Selection of numerical weather forecast features for PV power predictions with Random Forests

Björn Wolff[1], Oliver Kramer[2], Detlev Heinemann[1]

[1]Institute of Physics, Energy Meteorology
[2]Department of Computing Science, Computational Intelligence
Carl von Ossietzky University in Oldenburg, Germany

Abstract. The increasing volatility introduced to power grids by renewable energy sources makes it necessary that the accuracy of energy forecasts are improved. Photovoltaic (PV) power plants hold the biggest share of installed capacity of renewable energy in Germany, so that high quality PV power forecasts are vital for a cost efficient operation of the underlying electrical grid. In this paper, we evaluate multiple Numerical Weather Prediction (NWP) parameters for their ability to improve PV power forecasting features. The importance of features is decided by a Random Forest algorithm. Furthermore, the resulting top ranked features are tested by performing PV power forecasts with Support Vector Regression, Random Forest, and linear regression models.

Keywords: PV power forecasting, Random Forest, feature importance, Support Vector Regression

1 Introduction

One major part of ensuring the stability of a electricity grid is keeping a fixed utility frequency, which rises with production and lowers with the consumption of energy. Until about a decade ago, transmission system operators solved this problem by simply matching the energy output and dispatch of power plants to the demand of its consumers. Today, the increasing share of weather dependent renewable energy sources on the consumer's side has been introducing new volatility to the grid, thus, keeping stability intact is becoming a more complex task. In Germany, photovoltaic (PV) systems take up the highest share of installed capacity of renewable energy. PV is able to reach power output rates of almost 26 GW at midday with a total nominal power of around 40 GW_{peak} [**?**]. To generate the same amount of energy one would need about 20 of the bigger nuclear power plants in Germany, whereas today only eight plants with varying capacities are still active. To integrate this amount of energy, and still ensure grid stability, in a cost efficient way, i.e., without relying on too much expensive reserve energy, accurate power forecasts for PV are necessary.

One widely used approach for PV power forecasting is based on parametric modeling of PV power using measurements and numerical weather predictions

(NWP) on varying spatial and temporal scales. A comparison of models was compiled by Pelland at el. [?]. In recent years, more and more statistical learning algorithms were used for wind and solar energy predictions. One popular approach for forecasting solar radiation are artificial neural networks (ANN). A compilation of different ANN models can be found in Mellit [?]. More recent works that focus on ANN are extending the models with preprocessing steps such as weather type classifications [?], feature selection implemented with, e.g., Genetic Algorithms [?] or increasing the input data by adding meteorological data from numerical weather predictions [?]. There are also many works successfully applying and comparing different statistical modeling approaches for PV energy production forecasting [?,?,?].

In comparison to ANN, Support Vector Regression (SVR) is still less common for PV power and irradiance forecasts but shows good potential in some comparisons to other statistical learning methods [?] with numerical weather prediction features as input [?] on single-site and regional solar radiation values [?]. There are only a few works that use Random Forest (RF) as a predictor in the field of renewable energy forecasts. One example where RF is successfully used for electric load forecasting can be found in Jurado et al. [?].

Here, we evaluate Random Forest's ability to assess the importance of numerical weather prediction features for PV power forecasting. The datasets used in this evaluation and the preprocessing of our data is described in Section 2. In Section 3, we give a short overview on the statistical learning methods utilized in this work for PV power predictions, i.e., Random Forest and Support Vector Regression. The selection of numerical weather forecast features is conducted in Section 4 and the most important features are evaluated with different regression models in Section 5.

2 Datasets

In this work, we use the data of 93 PV systems spread across Germany (see Figure 1) with varying installed capacities from rooftop installations to big PV parks. These systems are monitored by our industry partner meteocontrol GmbH[1] and are randomly selected from a larger dataset containing PV power measurements from 2012-01-01 to 2013-12-31. The temporal resolution of these measurements is 15 minutes and the specifications, i.e., orientation, tilt, and power capacity are known.

2.1 Numerical Weather Predictions

The focus of this work lies on evaluating the usefulness of different weather parameters of a numerical weather prediction (NWP) system for PV power forecasting. NWP models are systems of differential equations that use the laws of physic, fluid dynamics, and chemistry as well as weather observations and measurements as a basis to predict changes in weather situations starting from a

[1] meteocontrol GmbH: `www.meteocontrol.com`

recent initial state. These calculations are performed on a fixed horizontal and vertical grid.

