Short-Term Wind Energy Forecasting Using Support Vector Regression

Oliver Kramer, Fabian Gieseke

Abstract Wind energy prediction has an important part to play in a smart energy grid for load balancing and capacity planning. In this paper we explore, if wind measurements based on the existing infrastructure of windmills in neighbored wind parks can be learned with a soft computing approach for wind energy prediction in the ten-minute to six-hour range. For this sake we employ Support Vector Regression (SVR) for time series forecasting, and run experimental analyses on real-world wind data from the NREL western wind resource dataset. In the experimental part of the paper we concentrate on loss function parameterization of SVR. We try to answer how far ahead a reliable wind forecast is possible, and how much information from the past is necessary. We demonstrate the capabilities of SVR-based wind energy forecast on the micro-scale level of one wind grid point, and on the larger scale of a whole wind park.

1 Introduction

Wind energy forecasting is an important aspect for balancing authorities in a smart grid. Up to now, the integration of decentralized energy into the grid is as good as ignored. It is estimated that the stability of the energy grid decreases, if the amount of ignored renewable energy exceeds about 15% to 20%. But wind resources are steadily increasing. For a reasonable integration of volatile resources like wind, a precise prediction for subhourly scheduling becomes necessary. Precise forecast will allow balancing and integrating of multiple volatile power sources at all levels of the transmission and distribution grid [10]. Soft computing can play an important role

Oliver Kramer, ICSI Berkeley, 1947 Center St,
Berkeley CA 94704, USA, e-mail: okramer@icsi.berkeley.edu

Fabian Gieseke, TU Dortmund, Otto-Hahn-Str. 14,
44227 Dortmund, Germany, e-mail: fabian.gieseke@tu-dortmund.de

in forecasting, and control of smart grids. They have proven success in many applications, e.g., the visualization of network traffic data for intrusion detection with neural techniques [5]. Also in energy and environmental sciences soft computing methods have successfully been applied, ranging from detection of lifetime building thermal insulation failures [14] to the identification of typical meteorological days [2].

State-of-the-art techniques in regression have already been applied to energy forecasting. But the results are often limited to simplified case-studies of particular windmills, neglecting parameter studies, or analyses of how far a regression method can reliably predict wind on a short-term level. In this paper we basically investigate the questions (1) can prediction of wind energy exclusively be based on the existing infrastructure of windmills and their wind speed measurements, and (2) what are the limitations of state-of-the-art regression techniques for wind resource time series forecasting. To answer these question we will conduct experiments based on real-world wind data from the western wind data resource of the National Renewable Energy Laboratory (NREL) [7, 13] employing a state-of-the art kernel regression method: support vector regression (SVR) by Vapnik [18]. Our analysis will be based on a direct mapping of wind speed measurements on produced wind energy.

Section 2 formalizes the regression problem, and illustrates the data scenario we plan to investigate. In Section 3 we will give a short overview of related work on wind resource forecasting, while Section 4 gives a brief introduction to SVR. Section 5 presents an experimental analysis of SVR on different data scenarios that help to understand the capabilities of SVR in the wind forecast scenario. The analysis concentrates on the choice of the loss parameter ε, and the question how far into the future predictions are possible, and how much data are necessary from the past. In Section 6 we summarize the results and discuss prospective research questions.

2 Problem Description

2.1 Formalization

We formulate the wind forecasting task as regression problem. We assume that a time series of N wind measurements of K wind grid points $\mathbf{x}(t) = (x_1(t), \dots, x_K(t))^T$ with time t and $1 \leq t \leq N$ is given, complemented by corresponding measurements $\mathbf{y}(t) = (y_1(t), \dots, y_K(t))^T$ of wind production. The task is to predict the wind production $\mathbf{y}_t$ at time $t = t_i + \theta$ based on the wind measurements at time $t_i, t_i - 1, t_i - 2, \dots, t_i - \mu$, with $\mu \in \mathbb{N}$ past observations. The following questions arise:

- how much data from the past do we need (i.e., how to choose μ to reduce the validation error),
- how far can we look into the future (i.e., how does the validation error depend on θ), and

- how many windmills do we need for reliable prediction (i.e., how to choose K, and where do the K windmills have to be located for an optimal prediction).

In this work we concentrate on the production of a single windmill in Section 5.2, and on the large-scale level of a whole wind park in Section 5.3. The third question, how many, and *which* windmills to select in the optimal case, will be subject to future work.

