TIME SERIES ANALYSIS USING VECTOR AUTO REGRESSIVE (VAR) MODEL OF WIND SPEEDS IN BANGUI BAY AND SELECTED WEATHER VARIABLES IN LAOAG CITY, PHILIPPINES

Cherie B. Orpia
Mariano Marcos State University
cherie.orpia@gmail.com

Dennis S. Mapa, Ph.D
School of Statistics/ School of Economics, University of the Philippines – Diliman
cdsmapa@yahoo.com

Julius C. Orpia
SN Aboitiz Power, Magat, Inc.
julius.orpia@gmail.com

ABSTRACT

Wind energy is the fastest growing renewable energy technology. Wind turbines do not produce any form of pollution and when strategically placed, it naturally blends with the natural landscape. In the long run, the cost of electricity using wind turbines is cheaper than conventional power plants since it doesn't consume fossil fuel. Wind speed modelling and forecasting are important in the wind energy industry starting from the feasibility stage to actual operation. Forecasting wind speed is vital in the decision-making process related to wind turbine sizes, revenues, maintenance scheduling and actual operational control systems. This paper models and forecasts wind speeds of turbines in the Northwind Bangui Bay wind farm using the Vector Auto Regressive (VAR) model. The explanatory variables used are local wind speed (Laoag), humidity, temperature and pressure generated from the meteorological station in Laoag City. Wind speeds of turbines and other weather factors were found to be stationary using Augmented Dickey-Fuller (ADF) test. The use of VAR model, from daily time series data, reveals that wind speeds of the turbines can be explained by the past wind speed, the wind speed in Laoag, humidity, temperature and pressure. Results of the analysis, using the forecast error variance decomposition, show that wind speed in Laoag, temperature and humidity are important determinants of the wind speeds of the turbines.

Keywords: Vector auto regressive (VAR) model, variance decomposition

1 INTRODUCTION

The installation of additional sustainable energy sources is needed in order to address the growing needs of mankind. This is also in parallel with the increasing concerns of taking care of the environment and the minimal use of natural resources. Thus, there is an urgent need for developing renewable energy.

Wind energy is briskly being the current trend in renewable energy. It both addresses rising energy demands while being mature-friendly. On a long term basis, electricity costs coming from wind turbines are much cheaper than conventional power plants since it doesn't consume fossil fuel.

The ASEAN region has a very good potential for wind energy as indicated by the NREL wind atlas study. The member developing countries just need to tap into wind energy technology to be able to harness this potential.

Wind data gathering is essential for the wind farm starting from its feasibility to its actual operation. Prior to putting up a wind farm requires at least a year of meteorological study. Meteorological masts with heights similar to proposed wind turbine hub heights must be erected. Meteorological values, more specifically wind speed and wind direction, is necessary for the calculation of the wind farm's possible annual electrical generation profile. When the turbines are operational, on time wind data is essential for the wind turbine to be able to adjust itself to yaw towards the current wind direction, adjust its blades, vibration protection and generation mode. Production, wind turbine status and wind speed are recorded at regular time intervals.

Typhoon, substation and transmission line problems and maintenance, computer failure can render the turbines inoperational at a certain period from a few minutes to days. The importance of wind speed prediction could be essential on this stage to be able to accurately compute production loss.

As from data loss during operation, missing data is also inevitable during feasibility study stage due to faults on the power supply of the data gathering module, failing anemometers and wind vanes, and electrical communication problems.

A quick, but unreliable estimate of the actual production is to adopt the last wind speed value prior to the data blackout. This should not be the case since it is known that wind speed is intermittent. Thus, a statistically accurate wind speed forecasting is needed for this purpose.
Wind speed is intermittent and is widely known to be highly unpredictable. However, wind is just one of the common weather factors usually gathered in a weather station (e.g., PAGASA). To be able to forecast wind speed, its statistical model could be derived based on its relationship with the other weather factors.

The study therefore aims to impute missing values, forecast future wind speeds using the VAR model and determine the factors that affect wind speed using VAR

2 ABOUT NORTHWIND BANGUI BAY WIND POWER PROJECT

The Bangui Bay Wind Power Project is the first commercial wind farm in the ASEAN region. It is composed of twenty 1.65MW wind turbines distanced 326 meters apart and 40 meters after the shoreline of Bangui Bay at the Northwest tip of Ilocos Norte. It is at foreshore facing the sea with the prevalent wind direction from north-northeast. It was initially projected to supply 50% of the annual energy demand of the province of Ilocos Norte delivered through the Ilocos Norte Electric Cooperative (INEC).

3 REVIEW OF RELATED LITERATURE

There are many missing values in the wind speed data. These missing wind speed values are caused by regular wind farm shut down for maintenance, turbine repair, electrical equipment problems, and communication problems with the wind turbines

There are three common imputation methods used in wind speed. These are (1) Nearby window average imputation, (2) Jones imputation using Kalman filter and (3) EM algorithm imputation. Nearby wind method has the following principle in imputing: it uses the average value of one value before the missing value begins and one value after the missing value ends. A similar imputation technique done by Webb (1999) is by using the actual values coming from the nearest source station to the target station.

Imputation using Kalman filter is introduced by Richard H. Jones when he considered a state space model using Kalman recursive estimation for time series with missing values in 1980. He also has been able to use ARMA and ARIMA models for state-space representations.

EM Algorithm is an iterative process of ML estimation in incomplete data (Schafer 1999). Two steps are done per iteration: the expectation step where the idea is to fill in the missing data X_{miss} based on an initial estimate of the parameter μ; and then the maximization step, where μ is re-estimated based on X_{obs} and the filled-in X_{miss}. This is repeated iteratively until the estimates converge.

Hu (2002) made a study in missing value imputation on wind speeds and she concluded that EM algorithm imputation (Schafer 1999) is better compared to Nearby window average imputation and Jones' imputation using Kalman Filter (Jones 1980).

