Accurate Medium-Term Wind Power Forecasting in a Censored Classification Framework

Carsten Croonenbroeck

Christian Møller Dahl

European University Viadrina Frankfurt (Oder)

Department of Business Administration and Economics

Discussion Paper No. 351

April 2014

ISSN 1860 0921

Accurate Medium-Term Wind Power Forecasting in a Censored Classification Framework

April 14, 2014

Carsten Croonenbroeck[a] and Christian Møller Dahl[b]

Abstract

We provide a wind power forecasting methodology that exploits many of the actual data's statistical features, in particular both-sided censoring. While other tools ignore many of the important "stylized facts" or provide forecasts for short-term horizons only, our approach focuses on medium-term forecasts, which are especially necessary for practitioners in the forward electricity markets of many power trading places; for example, NASDAQ OMX Commodities (formerly Nord Pool OMX Commodities) in northern Europe. We show that our model produces turbine-specific forecasts that are significantly more accurate in comparison to established benchmark models and present an application that illustrates the financial impact of more accurate forecasts obtained using our methodology.

JEL classification: C34, E27, Q47
Keywords: Censored Regression, Wind Energy, Forecasting

Addresses:

[a] Corresponding Author: Carsten Croonenbroeck, European University Viadrina, Chair of Economics and Economic Theory (Macroeconomics), Post Box 1786, 15207 Frankfurt (Oder), Germany, Tel. +49 (0)335 5534 2701, Fax +49 (0)335 5534 72701, croonenbroeck@europa-uni.de.

[b] Christian Møller Dahl, University of Southern Denmark, Department of Business and Economics, Campusvej 55, 5230 Odense M, Denmark, Tel. +45 65 50 32 67, Fax +45 65 50 32 37, cmd@sam.sdu.dk.

Accurate Medium-Term Wind Power Forecasting in a Censored Classification Framework

April 14, 2014

Abstract

We provide a wind power forecasting methodology that exploits many of the actual data's statistical features, in particular both-sided censoring. While other tools ignore many of the important "stylized facts" or provide forecasts for short-term horizons only, our approach focuses on medium-term forecasts, which are especially necessary for practitioners in the forward electricity markets of many power trading places; for example, NASDAQ OMX Commodities (formerly Nord Pool OMX Commodities) in northern Europe. We show that our model produces turbine-specific forecasts that are significantly more accurate in comparison to established benchmark models and present an application that illustrates the financial impact of more accurate forecasts obtained using our methodology.

JEL classification: C34, E27, Q47
Keywords: Censored Regression, Wind Energy, Forecasting

1 Introduction

Accurate medium-term forecasts (one to 36 hours ahead) are crucial in the wind energy industry for all market participants. Xiaohong et al. (2000) argue that power producers as well as power distributors (energy traders) require reliable forecasts to achieve market clearing and efficient pricing in the power markets. Transmission system operators (TSOs) are also in need of accurate forecasts of power production since they have to manage the network load, as Hemmingsson et al. (2006) point out.

This paper presents a forecasting model that focuses on the non-linear relationship between wind speed and wind power production, called the *power curve* as described by, for example, Burton et al. (2011). We utilize a distinct set of explanatory variables and observe the relationship as a two-sided censored set of data. Therefore, our model uses a piece of important additional ex-ante available information: The power range of the turbine(s) that determines the two censoring points. Our model returns an efficient and unbiased forecast of wind power production that is more precise than the models currently used. It should be emphasized that the model mainly aims at running at a turbine-specific level but can also be used in a more macro-oriented perspective WLOG; for example, at a wind park level including several turbines.

The paper is structured as follows: Section 2 presents a short overview of existing models and briefly describes how our model generalizes the idea on which one of these models is based. Section 3 shows how we estimate our model in detail. Section 4 sheds light on in-sample and out-of-sample forecasting precision in comparison, discusses goodness of fit, evaluates censoring forecasting power, and puts emphasis on the financial aspect of forecasting precision. Section 5 concludes.

2 Overview of Prevailing Models

In the comparably short history of wind power forecasting, basically three branches of methods have emerged:

1. Artificial Neural Network (ANN)-based methods. An early work is that of Beyer et al. (1994), which led to WPMS (Wind Power Management System), introduced by Ernst and Rohrig (2002). One of the main problems of these models is overfitting, which is being addressed by automatized specification finding as described by Jursa and Rohrig (2008). Similarly, Sánchez (2008) combines several ANN-based forecasting models to find the best forecast at

any one time.
2. Physics-based models. One of the first models based on this approach was introduced by Landberg (1999). The model is still being maintained and is called *Prediktor*, a commercial software solution. An overview of other developments based on the physics idea is presented by Lange and Focken (2006). Current contributions to Prediktor are being published irregularly in the overview reports by the project ANEMOS.plus. The most recent issue has been released by Giebel et al. (2011). However, these models depend on high-resolution local weather data,[1] so computation impact and data volume handled are comparably enormous if a forecast at a turbine-specific level is desired.
3. Methods based on stochastic modeling. Nielsen et al. (1999) capture the diurnal periodic compound of wind power production. The resulting model was called WPPT2 (Wind Power Prediction Tool, second version). In 2002, WPPT and Prediktor formed the project "Zephyr", trying to combine the advantages of both approaches. Zephyr is still being maintained, and current developments are reported by Giebel et al. (2011). Besides the cooperation with Zephyr, both Prediktor and WPPT are also updated separately. WPPT has been updated to a conditional parametric version, called WPPT4 by Nielsen et al. (2002). The most recent work on WPPT family models is reported by Nielsen et al. (2007). WPPT models find a broad basis of usage in Denmark, the worldwide leading country of wind power production.
However, more recent developments aim at taking more of the "stylized facts" of the data into account. Pinson (2012) exploits both-sided censoring in the data. Furthermore, he utilizes wind direction as an important explanatory variable and uses it for data-driven parametrization of an autoregressive model that produces precise forecasts in a very short time frame. We also take the data's both-sided censored structure into account and utilize wind direction, but do not set up an autoregressive model. Instead, we use our modeling structure to generalize WPPT and use that to calculate not short-term but medium-term forecasts that are very important in day-to-day electricity trading. As a by-product, our method provides a classifier that returns probability information on future censoring, a piece of information that can be very valuable to practitioners.

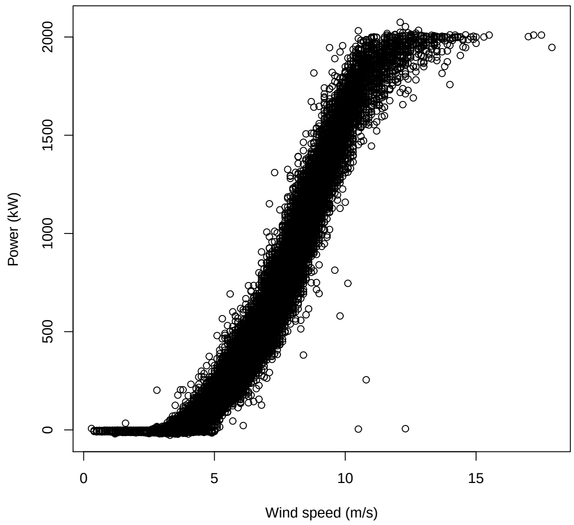


Figure 1. Example of a turbine's typical empirical power curve, i.e. the non-linear relationship between wind speed and wind power generated, Turbine B, time frame Oct. 01^{st}, 2007 to Sept. 30^{th}, 2009.

Our Model Suggestion

We focus on the interdependency of wind power and wind speed, the so-called *power curve*. Figure 1 presents an example of a turbine coded "Turbine B". Turbines A to H are all situated at specific locations in Germany that we cannot reveal, which is why we code the turbines "A" to "H". The interdependency exhibits the following properties, sometimes referred to as "stylized facts" of wind power production: The curve is non-linear, shows a sigmoidal shape, and is both-sided censored (lower censoring point at 0 kW, upper censoring point at 2,000 kW, which is the maximum power output of the turbines in our data set). Furthermore, there are hints suggesting conditional heteroscedasticity, i.e. a time-dependent volatility structure.

Wind power production possesses a diurnal periodic effect: The power production level of a certain time of day is highly correlated with the power production level of the same time one day earlier. WPPT addresses this interday persistence by means of a set of sine and cosine transformations. In addition to these transformations, WPPT uses lags of wind power as well as wind speed and the square of wind speed (due to the non-linear relationship) as the only explanatory variables in an overall linear modeling approach.

According to this, our model generalizes the WPPT idea regarding two major

[1]Typically, WAsP (Wind Atlas Analysis and Application Program) data is used.

aspects: First, our model takes the two-sided censoring into account, so it is a non-linear model. Second, we include wind direction as an additional explanatory variable. Wind comes from several directions. The relative frequencies of directions are far from being uniformly distributed, as Figure 2 shows: For the turbine investigated, wind seems to come from west-southwestern directions, mostly. Additionally, the correlation between wind power and wind speed seems to be dependent on wind direction. Figure 3 presents the correlation coefficients between wind power and wind speed conditioned on wind direction. It suggests that the correlation very much depends on wind direction. An explanation can be wind park-specific wake effects (shadowing), as analyzed by, e.g., González-Longatt et al. (2012). The four panels in Figure 4 show the distribution of wind speed by wind direction, overlain by the respective normal distribution densities of each direction's mean and standard deviation. While wind speeds from northern and eastern directions are rather mesokurtic, those from southern (weakly) and western (strongly) directions are leptokurtic. Figure 5 finally gives an example of a power curve including wind direction (azimuth) as a static 3D plot. It can be seen that wind direction fits into the dependency structure quite well. In a multivariate regression approach wind direction regularly proves to be a highly significant explanatory variable.