2.2 NREL Data

The data that are basis of our analysis are taken form the NREL western wind resources dataset [7, 13]. The western wind resources dataset is part of the Western Wind and Solar Integration Study, which is a large regional wind and solar integration study in the US. It was partly created with the help of numerical weather predictions. The data were sampled every ten minutes and every two kilometers. About 1.2 million grid points have been aggregated to 32,043 locations. Each grid point is estimated to hold ten Vestas 3 MW turbines, and therefore the 32,043 locations in total exhibit more than 960 GW of capacity. The set contains data of 2004, 2005 and 2006. Potter *et al.* [13] describe how the data for the Western Wind and Solar Integration Study have been created. The data have been measured every ten minutes, resulting in 52,560 measurements a year.

3 Related Work

Wind forecasting is an important task, and different approaches are known in literature. Costa *et al.* [3] review 30 years of short-term prediction concentrating on forecasting methods, mathematical, statistical and physical models, as well as meteorology. Negnevitsky *et al.* [12] review forecasting techniques used for power system applications with focus on electricity load, price forecasting and wind power prediction. They classify methods based on time frames, application specific areas and forecasting techniques. Milligan *et al.* [10] discuss, if wind is a capacity resource. They state that aggregation over a 750-km region leads to a reduction of the wind energy forecasting error by about 50%. Furthermore, they state that for a single wind power plant, predictions on a one- or two-hour basis can achieve an accuracy level of approximately 57% mean absolute error to installed wind capacity, increasing to 20% for day-ahead forecasts.

Machine learning approaches are successful methods for wind forecasting based on past observations. As an overview of all methods is not the scope of this paper, we restrict our overview to selected methods that are closely related to our approach. Many methods are based on neural networks. Shuhui Li *et al.* [9] estimate the wind energy production of single wind turbines at central and South West Services Fort Davis. They discuss the structure and number of neurons in a multi-layer perceptron for turbine power production of single windmills. Gong Li *et al.* [8] have introduced a robust two-step approach based on a Bayesian combination of three neural networks (e.g., backpropagation, and radial basis functions networks). They demonstrate the approach for a one-hour forecast of two wind sites in North Dakota. Preliminary work on SVR, and wind forecasting has recently been introduced. Mohandes *et al.* [11] compared an SVR approach for wind speed prediction to a multi-layer perceptron. The approach is based on mean daily wind speed data from Saudi Arabia. Shi *et al.* [15] proposed an approach that combines an evolutionary algorithm for parameter tuning with SVR-based prediction. The technique allows a six-hour prediction, and is experimentally evaluated on wind data from North China. Recently, Zhao *et al.* [19] compared SVR to backpropagation for a ten-minute prediction of wind speed. Further work concentrates on special aspects like prediction and diagnosis of wind turbine faults. Kusiak and Li [6] introduced an approach based on fault prediction on three levels, e.g., fault category and specific fault prediction in a five-minute to one-hour approach.

4 Support Vector Regression

As mentioned above, we make use of the support vector regression (SVR) [17, 18] model to address our regressions tasks. The approach is one of the state-of-the-art methods in regression. The goal of the learning process is to find a prediction function $\hat{f}: \mathscr{X} \to \mathbb{R}$ that assigns "good" predictions to unseen $x \in \mathscr{X}$ (e.g., $\mathscr{X} = \mathbb{R}^d$). Here, we only sketch the key ideas of this concept and refer to, e.g, Smola and Schölkopf [16] for a comprehensive overview. The SVR technique can be seen as a special case of regularization problems of the form

$$\inf_{f \in \mathscr{H}} \frac{1}{n} \sum_{i=1}^{n} L\big(y_i, f(\mathbf{x}_i)\big) + \lambda ||f||^2_{\mathscr{H}}, \tag{1}$$

where $\lambda > 0$ is a fixed user-defined real value, $L : \mathbb{R} \times \mathbb{R} \to [0, \infty)$ is a loss function and $||f||^2_{\mathscr{H}}$ is the squared norm in a so-called reproducing kernel Hilbert space $\mathscr{H} \subseteq \mathbb{R}^{\mathscr{X}} = \{f : \mathscr{X} \to \mathbb{R}\}$ induced by an associated kernel function $k : \mathscr{X} \times \mathscr{X} \to \mathbb{R}$ (which can be seen as similiarity measure between the patterns). Plugging in different loss functions leads to different (but related) regression models. The so-called ε-insensitive loss $L_\varepsilon(y,t) = \max(|t-y| - \varepsilon, 0)$ with $\varepsilon > 0$ leads to

$$\inf_{f \in \mathscr{H}} \frac{1}{n} \sum_{i=1}^{n} \max\big(|f(\mathbf{x}_i) - y_i| - \varepsilon, 0\big) + \lambda ||f||^2_{\mathscr{H}}, \tag{2}$$

and, hence, to the SVR approach.[1] Here, the first term corresponds to the "difference" between the values predicted by the function (i.e., model) and the correspond-

