



**HAL**  
open science

# A probabilistic approach to the estimation of regional photovoltaic power production

Yves-Marie Saint-Drenan, Garrett H. Good, Martin Braun

► **To cite this version:**

Yves-Marie Saint-Drenan, Garrett H. Good, Martin Braun. A probabilistic approach to the estimation of regional photovoltaic power production. *Solar Energy*, 2017, 147, pp.257-276. 10.1016/j.solener.2017.03.007 . hal-01506551

**HAL Id: hal-01506551**

**<https://minesparis-psl.hal.science/hal-01506551>**

Submitted on 27 Sep 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A Probabilistic Approach to the Estimation of Regional Photovoltaic Power Production

Y.M. Saint-Drenan<sup>1,2,\*</sup>, G. H. Good<sup>1</sup>, M. Braun<sup>1,3</sup>

<sup>1</sup>*Fraunhofer Institute for Wind Energy and Energy System Technology, Koenigstor 59, 34119 Kassel, Germany*

<sup>2</sup>*MINES ParisTech, PSL Research University, O.I.E. – Centre Observation, Impacts, Energy, CS 10207 – 06904 Sophia Antipolis CEDEX, France*

<sup>3</sup>*University of Kassel, Wilhelmshöher Allee 71 – 73, 34121 Kassel, Germany*

\* Corresponding author. Tel.: +49 561 7294 246; fax: +49 561 7294 260 E-mail address: yves-marie.saint-drenan@iwes.fraunhofer.de (Y.M. Saint-Drenan)

---

## Abstract

Forecasting the total photovoltaic (PV) power generated in the control areas of the transmission system operators (TSO) is an important step in the integration of the large amounts of PV energy into the German electricity supply system. A standard approach for evaluating the regional PV power generation from weather forecast consists in upscaling the forecast of a limited set of reference plants to the complete area. Previous studies shown that this method can lead to large errors when the set of reference plants has different characteristics or weather conditions than the set of unknown plants. In this paper, an alternative to the upscaling approach is proposed. In this method, called a probabilistic regional PV model, an average PV model with a very limited number of inputs (two module orientation angles) is used to calculate the power generation of the most frequent module orientation angles. The resulting power values are finally weighted according to their probability of occurrence to estimate the actual power generation. The implementation of this model thus only requires information on the location and peak capacity of the plant installed in a region and no PV plant measurement is necessary. The proposed method has been evaluated against the estimate of the total power generation provided by the German TSOs, which shows that an RMSE ranging from 4.2 to 4.9% can be obtained with this method using on IFS meteorological forecast. The regional power forecasted with the probabilistic approach was also compared to the day-ahead forecast disseminated by the TSO. This analysis shows that the forecast evaluated with the proposed approach has an RMSE less than 0.5% higher than the reference forecasts. This is considered a promising result given that the forecast evaluated with the probabilistic model is based on one single weather model and that – at the exception of the model calibration – no statistical post-processing method is used to optimize its performance.

---

## 1. Introduction

Over the last years, Germany has witnessed a rapid development of its photovoltaic (PV) capacity. In the beginning of 2016, more than 1.5 million PV plants with a total installed capacity of 36.66 GW<sub>p</sub> were connected to the German grid. With Germany having an energy demand varying between 35 and 75 GW [1] and a minute reserve power (tertiary

control reserve) of -5.5 and +7 GW [2], the accurate consideration of the PV power generation is thus crucial for the secure and economical operation of the power system.

For the integration of the large amount of PV power into the electricity supply system, the transmission system operators must assess and forecast the total PV power generated in their control areas. Since only a limited number of power measurements are available, an exact determination of the total power generated by the PV plants installed in a region is not possible. The actual value of the power generated by the numerous PV plants installed in a control area must instead be estimated using the limited information available for each PV plant (peak capacity, location, time of installation). This estimate, which is referred to in this paper as an estimate of the regional PV power generation, is very important, as it is used for the grid monitoring and as a basis for the PV power forecast.

The upscaling algorithm is currently the standard approach in Germany for estimating the regional PV power generation. The principle of the upscaling method is to assess the normalized power generation of any PV plant installed in a region by a spatial interpolation of the normalized power of a set of reference plants. This approach has been described by e.g. *Lorenz et al.* [3] and *Schierenbeck et al.* [4]. *Saint-Drenan et al.* conducted a detailed analysis of the sources of error of the upscaling method in [5], which shows that the uncertainty of this method results mainly from two issues:

- The spotty acquisition of the irradiation field by a limited number of point measurements, and,
- Differences between the characteristics of the reference and uncharacterized PV plants.

A methodology was proposed by *Shaker et al.* to identify the most adapted set of reference plants by a mixture of k-means clustering and PCA and to derive an estimate of the total PV power generation based on measurements from this set of reference plants using linear regression [27] or fuzzy-logic operators [28]. A prerequisite for the implementation of this approach is however that power measurements of all plants installed in the considered region are available over a training time period of at least 4 months. This requirement makes the implementation of this method difficult when the availability of PV power measurements is limited.

An alternative approach to the upscaling method and its modified versions [27], [28] with the aim of mitigating the two issues mentioned above without the need for PV power measurements is proposed in [6]. This method, called probabilistic approach, is based on two main ideas. Firstly, given that meteorological data is available region-wide (e.g. from NWP model or satellite derived irradiation), the irradiation and temperature of each PV plant installed in the considered region can be considered explicitly instead of interpolating data from a set of reference plants. Secondly, values derived from the statistical analysis of numerous PV plants are used as the parameters of a PV plant model, rather than considering only the parameters of the set of reference plants. The probabilistic approach thus avoids the first issue of the upscaling method by efficiently using the meteorological data available for the power estimation of each plant installed in the considered region. The problem of representativeness of the set of reference plants is also mitigated by the probabilistic approach by using statistical values for the model parameters.

A first validation of this approach is proposed in [6], where the results of the upscaling and probabilistic approaches are compared for several hundreds of PV plants located in Southern Germany. This analysis shows that the performance of the probabilistic approach is slightly better than that of the upscaling approach when the number of reference plants is high, and that the probabilistic approach outperforms the upscaling method when the number of reference plants is low. It however fails to clarify to what extent the results can be generalized for different regions, and especially for the control areas of the transmission system operators (TSOs), for which regional PV forecast are needed by grid operators. One major goal of this paper is thus to evaluate and analyse the performances of the probabilistic approach at TSO level, which is highly relevant for the grid integration of PV electricity.

In contrast to the work presented in [6], where power measurements of several hundreds of PV plant has been used to validate the probabilistic approach, assessing the performances of this method on control areas raises the problem of choosing the true value of the total power generated. Indeed, as previously mentioned, only a limited subset of the plants is measured and the access to these measurements is very difficult for most stakeholders. As a result, it is impossible to get the actual value of the aggregated power generation on that regional scale and the only possibility is to use a best guess, which has a given uncertainty. It was therefore decided to assess the performances of the proposed approach against TSO estimates of the PV power generation, which can be considered as the best estimate of this regional power generation. These values are evaluated by upscaling power measurements from numerous PV power plants installed in their control area.

The use of TSO estimates raises a number of questions related to the relevance of the validation presented in this paper. Since the reference power values are estimated by an upscaling method, it is questionable to which extent it is relevant to propose an alternative approach to the upscaling? To answer this first question, it is important to note that the power measurements used by the German TSOs are - to date - not available to any forecast provider. It is thus not possible to use the same information than the TSO for forecasting purposes. A forecast provider has thus to develop its forecast approach independently from the TSO by e.g. collecting its own power measurements. The access to power measurements being in Germany extremely difficult, this prerequisite can noticeably limit the number of companies being able to propose a forecast product. The methodology proposed in this paper being implementable without power measurements, its validation may be highly relevant to stakeholders interested by PV power forecast but without access to measurements. A second question may be what is the value of the validation proposed in this paper since the actual aggregated power generation is not known and an estimate is used as true value? In a rigorous scientific context, the relevance of a comparison of the output of the proposed method with the TSO estimates is unclear since the actual value is unknown and both estimates have a given uncertainty. However, the chosen validation reflects the situation faced by forecast providers during the evaluation of their model output: the performances of a PV power prediction are assessed by TSO using their estimates. Though the validation proposed in this paper is of limited relevance from a scientific perspective, it is nevertheless pertinent in an industrial context.

This paper is structured in three main parts. The first part describes the motivations and the principle of the probabilistic approach already introduced in [6] (section 2). The difference between the TSO estimates and the output of the probabilistic approach are analysed in the second part (section 3). The results are finally summarized and discussed in the concluding part of this paper (section 4).

## 2 Approach

### 2.1 Motivation

The motivation for the development of the probabilistic approach is to design a method mitigating the two problems identified in the upscaling method: The uncertainty resulting from the acquisition of the meteorological information by a coarse network of reference plants and the lack of representativeness of the set of reference plants.

The mitigation of these two problems by the probabilistic approach is based on two main ideas, (a) the efficient use of available meteorological data and (b) the use of statistical information on the PV plant parameter for the calculation of the PV power.

#### a) Efficient use of available meteorological data

As previously mentioned, one issue with the upscaling method is that the meteorological fields are only assessed by a limited number of points (set of reference PV plants). This approach is justified when an estimate is made from power measurements but when a NWP-based power forecast is made this approach is suboptimal since meteorological information is available region-wide and thus for each single plant installed in the considered region. Based on this consideration, the first idea behind the probabilistic approach consists of using all weather information available instead of only that at the locations of the reference PV plants.

#### b) Use of statistical information on PV plant parameters for the calculation of PV power

The second issue of the upscaling method is that an error may occur when non-representative PV plants are used as reference plants. *Lorenz and Heinemann* addressed this issue in [3]: “*In order to determine a suitable set of representative systems, in a first step the basic characteristics of the complete set of systems in an area have to be described*”. The recommendation of *Lorenz and Heinemann* is then to select a subset of representative reference plants among a larger set of plants using the characteristics of the complete set. This approach is problematic when the set of plants with available measurements is limited. Alternatively, if statistical information on the different parameters of a PV plant is available, it may be advantageous to directly use it for the calculation of the PV power. The second idea pursued in the development of the probabilistic approach is therefore to use statistical data on the parameters of PV plants for the calculation of the PV power, which should guarantee that the parameter used well describes the complete set of systems in an area.

A preliminary step to the implementation of the two ideas described above is the selection of a PV plant model. Here, the choice of the model has to compromise between a high accuracy and a limited number of parameters. The motivations for the choice of

the model as well as a description of the different parts of the model are given in section 2.2.

To implement the probabilistic approach, a statistical analysis of the model parameter is then necessary. At this step, one must take care that the data assessed during the statistical analysis can be used for simulating a PV plant. In particular, the nonlinearity of the model with the input parameter is taken into consideration. For example, the PV power simulated with a southward orientation is not equal to the mean of the PV power simulated with an eastward and westward orientation. To cope with this nonlinearity, the approach chosen is a statistical analysis to identify the parameter sets  $(A_i)_{i=1\dots n}$  that occur frequently, as well as their probability  $(w_i)_{i=1\dots n}$ . With this information, the power production of a PV plant can be estimated according to the following equation:

$$P_{PV}(x, t) = P_{peak} \sum_{i=1}^n w_i f_{PV}(x, t, G_{hor}(x, t), T_{air}(x, t), A_i) \quad (1)$$

Where

$P_{pv}(x,t)$	is an estimate of the power produced at time t by a PV plant located at x [kW]
$P_{peak}$	is the peak capacity of the PV plant [kW <sub>p</sub> ]
$G_{hor}(x,t)$	is the global horizontal irradiance at time t and location x [W/m <sup>2</sup> ]
$T_{air}(x,t)$	is the air temperature at time t and location x [°C]
$f_{PV}(\dots)$	represents the PV model used to calculate the normalized PV power [kW/kW <sub>p</sub> ]
$A_i$	represents the set of plant parameters needed by the PV model
$w_i$	is the probability of occurrence of a parameter set $A_i$

The choice of the term “probabilistic approach” for the method described in this section has been motivated by Eq. 1. Though a single (deterministic) power value is calculated for each plant, the proposed methodology is very similar to a probabilistic approach since all possible configurations are considered separately, resulting in an ensemble of estimates of the PV power production. A deterministic value is then evaluated from this ensemble by a weighted average of the ensemble members, the weights being the probabilities of occurrences of each scenario, which is evaluated in the statistical analysis.

The goal of the statistical analysis is therefore to identify sets of PV plant parameters  $A_i$  and their probability of occurrence  $w_i$ . The database used for this purpose as well as for the statistical analysis of each model parameter is presented in section 2.3. A detailed description of the calculation steps of the probabilistic model is finally given in section 2.4.

## 2.2 PV plant model selection

The first step needed for the implementation of the regional probabilistic PV model is the selection of a PV plant model. A model giving the output PV power generation as a function of the time, location, global horizontal irradiation and air temperature is needed. With the PV model being the cornerstone of the probabilistic approach, its accuracy directly impacts the performance of the algorithm. Under this perspective, the most accurate model should be preferred. Nevertheless, the choice of an accurate but too detailed model would imply the need to make a statistical analysis of some

parameters that are difficult to obtain. In this case, the situation can occur where a detailed model used with statistical information of low quality results in a lower quality than would a simpler model. With regards to this, the choice of a robust model with a limited number of parameters should be preferred. A compromise is thus required between minimizing the amount of information on the PV plant needed by the model and maximizing the model accuracy.

