Table of Contents

1 Executive Summary ........................................................................................................... 4

2 Introduction ................................................................................................................ ......4

   2.1 A Review of Wind Power Prediction (WPP) Methods ............................................ 4
   2.2 A Review of Solar Power Prediction Methods ...................................................... 5

3 The Mathematical Models Used for WPP and SPP of This Project ..............................6

   3.1 The AR Model ......................................................................................................... 6
   3.2 The LLR Model ..................................................................................................... 6
   3.3 The RBFNN Model .............................................................................................. 7
   3.4 The SVM Model .................................................................................................. 8

4 The WSVM-Based WPP Model .......................................................................................11

   4.1 Data Normalization ............................................................................................ 11
   4.2 Feature Representation ....................................................................................... 12
   4.3 Wind Speed-to-Wind Power Conversion ........................................................... 13
   4.4 Performance Evaluation ..................................................................................... 15

5 Typical WPP Results .........................................................................................................16

6 Data Representation and Preprocessing for SPP ...............................................................19

   6.1 2D Representation ............................................................................................. 19
   6.2 Correlation Analysis ......................................................................................... 20
   6.3 Normalization .................................................................................................. 21
   6.4 Performance Evaluation .................................................................................... 22

7 Typical SPP Results ...........................................................................................................23

8 Conclusion .................................................................................................................. .......27

References .................................................................................................................... ............27
## List of Figures

1. The one-hour prediction template ........................................................................................7
2. The structure of an RBFNN .................................................................................................8
3. The structure of an SVM ......................................................................................................9
4. The proposed wavelet kernel .............................................................................................10
5. The structure of the proposed WSVM-based WPP model ............................................11
6. Wind speed normalization ..................................................................................................12
7. Autocorrelations of input samples ....................................................................................12
8. Wind speed profile and corresponding wind power of 10 Vestas V-90 3-MW wind turbines obtained from the NREL database ........................................................................14
9. The process of generating the power curve .......................................................................15
10. One-hour ahead WPP using the WSVM Model and fixed-step scheme ......................16
11. Two-hours-ahead WPP using the WSVM model and fixed-step scheme ......................17
12. Three-hours-ahead WPP using the WSVM model and fixed step scheme ...................17
13. Comparison among the persistence, RBF-SVM, and WSVM models .........................18
14. Comparison between the RBF-SVM and WSVM models ...........................................18
15. Hourly solar radiation data in a 1D time plot .............................................................19
16. A 2D image view of the solar radiation data ..............................................................20
17. A 2D view of autocorrelations of the solar radiation .....................................................21
18. Comparison of the two methods for normalization of the radiation data ....................22
19. One-hour-ahead prediction in Denver using SVM .........................................................24
20. Correlation between the real and predicted solar radiation in Denver .........................24
21. One-hour-ahead prediction in Seattle with SVM ..........................................................25
22. Comparison of the MAEs and MAPEs of the AR, RBFNN, and SVM-based prediction models in Denver and Seattle .................................................................26
23. Comparison of the one-hour-ahead predicted values from the three models with the observations in two consecutive days in Seattle ..................................................26
1. Executive Summary

The information on the wind/solar resources, which determine the amount of electric power generated from the wind/solar hybrid power systems, is important for power management of the Energy-Plus Roadway/Traffic signal Light (EPRTL) units and the overall Roadway wind/solar Hybrid Power generation and distribution System (RHPS) microgrid. Several statistical time-series prediction models, including the Autoregressive (AR), local linear regression (LLR) model, radial basis function neural network (RBFNN), radial basis function support vector machine (RBF-SVM), and wavelet support vector machine (WSVM) models, have been developed for wind and solar power prediction [1], [2]. Simulation studies have been carried out to validate the proposed wind and solar power prediction models by using the data obtained from the National Renewable Energy Laboratory (NREL); results have shown that the proposed SVM models are preferable for a few hours-ahead wind and solar power prediction in terms of prediction accuracy and computational cost.

2. Introduction

Generating electricity from clean and renewable energy sources, e.g., wind and solar energy, has received more and more attention in the last decade. However, due to the intermittency of wind and solar energy sources, the increasing penetration of wind and solar power can create significant uncertainties in electricity generation, which has an adverse effect on the operation of electric power grids. To accommodate the increasing penetration of wind and solar power, it is essential to develop appropriate methods for wind and solar power prediction.

2.1 A Review of Wind Power Prediction (WPP) Methods

The existing WPP methods can be generally classified into two categories, i.e., physical model-based methods and statistical model-based methods. The physical model-based methods [3] take as the input not only the historical data of wind power but also meteorological information obtained from the Numerical Weather Prediction (NWP) model and other information, such as local weather conditions [4]-[6]. The statistical model-based methods [7], [8] predict wind power by using only measured historical and current values of wind power. No extra information is required. Compared to the statistical model-based methods, the physical model-based methods usually perform well for longer-term predictions (e.g., more than 6 hours ahead) but are more complicated and require vast computational resources and weather information. The statistical model-based methods are preferable for short-term prediction [9]. By combining the physical and statistical models, some hybrid model-based tools [10] were developed. They combine the benefits of both models to improve the performance of WPP [11]. This project focuses on the statistical model-based, short-term WPP.

