Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 48 (2014) 1617 – 1626

Energy Procedia

SHC 2013, International Conference on Solar Heating and Cooling for Buildings and Industry
September 23-25, 2013, Freiburg, Germany

A new method for fusion of measured and model-derived solar radiation time-series

Theresa Mieslinger[a,b], Felix Ament[b], Kaushal Chhatbar[a], Richard Meyer[a]

[a]*Suntrace GmbH, Brandstwiete 46, 20457 Hamburg, Germany*
[b]*University of Hamburg, Meteorlogical Institute, 20146 Hamburg, Germany*

Abstract

When planning solar plants project developers and financers consider the long-term average of solar radiation and its uncertainty as the most important site selection criteria. Accurate time series of at least 10 years length are needed to calculate realistic estimates of such long-term averages. Ground-based measurements usually are not available for long time periods, but satellite-derived data may show significant differences compared to more accurate ground-measured values with respect to instantaneous values and frequency distribution. To overcome these shortcomings a new method for adaption of satellite-derived solar radiation values to ground measured time-series is developed. One year of overlapping time period is used for training the adaption. A weighted polynomial fit with additional constraints is then applied to the remaining satellite-derived solar radiation values. The method is tested at two sites for four satellite data models. The results show improvements for all data sets. After applying the adaption also during periods where no ground-based measurements are available the bias on average is nearing 0 % for most models. This requires that the satellite-derived data realistically represent the interannual variability. Frequency distributions are also matching better. Especially this is the case for maxima, which can be critical for design purposes.

Selection and peer review by the scientific conference committee of SHC 2013 under responsibility of PSE AG

Keywords: Measured solar irradiance; adaption of satellite-derived irradiation data; data fusion

Nomenclature

BSRN	Baseline Surface Radiation Network
CSP	Concentrating Solar Power
DLR	German Aerospace Center
DNI	Direct Normal Irradiance

Selection and peer review by the scientific conference committee of SHC 2013 under responsibility of PSE AG
doi:10.1016/j.egypro.2014.02.182

DOM	Difference of high radiation values
EHF	Department of Energy and Semiconductor Research
GMS	GeoModel Solar
HC3	HelioClim-3
KSI	Kolmogorov-Smirnov Integral
MB	Mean Bias
RMSD	Root Mean Square Deviation

1. Motivation for correction of satellite-derived solar radiation data

Solar radiation at the Earth's surface can be measured on the ground or derived by models, when the atmospheric composition in the region is known. The constituents of atmosphere can be inferred using satellites. Satellite-derived data have the advantage that they cover longer historical periods, which is required to reach reliable long-term averages. However, the quality of such satellite data sets can be poor – especially for the direct normal irradiance (DNI) component. Ground-based measurements with appropriate instruments can be of much higher quality, as long as the instruments are well calibrated and maintained. However, measurement stations are scarce and usually not available for a specific site and for longer time periods. due to financial means and the limited timeframe of the planning phase of solar plants. Therefore, it is common to use such measurements over shorter time periods for adaption of satellite-derived solar radiation data.

Objective of such corrections is to minimize the mean bias (MB) and improve the frequency distribution of the satellite-derived time-series. As a measure for the frequency distribution the Kolmogorov-Smirnov Integral (KSI) is commonly used. [1] and [2] developed methods trying to optimize both these parameters. The first method uses a defined combination of parameters to efficiently reduce bias and the root mean square deviation (RMSD) on average. The second method, a so called feature transformation, uses properties of the cumulative distribution function of high quality ground measurements and transfers those to the satellite-derived time series which causes the distribution functions to approach better. Recently, a study [3] reported that both procedures do not necessarily minimize the bias of the adapted model-derived data sets, wherein deviations up to 20 % are observed. As such a large bias is crucial for calculating energy yield of any solar power plant there is high demand for a new method overcoming the above-mentioned shortcomings.

Besides the main goal of minimizing MB, the second priority is minimizing the deviation between satellite-derived and ground-based instantaneous values. Especially high solar radiation differences should be reduced because they are important for the design of solar systems. The objective is to decrease the difference of the highest 1 % of values (DOM)

2. Description of solar radiation data

In this study, both ground-measured and satellite-derived solar radiation data are used at two sites. Satellite-derived data from various providers is used to investigate the performance of the adaption procedure. Two Baseline Surface Radiation Network (BSRN) stations are selected to guarantee accurate ground measurements. The site specifications are shown in Table 1.

