International Journal of Combined Research & Development (IJCRD)
eISSN:2321-225X;pISSN:2321-2241 Volume: 3; Issue: 5; November -2014

High Leverage Points and Outliers in Application of Curve fitting and Surface fitting tools for Wind Speed Prediction

Krishna Kumar D[1], Prashanth P [2], Arjun N N [3], Sumanth Pareekshit V [4]
[1,2,3,4] Student , R V College of Engineering , Bangalore ,
Karnataka , India

Abstract: **Climate change is generally accepted as being the greatest environmental challenge facing our world today. Together with the need to ensure long-term security of energy supply, it imposes an obligation on all of us to consider ways of reducing our carbon footprint and sourcing more of our energy from renewable resource. Wind energy is one such source and this paper presents a method to predict the speed of wind, on which the wind energy generated, depends more efficiently and hence avoid both costly overproduction and underproduction. This can be achieved by using the statistical toolbox present in MATLAB wherein data in large numbers are collected, analyzed in the form of matrices for functional relationship. The results obtained are then compared with the actual values available for validation.**

Keywords- multivariate regression; time series; MATLAB; wind speed; prediction; models

I. INTRODUCTION

Intermittency of wind is the biggest challenge to implementing wind-energy a reliable autonomous source of electric power. Energy crisis, global warming, depletion of fossil fuels are the major factors looming the world today. Adequate utilization of renewable energy sources like wind, solar, biomass etc. proves to be the only alternative to overcome these problems. The wind power industry is very promising and it is necessary for the load dispatch centers that the wind farm power prediction be exact. Wind energy is directly dependent on the wind speed available at that location and this speed is extremely erratic, i.e. variations are very large. Hence, a model is needed for accurate prediction of wind speed.

Prediction models facilitate the integration of wind power with the grid and also forewarn the power system operators [1] by giving weather alerts. They also aid in power system operations planning, unit commitment and economic dispatch. Developing forecasting models is an overwhelming task, due to the random and stochastic nature of wind. Analysis of data or Data Analytics a process of inspecting, cleaning, transforming, and modeling data with the goal of discovering useful information, suggesting conclusions, and supporting decision making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, in different business, science, and social science domains. The main parameters

used in the process are atmospheric pressure, relative humidity and average temperature. The minute to minute values for the quantities were obtained from a weather station and analyzed. In our analysis wind speed was the parameter to be determined and it was as said earlier exhibited some correlation with Air temperature, Relative humidity and most importantly pressure which are our independent variables.

Various data analysis tools present in MATLAB were used for analysis. MATLAB deals with data in the form of matrices. The tools available in MATLAB for data analysis include functions like regress(), nlinfit() etc. along with tools like cftool[curve fitting tool] and sftool [surface fitting tool]. The four functions mentioned above were the ones used for multivariate analysis. Time series model was also implemented in MATLAB. In time series modeling, time was defined as a new variable along with temperature, pressure and relative humidity. The present value was predicted with the help of previous value and the error.

The models so arrived can be used for short term wind speed prediction and since wind speed is very uncertain the same model cannot however predict long term counterparts within acceptable limits. The models find their application mainly in state load regional dispatch center where wind speed from the model can be used to estimate wind energy that could be available in future can be predicted beforehand and better dispatch methodologies can be adopted [5].

II. PARAMETERS USED

There various variables on which wind speed depends upon and this paper present the statistical relationship among these variables using MATLAB which impact the wind speed in a particular region. The variables used are temperature, relative humidity and atmospheric pressure.

A. ***Atmospheric Pressure*** It is the force per unit area exerted on a surface by the weight of air above that surface in the atmosphere of Earth. In most circumstances atmospheric pressure is closely approximated by the hydrostatic pressure caused by the weight of air above the measurement point. On a given plane, low-pressure areas have less atmospheric mass above their location, whereas highpressure areas have more atmospheric mass above their location. Likewise, as elevation increases, there is less overlying atmospheric mass, so that atmospheric pressure decreases with increasing elevation. The unit of atmospheric pressure used in this paper is consistent throughout and it is mm of hg.