Ewing et. al. (2006), conducted a research on time series analysis of wind speed using VAR. He has been able to analyze the interdependence of the wind speeds gathered at the same meteorological station but at varying heights. Result indicated that there is no trending pattern with respect to the elevation of the wind speed measuring instrument.

4 METHODOLOGY

The statistical analysis used in this paper is a vector autoregressive (VAR) model. We used a six variable VAR that involves the following weather factors: (1) Wind Speed at hub height of a wind turbine in Bangui bay, meters per second (2) Wind Speed in Laoag City, meters per second (3) Air Pressure in Laoag City. mBar (4) Humidity in Laoag City, % Relative Humidity, (5) Precipitation, mm (6) Temperature, degrees Celsius.

The wind speed data in Bangui is computed as a daily average to coincide to the time interval of the weather data in Laoag which is daily. The data however have several missing values that need to be imputed first before analysis. After imputation, selected turbines will be chosen for VAR analysis.

In choosing which turbines will be analyzed for its wind speed, the authors preferred to choose four non-adjacent wind turbines with the least number of missing values.

4.1 MISSING VALUE ANALYSIS AND IMPUTATION

The EM algorithm is a general iterative algorithm in an incomplete data problem (Schafer, 1999). It is a two-step procedure, expectation step followed by maximization step. By filling the missing data X_{miss} based on an initial estimate of the parameter θ based on X_{obs} and the filled-in X_{miss}, iterated until the estimates converge. The distribution of the complete data X can be factored as

$$f(X|\theta) = f(X_{obs}|\theta)f(X_{miss}|X_{obs},\theta) \quad (1)$$

Letting $l(X|\theta) = \ln f(X|\theta)$,

The log likelihood will be

$$l(\theta|X) = l(\theta|X_{obs}) + f(X_{miss}|X_{obs},\theta) \quad (2)$$

Having X_{miss} unknown, we take the equation (2) in terms of $f(X_{miss}|X_{obs},\theta^t)$ where $\theta^t$is an estimate of an known parameter θ. We then get,

$$Q(\theta|\theta^t) = l(\theta|X_{obs}) + H(\theta|\theta^t)$$

Then, $Q(\theta|\theta^t) = \int l(\theta|X)f(X_{miss}|X_{obs},\theta^t)dX_{miss}$

And, $H(\theta|\theta^t) = \int[\ln f(X_{miss}|X_{obs},\theta)]f(X_{miss}|X_{obs},\theta^t)dX_{miss}$ (3)

Let θ^{t+1} be the value θ that maximizes $Q(\theta|\theta^t)$,then

$$Q(\theta^{t+1}|\theta^t) \geq Q(\theta^t|\theta^t) \quad (4)$$

The fact that $\ln(x) \geq x-1$, we have

$$\begin{aligned} H(\theta^t|\theta^t) - H(\theta^{t+1}|\theta^t) &= \int[\ln f(X_{miss}|X_{obs},\theta^t)]f(X_{miss}|X_{obs},\theta^t)dX_{miss} \\ &- \int[\ln f(X_{miss}|X_{obs},\theta^{t+1})]f(X_{miss}|X_{obs},\theta^t)dX_{miss} \\ &\geq -\int\left[\ln f\frac{(X_{miss}|X_{obs},\theta^{t+1})}{f(X_{miss}|X_{obs},\theta^t)} - 1\right]f(X_{miss}|X_{obs},\theta^t)dX_{miss} \\ &= -\left[\int f(X_{miss}|X_{obs},\theta^{t+1}) - f(X_{miss}|X_{obs},\theta^t)\right]dX_{miss} \\ &= 0 \end{aligned} \quad (5)$$

Thus

$$\begin{aligned} l(\theta^{t+1}|X_{obs}) - l(\theta^t|X_{obs}) &= Q(\theta^{t+1}|\theta^t) - H(\theta^{t+1}|\theta^t) - (Q(\theta^{t+1}|\theta^t) - -H(\theta^t|\theta^t)) \\ &= Q(\theta^{t+1}|\theta^t) - Q(\theta^{t+1}|\theta^t) + H(\theta^{t+1}|\theta^t) - H(\theta^t|\theta^t) \\ &\geq 0 \end{aligned} \quad (6)$$

That is

$$l(\theta^{t+1}|X_{obs}) \geq l(\theta^t|X_{obs}) \quad (7)$$

Thus maximizing $l(\theta^t|X_{obs})$ is sufficed to maximizing $Q(\theta|\theta^t)$. One iteration of the EM algorithm includes two steps:

1. E-step: the function $Q(\theta|\theta^t)$ is calculated by taking the expectation of $l(\theta|X)$ with the distribution $f(X_{miss}|X_{obs},\theta^t)$.
2. M-step: the parameter μ is found by maximizing $Q(\theta|\theta^t)$.

The two steps are iterated until the iterations converge. In SPSS, the EM algorithm by Schafer is used in the MVI procedure. Let the parameter $\theta = (\mu, \sum)$. For multivariate normal data, suppose there are *G* groups with distinct missing patterns. Then the observed-data log-likelihood can be expressed as

$$l(\theta|X_{obs}) = \sum_{g=1}^{G} l_g(\theta|X_{obs}), \tag{8}$$

where $l_g(\theta|X_{obs})$ is the observed-data log-likelihood from the *gth* group, and

$$l(\theta|X_{obs}) = \frac{n_g}{2}\ln|\Sigma_g \quad | - \frac{1}{2}\Sigma_{ig}^{G}[(x_{ig} - \mu_g)' \Sigma_g^{-1}((x_{ig} - \mu_g)], \tag{9}$$

where n_g is the number of observations in the g^{th} group, the summation is over observations in the g^{th} group, x_{ig} is a vector of observed values of x_g variables, μ_g is the corresponding mean vector, and Σg is the associated covariance matrix. The initial values for the first iteration are the sample means and sample variances from the observed data. The E-step uses the standard sweep operator on the covariance matrix of the observations to calculate the conditional expectation and variance of missing values. Suppose that *A* is a (*p £ p*) symmetric matrix with elements *aij* . The standard sweep operator *SWP* [*k*] operates on *A* by replacing it with another (*p x p*) symmetric matrix *B*, where the elements of *B* are given by

$$b_{kk} = -\frac{1}{a_{kk}};$$

$$b_{jk} = b_{kj} = \frac{a_{jk}}{a_{kk}}; \qquad \text{for } k \neq j;$$

$$b_{jl} = b_{lj} = a_{jl} - \frac{a_{jk}a_{kl}}{a_{kk}}; \qquad \text{for } k \neq j \text{ and } k \neq l;$$

