

Wind speed prediction using statistical regression and neural network

MAKARAND A KULKARNI^{1,*}, SUNIL PATIL², G V RAMA³ and P N SEN¹

¹*Department of Atmospheric and Space Sciences, University of Pune, Pune 411 007, India.*

²*Department of Physics, University of Pune, Pune 411 007, India.*

³*Meteorological Facility, Sriharikota 524 124, Nellore (Dt), A.P., India.*

**e-mail: makarand_kulkarni@hotmail.com*

Prediction of wind speed in the atmospheric boundary layer is important for wind energy assessment, satellite launching and aviation, etc. There are a few techniques available for wind speed prediction, which require a minimum number of input parameters. Four different statistical techniques, viz., curve fitting, Auto Regressive Integrated Moving Average Model (ARIMA), extrapolation with periodic function and Artificial Neural Networks (ANN) are employed to predict wind speed. These methods require wind speeds of previous hours as input. It has been found that wind speed can be predicted with a reasonable degree of accuracy using two methods, viz., extrapolation using periodic curve fitting and ANN and the other two methods are not very useful.

1. Introduction

Prediction of wind speed at the surface or near the surface is essential in many areas of science and technology, e.g., wind energy generation, aviation, space vehicle launching, weather forecasting, and agro-meteorology. The prediction of wind speed to a desired level of accuracy using least number of input parameters is always appreciated.

In order to issue weather forecasts the meteorologists generally solve some prognostic equations. These prognostic equations are either basic equations like the momentum equations, equation of mass continuity, thermodynamic energy equation, equation of moisture continuity or some derived equations like vorticity equations, balance equations, etc. These equations are solved numerically and are known as numerical modelling. But this type of modelling requires data of various weather parameters at equally spaced grid points. The various data required are wind speed, wind direction, temperature, atmospheric pressure, geopotential, etc. and some derived parameters. On the other hand, we have a long series of data recorded at

discrete points. We have data at a tower of 100 meter height and the only instruments available on the platform are cup anemometers for recording wind speed and direction. With these data no numerical modelling forecasts can be issued. In such situations some empirical methods like curve fitting to available data and extrapolation are well known techniques for prediction algorithms. Several studies have been reported for wind speed prediction using statistical and empirical techniques (e.g., Katz and Skaggs 1981; Mohandes *et al* 1998; Song 2000 and Zhang 2003, etc.). The objective of this paper is to present the study carried out for wind speed prediction using different statistical methods.

2. Data used for present study

Ten years of data (for the period 1992 to 2001) containing wind speed and direction recorded by cup anemometer at a level of 100 m from ground are used in the present study. These data were collected at an Indian coastal station, Sriharikota (latitude 13.70°N and longitude

Keywords. Wind speed prediction; artificial neural network; curve fitting; ARIMA.

Table 1. Errors in prediction of zonal component of wind speed for the month of January using Algebraic Curve Fitting.

| Degrees of the curve | One-step without synoptic system (m/s) | Two-step without synoptic system (m/s) | One-step with synoptic system (m/s) | Two-step with synoptic system (m/s) |
|----------------------|--|--|-------------------------------------|-------------------------------------|
| 1 | 2.4711 | 2.2875 | 1.7920 | 1.9032 |
| 2 | 2.4637 | 2.2834 | 1.5099 | 1.1627 |
| 3 | 2.4007 | 2.2777 | 1.3141 | 1.0857 |
| 4 | 2.4007 | 2.2726 | 1.0154 | 0.8314 |
| 5 | 2.4002 | 2.2679 | 0.9380 | 0.8098 |
| 6 | 2.4002 | 2.2679 | 0.8933 | 0.6243 |
| 7 | 2.4002 | 2.2674 | 0.8863 | 0.5945 |

80.20°E). Sriharikota is the space launching station of the Indian Space Research Organization, Government of India. The data recorded are 5 minutes averaged data. These data are further resolved in zonal component (u) and meridional component (v) because direct averaging of a vector quantity has no physical meaning. Hourly averaged datasets are then constructed by averaging u and v components of velocity to compensate the fluctuations.

