Support Vector Machines for Wind Energy Prediction in Smart Grids

Oliver Kramer[1], Nils André Treiber[1], Fabian Gieseke[1]

Abstract

In recent years, there has been a significant increase in energy produced by sustainable resources like wind- and solar power plants. This led to a shift from traditional energy systems to so-called smart grids (i.e., distributed systems of energy suppliers and consumers). While the sustainable energy resources are very appealing from an environmental point of view, their volatileness renders the integration into the overall energy system difficult. For this reason, short-term wind and solar energy prediction systems are essential for balance authorities to schedule spinning reserves and reserve energy. In this chapter, we build upon our previous work and provide a detailed practical analysis of several wind energy learning scenarios. Our approach makes use of support vector regression models, one of the state-of-the art techniques in the field of machine learning, to build effective predictors for single wind turbines based on data given for neighbored turbines.

1. Introduction

For the integration of wind, and photovoltaic power, a precise forecast of energy has an important part to play. Up to now, the integration of decentralized energy into the electricity grid is often ignored. Further, it is estimated that the grid gets unstable, if the amount of integrated renewable energy exceeds about 15 to 20%. Since the amount of integrated wind and solar energy resources is steadily increasing, a precise prediction for subhourly scheduling becomes necessary for a meaningful integration of such resources into the grid. Effective forecast systems will allow balancing and integrating multiple volatile power sources at all levels of the transmission and distribution grid (Milligan et al. 2009). The increasing number of sensors installed in wind turbines, and wind farms allows their detailed observation, and monitoring. The new data can be used for short-term prediction of wind energy. However, the efficient extraction of meaningful patterns, and the learning process of large wind energy data sets requires the application of powerful data analysis tools.

Among the most prominent techniques are the so-called support vector machines (Smola and Schölkopf 1998, Schölkopf and Smola 2001), which depict one of the state-of-the-art schemes for learning tasks like classification (i.e., automatic partition of objects into classes) and regression tasks (i.e., the prediction of a real value) (Hastie et al. 2009). The goal of the corresponding schemes is to generate high-quality models that can make reasonable predictions for new patterns, based on the knowledge given by the observed training data. In this chapter, we consider support vector machines to address the task of predicting the wind energy of single wind turbines based on time-series data that is available for neighbored turbines, and provide a detailed analysis of several learning scenarios, thus extending our previous work (Kramer and Gieseke 2011) in this context. Our main concern are short-term prediction systems that can provide valuable estimates for time-horizons up to 20, 60, and 120 minutes, respectively.

The chapter is organized as follows: After providing the related work in Section 2, we describe the wind

[1] Universität Oldenburg, Fakultät II, Uhlhornsweg 84, D-26111 Oldenburg
oliver.kramer/nils.andre.treiber/fabian.gieseke@uni-oldenburg.de

data considered for the evaluation in Section 3, followed by Section 4 that provides the related machine learning background. In Section 5, the experimental evaluation is given that encompasses several learning settings induced by a selection of single wind turbines. Conclusions are drawn in Section 6.

2. Related Work

Most wind prediction systems employ numerical weather predictions (Ernst et al. 2007), and models that map wind speed to electrical power output. The accuracy of the prediction can significantly be increased in case the models are combined to ensembles. Wind, and photovoltaic prediction play a very important role for balancing, and grid operations (Ernst et al. 2009). Costa et al. (2008) give a survey of 30 years of short-term prediction differentiating between mathematical, statistical and physical models. Negnevitsky et al. (2009) review forecasting techniques used for power system applications with a focus on electricity load, price forecasting and wind power prediction. Milligan et al. (2009) discuss the capacity property of wind energy. For a single wind turbine, predictions on one or two hours basis can achieve an accuracy level of approximately 5-7% mean absolute error to installed wind capacity, increasing to 20% for day-ahead forecasts.