The model selection is based on the work conducted by *Saint-Drenan* [6], where a sensitivity analysis revealed that the model parameters to which the PV power is the most sensitive are the module orientation angles and the power curve. In this previous work, a statistical analysis of power curves of several hundreds of PV plants also showed that large differences exist among single PV plants but that the difference between average power curves evaluated for difference classes of peak power is very small. Accordingly, we assume that a constant power curve can be used when the aggregated power of a large number of PV plants is evaluated. Based on this finding, the model used in [6] has been simplified. The different calculation steps of the simplified model used in the present analysis are described below.

With local shading being neglected, the plane of array (POA) irradiation can be estimated from the global irradiation and the sun position using a set of models commonly used for this purpose [7][8]. Here, the separation and transposition models proposed by *Skartveit et al.* [9] and *Perez et al.* [10] are each respectively used for estimating the plane of array irradiation from the global horizontal irradiation. The ground albedo is assumed to be constant and equal to 20%. The module azimuth and tilt angle are the two PV plant information required for this first step.

To estimate the POA irradiation effectively contributing to the photovoltaic effect (effective irradiation), optical losses are calculated using the formulation of *Martin and Ruiz* [11], where the fitting coefficients are assumed constant and fixed to the set of values representative for crystalline modules provided by the authors.

The module efficiency is evaluated using the model proposed by *Beyer et al.* in [12], which requires four parameters. Three parameters describe the dependency of the module efficiency with the POA irradiation for a module temperature equal to 25°C. The fourth parameter quantifies the temperature dependency of the module efficiency (Ross parameters), which is assumed to be linear [13]. The reference values of the *Beyer et al* model proposed by *Lorenz et al* in [14] are used for the simulation of the module efficiency (Figure 1 left) and a constant value equal to 0.030 °Cm<sup>2</sup>/W is assumed for the parameter of the Ross model. Based on the sensitivity analysis conducted in [6], the ohmic losses occurring in the DC cables are neglected. Further loss factors taking place in the DC part of the PV plant such as module mismatch, soiling, and shading are also neglected and it can therefore be assumed that the specific power calculated with the module model is equal to that fed into the inverter.

The efficiency of the inverter is evaluated using the model proposed by *Schmidt et Sauer* [15]. In this model, the inverter losses are assumed to be the sum of three loss terms (self consumption, voltage drop and ohmic losses), which are quantified by three parameters. The database of PV inverter provided by *Photon* [16] is used for the determination of the standard inverter parameters. Here, the parameters of the *Schmidt et Sauer* model corresponding to the mean power curve of inverters with an euro-

efficiency comprised between 75 and 90% is used as reference value for the PV inverter (Figure 1 right). Finally, the inverter model needs to consider the sizing ratio, which is equal to the rated power of the inverter divided by the module peak power. A constant sizing ratio of 0.85 is assumed.

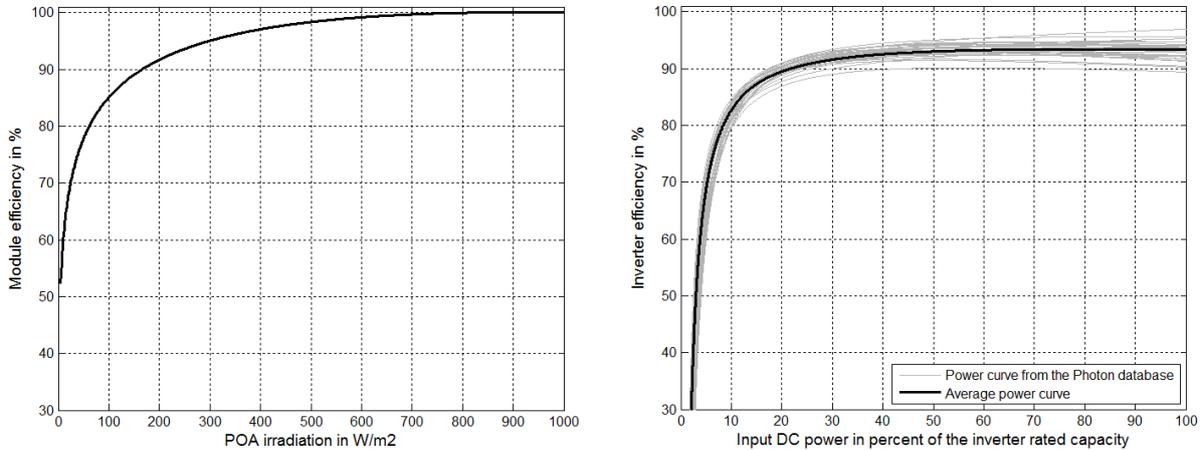


Figure 1: Left: reference PV module efficiency from [14]. Right: reference PV inverter efficiency evaluated from the Photon inverter database [16].

With the selected model, besides from the coordinates of the plant, the time, the global horizontal irradiance and air temperature which can be considered to be known (see section 1), the two parameters of the PV model are the module azimuth and tilt angles. As statistic analysis of these two angles is thus conducted in the following section to implement the regional model as defined in equation (1).

With the different loss factors being neglected, it is clear that the selected PV plant model reflects the power generation of a PV plant in optimal conditions. It can thus be expected that power values simulated with the model described previously over-estimate the actual power generation of many plants. This problem is addressed by calibrating the result of the regional model to the targeted power values. The constant derating factor fitted in this calibration accounts for the aggregated effect of the different neglected losses as well as for the difference between the performance ratios obtained with the model and those of the plants installed in the consideration [17].

It is also assumed here that a PV plant has a single fixed orientation. Special cases, where a PV plant is made of module on different orientations or where the modules are mounted on a solar tracker are thus not considered. It is assumed that the disregard of these special cases is not problematic for the application addressed in this work, as the large majority of plants have – or at least can be simulated with - a single fixed orientation. Finally, we have not considered the effect of inverter clipping on the power generation. Though this effect may have significant effect on single plants, we expect that this effect is smoothed by the aggregation of power generation of numerous plants.

Finally, it is worth mentioning that the effect of module degradation on the PV power generation is not considered in the chosen model. This effect has been intentionally omitted due to the uncertainty on the ageing rate [18][19]. We expect pragmatically that the effect of ageing is limited by the calibration procedure previously mentioned, whereas it is clear that the degradation occurring between the time period used for the calibration and that used for the validation will bring about an estimation error. A better

consideration of this loss mechanism may be considered in a future development of the proposed approach.

### **2.3 Estimation of the weights of the reference configurations (module orientation angles)**

The two parameters of the PV plant model introduced in section 2.2 being the module azimuth and tilt angle, the goal of the present section is to identify the most frequently occurring values of this pair of angles as well as their frequency of occurrence for the implementation of the regional PV power model described in equation (1). For this purpose, a database including the peak power, location and orientation angles of more than 35 000 plants is used. This database is resulting on the one hand from the collection of plant meta-information publicly available on the internet [20][21][22] and on the other hand from the characterisation of several hundreds of plants from historical data using the parameter estimation approach described in [23]. A detailed description of this database is available in [6].

Since characteristics of PV plants are known to be different for large and small plants [3], the dependence between the tilt angle and the nominal capacity was first tested by plotting all values of the tilt angles contained in the database against the corresponding nominal capacities (grey dots in Figure 2). Due to the different order of magnitude of the nominal PV plant capacity, the data are displayed with a semi-log scale. In order to better visualize the dependency between the two parameters, bin average and standard deviation are also represented (red and blue lines, respectively). The width of the bins is constant in the logarithmic scale.

A clear dependency between the tilt angle and the size of the PV plant (nominal capacity) can be observed. For PV plants smaller than 10 kW<sub>p</sub>, the bin-averages of the tilt angle increases from 31 to 35°. Between 10 and 100 kW<sub>p</sub>, bin averages of the tilt angle decrease from 32° to 16° and remain relatively constant until 1000 kW<sub>p</sub>. While the dependency of the average tilt angle with the nominal capacity is continuous until 1000 kW<sub>p</sub>, a strong discontinuity can be observed at 1000 kW<sub>p</sub>: the average tilt angle included in the classes 600-1000 and 1000-2000 kW<sub>p</sub> are 17° and 26° respectively. This discontinuity is also observable in the scatter points. The bin-average of the tilt angles for nominal capacity values greater than 1000 kW<sub>p</sub> is relatively constant and close to 26°.

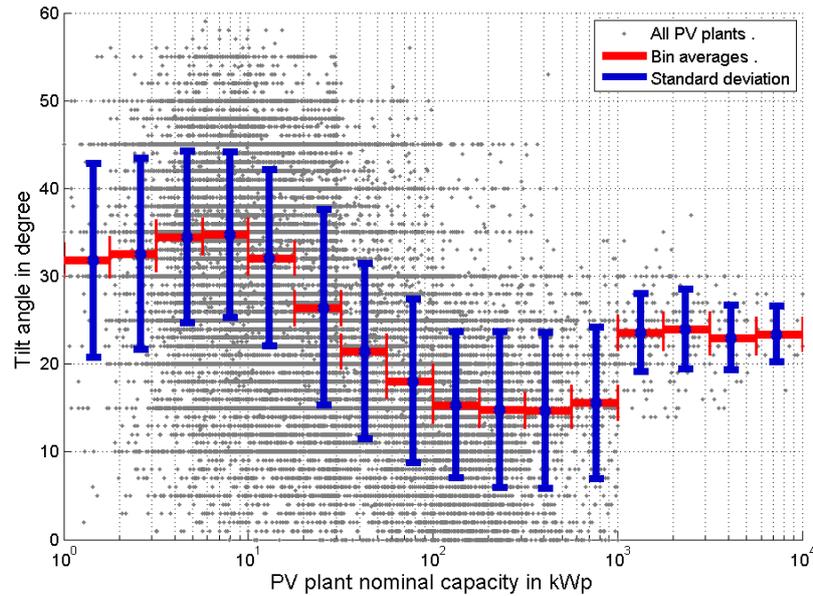


Figure 2: Scatter plot of the tilt angle against the PV plant nominal capacity. Grey points: all values; red lines: bin averages; blue lines: standard deviation for each bin.

In order to quantify the statistical dependency between the tilt angle and the size of a PV plant, 10 classes of nominal capacity were chosen to reflect the trend of the bin averages of the tilt angle observed in Figure 2 and to encompass the maximum number of values per class. Probability density distributions of the tilt angles have then been evaluated for each class. The results are displayed in Figure 3, where the colours of the different lines correspond to different classes of nominal capacity.

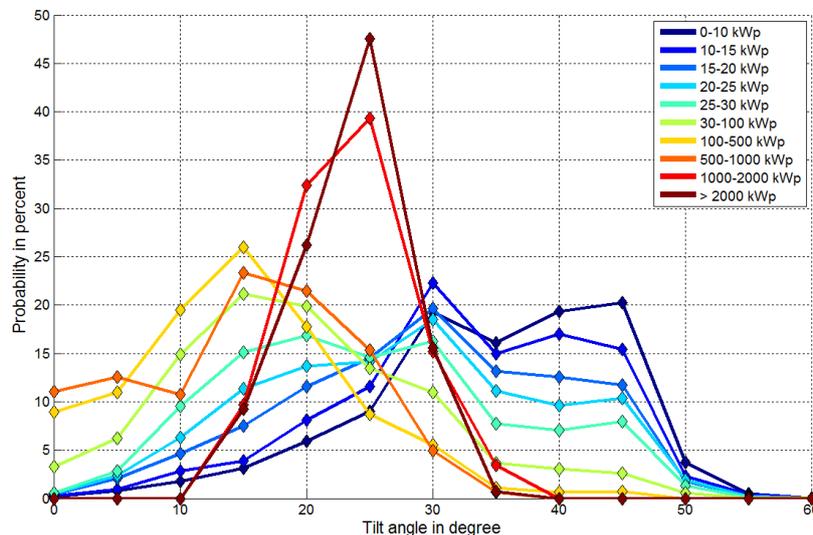
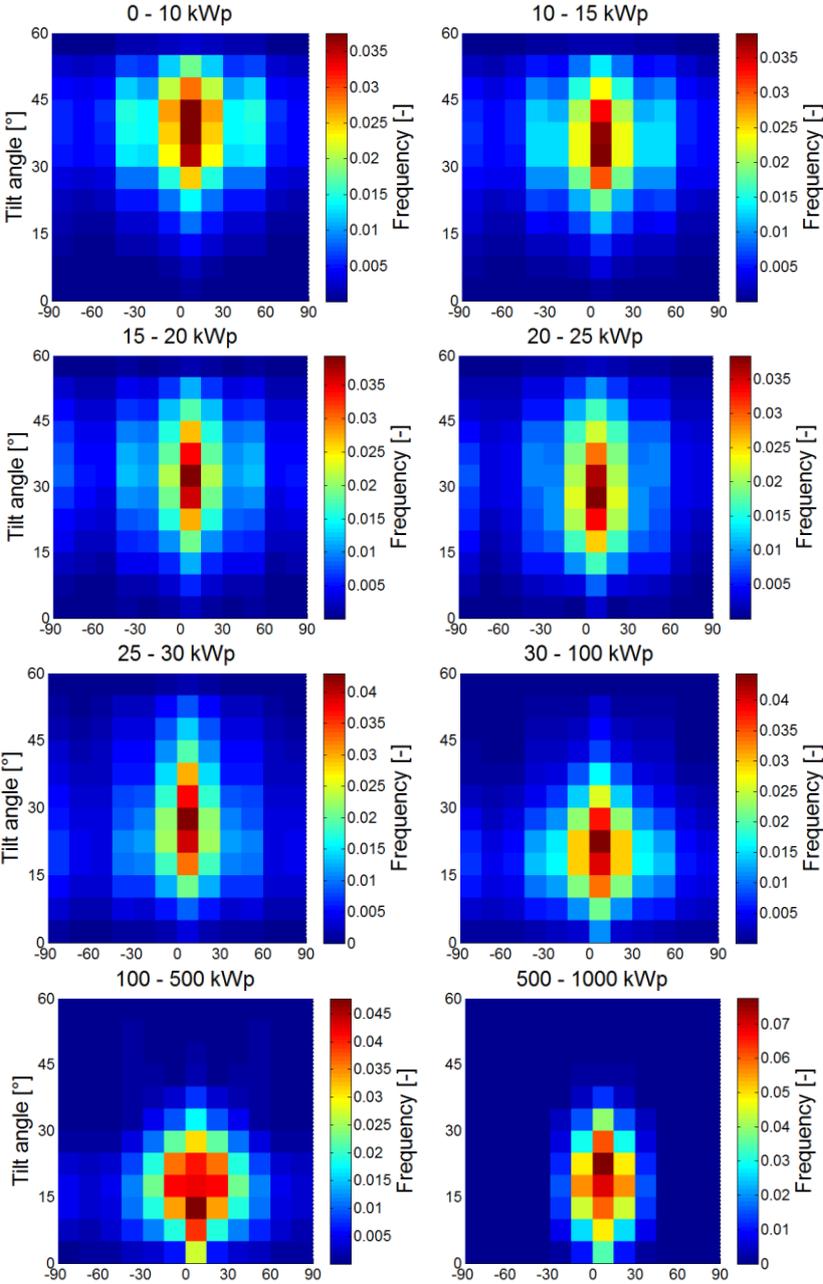


