Hybrid Machine Intelligent SVR Variants for Wind Forecasting and Ramp Events

Dhiman, Harsh S. ; Deb, Dipankar ; Guerrero, Josep M.

Published in:
Renewable & Sustainable Energy Reviews

DOI (link to publication from Publisher):
10.1016/j.rser.2019.04.002

Publication date:
2019

Document Version
Early version, also known as pre-print

Hybrid Machine Intelligent SVR Variants for Wind Forecasting and Ramp Events

Harsh S. Dhiman[a], Dipankar Deb[a], Josep M. Guerrero[b]

[a]*Department of Electrical Engineering*
Institute of Infrastructure Technology Research and Management, Ahmedabad, India 380026.
[b]*Department of Energy Technology*
Aalborg University, 9220 Aalborg East, Denmark

Abstract

Wind speed and power forecast is an essential component to ensure grid stability and reliability. The traditional forecasting methods fail to address the non-linearity in the wind speed time-series, thus paving way for machine intelligent algorithms. This paper discusses a hybrid machine intelligent wind forecasting model utilizing different variants of Support Vector Regression (SVR) built on wavelet transform. Various performance indices are evaluated to identify the possible best one among four different machine learning regressors for wind forecasting application. Apart from standard ε-SVR and LS-SVR, two new regression models, namely, ε-Twin Support vector regression (ε-TSVR) and Twin Support vector regression (TSVR) are used to forecast short-term wind speed, and are compared with Persistence model for four wind farm sites. The effect of larger dataset on forecasting performance is evaluated for two wind farm sites from USA and India. Further, wind power ramp events are investigated at different hub heights and the forecasting performance of different variants of SVR is compared for five wind farm sites.

Keywords: Wind forecasting, Wavelet transform, Twin Support vector regression, ε-Twin Support vector regression, Wind power ramp events

Email addresses: harsh.dhiman.17pe@iitram.ac.in (Harsh S. Dhiman), dipankardeb@iitram.ac.in (Dipankar Deb), joz@et.aau.dk (Josep M. Guerrero)

1. Introduction

Growing wind energy potential is attracting investments in the renewable energy market. With abundant wind availability, tapping power from wind is important. The demand for energy has pushed the envelope for renewable energy technologies, and Solar, wind, biomass being the pioneers, many developing countries are now focusing on utilizing the sustainable sources of energy. Wind energy brings a balance in the ecosystem by compensating the carbon footprints created by thermal power plants. Globally, wind energy brings job opportunities particularly in operations and maintenance (O&M) sector. According to Global wind energy council (GWEC) report [1], in 2017, with an installed capacity of 2.08 GW, wind sector in South Africa created 15,000 jobs while in Europe a total of 262,712 jobs were created. Lucrative tariff rates have ensured support for wind technology, both onshore and offshore.

Despite numerous advantages, wind sector leads to an imbalance in aquatic life, high initial investment costs and procedural obstacles in land acquisition. But, advanced manufacturing technologies have opened doors for rapid wind energy installations, and wind regime for offshore sites is found much stronger than onshore ones, which motivates investors to participate in bidding process. Threats posed by wind turbines include bird killings, high noise levels, opposition from local communities including farmers concerned about their livestock. Wind farms are being constructed keeping in mind the space constraints and recently a lot of focus is fed on Savonius style wind turbines (SSWT) operating under any wind direction. Roy et al. have discussed an inverse method based on differential evolution for determining optimal turbine dimensions [2]. Results reveal that area of SSWT is reduced by 9.8%. Further, a 2D computational fluid dynamics model is put forward by Gupta and Biswas to evaluate the steady state performance of a twisted three-bladed H-Darrieus rotor [3]. Considering wind as a stochastic variable, its accurate prediction can yield benefits to the plant operators. However, the error processing of forecasted wind speed/power and actual wind speed/power plays a crucial role in selecting appropriate forecasting algorithms. Machine learning models like Artificial neural networks (ANN), Support vector regression (SVR) [4]-[5], Gaussian process regression (GPR), Fuzzy logic and Extreme learning machine are widely used.

Recently a lot of impetus has been laid on hybrid wind forecasting that incorporates the advantageous aspects of individual methods. Earlier works in the field of hybrid wind forecasting include ARIMA-ANN model developed by Candenas and Rivera where for a fixed prediction horizon, wind forecasting is done [6]. Liu et al. have described a Support vector machine and Genetic algorithm (GA) based hybrid short-term forecasting technique using Wavelet transform for the decomposition of the wind signal and removal of any stochastic variations [7]. Zhang et al. have proposed a hybrid method based on gaussian process regression (GPR) and auto-regression (AR) and compared their wind speed forecast with results obtained through ANN, SVM and persistence models [8]. Mi et al. have described a hybrid model employing wavelet transform, extreme learning machine and outlier correction method to predict multi-step wind speed [9]. Wavelet and wavelet packet decomposition removes noise component from the wind series and extreme learning machine provides multi-step forecast on the sub-layers obtained in decomposition process.

Li et al. have discussed combined models based on variable weight and constant weight for short-term wind speed forecasting [10]. Jiang et al. have proposed a hybrid model employing fluctuations of adjacent wind turbines on target wind turbine and the relevant inputs are fed to the v-SVM model for forecasting short-term wind speed [11]. Azimi et al. have used data mining and wavelet analysis to perform k-means cluster selection of significant features from wind speed time series and the forecast is done using multilayer perceptron neural network (MLPNN) [12]. Jiang et al. have proposed correlation-aided discrete wavelet transform (DWT), least-square support vector machine (LSSVM) and generalized autoregressive conditional heteroscedastic (GARCH) model. The DWT is carried out to decompose original wind series into sub-series and a correlation coefficient is calculated between each sub-series and original dataset to select inputs for LSSVM model [13]. Further, a multi-step forecasting model based on a hybrid structure involving a modified BFGS neural network and wavelet decomposition based post processing technique is built by Liu et al. and is validated for four wind speed time series [14]. The effectiveness of wavelet filter based decomposition is observed by analyzing the cross-correlation coefficients between the instantaneous frequency components of sub-series. Tian et al. have proposed a hybrid preprocessing and satin bowerbird based multi-objective forecasting algorithm [15] wherein data preprocessing is based on complementary ensemble empirical mode decomposition (EEMD), sample entropy and variational mode decomposition. The proposed method is validated for eight datasets and is found to be superior to the benchmark models, but suffers from large computation time.

Wang et al. have implemented a novel hybrid model involving modern dragonfly algorithm (MODA), an optimization technique to tune the parameters and weights of elman neural network(ENN) to forecast three variables, that is, wind speed, electricity price and electrical load [16]. In order to remove noise and non-linear components from the wind speed time series, several decomposition algorithms like empirical mode decomposition (EMD), wavelet transform (WT) and EEMD are used. On similar grounds, Du et al. carried out multi-step ahead forecasting based on a Whale optimization algorithm-LSSVR model and have applied the same to forecast wind speed, electrical load and electricity price [17]. Six different datasets from China, Australia and Singapore are tested for the proposed approach and are compared with Generalized regression neural network (GRNN) and Back propagation neural network (BPNN). Results reveal that WOA-LSSVR model outperforms GRNN and BPNN models in terms of mean squared error, mean absolute error and mean absolute percentage error.

Further, Debanath et al. have presented a ANN model to predict the power and torque coefficients for a three-buck savonius type wind turbine. The model has three inputs: (i) overlap ratio, (ii) tip-speed ratio and (iii) angular velocity [18]. Results reveal that a two-hidden layer ANN outperforms single-layer and three-layer ANN topology. Wang et al. have proposed a newly developed hybrid wavelet neural network (WNN) model based multi-objective sine-cosine algorithm (MOSCA) optimization [16]. The model developed is tested for high accuracy and stability in order to ensure a reliable wind farm operation. Further, based on WNN-MOSCA model each sub-series is forecasted and a aggregated time series is obtained. The proposed WNN-MOSCA model is compared with ARIMA, persistence, WNN and GRNN models. However, the above mentioned forecasting models consume large computation time which is reduced via a hybrid SVR model and associated variants. Wavelet transform, a special Multi-Resolution Analysis (MRA) technique which fragments the input signal in time-frequency domain, is primarily used for power system transients like power ramp-up and ramp-down events that cause severe system jeopardy [19]. In this paper, we decompose the wind speed time-series signal using daubechies fourth order (db4) wavelet filter which ensures smooth and localized decomposition.

The main contribution of this manuscript is a hybrid model for wind forecasting based on wavelet transform and SVR variants. The hybrid model is then compared with persistence model based on several performance metrics and computation time. Effect of regularization on variants of SVR is assessed to evaluate the best hybrid model in terms of short-term forecasting. Further, wind ramp events are assessed for five wind farm sites under different variants of SVR along with frequency distribution at different hub heights. This paper is divided as follows. Section 2 describes various SVR variants and their problem formulation. Further, Section 3 discusses the framework for short-term wind speed forecasting and wind power ramp events. In Section 4, results and discussions are presented followed by Conclusions in Section 5.