Although state-of-the-art techniques in machine learning have already been applied to the task of wind energy forecasting, the results are often limited to simplified case-studies. Mohandes et al. (2004) compared a support vector regression approach for wind speed prediction to a neural approach (i.e., a multilayer perceptron). The prediction is based on mean daily wind speed data from Saudi Arabia. Shi et al. (2010) proposed an approach that combines an evolutionary algorithm for parameter tuning with support vector regression based prediction. The technique allows a six hour prediction, and is experimentally evaluated on wind data from North China. Recently, Zhao et al. (2010) compared support vector regression models to backpropagation for a ten minutes prediction of wind speed. Further work concentrates on special aspects like prediction and diagnosis of wind turbine faults. Kusiak and Li (2011) introduced an approach based on fault prediction on three levels, e.g., fault category and specific fault prediction in a five minutes to one hour approach.

3. Wind Data

The models are evaluated based on the National Renewable Energy Laboratory (NREL) western wind resources data set (Lew et al. 2008, Potter et al. 2008), which is part of the Western Wind and Solar Integration Study, a large regional wind and solar integration study initiated by the US. The data set has been designed to perform temporal and spatial comparisons like load correlation or estimation of production from hypothetical (i.e., simulated) wind farms for demand analysis and planning of storage based on wind variability.

The data set consists of three years of wind energy data from numerical simulations that are mainly based on real-world wind measurements. It consists of 32,043 grid points, each holding ten Vestas with 3 MW capacity, in the western area of the US, and can be downloaded from the NREL website.[2] The whole model employs a total capacity of 960 GW of wind energy. A GUI allows to select grid points, and to download corresponding times series data, see Figure 1 for an illustration. Based on a time-resolution of ten minutes, 52,560 entries per year and grid point are available for 2004, 2005, and 2006, respectively.

---

[2] http://www.nrel.gov/

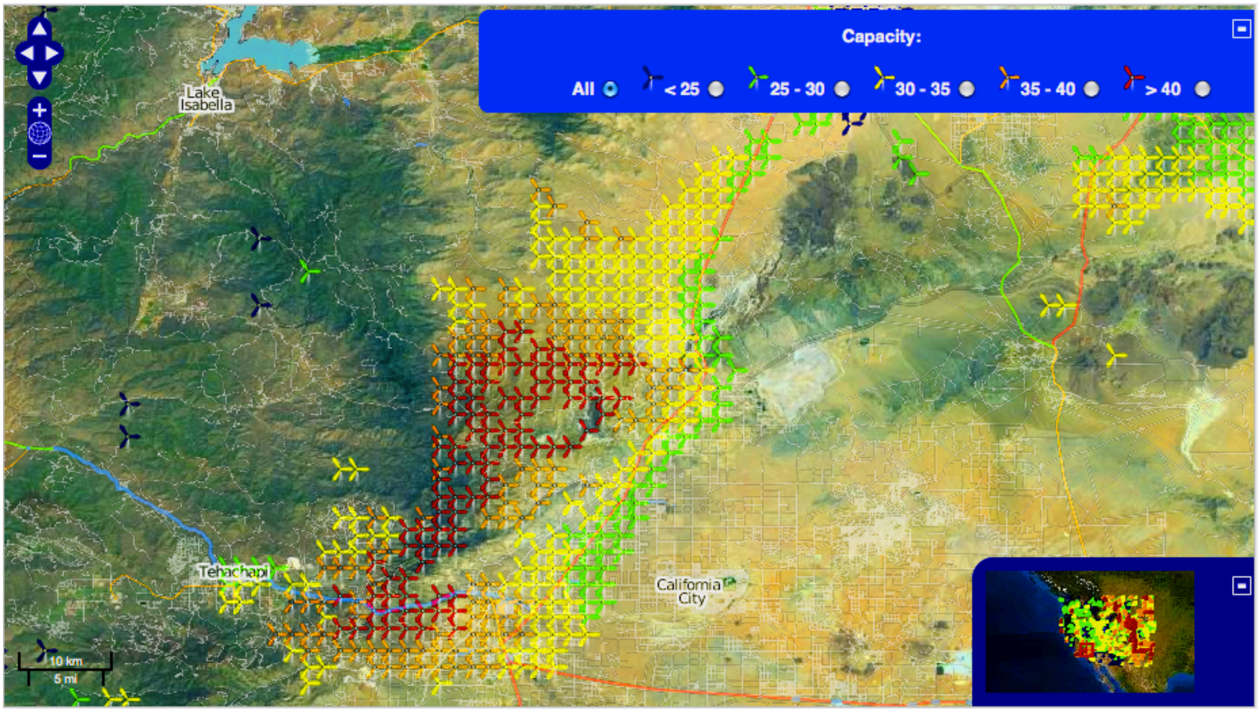


Figure 1: The figure shows a section of the Tehachapi wind park in California that consists of 50 single wind spots. The grid point colors indicate the capacities of the grid points (red = high, yellow = medium, green = low).