3 The Generalized WPPT Model

Forecasting wind power production is a two-stage process: First, the explanatory variables need to be forecasted, i.e. wind speed and wind direction. In real world applications, for example by practitioners in the industry (e.g. wind park operators), these data are typically purchased from an external source. There is a variety of services providing meteorology-based wind forecasts, for example WAsP, HIRLAM, or DWD. However, as these services are expensive (especially when data are required at a high frequency or at a high spacial resolution), a cost saving approach would be to perform univariate forecasts purely based on ARMA family stochastic processes.

Using ARMA models to forecast wind speed is well established in the literature, e.g. Giebel et al. (2011). Beblo and Schmid (2011) present evidence that the generalized ARFIMA(p, d, q) model (fractionally integrated ARMA) is able to capture the typical empirical properties of wind speed data quite well. They find best performance when choosing $p = q = 3$ and d around 0.4. However, Hyndman and Khandakar (2008) provide a flexible implementation of the (in

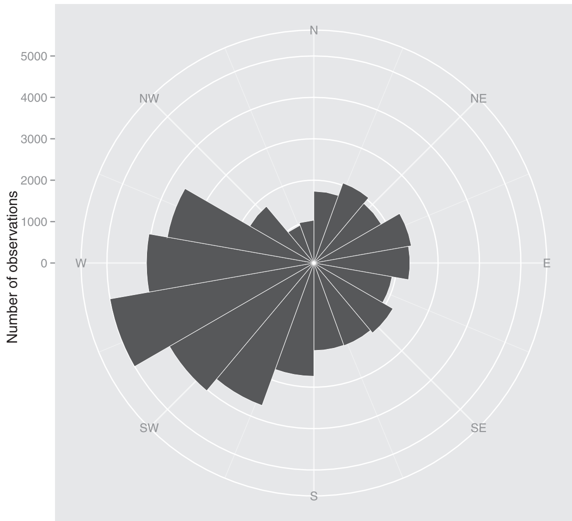


Figure 2. Absolute number of observations in several wind direction bins. Perceived wind directions at a turbine are not uniformly distributed. In this example, wind rarely comes from northern directions, Turbine B, time frame Oct. 01^{st}, 2007 to Sept. 30^{th}, 2009.

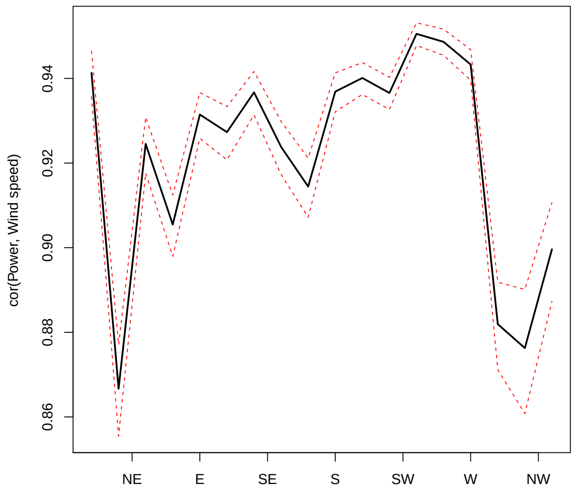


Figure 3. Correlation coefficients between wind power and wind speed, by wind direction, Turbine B, time frame Oct. 01^{st}, 2007 to Sept. 30^{th}, 2009. The curve consists of 18 azimuth segments, each 20 degrees "wide". The dashed red curves denote a 95% confidence band.

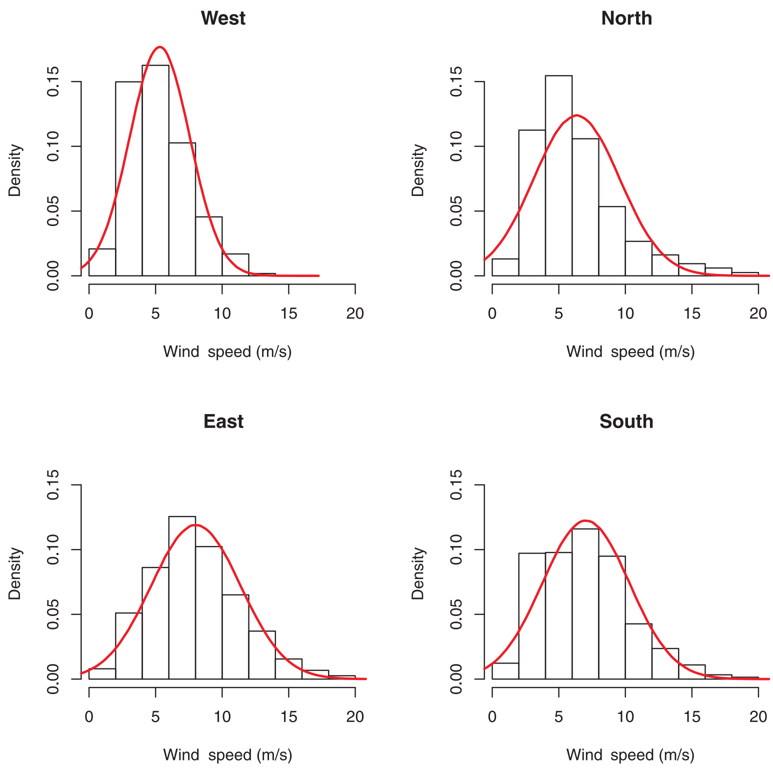


Figure 4. Histograms of wind speeds by wind direction, Turbine B, time frame Oct. 01^{st}, 2007 to Sept. 30^{th}, 2009. Red curves denote normal distribution densities at each direction's mean and standard deviation.

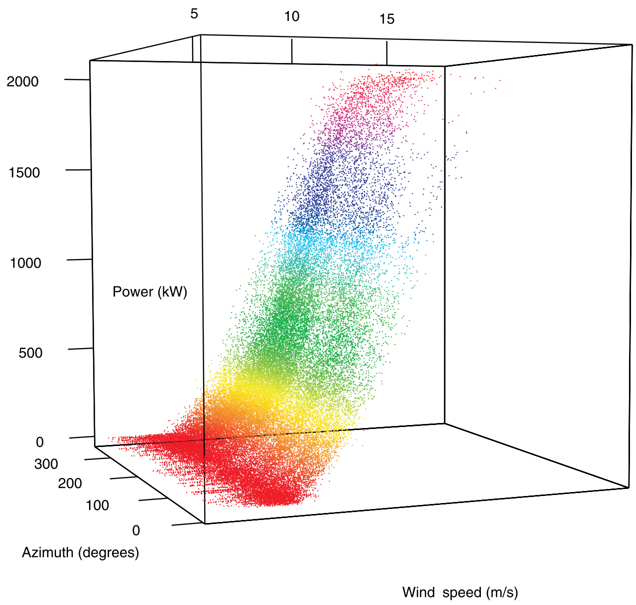


Figure 5. Empirical Power Curve, visualization extended by wind direction (azimuth), Turbine B, time frame Oct. 01^{st}, 2007 to Sept. 30^{th}, 2009. Identical colors denote identical "heights", i.e. identical power levels.

some way generalized) approach of Box and Jenkins (1970) to find proper specifications. As an example of the rather good forecasting performance, Figure 6 presents actual wind speeds vs. forecasts 12 hours ahead in a time frame of roughly eight weeks. $ME = -0.0639$ m/s, which is only $2.003 \cdot 10^{-4}$% of the total sum of errors. Analogously, wind direction forecasts imply an ME of 2.4909 degrees, $1.984 \cdot 10^{-4}$% of the total sum. Figure 7 shows a scatter plot of actual and forecasted wind speeds against each other. The blue line denotes an ideal 45° line, while the red line symbolizes the bivariate regression line. Although the lines look very much alike, a Wald test for ideal pairwise 1:1 fit of actual vs. forecasted wind speeds (intercept = 0, slope = 1) is rejected at an F-value of 62.9.

After all, we use this flexible framework to forecast wind speed data as an input for the actual wind power forecasting model. For the real world practitioner who chooses not to buy meteorology-based forecasts, the ARFIMA method is fast enough for real time forecasting, and the precision of the results is comparable to meteorology-based forecasts.

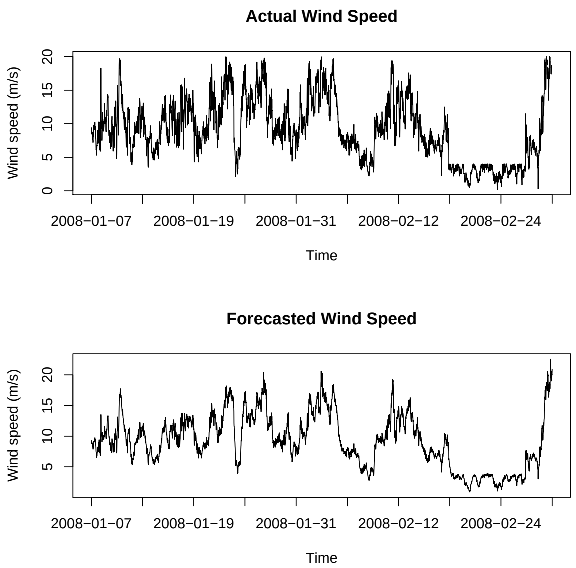


Figure 6. ARFIMA(*2, 0.4, 5*)-based wind speed forcasts, actual vs. fitted in an eight week time frame of Jan. 07^{th}, 2008 to Feb. 29^{th}, 2008, 72 steps (12 hours) ahead, Turbine B.