3. Statistical techniques and approach

3.1 Algebraic curve fitting

The hourly averaged time series with u and v components of wind speed is treated as independent variable while two time series generated by giving lag of 24 (u_{24} and v_{24}) and 48 (u_{48} and v_{48}) hours are treated as dependent variables. The polynomials of degrees ranging from 1 to 7 are fitted for u and v to corresponding u_{24} and v_{24} and the coefficients ($A_{i,24}$) are obtained for each month using Least Square method. The same procedure is repeated for u_{48} and v_{48} . The cases when some synoptic disturbances were present in the neighbourhood of the station have been studied separately. The two-step prediction is also tried for 48 hours by using the predicted values for 24 hours as input for the prediction of the next 24 hours.

3.2 Auto Regressive Integrated Moving Average Model (ARIMA)

The well known ARIMA (p, d, q) model (Box and Jenkins 1976) was tried for the forecast of wind speed. The parameter p signifies the amount of backtracking needed to estimate the next value and it is calculated by taking the first significant partial autocorrelation coefficients. The parameter q signifies the dependence of forecast on the difference between the actual and estimated values by autoregressive process and is calculated

by the first significant autocorrelation coefficients. The parameter d is the difference used to remove the non-stationarity in the trend type from original data, taken as zero in the present study, Box and Jenkins (1976) suggested that the repeated differencing and fitting of ARIMA model to the differenced series is suitable for economical forecasting, but for the time series, important for the physical sciences like meteorology, this operation ($d \neq 0$) is inappropriate (Katz and Skaggs 1981; Dahale and Singh 1993).

The prediction of wind speed is done with a complex boot strapping method. This predicts wind speed at an instant (say t) using wind speed data of a previous time (say $t-1$) and to predict at $t+1$ it uses the latest predicted value (at t) and so on. The coefficients are recalculated for every step.

3.3 Extrapolation using periodic curve fitting

When wind speeds of successive five days are plotted against local time they show a periodic behaviour (figure 1). From the figure it is clear that

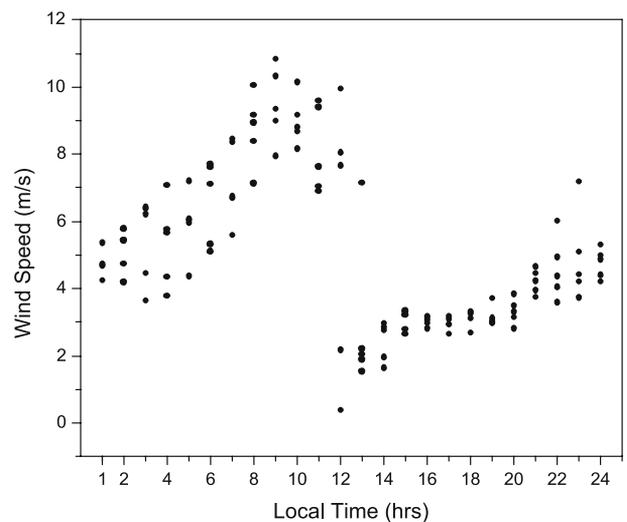


Figure 1. Wind speed as function of time.