B. *Relative Humidity*

It is the ratio of the partial pressure of water vapor in an air-water mixture to the saturated vapor pressure of water at a prescribed temperature. The relative humidity of air depends on temperature and the pressure of the system of interest. There is a considerable degree of difference between the relative humidity measured indoors and that measured outdoors in regions such as wind farms and those at altitudes of the shaft locations. As is known, there is no unit for relative humidity and the numbers used in this paper are expressed as a percentage.

C. *Temperature*

A temperature is a numerical measure of hot and cold. Its measurement is by

detection of heat radiation or particle velocity or kinetic energy, or by the bulk behaviour of a thermometric material. The kinetic theory indicates the absolute temperature as proportional to the average kinetic energy of the random microscopic motions of their constituent microscopic particles such as electrons, atoms, and molecules. The basic unit of temperature in the International System of Units (SI) is the Kelvin. It has the symbol K.

For everyday applications, it is often convenient to use the Celsius scale, in which 0°C corresponds very closely to the freezing point of water and 100°C is its boiling point at sea level.

D. *Wind Speed or Wind Velocity*

The equations are an exception to the prescribed is a fundamental atmospheric rate. Wind speed is caused by air moving from high pressure to low pressure. Wind speed affects weather forecasting, aircraft and maritime operations, construction projects, growth and metabolism rate of many plant species, and countless other implications. It is now commonly measured with an anemometer but can also be classified using the older Beaufort scale which is based on people's observation of specifically defined wind effects. The unit of wind speed used in this paper is consistently meters per second.

As stated earlier, the minute to minute values of the above parameters were obtained from a meteorological station. Now, using data analytics and statistics tools, they have to be made sense of in a way to fit a model to obtain predicted values.

III. MULTIVARIATE LINEAR REGRESSION

It is a generalization of linear regression by considering more than one independent variable, and a specific case of general linear models formed by restricting the number of dependent variables to one. Multivariate linear regression is often found to be more efficient and useful since it considers the effects of all the parameters that may have a significant effect on the dependent variable present.

Wind speed is the dependent variable and it is dependent on three quantities, i.e. Pressure, relative humidity and temperature. All three were used to develop models of regression to predict wind speed. The regress(), nlinfit(), cftool, sftool functions were used to fit the values into a system or a

model.

A. *regress()*

The syntax of the regress function is: b = regress(y,X)

b = regress(y,X) returns a p-by-1 vector b of coefficient estimates for a multilinear regression of the responses in y on the predictors in X. X is an n-by-p matrix of p predictors at each of n observations. y is an n-by-1 vector of observed responses[2]. Here 'X' is a matrix with the predictors, with or without intercept.

B. *nlinfit()*

The nlinfit is the acronym for Non-Linear fit of the curve and it is the function for nonlinear regression. The syntax for the command is: beta = nlinfit(X, y, Model, beta0)

beta = nlinfit(X, y, Model, beta0) returns a vector of estimated coefficients for the nonlinear regression of the responses in y on the predictors in X using the model specified by Model. The coefficients are estimated using iterative least squares estimation[3], with initial values specified by beta0. Here the model is

some fitting function to be defined beforehand in the command window of the MATLAB which better fits the input predictors with output.

C.*cftool*

The cftool is a curve fitting application present in MATLAB. The syntax for this command is: cftool

cftool(x, y) creates a surface fit[4]for the input x and y as the output. All the vectors x, y must be numeric and must be of same size, in the matrix form. cftool is a two dimensional data fitting tool, hence it can take only one predictor as an input.

D.*Sftool*

The sftool is a surface fitting application present in MATLAB. The syntax for this command is: sftool. Unlike cftool , sftool is a three dimensional data fitting tool[5] with two input predictors and an output. Hence, it fits a surface for the given data.