The statistical models can be further divided into linear and nonlinear models. The persistence and autoregressive moving average (ARMA) models are two traditional linear models used in WPP [12]. The commonly used nonlinear statistical WPP models are based on artificial intelligence techniques, such as artificial neural networks (ANNs) [13]-[15] and SVMs [17]. These nonlinear models were proven to be superior over linear models [14], [15]. Recent results have shown that the SVM-based models either compare favorably with [15] or outperform [16]
the ANN-based models in WPP. For example, SVM-based models have been found to take less computational time compared to the ANN-based models [18].

Wavelet analysis is a relatively new mathematical technique used to analyze the nature of signals. This technique has been recognized as a promising tool for nonstationary signal approximation and frequency analysis. In [19], the original wind speed series was decomposed with wavelets; the ARMA model was then used to predict the coefficients of different layers of the wavelets, which outperformed the method that uses the ARMA model directly to predict wind speed. De Aquino [20] used wavelet transform as a data preprocessing technique for an ANN for WPP, which is superior to the model using the ANN only. In [20], the inputs of the ANN are the wavelet coefficients derived from the multiresolution analysis [21].

The team for this project has developed a statistical WPP model by combining wavelet analysis with SVM. A new SVM kernel has been proposed based on the wavelet mother function in [22]. The new kernel can change between an RBF kernel and a Mexhat kernel and outperforms the RBF as well as the Mexhat. The resulting WSVM-based model is preferable for short-term WPP. The proposed model has been validated by using the data obtained from NREL and has been compared with the persistence and RBF-SVM models to show its superiority.

2.2 A Review of Solar Power Prediction (SPP) Methods

Similar to WPP, there are mainly two categories of SPP methods: physical model-based and statistical model-based methods. The physical model-based methods are based on physical processes occurring in the atmosphere and influencing solar radiation [23]. They are used to estimate direct and diffuse radiation with high spatial and temporal resolution [24], a clearness index, or a cloudiness index (CI) [25]. The CI can be further used to estimate radiation using the formula in [26]. The physical model-based methods do well in medium-term and long-term predictions.

The statistical model-based methods are based on time series analysis [27] and do better in short-term prediction. AR and ARMA [28] are among the linear models frequently used in solar power predictions [29], [30]. Nonlinear methods, such as the Takagi-Sugeno (TS) fuzzy model [31] and wavelet-based methods [32], have been shown to be superior to linear models. Some studies also indicate that ANNs can achieve a good performance in SPP [32], [33]. The ANN-based models involve modeling of daily or hourly solar radiation, clearness index [34], cloudiness index [35], [36], and effective transmission modeling [37]. Other studies also show that the SPP using multivariates, such as sun duration, temperature, wind speed, and relative humidity, can achieve much better performance than that using a univariate [38]. For example, Rivington [39] predicted solar radiation by using solar duration and air temperature. Ahhi et al. [33] used air temperature, wind speed, sun duration, and relative humidity as the inputs of an ANN to predict solar radiation. However, the effectiveness of using each individual meteorological variable has not been studied yet.

Nevertheless, solar radiation readings were taken as a one-dimensional (1D) time series in most of the existing work, which turned out to be inferior to a two-dimensional (2D) representation [40]. The 2D representation of solar radiation makes it possible to combine image processing methods with nonlinear prediction methods to improve the accuracy of SPP [41]. The team for
this project has developed four different methods with a 2D representation of solar radiation and other meteorological variables as the input for SPP, including AR model-based, LLR model-based, RBFNN-based, and SVM-based methods.

3. The Mathematical Models Used for WPP and SPP of This Project

3.1 The AR Model

Consider an arbitrary time series, \( X_t \), which can be converted to a mean-adjusted time series as follows:

\[
x_t = X_t - \bar{X}
\]  

where \( \bar{X} \) is the sample mean of the original time series \( X_t \), and \( x_t \) is the mean-adjusted time series. The AR model expresses a time series as a linear function of its past values. The order of the AR model indicates how many past values are used. An AR model with an order of \( p \), \( AR(p) \), can be written as:

\[
x_t = a_1 x_{t-1} + a_2 x_{t-2} + \cdots + a_p x_{t-p} + e_t
\]  

where \([x_{t-1}, x_{t-2}, \ldots, x_{t-p}]\) are the past values of the time series; \( a_i (i = 1, \cdots, p) \) is the autoregressive coefficient; and \( e_t \) is noise or error, which is assumed to be a normally distributed random number.

3.2 The LLR Model

The use of the LLR model was inspired by the locally high correlations of solar radiation. The idea is that the global nonlinearity of solar radiation can be approximated by multiple local linear models. The LLR differs from the AR model in its time-variant coefficients. These coefficients vary over time when using the LLR for SPP. Let \( i \) and \( j \) denote the index of hour and day, respectively, then a one-hour 2D linear prediction of solar radiation can be expressed by \( \hat{y}_{i,j} = X_{i,j}^T a_i \), where \( X_{i,j} = [x_{i-1,j}, x_{i,j-1}, x_{i,j-2}]^T \); \( a_i = [a_1^{(i)}, a_2^{(i)}, a_3^{(i)}]^T \) is the linear coefficient vector and \( x_{ij} \) corresponds to the radiation at the \( i \)th hour of the \( j \)th day. Then the error can be estimated as:

\[
e_i = \sum_{j=2}^{m} (y_{i,j} - X_{i,j}^T a_i)^2
\]  

