Renken, Volker ; Sorg, Michael ; Marschner, Volker ; Gerdes, Lewin ; Gerdes, Gerhard ; Fischer, Andreas

Geographical comparison between wind power, solar power and demand for the German regions and data filling concepts

Journal Article as: peer-reviewed accepted version (Postprint)

DOI of this document*(secondary publication): https://doi.org/10.26092/elib/3311

Publication date of this document: 13/09/2024

* for better findability or for reliable citation

Recommended Citation (primary publication/Version of Record) incl. DOI:

Volker Renken, Michael Sorg, Volker Marschner, Lewin Gerdes, Gerhard Gerdes, Andreas Fischer, Geographical comparison between wind power, solar power and demand for the German regions and data filling concepts, Renewable Energy, Volume 126, 2018, Pages 475-484, ISSN 0960-1481, https://doi.org/10.1016/j.renene.2018.03.046.

Geographical comparison between wind power, solar power and demand for the German regions and data filling concepts

Volker Renken [a, *], Michael Sorg [a], Volker Marschner [b], Lewin Gerdes [c], Gerhard Gerdes [c], Andreas Fischer [a]

[a] University of Bremen, Bremen Institute for Metrology, Automation and Quality Science, Linzer Str. 13, 28359, Bremen, Germany
[b] Deutsche WindGuard Systems GmbH, Bundesallee 67, 12161, Berlin, Germany
[c] Deutsche WindGuard GmbH, Oldenburger Straße 65, 26316, Varel, Germany

A R T I C L E   I N F O

Keywords:
Renewable energy
Geographical distribution
Solar
Wind
Data-based analysis
Model-based data filling

1. Introduction

The share of renewable energy for electricity supply within Germany has reached a value of 32% in 2015 [1]. For some regions the renewable power generation is much higher than the power demand which leads to transfer needs in the electrical transmission and distribution system. Due to the decentralised nature of more than 1 million renewable energy generators, the German grid, originally designed for central power production, has to undergo a drastic change that might need additional infrastructure.

There are still open questions regarding the expansion of the renewable energy capacities towards a 100% renewable energy system. Such a system needs additional infrastructure in the form of power lines, storage systems or load flexibility. In order to be able to assess these central questions, a complete data set with high spatial and temporal resolution is required. The temporal as well as the geographical compensation effects between the fluctuating energy inputs of wind and solar power on the one side and between the renewable energy input and the demand on the other side shall be investigated. Topical investigations are based on differences for the whole German energy sector on a federal state level. An investigation of central and decentral approaches regarding the addition of renewable energy towards a 100% renewable energy input in the year 2040 has been conducted by the Reiner Lemoine Institut [2]. Therefore, Germany is distributed in 14 regions of the larger federal states including the offshore wind energy regions. Both, central and decentral scenarios lead to similar economic cost factors that do not differ from today's energy costs. A sum of 60 GW residual thermal capacities will be needed and should be exchanged by bio mass and storage systems in future scenarios with a higher renewable energy share. The study shows that from an economical point of view, the storage systems are not needed until a renewable share of 70–80% is reached. Due to the limited spatial resolution, the calculated power amounts might be optimised by data with higher resolution and from time series.

It is shown in the wind energy report [3] that the distribution of the installed wind capacity is well known. But, the time series of the

* Corresponding author.
E-mail address: ren@bimaq.de (V. Renken).

generation power input across Germany are only available by the European Energy Exchange [4] or within the region of the transmission system operators (TSO) [5–8]. Studies with higher spatial resolution rely on simulation results using meteorological data as input and real measuring data is not considered. The importance of suitable weather models and their effects to power system operators regarding the forecasting of wind energy generation was shown [9]. The day-ahead forecast leads to a mean error of 2.3% with larger deviations in the colder months of the years 2012–2014. A prediction of the wind energy generation with Neuro-fuzzy methods on an average relative error of 5.9% has been done by Saleh [10]. Modelling the wind power production can be performed either by statistical or by physical models [11]. Statistical models, which only use the time series data of the energy generation, have the disadvantage that the effects from the load or from meteorological circumstances in extreme weather situations are not included. Thus, physical implementations have a benefit. Due to the complexity of the modelling purpose, there are several points to be considered. The wind power forecast with artificial neural networks and nearest neighbour search has been investigated by Jursa [12]. Furthermore, he shows that by using intelligent model structure optimisation and combining results from different model approaches can lead to improvements. Another important aspect is the German topography. The mean wind speed is higher at exposed locations than in lower regions. In addition, also the geographic roughness of the surrounding topography is important [13].

The fluctuation of wind and solar power modelling leads to uncertainties. Regarding the temporal fluctuation, a short term prediction leads to similar error values of 22% for wind and 17% for solar forecasts within 1 h. For a time frame of 4 h, the error exceeds up to 76% for wind and 31% for solar [14]. The spatial correlation for time series of wind and solar data in relation to the distances between the sources is discussed in Ref. [15]. It can be seen that for distances of over 200 km the correlation of wind sources is below 0.6 whereas it is over 0.8 for solar sources till 700 km distance. In comparison to the wind power modelling, which is the most difficult meteorological parameter to predict [16], solar generation modelling is easier. The main reason for solar uncertainties is found in cloud coverage and precipitation and is therefore dependent on geographic regions [14]. But in general, the behaviour of the solar radiation is smoother and thus more predictable. Prediction models for solar energy and combinations of different models for different weather situations have been shown by Schmelter [17]. Several influencing parameters are weighted and historic values are included through a Gaussian distribution. Using heuristic optimisation algorithm mean errors between 1.5 and 3.5% result depending on the time of year.

In summary, the situation for the German energy network regarding the integration of growing amounts of renewable inputs is challenging. Studies leading to answers to the open questions regarding the energy system are either based on simulations or have a limited spatial resolution. Therefore, investigations are needed which combine real measurement data and a higher spatial and temporal resolution. Already, in existing references on the federal state level, there are large regional differences leading to the assumption that distributing the generated and consumed energy on a lower regional level require a more comprehensive analysis of the energy system. Therefore, these aspects will be considered within this article.

The aim of the article is to provide of a data set strongly based on true measurement data with a higher spatial resolution out of 95 zip code regions and a temporal resolution of 15 min mean values, which supports answers to the infrastructural questions of the energy system. Furthermore, different modelling methods, with regard to their ability in filling existing gaps within the data base, will be considered. For that purpose, the different methods will be compared in order to create a full dataset of the wind generated energy in a high temporal and spatial resolution. In Section 2, an overview of the database is given, followed by an introduction of the modelling methods in Section 3. The analysis and evaluation of the models is presented in Section 4. Finally, the results of the fully distributed wind energy, the resulting residual power flow and a summary are given in Section 5 and 6, respectively.

2. Overview of the data basis

To gain meaningful results regarding infrastructural questions, the quality of the data basis has the most important impact. Therefore, in this work the spatial and temporal resolution of the measurement based data shall be used on a higher level as achieved as yet. The question to be solved here is, which resolution of the input energy and also the demand can be achieved. For that purpose, different database investigations and data acquisition activities have been performed. Basis of the evaluation are the time series data of wind and solar production and for electricity demand within the regions.

