Renewable Energy 87 (2016) 1016–1022

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene

Development of an algorithm to analyze the yield of photovoltaic systems

H. te Heesen [a,*], V. Herbort [b]

[a] University of Applied Sciences Trier, Environmental Campus Birkenfeld, Campusallee, 55761, Birkenfeld, Germany
[b] University of Applied Sciences Ulm, Albert-Einstein-Allee 55, 89081, Ulm, Germany

ARTICLE INFO

Article history:
Received 31 March 2015
Received in revised form
9 July 2015
Accepted 22 July 2015
Available online 4 August 2015

Keywords:
Performance
Photovoltaic
Algorithm
Specific yield

ABSTRACT

The global photovoltaic market has developed very fast in the last years. In order to help PV system operators to decide whether their PV system is running well or not, we present a data cleansing algorithm to evaluate the specific energy yield of photovoltaic (PV) systems. The aim of our algorithm is to identify outlying yield and insolation values by using statistical methods and to provide a normal distribution of yield data after the final step. Therefore, PV systems are separated in major and subregions. The specific energy yield and insolation values of all systems in a region are analyzed. Outlying values are neglected in further steps. The relevant key figure – the skew of the yield data distribution – converges within four passes of data cleansing steps and the confidence interval of the specific yield reaches 95 % if at least 50 PV systems are evaluable in one region.

1. Introduction

In recent years, the global photovoltaic market has developed extraordinarily. At least 38.4 GWp of PV systems have been newly-installed in 2013. The cumulative installed capacity has reached 138.9 GW at the end of 2013. The global PV cumulative scenario of EPIA predicts a capacity between 321 GWp and 430 GWp up to 2018 [3].

There are two key figures to classify the quality and the performance of PV systems – the specific energy yield measured in kWh/kWp and the performance ratio (PR) [5]. The specific yield is the preferred ratio for the evaluation of the energy production as the PV system operator has typically easy access to the annual or even monthly yield of the PV plant by referring to the measurement of the energy meter or to the invoice of the energy utility.

Several groups have analyzed the performance of PV systems. On the one hand there are systematical approaches to assess the performance of PV plants and on the other hand there are evaluations of PV systems in specific regions or countries. Tsafarikis presents a data analysis method in which the energy yield is related to reference energy yield as a scatterplot using the mean square error [12]. Nordmann shows a statistical evaluation of the operation of PV systems in the report IEA PVPS Task 13 [9]. Reich provides a research on the performance ratio of PV systems [10]. A historical review of the performance of PV systems is shown by Woyte [16]. Country-specific surveys have been published e. g. for Germany [4], the Netherlands [6], France/Belgium [7], and India [11].

Some publications lack a proof that the data analysis provides statistical significant results. Also, no data cleansing has been discussed. If no data quality inspection is done, then the conclusion of the performance analysis might be questionable. In Germany, the mean annual specific energy yield of all PV systems for 2014 is 914 kWh/kWp (installed PV capacity 38.236 GWp, total energy from PV systems in Germany 34.93TWh (data from Ref. [1]). We will show in this publication that this specific yield does not represent the maximum performance of well-operated PV systems.

We will present an algorithm that provides a PV system performance analysis based on statistical methods. The specific energy yield will be the key figure for the evaluation of PV systems performance. Our aim is to calculate the technical potential yield of PV systems in case of continuous operation and maintenance. This will help PV system operators to decide whether their system is running well or might have minor or major technical malfunctions.

2. Methods

In this section, the data sources are described which were used

* Corresponding author.
E-mail addresses: h.teheesen@umwelt-campus.de (H. te Heesen), herbort@hs-ulm.de (V. Herbort).

http://dx.doi.org/10.1016/j.renene.2015.07.058
0960-1481/

to evaluate the yield data of photovoltaic systems. The operators of PV plants are responsible for the quality and maintenance of the measuring data of the monitoring system. However, the measurement data are often inadequately maintained, so that a large effort is necessary to identify misconfiguration, data errors, and malfunctions automatically. Only well-maintained solar systems shall be taken into account to analyze the yield.

2.1. Data basis

Starting point of the analysis are measured data of PV systems that are remotely monitored with the monitoring system of the company meteocontrol GmbH. Until December 31, 2014 in total 36,733 systems are monitored globally with a total peak power of 9.7 GWp. In Germany, 24,921 installations with a capacity of 5.5 GWp are monitored, representing a market share of about 14.3% (total PV capacity in Germany at the end of 2014 according to the German Federal Network Agency: 38.359 GWp). 24.6% of the monitored PV systems have an installed insolation sensor.

The central component is the data logger, which records the measuring data of the individual components and transmits the data via the internet to a central server. For this paper, the yield data, which are determined by the counter or from measured values of the inverters, as well as insolation information, are used. Irradiation sensors or pyranometers, mounted in the inclined module level, record the radiation data locally. If no radiation sensor is installed, then the solar energy data from satellite measurements of meteosat satellites is used [8]. The accuracy of insolation data from satellite measurements is comparable to weather stations [2]. In addition, the configuration of each PV system is known. Besides the postal code of the plants, the orientation and inclination of the solar modules are included in the analysis. The yield and insolation values are measured in 15-min intervals for each PV system. These data are aggregated for the evaluation to monthly values. Systems with a Southern orientation (orientation angle between 135° and 225°) and a similar inclination (module inclination angle between 20° and 40°) are selected as these PV systems show the highest yield in Germany.

