Efficiency of Wind Power Production and its Determinants

Simone Pieralli, Matthias Ritter, Martin Odening

Paper prepared for presentation at the 2015 Agricultural & Applied Economics Association and Western Agricultural Economics Association Annual Meeting, San Francisco, CA, July 26-28

Efficiency of Wind Power Production and its Determinants*

Simone Pieralli[a**], Matthias Ritter[a], Martin Odening[a]

Abstract

This article examines the efficiency of wind energy production. Using non-convex efficiency analysis, we quantify production losses for 19 wind turbines in four wind parks across Germany. In a second stage regression, we adapt the linear regression results of Kneip, Simar, and Wilson (2014) to explain electricity losses by means of a bias-corrected truncated regression analysis. The results show that electricity losses amount to 27% of the maximal producible electricity. Most of these losses are from changing wind conditions, while 6% are from turbine errors.

Key words: wind energy, efficiency, free disposal hull, bias correction.

JEL codes: D20, D21, Q42.

1 Introduction

Renewable energy production has experienced rapid growth over the last two decades and this growth is likely to continue. Wind energy production contributed a significant share to this expansion and has attracted institutional investors. The profitability of wind energy production is determined by generation costs, energy prices, and turbine productivity. In the past, investments in wind parks were able to attain comparably high returns on investments. In many countries, such as Germany and Spain, producers receive guaranteed prices for wind energy that are above market prices. Generation costs are also fairly stable since operating costs are relatively low and installation costs are rather transparent. Thus, productivity is the crucial driver for the profitability of wind energy production. Productivity, in turn, heavily depends on wind

*Financial support from the German Research Foundation via CRC 649 "Economic Risk" is gratefully acknowledged.
[a] Humboldt-Universität zu Berlin, Department of Agricultural Economics, Philippstr. 13, 10115 Berlin, Germany.
** Corresponding author: simone.pieralli@agrar.hu-berlin.de, Tel.: +49 (0) 30209346847, Fax: +49 (0) 30209346841.

conditions, i.e., wind speed and its variability, at the production site. In fact, a careful assessment of wind conditions precedes any investment in wind parks. Given the importance of wind production, it is not surprising that a lot of effort has been devoted to developing models to predict how much of the installed capacity will actually be used during the investment period (e.g., Kusiak et al., 2013).

A second determinant of productivity, however, has received very little attention in the literature, namely, the efficiency of wind energy production which is the distance between the actual energy and maximal energy output under a certain level of production factors. In the context of wind energy, the maximal producible power as a function of wind speed is depicted by a power curve. Power curves are usually calculated by turbine producers for a specific turbine type under ideal conditions.[1] In reality, wind production does not take place under ideal conditions and therefore actual energy production regularly deviates from the power curve. For example, shortfalls can be caused by rainfall, icing, suboptimal adjustment of the pitch angle and nacelle position under changing wind conditions, technical failures, and scheduled maintenance. Under marginal wind conditions or a scenario of declining subsidies, these production inefficiencies can diminish the profitability of wind power plants.

A few empirical papers analyze the productivity and efficiency of wind power generation. Homola et al. (2009) analyze wind park data in Norway and suggest a correction for power curve estimation. Ilinca (2011) estimates that power losses due to icing conditions amount to as much as 50% of total annual production. Hughes (2012) and Staffell and Green (2014) indicate declining turbine performance due to increasing age of turbines in Denmark and the UK. Some other papers apply nonparametric methods to estimate the wind energy production frontier. Kusiak et al. (2012) use data envelopment analysis (DEA) to assess the performance of wind turbines in the presence of faults. They identify turbine downtime as the major reason for power curtailment. Iribarren et al. (2014) analyze the entire process of wind energy production and include further production factors, such as land and investment cost in their DEA model. To the best of our knowledge, Carvalho et al. (2009) is the only study that applies DEA to estimate a power curve based on high frequency production data.

The objective of this paper is twofold. First, we estimate the wind energy production frontier based on production data and quantify production losses that occur relative to this benchmark. In contrast to most other wind efficiency studies, we base frontier estimations directly on high frequency production data and do not aggregate the data for a single turbine or wind park. This sheds light on the emergence of production losses over time and avoids information losses through data smoothing. Carvalho et al. (2009) pursue a similar approach; however, they use a DEA model and thus estimate efficiency by assuming a convex production technology. This approach ignores the non-convex shape of a typical wind power curve and thus overestimates

[1] The industry standard for power curve estimation is IEC 61400-12-1 (Wächter et al., 2009; Homola et al., 2009).

inefficiency over some range of the production frontier. To avoid this flaw of DEA, we resort to a free disposal hull (FDH) estimation of the frontier, which does not assume convexity.

The second objective of this paper is to explain the magnitude of the observed production losses and to trace them back to factors which may or may not be under the control of wind park operators. To this end, we apply a truncated regression model that accounts for biases in the regression of estimated efficiency scores in the first step of our analysis (Kneip et al., 2014). From an applied perspective, our findings help improve the assessment of wind energy production under real world conditions.

In the following section, we explain in greater detail how we estimate the wind energy production frontier and derive the corresponding production losses. Moreover, we present the bias correcting regression model. These methods are then applied to high frequency production data from four wind parks in Germany. Section 3 describes the data base and Section 4 presents results. The final section summarizes and draws conclusion for improving the productivity of wind energy generation.

2 Methodology

The amount of wind's kinetic energy (E_k) available to be converted into electricity can be described by the following function (Hennessey, 1977; Gunturu and Schlosser, 2012):

$$E_k = 0.5\pi r^2 d w^3, \tag{1}$$

where r is the rotor size, so that the rotor swept area is πr^2, d is the air density, and w denotes wind speed. Air density is directly proportional to air pressure and inversely proportional to air temperature. Kinetic wind energy increases with wind speed and air density. It is important to note that according to Eq. (1), kinetic wind energy is a cubic function of wind speed. This characteristic results in a non-convex technology for wind speeds lower than the rated wind speed and has implications for the estimation of the production frontier. Air density is directly proportional to air pressure and inversely proportional to air temperature. Density is higher in the winter when the temperature is colder and is lower in the summer when the temperature is warmer. Air pressure causes variability in this general trend: High air pressure increases air density and low air pressure decreases density. However, only a portion of the wind's kinetic energy can be transformed into electricity. The efficiency of this transformation process depends on various technical and managerial factors and is the subject of this study.

In general terms, the production process is characterized by a production technology, which is defined as the set of all inputs (in our case: wind speed and air density) that are feasible to produce electric power:

$$T = \{w, d, e\colon\ (w, d) \text{ can produce } e\}, \tag{2}$$

where w is wind speed, d is air density, and e is wind electricity.