2. Support Vector Regression

Support vector regression (SVR) works on the principle of structural risk minimization (SRM) from statistical learning theory [20], [21]. The core idea of the SRM theory is to arrive at a hypotheses h which can yield lowest true error for the unseen and random sample testing data [22]. Apart from SVR, a universal machine intelligent technique called Artificial neural network (ANN) with applications in character recognition, image compression and stock market prediction, is studied [23]. Shirzad et al. have compared the performance of ANN and SVR to predict the Pipe Burst Rate (PBR) in Water Distribution Networks (WDNs) [24]. It was observed that ANN is a better predictor than SVR but cannot be generalized as it is not consistent with physical behavior. SVR has an advantage over ANN with respect to the number of parameters involved in training phase. The computation time is another important factor for carrying out regression analysis.

Consider a set of training data (historical data) $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n) \subset X \times \mathrm{R}$, where X denotes the input feature space of dimension $\mathrm{R^n}$. Let $Y = (y_1, y_2, \ldots, y_i)$ denote the set representing the training output or response, where $i = 1, 2, \ldots, n$ and $y_i \in \mathrm{R}$.

2.1. ε-support vector regression

ε-SVR aims to find a regressor

$$f(x) = w^T x + b, with\ w \in X, b \in \mathrm{R} \tag{1}$$

which represents a linear regression function for prediction, where $x \in X$ is the input set containing all the features, w is the weight coefficient related to each input vector x_i and b is the bias term.

The aim is to find out $f(x)$ with maximum deviation ε from the respective feature sets or classes while being as flat as possible. In order to achieve the flatness of the desired regressor, the square of the norm of weight vector w needs to be minimized. Thus we can formulate the SVR problem into a convex optimization problem [25] given as

$$\min \frac{1}{2} \parallel w \parallel^2 + C(e^T \chi + e^T \chi^*), \tag{2}$$

$$\textbf{subject to}\ \ y - w^T x - eb \leq e\varepsilon + \chi, \chi \geq 0, \tag{3}$$

$$w^T x + eb - y \leq e\varepsilon + \chi^*, \chi^* \geq 0,$$

where C is the regularization factor that reflects the trade-off between the flatness of regressor $f(x)$ and the maximum deviation ε which could be tolerated. The variables χ, χ^* are the slack variables introduced as a soft margin to the tolerable error ε and e is a vector of ones of appropriate dimensions $(n \times 1)$. However, this is not the case always, as the feature sets might not be linearly separable. To handle such nonlinearities in the feature sets, kernel trick or often called as kernel functions are used to transform data to a higher dimensional space. After transformation via suitable mapping function $\phi : \mathrm{R^n} \to Z$, the data becomes linearly separable in the target space (high dimensional space), that is, Z. The inner product $\langle w^T, \phi(x)\rangle$ in the target space can be represented by using kernel function. Kernel functions are similarity functions which satisfy Mercer's theorem such that $k(x_i, x_j) = \langle \phi(x_i), \phi(x_j)\rangle$, are the elements of the kernel matrix K. Several kernel functions are available in literature like linear, polynomial with degree d, gaussian, Radial Basis Function (RBF) with bandwidth of the function σ and exponential function.

The SVR optimization problem can be extended into its dual form as follows:

$$\min \frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i - \alpha_i^*)^T k(x_i, x_j)(\alpha_j - \alpha_j^*) + e^T\varepsilon\sum_{i=1}^{n}(\alpha + \alpha^*) - \sum_{i=1}^{n} y_i(\alpha - \alpha^*) \quad (4)$$

$$\textbf{s.t.} \quad e^T\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0, \quad 0 \leq \alpha, \quad \alpha^* \leq Ce,$$

where α and α^* represent positive and negative Lagrange multipliers such that $\alpha_i\alpha_i^* = 0$, $i = 1, 2, \ldots, n$. The regressor $f(x)$ can be written as

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)k(x, x_i) + b. \quad (5)$$

The complexity of this regressor is independent of the dimensionality of the feature set but only depends on the number of support vectors which are nothing but the data points which separate the feature sets from each other. However the performance of the SVR also depends on the choice of kernel function and helps in reducing the computation time of the regression.

2.2. Least square support vector regression

Least-square support vector regression (LS-SVR) originally derived from least-square support vector classifiers (LS-SVC) proposed by [26] where equality constraints are chosen and the square of the error term ε is minimized. The LS-SVR regression problem is formulated as

$$f(x) = w^T \phi(x) + b, \tag{6}$$

where w is the weight coefficient vector of dimension $(n \times 1)$ and $x_i \in \mathrm{R^n}$, $y \in \mathrm{R}$. The objective function to be minimized for LS-SVR is given as

$$\min \ \frac{1}{2} \parallel w \parallel^2 + \frac{1}{2}\gamma \sum_{i=1}^{n} \varepsilon_i^2 \tag{7}$$

$$\textbf{s.t.} \ \ y_i =^T \phi(x_i) + b + \varepsilon_i, \quad (i = 1, 2, \ldots, n), \tag{8}$$

where γ is the margin parameter and ε_i is the error term corresponding to each x_i. The optimization problem can be transformed by introducing Lagrange multipliers and is given as

$$L(w, b, \varepsilon, \alpha) = \frac{1}{2} \parallel w \parallel^2 + \frac{1}{2}\gamma \sum_{i=1}^{n} \varepsilon_i^2 - \sum_{i=1}^{n} \alpha_i (w^T \phi(x_i) + b + \varepsilon - y_i). \tag{9}$$

The Karush-Kuhn-Tucker (KKT) conditions for the optimization problem (9) can be obtained by partially differentiating the Lagrangian function with respect to $w, b, \varepsilon, \alpha$ which gives the solution in the matrix form

$$\begin{bmatrix} k(x, x^T) + \gamma^{-1} I & e \\ e^T & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} y \\ 0 \end{bmatrix}, \tag{10}$$

$$f_{LS-SVR}(x) = \sum_{i=1}^{n} \alpha_i k(x, x_i) + b, \tag{11}$$

where I is the identity matrix of appropriate dimension. The regressor obtained by LS-SVR is given by (11) and solves the optimization problem of smaller size than classical ε-SVR thus taking less computation time.

2.3. Twin support vector regression

Xinjun introduced an efficient way to solve the regression through support vector machines through a Twin Support Vector Regression (TSVR) that aims to derive two non-parallel hyperplanes around the data points [27]. Similar to ε-SVR, TSVR finds two ε-insensitive functions, that is, up-bound and down-bound regressors. Further TSVR solves the convex optimization problem having size smaller than the conventional ε-SVR thus reducing significant time on CPU. The mathematical formulation of TSVR is

$$\min \; \frac{1}{2}\sum_{i=1}^{n}(y_i - e\varepsilon_1 - (x_i w_1 + eb_1))^T(y_i - e\varepsilon_1 - (x_i w_1 + eb_1)) + C_1 e^T \sum_{i=1}^{n}\xi_i, \quad \textbf{s.t.}\; y_i - (x_i w_1 + eb_1) \geq e\varepsilon_1 - \xi_i, \tag{12}$$

$$\min \; \frac{1}{2}\sum_{i=1}^{n}(y_i - e\varepsilon_2 - (x_i w_1 + eb_2))^T(y_i - e\varepsilon_2 - (x_i w_2 + eb_2)) + C_2 e^T \sum_{i=1}^{n}\eta_i,, \quad \textbf{s.t.}\; (x_i w_2 + eb_2) - y_i \geq e\varepsilon_2 - \eta_i, \tag{13}$$

where $C_1, C_2 > 0$ and $\varepsilon_1, \varepsilon_2 \geq 0$ are the TSVR hyperparameters and ξ_i, η_i are the slack variables introduced as a soft margin to the error ε in optimization problem. The dual optimization problem formulation of TSVR is given by introducing a Lagrangian function [27]. Let $X = (x_1, x_2, \ldots, x_n)$ denote the set of input vectors, $Y = (y_1, y_2, \ldots, y_n)$ be the set of output vectors, where $y_i \in \mathrm{R}$ and α, γ are the lagrangian multipliers.

