Analysis and forecasting of wind velocity in Chetumal, Quintana Roo, using the single exponential smoothing method

E. Cadenas a, O.A. Jaramillo b, W. Rivera b,*

a Facultad de Ingeniería Mecánica, Universidad Michoacana de San Nicolás de Hidalgo, Santiago Tapia No. 403, Centro, México
b Centro de Investigación en Energía, Universidad Nacional Autónoma de México, Apartado Postal 34, Temixco 62580, Morelos, México

Abstract

In this paper the analysis and forecasting of wind velocities in Chetumal, Quintana Roo, Mexico is presented. Measurements were made by the Instituto de Investigaciones Eléctricas (IIE) during two years, from 2004 to 2005. This location exemplifies the wind energy generation potential in the Caribbean coast of Mexico that could be employed in the hotel industry in the next decade. The wind speed and wind direction were measured at 10 m above ground level. Sensors with high accuracy and a low starting threshold were used. The wind velocity was recorded using a data acquisition system supplied by a 10 W photovoltaic panel. The wind speed values were measured with a frequency of 1 Hz and the average wind speed was recorded considering regular intervals of 10 min. First a statistical analysis of the time series was made in the first part of the paper through conventional and robust measures. Also the forecasting of the last day of measurements was made utilizing the single exponential smoothing method (SES). The results showed a very good accuracy of the data with this technique for an α value of 0.9. Finally the SES method was compared with the artificial neural network (ANN) method showing the former better results.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

According to the World Wind Energy Association, by the end of June of 2008 the cumulative installed capacity was close to 95,600 MW. The currently installed wind power capacity generates 200 TWh per year, equaling 1.3% of the global electricity consumption. Wind power is now established as an energy source in over 50 countries around the world, it is important to indicate that in some countries and regions, wind energy already contributes 40% and more. Europe has the lead in wind energy development; however wind energy markets in the United States, China, and India are growing very fast.

Since January 2007, with the operation of the wind farm La Venta II with an installed capacity of 83.3 MW, Mexico joins the list of countries that produce electricity from wind on a commercial scale. Mexico is ranked in second place in Latin America after Brazil that presents 247 MW wind power installed.

As wind energy can be considered as a significant contribution to the electric generation in various places around the world, the need for accurate predictions of available wind electric potential for a variety of time scales is increasing in importance. Accurate wind forecasts are required to best integrate wind electric potential into scheduling and dispatch decisions made by an energy provider. Accurate wind forecasts will also help to remove barriers related with the wind intermittence and unpredictability that make some energy providers reluctant to pursue wind as an energy resource.

Wind is considered as one of the most difficult meteorological parameter to forecast. Wind is the result of the complex interactions such as pressure and temperature differences of the global environment, the rotation of the earth, and local characteristics of the surface. The forecasting technique employed depends on the available information and the time scale in question, and thus its application. Some different techniques are applied to forecast periods in the range of a few hours or days. The approaches that are found in the literature include Numerical Weather Prediction and Mesoscale models, Generalized Equivalent Markov models and time series approaches. The latter incorporates various techniques such as ARMA models, bilinear and smooth threshold autoregressive models. Recently, artificial intelligence techniques have been proposed. They include the use of Multi-Layered Perceptrons, Radial Basis Functions and Recurrent Neural Networks as well as Fuzzy Logic and the combination of a Fuzzy Classifier with a Temporal Neural Network [1].

In 1996 Kariniotakis et al. [2] reported a model based on neural networks for the prediction of power output profile of a wind park.
In the model are employed simple methods like persistence, as well as classical methods and it was implemented into a real control system of an autonomous wind-diesel power system.

In 1998 Alexiadis et al. [3] studied the operation of power systems with integrated wind parks. They proposed artificial neural networks models for forecasting average values of 10 min or 1 h. The methods were tested using data collected over seven years at six different sites on islands of the South and Central Aegean Sea in Greece.

At the end of 1999, Sfetsos [1] reported a comparison of various forecasting approaches, using time series analysis, on mean hourly wind speed data. In his work includes the traditional linear (ARMA) models and the commonly used feed forward and recurrent neural networks, other approaches are also examined including the Adaptive Neuro-Fuzzy Inference Systems (ANFIS) and Neural Logic Networks.

In 2006, Chan et al. [4] discussed the problem of “over-fitting” and some common generalization learning techniques in the ANN literature, as well as introducing a new Genetic Algorithm-based regularization method called “GARNET” for short-term load forecasting. Their work included four generalization learning techniques, namely: Early-Stopping, Bayesian Regularization, Adaptive-Regularization and GARNET. Those techniques are applied to train Multi-Layer Perceptrons networks (MLP) for day-ahead load forecasting on limited amount of hourly data from a US utility. They showed that forecasters trained by these four methods consistently produce lower prediction error than those trained by the standard error minimization method.