Figure 3: Probability density distribution of the tilt angle for different classes of the nominal PV plant capacity (line colours)

The trend observed in the Figures 2 and 3 can be explained as follows. For small to medium plants, PV modules are directly mounted on building roofs so that the tilt angle is determined by the roof inclination of the building. In contrast, large plants are more frequently mounted on a rack system with an optimal tilt angle. This explains why the tilt angle is suboptimal for small plants but becomes optimal, as plants get larger. A more detailed analysis of the effect of the plant size on the module tilt angle can be found in [6].

The calculated probability distributions of the tilt angles represented in Figure 4 shows the importance of considering the size of the PV plants when estimating the statistical properties of the module orientation angles. The displayed values are however not sufficient for implementing the model given in Equation 1 since the frequency of occurrence for each combination of tilt and azimuth angle is necessary. A joint probability distribution of the two module orientation angles has thus been evaluated for each class of the peak capacity (Figure 4). For this purpose, azimuth angles ranging from  $-90$  to  $90^\circ$  with a step of  $5^\circ$  and tilt angles ranging from  $0$  to  $60^\circ$  with a step of  $5^\circ$  have been considered.



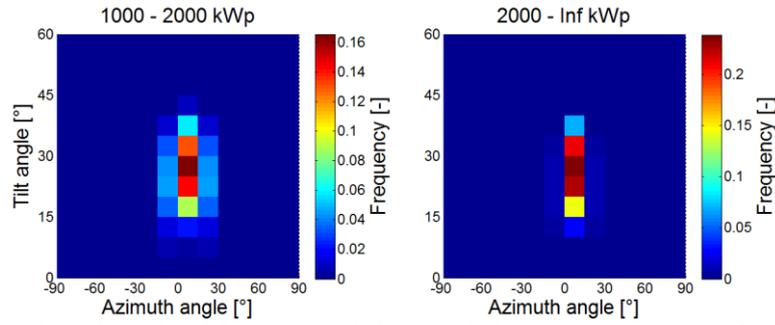


Figure 4: Joint probability distributions of the module orientation angles for 10 classes of PV plants

## 2.4 Conclusion

The set of model parameter ( $A_i$ ) introduced in Equation 1 includes the 432 combinations of tilt and azimuth angles considered for the calculation of the joint probability distributions (see Figure 4). This set of parameter is the same for all plant. The weights  $w_i$  introduced in Equation 1 are the probabilities of occurrence displayed in Figure 4. According to the results described in section 2.3, the set of weights ( $w_i$ ) will be different depending on the size of the PV plant ( $i$ ). It is thus necessary to differentiate the PV plants installed in the considered region according to their peak capacity when implementing equation 1 for the estimation of the regional PV power generation. This is possible in Germany using the register of RES plants [24], where the location and peak capacity of each plant is available.

## 3 Implementation and validation of the probabilistic model

### 3.1 Experimental setup

The proposed methodology is validated against the estimates of the total PV power generation of the four Germany transmission system operators in their control area. Time series of the total PV power generation with a time resolution of 15min have been collected on the websites of the four German TSOs [23] for the years 2014 and 2015. These data are estimates of the actual PV power generation made by the TSOs by generally upscaling the power measurements of a set of reference plants. Data of the year 2014 are used for the calibration of the model (section 3.2) and data for the year 2015 for the model validation (section 3.3). It would have been optimal to use the actual power generation instead of the estimates provided by the TSO for the validation. Unfortunately, the numerous plants installed in Germany being not measured, the true value is not available and TSO data represent the best guess of the total PV power generation. Though this approach makes a definitive conclusion on the performance of the proposed model difficult (a given uncertainty can be expected for the TSO estimates), it is representative of a real-world situation. Indeed, TSOs evaluate the skill of their forecast against their estimate.

The location and peak capacities of all plants installed in Germany have been collected from the register of plants funded by the German renewable energy act [22], which provides the plant locations ( $x$ ) and the installed capacity ( $P_{\text{peak},i}$ ) in Eq. 1. The weights ( $w_i$ ) are chosen under consideration of the installed capacity of each installed plant using the results of the statistical analysis presented in section 2.3.

Since day-ahead forecast is one of the most important applications of regional model for the integration of the PV energy in the electricity supply system, we decided to use NWP forecast data for the global horizontal irradiation and air temperature needed for the implementation of Equation 1. We used the 2-meter temperature and the solar surface radiation downwards parameter from the IFS model (ECMWF) for each plant location listed in the RES database. These parameters have been extracted from the model run starting at 00:00 and for a time lead comprised between 24 and 48 hours. The two meteorological quantities have then been linearly interpolated to match with the time resolution of the TSO estimates.

Since a day-ahead forecast is evaluated with the proposed methodology, publicly available day-ahead forecast provided by the German TSOs have also been collected for the year 2015, which allows a comparison of the forecast error obtained with the proposed methodology with that of a third party (reference). This comparison is conducted in section 3.3.

Since the effects of snow and shading are not considered in the model presented in section 2, the calibration and validation are conducted without winter months (April to October of the respective years).

### **3.2 Model calibration**

The goal of the model calibration is to determine a derating factor accounting for the different loss processes neglected in the model selection. This derating factor also accounts for systematic modelling errors. Though this step is motivated by the selection of the model in section 2.2, it can also be considered as a model output statistic aimed at improving the performance of the model.

For estimating the derating factors for the four German control areas, the rough output of the probabilistic model is compared with the TSO estimates for the year 2014 (data of the year 2015 being kept for the validation). This comparison is illustrated in Figure 5, where the TSO estimates are plotted against the output of the probabilistic approach for the four control areas (light grey dots).

We can observe in Figure 5 that - as expected (see Section 2.2) - the rough output of the probabilistic model overestimate the actual PV power generation of the four German control areas. Derating factors have been evaluated by a simple regression with the data displayed in Figure 5. The black lines in Figure 5 represent these derating factors. With values equal to 0.82205 and 0.81826, the derating factors are very similar for the Amprion and 50 Hertz control areas. The smallest and largest derating factors are found for the TenneT (0.72858) and TransNetBW (0.88464) control areas respectively.

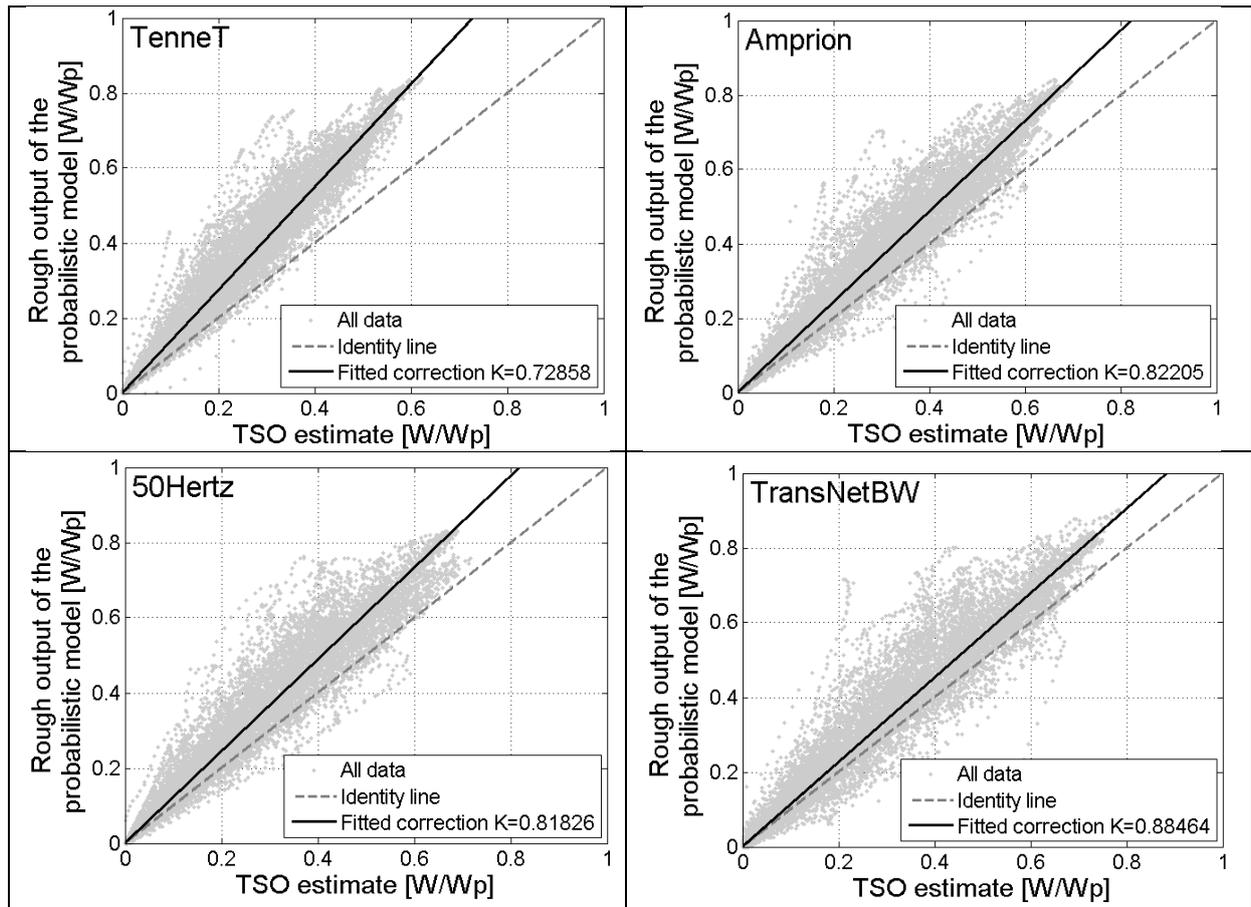


Figure 5: Scatter plot of the rough output of the probabilistic model against the TSO estimates (light grey dots) and fitted correction (black lines) for the four German control areas.

It is interesting to note that the larger is the control area, the smaller is the fitted derating factor. It appears thus that the derating factor account for a decrease of the regional power generation resulting from a smoothing effect that is underestimated by the probabilistic model. Indeed, the maximal values found with the probabilistic model are all approximately equal to 0.8 for the four control areas while the maximal values of the TSO estimates are equal to ca. 0.6, 0.65, 0.65 and 0.75 for the TenneT, Amprion, 50 Hertz and TransNetBW control area respectively (sorted by increasing size of the control area). The underestimation of the power decrease due to smoothing effect with the probabilistic model is probably stemming from the chosen weather data, whose spatial variations are smoother than the actual ones.

The evaluated derating factors are now used to calibrate the output of the probabilistic model of the validation data set, which is compared to the TSO estimates in the following section.

It is here assumed that the derating factors assessed with the data of 2014 are valid for the year 2015. The validity of this assumption is not obvious since several issues change the characteristics of the match between TSO estimate and output of the probabilistic method with time. Firstly, the statistics of the PV plant orientation angles can change as new plants are installed. We consider that this effect can be disregarded considering the newly installed PV capacity between 2014 and 2015 with respect to the installed capacity in 2014 (see Annexes 1 to 4). Secondly, improvements of the weather model

can result in a change of the characteristics of the output of the probabilistic model with time, which would in turn make the derating factor evaluated with 2014 data suboptimal for 2015. Thirdly and finally, changes in the algorithms used by the TSO for estimating the total power generated in their control area can also result in a difference between the optimal derating factors for the two time periods.