[1] Note that in the latter formulation, the offset b is omitted for simplicity.

ing real values given in the training set (residuals). The second term corresponds to the "complexity" of the model. Ideally, one would like to have a model that fits the data well, and that is not too "complex" at the same time to avoid overfitting.

The ε-insensitive loss does not take into account small residual errors. The choice of ε defines the magnitude of errors that can be neglected. We will experimentally analyze various settings for ε in Section 5.1. Additionally, we will consider the square-loss $L_2(y,t) = (y-t)^2$ instead of the ε-insensitive loss for our experimental evaluation.

5 Experimental Analysis

In the following, we will experimentally analyze forecasts with SVR based on the NREL western wind resources datasets. The analysis concentrates on wind grid points in the wind park of Tehachapi in California, USA. We employ the following experimental settings. The SVR is trained on $1/10$-th of the observations from 2006. As core of the SVR implementation we employ is LIBSVM [1]. In the experiments we make use of an RBF-kernel with kernel width σ. Furthermore, we employ grid search in the parameter space of λ, and σ of the RBF-kernel. Grid search makes use of a test dataset based on the second $1/10$-th of the one-year data, and tests the following values: 2^α with $\alpha = -15, \ldots, 15$. For time-critical applications we recommend to narrow the grid search bounds, as the successful parameters often lie in the range between $\sigma = 2^\alpha$ with $\alpha = -10, \ldots, -5$, and $\lambda = 2^\alpha$ with $\alpha = 5, \ldots, 10$ for the NREL wind data. The final validation error is computed based on the second $1/5$-th of the corresponding datasets, using the L_ε, and the square-loss L_2.

5.1 Loss Function Parameter Study

We start the analysis with tests of different loss function parameters for the SVR training process. The results will determine the choice of the ε-value in the remainder of this work. Table 1 shows the analysis of five values for ε that determine the magnitude of residual errors not contributing to the overall error during training.

Table 1 Analysis of loss function parameter ε on the validation error measures with L_ε and L_2 loss.

loss	0.01	0.1	0.5	1.0	2.0
L_ε	2.128	2.046	1.795	1.538	1.188
L_2	15.013	14.984	**14.365**	14.571	15.383

For comparison we state the L_ε, and the L_2 loss on the validation set. The experiments are based on a 30-minute forecast of wind based on two time steps (data of the last 20 min) from the past measurements of 15 wind grid points. The forecast is computed for the energy production of one wind grid point in the middle of the Tehachapi wind park. The results show that – as expected – the L_ε-error decreases with increasing tolerance threshold ε. But the L_2 loss has a minimum at $\varepsilon = 0.5$. We assume that this setting is a reasonable choice for the following experiments.

5.2 Small-Scale Analysis: Wind Grid Point Level

The question is how far we can look into the future, and how much information from the past is necessary for a reliable forecast. Intuitively, we would expect that a static shot of the wind situation results in a loss of information, as no development, e.g., no change with regard to successive time steps is put into the model. Nevertheless, this intuition can be misleading as hidden correlations and dependencies may exist (e.g. relations like "strong wind measured by a northern windmill, and weak wind by a southern means that the wind comes from the north"). In the following, we do not rely on any assumption. We analyze the influence of the number of past time steps on the prediction error for an increasing number of steps we look ahead.

Table 2 Forecasts for a single wind grid point in Tehachapi based on wind measurements of 15 grid points of Tehachapi and neighbored parks within a range of ca. 50 miles. The figures show the validation error with regard to increasing steps into the future (lines, top to bottom), and an increasing number of past measurements (columns, left to right).

