Early achievements towards an automatic assessment of the uncertainty in solar irradiation using web services

Lucien Wald¹, Philippe Blanc¹, Bella Espinar¹, Benoît Gschwind¹, Mireille Lefèvre¹, Lionel Ménard¹, Fabien Wald¹, Carsten Hoyer-Klick², Marion Schroedter-Homscheidt³, Claire Thomas⁴

Abstract
A long standing effort aims at creating and operating an automatic procedure for the assessment of the uncertainty in solar radiation data and the delivery of this information to users. This communication presents the early achievements. The overall objective can be seen as composed of three specific objectives: 1) assessing the uncertainty of the satellite-derived irradiation compared to a reference; 2) as the uncertainty depends on atmospheric conditions, devising a model that can predict uncertainty at any location and any time; 3) delivering information to users. As irradiation data are already delivered through the SoDa Service and the IEA SHC#36 / MESSOR portal, it is natural to use these dissemination media and to embed the uncertainty quantities in the data flow. The strategy is to implement the overall procedure by the means of interoperable Web services. Two ensembles of Web services are being developed: one for the assessment protocol, the other for the model of uncertainty.

1. Introduction
The solar radiation reaching the ground level on horizontal surfaces is of paramount importance in many applications, from climate to health (ESRA, 2000). The amount of power available on a surface of 1 m² integrated over the whole spectrum of the solar radiation is called surface solar irradiance (SSI). SSI is expressed in W/m². The SSI is an Essential Climate Variable as designed by the Global Climate Observing System in August 2010 (GCOS, 2010), meaning that it is a parameter of key importance for understanding and monitoring the global climate system. Meteorological networks measure the density of energy received on a horizontal plane at ground level during a certain period, e.g. 15 min, an hour or a day. This is the 15-min, hourly or daily irradiation, also called 15-min, hourly or daily solar exposure, and expressed in MJ/m² or J/cm². The hourly mean of SSI is derived from the hourly irradiation by dividing by the number of seconds in 1 h, i.e. 3600 s. By convention, the daily mean of SSI is computed from the daily irradiation by dividing it by the number of seconds in 24 h, i.e., 86400 s.

Accurate assessments of irradiation can now be drawn from satellite data (Blanc et al., 2011a). Stations at ground level measuring irradiation on the long-term (i.e. more than 10 years) are rare and satellites are an accurate way to complement or supplement them. There are several advantages in satellite-derived assessments of irradiation. They cover large areas, such as for example Europe, Africa and

¹ Center for Energy and Processes, MINES ParisTech, BP 207, 06904 Sophia Antipolis cedex, France
² German Aerospace Center DLR, Institute of Technical Thermodynamics, Pfaffenwaldring 38-40, 70569 Stuttgart, Germany
³ German Aerospace Center DLR, Münchner Strasse 20, 82234 Wessling, Germany
⁴ Transvalor, Parc de Haute Technologie Sophia Antipolis, 694, av. du Dr. Maurice Donat, 06255 Mougins Cedex, France
the Atlantic Ocean at one glance. Fig. 1 shows an example of the synoptic view offered by the Meteosat series of geostationary satellites. Such images clearly depict clouds and more generally the optical state of the atmosphere. These satellites take images every 15 min. It means that we have an assessment of the 15-min irradiation received by a pixel every 15 min, where a pixel corresponds to a ground surface of 3x3 km² in the subsatellite (or nadir) point. This makes a wealth of information that cannot be supplied by ground-based meteorological networks.

