SFB 649 Discussion Paper 2014-052

Designing an Index for Assessing Wind Energy Potential

Matthias Ritter*
Zhiwei Shen*
Brenda López Cabrera*
Martin Odening*
Lars Deckert**

* Humboldt-Universität zu Berlin, Germany
** 4initia GmbH, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

Designing an Index for Assessing Wind Energy Potential*

Matthias Ritter[a**] Zhiwei Shen[a] Brenda López Cabrera[b]
Martin Odening[a] Lars Deckert[c]

September 23, 2014

To meet the increasing global demand for renewable energy such as wind energy, more and more new wind parks are installed worldwide. Finding a suitable location, however, requires a detailed and often costly analysis of the local wind conditions. Plain average wind speed maps cannot provide a precise forecast of wind power because of the non-linear relationship between wind speed and production. In this paper, we suggest a new approach of assessing the local wind energy potential: Meteorological reanalysis data are applied to obtain long-term low-scale wind speed data at turbine location and hub height; then, with actual high-frequency production data, the relation between wind data and energy production is determined via a five parameter logistic function. The resulting wind energy index allows for a turbine-specific estimation of the expected wind power at an unobserved location. A map of wind power potential for whole Germany exemplifies the approach.

Keywords: Wind power, energy production, renewable energy, onshore wind, MERRA
JEL classification: Q42, Q47

1 Introduction

Because of increasing energy demand worldwide and the willingness to reduce greenhouse gas emissions, renewable energies such as wind energy are rapidly growing: The global cumulative installed capacity of wind energy increased from 6 GW in 1996 to 318 GW in 2013 and is expected to reach 596 GW in 2018 (GWEC, 2014).

*The financial support from the German Research Foundation via the CRC 649 'Economic Risk', Humboldt-Universität zu Berlin, is gratefully acknowledged.
[a]Humboldt-Universität zu Berlin, Department of Agricultural Economics, Philippstr. 13, 10115 Berlin, Germany
**Corresponding author: Matthias.Ritter@agrar.hu-berlin.de
[b]Humboldt-Universität zu Berlin, School of Business & Economics, Ladislaus von Bortkiewicz Chair of Statistics, Spandauer Straße 1, 10178 Berlin, Germany
[c]4initia GmbH, Reinhardtstraße 46, 10117 Berlin, Germany

Planning a new wind farm starts with the search for a suitable location. Besides the question of constructible surface and legal aspects, the geographical wind conditions and timing play an important role. Timing sustainably influences the financial success of a wind farm project because revenues generated from renewable energies are generally based on a regulated country-specific feed-in tariff system. The compensation paid to the operators decreases on an annual basis depending on the date of commissioning to reduce governmental subsidies. Therefore, delaying a project increases the costs and the uncertainty of the expected outcome.

Finding a suitable position by measuring the wind speed at different locations and heights is very time-consuming and costly. Hence, the expected energy production at possible locations has to be derived in a different manner.

Many studies deal with deriving detailed long-term wind speed maps for individual countries (e.g., USA (Archer and Jacobson, 2003), Spain (Gastón et al., 2008), Germany (Deutscher Wetterdienst, 2009), or Greece (Kotroni et al., 2014)), continents (e.g., Europe (Troen and Petersen, 1989)) or even the world (e.g. Archer and Jacobson, 2005). These maps are a rough indicator for the average local wind conditions, but for deriving the expected wind energy production, long-term average wind speeds are inadequate. The reason is the non-linear relationship between wind speed and production: It is possible that a stable wind speed of around 3 m/s over the year, which is smaller than the typical cut-in speed where the turbines start, leads to zero production. A wind speed with high fluctuations around the mean of 3 m/s, however, leads to a much higher production.

To overcome this problem, a long record of high-frequency wind speed at the turbine location and hub height is needed. Then, the wind power production can be estimated by transforming the high-frequency wind speed to the wind power production via a wind power curve (e.g. Brown et al., 1984; Sanchez, 2006). However, from the perspective of installing a turbine at a new location, the requirement of long-term high-frequency wind data can hardly be fulfilled. The wind power curve given by the turbine producer requires instantaneous mast wind speed to derive the production, which in most cases, are not recorded. Hence, the wind production cannot be estimated via the wind power curve, and the linkage between wind speed at a higher scale (e.g., hourly averages) and the true production deserves further investigation.

In this paper, we propose a new way to estimate the long-term wind energy potential of a new location by applying an index, which mainly consists of two steps: First, we derive lower scale wind speed data at the turbine location at hub height by processing meteorological reanalysis data. These data are available all over the world at low spatial and temporal scales, so that our approach is globally feasible. Second, we estimate an analytic production function based on real production data, which converts the meteorological reanalysis data into production data. Based on local wind speed data derived for an unobserved location, this production function gives an estimate of the low-scale energy production. By aggregating the estimated production to a larger time scale and long-term historical data, the proposed wind energy index is able to assess the long-term wind energy potential for any location.

The paper is organized as follows. In the next section, we describe in detail how the wind speed at the turbine location is derived, how the production function is estimated, and how the wind energy index is constructed. In Section 3, we apply our approach to data for Germany and evaluate it. This section concludes with an energy production

map of Germany, which estimates the long-term wind energy potential of each location. Section 4, finally, provides further discussion and conclusions.

2 Methods

2.1 Framework

To measure the potential of wind power production at a specific location, we develop a quantitative and objective wind energy index that represents the actual wind energy production of a certain turbine type. To obtain such an index, some steps have to be conducted.

First, the type of database has to be chosen, which is used to calculate the wind energy index. One possibility is energy production data from wind farms in the neighbourhood with similar wind conditions and turbine characteristics. Alternatively, wind speed data can be applied directly. When selecting wind speed data as the database, they have to be transferred to the wind turbine position as they are usually not available for every location. This means that the data have to be horizontally interpolated to the turbine location and vertically extrapolated to the turbine height. The most crucial decision is then how to transform local wind speed data to a wind energy index that reflects the actual wind energy production.

The aforementioned steps are described in greater detail in the following sections.

2.2 Database

In principle, the analysis can be built on production data or wind speed data. Production data of nearby wind farms have the advantage that they reflect the true fluctuations and no transformation, which might cause an estimation bias, is needed. Nevertheless, equal geographical and technical conditions have to be assumed.[1]

Another way of analyzing the energy potential is deriving a wind energy index based on wind speed data, which are better available than production data. The most common dataset used in the analysis of wind resource is weather station data because it objectively measures the actual wind speed at certain locations. Using weather station data for this aim, however, comes under criticism: the availability of such data is often limited; the historical data records might not be complete; weather stations are not located at realistic locations of wind farms; the time series record of weather station data is frequently no more than 25 years (Kubik et al., 2013a). In Germany, free wind speed data are available since 1996 for 64 weather stations with three measurements per day (6 am, 12 am, 18 pm) from DWD[2]. The data, however, are measured in Beaufort unit, which is a very rough scale.

An alternative dataset that has been recommended in the wind power analysis is reanalysis data, such as the Modern-Era Retrospective Analysis for Research and Applications

[1] In Germany, the BDB (Betreiber-Datenbasis) index is used to measure monthly fluctuations of wind energy production in 25 regions. It is, however, often criticized because of its in-transparency and unreliability as the wind conditions are not homogeneous in the 25 regions. Moreover, it remains unclear how this index can be used to estimate the potential of an unobserved location. For more details, we refer to Betreiber-Datenbasis (2011) and Bundesverband WindEnergie (2013).

[2] DWD: German Meteorological Service, `www.dwd.de`

(MERRA) data provided by NASA (Carta et al., 2013; Kubik et al., 2013a; Staffell and Green, 2014). MERRA reanalysis data reconstruct the atmospheric state by integrating data from different sources such as conventional and satellite data (Rienecker et al., 2011; Gunturu and Schlosser, 2012). They offer a complete worldwide grid of wind data at a spatial resolution of 1/2° latitude and 2/3° longitude (around 45 km × 54 km in Germany) and an hourly temporal resolution since 1979. The wind data consist of a northward and an eastward wind component at three different heights (2m, 10m, 50m above ground), which are helpful to derive the wind speed and wind direction at turbine height. Thus, reanalysis data could mitigate the problems that plague available weather station data.

Alternative reanalysis data sources such as NCEP/NCAR (National Center for Environmental Prediction/National Center for Atmospheric Research), ERA-Interim (European Centre for Medium-Range Weather Forecasts Re-Analysis, or CFSR (Climate Forecast System Reanalysis) are also possible, but so far, there is no consensus on the superiority of one particular reanalysis model (Liléo and Petrik, 2011; Jimenez et al., 2012; Carvalho et al., 2014). Further comparisons among these candidates are needed to determine the correct wind power potential. In the following, we apply MERRA data to obtain wind speed data at an unobserved location.

2.3 Horizontal interpolation

Every location lies within a rectangular spanned by the four nearest MERRA grid points. The wind speeds at these four points, i.e., the eastward and northward components u_h and v_h in 2m, 10m, and 50m, are interpolated to the turbine's location weighted by their horizontal distance (inverse distance weighting). This approach assumes that the influence decreases with increasing distance. Given the rather short distances (maximum distance to the nearest grid point is around 35 km) and the regular pattern of the MERRA grid, inverse distance weighting is a reasonable candidate. Nevertheless, alternative interpolation methods such as Kriging, polynomial, or spline interpolation are possible (Luo et al., 2008).

After interpolating, the two components for each height are combined to obtain absolute values of the wind speed at the turbine's location at the three heights using the Pythagorean theorem:

$$V_h = \sqrt{u_h^2 + v_h^2},\ h = 2, 10, 50. \tag{1}$$

At this point, it is still possible to calculate the wind direction at height h, φ_{V_h}, at the turbine's location by the following equation:

$$\varphi_{V_h} = \tan^{-1}\left(\frac{v_h}{u_h}\right). \tag{2}$$

Because most wind turbines can rotate towards the wind direction (Caporin and Preś, 2012), we neglect the wind direction in the following analyses and focus only on the wind speed.

2.4 Vertical extrapolation

The hub height of a typical wind turbine is much higher than 2m, 10m, and 50m, where wind speeds are provided by MERRA. Hence, the available wind speeds at height h need to be extrapolated to the turbine height z.

One extrapolating method is given by the power law (e.g. Brown et al., 1984; Jung et al., 2013; Kubik et al., 2013b):

$$V_z = V_h \left(\frac{z}{h}\right)^{\alpha}, \tag{3}$$

where V_z and V_h denote the wind speeds at heights z and h, respectively. The wind shear coefficient α depends on the stability of the atmosphere and can be derived empirically, but the results are very sensitive to a correct modelling and the right assumptions (Fırtın et al., 2011). The power law gives a "reasonable first approximation" (Brown et al., 1984, p. 1190). However, the procedure commonly used in the literature and applied in this study is the log wind profile (e.g., Stull, 1988; Gunturu and Schlosser, 2012):

$$V_z = \left(\frac{u_*}{\kappa}\right) \log\left[\frac{(z-d)}{z_0}\right], \tag{4}$$

where V_z denotes the wind speed at height z, u_* the friction velocity, κ the von Kármán constant ($\sim$0.41) used for fluid modelling, d the displacement height, and z_0 the surface roughness. The three unknown parameters u_*, d, and z_0, can be calculated by solving the three dimensional equation system for the wind speeds at 2m, 10m, and 50m.[3] By plugging in the turbine height for z, the desired wind speed at turbine height z, V_z, can be obtained.

2.5 Conversion of wind speed to production

When the wind speed at the turbine position is derived, the most crucial step is the conversion into produced energy. One way is applying a physical transformation such as the wind power density (WPD), which describes how much of the kinetic energy of the wind per area can be transformed into energy production (Hennessey Jr, 1977). It is defined as

$$\text{WPD} = \frac{1}{2}\, \rho\, C_P\, V_z^3, \tag{5}$$

where V_z denotes the wind speed at turbine height z, ρ the air density, and C_P the Betz limit (=16/27), which describes the maximum amount of energy a turbine can theoretically extract from the wind. The unit of the WPD is W/m^2. By multiplying the WPD with the circular area spanned by the rotor blades (= diameter $\times\, \pi$), the wind power potential for a turbine can be achieved. Empirical evidence shows, however, that WPD overestimates the real on-site production and is only an illustrative point (Gunturu and Schlosser, 2012).

[3] A very efficient way of solving this equation system is the Newton-Raphson method described under `http://www.met.reading.ac.uk/~marc/it/wind/`.

Another approach is to estimate a wind energy production function. For every turbine type, the producer offers a power curve which describes the amount of energy that can be produced depending on the current wind speed. Unfortunately, this power curve cannot be used for our purpose for two reasons: First, data for instantaneous wind is not available, only hourly average wind. Because of the non-linear relation between wind speed and energy production, the hourly wind production cannot be generated by inserting hourly average wind speed into the power curve (Brown et al., 1984; Sinden, 2007). Second, the power curve captures the relation between true mast wind speed and production. However, processed MERRA data and mast wind data are different and cannot simply replace each other. For these reasons, a new production function linking hourly MERRA wind speed data and production data is needed.

In this paper, we examine the relation between the observed wind speed and the resulting production from a statistical point of view and estimate the underlying function. A natural candidate is a **3rd order polynomial** (cf. Llombart et al., 2006) because of the cubic relation between wind speed and energy production (compare the wind power density in Eq. (5)):

$$f(x; a, b, c, d) = ax^3 + bx^2 + cx + d, \tag{6}$$

where $a, b, c, d \in \mathbb{R}$. This function, however, is unbounded whereas production data is bounded by zero and the maximal production C depending on the rated capacity. Hence, a second candidate is a **piecewise defined function** which bounds the 3rd order polynomial at the thresholds x_1 and x_2, $0 \leq x_1 \leq x_2$ (Chang et al., 2003):

$$f(x; a, b, c, d, x_1, x_2, C) = \begin{cases} 0 & 0 \leq x < x_1 \\ ax^3 + bx^2 + cx + d & x_1 \leq x \leq x_2 \\ C & x > x_2 \end{cases} \tag{7}$$

To smooth the transitions at the thresholds for a more realistic shape, we additionally assume continuity and differentiability, i.e., $f(x_1) = 0$, $f(x_2) = C$, $f'(x_1) = 0$, and $f'(x_2) = 0$. The thresholds x_1 and x_2 are estimated from the data.

Both aforementioned functions imply that the relation between wind speed and production is point symmetric. Another function type capturing the boundedness and the typical "S" shape of the production function is the class of logistic functions. A special type of logistic function also allowing for asymmetry is the **five parameter logistic (5PL) function** (Gottschalk and Dunn, 2005):

$$f(x; a, b, c, d, g) = d + \frac{a - d}{\left(1 + (\frac{x}{c})^b\right)^g} \tag{8}$$

with $a, b, d \in \mathbb{R}$ and $c, g \in \mathbb{R}^+$. The parameters d and a describe the lower and upper bounds, respectively, and are set to the minimal and maximal production. The parameters b, c and g determine the slope of the function, where g particularly controls the asymmetry (symmetric for $g = 1$).

When the best function type is determined and fitted to the available production data, it can be used to estimate the production at a new location where only wind speed data are available.

2.6 Wind energy index

The index we suggest to estimate the production potential at a certain location translates the derived wind speed at this location into the expected wind energy is defined as follows:

$$I(\tau_1, \tau_2) = \sum_{t=\tau_1}^{\tau_2} f(V_z(t)), \tag{9}$$

where $V_z(t)$ indicates the hourly wind speed at turbine location and turbine height obtained according to Sections 2.2–2.4 and $f(\cdot)$ is the best fitting function from Section 2.5. τ_1 and τ_2 denote the start and the end date of the index accumulation. The estimated hourly production can be summed up for different time horizons such as daily, monthly, or yearly, depending on the aim and the data availability. For the yearly index, for example, the time t changes in hourly steps, i.e., $t = 1, \ldots, 8760$ for a common year. To estimate the long-term potential of a location, we average the values of the yearly index over an adequately long period.

2.7 Validation

To evaluate the performance of our models, we compare the simulated production from Eq. (9) with the true production on different aggregation levels. First, we calculate Pearson's correlation coefficient to examine their dependency. Second, we measure the estimation accuracy by the root-mean-square error (RMSE) defined as:

$$\mathrm{RMSE}^{\Delta\tau} = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} \left(\hat{I}_i^{\Delta\tau} - I_i^{\Delta\tau}\right)^2}, \tag{10}$$

where $\hat{I}_i^{\Delta\tau}$ and $I_i^{\Delta\tau}$ are the estimated and the true productions for time period i, respectively. $\Delta\tau$ indicates the level of aggregation, i.e., hourly, daily, or monthly, and N the number of observations on this aggregation level. Because our production data record is not long enough, we do not compare the results on a yearly scale.