In 2007 Bilgili et al. [5] studied ANN to predict the mean monthly wind speed of any target station using the mean monthly wind speeds of neighboring stations which are indicated as reference stations. They reported that resilient propagation (RP) learning algorithm was applied in their simulation.

Mabel and Fernández [6] reported in 2008 a case study of wind power generation based on ANN. The model reported was developed with the help of neural network methodology. It involves three input variables—wind speed, relative humidity and generation hours and one output variable-energy output of wind farms.

In 2008 Costa et al. [7] made a review of history of wind power short-term prediction. In their work included the first ideas and sketches of the theme to actual state of the art on models and tools.

The techniques for wind speed forecasting suppose that the time series conformed for the measurements are a sum of different patterns and a random error. The objective of the most of the techniques used in forecasting is the patterns separation (horizontal, seasonal, cyclical and trend) that conform the series. Recently diverse statistical techniques and artificial intelligence have been used for the time series forecasting [8–11], even more, some of this techniques have been combined with the purpose of reducing the forecasting errors generating more accuracy predictions [12–14].

There are other statistical techniques denominated exponential smoothing that besides to forecast, smooth the analyzed function, allowing a more convenient presentation of the data and eliminating until certain degree, the random errors that appear. These techniques are based on the intuitive application of movable averages, in where the tool to smooth the function is the mean of the last observations.

The exponential smoothing methods have been used in diverse fields [15,16], however, they practically have not been used in wind speed forecasting [17], in spite of the potential that they have in the forecasting where the last observations have more weight in the forecasting of the present ones, as the case of wind speed short term measurements [18].

The main advantage of the exponential smoothing methods is their robustness [19], that allows a fast and efficient implementation of the technique together with the descriptive and the inferential statistic.

The Instituto de Investigaciones Eléctricas, IIE [20] has carried out studies of the wind potential assessment of México. These studies have been developed through measurement programs in various locations throughout the country. Some important locations are: La Ventosa in Oaxaca [21], Guerrero Negro in Baja California Sur (BCS) [22], Cerro La Virgen in Zacatecas, Laguna Verde in Veracruz, La Rivera Maya in Quintana Roo, and in the village of Chetumal in the same state of Quinta Roo.

### Table 1

<table>
<thead>
<tr>
<th>Specification</th>
<th>Anemometer</th>
<th>Wind Vane</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measuring rank</td>
<td>0.78–45 m/s</td>
<td>0–360</td>
</tr>
<tr>
<td>Exactness</td>
<td>±5%</td>
<td>±5%</td>
</tr>
<tr>
<td>Resolution</td>
<td>0.78 m/s</td>
<td>1 m/s</td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>Data</th>
<th>Min.</th>
<th>Max.</th>
<th>Mean</th>
<th>Median</th>
<th>St. Dev</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4464</td>
<td>6.10</td>
<td>2.25</td>
<td>2.25</td>
<td>1.20</td>
<td>1.40</td>
<td>3.00</td>
</tr>
</tbody>
</table>

Fig. 1. Time Series of the wind speed of Chetumal, Quintana Roo.

Fig. 2. Wind speed frequency and Normal distribution curve.
the wind potential assessments, the most important regions in México are La Ventosa in Oaxaca, the North-Pacific Region (including Baja California and Baja California Sur), the State of Veracruz and the State of Quintana Roo.

Chetumal is the capital city of the Mexican State of Quintana Roo. This city is situated at the eastern coast of the Yucatán Peninsula (Coordinates: 18°30'N, 88°20'W) just north of Belize; it lies only 10 m above sea level. The warm waters of the Caribbean Sea contribute to the climate, which is generally warm and humid. The average temperature range is 25.5–26.5°C, with maximum high temperatures between 36°C and 38°C and low temperatures ranging from 12°C to 14°C. The highest monthly average rainfall occurs in September.

In the south of Chetumal and the Riviera Maya in México, about 36 ecotourism developments represent the main attraction of this area, leaving an important economy. The objective of these projects is to diversify the tourism in the southern state and in parallel to encourage productive projects in communities. Investment funds for ecotourism projects, is the result of agreements with the three levels of government and coordinated by various state institutions such as the Ministry of Planning and Development (Seplader), Ministry of Tourism (Sedetur) and the Ministry of Development Agricultural, Rural and Indigenous (Sedari).

