Renewable Energy 85 (2016) 790–809

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene

Wind speed forecasting for wind farms: A method based on support vector regression

G. Santamaría-Bonfil [a, *], A. Reyes-Ballesteros [b], C. Gershenson [a]

[a] Departamento de Ciencias de la Computación, Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS), Universidad Nacional Autónoma de México (UNAM), Circuito Escolar S/N, Ciudad Universitaria, Coyoacan, D.F., 04510, Mexico

[b] Instituto de Investigaciones Eléctricas (IIE), Reforma 113, Col. Palmira, Cuernavaca, Morelos, 62490, Mexico

ARTICLE INFO

Article history:
Received 5 September 2014
Received in revised form
24 June 2015
Accepted 4 July 2015
Available online xxx

Keywords:
Wind speed forecasting
Phase space reconstruction
Support vector regression
Genetic algorithms
Non-linear analysis

ABSTRACT

In this paper, a hybrid methodology based on Support Vector Regression for wind speed forecasting is proposed. Using the autoregressive model called Time Delay Coordinates, feature selection is performed by the Phase Space Reconstruction procedure. Then, a Support Vector Regression model is trained using univariate wind speed time series. Parameters of Support Vector Regression are tuned by a genetic algorithm. The proposed method is compared against the persistence model, and autoregressive models (AR, ARMA, and ARIMA) tuned by Akaike's Information Criterion and Ordinary Least Squares method. The stationary transformation of time series is also evaluated for the proposed method. Using historical wind speed data from the Mexican Wind Energy Technology Center (CERTE) located at La Ventosa, Oaxaca, México, the accuracy of the proposed forecasting method is evaluated for a whole range of short termforecasting horizons (from 1 to 24 h ahead). Results show that, forecasts made with our method are more accurate for medium (5–23 h ahead) short term WSF and WPF than those made with persistence and autoregressive models.

1. Introduction

Wind Speed Forecasting (WSF) is particularly important for wind farms due to cost-related issues, dispatch planning, and energy markets operations [1,2]. These predictions are employed for optimal operation policies and operative costs [3,4], load balancing [1,5], site and capacity planning [6,7], and unit commitment for electricity markets [1–3]. Tipically, wind farm energy production is estimated using a fixed weighted measure of the wind farm's nominal power and forecasts from historical atmospheric data [8,9]. Further, it has been stated that wind speed is one (if not the most) important variable related to wind power generation [10]. Fig. 1 displays the power curve related to wind power generation for CERTE's wind turbine.[1] While energy demand can be forecasted, inaccurate WSF will become a potential point of failure when scheduling generation units (*i.e.* ramp rates) to satisfy energy demand [2,11,12]. Evenmore, WSF is of such criticallity that, in countries with large wind power generation, producers have the legal requirement to provide the energy markets with short and mid-term production forecasting [13].

Recently, Support Vector Regression (SVR) has been used for prediction of wind speed and other atmospheric variables with positive results [4,14–21]. SVR is based on the Structural Error Minimization principle; it is also equipped with the 'Kernel Trick' and other optimization features which allow it to perform a noise-robust, non-linear regression. SVR stability and accuracy depend on several aspects, some of the most important are *Parameter Tuning* (PT) and *Feature Selection* (FS). The former is a procedure which consists in properly selecting the kernel function and its parameters, and the penalization term [22]. Commonly tuned by an exhaustive search technique, deterministic and stochastic methods have also been proposed, where Genetic Algorithms (GA) have obtained good results. The latter consist in selecting the most important model variables to describe process behavior [23]. In the current setup, one is faced with the problem to estimate wind speed behavior as accurately as possible from only measures of itself. Typically, autoregressive models are used as a statistical proxy of dynamical systems by employing as variables past

* Corresponding author.
E-mail addresses: guillermo.santamaria@iimas.unam.mx (G. Santamaría-Bonfil), areyes@iie.org.mx (A. Reyes-Ballesteros), cgg@unam.mx (C. Gershenson).

[1] http://www.meti.go.jp/meti_lib/report/2012fy/E003282.pdf, last accessed: February 27, 2015.

http://dx.doi.org/10.1016/j.renene.2015.07.004
0960-1481/

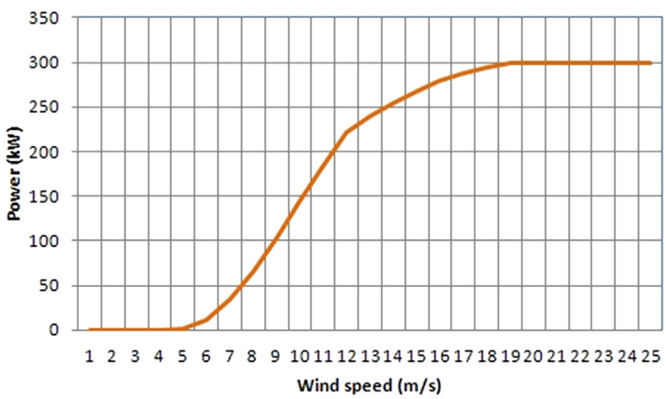

Fig. 1. KWT-300 power curve.

observations and stochastic shocks. From this family type of models, one which is employed to analyze non-linear chaotic univariate time series is Time Delay Coordinates (TDC) [24]. The embodied philosophy of TDC is that the non-measured variables of the system can be recovered from those measured, due the influence of the former over the latter [25]. If the studied process is chaotic, by employing the TDC model and the Phase Space Reconstruction (PSR) procedure, an approximate reconstruction of the studied phenomenon feature space can be obtained from a univariate time series [10,24,26,27].