As mentioned above, wind speed is monotonically related to the amount of electrical power produced, but the rate of transformation is non-constant and increasing up to the rated wind speed. However, to preserve the machine equipment from destructive centrifugal forces, the rotational speed and thus power production are limited for wind speeds greater than the rated wind speed. These features of the production technology process can be captured by a non-convex free disposal hull (FDH) for a sample of n observation points $\{w_i, d_i, e_i\}_{i=1}^{n}$:

$$\hat{T}_{\mathrm{FDH}} = \{w, d, e : w \geq w_i, d \geq d_i, e \leq e_i, \forall i = 1, \ldots, n\}. \tag{3}$$

The FDH technology set creates an outer envelope of the data points included in technology T without assuming convexity. As a measure of the efficiency of the turbines in exploiting wind and air density conditions, we measure the nonparametric distance between each point and the frontier envelope. Since inputs cannot be controlled by producers and instead are determined by nature, it is reasonable to measure distance in the direction of the outputs. We define this efficiency measure for unit $(w_0,\ d_0, e_0)$ as follows:

$$\hat{\lambda}_{\mathrm{FDH}}(w_0, d_0, e_0) = \sup\{\lambda : (w_0, d_0, \lambda e_0)\ \in \hat{T}_{\mathrm{FDH}}\}. \tag{4}$$

An estimate $\hat{\lambda}_{\mathrm{FDH}}$ can be computed by means of a sorting algorithm that identifies all units that dominate unit $(w_0,\ d_0,\ e_0)$, i.e., all units that use less or equal inputs to produce equal or more output than unit $(w_0,\ d_0,\ e_0)$.[2] The dominating unit with the highest output is the potential producible electric power, $\hat{e}_0$. The efficiency measure $\hat{\lambda}_{\mathrm{FDH}}$ is then a ratio of the potential producible electric power $\hat{e}_0$ and the actual electric power produced e_0. This measure of inefficiency is a conservative measure compared to convex hull measures of inefficiency, such as DEA. Given this efficiency measure, we define the production loss of electric power EL for every observation as the difference between the production potential and the electric production observed:

$$\widehat{\mathrm{EL}}_0 = \hat{\lambda}_{\mathrm{FDH}} e_0 - e_0 = \ \hat{e}_0 - e_0. \tag{5}$$

Previous efficiency analyses on wind energy production consider single turbines or wind parks as “decision making units” and calculate efficiency scores for these units (e.g., Iribarren et al., 2014; Iglesias et al., 2010). This kind of analysis requires aggregating inputs and outputs to annual values. Production factors may include capital and labor. Here, we pursue a different approach. Efficiency scores are assigned to production intervals consisting of ten minutes. Each observation in our sample relates the electric power produced to the average wind speed and average air density in a ten-minute interval. Thus, we derive an efficient production function that

[2] We are thankful to Professor Dr. Simar who provided the sorting algorithm to calculate the FDH efficiency scores. However, we adapted the algorithm received to calculate the frontier in case of missing or infinite values. Efficiency scores λ can be alternatively determined by a mixed integer program (Deprins et al., 1984).

characterizes the technology of the wind turbine under differing wind and air density conditions. This production function resembles a power curve. Production factors other than wind and air density are not considered. Pooling high frequency production data from different wind parks in an efficiency analysis makes sense if electricity is produced with the same technology, i.e., turbine type and rotor size, which is the case in our study. Productivity under this perspective can be understood as electricity output under given weather conditions. This definition differs from productivity of a wind turbine or wind farm per unit of time, such as annual production. In our analysis, we may find that a wind turbine is efficient because it converts wind energy optimally into electricity, but that the produced power is low due to low inputs, i.e., unfavorable wind conditions. Thus, our efficiency analysis cannot support decisions on the location of wind parks or the choice among production technologies. Instead, it provides useful information on the magnitude of production losses due to suboptimal utilization of wind energy. Such production losses may be caused by unfavorable weather conditions (apart from wind speed and air density), such as icing or turbine faults. Though most of these factors are out of immediate managerial control, it is helpful to understand their contribution to electricity losses.

The second step of our analysis focuses on explaining observed production inefficiency by regressing the estimated electricity loss $\widehat{\mathrm{EL}}$ on a set of explanatory variables $\boldsymbol{v}$. While this two-step procedure is standard for nonparametric efficiency analyses, it ignores a methodical problem (Kneip et al., 2014). The fact that electricity losses $\widehat{\mathrm{EL}}$ result from a nonparametric estimation of the technology frontier entails problems to use them as a dependent variable in a second stage regression. In fact, the estimated efficiencies are biased measures of the true electricity losses because in a full population sample, there could be observations lying above the sample frontier. Thus, the estimated inefficiency measures derived from the sample frontier represent a lower bound of the true measures. This bias will lead to a biased estimation of the regression coefficients $\boldsymbol{\theta}$. To account for this, we adapt the procedure proposed by Kneip et al. (2014). A further aspect that has to be accounted for in the regression is the limited range of the calculated electricity losses: They are non-negative and cannot exceed the maximum capacity of the turbine (2.365 MW in our sample). Thus, we apply a truncated regression model. Assuming that the latent variable EL is normally distributed,

$$\mathrm{EL}_i|\boldsymbol{v}_i \sim N(\boldsymbol{v}_i'\boldsymbol{\theta}, \sigma^2), \tag{6}$$

the estimates of the parameters $\widehat{\boldsymbol{\theta}}$ can be obtained by maximizing the following likelihood function:

$$\mathcal{L}_1 = \prod_{\mathrm{i}=1}^{\mathrm{n}} \left(\frac{\frac{1}{\sigma}\varphi\left(\frac{\widehat{\mathrm{EL}}_i - \boldsymbol{v}_i\boldsymbol{\theta}}{\sigma}\right)}{\Phi\left(\frac{\mathrm{UB}-\boldsymbol{v}_i\boldsymbol{\theta}}{\sigma}\right) - \Phi\left(\frac{\mathrm{LB}-\boldsymbol{v}_i\boldsymbol{\theta}}{\sigma}\right)} \right) \tag{7}$$

where $\varphi(\cdot)$ is the standard normal density function and $\Phi(\cdot)$ is the cumulative standard normal density function. The arguments of the normal density function and cumulative density function are derived from the conditional truncation points of the regression model. UB and LB in the argument of the cumulative distribution function are the upper bound (2,365) and lower bound (0) of the dependent variable, respectively, and σ is the variance of the error term.

