ARCHIVES OF ELECTRICAL ENGINEERING VOL. 70(4), pp. 801 –817 (2021)

DOI 10.24425/aee.2021.138262

Short-term wind power combined prediction based on EWT-SMMKL methods

Jun Li, Liancai Ma

Lanzhou Jiaotong University
Lanzhou, Gansu 730070, China
e-mail: lijun691201@mail.lzjtu.cn, 568546966@qq.com

(Received: 08.10.2020, revised: 02.05.2021)

Abstract: Since wind power generation has strong randomness and is difficult to predict, a class of combined prediction methods based on empirical wavelet transform (EWT) and soft margin multiple kernel learning (SMMKL) is proposed in this paper. As a new approach to build adaptive wavelets, the main idea is to extract the different modes of signals by designing an appropriate wavelet filter bank. The SMMKL method effectively avoids the disadvantage of the hard margin MKL method of selecting only a few base kernels and discarding other useful basis kernels when solving for the objective function. Firstly, the EWT method is used to decompose the time series data. Secondly, different SMMKL forecasting models are constructed for the sub-sequences formed by each mode component signal. The training processes of the forecasting model are respectively implemented by two different methods, i.e., the hinge loss soft margin MKL and the square hinge loss soft margin MKL. Simultaneously, the ultimate forecasting results can be obtained by the superposition of the corresponding forecasting model. In order to verify the effectiveness of the proposed method, it was applied to an actual wind speed data set from National Renewable Energy Laboratory (NREL) for short-term wind power single-step or multi-step time series indirectly forecasting. Compared with a radial basic function (RBF) kernel-based support vector machine (SVM), using SimpleMKL under the same condition, the experimental results show that the proposed EWT-SMMKL methods based on two different algorithms have higher forecasting accuracy, and the combined models show effectiveness.

Key words: combined model, empirical wavelet transform, prediction, soft margin multiple kernel learning, wind power

1. Introduction

In recent years, as one of the most promising renewable energy technologies, wind power has received increasing attention from energy enterprise. However, due to the randomness and intermittent nature of wind energy resources, wind power can be affected by many factors, and

the connection of large wind farms to the grid can have a significant impact on the planning and operation of the grid. Therefore, accurate short-term wind power prediction is one of the effective means to determine a reasonable dispatch plan and ensure the safe and economic operation of the power grid [1–3].

Some linear regression-based forecasting methods, such as physical methods and statistical methods [4], have difficulties to describe the inherent nonlinear characteristics of wind power time series due to the randomness and intermittence of wind energy resources. Currently, a class of intelligent forecasting methods including artificial neural networks [5] and support vector machine (SVM) [6] are widely used in wind power prediction. However, confined by inherent structure, these methods often suffer from disadvantages such as being easy to trap in local optimum, over-fitting and showing difficulty in determining the parameters of the kernel function as well as they cannot provide better prediction results.

In recent years, some advanced statistical methods have been applied to wind power forecasting. In [7], a hybrid short-term wind power forecasting model combining improved variational mode decomposition and sample entropy (IVMD-SE) data preprocessing and correntropy long short-term memory (LSTM) that is insensitive to outliers is proposed. Experimental results show that the method can significantly improve the prediction accuracy and robustness. In the research of Sinvaldo *et al*. [8], the authors studied the influence of the instability of wind energy resources on the prediction results, combining the LSTM network, and two signal decomposition strategies. Variational mode decomposition (VMD) and singular spectral analysis (SSA) are further proposed, to obtain more reliable prediction results. Ramon *et al*. [9] introduced a new decomposition ensemble learning method based on complete ensemble empirical mode decomposition (CEEMD) and stacking-ensemble learning (STACK), and the multi-step prediction strategy is used to predict the wind energy of the wind farm in Brazil. For wind power prediction, Yldz *et al*. [10] designed a model combining VMD and an improved residual-based deep convolutional neural network (CNN), the experimental results verify the effectiveness of the prediction model. In addition, as can be observed from the methods presented in [11–13], compared with a single prediction method, satisfactory prediction results can be obtained by using different decomposition methods for wind power with non-linear and non-stationary characteristics. Therefore, this paper attempts to reduce the non-stationarity of wind speed data by pre-processing them using empirical wavelet transform (EWT) [14]. EWT is an adaptive signal decomposition method based on wavelet analysis proposed by Gilles, which has a detailed mathematical theory basis and can avoid the phenomenon of mode mixing and false modes in an empirical mode decomposition (EMD) [15] method.

Kernel learning methods are effective ways to solve the problem of nonlinear pattern analysis problems. The multiple kernel learning (MKL) [16,17] model offers greater flexibility in mapping different features of heterogeneous data by the most appropriate kernel function, which ultimately leads to a more accurate representation of the data in a new combinatorial space, thus improving the classification accuracy or prediction accuracy of the sample data [18]. However, in practical application, the traditional MKL method cannot achieve good prediction results because it only selects the smallest base kernel and discards other useful ones. These methods can be regarded as hard margin MKL, and SimpleMKL [19] is a kind of hard margin MKL. The SimpleMKL obtains sparse kernel combination through additional weight constraint and weighted L2 norm regularization. Soft margin multiple kernel learning (SMMKL) [20] can be regarded as an

extension of a soft margin SVM [21]. This method effectively avoids the drawback of the hard margin MKL method (e.g. SimpleMKL) in solving the objective function by selecting only a few basis kernels and discarding other useful ones, and improves the prediction accuracy. At present, soft margin MKL method has been successfully applied to classification [20].

Based on the above analysis, a new EWT-SMMKL method for short-term wind power prediction is proposed in this paper. EWT is used to decompose the original wind speed series with non-stationary characteristics, SMMKL is used to predict the wind speed, and the final predicted value is obtained through the wind speed and wind power conversion curve. The main contributions of this paper are summarized as follows.

1. A new prediction model (EWT-SMMKL) is proposed. Firstly, considering the strong nonlinear and non-stationary characteristics of wind speed or wind power data, the EWT is used to preprocess the original time series data to improve the predictability.
2. Secondly, taking into account the difficulties of selecting the parameters of the kernel functions in SVM as well as only selecting the minimum number of base kernels in the traditional MKL, the SMMKL method is used to establish the prediction model to further improve prediction accuracy.