2.2. Data quality and significance

The distribution of the specific energy yield and the insolation of the PV systems must be cleaned from implausible values. Also, systems with technical failures or data communication malfunctions must be identified automatically and neglected in the analysis. The aim of the algorithm will be to eliminate outliers automatically and thereby ensure a normal distribution of the yield. A suitable method for this is the use of the interquartile range [13].

Exploratory data analysis shows that there are PV systems returning implausible or no values. To allow a correct result, the monthly values of systems, which show extreme or even no values, are initially filtered. The median and the quartiles are relevant for the evaluation metrics. The median m is at the position of assorted data with 50% of the measured values smaller and 50% of the measured values larger than the median. The lower quartile q_l (or first quartile) is the measured value with 25% of the measured values smaller and 75% of the measured values larger than the first quartile. Accordingly, based on the upper quartile q_u (or third quartile) 75% of the measured values are smaller and 25% are larger than the third quartile. The distance between the first and the third quartile is called the interquartile range (IQR). Tukey defines values which lie outside of the range of ± 1.5 IQR from the lower resp. the upper quartile as outliers. This range σ is called the whisker distance.

$$x_l = q_l - 1.5 \cdot IQR \quad \text{lower threshold}, \tag{1}$$

$$x_u = q_u + 1.5 \cdot IQR \quad \text{upper threshold}, \tag{2}$$

$$\sigma = x_u - x_l \quad \text{whisker distance}. \tag{3}$$

If a normal distribution is given, the median m is equal to the mode and equal to the arithmetic mean $\bar{x}$

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \tag{4}$$

with the number of systems n and the specific yield x_i of system i. In order to obtain a data approximation corresponding to a normal distribution, the difference between the median and the arithmetic mean must be zero. Criterion for the quality of the algorithm is thus the deviation of the two values. Second criterion is the symmetry of the distribution which is represented by the skewness γ

$$\gamma = \frac{1}{n} \cdot \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s} \right)^3 \tag{5}$$

with the standard deviation s

$$s = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2}. \tag{6}$$

If the distribution is symmetric, then the mean $\bar{x}$ is equal to the median m and the distribution will have zero skewness γ.

The significance of the yield distribution is represented by the confidence interval $D(z)$

$$z^2 = \frac{n \cdot \varepsilon}{\sigma(1-\sigma)} \tag{7}$$

with the probability of the standard normal distribution z, the number of evaluated PV systems n, the tolerated error ε of the calculated yield from the true yield, and the whisker distance σ. Typically, the error ε of the yield is about 5% according to the measurement error of the PV components taken from different manufacturers' instructions.

3. Results

First, the data cleansing algorithm is introduced. Next, a data analysis is presented to show that the algorithm leads to an energy yield distribution which represents the yield of PV systems. This yield should be reached if the system is constructed and operated well.

3.1. Algorithm

The initial specific energy yield distribution is shown in Fig. 1.

The concept of the algorithm is shown in Fig. 2. Before starting the data cleansing, PV systems with implausible values are identified and deleted (Step 1). Implausible values for PV systems are negative yield or insolation values. Also, if the value is zero or if no value is given, then these PV systems are not taken into further consideration.

The first step of the algorithm is the separation of PV systems in major regions (see Fig. 3). This major region should have similar insolation and climate conditions. If the insolation differs to much within this region, then the distribution of the specific energy yield and the insolation will be too large and thus the algorithm will not

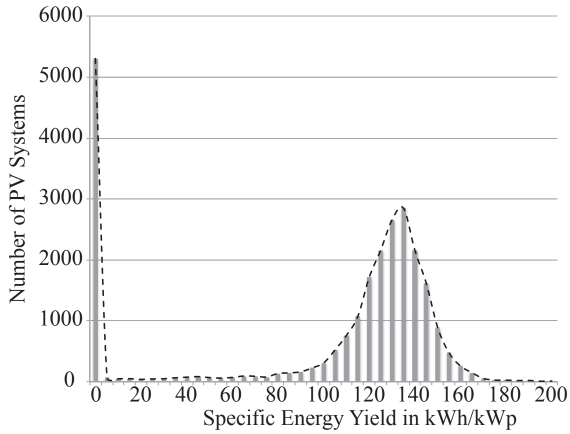

Fig. 1. Histogram of the specific energy yield of PV systems in Germany for July 2014. 5311 systems do not provide any yield information. The mean yield is 105,2 kWh/kWp, the median yield is 122.6 kWh/kWp and the mode (most often yield) is between 130 kWh/kWp and 135 kWh/kWp.