Figure 1
Example of an image taken by the satellite Meteosat on 11 November 2003, at 1100 UTC. Reflectance increases from black to white. Clouds are clearly visible. Copyright Eumetsat, 2003

Several initiatives were launched in order to create databases of irradiation by processing satellite images (Cros/Wald, 2004; Wald, 2006). Among them, is the SOLEMI project initiated by the German Aerospace Center DLR (Meyer et al., 2004). Another initiative is the HelioClim project launched by MINES ParisTech in 1997, to increase knowledge on irradiation and to offer irradiation values for any site and any instant over Europe, Africa and the Atlantic Ocean and long period of time, to a wide audience (Rigollier/Wald, 1999; Blanc et al., 2011a). The HelioClim-1 database offers daily irradiation for the period 1985–2005. It has been created from archives of images of the Meteosat First Generation satellites. The HelioClim-3 database represents a step forward regarding spatial and time resolution, improving assessment of uncertainty on SSI, and reducing time needed to access recent data. It exploits the enhanced capabilities of the series of satellites Meteosat Second Generation to deliver values of irradiation every 15 min with a spatial resolution of 3 km at nadir. To achieve this, MINES ParisTech has set up a routine operation for converting Meteosat data into 15-min irradiation on a near-real time basis. The SoDa Service (www.soda-is.com) was selected as the means to disseminate the HelioClim databases in order to benefit from the notoriety of this service. SoDa was initially a European project for the integration and exploitation of networked solar radiation databases (Wald et al., 2002). It was then established as an operational service in 2003 (Gschwind et al., 2005, 2006) and is now widely used by communities interested in solar radiation: more than 50 000 users in 2010. Data can be retrieved by users using a standard Internet browser. The high number of requests made to the HelioClim databases, approx. 2 millions in 2010, indicates clearly the high level of interest in irradiation data; the major demand comes from the energy sector.
2. A challenge: providing assessment of uncertainty

Concern about uncertainty is growing in environment. This concern includes radiation data as solar radiation plays a major role in environment. For example, errors in solar radiation may impact assessment of terrestrial or oceanic biomass. In another example, if erroneous data are used, wrong performances will be obtained in assessing the yield of a solar-powered system, leading to wrong decision in investment or energy policies as demonstrated in the ENVISOLAR project (Schroedter-Homscheidt et al., 2006). Energy producers are aware of this in the last years and have pushed data providers to make efforts to assess and document uncertainty in satellite-derived irradiation.

Supplying irradiation data together with associated uncertainty is a new challenge for data providers. Methods converting satellite data are fairly new, say 30 years, and till 2008, no common benchmark was available permitting to compare outcomes of these methods. A number of comparisons were made between results of methods and measurements by ground stations (Abdel Wahab et al., 2009; Blanc et al., 2011b; Cano et al., 1986; Diabaté/Wald, 1995; Ineichen et al., 2009; Lefèvre et al., 2007; Moradi et al., 2009; Njomo/Wald, 2006; Perez et al., 1997, 2002; Raschke et al., 1987; Rigollier et al., 2004; Ros sow/Zhang, 1995; Wang et al., 2011). It is not easy to compare the results of these comparisons one to each other. A first challenge lies in the quantities used to quantify the uncertainty. There are a number of differences in spatial and temporal resolutions which induce differences in results (Blanc et al., 2011a; England/Hunt, 1984; Pinker/Laszlo, 1991; Zelenka et al., 1999). The lack of ground-based measuring stations offering long time series of reliable data is another difficulty. It is fairly easy to perform comparisons in, e.g. France or Germany, where a large number of good quality measuring stations are operated; it is much more difficult to find good measurements in many other countries, and even impossible over oceans. The credibility of the uncertainty assessment increases as the number of stations used for its computation increases. Thus, the value of uncertainty assessment made with a limited number of stations is questionable, especially in areas far from these stations. Finally, on pure technical consideration, how uncertainty should be presented and delivered to users should be dealt with.

The authors are tackling these issues in a long standing effort. They are working together aiming at creating and operating an automatic procedure for the assessment of the uncertainty and the delivery of this information to users. The efforts have begun in 2004 during the Envisolar project co-funded by the European Space Agency (Schroedter-Homscheidt et al., 2006). They benefit from other international projects such as the MESOR, MACC and ENDORSE projects co-funded by the European Commission, the Task SHC#36 of the International Energy Agency (IEA) (Stackhouse et al., 2006), and the GEOSS (Global Earth Observation System of Systems).