Let *B* = *SWP*[*k*]*A*. For example, assume x_t is a time series following the model

$$(1 - \phi B)x_t = \mu + \epsilon_t \qquad \text{for t=1, …, b,} \tag{10}$$

where $|\phi| < 1$ is white noise with mean zero and variance σ^2. Let $\theta = (\mu, \phi, \sigma)$. The ML estimate is $\hat{\theta} = (\hat{\mu}, \hat{\phi}, \hat{\sigma})$. The ML estimate is $\theta = (\mu, \phi, \sigma)$. Hence the variance and covariance of missing values can be estimated by $\hat{\theta}$. Suppose that x_j is missing, and that x_{j-1} and xj+1are present. The covariance matrix of x_{j-1}; xj and x_{j+1} is

$$A = \frac{\sigma^2}{1 - \sigma^2}\begin{bmatrix} 1 & \phi & \phi^2 \\ \phi & 1 & \phi \\ \phi^2 & \phi & 1 \end{bmatrix}$$

Sweeping on var (x_{j-1}). i.e. row and column we get

$$A_{j-1} \equiv \text{SWP}[1]\text{A} = \begin{bmatrix} -\frac{1-\phi^2}{\sigma^2} & \phi & \phi^2 \\ \phi & \sigma^2 & \sigma^2\phi \\ \phi^2 & \sigma^2\phi & \sigma^2(1+\phi^4) \end{bmatrix}$$

Then sweeping on *var* (x_{j+1}). i.e., row and column 3,

$$SWP[3]A_{j-1} = \frac{1}{1+\sigma^2}\begin{bmatrix} -\frac{1}{\sigma^2} & \phi & -\frac{\phi^2}{\sigma^2} \\ \phi & \sigma^2 & \phi \\ \frac{\phi^2}{\sigma^2} & \phi & \frac{1}{\sigma^2} \end{bmatrix} \tag{13}$$

From the previous equation, we get

$$Var(x_j|x_{j-1}, x_{j+1}, \theta) = \frac{\sigma^2}{1+\phi^2} \tag{14}$$

$$E(x_j|x_{j-1}, x_{j+1}, \theta) = \mu + \frac{\phi^2}{1+\phi^2}(x_{j-1} - \mu) + \frac{\phi^2}{1+\phi^2}(x_{j+1} - \mu)$$

$$E\left(x_j \middle| x_{j-1}, x_{j+1}, \theta\right) = \mu\left(1 - \frac{2\phi}{1+\phi^2}\right) + \frac{\phi}{1+\phi^2}\left(x_{j-1+}x_{j+1}\right) \tag{15}$$

4.1.1 THE VAR MODEL

The VAR approach is preferred to other econometric models because it treats every endogenous variable in the system as a function of the lagged values of all endogenous variables in the system. In our case, the VAR model wind speed of a specific wind turbine will be based on endogenously determined variables. These variables will be the wind speed values of the four wind turbines with minimal missing values, and various weather values in Laoag like wind speed, pressure, humidity, precipitation and mean temperature. The Laoag City weather values are treated as endogenous because it is not far from Bangui and the Laoag City weather situation is commonly used as a representative for the entire province of Ilocos Norte.

While the VAR and impulse analysis are well suited for dynamic analysis, the impulse method is criticized because its results is based on "orthogonality assumption". If there are contemporaneously correlated two (or more) of the error terms in the various equations contained in the VAR system, then, the robustness of the impulse responses are questionable. In fact, the impulse responses may display noticeably different patterns (Lutkenpohl 1991). The conventional orthogonalized impulse response employs a Cholesky decomposition of the positive definite covariance matrix of the shocks (Mills 1999, Enders 2004). This restriction forces a shock to (at least) one series to have no contemporaneous effect on some other series.

In a six-variable case order VAR model, we have,

$$y_t = \beta_{10} - \beta_{12}z_t - \beta_{13}w_t - \beta_{14}u_t - \beta_{15}s_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \gamma_{13}w_{t-1} + \gamma_{14}u_{t-1} + \gamma_{15}s_{t-1} + \varepsilon_{yt}$$

$$z_t = \beta_{20} - \beta_{21}y_t - \beta_{23}w_t - \beta_{24}u_t - \beta_{25}s_t + \gamma_{21}y_{t-1} + \gamma_{22}y_{t-1} + \gamma_{23}w_{t-1} + \gamma_{24}u_{t-1} + \gamma_{25}s_{t-1} + \varepsilon_{zt}$$

$$w_t = \beta_{30} - \beta_{31}y_t - \beta_{32}z_t - \beta_{34}u_t - \beta_{35}s_t + \gamma_{31}y_{t-1} + \gamma_{32}y_{t-1} + \gamma_{33}z_{t-1} + \gamma_{34}u_{t-1} + \gamma_{35}s_{t-1} + \varepsilon_{wt}$$

$$u_t = \beta_{40} - \beta_{41}y_t - \beta_{42}z_t - \beta_{43}w_t - \beta_{45}s_t + \gamma_{41}y_{t-1} + \gamma_{42}y_{t-1} + \gamma_{43}z_{t-1} + \gamma_{44}u_{t-1} + \gamma_{45}s_{t-1} + \varepsilon_{ut}$$

$$s_t = \beta_{50} - \beta_{51}y_t - \beta_{52}z_t - \beta_{53}w_t - \beta_{54}u_t + \gamma_{51}y_{t-1} + \gamma_{52}y_{t-1} + \gamma_{53}z_{t-1} + \gamma_{54}u_{t-1} + \gamma_{55}s_{t-1} + \varepsilon_{st}$$

$$q_t = \beta_{61} - \beta_{61}y_t + \beta_{62}z_t + \beta_{63}w_t + \beta_{64}u_t + \gamma_{61}y_{t-1} + \gamma_{62}y_{t-1} + \gamma_{63}z_{t-1} + \gamma_{64}u_{t-1} + \gamma_{65}s_{t-1} + +\varepsilon_{qt}$$