Table 1. Site specifications

Site	Site code	Country	Latitude (°)	Longitude (°)	Altitude (*m*)
Plataforma Solar de Almería	ESPSA	Spain	37.09	-2.36	492
Tamanrasset	DZTAM	Algeria	22.78	5.51	1385

2.1. Ground-based measurements

The first selected site is the Plataforma Solar de Almeria, a research platform in southern Spain further referred to as ESPSA. The site is located in the Tabernas Desert in even terrain surrounded by mountains. Ground measurements are available from 2002 to the end of 2008. The time periods covered by all data sets are displayed in Fig. 1. Only complete years with a temporal resolution of 60 min are used in this study. The second test site with site code DZTAM is the BSRN station in Tamanrasset in southern Algeria. Data is available for 2001 until end of 2010. The site is located in the Saharan desert in a mountainous region.

2.2. Satellite-derived data sets

The German Aerospace Center (DLR) provides radiation data in the course of the Solar Energy Mining (SOLEMI) service [4]. In this study, DLR made solar radiation time-series available for the time period 1991 up to and including 2006 for ESPSA as well as for DZTAM. The data for both sites has a nominal spatial resolution of about 2.5 km and the half-hourly measurements are averaged to hourly values.

The Department of Energy and Semiconductor Research (EHF) at University of Oldenburg produce satellite-derived solar irradiance time-series and maps from Meteosat images since 1995. From 2005 onwards data from MSG is used as input to the Heliosat-2 method and the SOLIS clear sky model [5]. This combination leads to radiation data with a spatial resolution of 1 km^2 at sub satellite point and values produced every 15 minutes. Available data for this study covers the years 2005 to 2007 for the site ESPSA and 2005 to 2008 for DZTAM.

GeoModel Solar (GMS) uses satellite data from Meteosat and GOES, as well as outputs of atmospheric models like Global Forecast System to generate radiation data covering almost the whole earth. The high resolution Digital Elevation Model SRTM-3 retrieved from the Shuttle Radar Topography Mission, makes it possible to disaggregate the spatial resolution. The solar radiation data considered in this study has a resolution of 250 m and covers the years 2005 to 2010 for both sites. The model used for converting satellite images to radiation data is based on the Heliosat method. Great improvement is achieved by including daily aerosol and water vapor data, which are highly variable instead of climatological values [6]. The additional use of multispectral channels helps to treat snow covered and coastal regions more precisely.

The HelioClim databases supply solar radiation data for Europe, Africa and West Asia. The Center for Energy and Processes, Ecole des Mines de Paris/Armines, collects Meteosat satellite images and converts them to radiation data via the Heliosat-2 method. The HelioClim-3 database (HC3) provides data from February 2004 to present with a spatial resolution of approximately 5 km and a temporal resolution of 15 min. For this study, derived DNI values from 2005 to 2010 are available for both sites and all values are averaged hourly.

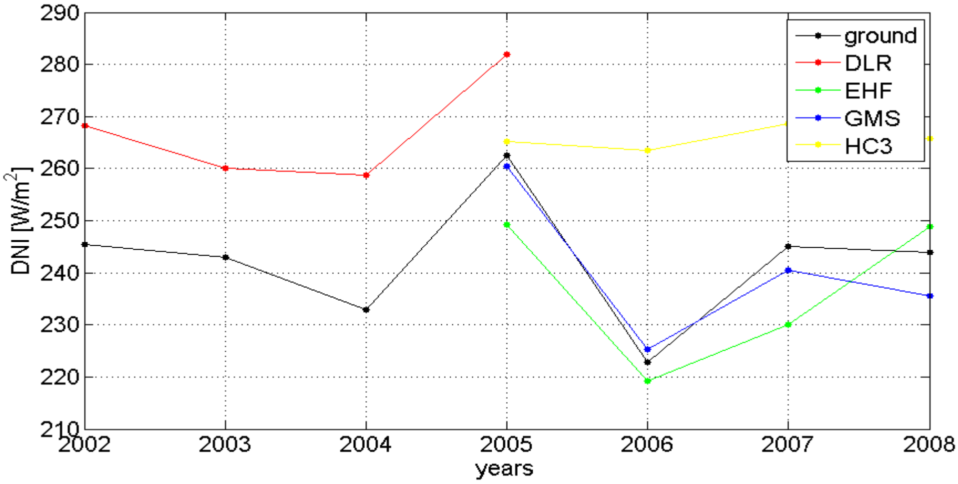

Fig. 1: Time periods of ground-based measurements and satellite-derived data sets used in this study.