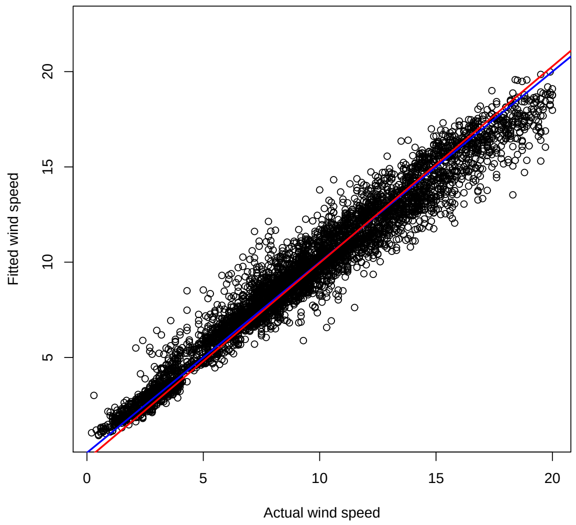


Figure 7. Actual vs. forecasted wind speeds. The blue line is an ideal 45° line, the red is the actual OLS line. Turbine B, time frame Oct. 01^{st}, 2007 to Sept. 30^{th}, 2009.

Censored Regression

As mentioned in section 2, one of the ways in which our model (henceforth "GWPPT", *Generalized* WPPT) generalizes the idea of WPPT is that it makes use of a piece of additional ex ante known information: The fact that produced power lies inside predetermined boundaries. We assume that the data investigated lie within an interval of $p_t \in [l, u]$ kW $\forall\ t$ per turbine, where p_t is the power produced at time t.
As the data are two-sided censored we impose a censored regression model (a generalized Probit model by Rosett and Nelson, 1975) as estimator for the forecasting procedure. The model imposes the following structure on wind energy production:

$$p_t^* = \mu(\mathbf{x}_t) + \varepsilon_t, \tag{1}$$

where $\mathbf{x}_t$ is a vector of explanatory variables, μ is an assumed linear function of $\mathbf{x}_t$, and ε_t is an assumed Gaussian error term. For WPPT, it is simply assumed that $p_t = p_t^*$, whereas GWPPT generalizes that idea and imposes a censored data structure, so that

$$p_t = \begin{cases} l, & p_t^* \leq l \\ p_t^*, & p_t^* \in (l, u) \\ u, & p_t^* \geq u. \end{cases} \tag{2}$$

l and u are the lower and upper censoring points and

$$\mu(\mathbf{x}_t) = \sum_{j=1}^{m} \beta_j x_{t,j} \tag{3}$$

with $\mathbf{x}_t = x_{1,t}, \ldots, x_{m,t}$ and $x_{1,t} = 1\ \forall\ t = 1, \ldots, T$.

Estimation of the parameters is performed via a combined Least-Squares-/Maximum-Likelihood-Approach.[2] Chung and Goldberger (1984) point out that (generalized) Probit models provide asymptotically unbiased estimates, even in the presence of heteroscedasticity. However, Arabmazar and Schmidt (1981) show that censored regression is prone to inconsistency when the data investigated is heteroscedastic and a comparably large fraction of the data is

[2]Another method for estimation is, for instance, the Markov Chain Monte Carlo method (MCMC) described by Gill (2008).

censored (about half of the observations or more). Still, the magnitude of the inconsistency is smaller than for a simply truncated model. Also, far fewer than half of the observations need to be censored for GWPPT, so we assume the inconsistency not to be too severe. Nevertheless, future models need to take heteroscedasticity into account to achieve guaranteed consistent estimates.

Specification

As the model's specification we basically put wind direction into the WPPT model as an additional variable and then run it through a censored regression estimator. In lag notation (using k as a lag parameter, i.e. the forecasting horizon is k steps ahead), the specification is defined as

$$p_t^* = m + a_1 \cdot p_{t-k} + a_2 \cdot p_{t-(k+1)} + b_1 \cdot w_{t|t-k} + b_2 \cdot (w_{t|t-k})^2 + c_1 \cdot v_{t|t-k} + d_1^c \cdot \cos\left(\frac{2\pi d_t}{144}\right) + d_2^c \cdot \cos\left(\frac{4\pi d_t}{144}\right) + d_1^s \cdot \sin\left(\frac{2\pi d_t}{144}\right) + d_2^s \cdot \sin\left(\frac{4\pi d_t}{144}\right) + \varepsilon_t, \quad (4)$$

where p_t^* is power produced at time t, $w_{t|t-k}$ is wind speed at time t given at time $t-k$, v_t is wind direction at time t, and d_t is time of day for observation t. The right-hand side of this model equation represents the function $\mu(\cdot)$ in equation (1). (G)WPPT uses the Fourier series (the four sine and cosine terms) to control for daytime. In that way, it captures the diurnal periodic compound of wind power production mentioned above: Since our data is based on an interval of ten minutes, there are 144 obs. per day, so the sine and cosine terms are modulated to oscillate once per day.

Contrary to the original WPPT specification, we set parameters a_1 and a_2 to zero, i.e. we do not utilize the lagged dependent variable in the specification. WPPT was developed to produce short-term horizon forecasts. That model benefits from an included lagged dependent variable because it helps to capture persistency effects in the short run. In our medium-term scenario, however, we discovered how to attain better results not taking the lagged dependent variable into account, especially if the forecasting horizon extends up to several hours. To smoothen the parametric conditions between short- and medium-term scenarios, Tastu et al. (2010) show how to condition a model's parametrization on the forecasting horizon.

For the WPPT model, the above completes the forecasting process, since for WPPT, $p_t = p_t^*$. However, a real world operator does know the turbine's

power range. Since WPPT frequently returns $p_t^* < l$ or $p_t^* > u$, those forecasts are assumed to be "flattened" to limit the extent of forecasting errors: $p_t = l \quad \forall \quad p_t^* < l$ and $p_t = u \quad \forall \quad p_t^* > u$.
For GWPPT instead, we estimate the conditional mean of p_t according to p_t^* as modeled by equation (2). After estimating the parameters of the GWPPT model we make use of the assumption of normally distributed errors and extract the forecasted power production as the conditional mean $\widehat{p_t}\left(\mathbf{x}_t\right) := E\left[p_t^*|\mathbf{x}_t\right]$, according to equation (1). Using this framework we obtain the forecasted power level as

$$\widehat{p_t} = (\Phi(f_2) - \Phi(f_1)) \cdot p_t^* + (\phi(f_1) - \phi(f_2)) \cdot \widehat{\sigma} + u \cdot (1 - \Phi(f_2)), \tag{5}$$

where

$$f_1 = \frac{l - p_t^*}{\widehat{\sigma}}, \tag{6}$$

$$f_2 = \frac{u - p_t^*}{\widehat{\sigma}}, \tag{7}$$

and $\phi(\cdot)$ and $\Phi(\cdot)$ denote normal PDF (Probability Density Function) and CDF (Cumulative Distribution Function), respectively.
As a by-product of the conditional mean function (5), we obtain CDF and PDF for upper and lower censoring, respectively. $\Phi(f_1)$ should be close to one whenever actual power output data is rather small, i.e. the forecasting algorithm predicts lower censoring. $\Phi(f_2)$ should be close to zero in those cases and vice versa. As an aggregated and more intuitive measure, we calculate a standardized sum of these indicators and subtract it from one:

$$C = 1 - \frac{\Phi(f_1) + \Phi(f_2)}{2}. \tag{8}$$

Therefore, the resulting "probability" measure should be close to one whenever the algorithm forecasts upper censoring and close to zero when lower censoring is forecasted. When no censoring is forecasted, the measure should be close to 0.5.
This measure (or, alternatively, a PDF-based measure) can be extended to serve as a censoring classifier which can be used as a censoring forecaster. The classifier can be applied to upper censoring or lower censoring separately, if needed. For example, whenever C from equation (8) exceeds a given threshold (e.g., 0.9), it can be assumed that the given forecast will be censored at the upper

bound. This way, a very simple classifier can forecast the class of censoring, and equation (8) provides the respective probability. We investigate the predictive power of this classifier in the text below and find that it is surprisingly precise, given its simplicity.

4 GWPPT in Practice

As a common measure for the comparison of wind power production forecasts, Giebel et al. (2011) suggest using standardized Root Mean Squared Errors (RMSE), henceforth called sRMSE. The method is widely accepted in the literature, so we use sRMSE to compare GWPPT to WPPT and to a naïve ($\widehat{p_t} = p_{t-k}$) forecast, which is a common benchmark in the literature (e.g. Costa et al., 2008, or Giebel et al., 2011). For small k, the naïve forecaster is known to be hard to outperform because of the high degree of persistence in the actual power levels. Pinson (2012) argues:

> *...the persistence benchmark indeed seems competitive: The more advanced approaches only propose overall improvements up to 5%.*

Furthermore, it is rather common to compare a forecast to the theoretical power curve the turbine producer provides. The International Electrotechnical Commission (IEC) obliges any producer to publish a characteristic profile of the system. In our case, the system is the Vestas V90-2MW turbine. The theoretical power curve is plotted in the product brochure[3], and actual values can be taken from the technical description. [4] However, as for our turbines, this curve seems to be systematically biased since it underestimates the power produced for almost all wind speeds, as the example of Figure 8 reveals. As can be seen, the theoretical profile hardly provides a good description of actual observations. Reasons might be above-average air densities at the turbine's location or just a prudent estimation of the theoretical power output. In any case, this curve will obviously not be a good forecaster, so we do not take this benchmark into further consideration.