Equation 16 is the structural equation of the VAR. The equations are not in reduced form since, for example, the variables on the left hand side have contemporaneous effects on the variables on the right side.

Where: y_t is the wind speed of the turbine in the study,
z_t is the wind speed in Laoag City,
w_t is for temperature,
u_t is for humidity,
s_t is for temperature,
q_t is for pressure

The ε_{yt} are white noise disturbance terms with mean 0 and standard deviation σ_y. Parameters β_{12}, β_{13}, β_{14}, β_{15}, β_{16} measure the contemporaneous effects while the lag 1 effect is measured by the γ's. With the time variable t on the left hand side and then putting in matrix form,

$$\begin{bmatrix} 1 & \beta_{12} & \beta_{13} & \beta_{14} & \beta_{15} & \beta_{16} \\ \beta_{21} & 1 & \beta_{23} & \beta_{24} & \beta_{25} & \beta_{26} \\ \beta_{31} & \beta_{32} & 1 & \beta_{34} & \beta_{35} & \beta_{36} \\ \beta_{41} & \beta_{42} & \beta_{43} & 1 & \beta_{45} & \beta_{46} \\ \beta_{51} & \beta_{32} & \beta_{53} & \beta_{54} & 1 & \beta_{56} \\ \beta_{61} & \beta_{62} & \beta_{63} & \beta_{64} & \beta_{65} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \\ w_t \\ u_t \\ s_t \\ q_t \end{bmatrix} = \begin{bmatrix} \beta_{10} \\ \beta_{20} \\ \beta_{30} \\ \beta_{40} \\ \beta_{50} \\ \beta_{60} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} & \gamma_{15} & \gamma_{16} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} & \gamma_{25} & \gamma_{26} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} & \gamma_{35} & \gamma_{36} \\ \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} & \gamma_{45} & \gamma_{46} \\ \gamma_{51} & \gamma_{32} & \gamma_{53} & \gamma_{54} & \gamma_{55} & \gamma_{56} \\ \gamma_{61} & \gamma_{62} & \gamma_{63} & \gamma_{64} & \gamma_{65} & \gamma_{66} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \\ w_{t-1} \\ u_{t-1} \\ s_{t-1} \\ q_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \\ \varepsilon_{wt} \\ \varepsilon_{ut} \\ \varepsilon_{st} \\ \varepsilon_{qt} \end{bmatrix} \quad 17)$$

simplifying,

$$B\underline{x}_t = \Gamma_0 + \Gamma_1\underline{x}_{t-1} + \underline{\varepsilon}_t$$

Isolating $\underline{x}_t$

$$\underline{x}_t = B^1\Gamma_0 + B^1\Gamma_1\underline{x}_{t-1} + B^1\underline{\varepsilon}_t \quad (18)$$

$$\underline{x}_t = A_0 + A_1\underline{x}_{t-1} + \underline{\varepsilon}_t$$

where:

$$\underline{x}_t = \begin{bmatrix} y_t \\ z_t \\ w_t \\ u_t \\ s_t \\ q_t \end{bmatrix}, \quad B = \begin{bmatrix} 1 & \beta_{12} & \beta_{13} & \beta_{14} & \beta_{15} & \beta_{16} \\ \beta_{21} & 1 & \beta_{23} & \beta_{24} & \beta_{25} & \beta_{26} \\ \beta_{31} & \beta_{32} & 1 & \beta_{34} & \beta_{35} & \beta_{36} \\ \beta_{41} & \beta_{42} & \beta_{43} & 1 & \beta_{45} & \beta_{46} \\ \beta_{51} & \beta_{32} & \beta_{53} & \beta_{54} & 1 & \beta_{56} \\ \beta_{61} & \beta_{62} & \beta_{63} & \beta_{64} & \beta_{65} & 1 \end{bmatrix},$$

$$\Gamma_0 = \begin{bmatrix} \beta_{10} \\ \beta_{20} \\ \beta_{30} \\ \beta_{40} \\ \beta_{50} \\ \beta_{60} \end{bmatrix}, \quad \Gamma_1 = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} & \gamma_{15} & \gamma_{16} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} & \gamma_{25} & \gamma_{26} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} & \gamma_{35} & \gamma_{36} \\ \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} & \gamma_{45} & \gamma_{46} \\ \gamma_{51} & \gamma_{32} & \gamma_{53} & \gamma_{54} & \gamma_{55} & \gamma_{56} \\ \gamma_{61} & \gamma_{62} & \gamma_{63} & \gamma_{64} & \gamma_{65} & \gamma_{66} \end{bmatrix}, \quad \varepsilon_{t=} \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \\ \varepsilon_{wt} \\ \varepsilon_{ut} \\ \varepsilon_{st} \\ \varepsilon_{qt} \end{bmatrix}$$