2. Methodology

2.1. EWT method

EWT is a type of adaptive wavelet decomposition method which can extract the mode signal components of the compactly supported Fourier spectrum by establishing appropriate orthogonal wavelet filter banks and performing adaptive segmentation on the Fourier spectrum of the original signal [14]. The purpose of EWT is to decompose the signal $f(t)$ into the sum of the $N+1$ intrinsic mode function (IMF) $f_k(t)$.

The signal to be analyzed, $f(t)$ is transformed by the Fourier transform, which is marked as $\hat{f}(\omega)$. It is assumed that the Fourier support interval $[0, \pi]$ is segmented into L continuous segments, the boundaries ω_l of each segment as the center between two consecutive maxima, and each segment is denoted $\Lambda_l = [\omega_{l-1}, \omega_l]$ ($\omega_0 = 0$, $\omega_L = \pi$). A transition phase at ω_l points is defined, and its width is $T_l = 2\tau_l$. $\tau_l = \gamma\omega_l$ ($0 < \gamma < 1$), where τ_l is transition zone variables and γ is the coefficient.

The empirical wavelet is defined as a band-pass filter on each partition interval Λ_l, which can be obtained by the construction of Littlewood-Paley and Meyer wavelets. Similarly to the classical wavelet transform, the detail coefficients $W_f^e(l,t)$ and the approximation coefficients $W_f^e(0,t)$ are represented by (1) and (2), respectively, as:

$$W_f^e(l,t) = \langle f, \psi_l \rangle = \int f(\tau)\overline{\psi_l(\tau - t)}\,\mathrm{d}\tau = F^{-1}\left(\hat{f}(\omega)\overline{\hat{\psi}_l(\omega)}\right), \tag{1}$$

$$W_f^e(0,t) = \langle f, \phi_1 \rangle = \int f(\tau)\overline{\phi_1(\tau - t)}\,\mathrm{d}\tau = F^{-1}\left(\hat{f}(\omega)\overline{\hat{\phi}_1(\omega)}\right), \tag{2}$$

where: $\phi_l(t)$ is an empirical scaling function and $\psi_l(t)$ is an empirical wavelet, the detailed expression of the two can refer to [14]. $\hat{\psi}_l(\omega)$ and $\hat{\phi}_1(\omega)$ are the Fourier transforms of $\psi_l(t)$ and

$\phi_l(t)$; $\overline{\hat{\psi}_l(\omega)}$ and $\overline{\hat{\phi}_1(\omega)}$ are the complex conjugate of $\psi_l(t)$ and $\phi_l(t)$. $F^{-1}(\cdot)$ denotes the inverse Fourier transform and $(\cdot)$ denotes the complex conjugation.

The signal reconstruction and empirical mode f_n are given by (3) and (4):

$$\begin{aligned} f(t) &= W_f^e(0,t) * \phi_1(t) + \sum_{l=1}^{L} W_f^e(l,t) * \psi_l(t) \\ &= F^{-1}\left[(0,t) * \hat{\phi}_1(\omega) + \sum_{l=1}^{L}(l,t) * \hat{\psi}_l(\omega)\right], \end{aligned} \tag{3}$$

$$f_0(t) = W_f^e(0,t) * \phi_1(t), \qquad f_l(t) = W_f^e(l,t) * \psi_l(t), \tag{4}$$

where $*$ is the convolution symbol.

2.2. Hard margin MKL method

Given a set of training data, $S = \{(\boldsymbol{x}_i, \boldsymbol{y}_i)\,|i = 1, \ldots, N\}$, $\boldsymbol{x}_i \in R^D$. When using a v-SVM model [22], similar to [20], the primal MKL problem with L2-norm regularization is written as:

$$\begin{aligned} &\min_{f_m, b, \xi_i, \xi_i^*} \frac{1}{2}\left(\sum_{m=1}^{M} \|f_m\|_{H_m}\right)^2 + C\left[v\varepsilon + \frac{1}{N}\sum_{i=1}^{l}\left(\xi_i + \xi_i^*\right)\right], \\ &\text{s.t.} \quad y_i - \left(\sum_{m=1}^{M} f_m(\boldsymbol{x}_i) + b\right) \le \varepsilon + \xi_i \quad \xi \ge 0, \\ &\left(\sum_{m=1}^{M} f_m(\boldsymbol{x}_i) + b\right) - y_i \ge \varepsilon + \xi_i^*, \quad \xi_i^* \ge 0, \quad \varepsilon \ge 0, \end{aligned} \tag{5}$$

where: $f_m(\boldsymbol{x}_i) = \langle f_m, \varphi(\boldsymbol{x}_i)\rangle$, ξ_i, ξ_i^* represent the slack variables, C is the regularization parameter. ε is the insensitive loss function, v is an upper bound on the fraction of training errors and a lower bound on the fraction of support vectors ($0 \le v < 1$), b is the bias. M is the total number of the base kernel. $\boldsymbol{\alpha}$, $\boldsymbol{\alpha}^*$ represent the Lagrange multiplier, $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_l]^T$, $\boldsymbol{\alpha}^* = \left[\alpha_1^*, \ldots, \alpha_l^*\right]^T$.

By using the Lagrange transform, the dual problem of hard margin MKL is

$$\max_{\boldsymbol{\alpha}, \tau} \ \tau\colon \mathrm{SVM}\{k_m, \boldsymbol{\alpha}\} \ge \tau \quad \forall m = 1, \ldots, M, \tag{6}$$

where:

$$\mathrm{SVM}\{k_m, \boldsymbol{\alpha}\} = -\frac{1}{2}\sum_{i,j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k_m(\boldsymbol{x}_i, \boldsymbol{x}_j) + \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)y_i,$$

$$k_m\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = \varphi_m\left(\boldsymbol{x}_i\right)^T \varphi_m\left(\boldsymbol{x}_j\right),$$

$\varphi_m(\boldsymbol{x})$ is the mapping function to map the data x from X to the reproducing kernel Hilbert space H, $k_1, \ldots, k_M$ denotes the M base kernels.