### 3.3 Model validation

The values of the calibrated output of the probabilistic model are compared with the TSO estimates of the actual PV power generation in the four scatter plots of Figure 6. It can be observed that the scatter points are centred on the identity line, which shows that the calibration indeed removed the bias observed with uncalibrated data in Figure 5.

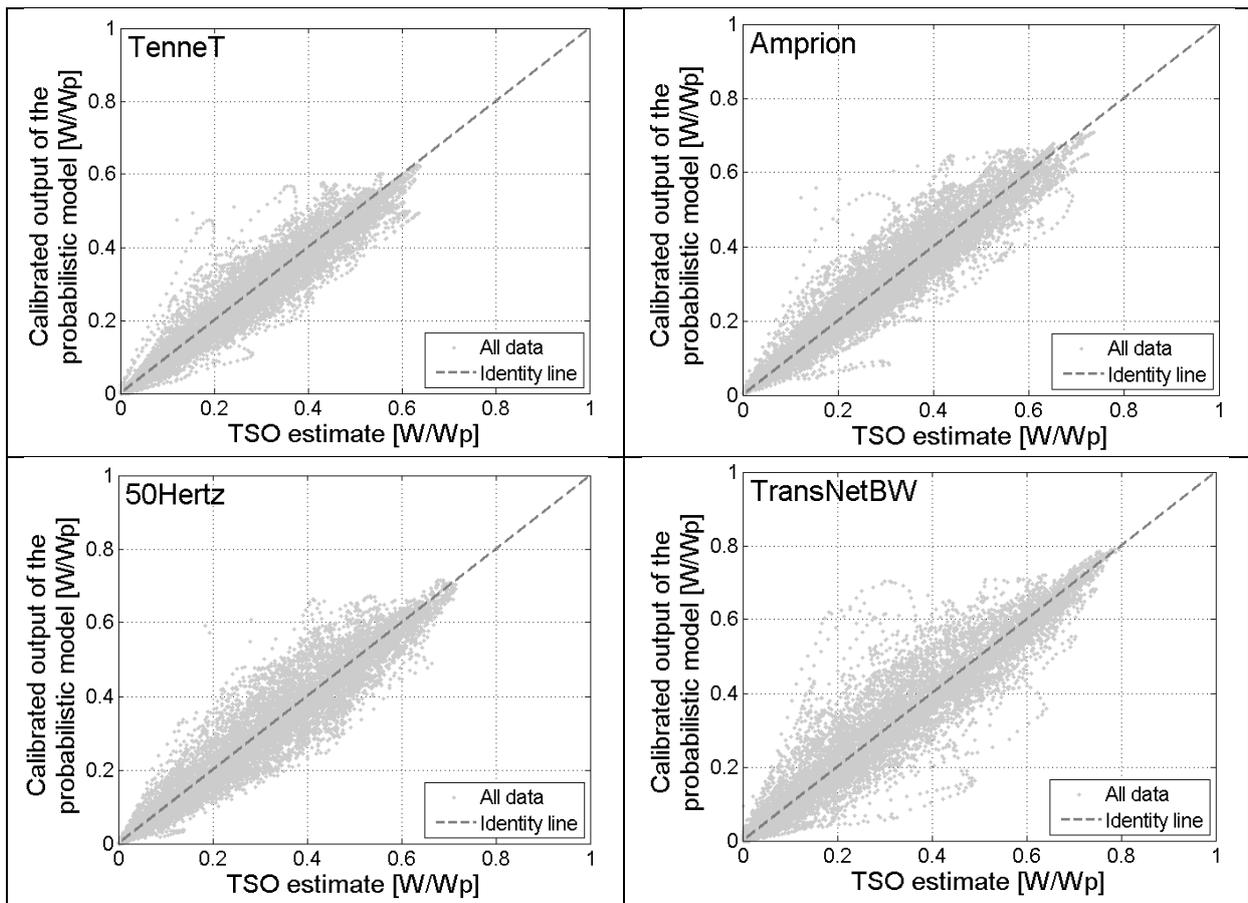


Figure 6: Scatter plot of the calibrated output of the probabilistic model against the TSO estimates (light grey dots) for the validation time period.

A comparison of the time series of the forecast disseminated by TenneT and the calibrated output of the probabilistic model with TenneT estimate is displayed for the time period 14-21/08/2015 in Figure 7. It can be observed that the output of the probabilistic approach matches well with the TSO estimate and that there are even days when it outperform the forecast provided by the TSOs.

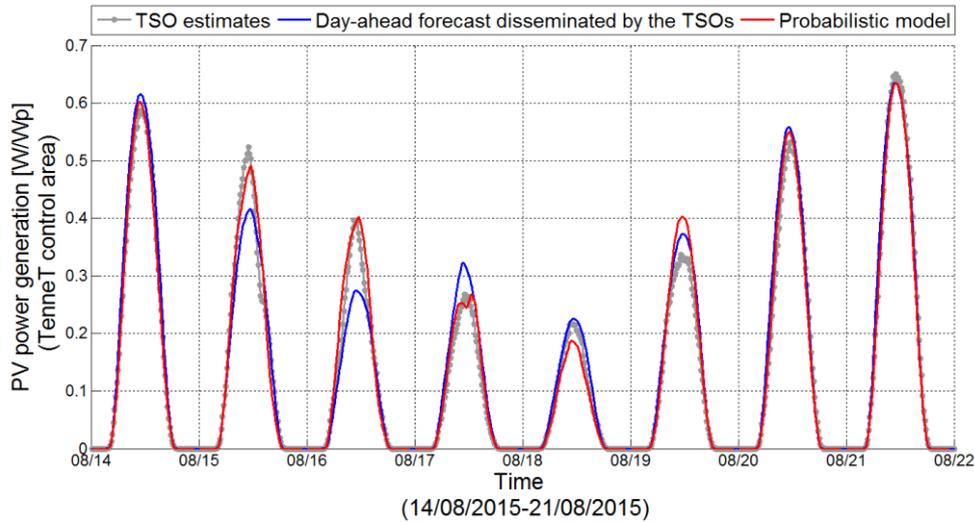


Figure 7: Comparison of the time series of the forecast disseminated by the TSO (blue) and the calibrated output of the probabilistic model (red) with the TSO estimates (grey) for the TenneT control area between 14/08/2015 and 21/08/2015

Different error metrics have been evaluated to assess the performance of the probabilistic approach, which are given in Table 3. The results are expressed in percent of the installed capacity. In addition, the root mean square errors of the probabilistic method and the forecast disseminated by the TSOs are displayed in Figure 8. Since the root mean square error (RMSE) is mainly used by the TSOs, the following analysis is focusing on this metric.

Table 3: Average power generation, bias, MAE and RMSE of the calibrated regional PV power calculation for the four German control areas evaluated with the TSO estimates of the actual PV power generation (in % of the installed capacity).

		TenneT	Amprion	50 Hertz	TransNetBW
Average power generation		0.2167 W/W <sub>p</sub>	0.2425 W/W <sub>p</sub>	0.2477 W/W <sub>p</sub>	0.2802 W/W <sub>p</sub>
Bias	Probabilistic approach	-0.61 %	0.28 %	-0.46 %	0.16 %
	TSO forecasts	1.17 %	0.27 %	-0.08 %	0.23 %
MAE	Probabilistic approach	2.77 %	3.16 %	3.27 %	3.83 %
	TSO forecasts	2.55 %	4.04 %	2.92 %	2.82 %
RMSE	Probabilistic approach	4.16 %	4.76 %	4.87 %	6.13 %
	TSO forecasts	3.73 %	5.99 %	4.51 %	4.54 %

In Figure 8, the TSOs are sorted by increasing size of the control area. It can be observed that the RMSE increases as the size of the control area decreases. It can also be noted in Table 3 and Figure 8 that at the exception of TransNetBW, RMSE values obtained with the probabilistic model are all below 5% and even very close to 4% for TenneT. The reason for the higher RMSE value obtained for the TransNetBW control area may on the one hand be explained by the smaller spatial extension of the area in comparison to

other control areas. On the other hand, as can be seen in Annex 4, the last input of the TransNetBW EEG register has been recorded in July 2014 while registers of the other TSO are more recent with last input recorded between April and May 2016 (see Annexes 1, 2 and 3). The higher error obtained for TransNetBW may thus also be resulting from a mismatch between the set of plants used for the calculation and the actual ones. Accordingly the higher RMSE value obtained for TransNetBW should be considered with some caution and cannot be used as an indicator of the performance of the proposed approach. The results obtained for the TransNetBW control area are therefore not further considered in the continuation of this analysis.

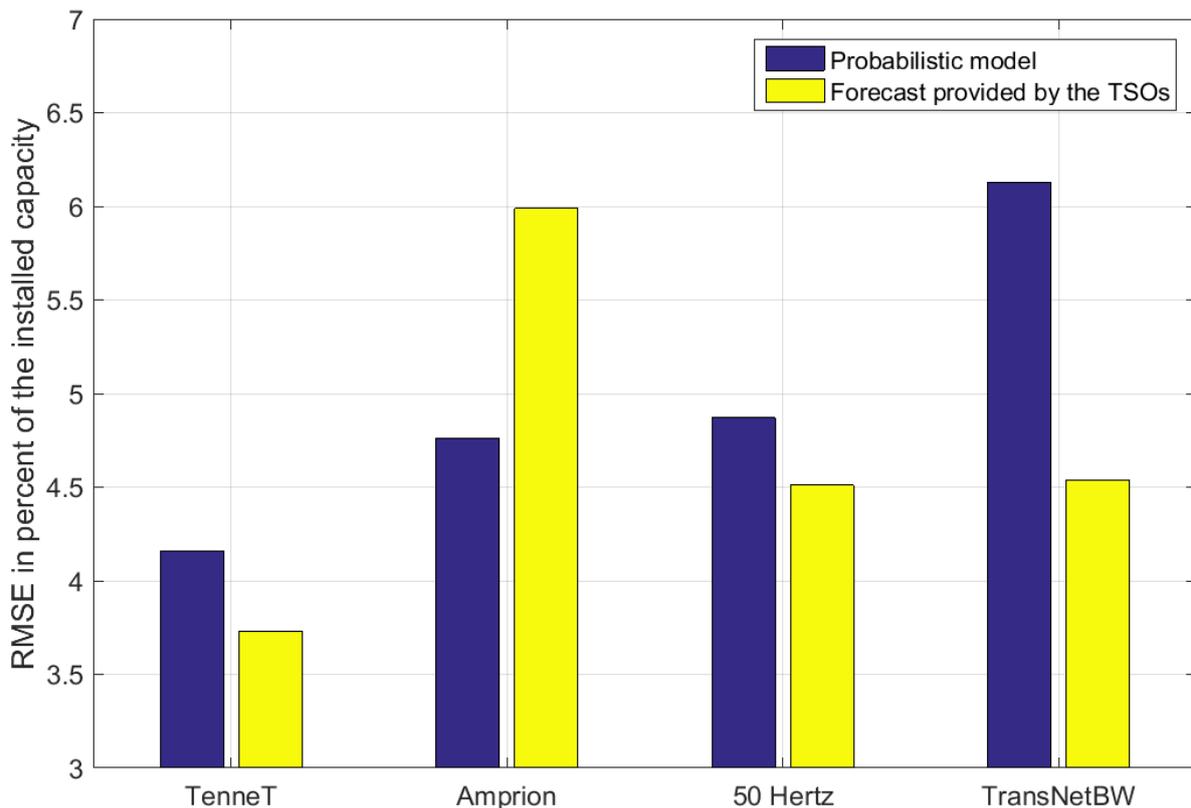


Figure 8 RMSE of the forecast calculated with the probabilistic model (blue bars) and RMSE of the day-ahead forecast disseminated by the TSOs (yellow bars). (More detailed analyses of the errors are given in Annexes 1, 2, 3 and 4)

The performances of the probabilistic model are then compared to those obtained with the day-ahead forecast provided by the TSOs. The dissemination of a day-ahead forecast by the TSO is made mandatory by the EEG act. It is however unclear which forecast products are delivered and if those are representative for the forecast used operationally by the TSOs. This benchmark has to be interpreted with caution, since the TSO forecast methodology is unknown. It can for example be noted that the RMSE of the forecast provided by Amprion is noticeably higher than those of the three other TSOs. This value is probably corresponding to an under-average forecast and will not be further considered in the present analysis. Given the relatively low RMSE values obtained for TenneT and 50 Hertz, we consider that the data provided by the TSO are corresponding to outputs of cutting-edge forecast systems. The comparison will thus be limited to these two control areas.

It can be observed in Table 3 and Figure 8 that the error obtained with the probabilistic model is only 0.43 and 0.36 higher than that of the day-ahead forecast provided by the TenneT and 50 Hertz respectively. This can be considered as a positive result considering that the forecast evaluated with the probabilistic model is based on one single weather model and that – with the exception of the model calibration – no post-processing scheme has been used. Indeed, most commercial forecast providers optimize the skill of their forecast by means of statistical post-processing and/or by implementing a multi-model approach.