The reduced form representation of the VAR (1) model is expressed as in equation (19). Generalizing this reduced form of VAR (1),

$$\underline{x}_t = A_0 + A_1\underline{x}_{t-1} + A_2\underline{x}_{t-2} + \ldots + A_p\underline{x}_{t-p} + \underline{e}_t \quad (19)$$

where $\underline{x}_t$ is a (k x 1) vector of endogenous variables; $\underline{\mathbf{A}}$ is a matrix of coefficients to be estimated; $\underline{e}_t$ is a (k x 1) vector of forecast error that may be correlated contemporaneously but uncorrelated with their own lagged values and uncorrelated with all right hand side variables. It is assumed that $\underline{e}_t$ is distributed normally with $\underline{\mathbf{0}}$ mean and $\underline{\Sigma}$ covariance matrix. The order of the VAR model (p) is determined using the information criteria (Akaike, Schwarz and the Hannan-Quinn)

4.1.2 THE IMPULSE RESPONSE FUNCTION

Generalized impulse analysis technique should immediately follow the corresponding estimation of VAR. This technique will allow the simulation of how wind speed at one turbine responds to a sudden change in weather factors in Laoag City over future time horizon. These sudden changes may be referred to as shocks or innovations or the unexpected changes in a variable (Harvey 1994). This innovation is commonly referred to as impulse, to reflect primarily the notion of a one-time shock occurring at some point in the timeline. The impulse value, chosen as standard error equal to 1 is non zero in the initial impact period (t=1) and zero elsewhere (t<>1). The dynamic nature of the relationship between wind speeds at various locations would likely lead to the impulse having an impact in future periods as well as the period of the shock.

Impulse response analysis provides useful information as to how wind speed at a wind turbine location is like to respond to shocks caused by weather factors measured at a nearby weather station. Consider the following moving average representation of the multiple-equation, VAR (m) model where the constant terms may be ignored:

$$X_t = \Psi(L)v_t \qquad (20)$$

Let $E(v_t v'_t) = \Sigma v$ such that shocks are contemporaneously correlated. The generalized impulse response function of X_i to a unit (one standard deviation) shock in X_j is given by:

$$\Psi_{ij,h} = (\sigma_{ii})^{-1/2} (e'_j \Sigma_v e_i) \qquad (21)$$

where σ_{ii} is the i^{th} diagonal element of Σ_v,

e_i is a selection vector with the i^{th} element equal to one and all other elements equal to zero, and h is the horizon.

4.1.3 FORECASTING WIND SPEED USING VAR

The data used in the study ranged within January 1, 2006 to October 27, 2010. To be able to confirm the accuracy of the forecast model, only the data up to Dec, 2009 will be used in forecasting 2010 data while the data from January 2010 onwards will be used to validate the forecasted 2010 data.

5 FINDINGS

The wind turbines in Bangui gather wind speed through anemometers located at the top of the nacelle. Aside from instantaneous data needed for power production, the wind turbine computer logs data on close constant time intervals. The wind speed data is collected on all 15 turbines and stored in a server located at the substation office. Data can be retrieved from the server through a LAN connected personal computer. Wind speeds are measured in terms of meters per second by an anemometer located at the top of the turbine nacelle. Occasionally, the effects of turbine shutdown, communication link problem and blackouts could render either blank or zero values on the recorded wind speed data. The data set analyzed in this paper consists of daily average wind speeds collected on a selected Northwind phase I wind turbines spanning from January 1, 2006 to October 27, 2010. Outside this time frame, the said company has had technical problems that had affected data archiving.

Wind turbine numbers 2,6,8,12 (T2,T6,T8,T12) has been selected in the study for the modelling. These turbines are nonadjacent, and have the least missing values.

This data set has several missing data and has been imputed. It also has zero wind speed values that means either zero wind speed or due to equipment malfunction. A Northwind engineer was consulted to identify zero readings that indicate equipment malfunction and replaced with blank that needed further imputation. Alongside with the wind speed gathered in Northwind, various weather data have also been taken at the closest weather data collection point, which is Laoag City via www.tutiempo.net.

A casual review of figure 1 indicates that wind speed series has a pattern that is annually occurring; it peaks during the North-easterly wind season (October-February) and troughs on the May to July.

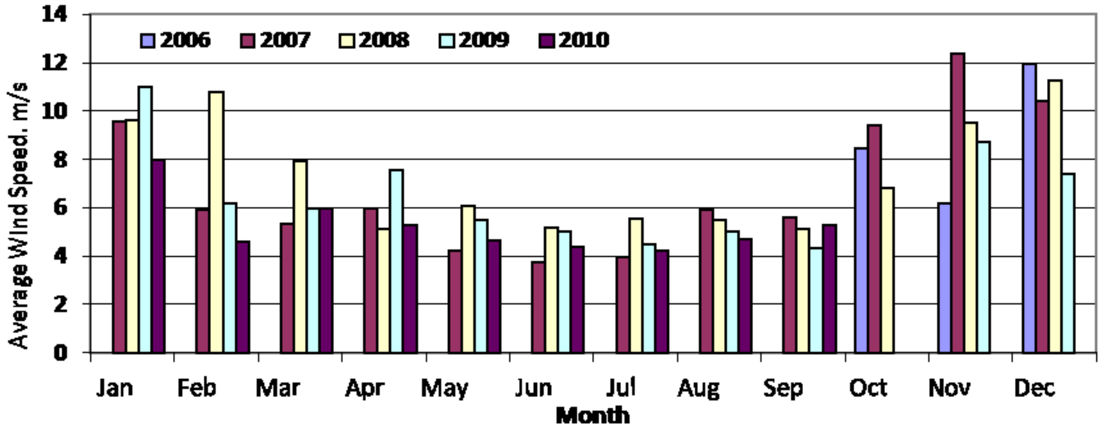


Figure 1. Wind farm's monthly average wind speed

5.1 AUGMENTED FICKEY-FULLER TEST (10% LEVEL)

Prior to building the VAR model, the time series of the variables were first tested with unit roots using Augmented Dickey-Fuller (ADF) Test. Test results showed that all the variables are stationary.