The first aim is the identification of the electricity demand in each zip code region as an annual time series. This analysis was performed based on time series data from the four major transmission system operators, by annual consumption data for each distribution system operator (DSO), their supply region by zip code and area. The data was gathered from the transparency web pages of the transmission system operators (TSO), a commercially available data base for the DSO's spatial distribution. Additionally, for the spatial assignment annual information from the individual DSO web pages and their geographic shares in each postal region is applied and combined with the TSO data.

In Fig. 1, the annual electricity demand per 2-digit postal region for 2014 is shown. The demand differs widely between the different regions. It can be clearly seen, that the western areas show larger demands than the north-eastern or the south-eastern part. Because

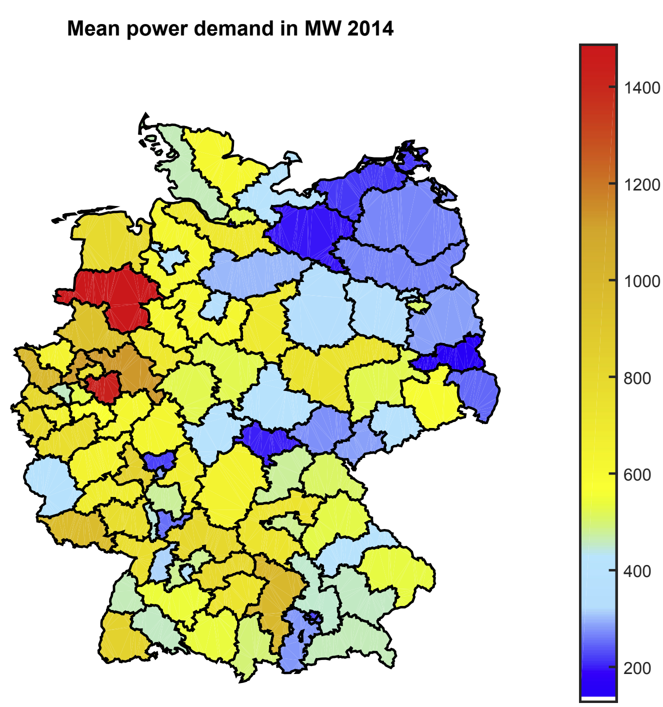

Fig. 1. Power demand within 95 postal regions in Germany.

of the diversity of the energy transmission and distribution between around 800 single companies, the generated data set is as comprehensive as possible and supports the planned further investigations.

The second aim comprises the collection of time series data for the wind and solar energy production. Regarding the distributed solar energy generation, the necessary data is already captured by SMA Solar Technology AG within its Sunny portal [18]. Fortunately, this data base, representing 13% of the German solar plants, could be integrated into the project GEOWISOL, underlying this work. The resolution of the data consists of 15 min mean values of the relative generation for all 95 two-digit postal regions. The mean solar power input for 2014 is shown in Fig. 2 (left). It can be recognised, that the most solar energy is available in the south east of Germany, but also in the eastern and north-western parts.

The data acquisition for wind data is more difficult and results in a data base with some gaps. This can be traced back to the different availability of wind energy systems over Germany and to the necessity of anonymity regarding single wind energy plants. Furthermore, compared to the solar sector, there is a greater number of companies operating the plants. Because of that, only regions can be integrated into the database if the number of wind farms and single units were large enough to allow anonymity. In Fig. 2 (right) the integrated regions can be seen, as they are mainly in the north-western part of Germany. Note that the entire data is not available and an expansion is needed to calculate the full wind energy generation for all German regions. The data represents only about 5% of the total German wind installation and is partly available for 52 of the 95 postal regions. But, after an expansion step for each of these 52 regions, the resulting regions represent about 85% of the total wind energy installation. Therefore, with the second expansion step, the calculation of the input power for the regions without input data will be done either by data of neighboured regions or by additional meteorological data. At first, the relative wind generation will be calculated and in the second step, with the support of the installed quantities per region, the absolute wind generation input will be determined.

For a complete investigation, it is essential to fill these regional data gaps. Therefore, different strategies will be discussed and compared within Section 3. Especially the south-eastern region is difficult to fill because of partly missing direct neighboured regions to gather data from. Due to the complexity of the energy system and different topographical situations, the usage of the existing data of neighboured regions for a reasoning process is not enough. It might lead to better results if additional data will be included in the models, for example meteorological data. For that purpose, measuring data from the Climate Data Centre of the German Weather Service (Deutscher Wetterdienst - DWD) [19] has been investigated in order to compare the modelling behaviour and to optimise the results. The data is based on hourly values from single weather stations across Germany and is aggregated by the mean values of 5 next neighbours to the central position of each zip code region. That way, two data sets are generated with an hourly resolution for the solar radiation and the wind speed at a height of 10 m. The mean values of all 95 zip code regions are shown in Fig. 3. Note that, the relative meteorological values are shown here in comparison to the absolute power generation per region.

As the DWD wind data has limited validity for the generated wind power in larger height, also GEOWISOL generated wind data sets will be used in the further investigation of Section 3.2.1. This data has been calculated by the wind speed data of the power plants. Therefore, for that data set also 5% of all German wind power stations have been aggregated to 52 values for the available postal regions.

3. Methods for filling gaps

As discussed in Section 2, the wind energy generation data has gaps that need to be filled in an appropriate way to come to a reliable database for both fluctuating energy sources. For this purpose, the investigation will be conducted by two general modelling approaches:

- heuristic models,
- wind speed models or
- a combination of both types.

The simple heuristic models use the available data of other regions. The wind speed models represent a relation between meteorological data and energy data, while the combination of the