#### **4 Discussion and Conclusion**

In this paper, an alternative to the upscaling approach for estimating the aggregated PV power generation of PV plants installed in a region was presented. In this method, called a probabilistic regional PV model, an average PV model with a very limited number of inputs (two module orientation angles) is used to calculate the power generation of the most frequent module orientation angles. The resulting power values are finally weighted according to their probability of occurrence to estimate the actual power generation. The probability of occurrence of the module orientation is assessed on the basis of a database with more than 35 000 PV plants. The implementation of this model only requires information on the location and peak capacity of the plant installed in a region and no PV plant measurement is necessary.

The advantage of this method is that a value of the PV power generation can be estimated for each PV plant installed in the region of interest with local meteorological data. The interpolation error made in the upscaling method is thus avoided with the proposed methodology. Furthermore, it is hoped that the problem of representativity faced by the upscaling method [24] are mitigated by the method proposed in this paper by using statistical distribution of module orientation angles as input. Lastly, the proposed approach allows forecasting the aggregated PV power generation of any region without PV power measurements, which characteristic is highly valuable for numerous stakeholders without access to the numerous PV power measurement data needed to apply the upscaling method.

The validation of the proposed model against TSO estimates shown that the RMSE obtained with the probabilistic model is below 5%. The forecast made with the probabilistic approach was found to be less than 0.5% higher than the forecast delivered by the TSOs. This is considered a positive result considering that this forecast is evaluated with only one single weather model and that we do not use a MOS scheme while commercial forecast providers generally optimize their forecast accuracy by using a multi-model approach or/and by implementing statistical post-processing techniques. The proposed methodology can thus be considered as an interesting alternative to the upscaling approach in situations where no or little power measurements are available. As shown in [5] and [6], the upscaling method yields very accurate results when it is used with numerous PV power measurements. In that case, the added value of the proposed methodology is less obvious but it can be expected that a combination of the two approaches may be interesting to e.g. limit the effect of unrepresentative reference plants.

Though the validation of the probabilistic approach is made using PV power forecast, the focus of this paper is on the calculation of the aggregated power generation of a fleet of PV plants and not on the reduction of the PV power forecast error. It is solely validated that the proposed methodology can be implemented to model a fleet of PV plant in a forecast context. The improvement of the forecast skill may be considered using adapted approaches such as e.g. the explicit consideration of systematic forecast errors with fog and low stratus [29].

It is finally interesting to note that, thanks to its flexibility, the proposed approach offer a wide scope for improvements and further developments.

It can at first glance be expected that the performance of the method increases with the quality of the statistical information used (average PV plant model and distribution of the module orientation angles). An improvement of the statistical analysis presented in this paper by using more PV plants may thus represent a possible continuation of this work. Alternatively, module orientation angles assessed from LiDAR data [25][26] could also be used instead –or together with the statistical data in the future.

It should however be noted that the performance of a forecast system are evaluated by the TSO against an estimate of the actual generation, which is based on the upscaling of a limited set of reference plants. Situations can thus occur where an improvement of the probabilistic forecast by e.g. using LiDAR data or improved statistics is interpreted as a degradation of the forecast skill. This would happen if the difference between the forecast and the TSO estimates increases. Such situations are probable given that medium to large reference plants are mainly used by TSO for estimating the actual PV production while the probabilistic approach is taking into account characteristics of all plant sizes, including small ones. Should the goal of a forecast be to match with the TSO estimate, a more pragmatic approach may consist in choosing the statistical input of the probabilistic method giving the best match between model output and TSO estimates. This is a difficult issue given the number of parameters and the ill-conditioned nature of the problem. However, it could be addressed by implementing inverse modelling techniques, which could also represent a possible continuation of the work presented in this paper.

Finally, it should be noted that winter months have been excluded for the present analysis. The reason for that is that the effects of snow and shading on the PV power generation are not considered in the model described in this paper. A consideration of these effects could also be addressed in a further development of the proposed methodology.

## **Acknowledgement**

This work was completed within the framework of the EWeLiNE project (Erstellung innovativer Wetter- und Leistungsprognosemodelle für die Netzintegration wetterabhängiger Energieträger) funded by the German Federal Ministry for Economic Affairs and Energy. The authors would like to thank the German TSOs Amprion, TenneT and 50 Hertz for their support in understanding the integration mechanisms of RES energy in the electricity supply system and all the related data. We would like to thank the German weather service DWD for the pleasant cooperation and productive discussions. The authors are indebted to Prof. Lucien Wald for his time and valuable

feedback. We would also like to thank the Transvalor team, who take care of the SoDa Service that makes the access to the HelioClim databases efficient and user-friendly.

## References

- [1] Consumption data: hourly load values for Germany, European network of transmission system operators for electricity (ENTSO-E), Retrieved in October 2015 from: [www.entsoe.eu/data/data-portal/consumption/Pages/default.aspx](http://www.entsoe.eu/data/data-portal/consumption/Pages/default.aspx)
- [2] Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen, Monitoringbericht 2014), Retrieved on Februar 18, 2014 from: [www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Allgemeines/Bundesnetzagentur/Publikationen/Berichte/2014/Monitoringbericht\\_2014\\_BF.pdf](http://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Allgemeines/Bundesnetzagentur/Publikationen/Berichte/2014/Monitoringbericht_2014_BF.pdf)
- [3] Lorenz, E., Heinemann, D.: 'Prediction of Solar Irradiance and Photovoltaic Power' in *Comprehensive Renewable Energy*, Volume 1, pp239 - 292. DOI: 10.1002/pip.1224
- [4] Schierenbeck, S., Graeber, D. Semmig A., and Weber A., 2010, Ein distanzbasiertes Hochrechnungsverfahren für die Einspeisung aus Photovoltaik, *Energiewirtschaftliche Tagesfragen*
- [5] Saint-Drenan, Y-M., G.H. Good, M. Braun (2016), "Analysis of the uncertainty in the estimates of regional PV power generations evaluated with the upscaling method", *Solar Energy*, 135, 536-550
- [6] A Probabilistic Approach to the Estimation of Regional Photovoltaic Power Generation using Meteorological Data, Application of the Approach to the German Case, PhD Thesis, Yves-Marie Saint-Drenan, Kassel, August 2015
- [7] Iqbal, M., 1983. *An Introduction to Solar Radiation*. Academic Press, New York.
- [8] Quaschnig, V., 2009. *Regenerative Energiesysteme*, sixth ed. Hanser Verlag München.
- [9] Skartveit, A., Olseth, J.A., Tuft, M.E., 1998. An hourly diffuse fraction model with correction for variability and surface albedo. *Sol. Energy* 63, 173-183.
- [10] Perez, R., Seals, R., Michalsky, J., 1991. An all-weather model for sky luminance distribution-preliminary configuration and validation. *Sol. Energy* 50 (3), 235-245.
- [11] Martin, N., Ruiz, J.M., 2001. Calculation of the PV modules angular losses under field conditions by means of an analytical model. *Sol. Energy Mater. Sol. Cells* 70, 25-38.
- [12] Beyer, H.G. et al.: Identification of a general model for the MPP performance of PV-modules for the application in a procedure for the performance check of grid connected systems, 19th E. PV. *Solar Energy Conference*, Paris, 7.-11.06.2004
- [13] Ross, R.G., 1976. Interface design considerations for terrestrial solar cell modules, *Proceedings of the 12th IEEE photovoltaic specialist's conference*, Baton Rouge, LA, pp: 801-806
- [14] Lorenz, E., Hurka, J., Karampela, G., Heinemann, D., Beyer, H.G., Schneider, M.: 'Qualified forecast of ensemble power production by spatially dispersed grid-connected PV systems', 23rd *European Photovoltaic Solar Energy Conference*, Valencia, Spain, 1-5 September 2008
- [15] Schmidt, H., Sauer, D. U.: "Wechselrichter-Wirkungsgrade – praxisgerechte Modellierung und Abschätzung", *Sonnenenergie* 4/96, Seiten 43-47
- [16] Database of inverter characteristics from Photon. Retrieved in January 2014 from [www.photon.info/photon\\_site\\_db\\_solarmodule\\_de.photon](http://www.photon.info/photon_site_db_solarmodule_de.photon)

[17] Sark, W.G.J.H.M. van; Reich, N.H.; Müller, B.; Armbruster, A.; Kiefer, K.; Reise, C., *Review of PV performance ratio development*, World Renewable Energy Forum, WREF 2012. Vol.6 : Including World Renewable Energy Congress XII and Colorado Renewable Energy Society (CRES) Annual Conference, Red Hook, NY: Curran, 2012, pp. 4795-4800

[18] D. C. Jordan, Kurtz S. R., *Photovoltaic degradation rates – an analytical review*, Progress in Photovoltaics: Research and Applications. Vol. 21, Issue 1, pp. 12-29. 2013.

[19] K Kiefer, Dirnberger D., Müller B., Heydenreich W., Kröger-Vodde A., *A degradation analysis of PV power plant*, Proceedings of the 25th European Photovoltaic Solar Energy Conference, 5032-5037

[20] *Sonnenertrag.eu*. Free non-party and independent energy database for photovoltaic facilities including detailed information for several thousands of PV plants, Retrieved in Juni 2014 from [www.sonnenertrag.eu](http://www.sonnenertrag.eu)

[21] *PV-log*. Free online database of photovoltaic plants including detailed information from several thousands of plants. Retrieved in July 2014 from: [www.pv-log.com](http://www.pv-log.com)

[22] *Solarstron.Ertragsdatenbank des Solarenergie-Förderverein Deutschlands e.V. (SFV)* Retrieved in Juni 2014 from: [www.pv-ertraege.de](http://www.pv-ertraege.de)

[23] Saint-Drenan, Yves-Marie, Bofinger, Stefan, Fritz, Rafael, Vogt, Stefan, Good, Garrett, Dobschinski, Jan, (2015). "An empirical approach to parameterizing photovoltaic plants for power forecasting and simulation". *Sol. Energy* 120, 479–493

[22] Register of the power plants funded by the German renewable energy act (EEG) disseminated by the four German transmission system operators. Retrieved in May 2016 from:  
[https://eeg-kwkg.tennet.eu/EEG\\_Auskunft/eeg-einspeisung.do](https://eeg-kwkg.tennet.eu/EEG_Auskunft/eeg-einspeisung.do)  
<http://www.amprion.net/eeg-anlagenstammdaten-aktuell>  
<http://www.50hertz.com/de/EEG/Veroeffentlichung-EEG-Daten/EEG-Anlagenstammdaten>  
<https://www.transnetbw.de/de/eeg-kwk-g/eeg/eeg-anlagendaten>

[23] Estimates and day-ahead forecast of the PV power generation disseminated by the four German transmission system operators. Retrieved in September 2016 from:  
<https://www.tennetso.de/site/Transparenz/veroeffentlichungen/netzkennzahlen/tatsaechliche-und-prognostizierte-solarenergieeinspeisung>  
<http://www.amprion.net/photovoltaikeinspeisung>  
<http://www.50hertz.com/de/Kennzahlen/Photovoltaik>  
<https://www.transnetbw.de/de/kennzahlen/erneuerbare-energien/fotovoltaik/>

[24] Saint-Drenan, Yves-Marie, Good, Garrett, Braun, Martin, Freisinger Thomas, "Analysis of the uncertainty in the estimates of regional PV power generation evaluated with the upscaling method", *Solar Energy*, Volume 135, October 2016, Pages 536–550

[25] R.O.C., Tse, M., Dakowicz, C.M., Gold, D., Kidner, (2005), „Building reconstruction using LIDAR data“, *Proceedings 4th ISPRS Workshop on Dynamic and Multi-dimensional GIS*, 156-161

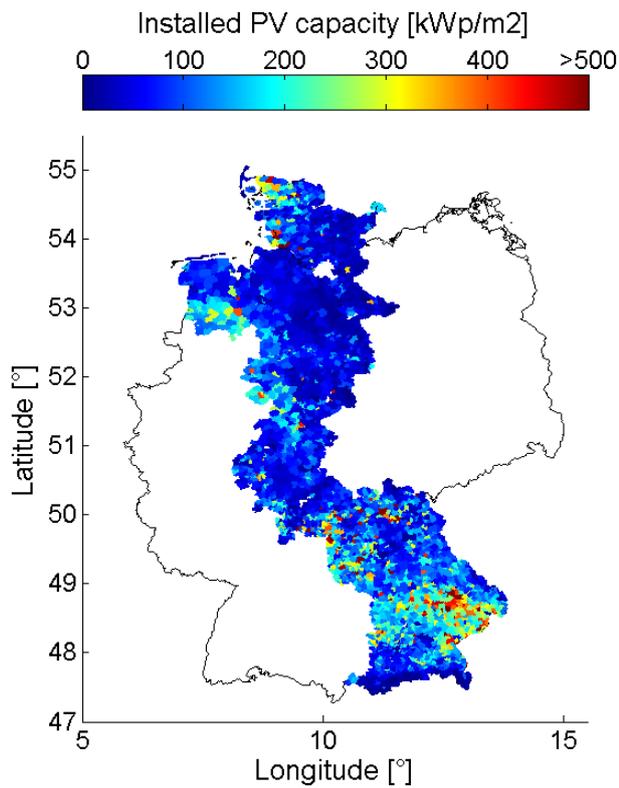
[26] M., Kada, L., McKinley (2009), "3D building reconstruction from LiDAR based on a cell decomposition approach", *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVIII, Part 3/W4 (pp. 47-52)

[27] H. Shaker, H. Zareipour, and D. Wood, "A data-driven approach for estimating the power generation of invisible solar sites," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, 2015.