Combining the KKT conditions [27] and optimization problem described by (12), the dual can be reformulated as

$$\max \; -\frac{1}{2}\alpha^T Q(Q^T Q)^{-1}Q^T\alpha + t^T Q(Q^T Q)^{-1}Q^T\alpha - t^T\alpha \quad \textbf{s.t.}\; \alpha \in [0, C_1] \tag{14}$$

$$\max \; -\frac{1}{2}\gamma^T Q(Q^T Q)^{-1}Q^T\gamma + m^T Q(Q^T Q)^{-1}Q^T\gamma - m^T\gamma \quad \textbf{s.t.}\; \gamma \in [0, C_2], \tag{15}$$

where $Q = [X\ e]$, $t = Y - e\varepsilon_1$, $m = Y + e\varepsilon_2$ and $u_2 = (Q^T Q)^{-1}Q^T(m - \gamma)$. Equations (14-15) refer to the dual of original convex optimization problem where the size of the former is smaller than classical SVR thereby making it faster than it. The final regressor for predicting raw data points is given as

$$f_{TSVR}(x) = \frac{1}{2}((w_1 + w_2)^T x + (b_1 + b_2)). \tag{16}$$

2.4. ε-Twin support vector regression

Derived from Twin support vector machine discussed in previous section, Shao et al. [28] propose a novel regressor- ε-Twin support vector regression (ε-TSVR) that determines the pair of ε-insensitive functions by solving two convex optimization problems. In terms of the objective function to be minimized, ε-TSVR considers an added regularization term that solves the ill-conditioning problem of Q^TQ. The formulation of primal objective functions for ε-TSVR are

$$\min \ \frac{1}{2}C_3(w_1^T w_1 + b_1^2) + \frac{1}{2}\xi *^T \xi + C_1 e^T \xi, \quad \textbf{s.t.} \ Y - (Xw_1 + eb_1) = \xi^*, \tag{17}$$

$$Y - (Xw_1 + eb_1) \geq -e\varepsilon_1 - \xi, \xi \geq 0, \tag{18}$$

$$\min \ \frac{1}{2}C_4(w_2^T w_2 + b_2^2) + \frac{1}{2}\xi *^T \xi + C_2 e^T \eta, \quad \textbf{s.t.} \ Y - (Xw_2 + eb_2) = \eta^*, \tag{19}$$

$$Y - (Xw_2 + eb_2) \geq -e\varepsilon_2 - \eta, \eta \geq 0, \tag{20}$$

In the optimization problem $C_1, C_2, \varepsilon_1, \varepsilon_2$ are the hyperparameters that determine the regression performance. The Lagrangian function for the above two primal problems can be written as

$$\begin{aligned} L(w_1, b_1, \xi, \alpha, \beta) &= \frac{1}{2}(Y - (Xw_1 + eb_1))^T(Y - (Xw_1 + eb_1)) \\ &+ \frac{1}{2}C_3(w_1^T w_1 + b_1^2) + C_1 e^T \xi \\ &- \alpha^T(Y - (Xw_1 + eb_1) + e\varepsilon_1 + \xi) - \beta^T \xi, \end{aligned} \tag{21}$$

where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ and $\beta = (\beta_1, \beta_2, \ldots, \beta_n)$ are the Lagrangian multipliers. In order to obtain the dual of the above stated primal objective functions, KKT conditions are given by

$$\begin{cases} \dfrac{\partial L}{\partial w_1} = 0 \Rightarrow -X^T(Y - Xw_1 - eb_2 - e\varepsilon_1) + X^T\alpha + C_3 w_1 = 0 \\ \dfrac{\partial L}{\partial b_1} = 0 \Rightarrow -e^T(Y - Xw_1 - e\varepsilon_1 - eb_2) + e^T\alpha + C_3 b_1 = 0 \\ \dfrac{\partial L}{\partial \xi} = 0 \Rightarrow C_1 e^T - \alpha - \beta = 0 \\ \dfrac{\partial L}{\partial \alpha} = 0 \Rightarrow Y - (Xw_1 + eb_1) \geq -e\varepsilon - \xi, \quad \xi \geq 0, \end{cases}$$

$$\alpha^T(Y-(Xw_1+eb_1)+e\varepsilon_1+\xi)=0. \;\; \alpha=0, \;\; \beta^T\xi=0, \;\; \beta\geq 0, \tag{22}$$

where $\alpha\in[0,C_1e]$ for $\beta\geq 0$. The above KKT conditions can be combined and can be written as

$$-\begin{bmatrix}X^T\\ e^T\end{bmatrix}Y+\left(\begin{bmatrix}X^T\\ e^T\end{bmatrix}\begin{bmatrix}X & e\end{bmatrix}+C_3I\right)\begin{bmatrix}w_1\\ b_1\end{bmatrix}+\begin{bmatrix}X^T\\ e^T\end{bmatrix}\alpha=0. \tag{23}$$

Let us define

$$Q=\begin{bmatrix}X & e\end{bmatrix}, \;\; u_1=\begin{bmatrix}w_1^T & b_1\end{bmatrix}^T, \tag{24}$$

and rewriting (23) as

$$-Q^TY+(Q^TQ+C_3I)u_1+Q^T\alpha=0. \tag{25}$$

Further we can write $u_1=(Q^TQ+C_3I)^{-1}Q^T(Y-\alpha)$. The dual optimization objective function for the above primal can be written as

$$\begin{aligned}\max \; -\frac{1}{2}\alpha^TQ(Q^TQ+C_3I)^{-1}Q^T\alpha^T+Y^TQ(Q^TQ+C_3I)^{-1}Q^T\alpha\\ -(e^T\varepsilon_1+Y^T)\alpha, \quad \textbf{s.t. } \alpha\in[0,C_1].\end{aligned} \tag{26}$$

Similarly the other dual can be obtained as

$$\begin{aligned}\max \; -\frac{1}{2}\gamma^TQ(Q^TQ+C_4I)^{-1}Q^T\gamma^T+Y^TQ(Q^TQ+C_4I)^{-1}Q^T\gamma\\ +(-e^T\varepsilon_2+Y^T)\gamma, \quad \textbf{s.t. } \gamma\in[0,C_2].\end{aligned} \tag{27}$$

The equations (26) and (27) are the duals of the primal objective optimization function when the feature set X is linearly separable in n-dimensional space. The end regressor $f(x)$ which is the mean of two functions $f_1(x)$ and $f_2(x)$, is

$$f(x)=\frac{1}{2}(f_1(x)+f_2(x))=\frac{1}{2}((w_1+w_2)^Tx+(b_1+b_2)). \tag{28}$$

2.4.1. Kernel ε-Twin support vector regression

However to extend this study to the non-linear regression the input set is transformed into higher dimension using a suitable mapping function $\phi:\mathrm{R^n}\rightarrow$

$\mathrm{R^k}$, where k is the dimension in target space. In order to avoid the selection of appropriate mapping function, kernel functions are used to transform the data into higher dimension space. As in [28], the convex optimization problem considering kernel function $K(X, X^T)$ is given as

$$\min \ \frac{1}{2}C_3(w_1^T w_1 + b_1^2) + \frac{1}{2}\xi *^T \xi + C_1 e^T \xi,$$
$$\textbf{s.t.} \ Y - (K(X, X^T)w_1 + eb_1) = \xi^*. \tag{29}$$

$$Y - (K(X, X^T)w_1 + eb_1) \geq -e\varepsilon_1 - \xi, \xi \geq 0, \tag{30}$$

$$\min \ \frac{1}{2}C_4(w_2^T w_2 + b_2^2) + \frac{1}{2}\xi *^T \xi + C_2 e^T \eta,$$
$$\textbf{s.t.} \ Y - (K(X, X^T)w_2 + eb_2) = \eta^*. \tag{31}$$

$$Y - (K(X, X^T)w_2 + eb_2) \geq -e\varepsilon_2 - \eta, \eta \geq 0, \tag{32}$$

where C_1, C_2, C_3, C_4 are the hyperparameters for kenrel-based ε-TSVR. The duals of the primal optimization problems are given as

$$\max \ -\frac{1}{2}\alpha^T S(S^T S + C_3 I)^{-1} S^T \alpha^T + Y^T S(S^T s + C_3 I)^{-1} S^T \alpha$$
$$-(e^T \varepsilon_1 + Y^T)\alpha, \quad \textbf{s.t.} \ \alpha \in [0, C_1], \tag{33}$$

$$\max \ -\frac{1}{2}\gamma^T S(S^T S + C_4 I)^{-1} S^T \gamma^T + Y^T S(S^T S + C_4 I)^{-1} S^T \gamma$$
$$+(-e^T \varepsilon_2 + Y^T)\gamma, \quad \textbf{s.t.} \ \gamma \in [0, C_2], \tag{34}$$

where $S = [K(X, X^T) \quad e]$ and α, γ are the Lagrangian multipliers. The end regressor $f_{\varepsilon-TSVR}(x)$ is given as the mean of the two functions, given as

$$f_{\varepsilon-TSVR}(x) = \frac{1}{2}((w_1^T + w_2^T)K(X, X^T) + (b_1 + b_2)). \tag{35}$$

It should be noted that, by varying the value of C_3 in (33), the regression accuracy can be improved and is validated for one of the datasets in Section 4.