4. Support Vector Regression Models

One of the most popular tools in the field of machine learning are support vector machines, which can be used for classification, regression, and a variety of other learning settings (Hastie et al. 2009, Schölkopf and Smola 2009, Smola and Schölkopf 1998). In general, the basis for training appropriate models is a set $T = \{(x_1, y_1), \dots, (x_N, y_N)\} \sqsubset \mathbb{R}^d \times \mathcal{Y}$ consisting of labeled patterns. For classification settings, the space $\mathcal{Y}$ of labels is discrete (e.g., $\mathcal{Y} = \{-1, +1\}$ for the binary case). For regression scenarios, the space $\mathcal{Y}$ is given by $\mathbb{R}$; here, the goal of the learning process consists in finding a prediction function $f: \mathcal{X} \to \mathbb{R}$ that maps unseen patterns $x \in \mathcal{X}$ to reasonable real-valued labels. These models can be seen as special instances of problems having the form

$$inf_{f\in\mathcal{H},b\in\mathbb{R}} \frac{1}{N} \sum_{i=1}^{n} L\big(y_i, f(x_i + b)\big) + \lambda||f||_{\mathcal{H}}^2$$

where $\lambda > 0$ is a user-defined real-valued parameter, $L: \mathbb{R}\times\mathbb{R} \to [0, \infty)$ a loss function and $||f|_{\mathcal{H}}^2$ the squared norm in a so-called reproducing kernel Hilbert space $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}} = \{f: \mathcal{X} \to \mathbb{R}\}$ induced by an associated kernel function $k: \mathcal{X}\times\mathcal{X} \to \mathbb{R}$. The space $\mathcal{H}$ contains all considered models, and the term $|f|_{\mathcal{H}}^2$ is a measure for the "complexity" of a particular model f (Hastie et al. 2009, Schölkopf and Smola 2009, Smola and Schölkopf 1998). Ideally, one would like to generate models that represent the training data well and that are, at the same time, not too complex to avoid overfitting. This idea is reflected by the two objectives present in the above optimization task: The first term measures how well the model f fits to the data (according to the definition of the loss function L), whereas the second term measures the complexity of the model. The parameter λ is called regularization parameter and determines the trade-off between these two objectives, see Figure 2 for an illustration.

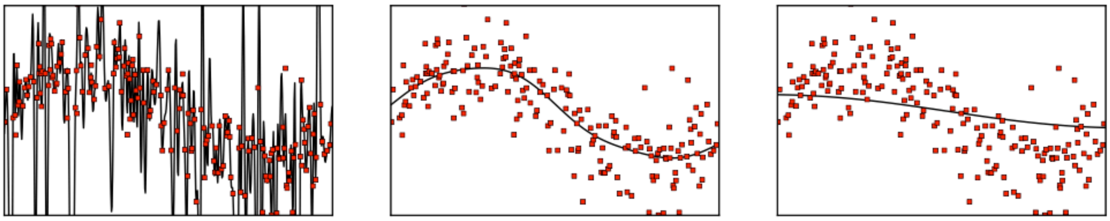


Figure 2: The model paramter $\boldsymbol{\lambda}$ determines the complexity of the model. Given a small value, the model fits the training data well but might be inappropriate for new unseen data, see Figure (a). On the other hand, large values for $\boldsymbol{\lambda}$ might yield too simple models that cannot reflect the structure of the data, see Figure (c). A reasonable trade-off between data fit and model complexity is given in Figure (b).