To correct for the estimation bias, we divide the sample into two parts and recalculate the efficiency losses $\widehat{\mathbf{EL}}_1$ and $\widehat{\mathbf{EL}}_2$, as in Kneip et al. (2014).[3] We stack these variables together to create a column vector of n elements $\widehat{\mathbf{EL}}^s = \begin{pmatrix} \widehat{\mathbf{EL}}_1 \\ \widehat{\mathbf{EL}}_2 \end{pmatrix}$. Accordingly, we stack the respective m explanatory variables to obtain an $n \times m$ matrix $\boldsymbol{v}^s = \begin{pmatrix} \boldsymbol{v_1} \\ \boldsymbol{v_2} \end{pmatrix}$, in which the order of the observations is rearranged to match the dependent variable. Similarly to Eq. (7), we calculate a new estimator $\widehat{\widehat{\boldsymbol{\theta}}}^s$ in a regression of $\widehat{\mathbf{EL}}^s$ on $\boldsymbol{v_s}$ by maximizing the following likelihood function:

$$\mathcal{L}_2 = \prod_{\mathrm{i}=1}^{\mathrm{n}} \left(\frac{\frac{1}{\sigma}\varphi\left(\frac{\widehat{\mathrm{EL}}_i^S - v_i^S \theta^S}{\sigma}\right)}{\Phi\left(\frac{\mathrm{UB} - v_i^S \theta^S}{\sigma}\right) - \Phi\left(\frac{\mathrm{LB} - v_i^S \theta^S}{\sigma}\right)} \right). \tag{8}$$

Under Theorem 5.2 in Kneip et al. (2014), the convergence result for an FDH convergence rate of $n^{-\xi}$ with $\xi \geq 1/3$ can be obtained by:

$$\sqrt{n}\left[\widehat{\boldsymbol{\theta}} - \left(2^{\xi} - 1\right)^{-1}\left(\widehat{\widehat{\boldsymbol{\theta}}}^s - \widehat{\boldsymbol{\theta}}\right) - \boldsymbol{\theta}\right] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2 \boldsymbol{Q}) \tag{9}$$

as $n \to \infty$. In Eq. (9), $\widehat{\boldsymbol{\theta}}$ is the biased estimate and $\left(2^{\xi} - 1\right)^{-1}\left(\widehat{\widehat{\boldsymbol{\theta}}}^s - \widehat{\boldsymbol{\theta}}\right)$ is the bias correction. This correction depends on ξ, which is the inverse sum of the number of inputs (p) and outputs (q), i.e., $\xi = 1/(p + q)$. That is, the higher the number of inputs and outputs, the higher the bias correction.

Asymptotic confidence intervals for the vector of parameters can be determined from Eq. (9) by:

$$\widehat{\boldsymbol{\theta}} - \left(2^{\xi} - 1\right)^{-1}(\widehat{\widehat{\boldsymbol{\theta}}}^s - \widehat{\boldsymbol{\theta}}) \pm z_{1-\frac{\alpha}{2}} \hat{\sigma}_n \hat{q}_{mm}/\sqrt{n} \tag{10}$$

where $\hat{q}_{mm} = (Z'Z)^{-1}/n$ are diagonal elements of the matrix $\boldsymbol{Q}$.

3 Data and model variables

The production data in this study are for four wind parks situated in different regions in western, central, and eastern Germany.[4] There is a total of 19 turbines in the sample, with each wind park containing up to 7 turbines, which are all of the same type and capacity, namely 2.365 MW. The

[3] To maintain a representative sample in this estimation, we assign observations in odd positions (1, 3, 5, etc.) to the first sub-sample and in even positions (2, 4, 6, etc.) to the second sub-sample.

[4] The authors thank 4initia GmbH for providing the data. The names and exact locations of the wind parks are concealed for confidentiality reasons.

average amount of power produced in kilowatts is reported for 10-minute intervals from July 1st, 2013 to June 30th, 2014. There are 989,175 observations for the 19 turbines in the sample. The dataset also includes observations of average mast wind for each 10-minute interval which constitutes the first (non-controllable) input in our efficiency analysis. It should be noted that the averaging of electricity output and wind speed over a ten-minute interval may lead to measurement errors in case of short-term fluctuations in wind speed. Since wind electricity is a non-linear function of wind speed, the mean value of wind electricity differs from the wind electricity rated at mean wind speed due to Jensen's inequality. This flaw of standard methods of power curve estimation is well-known and has led to modifications, such as dynamical power curve estimation (cf. Gottschall and Peinke, 2008; Homola et al., 2009). We do not adjust the measurement in our efficiency analysis. Instead, we include wind variability as an explanatory variable for inefficiency in the regression analysis.

The second input for wind electricity production is air density. We obtain this variable from a reanalysis dataset often used in wind power analysis, namely the Modern-Era Retrospective Analysis for Research and Applications (MERRA) data provided by NASA (Rienecker et al., 2011). MERRA reanalysis data reconstruct the atmospheric state by integrating data from different sources, such as conventional and satellite data. They offer a complete worldwide grid of weather data at a spatial resolution of 1/2° latitude and 2/3° longitude (about 45 km × 54 km in Germany). We interpolate the surface air density data of the four nearest grid points weighted by their distance to each wind park.[5] The data are available daily at one-hour intervals, beginning at 12:30 a.m., which we linearly interpolate to obtain observations for each 10-minute interval.

The summary statistics of input and output variables used for the estimation of the FDH and production losses are presented in Table 1. Summary statistics of the electricity produced in all single turbines is provided in Table A.1 in the appendix.

Table 1: Summary statistics of input and output variables

	Mean	Standard deviation	Minimum	Maximum
Inputs				
Wind speed (m/s)	5.90	3.04	0.00	28.20
Air density (kg/m³)	1.20	0.03	1.10	1.34
Output				
Electricity produced (kW)	507.46	622.76	0.00	2,365.00
Number of observations	989,175			

[5] The air density variable is called RHOA in the "MERRA IAU 2d surface turbulent flux diagnostics (AT1NXFLX)". More details on MERRA products can be found in Lucchesi (2012).

The upper part of Figure 1 depicts the average electricity production in the 19 turbines plotted against wind speed during the 10-minute intervals.[6] This figure provides a first impression of the range of productivity in observed wind electricity given a certain level of wind speed. Apparently, the distance between the lowest and highest output varies with the wind speed. The highest variation in productivity can be observed at moderate wind speeds between 5 and 12 m/s, whereas the productivity is less variable for calm wind conditions as well as for observations above the rated wind speed. This is plausible since the production potential under calm wind conditions is low for technical reasons. On the other hand, the frequency of observations at moderate wind speed is high (cf. Fig. 1(b)), so that heterogeneity in outcomes is expected to be high in the presence of other production factors or measurement errors. The wind distribution in Fig. 1(b) reveals that the behavior at moderate wind speeds is important for the overall efficiency of wind turbines.