where \( m \) is the number of training days. The coefficients that minimize the error in Eq. (3) can be found from the solution of \( \partial e_i / \partial a_i = 0 \), which yields the following equation:

\[
\begin{pmatrix}
R_{11}^{(i)} & R_{12}^{(i)} & R_{13}^{(i)} \\
R_{21}^{(i)} & R_{22}^{(i)} & R_{23}^{(i)} \\
\vdots & \vdots & \vdots \\
R_{m1}^{(i)} & R_{m2}^{(i)} & R_{m3}^{(i)}
\end{pmatrix}
\begin{pmatrix}
a_1^{(i)} \\
2
\end{pmatrix}
= 
\begin{pmatrix}
r_1^{(i)} \\
r_2^{(i)} \\
\vdots \\
r_m^{(i)}
\end{pmatrix}
\]
where \( R_{jk}^{(i)} \) is the correlation between \( x_{ij} \) and \( x_{ik} \) within the prediction template [41].  Fig. 1 shows the one-hour-prediction template at the \( i^{th} \) hour, which contains \( x_{ij}, x_{i-1,j}, and x_{i-2,j} \).  \( r_k^{(i)} \) corresponds to the correlation between the predicted value and each sample in the prediction template.  For instance, \( r_1^{(i)} \) is the correlation between \( x_{ij} \) and \( x_{i-1,j} \); \( r_2^{(i)} \) is the correlation between \( x_{ij} \) and \( x_{i-1,j} \); etc.  Then the coefficients at the \( i^{th} \) hour can be obtained:

\[
a_i = (R^{(i)})^+ \cdot r^{(i)} 
\]

where \((R^{(i)})^+\) is the pseudo-inverse matrix of \( R^{(i)} \).  If the index \( i \) is changed from one hour to another, then the time-variant coefficient matrix consisting of multiple \( a_i \) can be obtained.  It should be noticed that the LLR model becomes the AR model if the coefficients are time-invariant.  In other words, the AR model is a special case of the LLR model.

3.3 The RBFNN Model

RBFNNs are a class of feed-forward ANNs constructed based on the function approximation theory.  Fig. 2 shows the structure of an RBFNN.  It has three functionally distinct layers.  The input layer is simply a set of sensory units.  The second layer is a hidden layer of sufficient dimension, which performs a nonlinear transformation from the input space to a higher-dimensional hidden-unit space.  The third layer performs a linear transformation from the hidden-unit space to the output space.  The output of the RBFNN is given by:

\[
\hat{y} = \sum_{i=1}^{n} w_i \phi(x, c_i, \sigma_i) + w_0
\]

where \( n \) is the number of neurons (i.e., RBF units) in the hidden layer, \( w_0 \) is a bias term, \( w_i \) is the weight between the hidden and output layers, and \( \phi(\cdot) \) is the activation function in the hidden layer.  In this work, the function \( \phi(\cdot) \) is defined as:

\[
\phi(x, c_i, \sigma_i) = \exp \left( -\frac{\|x-c_i\|^2}{2\sigma_i^2} \right)
\]

where \( c_i \) and \( \sigma_i \) are the center and width of the RBF function, respectively.  The values of \( c_i \) and \( \sigma_i \) can be determined by different methods.  The simplest method is to randomly choose a subset of the data points as the RBF centers.  A more sophisticated approach is to cluster the data into an appropriate number of clusters, whose centers are then used as the centers of the RBF units.
In this project, a local Gaussian mixture model [42] with spherical covariance structure was created to determine the RBF centers by a \( K \)-means clustering algorithm [43]. The Gaussian mixture model was trained by using the Expectation Maximum (EM) algorithm [44]; the resulting centers were then transferred to the RBFNN.

It has been shown [45] that setting the widths of the RBF functions equal to the variances of the corresponding mixture model tends to give poor results, because the widths are too small and there is insufficient overlap between the RBF functions. In this work, all the widths are set at the same value, which is proportional to the maximum Euclidean distance, \( d_{\text{max}} \), between RBF centers.

\[
\sigma_i = k \cdot d_{\text{max}}
\]  

where \( k \) is a nonnegative scalar with the typical value in the range of \([0.1, 0.2]\) [46]. Given a data set \( X \), Eq. (6) can be further written as:

\[
\hat{Y} = \Phi \cdot W
\]  

where \( W = [w_0, w_1, \cdots, w_n] \) is the vector of the output weights and bias term, and \( \Phi \) is the matrix of hidden-layer activations due to the input data \( X \). A sum-of-squares error function is defined by:

\[
E = \frac{1}{2} \| \hat{Y} - Y \|^2
\]  

Since this error function is a quadratic function of the vector \( W \), pseudo-inverse can be used to determine the optimal \( W \) to minimize the value of the error function.

\[
W = \Phi^+ \cdot Y
\]  

where \( \Phi^+ = (\Phi^T \Phi)^{-1} \Phi^T \).

