Wind speed forecasting in the South Coast of Oaxaca, México

Erasmo Cadenas\textsuperscript{a}, Wilfrido Rivera\textsuperscript{b,*}

\textsuperscript{a}Facultad de Ingeniería Mecánica, Universidad Michoacana de San Nicolás de Hidalgo, Santiago Tapia No. 403, Centro, México
\textsuperscript{b}Centro de Investigación en Energía of the Universidad Nacional Autónoma de México (UNAM), Apartado Postal 34, Temixco 62580, Morelos, México

Received 1 November 2005; accepted 18 October 2006
Available online 11 December 2006

Abstract

Comparison of two techniques for wind speed forecasting in the South Coast of the state of Oaxaca, Mexico is presented in this paper. The Autoregressive Integrated Moving Average (ARIMA) and the Artificial Neural Networks (ANN) methods are applied to a time series conformed by 7 years of wind speed measurements. Six years were used in the formulation of the models and the last year was used to validate and compare the effectiveness of the generated prediction by the techniques mentioned above. Seasonal ARIMA models present a better sensitivity to the adjustment and prediction of the wind speed for this case in particular. Nevertheless, it was shown both developed models can be used to predict in a reasonable way, the monthly electricity production of the wind power stations in La Venta, Oaxaca, Mexico to support the operators of the Electric Utility Control Centre.

\textcopyright{} 2006 Elsevier Ltd. All rights reserved.

Keywords: Wind speed; Prediction; ARIMA; Neural networks

1. Introduction

Nowadays, the worldwide installed wind power capacity is approximately 46,048 MW, with a growth of 17\% in 2004. Countries like the United States of America, Germany, Spain and Denmark will continue developing strongly this source of energy, which has
already arised the interest of several countries such as the UK, Japan, India and Egypt, Africa has 40 MW installed, whereas Costa Rica is the leading country in Latin America with 71 MW [1].

Mexico has around 5 MW installed, nevertheless, in 2006 a 100 MW project will be realised in the Isthmus of Tehuantepec region in the state of Oaxaca, that reflects the government interest in the use of this kind of energy at a large scale [2].

The state of Oaxaca is located in the south part of Mexico, with a surface area of approximately 95,364 km², and it is the fifth largest in the country. Its orography includes mountainous zones, flat plateaus, valleys and coasts, with a great variety of climates, tropical throughout the coast and temperate in the interior. Rainfall appears generally at the end of April and continues until the end of October. The average temperature oscillates between 26 and 28 °C throughout the coast, from 20 to 22 °C in central valleys and from 12 to 15 °C in the mountains [3].

La Venta is located 60 km NNE of the port of Salina Cruz, Oaxaca, and it is a zone recognized by its strong and persistent winds, reason why, the region has been object of several studies. In 1988, Steenburgh et al. [4], describes the development of a model that analyses the dynamic behaviour of the wind in the Gulf of Tehuantepec. In 2005, Jaramillo et al. [5], proposed a Weibull&Weibull function that describes the bimodality of the wind behaviour in this place.

A recent study has determined a wind potential of around 6000 MW in the most productive zones but greater than 30,000 MW in all the state [3].

The exploitation of this potential will imply the incorporation of greater volumes of energy to the power grid of the Interconnected National System (INS), energy whose intermittent and random character must be considered by the operators of the Electric Utility Control Centre (EUCC).

The models for wind prediction and power generation are valuable support tools for the operation of the EUCC [6]. Its importance has been recognized and valued to such a point that in countries with a large wind power development like Spain it is a legal requirement for the producers to give the short and mid-term production forecasting to the EUCC for the energy supply [7].

In this study, two techniques used for wind prediction for the region of La Venta are analysed with the objective that they may be applied to the future of wind power developments that will take place in the zone.

2. Data measurement

The Comisión Federal de Electricidad (CFE) has made wind speed measurements since 1994, through a network of measurement stations located in the place of interest. The sensors were located at different heights in the measurement towers (20, 30, 40 m from land level). Their characteristics are shown in Table 1.

The information generated by the sensors is accumulated in the data acquisition systems through chips or memory cards that later are unloaded in a computer in order to be processed.

Fig. 1 shows the monthly behaviour of the wind speed in La Venta for the period from June 1994 to May 2000. A seasonal behaviour in the series is observed, the strongest winds appear at the end of every year and are weakest in the middle.
3. Time series model

A time series model \( (y_t) \) reproduces the patterns of the previous movements of a variable and uses this information to predict its future movements. In this way, it is possible to construct a simplified model of the series that represent its randomness so that it is useful for prediction [8].

