ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8. Available from http://www.i6doc.com/en/.

Comparison of Numerical Models and Statistical Learning for Wind Speed Prediction

Nils André Treiber[1], Stephan Späth[2], Justin Heinermann[1], Lueder von Bremen[2] and Oliver Kramer[1]

[1]Computational Intelligence Group, Carl von Ossietzky University, 26111 Oldenburg, Germany.

[2]ForWind - Center for Wind Energy Research, Carl von Ossietzky University, 26129 Oldenburg, Germany.

Abstract. After decades of dominating wind forecasts based on numerical weather predictions, statistical models gained attention for shortest-term forecast horizons in the recent past. A rigorous experimental comparison between both model types is rare. In this paper, we compare COSMO-DE EPS forecasts from the German Meteorological Service (DWD) post-processed with non-homogeneous Gaussian regression to a multivariate support vector regression model. Further, a hybrid model is introduced that employs a weighted prediction of both approaches.

1 Introduction

A future energy grid with high penetration of wind energy can only be balanced with precise forecasts of upcoming wind speed and resulting wind power. To improve wind power predictions, various methods have been developed and applied in the past. Most approaches are based on numerical weather prediction models [2], which describe the dynamics of the atmosphere by solving differential equations. Numerical weather predictions are used for short- and long-term predictions and usually cover a period in the range of hours to a few days. The second type of prediction methods is based on statistical learning. Generally, these models are based on machine learning algorithms deriving functional dependencies directly from the observations. The advantage of statistical models is that they can be run on standard personal computers within few minutes.

2 Numerical Weather Prediction Model and Calibration

We use operational forecasts of the COSMO-DE EPS, a high-resolution local ensemble prediction system run by the German Meteorological Service (DWD) [1]. The model consists of 20 ensemble members and the forecasts are initialized at 00:00 UTC. The maximum forecast lead time is 21 hours, and the output is available hourly. In this study, we use the ensemble mean as a point forecast. The ensemble mean forecast is expected to be as good or even better than a deterministic forecast. The model solves the basic atmospheric equations numerically. Model errors are induced for example by insufficient knowledge of the initial state of the integration and by the fact, that many physical processes

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8. Available from http://www.i6doc.com/en/.

cannot be explicitly modeled, because the computational cost would be too high and are thus parametrized. To reduce model errors, we employ statistical post-processing, which is often called calibration.

Calibration

A well-established method for correcting spread and bias of ensemble forecasts is non-homogeneous Gaussian regression (NGR), which was developed by Gneiting *et al.* [3]. NGR transforms the ensemble forecast to a probability density. Thorarinsdottir and Gneiting have shown, that a truncated Gaussian distribution can be used to model wind speed forecasts [6]. Location (ν) and scale parameter (ρ) of this predictive distribution are estimated from ensemble mean forecasts ($\overline{x}$) and ensemble standard deviation (S) by linear regression: $\nu = a + b\overline{x}$ and $\rho^2 = c + dS^2$.
The regression parameters a, b, c and d are estimated by minimization of a proper scoring rule over a training period. Here we follow [6] and use multidimensional minimization of the continuous ranked probability score (crps). The crps of the cumulative distribution function of the prediction f and the observation y is defined as [4]:

$$\mathrm{crps}(f, y) = \int_{-\infty}^{\infty} [f(t) - H(t-y)]^2 \, dt \tag{1}$$

H is the Heaviside function. For the multidimensional minimization, we minimize the mean crps over a N-day training period.

$$\mathrm{CRPS}(f, y) = \frac{1}{N} \cdot \sum_{i=1}^{N} \mathrm{crps}(f(a, b, c, d), y) \tag{2}$$

The minimization is done using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm as implemented in the GNU R function OPTIM [11]. The chosen training period is 100 days and the training is computed for every lead time separately.

3 Statistical Model: Spatio-Temporal Time Series Model

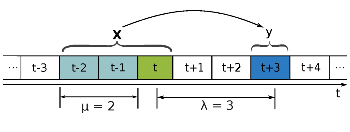


Fig. 1: The pattern $\mathbf{x}$ is mapped to label y. The time horizon of the prediction is λ.