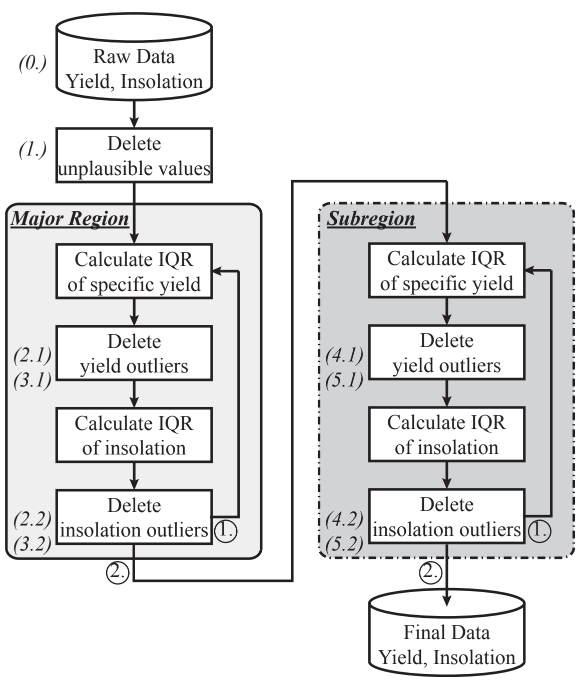

Fig. 2. Data cleansing algorithm to identify and delete values of PV systems with outlying yield and insolation values. First, the systems with implausible values (yield and insolation smaller or equal to zero) are deleted. Next, systems within a major region are selected. Median, lower and upper quartile and the interquartile range (IQR) of the specific energy yield are calculated for these systems. PV systems with outlying values (see Ref. [13]) are deleted. The same procedure is done for the insolation. These four steps are repeated once. Afterwards, the resulting PV systems are separated into smaller regions. The steps described for the major regions are repeated twice for the systems in the smaller regions. The remaining PV systems are used for the yield analysis.

be able to identify well-running PV systems automatically. In Germany, the first number of the postal code represents a major region with similar insolation conditions.

The key ratios (median, lower and upper quartile, interquartile range) of the specific energy yield of all PV systems in each major region (Step 2.1) are calculated. All systems, whose specific yield is within the threshold (larger or equal to x_l and smaller or equal to x_u) are selected for the next step. The PV systems, whose yield is

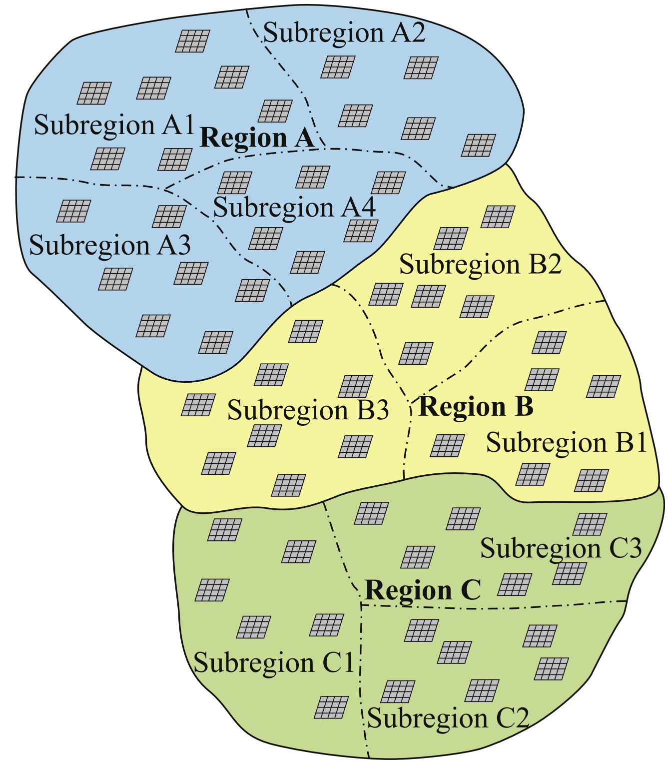

Fig. 3. Concept of spatial partition. The considered region is separated in three major regions A, B, and C with similar insolation and climate conditions. Each major region is split into several subregions.

lying outside this whisker distance σ, are deleted and not taken into further consideration (Step 2.2). Next, the same calculations and selections are performed for the insolation of the PV systems (steps 3.1 and 3.2). After the deletion of PV systems with outlying insolation values, steps 2.1 to 3.2 are repeated once. Finally, a list of PV systems for each major region is generated.

The major regions are separated in several subregions (see Fig. 3). The second part of the algorithm is executed for each subregion. The PV systems in each subregion and the adjacent subregions are selected. For instance, all systems in the subregions A1, A3, A4, and B3 are selected for the execution to calculate the key ratios for subregion A3. The data cleansing steps for each subregion are identical to the steps for the major regions. First, the IQR and the whisker range σ of the specific energy yield is calculated for the PV systems in the subregion and the neighboring subregions (Step 4.1). Then, the PV systems with outlying yield values within each subregion are deleted (Step 4.2). Next, the IQR and the whisker range σ of the insolation is calculated (Step 5.1) and the systems with outlying values are deleted (Step 5.2). As described above, these four steps 4.1 to 5.2 are repeated once for each subregion. The remaining systems for the subregions are stored in a data base for further analysis.