Here, we use data from the European Centre for Medium-Range Weather Forecasts (ECMWF) [2]. The ECMWF's Integrated Forecasting System (IFS) provides single level forecasts covering over 120 parameters in a temporal resolution of three hours with two major forecasting runs starting at 00 UTC and 12 UTC. Single level forecasts are forecasts that are averaged over different height levels or forecasts for specific heights (e.g., 2 metre temperature). These runs are not instantly available at these times, so that we always use forecasts from the 12 UTC run of the previous day to ensure their availability at the point of time of our prediction. In Figure 1, the grid points of the ECMWF IFS model are shown with each grid field covering an area of about 12.5km × 12.5km. As we only use single level forecasts, there is no need for the vertical expansion of the grid in our case.

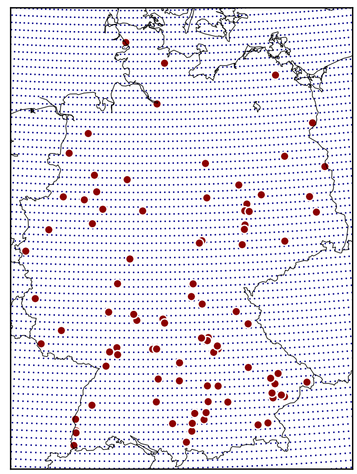


Fig. 1: Locations of 93 PV systems in Germany (red) and grid points of ECMWF's NWP model (blue)

2.2 Data Preprocessing

The NWP model output is interpolated both temporally and spatially to match the PV systems' measurements.

Spatial Interpolation In previous works (see Lorenz et al.[?]), a positive effect of averaging NWP radiation forecasts of multiple surrounding grid points instead of using the geographically nearest forecast was observed. The best results were achieved by averaging a 4×4 grid around a PV system's location. Due to the easier implementation and faster calculation, the spatial interpolation in this work is done with a distance weighted k-nearest neighbor regression model [?] with $k = 16$ in regard to the 4×4 grid. For consistency, this method of spatial interpolation is further applied for all other NWP forecasts used in this work.

Temporal Interpolation In case of irradiance forecast parameters, we are utilizing a clear sky model of Dumortier, described in Fontoynont et al.[?], to interpolate the data from three hours to 15 minutes. For all other parameters, we apply a normal linear interpolation. The clear sky interpolation is working in three steps:

1. Calculating clear-sky index k^*_{3h} for 3 hour values:

$$k^*_{3h} = \frac{I_{forec,3h}}{I_{clearsky,3h}}$$

[2] European Centre for Medium-Range Weather Forecasts: `www.ecmwf.int`

2. Linear interpolation of k^*_{3h} to 15 minute values results in k^*_{15min}.
3. Using k^*_{15min} as a factor for $I_{clearsky,15min}$ with 15 minute resolution:

$$I_{forec,15min} = k^*_{15min} \cdot I_{clearsky,15min}.$$

With these interpolations, we are now able to generate PV power forecasts with a resolution of 15 minutes for each of PV system using the following statistical learning approaches.

3 Statistical Learning Methods

In this section, we introduce the applied statistical learning methods, i.e., Random Forest for feature selection and Support Vector Regression for modeling PV power forecasts. While the two methods construct a regression function f mapping patterns X_i, consisting of one or more features, to labels y_i in different ways, both algorithms need a dataset Z containing known pattern and label combinations for training purposes. The mapping can be defined as follows:

$$\begin{pmatrix} X_1 \\ X_2 \\ \dots \\ X_N \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1d} \\ x_{21} & x_{22} & \dots & x_{2d} \\ \dots & \dots & \dots & \dots \\ x_{N1} & x_{N2} & \dots & x_{Nd} \end{pmatrix} \xrightarrow{f} \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_N \end{pmatrix},$$

with pattern $X_i \in \mathbb{R}^d$, where d denotes the number of features, label $y_i \in \mathbb{R}$, and training set size $N \in \mathbb{N}$.

