



## Use of Bayesian networks classifiers for long-term mean wind turbine energy output estimation at a potential wind energy conversion site

José A. Carta<sup>a,\*</sup>, Sergio Velázquez<sup>b</sup>, J.M. Matías<sup>c</sup>

<sup>a</sup> Department of Mechanical Engineering, University of Las Palmas de Gran Canaria, Campus de Tafira s/n, 35017 Las Palmas de Gran Canaria, Canary Islands, Spain

<sup>b</sup> Department of Electronics and Automatics Engineering, University of Las Palmas de Gran Canaria, Campus de Tafira s/n, 35017 Las Palmas de Gran Canaria, Canary Islands, Spain

<sup>c</sup> Department of Statistics, University of Vigo, Lagoas Marcosende, 36200 Vigo, Spain

### ARTICLE INFO

#### Article history:

Received 20 January 2010

Accepted 4 September 2010

#### Keywords:

Bayesian network

Method of bins

Wind speed

Short-term records

Long-term records

Measure–correlate–predict

### ABSTRACT

Due to the interannual variability of wind speed a feasibility analysis for the installation of a Wind Energy Conversion System at a particular site requires estimation of the long-term mean wind turbine energy output. A method is proposed in this paper which, based on probabilistic Bayesian networks (BNs), enables estimation of the long-term mean wind speed histogram for a site where few measurements of the wind resource are available. For this purpose, the proposed method allows the use of multiple reference stations with a long history of wind speed and wind direction measurements. That is to say, the model that is proposed in this paper is able to involve and make use of regional information about the wind resource. With the estimated long-term wind speed histogram and the power curve of a wind turbine it is possible to use the method of bins to determine the long-term mean energy output for that wind turbine. The intelligent system employed, the knowledgebase of which is a joint probability function of all the model variables, uses efficient calculation techniques for conditional probabilities to perform the reasoning. This enables automatic model learning and inference to be performed efficiently based on the available evidence. The proposed model is applied in this paper to wind speeds and wind directions recorded at four weather stations located in the Canary Islands (Spain). Ten years of mean hourly wind speed and direction data are available for these stations. One of the conclusions reached is that the BN with three reference stations gave fewer errors between the real and estimated long-term mean wind turbine energy output than when using two measure–correlate–predict algorithms which were evaluated and which use a linear regression between the candidate station and one reference station.

© 2010 Elsevier Ltd. All rights reserved.

### 1. Introduction

As stated by Hiester and Pennell [1], the interannual variability of wind speed at a potential wind energy conversion site is a very important issue. The energy that can be obtained with a wind turbine and, therefore, the economic feasibility of the project, will be very sensitive to the wind speed acting on the wind turbine over the working life of the facility.

On occasions no historical series of wind data measurements for a potential candidate site are available that would enable an assessment of the economic feasibility of the investment planned for the installation of a Wind Energy Conversion System (WECS). One option that can be employed to make up for the drawback of a lack of such data is to undertake a wind data measurement campaign over a sufficiently long period of time. According to Hiester and Pennell [1], it is difficult to accurately estimate the mean values of the behaviour of the wind at a potential candidate

site with less than 10 years worth of data. This option therefore entails an increase in costs to cover the measurement campaign and, importantly, postponing the decision-taking process for a normally unacceptably length of time.

Another option entails the estimation by statistical methods of the interannual mean wind behaviour at the candidate site. Such methods rely on the existence of reference climatological stations installed at nearby sites, for which long-term wind resource measurements are available. These methods also require a wind data measurement campaign to be undertaken at the candidate site for a relatively short period of time (normally 1 year). Additionally, a portion of the reference station's wind data has to coincide in terms of length of time and dates with the measurement period of the candidate site.

For several authors [2–9], the first concern about a site under consideration for a wind power station is with the long-term (many years) mean wind speed. With this in mind, various methods have been proposed to estimate the long-term mean wind speed. Putnam [2] used a procedure that estimated the long-term mean wind speed  $\bar{V}_c^{LT}$  at a candidate site that was based on knowledge of the

\* Corresponding author. Tel.: +34 928 45 14 83; fax: +34 928 45 14 84.

E-mail address: [jcarta@dim.ulpgc.es](mailto:jcarta@dim.ulpgc.es) (J.A. Carta).

mean wind speed determined over a short-term period  $\overline{V}_c^{ST}$ . This method is known as the method of ratios [3–5] and is given by Eq. (1)

$$\overline{V}_c^{LT} = \overline{V}_c^{ST} [\overline{V}_r^{LT} / \overline{V}_r^{ST}] \quad (1)$$

$\overline{V}_r^{LT}$  is the long-term mean for the reference site and  $\overline{V}_r^{ST}$  is the short-term mean for the reference site (for the period coinciding with that of  $\overline{V}_c^{ST}$ ).

Conrad and Pollak [5] listed the requirements that need to be met so that the method of ratios can be applied. The main requirement is equivalent, from a mathematical point of view, to the demand for a high spatial correlation between the candidate and reference site wind speeds. Corotis [1,3,4,6,7], using as a basis the hypothesis that the mean wind speeds at the candidate and reference site are distributed in accordance with a normal law, proposes a method that explicitly includes the spatial correlation coefficient ( $\rho^{LT}$ ) between both sites and the standard deviations for the candidate site  $\sigma_c^{LT}$  and for the reference site  $\sigma_r^{LT}$ , Eq. (2)

$$\overline{V}_c^{LT} = \overline{V}_c^{ST} + \rho^{LT} [\overline{V}_r^{LT} - \overline{V}_r^{ST}] [\sigma_c^{LT} / \sigma_r^{LT}] \quad (2)$$

The values of  $\rho^{LT}$ ,  $\sigma_c^{LT}$  are unknown and have to be estimated. Working on the hypothesis that the correlation varies little over time, the short-term rather than long-term correlation is normally used. Daniels and Schroeder [6] used a series of hypotheses to estimate  $\sigma_c^{LT}$ .