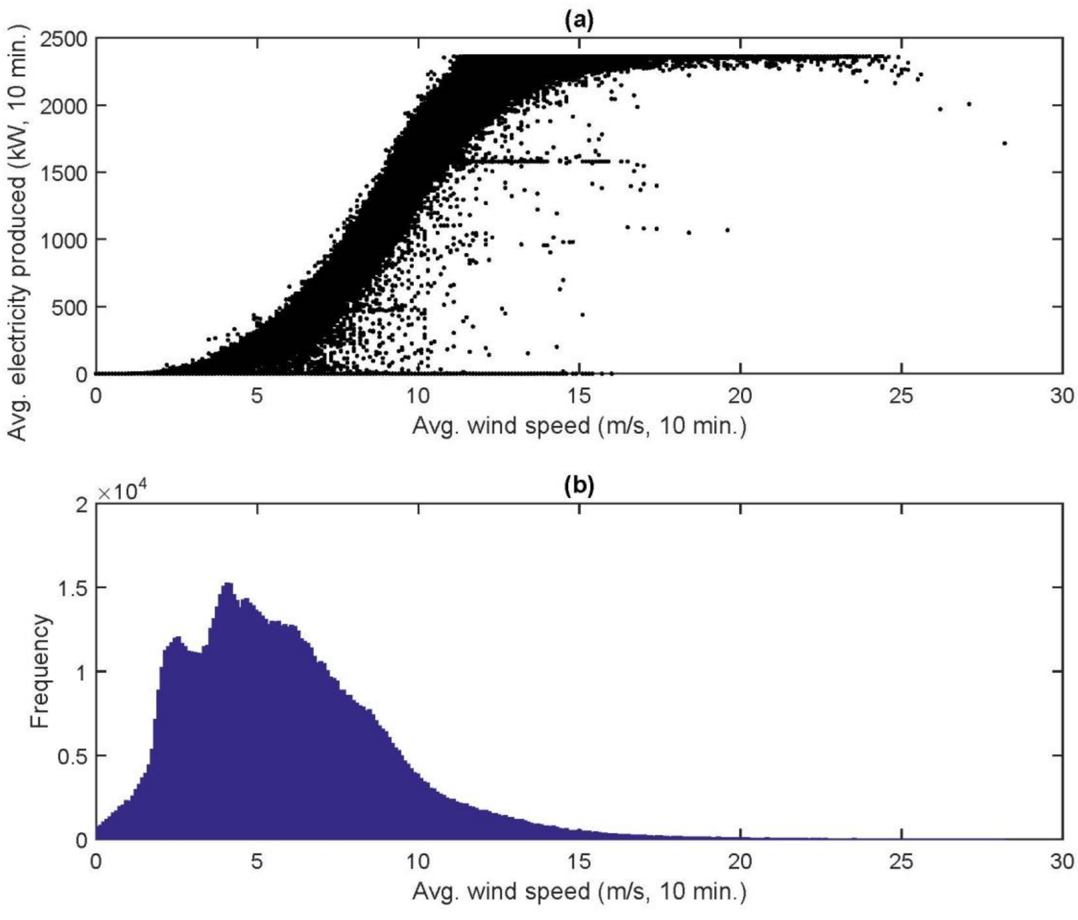

Fig. 1. (a) Power produced against wind speed; (b) Frequency of observed wind speed

[6] Note that production is measured in kW and not in kWh. This allows for a direct comparison with the installed capacity, but needs careful interpretation in the context of production losses. Production losses reported in kW for a 10-minute interval can be converted to kWh by dividing by six.

Our data set provides further information that can be used to derive explanatory variables in the second stage regression. We hypothesize that variability of wind conditions, i.e., wind speed and wind direction, affect measured productivity for two reasons. First, changing wind conditions require adjustments of the turbine's operation, e.g., the rotor pitch angle or the nacelle position. Because these adjustments are not frictionless and are realized with some delay, they decrease power generation compared to a situation of stable wind conditions. Here, we use the range of the wind speed in a 10-minute interval as an indicator of wind speed variability. Moreover, as mentioned above, changing wind speed will lead to errors in measurement of the average power production during 10-minute intervals. We further control for the speed of adjustment of the machine to different wind speeds by considering the difference in the average rotor speed between two consecutive 10-minute intervals. Changes in the wind direction are approximated by the absolute change (either to the left or right) in nacelle positions between two subsequent observations. The impact of this variable on productivity, however, is not clear a priori. On the one hand, higher direction variability is supposed to decrease the capability of a stable production of electric energy. On the other hand, the capability to adapt to changing wind directions can be regarded as an efficiency improving feature of the turbine. To ensure convergence of the maximum likelihood estimation, we use the cubic root of the absolute change.

Moreover, a detailed report for the turbine status is available at each point in time. This includes the occurrence of various error types with their starting times. In the data processing, we combine this information with the 10-minute interval data, i.e., we assign an error to all 10-minute intervals between the beginning of an error and the restart of the turbine. If there are concurrent errors, we consider only the error that occurred first. In the regression model, we consider three error categories that result in the disruptions of electricity production: the presence of an ice alert on the turbine, the presence of turbine maintenance, and a residual error category that includes all other errors. To put this in perspective, errors occur in about 3% of the observations in our sample, 30,363 observations. About half of these errors, 15,077, are ice alerts, with the remaining errors resulting from machine errors (9,775 cases or 32%) and maintenance (5,511 cases or 18%). All dummy variables for errors are interacted with wind speed to weigh the occurrence of an error by the wind energy.

Finally, to account for location specific heterogeneity, we include a full set of 19 turbine dummies ($\boldsymbol{D}^{\text{turbine}}$) which capture, for example, exposition to particular wind conditions due to the position of the turbine within a wind park.

We specify the regression model of Eq. (6) for the production loss EL by:

$$\text{EL} = \boldsymbol{D}^{\text{turbine}}\boldsymbol{\alpha} + \boldsymbol{Z\beta} + w\boldsymbol{H\delta} + \epsilon \qquad (11)$$

where $\boldsymbol{Z} = [\text{Wind range}, \text{Nacelle position change}, \text{Rotor speed difference}]$

and $\boldsymbol{H} = [\text{Icing error}, \text{Maintenance error}, \text{Machine error}]$

Summary statistics of the dependent and explanatory variables used in the second stage are in Table 2.

Table 2: Summary statistics of second stage regression variables

	Mean	Standard deviation	Minimum	Maximum
Second stage variables				
Electricity loss (kW)	190.78	150.55	0.00	2,365.00
Wind range (m/s)	3.51	2.09	0.00	30.60
Wind speed (m/s)	5.90	3.04	0.00	28.20
Nacelle position change (°, $\sqrt[3]{\lvert x \rvert}$)	1.18	0.91	0.00	6.70
Rotor speed difference (rpm)	0.0004	1.02	-17.36	16.65
Number of observations	989,175			

4 Results

The resulting estimate of the FDH technology is depicted in Fig. 2. As required by the free disposability assumption on the technology, increasing amounts of both inputs are related to higher potential electricity production. The impact of the two factors on electricity production, however, is different: Low wind speed renders any air density amount unimportant for power production, whereas low air density diminishes electricity production only marginally. For this reason, we focus on wind speed as the most important production factor in the subsequent figures, i.e., we focus on the power curve.