3. Framework of hybrid forecasting model

The present study deals with short-term wind speed prediction using a hybrid method involving wavelet transform and support vector regression. Hybrid methods hold an advantage over individual methods in terms of filtering any stochastic volatility. The error in wind speed prediction depends on the prediction horizon, i.e. the time frame for which the forecasting is supposed to be carried out. For market clearing operations and economic load dispatch usually short-term wind speed prediction ranging from 30 minutes to 3 hours is a preferred choice. A hybrid method involving wavelet transform and SVR variants, is used for short-term forecast for different wind sites. The wind forecasting is carried out using the hybrid model, that is, Wavelet-SVR, Wavelet-LSSVR, Wavelet-TSVR and Wavelet-ε-TSVR. The forecasting accuracy is evaluated by computing various performance metrics like Root mean squared error (RMSE), Mean absolute error (MAE), Sum of squared residuals (SSR) and Sum of squared deviation of testing samples (SST), Sum of squared error of testing samples (SSE) Index of agreement (IOA), Theil's U1 and U2 statistic [29]. Mathematically these metrics are expressed as

$$
\begin{aligned}
RMSE &= \left[\frac{1}{n}\sum_{i=1}^{n}(\hat{x_i}-x_i)^2\right]^{1/2}, \quad MAE = \left[\frac{1}{n}\sum_{i=1}^{n}|\hat{x_i}-x_i|\right] \\
SSR/SST &= \frac{\sum_{i=1}^{n}(\hat{x_i}-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x_i})^2}, \quad SSE/SST = \frac{\sum_{i=1}^{n}(\hat{x_i}-x_i)^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}, \\
IOA &= 1-\sum_{i=1}^{n}\left(\hat{x}_i-x_i\right)^2 \Big/ \sum_{i=1}^{n}\left(|\hat{x}_i-\overline{x}|+|x_i+\overline{x}|\right)^2 \\
U1 &= \sqrt{\frac{1}{n}\times\sum_{i=1}^{n}\left(\hat{x}_i-x_i\right)^2} \Bigg/ \left(\sqrt{\frac{1}{n}\times\sum_{i=1}^{n}x_i^2}+\sqrt{\frac{1}{n}\times\sum_{i=1}^{n}\hat{x}_i^2}\right) \\
U2 &= \sqrt{\frac{1}{n}\times\sum_{i=1}^{n}\left(\left(x_{i+1}-\hat{x}_{i+1}\right)/x_i\right)^2} \Bigg/ \sqrt{\frac{1}{n}\times\sum_{i=1}^{n}\left(\left(x_{i+1}-\hat{x}_i\right)/x_i\right)^2}
\end{aligned}
$$

where $\hat{x_i}, x_i, \bar{x}$ are the predicted, actual and mean values of the testing samples.

Figure 1 shows the block diagram of forecasting through hybrid wavelet-SVR method. First, the original wind speed time series is decomposed into low frequency and high frequency components. Further, the appropriate decomposition signals are selected as inputs to the SVR forecasting model. The wavelet filter chosen was daubechies 'db4' with 5-level decomposition. Wavelet transform can be categorized as continuous (CWT) and discrete wavelet transform (DWT).

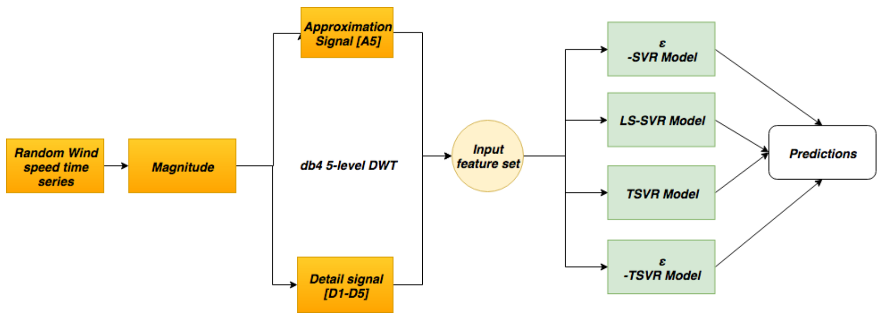

Figure 1: Wind forecasting using Wavelet transform and SVR

Computationally DWT is more rich than CWT due to which former finds more use in signal processing. Mathematically CWT and DWT are expressed as

$$B(a,b) \quad = \quad = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} r(x)\phi\left(\frac{x-b}{a}\right), \tag{36}$$

$$B(u,v) \quad = \quad 2^{-u/2}\sum_{t=0}^{N-1} r(t)\phi\left(\frac{t-v.2^u}{2^u}\right), \tag{37}$$

where $r(t)$ is the wind speed time series and N is its length, $\phi(.)$ is the mother wavelet function, and scaling and translation parameters are functions of integers u and v. The WT process involves successive decomposition of approximation signal obtained at each stage. The two signals obtained at each decomposition stage are approximate and detail signals, former containing low-frequency components and later high-frequency components. The approximate (A5) and detail signals (D1, D2, D3, D4 and D5) together form a matrix of input features and wind speed is the output used in short-term wind forecasting algorithm (here SVR and its variants).

3.1. Description of Datasets

To test the hybrid wavelet-SVR wind farm sites from Spain, Western Massachusetts (USA), South Dakota (USA), Victoria (Australia) and India are chosen with their descriptive statistics being listed in Table 1, and are selected to test the forecasting performance based on wavelet-SVR and its variants. Figure 2 shows the wind speed variations for these wind farm sites.

- Paxton, MA: The wind site is located in western Massachusetts with $42°18'11.6''$ and $71°53'50.9''$ as its coordinates. The wind speed is measured every 10 minutes with cup anemometers installed at a height of 78 m above the ground. The wind speed data ranges from January 1, 2011 to January 7, 2011 22:30 hrs.

- Sotavento, Spain: The wind farm is located in Sotavento, Galicia, Spain with latitude $43°21'35.9''$ and longitude $-7°52'47.9''$. The dataset chosen is for the month of October 2017 where the wind speed is measured hourly.

- Blandford, MA: Blandford is situated at 42.223° N and 72.968° E with wind speed recorded at a height of 60 m above the ground with a cup anemometer at every 10 minute. The wind speed data ranges from January 1, 2011 to January 7, 2011 22:30 hrs.

- Bishop & Clerks, MA: Wind monitoring site is located at 41.574° N and 70.249° E with anemometer installed at height of 15 m above ground. The data ranges from January 1 2011 to January 7, 2011 22:30 hrs and is recorded every 10 minutes.

- Beresford, South Dakota: The wind site is located at 43.088° N and 96.786° E and ranges from March 1, 2006 22:20 hrs to March 8, 2006 20:50 hrs. Wind speed is recorded every 10 minutes at a height of 20 m.

- AGL Macarthur, Victoria, Australia: Macarthur wind farm is located at 38.049° S and 142.190° E with 420 MW installed capacity featuring 140 V112-3.0 Vestas wind turbines. The hourly wind speed data is taken from February 26, 2019 00:00 hrs to March 5, 2019 23:00 hrs [30].

- Muppandal, Kanyakumari, India: Located in Kanyakumari, Tamil Nadu, it has a capacity of 1500 MW. Wind speed data for the month of January 2019 is chosen. The samples are recorded at 10 minute intervals [31].

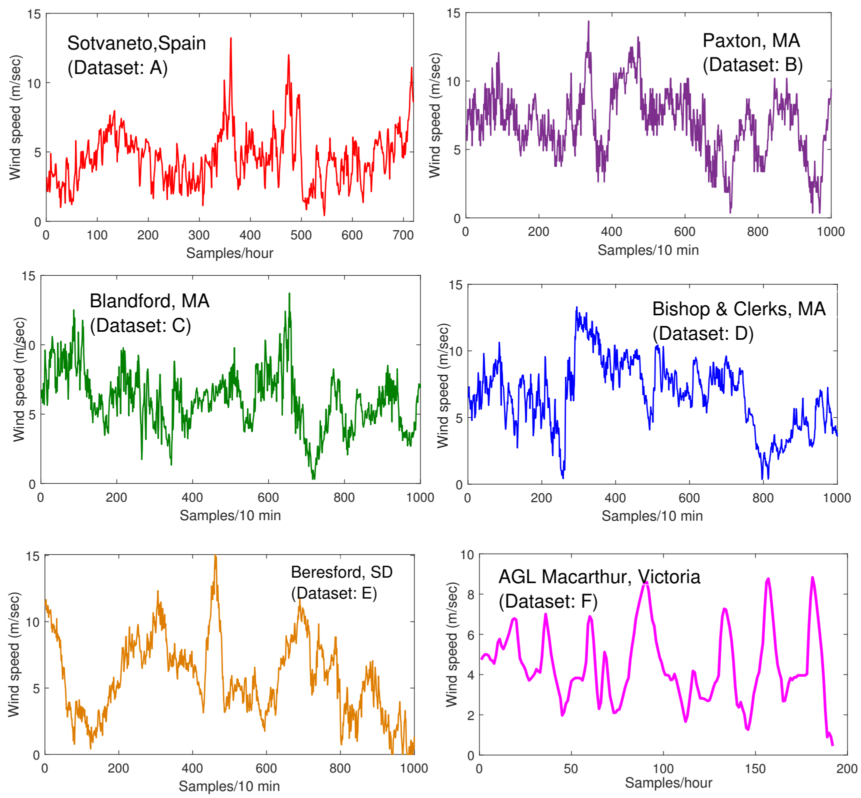

Figure 2: Wind speed for datasets A through F

Table 1: Descriptive statistics for wind speed at various wind farm sites

Wind farm (Dataset)	Max (m/sec)	Min (m/sec)	Mean (m/sec)	Std Dev
Sotavento, Spain (A)	13.23	0.41	4.6072	1.9395
Paxton, MA (B)	14.39	0.35	6.9209	2.3734
Blandford, MA (C)	13.73	0.30	6.0553	2.1242
Bishop & Clerks, MA (D)	13.31	0.36	6.7065	2.5923
Beresford, SD (E)	15.06	0.58	5.4729	2.9828
AGL Macarthur (F)	9.05	1.92	6.2926	1.5035
Muppandal, Kanyakumari (G)	8.48	0.71	4.8878	1.4641