Other methods have been proposed in the scientific literature to tackle the problem arising from a site with few measurements of the wind resource [10–13]. However, the statistical method that has been most widely used over the last 20 years has been the measure–correlate–predict method (MCP) [14–18]. There are several variants of the method, but they are all based on the existence of statistical relationships between two time-series data sets and try to model the relationship between wind data (speed and direction) measured at the candidate site over a short period of time (normally 1 year) and wind data simultaneously measured at a reference site. By introducing into the model the long-term wind data (normally from several years) measured at the reference site, the long-term wind data series are estimated for the candidate site.

In the last few decades, the ‘data mining’ has become an increasingly common term [19] used to refer to the automatic or semi-automatic (assisted) process of extracting useful and understandable knowledge from a dataset [20]. Data mining incorporates numerous data analysis and model extraction techniques. This paper proposes the use of one of these techniques, probabilistic Bayesian networks (BNs) [19,21–23], to estimate the long-term mean wind speed frequency distribution (called a histogram) for a site with few measurements of the wind resource. The method proposed enables the use of multiple reference stations with a long history of wind speed and direction measurements. That is, the proposed model is able to make use of and involve regional information of the wind resource. The model has been implemented in Weka (Waikato Environment for Knowledge Analysis), free software available under the GNU General Public License [19] and developed by Waikato University (New Zealand). The intelligent system employed, the knowledgebase of which is a joint probability function of all the model variables, uses efficient calculation techniques for conditional probabilities to perform the reasoning. This enables automatic model learning and inference to be performed efficiently based on the available evidence. With the estimated long-term wind speed histogram and the power curve of a wind turbine it is possible to use the method of bins [24] or a probability density function that is fitted to the histogram [25–27] to determine the long-term mean energy output for that wind turbine. The proposed model is applied in this paper to wind speeds and wind directions recorded at four weather stations located on

the Canary Islands (Spain). Ten years of mean hourly wind speed and direction data are available for these stations. To estimate the goodness-of-fit of the proposed estimation models a comparison has been made between the errors generated with these models and those generated with two measure–correlate–predict (MCP) algorithms which have been evaluated [13,14]: a linear regression method and a vector regression method.

For this purpose different metrics have been used. These metrics characterise the estimation of (1) the correct cumulative relative frequency histogram of the wind speed, (2) the correct mean wind power density and (3) the correct mean annual energy production. For this final analysis we have also used two commercial wind turbines of 330 and 800 kW rated power [28].

## 2. MCP methods used for comparison

The methods of Derrick [14] and Niesel et al. [13] have been chosen to compare their results with the alternative method based on BN as proposed by the authors.

### 2.1. Derrick method

Derrick [14] uses a linear regression to characterise the long-term (LT) relationship between the wind speeds of the candidate site ( $c$ ) and of the reference site ( $r$ ), Eq. (3). This method will be referred to in this paper as the linear regression method (LR).

$$(v_j)_c^{LT} = a(v_j)_r^{LT} + b \quad (3)$$

$(v_j)_c^{LT}$  are the estimated wind speeds at the candidate site,  $(v_j)_r^{LT}$  are the measured wind speeds at the reference site,  $a$  and  $b$  are the slope and ordinate at the offset obtained from the linear regression [29], Eq. (4).

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n (v_i)_r^{ST} & \sum_{i=1}^n (v_i)_c^{ST} \\ \sum_{i=1}^n (v_i)_r^{ST} & n \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^n (v_i)_c^{ST} (v_i)_r^{ST} \\ \sum_{i=1}^n (v_i)_c^{ST} \end{bmatrix} \quad (4)$$

In Eq. (4),  $(v_i)_r^{ST}$  and  $(v_i)_c^{ST}$  are the short-term (ST) mean hourly wind speeds at the reference site and the candidate site, respectively.  $n$  is the number of wind speed data.

Generally, the MCP method ‘bins’ or groups together the hourly wind speeds in terms of the wind direction at the reference site. For this purpose the data are divided into a specific number of wind direction sectors (normally 8 sectors of 45° or 12 of 30°). Then, the parameters  $a$  and  $b$  of the model are estimated, Eq. (4), for each wind direction sector. That is,  $a$  and  $b$  depend on the wind speed and wind direction. Thus,  $n$  in Eq. (4) is the number of wind speeds in the wind direction sector of the reference site that is being analysed.

The Mathcad Software 2001i of MathSoft Engineering & Education, Inc. [30] was used for the programming of this method.

### 2.2. Niesel et al. method

Niesel et al. [13] propose decomposing the wind speeds of the reference site and candidate site using two Cartesian axes ( $x, y$ ) [31] and performing a two-dimensional linear regression, Eq. (5). In this paper, this method is referred to as the vector regression method (VR).

$$\begin{bmatrix} \{(v_x)_i\}_c^{LT} \\ \{(v_y)_i\}_c^{LT} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \{(v_x)_i\}_r^{LT} \\ \{(v_y)_i\}_r^{LT} \end{bmatrix} \quad (5)$$

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات