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/

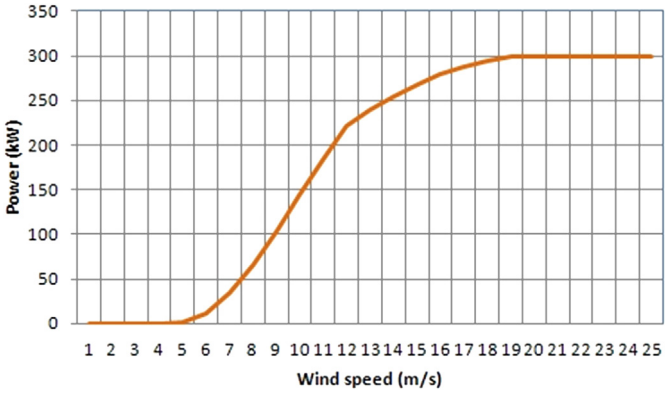

Fig. 1. KWT-300 power curve.

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 $\hat{x}_{(d+1,h)} = x_{(d,h)}$, $h = 1, \ldots, 24$, 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

actual observations x_t respect p AR components and a Moving Average (MA) process [30].

Altogether, an *ARMA*(p,q) model, is defined as:

$$x_t = \mu + \sum_{i=1}^{p} \phi_i x_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t, \tag{2}$$

where ϕ_i and θ_i are weight coefficients which reflect the influence of past p observations and q stochastic perturbations on the current value.

ARIMA models proposed by Box et al. [30] are an ARMA extension to deal with non stationarity in data by making data ergodic through d differentiation steps; then, an ARMA model like 2 is estimated and used. Therefore, the model is denoted as *ARIMA*(p,d,q).

Moreover, in accordance to the Box–Jenkins (B–J) methodology [30], an optimal ARIMA model is obtained by the optimization of (p, q) orders. In this paper, AR and MA orders were approximated by Akaike's Information Criterion [31].

The first kernel is parameterless, the second employes parameter d to determines the polynomial degree and r as a constant. The third function is a radial function, where γ is a scaling factor of the Euclidian distance between patterns.

2.3. Support vector regression

Recently, more robust semi-parametric methods like SVR have been successfully applied to the prediction of WS and other time series [4,14–22]. SVR, an extension of Support Vector Machines (SVM), was proposed by Drucker et al. [32]. SVR pursues the best trade-off between the model's Empirical Error and the model complexity [33]. This compromise is achieved by constraining SVR regression function $f(\cdot)$ to the hyperplanes function class, and employing a margin, also called *insensitive tube*, around the hyperplane. Moreover, $f(\cdot)$ only depends on a reduced set of the training data called the *Support Vectors* (SV), those which correspond to the active constraints in the optimization problem.

Formally, given a data set of the form $(\mathbf{x}_i, y_i) \in \mathbb{R}^N \times \mathbb{R}$, the SVR dual optimization problem is formulated as:

$$\begin{aligned} &\textit{maximize}\ \ W(\alpha,\alpha^*) = -\frac{1}{2}\sum_{i,j=1}^{m}(\alpha_i-\alpha_i^*)\left(\alpha_j-\alpha_j^*\right)\left\langle \phi(x_i),\phi(x_j)\right\rangle - \varepsilon\sum_{i=1}^{m}(\alpha_i^*+\alpha_i) + \sum_{i=1}^{m}(\alpha_i^*-\alpha_i)y_i, \\ &\textit{Subject to}\sum_{i=1}^{m}(\alpha_i-\alpha_i^*) = 0 \\ &0 \le \alpha_i \le C,\ \forall i = 1,\ldots,m \\ &0 \le \alpha_i^* \le C,\ \forall i = 1,\ldots,m, \end{aligned} \tag{3}$$

where C is the complexity penalization term, and α,α^* correspond to the dual variables for the active constraints [33].

Evenmore, SVR is able to perform a non-linear regression due the *kernel trick*. Colloquially, it consist in providing SVR with a specific kernel function which maps data from the *input space* to a high dimensional *feature space* where a linear regression is performed. Typical kernel functions are Linear, Polynomial, and Gaussian, those which are depicted in Table 1.

Once Eq. (3) is solved and the hyperplane function found, a future value can be predicted employing Eq. (4).

$$f(\mathbf{x},\alpha,\alpha^*) = \sum_{i=1}^{s}(\alpha_i-\alpha_i^*)k(\mathbf{x}_i,\mathbf{x}) + b. \tag{4}$$

Table 1
Classical kernel functions.

Linear	$K_{linear}(\mathbf{x}_i,\mathbf{x}_j)=\mathbf{x}_i{}^*\mathbf{x}_j$
Polynomial	$K_{polyn}(\mathbf{x}_i,\mathbf{x}_j)=((\mathbf{x}_i{}^*\mathbf{x}_j)+r)^d$
Gaussian	$K_{Gauss}(\mathbf{x}_i,\mathbf{x}_j)=exp(-\gamma\|\mathbf{x}_i{}^*\mathbf{x}_j\|^2)$

3. Proposed method *PSR–SVR$_{GA}$*

In the current WSF setup, we want to fit a hyperplane (i.e. SVR model) to wind speed data in order to use it as a proxy of the evolution of WS phenomenon. Even while SVR is one of the most renowned machine learning methods, there are several opportunity areas to increase the model stability and accuracy. For example: the selection of relevant variables, or the selection of a kernel map and its parameter tuning. Moreover, typical machine learning and classical B–J methodologies consider one or more of these procedures: Data pre-processing, Feature Selection, Model Parameter Tuning, and Model Validation [10,19,21,30,34]. Inspired on these previous works, and given wind speed characteristics like its non-linearity, non-stationarity, high fluctuations and irregularity, the methodology *PSR–SVR$_{GA}$* is proposed.

As the classical B–J methodology, *PSR–SVR$_{GA}$* consists of four steps. First, data pre-processing following human expert heuristics, and data scaling is applied. Then, a chaos theory model called Time Delay Coordinates (TDC) is used to represent wind speed phenomena. Next, a proper TDC model is estimated by means of Phase Space Reconstruction (PSR). SVR model is trained, validated, and tuned using a Genetic Algorithm (GA) [22]. Finally, the optimal SVR model is used for WSF. *PSR–SVR$_{GA}$* is shown in Fig. 2. Subsection 3.1 presents the details of the TDC model and the PSR process. Subsection 3.2 briefly describe the GA and the genetic operators used.

3.1. Feature selection through chaos theory

Classical univariate autoregressive models establish that any phenomenon can be expressed as linear combination of its own past values given that the studied stochastic process is stationary, weakly dependent, and homoscedastic [30]. However, it has been documented that wind speed is heteroscedastic, non-stationary, and highly non-linear [2,5,13,19,35,36]. Furthermore, wind speed behavior is not only affected by itself, several other atmospheric, geographical, and physical variables influence its evolution [2,37].

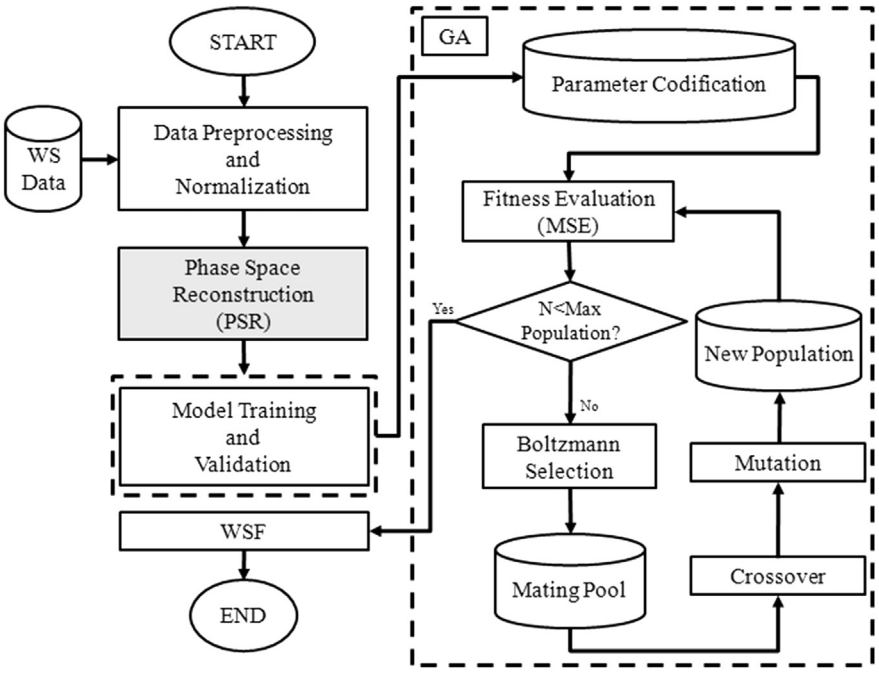

Fig. 2. The diagram presents the $PSR{-}SVR_{GA}$ method. First data is loaded, preprocessed, and normalized; then, the Phase Space Reconstruction method is applied to embed the univariate treated data into a high-dimensional feature space. Next, an SVR model is trained and tuned through a Genetic Algorithm. Lastly, WSF is performed for the specified horizon.

Recently, methods based on Chaos Theory (CT) have been proposed for WSF [10,27]. The latter provide a framework to analyze and describe non-linear phenomena which display a chaotic behavior. A system can be defined as deterministic chaotic if it is sensitive to initial conditions, aperiodic, and bounded [26]. Sensitivity to initial conditions can be defined as having at least one positive Lyapunov exponent [26,38], i.e. close trajectories diverge faster than exponentially. PSR is a CT method which elaborates models and theoretical constraints in order to reconstruct a system's inaccessible internal state from only one component. The most used model for the reconstruction of phase space is Time Delay Coordinates (TDC). According to [39] the TDC parameters can be obtained by the Mutual Information method, and False Nearest Neighbors.

The reconstruction constraints, the TDC model, and the methods to estimate its parameters are detailed in the next subsections.

3.1.1. Takens' theorem, time delay coordinates, and phase space reconstruction procedure

A univariate time series $\{x_i\}_{i=1}^{N}$ per se do not detail the whole system state, still, if it satisfies the chaotic definition and available data is long enough, it can be used to approximate the process state space. In accordance to Takens' Theorem [26,27], this reconstruction is topologically equivalent if it satisfies the relation $d_e{>}2m{+}1$, where d_e is the embedding dimension and m is the true process dimensionality.

If Takens' theorem is met, a $\mathbb{R}^{d_e}$ phase space may be reconstructed through the use of an autoregressive model called Time Delay Coordinates. In accordance with Takens', TDC allows to embed a time series in a higher dimensional feature space, by mapping the univariate time series into M vector states of the form:

$$\mathbf{x}_t = \left[x_t, x_{t-\tau}, \ldots, x_{t-(d-1)\tau}\right]^{d_e} t = 1, \ldots, M, \tag{5}$$

where τ is a sampling factor, and d_e denotes the number of variables of the reconstructed space.

Therefore, the PSR procedure is defined as finding the appropriate values of τ and d_e, in order to reconstruct a topological equivalent space of the data.

3.1.2. TDC delay factor τ via mutual information

A time delay factor τ is employed to map univariate data into a higher dimensional state space where each point is Independent and Identically Distributed (I.I.D.) [26,34]. However, if τ is too small data points in the new space will be highly correlated and mutually independence can not be assured. Still, if τ is too large, data points independence will be trivial [26]. Moreover, the optimal sampling frequency increases smoothness in data embedding and the identification of d_e.

The most common method used for the estimation of τ is called Mutual Information (MI) [26]. This technique is based on Shannon's entropy which is used to quantify information gain among two random variables. In this sense, original series $\{x_t\}_{t=1}^{N-\tau}$ is compared with a τ-delayed version of it $\{x_{t+\tau}\}_{t=1}^{N-\tau}$, and the information between these time series is calculated as

$$I(\tau) = \sum_{t=1}^{N-\tau} P(x_t, x_{t+\tau}) log_2 \left[\frac{P(x_t, x_{t+\tau})}{P(x_t)P(x_{t+\tau})}\right], \tag{6}$$

where $P(x_t,x_{t+\tau})$ is the joint probability of events x_t and $x_{t+\tau}$, and $P(x_t)$ and $P(x_{t+\tau})$ correspond to the marginal probabilities. Eq. (6) is then iterated for $1{\leq}\tau_i{\leq}\tau_{max}$. Finally, MI determines that the optimal τ value is the first minimum value of $I(\tau)$.

3.1.3. TDC embedding dimensions d_e via false nearest neighbors

Typically, an optimal PSR d_e parameter is estimated by successively embedding data into higher dimensional state spaces, and then, checking result consistency [26]. In this regard, False Nearest Neighbors (FNN) is the most common method used for the estimation of d_e [39]. The FNN procedure consists of comparing each

embedded point against its nearest neighbor for d_e and d_{e+1}. The idea is that if two points are *true* neighbors in d_e, they will continue to be in d_{e+1}. Moreover, FNN establishes that d_e^* is optimal if it minimizes the number of false neighbors in the reconstructed space. As in the case of the delay factor, a very small d_e will occlude original data behavior, while a very large d_e will destroy any relationship among data. Therefore, the first FNN minimum is used as d_e^*.

FNN has two criteria for counting points as false neighbors, if any of these is not met, points are considered to be *false* neighbors. FNN criteria are explained below:

1. Calculate the Euclidian distance for a point $\mathbf{x}_i$ and its nearest neighbor $\mathbf{x}_j^{NN}$ embedded in d_e, $\Delta_1 = \left\|\mathbf{x}_i - \mathbf{x}_j^{NN}\right\|^2$. Then, both vectors are embedded into a d_{e+1} space, and its Euclidian distance computed. If the normalized difference between Δ_1 and Δ_2 is above a threshold ϵ, points are considered *false* neighbors, else the next criterium requires to be satisfied.

$$C_1 = \sqrt{\frac{\Delta_2 - \Delta_1}{\Delta_1}} = \frac{|y_{t+d_\tau} - y_{t'+d_\tau}|}{\Delta_1} \geq \varepsilon. \tag{7}$$

2. If the difference between two points embedded on dimension d_e is beyond data standard deviation, these are considered to be *false* neighbors

$$\frac{\Delta_2}{\sigma_A} \geq A_{Tol}, \tag{8}$$

where σ_A is the standard deviation of time series, and A_{Tol} is a threshold employed to describe model's attractor size.

3.2. Parameter tuning through genetic algorithms

It has been documented that stability and accuracy of SVR highly depends on its parameter tuning [21,22]. Commonly, SVR is tuned by a brute force search method called Grid Search (GS). However, GS suffers of a high computational cost, *a priori* problem-knowledge requirements, and it is inefficient for tuning more than three parameters. Recently, Genetic Algorithms (GA) [40] have been proposed succesfully for the SVR parameter tuning problem [21,22].

GA are well-known optimization methods inspired by Darwinian evolution [40]. Moreover, GA have a good performance over problems with non-linear fitness landscapes. Typically, a GA is composed by four genetic operators: a fitness function, a selection method, a crossover and a mutation operator; our proposed method employ these four operators. First, SVR parameters which are the kernel function and its parameters, and the C trade-off constant (see Table 1) are coded into a hybrid chromosomal structure like the one proposed in Ref. [22]. Then, a random population is generated and the fitness of each SVR configuration is calculated. Using Cross Validation, a statistical method for testing the generalization capabilities of a certain model over a data set, altogether with the Mean Squared Error ($MSE = \frac{\sum_{i=1}^{n}(y_t - \widehat{y}_t)^2}{n}$, where y_t is the observed wind speed and $\widehat{y}_t$ is the forecasted wind speed) the quality of each solution is obtained. In accordance to literature [22], and the available data, a 10-fold cross validation is used. Next, the Boltzmann Selection (BS) operator is employed to choose the surviving solutions of the current population [22]. BS is a selection operator which is based on the Boltzmann distribution and a linear cooling schedule. By relating the current GA generation with the system's temperature, BS allows to employ Simulated Annealing optimization criteria into GA allowing bad solutions to be accepted in order to escape from local optimums. Lastly, an n-points crossover and a uniform mutation operators are used to form the new population. This procedure is iterated until a stopping criterion is satisfied.

Chromosomal structure and the Boltzmann operator are shown in Figs. 3 and 4, respectively.

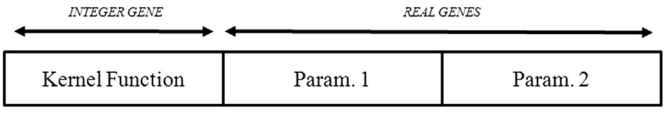

Fig. 3. The proposed GA chromosome structure. SVR parameters are coded into this structure: first, the kernel type is coded in the integer gene; then, kernel's parameters are coded into the real genes.

4. Experiments

This section presents the description of the experimental setup, results, and analysis. First, a brief description of the wind facility and how data is obtained is presented. Next, data pre-processing, data descriptive statistics, forecasting horizons, configuration of each tested method, and quality metrics are presented. Finally, experimental results and analysis are shown.

4.1. Data description

Through the sponsorship of the Global Environment Facility (GEF), and the United Nations Development Programme (UNDP), the Electrical Research Institute (IIE) built the Wind Energy Technology Center (CERTE) in La Ventosa, which is located at the Isthmus of Tehuantepec, Oaxaca.[2] Due to México's renewable energy policies [41,42], CERTE was built as Mexico's first wind energy small producer, and its main objectives consist in provide a framework for wind energy research and technological development.

La Ventosa county is characterized by a wind power density above 800 Watts per squared meter (800 W/m^2), and a mean wind speed (at 50 m height) greater than 8.5 m per second (50 m/s) [43]. Furthermore, it has been found that La Ventosa wind speed distribution is explained better by a bimodal probability distribution than the Weibull distribution [6,7]. In accordance to these features, and the international standard IEC 61400-1 [44], this site is suitable for testing Class I and Class II + S (special) wind turbines [45]. Therefore, a special class of wind turbine (KWT300)[3] was installed at CERTE's facilities during 2009. This turbine has a height of 40 m and is appropriate for local electricity suppliers, or where weather extreme conditions are present (e.g. seismic hazard is high, typhoons, lightning).[4] CERTE's installed capacity is 300 kW which corresponds to the energy provided by the KWT300 wind turbine [45]. This center possess two anemometric measuring towers at 80 and 40 m height. Data from wind speed, wind direction, humidity,

[2] http://www.iie.org.mx:8080/SitioGENC/producto02.html, last accessed: February 16, 2015.

[3] http://www.wind-energy-market.com/en/wind-turbines/big-plants/details/details/bp/kwt300/, last accessed: February 16, 2015.

[4] http://www.komaihaltec.co.jp/ENGLISH/PDF/KWT300_Brochure_2.pdf, last accessed: February 16, 2015.

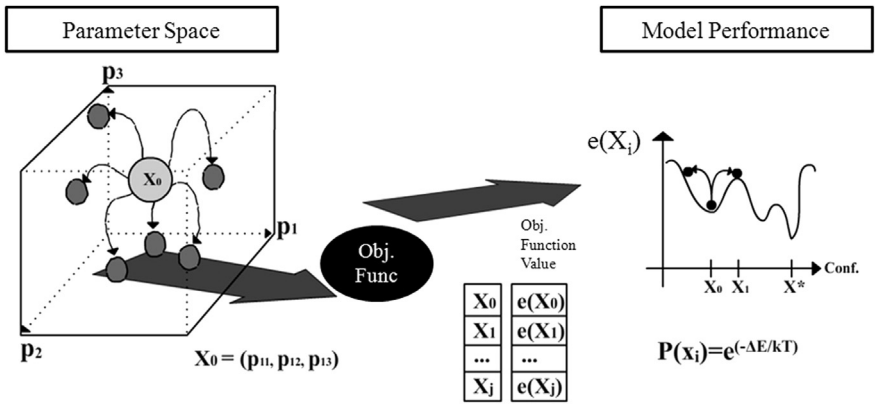

Fig. 4. GA Boltzmann Selection is depicted. Initially, a population of solutions is created. Next, fitness of each solution is determined by the objective function. Then, an SVR configuration is accepted if its quality is better than current best solution. Bad solutions are accepted in order to escape from local optima using Boltzmann distribution.

solar radiation, temperature, atmospheric pressure, and heat radiation is measured and stored in a central computer. The purpose of this paper is to provide a new model for univariate wind speed forecasting in wind farms, consequently, only wind speed data from the sensors at 40 m is employed.

4.1.1. Data pre-processing

Wind speed data from CERTE is composed by measures from January 1 2012 to August 28 2013. Data is measured every 10 min. First, outliers were removed in accordance to expert criteria: data outside typical range values were removed, then, wind speed values ≤ 0 were substituted with 0.00001; for missing values data was interpolated between previous and next tick; registers with wind speed missing data were removed. Short-term WSF (ranging from 1 h to several days ahead) is highly important for wind farms operations like wind turbines control, load and grid balancing, ramp events forecasting, and unit commitment for day-ahead markets [1,2,4,10,11,20,29]. Therefore, treated time series data was transformed from 10 min ticks to 1 h ticks by averaging it. The resulting 14,437 data points were then split into the training and testing sets. The training set contains the first 11, 550 (80%) consecutive data points, from January 1 2012 to the first 2 h of April 30 2013. The testing set contains the remaining 2887 (20%) data points from the third hour of April 30 2013 to August 28 2013. Additionally, when the proposed model of this paper was used, data was scaled between 0 and 1 to enhance SVR training time requirements. Data description is detailed in Table 2.

Moreover, CERTE energy production time series is obtained using wind speed data (i.e. realized) and KWT300 power curve. Then, the resulting Wind Power (WP) time series is splitted into training (80%) and testing (20%) sets. Descriptive statistics are shown in Table 2.

In order to assess estability of the tested methods, we splitted the whole data in three consecutive subsamples time series. Dates and typical descriptive statistics for WS subsamples are described in Table 3. In the mentioned table, SS stands for SubSample. Also, in the case of SS, logarithmic returns where calculated to show variance in WS as time goes by. Logarithm returns are commonly used as a proxy of the rate of change in financial time series analysis [22]; it is defined as $r(t)=[x(t)/x(t-1)]$. Full, subsampled, and logarithmic returns time series are depicted in Fig. 5. As can be observed in this figure, SS1 testing data has the largest rate of changes from all SS. Although, as measured by the standard deviation (and shown in Table 3), SS2 has the largest σ^2. This discrepancy is related to the definition of the standard deviation, which highly penalize larger deviations from the mean. SS3 has a smaller σ^2, however, as is shown in Fig. 5 bullet point G, there are more changes of smaller size.

The training of the proposed model was performed through a 10-fold cross validation procedure [22]. For every WS time series case, testing was performed consecutively using the remaining 20% of data points.

4.1.2. Data descriptive statistics

Previously to the introduction of traditional and non-linear descriptive statistics (and their results), it is worth noting that the following analysis is carried only in WS time series. The reason behind this is that we are interested in modelling wind speed process rather than forecast wind power. Therefore, we focus on wind speed probability distribution, chaotic behavior, and

Table 2
Wind speed and Wind Power data descriptive statistics.

	WS train data	WS test data	WP train data	WP test data
Sample Size	11,550	2887	11,550	2887
Sample Date Ranges	01/01/12–04/30/13	04/30/13–08/28/13	01/01/12–04/30/13	04/30/13–08/28/13
Max-Min Values	28.80–0	18.835–0	300-0	300-0
Mean	8.29	6.28	123.46	75.17
Standard Deviation	4.79	3.44	115.32	89.84
Skewness	0.55	0.40	0.34	1.02
Kurtosis	−0.16	−0.45	−1.49	−0.24
J-B Test	595.25	103.78	–	–
Lyapunov Exponent	0.04	−0.03	–	–
Emergence	0.82	0.85	–	–
Self-Organization	0.17	0.14	–	–
Complexity	0.58	0.49	–	–

Table 3
Wind speed data descriptive statistics for each subsample.

	SS 1,Train data	SS 1, test data	SS 2,Train data	SS 2, test data	SS 3,Train data	SS 3, test data
Size	3850	962	3850	962	3851	963
Date Ranges	01/01/12–06/09/12	06/09/12–07/19/12	07/09/12–12/12/12	12/12/12–02/08/13	02/08/13–07/19/13	07/19/13–08/28/2013
Max-Min Values	26.8–0	17.8–0	25.76–0	24–0	23–0	15.2–0.1
Mean	8.25	5.8	8	11	7.20	7.43
Stand. Dev.	4.66	3.72	4.2	5.8	4.69	3.32
Skewness	0.52	0.62	0.4	−0.05	0.94	−0.69
Kurtosis	−0.25	−0.45	0.4	−0.9	0.35	−0.871

components interaction analysis.

First, *Maximum* and *Minimum* values, *Mean*, and *Standard Deviation* are obtained. Next, *Skewness* is used to determine how symmetric is the data probability distribution, *Kurtosis* measures peakedness and extreme values frequency of the distribution. Additionally, the *Jarque-Bera (J-B) test* [46] a statistical test which proposes as its null hypothesis that sample's probability distribution is Gaussian, is evaluated.

One of the main PSR requirements is provided data must come from the observation of a chaotic process. In order to support PSR usage, we applied a statistical test called *Neural Networks (NN) Chaos Test*, which is a nonparametric statistical framework based on artificial NN for chaotic testing [47]. It is well-documented [26] that positive Lyapunov exponents typically characterize a chaotic process. Consequently, *NN Chaos Test* evaluates as its null hypothesis that data do not proceed from a chaotic process by approximating Lyapunov exponents values through a neural network model.

Recently, it has been suggested that weather non-predictable behavior is more related to the interaction of atmospheric variables rather than to deterministic chaos [48]. In this sense, the information-based framework proposed in Ref. [49], is useful to characterize a system according to the interactions of its components. Specifically Emergence (E), Self-Organization (SO), and Complexity (C) are measured from data. These measures are based on Shannon's information. Their value ranges between $0 \leq E, SO, C \leq 1$. Emergence is used to measure how much new information arises from the evolution of the system; a high E implies a system with high variance, where an $E \sim 0$ implies a static system. Self-Organization is used to measure the organization of the system, where a high SO implies a highly organized system, and low SO is correlated to a disordered system with high entropy. Lastly, Complexity presents an overall measure of the system's balance between chaos and order. Further, it has been stated that a system with high complexity is characterized by complex patterns which are prone to be identified; where a system with low complexity can be a) *fully deterministic*, in which case a simple model is enough to describe its behavior, or b) *completely random*, in which case identify any pattern is a futile task. Using E, S, and C, we characterize the data in terms of system information novelty, organization, and the overall complexity of the interactions of WS components.

The results for the applied statistics for training and testing data are shown in Table 2. In accordance to these, the distribution's skewness for both data sets (training and testing) is approximately symmetric. Moreover, Kurtosis statistics show that samples distribution is highly concentrated around the mean and extreme values occur less often than in a Gaussian distribution. On the other hand, the J-B test corroborates the findings of Kurtosis, by rejecting normality of data probability distribution with a significance level of 0.001%. *NN Chaos Test* does not reject the null hypothesis with a p-value of 0.9837 on the training data, while for the testing data it is rejected with a p-value of 0.0049. Nevertheless, it is well known that chaos theory methods are clearly influenced by small data sets [26]. Evenmore, this test found that for the whole data set the null hypothesis is not rejected with a contundent p-value of 0.999 and a Lyapunov exponent of 0.0859.

Information-based measures for training and testing wind speed data showed consistent results for E, SO, and C. Evenmore, results for the whole data set were high similar to those of the training set. Therefore, the analysis is made upon the whole set results. According to these the system is highly variable (Emergence of 0.825), a necessary condition for exhibiting chaos. Also, the system has a low level of self-organization (0.175), this reflects a process of low regularity. Although the values for E are high while low for SO, according to [49] the complexity values calculated for the data reveal that the data have *fair complexity* (C of 0.577). This type of system present identifiable patterns, nevertheless, they are occluded because of the high emergence of new states. This indicates that predicting wind speed patterns is non-trivial because of a high information novelty.

Provided the former statistical and information-based analysis, data is considered to come from a process at the edge of chaos. The corresponding J-B test value, Lyapunov exponents, E, SO, and C are shown in Table 2.

4.1.3. Forecasting horizon

Wind power production requires wind speed to be forecasted in order to manage electric production and distribution. Evenmore, due to physical thresholds, wind turbines only operate under certain wind speed range. Therefore, future wind speed predictions are required to maximize efficiency and minimize production/maintenance costs. Typically, short-term wind speed forecasting is required to be predicted from 1 h to 24 h ahead.

Therefore, our experimental setup considered the aforementioned short-term predictive range: from one step ahead ($h=1$) to one day ahead ($h=24$).

4.2. Experimental parameter settings

Classical univariate models and the proposed $PSR{-}SVR_{GA}$ model require some parameters to be defined beforehand. Each setup for the tested methods is presented below.

4.2.1. Classical and TDC model orders

Classical models and $PSR{-}SVR_{GA}$ require the identification of the dependent variables of the autoregressive models. For the classical models, AR and MA orders were obtained through Akaike's Information Criteria (AIC) [30]. In the case of the TDC model, time delay τ and the embedding dimension d_e were obtained by the PSR method [24]. Data integration (I) order was obtained by differencing the time series, and then used by the ARIMA method. We compared our method with and without the latter transformation. We denote with the greek letter α the instance which does not employ the integration transformation, and with β the instance of the $PSR{-}SVR_{GA}$ which uses it.

A final note about subsampled data, and its repercutions on the

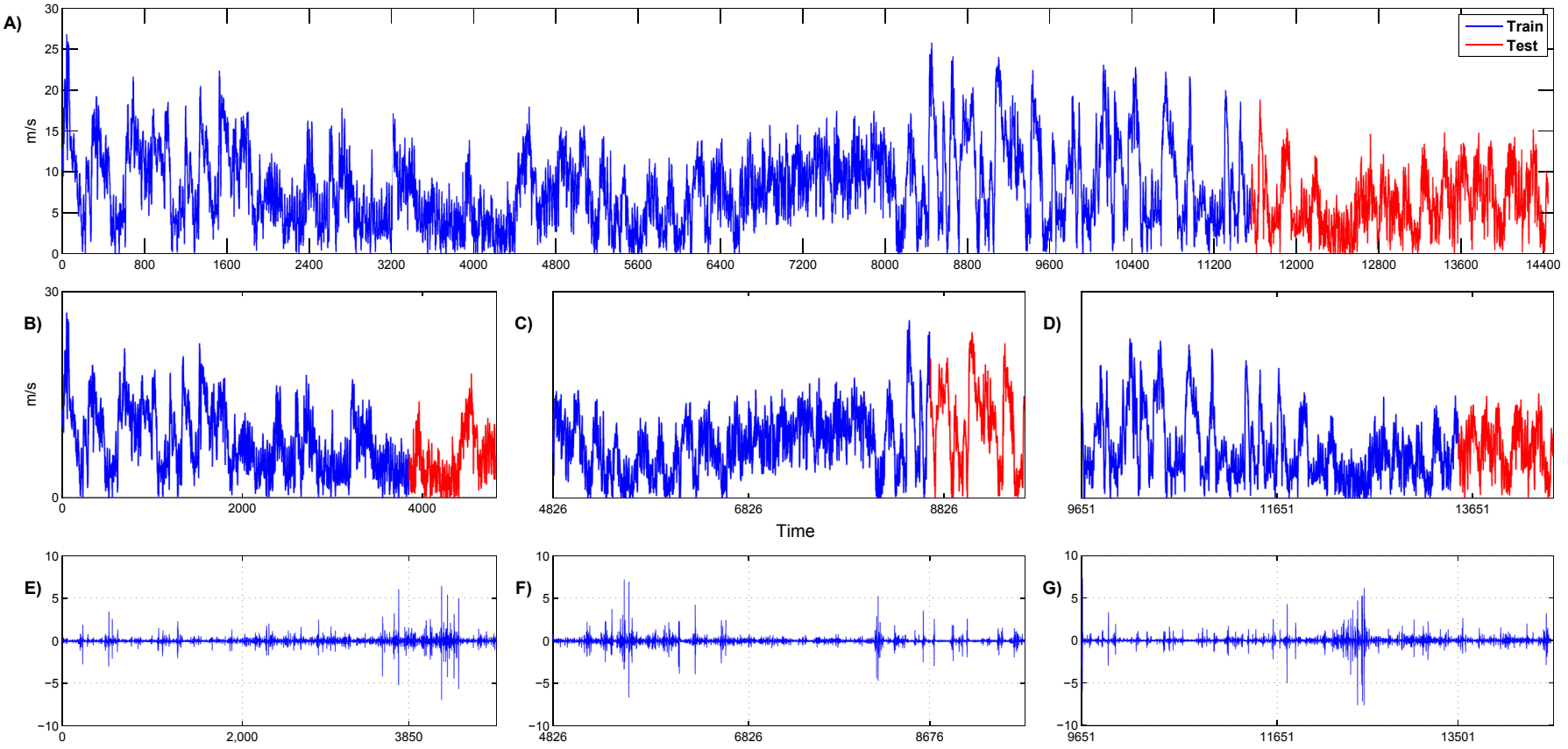

Fig. 5. Full, Subsampled, and logarithmic returns time series. Blue corresponds to training data, while red is employed for test data. Bullet point A presents the complete time series. Bullet points B, C, and D show subsampled sets 1, 2 and 3, respectively. Bullet points E, F, and G presents the logarithmic returns for SS 1, 2, and 3, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Order configuration for classical and the proposed methods.

Order	Classical Model			TDC Model	
	AR	MA	I	τ	d_e
AR	2	–	–	–	–
ARMA	1	27	–	–	–
ARIMA	30	5	1	–	–
$PSR-SVR_{GB}\alpha$	–	–	–	19	6
$PSR-SVR_{GB}\beta$	–	–	1	19	6

Table 5
$PSR-SVR_{GA}$ training parameters.

Parameter	Method	Value
MI: Max τ	PSR	50
FNN: Max d_e	PSR	50
FNN: ε	PSR	10
FNN: A_{Tol}	PSR	2.5
Fitness Function	GA	10 Fold Cross Validation
Selection Operator	GA	Boltzmann Selection
Crossover Operator	GA	n-points
Mutation Operator	GA	Uniform
Crossover/Mutation Rate	GA	0.8/0.2
Insensitive ε tube size	SVR	0.1
K_{Poly}: polynomial degree d	SVR	2–4
K_{Gauss}: rescaling factor γ	SVR	0.0001–10
Error trade-off constant C	SVR	0.0001–10

optimal orders of ARIMA and $PSR-SVR_{GA}\alpha$ models, should be made. For the former, the model detailed in Table 4 is not stationary, hence, AR and MA orders were once again calculated. For the latter, we recalculated the TDC model with subsampled data. The resulting TDC models, slightly varied for all subsamples (i.e. SS1, τ=13, d_e=6; SS2, τ=17, d_e=6); SS3, 7). However, it is well known that non-linear methods (e.g. MI [50], FNN [39]) performance improves with the availability of larger time series. Therefore, we conclude these were projections of the true TDC, and employed the one detailed on Table 4.

Table 4 shows the AR, MA, I, τ, and d_e orders identified for the models.

4.2.2. $PSR-SVR_{GA}$ configuration

The novel methodology proposed in this work for wind speed forecasting is composed by several methods: Phase Space Reconstruction, a Hybrid GA, and SVR. Each one of these has several parameters that need to be set beforehand. Table 5 shows the parameter's name, the method in which is employed, and the value or range of values defined for training the forecasting model.

4.3. Quality metrics

Performance evaluation of persistence, autoregressive and our proposed methods was done in terms of WSF and WPF. In doing so, we provide a completer overview of the performance of the evaluated methods, and their utility for wind farms. First, the metrics employed for WSF are described. Then, WPF performance metrics are introduced. Finally, an improvement index for comparing results provided by WSF and WPF quality metrics is detailed.

4.3.1. WSF quality metrics

Performance of WSF methods can be measured in terms of accuracy, which is related to the exactness of the predicted wind speed respect to the realized. On the other hand, directional-based measures describe forecaster's phase errors (wind speed magnitude is forecasted correctly but with a time delay), and are useful for economic interpretations like utilities integration costs or wind generator ramp policies [29,51–54].

In order to compare quality of the proposed method against the persistence model and classical methods, several metrics for different forecasting features are employed. For methods accuracy performance, the Mean Absolute Error (MAE) [2], Mean Bias Error [10,53,55,56], Root Mean Squared Error (RMSE) [1,2], and Mean Absolute Scaled Error (MASE) [57–59] were employed. For methods directional performance, the Directional Accuracy (DA) was employed [52,60,61]. These metrics are depicted by Eqs. (9)–(13), where x_i stands for the observed wind speed and $\widehat{x}_i$ for the estimated wind speed, respectively.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|x_i - \widehat{x}_i\right|. \tag{9}$$

$$MBE = \frac{1}{N}\sum_{i=1}^{N}(x_i - \widehat{x}_i) \tag{10}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \widehat{x}_i)^2}. \tag{11}$$

$$MASE = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{|x_i - \widehat{x}_i|}{\frac{1}{N-1}\sum_{i=2}^{N}\left|x_i - x_{i-1}\right|}\right). \tag{12}$$

$$\begin{aligned} DA(\%) &= \frac{100}{N}\sum_{i=1}^{N}a_i, \\ Err_1 &= sign(x_{i+1} - x_i) \\ Err_2 &= sign(\widehat{x}_{i+1} - x_i) \\ where\ a_i &= \begin{cases} 1 & Err_1 = Err_2 \\ 0 & Otherwise \end{cases} \end{aligned} \tag{13}$$

MAE and RMSE are classical accuracy-based measures for qualifying the performance of wind speed forecasters. RMSE assumes that the errors are unbiased and follow a normal distribution, while MAE is suitable to describe uniformly distributed errors. However, they have several drawbacks like heavier penalties on positive errors than on negative errors [59]. Another accuracy-based metric closely related to MAE is the MBE [29,55]. This measure is used to evaluate the average bias of forecasting models (*i.e.* over- or under-estimation). Typically, commercial consumers of wind energy methods prefer under-estimated models for WSF [53]. However, MBE should be used altogether with other performance measures (e.g. MAE) due its inconsistent penalization to error magnitude [55]. Further, MAE, RMSE, and MBE are scale-dependent measures not suited for comparing time series with different scales [57,62]. Recently, the Mean Absolute Scaled Error (MASE) have been proposed for the evaluation of wind speed and wind power forecasters [58,59]. It has been stated that MASE performance is independent to data scaling, and it can be used to compare different forecasters across multiple time series [57,59]. MASE scale-independence is achieved by scaling the in-sample MAE of a forecaster respect to average error of a one-step naïve method [57]. In the 1-h ahead WSF context, MASE provides a direct comparison of the error of a forecasting method respect to the average error of the PM. If MASE>1, the performance of the forecasting method is worse than the average of PM. Otherwise, the tested forecasting method is more accurate than PM [57]. Although robuster, MASE has not been

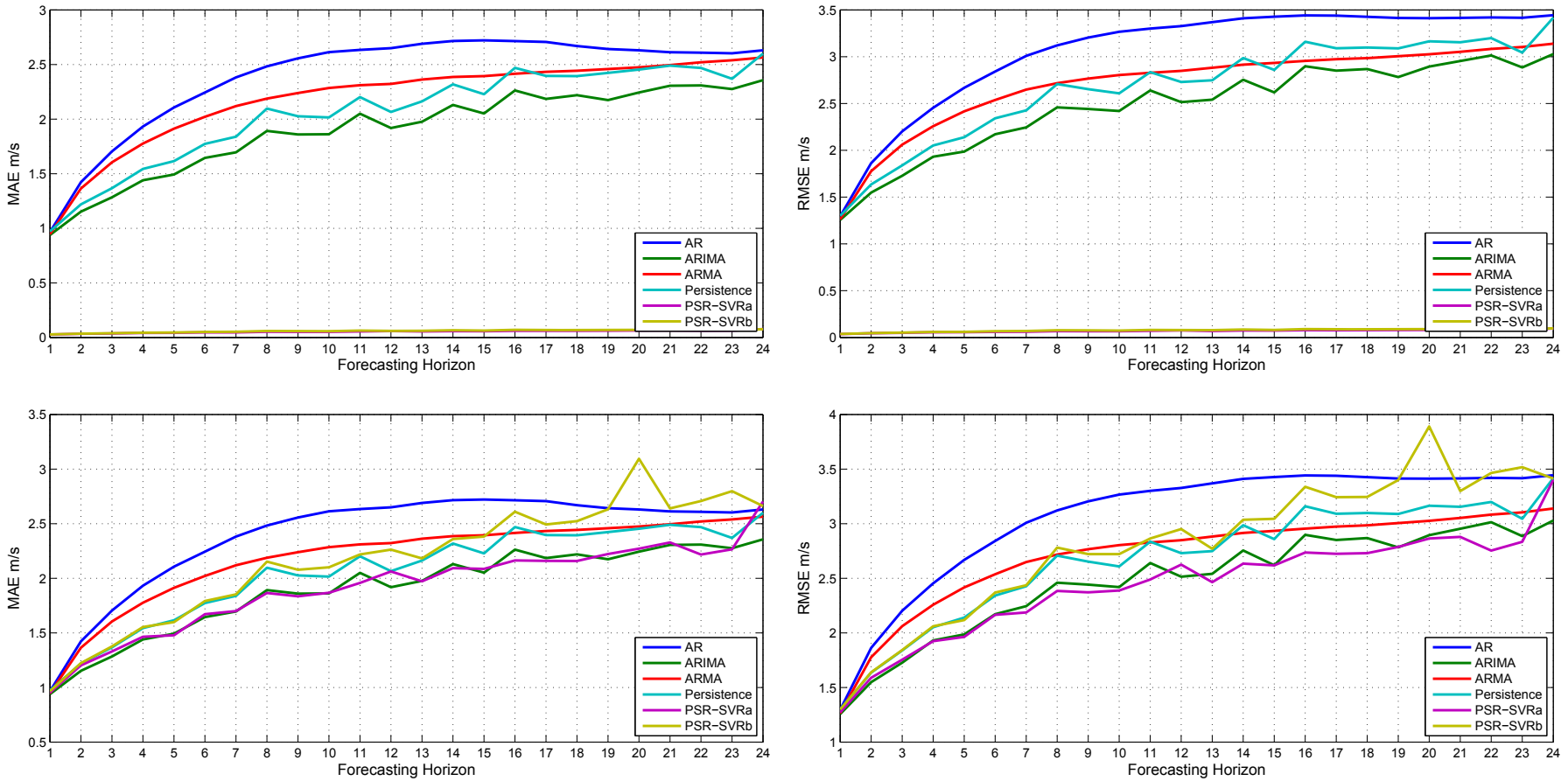

Fig. 6. MAE and RMSE Results. Top of the figure shows results for scaled forecasts, while in the bottom, graphics present results for unscaled forecasts.

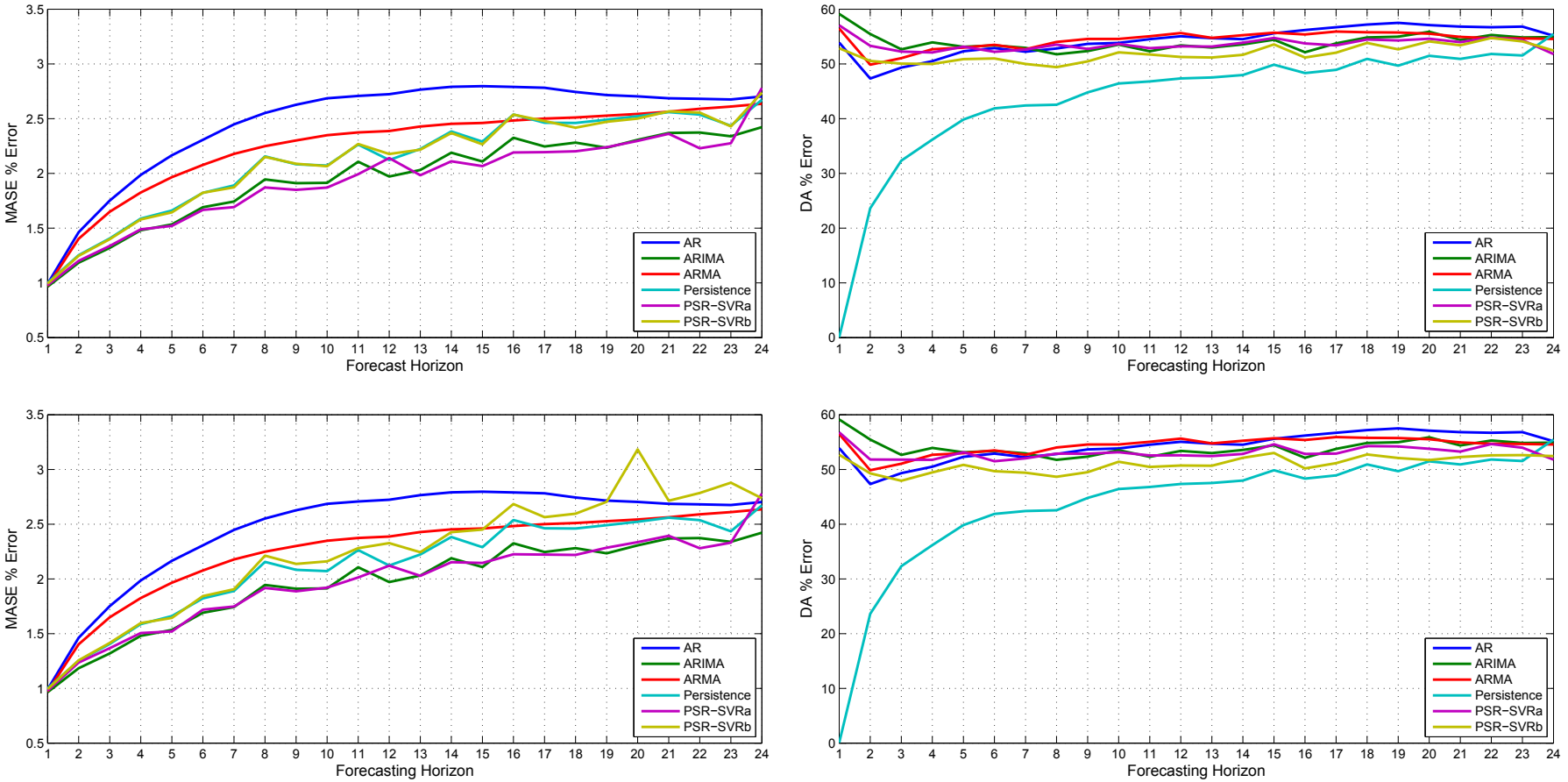

Fig. 7. MASE and DA Results. Top of the figure shows results for scaled forecasts, while in bottom unscaled forecasts are displayed.

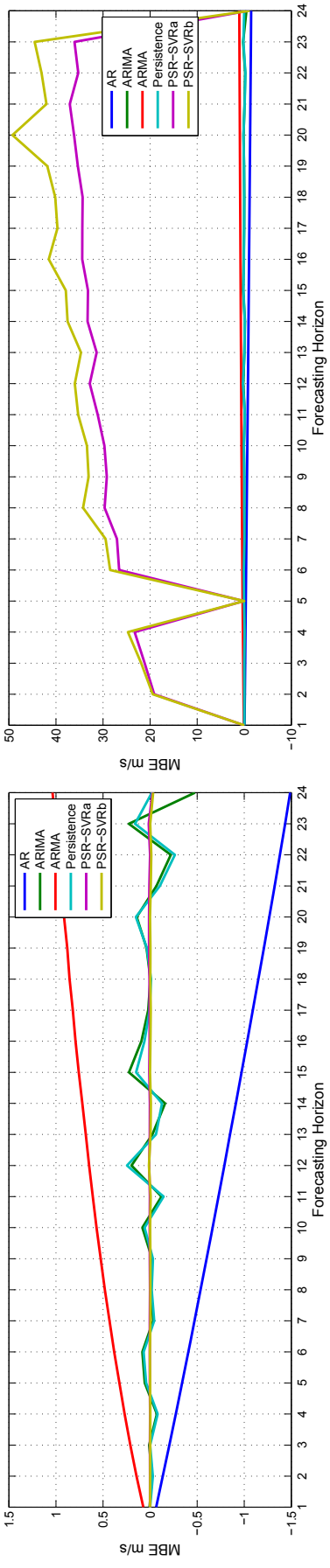

Fig. 8. MBE Results. Left side of the figure shows results for scaled forecasts while in the right results for unscaled forecasts are presented.

extensively used in wind energy literature [63].

Niether MASE, nor MAE, MBE nor RMSE account for phase errors of the forecasting method by itself [53]. It has been stated that directional accuracy must be considered in WSF for utilities, and grid integration purposes [52]. Further, directional errors of WSF are directly transferred to the power prediction, and affect wind genetor ramping policies [12,53,54].

Therefore, a good forecasting method must ensure low MAE, MBE, RMSE, and MASE, while achieving high DA values.

4.3.2. WPF quality metrics

Once wind speed is forecasted, wind power time series is computed using KWT300 power curve. Employing the wind power data, the error for the forecasting horizon h at the time step t is defined as

$$\varepsilon(t+h|t) = \frac{1}{P_{inst}} P(t+h) - \widehat{P}(t+h|t),$$

where $P(t)$ and $\widehat{P}(t)$ stands for the realized and forecasted wind power produced. P_{inst} corresponds to CERTE's installed capacity (i.e. 300 kW).