The present study uses the Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Networks (ANN) methodologies for the treatment of time series to obtain the prediction models. ARIMA models have been used in a great number of prediction problems related with time series because its robustness makes its implementation and understanding easy. Nevertheless, the difficulty of its adaptation with atypical values, influences the estimation of the future values. Another disadvantage of the stochastic models is that the order is generally high [9,10]. Because of this, the use of the ANN technique becomes necessary in the comparison and validation of the statistical model. Neural Networks have been used in applications related to wind prediction [11,12], in which most of the suggested models use networks multi-layer perceptron (MLP).
George Box and Gwilym Jenkins helped ARIMA models to become popular at the beginning of the 1970s, and their names are frequently associated with the general models of ARIMA applied to the analysis of time series and prediction.

A great number of ARIMA models exist. The general non-seasonal model is known as ARIMA \((p, d, q)\), where \(p\) is the order of the autoregressive part of the model, \(d\) the order of differencing done to the data to make it stationary, \(q\) the order of the moving average part of the model.

The linear expression that defines the previous notation is the following:

\[
y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j e_{t-j} + \varepsilon_t, \tag{1}
\]

where \(\phi_i\) is the \(i\)th autoregressive parameter, \(\theta_j\) the \(j\)th moving average parameter, \(\varepsilon_t\) the error term at time \(t\).

The ARIMA models are compound of a seasonal and non-seasonal part and are represented by the following way:

\[\text{ARIMA}(p, d, q)(P, D, Q),\]

where \(P, D, Q\) is the seasonal part of the model, \(S\) the number of periods per season.

Lineal representation can be made adding the seasonal part to the expression (1).

ARIMA models are used in diverse fields of knowledge such as engineering and economy among others. In specific cases like the prediction of the demand of power, the wind speed, and the behaviour of stock-market values, anything possible to be represented by means of a time series and has a reasonable number of measurements is feasible to be treated by means of this technique.

ANN are based on simple mathematical models. When they are applied to time series they provide a method of non-linear prediction. The prediction with ANN generally requires a vast number of observations, nevertheless, it allows the adjustment of more complicated models. Also, it adopts a different terminology from the one used in the traditional forecasting methods; for example, in a prediction model represented by a network, the parameters of the model are the weights of the network and the estimation process of the parameters is defined as network training.

The necessary components to establish a neural network are the following:

(a) Its architecture (the number of layers and units in the network and connections among them).
(b) The activation function (that describes as each unit combines its inputs to obtain the desired outputs).
(c) The cost function (a measurement of the accuracy of the prediction like the average squared error).
(d) The training algorithm to find the values of the parameters that diminish the cost function.

In the case of the ANN used in the present study, ADALINE (ADAptive LINear Element), its main application field is in the elimination of noise and signal prediction.
carrying digital information, one of the outstanding applications in the industrial field for this type of networks is the elimination of echoes in telephonic circuits [11].

3.1. Box–Jenkins methodology

In order to model the time series of La Venta, the Box–Jenkins methodology was followed. This methodology is basically divided in the following four phases:

(a) Identification: It is the stage of preparation of the data, the necessity or not to transform the data is identified with the purpose of stabilizing the variance. In this phase, the data are also differenced to obtain the stationary series. Fig. 2 shows the transformation made to the series in Fig. 1, with the purpose of stabilizing the mean and variance and to identify the potential models when the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) are obtained.

The analysis of the ACF and PACF was made in this stage as shown in Fig. 3, where a strong correlation is detected with the 1st lag, as well as with the 12th, indicating a seasonal behaviour of the series.

(b) Estimation and test: The parameters in potential models were considered and the best model was selected using a suitable criterion.

(c) Tests were made to the residues through the ACF and PACF techniques: Normality test and the statistical test “t” are applied to the residues in order to determine if they are white noise Fig. 4 [17].

(d) Forecasting: A model is considered adequate for prediction once it approved the statistical tests.