Alternatively, the dual problem in (14) of hard margin MKL can also be written as

$$\max_{\boldsymbol{\alpha}} \min_{\boldsymbol{\mu}} \ \mu_m \mathrm{SVM}\{k_m, \boldsymbol{\alpha}\}, \tag{7}$$

where: $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_M]^T$, μ_m is a coefficient that measures the importance of the m-th base kernel

$$U = \left\{ \boldsymbol{\mu} \middle| \sum_{m=1}^{M} \mu_m = 1, \mathbf{0} \leq \boldsymbol{\mu} \right\}.$$

Then the final output of hard margin MKL is given by

$$f(\boldsymbol{x}) = \sum_{i=1}^{N} \sum_{m=1}^{M} \left(\alpha_i - \alpha_i^*\right) \mu_m k_m (\boldsymbol{x}_i, \boldsymbol{x}) + b. \tag{8}$$

2.3. SMMKL method

Inspired by the soft margin SVM, in this section, the kernel slack variables are introduced for each base kernel. The kernel slack variable ξ_m is defined as the difference between the target margin τ and the SVM dual objective SVM$\{k_m, \boldsymbol{\alpha}\}$ for the given kernel k_m, that is

$$\xi_m = \tau - \text{SVM}\{k_m, \boldsymbol{\alpha}\} \quad \forall m = 1, \ldots, M. \tag{9}$$

Then, the loss function is defined as $Z_m = l(\xi_m), \forall m = 1, \ldots, M$, where $l(\cdot)$ is any general loss function. Two loss functions are used in the paper, they include the hinge loss (i.e., $l(\xi_m) = \max(0, \xi_m)$) and the square hinge loss, (i.e., $l(\xi_m) = \max(0, \xi_m)^2$).

In addition, the hinge loss soft margin MKL and the square hinge loss soft margin MKL are respectively abbreviated as SM1MKL and SM2MKL described in the following sections, and the naming methods of the two soft margin SVM are the same as here.

2.3.1. SM1MKL

Based on the definition of the kernel slack variable for each base kernel, when the kernel slack variable is introduced for hinge loss, the following objective function for the primal form of SM1MKL is as follows [20]:

$$\begin{aligned} &\min_{\boldsymbol{\mu}, f_m, b, \xi_i, \xi_i^*} \frac{1}{2} \sum_{m=1}^{M} \frac{\|f_m\|_{H_m}^2}{\mu_m} + C \left[v\varepsilon + \frac{1}{N} \sum_{i=1}^{N} (\xi_i + \xi_i^*) \right], \\ &\text{s.t.} \quad y_i - \left(\sum_{m=1}^{M} f_m(\boldsymbol{x}_i) + b \right) \leq \varepsilon + \xi_i, \quad \xi_i \geq 0, \\ &\left(\sum_{m=1}^{M} f_m(\boldsymbol{x}_i) + b \right) - y_i \geq \varepsilon + \xi_i^*, \quad \xi_i^* \geq 0, \quad \varepsilon \geq 0. \end{aligned} \tag{10}$$

The Lagrange transformation is performed on (10).

Set the partial derivatives of the original variables f_m, b, ξ_i, ξ_i^*, εv and μ_m in a Lagrange form to zeros, and substitute the primal variables into a Lagrangian function using corresponding Karush–Kuhn–Tucker (KKT) conditions. The corresponding objective function for the SM1MKL can be further described by the following:

$$\begin{aligned} &\min_{\tau, \alpha, \xi_m} -\tau + \theta \sum_{m=1}^{M} \xi_m, \\ &\text{s.t.} \quad \text{SVM}\{k_m, \boldsymbol{\alpha}\} \geq \tau - \xi_m, \quad \xi_m \geq 0, \quad m = 1, \ldots, M. \end{aligned} \tag{11}$$

The objective function of SM1MKL is to obtain the maximum margin τ as a certain error exit in the m-th base kernel, and the parameter θ is used to balance the loss term represented by the slack variables $\xi'_m s$ and the margin τ. By using the strong duality theorem, the constrained optimization problem of solving SM1MKL is transformed into solving the following optimization problem:

$$\min_{\mu} \max_{\alpha} J(\boldsymbol{\mu}, \boldsymbol{\alpha}), \tag{12}$$

where:

$$J(\boldsymbol{\mu}, \boldsymbol{\alpha}) = -\frac{1}{2} \sum_{m=1}^{M} \sum_{i,j=1}^{N} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \mu_m k_m(\boldsymbol{x}_i, \boldsymbol{x}_j) + \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) y_i \,,$$

$$U_1 = \left\{ \boldsymbol{\mu} \Big| \sum_{m=1}^{M} \mu_m = 1, \quad \mathbf{0} \le \boldsymbol{\mu} \le \theta \mathbf{1} \right\},$$

and $\mathbf{1} \in \mathbb{R}^m$ denote the vector of all ones.

The objective function of SM1MKL is the same as the objective function in [21], thus it is convex, which can be solved by using the block-wise coordinate descent algorithm [23], that is

$$\min_{\mu} \sum_{m=1}^{M} \frac{a_m}{\mu_m}, \tag{13}$$

where

$$a_m = \frac{1}{2} \sum_{i,j=1}^{N} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \mu_m^2 k_m(\boldsymbol{x}_i, \boldsymbol{x}_j).$$

Suppose $a_m > 0$, $m = 1, \ldots, M$ and $a_1 \ge a_2 \ge \cdots \ge a_M > 0$, by introducing Lagrange multipliers λ, η_m, ξ_m for the constraints in (13), the following Lagrange function can be obtained:

$$L = \sum_{m=1}^{M} \frac{a_m}{\mu_m} - \sum_{m=1}^{M} \mu_m \eta_m - \sum_{m=1}^{M} \xi_m (\theta - \mu_m) + \lambda \left(\sum_{m=1}^{M} \mu_m - 1 \right). \tag{14}$$

The partial derivative of L with respect to μ_m is set to zero. According to the complementary KKT condition, for $0 < \mu_m < \theta$, the following formula can be obtained:

$$-\frac{a_m}{\mu_m^2} + \lambda = 0, \quad \text{or} \quad \mu_m = \sqrt{\frac{a_m}{\lambda}}\,. \tag{15}$$

According to (15), for all cases of $a_m > 0$, the constraint $\mu_m \ge 0$ can be replaced by $\mu_m > 0$. Define ω, which means the number of elements in μ whose value is strictly equal to θ, the solution of the above problem can be written as:

$$\mu_m = \begin{cases} \theta, & m \le \omega \\ \dfrac{(1 - \omega\theta)\sqrt{a_m}}{\sum_{p=\omega+1}^{M} \sqrt{a_p}}, & m > \omega \end{cases}. \tag{16}$$

By following [20], and combining with the above analysis, we get the algorithm of SM1MKL that is briefly summarized in the following steps:

Step 1 – Initialize $\boldsymbol{\mu}^1 = 1/M, t = 1$.