3.1 Random Forest

The concept behind the construction of a Random Forest (RF) [**?**] regressor is Bagging (Bootstrap aggregation). Bagging makes it possible to "average many noisy but approximately unbiased models, and hence reduce the variance" [**?**] by combining multiple single models. With RF this variance reduction through correlation reduction of the trees is further improved by selecting input features in each node (one element of the tree) splitting step at random.

To train a Random Forest model, bootstrapping (random sampling with replacement) is applied on the training set Z to retrieve B, the number of trees in the *forest*, subsets $Z^{*b} = \{(\mathbf{x}^*_1, y^*_1), (\mathbf{x}^*_2, y^*_2), \dots, (\mathbf{x}^*_N, y^*_N)\}$, $b = 1, 2, \dots, B$. On each of these subsets a RF tree T_b is grown by applying the following steps until a stop criterion is reached (e.g., a minimum number of samples belonging to a newly created node), according to Hastie et al.[**?**]:

1. Randomly select $m \leq d$ features of the input pattern.
2. Calculate the best feature for splitting, i.e., the feature that "maximizes the decrease of some impurity measure" [**?**].
3. Split the node according to the selected feature into two new nodes.

After finishing the tree-growing process, the algorithm outputs an ensemble of trees $\{T_b \mid b = 1, 2, \ldots, B\}$ and the Random Forest regression function is

$$f_{rf}^{B}(\mathbf{x}) = \frac{1}{B}\sum_{b=1}^{B} T_b(x). \tag{1}$$

In our implementation, we use the Random Forest regressor method of the *scikit-learn* [**?**] Python package with its standard settings except for the number of trees. We increased the number of trees from ten to 64 to improve tree diversity but still keep the computation time at a minimum. This is as well the minimal recommendation of Oshiro et al. [**?**], even though they have a different field of use. As a impurity measure (step 2 of the tree-growing algorithm) to decide on the best splitting criterion/feature, this implementation uses the mean square error (mse)

$$E_{mse} = \frac{1}{N}\sum_{i=1}^{N}(z - z')^2. \tag{2}$$

The E_{mse} for each possible feature split is calculated and the feature with the highest E_{mse} decrease is selected.

3.2 Support Vector Regression

The basic idea of Support Vector Regression [**?**] is to find the regression function f that maps patterns to labels by solving the optimization problem

$$\inf_{f \in \mathcal{H}, b \in \mathbb{R}} \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}_\epsilon\big(y_i, f(\mathbf{x}_i + b)\big) + \lambda ||f||_{\mathcal{H}}^2. \tag{3}$$

Here, $\lambda \in \mathbb{R} > 0$ is a fixed user-defined cost parameter and the ϵ-insensitive loss function $\mathcal{L}_\epsilon$ is defined as $\mathcal{L}_\epsilon(y,t) = \max(0, |y - t| - \epsilon), \epsilon \in \mathbb{R} > 0$. $||f||_{\mathcal{H}}^2$ describes the squared norm in a so-called reproducing kernel Hilbert space $\mathcal{H}$ 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 regression model f [**?**]. Because of good results in related publications, e.g., [**?**,**?**,**?**], we use the radial basis function (RBF) kernel requiring another parameter $\gamma \in \mathbb{R} > 0$.

In summary, there are three user-defined parameters, i.e., λ, ϵ, and γ that we optimize by applying grid search with 80 different parameter combinations. Again, we use the SVR implementation of *scikit-learn* which is based on *LIBSVM* [**?**].

4 Feature Selection

The main goal of this work is to find additional NWP weather parameters that improve PV power forecasting. Using all available weather parameters of the