Rather than employing the WSF qualitative metrics, we employed the NMAE, NMBE, and NRMSE [29,64] to evaluate the forecasting abilities of benchmark and the proposed methods in terms of WP. It has been documented that these metrics are robuster, are directly related to the facility produced energy, and provide results independent to wind farm size [64]. Moreover, they yield further insight in the economicimpact of the models errors as a function of the site installed capacity [29]. These metrics are described by Eqs. (14)–(16).

$$NMAE = \frac{1}{N}\sum_{i=1}^{N}\left|\varepsilon(t+h|t)\right|. \tag{14}$$

$$NMBE = \frac{1}{N}\sum_{i=1}^{N}\varepsilon(t+h|t) \tag{15}$$

$$NRMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\varepsilon(t+h|t))^2}. \tag{16}$$

4.3.3. Methods comparison

As a mean to quantify the improvement of a specific method against the reference method (i.e. PM/D2D method), we employed the Improvement (Impr) metric proposed by Ref. [29]. This metric is defined as follows

$$Impr_{ref,EC} = 100*\left(\frac{EC_{ref}(h) - EC(h)}{EC_{ref}(h)}\right), \tag{17}$$

where *ref* stands for the reference method (e.g. Persistence), EC_{ref} and EC stands for the Evaluated Criterion result (e.g. MAE, RMSE, and so on) for the reference and compared methods, respectively. It is worth mentioning that, when DA is employed, EC_{ref} becomes the subtrahend instead of the minuend, in the dividend part or Eq. (17).

4.4. Results

In order to analyze $PSR{-}SVR_{GA}$ forecasting capacities, it is compared against the persistence model and classical time series

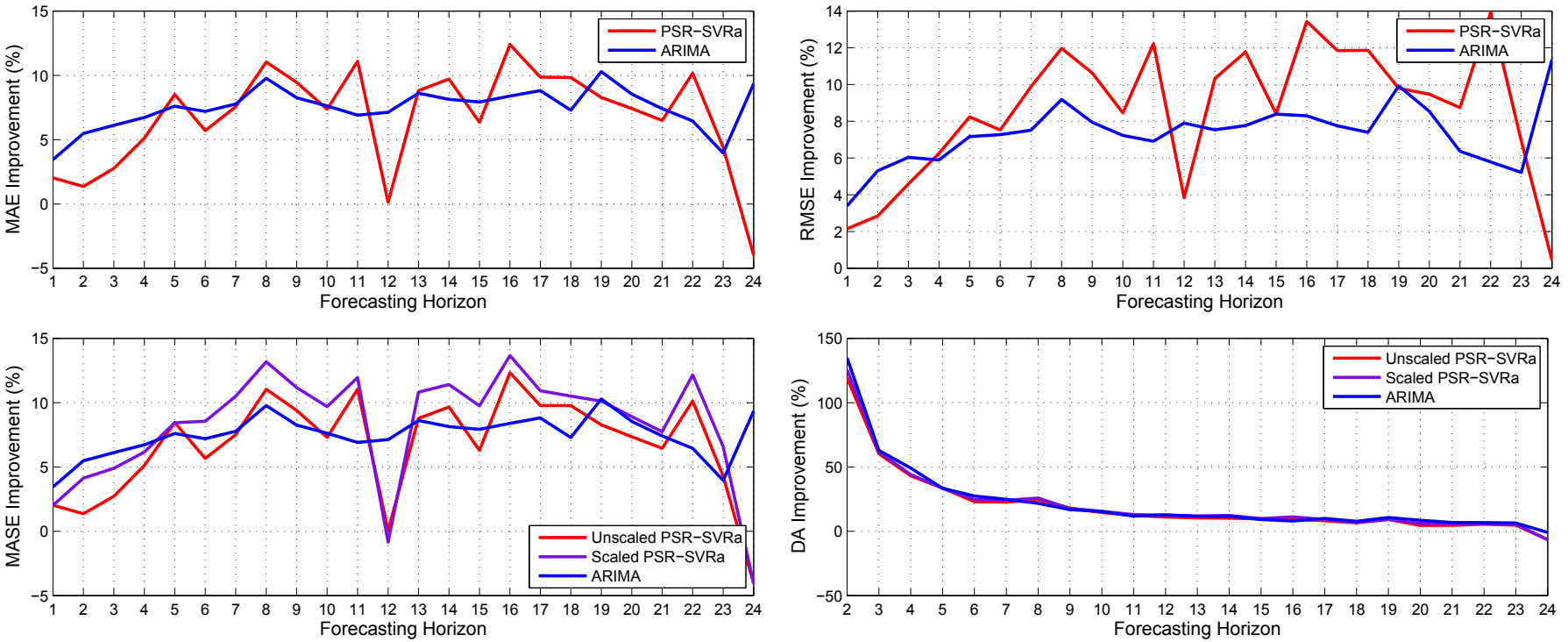

Fig. 9. WSF Improvement of Models. From left to right, from top to bottom, MAE, RMSE, MASE, and DA results are shown. In the case of MASE and DA, scaled results are shown un purple whereas unscaled are displayed in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

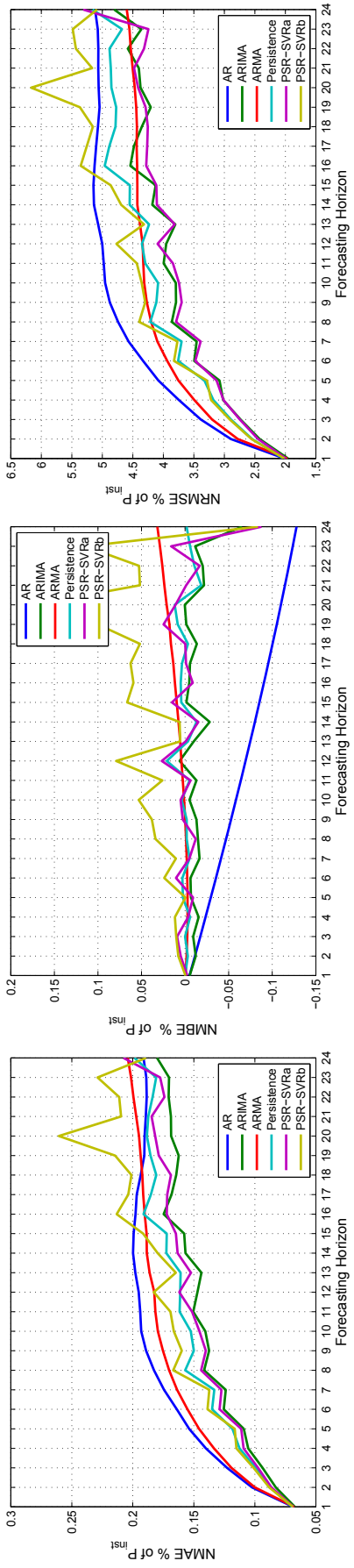

Fig. 10. WPF results. From left to right, NMAE, NMBE, and NRMSE measurements for the full time series setup are shown.

methods. Results are analyzed in terms of WSF and WPF for consecutive 24 h ahead (i.e. day ahead). Evenmore, data integration applied by ARIMA to ensure stationarity in data is also studied within our proposed methodology.

WSF results are produced for the full time series setup. These results are analyzed in regards to the bias, precision, and variance of forecasting methods as measured by WSF quality metrics. Further, these results are employed to assess the sensitivity of WSF metrics respect to the scaling preprocess required by $PSR{-}SVR_{GA}$. Improvement of the models are then shown. Improvements are analyzed and displayed for MAE and RMSE in terms of the unscaled predictions. For MASE and DA, improvements for scaled and unscaled forecasts are shown. Moreover, for reasons that will become clear later, comparison of WSF improvements are only made for ARIMA and $PSR{-}SVR_{GA}\alpha$ results. Next, a final note about the effects of the stationary transformation applied to the proposed method is made.

WPF results are produced for the full time series and subsamples setup. For the former, results for WPF quality measures are employed to evaluate the proposed and studied methods. For the latter, results are used to assess generalization capabilities of proposed and classical forecasting methods conserning to the available training data. WPF improvements are displayed for ARIMA and $PSR{-}SVR_{GA}\alpha$ models.

All experimentation was developed using Matlab 2009a. The hybrid GA was manually coded, while LibSVM was used as SVR library [65]. CRP Toolbox for Matlab was employed for the estimation of the TDC model (i.e. τ, d_e) [66]. Lastly, Matlab 2009a System Identification Toolbox was employed for AR-based methods.

4.4.1. WSF analysis

PM/D2D, AR-like methods, and the proposed $PSR{-}SVR_{GA}$ were employed for WSF for consecutive 24 h ahead. Results in terms of MAE, and RMSE are shown in Fig. 6; for MASE, and DA, Fig. 7 displays the results; Fig. 8 presents results for MBE. Figs. 6 and 7 presents unscaled data in its top part, whereas unscaled is shown in bottom. Fig. 8 presents scaled results on the left side, whereas unscaled are displayed on the right. For each figure, *X* axis corresponds to the forecasting horizon, while *Y* axis shows units in which each metric describe the incurred errors (i.e. MAE (m/s), MBE (m/s), RMSE (m/s), MASE (%), and DA (%)). Improvements comparison in terms of MAE, RMSE, MASE, and DA are shown in Fig. 9.

4.4.1.1. Scaled Vs un-scaled results. As can be observed in Figs. 6–8, MAE, MBE and RMSE are not adequate for comparing results with different scales. For all of these measures, the estimated error is approximately zero for the whole range of forecasting horizons. In the case of unscaled results, MAE and RMSE are closely simmilar to those reported by MASE. These measures are considered for the improvement comparison of models. It is also worth noting the inconsistent penalization of incurred errors as measured by MBE for both, scaled (i.e. error penalization is notably biased by data scaling) and unscaled (e.g. for the 20 h ahead MBE error is as large as 50 m/s, which clearly contrast with MAE, RMSE, and MASE results) forecasts. In consequence, MBE results are discarded for the rest of this WSF analysis.

On the other hand, MASE and DA results (Fig. 9) show consistency for scaled and unscaled results. The largest difference between $PSR{-}SVR_{GA}\alpha$ MASE results is of 0.08%, whereas for $PSR{-}SVR_{GA}\beta$ is of 0.68% for 20 h ahead. In the case of DA, differences are imperceptible. Additionally, it is interesting how DA of PM quickly (i.e. logarithmic rate) increases for increasing forecasting horizons. These empirical results show that directional changes (i.e. increase or decrease in wind speed), have a higher correlation for longer forecasting horizons in contrast with shorter.

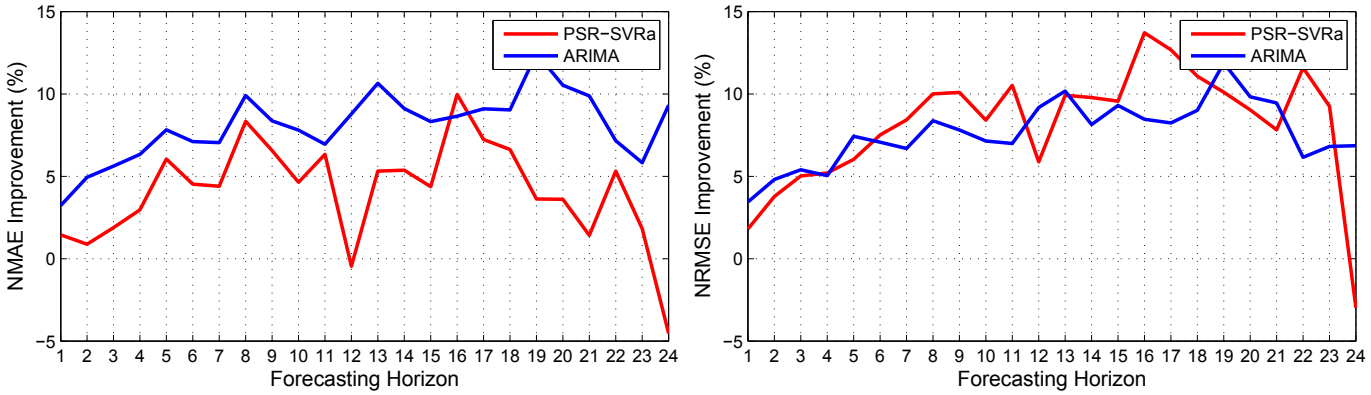

Fig. 11. WPF Improvement of Models in terms of NMAE and NRMSE (from left to right, respectively). As in the case of WSF, only results for ARIMA (blue) and $PSR{-}SVR_{GA}\alpha$ (red) are depicted. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Therefore, due the high regularity of MASE and DA measurements, scaled results for these metrics are also considered in the improvement comparison.

4.4.1.2. Improvement comparison of models. Improvement results for MAE, RMSE, MASE and DA are presented in Fig. 9. For readability, DA results for 1 h ahead are dismissed (i.e. improvement is near four orders of magnitude). From visual inspection at Fig. 9 it is clear that, in average, AR, ARMA and $PSR{-}SVR_{GA}\beta$ do perform worse than PM, whereas ARIMA and $PSR{-}SVR_{GA}\alpha$ perform better in terms of WSF. Hence, comparisons over the reference model are constrained to ARIMA, and $PSR{-}SVR_{GA}\alpha$ results.