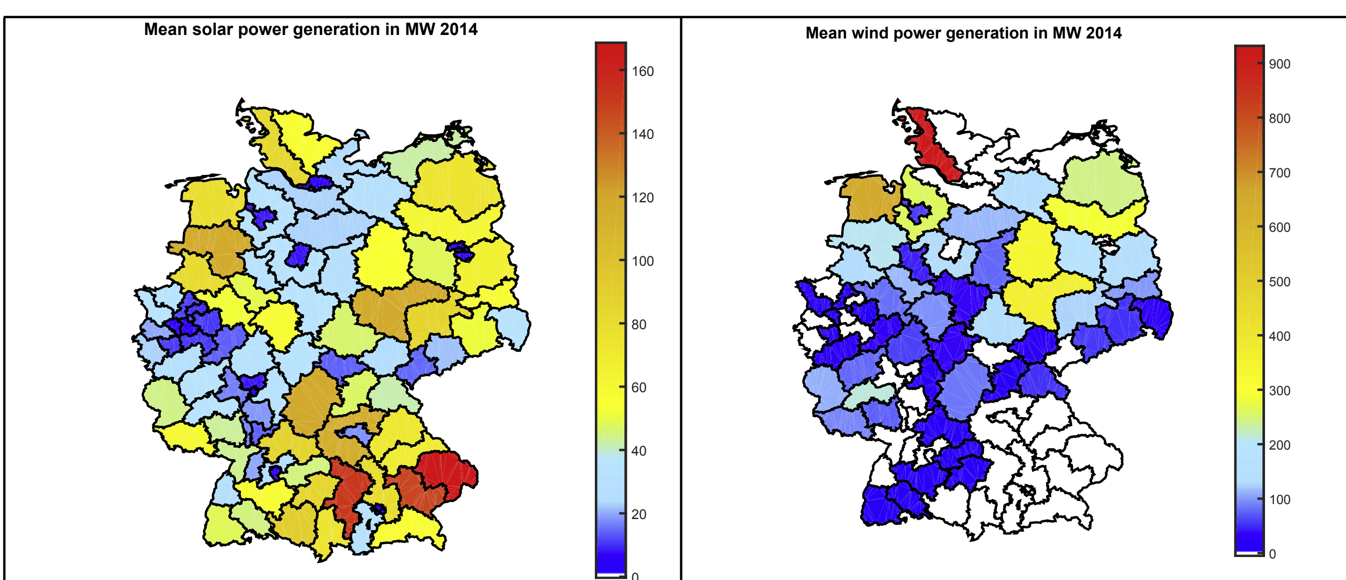

Fig. 2. Mean solar (left) and wind (right) power distribution within 95 postal regions in Germany.

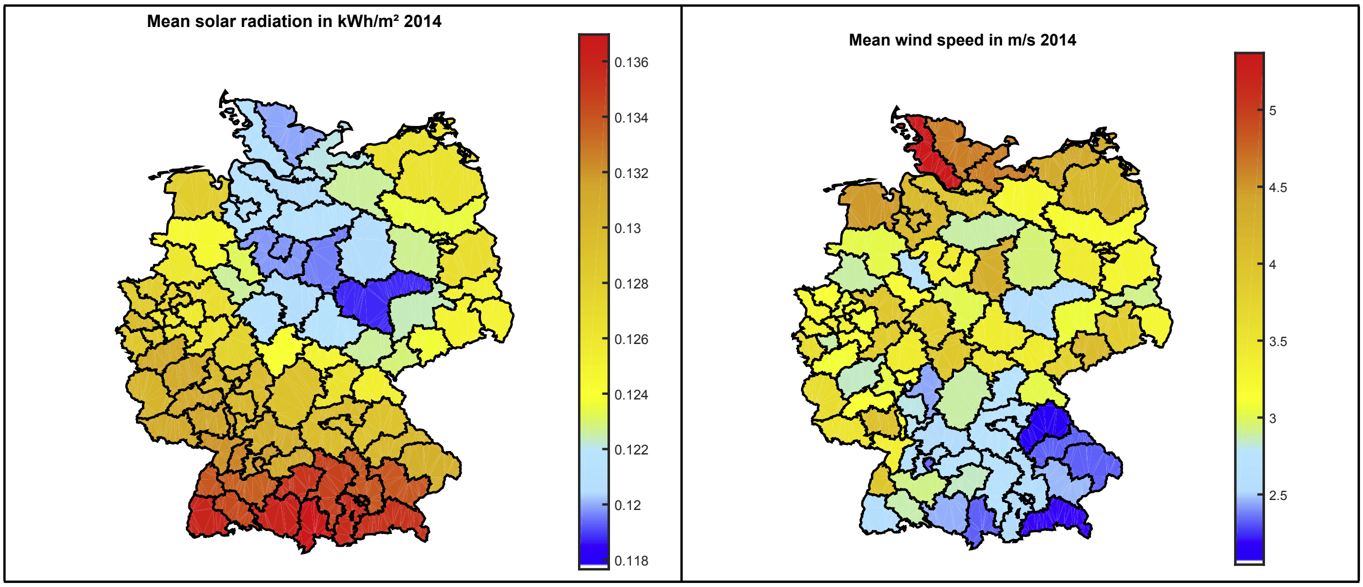

Fig. 3. Mean meteorological solar radiation (left) and wind speed (right) distributions within 95 postal regions in Germany, aggregated from data of the Climate Data Centre [19].

two models uses both the results.

3.1. Heuristic models

In this section, a number of simple heuristic models (polynomial method, method of k nearest neighbours and radial basis functions) will be investigated to fill the wind generation data gaps. The correlation of the calculated values to the target values of all regions will give an idea of the suitability of the different methods. Additionally, the modelling procedure will be used for all four introduced data sets (wind generation, solar generation, wind speed and solar radiation) for comparison. Therefore, the other data sets will be reduced so that only 52 zip code regions with wind generation data are considered.

3.1.1. Definition of heuristic models

The different model types will be investigated with regard to their ability in calculating output values that correlate to target values. In the basic approach, the target value of the model represents the power generation P, and the input values are latitude lat and longitude lon of each region.

Polynomial method

Polynomial methods offer a simple possibility to reproduce the relationship between input values and output values and often lead to sufficient results. The order of the polynomial determines the maximal appearance of the independent variables within each factor. The general formula for two input variables can be seen as follows:

$$\begin{aligned}\widehat{P} &= P(lat, lon) \\ &= a_{0,0} + a_{1,0}lat + a_{0,1}lon + a_{2,0}lat^2 + a_{1,1}lat \cdot lon + a_{0,2}lon^2 \\ &\quad + a_{3,0}lat^3 + a_{2,1}lat^2 \cdot lon + a_{1,2}lat \cdot lon^2 + a_{0,3}lon^3 + a_{n,0}lat^n \\ &\quad + a_{n-11}lat^{n-1} \cdot lon + \ldots + a_{1,n-1}lat \cdot lon^{n-1} + a_{0,n}lon^n. \end{aligned} \tag{1}$$

Higher order approximation promises good results for interpolation tasks, lower order polynomials might be better for extrapolation topics, because higher orders drift away beyond the borders.

Method of nearest neighbours

The method of nearest neighbours is based on the assumption that similar input combinations lead to similar output values. It is an instance based method. Thus, it directly uses all the instances of available data to calculate the estimated value. To optimise the models and to avoid overestimating single data points, it is possible to consider an average of 'k' immediate neighbours, for example, an average of 5 immediate neighbours. With respect to the total distance from i to k immediate neighbours, the output values can also be weighed. For example, the reciprocal value $w_i = 1/d_i$, for each single distance d_i of the previously sorted k immediate neighbours

$$\widehat{P} = \frac{\sum_{i=1}^{k} w_i \cdot P_i}{\sum_{i=1}^{k} w_i}. \tag{2}$$