3. Methodology

3.1. Characterization of input data

A first comparison of satellite derived and ground measured DNI values in Fig. 1 shows that there are significant annual differences in the data sets. Data from DLR for example has a systematic overestimation of roughly 10 % compared to the ground measurements. All satellite data sets have the same year-to-year variation as the ground-measured data. However, HC3 has a much weaker correlation compared to other satellite data sets.

On a more precise timescale two major effects causing differences in satellite and ground-based measurements of DNI can be detected. Firstly, a general over- or underestimation resulting from radiation transfer models used for calculating DNI values from satellite data. Input parameters to the models such as water vapour content are only measured as certain points and are interpolated to receive a worldwide pattern. Local topographic and atmospheric characteristics are only considered to a certain degree. However, those input parameters deviate systematically at a specific site.

Secondly, there can be a shift in the hourly values often caused by clouds. This so called parallax error is introduced when satellite observed cloud positions don't match the actual locations. In the following those random deviations as well as the systematically deviations are described by statistical means.

3.2. Statistical means to quantify deviations

A systematic deviation between two data sets can be expressed by the mean bias (MB) and relative mean bias (rMB) respectively:

$$MB = \frac{1}{N}\sum_{i=1}^{N} G_{sat,i} - G_{grd,i} \quad (1) \qquad\qquad rMB = \frac{MB}{\frac{1}{N}\sum_{i=1}^{N} G_{grd,i}} \quad (2)$$

$G_{sat,i}$ represents a single irradiation value from the satellite and $G_{grd,i}$ is the corresponding value measured from the ground. N gives the number of samples.

Another common measure of deviation is the root mean square deviation (RMSD) or relative root mean square deviation (rRMSD). It is the root of the mean squared difference of two data sets:

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(G_{sat,i} - G_{grd,i})^2} \quad (3) \qquad\qquad rRMSD = \frac{RMSD}{\frac{1}{N}\sum_{i=1}^{N} G_{grd,i}} \quad (4)$$

Deviation can also be expressed in terms of MB and the standard deviation of the difference (SDD).

$$RMSD = \sqrt{MB^2 + SDD^2} \quad (5)$$

While the systematic deviation of satellite data from ground-measured values is represented by the MB, the SDD parameter represents the additional random deviations. Those mainly result from parallax errors but also because it is difficult to estimate the vertical thickness of clouds from satellite data.

Furthermore, solar irradiation and efficiency of a solar thermal power plant have a non-linear correlation. The frequency distribution of solar irradiation determines upon type and size of the power plant. Radiation values up to 200 W/m^2 are of low significance because the start-up of a turbine and other auxiliary sources need more energy than that generated from the plant. In addition, the distribution of high radiation values defines the size of installed storage capacity.

Therefore, another measure is needed to quantify the difference of the frequency distributions in addition to the usual statistical parameters MB and RMSD. In this study, the Kolmogorov-Smirnov Integral (KSI) shall serve this purpose. It is a useful parameter to determine the similarity of two distributions and was first introduced by [7]. To calculate the KSI a critical value V_c depending on the number of samples N is defined for a 99 % level of significance. Following, a critical area A_c can be calculated representing the integral over V_c for a defined range x_{min} to x_{max}.

$$V_c = \frac{1.36}{\sqrt{N}}, \forall N \geq 35 \quad (6) \qquad\qquad A_c = \int_{x_{min}}^{x_{max}} V_c \, dx \quad (7)$$

The KSI is the integral over the absolute difference *D(x)* in the cumulative distribution functions divided by the critical area:

$$KSI = \frac{\int_{x_{min}}^{x_{max}} D(x)\, dx}{A_c} * 100 \quad (8)$$

3.3. Method for data adaption

In this chapter, a new approach for improving satellite-derived solar radiation time series with ground measurements is introduced. The last section showed that deviations in between two data sets are due to systematic and random errors. Latter are not conceivable, but the systematic part shall be quantified using a regression model. Beginning with the general form of a regression model, additional conditions shall be integrated into the model.