IV.METHODOLOGY

A.*RAW Data Collection*

The data was collected from ARS Bagalkot, Karnataka. The data was collected by automatically measuring parameters like Air Temperature, Relative Humidity, pressure, wind speed and the wind direction. The Data consisted of values for different altitudes. For analysis one particular altitude values were considered. The choice is made based on the fact that wind velocity is higher in upper strata of atmosphere thereby having more potential to generate wind energy[6].

B.*Data matrices*

The data with parameters like air temperature, pressure, relative humidity, wind speed were defined in the command wind of the MATLAB in the form of matrix imported from Microsoft excel. Then the individual physical parameters were extracted from the matrix and were assigned with a specific name for the purpose of identification.

C.*Implementation of the models*

The correlation between the physical parameters was obtained by using the stats data toolbox present in MATLAB.

The functions used as stated earlier were regress(), nlinfit(), cftool, sftool. Four models were implemented in MATLAB using as many functions. The four models implemented were multivariate regression models. A time series model was also implemented for the purpose tracking the peak overshoots and undershoots. Here time was defined as a variable along with temperature, pressure and relative humidity and the previous value of the wind speed was used as the input. regress() function was used for the determination of the error equation.

V. MODELS

A.*Model regress:*

Using the MATLAB function regress(), the model regress was implemented.

B=regress(y, X)

Where y = column array of output values

X = multi-column array of input values

B = Co-efficient matrix for modified multivariate equation

The parameters taken as inputs were combinations of the measured quantities (Pressure, Temperature and Relative Humidity). This command in Matlab

permits the selection of product terms of independent variables also. Hence, a number of models have been tested and the best model has been chosen. The plot is shown in Fig 1. It was constructed by taken into consideration the inputs of temperature, pressure, relative humidity, product of temperature and relative humidity, product of temperature and pressure. Co-efficients of other combinations of input variables were set to zero.

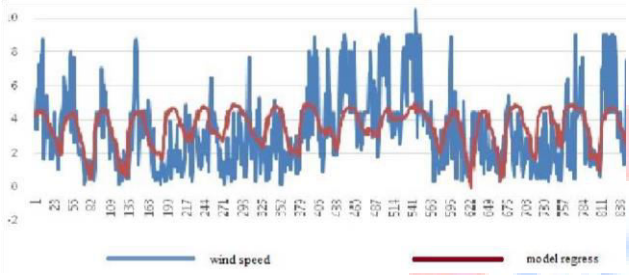

Fig 1

The following is the output obtained after modelling:

Wind Speed = -30.079*Temperature + 0.1257*Relative Humidity - 0.0353*Pressure - 0.0061*Temperature*Relative Humidity + 0.0439*Temperature*Pressure

*B.**Model nlinfit:***

Using the function nlinfit(), model nlinfit was implemented.

Beta = nlinfit(X,y,model,beta0)

Where y = column array of output values X = multi-column array of input values

model = Equation to be fit for the curve consisting of variables of X and co-efficients to be determine

beta0 = Initial values of co-efficients in the given model

beta = Co-efficient matrix corresponding to the given model

This function gave us the additional flexibility of choosing the equation to be fit allowing theoretical observations to be implemented directly. One such observation applied was the proportionality between wind speed and square root of pressure. Various combinations were tried out and the best of the lot was chosen. Figure 2 shows the plot.