Currently, the association called “The Island of Chetumal”, which brings together local and international investment, is developing various ecotourism projects on the shores of Lake Bacalar and Rio Hondo. The hotels are been developed under the concept of ecotourism which use solar or wind energy to meet the electricity demand. It is possible to consider the use of small wind turbines (20 kW or less) to generate electricity required for the hotels. The analysis and forecasting of wind speed in Chetumal is very important to predict the wind power variations and performance of the individual wind turbines.

In our research, the wind speed and wind direction were measured at 10 m above ground level where the wind speed values were measured with a frequency of 0.5 Hz and the average wind speed was recorded considering regular intervals of 10 min. Sensors with high accuracy and a low starting threshold were used. Anemometers with three cone-shaped cups were used to measure wind speed. This type of anemometer was selected because its design has been shown to exert a more uniform torque throughout a revolution. Conventional wind vanes were used to measure the wind direction or the azimuth angle of the wind. The starting threshold was 0.5 m/s for the anemometer and the wind vane. The wind velocity was recorded using a data acquisition system supplied by a 10 W photovoltaic panel. It is worth mentioning that the average values were calculated for all parameters on a ten-minute basis, which is now the international standard period for wind measurement. Except for wind direction, the average is defined as the mean of all samples. For wind direction, the average should be a unit vector (resultant) value. Average data are used in reporting wind speed variability, as well as wind speed and direction frequency distributions.
For any wind development project, there is a need for information on the 10-minute average wind speed for the development site to determine suitability for wind turbine technologies in accordance with loading guidelines, such as those of the International Electrotechnical Commission (IEC 61400-1 and IEC 61400-2). Short-term wind speed variations of interest include turbulence and gusts. Short-term variations usually mean variations over time intervals of 10 min or less. Ten-minute averages are typically determined using a sampling rate of about 1 s. It is generally accepted that variations in wind speed with periods from less than a second to 10 min and that have a stochastic character are considered to represent turbulence. For wind energy applications, turbulent fluctuations in the flow need to be quantified for the turbine design considerations based on maximum load and fatigue prediction, structural excitations, control, system operation, and power quality.

2. Single exponential smoothing method

In conventional moving average forecast, the mean of the past n observations is used as a forecast. This implies equal weights (equal to 1/n) for all n data points. However, with forecasting, the most recent observations will usually provide the best guide to predict the future. Furthermore a weighting scheme that has decreasing weights as the observations get older can give better results.

If the weights decrease exponentially when the observations get older the exponential smoothing methods can be very useful for forecasting.

Considering that $F_t$ is the forecast at some point of the time series and $Y_t$ is the available observation, thus the forecast errors is found to be:

$$e_t = Y_t - F_t$$  \(1\)

The method of single exponential smoothing method (SES) takes the forecast for the previous period and adjusts it using the forecast error. Therefore, the forecast for the next period is:

$$F_{t+1} = F_t + \alpha(Y_t - F_t)$$  \(2\)

where $\alpha$ is a constant between 0 and 1.

Of this way, it can be seen that the new forecast is simply the old forecast plus an adjustment for the error that occurred in the last forecast. When $\alpha$ has a value close to 1 means that the new forecast will include a substantial adjustment for the error in the previous forecast. Conversely, if $\alpha$ is close to 0 means that the new forecast will include very little adjustment.

3. Statistics measures to determine the accuracy of the forecast

In order to determine quantitatively the best model, five forecast error measures were employed for model evaluation and
model comparison, being these: the mean error (ME), the mean square error (MSE), the mean absolute error (MAE), the mean percentage error (MPE) and the mean absolute percentage error (MAPE).

The standard statistical error measures can be defined as:

\[
\text{MSE} = \frac{1}{n} \sum_{t=1}^{n} e_{t}^2
\]

The mean absolute error as:

\[
\text{MAE} = \frac{1}{n} \sum_{t=1}^{n} |e_{t}|
\]

The mean percentage error as:

\[
\text{MPE} = \frac{1}{n} \sum_{t=1}^{n} PE_{t}
\]

And the mean absolute percentage error as:

\[
\text{MAPE} = \frac{1}{n} \sum_{t=1}^{n} |PE_{t}|
\]

Where \(n\) is the number of periods of time.

4. Data measurement

The Instituto de Investigaciones Eléctricas (IIE) made the measurements of the wind speed for two years from 2004 to 2005, through a network of measurement stations located in the interest places. The sensors were located in the measurement towers at 10 m above the ground 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 for their processing.

The data acquisition measures the wind velocity every 1 s and every 10 min the mean value was estimated. In total 52,560 wind velocity data were measurement through the period.