5.2 THE VAR MODEL

The paper examines how wind speed at a certain location responds to shock being made by various weather indicators in an off-site place, i.e., Laoag City, as well as identifying interdependencies and cross variable analysis. To be able to estimate this, a multiple equation, Vector Autoregressive (VAR) modelling will be used. For this type of data, it is appropriate to use this method, alongside with impulse response analysis (Mills 1999). The structure of the research model makes minimal theoretical demands when VAR technology is used.

The result of the VAR (3) model using the daily time series data on the wind speed on a randomly selected turbine location (T2, T6, T8 and T12), the humidity, temperature, precipitation, pressure and wind speed in Laoag City. The paper is interested in the first equation of the VAR where the dependent variables are T2, T6, T8 and T12. The wind speed at time *t* can be explained by lag 1 to lag 3 values of the weather factors in Laoag. The significance level is explained in table 1.

It can be seen on the table above that through VAR analysis, wind speeds in T2 can be explained by AR (1), AR (2) and AR (3) of itself at very high levels of significance. The same explanation can be seen through the other turbines in the study (T6, T8 and T12).

The wind speeds at individual turbines can be explained significantly at 1% by Lag (1) of the wind speed in Laoag. The results are varying at AR (2) from not significant to significant.

However the AR (3) results range from 1% to 10% significant. Thus, turbine wind speeds can be explained at AR (1) and AR (3) values for wind speeds in Laoag.

Humidity can explain at 1% level of significance of the turbine wind speeds at AR (1) and AR (2) values. At AR (3) values, several turbines can still be explained by humidity at 1% to 10% level. Temperature and pressure can be easily explained at the 1% level of significance of the turbine wind speeds at both AR (1) and AR (2) but not significant at AR (3).

Precipitation cannot easily explain wind speeds as the VAR results give inconsistent results and most of them are not significant at any lag values.

To check the volatility of the VAR model, a VAR residual heteroskedasticity test was conducted and the result showed that the wind speed of the four turbines included in the study have non constant variance. Thus, MGARCH was implemented. The MGARCH result shows that for T2, T6, T8 and T12 wind speeds, the GARCH-in-mean terms σ^2 are 0.2208, 0.39, 0.57 and 0.56 respectively, which are all positive and statistically significant at 5% level.

Table 1. Significance level of the VAR results

	T2	T6	T8	T12
T2/T6/T8/T12 (-1)	1%	1%	1%	1%
T2/T6/T8/T12 (-2)	1%	1%	1%	1%
T2/T6/T8/T12 (-3)	1%	1%	1%	1%
WindSpeedLaoag (-1)	1%	1%	1%	1%
WindSpeedLaoag (-2)	10%	5%	Ns	ns
WindSpeedLaoag (-3)	5%	10%	1%	1%
Humidity (-1)	1%	1%	1%	1%
Humidity (-2)	1%	1%	1%	1%
Humidity (-3)	5%	1%	10%	10%
MeanTemperature (-1)	1%	1%	1%	1%
MeanTemperature (-2)	1%	1%	1%	1%
MeanTemperature (-3)	Ns	ns	Ns	ns
Precipitation (-1)	Ns	1%	Ns	ns
Precipitation (-2)	Ns	10%	Ns	ns
Precipitation (-3)	Ns	ns	Ns	ns
Pressure (-1)	1%	1%	1%	1%
Pressure (-2)	1%	1%	1%	1%
Pressure (-3)	Ns	ns	Ns	ns
C	1%	5%	1%	1%

5.3 THE IMPULSE RESPONSE FUNCTION

It can be seen on figure 2 that the wind speeds in wind turbines responds proportionately to the wind speed in Laoag City. The shock can also be seen peaking on day 3.

In figure 3-4 the impulse caused temperature and humidity are quite similar and is felt by the turbine wind speed lasting even beyond two weeks. This reaction of the wind speed in turbines to temperature and humidity, however, is reverse, i.e., as the temperature and humidity goes up wind speed goes down and vice versa. This result could yield to further analysis of the effects of rising surface temperatures caused by climate change to the wind speed, of which, in turn, would have an effect on the power output. Several research has already analyzed these impact. In Ontario, for example, a research has found out that there will be a wind speed reduction of 5% from present to 2071 (Yao et. al. 2012). Conversely, the proliferation of wind turbine claims that it does not affect the global climate, i.e, to effects on temperatures and wind speeds (Marvel et. al. 2013).

Impulse caused precipitation however, gives a no to very little impact on the wind speed (figure 5). Moreover, the impulse due to pressure gives a short term impact to wind speed lasting only 2-3 days (figure 6)