As for the data we have, $p_t \in [0; 2,000]$ kW $\forall\ t$ per turbine, since 2,000 kW (or 2 MW) is the maximum power output of Vestas V90-2MW turbines. For all turbines to which we have access, sensor data in the time window October

[3] Cf. http://nozebra.ipapercms.dk/Vestas/Communication/Productbrochure/2MWbrochure/2MWProductBrochure/

[4] Cf. www.horizonwindfarms.com/northeast-region/documents/under-dev/arkwright/Exhibit7_GeneralSpecifications.pdf

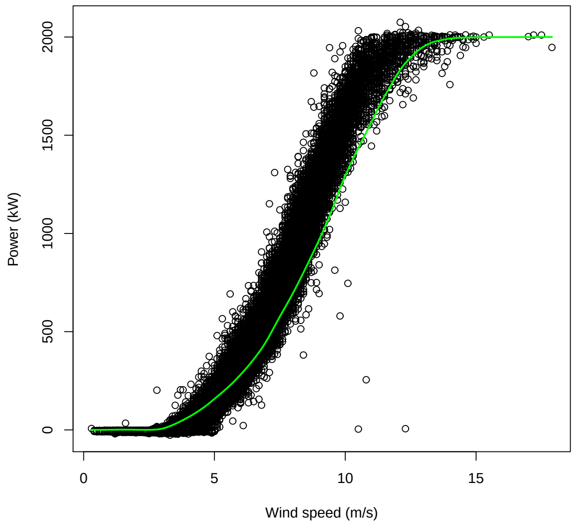


Figure 8. Theoretical Power Curve as given by the International Electrotechnical Commission, reported by turbine producer Vestas (Turbine B), Turbine B, time frame Oct. 01^{st}, 2007 to Sept. 30^{th}, 2009.

1^{st} 2007 to September 30^{th} 2009, i.e. for two full years, have been observed. Observations are reported at a ten-minute frequency, so we have 105,264 observations of variables such as wind speed, wind direction, actually produced power and others.

In-Sample Performance

As a preparation for the analysis of real world scenarios, we first analyze the model's in-sample performance, i.e. its behavior under actually observed conditions. Therefore, we pick a representative, but shorter, time frame out of the total sample and feed the forecasting model not *estimated* explanatory variables, but *observed* ones. All forecasts are calculated for a 12-hour-horizon, i.e. 72 steps ahead. If, in comparison to forecasts given estimated explanatory variables, GWPPT performs better than WPPT, we can switch to a more realistic scenario and analyze the total data set out-of-sample. Additionally, given this small "sandbox" framework, we analyze the model's performance conditioning on wind direction, which is our additional explanatory variable not utilized in the WPPT benchmark model.
Figure 9 shows the power curve performance of the GWPPT vs. WPPT models

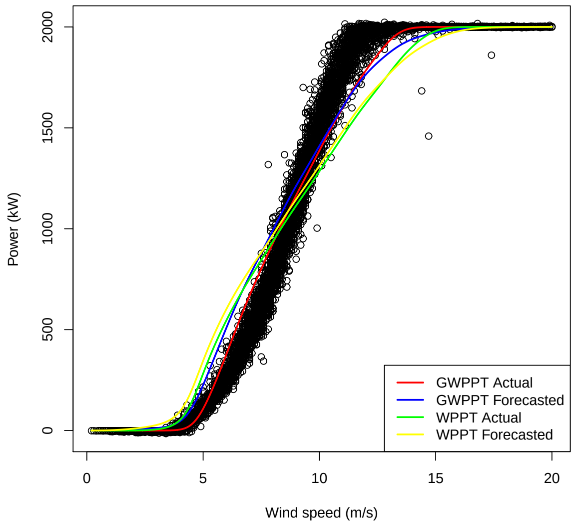


Figure 9. WPPT and GWPPT performance in a power curve, actual vs. forecasted explanatory variables, Turbine A, time frame Jan. 07^{th}, 2008 to Feb. 29^{th}, 2008.

in a roughly eight-week time frame, using both actual and forecasted explanatory variables, respectively. Both GWPPT curves are comparably close by and diverge from the data support area much less than the WPPT curves do (overestimation for 4-8 m/s, underestimation for 11-14 m/s). The same holds true for estimations conditioned on wind direction. For example, Figure 10 presents curves for eastern wind directions only. Table 1 presents RMSE, sRMSE, and relative RMSE differences for several turbines, by wind direction. It provides numerical evidence that WPPT is outperformed by GWPPT in-sample by far. Table 2 provides numbers of observations in total, numbers of lower censoring observations and numbers of upper censoring observations (and percentages, respectively) per quadrant for these turbines, in-sample (and out-of-sample).

Results so far show that GWPPT performs a lot better than WPPT in-sample, that wind direction matters, and also that GWPPT performs better than WPPT when controlling for wind direction. Since everything else is identical in that case, the remaining improvement must be due to GWPPT censoring. However, in-sample forecasting is not realistic. Out-of-sample improvements are not as significant, but still very important, as table 3 shows. Again, table 2 provides numbers of observations in total, numbers of lower censoring observations, and numbers of upper censoring observations (and percentages, respectively) per quadrant for these turbines, out-of-sample (and in-sample).
We will move on to a more realistic out-of-sample evaluation in the text below.

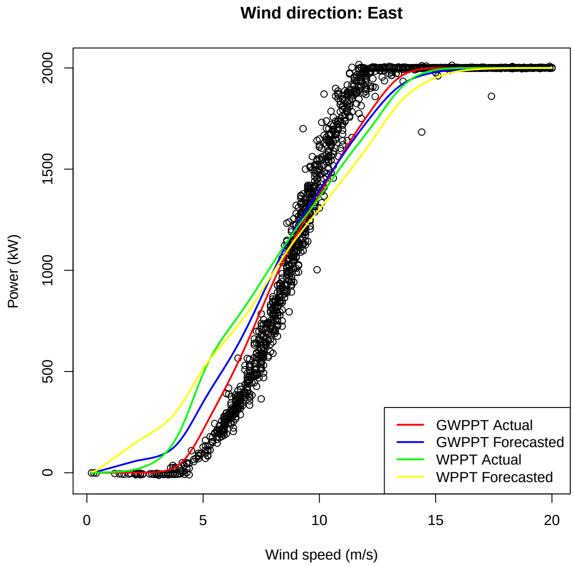


Figure 10. WPPT and GWPPT performance in a power curve, actual vs. forecasted explanatory variables, eastern wind directions only, Turbine A, time frame Jan. 07^{th}, 2008 to Feb. 29^{th}, 2008.

	A: RMSE	A: sRMSE	A: rel. RMSE	B: RMSE	B: sRMSE	B: rel. RMSE	C: RMSE	C: sRMSE	C: rel. RMSE
North:									
GWPPT	68.3985	0.0342	.	126.4946	0.0633	.	95.9762	0.0480	.
WPPT	138.6348	0.0693	202.69%	229.5685	0.1148	181.48%	216.3064	0.1082	225.38%
Persistence	363.3667	0.1817	531.24%	759.1176	0.3796	600.11%	625.5123	0.3128	651.73%
East:									
GWPPT	124.5942	0.0623	.	141.5651	0.0708	.	118.0012	0.0590	.
WPPT	230.0630	0.1150	184.65%	296.9725	0.1485	209.77%	224.5130	0.1123	190.26%
Persistence	747.4903	0.3737	599.93%	808.2365	0.4041	570.93%	603.5360	0.3018	511.46%
South:									
GWPPT	118.5263	0.0593	.	88.5933	0.0443	.	109.9002	0.0550	.
WPPT	217.6128	0.1088	183.59%	309.7825	0.1549	349.66%	232.1607	0.1161	211.24%
Persistence	793.3232	0.3967	669.32%	624.1250	0.3121	704.48%	428.8351	0.2144	390.20%
West:									
GWPPT	91.6265	0.0458	.	196.4569	0.0982	.	107.3532	0.0537	.
WPPT	168.3580	0.0842	183.74%	306.0273	0.1530	155.77%	210.3827	0.1052	195.97%
Persistence	363.8412	0.1819	397.09%	784.1360	0.3920	399.14%	412.4097	0.2062	384.16%

Table 1. WPPT and GWPPT performance at an actual explanatory variables scenario (in-sample) for turbines A, B, and C, time frame Jan. 07^{th}, 2008 to Feb. 29^{th}, 2008. RMSE = Root Mean Squared Error, sRMSE = standardized RMSE, rel. RMSE = relative increase of RMSE vs. RMSE of GWPPT.