![Fig. 2. The structure of an RBFNN.](image)

### 3.4 The SVM Model

SVM belongs to the class of kernel methods. The use of SVM for time series prediction can be expressed as follows.
\[ \hat{y}_t = w^T \Phi(x_t) + b \]  

(12)

where \( \hat{y}_t \in R \) is the predicted value of the time series; \( x_t \in R^D \) is the input regression vector consisting of historical data of the time series and \( x_t = [y_{t-D}, y_{t-D+1}, \ldots, y_{t-1}]^T \); \( b \in R \) is a bias term; \( w \in R^M \) is the weight vector; and \( \Phi: R^D \rightarrow R^M (M \geq D) \) is a nonlinear feature map, which transforms the input vector \( x_t \in R^D \) to a higher-dimensional vector \( \Phi(x_t) \in R^M \). Fig. 3 shows the structure of an SVM, where \( x_{t,i} (i = 1, 2, \cdots, D) \) and \( \Phi_j(x_t) (j = 1, 2, \cdots, M) \) denote the \( i \)-th and \( j \)-th element of \( x_t \) and \( \Phi(x_t) \), respectively.

In an SVM, the historical data of the time series is mapped into a higher-dimensional feature space via a nonlinear mapping \( \Phi \); then linear regression is used in the high-dimensional feature space to train SVM and to predict the time series, which is equivalent to solving a nonlinear regression problem in the low-dimensional space of the original time series [17]. The key issue to solving such a prediction problem is to find the optimal values of the SVM parameters \( w \) and \( b \). This can be done by solving a constrained optimization problem.

\[
\begin{align*}
\min \quad & \frac{1}{2} w^T w + \gamma \sum_{i=1}^{N} e_i^2 \\
\text{s.t.} \quad & y_i = w^T \Phi(x_i) + b + e_i, \quad t = 1, 2, \ldots, N
\end{align*}
\]  

(13)

where \( y_t \) is the real value of \( \hat{y}_t \); \( e_t \) is the prediction error; \( \gamma \) is a regularization parameter, which balances the fitting in the training stage and generalization in the implementation stage. A too large or too small \( \gamma \) might deteriorate the generalization ability of the SVM in the implementation stage. Eq. (13) is solved by using Lagrange multipliers, and the solution is expressed in its dual form. Then the SVM of Eq. (12) can be represented by the following:

\[ \hat{y}(x_t) = \sum_{i=1}^{N} \alpha_i K(x_t, x_i) + b \]  

(14)

where \( \alpha_i (i = 1, \cdots, N) \) are the nonnegative Lagrange multipliers of Eq. (13); \( K(x_t, x_i) = \Phi(x_t) \Phi(x_i) \) are the positive-definite kernel functions. In SVM, all of the data samples are support vectors (SVs); the nonnegative multiplier \( \alpha_i \) represent the contribution of the SVs to the predicted value, namely, a larger \( \alpha_i \) indicates that its corresponding SV is more important.

![Fig. 3. The structure of an SVM.](image)
Commonly used kernel functions include linear, polynomial, and RBF kernels. An SVM with the following RBF kernel [47] is used in this work.

$$K(x_i,x_j) = \exp \left( -\frac{\|x_i - x_j\|^2}{\sigma^2} \right)$$  \hspace{1cm} (15)

where $\sigma$ is the width of the RBF kernel, which determines the influence area of the SVs over the data space. The resulting SVM is called RBF-SVM.

Based on the wavelet function in [22], a new wavelet mother function is proposed as follows.

$$h(x) = \cos \left( k \cdot \frac{x}{a} \right) \cdot \exp \left( -\frac{x^2}{a^2} \right)$$ \hspace{1cm} (16)

where $a$ is a parameter of the Gaussian kernel and $k$ is a new parameter which controls the kernel shape. Fig. 4 shows the proposed kernel function $h(x)$ when $k$ varies from 0 to 1.5 ($a = 1$). When $k = 0$, $h(x)$ is an RBF kernel; when $k = 1.5$, $h(x)$ approximates the Mexhat kernel in the range of [-1, 1]; when $k = 1.75$, $h(x)$ is exactly the kernel proposed by Szu [22]. The proposed wavelet kernel is then obtained by substituting Eq. (16) into the following function.

$$K(z_i,z'_j) = \prod_{i=1}^{D} h \left( \frac{z_i - z'_i}{a} \right)$$ \hspace{1cm} (17)

Equation (17) represents a multidimensional wavelet function. Substitute Eq. (17) into Eq. (14); then the SVM can be obtained:

$$\hat{y}(x_i) = \sum_{i=1}^{N} \alpha_i \prod_{j=1}^{D} h \left( \frac{x_{i,j} - x_{j,i}}{a_i} \right) + b$$ \hspace{1cm} (18)

Fig. 4. The proposed wavelet kernel.
where $x_{i,j}$ and $x_{i,j}$ denote the $j$th elements of $x_i$ and the $i$th training sample $x_i$. The SVM in Eq. (18) is called the WVSM because it uses wavelet functions as kernels; and the principle of wavelet analysis to approximate the time series in the wavelet kernel basis, where the wavelet coefficients are the nonnegative weights and bias in the WSVM. Therefore, finding the optimal weights and bias for the WSVM is equivalent to determining the wavelet coefficients in the kernel basis.

4. The WSVM-Based WPP Model

The proposed WPP model consists of three components, which are preprocessing, WSVM-based wind speed prediction, and wind speed-to-wind power conversion, as shown in Fig. 5. The preprocessing includes data normalization and feature representation. The WSVM plays a key role in the whole WPP model, on which the performance depends greatly. The output of the WSVM is wind speed, which is converted into wind power according to the power-wind speed curves of the wind turbine generators (WTGs).

Fig. 5. The structure of the proposed WSVM-based WPP model.

4.1 Data Normalization

To avoid tuning the SVM parameters in Eq. (14) due to large variations of the input variables (i.e., wind speeds), the inputs are normalized by using the sigmoid function.

$$y = \frac{1}{1 + e^{-\frac{y - \mu}{s}}}$$  \hspace{1cm} (19)

where $\mu = v_{\text{norm}}$ is the nominal wind speed; and $s = (v_{\text{cut-out}} + v_{\text{cut-in}})/3$ and $v_{\text{cut-out}}$ and $v_{\text{cut-in}}$ are the cut-in and cut-out wind speeds, respectively. There are two reasons of using the sigmoid function for data normalization. First, the sigmoid function can strictly map the original inputs (i.e., the real wind speeds) to the range of $[0, 1]$, as shown in Fig. 6, where the normalized values of 0.06 and 0.91 correspond to the original cut-in and cut-out wind speeds of 3.5 m/s and 25 m/s, respectively. Second, the normalization using $\mu$ and $s$ make the data translation, rotation, and scaling invariant.
4.2 Feature Representation

Feature representation, which aims to extract certain characteristics from the original data, plays a key role in determining the performance of WPP. Improper features obtained from bad feature extraction will lead to poor regression in the WSVM. In this work, wind speed is selected as an intermediate variable, which is predicted by the proposed WSVM algorithm. The predicted wind speed is then used to calculate the wind power according to the power-wind speed characteristics of the WTGs. The reason for using wind speed as an intermediate variable instead of predicting
wind power directly is that wind speed is a continuous variable while wind power discontinues at certain wind speeds (e.g., the cut-in, rated, and cut-off wind speeds). It is easier and more accurate to predict wind speed than wind power.

The WSVM input is expressed in time series form as $x_t = [y_{t-D}, y_{t-D+1}, \ldots, y_{t-1}]^T$, where $D$ is called an embedding dimension [18] and is determined from the autocorrelation coefficients of the data samples as follows:

$$r_k = \frac{1}{(N - k)s_y^2} \sum_{i=k}^{N} (y_i - \mu_y)(y_{i-k} - \mu_y)$$

where $\mu_y$ and $s_y$ are the mean and standard deviation of the training data. Fig. 7 shows the autocorrelations of the original data. As shown in Fig. 7, the original data is highly linearly correlated. Given a threshold $r_T$ of the autocorrelation coefficients, the embedding dimension can be determined. For example, if $r_T = 0.85$, then the former six samples are used as the input of the WSVM, (i.e., $x_t = [y_{t-6}, y_{t-5}, \ldots, y_{t-1}]^T$).

4.3 Wind Speed-to-Wind Power Conversion

According to the predicted wind speed, the wind power is obtained from the wind turbines’ power-wind speed characteristics. Fig. 8 shows a wind speed profile and the corresponding total power of 10 Vestas V-90 3-MW wind turbines obtained from the NREL database, where the cut-in and cut-out wind speeds make the wind power curve discontinuous although the wind speed curve is continuous.

The wind speed-to-wind power conversion should take into account the wind turbine hysteresis effect [48]. This effect occurs during the period between the shutdown and restart of a wind turbine. The former event can be triggered when one of the cut-out criteria is met, while the latter happens when the wind speed drops below a certain value. As shown in Fig. 8, the wind turbine is shut down at $9.405 \times 10^4$ minutes when the wind speed exceeds 30 m/s and is not turned on until the wind speed drops below 25 m/s. Once the wind speed is obtained, the following function is used to determine the wind power.

$$\hat{P}(\hat{v}) = \begin{cases} 0 & \hat{v} < v_{\text{cut\_in}} \\ C(\hat{v}) & v_{\text{cut\_in}} \leq \hat{v} \leq v_{\text{norm}} \\ P_{\text{norm}} & v_{\text{norm}} \leq \hat{v} \leq v_{\text{cut\_out}} \\ 0 & \hat{v} > v_{\text{cut\_out}} \end{cases}$$

where $\hat{v}$ is the predicted wind speed and $C(\cdot)$ is obtained from the wind turbine power-wind speed curve (or power curve).

The function $C(\cdot)$ can be determined from the power curves provided by the wind turbine manufacturers. However, recent research has shown that there are advantages to determining an equivalent power curve (EPC) from the measured wind speed and power [49]. In this work, the power curve is derived from the distribution of the data. Fig. 9 shows the process of generating the power curve. First, the wind speed data is allocated into multiple small intervals, where the length of each wind speed interval is 0.2 m/s, e.g., the interval [10.6, 10.8] m/s in Fig. 9(a). The mean $\mu_p$ and standard deviation $\sigma_p$ of the corresponding wind powers are then calculated for the
data in each wind speed interval. The wind speed samples the corresponding wind powers located far away from the center, (e.g, the samples (labeled as star) outside the range of $[\mu_P-\sigma_P, \mu_P+\sigma_P]$ in Fig. 9(b)) are discarded.