The consideration of PV systems from adjacent subregions is necessary to avoid the formation of "hotspot subregions". A hotspot region means that the yield in one subregion differs significantly from the neighboring subregions. The yield respectively the insolation from PV systems within the subregion and the adjoined subregions is calculated. All PV systems in the subregion, whose yield/insolation is lying outside the whisker range σ, are neglected in the next data cleansing step. It might happen in rare cases that the algorithm cannot select PV systems with reasonable values due to an artifact in the raw data. For instance if there is one system operator in a subregion being responsible for a couple of systems with a systematical misconfiguration then the algorithm cannot automatically identify this systematical problem in the raw data. This can lead to an incorrect selection of systems and thus to an over- or underestimation of the specific energy yield in this subregion.

3.2. Data analysis

The results of the algorithm is shown in this subsection. The specific energy yield, the insolation and the postal code is taken into consideration from PV systems in Germany for July 2014. The systems are separated into ten major regions (first digit of the postal code) and the associated subregions (see Fig. 4).

First, we show the results for the data cleansing algorithm with the 1-digit postal code as the major regions and the 2-digit postal code regions including the adjoined 2-digit postal code regions as the subregions. Fig. 5 represents the distribution of the total number of PV systems for each 2-digit postal code region after each step of the data cleansing algorithm. There are 260.4 PV systems in average in each subregion in the beginning. The mean number of PV systems drops to 187.0 PV systems per subregion after deleting the PV systems with no yield or insolation data. Finally, about 160 PV systems remain in average in each subregion after the final step of the algorithm. These PV systems can be used to evaluate the nationwide specific energy yield.

One key ratio to decide whether the distribution of yield data

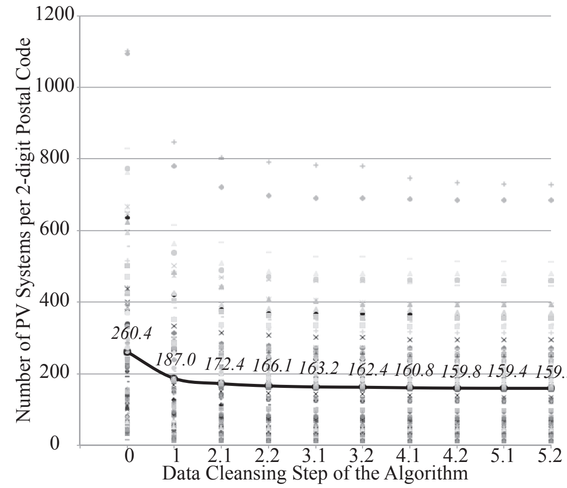

Fig. 5. Distribution of the number of PV systems per 2-digit postal code after each data cleansing step. The data points represent each 2-digit postal code region. The black line shows the mean number of PV systems per region. The data cleansing algorithm provides a stable number of evaluable PV systems in the regions. See Table 1 for data.

matches a normal distribution is the difference between the median and the mean of the specific energy yield. Fig. 6 shows this distribution of the deviation of median and mean for each 2-digit postal code region in Germany as well as the average deviation of all subregions. The deviation drops from 22.62 kWh/kWp (raw data) to −2.16 kWh/kWp (deletion of zero values). The deviation converges during the next steps of the algorithm to a constant level of about 1.03 kWh/kWp.

The skew of a distribution is the second ratio representing the deviation from a normal distribution. Fig. 7 shows the convergence of the skew after each data cleansing step. A wide spread of data points can be identified before and after the deletion of zero values. The deletion of PV systems with outlying values (steps 2.1 to 5.2 of the algorithm) reduces the skew significantly. Finally, there is a small skew of −0.37 after the final data cleansing step.

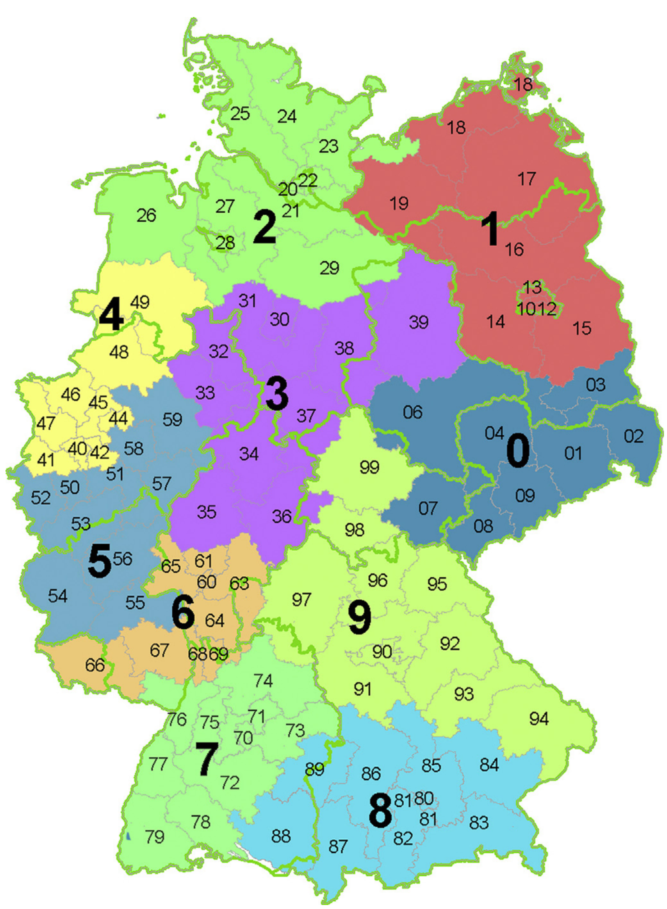

Fig. 4. Postal codes in Germany. The major postal code regions (first digit of the five-digit postal code) are accentuated in different colors. The green lines represent the borders of the states. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