	A (total)	A (%)	B (total)	B (%)	C (total)	C (%)
North:						
Total	1260		4064		2201	
Lower censored	361	28.65%	416	10.24%	434	19.72%
Upper censored	0	0%	842	20.72%	12	0.55%
In-Sample						
Lower censored	174	13.81%	425	10.46%	509	23.13%
Upper censored	0	0%	241	5.93%	0	0%
Out-of-Sample						
Lower censored	124	9.84%	345	8.49%	272	12.36%
Upper censored	0	0%	333	8.19%	0	0%
East:						
Total	3778		1503		1716	
Lower censored	492	13.02%	78	5.19%	339	19.76%
Upper censored	309	8.18%	435	28.94%	39	2.27%
In-Sample						
Lower censored	385	10.19%	83	5.52%	392	22.84%
Upper censored	148	3.92%	129	8.58%	2	0.12%
Out-of-Sample						
Lower censored	182	4.82%	77	5.12%	406	23.66%
Upper censored	118	3.12%	108	7.19%	0	0%
South:						
Total	1347		1241		1584	
Lower censored	202	15.00%	607	48.91%	492	31.06%
Upper censored	126	9.35%	71	5.72%	2	0.13%
In-Sample						
Lower censored	114	8.46%	610	49.15%	523	33.02%
Upper censored	71	5.27%	29	2.34%	0	0%
Out-of-Sample						
Lower censored	98	7.28%	516	41.58%	565	35.67%
Upper censored	8	0.59%	2	0.16%	0	0%
West:						
Total	1542		1118		2426	
Lower censored	319	20.69%	322	28.80%	726	29.93%
Upper censored	14	0.91%	32	2.86%	0	0%
In-Sample						
Lower censored	211	13.68%	409	36.58%	766	31.57%
Upper censored	2	0.13%	67	5.99%	0	0%
Out-of-Sample						
Lower censored	229	14.85%	378	33.81%	503	20.73%
Upper censored	0	0%	44	3.94%	0	0%

Table 2. Numbers of total observations, numbers of lower censored observations and numbers of upper censored observations (and percentages, respectively) per quadrant for turbines A, B, and C, time frame Jan. 07^{th}, 2008 to Feb. 29^{th}, 2008. Actual, in-sample GWPPT and out-of-sample GWPPT.

	A: RMSE	A: sRMSE	A: rel. RMSE	B: RMSE	B: sRMSE	B: rel. RMSE	C: RMSE	C: sRMSE	C: rel. RMSE
North:									
GWPPT	365.3935	0.1827	.	714.0559	0.3570	.	617.8390	0.3089	.
WPPT	371.8285	0.1859	101.7611%	710.0700	0.3550	99.4418%	604.5562	0.3022	97.8501%
Persistence	363.3667	0.1817	99.4453%	759.1176	0.3795	106.3107%	625.5123	0.3128	101.2420%
East:									
GWPPT	723.6864	0.3618	.	774.8392	0.3874	.	575.3244	0.2877	.
WPPT	712.9572	0.3565	98.5174%	759.7874	0.3799	98.0574%	576.9933	0.2885	100.2901%
Persistence	747.4903	0.3737	103.2893%	808.2365	0.4041	104.3102%	603.5360	0.3018	104.9036%
South:									
GWPPT	716.5345	0.3583	.	566.4219	0.2832	.	430.7469	0.2154	.
WPPT	701.9820	0.3501	97.9690%	594.0912	0.2970	104.8849%	447.7656	0.2239	103.9510%
Persistence	793.3232	0.3967	110.7167%	624.1250	0.3121	110.1873%	428.8351	0.2144	99.5562%
West:									
GWPPT	338.1057	0.1690	.	210.1800	0.1051	.	435.5251	0.2178	.
WPPT	366.7044	0.1833	108.4585%	264.8417	0.1324	126.0071%	455.4264	0.2277	104.5695%
Persistence	363.8412	0.1819	107.6117%	273.7148	0.1369	130.2301%	412.4097	0.2062	94.6925%

Table 3. WPPT and GWPPT performance at a forecasted explanatory variables scenario (out-of-sample) for turbines A, B, and C, time frame Jan. 07^{th}, 2008 to Feb. 29^{th}, 2008. RMSE = Root Mean Squared Error, sRMSE = standardized RMSE, rel. RMSE = relative increase of RMSE vs. RMSE of GWPPT.

sRMSE Comparison - An Evaluation of Forecasting Precision

Since we have data at a ten-minute frequency and the most important forecasting horizons range from 12 to 36 hours ahead, we calculate out-of-sample forecasts up to 216 steps (36 hours) ahead using a rolling window of fixed size throughout the whole data set. Figure 11 presents the development of sRMSE at turbine A between about 12 and 36 hours ahead. As can be seen, WPPT clearly outperforms the persistence forecaster. Moreover, GWPPT performs somewhat better than WPPT, i.e. sRMSE are lower for all forecasting horizons. We double-check the robustness of these findings with turbines B to H, and results are very similar. Turbines A to C are all located at one wind park, turbines D to F are located at another park, and turbines G and H belong to a third wind park. sRMSE plots are shown in Figures 12a to 13c. Table 4 presents numerical evidence of relative forecasting power improvement, i.e. relative reduction in sRMSE from one forecaster to another.[5] Panel A shows that the GWPPT forecaster outperforms the persistence forecaster in most of the cases, except for the 1 step forecasting horizon. In panel B it can be seen that GWPPT severely improves forecasting performance (i.e. lower sRMSE) over the plain WPPT model. Especially for turbines D to F, WPPT seems to be prone to heavy censoring and therefore performs worse than persistence even for longer forecasting horizons as can be seen in panel C. GWPPT can do a lot better here.

Compared to WPPT, GWPPT improves forecasting precision by more than five percentage points in most cases. While this may seem to be only a limited success, it should be kept in mind that in the competition of wind power forecasting, many models (such as the rather sophisticated WPPT) are already very mature and prove to be tough competitors.

Still, numerically small improvements in terms of sRMSE percentage points usually result in quite substantial monetary benefits. Carl Hilger, Operations Manager of Eltra (the antecessor of Energinet.dk), is cited by Giebel et al. (2011) for this statement:

> *»If only we improved the quality of wind power forecasts with one percentage point, we would have a profit of two million Danish crowns.« Similar orders of magnitude are quoted infrequently by other utilities or traders, but usually not for publication.*

[5] For convenience, positive values mean “better” in this setting.

	1 step	72 steps	144 steps	216 steps
Panel A: GWPPT vs. persistence				
Turbine A	-1.7463%	5.4020%	5.7217%	5.4997%
Turbine B	-2.2748%	3.6417%	3.9640%	3.5651%
Turbine C	-0.7451%	3.0983%	1.2932%	0.4890%
Turbine D	-15.9569%	5.4569%	3.3089%	-0.5181%
Turbine E	-16.7833%	4.3372%	2.6988%	-0.3105%
Turbine F	-16.2344%	5.3569%	3.3826%	-0.3068%
Turbine G	-4.9628%	2.9315%	3.7986%	3.8203%
Turbine H	-2.0976%	5.1821%	5.0271%	3.7107%
Panel B: GWPPT vs. WPPT				
	1 step	72 steps	144 steps	216 steps
Turbine A	-0.4729%	0.9508%	1.2832%	0.8249%
Turbine B	-1.3476%	0.3625%	-1.5940%	-2.7703%
Turbine C	1.3202%	-0.8390%	-0.4717%	-1.1052%
Turbine D	3.1559%	5.1436%	5.7490%	6.3705%
Turbine E	2.9162%	5.9046%	6.1230%	6.0821%
Turbine F	3.7082%	5.9356%	5.4203%	7.1821%
Turbine G	-0.3587%	0.7765%	-0.7307%	-1.3308%
Turbine H	13.7785%	7.5781%	3.4280%	5.0408%
Panel C: WPPT vs. persistence				
	1 step	72 steps	144 steps	216 steps
Turbine A	-1.2674%	4.4939%	4.4962%	4.7136%
Turbine B	-0.9149%	3.2911%	5.4708%	6.1646%
Turbine C	-2.0929%	3.9046%	1.7566%	1.5767%
Turbine D	-19.7357%	0.3303%	-2.5889%	-7.3573%
Turbine E	-20.2912%	-1.6658%	-3.6475%	-6.8066%
Turbine F	-20.7106%	-0.6152%	-2.1545%	-8.0685%
Turbine G	-4.5877%	2.1718%	4.4964%	5.0835%
Turbine H	-18.4131%	-2.5925%	1.6559%	-1.4007%

Table 4. Relative Forecasting Power improvement by sRMSE. Panel A provides pairwise comparisons of GWPPT vs. persistence forecasts for several forecasting horizons (columns) and several example turbines (rows). Panel B does the same for pairwise GWPPT vs. WPPT comparisons and panel C compares WPPT vs. persistence forecasts. Positive values denote “better” forecasts, i.e. lower sRMSE. Estimation sample spans from Oct. 01^{st}, 2007 to Dec. 12^{th}, 2008.

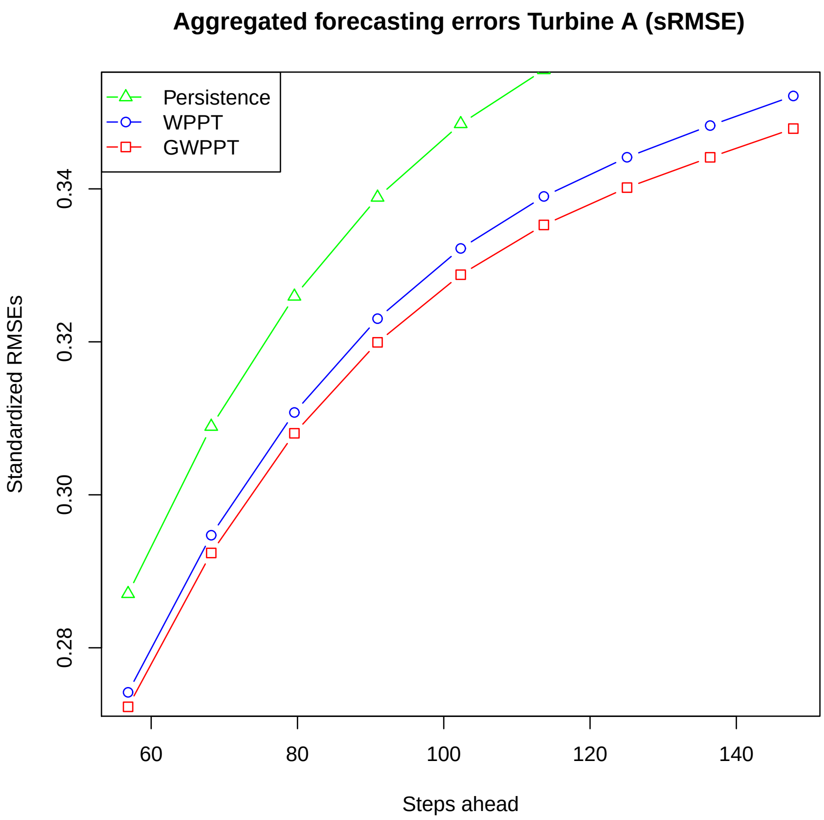


Figure 11. Standardized RMSE profile for several forecasting horizons, persistence vs. WPPT vs. GWPPT predictor, Turbine A, estimation sample spans from Oct. 01^{st}, 2007 to Dec. 12^{th}, 2008.