For calculating the distance of the target instance to the k neighboured instances in this basic case the Euclidean distance is used:

$$d_i(lat_t, lon_t) = \sqrt{(lat_t - lat_i)^2 + (lon_t - lon_i)^2}. \tag{3}$$

Neural networks with radial basis functions

Radial basis functions (RBF) are a commonly used type of neural networks. It combines the advantages of neural networks in a way that it implicitly integrates the links that are not directly visible into the database on one hand and open structure of polynomial functions on the other hand. The radial basis toolbox was developed for the Software Matlab and was already used in different production engineering tasks [20,21]. In this method, a number of neurons will be equally distributed over each input dimension. If the network for example consists of 2 input dimensions with 3 radial basis function centre points the whole model exhibits $n = 2^3 = 9$ neurons. The approximated value can be calculated from the sum of all single neurons weighted by the factor a_i:

$$\widehat{P}(x) = \sum_{i=1}^{n} a_i \Phi_i. \tag{4}$$

The function value Φ_i of each radial basis function depends on the Euclidian distance to centre points x_0 by an exponential function with an RBF width σ:

$$\Phi_i(x) = e^{-\frac{(x-x_{0i})^2}{2\sigma_i}}. \quad (5)$$

The model parameters a_i can be calculated within the toolbox by using the least squares method. Compared to a classical multilayer perceptron neural network, that in optimal cases has a simpler structure and good approximation results, RBF are expected to increase the reliability and confirmability [22]. The different model types will be investigated by the leave-one-out method. Thus, for training the model, *n-1* data sets will be used and afterwards the calculated value of each model will be compared to the target value of the expulsed data set *n*.

3.1.2. Analysis of heuristic models

The methods introduced within Section 3.1.1 might have different properties for different data sets as well as for different regions (e. g. interpolation and extrapolation) or meteorological circumstances. Therefore, the methods will be analysed in this section based on their ability to predict output values for unknown regions by validating them. This will be done in the way that the model is set up by incorporating a reduced data set and is then tested by comparing the model results with the original value of the excluded data instance. By this procedure, an unwanted influence by the test value on the model is suspended. To compare the correlation between model output and the original value the Pearson correlation coefficient *p* is used. For comparability reasons all four data sets (wind speed, wind generation, solar radiation and solar generation) were reduced to the 52 zip code regions of the sparsest data set (wind power generation). For each time step of the data sets a model is generated and the model result will be compared to the original value. For the following results seven different model types have been investigated: polynomials of order 1 to order 4 (P1, P2, P3, P4), RBF model with 3 equally distributed RBF centre points and *k* nearest neighbours models with 3 (KNN3) and 5 (KNN5) neighbours.

Method comparison

The correlation coefficients for the four data sets are given in Fig. 4 distinguished by the used method and supplemented by the mean value of all methods. It can be seen that the solar radiation can be predicted with the highest probability and a mean correlation of 0.99. The solar generation follows with 0.97, while wind speed and wind generation exhibit lower correlation values of 0.90 and 0.84, respectively. The reason for this difference between wind and solar values can be found in the variances in the underlying weather phenomena. Solar radiation has smoother changes at typical weather situations compared to the wind speed which varies in shorter time frames. Following on that, neighboured postal regions have different values more often and prediction from neighboured values is more difficult.

Looking at the single methods, it can be derived that the methods differ in a small range within one data set. The difference between the best and the worst method is below 1%. Therefore, no clear advice can be given regarding which method should be used. Generally, for a modelling task with good correlation like the solar radiation the complex models with lower balancing behaviour lead to better results than the simpler ones. For example, the third order polynomial is the best polynomial approach and the KNN method with 3 neighbours is better than the one with 5 neighbours. Within the less correlated wind generation first or second order polynomials are better than the higher orders and it is possible to use more neighbours in the KNN approach. The RBF model shows a comparable performance as the polynomial approaches.

Geographic analysis

As the wind generation is the most interesting data set for the data filling problem it will be discussed in detail regarding geographic and temporal aspects in the following sections. To judge the methods based on their distinguishability the correlation factors of the 52 regions will be considered based on their standard deviation and uncertainty, which is decreasing with the sample size.

In Fig. 5 is shown that all tested simple heuristic methods have a good mean correlation coefficient above 80%. The methods KNN5 and P2 exhibit the highest values, but there is no significance for preferring a single method. Due to the shown error bars by applying the single standard deviation as confidence interval, none of the methods can be preferred against the other methods. Because of the always existing overlap between all methods a disjuncture is not given. On the right side of Fig. 5, the geographic distribution for the correlation coefficients of the wind generation is given. As expected, the regions with large distances to their neighbours and with border geographical position exhibit a lower correlation. However, in the upper right map for wind generation, there are some regions in the middle of Germany with low correlation values.

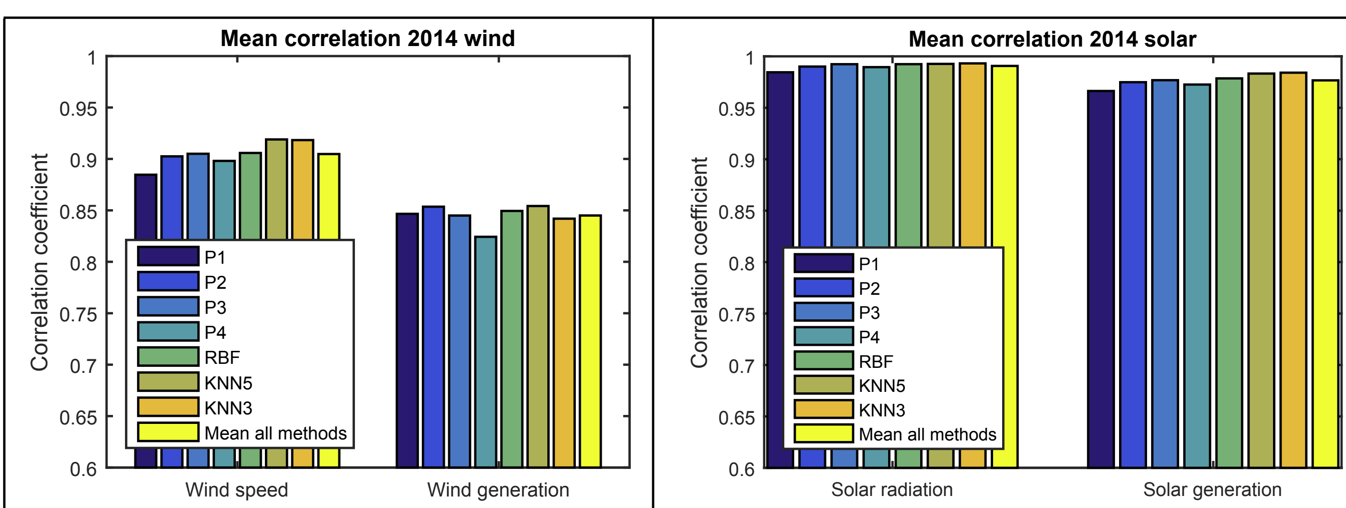

Fig. 4. Mean correlation by modelling method for wind speed and wind generation (left) as well as for solar radiation and solar generation (right).