Fig. 3 depicts two power curves. The broken line represents a power curve that we calculated following the industry standard IEC 61400-12-1 (Homola et al., 2009). It reflects the average produced power for wind speed bins of 0.1 m/s width, where wind speeds are adjusted to an air density of 1.225 kg/m³ and erroneous observations are excluded. The solid line is a cross-section of the FDH at the same air density of 1.225 kg/m³. It differs from the standard power curve in two ways. First, it represents an envelope of the production data instead of an average. Second, the estimation is based on all (non-filtered) observations. Apparently, both curves differ across a wide range of wind speeds. The difference between the curves amounts to 183 kW per observation, which corresponds to 8% of the rated capacity. Fig. 3 shows that the distance between the best performing observations and the average observations for a given wind speed is higher for wind speeds between 8 and 12 m/s, compared to lower wind speeds. For wind speeds larger than 15 m/s, both functions converge. Note that there is a deviation between the FDH and the standard power curve for extreme wind speeds. Storm control in modern wind turbines results in a decline in power for very high wind speeds. Due to the free disposability assumption, this particular feature of wind electricity production cannot be mimicked by the FDH. In these cases, our method overestimates electricity losses. However, in our dataset only eight observations have wind speeds greater than 25 m/s (cf. Fig. 1(a)).

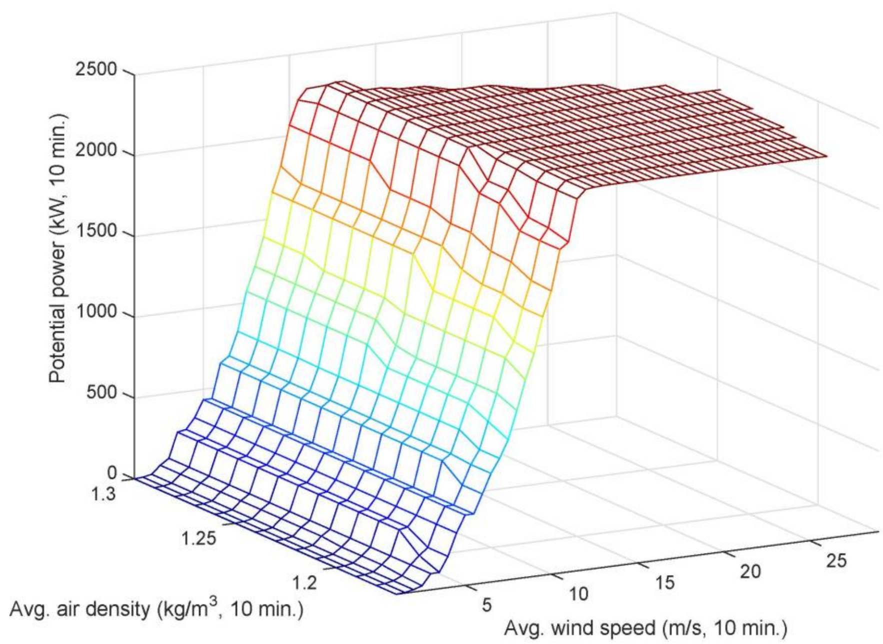

Fig. 2. Estimated free-disposal hull technology

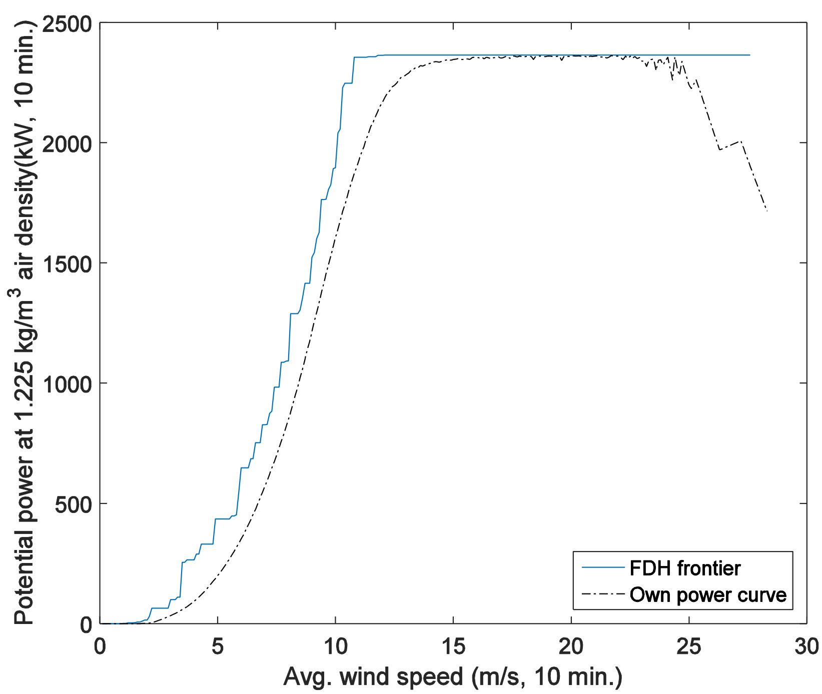

Fig. 3. Power curve vs. production frontier

The difference for every observation between the potential power production represented by the FDH estimated frontier in Fig. 3 and the actual power production defines an electricity loss (EL) that is plotted against wind speed and air density in Fig. 4. Fig. 4(a) shows that the bulk of electricity losses occurs when approaching the rated wind speed. One can further realize a bifurcation of electricity losses. While most losses decline for wind speeds greater than 10 m/s, there are a few observations that increase linearly with wind speed and are capped at the rated power capacity. The latter losses represent faults that occur when the turbine is partly or completely out of operation. In Fig. 4, observations with and without an error in the status code are represented by grey and black dots, respectively. Fig. 4(b) shows that there is variability in electricity losses for different air densities. Note that for air densities beyond 1.3 kg/m³ only observations occur that are affected by errors. These errors result from ice alerts caused by low temperatures, which correspond to high air density.

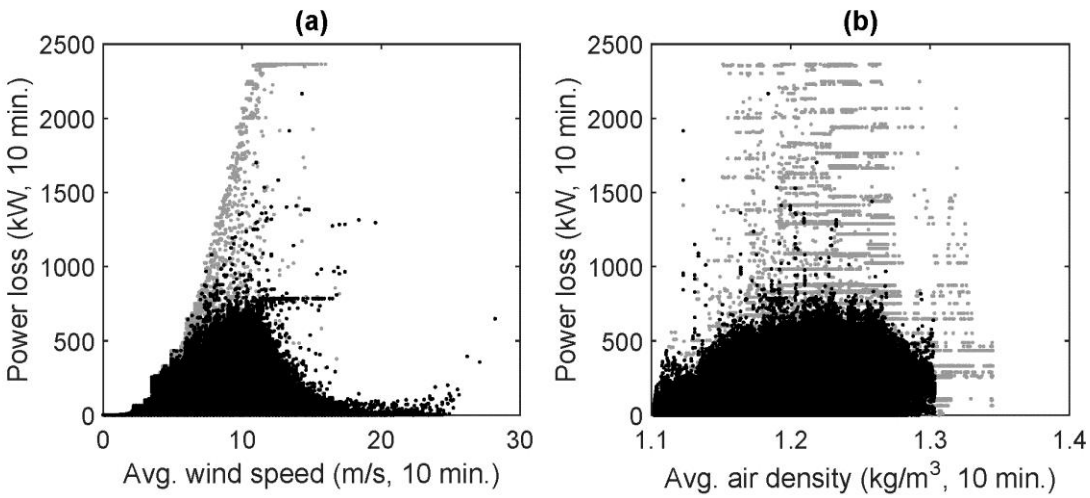

Fig. 4. (a) Power losses and wind speed; (b) Power losses and air density; observations with (gray) and without (black) turbine errors

However, as mentioned for Fig. 4(a), it is also the case for Fig. 4(b) that the frequency of production losses is not visible because of overlapping data points. To illustrate this more clearly, Fig. 5 depicts the sum of the losses calculated in correspondence of wind speeds at 0.1 m/s intervals. Thus, Fig. 5 accounts for the severity and frequency of losses occurring at different wind speeds. It can be seen that the highest losses occur for wind speeds between 4 and 9 m/s, whereas for lower and higher wind speeds, cumulative losses are much smaller. This reflects the distribution of wind speeds (Fig. 1(b)).