We complement this statement by a model calculation and find that the financial impact of an improvement of forecasting precision by a few percentage points can be tremendous.

A Time Series Comparison

Figure 14 presents a comparison of actual power production against forecasted production using WPPT and GWPPT, 72 steps (i.e. 12 hours) ahead on turbine A. The time series plot consists of the first 6,000 obs. of the data set (about 6 weeks) to make the effects visible in greater detail. Whenever actual power is high, the blue WPPT curve is located below the red GWPPT curve and vice versa and whenever actual power is low, the blue WPPT curve lies above the red GWPPT curve. This behavior suggests that not respecting censoring, as it is done for WPPT, leads to biasedness in the boundary areas. Furthermore, there are several differences between WPPT and GWPPT in addition to the censoring point areas. In the area of obs. 3,800, for example, actual power experiences a drastic drop. WPPT follows that drop to a degree, but in a far more restrained manner than the actual power drop turns out. GWPPT, however, still does not follow the drop as drastically as the actual power level, but more clearly than WPPT does. Another example of this behavior can be seen in the area of obs. 4,400. Possible explanations for this behavior can be the inclusion of wind direction as an additional explanatory variable, but also the different slope of the estimator due to the censored regression model.

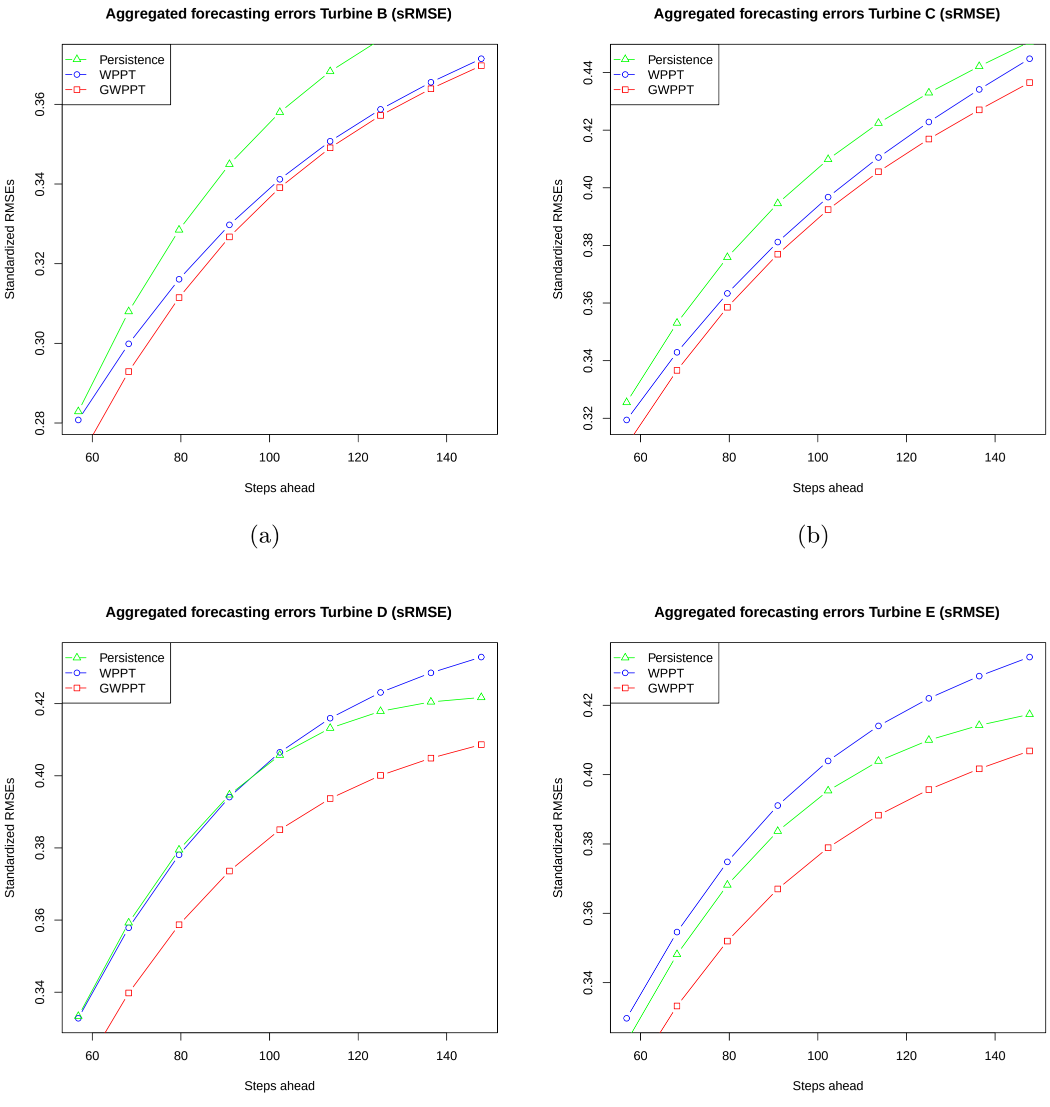
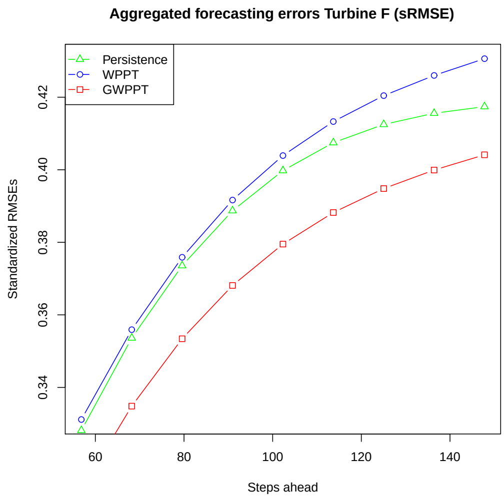
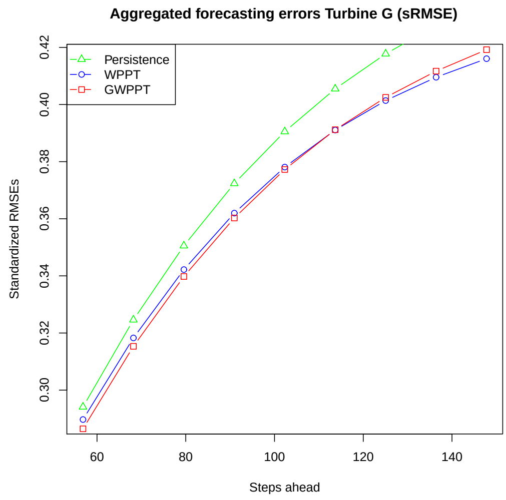


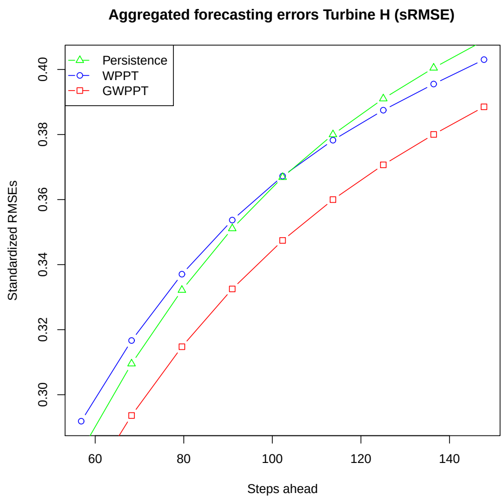
Figure 12. Standardized RMSEs, Turbines B to E, estimation sample spans from Oct. 01^{st}, 2007 to Dec. 12^{th}, 2008.



(a)



(b)



(c)

Figure 13. Standardized RMSE, Turbines F to H, estimation sample spans from Oct. 01^{st}, 2007 to Dec. 12^{th}, 2008.