In general, MAE, RMSE, and MASE improvements over PM for 1–4 h ahead are greater for ARIMA than the proposed method. For 5–22 h ahead, $PSR{-}SVR_{GA}\alpha$ achieves greater improvements than ARIMA over the reference method. For 23 h and day ahead horizons, ARIMA is better. Particularly, for 12 h and day ahead horizons, $PSR{-}SVR_{GA}\alpha$ performed worse than the reference method. Decrease in improvement for the former case, given the average performance of previous and next forecast horizons, is a consequence of premature convergence in the stochastic optimization of SVR parameters. Some specific findings for these WSF quality metrics are described next:

- As measured by MAE improvements, ARIMA is more accurate than $PSR{-}SVR_{GA}\alpha$ for the first and last forecast horizons. However, as measured by RMSE, the proposed method variance is lesser than ARIMA.
- In the case of MASE, improvements for scaled $PSR{-}SVR_{GA}\alpha$ forecasts are higher than those obtained by the unscaled results, been the largest difference 3.5%.

In general for DA, higher improvements are gained for the first 4 h ahead. For the next 20 forecast horizons, improvement decays from 33% to −1% in the case of ARIMA, and −6% in the case $PSR{-}SVR_{GA}\alpha$.

4.4.1.3. Data integration for $PSR{-}SVR_{GA}$. As is shown in Figs. 6–8, data integration reduces the performance of the proposed method. For shorther forecast horizons $PSR{-}SVR_{GA}\beta$ accuracy is as good as the reference method, whereas $PSR{-}SVR_{GA}\alpha$ is better than PM. As the forecast horizon increases, the performance is worsen for the former against the latter. Day ahead results are the only case where $PSR{-}SVR_{GA}\beta$ is better than $PSR{-}SVR_{GA}\alpha$. Although, given the variance as measured by RMSE, is higher for the former than the latter, there is no reason to think $PSR{-}SVR_{GA}\beta$ will forecast better the day ahead.

4.4.2. WPF analysis

Results for 1 h to 24 h WPF in terms of NMAE, NMBE, and NRMSE are shown in Fig. 10. Each plot in figure (from left to right) corresponds to one of the WPF quality metrics as described in section 4.3.2. In this figure, *X* axis corresponds to the forecasting horizon, while *Y* axis shows error percentage in terms of CERTE's installed capacity (i.e. P_{inst}). Improvement comparison for NMAE and NRMSE are presented in Fig. 11.

Further, model WPF capabilities are analyzed in terms of generalization and error distribution. For the first, sensibility to training data is assessed by splitting in three subsamples the whole WS time series as described in section 4.1.1. The corresponding subsample results in terms of NMAE, NMBE, and NRMSE, are presented in Fig. 12; *X* axis displays the forecasting horizon, while *Y* axis shows error percentage in terms of CERTE's installed capacity (i.e. P_{inst}). For the second, error distribution histograms as suggested by Refs. [29,64] are employed. Error distributions for 1, 5 and 24 h ahead are obtained. Bins represents 10% of the site installed capacity (i.e. 30 kW); *X* axis displays the forecasting horizon, while *Y* axis shows the frequency of errors per bin.

4.4.2.1. WPF results description. In accordance to literature, NMAE errors are directly associated to the produced energy. The evidence presented in Fig. 10 shows that, ARIMA is the most accurate model forecasting WP using WS forecasts. In terms of NMAE, $PSR{-}SVR_{GA}\alpha$ in general, performs better than the reference method. Although, for 12 and 24 h ahead performance is inferior. It is interesting that, for WPF quality metrics results for the 12 h ahead for $PSR{-}SVR_{GA}\alpha$ show consistency in regards to the neighborhood of horizons results. Nonetheless, excluding day ahead results, NMAE average error for ARIMA (14.05%) and $PSR{-}SVR_{GA}\alpha$ (14.6%) are very close.

As stated by Ref. [64], NMBE is related to the bias of the models. As shown by Fig. 10, in general AR model is the most biased model, with a systematically understimation of the WP process. On the other hand, $PSR{-}SVR_{GA}\beta$ obtains an overestimation of WP process with an average NMBE of 0.04%. Nevertheless, in accordance [29], none of the methods incur in systematic errors.

NMRSE describes the variance of the tested methods. As can be observed in Fig. 10, in average $PSR{-}SVR_{GA}\alpha$ displays the lowest NRMSE for the 24 h ahead with an error of 3.8% of the P_{inst}. On the other hand, $PSR{-}SVR_{GA}\beta$ variance is only as good as the reference method for the first 5 h ahead; after, its performance decay with a maximum for the 21 horizon of 6% of the. P_{inst}.

4.4.2.2. Improvement comparison of models. Improvement results for NMAE, and NRMSE are presented in Fig. 11. From visual inspection at 10 it is clear that, in average, AR, ARMA and $PSR{-}SVR_{GA}\beta$

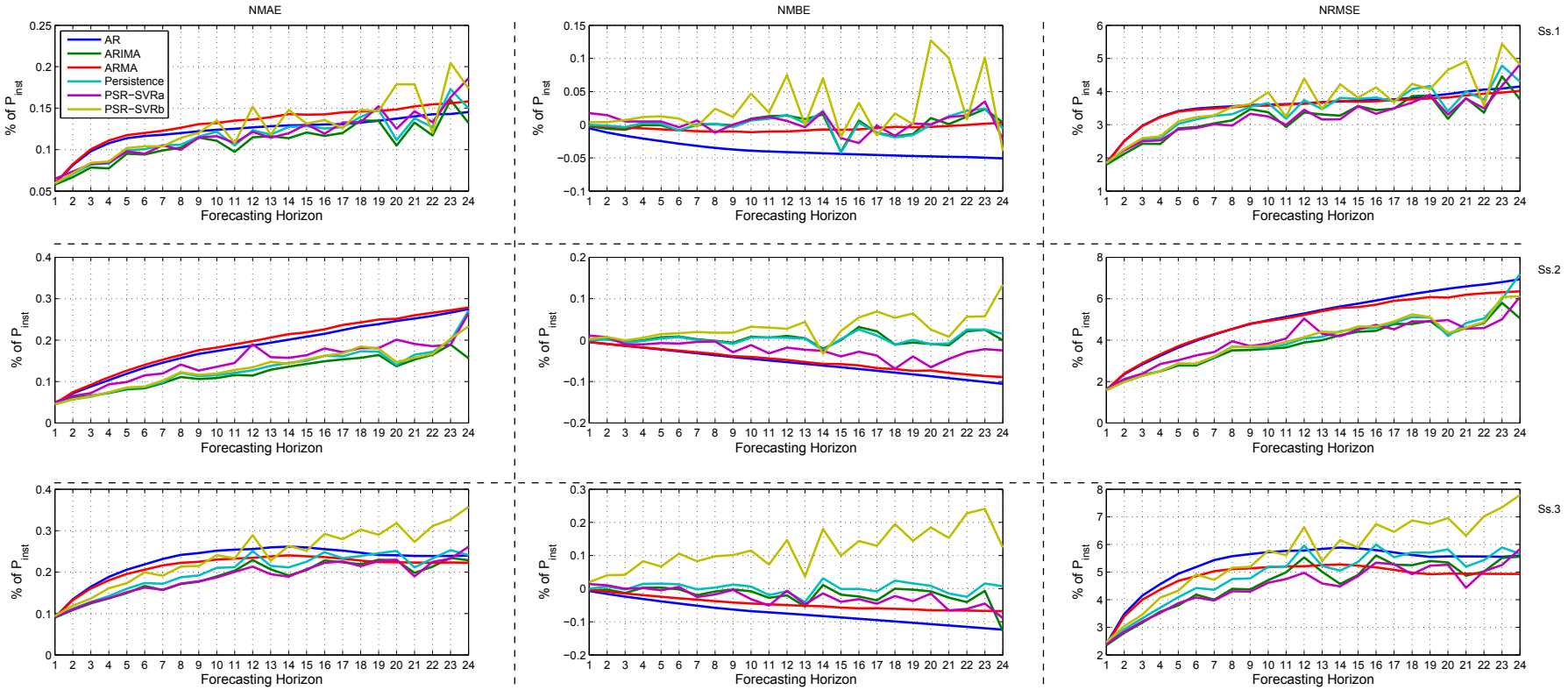

Fig. 12. WPF results for three subsamples. From left to right, NMAE, NMBE, and NRSME measures are shown, respectively. From top to bottom, results for subsample 1 to 3 are presented.

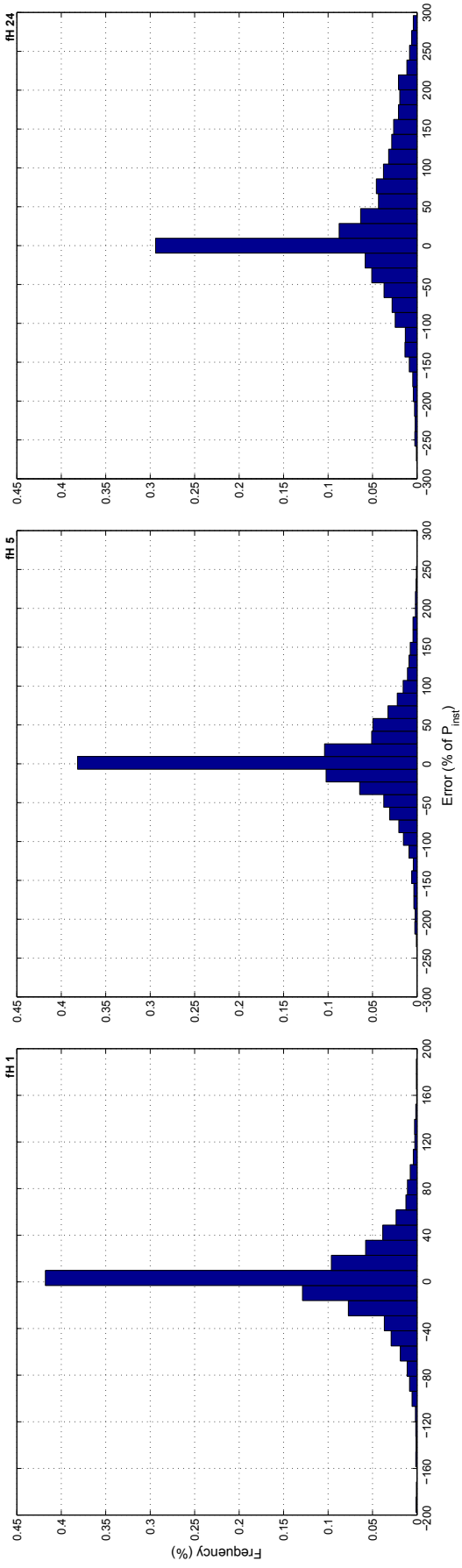

Fig. 13. Analysis of Error Distribution. From left to right, error distributions obtained by the $PSR{-}SVR_{GA}\alpha$ for 1, 5 and 24 h ahead are shown.

do perform worse than PM, whereas ARIMA and $PSR{-}SVR_{GA}\alpha$ perform better in terms of WPF. Hence, comparisons over the reference model are constrained to ARIMA, and $PSR{-}SVR_{GA}\alpha$ results.

In general, NMAE improvements over the reference method are greater for ARIMA rather than $PSR{-}SVR_{GA}\alpha$. In particular, ARIMA gain is greater for 12 (9.2%) and 24 (13.8%) hours ahead. In terms of NRMSE improvements over PM for 1–4 h ahead are greater for ARIMA than the proposed method. Although, the maximum improvement difference between ARIMA and the proposed method, is 1.6% for the 1 h horizon. For 5–23 h ahead, $PSR{-}SVR_{GA}\alpha$ achieves greater improvements than ARIMA over the reference method with an average of 9.5%. Again, for 12 h and day ahead horizons, $PSR{-}SVR_{GA}\alpha$ performed worse than the reference method. It is worth noting that, both are multiples of 12. This suggests that, the approximated TDC model used for SVR does not fully capture the conditional probability of these forecasting horizons.