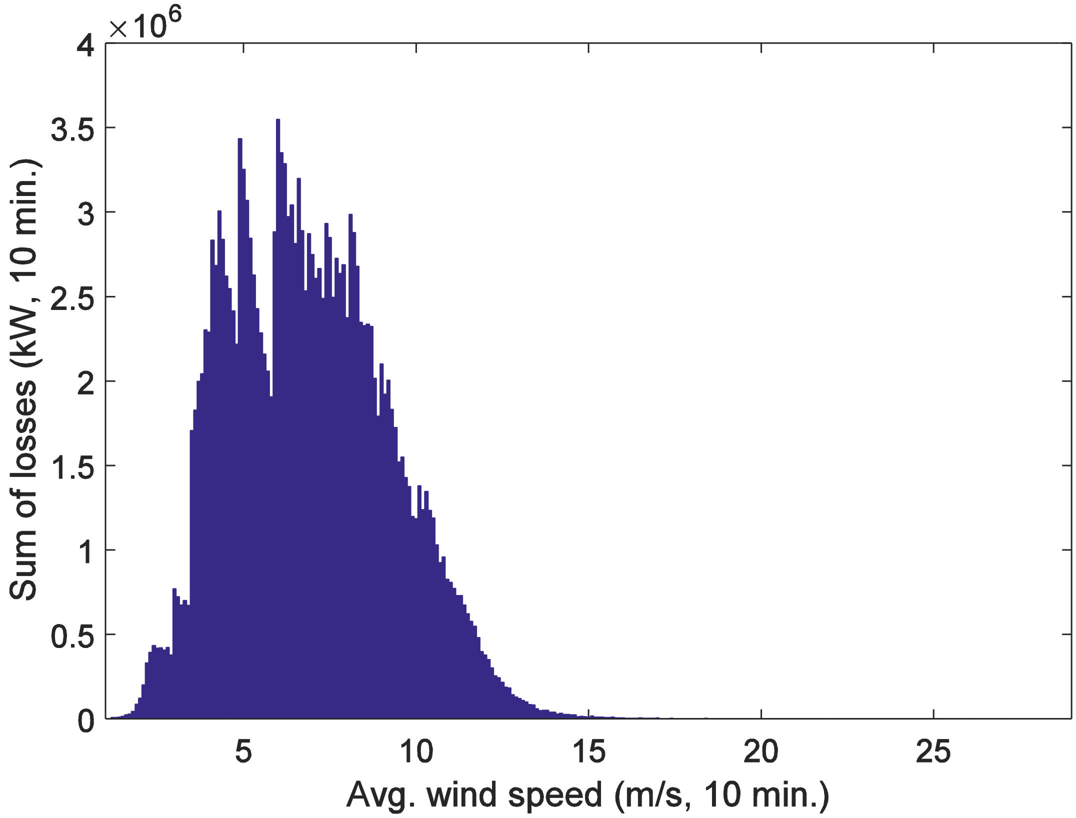

Fig. 5. (a) Cumulative power losses against wind speed

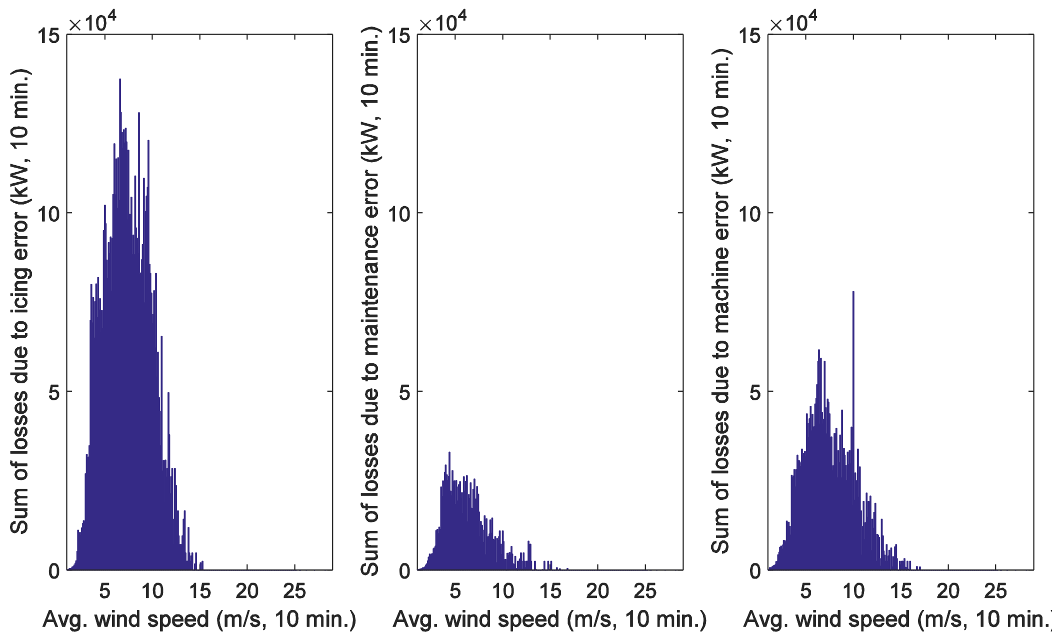

Fig. 5. (b) Cumulative power losses against wind speed for different turbine error groups

The quantitative analysis of electricity losses in Table 3 shows that an average loss in a 10-minute interval is 191 kW, which is associated with an average production of approximately 507 kW. This means that average losses amount to 27% of the amount of electricity that could be produced. If one aggregates this finding over a year, these losses amount to 9,930 MW per turbine, which translates to 1,655 MWh. Multiplied by an average spot price of 35.27 €/MWh attainable on the German electricity spot market, this value results in a yearly loss of 58,372 € per turbine.