No.	Feature name	Unit
1	100 metre U wind component	ms^{-1}
2	100 metre V wind component	ms^{-1}
3	10 metre U wind component	ms^{-1}
4	10 metre V wind component	ms^{-1}
5	2 metre dewpoint temperature	K
6	2 metre temperature	K
7	Clear-sky direct solar radiation at surface*	Jm^{-2}
8	Cloud base height	m
9	Evaporation	m of water equivalent
10	High cloud cover	$(0-1)$
11	Large-scale precipitation	m
12	Low cloud cover	$(0-1)$
13	Medium cloud cover	$(0-1)$
14	Snow density	kgm^{-3}
15	Snow depth	m of water equivalent
16	Snow evaporation	m of water equivalent
17	Snowfall	m
18	Surface net solar radiation, clear sky*	Jm^{-2}
19	Surface solar radiation*	Jm^{-2}
20	Surface solar radiation downwards*	Jm^{-2}
21	Total cloud cover	$(0-1)$
22	Total column ice water	kgm^{-2}
23	Total column liquid water	kgm^{-2}
24	Total column rain water	kgm^{-2}
25	Total column snow water	kgm^{-2}
26	Total column water	kgm^{-2}
27	Total column water vapour	kgm^{-2}
28	Total precipitation	m
29	Total sky direct solar radiation at surface*	Jm^{-2}
30	Zero degree level	m

Table 1: Selection of ECMWF weather forecast feature list. Marked features (*) are interpolated with a clear sky interpolation method instead of a linear interpolation.

ECMWF's model would result in feature spaces with over 120 dimensions. This is not feasible for short-term PV power forecasting with forecast horizons of 15 minutes as the time to calculate a forecast would simply take too much time. As a result, we want to select less features that still hold the highest possible additional information for our models.

In literature (e.g., in Guyon and Elisseeff[**?**]), there are two major classes of feature selection methods: wrapper and filter. While wrapper methods embed a regression model into another optimization algorithm, e.g., genetic algorithms, that try different feature combinations until a stop-criterion is reached and select the best solution that was found along the search, filter methods rank single features by some measure that describe their usefulness for the given task.

As one iteration of a wrapper algorithm would take a lot of time in our scenario with 93 PV systems, we decided to use a filter model to achieve a fast method for evaluating the benefit of single weather parameters. Instead of using common measures like Pearson's correlation coefficient, we apply a comparatively new approach that is the build-in feature importance ranking of the Random Forest algorithm.

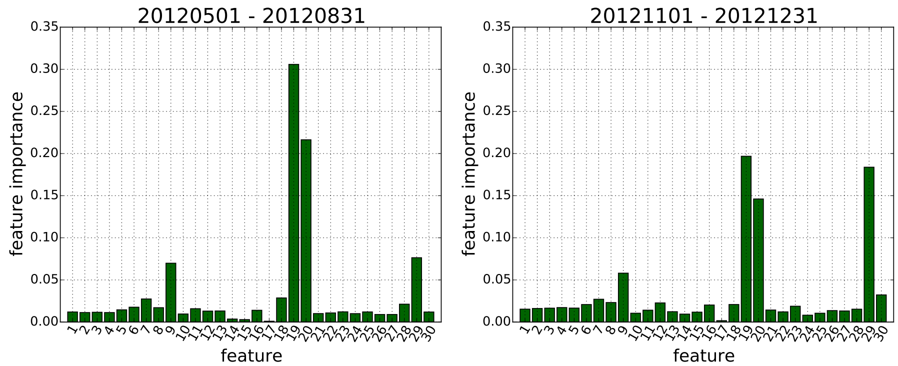
Preselection of Features With the help of experts in the field of meteorology, we removed all NWP parameters that are irrelevant for PV power forecasting. Thus, from the about 120 available IFS single level forecast parameters, 30 features remain for our evaluation. These 30 features are listed in Table 1.

Feature Selection with Random Forest To rank the preselected features, we use Random Forest's feature importance capabilities. The feature importance is achieved by traversing all trees with the training dataset and summing the impurity criterion (in this case the E_{mse} value) of each node. The E_{mse} is weighted with the number of samples that were routed to a specific node. Thus, features that were choosen often and early in a trees hierarchy will also receive high importance values. The output of this feature importance algorithm is normalized to one. Details on Random Forest's feature importance are well discussed in Genuer et al. [**?**].

Now, all preselected features are presented to the RF algorithm with the goal of calculating a regression function to predict power output of each single PV system. For this, we use the daytime (cosine of the solar zenith angle above 80 degree) feature values of the previous 65 days for training as a similar configuration showed good results in previous evaluations [**?**]. Each of the 93 PV systems is trained independently for each day of the year 2012 starting at 2012-03-10. By doing this, we receive the feature importance of every day separately and are able to evaluate seasonal or daily changes in feature importances.

In Figure 2, the feature importance of all days is averaged over all PV systems for summer and winter months. In summer, all features are dominated by the forecast of surface solar radiation (19) and surface solar radiation downwards (20) with over 50% feature importance. Only the parameters total sky direct solar radiation at surface (29) and evaporation (9) seem to hold some information for the resulting PV power output with an importance of about 7.5% each. This changes when looking at the winter months. The feature importances of 19 and 20 decrease by about one third while the importance of many other features increase. Noticeably, the zero degree level (30) and low cloud cover (12) as well as snow depth (14) and snow evaporation (15) become more important for the forecasting of PV power generation in winter.

In our following experiments, we want to evaluate whether adding weather parameters to the prediction pattern can improve the quality of PV power forecasts. In this first approach, we do not differentiate between seasons and use the



(a) summer

(b) winter

Fig. 2: Random Forest feature importances of 30 NWP weather parameters averaged for (a) summer and (b) winter months for all 93 PV systems.

Rank	No.	Feature name	Feature importance
1	19	Surface solar radiation	0.244
2	20	Surface solar radiation downwards	0.178
3	29	Total sky direct solar radiation at surface	0.171
4	9	Evaporation	0.050
5	7	Clear-sky direct solar radiation at surface	0.026
6	18	Surface net solar radiation, clear sky	0.023
7	8	Cloud base height	0.020
8	30	Zero degree level	0.019
9	6	2 metre temperature	0.018
10	16	Snow evaporation	0.017

Table 2: Top ten most important ECMWF weather forecast features ranked by Random Forest feature importance.

ten highest ranked features of the average feature importance for 2012. These features are listed ranked by their RF feature importance values in Table 2.

5 Prediction Comparison

Now, we use the most important features obtained in Section 4 for PV power forecasts. In this evaluation, we train our models at each day of the year 2013 and our training set consists of the previous 65 days as before. The training and forecasting is done on each PV system separately. Aside from the Random Forest predictor used to determine the feature importance, we test a linear regression and Support Vector Regression approach and evaluate if the Random Forest feature importance is applicable for different learning alogrithms as well.

To measure the quality of the different predictors, we use the E_{rmse}, the root of the E_{mse} introduced in Equation 2. The E_{rmse} is a good real life measure as it penalizes high deviations, that would have a higher impact on grid stabilizing actions in our case, more than small ones.

Instead of testing all possible combinations, we decided to iteratively add the top ten features to our pattern. Our base pattern (Equation 4) consists of the latest available measurement values P_{meas} according to the considered forecast horizon Δt. The different applied pattern/label matchings have the following structure:

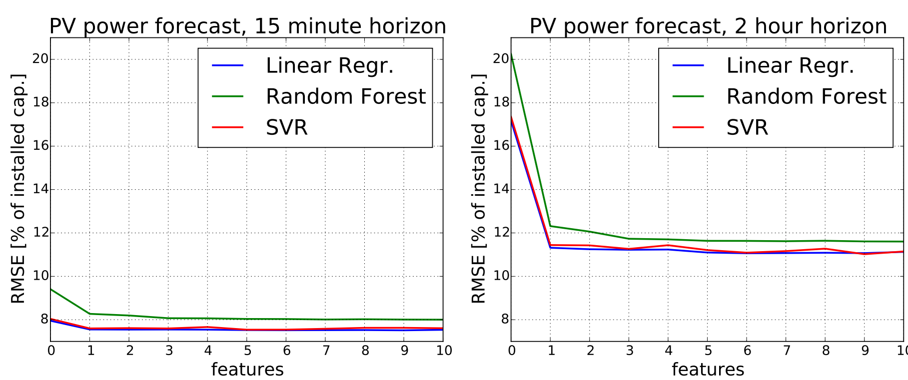
$$(P_{meas}(t - \Delta t)) \rightarrow P_{meas}(t) \quad (4)$$

$$(P_{meas}(t - \Delta t), Feature_1(t)) \rightarrow P_{meas}(t) \quad (5)$$

$$\dots$$

$$(P_{meas}(t - \Delta t), Feature_1(t), Feature_2(t), \dots, Feature_{10}(t)) \rightarrow P_{meas}(t) \quad (6)$$

Figure 3 shows the results of our forecasts for the linear regression, SVR, and RF models. First, we look at the shortest forecast horizon of 15 minutes in Figure 3(a). The E_{rmse} values of the forecast models using only measurement data (features 0) are already good and can not profit much from additional features. Especially after adding the NWP forecast of surface solar radiation, which was deemed the most important feature in our importance tests, there is almost no further improvement of the E_{rmse} for the linear and SVR model anymore. Only Random Forest forecasts acquire their best results with the top five features but the E_{rmse} values are still higher than that of the other two models.