	1		2		3		6		12	
steps	L_ε	L_2	L_ε	L_2	L_ε	L_2	L_ε	L_2	L_ε	L_2
1	1.734	15.040	1.679	13.526	1.714	15.384	1.690	13.558	1.807	13.592
2	1.765	14.654	1.767	15.698	1.797	16.022	1.798	14.790	1.860	14.193
3	1.869	17.128	1.868	16.605	1.823	15.571	1.919	16.414	1.955	15.903
6	2.220	20.526	2.149	18.836	2.233	19.996	2.248	19.185	2.259	18.852
12	2.984	30.821	2.884	28.675	2.838	28.798	2.865	27.688	2.814	26.628

Table 2 shows the validation error for the energy forecast of a wind grid point in Tehachapi. It is based on 15 grid points from Tehachapi and neighbored wind parks within the range of about 50 miles. The figures show the validation error, i.e., L_ε- and L_2-loss on the validation set. From top to bottom the lines show predictions going further into the future. From left to right the figures show predictions that take more past time steps into account. One time step corresponds to ten minutes. The results show that the error is increasing the further we try to predict the future energy production. L_ε- and L_2-loss are strongly correlated. Furthermore, the figures confirm a trend that is consistent with our expectations: the more past is taken into account the better the predictions become.

Forecasts of sudden changes, e.g., caused by storm fronts passing, belong to the most important aspects. As a measure for the quality of forecasts is no easy undertaking, we employ a visual interpretation of two typical forecasts in the following. Figure 1 shows two randomly chosen wind time series from 2006 that are not basis of the training and testing process. The plots show the actual wind (blue/solid lines), and the forecasts based on a trained SVR model. Both plots on the left show the ten-minute forecasts, the plots on the right show the two-hour forecasts. Red (dark dotted) lines show the forecast based on the data from the last two hours (i.e., based on $12 \cdot 15$-dimensional vectors), while green (bright dotted) lines show the forecasts only based on the last measurements ten minutes ago (i.e., based on 15-dimensional vectors). In both situations we can observe that the ten-minute ahead forecasts lead to very accurate results. In particular the forecast based on the last ten minutes leads to a reliable prediction. More deviations from the true curve can be observed, if we use the last two hours for predictions. It is known that too much additional data can act like noise and disturb the prediction [4]. The situation changes on the two-hour level, where the forecast based on wind measurements from the last two hours leads to a higher accuracy. The forecast based on the ten-minute level is much less reliable and leads to larger deviations.

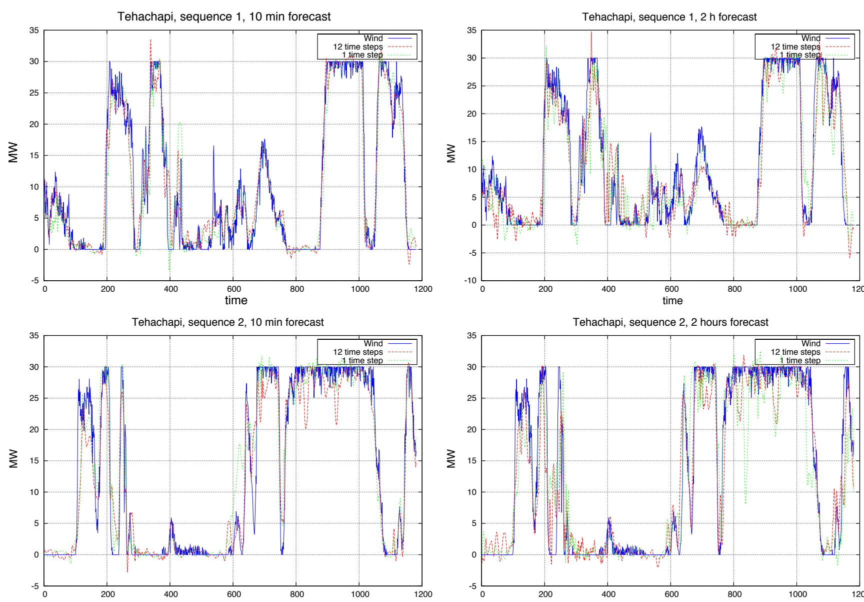

Fig. 1 Ten-minute (left figures) and two-hour (right figures) ahead wind energy forecasts for one Tehachapi wind grid point. Each figure shows the forecast based on ten minutes, and two hours of past wind measurements.

5.3 Large-Scale Analysis: Wind Park Level

Large-scale forecasting on the level of wind parks has an important part to play for global control strategies. Besides the approach to aggregate the forecasts of all windmills of the whole park, the sum of energy can be taken into account. In the following, we conduct the prediction analysis on the level of a wind park near Salt Lake City that consists of 28 wind grid points. For the forecasts we employ a set of 100 randomly chosen wind grid points in the whole western area.