When the production function for a certain turbine type is estimated based on all data available, we assume that it is valid for all locations with the same turbine type in Germany. To test if this assumption holds true, we perform a leave-one-out cross validation (e.g. Arlot and Celisse, 2010): Instead of using all n locations for fitting the production function, we take only $n - 1$ locations. The left-out location then simulates a new, unobserved location and is used to test the estimated function. This procedure is repeated n times so that each location is once the left-out location.

3 Empirical analysis

3.1 Wind farm data

We use data for wind energy production at seven German wind parks A–G summarized in Table 1.[4] The wind parks are situated in different regions of Germany (see Fig. 11) and

[4]The production data are provided by 4inita GmbH.

Wind park	# Turb.	Height	Start	End
A	1	140	15.11.2011	30.06.2014
B	4	79	25.07.2012	30.06.2014
C	6	138	01.01.2013	30.06.2014
D	3	138	01.01.2013	30.06.2014
E	6	138	16.01.2012	30.06.2014
F	5	108	01.01.2013	30.06.2014
G	8	138	20.12.2011	30.06.2014

Table 1: List of available production data

consist of a different number of turbines, which all have the same type and a capacity of 2.3 MW.[5] The data are reported in an interval of 10 minutes and last minimum 1.5 years.

We cleaned the data according to the error code provided by each turbine, i.e., the 10 min production is set to "NaN" in case of an error. By this procedure, we manage to estimate the true relation between wind speed and production regardless of technical issues. When summing up the 10 min data to hourly production, we allow for 1/3 missing values until we set the hourly value to "NaN". The same rule is applied for aggregating to daily and monthly scales.

The number of turbines varies from 1 to 8 among the wind parks (see Table 1). Because the turbines influence each other's wind conditions and efficiency, we average the production of all turbines in a wind park to obtain a time series representative for the whole park. Table 2 shows the average production for each wind park's average turbine on different time scales. The average monthly production ranges from 280 MWh to 492 MWh indicating different topographical wind conditions. Moreover, the share of missing values lies between 0 % and 4 % for the monthly production data.

As an example, a month with typical hourly production is shown in Fig. 1, as well as the corresponding hourly wind speed at turbine height. This so-called mast wind speed is also provided for each turbine in steps of 10 minutes. Fig. 2 depicts the relation between the hourly average mast wind and the hourly total production for a certain turbine of wind park A. The typical "S" shape known from power curves is also visible here: The production is zero for low wind speeds; then, the production increases up to the capacity. After the capacity is reached, the production stays constant with increasing wind. For very high wind speeds, the production even decreases because the turbine is disconnected to prevent damages. The points further away from the curve might come from measurement errors of the wind speeds or technical problems not captured by the error code. As we later fit the production function to the MERRA wind speed, we do not further investigate the outliers of the mast wind data.

3.2 MERRA data

The MERRA data used in this study come from the "MERRA IAU 2d atmosperic single level diagnostics (AT1NXSLV)" and are available at times 0:30, 1:30, 2:30, ... for each day since 1979 (Lucchesi, 2012). We use the variables `U2M`, `V2M`, `U10M`, `V10M`, `U50M`, and `V50M`,

[5]The names and exact locations of the wind parks are concealed here for confidentiality reasons.

	Hourly		Daily		Monthly	
	Mean	NaN	Mean	NaN	Mean	NaN
A	0.67	0.22 %	16.06	0.10 %	492.29	3.13 %
B	0.38	0.83 %	9.21	1.13 %	279.87	4.17 %
C	0.58	3.01 %	13.91	3.48 %	421.84	0.00 %
D	0.50	0.64 %	11.94	0.73 %	362.04	0.00 %
E	0.53	1.90 %	12.55	2.68 %	383.11	3.33 %
F	0.50	0.39 %	11.94	0.37 %	361.82	0.00 %
G	0.43	2.20 %	10.41	3.14 %	313.23	3.23 %

Table 2: Mean production values (in MWh) and share of missing values on different time scales

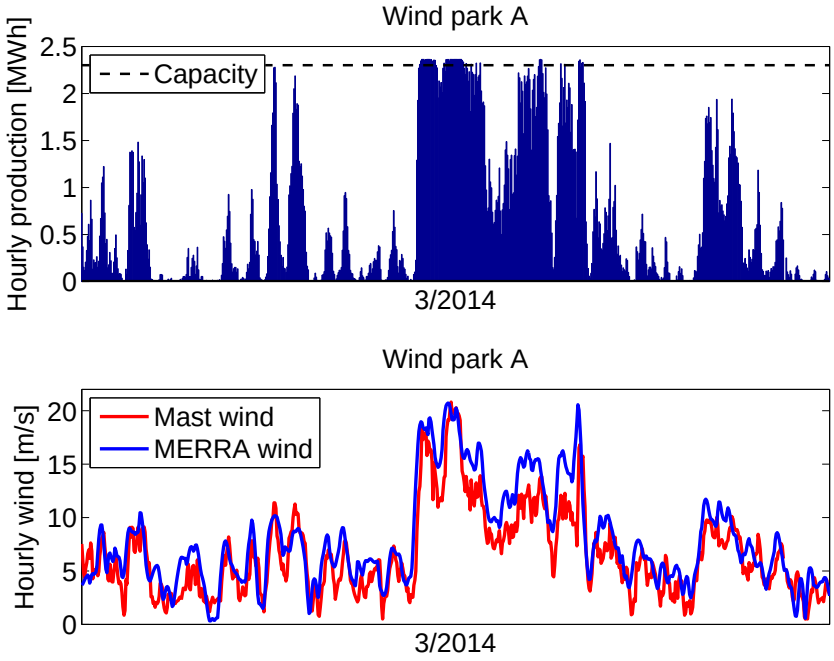

Figure 1: Hourly production and hourly average wind speed (mast and MERRA) for an exemplary month of wind park A

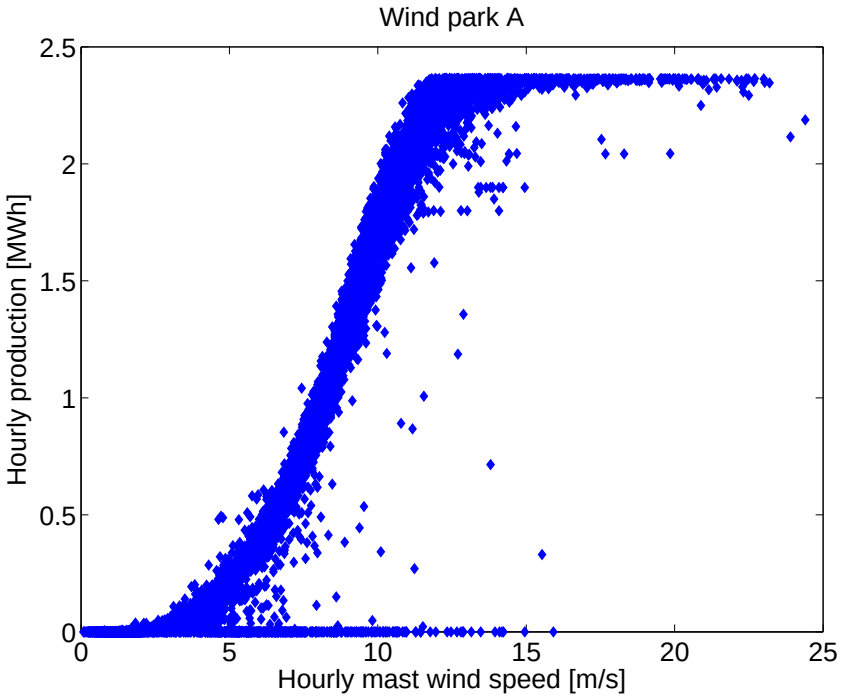

Figure 2: Hourly average mast wind compared with hourly production

which indicate the eastward and northward wind speeds measured in m/s at heights of 2m, 10m, and 50m above surface. To cover whole Germany, all grid points with latitude between 5.33° E and 16° E and longitude between 47° N and 56° N are used. The grid points in Germany are depicted in Fig. 11.