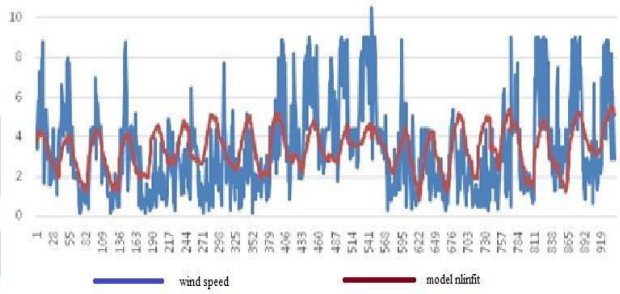

Fig 2

The following is the output obtained after modelling:

Wind Speed = 0.3143*Temperature + 0.0408*Relative Humidity - 0.2633*sqrt (Pressure)

*C.**Model cftool:***

The input parameters include predictor data along x-axis and response data taken along y-axis. There is additional option for including weights in the data points. Power fit in cftool yielded the result shown in figure 3.

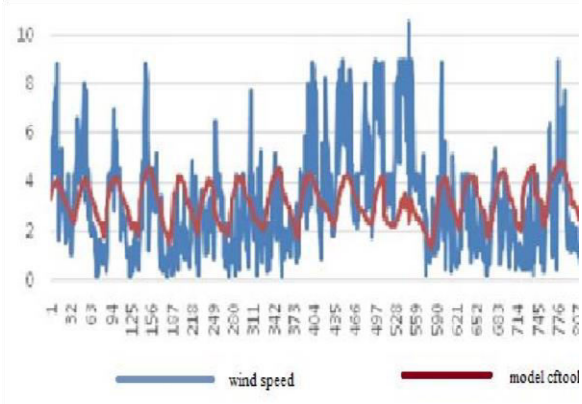

Figure: 3

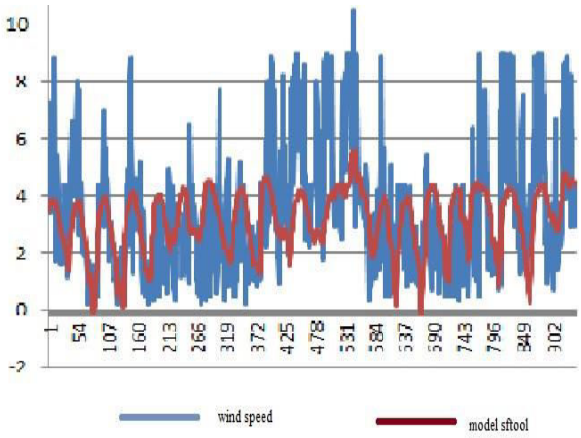

Figure: 5

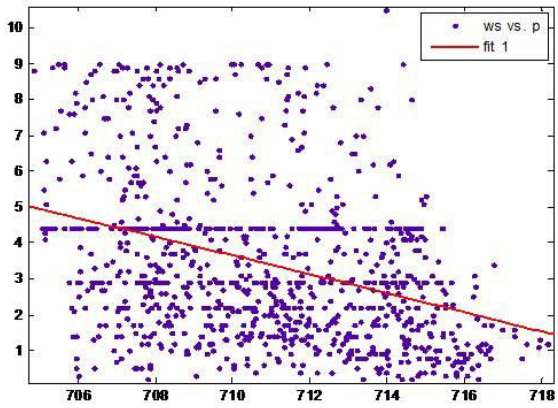

Figure: 4

Figure 4 shows the plot of the model cftool implemented using cftool.

The following is the equation obtained from power fit using cftool:

Wind Speed = -9.565e-010 *(pressure)^3.757+53.08

*D.**Model sftool:***

The inputs yielding best results were those of pressure and relative humidity and the surface fit is as shown in Figure 5. The output graph obtained was slightly different than that obtained from the previous functions and is given by Figure 6. The difference noted was the increased correlation between the inputs and lower values of wind speed.