Data analysis showed that low intensities of radiation are often overestimated with simultaneous underestimation of high intensities and vice versa. Therefore, it is more reasonable to take a polynomial p of 3rd degree instead of a simple linear regression function. The corresponding function has the following form:

$$y(x) = p_1 x^3 + p_2 x^2 + p_3 x + p_4 \quad (9)$$

The vector $p(p_1,p_2,p_3,p_4)$ will modify the satellite time series in a way that it will better match with the ground measured 'real' radiation values. Before solving the equation more conditions are defined. These conditions result from the earlier defined properties of the two data sets. Night values are supposed to be zero. Moreover, negative radiation values do not exist. From this follows that p_4 must be zero.

Second condition is that the modified time series shall be free of bias or at least be within the average level of uncertainty of the measurements. From a statistical point of view ground measured radiation values are taken as the truth so that finally, the following equation can be applied:

$$\overline{y(x)} - \overline{y_{grd}(x)} = 0 \quad (10)$$

y_{grd} represents the ground measured values and the line above variables always defines the mean of the same.

Hence, the degrees of freedom of this function are reduced to two since $p_4 = 0$ and p_1 can be written as a function of p_2 and p_3. The cost function J shall be

$$J(p_2, p_3) = \frac{1}{2N} \sum_{i=1}^{N} (y - y_{grd})^2 \quad (11)$$

Minimizing J with respect to p_2 and p_3 results in a biasfree adaption algorithm where the squared difference of input values additionally causes major deviations, hence getting more intense corrected than minor ones.

Adapting satellite data with this algorithm already shows some improvement. However, high intensity values appear only at rare intervals and therefore have minor influence on the adapting function. An additional weighting function can help to put more weight on the high values. A simple linear weighting is sufficient and most convenient in this case because low radiation values are negligible and high ones also have higher impact on the total sum of irradiance as well as the energy output and they additionally define the storage size.

The following weighting function is used in this study.

$$w(x) = \frac{x}{max(x)} \quad (12)$$

The weighting is integrated into the derivative of the cost function. The maximum of ground-measured radiation values at most sites on Earth reaches about 1100 W/m^2.

4. Validation of the method at two sites

The adaption algorithm was tested at two sites, ESPSA and DZTAM, and for all satellite-derived data sets DLR, EHF, GMS and HC3. In the first step one year of overlapping data time series is used for training the algorithm. Taking a whole year is sufficient as all azimuth angles are covered and it therefore includes the usual seasonal variability [1]. In the second step, the calculated polynomial coefficients are applied to the respective remaining years, the validation period. In the following, one example for the adaption procedure shall be given before taking a look at the results from all test cases. In Fig. 2 satellite–derived DNI values from EHF at site ESPSA for the year 2005 are plotted against ground based measurements. It can be seen that the random scattering is very high causing a general high SDD. Applying the polynomial regression slightly changes the distribution, which can be better seen in Fig. 3 for the same data. The frequency distribution of the adapted time series approximates to ground based values for all ranges of values. The corresponding cumulative distribution function is very important for engineers when calculating the energy output of a solar power plant. The difference of the cumulative distribution functions is directly linked to the KSI.

In this study bias was set to be the main parameter defining a good data set. The results at ESPSA concerning relative bias for all training years and all satellite data providers are displayed in Fig. 4. All parameters discussed in the following refer to the 'testing period', i.e. the parameters are calculated from hourly data for all available years but the one used for training the algorithm. Concerning rMB, the adaption of DLR data to ground measurements leads to an improvement of 4.5 % on an average at site ESPSA and 2.4 % at DZTAM. Data from EHF and GMS

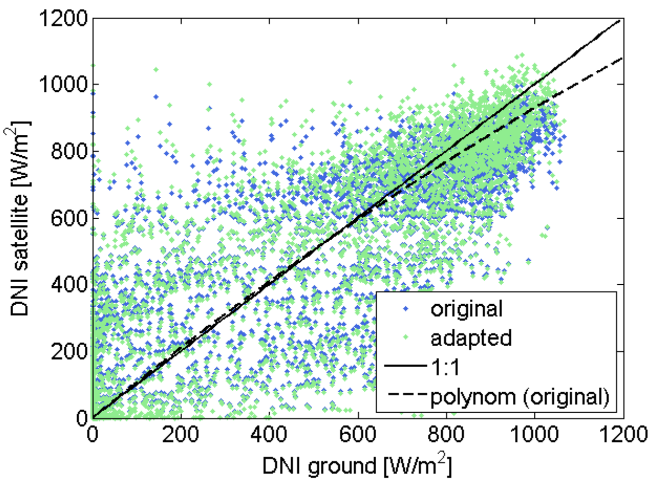

Fig. 2: Scatter plot of ground measured and original satellite-derived DNI values from EHF in blue and after applying the adaption algorithm in green. Data represents the year 2005 at site ESPSA. Only daytime values with corresponding sun elevation angles greater than zero are considered. The dashed line displays the polynomial applied to the satellite-derived data.