As statistical model we use the approach introduced by Kramer and Gieseke [5] and further analyzed by Treiber *et al.* [8, 7]. This model makes predictions with past wind power measurements. For this task, we formulate the prediction as regression problem. Let us first assume, we intend to predict the wind speed of a single station with its time series. The wind speed measurement $\mathbf{x} = p(t)$ (pattern) is mapped to the wind speed at a target time $y = p_T(t + \lambda)$ (label) with $\lambda \in \mathbb{N}^+$ being the

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8. Available from http://www.i6doc.com/en/.

forecast horizon. For our regression model we assume to have N of such pattern-label pairs that are the basis of the training set $T = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$ and allow, via a regression, to predict the label for a unknown pattern $\mathbf{x}'$ (the present wind scenario). One can assume that this model yields better predictions if more information of the time series will be used. For this reason, we extend the pattern $\mathbf{x}$ by considering past measurements $p_T(t-1), \ldots, p_T(t-\mu)$ with $\mu \in \mathbb{N}^+$. Since we aim to catch spatio-temporal correlations, we add further information to our patterns from m neighboring stations. Their attributes can be composed in the same way like for the target stations.

Finally, we learn a prediction function $f(\mathbf{x}) \to y$ by employing a Support Vector Regression (SVR) [9] with an RBF-kernel $k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{||\mathbf{x}-\mathbf{x}'||^2}{2\sigma^2}\right)$.

4 Experimental Comparison

4.1 Wind Data

We use 10 m wind speed measurements from 253 synoptic stations in Germany. The measurement values are available hourly. The measurements are conducted either by sonic or cup anemometers. Some basic quality control is applied to the data, e.g., checks for negative wind speeds. The applied data of measured wind speeds covers a time interval from 01/01/2009 to 8/30/2013. The prediction of the Cosmo-EPS model and the calibrated adjustment are available from 12/01/2011 to 08/30/2013. Thus, we decided to train the statistical model on data from 01/01/2009 to 11/30/2011.

4.2 Experimental Settings

For the parameter tuning of the SVR, we employ a 3-fold cross-validation on the training set. We test via grid search values of the trade-off parameter $C \in \{10, 100, 500, 1000, 5000\}$ and the kernel bandwidth $\sigma = 10^{-i}$ with $i = 3, 4, 5, 6$. Due to the failure or the maintenance of the anemometers, the measurements are often not complete. Since the statistical models cannot naturally cope with missing data, imputation methods usually have to be applied. However, in our experimental study, we only use complete pattern-label pairs for the training of the SVR model. The advantage is that we avoid interactions between the imputation technique and the later forecast results. Finally, we only make predictions for stations with at least 1000 training pairs for the statistical model. In order not to skew the results by a too low sample size, we enforce that on at least $l_{\min} = 100$ days the predictions are compared. This constraint and the one corresponding to the SVR training are fulfilled for $z = 237$ stations, which are analyzed in the following. For the spatio-temporal model, we consider all stations that are located in a radius of 150 km from the target station. Typically, this are 20-40 stations.

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8. Available from http://www.i6doc.com/en/.

4.3 Verification for Different Forecast Lead Times

In the first part of the experimental analysis, we want to compare the different forecast types, i.e., the uncalibrated, the calibrated and the SVR model. We are interested in the prediction accuracy and its behavior with regard to an increasing prediction horizon. In the experimental study, we have to notice that the mean wind speed differs from station to station. We normalize the error for each station that is normalized with the mean observed wind speed $\overline{y^j}$. This results in the following error measure:

$$\overline{\text{RMSRE}} = \frac{1}{237} \sum_{j=1}^{237} \text{RMSRE}_j \qquad (3)$$

with the single station error:

$$\text{RMSRE}_j = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(\frac{f_i^j - y_i^j}{\overline{y^j}} \right)^2} \qquad (4)$$