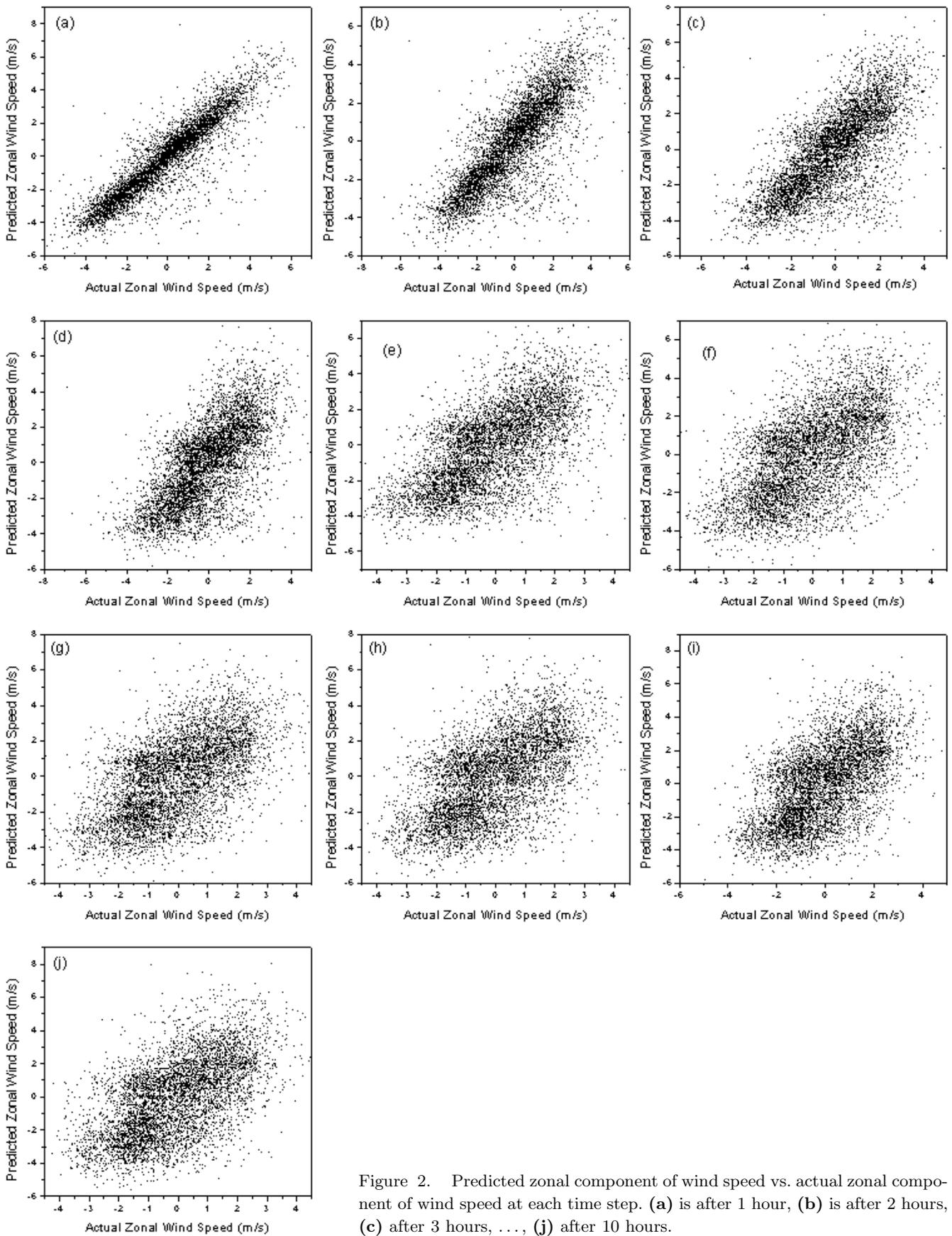


Figure 2. Predicted zonal component of wind speed vs. actual zonal component of wind speed at each time step. (a) is after 1 hour, (b) is after 2 hours, (c) after 3 hours, . . . , (j) after 10 hours.

the wind speeds show some periodic/oscillatory behaviour. It is therefore appropriate that an empirical formula containing periodic function, like *sine* can be used for prediction of wind speed. Accordingly an empirical formula of wind speed at an instant of time t is,

$$Y_t = A + BY_{(t-1)} \frac{\sin(ct + d)}{(ct + d)}, \quad (1)$$

where Y_{t-1} is the wind speed of previous hour and A, B, c and d are constants.