3.2.1 Wind speed and direction

The eastward and northward wind components u_z and v_z at height z, respectively, allow for deriving the wind speed and the wind direction. Fig. 3 (left) depicts the histogram of the wind speeds at 2m height at the MERRA grid point 13.3° E/52.5° N, which is located in Berlin, for 2004–2013. It shows that wind speeds around 3 m/s are most common at this location. Moreover, the wind speed follows a Weibull distribution, which is the standard distribution for modelling wind speeds (e.g. Hennessey Jr, 1977; Gunturu and Schlosser, 2012). Fig. 3 (right) depicts an angle histogram of the wind direction in Berlin in 2004–2013. It shows that most of the wind comes from west in Berlin (180°).[6]

3.2.2 MERRA wind vs. mast wind

The wind speeds derived from MERRA data are used to replace the mast wind speeds which are not available for new locations, at least at an early planning stage. However, we can compare the mast wind speeds available in our dataset with the MERRA wind speeds for the same location. Fig. 4 and Table 3 illustrate the relationship between the hourly average mast and MERRA wind speeds. The correlation is always higher than 0.81 at

[6]This information can be used to adjust a new wind park so that turbulences are minimized and the efficiency is maximized for the main wind direction.

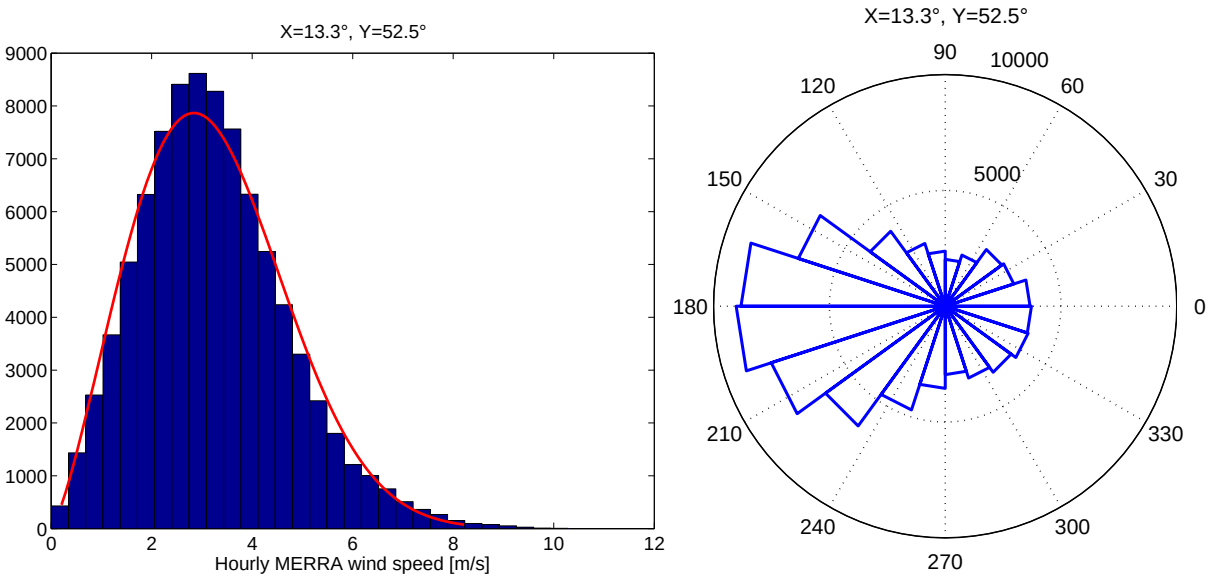

Figure 3: Histogram of wind speeds (left) with fitted Weibull distribution and angle histogram of wind direction (right); data 01/01/2004–31/12/2013, Berlin, 2M

	A	B	C	D	E	F	G
Correlation	0.8515	0.8338	0.8621	0.8127	0.8471	0.8142	0.8547

Table 3: Correlation between mast wind and MERRA wind

all wind parks which indicates that MERRA data is a suitable substitute for local wind data. The figure for wind park A, however, also reveals that there might be a systematic bias, i.e., that the processed MERRA wind speeds generally overestimate the mast wind speeds, which can also be conjectured from Fig. 1. By multiplying the MERRA data with a location-dependant factor, this bias could be mitigated, but this factor is unavailable for an unobserved location. Nevertheless, we can disregard this problem as we directly estimate the relation between the production and the MERRA data and not the mast wind data.

3.3 Relation between MERRA wind and production

3.3.1 MERRA wind production function

In this section, we compare the different functions types introduced in Section 2.5 to describe the relationship between average hourly MERRA wind speed and the true hourly energy production. Fig. 5 (left) reveals the weakness of the 3rd order polynomial exemplarily for wind park A: The overall fitting is good, but the shape at the boundaries does not reflect the traits of wind power production because the production does neither increase for very low wind nor fall immediately after the maximum. These drawbacks are overcome by the piecewise defined function in Fig. 5 (right), but despite the additional assumptions, the overall fitting does not improve (see Table 4). The 5PL function (Fig. 6) reflects best the actual behaviour of production data. It allows for asymmetry, is easy

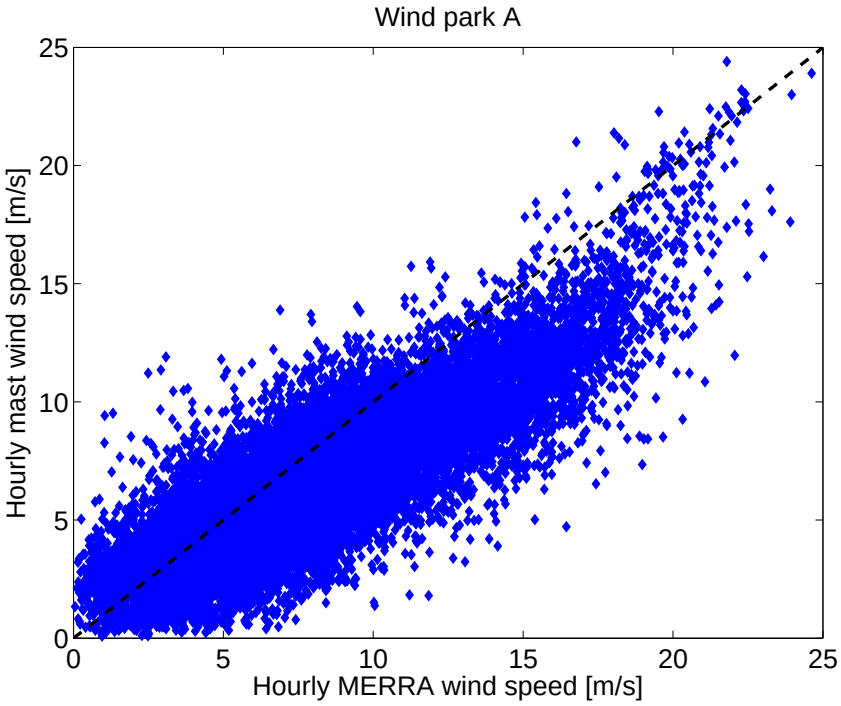

Figure 4: Comparison of hourly MERRA wind speed with hourly mast wind speed

R^2	A	B	C	D	E	F	G
3rd order polynomial	0.698	0.758	0.709	0.656	0.708	0.681	0.734
Piecewise function	0.697	0.751	0.709	0.654	0.706	0.680	0.732
5 parameter logistic	0.698	0.760	0.710	0.656	0.708	0.683	0.735

Table 4: Goodness of fit (measured by R^2) for different functions

to calibrate and shows the best fit among the considered candidates (see Table 4). For these reasons, we continue with using the 5PL function for the wind energy index. The estimated parameters for each wind park are given in Table 5. It can be seen that the parameter g is not equal to 1, hence the functions are asymmetric. The parameter a indicates the maximal production in MWh, d the minimal production (0).