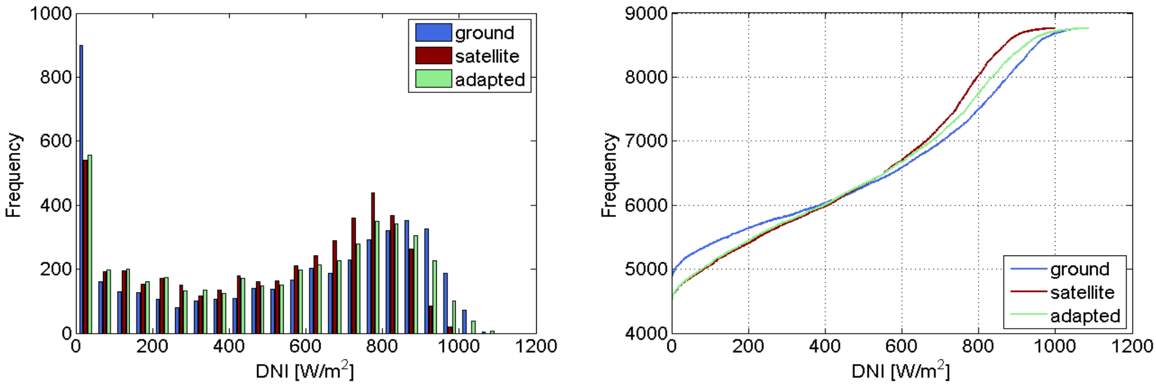

Fig. 3: Frequency distribution (left) and cumulative frequency distribution (right) of ground measured in blue, original satellite derived DNI values from EHF in red and after applying the adaption algorithm in green. Data represents daytime values from 2005 at site ESPSA.

both show similar good results at the two sites. The average adapted rMB is about 0 % over all data providers, although there are cases where the original relative bias is very high with up to -27 %. In several cases at site ESPSA the adaption leads to a marginal rMB. The reason for that could be that the original data sets from EHF and GMS already have a very low bias. Here, modifying the data can also lead to worse results. The data set HC3 was already detected to be the least fitting to the ground measurements. For the year 2006 at site ESPSA the adaption leads to a change in rMB from 11 % to -8 %. However, at site DZTAM an improvement in rMB from around 24 % to 1 – 9 % could be achieved on average over all years. The year 2001 from DLR data at DZTAM is the only year in which the adaption leads to worse results.

Averaged over all test cases the adaption leads to a significant reduction of rMB for the data sets from all providers. The standard deviation over all test cases though increases in general. A reason for this effect might be the fact that the characteristics from a single year are transferred to a longer time period. A larger sample might reduce this effect. In general, satellite-derived solar radiation data have inherent uncertainties. Ineichen, 2011 [8], states the uncertainties of satellite radiation products dependent on the provider and their model used, as well as on region of interest. Following Ineichen, the bias of DNI is around 2 % and the standard deviation of mean bias of 5 %. By using the adaption method presented in this paper, the overall uncertainty of the data can be reduced by minimizing the systematic part. By using various training years for the adaption and multiple satellite-derived solar radiation data at two sites it is found that the mean bias is 1.2 % and the standard deviation 2.3 %. However, leaving out the data set HC3 which was found to be of minor quality the mean bias even is reduced to 0.01 % with a standard deviation of 0.06 %. As a result, the overall uncertainty of the method using one year of overlap is found to be 2.6 % and without HC3 data even 0.07 %.

4.1. Improvements in SDD

As explained before, the total deviation of two data sets consists of a systematic part expressed by the bias and an additional random deviation given by the SDD. Data from HC3 improves concerning rSDD for every base year used during the adaption. The average reduction at site ESPSA is about 8 % and at DZTAM 13 %. At site ESPSA data from EHF and GMS vary slightly for the tested cases in the range of ± 1 %. Hence, on average rSDD does not change through the adaption. However, at site DZTAM both data sets improve when applying the adaption. For EHF, rSDD is reduced by 7 % on average, for GMS the improvement is about 1 %. Data from DLR at site ESPSA is reduced by approximately 2 % for each year, while rSDD at site DZTAM enlarges by the same amount. The RMSD parameter basically follows the SDD values because systematic deviations are small compared to the predominant random ones. Therefore, it is not explicitly discussed.