The deviation of predictions obtained via the model and the true data samples is penalized by the loss function. Typical choices are the square loss $L_\varepsilon(y,t) = (t-y)^2$ and the ε-insensitive loss $L_\varepsilon(y,t)$= $\max(|t - y| - \varepsilon, 0)$ with $\varepsilon > 0$, see Figure 3.

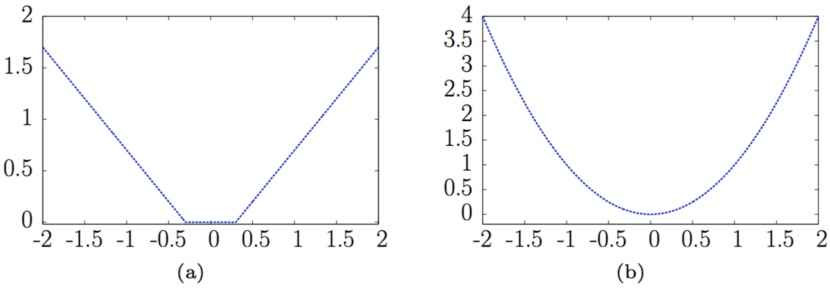


Figure 3: The $\boldsymbol{\varepsilon}$-insensitive (a) and the square loss (b).

The latter one does not take into account small residual errors and the choice of ε defines the magnitude of errors that can be neglected. Plugging in this loss function into the above objective leads to

$$inf_{f \in \mathcal{H}} \frac{1}{N} \sum_{i=1}^{n} \max(|f(x_i) + b - y_i| - \varepsilon, 0) + \lambda ||f||_{\mathcal{H}}^2$$

and the resulting models are called support vector regression (SVR) models.

5 Experimental Analysis

We formulate the wind forecasting task as regression problem, and assume that a time series $x_1, \ldots, x_N \in \mathbb{R}^d$ of f wind measurements of K wind grid points, and corresponding measurements $y_1, \ldots, y_N \in \mathbb{R}$ of wind energy production of a target point is given. The task is to predict the production y_t at time $t = t_i + \Theta$ with $\Theta \in \mathbb{N}$ based on past wind measurements at time $t_i - 1, t_i - 2, \ldots, t_i - \mu$, with $\mu \in \mathbb{N}$ past observations. We investigate the questions, if prediction of wind energy can exclusively be based on the existing infrastructure of wind turbines for three different time horizons, i.e., 20 minutes, one hour, and two hours.

5.1 Experimental Settings

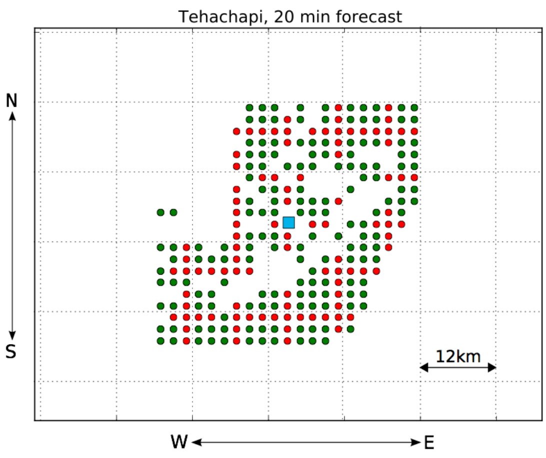
Our experiments show the behavior of SVR wind energy prediction for three different time horizons based on the NREL western wind resources data set we have described in Section 3. The analysis concentrates on wind grid points in the wind park of Tehachapi in California, USA. Each time horizon requires a selection of neighbored grid points with adequate distances. To automatize this selection, we divide the environment of the target grid point into separate cells using a quadratic lattice, which geometries depend on the choice of the time horizon, e.g., an increasing horizon potentially requires larger neighborhoods. Finally, for each grid cell of the lattice the grid point closest to its center is selected. In our experiments the size of the lattice and the number of employed grid cells for every setup have been determined with a manual parameter study. For a reliable forecast it is important to train the SVR model with a huge variety of wind situations. Therefore we chose a training set length of nine months in 2006. The SVR models are based on a RBF-kernel