4.4.2.3. Generalization assessment of models. It is well documented elsewhere that, statistical learning risk improves for larger data sets. Moreover, smaller data sets may restrain function estimation to smaller scale WS processes, occluding *the big picture*. In order to assess the impact of available data in the generalization of PM, AR-like, and $PSR{-}SVR_{GA}$ methods, CERTE data was splitted in 3 sub-samples. Fig. 12 presents the results. Columns correspond to WPF quality metrics (NMAE, NMBE, and NRSME, respectively), whereas Rows correspond to subsampled data (1–3, respectively). From an exploratory data analysis view, the next conclusions are obtained:

- Performance of AR, and ARMA methods across all metrics for all SS remains constant. In terms of NMAE, and NRSME, these methods performed better for longer horizons (i.e. 18) than PM/D2D for SS1 and 3. It can be observed, for NMBE of all SS that, AR and ARMA models underestimate the true process of WP.
- Performance of reference models is high for all measures in all SS cases. Particularly, for the SS2 case, it achieves marginally worse results for NMAE than the best method; in terms of NRMSE, it has the best performance in average.
- ARIMA achieved a better performance over PM/D2D methods in all SS for NMAE. In terms of NMBE, it tends to obtain slightly underestimated models. For NRMSE, ARIMA results are better for SS2 than the other two subsamples.
- The achieved performance of $PSR{-}SVR_{GA}\alpha$ in terms of NMAE is better for SS3, followed by SS1, and SS2, respectively. In terms of NMBE, it tends to slightly underestimate the process for SS2 and SS3. For NRMSE, it achieves the best performance in SS1 and SS3.
- Performance of $PSR{-}SVR_{GA}\beta$ in terms of NMAE is (surprisingly!) better than $PSR{-}SVR_{GA}\alpha$ for SS2. It seems that, for time series with larger but infrequent changes, ergodic transformation is benefical for the proposed method. However, this transformation leads to overestimated functions of the WP process, as is shown by NMBE results for all SS. For NRMSE, variance of the model increases for longer forecasting horizons.
- AR-like methods and reference models are better modelers of highly persistence processes using lesser data in respect to $PSR{-}SVR_{GA}$ methods. Although, there is a clear impact in the performance of the proposed method for smaller training sample, in the case where more changes (of smaller magnitude) are present, $PSR{-}SVR_{GA}\alpha$ outperforms other methods.

4.4.2.4. Analysis of error distribution. Finally, Fig. 13 presents error distributions for one, five and 24 h ahead predictions of the full

time series setup. These histograms correspond to error distribution for the $PSR{-}SVR_{GA}\alpha$ method. Comparing the three histograms, it is notorious that the error distributions is more symetric for one and 5-h ahead, than day ahead forecasts. In the three cases, distribution is centered around 0, however, for the day ahead horizon histogram is highly skewed over possitive errors. Further, the error distribution of one and 5-h ahead, are more leptokurtic than for 24 h. Conclusions derived for these histograms are detailed below:

- Robustness
  - For 1 h ahead, errors are less than 10% of P_{inst} 76% of the times,
  - For 5 h ahead, errors are less than 10% of P_{inst} 63% of the times,
  - For the day ahead, errors are less than 10% of P_{inst} 45% of the times.
- Large Errors
  - For 1 h ahead, errors are more than 20% of P_{inst} only 9% of the times,
  - For 5 h ahead, errors are more than 20% of P_{inst} only 20% of the times,
  - For the day ahead, errors are more than 20% of P_{inst} only 40% of the times.

5. Conclusions

Wind speed forecasting is a key component for wind energy production. Nevertheless, WSF is a hard task due its intermittence, high variability and non-linearity. In this work, a new methodology for WSF called $PSR{-}SVR_{GA}$ is presented. $PSR{-}SVR_{GA}$ is a methodology for univariate time series forecasting. Employing a chaos theory model called TDC, WS data is embedded into a reconstructed phase space where regression by SVR is performed. Then, a proper SVR model is selected by means of a GA; the genetic operator called Boltzmann selection is used to avoid premature GA convergence. Previous to experimentation, a statistical analysis over CERTE's wind data was performed. Results from this analysis showed that wind speed for studied location has a non-Gaussian distribution. Also, the analysis showed that data have positive Lyapunov exponents, a characteristic feature of chaotic processes. Furthermore, the information-based metrics which analyze the complexity of the system revealed that the system behavior can be cataloged as being on the chaotic side but close to the edge of chaos. Therefore, these evidences support the use of TDC model and PSR in the proposed method.

The mettle of $PSR{-}SVR_{GA}$ was tested against PM/D2D, AR, ARMA, and ARIMA for a range of 24 h ahead in terms of WSF and WPF. For WSF results, $PSR{-}SVR_{GA}$ is more accurate than PM/D2D, AR, and ARMA, for almost all horizons. Although, $PSR{-}SVR_{GA}\alpha$ performance for the day ahead horizon is an open opportunity for further improvements. Moreover, ARIMA is more accurate for the first 5 h ahead, whereas for the subsequent 16 h, $PSR{-}SVR_{GA}$ is better. In terms of phase errors, improvements in directional changes are only gained for the first hours ahead over the refence method, been marginal the gain for larger horizons. In this sense, wind speed ramp events are an opportunity area for the proposed method. For WPF results, the same performance between $PSR{-}SVR_{GA}$ and AR, ARMA, and reference methods is obtained. Improvements over PM/D2D in terms of accuracy, are larger for ARIMA than the proposed method. On constrast, improvements over the reference methods in terms of variance are larger for $PSR{-}SVR_{GA}$. In this sense, the proposed method would be useful for the mitigation of large fluctuations in wind power production.

The present study only uses a univariate time series. Still, other variables like wind direction, humidity, solar radiation, and so on, are available. Moreover, only a model of $PSR{-}SVR_{GA}$ is employed to forecast multiple steps ahead. Regarding this, an ensemble of SVR models for each individual horizon could improve the prediction accuracy. The incorporation of these variants into the model may improve accuracy in terms of WSF and WPF. Additionally, further comparisons against other reference and NWP methods are required. Bringing comparisons with NWP to the table, should provide further insight about the relation between the physical model, and the TDC model found by the PSR procedure. Therefore, future work will be delved in these opportunity areas.

Acknowledgments

This research was supported by the Universidad Autónoma de México (UNAM) under grant CJIC/CTIC/0706/2014 and the Mexican Ministry of Energy (SENER) and the Interamerican Developement Bank (IDB) through the Energy Sustaintability Fund (FSE) CONACYT-SENER.

Nomenclature

AIC	Akaike's Information Criteria
AR	autoregressive models
ARIMA	autoregressive integrated moving average models
ARMA	autoregressive moving average models
B–J	Box–Jenkins ARIMA methodology
D2D	day-to-day persistence method
DA	directional accuracy
FNN	false nearest neighbors
FS	feature selection
GA	genetic algorithms
GS	grid search method
I.I.D.	independent and identically distributed random variables
Impr	forecasting improvement
MAE	mean absolute error, also called mean absolute deviation
MASE	mean absolute scaled error
MBE	mean bias error
MI	mutual information
NN	neural networks
OLS	ordinary least squares
P inst	power installed
PM	persistence method
PSR	phase space reconstruction
PT	parameter tuning
RMSE	root mean squared error
SV	support vectors
SVM	support vector machines
SVR	support vector regression
TDC	time delay coordinates
WPF	wind power forecasting
WSF	wind speed forecasting

References

[1] R. Kavasseri, K. Seetharaman, Day-ahead wind speed forecasting using f-ARIMA models, Renew. Energy. ISSN: 09601481 34 (5) (2009) 1388–1393, http://dx.doi.org/10.1016/j.renene.2008.09.006.

[2] A. Foley, P. Leahy, A. Marvuglia, E. McKeogh, Current methods and advances in forecasting of wind power generation, Renew. Energy. ISSN: 09601481 37 (1) (2012) 1–8, http://dx.doi.org/10.1016/j.renene.2011.05.033.

[3] H. Holttinen, Optimal electricity market for wind power, Energy Policy. ISSN: 03014215 33 (2005) 2052–2063, http://dx.doi.org/10.1016/j.enpol.2004.04.001.

[4] J. Wang, S. Qin, Q. Zhou, H. Jiang, Medium-term wind speeds forecasting utilizing hybrid models for three different sites in Xinjiang, China, Renew. Energy. ISSN: 09601481 76 (2015) 91–101, http://dx.doi.org/10.1016/j.renene.2014.11.011.

[5] E. Cadenas, O. a. Jaramillo, W. Rivera, Analysis and forecasting of wind velocity in chetumal, quintana roo, using the single exponential smoothing method, Renew. Energy. ISSN: 09601481 35 (5) (2010) 925–930, http://dx.doi.org/10.1016/j.renene.2009.10.037.