Step 2 – Obtain $\boldsymbol{\alpha}^t$ by solving the sub-problem in (12) using the standard Quadratic Programming (QP) solver with $\boldsymbol{\mu}^t$.

Step 3 – Calculate a_m, $m = 1, \ldots, M$ and update $\boldsymbol{\mu}^{t+1}$. $\boldsymbol{\mu}^{t+1}$ can be updated by solving the sub-problem in (13).

Step 4 – $t = t + 1$, determine whether the stop condition of the algorithm is satisfied. If it is satisfied, the algorithm stops, the optimal solutions $\boldsymbol{\alpha}, \boldsymbol{\mu}$ are given; otherwise, return to Step 2. Define $obj(t)$ as the solution of (12) obtained in the t-th cycle, and set the stop condition:

$$abs\left(\left(obj(t) - obj(t-1)\right)/obj(t-1)\right) \leq 1e-3.$$

Step 5 – Compute the output of SM1MKL.

$$y_t = f(\boldsymbol{x}_t) = \sum_{i=1}^{N}\sum_{m=1}^{M}(\alpha_i - \alpha_i^*)\mu_m k_m(\boldsymbol{x}_i, \boldsymbol{x}_t) + b. \tag{17}$$

2.3.2. SM2MKL

The original form of the objective function for SM2MKL is as follows [20]:

$$\begin{aligned}
&\min_{\boldsymbol{\mu}, f_m, b, \xi_i, \xi_i^*} \frac{1}{2}\sum_{m=1}^{M}\frac{\|f_m\|_{H_m}^2}{\mu_m} + C\left[v\varepsilon + \frac{1}{N}\sum_{i=1}^{N}(\xi_i + \xi_i^*)\right] + \frac{1}{2\theta}\sum_{m=1}^{M}\mu_m^2,\\
&\text{s.t. } y_i - \left(\sum_{m=1}^{M} f_m(\boldsymbol{x}_i) + b\right) \leq \varepsilon + \xi_i, \quad \xi_i \geq 0,\\
&\left(\sum_{m=1}^{M} f_m(\boldsymbol{x}_i) + b\right) - y_i \geq \varepsilon + \xi_i^*, \quad \xi_i^* \geq 0, \quad \varepsilon \geq 0.
\end{aligned} \tag{18}$$

Similarly to SM1MKL, the following constrained optimization problems of SM2MKL can be expressed as:

$$\begin{aligned}
&\min_{\tau, \alpha, \xi_m} -\tau + \frac{\theta}{2}\sum_{m=1}^{M}\xi_m^2,\\
&\text{s.t. } \mathrm{SVM}\{k_m, \boldsymbol{\alpha}\} \geq \tau - \xi_m, \quad m = 1, \ldots, M.
\end{aligned} \tag{19}$$

Similarly to (13), the objective function problem for solving SM2MKL is transformed into the following problem:

$$\min_{\mu}\max_{\alpha} J(\boldsymbol{\mu}, \boldsymbol{\alpha}) + \frac{1}{2\theta}\sum_{m=1}^{M}\mu_m^2 \tag{20}$$

and

$$U_2 = \left\{\boldsymbol{\mu}\middle|\sum_{m=1}^{M}\mu_m = 1, \quad 0 \leq \boldsymbol{\mu}\right\}.$$

With a fixed $\boldsymbol{\mu}$, the optimization problem with respect to $\boldsymbol{\alpha}$ is a standard QP problem, which can be optimized with the QP solver. When $\boldsymbol{\alpha}$ is fixed, the kernel combination coefficient can be updated using an algorithm based on projection gradient descent. Following [22], the gradient $\boldsymbol{p}^t$ of the optimization problem in (20) with respect to $\boldsymbol{\mu}$ can be computed as:

$$p_m = -h_m + \frac{1}{\theta}\mu_m, \quad m = 1, \ldots, M, \tag{21}$$

where

$$h_m = \frac{1}{2}\sum_{i,j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k_m(\boldsymbol{x}_i, \boldsymbol{x}_j).$$

The coefficient $\boldsymbol{\mu}$ is updated by using the coefficients $\boldsymbol{\mu}^t$ at the current iteration, that is

$$\boldsymbol{\mu}_{\text{sub}}^* = \Pi_{U_2}\left(\boldsymbol{\mu}^t - \eta_t \boldsymbol{p}^t\right), \tag{22}$$

where: $\Pi_{U_2}(\cdot)$ is the simplex projection operation and η_t is the updating step size.

The procedure of an SM2MKL algorithm is summarized as follows:

Step 1 – Initialize $\boldsymbol{\mu}^1 = 1/M$, $t = 1$.

Step 2 – Obtain $\boldsymbol{\alpha}^t$ by solving the sub-problem in (12) using the standard QP solver with $\boldsymbol{\mu}^t$.

Step 3 – Calculate $\boldsymbol{\mu}_{\text{sub}}^*$ that can reduce the objective function value for the problem in (20) and update $\boldsymbol{\mu}^{t+1} = \boldsymbol{\mu}_{\text{sub}}^*$.

Step 4 – $t = t + 1$, determine whether the stop condition of the algorithm is satisfied, where the stop condition is the same as the algorithm of SM1MKL. If it is satisfied, the algorithm stops, the optimal solutions $\boldsymbol{\alpha}$, $\boldsymbol{\mu}$ are given; otherwise, return to Step 2.

Step 5 – Compute the output of SM2MKL according to (17).

3. EWT-SMMKL model

Combining the advantages of EWT and SMMKL methods, a new combination prediction method for short-term wind power prediction is proposed in the paper. The overall combined prediction process is shown in Fig. 1.