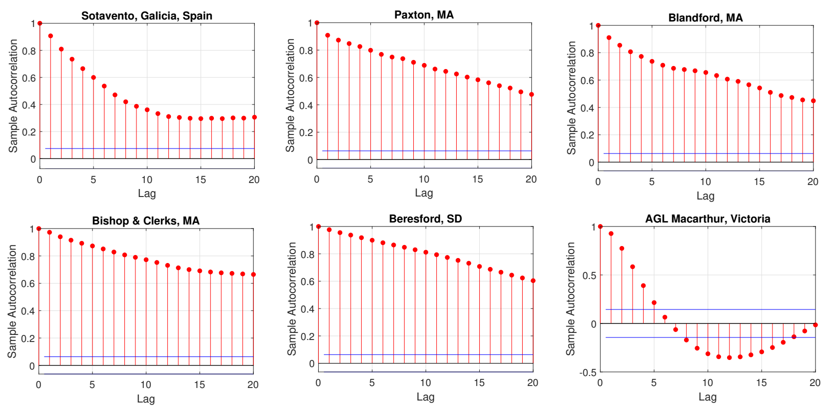

Figure 3: Periodicity of wind speed time-series for Sotavento, Paxton, Blandford, Bishop & Clerks, Beresford and AGL Macarthur

Figure 3 illustrates the periodicity of the wind speed time series for all datasets. The auto-correlation plots depict the correlation of time series samples with itself at different lag order. For datasets A to E, we find that lag order of 1 and 2 are significantly dominant, indicating strong correlation. However, the auto-correlation for dataset F is negative for lag order 7.

3.2. Forecasting performance during Wind power ramp events

Wind power intermittency owing to sudden wind speed variations, is a critical event in case of grid connected power plants, leading to severe consequences like low system reliability, high reserve capacity and high operational costs. A wind power ramp event is defined as rate of change in wind power generated by a wind turbine or wind farm over a short period of time exceeding a predefined threshold value (normally 50%) [32]. According to [33], power ramp event is said to occur if the change in power signal $|P(t + \Delta t) - P(t)|$ is greater than a said threshold ΔP_{ramp}. Intermittent nature wind speed leads to installation of energy storage systems in the wind farms to tackle peak demand scenarios, thus constant charging and discharging of batteries during multiple ramp events degrades their life [34]. In order to analyze the power ramp up or down events, setting the threshold power is an important task. For a given wind turbine, let us say the ramp threshold power is $r\%$ of the nominal wind power. Then we can define two ramp thresholds, that is,

$$\Delta P_{ramp} \quad = \quad \begin{cases} +r\% \quad of \quad P_{nominal} & = P_{th}^{u}, \\ -r\% \quad of \quad P_{nominal} & = P_{th}^{l}, \end{cases} \tag{38}$$

where P_{th}^{u} and P_{th}^{l} are the upper and lower ramp thresholds respectively depicting ramp-up and ramp-down events in a given short period of time. We now compare different forecasting methods during power ramp events and analyze the critical conditions prevailing during such events. The forecasting methods implemented are hybrid models based on wavelet transform and ε-SVR, LS-SVR, TSVR and ε-TSVR.

3.3. Ramp event error analysis for ε-SVR and LS-SVR

Consider a power ramp-up event at points g and h as shown in Figure 4. Let the wind power at point g be P_g and at point h be P_h, and the difference $\Delta P_{gh} = P_h - P_g$ denotes change in wind power over a short time interval ΔT.

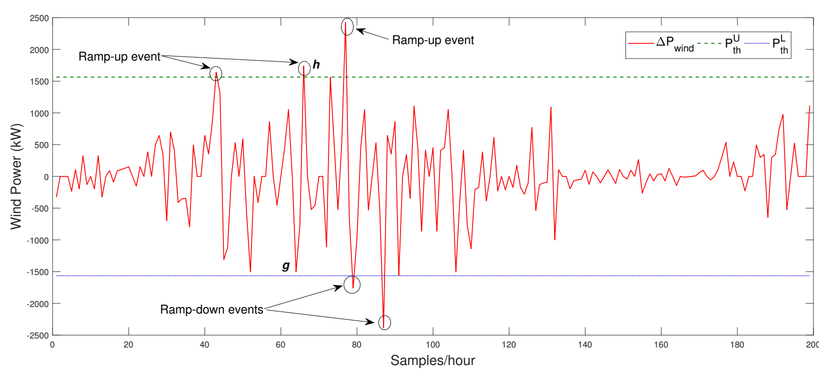

Figure 4: Schematic representation of wind power ramp events

According to ε-SVR and LS-SVR, the forecasted values are

$$\hat{P}_{g1} = (\alpha_g - \alpha_g^*)k(x, x_g) + b, \tag{39}$$

$$\hat{P}_{h1} = (\alpha_h - \alpha_h^*)k(x, x_h) + b, \tag{40}$$

$$\hat{P}_{g2} = \eta_g k(x, x_g) + b_1, \tag{41}$$

$$\hat{P}_{h2} = \eta_h k(x, x_h) + b_1, \tag{42}$$

at ramp points g and h such that α_g, η_g are the Lagrangian multipliers, $\hat{P}_{g1}, \hat{P}_{h1}$ and $\hat{P}_{g2}, \hat{P}_{h2}$ are the predicted values based on ε-SVR and LS-SVR models respectively. Error in predicted value $\hat{P}_{g1}$ and actual value P_g is given as

$$e_{SVR} = \hat{P}_{h1} - P_h - \hat{P}_{g1} - P_g, \tag{43}$$

$$e_{LS-SVR} = \hat{P}_{h2} - P_h - \hat{P}_{g2} - P_g, \tag{44}$$

where e_{SVR} and e_{LS-SVR} are the errors based on ε-SVR and LS-SVR models respectively. If LS-SVR outperforms ε-SVR, we have $e_{LS-SVR} < e_{SVR}$, that is,

$$\hat{P}_{h1} - P_h - \hat{P}_{g1} + P_g > \hat{P}_{h2} - P_h - \hat{P}_{g2} + P_g. \tag{45}$$

Let us define $\beta_h = \alpha_h - \alpha_h^*$ and $\beta_g = \alpha_g - \alpha_g^*$, and by simplifying (45), we get

$$\beta_h k(x, x_h) + b - \beta_g k(x, x_g) - b > \eta_h k(x, x_h) + b_1 - \eta_g k(x, x_g) - b_1, \tag{46}$$

since the kernel matrix elements $k(x, x_i)$ are equal for ε-SVR and LS-SVR, the equation can be further simplified as,

$$k(x, x_h)\Big(\beta_h - \eta_h\Big) - k(x, x_g)\Big(\beta_g - \eta_g\Big) > 0, \tag{47}$$

Thus if condition in (47) is satisfied, LS-SVR will outperform ε-SVR during ramp events.

