being uses for this purpose.

In order to find out the best-fitted model for the current study and to reveal the best methods and their correlation, in this study Multiple Linear Regression Analysis (LR), Seasonal Autoregressive Moving Average (SARIMA) and Neural Network Analysis(NAA) have been performed.

* **Corresponding author :** Dr. Apurva K. Pathak, Department of Pathology & Microbiology, Modern Dental College & Research Center, Indore-453112 (M.P.); India; E-mail: *pathak.apurva@gmail.com*

Materials and Method

The present aerobiological investigation was carried out at three places of the city Jabalpur (Latitude: 23.2; Longitude: 79.95; Altitude: 391.). The Ranjhi, Deotal and Rani Durgawati University campus which are five kilometer distantly placed are chosen for this investigation. The open ground area of these sites have been chosen in order to avoid any kind of influence arises in due course of community activities. Duplicate samples in time and place must always be taken therefore, aerobiological sampling has been done in duplicate and fortnightly in order to cover all the major season for a period of two years. The metrological data were collected from weather station Jabalpur.

Isolation from Air

The Andersen two-stage viable (microbial) particle sampler (2-STG) has been developed for monitoring bioaerosols. It is a multi-orifice, cascade impactor with 400 holes per stage, drawing air at a flow rate of 28.3 L min^{-1}. The different stages separate the airborne particles in size fractions (Andersen, 1958, 1966).

For this study air sampling was done on Sabouraud Dextrose Agar Medium at one meter height from the ground. The sampler was operated for ten minutes at the site. For enumeration and identification of total viable type of fungal population present in air, the SDA medium plate kept on upper stage of the sampler. Processing of samples was done by incubating plates into B.O.D. incubator at 27 $\pm$ 2 ^{0}C for 2 to 7 days. The colonies developed were counted, isolated and identified for fungal colonies with the help of standard literature (Clements and Shear, 1954; Ellis, 1971; Subramanian, 1971; Barnett and Hunt, 1972; Dennis and Cramer, 1978).

Statistical Analysis

The number of samples collected will influence the precision of the exposure estimate and the associated confidence limits. In order to analyze the effect of various environmental factors on the forecasting of airborne fungal population, Seasonal Autoregressive Moving Average, Multiple Linear Regression Analysis and Neural Network Application was done by using NCSS, SPSS 12.0 for window and Alyuda Forecaster XL v2.3 (Galan *et al*, 2001; Laaidi *et al.*, 2003; Katyal *et al.*1997, Luk *et al.*, 2000, Moseholm *et al.*, 1987). Outputs of the forecasted / predicted data were further analyzed by using correlation coefficient and $ø^2$ (Chi-Square) test. All the data were presented in the form of table and figures.

Results

Fungal spores are playing an important role in etiology of allergenic disorder. The quantity and quality of airborne spores are largely influenced by environmental factors like temperature, humidity and seasonal variation. Correlation and variance analysis shows a stronger relationship between the quantity of spores and the environmental parameters (Verma and George, 2000).

In this study, highest concentrations of spores were calculated for the month of July by all the methods applied, non-viable spore counts of fungal spores are not recorded it by the previous workers (Verma, 1990; Hollins *et al.*, 2004). SARIMA model forecasted and Linear Regression Analysis were predicted the lowest concentration of air borne fungal spores for the month of May. According to Neural Network Application, lowest concentration of fungal spore might be observed in the month of March (Table No. 4 & Figure.1).

The Mean value and the Standard Deviation of Linear Regression Model were calculated lowest in the entire model studied, while the range of it is highest. Highest Mean value and Std. Deviation with lowest range observed for SARIMA model (Table No. 1).

The ø^2 (Chi-Square) test revealed that all the data created after the calculation is 100% significant, all the variable tested have expected frequencies less than 5 (Table No. 2).Inter-item Correlation Matrix shows that the deference between two extreme Models is not more than 7%, the Linear Regression Model is very close to Neural Network Application. (Table No. 3)

Table 1 : Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Forecast SARIMA	24	94.14	118.31	19.52	388.24
Forecast NNA	24	92.03	101.18	18.94	395.19
Predicted Value L R	24	85.87	96.27	17	398

Table 2 : Test Statistics

	Forecast SARIMA	Forecast NNA	Predicted Value L R
Chi-Square(a)	0	0	0
df	23	23	23
Asymp. Sig.	1	1	1

(a). 24 cells (100.0%) have expected frequencies less than 5. The minimum expected cell frequency is 1.0.