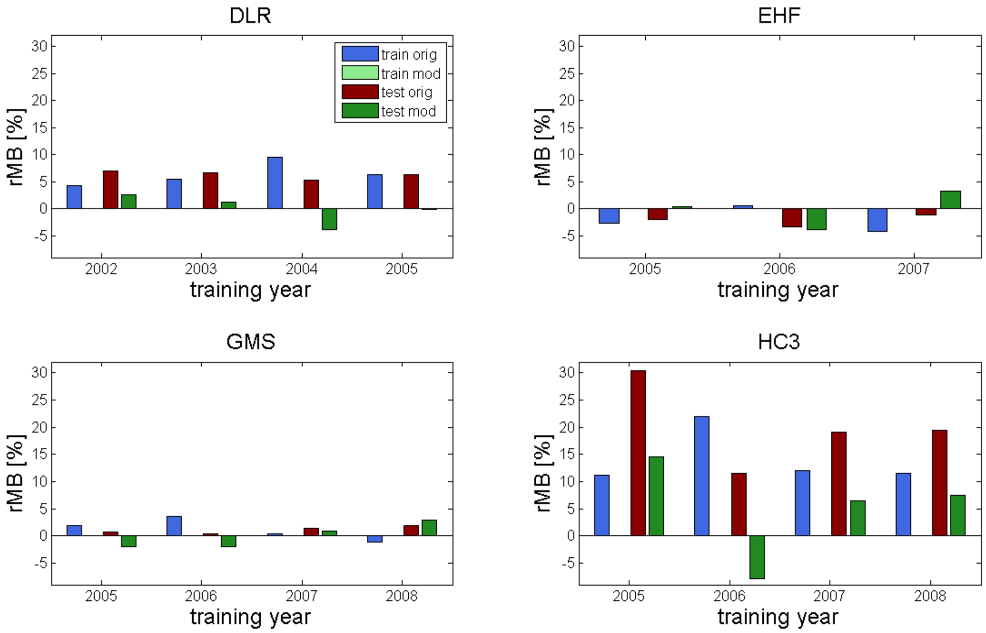

Fig. 4: Relative bias of original DNI values and modified ones each for training and testing period with different years used as training period. Satellite data from DLR, EHF, GMS and HC3 at site ESPSA are used.

4.2. Improvements in KSI

The KSI improved in all tested cases, between 30 % and 89 % on average . The only exception to the above statement is the year 2001 of DLR data at site DZTAM. However, this specific data set was already detected during the discussion of rMB, leading to unexpected results. For all other years KSI improves by approximately 32 % on average at both sites for data from DLR. Data from GMS and HC3 on average improve by around 75 % at site DZTAM. At site ESPSA KSI is reduced by 68 % for data from GMS and by 89 % for data from HC3. The KSI of data from EHF also significantly improves about 40 % on average at site ESPSA. Notably is the improvement of 380 % at site DZTAM. This shows that the adaption is able to minimize the difference in cumulative distribution functions of satellite-derived and ground-measured data even if the original satellite data is of minor quality.

4.3. Improvements in DOM

The difference of high DNI values expressed by the introduced parameter DOM varies strongly in between the years for the site ESPSA. On average high radiation values show less deviation at both sites for satellite data sets from EHF and GMS. The difference at site ESPSA is reduced by approximately 25 %. However, for data from GMS in two out of four cases the initial overestimation from original satellite-derived of ground-measured solar radiation data is changed to major underestimation. For data from DLR the original DOM is very low compared to other datasets. Here, it can happen that applying the adaption leads to worse results. For HC3 data at site DZTAM the adaption cannot improve the data, the DOM increases by up to 40 W/m^2 (~20 %). However, in total the adaption results in improved DOM on average for EHF and GMS data at both sites and data from DLR and HC3 can be each improved at one of the two sites.

5. Conclusions and outlook

A new method for adaption of satellite-derived solar radiation data to ground-measured data is developed in the course of this study. In addition, the method is tested at the two sites ESPSA and DZTAM and for the four satellite data providers DLR, EHF, GMS and HC3.