In the paper, according to the way of time series modeling, the wind speed prediction model is established as follows:

$$\hat{y}(t + \Delta) = f(\boldsymbol{x}_t) = \sum_{i=1}^{N}\sum_{m=1}^{M}\left(\alpha_i^* - \hat{\alpha}_i^*\right)\mu_m^* k_m\left(\boldsymbol{x}_i, \boldsymbol{x}_t\right) + b^*, \tag{23}$$

where: Δ represents the prediction step size, $\boldsymbol{x}_t = (y_{t-1}, \ldots, y_{t-D})$ is the multi-dimensional input vector, i.e., historical wind speed data, D is the embedding dimension. M is the number of the base kernel, N is the number of wind speed data sets and $\hat{y}(t + \Delta)$ is the corresponding predicted output.

The specific implementation steps are as follows:

Step 1 – According to (1) and (2), the original wind speed sequence with strong randomness and non-linearity is decomposed into a series of different mode components, i.e., stationary subseries using an EWT algorithm.

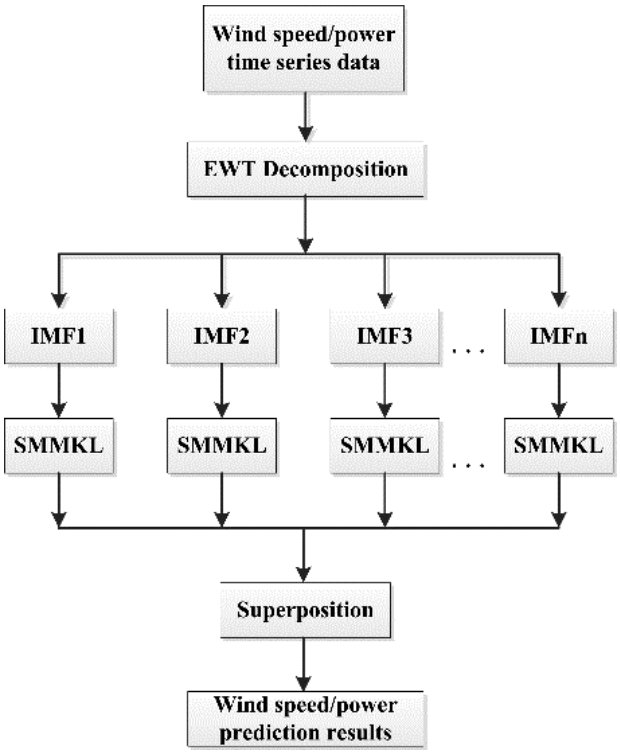

Fig. 1. Flow chart of EWT-SMMKL combined prediction

Step 2 – Establish different prediction models using the proposed SMMKL methods for each sub-sequence and the corresponding predicted values can be obtained according to (23).

Step 3 – Finally, based on the wind speed prediction output of each sub-series the final predictions are superimposed according to (3) and (4).

Step 4 – Consistent with the solution method in [25], the wind power predictions can be obtained by wind speed-to-wind power conversion curve.

4. Experimental analysis

In the following case, to measure the prediction performance using different MKL methods for wind power prediction, three polynomial functions and six radial basic function (RBF) kernel functions are selected as basic kernel functions when using SM1MKL, SM2MKL, SimpleMKL algorithms, and the kernel functions' form are as follows:

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \left((\boldsymbol{x}_i \cdot \boldsymbol{x}_j) + 1\right)^p, \tag{24}$$

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left\{-\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\| / 2\delta^2\right\}. \tag{25}$$

The degree of the polynomial kernel function is selected from $p = \{1, 2, 3\}$, the RBF kernel function using six different bandwidth parameters from $2\delta^2 \in \{0.01, 0.1, 1, 10, 50, 100\}$, and the regularization parameter $C = 100$. The parameter θ is introduced for SM1MKL and SM2MKL, for SM1MKL, $\theta \in \{1/M, 0.1, 0.2, \ldots, 1\}$; for SM2MKL, $\theta \in \left\{10^{-5}, \ldots, 10^{4}, 10^{5}\right\}$. Moreover, the RBF kernel function is selected as the kernel function of an SVM, and the algorithm implementing the SVM uses LIBSVM software [24]. The optimal parameters are determined by using five-fold cross validation on the training set.

In addition, to evaluate the performance of the model, the mean absolute error (MAE), root mean square error (RMSE), standard deviation (Std) [25] of the MAE are used, which are expressed as follows:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}(i) - y(i)|, \tag{26}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}(i) - y(i))^2}, \tag{27}$$

$$\mathrm{Std} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\hat{y}(i) - y(i) - E_{\mathrm{MAE}})^2}, \tag{28}$$

where: $y(i)$ is the actual value, $\hat{y}(i)$ is the prediction value of the prediction model and N is the number of samples in the test set.

4.1. Data

The Western dataset provided by National Renewable Energy Laboratory (NREL) [26] in the United States was used for the experiments in this section and the data were sampled at a time interval of 10 min and a spatial interval of 2 km. Sixty-eight grid points (i.e., 680 Vestas v-90 3-MW wind turbines) located 10 miles west of Denver were selected for the experiment, where each data sample contains the average of the wind speed and wind power provided by 68 grid points at the same time. The average of two consecutive samples from the original dataset was used as the selected new data, and the data were sampled at an interval of 20 min. The test dataset contained wind speed data from December 1 to December 7, 2004, and the wind speed data contained in the previous 14 days were used as training.

4.2. Experiments analysis and results

First, the original wind speed of the training data set is transformed using the EWT method, and by choosing the relative amplitude ratio $\delta = 0.5$, the wavelet number $L = 4$ can be calculated. Therefore, the output of EWT consists of a band-pass filter bank with one scale function and four wavelet functions. Figure 2 shows the extracted five mode component signals after EWT.

Secondly, for each component signal of wind speed (F0-F4), a time series modelling method was used to establish a multi-step prediction model for the SMMKL method 60-minutes in advance. The embedded dimension of each sub-series model is set as $D = 6$. With different kernel functions selected, Table 1 shows a comparison of the Std value for different MKL methods for

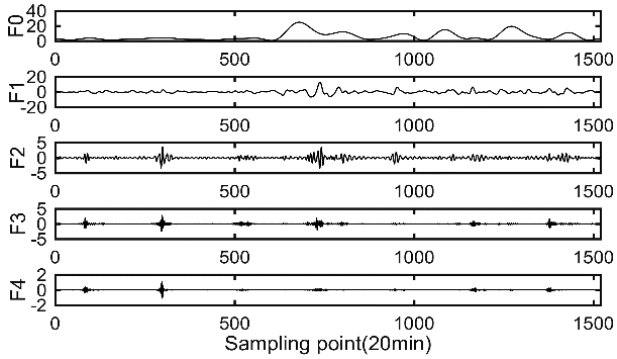

Fig. 2. Mode signals extracted by the EWT

wind speed prediction 60-min ahead with different combinations of kernel functions. According to the results in Table 1, the SMMKL method exhibits higher robustness than the SimpleMKL method, and the best prediction accuracy is obtained when three polynomial kernel functions and six RBF kernel functions are selected as the kernel functions combination of the SM1MKL method.