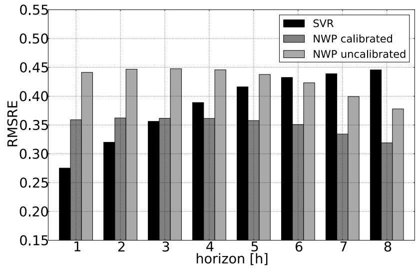


Fig. 2: Comparison of the three models

of station j. Fig. 2 visualizes the results and shows that the statistical SVR predictions are superior for a forecast horizon up to three hours. We can observe that the accuracy for the statistical model initially grows linearly with the forecast horizon. The calibrated NWP model has a relatively high error already at lead time +1 h. Its forecast error does not grow with lead time, it even decreases due to the diurnal cycle of wind speed which leads to different magnitudes of errors throughout the day. The same holds for the uncalibrated NWP forecasts, which perform significantly worse than the calibrated ones.

4.4 Direct Comparison for Shortest-term Predictions

In the previous section, we observed that the calibrated forecasts are more precise with regard to the RMSRE than the raw predictions for every forecast horizon. Now, we compare this model with the statistical one. The calibrated model serves as reference in our studies. To analyze the quality of the two different prediction models, we first compute the root mean square error RMSE_j for each station j:

$$\text{RMSE}_j = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(f_i^j - y_i^j \right)^2} \qquad (5)$$

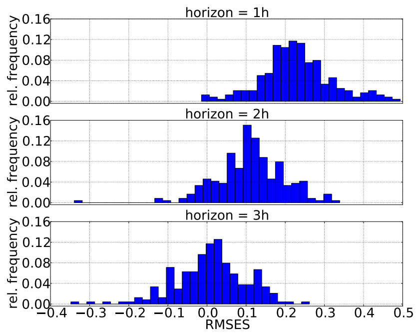


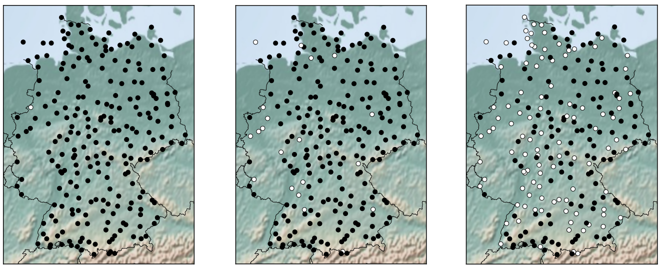
Fig. 3: Visualization of the distribution of the RMSES for lead times +1 to +3 h.

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8.
Available from http://www.i6doc.com/en/.

With these values, we compute the relative skill score RMSES_j for each station j based on the RMSE_j for the statistical model ($\text{RMSE}_j^{\text{SVR}}$) and for the calibrated model ($\text{RMSE}_j^{\text{NWPcal}}$):

$$\text{RMSES}_j = 1 - \frac{\text{RMSE}_j^{\text{SVR}}}{\text{RMSE}_j^{\text{NWPcal}}} \tag{6}$$

The RMSES_j indicates how the statistical model behaves with regard to the calibrated model. For example, a positive value for the RMSES points out that the statistical model outperforms the other model and the other way round. In Fig. 3, the histograms of the individual RMSES are shown for three horizons, i.e., one to three hours. One can observe that an increasing prediction horizon causes a shift of the distribution to the left. At +1 h lead time, the calibrated NWP forecast achieves a higher accuracy only for one station. For a horizon of three hours, the distribution gets more asymmetric. The next step is an analysis, for which locations the statistical model generates better predictions. Fig. 4 shows a map of all stations. In the two hour ahead scenario, there is a greater accumulation of stations in the western part of Germany, where the statistical predictions are more inaccurate, while in the three hour ahead predictions the stations are homogenously distributed. A further geographical dependency on the superiority of any model cannot be observed.



(a) horizon = 1 h (b) horizon = 2 h (c) horizon = 3 h

Fig. 4: Black points represent stations, whose statistical predictions are better than the numerical ones. For a forecast horizon of one hour $\lambda = 1$, the numerical prediction is better for one station. For $\lambda = 2$, 20 stations and for $\lambda = 3$, 102 stations achieve a higher accuracy with the numerical model.