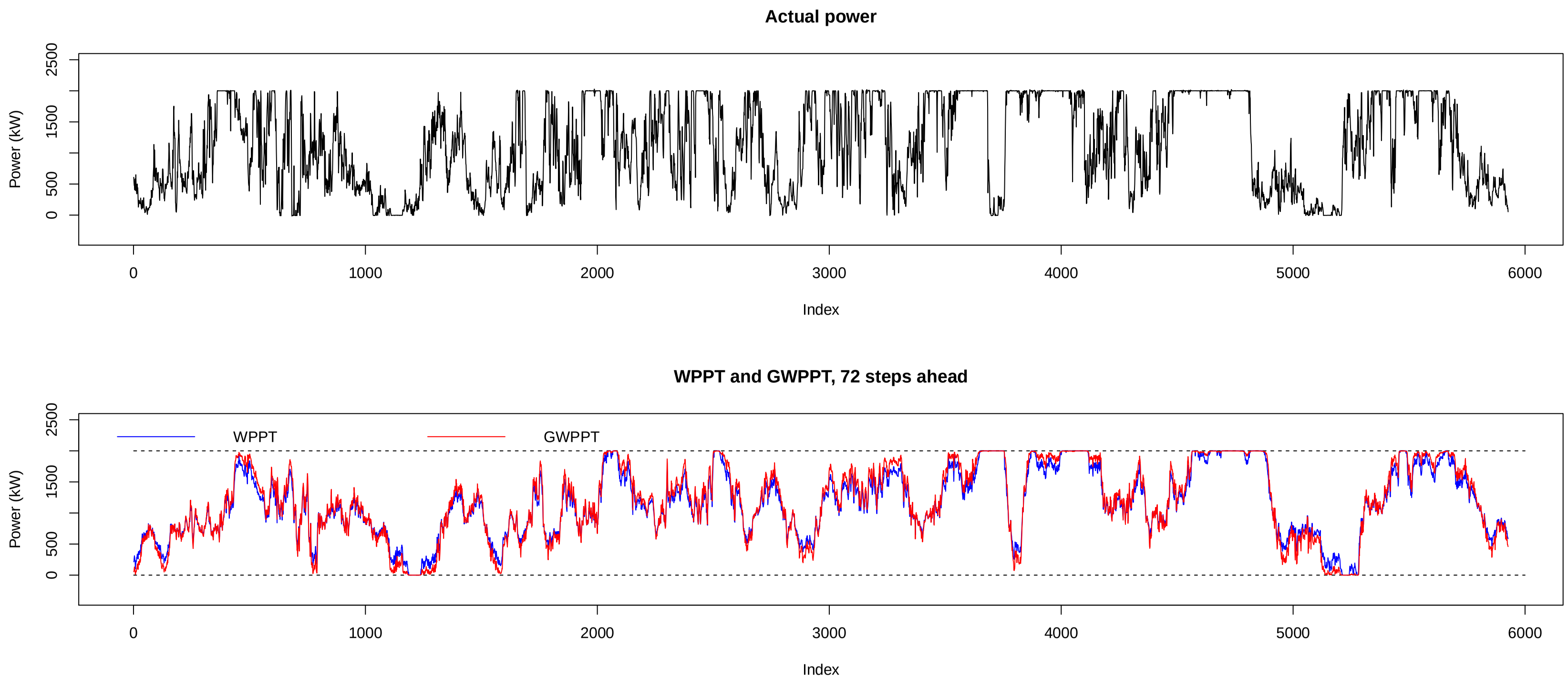


Figure 14. Comparison of forecasted wind power using WPPT and GWPPT against actually produced power over time, Turbine A, time frame Oct. 01^{st}, 2007 to Nov. 13^{th}, 2007.

	1 step	72 steps	144 steps	216 steps
Correct Censoring	10.06%	4.49%	2.89%	2.33%
Correct non-Censoring	77.54%	71.28%	70.00%	69.70%
False Positive	0.40%	6.76%	8.25%	8.74%
False Negative	12.00%	17.47%	18.86%	19.23%
Sums	100.00%	100.00%	100.00%	100.00%

Table 5. False and correct censoring classification for several forecasting horizons (columns), Turbine A, estimation sample spans from Oct. 01^{st}, 2007 to Dec. 12^{th}, 2008.

Evaluating Censoring Forecasts

As each single forecast returned by GWPPT may or may not be censored, there can be

1. correct censoring (correct positive, i.e. a forecast is censored when it in fact has to be censored [that is, actual power levels reach 2,000 or 0 kW and so does the respective forecast]),
2. type one errors (false negative, i.e. not censored although actual power levels reach 2,000 or 0 kW),
3. type two errors (false positive, i.e. censored although actual power levels are within the $[0; 2,000]$ kW boundary), and
4. correct decisions not to censor (correct negative, i.e. not censored while actual power levels are in fact within the $[0; 2,000]$ kW boundary).

Table 5 presents generic results for turbine A. As expected, the relative number of aggregated false censoring classification increases when the forecasting horizon increases. Still, the large magnitude of correct censoring decisions emphasizes the importance of respecting the power range of the turbines while forecasting.

As pointed out in section 3, a censoring forecaster comes along with the methodology. Equation 8 can be used as a measure of censoring probability. Figure 15 shows actual data upper censoring and such forecasts that are classified as upper censoring in comparison. In this figure, one means "upper censoring", and zero means "no upper censoring". 81% of all upper censoring classifications are ex-post correctly classified. For lower censoring, even 94.2% of all classifications are correct. For total censoring (lower *and* upper), 87.6% of all censoring forecasts are correct, which is in accordance with the findings

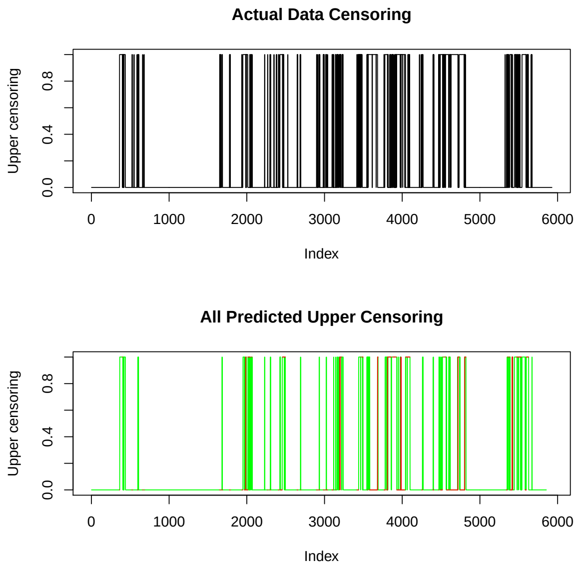


Figure 15. Upper censoring forecasts. Values of one refer to upper censoring forecasts, values of zero refer to no forecasted upper censoring. Actual data upper censoring is displayed for reference (upper panel). Turbine B, time frame Oct. 01^{st}, 2007 to Nov. 13^{th}, 2007. Green lines denote correct upper censoring, red lines denote false classification of upper censoring.

presented in table 5.

As a consequence, we conclude that the method introduced provides a classifier for lower and upper censoring which forecasts censoring rather precisely. If only censoring is to be forecasted (either "censoring: yes/no" or censoring probability), but not power itself, the classifier which comes as a by-product of our method can be applied.

A Peek on Goodness of Fit

Good fit is not a goal of this paper. A virtually perfectly fitting curve can be found easily, e.g. by using simple kernel smoothing methodology. However, as overfitting usually leads to bad forecasting performance, fitting is not the subject here, but forecasting performance is. Still, we briefly look at the estimators' goodness of fit.

Figure 16 presents simplified (i.e. bivariate), yet typical, visualizations of WPPT and GWPPT estimators in comparison. First of all, the blue WPPT estimator ($R^2 = 0.8711$) comes with very wide confidence bands, represented by the dashed blue lines.[6] Especially the right-hand side of the confidence interval seems to be unnecessarily wide, most likely caused by outliers (as compared to the most concentrated support area of the data, i.e. the wide "band" of observations) at the topmost left area and the fact that there are many observations located to the right of the estimator in the area around 4 to 8 m/s. The red GWPPT estimator (Pseudo-$R^2 = 0.9609$), however, presents rather narrow confidence bands. Bands are asymmetric and get very close to the actual estimator in the inner area, i.e. between about 5 and 10 m/s. As expected, the distance between right-hand side confidence band and estimator becomes zero where the estimator follows the lower censoring of the data. The left-hand side confidence band does just the same when there is upper censoring. Due to the conditional mean being used as described in section 3, the GWPPT curve is non-linear, which makes a comparably good fit quite possible. However, the most important advantage of censored regression over a purely linear model is that taking censoring into account results in a "steeper" slope of the curve, enabling it to fit the data much better than the linear model can.

Financial Impact

Holttinen (2005) gives insight into the Dutch electricity market and the asymmetry of the loss function of wind power trading. To sum up, Holttinen states that in his somewhat simplified example, power sold by contract in advance is traded for around $P_C = 100$ EUR/MWh. Any additional MWh delivered beyond the contracted amount returns a revenue of $P_S = 16$ EUR/MWh (e.g. at the spot market). Any MWh contracted but not delivered needs to be repayed in addition to a contract penalty of $P_P = 20$ EUR. Therefore, the opportunity loss of a produced but not contracted MWh is 84 EUR, while the loss of a contracted but not delivered MWh is 20 EUR. In the end, the economic loss of forecasting impreciseness is

$$L_t\left(P_C, P_S, P_P, e_t\right) = \begin{cases} (P_C - P_S) \cdot e_t & e_t \geq 0 \\ P_P \cdot |e_t| & e_t < 0, \end{cases} \tag{9}$$

[6] Confidence bands stem from parametric bootstrap methodology with $M = 19$ simulation rounds.