Table 1. Comparison of Std of different MKL methods for 60-min ahead power prediction under different kernel function combinations

Kernel function	**SimpleMKL**	**SM1MKL**	**SM2MKL**
Three polynomial kernel	4.2805	3.8735	4.0088
Six RBF kernel	4.2931	3.8639	3.9729
Three polynomial mix six RBF kernel	3.9531	3.4033	3.6493

Figure 3 shows the distribution comparison of kernel weight using SimpleMKL, SM1MKL and SM2MKL algorithms when establishing the prediction model based on F0 component for

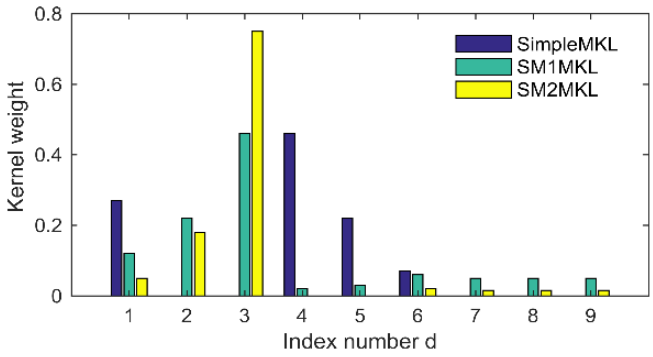

Fig. 3. The distribution of the weight coefficient of each base kernel

60-min advance prediction. Among the index numbers of d indicated by the x-axis, the first three indexes correspond to the polynomial kernel and the other indexes correspond to the RBF kernel. It can be observed that the kernel weights of the SimpleMKL method are sparse, and some useful base kernels may be discarded.

Finally, the wind speed predictions are output by superimposing the wind speed predictions of each sub-series and the final power predictions are obtained by wind speed-to-wind power conversion [25]. Figure 4 to Fig. 6 show the comparison of the multi-step prediction results, errors and normalized errors of the prediction models based on different algorithms for 60-min ahead wind power prediction on the test set, where the normalized error is defined as: error/$P_{\text{norm}} \times 100\%$. From the results in Fig. 4 to Fig. 6, it can be concluded that the combined prediction methods based on EWT and SMMKL can accurately predict the actual wind power value with low fluctuations in prediction error, shows a good prediction effect, with the EWT-SMMKL method achieving the best prediction accuracy.

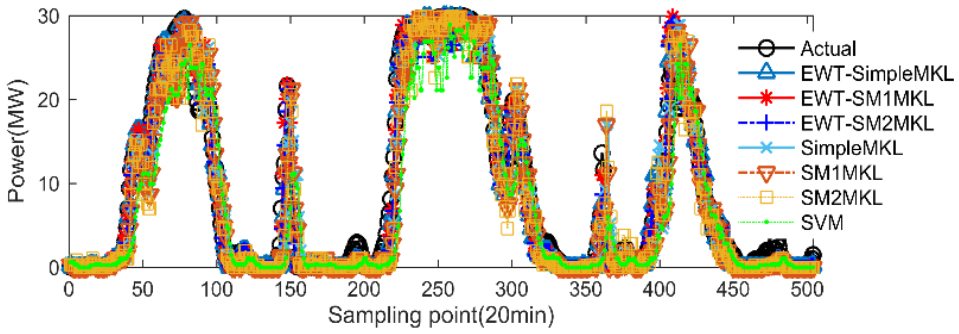

Fig. 4. Comparison of results using EWT-SMMKL methods for 60-min ahead prediction

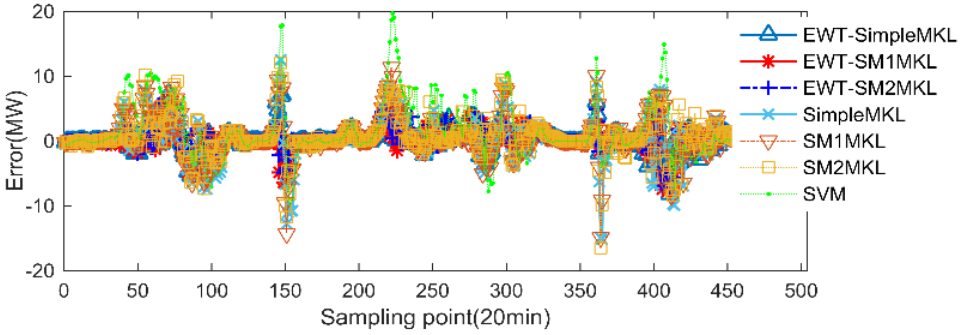

Fig. 5. Comparison of errors using EWT-SMMKL methods for 60-min ahead prediction

Table 2 to Table 4 shows a comparison of the errors when predicting wind power 20, 40 and 60 minutes in advance based on different algorithms on the test dataset. To verify the prediction performance of the EWT-SMMKL model, this experiment also compares its prediction results

with those of SVM and other methods. As can be seen from Table 2 to Table 4, the two EWT-SMMKL methods exhibit better predictive performance compared to the single MKL method.