(a) 15 minutes ahead (b) 2 hours ahead

Fig. 3: Comparison of Linear Regression, Random Forest, and Support Vector Regression with increasing input feature count for a prediction horizon of (a) 15 minutes and (b) two hours ahead.

When the temporal difference between the most recent measurement and the forecasted timestep gets bigger, the less important the measurements get. This is demonstrated in Figure 3(b) for a two hour forecast horizon. The initial measurement-based forecasts achieve a higher E_{rmse} for a forecast horizon of two hours than in the 15 minutes case. As before, adding only the most important feature is already enough to reach the lowest E_{rmse} value for the linear regression and SVR (only very slight improvements afterwards). Again, the Random Forest model needs more information, i.e., more features to reach its best forecast with the lowest E_{rmse}.

In both scenarios, the difference between a simple linear regression and the more sophisticated Support Vector Regression with optimized user-defined parameters is small. In Figure 4, we look at the most important feature, i.e., sur-

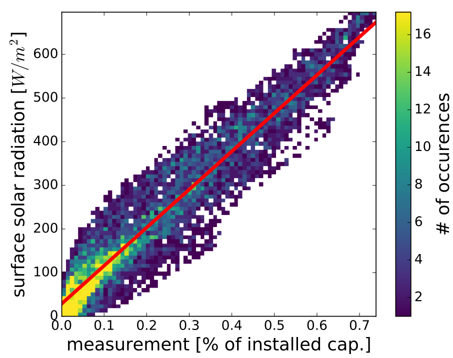
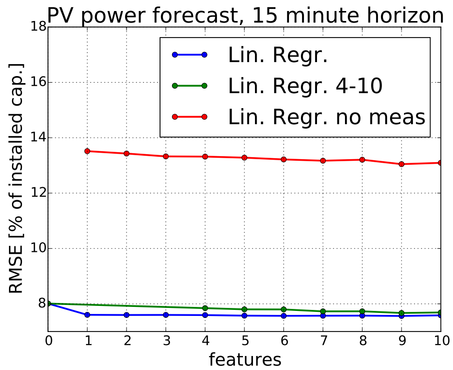


Fig. 4: Scatter plot of measurements and corresponding NWP forecast of surface solar radiation. The red line indicates the linear regression curve of these two variables.

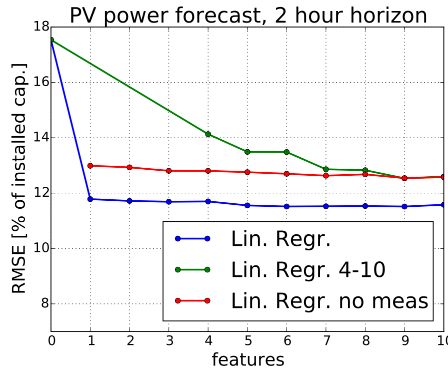
face solar radiation, which allows the highest E_{rmse} improvement, in more detail. Here, the average of all stations' measurements and the respective average of the surface solar radiation forecasts is compared. The distribution shows that these parameters hold a strong linear dependency, which is highlighted by a linear regression fit in this figure. This indicates that the SVR is not able to benefit from its ability to model non-linear relations and, therefore, is uncapable of achieving lower E_{rmse} values than a linear regression.

Due to the fact that there was almost no improvement in adding more than one NWP feature, we try different input combinations in Figure 5. Because there was not much of a difference between the results of the linear and the SVR model as well as the shorter computation times of a linear regression, we decided on testing these combinations only with the linear model. We test three different models:

– *Model 0*: Linear Regression with measurements and all top ranked features as before (Linear Regr.)