![Fig. 2. Time series data stationary.](image-url)
3.1.1. Proposed model

According to the development of the analysis described in the previous sequence, the representative model for the analysed series is ARIMA(0,1,1)(0,1,1)\(_{12}\), that is, it was necessary to obtain both a seasonal difference and a non-seasonal one. The ACF of Fig. 3 indicates a strong relation between the last observation and the lags \(y_{t-1}\) and \(y_{t-12}\); the model represented in notation of operators is shown by Eq. (2) and (3) and the model of the synthetic series in Fig. 5.

\[
(1 - B)(1 - B^{12})y_t = (1 - \theta_1 B)(1 - \Theta_1 B^{12})e_t, \tag{2}
\]

\[
Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + e_t - 0.9975e_{t-1} - 0.7976e_{t-12} + 0.7956e_{t-13}, \tag{3}
\]

where \(By_t = y_{t-1}\), \(Be_t = e_{t-1}\), \(\theta_1\) the moving average parameter, \(\Theta_1\) the seasonal parameter of the moving averages.

Fig. 5 shows a suitable adjustment, which gets stronger in the last year; it is also appraised that the series does not follow the atypical behaviour of series of the 4th year.
3.2. Neural network model

Nowadays a considerable number of studies are made and published in the field of neural computation, the advance of the computational technology has made the reappearance of these techniques possible. In the field of wind speed prediction, the MLP is the most used [13,14].

In the present study, a variant of this architecture is considered using a ADALINE (ADAptive LINear Element) network which modifies the learning mechanism of the Perceptron.

3.2.1. ADALINE Networks

ADALINE Networks were developed by Widrow [15] in the University of Stanford subsequent to the development of the Perceptron. The ADALINE networks are similar to the Perceptron, but their transfer function is linear, which allows their outputs to take on any value, whereas the Perceptron output is limited to either 0 or 1. Both the ADALINE and the Perceptron can only solve linearly separable problems. However, here the least mean squared (LMS) learning rule, which is much more powerful than the Perceptron learning rule, is used. The LMS, or Widrow-Hoff, learning rule minimizes the mean squared error (MSE) and thus moves the decision boundaries as far as it can from the training patterns [16]. The structure of the ADALINE network is shown in Fig. 6.

The ADALINE uses a denominated adaptive linear element controller (ALC) that obtains a linear exit that can be applied to another element. The ALC makes the calculation of the weighted sum of the inputs:

\[ s = w_0 + \sum_{j=1}^{N} w_j x_j. \]  

(4)
As in the case of the Perceptron, the threshold of the transfer function is represented through a fictitious connection of weight \( w_0 \). If it is considered that for this input it takes the value of \( x_0 = 1 \), the previous expression can be written as

\[
s = \sum_{j=0}^{N} w_j x_j = XW^T. \tag{5}
\]

### 3.2.2. Learning process

ADALINE uses learning OFF LINE with denominated supervision LMS or rule of the minimum average squared error, also well-known as delta rule because it tries to diminish a delta or difference between the observed value and the desired one in the output of the network.

The learning rule LMS diminishes the average squared error, defined as

\[
\langle \varepsilon_k^2 \rangle = \frac{1}{2L} \sum_{k=1}^{L} \varepsilon_k^2, \tag{6}
\]

where \( L \) is the number of input vectors (patterns) that form the training set, and \( \varepsilon_k \) the difference between the desired output and the one obtained. In the case of ADALINE, it is expressed as \( \varepsilon_k = (d_k - s_k) \), being \( s_k \) the output of the ALC,

\[
s_k = X_k W^T = \sum_{j=0}^{N} w_j x_{kj} \tag{7}
\]

and \( d_k \) the desired output.

The error function is a mathematical function defined in the multidimensional space of weights for a set of given patterns. It is a surface that will have many minimums (global and local), and the learning rule is going to search for the point in space, which is the global minimum of this surface.
Fig. 7 shows the evolution of the average squared error, used in the convergence of the ANN model generated for the wind speed forecasting of La Venta, Oaxaca.

The changes in weights are proportional to the descendent gradient of the error function, which is defined as the rate of learning:

$$w_i(t + 1) = w_i(t) + \alpha(d_k - s_k)x_{ki},$$

(8)

where $\alpha$ represents the proportionality constant or rate of learning.

3.2.3. Neural network structure

In order to conform the neural network used in the forecast, correlation and PACF were used to detect the existing relation between the actual observation and the previous ones, determining therefore the number of input and output neurons and the vectors used during the training.