Second, for each wind speed interval after discarding the scattered data, the overall range of wind power is equally divided into 10 intervals. The power occurrence frequency $F_i$ (i.e, the number of samples in the $i^{th}$ power interval) are calculated for each power interval, where $i = 1, \ldots, 10$. The normalized $F_i$ is taken as the power occurrence probability $\rho_i$ (i.e., $\rho_i = F_i/\Sigma F_i$). Fig. 9(c) shows the values of $\rho_i$ ($i = 1, \ldots, 10$) in different intervals of wind power, which is approximately a normal distribution. The ten $\rho_i$s are then sorted in a decreasing order. Given a threshold $\rho_T$, the first $m$ ($m < 10$) intervals are selected according to the following criterion.

$$\sum_{i=1}^{m} \rho_i \geq \rho_T \quad (\rho_1 \geq \rho_2 \geq \ldots \geq \rho_m)$$

Third, only the data in the selected power intervals are used. As shown in Fig. 9(c), the samples distributed between the separation lines in Fig. 9(b) have high and low density values, respectively. Therefore, according to Eq. (22), the data with low densities in Fig. 9(c), which correspond to the data marked as dots in Fig. 9(b), are discarded; while the data marked as circles is selected. Finally, the wind power $p$ for the wind speed interval is calculated as:

$$p = \sum_{i=1}^{m} p_i \rho_i$$

where $p_i$ is the average wind power of the $i^{th}$ power interval. Equation (23) is used to calculate the wind power for each wind speed interval. The power curve can then be obtained by sliding the wind power over all wind speed intervals, as shown in Fig. 9(a).
4.4 Performance Evaluation

The Mean Absolute Error (MAE), Mean Absolute Percent Error (MAPE), and the standard deviation (Std) of MAE are used to measure the prediction performance. The definitions are expressed as:

\[
MAE = \frac{1}{N} \sum_{t=1}^{N} \left| \hat{p}_{t+h} - p_{t+h} \right| \tag{24}
\]

\[
MAPE = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{\hat{p}_{t+h} - p_{t+h}}{P_{\text{norm}}} \right| \tag{25}
\]

\[
Std = \sqrt{\frac{1}{N-1} \sum_{t=1}^{N} \left( \hat{p}_{t+h} - p_{t+h} - MAE \right)^2} \tag{26}
\]

where \( h \) is the predict horizon, \( P_{\text{norm}} \) is the nominal wind power, \( \hat{p}_{t+h} \) and \( p_{t+h} \) are the \( h \) step-ahead predicted and actual wind power, respectively. Small values of MAE and Std imply the superior prediction performance of the model.
5. Typical WPP Results

The Western Dataset [50] created by 3TIER with oversight and assistance from NREL was used to validate the proposed WPP model. In the Western Dataset, NWP models were used to essentially recreate the historical weather for the western U.S. for the years of 2004, 2005, and 2006. The modeled data was sampled every 10 minutes temporally and every 2 kilometers spatially. The power output of 10 wind turbines at 100 meters above ground level on each grid point was modeled by 3TIER using a technique called Statistical Correction to Output from a Record Extension (SCORE) [51], which replicates the stochastic nature of the wind plant output. The hysteresis effect of the wind turbines was modeled by NREL to further replicate the real operation of wind plants. The data includes wind speed, rated power, SCORE-lite power, etc. Sixty-eight wind turbines in a wind plant 10 miles west of Denver, Colorado, were selected for simulation studies. Each data sample contains the average values of the wind speed and power among the 68 wind turbines at the same time. The time interval between two nearest samples (called the time resolution) is 20 minutes.

The one hour (1h), two hour (2h), and three hour (3h)-ahead WPP results using the proposed WSVM-based model are shown in Figs. 10, 11, and 12, respectively, where the normalized error is defined as $\frac{\text{error}}{P_{\text{norm}}} \times 100\%$. In all cases, the predicted wind power follows closely the actual wind power. The normalized errors of most samples fall between -10% and 10%. More than 70% of the normalized errors are less than 5% in the case of 1h-ahead WPP. Approximately 60% and 50% normalized errors are less than 5% in 2h- and 3h-ahead predictions, respectively. A large error occurs when the wind speed changes drastically. Moreover, from the perspective of statistics, the larger the prediction horizon, the more uncorrelated data is used which leads to a larger prediction error. Therefore, the performance of the proposed model degrades with the increase of the prediction horizon.

![Fig. 10. One-hour-ahead WPP using the WSVM model and fixed-step scheme.](image-url)
The proposed WSVM-based model was compared with the persistence and RBF-SVM models to further evaluate its performance. The persistence model is a classical benchmark model in which the predicted values at any future times within the prediction horizon are set at the current value. The results are shown in Figs. 13 and 14. As shown in Fig. 13, both SVM-based models significantly outperform the persistence model in terms of prediction accuracy. Furthermore, the
WSVM model is always better than the RBF-SVM model. Fig. 14 compares the actual wind power with the predicted wind powers from the WSVM and RBF-SVM models. The predicted value using the WSVM model follows more closely the observation than that using the RBF-SVM model. The possible reason is that the proposed kernel is better than the RBF kernel.

![Graph comparing MAPE and prediction horizon for different models](image1)

**Fig. 13.** Comparison among the persistence, RBF-SVM and WSVM, models.

![Graph comparing 1h ahead prediction for two models](image2)

**Fig. 14.** Comparison between the RBF-SVM and WSVM models.
6. Data Representation and Preprocessing for SPP

The National Solar Radiation Database (NSRDB) [52] was used to validate the effectiveness of the proposed SPP methods. The NSRDB was produced by NREL in collaboration with other partners. The NSRDB contains 47 variables, including hourly solar radiation and other meteorological data for 1,454 locations in the United States. All the data was recorded from 1991 to 2005. In this report, the data from San Francisco (Station ID: 724940), Kansas City (Station ID: 724460), and Boston (Station ID: 725090) were selected for simulation studies. The San Francisco data is used in the following illustration.