3.3.2 In-sample estimation

When plugging in the MERRA wind speeds into the fitted 5PL function, we obtain estimated values for the hourly production, which we call hourly "MERRA production". Figures 7–9 compare the MERRA production with the true production for wind park A on the hourly, daily, and monthly scale. It can be seen that the fit becomes better for higher scales, which is also confirmed by an increase of the correlation from 0.82 (hourly) to 0.92 (daily) and 0.98 (monthly) (see Table 6). This can be explained by an averaging effect of estimation errors. The RMSE increases from 0.39 (hourly) to 5.2 (daily) and 38.5 (monthly), but this increase results from different magnitudes of the production on different time scales: The RMSE for hourly production for wind park A corresponds to 57% of the hourly production (0.39/0.69), whereas the RMSEs for the daily and monthly

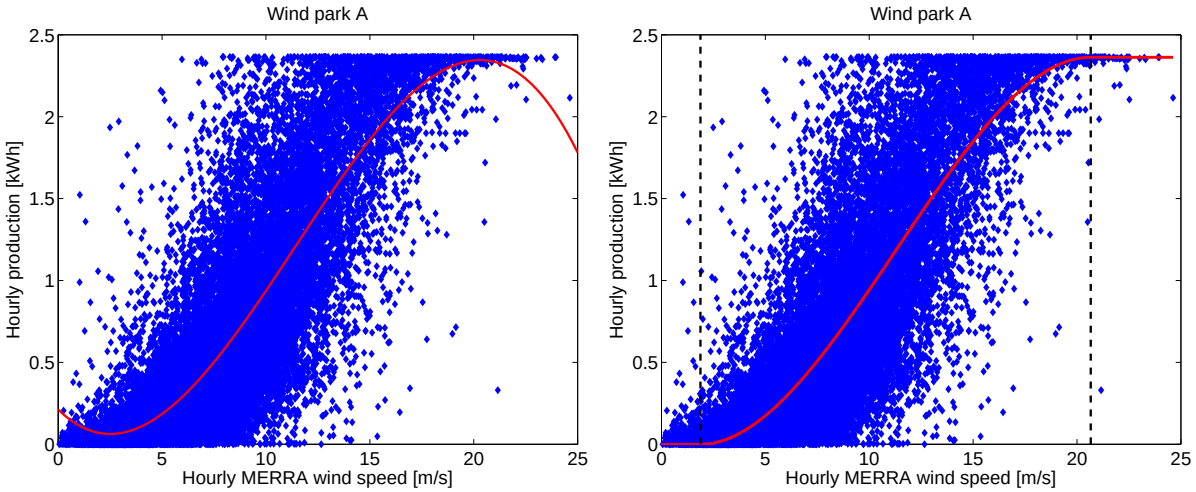

Figure 5: Fitting the hourly production for hourly MERRA wind speed using the 3rd order polynomial (left) and the piecewise function (right)

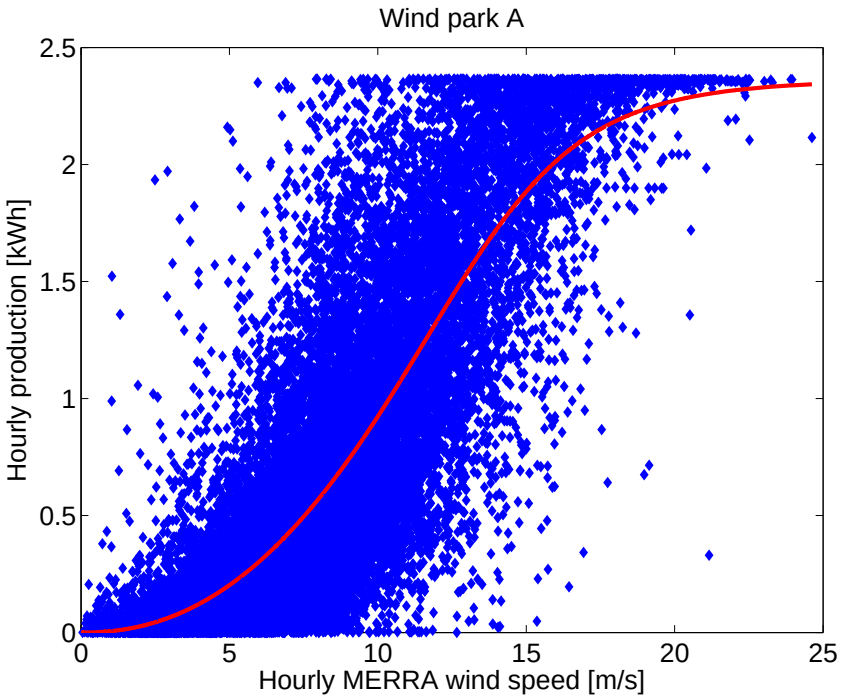

Figure 6: Fitting the hourly production for hourly MERRA wind speed using the 5PL function

	a	b	c	d	g
A	2.365	-7.219	15.200	0	0.306
B	2.361	-15.855	18.247	0	0.144
C	2.363	-6.039	16.718	0	0.377
D	2.264	-6.691	16.519	0	0.334
E	2.364	-7.229	15.052	0	0.340
F	2.361	-7.112	12.956	0	0.383
G	2.360	-6.315	17.020	0	0.370
All	2.365	-6.627	15.945	0	0.351

Table 5: Estimated parameters of the 5PL function

	Hourly			Daily			Monthly		
	Mean	Corr.	RMSE	Mean	Corr.	RMSE	Mean	Corr.	RMSE
A	0.69	0.82	0.39	16.48	0.92	5.18	500.73	0.98	38.46
B	0.38	0.87	0.21	9.14	0.96	2.42	273.10	0.98	18.49
C	0.57	0.84	0.31	13.77	0.94	3.74	417.59	0.97	30.38
D	0.51	0.81	0.32	12.17	0.93	4.06	369.27	0.95	49.49
E	0.54	0.83	0.34	12.93	0.94	4.28	396.47	0.97	36.61
F	0.51	0.82	0.35	12.16	0.89	5.64	368.90	0.95	57.39
G	0.44	0.84	0.26	10.49	0.94	3.26	322.60	0.96	33.06

Table 6: Results of in-sample estimation

production correspond to 31% (5.18/16.48) and 8% (38.46/500.73) of the total average production in these periods. The good fit on the monthly scale is also visible from Fig. 10 where the monthly true and MERRA productions are depicted for more than 2.5 years for wind park A.

The average ratio of the RMSE to the monthly production for all wind parks lies around 10%. Because of the better results on higher time scales, we conjecture that the yearly scale, which we use to assess the wind energy potential, leads to an even better approximation.

3.3.3 Out-of-sample estimation

In this section, we evaluate the performance of our approach in estimating the production at an unobserved location (out-of-sample) by conducting a leave-one-out cross validation. To obtain the estimated production at one location, the parameters of the MERRA wind production function (Eq. (8)) are estimated using a training dataset of six out of seven wind parks.[7] Instead of a joint estimation for all six wind parks, one could also estimate the production function of the nearest neighbour only. Given different geographical con-

[7]We are aware that different lengths of the data record put different weights on the wind parks for the joint estimation of the production function (see Table 1). Nevertheless, we prefer not to shorten the data, but to use as many data as possible.