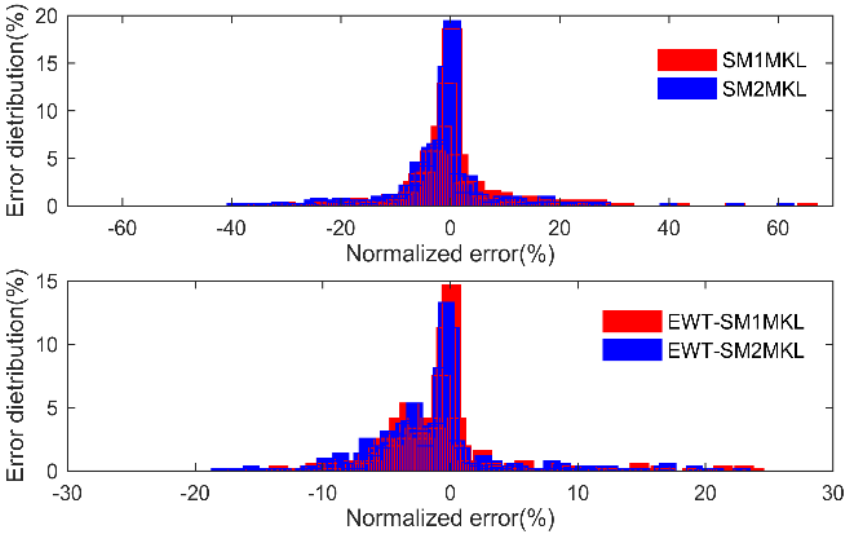

Fig. 6. Comparison of normalized errors using different SMMKL methods for 60-min ahead prediction

Table 2. Performance comparison of different methods for power prediction 20-min ahead of test set

Method	MAE	Std	RMSE
SVM	1.2031	2.3722	1.8222
SM1SVM	1.1140	2.2309	1.9397
SM2SVM	1.1099	2.2242	1.9265
SimpleMKL	1.1511	2.2043	1.8130
SM1MKL	1.0093	1.9966	1.7499
SM2MKL	1.1015	2.2719	1.7655
EWT-SM1SVM	1.2142	2.4984	1.8881
EWT-SM2SVM	1.1587	2.3932	1.8511
EWT-SimpleMKL	0.9619	1.9860	1.5169
EWT-SM1MKL	0.9479	1.9641	1.5167
EWT-SM2MKL	0.9492	1.9528	1.5035

Figure 7 further shows a comparison of the MAE values with the prediction step size for the EWT-SMMKL method after the multi-step prediction of wind power from 20 minutes to

180 minutes in advance. The results in Fig. 5 show that the combined forecasting method based on EWT-SMMKL has better prediction performance than SMMKL and other methods, with a significantly lower MAE value, and that the EWT-SM1MKL method has the best forecasting results.

Table 3. Performance comparison of different methods for power prediction 40-min ahead of test set

Method	**MAE**	**Std**	**RMSE**
SVM	1.6062	3.2843	2.6544
SM1SVM	1.5487	3.0849	2.6349
SM2SVM	1.5512	3.0934	2.6396
SimpleMKL	1.5340	3.1762	2.5572
SM1MKL	1.3004	2.6357	2.2991
SM2MKL	1.3963	2.9328	2.3493
EWT-SM1SVM	1.5201	3.1587	2.4166
EWT-SM2SVM	1.4676	3.0236	2.3802
EWT-SimpleMKL	0.9834	2.0079	1.5710
EWT-SM1MKL	0.9488	1.9565	1.5173
EWT-SM2MKL	0.9522	1.9733	1.5191

Table 4. Performance comparison of different methods for power prediction 60-min ahead of test set

Method	**MAE**	**Std**	**RMSE**
SVM	1.9889	4.1172	3.4167
SM1SVM	1.9504	3.9266	3.3217
SM2SVM	1.9515	3.9260	3.3219
SimpleMKL	1.8977	3.9531	3.2157
SM1MKL	1.6878	3.4033	2.9254
SM2MKL	1.7233	3.6493	2.9638
EWT-SM1SVM	1.8061	3.8270	2.9560
EWT-SM2SVM	1.7435	3.6402	2.9112
EWT-SimpleMKL	1.1198	2.2427	1.7388
EWT-SM1MKL	1.0269	2.0975	1.6674
EWT-SM2MKL	1.0793	2.2001	1.6213

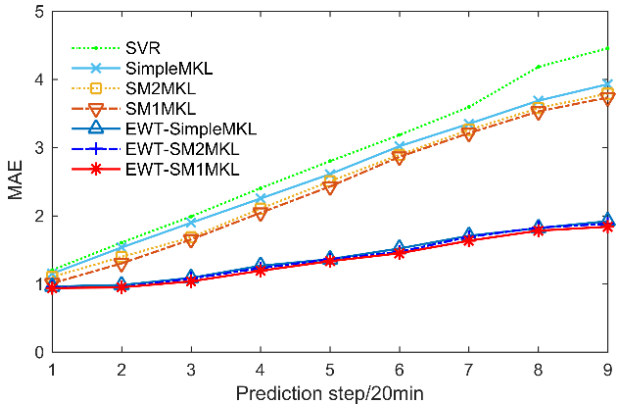

Fig. 7. Comparison of prediction errors in terms of MAE based on different methods for different steps

5. Conclusions

For short-term wind power prediction, a combined prediction method based on EWT-SMMKL is proposed. After theoretical analysis and the establishment of comparative experiments, the following conclusions are drawn.

1. EWT can effectively improve the non-stationarity of nonlinear wind speed data and improve the prediction accuracy of short-term wind power prediction based on decomposition results.
2. The SMMKL method not only retains the advantages of the hard margin MKL method in that the prediction results do not depend on the kernel function and its parameter selection, but also selects as many useful base kernels as possible for prediction, providing better prediction performance in short-term wind power prediction.
3. The prediction performance of the EWT-SMMKL method is better than SVM and other single MKL methods in dealing with short-term wind power prediction. Based on the above analysis, the proposed short-term wind power prediction method is feasible.

The analysis based on the above results shows that the proposed short-term wind power forecasting method is feasible. In future works, combining the proposed SMMKL method, the researchers plan to use new decomposition approaches, such as VMD to preprocess the data needed to further improve prediction accuracy.

Acknowledgements

This work was financially supported by the Natural Science Foundation of China (51467008).

References

[1] Wang Q., Martinez-Anido C.B., Wu H.Y., Florita A.R., Hodge B.M., *Quantifying the economic and grid reliability impacts of improved wind power prediction*, IEEE Transactions on Sustainable Energy, vol. 7, no. 4, pp. 1525–1537 (2016), DOI: 10.1109/TSTE.2016.2560628.

[2] Liu H.Q., Li W.J., Li Y.C., *Ultra-short-term wind power prediction based on copula function and bivariate EMD decomposition algorithm*, Archives of Electrical Engineering, vol. 69, no. 2, pp. 271–286 (2020), DOI: 10.24425/aee.2020.133025.