Table 3 shows the experimental results of the analysis with regard to various combinations of time steps from the past ([1,3]: ten and 30 minutes, [3,6]: 30 and 60 minutes, and [6,12]: 60 and 120 minutes), and the steps we try to look into the future (from ten minutes to six hours). Similar to the previous section, the results show the corresponding validation error. Based on these figures we can observe the trend that the best forecast is achieved for the experiments predicting one hour ahead. Looking further into the future decreases the forecasts, but still results in an acceptable validation error. Short-term forecasts do also not result in the best validation errors. This is probably due to the fact that most of the windmills used for prediction are too far away to determine ten-minute forecasts. They are spread across the whole western area of the US. For the one-hour ahead forecast the past information from the last ten minutes and 30 minutes achieves the best validation error. But employing other combinations of time steps does not deteriorate the results significantly.

Table 3 Forecast of wind energy production of wind park near Salt Lake City. The figures show the validation error for increasing time steps (lines, top to bottom) with regard to various time steps from the past (columns) taken into account.

	[1,3]		[3,6]		[6,12]	
steps	L_ε	L_2	L_ε	L_2	L_ε	L_2
1	57.549	9,044.233	57.218	9,271.327	58.313	9,148.557
6	58.786	9,932.734	58.047	9,355.095	57.745	9,433.448
12	56.113	8,774.924	56.879	8,899.538	56.649	8,822.972
24	58.448	9,250.796	57.700	8,965.454	56.869	8,804.929
36	58.598	9,599.905	59.171	9,436.259	58.992	9,968.387

Figure 2 shows a visualization of two random sequences and the corresponding ten-minute, and six-hour ahead forecasts. The curves show the real wind that was blowing, and the forecasts, each based on two past time steps. The plots show that all forecasts achieve a relatively high prediction accuracy that should be satisfying for most balancing activities in a smart grid. The predictions based on the last two hours are even more reliable based on a ten-minute forecast than the predictions based on the last 30 minutes. Also for the six-hour ahead forecast the prediction based on the [6,12]-dataset results in the best curve. Local deviations from the true curve are more frequent in the case of the [1,3]-dataset forecast.

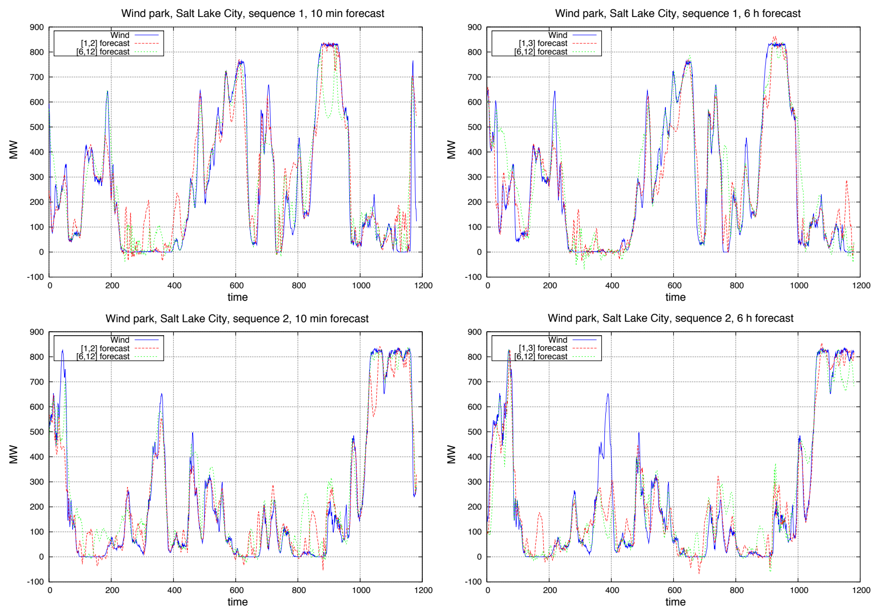

Fig. 2 Ten-minute (left figures) and six-hour (right figures) ahead forecasts for a wind park south-east of Salt Lake City on two randomly selected sequences (upper part and lower part). Also in case of the six-hour forecast the achieved accuracy is satisfactory.