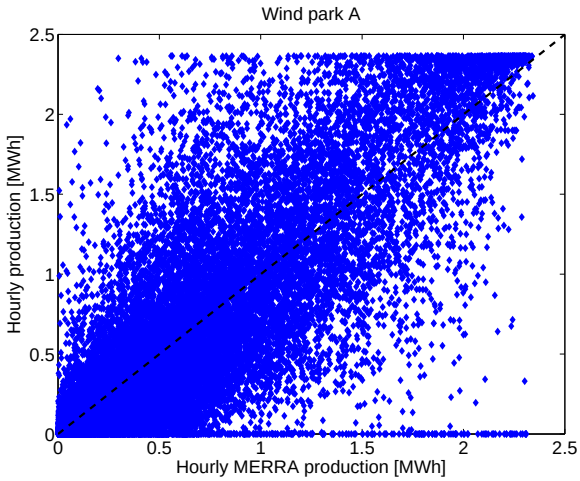

Figure 7: Hourly MERRA production vs. true hourly production

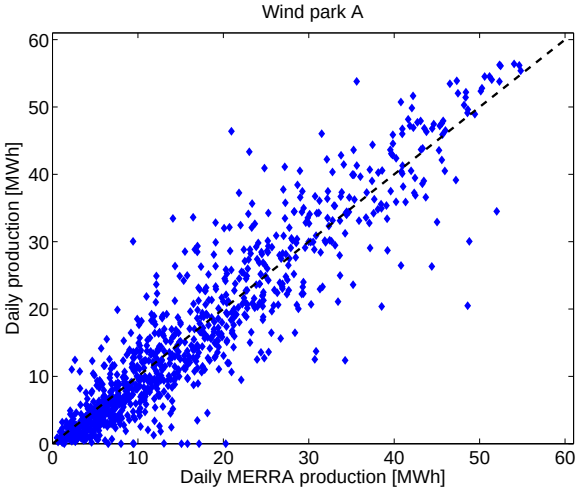

Figure 8: Daily MERRA production vs. true daily production

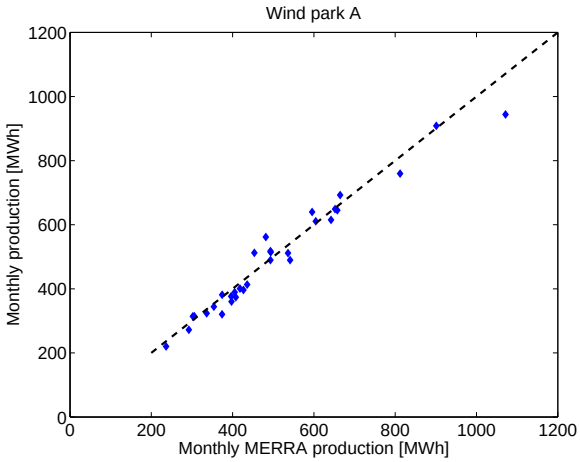

Figure 9: Monthly MERRA production vs. true monthly production

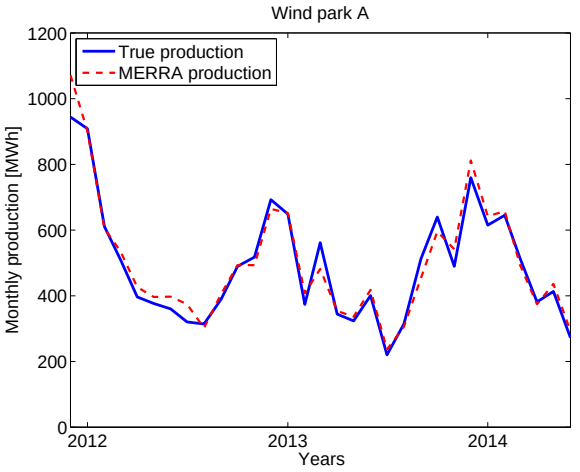

Figure 10: Temporal development of monthly MERRA production and true monthly production

	Hourly			Daily			Monthly		
	Mean	Corr.	RMSE	Mean	Corr.	RMSE	Mean	Corr.	RMSE
A	0.57	0.82	0.41	13.71	0.92	5.88	416.63	0.98	83.51
B	0.50	0.87	0.26	12.03	0.95	4.15	359.47	0.98	92.99
C	0.63	0.84	0.32	15.10	0.94	4.00	458.16	0.97	48.23
D	0.55	0.81	0.33	13.11	0.93	4.18	397.91	0.95	58.08
E	0.50	0.83	0.34	11.98	0.94	4.57	367.40	0.97	45.10
F	0.38	0.81	0.39	9.09	0.89	7.22	275.85	0.94	116.98
G	0.52	0.84	0.29	12.47	0.94	4.12	383.24	0.96	74.66

Table 7: Results of out-of-sample estimation

ditions, however, this entails the risk of a misspecified production function. Therefore, we use all available (training) data to estimate the average relationship.

With the estimated relation and local MERRA wind data, the MERRA production can be derived for the left-out location and compared with the true out-of-sample production data from the seventh location using again correlation and RMSE on different scales (see Table 7). As expected, the RMSE for the out-of sample estimation increases compared to in-sample estimation. The change is also observable in the means. The higher errors for the wind parks B and F on the monthly scale can possibly be explained by the turbine heights, which differ from those in the other wind parks (see Table 1). Moreover, the in-sample fitting for wind park F was already rather poor and wind park B is situated further away from the other wind parks. Opposite to the RMSE, the correlation remains almost equal compared to in-sample estimation because the wind speed is the main driver of direction of variation regardless of the production function.

3.4 Wind energy potential in Germany

The main advantage of MERRA data is the availability of long-term wind speed data on a global grid. With these data and the aforementioned approach, we can estimate the wind energy potential for every location in Germany by averaging the yearly wind energy index based on historical wind speed data. The adequate length of times series is widely discussed in the literature and ranges from minimum ten years (Jimenez et al., 2012) over 15–20 years (Liléo et al., 2013) to 25 years (Brower et al., 2013) to balance large fluctuations and not to be biased by structural breaks in the wind speed data due to climate change or reanalysis data developments.

In this context, we choose a time horizon of twenty years, i.e., 1994–2013. Rather stationary wind conditions can be anticipated for this period: The estimated yearly production shows a significant trend only for 30 % of the grid points at the 5 % significance level according to the Mann-Kendall test. When assessing the potential of a specific location, however, it has to be investigated if a trend has to be considered.

Fig. 11 (left) shows a map of the averaged yearly index, i.e., the expected yearly production, for each MERRA grid point and their interpolation.[8] It depicts a rather low

[8]Our approach allows for calculating MERRA production at an arbitrary resolution. To decrease the computational effort, we refrain from that and interpolate between the grid points using Natural Neighbour interpolation in ArcGIS 10.2.

potential in southern Germany, but a high potential near the sea. Of course, this map describes only the production potential depending on the wind speed. The geographical and structural situation such as the existence of cities or lakes has to be considered as well for the actual planning. Moreover, the map is turbine-specific, hence we assume the same technology as used in the wind parks under consideration. Opposite to classical wind maps, it provides the estimated amount of energy that can be produced under the local wind conditions.

Fig. 11 (right) depicts the coefficient of variation, i.e., the standard deviation of each location normalized by the location's mean. This value is an important indicator for the (model) risk involved in installing a new wind park. It follows that the risk is much lower near the coast with fluctuations around 5% compared to the south with fluctuations around 10%.

4 Discussion and conclusion

In this paper, we provide a novel and transparent approach to estimate the long-term wind energy potential at an unobserved location by applying a newly developed wind energy index. The production data available for a certain turbine type is used to estimate a general production function which can then be applied to wind data at any new location. The wind energy index provides the expected long-term energy production for this location and a certain turbine type. The resulting wind energy production map for Germany is useful for governments, practitioners, or investors who are involved the value chain of a wind farm investment.

Therefore, our approach could meet the following needs: First, it allows a pre-assessment of the suitability of potential location at no costs before analyzing the production potential in greater detail by means of site-specific wind measurements. Second, it could fill the gap of missing standards of assessing the wind energy potential from a legal point of view. Third, it could assist in creating a transparent approach for the valuation of wind production derivatives.

However, to achieve any of the aforementioned potentials, this approach has to be adapted to other turbine types, which is possible as long as real production data for this type is available at least at one other location. Moreover, it can be transferred to other regions in the world as MERRA data are globally available.