Because of the amount of data was too big, the sample was divided per month given a total of 4464 data. Fig. 1 shows the time series of the wind speed for January which is taken as an example for the analysis.

5. Data analysis

Table 2 shows the statistics parameters of the wind speed measurements in Chetumal. From the central tendency parameters it can be seen that the sample has a Normal behavior.

The frequency distribution and Normal behavior curve of the data are shown in Fig. 2. It can be observed that the wind velocities varied in a range from 0 m/s to 6 m/s. The 75% of the wind velocities are over 1.3 m/s.

Fig. 3 shows the data spread by means of a box diagram. This figure shows the quartiles and it can be seen that the median is in the middle of the rectangle showing the symmetric of the data. Also it can be observed a higher spread in the upper quartile.

From the analysis of Fig. 3 it is clear that the time series should be divided into for sections as it is shown in Fig. 4 in order to develop separated models.

To continue with the data analysis, the moving average centered technique was applied for smoothing the data to find and to identify tendency, trend-cycle or seasonal components.

Fig. 5 shows a plot of the moving average centered in one thousands data. It can be observed that the lower wind velocities
occur at the beginnings and end of year and the maximum velocities in the middle of this one.

Figs. 6 and 7 show the moving average plots centered in one thousands five hundred and two thousands five hundred respectively. In Fig. 6 it can be seen a negative tendency of the series and Fig. 7 a regular behavior of this one.

Finally, Fig. 8 shows the overall behavior of the series through the monthly averages. It can be observed that January, November and December are the months with the lowest average wind velocities. March, April, May, June and July with the higher average velocities and in particular June with the highest average velocities.

As it was observed from the previous analysis, there exist in some cases a trend in downward direction of the data, for this reason the mean method is not a precise method to be used for forecasting purposes, furthermore the exponential smoothing method was utilized.

In order to do the forecasting, a data sample of wind velocities every 10 min was taken from the last day of the time series. Fig. 9 shows the time series of the data selected.

Fig. 10 shows the forecasting of the selected data generated by Eq. (2) for \( \alpha \) values of 0.1, 0.5 and 0.9 compared with the real series of time. It can be observed that for the \( \alpha \) value of 0.1 just a rough tendency was obtained, however for \( \alpha \) values of 0.5 and 0.9 the forecasting shows good results.

In order to find quantitatively the best value of \( \alpha \) that minimizes the differences between the real data and the forecasting values, the SES method was used changing the \( \alpha \) value from 0.1 to 0.9 in Eq. (2) and the five statistical error measures were calculated. Table 3 shows the statistical error measures and it can be observed that the \( \alpha \) value of 0.9 minimizes the error measurements.

It is important to point that there exists another exponential methods for forecasting such as the adaptive-response-rate single exponential smoothing (ARRSES) which allows the \( \alpha \) value to be modified in a controlled manner, as changes in the pattern of data occur, however, this method was not selected in the present study due to its higher complexity.

In order to compare the results obtained with the SES with other forecasting methods, the artificial neural network (ANN) model proposed by Cadenas and Rivera [9] was utilized with the same data input.

Fig. 11 compares the SES and ANN methods against the real data. In this figure it can be seen qualitatively that both methods adjust quite well to the real data, however, it is not possible to see which of the two models is better, furthermore both models were evaluated obtaining their statistical error measurements.

Table 4 shows the five statistical errors measures for the SES and ANN models and it can be seen that in all the cases the SES has the lowest error measures.

Fig. 12 compares the exponential smoothing method and the artificial neuron network methods against the last real 20 data of the short term wind speed forecasting in Chetumal. It can be seen that the SES method adjusts better to the real data and only a small delay can be observed between the two curves. The ANN follows the same tendency than the real data however the difference of the wind speed values are higher than with the SES method.

6. Conclusions

In this paper the analysis and forecasting of wind velocities in Chetumal, Quintana Roo, Mexico was presented. First a statistical analysis of the time series was made in the first part of the paper through conventional and robust measures. Also the forecasting of the last day of measurements was made utilizing the exponential smoothing method. The results showed that the exponential smoothing method is a good alternative for the wind forecasting for \( \alpha \) values close to 1 for the analyzed data. For short term wind speed forecasting this technique responds of a satisfactory way to the necessities of precision and accuracy required to support the operators of the Electric Utility Control Centre. Because of its simplicity and exactness, compared with another techniques such ANN, this technique showed to be very useful for wind speed forecasting.

Acknowledgments

We thank technical support from José de Jesús Quiñones Aguilar and Carlos Alberto Perez-Rábago. This work has been partially supported by DGAPA-UNAM through Grant PAPIIT IN-106207-3.

References