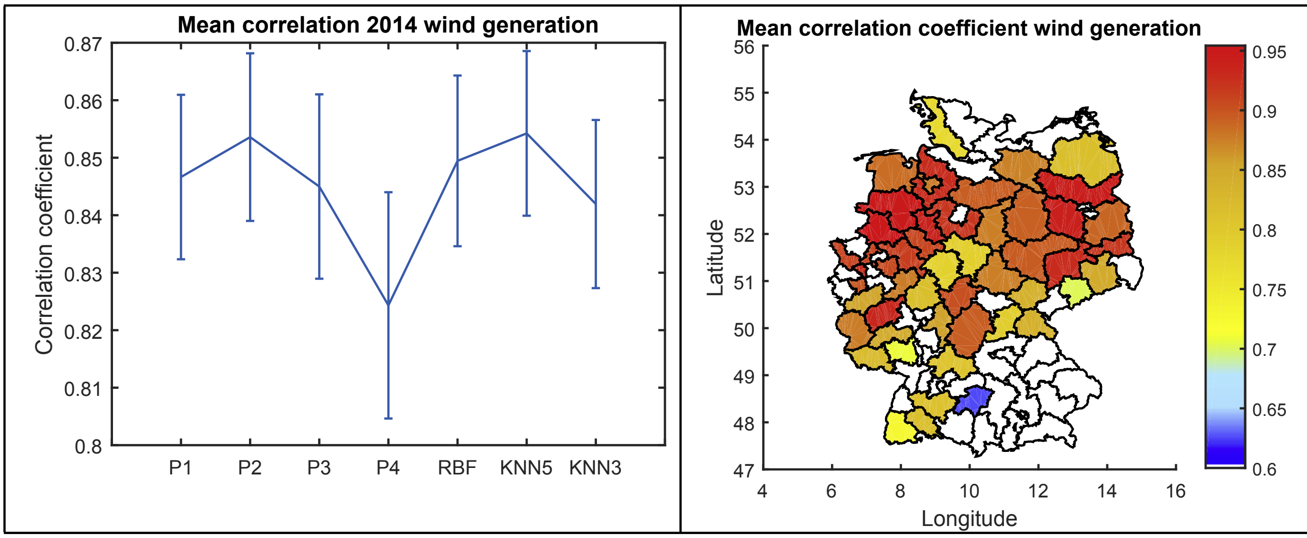

Fig. 5. Mean Correlation coefficients of the single regions by methods regarding mean value and uncertainty (left) and their geographical distribution (right).

An idea for filling the empty regions is to use a hybrid modelling approach. For that purpose, information of these regions without data is important. If there is a similar behaviour between the wind speed and the wind generation, the wind speed data set can be used to determine the best filling method. Therefore, the correlation of the regional correlations values as shown for wind generation in Fig. 5 (left) has been calculated also for the other data sets distributed by the same regions. This correlation of the data sets between wind speed and wind generation exhibits a value of 46%. It is larger than the correlation to solar radiation (25%) and to solar generation (23%), but the level is too low to see potential within this method. These low values lead to a need of optimising the model or adding extra data inputs, as the correlation between the 10 m wind and the power generation is too low. As the models are all applying similar results an integration of wind data in a height of 100 m or the combination with air density and air density differences in the observed regions is intended to generate better results.

Temporal analysis

To get a feeling about the temporal effects, the models are not only investigated for a full year, but also regarding smaller time intervals of months, weeks and days of the year. As a result, coefficients decrease with decreasing vector size during the calculation of the correlation coefficients.

In Fig. 6 the temporal effect is given. It can be noticed on the left side that a high mean correlation of above 80% is given for intervals larger than one month. The correlation decreases to 60% when only one day is considered. On the right side of Fig. 6 the weekly development of the correlation coefficients over the year can be seen. It can be determined that fluctuations between 58% and 92% appear similar over all methods. As a result, there is an obvious variation of the correlation coefficients over time. The mean correlation decreases at smaller time intervals. To reach a correlation of above 80%, time intervals of one month or longer are needed.

3.2. Wind speed models

Besides using the shown simple heuristic models, the ability of filling the data set based on wind speed data will be tested and compared. For that matter, a function will be defined from the data and the relation from relative wind speed to wind generation is applied.

3.2.1. Definition of wind speed models

Apart from using only the available neighboured data values of the aiming wind generation power, it seems promising to calculate the wind generation from the wind speed. Two wind speed data sources are available: firstly the data of the DWD as already shown in Fig. 3 and secondly the wind speed data of the GEOWISOL plants, but only for the 52 zip code with generation data. Both data sets have individual advantages. The DWD data set is available for all 95 zip code regions, but it is measured at a height of 10 m where large influences from the surrounding topographical roughness exist, leading to a limited correlation of the wind speed at the generation height of the wind power turbines. The GEOWISOL wind data is measured directly at the height of the turbines and is therefore suitable for the wind power prediction by the plants. But it has the limit that the data is only available for 52 regions. The relation between wind speed and wind power generation is given by a cubic function as shown in the following equation [13]:

$$P \sim v^3. \tag{6}$$

Therefore, a wind speed model predicting the generated power P from the speed v will follow a cubic function for small values. For higher wind speeds, wind power plants have to limit the generated energy regarding their power specifications, thus a saturation effect will be expected.

Furthermore there are limiting conditions that lead to a smoothing decrease when the nominal power at a nominal wind speed is reached. Above the nominal wind speed, the power output remains constant (cut-off wind speed). The power output drops to zero at very high wind speeds above cut-off wind speed, because of plant protection reasons. It is aimed to define a function for this relation that can be used afterwards to calculate the wind power P from the wind speed v base on both data sets, DWD and GEOWISOL. As function for wind turbine power curves, the logistic function is a valuable heuristic approximation [23].

$$P(v) = 100 \cdot \frac{1 + m \cdot e^{-v/\tau}}{1 + n \cdot e^{-v/\tau}}. \tag{7}$$

It can be approximated in form of the given equation with three parameters m, n and τ; here the parameter m (e.g. 0.001) determines the ordinate axis start value of the curve, parameter n (e.g. 100) determines the form of the curve and the turning point on the

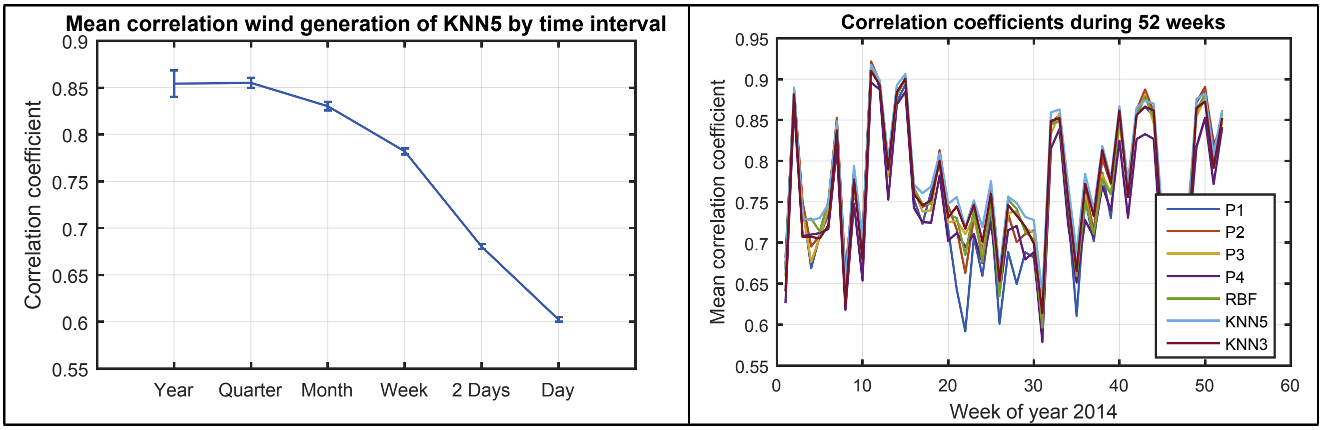

Fig. 6. Mean Correlation coefficients of model KNN5 by considered time interval (left) and the weekly temporal development of the correlation coefficients of the different methods (right). Geographical results.

abscissa axis, finally parameter τ (e. g. 2) gives the relation to the input parameter range of maximal 25 m/s. The results can be compared with the existing wind power values of the considered regions and a correlation factor will be calculated.