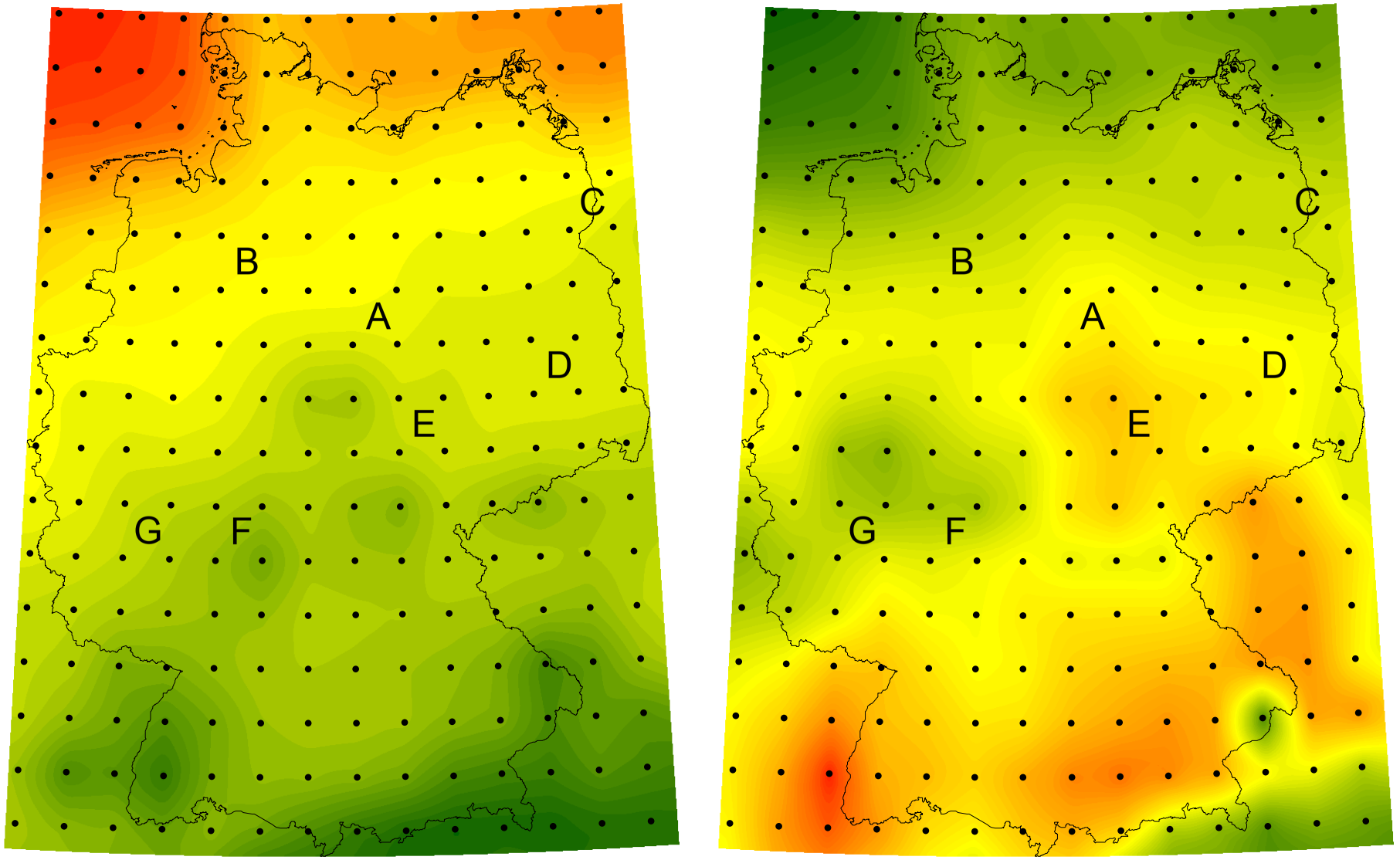

Figure 11: Map of estimated yearly production (left) and coefficient of variation (right) over 1994–2013; dots indicate MERRA grid
Annotation: The estimated annual production (left) varies from 1,250 MWh (dark green) to 10,250 MWh (dark red) in steps of 250 MWh, the coefficient of variation (right) from 5% (dark green) to 11.1% (dark red) in steps of 0.1%.

References

Archer, C. L. and Jacobson, M. Z. (2003). Spatial and temporal distributions of US winds and wind power at 80 m derived from measurements. *Journal of Geophysical Research: Atmospheres*, 108(D9).

Archer, C. L. and Jacobson, M. Z. (2005). Evaluation of global wind power. *Journal of Geophysical Research: Atmospheres*, 110(D12).

Arlot, S. and Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4:40–79.

Betreiber-Datenbasis (2011). *Einführung in den BDB-Index der Betreiber Datenbasis (BDB)*. `http://www.btrdb.de/PDF/IndexV2011Einfuehrung.pdf`.

Brower, M. C., Barton, M. S., Lledó, L., and Dubois, J. (2013). A study of wind speed variability using global reanalysis data. Technical report, AWS True Power.

Brown, B. G., Katz, R. W., and Murphy, A. H. (1984). Time series models to simulate and forecast wind speed and wind power. *Journal of climate and applied meteorology*, 23(8):1184–1195.

Bundesverband WindEnergie (2013). *Ergebnisbericht der Arbeitsgruppe Langzeitbezug des BWE-Windgutachterbeirates*.

Caporin, M. and Preś, J. (2012). Modelling and forecasting wind speed intensity for weather risk management. *Computational Statistics & Data Analysis*, 56(11):3459–3476.

Carta, J. A., Velázquez, S., and Cabrera, P. (2013). A review of measure-correlate-predict (MCP) methods used to estimate long-term wind characteristics at a target site. *Renewable and Sustainable Energy Reviews*, 27(0):362–400.

Carvalho, D., Rocha, A., Gómez-Gesteira, M., and Santos, C. S. (2014). WRF wind simulation and wind energy production estimates forced by different reanalyses: Comparison with observed data for Portugal. *Applied Energy*, 117(0):116–126.

Chang, T.-J., Wu, Y.-T., Hsu, H.-Y., Chu, C.-R., and Liao, C.-M. (2003). Assessment of wind characteristics and wind turbine characteristics in Taiwan. *Renewable Energy*, 28(6):851–871.

Deutscher Wetterdienst (2009). Windgeschwindigkeiten in der Bundesrepublik Deutschland. `http://www.dwd.de/windkarten`.

Fırtın, E., Güler, Ö., and Akdağ, S. A. (2011). Investigation of wind shear coefficients and their effect on electrical energy generation. *Applied Energy*, 88(11):4097–4105.

Gastón, M., Pascal, E., Frías, L., Martí, I., Irigoyen, U., Cantero, E., Lozano, S., and Loureiro, Y. (2008). Wind resources map of Spain at mesoscale. Methodology and validation. In *Proceedings of European wind energy conference*.

Gottschalk, P. G. and Dunn, J. R. (2005). The five-parameter logistic: a characterization and comparison with the four-parameter logistic. *Analytical biochemistry*, 343(1):54–65.

Gunturu, U. B. and Schlosser, C. A. (2012). Characterization of wind power resource in the United States. *Atmospheric Chemistry and Physics*, 12(20):9687–9702.

GWEC (2014). Global wind report – annual market update 2013. *Global Wind Energy Council*.

Hennessey Jr, J. P. (1977). Some aspects of wind power statistics. *Journal of applied meteorology*, 16(2):119–128.

Jimenez, B., Monnich, K., and Durante, F. (2012). Comparison between NCEP/NCAR and MERRA reanalysis data for long term correction in wind energy assessment. In *Proceedings of the EWEA Annual Event 2012, Copenhagen, Denmark*.

Jung, S., Vanli, O. A., and Kwon, S.-D. (2013). Wind energy potential assessment considering the uncertainties due to limited data. *Applied Energy*, 102(0):1492–1503. Special Issue on Advances in sustainable biofuel production and use - {XIX} International Symposium on Alcohol Fuels - {ISAF}.

Kotroni, V., Lagouvardos, K., and Lykoudis, S. (2014). High-resolution model-based wind atlas for Greece. *Renewable and Sustainable Energy Reviews*, 30(0):479–489.

Kubik, M., Brayshaw, D. J., Coker, P. J., and Barlow, J. F. (2013a). Exploring the role of reanalysis data in simulating regional wind generation variability over Northern Ireland. *Renewable Energy*, 57:558–561.

Kubik, M., Coker, P., Barlow, J., and Hunt, C. (2013b). A study into the accuracy of using meteorological wind data to estimate turbine generation output. *Renewable Energy*, 51(0):153–158.

Liléo, S., Berge, E., Undheim, O., Klinkert, R., and Bredesen, R. E. (2013). Long-term correction of wind measurements: State-of-the-art, guidelines and future work. *Elforsk report*, 13:18.

Liléo, S. and Petrik, O. (2011). Investigation on the use of NCEP/NCAR, MERRA and NCEP/CFSR reanalysis data in wind resource analysis. In *Proceedings of the EWEA Annual Event 2011, Brussels, Belgium*.

Llombart, A., Pueyo, C., Fandos, J., and Guerrero, J. (2006). Robust data filtering in wind power systems. *European wind energy conference EWEC, Athens*.

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`.

Luo, W., Taylor, M., and Parker, S. (2008). A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from England and Wales. *International Journal of Climatology*, 28(7):947–959.

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., and Woollen, J. (2011). MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *Journal of Climate*, 24(14).

Sanchez, I. (2006). Short-term prediction of wind energy production. *International Journal of Forecasting*, 22(1):43–56.

Sinden, G. (2007). Characteristics of the UK wind resource: Long-term patterns and relationship to electricity demand. *Energy Policy*, 35(1):112–127.

Staffell, I. and Green, R. (2014). How does wind farm performance decline with age? *Renewable Energy*, 66(0):775–786.

Stull, R. B. (1988). *An introduction to boundary layer meteorology*, volume 13. Springer.