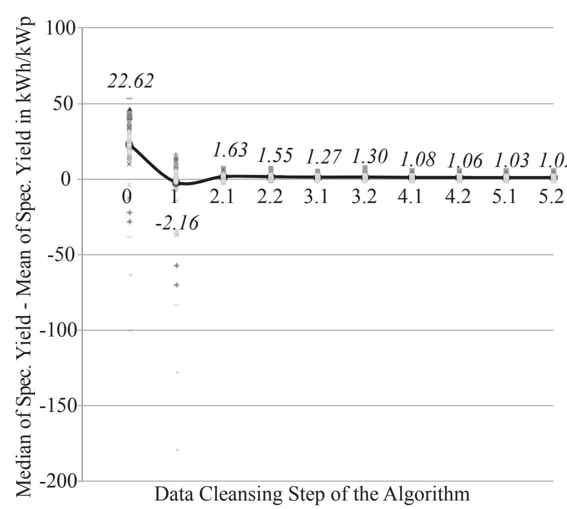

Fig. 6. Distribution of the difference between specific yield median and the mean specific yield in each 2-digit postal code region in Germany. The black line represents the mean deviation in all regions. The raw data (step 0) shows a large spread of the deviation. The difference of the median and the mean as a key ratio to represent a normal distribution drops significantly after the first step of the data cleansing algorithm (step 1). The next steps of the algorithm (step 2.1 to 5.2) provide a step-by-step decrease of the spread of values. See Table 1 for data.

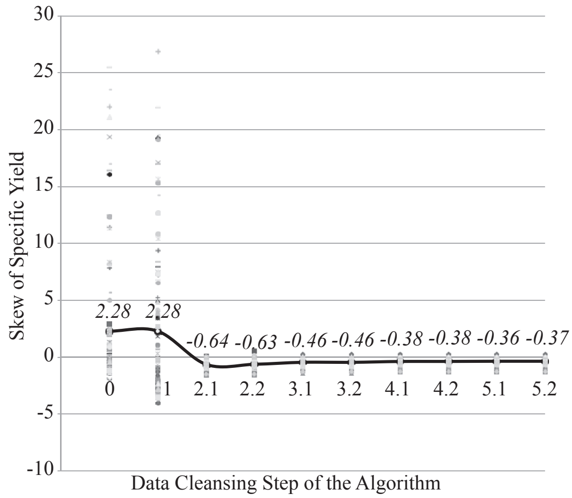

Fig. 7. Distribution of the skew of the specific energy yield in each 2-digit postal code region. The black line represents the mean skew in all 2-digit postal code regions. The skew represents the symmetry of the yield distribution and converges towards zero. See Table 1 for data.

The comparison of three different ways how to apply the data cleansing algorithm is shown in Fig. 8. The skew after each step according to the 1-digit postal code regions and the 2-digit postal code regions including the neighboring regions is shown in Fig. 7. If only major regions are considered in the algorithm (application of steps 4.1 to 5.2 for the major regions instead of subregions), then the convergence of the skew is smaller. Non-consideration of the neighboring regions in the steps 4.1 to 5.2 regarding the subregion analysis shows an irregular development of the skew. No convergence can be identified in this case although the absolute skew of the specific yield is smaller than the skew of the yield considering adjoined regions.

The problem of nonconsideration of neighboring subregions is, that subregions with “hotspot” values might appear. Tables 2 and 3 show exemplary a hotspot subregion (2-digit postal code 58). The upper tables show the results after the data cleansing without taking adjoined subregions into account, the lower tables represent the specific yield of the subregions under consideration of the neighboring subregions during the steps 4.1 to 5.2. There are minor differences of the median of the specific energy yield regarding the consideration of neighboring subregions. But the comparison of the relative deviation of the specific yield referring to the specific yield in subregion 58 shows that a hotspot region is formed if the neighboring regions are neglected. Consideration of the adjoined subregions leads to a more homogenous distribution of the specific yield.