The characteristics of the used neuronal network are the following:

(a) Number of input neurons: 3.
(b) Hidden layer does not exist.
(c) Number of output neurons: 1.
(d) The inputs were defined by means of ACF, being the conformed vectors like function of $(y_{t-1}, y_{t-12}, y_{t-13})$.
(e) Number of training vectors: 59.
(f) Value of the learning rate (0.01–0.25).

Fig. 8 shows the generated synthetic series with ADALINE. Significant differences with respect to the series generated with ARIMA are not appraised. In the same way, in the last year a precise adjustment is obtained.
4. Comparative analysis of the models

The 7th year of the original time series was used to compare the performance of the ADALINE and SARIMA models.

Fig. 9 shows the results obtained in this comparison. As it can be observed both models seem to work in a reasonable way considering the wind speed randomness. However, a greater sensitivity in the adjustment of the curve on the part of seasonal model SARIMA is appraised in most part of the year.

In order to determine quantitatively which is the best model, three forecast error measures were employed for model evaluation and model comparison, the MSE, the mean absolute error (MAE), and the mean absolute percentage error (MAPE).

If $y_t$ is the actual observation for a time period $t$ and $F_t$ is the forecast for the same period, then the error is defined as

$$e_t = y_t - F_t.$$  \hspace{1cm} (9)

The standard statistical error measures can be defined as

$$\text{MSE} = \frac{1}{n} \sum_{t=1}^{n} e_t^2.$$  \hspace{1cm} (10)

And the MAE as

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^{n} |e_t|.$$  \hspace{1cm} (11)
To define MAPE, first we need to define a relative or percentage error as

\[ PE_t = \left( \frac{y_t - F_t}{y_t} \right) \times 100. \] (12)

Then the MAPE is

\[ MAPE = \frac{1}{n} \sum_{t=1}^{n} |PE_t|. \] (13)

Additionally, it is important to compare the aforementioned models with the reference model Naïve. This method uses the most recent observation available as a forecast. The Naïve method provides a measure of the improvement attainable through the use of a more sophisticated forecasting method. This type of comparison is much more useful than simply computing MSE, MAE or MAPE, since it provides a basis to evaluate the relative accuracy of those results.

Finally it would be helpful to have a measure that considers the disproportionate cost of large errors and provides a relative basis for comparison with Naïve methods. A measure that has these characteristics is the U-statistic developed by Theil [18].

This statistic allows a relative comparison of formal forecasting methods with Naïve approaches and also squares the errors involved so that large errors are given much more weight than small errors.

Mathematically, Theil’s U-statistic is defined as

\[ U = \sqrt{\frac{\sum_{t=1}^{n-1} (FPE_{t+1} - APE_{t+1})^2}{\sum_{t=1}^{n-1} (APE_{t+1})^2}}. \] (14)
where
\[
F_{\text{PE},t+1} = \frac{F_{t+1} - y_t}{y_t} \quad \text{(forecast relative change)}
\]
and
\[
A_{\text{PE},t+1} = \frac{y_{t+1} - y_t}{y_t} \quad \text{(actual relative change)}.
\]

The ranges of the $U$-statistic can thus be summarized as follows:

$U = 1$: the Naïve method is as good as the forecasting technique being evaluated.

$U < 1$: the forecasting technique being used is better than the Naïve method. The smaller the $U$-statistic, the better the forecasting technique is relative to the Naïve method.

$U > 1$: there is no point in using a formal forecasting method, since using a Naïve method will produce better results.

Table 2 shows the results obtained for the statistical errors with the different models.

From Table 2 it is clear that the SARIMA model has lower statistical errors than those compared with ADALINE, as it was observed qualitatively in Fig. 9. However, both models can be used in a reasonable way since in both cases the values of the Theil’s $U$ are lower than the value of the comparative Naïve method.

5. Conclusions

Seasonal ARIMA models present a better sensitivity to the adjustment and prediction of the wind speed for this case in particular. Nevertheless, it is probable that when increasing the number of training vectors for the ANN model, its performance will improve the adjustment. Finally it is advisable to observe that the developed models can be used to predict, in a reasonable way, the monthly power production of wind power stations in La Venta, Oaxaca, Mexico to support the operators of the Electric Utility Control Centre.

Acknowledgements

We are thankful to Subgerencia de Proyectos Geotermoelectricos of the Comision Federal de Electricidad (CFE), for the support in accomplishment of the present study, specially to Ing. Roberto Tovar Cadenas, project manager and to Ing. Carlos Sanchez Cornejo for its valuable help.
References