(a) 15 minutes ahead



(b) 2 hours ahead

Fig. 5: Comparison of Linear Regression forecasts with different input feature combinations for a prediction horizon of (a) 15 minutes and (b) two hours ahead

- *Model 1*: Linear Regression with measurements and without the top three ranked (radiation) features (Linear Regr. 4-10)
- *Model 2*: Linear Regression without measurements and with all top ranked features (Linear Regr. no meas)

In case of our shortest forecast horizon of 15 minutes (Figure 5(a)), the models using measurements are substantially better than the model without measurement values. While there are no improvements of the E_{rmse} after adding feature 1 for Model 0, the three missing features of Model 1 can be (almost completely) recuperated by adding more NWP features. Although the forecasts of Model 2 are inferior to measurement-based ones, an improvement is achieved by adding more of the top ten features. For a forecast horizon of two hours ahead, measurements are not that important anymore (see Figure 5 (b)). Now, Model 2, as the forecast does not change with time due to the same NWP forecast used, is competitive to Model 0 and even better than Model 1, despite the lack of measurements. There are slight differences of Model 2's E_{rmse} values for 15 minutes and two hours as we filter the time series in a way that only time steps where all models generate useable outputs are considered, so that the E_{rmse} values are calculated on different slightly timeseries. Model 1 is profiting from additional NWP features and able to compensate less information about the predicted radiation with increasing feature count. Utilizing all information (Model 0) is still generating the best forecasts, but with an increasing forecast horizon the difference to the other models is increasingly vanishing.

6 Conclusion

The expansion of PV power in the german grid makes it necessary that its forecasts become more accurate. To address this task, we evaluate the benefit of using additional weather forecast parameters of the ECMWF's NWP model for PV power predictions applying different regression models. The importance of all features that are related to the output of a PV system are assessed via Random Forest's feature importance algorithm. While there are differences between seasons and weather situations, the averaged results of the feature importance evaluation show that over 50% of importance are shared by irradiance weather parameters. This is later on confirmed by our analysis of PV power forecasts, that are using the highest ranked features, showing that the quality of power output forecasts for more than a few minutes ahead is mostly depending on irradiance forecasts. Otherwise, for shorter prediction horizons (15 to 30 minutes), PV power measurements are essential for generating high quality forecasts. These results are seen with all of the applied regression models, i.e., Support Vector Regression (SVR), Random Forest, and linear regression. As irradiance and a PV module's power output are highly linear correlated, SVR is not able to create better forecasts than the linear regression approach.

In further works, we want to investigate whether there are weather situations where additional features actually increase the forecast quality, e.g., snow and fog.

Acknowledgements

We like to thank meteocontrol GmbH for providing the PV system measurements as well as the ECMWF for numerical weather prediction forecasts that are basis of our experimental analysis. Björn Wolff is funded by the PhD program "System Integration of Renewable Energies" (SEE) of the University of Oldenburg promoted by the Lower Saxony Ministry for Science and Culture (MWK).

References

1. T. Bailey and A. K. Jain. A note on distance-weighted k-nearest neighbor rules. *IEEE Transactions on Systems, Man and Cybernetics*, 8:311–313, 1978.
2. M. Brabec, E. Pelikán, P. Krc, K. Eben, and P. Musilek. Statistical modeling of energy production by photovoltaic farms. In *Electric Power and Energy Conference (EPEC), 2010 IEEE*, pages 1–6, Aug 2010.
3. L. Breiman. Random forests. *Mach. Learn.*, 45(1):5–32, Oct. 2001.
4. C.-C. Chang and C.-J. Lin. LIBSVM – A library for support vector machines. Website, 2015. last checked: 2016-02-11.
5. C. Chen, S. Duan, T. Cai, and B. Liu. Online 24-h solar power forecasting based on weather type classification using artificial neural network. *Solar Energy*, 85(11):2856 – 2870, 2011.
6. M. Fontoynont, D. Dumortier, D. Heinnemann, A. Hammer, J. Olseth, A. Skarveit, P. Ineichen, C. Reise, J. Page, L. Roche, and others. Satellight: a WWW server which provides high quality daylight and solar radiation data for Western and Central Europe. In *9th Conference on Satellite Meteorology and Oceanography*, pages 434–437. American Meteorological Society Ed., Boston, Massachusetts, USA, 1998.