Table 3: Total loss and number of observations for different error groups

	Observations	%	FDH Loss (kW)	%	FDH Loss/ observation
Icing error	15,077	1.52	7,670,313	4.06	508.74
Maintenance error	5,511	0.56	1,379,068	0.73	250.24
Machine error	9,775	0.99	3,141,902	1.66	321.42
All Errors	30,363	3.07	12,191,283	6.46	401.52
No Error	958,812	96.93	176,525,730	93.54	184.11
Total	989,175	100.00	188,717,013	100.00	190.78
Total FDH potential production			690,683,462		

Table 3 separates losses for the three different error groups. Not surprisingly, in case of an error, the average loss is much higher than without an error (401.52 kW compared to 184.11 kW). However, only 6.46% of the total losses are caused by errors because they occur for only 3.07% of the observations. Among the three error groups, ice has the largest share, with 4.06% of the total loss. The average loss in case of an icing error equals 508.74 kW, which is very similar to the overall mean of electricity produced (507.46 kW, Table 1). The average loss of 250.24 kW in case of a maintenance error is much lower and is not far from the average loss without an error (184.11 kW). This indicates that maintenance, if possible, should be conducted in periods where few losses are expected, i.e., during light wind conditions (Fig. 5(b)). The average loss of 321.42 kW in case of a machine error is higher than losses resulting from a maintenance error, but is lower than losses from icing.

The results of the second stage regressions are presented in Table 4. We ran three models of different complexity. Model 1 is a parsimonious model that includes the variability of wind speed and direction, error dummies, and turbine dummies as regressors. All coefficients are highly significant which is not surprising given the large number of observations. The coefficient of the wind speed range is negative. This can be explained by the concavity of some parts of the production frontier: If the wind speed is constantly at or below the cut-in speed, no electricity will be generated. On the other hand, a mean preserving spread of the wind speed will result in some positive output during the observed 10-minute interval and thus reduce observed losses. The coefficient of the wind range should be interpreted in conjunction with the effect of a change of the rotor speed, since the latter is a response to the former and both variables are not independent. We find that a higher change of the rotor speed over time is positively related to electricity losses, which reflects the physical energy required to overcome the inertia of the rotor blades. Surprisingly, adjustments of the nacelle position have an opposite effect, i.e., they reduce production losses. As expected, all error dummies interacted with wind speed are positively related to power losses and the magnitude of the coefficients confirms the earlier analysis, i.e., icing errors have the highest impact among the three considered error types. Turbine dummy variables have a positive sign, which indicates a positive average loss depending on turbine specific wind conditions and the position of the turbine in the park.

The fit of the predicted losses to the observed losses in Model 1, calculated via Pearson's ρ^2, is modest (ρ^2 = 0.249). Though the fit or predictive power is not the key issue in second stage regression analyses of technical efficiency, we modify the base regression model to attain a better fit to the data. There is a controversial discussion on whether input factors that are used in the nonparametric estimation of the production frontier should also enter the second stage regression (Simar and Wilson, 2007; Kneip et al., 2014). This amounts to the violation of a separability condition on the production technology. Disregarding these potential theoretical flaws, we include wind speed in the set of regressors in Model 2. This can be justified by the fact that potential power output varies with wind speed and thus the impact on wind electricity losses is different depending on the speed. In fact, the sign of the wind speed is positive while the signs of all other variables remain the same as in the base model. The inclusion of this variable increases ρ^2 considerably, from 0.249 to 0.41.

Inspection of Fig. 4(a) suggests that the impact of factors on realized production losses depends on whether the wind speed is above or below the rated wind speed. To distinguish these two regimes, in Model 3 we interact all variables with a dummy variable (DR) indicating a wind speed at or above the rated wind speed of 11.5 m/s.[7] Separating these two wind regimes increases ρ^2 further to 0.63. Again, the sign of the coefficients remain the same as before, apart from the rotor speed variable which is now negative. The coefficients associated with the dummy variable for the high wind speed regime capture the difference between the regression coefficients for wind speed above the rated speed relative to the coefficients below the rated wind speed. This difference is positive for wind variability. Varying wind speeds around a high mean value typically represents gusty and turbulent wind conditions, which may disturb power production, e.g., via storm control, while maximal capacity cannot be further increased. For the error variables, this means that their impact on power losses is even stronger than in the low wind regime. The effect of the wind speed becomes negative under the high wind regime. This is plausible since losses are lower for increasing wind speed in the high wind regime. Comparing the turbine dummies reveals considerable differences in productivity at different locations. The

[7] This wind speed is the empirically recovered rated wind speed in our sample, that is the first occurrence of an average production of 2,365 kW in a 10 minutes interval is at 11.5 m/s. A higher wind speed than rated wind speed occurs in approximately 5% of the observations.

best performing turbine in our sample (A1) is located in a wind park in a position that is free from obstacles in most recurring wind directions, such as other wind turbines or hills. None of the other turbines in this study have the same advantage.

Table 4: Second stage regression results

Dependent variable:	Truncated regression model					
Electricity loss (kW)	Model 1		Model 2		Model 3	
Wind range	-1.101	***	-11.792	***	-0.332	***
Nacelle position change($\sqrt[3]{\lvert x \rvert}$)	-12.361	***	-8.549	***	-6.817	***
Icing error*wind speed	76.346	***	90.358	***	82.119	***
Maintenance error*wind speed	25.436	***	45.661	***	38.673	***
Machine error*wind speed	58.217	***	72.398	***	58.849	***
Rotor speed difference	5.124	***	2.454	***	-0.589	***
Wind speed			16.568	***	34.923	***
Wind range *DR					12.506	***
Nacelle position change *DR					75.149	***
Icing error*Wind speed*DR					154.512	***
Maint. error*Wind speed*DR					64.504	***
Machine error*Wind speed*DR					75.706	***
Rot. speed diff. *DR					14.978	**
Wind speed*DR					-49.938	***
A1	434.145	***	287.684	***	124.344	***
B1	437.149	***	308.448	***	140.722	***
B2	431.775	***	305.745	***	139.863	***
B3	430.911	***	309.061	***	147.127	***
B4	434.334	***	299.450	***	129.991	***
B5	451.763	***	308.846	***	124.861	***
B6	438.381	***	301.511	***	131.402	***
B7	429.294	***	305.203	***	144.920	***
C1	424.015	***	296.618	***	139.792	***
C2	420.881	***	293.061	***	138.054	***
C3	423.832	***	301.221	***	145.960	***
C4	428.083	***	301.607	***	142.611	***
C5	417.072	***	286.462	***	136.712	***
C6	420.409	***	289.573	***	134.045	***
D1	422.769	***	289.764	***	133.259	***
D2	419.365	***	291.877	***	137.401	***
D3	431.860	***	290.234	***	124.822	***
D4	419.091	***	294.521	***	136.140	***
D5	411.033	***	283.066	***	138.753	***
Pearson's ρ^2	0.249		0.409		0.631	

** and *** denote statistical significance at the 5 and 1 percent level, respectively.

5 Conclusions

This article analyzes the productivity and efficiency of wind electricity generation under real world conditions. Based on a sample of 19 wind turbines located in four wind parks in Germany, we calculate an efficient production frontier that represents the maximum producible electricity at a certain level of wind speed and air density. This method results in a production frontier that is similar to a power curve. With this frontier, we can quantify electricity losses that have been realized under various wind conditions compared with the efficient frontier benchmark. By construction, the production frontier dominates the standard power curve since the latter represents average production under ideal conditions. We find that the average difference between the frontier and the standard power curve is 183 kW. Thus, the standard power curve can be regarded as a conservative estimate of electricity output, conditional on exogenous wind input.