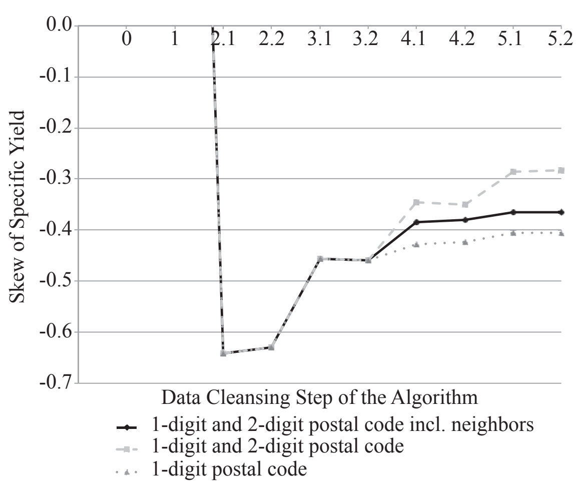

Fig. 8. Skew of the specific energy yield according to the application of the data cleansing algorithm. The skew according the 1-digit postal code (steps 2.1 to 3.2) and the 2-digit postal code including the neighboring regions (steps 4.1 to 5.2) is shown in diamonds and a solid black line. The light gray squares and the dashed line represent the skew of the specific yield without the consideration of neighboring regions during the steps 4.1 to 5.2. Finally, the dark gray triangles and the dotted line show the skew by applying the data cleansing algorithm just to the 1-digit postal code for four times.

Table 1
Data table to Figs. 5 to 7 for each data cleansing step of the algorithm. Fig. 5: Number of PV systems, Fig. 6: Median of specific yield − mean of specific yield, Fig. 7: Skew of specific yield.

Step	Fig. 5	Fig. 6	Fig. 7
0	260.4	22.62 kWh/kWp	2.28
1	187.0	−2.16 kWh/kWp	2.28
2.1	172.4	1.63 kWh/kWp	−0.64
2.2	166.1	1.55 kWh/kWp	−0.63
3.1	163.2	1.27 kWh/kWp	−0.46
3.2	162.4	1.30 kWh/kWp	−0.46
4.1	160.8	1.08 kWh/kWp	−0.38
4.2	159.8	1.06 kWh/kWp	−0.38
5.1	159.4	1.03 kWh/kWp	−0.37
5.2	159.3	1.03 kWh/kWp	−0.37

Table 2
Exemplary hotspot of the specific yield for the postal code region 58. All values are given in kWh/kWp. The 2-digit postal code is shown in italic. See Fig. 4 for the postal code regions. The values in the upper table show the median of the specific yield for the neighboring 2-digit postal codes after the application of the data cleansing algorithm without the consideration of the neighboring postal code regions during the steps 4.1 to 5.2. The lower table represents the median specific yield from the data cleansing algorithm including the neighboring regions during the steps 4.1 to 5.2.

45: 126.3	*44*: 130.3	
42: 129.7	***58*: 126.0**	*59*: 132.6
	51: 127.6	*57*: 129.0
45: 126.3	*44*: 129.2	
42: 129.7	***58*: 128.3**	*59*: 132.2
	51: 127.6	*57*: 129.0

Table 3
Relative deviations in percentage of the specific yield median from Table 1. The application of the algorithm with the consideration of the influence of neighboring regions leads to a hotspot region. The deviation of the median compared to the neighboring regions is significantly larger if the adjoining regions are not taken into account during the data cleansing algorithm.

45: 0.2%	*44*: 3.4%	
42: 2.9%	*58*:—	*59*: 5.2%
	51: 1.2%	*57*: 2.4%
45: 1.6%	*4*: 0.7%	
42: 1.1%	*58*:—	*59*: 3.2%
	51: 0.6%	*57*: 0.6%

4. Discussion

Figs. 7 and 8 show that the data cleansing algorithm leads to a normal distribution of the specific energy yield on the subregional level. The occurrence of hotspot regions (subregion with a specific yield which differs significantly from the yield of the neighboring regions) can be prevented. Fig. 9 compares the frequency distribution of all PV systems before applying the algorithm and after

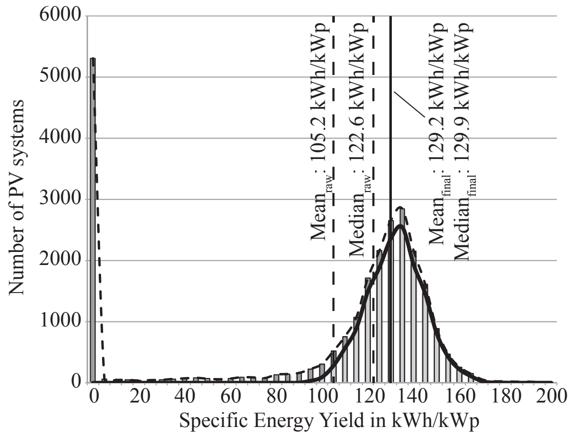

Fig. 9. Frequency distribution of the specific energy yield of the raw data (dark gray bars and dashed line) and after the final step of the data cleansing algorithm (light gray bars and solid line) for all systems in Germany. The PV systems with outlying yield values due to technical malfunctions are automatically identified and deleted by the algorithm. The final distribution corresponds to a Gaussian one – median and mean of the yield are nearly equal.