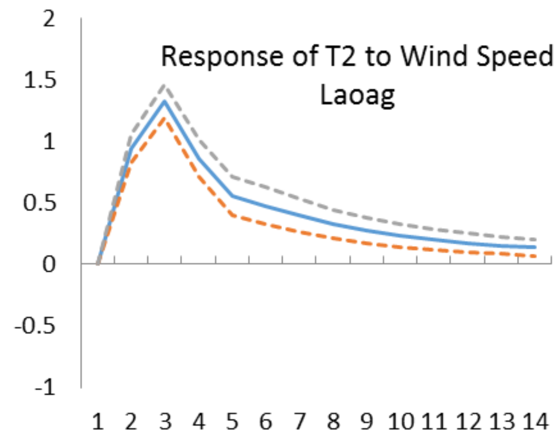


Fig.2. Impulse response of Wind speeds in T2 to the wind speed in Laoag

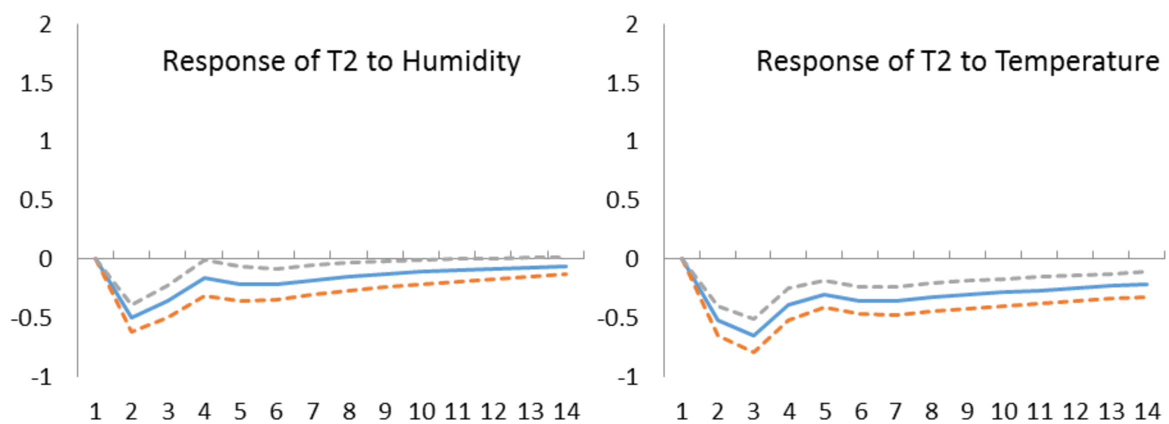
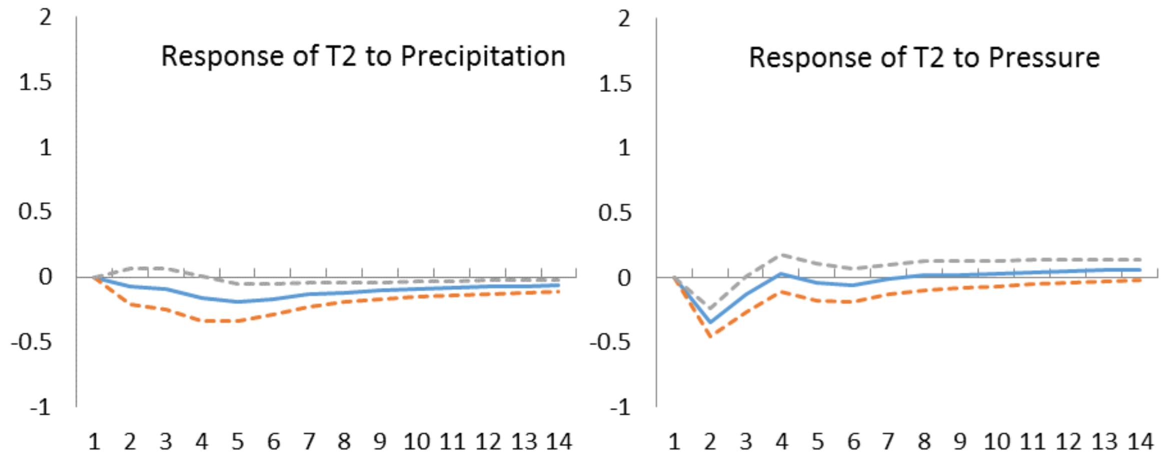


Fig.3-4. Impulse response of Wind speeds in T2 to the humidity and temperature



Figs5-6. Impulse response of Wind speeds in T2 to the temperature and precipitation

5.4 THE FORECAST ERROR VARIANCE DECOMPOSITION

While the impulse response functions trace the effects of a shock to one endogenous variable in the other variables in the VAR model, the Forecast Error Variance Decomposition tells us the proportion of the movements in the series (error of turbine wind speeds at period (*t*+1). The shock caused by the wind speed in Laoag yields only from 10.25% in T2 to 13.38% in T12. Shock effect of humidity and mean temperature to the wind speed in the turbines only ranges from 2.52% to 3.05% during this time. Effect of precipitation and pressure are negligible. However, at period (*t*+2) about 63-66% of T2, T6, T8 or T12 wind speeds) due to its "own" shocks versus the shocks to the other variables (Laoag wind speeds, pressure, temperature, precipitation and humidity). The variance decomposition provides information about the relative importance of each random innovation in affecting the variables in the VAR model.

The forecast error variance decomposition in the wind speed in the turbines shows how much future error variance of the wind speed can be explained by shocks to the wind speed itself and the weather factors in Laoag. Analysis shows that the shock to the wind speed of the turbine itself (or "own shock") can explain almost all of it, with at least 81.27% in T8 to 83.29% in T2, of the variance of the forecast.The forecast error variance of the turbine wind speed can only be explained by its own shock while the forecast error variance of the wind speed in Laoag can now explain 23-28% of the turbine wind speed. The temperature can now explain 5-6% of the turbine wind speed. The pressure this time can explain just 2-3% of the turbine wind speed. Precipitation and pressure remained negligible at 1% or less.

At period (*t*+3) the total variance explained by wind speeds own shock is reduced to 61-64%, while the wind speed in Laoag ranged at 27-34% and temperature ranged at 5-5% to 6.8%. Humidity is pegged at less than 3%, while other factors stayed at around 1%. From this we can say that, arranged in decreasing effect, the wind speed in Laoag, temperature and humidity are important determinants of the wind speeds in the turbines.

5.5 FORECASTING USING VAR

The study used the data values from 2006 to 2009 to be able to forecast values in 2010. Basing on figure 7, the forecasted values appear to have similar nature with the analyzed data, that is, exhibiting the same periods of peaks and troughs.