Troen, I. and Petersen, E. L. (1989). *European wind atlas*. Risø National Laboratory, Roskilde.

SFB 649 Discussion Paper Series 2014

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Principal Component Analysis in an Asymmetric Norm" by Ngoc Mai Tran, Maria Osipenko and Wolfgang Karl Härdle, January 2014.
002 "A Simultaneous Confidence Corridor for Varying Coefficient Regression with Sparse Functional Data" by Lijie Gu, Li Wang, Wolfgang Karl Härdle and Lijian Yang, January 2014.
003 "An Extended Single Index Model with Missing Response at Random" by Qihua Wang, Tao Zhang, Wolfgang Karl Härdle, January 2014.
004 "Structural Vector Autoregressive Analysis in a Data Rich Environment: A Survey" by Helmut Lütkepohl, January 2014.
005 "Functional stable limit theorems for efficient spectral covolatility estimators" by Randolf Altmeyer and Markus Bibinger, January 2014.
006 "A consistent two-factor model for pricing temperature derivatives" by Andreas Groll, Brenda López-Cabrera and Thilo Meyer-Brandis, January 2014.
007 "Confidence Bands for Impulse Responses: Bonferroni versus Wald" by Helmut Lütkepohl, Anna Staszewska-Bystrova and Peter Winker, January 2014.
008 "Simultaneous Confidence Corridors and Variable Selection for Generalized Additive Models" by Shuzhuan Zheng, Rong Liu, Lijian Yang and Wolfgang Karl Härdle, January 2014.
009 "Structural Vector Autoregressions: Checking Identifying Long-run Restrictions via Heteroskedasticity" by Helmut Lütkepohl and Anton Velinov, January 2014.
010 "Efficient Iterative Maximum Likelihood Estimation of High-Parameterized Time Series Models" by Nikolaus Hautsch, Ostap Okhrin and Alexander Ristig, January 2014.
011 "Fiscal Devaluation in a Monetary Union" by Philipp Engler, Giovanni Ganelli, Juha Tervala and Simon Voigts, January 2014.
012 "Nonparametric Estimates for Conditional Quantiles of Time Series" by Jürgen Franke, Peter Mwita and Weining Wang, January 2014.
013 "Product Market Deregulation and Employment Outcomes: Evidence from the German Retail Sector" by Charlotte Senftleben-König, January 2014.
014 "Estimation procedures for exchangeable Marshall copulas with hydrological application" by Fabrizio Durante and Ostap Okhrin, January 2014.
015 "Ladislaus von Bortkiewicz - statistician, economist, and a European intellectual" by Wolfgang Karl Härdle and Annette B. Vogt, February 2014.
016 "An Application of Principal Component Analysis on Multivariate Time-Stationary Spatio-Temporal Data" by Stephan Stahlschmidt, Wolfgang Karl Härdle and Helmut Thome, February 2014.
017 "The composition of government spending and the multiplier at the Zero Lower Bound" by Julien Albertini, Arthur Poirier and Jordan Roulleau-Pasdeloup, February 2014.
018 "Interacting Product and Labor Market Regulation and the Impact of Immigration on Native Wages" by Susanne Prantl and Alexandra Spitz-Oener, February 2014.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

SFB 649 Discussion Paper Series 2014

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

019 "Unemployment benefits extensions at the zero lower bound on nominal interest rate" by Julien Albertini and Arthur Poirier, February 2014.
020 "Modelling spatio-temporal variability of temperature" by Xiaofeng Cao, Ostap Okhrin, Martin Odening and Matthias Ritter, February 2014.
021 "Do Maternal Health Problems Influence Child's Worrying Status? Evidence from British Cohort Study" by Xianhua Dai, Wolfgang Karl Härdle and Keming Yu, February 2014.
022 "Nonparametric Test for a Constant Beta over a Fixed Time Interval" by Markus Reiß, Viktor Todorov and George Tauchen, February 2014.
023 "Inflation Expectations Spillovers between the United States and Euro Area" by Aleksei Netšunajev and Lars Winkelmann, March 2014.
024 "Peer Effects and Students' Self-Control" by Berno Buechel, Lydia Mechtenberg and Julia Petersen, April 2014.
025 "Is there a demand for multi-year crop insurance?" by Maria Osipenko, Zhiwei Shen and Martin Odening, April 2014.
026 "Credit Risk Calibration based on CDS Spreads" by Shih-Kang Chao, Wolfgang Karl Härdle and Hien Pham-Thu, May 2014.
027 "Stale Forward Guidance" by Gunda-Alexandra Detmers and Dieter Nautz, May 2014.
028 "Confidence Corridors for Multivariate Generalized Quantile Regression" by Shih-Kang Chao, Katharina Proksch, Holger Dette and Wolfgang Härdle, May 2014.
029 "Information Risk, Market Stress and Institutional Herding in Financial Markets: New Evidence Through the Lens of a Simulated Model" by Christopher Boortz, Stephanie Kremer, Simon Jurkatis and Dieter Nautz, May 2014.
030 "Forecasting Generalized Quantiles of Electricity Demand: A Functional Data Approach" by Brenda López Cabrera and Franziska Schulz, May 2014.
031 "Structural Vector Autoregressions with Smooth Transition in Variances – The Interaction Between U.S. Monetary Policy and the Stock Market" by Helmut Lütkepohl and Aleksei Netsunajev, June 2014.
032 "TEDAS - Tail Event Driven ASset Allocation" by Wolfgang Karl Härdle, Sergey Nasekin, David Lee Kuo Chuen and Phoon Kok Fai, June 2014.
033 "Discount Factor Shocks and Labor Market Dynamics" by Julien Albertini and Arthur Poirier, June 2014.
034 "Risky Linear Approximations" by Alexander Meyer-Gohde, July 2014
035 "Adaptive Order Flow Forecasting with Multiplicative Error Models" by Wolfgang Karl Härdle, Andrija Mihoci and Christopher Hian-Ann Ting, July 2014
036 "Portfolio Decisions and Brain Reactions via the CEAD method" by Piotr Majer, Peter N.C. Mohr, Hauke R. Heekeren and Wolfgang K. Härdle, July 2014
037 "Common price and volatility jumps in noisy high-frequency data" by Markus Bibinger and Lars Winkelmann, July 2014
038 "Spatial Wage Inequality and Technological Change" by Charlotte Senftleben-König and Hanna Wielandt, August 2014
039 "The integration of credit default swap markets in the pre and post-subprime crisis in common stochastic trends" by Cathy Yi-Hsuan Chen, Wolfgang Karl Härdle, Hien Pham-Thu, August 2014

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

SFB 649 Discussion Paper Series 2014

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

040 "Localising Forward Intensities for Multiperiod Corporate Default" by Dedy Dwi Prastyo and Wolfgang Karl Härdle, August 2014.
041 "Certification and Market Transparency" by Konrad Stahl and Roland Strausz, September 2014.
042 "Beyond dimension two: A test for higher-order tail risk" by Carsten Bormann, Melanie Schienle and Julia Schaumburg, September 2014.
043 "Semiparametric Estimation with Generated Covariates" by Enno Mammen, Christoph Rothe and Melanie Schienle, September 2014.
044 "On the Timing of Climate Agreements" by Robert C. Schmidt and Roland Strausz, September 2014.
045 "Optimal Sales Contracts with Withdrawal Rights" by Daniel Krähmer and Roland Strausz, September 2014.
046 "Ex post information rents in sequential screening" by Daniel Krähmer and Roland Strausz, September 2014.
047 "Similarities and Differences between U.S. and German Regulation of the Use of Derivatives and Leverage by Mutual Funds – What Can Regulators Learn from Each Other?" by Dominika Paula Gałkiewicz, September 2014.
048 "That's how we roll: an experiment on rollover risk" by Ciril Bosch-Rosa, September 2014.
049 "Comparing Solution Methods for DSGE Models with Labor Market Search" by Hong Lan, September 2014.
050 "Volatility Modelling of CO2 Emission Allowance Spot Prices with Regime-Switching GARCH Models" by Thijs Benschopa, Brenda López Cabrera, September 2014.
051 "Corporate Cash Hoarding in a Model with Liquidity Constraints" by Falk Mazelis, September 2014.
052 "Designing an Index for Assessing Wind Energy Potential" by Matthias Ritter, Zhiwei Shen, Brenda López Cabrera, Martin Odening, Lars Deckert, September 2014.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".