3.2.2. *Analysis of wind speed models*

The two data sets of the 52 considered regions are compared. In Fig. 7 on the left side, the wind power is shown dependent on the DWD wind speed. The graph shows the expected form of a cubic increase at low and middle wind speeds and a damping effect towards the nominal wind power for higher wind speeds. Especially for middle wind speeds, there are large deviations for the wind power that can be motivated by different data source locations and heights. On the right side of Fig. 7 the relation to the GEOWISOL wind speed is given. It can be seen that the fit is much better to the wind power with a correlation of 95% compared to 75% for the DWD data. Therefore, it can be stated that the available DWD data is not well suitable for modelling of the missing wind energy generation. That means, the DWD data should not be used for the data filling, instead, data of neighboured regions should be preferred. But there is a high correlation of the GEOWISOL wind speed data, which are gained directly from the power plants, to the wind generation. These high correlation coefficients offer predictability options that are much better than using only the data of neighboured regions. Unfortunately, the GEOWISOL wind data is not available for the unknown regions, therefore this really high correlation cannot be used for the data filling, but it shows the potentials that can be exploited when better wind speed data could be acquired for the investigation.

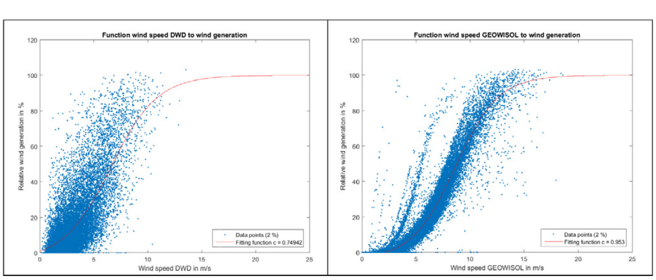

Fig. 7. The wind power generation within the 52 zip code regions in relation from the DWD data (left) and from GEOWISOL plant data (right). Note that in sake of visibility the amount of data points is reduced to 2% randomly.

3.3. *Combination*

Finally, the combination of both approaches, using the data of neighboured regions and the wind generation function shall be investigated. The combination of the calculated values from the wind speed models and the heuristic models will be validated. As an example, the results from method KNN5 will be used in the combining method. Therefore, the mean value of both results from the models will be calculated as shown in Fig. 8 on the left side. The correlation of the resulting time depending vector of wind generation values to the available original values is used to assess the performance of this method. From the correlation values in Fig. 8 on the right side, it can be noticed that the combination leads to a similar correlation than the better single method. The combination of KNN5 wind generation data and GEOWISOL wind data leads to a correlation of 94.3% compared to 95.3% of the single correlation of GEOWISOL wind data. The combination of KNN5 and DWD wind data gets a correlation of 83.8% as against 83.2% of single KNN5 method. As a result, the combination of neighboured data and meteorological data does not lead to an increase of the correlation. A moderate increase of the correlation might be possible when using weight factors instead of using only the mean value.

4. Results

As all methods lead to similar correlation coefficients the method KNN5 is used for the following conclusions, exemplary. It does not matter which method is used for calculating the relative wind generation of the unknown regions with respect to the uncertainties of the values.

After the application of KNN5 method for the regions without data, from the relative wind generation data the absolute wind generation will be calculated. Therefore, for each region, the 5 nearest data filled neighbours are chosen and a weighted mean value of the relative wind generation is calculated. From the result the relative value will be multiplied by the total region installation, leading to the absolute wind generation power. In Fig. 9, the concluded wind energy distribution is shown on the left side. Resulting from that distribution also the residual power $P_{residual}$ as difference between the load P_{demand} and the sum of the fluctuation input powers of wind P_{wind} and solar P_{solar} can be calculated: $P_{residual} = P_{demand} - P_{wind} - P_{solar}$.

The residual power is shown on the right side of Fig. 9. In the northern and eastern parts of Germany are the regions with the lowest values and in the westerns part is the largest demand to see.

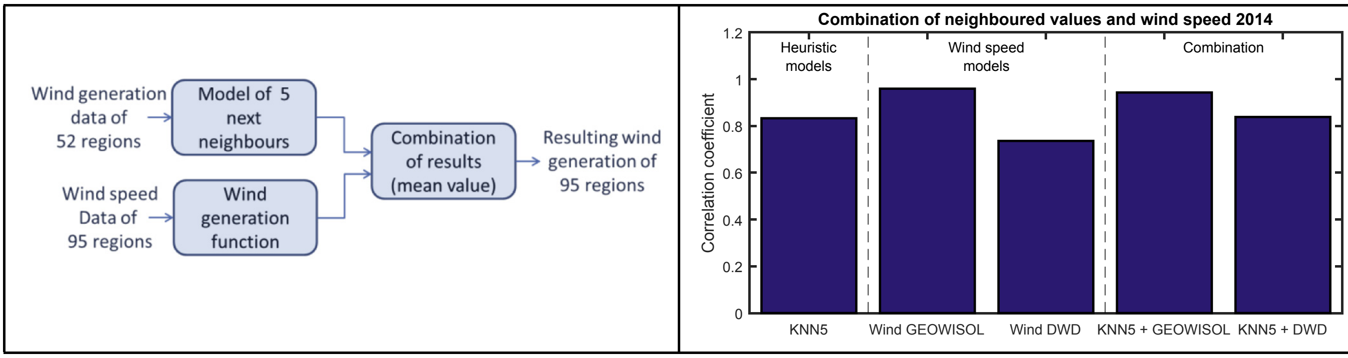

Fig. 8. The strategy of combining neighboured wind generation data and wind speed data (left) and calculated correlation values of single and combined methods (right).