6 Conclusion

Wind production forecasting is an important aspect for a stable grid. The integrity and stability can be improved the better a forecast of volatile energy sources is possible. We have demonstrated that SVR is a successful method for the prediction of wind energy production only based on wind measurements from windmills, in particular without further meteorological data or weather forecasts. SVR turns out to be a fast and robust time series prediction technique. For the wind resource scenarios we have found recommendable parameters in case the ε-loss is employed. For fast training of the SVR model, typical bounds can be identified, and the region of interest for grid search can be narrowed. The experiments have shown that a reliable forecast one the level of grid points is possible on the two-hour level, while the ten-minute prediction leads to almost exact results. On the level of a whole wind park, the results have shown that even a reasonable six-hour forecast is possible.

As a next step we plan to identify relevant prediction wind spots in a feature selection approach. For this sake we plan to employ evolution strategies. Energy production forecasting is not only necessary for wind data. In the future, we will try to extend the prediction to solar energy, and to prediction of energy consumption on the demand side.

References

1. C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.
2. E. Corchado, A. Árroyo, and V. Tricio. Soft computing models to identify typical meteorological days. In *Logic Journal of the IGPL*. Oxford University Press, 2010.
3. A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa. A review on the young history of the wind power short-term prediction. *Renewable and Sustainable Energy Reviews*, 12(6):1725–1744, August 2008.
4. P. F. Evangelista, M. J. Embrechts, and B. K. Szymanski. Taming the curse of dimensionality in kernels and novelty detection. In *Applied Soft Computing Technologies: The Challenge of Complexity*, pages 431–444. Springer Verlag, 2006.
5. Á. Herrero, E. Corchado, P. Gastaldo, and R. Zunino. Neural projection techniques for the visual inspection of network traffic. *Neurocomputing*, 72(16-18):3649–3658, 2009.
6. A. Kusiak and W. Li. The prediction and diagnosis of wind turbine faults. *Renewable Energy*, 36(1):16–23, 2011.
7. D. Lew, M. Milligan, G. Jordan, L. Freeman, N. Miller, K. Clark, and R. Piwko. How do wind and solar power affect grid operations: The western wind and solar integration study. In *8th International Workshop on Large Scale Integration of Wind Power and on Transmission Networks for Offshore Wind Farms*, 2009.
8. G. Li, J. Shi, and J. Zhou. Bayesian adaptive combination of short-term wind speed forecasts from neural network models. *Renewable Energy*, 36(1):352–359, January 2011.
9. S. Li, D. C. Wunsch, E. A. Ohair, and M. G. Giesselmann. Using neural networks to estimate wind turbine. *Journal of Guidance, Control, and Dynamics*, 16(3):276–282, 2001.
10. M. Milligan, K. Porter, E. DeMeo, P. Denholm, H. Holttinen, B. Kirby, N. Mille, A. Mills, M. OMalley, M. Schuerger, and L. Soder. Wind power myths debunked. *IEEE Power and Energy Society*, February, 2009.
11. M. Mohandes, T. Halawani, S. Rehman, and A. A. Hussain. Support vector machines for wind speed prediction. *Renewable Energy*, 29(6):939–947, May 2004.
12. M. Negnevitsky, P. Mandal, and A. Srivastava. Machine learning applications for load, price and wind power prediction in power systems. In *Intelligent System Applications to Power Systems (ISAP)*, pages 1–6, 2009.
13. C. W. Potter, D. Lew, J. McCaa, S. Cheng, S. Eichelberger, and E. Grimit. Creating the dataset for the western wind and solar integration study (u.s.a.). In *7th International Workshop on Large Scale Integration of Wind Power and on Transmission Networks for Offshore Wind Farms*, 2008.
14. J. Sedano, L. Curiel, E. Corchado, E. de la Cal, and J. R. Villar. A soft computing method for detecting lifetime building thermal insulation failures. *Integrated Computer-Aided Engineering*, 17:103–115, April 2010.
15. J. Shi, Y. Yang, P. Wang, Y. Liu, and S. Han. Genetic algorithm-piecewise support vector machine model for short term wind power prediction. In *Proceedings of the 8th World Congress on Intelligent Control and Automation*, pages 2254–2258, 2010.
16. A. J. Smola, B. Schölkopf, and B. Scholkopf. A tutorial on support vector regression, 1998.
17. I. Steinwart and A. Christmann. *Support Vector Machines*. Springer-Verlag, New York, NY, USA, 2008.
18. V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, New York, 1995.
19. P. Zhao, J. Xia, Y. Dai, and J. He. Wind speed prediction using support vector regression. In *Industrial Electronics and Applications (ICIEA)*, pages 882 – 886, 2010.