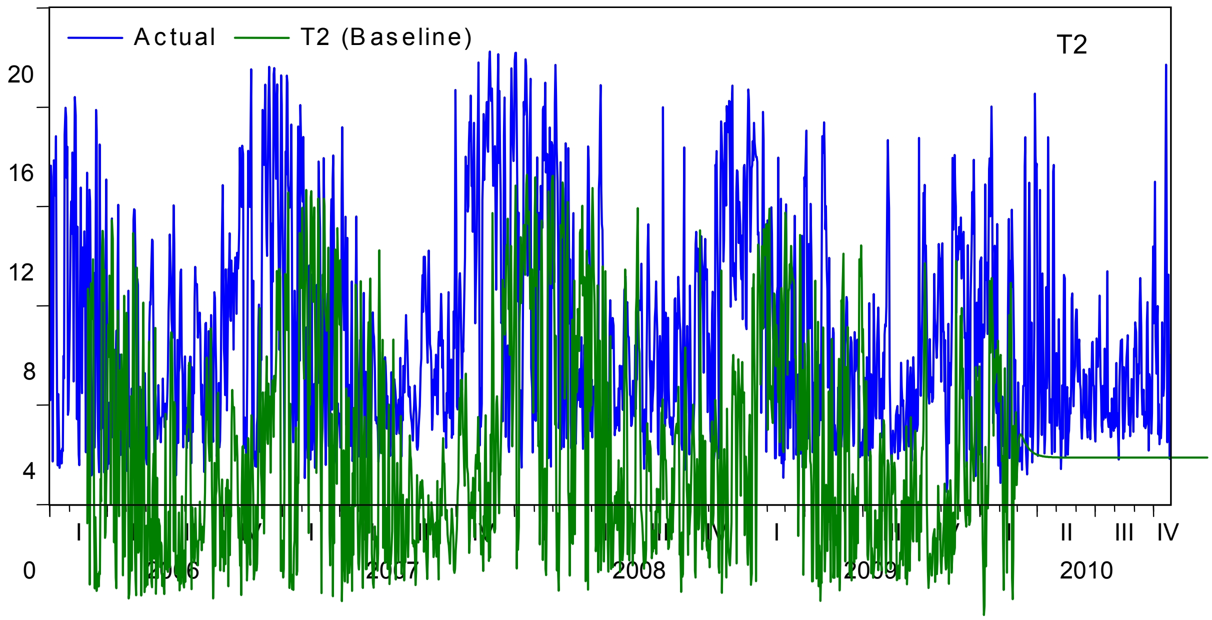


Figure 7.Windspeeds in T2 in 2006-2009 (green) and the forecasted values in 2010 (blue)

6 CONCLUSION AND RECOMMENDATION

This study examines the dependence of the time series model of wind turbine wind speed the weather factors gathered at the closest weather station. The econometric-based model using vector autoregressive (VAR) shows that the weather factors like temperature, pressure, humidity and wind speed highly affects the time series model of the wind speed in a wind turbine.
The wind speeds and the other weathers factors were found to be stationary using Augmented Dickey-Fuller test.

The use of VAR(3) model using daily time series data reveals that wind speeds on the turbines can be explained at AR(1) of its own wind speed, the wind speed in Laoag, Humidity, Temperature and Pressure.

With the use of impulse response function, the wind speed on the turbines responds very well with the wind speed of Laoag. The wind speed on the turbines also responds very well with temperature and pressure. Any increase in the latter would incur a decrease in the wind speed in the turbine and vice versa. Humidity affects wind speed on the same way although the shock caused by the humidity only lasts 2-3 days.

Results of the forecast error variance decomposition show that wind speed in Laoag, temperature and humidity are important determinants of the wind speeds in the turbines.

Further studies using VAR could also include wind direction and wind gusts to further confirm the result of this study.

REFERENCES

Akaike, H, 1974. A new look at the statistical model identification.IEEE Trans. Automat. Control. vAC-19. 716-723.

Azami Zahari, Siti Khadijah Najid, Ahmad Mahir Razali, &Kamaruzzaman Sopian, 2009. Wind Speed Analysis in the East Coast of Malaysia.European Journal of Scientific Research. Issue 2, Vol. 3.

Durban, M &Glasbey, CA, 2001. Weather Modelling using a multivariate latent Gaussian. Agricultural and Forest Meteorology, 109,187-201. JCMB, King's Buildings, Edinburgh, EH9 3JZ, Scotland. Retrieved at http://www.bioss.ac.uk/staff/chris/rain.pdf (March 30, 2012).

Enders, W, 2004. Applied Econometric Time Series. 2nd Ed. John Wiley.

Ewing, Bradley T, Kruse, Jamie Brown, Schroeder, John L& Smith, Douglas A, 2006. Time Series Analysis of Windspeed Using VAR and the Generalized Impulse Response Technique. Revised March 2006. Retrieved at http://www.ecu.edu/cs-educ/econ/upload/ecu0607.pdf (March 30, 2012).

Hannan, EJ&Quin, GG, 1979. The determination of the order of an autoregression. J.R. Statistic. Soc. B, 41, 190-195.

Hu, Mingyan, 2002. Wind Speed Analysis for Lake Okeechobee. Unpublished thesis. Florida Atlantic University.

ISSN 1450-216X Vol.32 No.2 (2009), pp.208-215.© EuroJournals Publishing, Inc. 2009. http://www.eurojournals.com/ejsr.htm. Retrieved at http://www.eurojournals.com/ejsr_32_2_09.pdf (March 30, 2012).

Jones, RH, 1980. Maximum Likelihood Fitting of ARM Models to Time Series with missing Observations. Technometrics 22(3): 389-395.

Marvel, K, Kravitz, B&Caldeira, K, 2012. Geophysical limits to global wind power. Nature Climate Change 3, 118–121.

Mills,T, 1999. The Econometric Modelling for Financial Time Series. 2nd ed. Cambridge University Press. Cambridge.

Schafer, JL, 1999. Analysis of Incomplete Multivariate Data. Chapman & Hall, Boca Raton.

Tastu, J, Pinson, P, Kotwa, E&Madsen, H, 2010. Spatio temporal and modelling of short term wind power forecast errors. Wind Energy. February 16, 2010.

Webb, A, 1999. Statistical Pattern Recognition.New York, Oxford University press.

Yao, Y, Huang, GH,&Lin, Q, 2013. Climate Change impacts on Ontario wind power resource. Environmental Systems Research Journal.