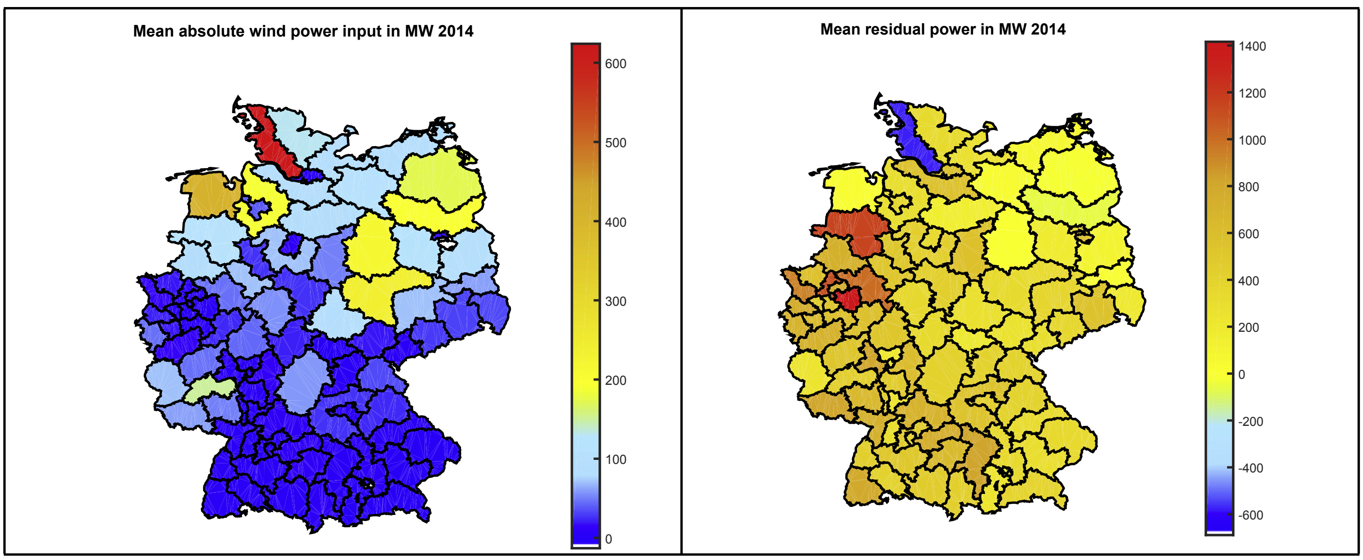

Fig. 9. The mean wind power generation with filled gaps (left) and resulting residual power for all 95 zip code regions (right).

Now, the resulting database can be used to investigate a number of questions regarding the power system infrastructure and optimisation problems.

4.1. Definition of use cases from residual power flow

The residual power of all postal regions defines the surplus and sub-offerings of the energy amounts within the regions. The optimal case would be when the needed energy is produced in the same region. But, mostly the generation and demand do not fit. Therefore, a transportation model is generated, in order to determine the optimal transfer directions and ratios. The optimal power flow is defined as a transportation problem that minimises the transportation costs regarding the distances between the source and the sink as well as the amounts. The transportation problem was introduced by Kantorovich [24] and Hitchcock [25] and will be used for the use case analysis. It is based on a linear optimisation method and the minimisation will be solved by the simplex algorithm [26]. As prerequisite, it has to be given that the sum of source power fits to the sink power at every time step. Therefore, at each time step the generation amount will be increased with a constant expansion factor to equal the sums. In the use cases the power flows as mean values for the whole year 2014 as well as for the sunniest day and for the windiest day will be discussed.

4.2. Results of power flow analysis

The power flows resulting from the transportation optimisation lead to the transfer needs of electrical power between the different German regions. Therefore, the capacities of an optimal distribution of the electrical energy between different zip code regions can be discussed. Especially extraordinary weather situations define the needs onto the power system. In Fig. 10, calculations of the optimal power flows between 10 German zip code regions are shown. In sake of visibility 2 digit zip code regions are combined to regions by the first digit zip code numbers. The input power of each region has been increased in a way that the level of generation equals the entire demand for Germany. Therefore, the investigation is a scenario for a time when the energy demand is completely covered by renewable energy. After the optimisation step, the models calculated resulting optimal power flows, leading to minimal costs. On the left side of Fig. 10 the resulting mean power flows of the year 2014 between the regions are shown. It is evident that the optimal power flow leads from the north-eastern regions of Germany to the south-western regions.

The middle illustration of Fig. 10 shows the day with the highest solar input within 2014 and the right illustration the day with the highest wind input. It can be detected that these two situations exhibit completely different requirements to the power system. A solar power dominated day leads to energy transport need from the

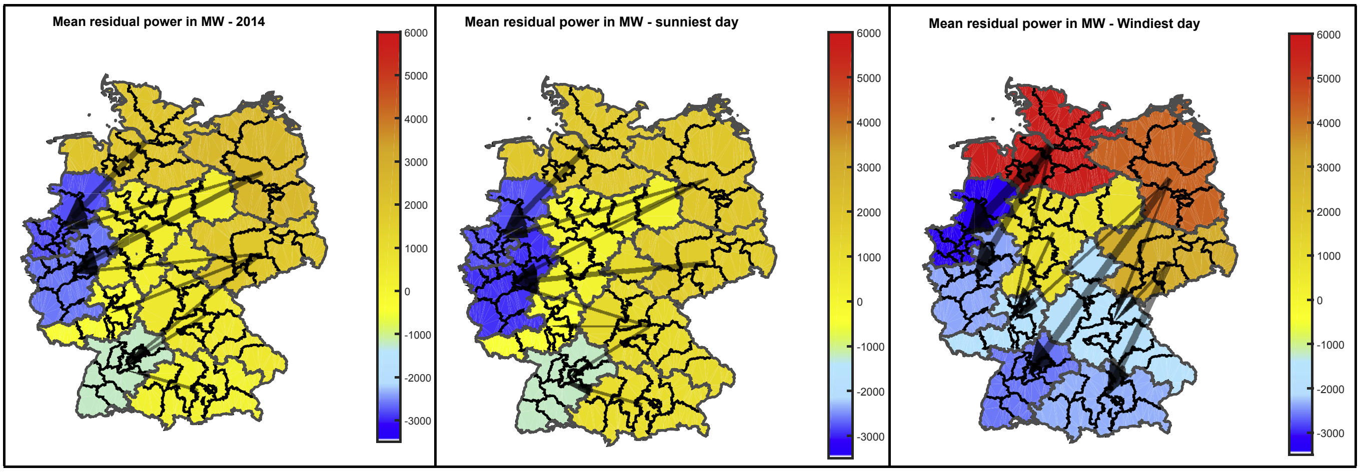

Fig. 10. Residual loads with optimal power flow regarding the mean of 2014 (left), the day with highest solar input (middle) and the day with highest wind input (right).

eastern to western regions, while a wind dominated day leads to transport needs from the northern to the southern regions. That means in different weather situations different infrastructure is needed. On windy days large transfer capacities from north and north-east to the west are needed, while on sunny days these capacities are used moderate, but larger transfer needs from east and south-east to the west are required. A further necessity on infrastructure with storage function or a demand side management function can also be conducted from the use case investigation. While large transfer needs appear, these needs can also be mitigated by a growing usage of storage systems and rescheduling of power demand in the regions of the largest sources as well as of the largest sinks. A distinct heterogeneity between the regions is recognisable. The capacities of the power line from east to west need to be enlarged for sunny days and from north to south for windy days, when the expansion of the renewable energy generation progresses.