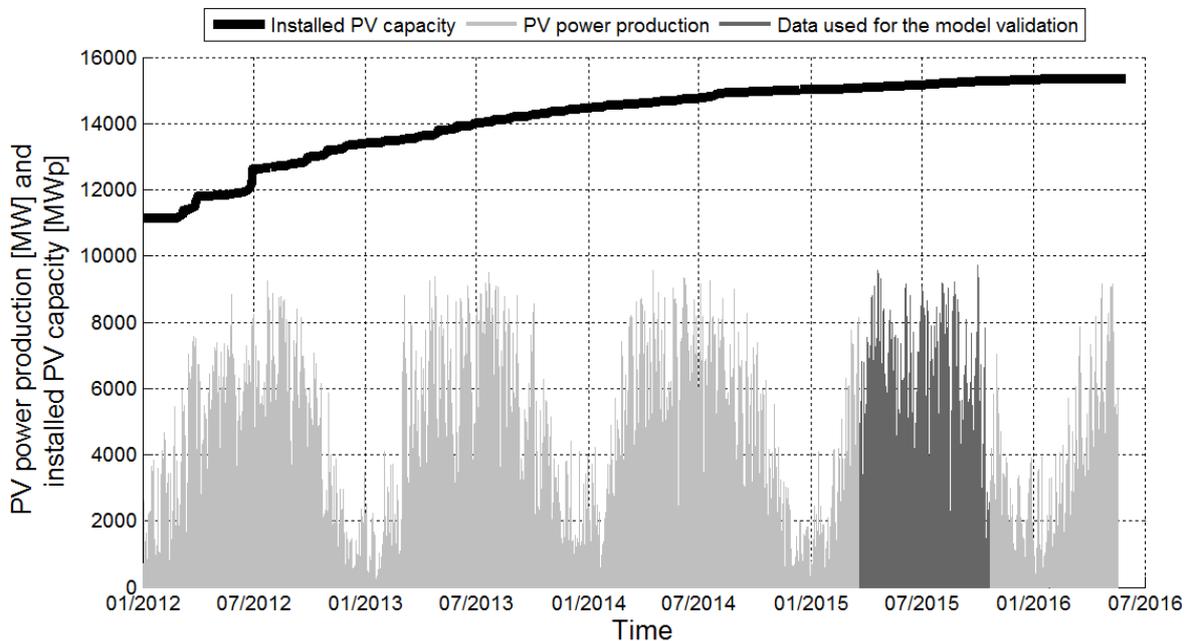
[28] H. Shaker, H. Zareipour, and D. Wood, "Estimating power generation of invisible solar sites using publicly available data," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, 2016.

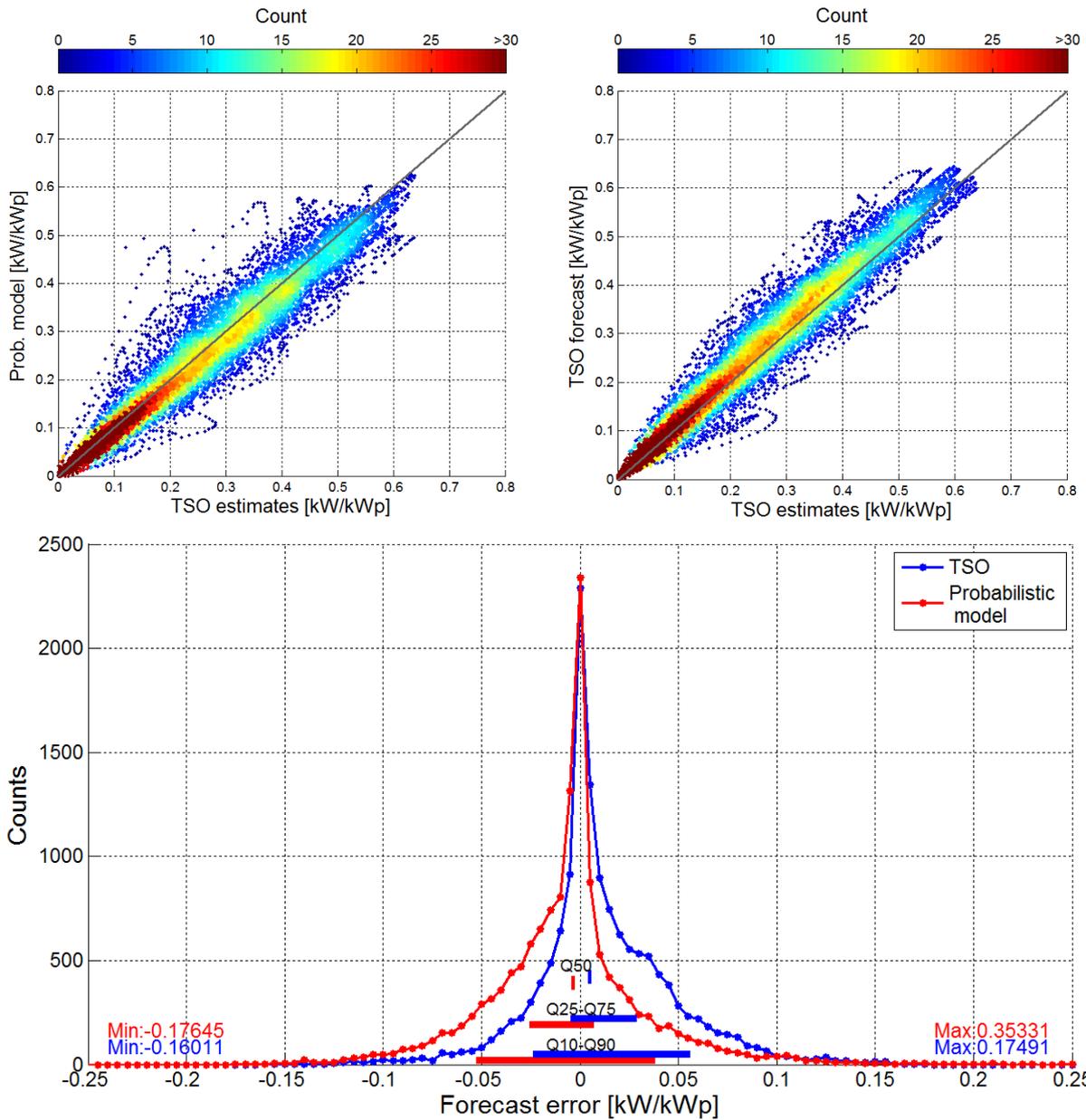
[29] Carmen Köhler, Andrea Steiner, Yves-Marie Saint Drenan, Dominique Ernst, Anja Bergmann-Dick, Mathias Zirkelbach, Zied Ben Bouallègue, Isabel Metzinger, Bodo Ritter, (2016) "Critical Weather Situations for Renewable Energies -Part B: Low Stratus Risk for Solar Power", *Renewable Energy*, Volume 101, February 2017, Pages 794–803

# Annex 1 - Results for the TenneT control area



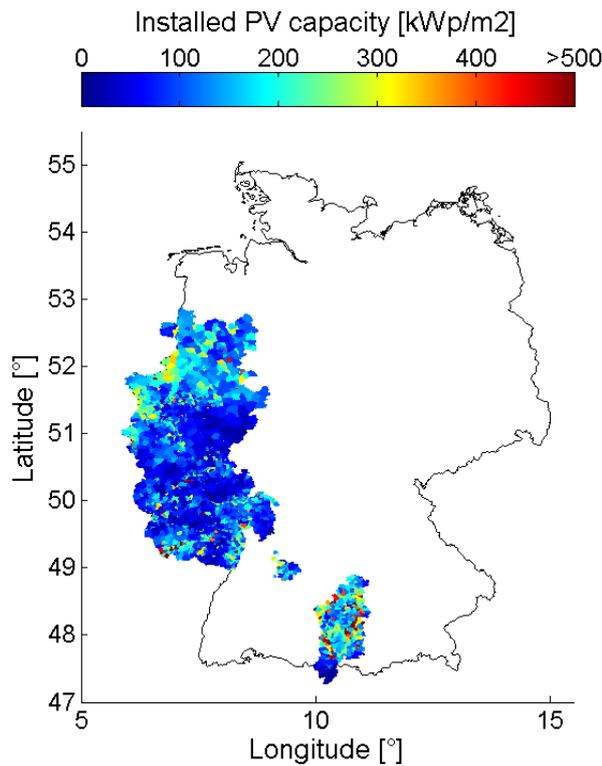
<b>Main characteristics of the TenneT control area</b>		
<i>(Update time: 05/2016)</i>		
Area		140.000 km <sup>2</sup>
Time of the last input		12-May-2016
Number of plants	0-30kW <sub>p</sub>	595 415 (88.7 %)
	30-100 kW <sub>p</sub>	63 347 (9.4 %)
	100-1000 kW <sub>p</sub>	11 346 (1.7 %)
	>1000 kW <sub>p</sub>	1 143 (0.2 %)
	<b>Total</b>	<b>671 251</b> <b>(100 %)</b>
Installed capacity	0-30kW <sub>p</sub>	6 569.2 MW <sub>p</sub> (42.8 %)
	30-100 kW <sub>p</sub>	3 075.3 MW <sub>p</sub> (20.0 %)
	100-1000 kW <sub>p</sub>	2 591.7 MW <sub>p</sub> (16.9 %)
	>1000 kW <sub>p</sub>	3 115.9 MW <sub>p</sub> (20.3 %)
	<b>Total</b>	<b>15 352.0 MW<sub>p</sub></b> <b>(100 %)</b>