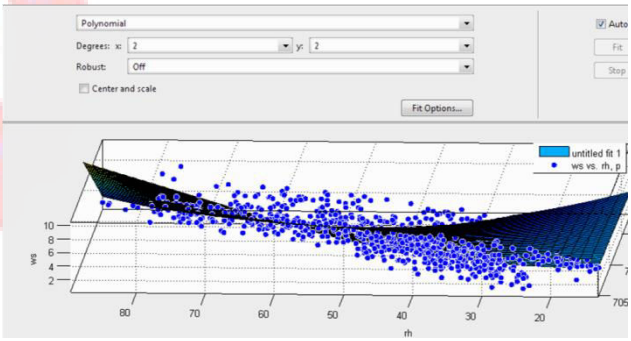

Figure : 6

The following is the equation obtained after modelling using sftool:

Wind Speed= -1525 +13.35*Relative Humidity +4.131*Pressure + 0.00332*(Relative Humidity)^2 - 0.01909*Relative Humidity*Pressure - 0.00278*(Pressure)

*E.**Time series model***

Even in the best fit, the generated curve matched the original curve in phase alone but not in magnitude. Observing all the results, we came to the

conclusion that an open loop system will never give us the required correlation. Hence, feedback systems were incorporated in order to account for the mismatched magnitudes. Thus, time series models were implemented to improve the response.

In time series modelling, time was defined as a new variable along with temperature, pressure and relative humidity. The present value was predicted with the help of previous value and the error. The input vector 'X' was defined for an intercept along with temperature and pressure.

This time series model solved the problem of mismatch magnitudes and the output curve showed better correlation with the input. But, the output had very large offshoots causing large errors. The output was enhanced by providing error correction by shifting the output waveform by trial and error. Figure 7 shows plot.

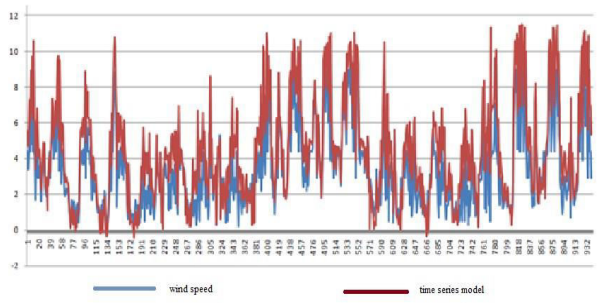

Figure : 7

The following is the equation obtained after modelling using time series:

Wind Speed(time+1) = Wind Speed(time) + 0.2018 * Temperature - 2.0873

VI. RESULTS AND DISCUSSIONS

As seen, the predicted values form a graph that is following the trend of the plot formed by the actual values. The ups and downs are followed, but the numerical data differs in case of multivariate regression models. In case of time series model both trend and the magnitude was matched to a greater extent. Table 1 shows the values of sum of squared errors for the various models implemented.

Model	SSE
Model regress	3763.851
Model nlinfit	3796.608
Model cftool	4178.766
Model sftool	3746.625
Time series model	3568.12

Table 1

VII. CONCLUSION

From table 1 it can be seen that the time series model has the least sum of squared errors taken for about 942 points. Hence the model with feedback elements yielded the best result. The time series element solved the magnitude mismatch problem and predicted wind speed with the highest accuracy as can be seen in the output.

REFERENCES

[1] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, "Current methods and advances in forecasting of wind power generation," *Renewable Energy*, vol. 37, pp. 1-8, Jan 2012.

[2] Chatterjee, S., and A. S. Hadi. "Influential Observations, High Leverage Points, and Outliers in Linear Regression."*Statistical Science.* Vol. 1, 1986, pp. 379–416.

[3] Seber, G. A. F., and C. J. Wild. " Nonlinear Regression ". *Hoboken, NJ*: Wiley- Interscience, 2003.

[4] DuMouchel, W. H., and F. L. O'Brien. "Integrating a Robust Option into a Multiple Regression Computing Environment." *Computer Science and Statistics: Proceedings of the 21st Symposium on the Interface*. Alexandria, VA: American Statistical Association, 1989.

[5] Holland, P. W., and R. E. Welsch. "Robust Regression Using Iteratively Reweighted Least-Squares."*Communications in Statistics: Theory and Methods*, A6, 1977, pp. 813–827