5. Summary

This work shows the distribution of power generation and demand of the German electrical power system. The integration of the fluctuating wind and solar energy input is given by temporal and spatial distributed measurement values. The application of real measurement values in that high resolution goes beyond existing studies, which are basing on simulations or possess a limited spatial resolution. Using the distribution to 95 zip code regions, more effects can be considered concerning different behaviours of the different geographical regions. A number of modelling methods were investigated regarding their suitability in filling gaps of regions with absent data. All considered methods lead to a correlation coefficient of about 80%, but a single method cannot be identified that is significantly more accurate in all cases than the others.

Furthermore, the inclusion of wind data into the modelling method is discussed. Using wind speed measurement at the height of the wind energy turbines lead to a high correlation of above 95% but cannot be used for the data filling algorithm. Using data of DWD measurement stations near the ground do not lead to better results. Even in combination with the wind generation data of the next neighbours, it does not lead to a significant higher correlation than the simple modelling approaches.

With the filled data set, investigations can be conducted regarding the resulting energy flow between the regions within the use cases. It can be detected that situation with dominating wind or solar input lead to an obviously different power flow. For stabilisation of the power system it is intended to bring the gained results to the transmission system operator and the distribution system operator level, in order to make the data available for system management.

Future work will deal with the optimisation of data filling procedures and with the analysis of the data set regarding to the open questions of the expansion process of renewable energy. One aim is to carry out an uncertainty analysis and to enlarge the correlation by acquiring additional data and replenish the models with more suitable meteorological parameters. Furthermore the combination with additional parameters (e. g. air temperature or air density) promises a reduction of modelling errors. Finally, the power flow shall be aligned with the real power flow considering the existing structure of the energy system.

Acknowledgement

The project “Effects of the GEOgraphical distribution and temporal correlation of WInd and SOLar input on the power supply system (GEOWISOL)” has been set up to solve this topic. The work leading to these results has received funding from the BMWi by reference number 0325695.

References

[1] BMWi, Erneuerbare Energien in Deutschland - Daten zur Entwicklung im Jahr 2015, Bundesministerium für Bildung und Forschung, Berlin, 2016.
[2] Reiner Lemoine Institut, Vergleich und Optimierung von zentral und dezentral orientierten Ausbaupfaden zu einer Stromversorgung aus erneuerbaren Energien in Deutschland, 2013 (Berlin).
[3] K. Rohrig, Windenergie Report Deutschland 2014, Fraunhofer IWES, Kassel: Fraunhofer Verlag, 2015.
[4] EEX Transparency Platform, Solar & Wind Power Production, 2017 [cited 2017 18.05.2017]; Available from: https://www.eex-transparency.com/homepage/power/germany/production/usage/solar-wind-power-production.
[5] T. Tenne, Network Figures Overview, 2017 [cited 2017 18.05.2017]; Available from: http://www.tennettso.de/site/en/Transparency/publications/network-figures/overview.
[6] Amprion, Grid Data, 2017 [cited 2017 18.05.2017]; Available from: http://amprion.net/en/grid-data.
[7] 50Hertz, Grid Data, 2017 [cited 2017 18.05.2017]; Available from: http://www.50hertz.com/en/Grid-Data.
[8] TransnetBW, Key Figures, 2017 [cited 2017 18.05.2017]; Available from: https://www.transnetbw.com/en/transparency/market-data/key-figures.
[9] A. Steiner, C. Köhler, I. Metzinger, A. Braun, M. Zirkelbach, D. Ernst, P. Tran, B. Ritter, Critical weather situations for renewable energies—Part A: cyclone detection for wind power, Renew. Energy 101 (2017) 41—50.
[10] A.E. Saleh, M.S. Moustafa, K.M. Abo-Al-Ez, A.A. Abdullah, A hybrid neuro-fuzzy power prediction system for wind energy generation, Int. J. Electr. Power &

Energy Syst. 74 (2016) 384–395.
[11] J. Olauson, M. Bergkvist, Modelling the Swedish wind power production using MERRA reanalysis data, Renew. Energy 76 (2015) 717–725.
[12] R. Jursa, K. Rohrig, Short-term wind power forecasting using evolutionary algorithms for the automated specification of artificial intelligence models, Int. J. Forecast. 24 (4) (2008) 694–709.
[13] V. Quaschning, Regenerative Energiesysteme: Technologie-berechnung-simulation, Carl Hanser Verlag GmbH Co KG, 2015.
[14] J. Widén, N. Carpman, V. Castellucci, D. Lingfors, J. Olauson, F. Remouit, M. Bergkvist, M. Grabbe, R. Waters, Variability assessment and forecasting of renewables: a review for solar, wind, wave and tidal resources, Renew. Sustain. Energy Rev. 44 (2015) 356–375.
[15] G. Reikard, B. Robertson, J.-R. Bidlot, Combining wave energy with wind and solar: short-term forecasting, Renew. Energy 81 (2015) 442–456.
[16] Z. Zheng, Y. Chen, M. Huo, B. Zhao, An overview: the development of prediction technology of wind and photovoltaic power generation, Energy Procedia 12 (2011) 601–608.
[17] J. Schmelter, U. Focken, Operationelle Erfahrungen mit kombinierten Solarleistungsvorhersagen für deutsche ÜNBs und VNBs, 26, in: Symposium Photovoltaische Solarenergie, Bad Staffelstein, March. 2011.
[18] SMA, Performance of Photovoltaics (PV) in Germany, 2017 [cited 2017 18.05.2017]; Available from: http://www.sma.de/en/company/pv-electricity-produced-in-germany.html.
[19] DWD, CDC (Climate Data Center), 2017 [cited 2017 18.05.2017]; Available from: http://www.dwd.de/EN/climate_environment/cdc/cdc_node.html.
[20] C. Ament, in: Direkte Regelung der Werkstückqualität in der Fertigung - Konzepte für ein neues Paradigma, Universität Bremen, Fachbereich Produktionstechnik, 2005.
[21] P. Zhang, A. von Freyberg, A closed loop quality control system for laser chemical machining, in: 16th International Conference & Exhibition of the European Society for Precision Engineering and Nanotechnology (Euspen), 2016, pp. 505–506.
[22] M. Dijkman, Automated compensation of distortion in the production process of bearing rings, in: Universität Bremen, Fachbereich Produktionstechnik, 2009.
[23] A. Kusiak, H. Zheng, Z. Song, On-line monitoring of power curves, Renew. Energy 34 (6) (2009) 1487–1493.
[24] L. Kantarovich, Mathematical Methods in the Organization and Planning of Production, Publication House of the Leningrad State University, 1939 [Translated in Management Sc., vol. **66**, 366–422].
[25] F.L. Hitchcock, The distribution of a product from several sources to numerous localities, Stud. Appl. Math. 20 (1–4) (1941) 224–230.
[26] S.S. Rao, Linear programming I: simplex method, in: Engineering Optimization: Theory and Practice, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2009.