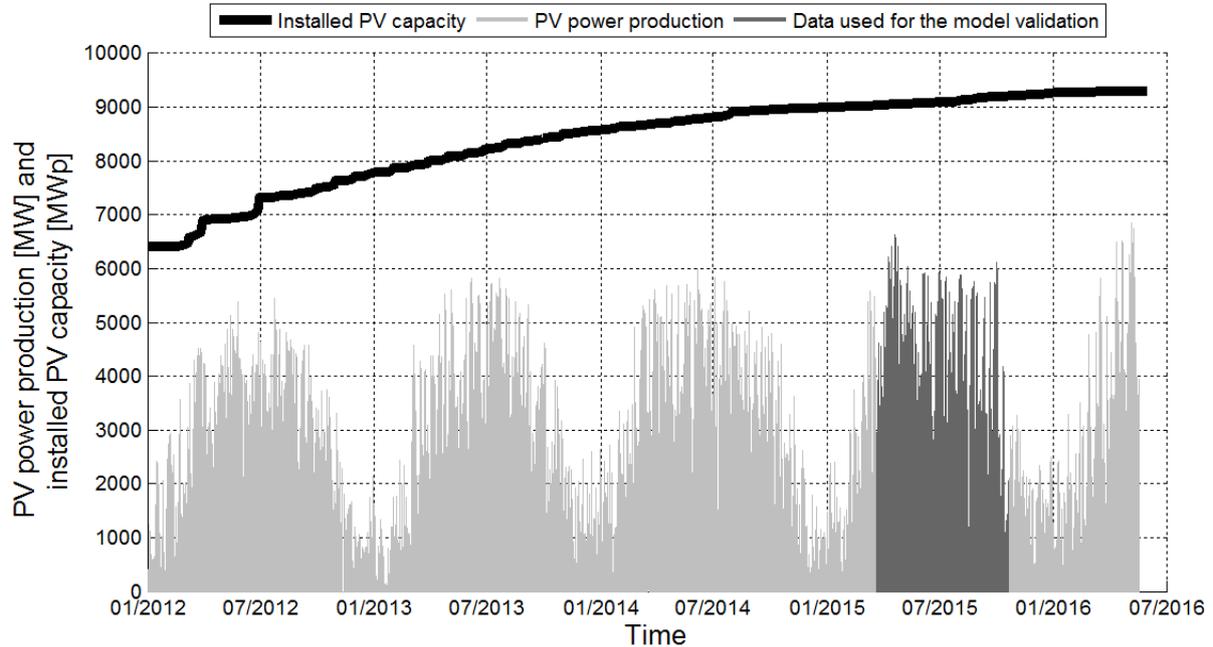


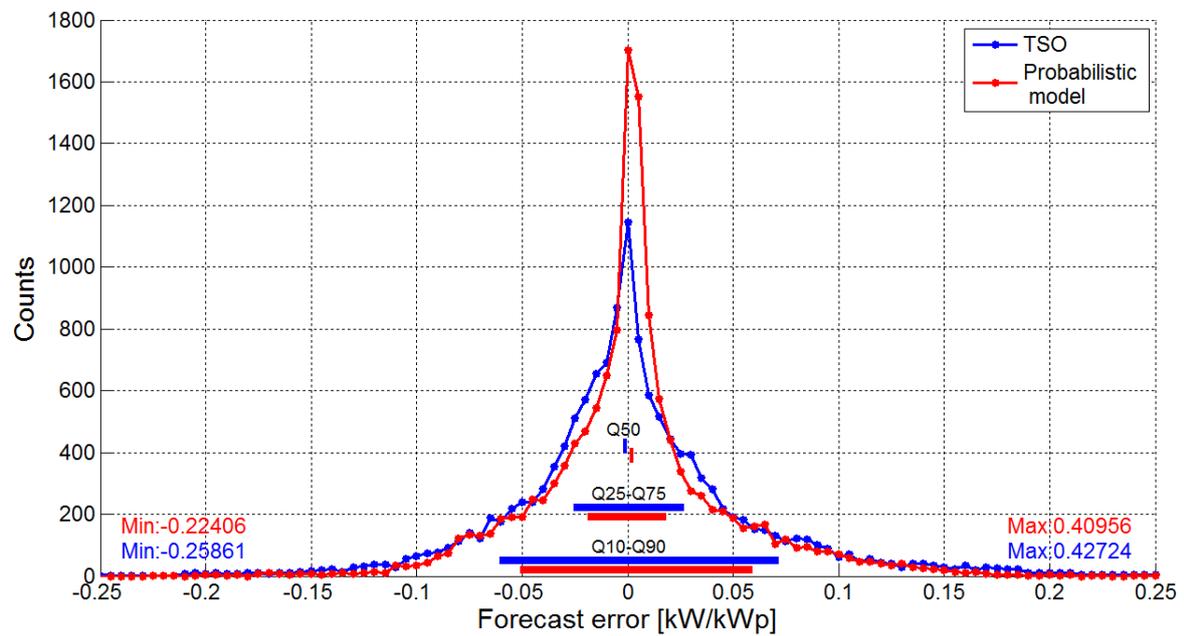
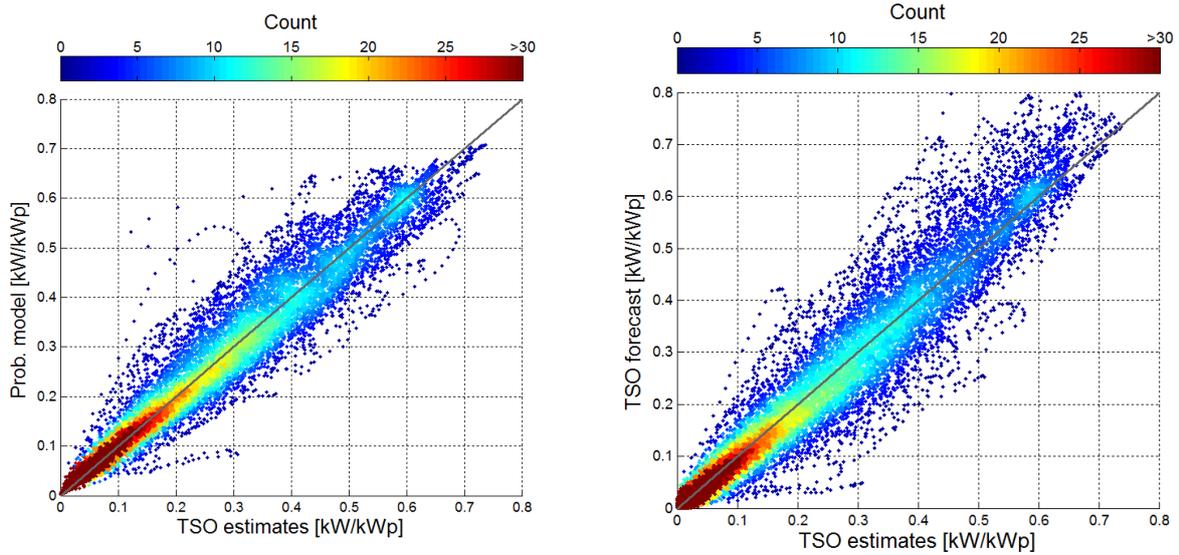
	<b>Probabilistic model</b>	<b>TSO forecast</b>
<i>Bias</i>	-0.61 %	1.17 %
<i>MAE</i>	2.77 %	2.55 %
<i>RMSE</i>	4.16 %	3.73 %
<i>min / max</i> $\Delta$	-17.65 / 35.33 % 52.98 %	-16.01 / 17.49 % 33.50 %
<i>Q10% / Q90%</i> <i>IQ</i>	-5.25 / 3.82 % 9.07 %	-2.39 / 5.61 % 8.00 %
<i>Q25% / Q75%</i> <i>IQ</i>	-2.54 / 0.75 % 3.29 %	-0.47 / 2.93 % 3.4 %
<i>Median</i>	-0.34 %	0.51 %
<i>Correlation coefficient</i>	0.96918	0.97953

## Annex 2 – Results for the Amprion control area



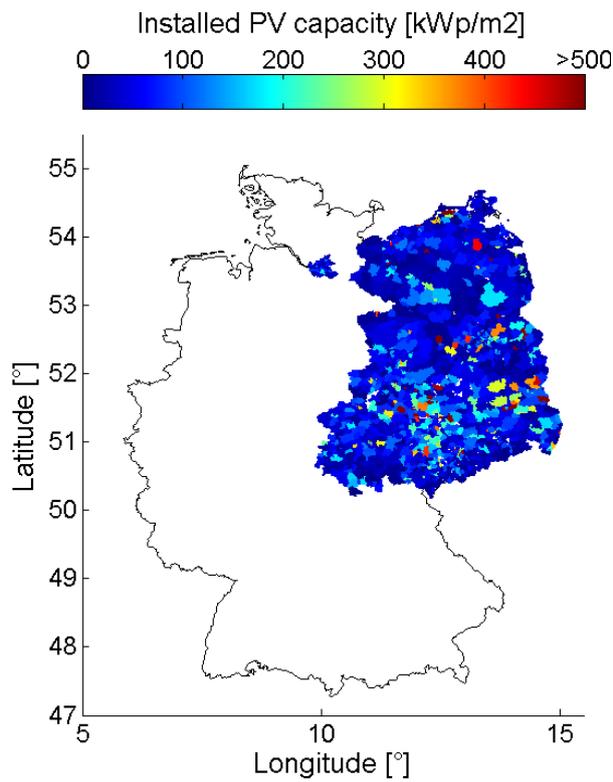
<b>Main characteristics of the Amprion control area</b>		
<i>(Update time: 05/2016)</i>		
Area		73.100 km <sup>2</sup>
Time of the last input		03-May-2016
Number of plants	0-30kW <sub>p</sub>	411 271 (89.3 %)
	30-100 kW <sub>p</sub>	40 387 (8.8 %)
	100-1000 kW <sub>p</sub>	8 623 (1.9 %)
	>1000 kW <sub>p</sub>	501 (0.1 %)
	<b>Total</b>	<b>460 782</b> <b>(100 %)</b>
Installed capacity	0-30kW <sub>p</sub>	4 004.5 MW <sub>p</sub> (43.1 %)
	30-100 kW <sub>p</sub>	2 039.2 MW <sub>p</sub> (21.9 %)
	100-1000 kW <sub>p</sub>	1 927.7 MW <sub>p</sub> (20.7 %)
	>1000 kW <sub>p</sub>	1 329.2 MW <sub>p</sub> (14.3 %)
	<b>Total</b>	<b>9 300.6 MW<sub>p</sub></b> <b>(100 %)</b>



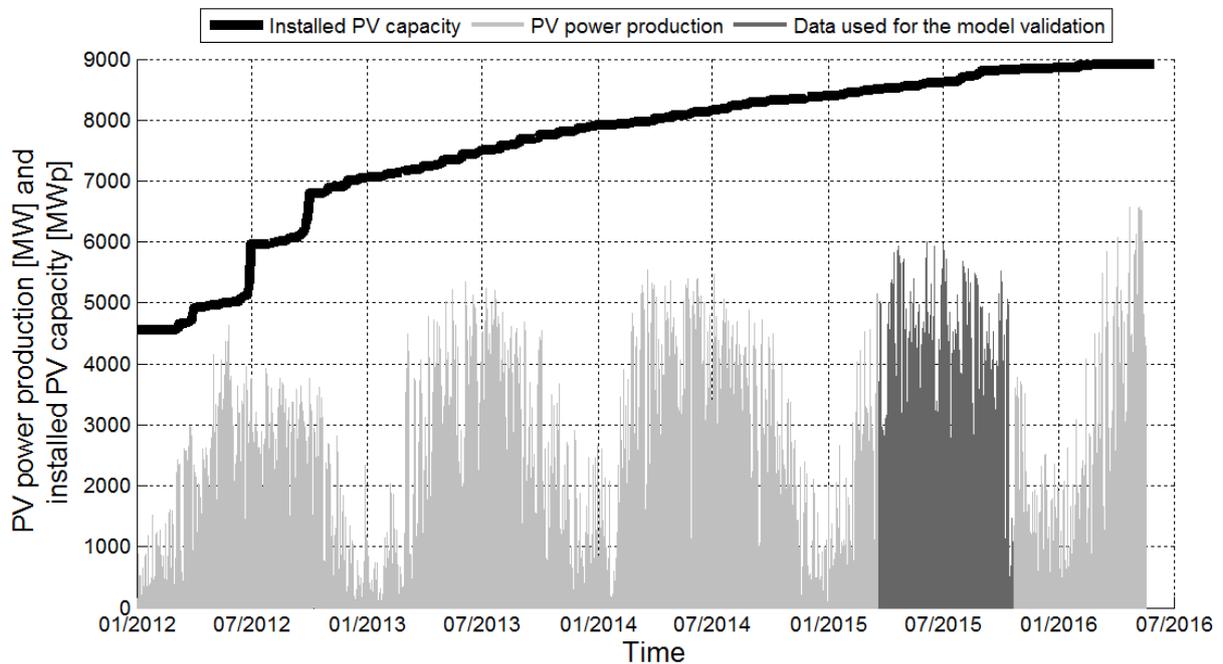


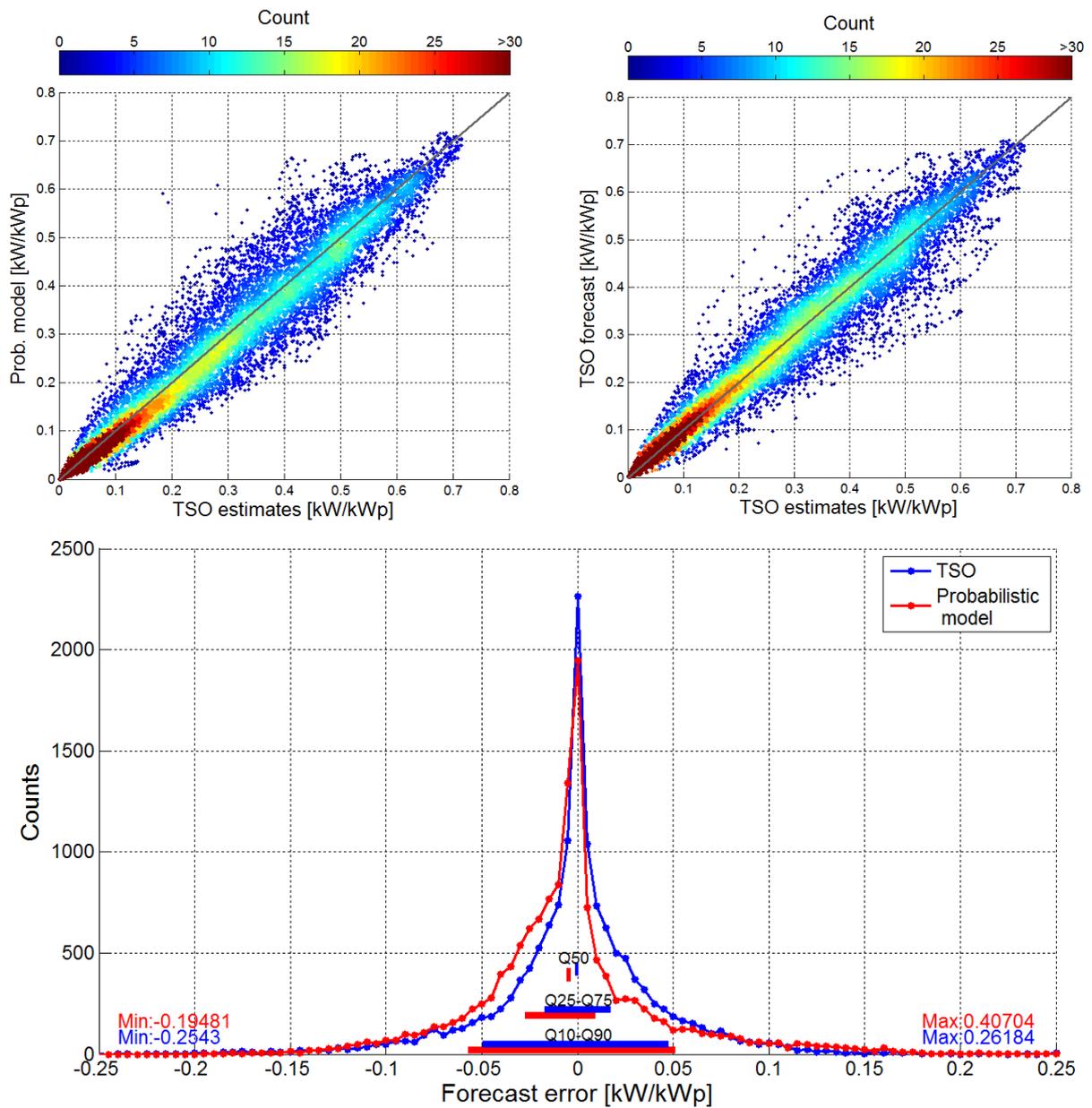
	<b>Probabilistic model</b>	<b>TSO forecast</b>
<i>Bias</i>	0.28 %	0.27 %
<i>MAE</i>	3.16 %	4.04 %
<i>RMSE</i>	4.76 %	5.99 %
<i>min / max</i> $\Delta$	-22.41 / 40.96 % 63.36 %	-25.86 / 42.72 % 68.58 %
<i>Q10% / Q90%</i> <i>IQ</i>	-5.1 / 5.95 % 11.05 %	-6.07 / 7.19 % 13.26 %
<i>Q25% / Q75%</i> <i>IQ</i>	-1.88 / 1.83 % 3.71 %	-2.56 / 2.67 % 5.23 %
<i>Median</i>	0.19 %	-0.11 %
<i>Correlation coefficient</i>	0.96713	0.95438