7. Y. Gala, A. Fernández, J. Díaz, and J. R. Dorronsoro. Support vector forecasting of solar radiation values. In *Hybrid Artificial Intelligent Systems*, volume 8073 of *Lecture Notes in Computer Science*, pages 51–60. Springer Berlin Heidelberg, 2013.
8. R. Genuer, J.-M. Poggi, and C. Tuleau-Malot. Variable selection using random forests. *Pattern Recognition Letters*, 31(14):2225 – 2236, 2010.
9. I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *J. Mach. Learn. Res.*, 3:1157–1182, Mar. 2003.
10. T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer, 2001.
11. S. Jurado, J. Peralta, A. Nebot, F. Mugica, and P. Cortez. Short-term electric load forecasting using computational intelligence methods. In *Fuzzy Systems (FUZZ), 2013 IEEE International Conference on*, pages 1–8, July 2013.
12. O. Kramer and F. Gieseke. Short-term wind energy forecasting using support vector regression. In *SOCO*, volume Volume 87/2011, pages 271–280. Springer, 2011.
13. P. Krömer, P. Musílek, E. Pelikán, P. Krc, P. Jurus, and K. Eben. Support vector regression of multiple predictive models of downward short-wave radiation. In *2014 International Joint Conference on Neural Networks, IJCNN 2014, Beijing, China, July 6-11, 2014*, pages 651–657, 2014.
14. E. Lorenz, J. Hurka, D. Heinemann, and H. G. Beyer. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2(1):2–10, March 2009.
15. G. Louppe, L. Wehenkel, A. Sutera, and P. Geurts. Understanding variable importances in forests of randomized trees. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 431–439. Curran Associates, Inc., 2013.
16. P. Mandal, S. T. S. Madhira, A. U. haque, J. Meng, and R. L. Pineda. Forecasting power output of solar photovoltaic system using wavelet transform and artificial intelligence techniques. *Procedia Computer Science*, 12:332–337, 2012.
17. R. Marquez and C. F. Coimbra. Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database. *Solar Energy*, 85(5):746 – 756, 2011.
18. A. Mellit. Artificial intelligence technique for modelling and forecasting of solar radiation data: A review. *International Journal of Artificial Intelligence and Soft Computing*, 1:52, 2008.
19. C. Monteiro, T. Santos, L. Fernandez-Jimenez, I. Ramirez-Rosado, and M. Terreros-Olarte. Short-term power forecasting model for photovoltaic plants based on historical similarity. *Energies*, 6(5):2624–2643, May 2013.
20. T. M. Oshiro, P. S. Perez, and J. A. Baranauskas. *Machine Learning and Data Mining in Pattern Recognition: 8th International Conference, MLDM 2012, Berlin, Germany, July 13-20, 2012. Proceedings*, chapter How Many Trees in a Random Forest?, pages 154–168. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.
21. H. T. Pedro and C. F. Coimbra. Assessment of forecasting techniques for solar power production with no exogenous inputs. *Solar Energy*, 86(7):2017 – 2028, 2012.
22. S. Pelland, J. Remund, J. Kleissl, T. Oozeki, and K. De Brabandere. Photovoltaic and solar forecasting: State of the art. `http://www.iea-pvps.org/index.php?id=1&eID=dam_frontend_push&docID=1690`, 2013.
23. scikit learn. Machine learning in python. Website, 2016. last checked: 2016-02-11.
24. A. J. Smola and B. Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14:199–222, 2004.

25. H. Wirth and K. Schneider. Recent facts about photovoltaics in germany. `http://www.ise.fraunhofer.de/en/renewable-energy-data/data-and-facts-about-pv`, Apr. 2016.
26. B. Wolff, J. Kühnert, E. Lorenz, O. Kramer, and D. Heinemann. Comparing support vector regression for PV power forecasting to a physical modeling approach using measurement, numerical weather prediction, and cloud motion data. *Solar Energy*, 135:197 – 208, 2016.
27. J. Zeng and W. Qiao. Short-term solar power prediction using a support vector machine. *Renewable Energy*, 52(0):118 – 127, 2013.