Table 3 : Inter-Item Correlation Matrix

	Forecast SARIMA	Forecast NNA	Predicted Value L R
Forecast SARIMA	1	0.93	0.926
Forecast NNA	0.93	1	0.989
Predicted Value L R	0.926	0.989	1

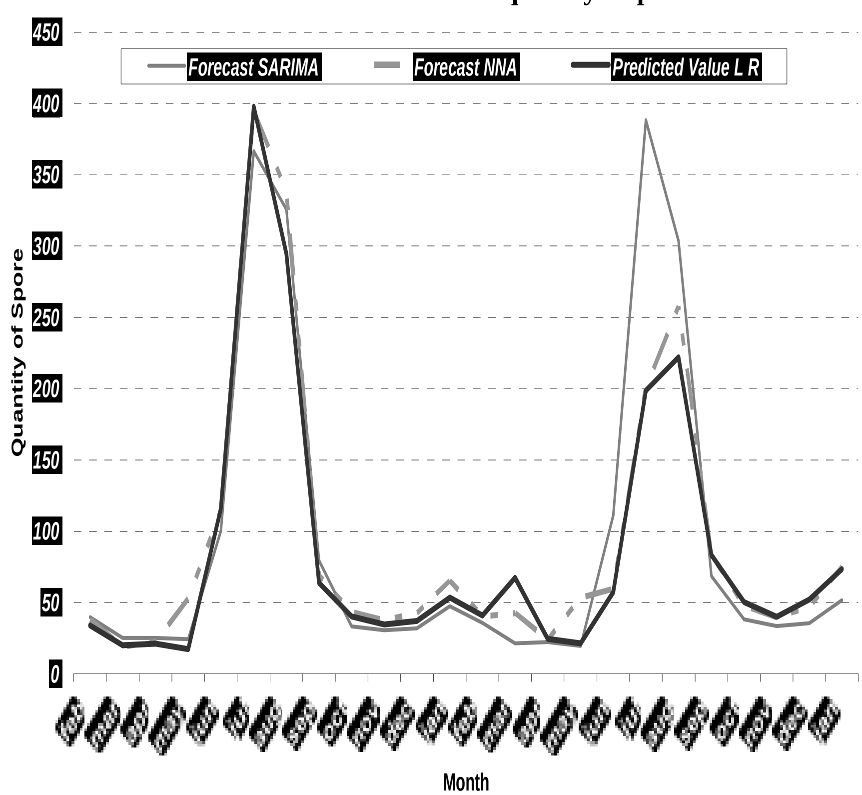
The covariance matrix is calculated and used in the analysis.

Table 4 : The table of the comparative analysis of Forecasting

Input Metrological parameters						**Output** Number of fungal spores.		
Season (Months)	Temperature max.	Temperature min.	Humidity morning	Humidity evening	rainfall	Forecast NNA	Forecast SARIMA	Predicted Value Lin. Reg.
FEB	27.2	12.5	64	33	5	38.08985	39.84259	34
MAR	35.3	18	40	20	0	18.94036	25.12859	20
APR	37	22.5	36	20	7.6	22.37466	25.26264	21.2
MAY	42.2	28	30	16	6.3	52.614	24.33178	17.43
JUN	35.6	26.5	71	57	219.6	114.5523	100.159	115.87
JUL	28.6	23.7	90	83	1019.9	395.1889	366.4896	397.63
AUG	28	23.4	91	83	709.2	338.5297	325.5435	294.4

SEP	29.8	23	80	61	50.4	69.13113	79.50752	63.8
OCT	31.8	19.6	71	46	3.9	43.19103	33.40152	40.3
NOV	28.9	14.2	59	45	0	37.71694	30.60402	34.67
DEC	26.1	9.9	64	47	0	42.69808	31.87143	37
JAN	21.7	10.1	76	51	33.1	65.19802	47.52415	53.37
FEB	27.7	13.4	69	38	16	40.28559	35.76258	41
MAR	30.3	17.7	60	40	102	42.66368	21.54719	67.33
APR	37.4	22.4	42	22	9.3	23.81179	22.42565	24.43
MAY	39.8	28.1	40	23	0.7	53.35419	19.51449	21.23
JUN	40.4	28.2	53	42	77	59.73846	111.1282	57.33
JUL	32.2	25	87	71	438.1	200.541	388.2384	198.7
AUG	29.8	24	90	81	495	257.5274	303.7957	222
SEP	32.7	23.6	81	68	100.9	83.65266	68.53873	83.3
OCT	32.3	20.4	71	52	28	47.57305	38.2189	50.33
NOV	28.7	14.3	68	52	0	39.39484	33.44109	40
DEC	25.9	12.3	74	58	24.6	47.07331	35.45285	52.2
JAN	23	12.4	87	63	70.4	74.88763	51.60417	73.47

Table 4 : forecasted/Predicted quantity of spore



Discussion

The differences observed in these models are due to the pattern by which the independent variables were taken into the account for the computation. The degrees of weightage given to different environmental factors varied with the models under consideration. In Linear Regression Analysis all the factors given equal importance, have same value during the calculation. SARIMA model is entirely dependent upon the seasonality the algorithmic calculation based on seasonality. In Neural Network Application, the weightage divided by the artificial intelligent method, in the method 75% weightage given to the season and rest of this distributed accordingly.