4.5 Combination of Numerical and Statistical Forecasts

In the following, we analyze, if the combination of both prediction models achieves a higher accuracy. To investigate this question, we employ a linear weighted combination of both models:

$$\hat{f} = \alpha \cdot f^{\text{NWPcal}} + (1 - \alpha) \cdot f^{\text{SVR}} \tag{7}$$

with coefficient $\alpha \in [0, 1]$. Fig. 5 shows a study of parameter α w.r.t. different values for α and various prediction horizons. We can observe that the

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8. Available from http://www.i6doc.com/en/.

hybridization of both models is always better than a single model, i.e., there is always a minimal error for $0 < \alpha < 1$ in comparison to both extremes $\alpha = 0$ and $\alpha = 1$. Further, we can observe a wider spread of errors on the left (higher weight of statistical model f^{SVR}) than on the right (higher weight of numerical model f^{NWPcal}).

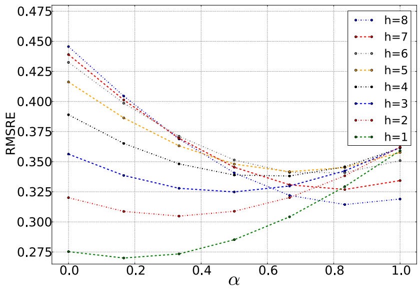


Fig. 5: RMSRE of linear combinations for all stations with a variable linear coefficient α, see Equation 7.

5 Conclusions

In this paper, statistical and numerical models for wind speed prediction are compared. We demonstrate that the multivariate spatio-temporal statistical model is superior for a horizon up to three hours. For a larger forecast horizon, the calibrated numerical model achieves a higher accuracy. Due to few time steps, for which both models yield complete predictions, only a linear hybridization has been analyzed. In the future, we will explore further types of hybridizations applicable on larger data sets.

References

[1] Ben Bouallègue, Z., Theis, S.E., Gebhardt, C. Enhancing COSMO-DE ensemble forecasts by inexpensive techniques. In *Meteorol. Zeitschrift*, 22(1), 49-59, 2013

[2] Ernst, B., Oakleaf, B., Ahlstrom, M., Lange, M., Moehrlen, C., Lange, B., Focken, U., Rohrig, K. Predicting the Wind. In *Power and Energy Magazine*, 5(6), 78-89, 2007.

[3] Gneiting, T., Raftery, A. E., Westveld III, A. H., Goldman, T. Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. In *Monthly Weather Review*, 133 (5), 1098-1118, 2005.

[4] Hersbach, H. Decomposition of the continuous ranked probability score for ensemble prediction systems. In *Weather and Forecasting*, 15(5), 559-570 2000

[5] Kramer, O., Gieseke, F. Short-Term Wind Energy Forecasting Using Support Vector Regression. In *6th International Conference on Soft Computing Models in Industrial and Environmental Applications*, 2011.

[6] Thorarinsdottir, T. L., Gneiting, T. Probabilistic forecasts of wind speed: ensemble model output statistics by using heteroscedastic censored regression. In *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 173, 371-388, 2010

[7] Treiber, N. A., Kramer, O. Evolutionary Turbine Selection for Wind Power Predictions. In *37th Annual German Conference on AI*, pages 267-272, 2014.

[8] Treiber, N. A., Heinermann, J., Kramer, O. Aggregation of Features for Wind Energy Prediction with Support Vector Regression and Nearest Neighbors. In *European Conference on Machine Learning, DARE Workshop*, 2013.

[9] Vapnik, V. The Nature of Statistical Learning Theory. Springer, New York, 1995.

[10] Wolpert, D. Stacked Generalization. In *Neural Networks, Vol. 5*, pp. 241-259, 1992

[11] R Core Team, R Foundation for Statistical Computing, Vienna, Austria. R: A Language and Environment for Statistical Computing. http://www.R-project.org/, 2014