6.1 2D Representation

To visualize the benefits of using 2D representation, one year of data (Jan. 1, 2004–Dec. 31, 2004) was first considered as a 1D time series and then as a 2D image formed in the raster scan form with the columns and rows corresponding to days and hours, respectively. Figs. 15 and 16 show the 1D and 2D representations of the solar radiation data, respectively.

In Fig. 15, it is visually difficult to grasp the solar radiation characteristics within a day although the seasonal behavior is obvious. In Fig. 16, daily and seasonal behavior of solar radiation can be easily interpreted, where a larger value in the range of [0, 1000] indicates stronger radiation. In winter, the dawn to dusk period is shorter than that of summer. While in summer, radiation at noon achieves the strongest of the whole year. Such a 2D representation provides significant insight into not only the radiation pattern as a function of time but also the horizontal and vertical correlations within the 2D data.

![Hourly solar radiation data in 1-D](image)

Fig. 15. Hourly solar radiation data in a 1D time plot.
6.2 Correlation Analysis

The embedding dimension of the input of the prediction model, i.e., the number of previous data samples used as input, is determined by the autocorrelation coefficients of the samples.

\[
r_k = \frac{1}{(N-k)s^2} \sum_{i=k}^{N} (x_i - \mu)(x_{i-k} - \mu)
\]

where \( \mu \) and \( s \) are the mean and variance of the samples, respectively, \( N \) is the number of samples of the series. Fig. 17 shows a 2D view of the autocorrelation coefficients of the solar radiation in 2004.

An important observation from Fig. 17 is that there are strong correlations between radiation not only in consecutive hours but also in some hours of consecutive days. The correlation between two consecutive days in the same hour is stronger than that between the current hour and two hours ahead of the same day. Therefore, when constructing a prediction model, the data from the previous day at the time of prediction, must be used with a higher priority than the data of the previous two hours. In this study, the former two days’ radiation data at the time of prediction and the data at the current time are used as the input for the prediction model.
6.3 Normalization

The data can be normalized to the range of \([0, 1]\) by using the sigmoid function in Eq. (19). Another method of data normalization is based on the concept of transmissivity [37], which is defined as the ratio between the radiation received on the ground surface and the incoming radiation (extraterrestrial radiation) at the top of atmosphere.

\[
\tau = \frac{R_g}{R_e}
\]  

(28)

where \(\tau\) is the transmissivity, \(R_g\) and \(R_e\) are the ground radiation and extraterrestrial radiation, respectively. The extraterrestrial solar radiation \(R_e\) can be accurately estimated using geometry factors (latitude and longitude), day of the year (DOY), and time of the day (TOD). Therefore, the actual ground radiation can be derived if the transmissivity is known.

The transmissivity takes time variations into account. Therefore, \(\tau\) not only reflects the radiation but also contains certain weather information. A larger \(\tau\) is equivalent to a clearer sky, which plays a key role in solar radiation. Due to its physical meaning, the normalization by transmissivity is superior to that by the sigmoid function. Fig. 18 compares the two methods of normalization for radiation data. Fig. 18(a) is the original radiation on May 28 and December 10, 2005, where the \(R_e\) curve indicates seasonal variations of the solar radiation. The radiation on May 28 is much stronger than that on December 10. The ground radiation \(R_g\) curve reflects the effect of the weather conditions on solar radiation. For example, May 28 could not be a clear day; otherwise, the ground radiation on May 28 should be much larger than that on December 10. Fig. 18(b) shows the normalized values of ground radiation by using the two methods. The sigmoid normalized ground radiation values \((y_n)\) in both days are similar, which fails to “discover” weather difference. The transmissivity helps “recognize” the weather condition,
which plays an important role in SPP. Therefore, in this report, the transmissivity is used for normalization of the radiation data, while other variables are normalized by the sigmoid function.

![Comparison of two methods for normalization of radiation data](image)

**Fig. 18.** Comparison of the two methods for normalization of the radiation data.

### 6.4 Performance Evaluation

The mean absolute error (MAE), mean absolute percentage error (MAPE), coefficient of determination ($R^2$), and correlation coefficient ($\rho$) are used to evaluate the performance of the SPP models. Definitions are expressed as follows.

\[
MAE = \frac{1}{N} \sum_{i=1}^{N} |\hat{y}_i - y_i| \tag{29}
\]

\[
MAPE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{\hat{y}_i - y_i}{y_i} \right| \tag{30}
\]

\[
R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2} \tag{31}
\]

\[
\rho = \frac{\sum_{i=1}^{N} (\hat{y}_i - \bar{y})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2} \sqrt{\sum_{i=1}^{N} (y_i - \bar{y})^2}} \tag{32}
\]

where $y_i$ and $\hat{y}_i$ are the observation and the predicted value, respectively, and $\bar{y}$ and $\bar{\hat{y}}$ represent the mean values of the observation and prediction, respectively. Smaller values of the MAE and
MAPE imply a superior prediction performance of the model. $R^2$ is a measure of the global fit of the model, and $\rho$ is a measure of linear correlation between two variables.