$$k(x_i, x_j) := \exp\left(-\sigma||x_i - x_j||^2\right)$$

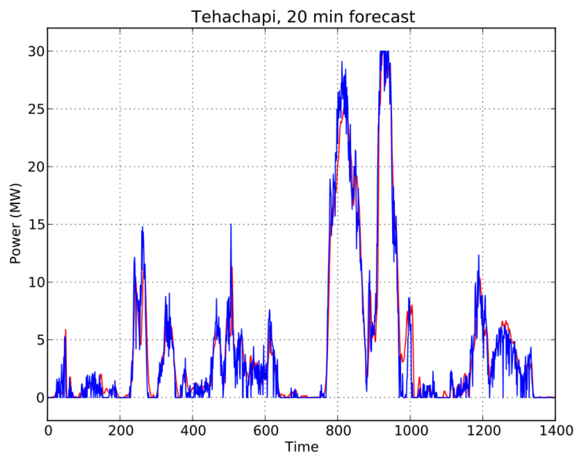
with bandwidth σ. For the training process we employ two-fold cross-validation, and grid search for parameters in the interval $\lambda = 10^i$ with $i = -4, -2, \dots, 4,$ and $\sigma = 10^j$ with $j = 0, -2, \dots, -10$. The intervals have been manually restricted to accelerate the training process like proposed in our previous work (Kramer and Gieseke 2011).

5.2 Time Horizon Analysis

The question is how far we can look into the future, and how much information from the past is necessary for a reliable forecast. After parameter tuning, and training of the SVR with optimal settings on the first nine months of 2006, we evaluated the SVR models on randomly chosen time series intervals of the last quarter of 2006. The intervals contain two major peaks (ramp events), which are very relevant for balancing the grid. Such events can be caused by storm fronts passing. Each evaluation interval consists of 1,400 time steps that correspond to approximately ten days. Figure 4 shows the experimental results of a 20-minute ahead prediction.



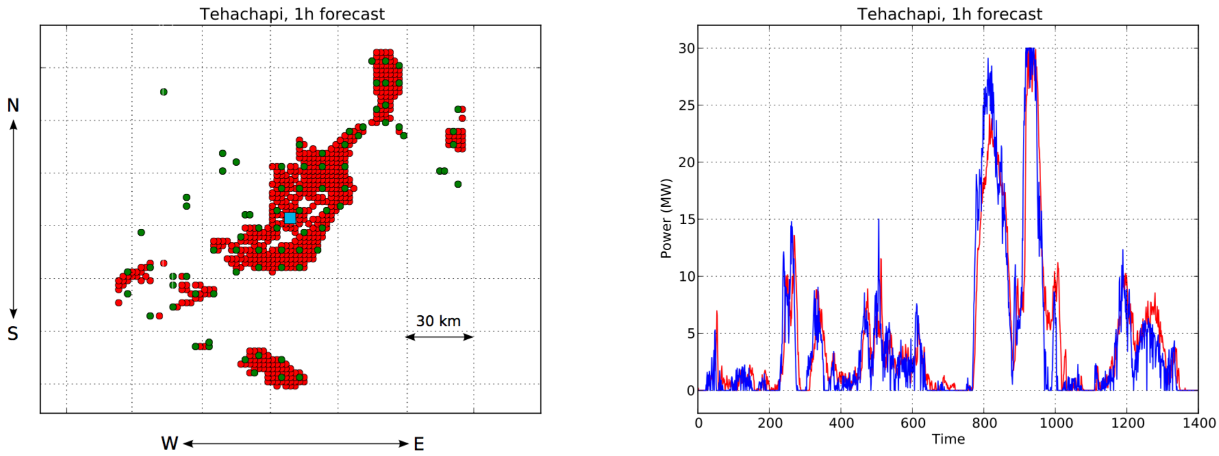
(a)



(b)

Figure 4: 20-minute forecast of a wind grid point in Tehachapi: (a) a selection of input grid points (green), (b) comparison of SVR prediction (red) with actual wind energy (blue). Prediction, and target wind energy curve match perfectly.