The overall objective can be seen as composed of three specific objectives: 1) assessing the uncertainty of the satellite-derived irradiation compared to a reference; 2) as the uncertainty depends on atmospheric conditions, devising a model that can predict uncertainty at any location and any time; 3) delivering information to users.

3. Defining a reference and a protocol for uncertainty assessment

The first specific objective is to assess the uncertainty of the satellite-derived irradiation compared to a reference. Discussions were held in the working groups of the IEA SHC#36 and MESOR projects. It was established that the reference should be made of an ad-hoc series of ground-based measurements of recognised and checked quality. The typical accuracy of irradiation measured in the global meteorological network is 3% to 5% in terms of root mean square error, provided that irradiation data has passed some quality check procedures. Therefore, under these conditions, the ground measurements can be seen as an accurate reference against which one may compare the irradiation derived from satellite.

The comparison is made by computing the difference between the two sets of measurements and analysing statistical quantities, such as mean bias error, standard-deviation of error, root-mean square error (RMSE), and correlation coefficient. The choice of these quantities results from consultation of practitioners in ENVISOLAR and MESOR projects (Hoyer-Klick et al., 2008) and is dictated by the
knowledge of these practitioners about uncertainty and statistical quantities. These quantities are actually those recommended by the ISO (1995). Other quantities were elaborated relating to agreement between the statistical distributions such as Kolmogorov-Smirnov test integral (KSI) proposed by Espinar et al. (2009); there are not adopted presently because of the lack of understanding by practitioners.

The protocol for uncertainty assessment is as follows. Firstly, select good quality data from ground-based stations with ad-hoc quality control procedures. Then, perform the comparison between these data and coincident satellite-derived irradiation. Compute and provide the mean irradiation in ground-based observations, number of observations, correlation coefficient, mean bias error, standard-deviation of error, RMSE, and the relative value, expressed in per cent, of the last three quantities to the mean irradiation. Drawing graphs, such as correlograms or scatterplots (fig. 2), is recommended.

Optionally, these operations should be performed for irradiation as well as for the clearness index, which is the ratio of the irradiation observed at ground level to the irradiation observed at the top of the atmosphere. Another option is to apply the protocol for selected cases, such as classes of solar zenithal angles or cloud coverage (e.g. clear-sky or overcast conditions). A third option is to perform the above-described operations for the different time scales (e.g. hourly, daily, monthly or yearly time scales) available within the data sets depending on the finer time scale, data missing and the duration of coincident periods.

The result of this benchmarking protocol describes the uncertainty observed for the few selected measuring stations during a limited period of time. This protocol permits to qualify in a similar manner, similar products from different suppliers and is a useful tool to select the most appropriate product for a specific application.

Several limitations exist that make the assessment of the quality of retrieved irradiances a difficult task. Blanc et al. (2011b) discuss the quality of the ground measurements that can be obtained from the World Radiation Data Center (WRDC) and their spatial and temporal scarcity. A severe limitation is due to the large differences in principles of measurements. Single point and temporally integrated data (ground measurements) are compared to spatially integrated and instantaneous data (satellite estimates). An assumption of ergodicity (e.g. here equivalence between the temporal and spatial averages)

![Figure 2](image.png)

Example of a correlogram (or scatterplot) between hourly irradiation measured by the meteorological station of Nice (France) and that derived from satellite images (HC-3).
is usually made. This assumption is correct only if the field is spatially homogeneous at the considered
time scale over an area much larger than a pixel. This is generally false when a significant physi-
ographic feature is present. Other local effects such as reflections on the surrounding slopes or the
shadows of clouds may add to the difficulty in comparison. Accordingly, an intrinsic discrepancy is
expected because of the natural variability of irradiation in space.