## Annex 3 – Results for the 50 Hertz control area



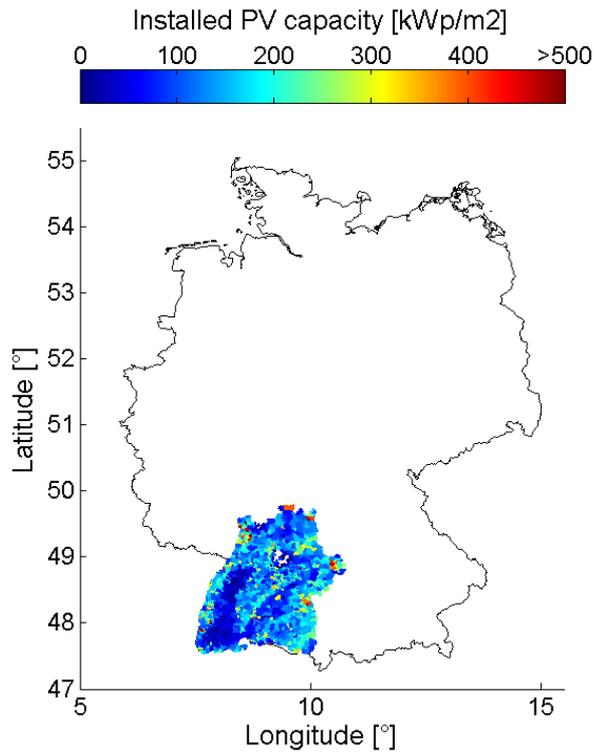
<b>Main characteristics of the 50 Hertz control area</b>		
<i>(Update time: 05/2016)</i>		
Area		109.360 km <sup>2</sup>
Time of the last input		20-Apr-2016
Number of plants	0-30kW <sub>p</sub>	118 296 (85.3%)
	30-100 kW <sub>p</sub>	12 910 (9.3%)
	100-1000 kW <sub>p</sub>	5923 (4.3%)
	>1000 kW <sub>p</sub>	1565 (1.1%)
	<b>Total</b>	<b>138 694</b> <b>(100%)</b>
Installed capacity	0-30kW <sub>p</sub>	983.2 MW <sub>p</sub> (11.0%)
	30-100 kW <sub>p</sub>	626.5 MW <sub>p</sub> (7.0%)
	100-1000 kW <sub>p</sub>	1 777.5 MW <sub>p</sub> (19.9%)
	>1000 kW <sub>p</sub>	5 533.2 MW <sub>p</sub> (62.0%)
	<b>Total</b>	<b>8920.4 MW<sub>p</sub></b> <b>(100%)</b>



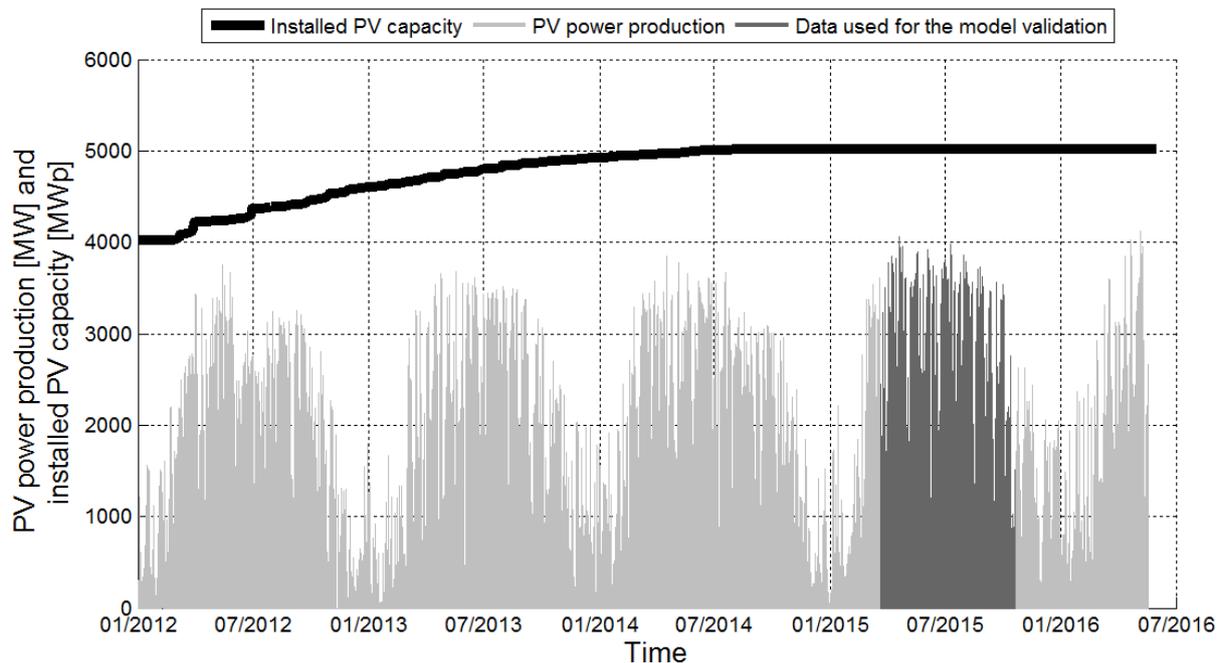


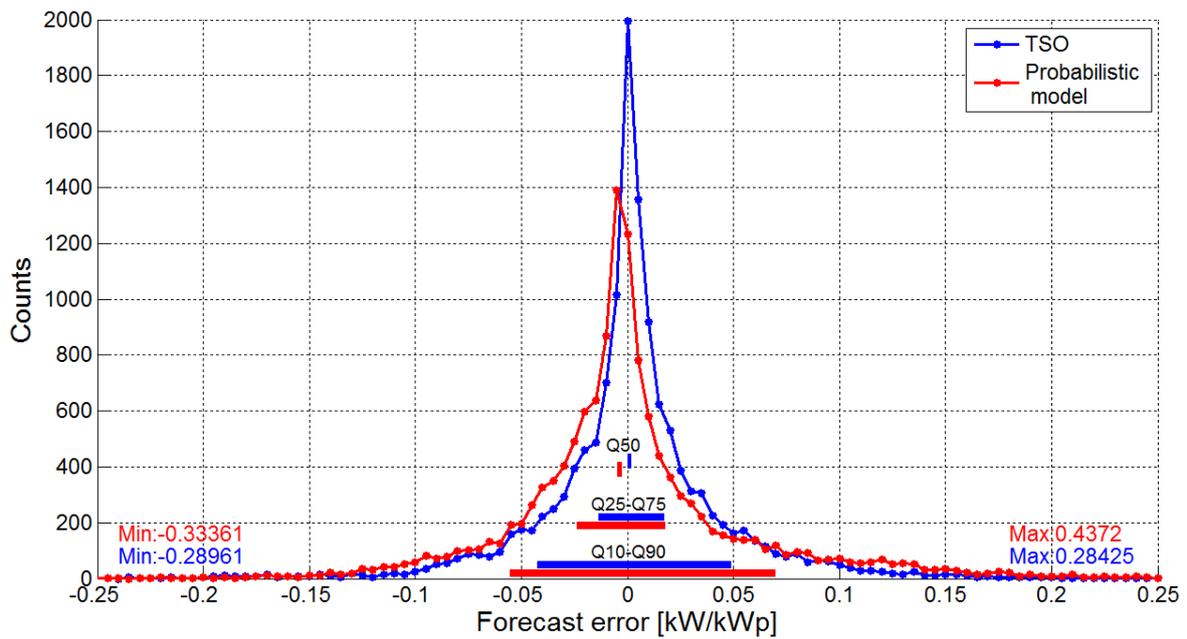
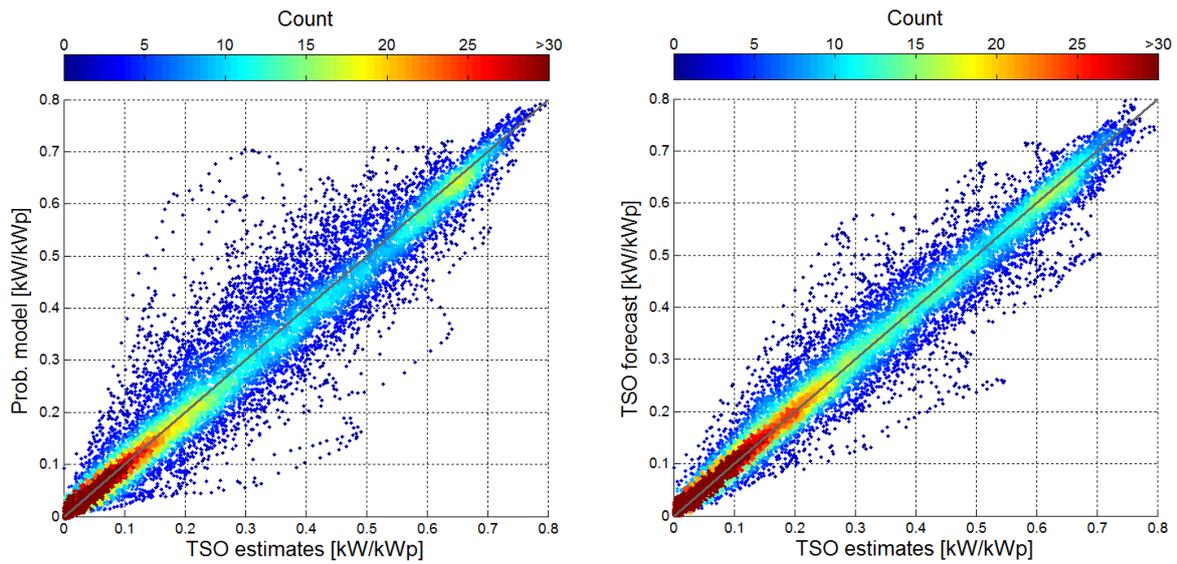
	<b>Probabilistic model</b>	<b>TSO forecast</b>
<i>Bias</i>	-0.47 %	-0.08 %
<i>MAE</i>	3.27 %	2.92 %
<i>RMSE</i>	4.87 %	4.51 %
<i>min / max</i> $\Delta$	-19.48 / 40.70 % 60.1852	-25.43 / 26.18 % 51.6139
<i>Q10% / Q90%</i> <i>IQ</i>	-5.72 / 5.14 % 10.86	-4.97 / 4.76 % 9.73
<i>Q25% / Q75%</i> <i>IQ</i>	-2.73 / 0.94 % 3.67	-1.72 / 1.76 % 3.48
<i>Median</i>	-0.44 %	-0.03 %
<i>Correlation coefficient</i>	0.96808	0.97212

## Annex 4 – Results for the TransNetBW control area



<b>Main characteristics of the TransNetBW control area</b> (Update time: 05/2016)		
Area		34.600 km <sup>2</sup>
Time of the last input		30-Sep-2014
Number of plants	0-30kW <sub>p</sub>	245 731 (87.7 %)
	30-100 kW <sub>p</sub>	29 345 (10.5 %)
	100-1000 kW <sub>p</sub>	4 816 (1.7 %)
	>1000 kW <sub>p</sub>	175 (0.1%)
	<b>Total</b>	<b>280 067</b> <b>(100 %)</b>
Installed capacity	0-30kW <sub>p</sub>	2 314.3 MW <sub>p</sub> (46.1 %)
	30-100 kW <sub>p</sub>	1 294.9 MW <sub>p</sub> (25.8 %)
	100-1000 kW <sub>p</sub>	988.9 MW <sub>p</sub> (19.7 %)
	>1000 kW <sub>p</sub>	422,5 MW <sub>p</sub> (8.4 %)
	<b>Total</b>	<b>5 020.6 MW<sub>p</sub></b> <b>(100 %)</b>





	<i>Probabilistic model</i>	<i>TSO forecast</i>
<i>Bias</i>	0.16 %	0.23 %
<i>MAE</i>	3.83 %	2.82 %
<i>RMSE</i>	6.13 %	4.54 %
<i>min / max</i> $\Delta$	-33.36 / 43.72 % 77.08 %	-28.96 / 28.42 % 57.39 %
<i>Q10% / Q90%</i> <i>IQ</i>	-5.55 / 6.98 % 12.53 %	-4.26 / 4.9 % 9.16 %
<i>Q25% / Q75%</i> <i>IQ</i>	-2.37 / 1.81 % 4.18 %	-1.36 / 1.74 % 3.1 %
<i>Median</i>	-0.35 %	0.12 %
<i>Correlation coefficient</i>	0.95954	0.97747