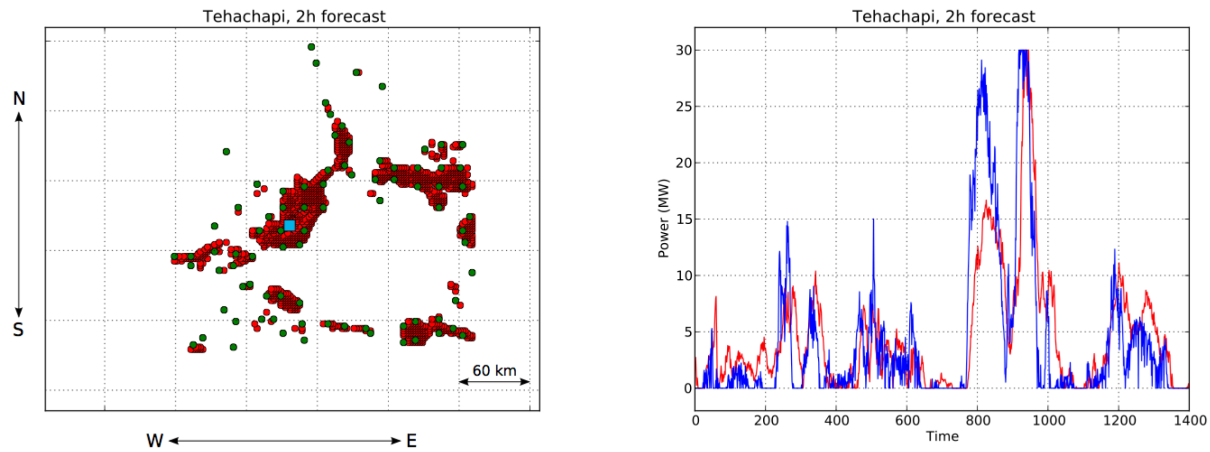
center, training input grids points are shown in green. Figure 4(b) compares the 20-minute SVR prediction (red) with the actual wind energy (blue). On the same prediction interval the corresponding 1-, and 2-hour orecasts are shown in Figures 5, and 6.



(a) (b)

Figure 5: 60-minute forecast of a wind grid point in Tehachapi: (a) selection of input grid points, (b) comparison of SVR prediction with actual wind energy. A larger neighborhoods of input grid points is required for an optimal prediction.

We can observe that the precision of the 20-minute ahead forecast is very high. The two peaks in the second half of the interval match almost perfectly. For larger time horizons, grid points spread in a larger neighborhood were necessary for optimal predictions. Although the accuracy is relatively high, small deviations from the target curve can be observed. This observation is even more evident in case of the 2-hour forecast with a tendency towards a temporal shift.



(a) (b)

Figure 6: 120-minute forecast of a wind grid point in Tehachapi: (a) selection of input grid points, (b) comparison of SVR prediction with actual wind energy. The quality of the SVR prediction may still be satisfying for many applications, but deviations from the target wind energy curve can be observed.

In (Kramer and Gieseke 2011) we have presented a numerical analysis of different time horizons, and past steps we used for training of the SVR model. We repeat these results in the following to illustrate numerically that with increasing time horizon the forecast deteriorates. The tendency for an improvement of predictions can be observed, if we take into account more past time steps.

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

Table 1: Analysis of loss function parameter $\boldsymbol{\varepsilon}$ on the validation error measure with $\boldsymbol{L_\varepsilon}$ and $\boldsymbol{L_2}$ loss.

We employ the L_ε-loss in the following. To identify an optimal setting for ε we chose a 30-minute forecast setting based on two past time steps, i.e., 20 minutes of 15 grid points from the environment of a Tehachapi wind grid point. 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. For comparison we state the L_ε-, and the L_2-loss on the validation set. The results have shown that the L_ε-error decreases with increasing tolerance threshold ε. But the L_2-loss has a minimum at $\varepsilon = 0.5$.

	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
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: 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 kilometers. The figure 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), cf.[5,6].