Our results show that production inefficiencies amount to a loss of 27% of the potential amount of producible electricity. In a subsequent step, we decompose these production losses. It is noteworthy that turbine errors are responsible for only 6% of the production losses, though they often cause a complete stop in production. The reason is that only 3% of the observations in our sample were affected by errors. Among the losses caused by turbine errors, icing had the highest impact. We employ a regression model to explain the occurrence of production losses in greater detail. Other than turbine errors, we find that changing wind conditions, i.e., variations in wind speed and direction, affect the efficiency of electricity production. Moreover, turbine specific effects exist which can likely be traced to the position of a turbine within a wind park.

From a methodological perspective, to the best of our knowledge, this paper represents the first empirical estimation of efficiency with high frequency wind electricity production data using non-convex analysis methods. Moreover, it is the first estimation of bias-corrected explanations of efficiency in production by adapting the linear regression convergence results in Kneip et al. (2014) to a truncated regression.

It is not straightforward to draw immediate managerial conclusions from the empirical findings since most factors (air density, wind speed and variability, temperature) are not controllable. Weather conditions are entirely exogenous and stochastic. Nevertheless, it is important to understand how vulnerable wind electricity production is under real world conditions. This analysis emphasizes that weather conditions should be carefully inspected when making the decision of where to locate a wind turbine. Moreover, our analysis is helpful to assess the trade-offs between the benefits from larger distances between turbines within a wind park and higher costs for land acquisition to build larger wind parks. Finally, our results highlight the gains arising from flexible adjustments of wind turbines to changing weather conditions. Technical progress, such as anti-icing and de-icing systems, aims to increase this flexibility. This study shows that even if such systems are costly, they could prevent important losses in the medium to long-run. Since our analysis considers only one particular turbine type we cannot draw conclusions on the impact of different technologies on the efficiency of wind energy production. A comparison of different wind power technologies is recommended as a subject of further research.

6 References

Carvalho, A., González, M.C., Costa, P., Martins, A. (2009): Issues on Performance of Wind Systems derived from Exploitation Data. Industrial Electronics, 2009. IECON '09. 35th Annual Conference of IEEE: 3599-3604.

Deprins, D., Simar, L., Tulkens, H. (1984): Measuring Labor Efficiency in Post Offices. In M. Marchand, P. Pestieau and H. Tulkens (eds.), The Performance of Public Enterprises: Concepts and Measurements. Amsterdam: North Holland.

Gottschall J., Peinke J. (2008): How to improve the estimation of power curves for wind turbines. Environmental Research Letters 3(1): 015005 (7pp).

Gunturu U. B., Schlosser, C. A. (2012): Characterization of wind power resource in the United States. Atmospheric Chemistry and Physics, 12: 9687-9702.

Hennessey Jr., J. (1977): Some Aspects of Wind Power Statistics. Journal of Applied Meteorology 16: 119-128.

Homola, M. C., Byström, J., Nicklasson, P. J., Sundsbø, P. A. (2009): An improved method for wind power estimation. Working Paper, Narvik University College, Norway, 1-13.

Hughes, G. (2012): The Performance of Wind Farms in the United Kingdom and Denmark. Renewable Energy Foundation, London.

Iglesias, G., Castellanos, P., Seijas, A., (2010). Measurement of productive efficiency with frontier methods: A case study for wind farms. Energy Economics 32 (5): 1199–1208.

Iribarren, D., Vázquez-Rowe, I., Rugani, B., Benetto, E. (2014): On the feasibility of using emergy analysis as a source of benchmarking criteria through data envelopment analysis: A case study for wind energy. Energy 67: 527-537.

Ilinca, A., (2011). Analysis and Mitigation of Icing Effects on Wind Turbines, Wind Turbines, Dr. Ibrahim AlBahadly (Ed.), ISBN: 978-953-307-221-0, InTech, Available from: http://www.intechopen.com/books/windturbines/analysis-and-mitigation-of-icing-effects-on-wind-turbines.

Kneip, A., Simar, L., Wilson, P. W. (2014): When Bias Kills the Variance: Central Limit Theorems for DEA and FDH Efficiency Scores. Econometric Theory, FirstView http://dx.doi.org/10.1017/S0266466614000413.

Kusiak, A., Verma, A., Wei, X. (2012): Wind turbine capacity frontier from SCADA. Wind Systems Magazine. September 2012: 36-39.

Kusiak, A., Zhang, Z., Verma, A. (2013): Prediction, operations, and condition monitoring in wind energy. Energy 60: 1-12.

Lucchesi, R. (2012). File specification for MERRA products (version 2.3). GMAO Office Note No. 1. http://gmao.gsfc.nasa.gov/pubs/docs/Lucchesi528.pdf.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., Woollen, J. (2011). MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. Journal of Climate 24(14): 3624-3648.

Simar, L., Wilson. P. W. (2007): Estimation and inference in two-sage semi-parametric models of production processes. Journal of Econometrics 136: 31-64.

Staffell, I., Green, R. (2014): How does wind farm performance decline with age?, Renewable Energy, 66: 775-786.

Wächter, M., Gottschall, J., Rettenmeier, A., Peinke, J. (2009): Power curve estimation using LIDAR measurements. In: Proceedings of the EWEA European Wind Energy Conference, Marseille, France, 16–19 March 2009.Vol.1: 4473-4477.

7 Appendix

Table A.1: Summary statistics of electric power production of individual turbines

	Mean	Standard deviation	Minimum	Maximum
Produced power (kW)	507.46	622.76	0.00	2,365.00
By turbine:				
A1	659.69	699.27	0.00	2,365.00
B1	406.14	496.44	0.00	2,365.00
B2	465.19	554.64	0.00	2,365.00
B3	438.52	541.08	0.00	2,365.00
B4	445.77	534.40	0.00	2,365.00
B5	455.51	540.41	0.00	2,365.00
B6	451.54	542.09	0.00	2,365.00
B7	460.88	553.64	0.00	2,365.00
C1	540.56	645.49	0.00	2,365.00
C2	542.04	659.38	0.00	2,365.00
C3	516.36	641.55	0.00	2,365.00
C4	532.76	632.25	0.00	2,365.00
C5	604.09	709.68	0.00	2,365.00
C6	553.62	666.26	0.00	2,365.00
D1	534.79	647.18	0.00	2,365.00
D2	504.27	643.73	0.00	2,364.00
D3	462.71	621.11	0.00	2,364.00
D4	485.51	644.01	0.00	2,364.00
D5	582.42	732.41	0.00	2,365.00
Number of observations	989,175			