3.4. Ramp event error analysis for TSVR and ε-TSVR

Similarly TSVR and ε-TSVR can be compared based on same approach. Let e_{TSVR} and $e_{\varepsilon-TSVR}$ denote the errors in the wind ramp power between points g and h based on TSVR and ε-TSVR respectively and are given as

$$\hat{P}_{g3} = \frac{1}{2}(w_1 + w_2)k(x, x_g) + \frac{1}{2}(b_1 + b_2), \tag{48}$$

$$\hat{P}_{h3} = \frac{1}{2}(w_1 + w_2)k(x, x_h) + \frac{1}{2}(b_1 + b_2), \tag{49}$$

$$\hat{P}_{g4} = \frac{1}{2}(u_1 + u_2)k(x, x_g) + \frac{1}{2}(b_3 + b_4), \tag{50}$$

$$\hat{P}_{h4} = \frac{1}{2}(u_1 + u_2)k(x, x_h) + \frac{1}{2}(b_3 + b_4), \tag{51}$$

where $\hat{P}_{g3}$,$\hat{P}_{h3}$ and $\hat{P}_{g4}$, $\hat{P}_{h4}$ are the predicted values of wind power using TSVR and ε-TSVR respectively. The forecasted ramp power $\hat{\Delta} P_{gh}$ is then compared for two methods. The error in ΔP_{gh} for TSVR and ε-TSVR is given as

$$e_{TSVR} = \hat{P}_{h3} - P_h - \hat{P}_{g3} + P_g, \tag{52}$$

$$e_{\varepsilon-TSVR} = \hat{P}_{h4} - P_h - \hat{P}_{g4} + P_g, \tag{53}$$

Comparing the two ramp power errors, if, $e_{TSVR} > e_{\varepsilon-TSVR}$ we get,

$$\hat{P}_{h3} - P_h - \hat{P}_{g3} + P_g \quad > \quad \hat{P}_{h4} - P_h - \hat{P}_{g4} + P_g. \tag{54}$$

Let us define $\frac{1}{2}(w_1 + w_2) = \hat{w}$ and $\frac{1}{2}(u_1 + u_2) = \hat{u}$ and simplifying (54), we get

$$\hat{w}\Big(k(x, x_h) - k(x, x_g)\Big) \quad > \quad \hat{u}\Big(k(x, x_h) - k(x, x_g)\Big). \tag{55}$$

As long as the condition (55) is satisfied, ε-TSVR outperforms TSVR during ramp events between points g and h. The next section discusses the forecasting errors during wind power ramp events.

4. Results and Discussion

A hybrid model is built on wavelet decomposition technique and machine intelligent SVR model where 80% of data is used for training and the rest for testing. TSVR and ε-TSVR forecasting models are evaluated via-a-vis ε-SVR and LS-SVR models. For ε-TSVR, we assume the regularization factor C_1=C_2 and C_3=C_4. Similarly for TSVR, we select C_1=C_2. The kernel function used for building the regression models is Radial basis function (RBF), with bandwidth σ, $k(x, x_i) = e^{\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right)}$. The hyperparameters C_1, C_2, C_3 and C_4 along with RBF bandwidth (σ) are chosen from a set 2^i, where $i = -9, -8..., 9, 10$. Optimal parameters can be tuned manually or by grid search algorithm. Datasets related to four wind farm sites labeled A, B, C and D are chosen to test the performance of hybrid forecasting model. Dataset A consists of 720 samples out of which 80% (576) are used for training process and 20% (144) are used for testing.

Similarly for datasets B, C and D, 800 samples are used for training and 200 for testing.

Table 2 depicts various performance indices for wavelet based hybrid SVR models. For dataset A, ε-TSVR and TSVR outperformed ε-SVR by 41.95% and 84.25% respectively in terms of RMSE. Similarly for dataset B, C and D, ε-TSVR outperforms ε-SVR by 3.537%, 63.03% and 59.60% respectively in terms of RMSE. Among all the models, ε-TSVR and TSVR outperform LS-SVR and ε-SVR quantitatively in terms of RMSE and MAE for all the datasets.

Further, in terms of speed of computation, LS-SVR spends minimum processor time owing to its smaller sized optimization problem. ε-TSVR takes less time than classical ε-SVR and TSVR, and among the four datasets, datasets B, C and D take more or less same computation time for the respective models. The ratios SSR/SST and SSE/SST give an estimate of goodness of fit among all the regression models. SSR/SST ratio value greater than 1 implies over-fitting during training process which is not desirable during testing phase. Among all the regressors, TSVR obtains the optimal SSR/SST and SSE/SST ratio. Further, forecasting is assessed from statistical point of view by determining the index of agreement (IOA), Theil's U1 and U2 statistic for all the models. From Table 2, we observe that ε-TSVR and TSVR models outperform ε-SVR, LS-SVR, and persistence models in terms of Theil's U1 and U2 statistic, thereby indicating the forecasting accuracy of the two models is superior to the rest.

Figure 5 shows the forecasting results of the four variants of SVR for four wind farm sites. In order to further validate the proposed hybrid forecasting model, the forecast accuracy of ε-TSVR and TSVR is tested using Diebold-Mariano (DM) test. The DM statistic test assumes a null hypothesis wherein two forecasting models have similar accuracy [35]. We compare the DM statistic of TSVR (Test 1) and ε-TSVR (Test 2) against classical ε-SVR model. The test is carried out at 1% significance level for datasets A, B, C, and D and results are highlighted in Table 3. Thus, by rejecting the null hypothesis from the DM test, we observe that both, TSVR and ε-TSVR models have significant forecast

Table 2: Performance metrics of wavelet- ε-SVR, LS-SVR, TSVR and ε-TSVR

Dataset	Model	RMSE (m/sec)	MAE	SSR/SST	SSE/SST	IOA	U1	U2	CPU time (secs)
	ε-SVR	0.1423	12.254	0.9434	0.064	1.0000	0.0233	1.1281	5.7964
A	LS-SVR	0.1097	8.4668	0.9417	0.0038	1.0000	0.0028	0.9665	0.5924
	TSVR	0.0224	1.4877	0.9821	0.0008	1.0000	0.0011	0.9496	3.1864
	ε-TSVR	0.0823	7.6699	0.9694	0.0021	1.0000	0.00116	0.0150	2.4421
	Persistence	0.7241	71.41	1.0192	0.2925	0.9991	0.0913	1.1331	0.3024
	ε-SVR	0.0424	6.8172	1.0055	0.0003	1.0000	0.0138	0.0645	11.244
B	LS-SVR	0.0329	4.6550	0.9862	0.0001	1.0000	0.00340	0.0466	0.8165
	TSVR	0.0148	2.0444	0.9965	0.00003	1.0000	0.0008	0.0072	6.2343
	ε-TSVR	0.0409	6.8764	0.9884	0.0003	1.0000	0.0021	0.0130	3.6897
	Persistence	0.7825	66.31	1.0197	0.1437	0.9998	0.0731	1.0000	0.2560
	ε-SVR	0.1791	24.211	0.9777	0.0151	0.9980	0.0453	1.3804	12.3536
C	LS-SVR	0.0733	8.9438	0.9929	0.0027	1.0000	0.0021	0.0231	0.8962
	TSVR	0.0100	1.2674	0.9976	0.0004	1.0000	0.0003	0.0046	6.2235
	ε-TSVR	0.0662	8.7964	1.0127	0.0022	1.0000	0.0040	0.0635	3.7219
	Persistence	0.6939	54.95	1.0501	0.2934	0.9998	0.0734	1.0769	0.0818
	ε-SVR	0.1901	28.8003	1.0538	0.0174	1.0000	0.0230	0.8751	11.041
D	LS-SVR	0.1427	20.5274	0.9798	0.0098	1.0000	0.0129	0.6855	1.2148
	TSVR	0.0740	9.4829	0.9816	0.0026	1.0000	0.0050	0.0746	6.7960
	ε-TSVR	0.0768	12.1187	0.9507	0.0028	1.0000	0.0047	0.0254	9.0181
	Persistence	0.4366	34.71	1.0535	0.1156	0.9998	0.0587	0.6667	0.0925

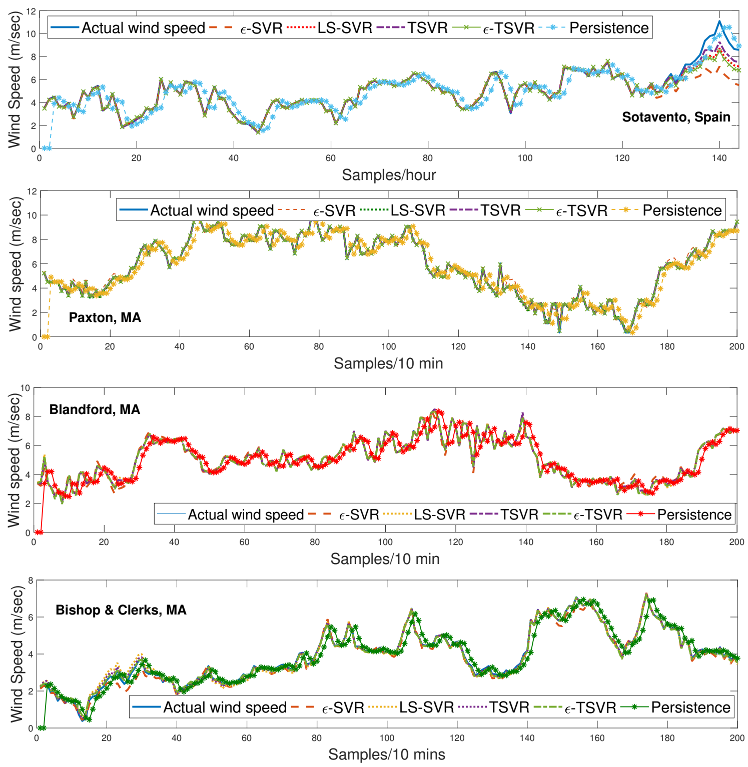

Figure 5: Forecasting results for ε-SVR, LSSVR, TSVR, ε-TSVR and Persistence model

superiority over ε-SVR model, proving the robustness of the hybrid SVR model and its variants over the persistence model.