To get the constants in equation (1), the usual numerical method of least square for curve fitting

has been used. But the partial derivatives of constructed function are nonlinear in nature. So it becomes very difficult to obtain the values of constants, numerically. The equation (1) can be simplified by using power series approximation of *sine* function in t . The approximated formula is as follows:

$$Y_t = a_0 + a_1Y_{(t-1)} + a_2Y_{(t-1)}^2 + a_4Y_{(t-1)}^4 + \dots + a_nY_{(t-1)}^n. \quad (2)$$

The coefficients $a_0, a_1, a_2, \dots, a_n$ can be easily obtained by using Least Square method. A simple

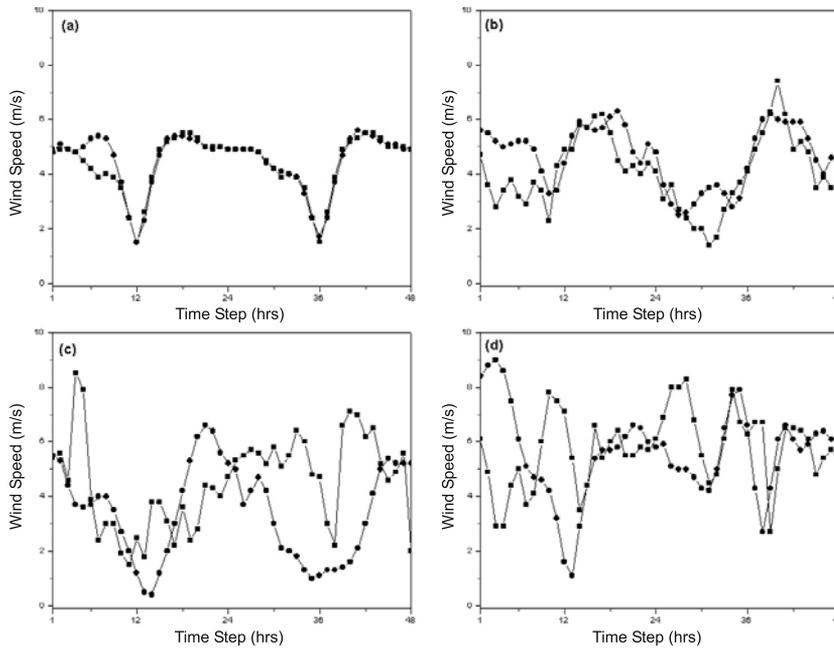


Figure 3. Prediction of wind speed using the periodic function for four random cases, and the last two days data are considered for input. Circles represent the predicted wind speed and squares represent the actual wind speed.

Table 2. Error (m/s) in forecasting wind speed using the periodic curve fitting.

| Using previous data of days | R.M.S. error in | | | |
|----------------------------------|-------------------------|------------------------------|--------------------|----------------------|
| | Zonal (u) component | Meridional (v) component | $\sqrt{u^2 + v^2}$ | Scalar wind velocity |
| 15 days | 2.11848 | 2.26790 | 2.02918 | 1.86848 |
| 12 days | 2.13730 | 2.28467 | 2.03539 | 1.87167 |
| 10 days | 2.14763 | 2.30938 | 2.03564 | 1.87281 |
| 7 days | 2.12947 | 2.30130 | 2.00312 | 1.83688 |
| 5 days | 2.09736 | 2.30254 | 1.97959 | 1.78151 |
| 2 days | 2.18937 | 2.33566 | 1.89914 | 1.82290 |
| After averaging above six curves | 2.98825 | 3.54873 | 2.37229 | 2.72853 |