[3] Waśkowicz B., *Statistical analysis and dimensioning of a wind farm energy storage system*, Archives of Electrical Engineering, vol. 66, no. 2, pp. 265–277 (2017), DOI: 10.1515/aee-2017-0020.

[4] Cassola F., Burlando M., *Wind speed and wind energy forecast through Kalman filtering of numerical weather prediction model output*, Applied Energy, vol. 99, no. 6, pp. 154–166 (2012), DOI: 10.1016/j.apenergy.2012.03.054.

[5] Li J., Li M., *Prediction of ultra-short-term wind power based on BBO-KELM method*, Journal of Renewable and Sustainable Energy, vol. 11, no. 5, 056104 (2019), DOI: 10.1063/1.5113555.

[6] Zhang Y.G., Wang P.H., Zhang C.H., Lei S., *Wind energy prediction with LS-SVM based on Lorenz perturbation*, The Journal of Engineering, vol. 2017, no. 13, pp.1724–1727 (2017), DOI: 10.1049/joe.2017.0626.

[7] Duan J., Wang P., Ma W., *Short-term wind power forecasting using the hybrid model of improved variational mode decomposition and Correntropy Long Short-term memory neural network*, Energy, vol. 214, 118980 (2021), DOI: 10.1016/j.energy.2020.118980.

[8] Moreno S.R., Silva R.G. D., Mariani V.C., *Multi-step wind speed forecasting based on hybrid multi-stage decomposition model and long short-term memory neural network*, Energy Conversion and Management, vol. 213, 112869 (2020), DOI: 10.1016/j.enconman.2020.112869.

[9] Ramon G.D., Matheus H.D.M.R., Sinvaldo R.M., *A novel decomposition-ensemble learning framework for multi-step ahead wind energy forecasting*, Energy, vol. 216, 119174 (2021), DOI: 10.1016/j.energy.2020.119174.

[10] Yldz C., Akgz H., Korkmaz D., *An improved residual-based convolutional neural network for very short-term wind power forecasting*, Energy Conversion and Management, vol. 228, no. 1, 113731 (2021), DOI: 10.1016/j.enconman.2020.113731.

[11] Ribeiro G.T., Mariani V.C., Coelho L.D.S., *Enhanced ensemble structures using wavelet neural networks applied to short-term load forecasting*, Engineering Applications of Artificial Intelligence, vol. 28, no. June, pp. 272–281 (2019), DOI: 10.1016/j.engappai.2019.03.012.

[12] Liu X., Zhou J., Qian H.M., *Short-term wind power forecasting by stacked recurrent neural networks with parametric sine activation function*, Electric Power Systems Research, vol. 192, 107011 (2021), DOI: 10.1016/j.epsr.2020.107011.

[13] Zhu R., Liao W., Wang Y., *Short-term prediction for wind power based on temporal convolutional network*, Energy Reports, vol. 6, pp. 424–429 (2019), DOI: 10.1016/j.egyr.2020.11.219.

[14] Gilles J., *Empirical wavelet transform*, IEEE Transactions on Signal Processing, vol. 61, no. 16, pp. 3999–4010 (2013), DOI: 10.1109/TSP.2013.2265222.

[15] Wang S.X., Zhang N., Wu L., Wang Y.M., *Wind speed prediction based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method*, Renewable Energy, vol. 94, pp. 629–636 (2016), DOI: 10.1016/j.renene.2016.03.103.

[16] Lanckriet G.R.G., Cristianini N., Bartlett P.L., Ghaoui L.E., Jordan M.I., *Learning the kernel matrix with semi-definite programming*, Journal of Machine learning research, vol. 5, pp. 323–330 (2002).

[17] Gönen M., Alpaydin E., *Multiple kernel learning algorithms*, Journal of Machine Learning Research, vol. 12, pp. 2211–2268 (2011).

[18] Wu D., Wang B.Y., Precup D., Boulet B., *Multiple kernel learning based transfer regression for electric load forecasting*, IEEE Transactions on Smart Grid, vol. 11, no. 2, pp. 1183–1192 (2020), DOI: 10.1109/TSG.2019.2933413.

[19] Rakotomamonjy A., Bach F.R., Canu S., Grandvalet Y., *Simplemkl*, Journal of Machine Learning Research, vol. 9, no. 3, pp. 2491–2521 (2008), DOI: 10.1007/s10846-008-9235-4.

[20] Xu X.X., Tsang I.W., Xu D., *Soft margin multiple kernel learning*, IEEE Transactions on Neural Networks and Learning Systems, vol. 24, no. 5, pp. 749–761 (2013), DOI: 10.1109/TNNLS.2012.2237183.

[21] Dong Y., Zhang T., Xi L., *Blind steganalysis method for JPEG steganography combined with the semisupervised learning and soft margin support vector machine*, Journal of Electronic Imaging, vol. 24, no. 5, pp. 013008.1-013008.8 (2015), DOI: 10.1117/1.JEI.24.1.013008.

[22] Chapelle O., Vapnik V., Bousquet O., *Choosing multiple parameters for support vector machines*, Machine Learning, vol. 46, no. 1–3, pp. 131–159 (2002), DOI: 10.1023/A:1012450327387.

[23] Kloft M., Brefeld U., Sonnenburg S., *Lp-norm multiple kernel learning*, Journal of Machine Learning Research, vol. 12, pp. 953–997 (2011), DOI: 10.14279/depositonce-2978.

[24] Chang C.C., Lin C.J., *LIBSVM: A library for support vector machines*, Acm Transactions on Intelligent Systems and Technology, vol. 2, no. 3, pp. 1–27 (2011), DOI: 10.1145/1961189.1961199.

[25] Zeng J., Qiao W., *Short-term wind power prediction using a wavelet support vector machine*, IEEE Transactions on Sustainable Energy, vol. 3, no. 2, pp. 255–264 (2012), DOI: 10.1109/TSTE.2011.2180029.

[26] Potter C.W., Lew D., McCaa J., *Creating the dataset for the western wind and solar integration study (USA)*, Wind Engineering, vol. 32, no. 4, pp. 325–338 (2008), DOI: 10.1260/0309-524X.32.4.325.