In order to further validate the effect of larger dataset on our hybrid model, we select wind speed data from Blandford, MA (Dataset:C) and Muppandal, Kanyakumari (Dataset:G). The training set comprises of 4000 and 3000 samples for dataset C and G respectively and testing set consists 1000 samples. The forecasting performance is depicted in Table 4 and is illustrated in Figure 6. From Table 4, we observe that, ε-TSVR and TSVR perform significantly better than ε-SVR and LS-SVR in terms of RMSE and MAE thus indicating their superiority. Further, in terms of computation speed, ε-TSVR saves 93% and 81% of time compared to ε-SVR for datasets C and G respectively.

Table 3: Diebold-Mariano test for datasets

Dataset	Diebold-Mariano Statistic	
	Test 1	Test 2
A	10.7291	9.7084
B	7.6321	7.4852
C	5.2699	5.2398
D	6.9036	6.6344

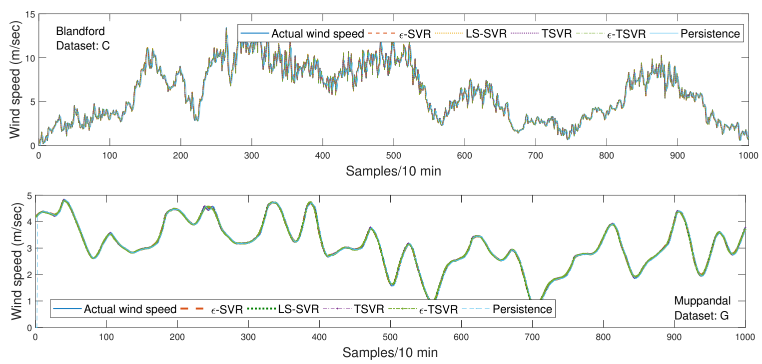

Figure 6: Forecasting results for larger datasets

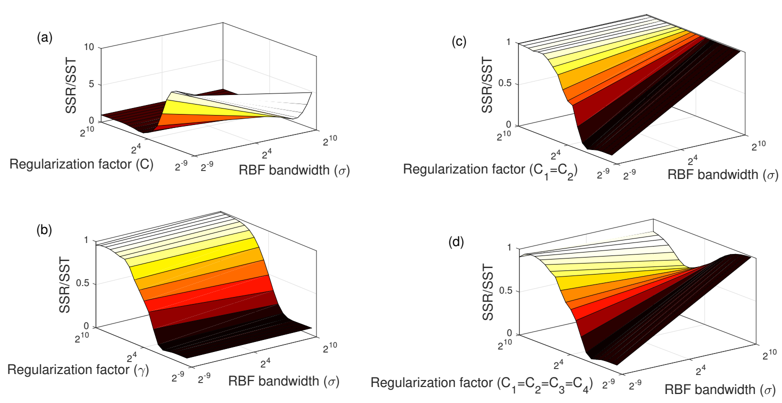

Figure 7: Variation of SSR/SST with RBF bandwidth (σ) and Regularization factor (C) for (a) ε-SVR (b) LS-SVR (c) TSVR and (d) ε-TSVR

Table 4: Performance metrics for a larger dataset

Dataset	Model	RMSE	MAE	SSR/SST	SSE/SST
	ε-SVR	0.0416	33.7801	1.0079	0.0001
	LS-SVR	0.0194	12.7811	1.0013	0.00003
	TSVR	0.0036	2.2739	0.99994	0.00001
	ε-TSVR	0.0127	8.9721	0.9985	0.00001
	Persistance	0.8553	640.65	0.9826	0.0755
C	Model	IOA	U1	U2	CPU time
	ε-SVR	1.0000	0.0031	0.1440	911.4211
	LS-SVR	1.0000	0.0014	0.0489	13.3221
	TSVR	1.0000	0.00002	0.0084	355.011
	ε-TSVR	1.0000	0.00094	0.1371	61.0762
	Persistence	1.0000	0.0633	0.7813	0.3024
Dataset	Model	RMSE	MAE	SSR/SST	SSE/SST
	ε-SVR	0.0283	20.3760	0.9823	0.0008
	LS-SVR	0.0170	12.6992	0.9851	0.00003
	TSVR	0.0143	9.6658	0.9871	0.00002
	ε-TSVR	0.0157	10.9721	0.9910	0.00001
	Persistance	0.2053	73.4500	1.0213	0.0523
G	Model	IOA	U1	U2	CPU time
	ε-SVR	1.0000	0.0041	0.6416	347.304
	LS-SVR	1.0000	0.0026	0.5163	5.9329
	TSVR	1.0000	0.0022	0.00485	70.2648
	ε-TSVR	1.0000	0.0165	0.1241	63.3087
	Persistence	1.0000	0.0318	0.6053	0.0131

Figure 7 shows the variation of SSR/SST ratio with RBF bandwidth (σ) and regularization factor (C) for four different variants of SVR. The ratio SSR/SST estimates whether the training data has been over trained or not. ε-TSVR and TSVR show better variation of SSR/SST ratio for testing samples than classical ε-SVR and LS-SVR with σ (keeping C constant) and C (keeping σ constant).

As we increase σ (from 2^{-9} to 2^{10}), the value of SSR/SST increases from 0 to 1 for ε-TSVR and TSVR and remains constant for LS-SVR. However, for ε-SVR, the SSR/SST value first decreases and then increases further after $\sigma = 2^4$.

In our study, we choose five wind farms namely, Sotavento (Spain), Paxton (MA) and Blandford (MA), Beresford (South Dakota) [36] and AGL Macarthur wind farm, Victoria, Australia to analyze the wind power ramp events. The threshold ramp power is chosen as 15% of nominal power (P_{nom}). Wind turbines (Vestas V112) from the Danish manufacturer Vestas with rated speed 12 m/sec are selected to study the wind power ramp event. Two ramp events, that is, power ramp-up and power ramp down events are studied. The nominal wind power of the given wind turbine is 3.6496 MW. The threshold limit for ramp power events is chosen as 15% of the nominal power.

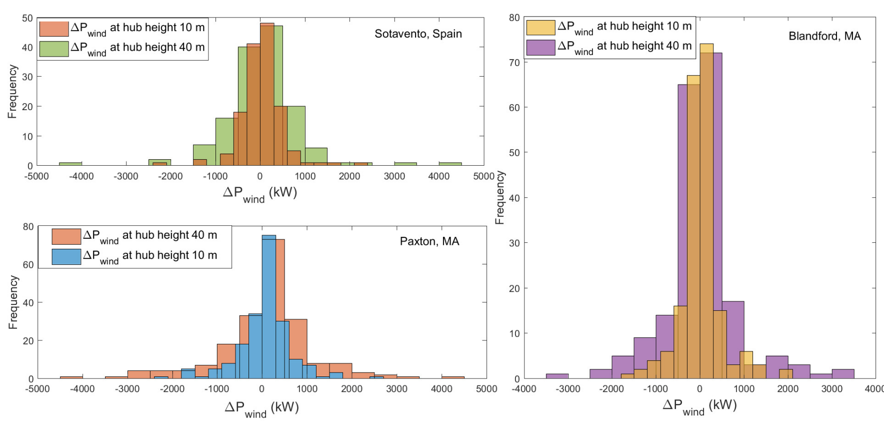

Figure 8: Frequency distribution of change in wind power with hub heights

Figure 8 shows the frequency distribution of change in wind power (ΔP_{wind}) in successive dispatch windows for different hub heights. The wind speed data available at hub height of 10 meters is transformed at a hub height of 40 m using the wind profile power law [37] given as

$$\frac{u_h}{u_r} = \left(\frac{z_h}{z_r}\right)^{\alpha}, \tag{56}$$

where u_h, and u_r are the wind speeds (in m/sec) at desired hub height and reference hub height, z_h, z_r are the hub heights (in meters) at desired level and reference level respectively and ($\alpha = 1/7$) is an empirically calculated constant dependent on atmospheric conditions [38].

Table 5 shows the Absolute error (AE) values computed for wind power ramp-up and ramp-down events for different wind farm sites. During ramp-up events for all the wind sites, ε-TSVR performs better than TSVR, LS-SVR and classical ε-SVR.