[6] O. Jaramillo, M. Borja, Bimodal versus Weibull Wind Speed Distributions: an Analysis of Wind Energy Potential in La Venta, Mexico, Wind Eng. ISSN: 0309-524X 28 (2) (2004a) 225–234, http://dx.doi.org/10.1260/0309524041211404.
[7] O. Jaramillo, M. Borja, Wind speed analysis in La Ventosa, Mexico: A bimodal probability distribution case, Renew. Energy. ISSN: 09601481 29 (2004b) 1613–1630, http://dx.doi.org/10.1016/j.renene.2004.02.001.
[8] M. Jafarian, A. Ranjbar, Fuzzy modeling techniques and artificial neural networks to estimate annual energy output of a wind turbine, Renew. Energy. ISSN: 09601481 35 (9) (2010) 2008–2014, http://dx.doi.org/10.1016/j.renene.2010.02.001.
[9] I. Segura-Heras, G. Escriva-Escrivá, M. Alcázar-Ortega, Wind farm electrical power production model for load flow analysis, Renew. Energy. ISSN: 09601481 36 (3) (2011) 1008–1013, http://dx.doi.org/10.1016/j.renene.2010.09.007.
[10] C. Skittides, W. Früh, Wind forecasting using Principal Component Analysis. Renew, Energy. ISSN: 09601481 69 (September 2014) 365–374, http://dx.doi.org/10.1016/j.renene.2014.03.068. http://linkinghub.elsevier.com/retrieve/pii/S0960148114002432.
[11] C. Monteiro, R. Bessa, V. Miranda, A. Botterud, J. Wang, G. Conzelmann, Wind Power Forecasting. Technical Report, Argonne National Laboratory, 2009.
[12] C. Ferreira, J. Gama, L. Matias, A. Botterud, J. Wang, A Survey on Wind Power Ramp Forecasting. Technical Report, Argonne National Laboratory, 2010. http://www.dis.anl.gov/pubs/69166.pdf.
[13] E. Cadenas, W. Rivera, Wind speed forecasting in the South Coast of Oaxaca, México, Renew. Energy. ISSN: 09601481 32 (12) (October 2007) 2116–2128, http://dx.doi.org/10.1016/j.renene.2006.10.005. http://linkinghub.elsevier.com/retrieve/pii/S0960148106002801.
[14] M. Mohandes, T. Halawani, S. Rehman, A. Hussain, Support vector machines for wind speed prediction, Renew. Energy. ISSN: 09601481 29 (6) (May 2004) 939–947, http://dx.doi.org/10.1016/j.renene.2003.11.009. http://linkinghub.elsevier.com/retrieve/pii/S0960148103003860.
[15] K. Larson, K. Westrick, Short-term wind forecasting using off-site observations, Wind Energy. ISSN: 10954244 9 (December 2005) 55–62, http://dx.doi.org/10.1002/we.179, 2006.
[16] S. Salcedo-Sanz, E. Ortiz-García, A. Pérez-Bellido, A. Portilla-Figueras, L. Prieto, Short term wind speed prediction based on evolutionary support vector regression algorithms, Expert Syst. Appl. 38 (4) (April 2011) 4052–4057, http://dx.doi.org/10.1016/j.eswa.2010.09.067. http://linkinghub.elsevier.com/retrieve/pii/S0957417410010249.
[17] A. Paniagua-Tineo, S. Salcedo-Sanz, C. Casanova-Mateo, E. Ortiz-García, M. Cony, E. Hernández-Martín, Prediction of daily maximum temperature using a support vector regression algorithm, Renew. Energy. ISSN: 09601481 36 (11) (November 2011) 3054–3060, http://dx.doi.org/10.1016/j.renene.2011.03.030. http://linkinghub.elsevier.com/retrieve/pii/S0960148111001443.
[18] J. Zeng, W. Qiao, Short-term solar power prediction using a support vector machine, Renew. Energy. ISSN: 09601481 52 (April 2013) 118–127, http://dx.doi.org/10.1016/j.renene.2012.10.009. http://linkinghub.elsevier.com/retrieve/pii/S0960148112006465.
[19] J. Hu, J. Wang, G. Zeng, A hybrid forecasting approach applied to wind speed time series, Renew. Energy. ISSN: 0960-1481 60 (2013) 185–194, http://dx.doi.org/10.1016/j.renene.2013.05.012.
[20] K. Chen, J. Yu, Short-term wind speed prediction using an unscented Kalman filter based state-space support vector regression approach, Appl. Energy. ISSN: 03062619 113 (2014) 690–705, http://dx.doi.org/10.1016/j.apenergy.2013.08.025.
[21] D. Liu, D. Niu, H. Wang, L. Fan, Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm, Renew. Energy. ISSN: 09601481 62 (February 2014) 592–597, http://dx.doi.org/10.1016/j.renene.2013.08.011. http://linkinghub.elsevier.com/retrieve/pii/S0960148113004138.
[22] G. Santamaría-Bonfil, J. Frausto-Solís, I. Vázquez-Rodarte, Volatility forecasting using support vector regression and a hybrid genetic algorithm, Comput. Econ. ISSN: 0927-7099 (December 2013) 1–23, http://dx.doi.org/10.1007/s10614-013-9411-x. http://link.springer.com/10.1007/s10614-013-9411-x.
[23] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (2003) 1157–1182. http://dl.acm.org/citation.cfm?id=944968.
[24] F. Takens, Detecting strange attractors in turbulence, Dyn. Syst. Turbul. Warwick 1980 (1981), http://dx.doi.org/10.1007/BFb0091924. http://link.springer.com/content/pdf/10.1007/BFb0091924.pdf.
[25] M. Small, D. Walker, A. Tordesillas, Verifying chaotic dynamics from experimental data, in: Int. Symp. Nonlinear Theory its Appl. (NOLTA), Majorca, Spain, Majorca, Spaon, 2012.
[26] M. Small, Applied Nonlinear Time Series Analysis: Applications in Physics, Physiology and Finance, World Scientific, 2005, ISBN 978-981-256-117-6.
[27] S. Goh, M. Chen, D. Popović, K. Aihara, D. Obradovic, D. Mandic, Complex-valued forecasting of wind profile, Renew. Energy. ISSN: 09601481 31 (2006) 1733–1750, http://dx.doi.org/10.1016/j.renene.2005.07.006.
[28] R. Abdel-Aal, M. Elhadidy, S. Shaahid, Modeling and forecasting the mean hourly wind speed time series using GMDH-based abductive networks, Renew. Energy. ISSN: 09601481 34 (7) (2009) 1686–1699, http://dx.doi.org/10.1016/j.renene.2009.01.001.
[29] H. Madsen, P. Pinson, G. Kariniotakis, Standardizing the performance evaluation of short-term wind power prediction models, Wind Energy 29 (6) (2005) 475–489. http://multi-science.metapress.com/index/B76710556042JU72.pdf.
[30] George Box, Gwilym Jenkins, Gregory Reinsel, Time Series Analysis: Forecasting and Control, fourth ed., Wiley, 2008, ISBN 978-0-470-27284-8.
[31] H. Akaike, Information Theory and an Extension of the Maximum Likelihood Principle, in: Sel. Pap. Hirotugu Akaike, Springer, New York, 1998, ISBN 978-1-4612-7248-9, pp. 199–213, http://dx.doi.org/10.1007/978-1-4612-1694-0_15.
[32] H. Drucker, C. Burges, L. Kaufman, A. Smola, V. Vapnik, Support vector regression machines, Adv. Neural Inf. Process. Syst. 9 (1997) 155–161 doi: 10.1.1.10.4845.
[33] B. Schölkopf, A. Smola, Learning with Kernels, The MIT Press, 2002, ISBN 9780262194754.
[34] S. Huang, P. Chuang, C. Wu, H. Lai, Chaos-based support vector regressions for exchange rate forecasting, Expert Syst. Appl. ISSN: 09574174 37 (12) (December 2010) 8590–8598, http://dx.doi.org/10.1016/j.eswa.2010.06.001. http://linkinghub.elsevier.com/retrieve/pii/S0957417410005051.
[35] E. Cadenas, W. Rivera, Short term wind speed forecasting in La Venta, Oaxaca, México, using artificial neural networks, Renew. Energy. ISSN: 09601481 34 (1) (January 2009) 274–278, http://dx.doi.org/10.1016/j.renene.2008.03.014. http://linkinghub.elsevier.com/retrieve/pii/S0960148108001171.
[36] E. Cadenas, W. Rivera, Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA-ANN model, Renew. Energy. ISSN: 09601481 35 (12) (December 2010) 2732–2738, http://dx.doi.org/10.1016/j.renene.2010.04.022. http://linkinghub.elsevier.com/retrieve/pii/S0960148110001898.
[37] I. Colak, S. Sagiroglu, M. Yesilbudak, Data mining and wind power prediction: A literature review, Renew. Energy. ISSN: 09601481 46 (October 2012) 241–247, http://dx.doi.org/10.1016/j.renene.2012.02.015. http://linkinghub.elsevier.com/retrieve/pii/S0960148112001541.
[38] S. Banerjee, Dynamics for Engineers, Wiley, 2005. ISBN 978-0470868430.
[39] M. Kennel, R. Brown, H. Abarbanel, Determining embedding dimension for phase-space reconstruction using a geometrical construction, Phys. Rev. A. ISSN: 1050-2947 45 (6) (March 1992) 3403–3411. http://www.ncbi.nlm.nih.gov/pubmed/9907388.
[40] J. Holland, Adaption in Natural and Artificial Systems, vol. 1, 1975, http://dx.doi.org/10.1162/1064546053278919. ISBN 0262581116.
[41] Y. Cancino-Solórzano, A. Gutiérrez-Trashorras, J. Xiberta-Bernat, Current state of wind energy in Mexico, achievements and perspectives, Renew. Sustain. Energy Rev. ISSN: 13640321 15 (2011) 3552–3557, http://dx.doi.org/10.1016/j.rser.2011.05.009, 2011.
[42] G. Aleman-Nava, V. Casiano-Flores, D. Cárdenas-Chávez, R. Díaz-Chavez, N. Scarlat, J. Mahlknecht, J. Dallemand, R. Parra, Renewable energy research progress in Mexico: A review, Renew. Sustain. Energy Rev. ISSN: 13640321 32 (2014) 140–153, http://dx.doi.org/10.1016/j.rser.2014.01.004.
[43] D. Elliott, M. Schwartz, G. Scott, S. Haymes, D. Heimiller, R. George, Wind Energy Resource Atlas of Oaxaca, 2004. NREL/TP-500-34519, http://www.nrel.gov/wind/pdfs/34519.pdf.
[44] International Electrotechnical Commission (IEC), IEC 61400-1 Wind Turbine Generator Systems – Part 1: Design Requirements, third ed., 2005.
[45] M. Borja, J. Huacuz, J. Lopez-Laing, J. Tejeda, Main results of the action plan for removing barriers to the implementation of wind power in Mexico, in: Eur. Wind Energy Assoc. Annu. Event. 2011. Brussels, 2011, pp. 1–6. Brussels, Belgium, http://proceedings.ewea.org/annual2011/allfiles2/1395_EWEA2011presentation.pdf.
[46] C. Jarque, A. Bera, A test for normality of observations and regression residuals, Stat. Rev. Int. Stat. ISSN: 0036-8075 55 (1987) 163–172, http://dx.doi.org/10.1126/science.ns-9.225.507.
[47] M. Shintani, O. Linton, Nonparametric neural network estimation of Lyapunov exponents and a direct test for chaos, J. Econom. ISSN: 03044076 120 (1) (May 2004) 1–33, http://dx.doi.org/10.1016/S0304-4076(03)00205-7. http://linkinghub.elsevier.com/retrieve/pii/S0304407603002057.
[48] C. Gershenson, The implications of interactions for science and philosophy, Found. Sci. ISSN: 1233-1821 18 (4) (October 2012) 781–790, http://dx.doi.org/10.1007/s10699-012-9305-8. http://link.springer.com/10.1007/s10699-012-9305-8.
[49] N. Fernandez, C. Maldonado, C. Gershenson, Information Measures of Complexity, Emergence, Self-organization, Homeostasis, and Autopoiesis, Springer Berlin Heidelberg, 2014 doi: 10.1007/978-3-642-53734-9_2. http://link.springer.com/chapter/10.1007/978-3-642-53734-9_2.
[50] Am Fraser, Hl Swinney, Independent coordinates for strange attractors from mutual information, Phys. Rev. A. ISSN: 1050-2947 33 (2) (February 1986) 1134–1140. http://www.ncbi.nlm.nih.gov/pubmed/9896728.
[51] M. Lange, On the uncertainty of wind power predictions – analysis of the forecast accuracy and statistical distribution of errors, J. Sol. Energy Eng. ISSN: 01996231 127 (May 2005) 177, http://dx.doi.org/10.1115/1.1862266, 2005.
[52] D. Sanders, M. Manfredo, K. Boris, Accuracy and efficiency in the U.S. Department of Energy's short-term supply forecasts, Energy Econ. ISSN: 01409883 30 (2008) 1192–1207, http://dx.doi.org/10.1016/j.eneco.2007.01.011.
[53] M. Bielecki, J. Kemper, T. Acker, A methodology for comprehensive characterization of errors in wind power forecasting, in: Es2010 Proc. Asme 4th Int. Conf. Energy Sustain, vol. 2, 2010, pp. 867–876, http://dx.doi.org/10.1115/ES2010-90381. <Go to ISI>://000283271300101.

[54] M. Yoder, A. Hering, W. Navidi, K. Larson, Short-term forecasting of categorical changes in wind power with Markov chain models, Wind Energy. ISSN: 10991824 17 (June 2013) 1425–1439, http://dx.doi.org/10.1002/we. 1641 http://onlinelibrary.wiley.com/doi/10.1002/we.1608/full.
[55] C. Willmott, K. Matsuura, Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, Clim. Res. ISSN: 0936577X 30 (2005) 79–82, http://dx.doi.org/10.3354/cr030079.
[56] J. Jung, R. Broadwater, Current status and future advances for wind speed and power forecasting, Renew. Sustain. Energy Rev. ISSN: 13640321 31 (2014) 762–777, http://dx.doi.org/10.1016/j.rser.2013.12.054.
[57] R. Hyndman, A. Koehler, Another look at measures of forecast accuracy, Int. J. Forecast. ISSN: 01692070 22 (4) (October 2006) 679–688, http://dx.doi.org/10.1016/j.ijforecast.2006.03.001. http://linkinghub.elsevier.com/retrieve/pii/S0169207006000239.
[58] S. Bivona, G. Bonanno, R. Burlon, D. Gurrera, C. Leone, Stochastic models for wind speed forecasting, Energy Convers. Manag. ISSN: 01968904 52 (2) (2011) 1157–1165, http://dx.doi.org/10.1016/j.enconman.2010.09.010.
[59] Z. Olaofe, A 5-day wind speed & power forecasts using a layer recurrent neural network (LRNN), Sustain. Energy Technol. Assess. ISSN: 22131388 6 (2014) 1–24, http://dx.doi.org/10.1016/j.seta.2013.12.001.
[60] O. Blaskowitz, H. Herwartz, On economic evaluation of directional forecasts, Int. J. Forecast 27 (4) (2011) 1058–1065, http://dx.doi.org/10.1016/j.ijforecast.2010.07.002. ISSN 01692070.
[61] Oliver Blaskowitz, Helmut Herwartz, Testing the value of directional forecasts in the presence of serial correlation, Int. J. Forecast. ISSN: 01692070 30 (1) (2014) 30–42, http://dx.doi.org/10.1016/j.ijforecast.2013.06.001.
[62] D. Montgomery, C. Jennings, M. Kulahci, Introduction to Time Series Analysis and Forecasting, Wiley, 2008, ISBN 978-0-471-65397-4.
[63] R. Weron, Electricity price forecasting : A review of the state-of-the-art with a look into the future, Int. J. Forecast. ISSN: 01692070 30 (4) (October 2014) 1030–1081, http://dx.doi.org/10.1016/j.ijforecast.2014.08.008. http://www.sciencedirect.com/science/article/pii/S0169207014001083.
[64] H. Madsen, G. Kariniotakis, H. Nielsen, T. Nielsen, P. Pinson, A Protocol for standardazing the performance evaluation of short term wind power prediction models, in: Proc. Glob. Wind. Conf. Exhib., Chicago, 2004.
[65] C. Chang, C. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. ISSN: 21576904 2 (2011) 1–27, http://dx.doi.org/10.1145/1961189.1961199. http://dl.acm.org/citation.cfm?doid=1961189.1961199.
[66] N. Marwan, J. Kurths, Nonlinear analysis of bivariate data with cross recurrence plots, Phys. Lett. A. ISSN: 03759601 302 (2002) 299–307, http://dx.doi.org/10.1016/S0375-9601(02)01170-2.