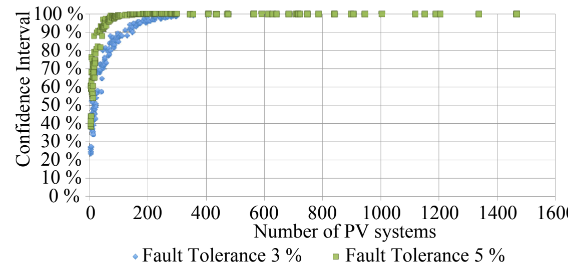

Fig. 11. Confidence interval of the specific energy yield accuracy according to the number of PV systems per subregion. If the fault tolerance for the yield evaluation is 5% (green squares), then the analysis of at least 50 PV systems located in the same region leads to a confidence interval of more than 95%. A smaller fault tolerance (3%) increases the number of PV systems (blue diamonds) which are necessary to achieve the wanted confidence interval. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

finishing the data cleansing process. The systems with outlying yield data are identified automatically and are deleted during the cleansing steps. Finally, the specific yield of the remaining PV systems corresponds to a Gaussian distribution. This distribution corresponds e. g. to yield analysis of other groups. Van Sark et al. have presented a Gaussian yield distribution for PV systems in the Netherlands [14,15].

The execution of the data cleansing algorithm to all PV systems in Germany for a calendar year from January to December provides an evaluable PV system density with a strong Southeast-Northwest divide (Fig. 10). The number of PV systems which remain after the algorithm run is larger in the Southern part of Germany due to a larger amount of total installed PV systems in these regions. The confidence interval which represents the accuracy of the specific energy yield after the final step of the algorithm is shown in Fig. 11. If the accuracy of the data should be at least 5%, then there must be at least 50 evaluable PV systems within a subregion to have a confidence level of more than 95%. The specific yield accuracy of 5% represents the typical error of measurement of the components. In particular, the inverters (as the central component providing energy production information) have a measuring inaccuracy of at least 5% taken from different manufacturers' instructions.

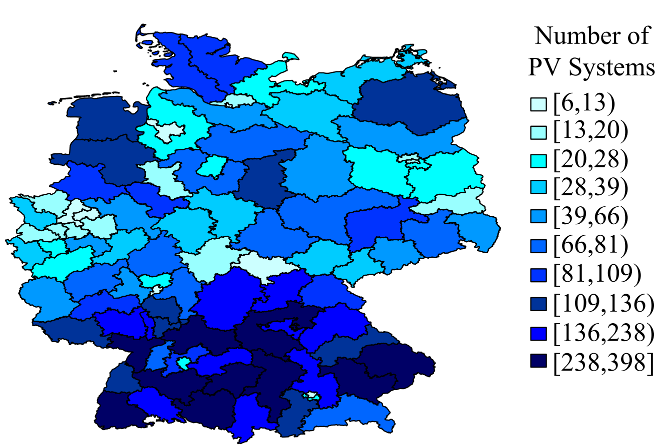

Fig. 10. Number of PV systems per 2-digit postal code in Germany after the final step of the data cleansing algorithm. There are more PV systems per region in the Southern part of Germany due to a higher PV system density in these regions.

The number of steps of the algorithm is sufficient to reach a high accuracy. The key figures (deviation of median/mean, skew of the data distribution) converge within the number of data cleansing steps. If more steps are applied, then the algorithm will not be able to identify more PV systems with outlying values as the outliers have already been deleted in previous steps and the data have already converged.

The key figure to rate the energy production of PV systems should not be the mean or the median of the specific yield in a subregion but the upper quartile. The median yield represents an adequate specific energy yield as 50% of the PV systems show a smaller yield and 50% show a larger yield. The consideration of the upper or 3rd quartile takes into account that there are still 25% PV systems, which reach a higher specific energy production. This means that PV systems within this subregion have the capability to achieve this specific energy yield if the PV system is planned, constructed and operated well. If the specific yield of a PV system is smaller than the upper quartile, then a technical reason can be identified why the PV systems do not reach the reference yield (e. g. temporary outage of a component, shadows on the PV modules due to surrounding objects, lower efficiency of the components than stated from the manufacturer).

5. Conclusion

A large amount of PV systems has been installed globally in the last years. We have presented a data cleansing algorithm, which is able to identify and select well-operated PV systems automatically. The grouping of PV systems to major regions in a first step and separation in subregions with similar insolation and climate conditions ensure reliable results. The key figures for the data evaluation of the algorithm converge within a couple of steps and show a confidence interval of 95% for at least 50 PV systems in a region with an accuracy of 5%.

Operators and owner of PV systems can rely on the yield data to evaluate the specific energy yield of their PV system in relation to well-running PV systems within the same region. If the yield of their PV system does not reach the calculated reference yield, then this can be a first hint that the PV system might have a technical problem or a malfunction.

The yield distribution of PV systems in Germany for 2013 is shown in Fig. 12 [4]. The specific energy yield shown in the map represents the upper quartile of the remaining PV systems after the data cleansing process.