Zelenka et al. (1999) have observed the local variability of the solar radiation using measurements
made by well-calibrated ground stations close to each other. They demonstrate that this variability
cannot be ignored. Expressing it as the ratio of the variance relative to the mean value over the area,
they found typical variability of 17% in hourly irradiation for an area of 10 km in radius. This means
that within a 10 x 10 km² area, irradiation measured by a series of similar inter-calibrated sensors
would exhibit the same mean value but would differ from hour-to-hour, with a relative variance equal
to 17%. Therefore, observing a difference hour-to-hour of 17% between a single pyranometer located
in a pixel cannot mean that the satellite-derived irradiation is of bad quality. These authors established
that a substantial part of the difference between satellite-derived data and ground-based measurements
comes from three different sources. One is the natural variability of the irradiation as discussed above.
The second one is the error in measurements, which may amount from 3% to 5%. Finally, the hypo-
theses of ergodicity and spatial homogeneity of the irradiation may induce an additional error of
3% of the mean irradiation.

4. Predicting uncertainty at any location and any time

The application of the protocol to several cases and databases in the MESOR and IEA SHC#36 pro-
jects, among others, reveals that the uncertainty depends on atmospheric conditions and other envi-
ronmental variables. Uncertainty cannot be described by a set of quantities that are constant in time
and space, and that could be derived from one or more benchmarks. Consequently, a model should be
devised that can predict uncertainty at any location and any time. This is the second specific objective.

Previous benchmarks reveal that the most important variables explaining the uncertainty are sun ze-
nithal angle, clearness index, and irradiation itself. All these variables can be known in a determinis-
tic way or already estimated and therefore can be used as inputs to a model predicting uncertainty. In case
of daily irradiation, there are only two variables: clearness index, and irradiation itself.

To our knowledge, there is no such published model. MINES ParisTech made a first attempt in
2006 for the HelioClim-1 database for daily irradiation. Based on results of multiple applications of
the benchmarking protocol, we established an empirical analytical model.

Let $K_c$ be the clear-sky index. It is the ratio of the irradiation $I$ at ground level to the irradiation $I_c$
that should be observed under a clear sky for this day and location. The model ESRA (Rigollier et al.
2000) is one of the models that can be used to predict $I_c$. Then, we define a systematic error (0.05) and
a random part in $K_c$:

$$\alpha = 0.05 + 0.04 / \sqrt{n_{days}}$$

where $n_{days}$ is the number of days used to compute the daily irradiation. In case of monthly means
of daily irradiation, e.g., $n_{days}$ will be equal to 30.

We define $r$ as the number of hourly values used to compute the daily irradiation divided by the
number of possible hours in the duration of this particular day. The error in $K_c$ is:

$$\text{error}_{K_c} = \alpha \exp[\ln(2) (1 - K_c)] \text{ if } r > 0.5$$

$$\text{error}_{K_c} = \alpha \exp[\ln(2) (1 - K_c)] \exp[1.54 (0.5 - r)] \text{ otherwise}$$

The error in daily irradiation is then:

$$\text{error} = I_c \text{ error}_{K_c}$$

The outcome error of this model was considered as equivalent to the RMSE. It was delivered to-
gether with the daily irradiation to the users on the SoDa Service. It helps creating awareness on un-
certainty among users and helps researchers to gain expertise in this domain.

Though the approach is the same, we have elaborated a more precise model for the HelioClim-3 da-
tabase taking into account comments from users. In particular, we have decided to deliver lower and
upper bounds in irradiation, such that there is a 68% chance that the actual value is comprised between
them.

We have selected 29 stations offering time series of hourly irradiation of good quality during the pe-
riod 2004-2005. These stations are mostly located in Europe with one in South Africa, one in South
Algeria and one in Israel. By using the protocol, we have estimated the differences for all stations
merged together. Prior to this, the HelioClim-3 estimates were corrected in elevation using the model
of Abdel Wahab et al. (2009) to adjust to the elevation above sea level of the ground stations.

In case of daily irradiation, we have created classes in clearness index, and HelioClim-3 daily mean
of irradiance. For each class, we have computed bias (mean value of the difference), standard-deviation,
and RMSE. All these quantities were expressed relatively to the HelioClim-3 mean irradi-
ance for this class.