In order to evaluate the improvement of one model to another, a parameter called skill is defined as follows:

$$skill = \frac{|e_1 - e_0|}{e_0} \times 100\%$$

(33)

where $e_1$ and $e_0$ are the MAE of the SPP using a new model and the reference model, respectively. A larger skill value indicates more superiority of the new model.

7. Typical SPP Results

Simulations were carried out for short-term SPP using the NSRDB. In this study, the testing set was selected as the data from September 1, 2005, to September 10, 2005, which has a moderate number of sunny and cloudy days. The training set contains the data from July 17 to October 20 in previous years and from July 17 to August 31 in 2005. The inputs of the SPP models include the latest observed solar radiation, radiation at the hour of prediction in the previous two consecutive days, and the latest meteorological features, including sky cover, wind speed, and relative humidity. In addition, since there is no radiation at night, only the observations from 5 a.m. to 9 p.m. were used. For a one-hour prediction, the radiation from 6 a.m. to 9 p.m. in a day was predicted. During testing, all of the predicted values were true out-of-sample forecasts, in which only the historical data samples were used. The predicted data was then compared to the actual measured value. The procedure was repeated for the next time step until it ran over the entire testing dataset.

Fig. 19 and 20 show the one-hour-ahead prediction results in Denver using the SVM model, where the normalized error is defined as $(1 - \hat{y}_t/y) \times 100\%$. As shown in Fig. 19, the SVM model works well especially during clear days or totally overcast days (the 45th–105th hours, 15th –30th hours), where the predicted values closely follow the observations. The prediction errors were relatively small, where the maximal error is 100 W/m². Large prediction errors mainly occur in those days when the ground radiation was drastically changed. For instance, the weather conditions during the 30th–45th and the 105th–120th hours make the prediction less accurate than in the 45th–105th hours. The error distribution, however, shows that the majority of prediction errors concentrate in a small range. More than 70% of the normalized errors are less than 10% in the case of one hour-ahead SPP. Fig. 20 shows the collaboration between the real and predicted solar radiation in Denver. As mentioned previously, $\rho$ is a measure of linear correlation between two variables. Since $|\rho| \leq 1$, $\rho = 0.97$ corresponds to the slope of the fitting line (1.03 in this case), which is close to 45 degrees. Therefore, the predicted values closely match the actual data along the diagonal axis, which indicates a successful prediction. Fig. 21 shows the one-hour-ahead prediction in Seattle by using the SVM model. Similarly, more than 50% normalized error is less than 10%. It also shows an accurate prediction during clear or totally cloudy days.
Fig. 19. One-hour-ahead prediction in Denver using SVM.

Fig. 20. Correlation between the real and predicted solar radiation in Denver.
The AR model, which turned out to be superior to the persistence model, was used as the reference model in this work. Furthermore, the SVM-based model is also compared to the RBFNN-based model. Fig. 22 compares AR, RBFNN, and SVM-based prediction models using the data in Denver and Seattle. As shown in Fig. 22, RBFNN and SVM achieved much better results than the AR model in terms of accuracy. In Denver’s one-hour-ahead prediction, the MAE value obtained by the AR model was 62W/m²; while the MAE values predicted with the RBFNN and SVM were 43W/m² and 33.7W/m², respectively. The inferiority of the AR model is largely due to the failure of the linear model to capture the nonlinear characteristic of the solar radiation. Moreover, it should be mentioned that the predicted accuracy improvement brought by the proposed SVM-based model, as illustrated in Fig. 22, is indeed better than the RBFNN model. Fig. 23 compares the predicted values among the three models with the observations during two consecutive days in Seattle (75th–105th hours in Fig. 21). Likewise, the SVM-based prediction value follows much closer to the observation value than either the AR model or the RBFNN model does. For example, on September 6, 2005, the prediction value obtained from the SVM-based model quickly captures the change trend in the morning (i.e., 22th–28th hours); it achieves smaller prediction errors than the other two models.

It is turned out that using the SVM-based prediction model achieved less prediction errors than RBFNN. SVM is superior to RBFNN in terms of generalization ability, which is determined by two aspects. One is the objective functions. SVM not only includes the error term that RBFNN has but also has an extra weights term, as shown in Eq. (13), that is designed to balance the training and fitting errors. The other aspect is the way of optimization. As the SVM approach automatically solves the neural network complexity problem, the size of the hidden layer is obtained as a result of the quadratic programming; while in RBFNN, the number of hidden neurons, which is an important parameter, should be provided before training.
Fig. 22. Comparison of the MAEs and MAPEs of the AR, RBFNN, and SVM-based prediction models in Denver and Seattle.

Fig. 23. Comparison of the one-hour-ahead predicted values from the three models with the observations in two consecutive days in Seattle.
8. Conclusion

This report has presented several statistical time-series prediction models, including the AR, LLR, RBFNN, RBF-SVM, and WSVM models, for wind and solar power prediction. These models have been evaluated by using the data obtained from NREL. Results have shown that SVM models are preferable for wind and solar power prediction in this project. The predicted wind and solar powers will provide necessary information for optimal power management of the EPRTL units and the overall RHPS microgrid to be developed in this project.

References