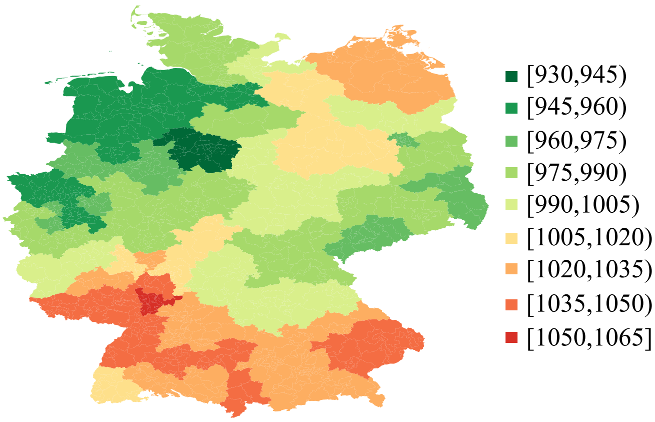

Fig. 12. Distribution of the specific energy yield in Germany in 2013. In Northwestern Germany PV systems have, due to lower solar radiation than the long-term average (2005–2012), lower energy yields than in previous years. PV systems located south of the line Saarbrücken-Frankfurt/Main-Regensburg achieve a specific annual yield of more than 1,020 kWh/kWp. The largest energy production can be observed in the Southwestern part of Germany (specific yield of more than 1,035 kWh/kWp) See Ref. [4] for details.

Acknowledgment

We thank the meteocontrol GmbH, Augsburg, Germany, for providing the monitoring data of PV systems.

References

[1] AGEE-Stat, Zeitreihen zur Entwicklung der erneuerbaren Energien in Deutschland, Federal Ministry for Economic Affairs and Energy, Germany, 2015.
[2] T. Behrendt, J. Kühnert, A. Hammer, E. Lorenz, J. Betcke, D. Heinemann, Spectrally resolved solar irradiance from satellite data to investigate the performance of thin film photovoltaics, in: Proc. 2010 EUMETSAT Meteorological Satellite Conference, Cordoba, 2010.
[3] European Photovoltaic Industry Association, Global Market Outlook for Photovoltaics 2014–2018, Renewable Energy House, Brussels, Belgium, 2014.
[4] H. te Heesen, V. Herbort, R. Pfatischer, Survey on yield of PV systems in Germany 2013, in: 29th European Photovoltaic Solar Energy Conference and Exhibition, 2014, pp. 2198–2201.
[5] E.N. IEC-CEI, Photovoltaic System Performance Monitoring-Guidelines for Measurement, Data Exchange and Analysis 61724, International Electrotechnical Commission, 1998.
[6] B. Kausika, W. Folkerts, G.J.H.M. van Sark, B. Siebenga, P. Hermans, A big data approach to the solar PV market: design and results of a pilot in the Netherlands, in: 29th European Photovoltaic Solar Energy Conference and Exhibition, 2014, pp. 4030–4033.
[7] J. Leloux, L. Narvarte, D. Trebosc, Performance analysis of 10,000 residential PV systems in France and Belgium, in: 26th European Photovoltaic Solar Energy Conference and Exhibition, 2001, pp. 3939–3946.
[8] E. Lorenz, J.W.H. Betcke, A. Drews, A.C. de Keizer, S. Stettler, M. Schneider, S. Bofinger, D. Heinemann, Intelligent Performance Check of PV System Operation Based on Satellite Data (PVSAT-2), final technical report. Tech. rep., Copernicus Institute, Utrecht University, 2004.
[9] T. Nordmann, L. Clavadetscher, W.G.J.H.M. van Sark, M. Green, Analysis of Long-term Performance of PV Systems, IEA PVPS Task 13, 2014. Subtask 1.
[10] N.H. Reich, B. Mueller, A. Armbruster, W.G.J.H.M. van Sark, K. Kiefer, C. Reise, Performance ratio revisited: is PR > 90% realistic? Prog. Photovoltaics Res. Appl. 20 (6) (2012) 717–726.
[11] V. Schacht, A. Häberle, W. Warmuth, I. Mitra, A. Kumar, A. Srivastava, PV performance benchmarking in India. Results from the project solar mapping and monitoring – SolMap, in: 29th European Photovoltaic Solar Energy Conference and Exhibition, 2014, pp. 2642–2645.
[12] O. Tsafarakis, W.G.J.H.M. van Sark, Development of a data analysis methodology to assess PV system performance, in: 29th European Photovoltaic Solar Energy Conference and Exhibition. Amsterdam, 2014, pp. 2908–2910.
[13] J.W. Tukey, Exploratory Data Analysis, vol. 18, Addison-Wesley, 1977.
[14] W. van Sark, L. Bosselaar, P. Gerrissen, K. Esmeijer, P. Moraitis, M. van den Donker, G. Emsbroek, et al., Update of the dutch pv specific yield for determination of pv contribution to renewable energy production: 25% more energy!, in: 29th European Photovoltaic Solar Energy Conference and Exhibition, 2014, pp. 4095–4097.
[15] W.G.J.H.M. van Sark, S. t Hart, M.M. de Jong, P. de Rijk, P. Moraitis, B. Kausika, H. van der Velde, Counting the sun – a dutch public awareness campaign on pv performance, in: 9th European Photovoltaic Solar Energy Conference and Exhibition, 2014, pp. 4161–4164.
[16] A. Woyte, M. Richter, d. Moser, S. Mau, N. Reich, U. Jahn, Monitoring of photovoltaic systems: good practices and systematic analysis, in: 28th European Photovoltaic Solar Energy Conference and Exhibition, 2013, 3696–3694.