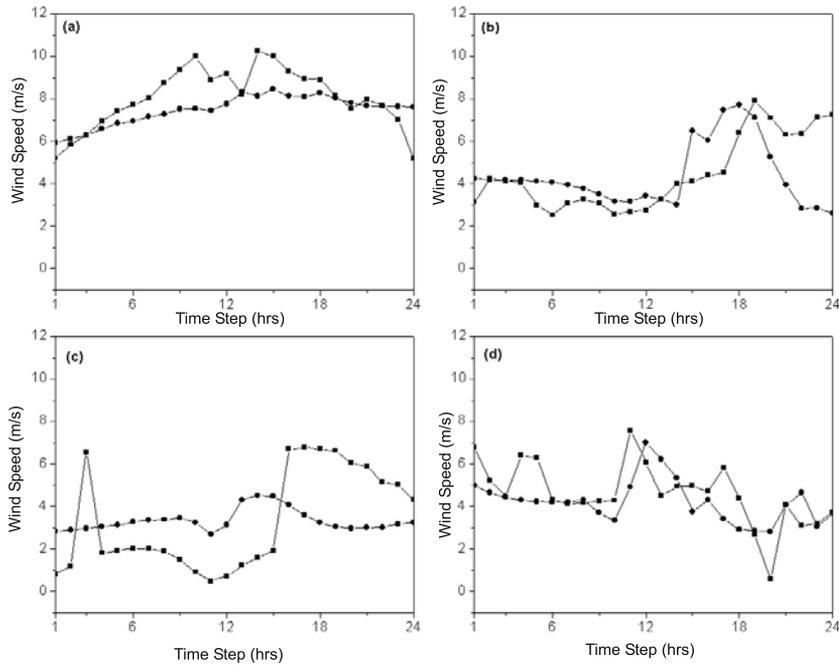


Figure 4. Prediction of wind speed using neural network for four random cases. Symbols are the same as in figure 3.

Table 3. Prediction of zonal component of wind speed using neural networks.

| Time step | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|----------------------------|--------|-------|-------|--------|------|------|-----|
| Actual wind speed (m/s) | 0.948 | 0.62 | 1.35 | 1.22 | 1.84 | 0.87 | 1 |
| Predicted wind speed (m/s) | 0.0324 | 0.336 | 1.048 | 1.6256 | 2.17 | 2.87 | 3.2 |

boot strapping method is used for the prediction of wind speed with same value of coefficients throughout the computation. The input data are varied from 15 days, 12 days, 10 days, 7 days, 5 days and 2 days, as periodicities like quasi-bi-weekly, weekly and synoptic scales are generally observed (Asnani 2005) in the atmosphere. The six different curves and consequently six sets of coefficients are available. Ensemble forecast by weighted averaging of all the six curves was also tried.

3.4 Neural networks

The feed forward network for simplicity and error back propagation method for learning with sigmoid as an activation function is used in the present study. The software used is Stuttgart neural network simulator. The normalized input data are used for training the network. Normalization is done by the following formula,

$$N = \frac{\text{Windspeed} - \min(\text{data})}{\max(\text{data}) - \min(\text{data})}. \quad (3)$$

The network (network 1) is designed by trial and error method; 6 neurons at the input, 3 neurons in the hidden layer and one at the output. A single pattern contains six inputs, having the previous 5 hours' wind speed and time (t) and one output, the normalized wind speed at time t . It has been observed that the inputs are highly correlated among themselves. Therefore the inputs are orthogonalized by the principal component analysis.

Another network (network 2) is designed with the first, second and third day wind speed (Y) data divided in 48 patterns of Y_t and t as the inputs, i.e., 2 neurons are at input, 2 neurons in the hidden layer and Y_{t+24} at the output to train the network. Then wind speed data of the third day are used for the prediction of wind speed of the fourth day, i.e., the validation.

4. Results and discussion

The Root Mean Square Error (RMSE) in prediction of zonal component of wind speed has been

computed for all the months. In this paper, the RMSE for the month of January as an example has been presented in table 1. The method used here is Algebraic Curve Fitting. The RMSE observed in prediction of wind speed using two steps (each of 24 hours) is less than that of single step (48 hours). The two-step process magnifies the step results in more accurate prediction than that of single step. It has been observed that the RMSE in the prediction of wind speed is less when there was some synoptic disturbance in the neighbourhood of the station. The number of cases is small when there was synoptic disturbance in the vicinity of the station. This may be the reason for small RMSE. The standard deviation of the scalar velocity component is 2.8 m/s. Results obtained from the second technique (ARIMA) are plotted in figure 2 (a, b, c, . . . , j) as predicted wind speed versus actual wind speed. Ideally it should be a straight line, which is more or less observed in the first few steps. The error increases with increase in steps as the complex boot strapping is involved in prediction where cumulative effect of error is reflected in the prediction of the subsequent time steps. It has been observed that the 52% predicted data contain error less than 1.5 m/s and RMSE is 2.12 m/s.