This paper proposes a new algorithm to the short-term WSF problem based on SVR. The algorithm developed here, named $PSR-SVR_{GA}$, uses the TDC model and the PSR procedure as an FS technique. Then, a genetic algorithm which uses the GA Boltzmann selection method [22] is employed to tune the SVR parameters. The proposed algorithm quality is compared against the Persistence method (PM) and classical time series models: AutoRegressive (AR), AutoRegressive Moving Average (ARMA), and AutoRegressive Integrated Moving Average (ARIMA). AR-like models were tuned by identifying the autoregressive and moving average orders through Akaike's Information Criteria (AIC). Then, order weights were optimized by the Ordinary Least Squares (OLS) method. Additionally, time series are integrated to ensure stationarity; transformed data is used by ARIMA and a variation of the proposed algorithm. The accuracy of the methods is analyzed in terms of WSF and Wind Power Forecasting (WPF). On one hand, WSF methods performance is evaluated based on five statistical measures: the Mean Absolute Error (MAE), Mean Bias Error (MBE), Root Mean Squared Error (RMSE), Mean Absolute Scaled Error (MASE), and Directional Accuracy (DA). On the other, WPF is analyzed in terms of the Normalized Mean Bias Error (NMBE), Normalized Mean Absolute Error (NMEA), and the Normalized Root Mean Squared Error (NRMSE). According to the analysis of the obtained results, the best model produced by the hybrid GA method is, in general, better to forecast wind speed and wind power than persistence method and AR and ARMA models.

Summarizing, the main contributions of our WSF methodology are:

- The usage of a non-linear method called PSR, which is designed to analyze and describe chaotic phenomena.
- A Genetic Algorithm is employed to select from a pool of kernel functions the most adequate function for WSF altogether with its parameters.
- A chaotic and complex analysis was performed over wind speed data to corroborate the chaotic nature of wind data, and therefore validating the usage of the PSR procedure.
- Further, we studied the influence of differentiation as a preprocessing treatment over the forecasting performance of the proposed method.
- A rigorous analysis was performed under a framework composed of WSF and WPF quality metrics.

This paper is organized as follows. Section 2 describes WSF classical time series methods and SVR state of the art. Section 3 presents the proposed method: first, the need for a forecasting methodology while using SVR is presented; next, the feature selection problem and how the PSR method is used is described; then, SVR parameter tuning and the hybrid genetic method are detailed. Section 4 presents the data description, the experimental setup, and our results. Section 5 presents the conclusions of this work. A nomenclature listing the abbreviations used is included before the references.

2. Background

WSF models are usually divided into physical-based models and statistical models [11]. The former are based on numerical weather models which employ several equations to describe the governing motions and forces affecting fluids. The latter analyze previous wind patterns over time and extrapolate them to predict future wind behavior. The scope of this work focuses on statistical methods.

2.1. Persistence models

Before we continue, its necessary to introduce the benchmark method for WSF, the Persistence model. PM states that due the high autocorrelation underlying WS behavior, any wind speed future value is equal to its last known value [28]. Despite its simplicity, PM achieves very good results in the WSF problem and is used to compare the quality of new forecasting approaches [2,29]. Typically, PM predicts a future WS value as $\hat{x}_{t+h} = x_t$, where $\hat{x}$ stands for the forecasted value, t for the current time step, and h for the forecasting horizon. In the case where the day-ahead forecasting is required, a persistence method called Day-to-Day (D2D) is used [28]. D2D method forecast a future value as $\hat{x}_{(d+1,h)} = x_{(d,h)},\ h = 1,\dots,24$, where d stands for the current day.

2.2. Classical time series forecasting models

Autoregressive (AR) models are commonly used for time series forecasting since they are able to capture persistence in a time series [30]. In simple terms, an $AR(p)$ model relates p past observations to the current value x_t as:

$$x_t = \mu + \sum_{i=1}^{p} \varphi_i x_{t-i} + \varepsilon_t, \tag{1}$$

where μ is the mean value, φ_i is a coefficient which reflects each past observation x_{t-i} influence on current value, and ε_t is the actual stochastic perturbation [30].

2.2.1. ARMA and ARIMA

AR models have been extended for more robust versions like the *Autoregressive Moving Average* models (ARMA) and the *Autoregressive Integrated Moving Average* models (ARIMA). These type of models, describe a univariate time series as the relation between