In most cases applying the developed adaption method to satellite-derived solar radiation data leads to an improvement, with the result that bias approximates zero. Especially for data sets with major deviations to corresponding ground measured values the algorithm works well. Approximately half of the modified time-series have a relative bias within the uncertainty of the reference instrument of ± 2 % which was set as a reference point for a good bias. By using various training years for the adaption and multiple satellite-derived solar radiation data at two sites it is found that the mean bias is 1.2 % and the standard deviation 2.3 %. However, leaving out the data set HC3, which was found to be of minor quality the mean bias even is reduced to 0.01 % with a standard deviation of 0.06 %. As a result, the overall uncertainty of the method is found to be 2.6 % and without HC3 data even 0.07 %.

However, the modification strongly depends on the base year used as input to the algorithm. A longer overlapping time period might help reducing the dependency on the base year and leads to more precise long-term estimate for solar radiation at a specific site. The HC3 data set does not well represent inter-annual varibaility. Therefore the adaption cannot work well and adapted HC3 still show major deviations for most parameters concerned. The bias improved by up to 19 % on average at site DZTAM but still could not approximate the goal of 0 %.

The calculations show that when forcing the bias to zero during the development of the adaption method, KSI of the adapted satellite-derived DNI values is also reduced significantly in all test cases. The improvement ranges from 30 % on average for DLR data to 89 % for HC3 data. The improvement of the standard deviation of the difference ranges from no change on average to 13 %. Due to the additional weighting function the difference of high radiation values can be reduced for data sets with major initial DOM. Around 70 % of the adapted radiation values meet the defined condition that the difference of the highest one percent of data shall be less than ± 5 % of the reference data set.

To validate the procedure, the developed adaption method should be applied to more sites to investigate the performance in different climate zones and different topographies. Even better results could be achieved by splitting the adaption based on one year of high quality measurements to seasonal dependent adaption, or depending on weather situations. Moreover, it could be interesting for solar power projects to test the possibility of using ground measurements from a nearby site for the adaption would save time and money when operating an own measurement station close or directly on site.

Acknowledgements

We thank BSRN for solar radiation data at Tammanrasset, Stefan Wilbert and Fabian Wolfertstetter from DLR-PSA for sharing the solar radiation measurements taken at PSA, which was partly funded by the BMU-founded project SESK (standardization of yield prognosis for solar thermal power plants; grant number 0325084), as well as the satellite-derived solar irradiance time-series data from DLR and the Energy Meteorology Group of the Carl von Ossietzky Universität Oldenburg. Thanks also go to GeoModel Solar for supplying satellite-derived time-series data. SolarGIS (c) 2013 GeoModel Solar and Dr. Etiene Wey from Transvalor and Prof. Dr. Lucien Wald from Mines ParisTech for supplying satellite-derived time-series data.

References

[1] K. Schumann, H. G. Beyer, K. Chhatbar, R. Meyer, "Improving satellite-derived solar resource analysis with parallel ground-based measurements.," ISES Solar World Congress, Kassel, Germany, 2011.

[2] J. R. Carow, "Anpassung langjähriger Satelliten-Strahlungszeitreihen an Bodenmesswerte eines Jahres," Diploma thesis, University of Applied Science Nordhausen, 2008.

[3] N. Killius and M. Schroedter-Homscheidt, "Data fusion of satellite and ground measurement time series," EU-Project Report ENDORSE D3.5 - Part B, Sep. 2012.

[4] Dumortier, D., Suri, M., & Hoyer-Click, C. (2009). Description of solar resource products, Summary of benchmarking results. MESOR.
[5] Hammer, A., Lorenz, E., & Heinemann, D. (2007). Fernerkundung der Solarstrahlung für Anwendungen in der Energietechnik. In DACH 2007. Presented at the DACH 2007, Hamburg, Germany.
[6] Cecebauer, T., & Suri, M. (2010). Accuracy improvements of satellite-derived solar resource based on GEMS re-analysis data. In Proceedings of SolarPACES 2010. Presented at the SolarPACES 2010, Perpignan, France.
[7] Espinar Bella, Lourdes Ramirez, Anja Drews, Hans Georg Beyer, Luis F. Zarzalejo, Jesús Polo, und Martin. „Analysis of Different Comparison Parameters Applied to Solar Radiation Data from Satellite and German Radiometric Stations". Solar Energy 83 (2009) (10. Juli 2008): 118–125.
[8] Ineichen, P. (2011). Five satellite products deriving beam and global irradiance validation on data from 23 ground stations. University of Geneva.