The RMSE appeared in prediction of wind speed by the third technique (extrapolation using periodic curve fitting) is given in table 2. In 62% and 70% cases error in the prediction is less than 1.5 m/s by using averaged coefficients from all six curves and only using coefficients from the previous two-day data respectively. Since the dataset is very large, the qualitative comparison between the actual and predicted wind speeds using the extrapolation technique has been presented for four randomly chosen datasets and are given in figure 3. The abscissa of figure 3 is the time step used for issuing forecast and the ordinate is the wind speed. It can be seen that the wind speed predicted is in reasonably good agreement with the actual ones. It also shows a similar trend in most of the cases.

Table 3 gives the actual zonal component of wind speed and the prediction at each time step by using the fourth technique (Neural Network, network 1). The error increases with time step. It suggests that boot strapping method is not applicable to neural networks. Though neural networks get nonlinear trend in the data series, the concept of neural networks is based on pattern recognition as the simple boot strapping doesn't allow retraining including predicted pattern. By using network 2, it is observed that the total RMSE in predicting the wind is 2.10 m/s and more than 65% predicted data have error less than 1.5 m/s. As in figure 3, the predicted and the actual values of wind speed for the four randomly chosen datasets are plotted

in figure 4 (a, b, c, d) for comparison. It is observed that the predicted values show good agreement with the actual one.

Algebraic curve fitting technique generally gives erroneous results as the curves represent the nature of wind speed in previous years and expect similar nature in the subsequent year in corresponding months. The RMSE using extrapolation technique is less than that of the ARIMA. The extrapolation technique assumes that the present wind speed is a function of both previous wind speed and local time whereas ARIMA assumes that the current wind speed is function of only previous wind speeds. An appropriate choice of function gives reasonable accuracy in the prediction of wind speed. Wind speed is also affected by air density, pressure gradient, temperature, vertical wind profile and season, etc. Under the complicated influence of these parameters, selection of an appropriate function for a regression model and its fitting becomes difficult. The neural network gives similar results as that of extrapolation, only with previous knowledge of wind speed.

5. Conclusions

It has been found that wind speed can be successfully predicted using only previous knowledge of the wind speed by regression techniques and neural network. The prediction of wind speed improves if wind speed is assumed to be a function of previous wind speed and local time. Boot strapping method is not useful for prediction of wind speed with neural networks as neural network is based on pattern recognition. It is well known that the wind speed is a function of several parameters like pressure gradient, air temperature, orography, etc. But the only data available for this study were the wind speed and direction. Therefore, no dynamical methods could be developed for the prediction of wind speed for the station.

Acknowledgements

We are thankful to the anonymous reviewers for their valuable comments. These comments have made significant improvements in the paper. We are also thankful to the Indian Space Research Organization for providing funds for this study.

References

- Asnani G C 2005 *Tropical Meteorology*; Pune, India, 2 5–77.
- Baillie R T 1996 Long Memory Processes and Fractional Integration in Economics; *Journal of Econometrics* **73** 5–59.

- Box G E P and Jenkins G M 1976 Time Series Analysis Forecasting and Control; Holden-Day, Oakland, 575.
- Dahale S D and Singh S V 1993 Modelling of Indian Monsoon Rainfall Series by Univariate Box-Jenkins Type of Models; *Adv. Atmos. Sci.* **10(2)** 211–220.
- Katz R W and Skaggs R H 1981 On use of autoregressive-moving average process to the meteorological time series; *Mon. Weather. Rev.* **109(2)** 479–484.
- Mohandes A M, Rehman S and Halawani T O 1998 A neural network approach for wind speed prediction; *Renewable Energy* **13(3)** 345–354.
- Song Y D 2000 A New Approach for Wind Speed Prediction; *Wind Engineering* **24(1)** 35–47.
- Zhang G P 2003 Time series forecasting using a hybrid ARIMA and neural network model; *Neurocomputing* **50** 159–175.

MS received 27 September 2007; revised 24 April 2008; accepted 8 May 2008