Table 2 shows the validation error for the energy forecast using the same setting (target and input grid points) as in the previous loss function study. The figure 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. The SVR has been trained on 1/10-th of the observations from 2006, and again employs an RBF-kernel, and grid search in the ranges $\sigma = 2^\alpha$ with $\alpha = -10, \dots, -5$, and $\lambda = 2^\alpha$ with $\alpha = 5, \dots, 10$ on the second 1/10-th of the one-year data. The final validation error is computed based on the second 1/5-th of the corresponding data sets, using the L_ε, and the square loss L_2. 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.

6. Conclusions

Wind energy prediction is an important aspect for a stable energy grid. The integrity and stability can be improved with a precise forecast of the volatile energy sources. We have demonstrated that SVR is a successful method for the prediction of wind energy production only based on wind measurements from wind turbines, in particular without further meteorological data or weather simulation models. SVR turns out to be a fast and robust time series prediction technique. For the wind resource scenarios we have found recommendable parameters for the ε-loss, and typical bounds for optimal kernel parameters. The lattice introduced in this chapter is a new contribution, and is a convenient way to select input grid points for the learning process. Our experiments have shown that a reliable forecast on the level of wind grid points is possible for the 20- minute and 2-hour ahead prediction. On the level of a whole wind park, the results have shown that even a reasonable six-hour forecast is possible (Kramer and Gieseke 2011).

Acknowledgements: This work has been supported in part by funds of the Deutsche Forschungsgemeinschaft (DFG) (Fabian Gieseke, grant KR 3695) and by EWE Next Energy, and the University of Oldenburg.

Bibliography

Costa, A. , Crespo, A., Navarro, J., Lizcano, G., Madsen, H., Feitosa, E. (2008): A review on the young history of the wind power short-term prediction. Renewable and Sustainable Energy Reviews, 12(6):1725–1744.

Ernst, B., Oakleaf, B., Ahlstrom, M., Lange, M., Moehrlen, C., Lange, B., Focken, U., Rohrig, K. (2007): Predicting the wind. Power and Energy Magazine, 5(6):78–89.

Ernst, B., Reyer, F., Vanzetta, J. (2009): Integration of wide-scale renewable resources into the power delivery system. In CIGRE/IEEE PES Joint Symposium, pages 1–9.

Hastie, T., Tibshirani, R., Friedman, J.(2009): The Elements of Statistical Learning. Springer, 2nd edition.

Kramer, O., Gieseke, F. (2011): Analysisofwindenergytimeserieswithkernelmethods and neural networks. In Proceedings of the 7th International Conference on Natural Computation, pages 2381–2385.

Kramer, O., Gieseke, F. (2011): Short-term wind energy forecasting using support vec- tor regression. In Proceedings of the International Conference on Soft Computing Models in Industrial and Environmental Applications, Advances in Intelligent and Soft Computing, pages 271–280.

Kusiak, A., Li, W. (2011): The prediction and diagnosis of wind turbine faults. Renewable Energy, 36(1):16–23.

Lew, D., Milligan, M., Jordon, G., Freeman, L., Miller, N., Clark, K., Piwko, R. (2009): 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.

Milligan, M., Porter, K., DeMeo, E., Denholm, P., Holttinen, H., Kirby, B., Mille, N., Mills, A., O'Malley, M., Schuerger, M., Soder, L. (2009): Wind power myths debunked. IEEE Power and Energy Society.

Mohandes, M., Halawani, T., Rehman, S., Hussain, A. A. (2004): Support vector machines for wind speed prediction. Renewable Energy, 29(6):939–947

Negnevitsky, M., Mandal, P., Srivastava, A. (2009): Machine learning applications for load, price and wind power prediction in power systems. In Intelligent System Applications to Power Systems (ISAP), pages 1–6.

Potter, C. W., Lew, D., McCaa, J., Cheng, S., Eichelberger, S., Grimit, E. (2008): Cre- ating 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.

Schölkopf, B., Smola, A. J. (2011): Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, Cambridge, MA, USA.

Shi, J., Yang, Y., Wang, P., Liu, Y., Han, S. (2010): 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.

Smola, A. J., Schölkopf, B. (1998): A tutorial on support vector regression.

Zhao, P., Xia, J., Dai, Y., He, J. (2010): Wind speed prediction using support vector regression. In Industrial Electronics and Applications (ICIEA), pages 882 – 886.