Thus, we obtained abaci whose inputs are clearness index, and HelioClim-3 daily mean of irradi-
ance. There are 6 bins in daily mean of irradiance, in W/m²: [0, 100, 200, 300, 400, 500, 600], and 8
bins in clearness index: [0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 1.1]. Each class contain relative bias: $bias$
standard-deviation: $std$, and RMSE: $rmse$. The computation of the lower and upper bounds $Elow$ and
$Eup$ for the daily mean of irradiance $E$ is as follows:

\[ \frac{Elow}{E} = 1 + bias - std / \sqrt{ndays} \]
\[ \frac{Eup}{E} = 1 + bias + std / \sqrt{nday} \]

if $(Elow / E) > 0.9$, or $(Eup / E) < 1.1$, then
\[ \frac{Elow}{E} = 1 + bias - RMSE / \sqrt{ndays} \]
\[ \frac{Eup}{E} = 1 + bias + RMSE / \sqrt{nday} \]

The case of irradiation for period less than a day, e.g., hour, 30 min, or 15 min is fairly similar, ex-
cept that we compute abaci whose inputs are sun zenithal angle, clearness index, and HelioClim-3
hourly (or 30-min, or 15-min) mean of irradiance. There are 5 bins in solar zenithal angle, in degree:
[0, 30, 40, 50, 60, 90], 6 bins in hourly mean of irradiance, in W/m²: [0, 200, 400, 600, 800, 1000,
1400], and 8 bins in clearness index: [0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 1.1]. Each class contain relative bias,
standard-deviation and RMSE.

Let $nmin$ denotes the number of minutes in the period of integration. $nmin$ is equal to 60 for hourly
irradiation, and 15 for 15-min irradiation. The computation of the lower and upper bounds $Elow$ and
$Eup$ for the hourly (or 30-min, or 15-min) mean of irradiance $E$ is as follows:

\[ \frac{Elow}{E} = \max(0, 1 + bias - std * \sqrt{60/nmin}) \]
\[ \frac{Eup}{E} = 1 + bias + std * \sqrt{60/nmin} \]

if $(|Elow + Eup/2 - E|) > 0.8 \ (Eup – Elow)$, then
\[ \frac{Elow}{E} = 1 + bias - RMSE * \sqrt{60/nmin} \]
\[ \frac{Eup}{E} = 1 + bias + RMSE * \sqrt{60/nmin} \]

These two quantities are delivered routinely together with the irradiation to the users on the SoDa
Service. It brings benefit to companies monitoring the electricity produced by solar-powered systems,
since after transformation by a model simulating the system of energy conversion it provides a corre-
sponding plausible range of energy to be produced. Any outlier may mean a problem in the system.
The model has been validated by further benchmarks, using other stations in other areas. It reveals to
be quite robust though it has been noted that it overestimates the uncertainty in several cases.

5. Delivering information to users

The third objective is to deliver information to users. The easiest way to do so is to present the uncer-
tainty values on the same Web page containing the irradiation. This is what was made at the earliest
stages of our project in the SoDa Service in 2006. Much more can be done today, considering the
needs of users to have their computers interacting with the computers providing access to data and the
current capabilities in interoperability. We do not discuss here the quantities to be provided that relate
to uncertainty or their format. As written before, choice of quantities was made as a function of the
knowledge and expectations of practitioners.
We see the delivery of information on uncertainty as a series of challenges, dealing on the one hand with the description of uncertainty and its coding, and on the other hand, with Web services capable of computing uncertainty, thus implementing the workflow by their composition (Gschwind et al., 2007b).

As irradiation data are already delivered through the SoDa Service and the IEA SHC#36 / MESOR portal, it is natural to use these dissemination means. The uncertainty quantities are embedded in the flow containing irradiation.

Several services providing irradiation and uncertainty were implemented in the SoDa Service using the proprietary SoDa XML presented at Envirolinfo in 2002 (Wald et al., 2002). In this case, there is no specific metadata devoted to the description of uncertainty. The SoDa XML has the advantage of being simple, thus allowing rapid development but is limited regarding interoperability and does not obey current standards promoted by GEOSS (Percivall et al., 2011; Robinson et al., 2009).