Table 5: Performance metric (AE) during wind power ramp events

Wind Farm	Model	Wind power ramp event	
		Ramp-up	Ramp-down
Sotavento, Spain	ε-SVR	0.7245	0.7626
	LS-SVR	0.4587	0.5303
	TSVR	0.3600	0.4191
	ε-TSVR	0.1414	0.2019
Paxton, MA	ε-SVR	0.0454	0.0174
	LS-SVR	0.0347	0.0133
	TSVR	0.0118	0.0055
	ε-TSVR	0.0018	0.0104
Blandford, MA	ε-SVR	0.0574	0.8058
	LS-SVR	0.0350	0.1650
	TSVR	0.0237	0.1454
	ε-TSVR	0.0161	0.1211
Beresford, South Dakota	ε-SVR	0.0657	0.1074
	LS-SVR	0.0500	0.1058
	TSVR	0.0074	0.0025
	ε-TSVR	0.0067	0.0021
AGL Macarthur, Victoria, Australia	ε-SVR	0.2540	0.5497
	LS-SVR	0.1221	0.3624
	TSVR	0.0669	0.1877
	ε-TSVR	0.0570	0.2821

From Figure 8, we see that the probability of wind power ramp event increases if wind speed is recorded at a hub height above the ground. The number of wind power ramp events for five wind farm sites are illustrated in Figure 9.

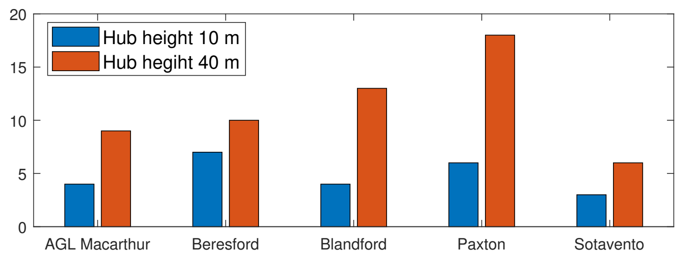

Figure 9: Frequency of number of wind power ramp events with different hub heights

5. Conclusion

In this paper, we study hybrid machine intelligent SVR models for short-term wind forecasting built on wavelet transform decomposition technique. A fourth order daubechies (db4) wavelet filter is chosen to carry out the wind speed time-series decomposition for four different wind farm sites. Among these regressors, the hybrid model based on TSVR and ε-TSVR proves a better short-term forecast choice based on the performance indices for the four datasets. Computationally, LS-SVR takes the minimum time on CPU, and ε-TSVR takes less computation time than ε-SVR owing to its smaller sized optimization problem. The wind speed forecasting accuracy for all the hybrid models can be further improved by optimally selecting the SVR hyperparameters: RBF bandwidth and regularization constants. Further, the wind power ramp events are studied and under certain conditions ε-SVR and LS-SVR, and TSVR and ε-TSVR forecast errors were compared. Among the regressors, ε-TSVR outperformed TSVR, LS-SVR and ε-SVR in terms of absolute error. The ramp events are analyzed for different hub heights and the number of recorded ramp events increased significantly with height. Thus, this machine intelligent hybrid methodology improves the forecasting performance of wind farms with uncertain wind conditions like ramp events.

References

[1] Global wind energy council report. http://files.gwec.net/files/GWR2017.pdf; 2017. (Accessed on 03/07/2019).

[2] Roy S, Das R, Saha UK. An inverse method for optimization of geometric parameters of a savonius-style wind turbine. Energy Conversion and Management 2018;155:116–27.

[3] Gupta R, Biswas A. Computational fluid dynamics analysis of a twisted three-bladed h-darrieus rotor. Journal of Renewable and Sustainable Energy 2010;2(4):043111.

[4] Mohandes M, Halawani T, Rehman S, Hussain AA. Support vector machines for wind speed prediction. Renewable Energy 2004;29(6):939–47.

[5] Zhou J, Shi J, Li G. Fine tuning support vector machines for short-term wind speed forecasting. Energy Conversion and Management 2011;52(4):1990–8.

[6] Cadenas E, Rivera W. Wind speed forecasting in three different regions of mexico, using a hybrid ARIMA–ANN model. Renewable Energy 2010;35(12):2732–8.

[7] Liu D, Niu D, Wang H, Fan L. Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. Renewable Energy 2014;62:592–7.

[8] Zhang C, Wei H, Zhao X, Liu T, Zhang K. A gaussian process regression based hybrid approach for short-term wind speed prediction. Energy Conversion and Management 2016;126:1084–92.

[9] Mi X, Liu H, Li Y. Wind speed forecasting method using wavelet, extreme learning machine and outlier correction algorithm. Energy Conversion and Management 2017;151:709–22.

[10] Li H, Wang J, Lu H, Guo Z. Research and application of a combined model based on variable weight for short term wind speed forecasting. Renewable Energy 2018;116:669–84.

[11] Jiang P, Wang Y, Wang J. Short-term wind speed forecasting using a hybrid model. Energy 2017;119:561–77.

[12] Azimi R, Ghofrani M, Ghayekhloo M. A hybrid wind power forecasting model based on data mining and wavelets analysis. Energy Conversion and Management 2016;127:208–25.

[13] Jiang Y, Huang G, Peng X, Li Y, Yang Q. A novel wind speed prediction method: Hybrid of correlation-aided DWT, LSSVM and GARCH. Journal of Wind Engineering and Industrial Aerodynamics 2018;174:28–38.

[14] Liu H, Duan Z, Han F, Li Y. Big multi-step wind speed forecasting model based on secondary decomposition, ensemble method and error correction algorithm. Energy Conversion and Management 2018;156:525–41.

[15] Tian C, Hao Y, Hu J. A novel wind speed forecasting system based on hybrid data preprocessing and multi-objective optimization. Applied Energy 2018;231:301–19.

[16] Wang J, Yang W, Du P, Li Y. Research and application of a hybrid forecasting framework based on multi-objective optimization for electrical power system. Energy 2018;148:59–78.

[17] Du P, Wang J, Yang W, Niu T. Multi-step ahead forecasting in electrical power system using a hybrid forecasting system. Renewable Energy 2018;122:533–50.

[18] Debnath BK, Das R. Prediction of performance coefficients of a three-bucket savonius rotor using artificial neural network. Journal of Renewable and Sustainable Energy 2010;2(4):043107.

[19] Silva M, Oleskovicz M, Coury D. A hybrid fault locator for three-terminal lines based on wavelet transforms. Electric Power Systems Research 2008;78(11):1980–8.

[20] Vapnik V, Lerner A. Pattern recognition using generalized portrait method. Automation and Remote Control 1963;24.

[21] Cortes C, Vapnik V. Machine Learning 1995;20(3):273–97.

[22] Joachims T. Text categorization with support vector machines: Learning with many relevant features. In: Machine Learning: ECML-98. Springer Berlin Heidelberg; 1998, p. 137–42.

[23] Schölkopf B, Smola A, Müller KR. Kernel principal component analysis. In: Lecture Notes in Computer Science. Springer Berlin Heidelberg; 1997, p. 583–8.

[24] Shirzad A, Tabesh M, Farmani R. A comparison between performance of support vector regression and artificial neural network in prediction of pipe burst rate in water distribution networks. KSCE Journal of Civil Engineering 2014;18(4):941–8.

[25] Vapnik VN. The Nature of Statistical Learning Theory. Springer New York; 2000.

[26] Suykens J, Vandewalle J. Neural Processing Letters 1999;9(3):293–300.

[27] Peng X. TSVR: An efficient twin support vector machine for regression. Neural Networks 2010;23(3):365–72.

[28] Shao YH, Zhang CH, Yang ZM, Jing L, Deng NY. An ε-twin support vector machine for regression. Neural Computing and Applications 2012;23(1):175–85.

[29] Hao Y, Tian C. The study and application of a novel hybrid system for air quality early-warning. Applied Soft Computing 2019;74:729–46.

[30] Weather history download macarthur - meteoblue. https://www.meteoblue.com/en/weather/archive/export/macarthur_australia_2159266; 2019. (Accessed on 03/09/2019).

[31] Merra - www.soda-pro.com. http://www.soda-pro.com/web-services/meteo-data/merra; 2019. (Accessed on 03/11/2019).

[32] Greaves B, Collins J, Parkes J, Tindal A. Temporal forecast uncertainty for ramp events. Wind Engineering 2009;33(4):309–19.

[33] Kamath C. Understanding wind ramp events through analysis of historical data. In: IEEE PES T D 2010. 2010, p. 1–6.

[34] Ren G, Liu J, Wan J, Guo Y, Yu D. Overview of wind power intermittency: Impacts, measurements, and mitigation solutions. Applied Energy 2017;204:47–65.

[35] Diebold FX. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold–mariano tests. Journal of Business & Economic Statistics 2015;33(1):1.

[36] Anemometer data (wind speed, direction) for beresford, south dakota (2006) - datasets - openei datasets. https://openei.org; 2019. (Accessed on 03/01/2019).

[37] Irwin JS. A theoretical variation of the wind profile power-law exponent as a function of surface roughness and stability. Atmospheric Environment (1967) 1979;13(1):191–4.

[38] Akpinar EK, Akpinar S. An assessment of wind turbine characteristics and wind energy characteristics for electricity production. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects 2006;28(10):941–53.