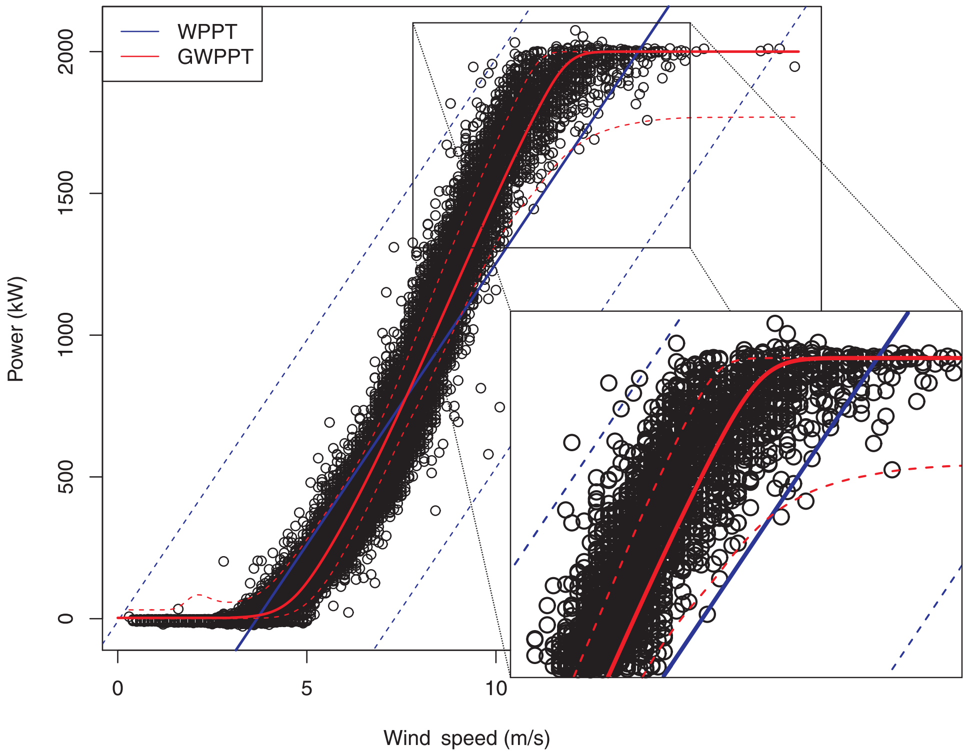


Figure 16. WPPT versus GWPPT estimator fit. The (blue) linear WPPT model fits the data rather badly, while the non-linear GWPPT model (red) provides a better fit as well as more narrow and asymmetric confidence bands (dashed, respectively), Turbine B, time frame Oct. 01^{st}, 2007 to Sept. 30^{th}, 2009.

Turbine	Monetary gain	Percentaged monetary gain
Turbine A	282,198.68 €	13.12%
Turbine B	301,475.47 €	12.52%
Turbine C	247,510.60 €	8.96%

Table 6. Absolute and relative monetary gain of GWPPT over WPPT for 72 steps ahead forecasts.

where e_t is the respective forecasting error. This model depicts a simplified (since static and linear) asymmetric loss function as also found in the NASDAQ OMX Commodities financial energy market (formerly Nord Pool). In practice, numbers may vary, but the example given is still a good approximation of average market conditions.

We run GWPPT as well as WPPT errors through that model and calculate the monetary losses of forecasters. In that way, we quantify the financial impact of using GWPPT instead of (hypothetically) using WPPT, i.e. we seek to find the cost reduction of increased forecast precision by using GWPPT. For the turbines located at one wind park, we calculate the monetary cost of impreciseness of WPPT and GWPPT, the difference is the monetary gain of GWPPT over WPPT. Table 6 presents the results over the whole span of observations, i.e. two full years, and at a forecasting horizon of 72 steps ahead. Since each turbine can gain around 250,000 euros (on average) during the two years of observation, it should be safe to say that applied to the whole wind park consisting of ten of these identical turbines, the park could gain at least an additional 1.25 million euros *per year* just by using GWPPT instead of state-of-the-art forecaster WPPT. This number emphasizes the financial impact of more precise forecasts even though the relative preciseness gain is only a few percentage points, as argued before.

5 Conclusion

This paper introduced a wind power forecasting model (GWPPT), which usually operates at a turbine-specific level, but can also be applied to wind parks and whole regions of countries. The method generalizes the idea behind a well established forecasting approach (WPPT) in two ways: It takes the two-sided censoring of empirical data into account and also introduces wind direction as an additional variable to the specification. We were able to show that GWPPT outperforms the WPPT model significantly and suggest that practitioners add our generalizations to their modeling tools. Furthermore, we provide a censoring classifier that comes as a by-product of our method. For the wind park investigated we could show that using GWPPT instead of WPPT can increase monetary profit by around 1.25 million euros a year, simply due to smaller overall forecasting impreciseness.

References

Arabmazar, A. and Schmidt, P. (1981), Further Evidence on the Robustness of the Tobit Estimator to Heteroskedasticity, *Journal of Econometrics*, 17, pp. 253–258.

Beblo, G. and Schmid, W. (2011), Modeling High Frequency Wind Speed Data, European University Viadrina Frankfurt (Oder), Department of Business Administration and Economics, Discussion Paper No. 294.

Beyer, H., Degner, T., Hausmann, J., Hoffmann, M., and Rakján, P. (1994), Short Term Prediction of Wind Speed and Power Output of a Wind Turbine with Neural Networks, *Proceedings of the EWEC 1994*, pp. 349–352.

Box, G. and Jenkins, G. (1970), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden Day.

Burton, T., Jenkins, N., Sharpe, D., and Bossanyi, E. (2011), *Wind Energy Handbook*, John Wiley & Sons, West Sussex, UK.

Chung, C.F. and Goldberger, A.S. (1984), Proportional Projections in Limited Dependent Variable Models, *Econometrica*, 52, pp. 531–534.

Costa, A., Crespo, A., Navarro, J., Lizcano, G., Madsen, H., and Feitosa, E. (2008), A Review on the Young History of the Wind Power Short Term Prediction, *Renewable and Sustainable Energy Reviews*, 12, pp. 1725–1744.

Ernst, B. and Rohrig, K. (2002), Online-Monitoring and Prediction of Wind Power in German Transmission System Operation Centres, *Proceedings of the first IEA Joint Action Symposium on Wind Forecast Techniques, Norrköping, Sweden*, pp. 1062–1065.

Giebel, G., Brownsword, R., Kariniotakis, G., Denhard, M., and Draxl, C. (2011), The State-Of-The-Art in Short-Term Prediction of Wind Power, Tech. rep., ANEMOS.plus, Risø DTU, Wind Energy Division.

Gill, J. (2008), *Bayesian Methods: A Social and Behavioral Sciences Approach*, Chapman and Hall/CRC, London.

González-Longatt, F., Wall, P., and Terzija, V. (2012), Wake Effect in Wind Farm Performance: Steady-state and Dynamic Behavior, *Renewable Energy*, 39(1), pp. 329–338.

Hemmingsson, M., Nielsen, A., Pedersen, K., and Rasmussen, J. (2006), Monitoring of Power System Events at Transmission and Distribution Level, *Transactions on Power Systems*, 21, pp. 1007–1008.

Holttinen, H. (2005), Optimal Electricity Market for Wind Power, *Energy Policy*, 33, pp. 2052–2063.

Hyndman, R. and Khandakar, Y. (2008), Automatic Time Series Forecasting: The Forecast Package for R, *Journal of Statistical Software*, 27(3), pp. 1–22.

Jursa, R. and Rohrig, K. (2008), Short-term Wind Power Forecasting Using Evolutionary Algorithms for the Automated Specification of Artificial Intelligence Models, *International Journal of Forecasting*, 24, pp. 694–709.

Landberg, L. (1999), Short-term Prediction of the Power Production from Wind Farms, *Journal of Wind Engineering and Industrial Aerodynamics*, 80, pp. 207–220.

Lange, M. and Focken, U. (2006), *Physical Approach to Short-Term Wind Power Prediction*, Springer Berlin Heidelberg.

Nielsen, H.A., Pinson, P., Christiansen, L.E., Nielsen, T.S., Madsen, H., Badger, J., Giebel, G., and Ravn, H.F. (2007), Improvement and Automation of Tools for Short Term Wind Power Forecasting, Tech. rep., Scientific Proceedings of the European Wind Energy Conference & Exhibition, Milan, Italy.

Nielsen, T.S., Madsen, H., and Nielsen, H.A. (2002), Prediction of Wind Power Using Time-Varying Coefficient-Functions, Tech. rep., 15th IFAC World Congress, Barcelona, Spain.

Nielsen, T.S., Madsen, H., Nielsen, H.A., and Tøfting, J. (1999), Using Meterological Forecasts in Online Predictions of Wind Power, Tech. rep., Eltra, Fredericia, Denmark.

Pinson, P. (2012), Very Short-Term Probabilistic Forecasting of Wind Power With Generalized Logit-Normal Distributions, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(4), pp. 555–576.

Rosett, R. and Nelson, F. (1975), Estimation of the Two-Limit Probit Regression Model, *Econometrica*, 43, pp. 141–146.

Sánchez, I. (2008), Adaptive Combination of Forecasts with Application to Wind Energy, *International Journal of Forecasting*, 24, pp. 679–693.

Tastu, J., Pinson, P., and Madsen, H. (2010), Multivariate Conditional Parametric Models for a Spatio-Temporal Analysis of Short-Term Wind Power Forecast Errors, *Proceedings of the European Wind Energy Conference, EWEC.*

Xiaohong, G., Xi-Ren, C., and Papalexopoulos, A. (2000), Forecasting Power Market Clearing Price and Quantity Using a Neural Network Method, *Power Engineering Society Summer Meeting*, 4, pp. 2183–2188.