The prediction level of all the models are reported during this study is more than 95%, using co-evaluative neural network models. Sanchez-Mesa *et al.* (2002) obtained the best forecasts for about 90% "good" classification. According to them, this new system based on neural network models is a step toward the automation of the pollen forecast process. Many of the workers used Multiple Linear model (Galan *et al.*, 2001), Partial Linear Regression (Laaidi *et al.*, 2003), ARIMA (Galan *et al*, 2001; Box and Jenkins, 1976), and NNA (Katyal *et al.*, 1997, Luk *et.al.*,2000, Moseholm *et.al.*,1987) to predict the pollen incidence reported 90% of accuracy.

Conclusion

Thus, for spore forecasting of short duration all the three models were effective and for long-term analysis, Linear Regression Models can be used. In regression tasks, the output values of neural networks are number values. Therefore, the quality of network functioning can be evaluated with the help of traditional statistical values, this make the output more accurate in NNA (Patterson, 1996).

References

Alyuda Forecaster XL2.3. (2006): Alyuda Research, Inc. www.alyuda.com.

Andersen A. A. (1958): A new sampler for collection, seizing & enumeration of viable airborne bacteria. *J. Bacterial.* **76,** 471 - 484.

Andersen A. A. (1966): A sampler for respiratory health hazards assessment. *Am. Ind Hyg. Assoc. J.* **27**, 260 - 265.

Atlas R.M. (1984): Microbiology, Fundamentals and Applications Macmillan Publishing Co., New York, London.

Bernett H.L. and Hunte B.B. (1972): Illustrated genera of imperfect fungi. Burgess Publishing Company, Minneapolis, Minnesota.

Box G. and Jenkins G. (1976): Time series analysis forecasting and control. Holden-Day, San Francisco.

Clements F.E. and Sher C.L. (1954): The genera of fungi. Heffner Publishing Company, New York. pp.496.

Dennis R.W.G. and Cramer J. (1978): British Ascomycetes. Inder A.R. Ganter Verlag. Kammanditgeselishaft, F.I. 9490, VADUZ, Germany. pp. 585.

Ellis M.B. (1971): Dematiaceous Hyphomycetes. Kew England, Common Wealth Mycological Institute. pp. 317.

Galan C., Carinanos P., Garcia-Mazo H., Alcazar P. and Dominguez-Vilches E. (2001): Model for forecasting Olea europaea L. airborne pollen in South-West Andalusia, Spain. *Int J. Biometeorol.*, 45(2), 59-63.

Hollins P. D., Kettlewell P. S., Atkinson M. D., Stephenson D. B., Corden J. M., Millington W.M. and Mullins J. (2004). Relationships between airborne fungal spore concentration of Cladosporium and the summer climate at two sites in Britain. *Int J Biometeorol.*, **48(3)**, 137-141.

Jones A.M. and Harrison R.M. (2003): The effects of meteorological factors on atmospheric bioaerosol concentrations – a review. *Science of the Total Environment.*, **326(1-3)**, 151-180.

Katyal R.K., Zhang Y., Jones R.H. and Dyer P. D. (1997): Atmospheric mold spore counts in relation to meteorological parameters. *Int. J. Biometeorol*., **41**, 17-22.

Laaidi M., Thibaudon M. and Besancenot J.P. (2003): Two statistical approaches to forecasting the start and duration of the pollen season of Ambrosia in the area of Lyon (France). *Int J Biometeorol*., **48(2)**, 65-73.

Luk K.C., Ball J.E. and Sharma A. (2000): A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting, *Journal of Hydrology,* **227**, 56-65.

Moseholm L., Weeke E.R. and Petersen B.N. (1987): Forecast of pollen concentrations of poaceae (grasses) in the air by time series analysis, *Pollen et Spores*, Vol. XXIX, No. 2-3, 305-322.

Patterson D.W. (1996): Artificial Neural Networks: theory and applications, Simon and Schuster, Singapore, pp. 477.

Sanchez-Mesa J.A., Galan C., Martinez-Heras J.A. and Hervas-Martinez C. (2002): The use of a neural network to forecast daily grass pollen concentration in a Mediterranean region: the southern part of the Iberian Peninsula. *Clin Exp Allergy.*, **32(11)**,1606-1612.

SPSS 12.0 for window© (2000): The apache software foundation. Release 1989-2003.

NCSS Software (2006): On www.ncss.com. NCSS, 329 North 1000 East, Kaysville, Utah, 84037.

Subramanian C.V. (1971): Hyphomycetes. New Delhi ICMR. pp. 930.

Verma K.S. and George A.M. (2000): Seasonal and diurnal variation of airborne fungal spores in Jabalpur. *Vasundhara the Earth*, Vol **1**, **2**, pp 23-26.

Verma K.S. (1990): Biocomponents in the atmosphere of Jabalpur. *Ind.J.Aerobiol*, **3 (1&2)**, pp. 48-51.