More advanced services providing the same information were implemented using WSDL; a revision of the XML schema “SolarResourceKnowledge-2.1.xsd” (Gschwind et al., 2007a) was made to include uncertainty quantities in the product delivered to users. For each instant of observation, e.g., every hour, an irradiation is provided together with the uncertainty attached to this irradiance, as well as a reliability code. The uncertainty can be expressed as a single value, e.g., standard-deviation, or as lower and upper bounds as seen above. These bounds are such that there is a 68% chance that the actual value is comprised between lower and upper bounds. The meaning of the reliability code depends on the type of products. This code denotes whether the irradiation results from an interpolation in time of two adjacent values (for example, the case of a gap during a day), or provides the number of valid 15-min values used to compute a daily average, or the number of valid daily values used to compute a monthly average. This is fully described in the exploitation metadata conveyed by the product. This schema is presently used to disseminate HelioClim and SOLEMI data through the MESOR portal.

Though operational, all these services must be considered as prototypes. They do not permit to achieve the third specific objective “delivering information to users” and further work is needed.

One of our goals is to create Web services permitting practitioners to upload their legacy data, or use archived data measured by ground-based stations in order to perform the protocol. In this way, we expect to increase the number of tests of performances made on the HelioClim and SOLEMI databases by users, therefore increasing the knowledge on the merits of these databases and the sharing of this knowledge. Another series of Web services should be developed implementing the model predicting the accuracy. Actually, this implementation will be part of the development of the Web services providing the irradiation from HelioClim or SOLEMI databases.

6. Conclusions and perspectives

We have presented here early achievements reached in this long-standing effort for the assessment of uncertainty of solar irradiation data estimated by the processing of satellite images and other Earth Observation data.

International efforts have resulted in a protocol for benchmarking such irradiation data. This was the first objective to attain. The reference is composed of good quality measurements made at ground level by qualified measuring stations. Quantities describing uncertainty are quite standard: bias, standard-deviation, RMSE. Similar irradiation data from different suppliers can be qualified in a similar manner by this protocol. It is a useful tool for users to select the most appropriate database for a specific application.

It is impossible to describe the uncertainty in solar irradiation with constant values in time nor in space, whether absolute or relative. The uncertainty is a function of atmospheric conditions and the irradiation itself, thus is changing rapidly. This communication has presented two models elaborated for the HelioClim databases that predict the uncertainty for any site and any instant. These models have proved to be robust. We may consider that we have reached the second specific objective. However, efforts should be made to better estimate the uncertainty. We believe that this objective can only be reached by improving the method estimating the irradiation from satellite images.
Delivering information on uncertainty to users is our current challenge. Interoperability is a key element in our strategy. Web services will be chained in order to obtain value-added Web services implementing the benchmarking and the uncertainty model. Interoperability is also a means to exploit new or more accurate services, replacing older ones. The Web services will be deployed in the Energy Community Portal (www.webservice-energy.org). The GEOSS-compliant CSW (catalogue service for the web) provided by the EC-funded project EnerGEO (energeo.researchstudio.at) will enable these services to be searched and discovered.

These activities are taking place under the GEOSS umbrella, thus contributing to the Architecture and Data Committee and to Community of Practice “energy”. A scenario will be developed in Phase-4 of the GEOSS Architecture Implementation Pilot (AIP-4) exploiting several services where data flow conveys uncertainty information. This research partly contributes to EC-co-funded projects MACC (2009-2011) and ENDORSE (2011-2013) and to IEA activities in SHC#46 (2011-2016). Cooperation with other projects such as the QA4EO project or EC-co-funded project UncertWeb (2010-2012) is a perspective.

Acknowledgements

The research leading to these results has partly received funding from the European Union’s Seventh Framework Programme (FP7/2007-2013) under Grant Agreements no 218793 (MACC project) and no. 262892 (ENDORSE project).

Bibliography


