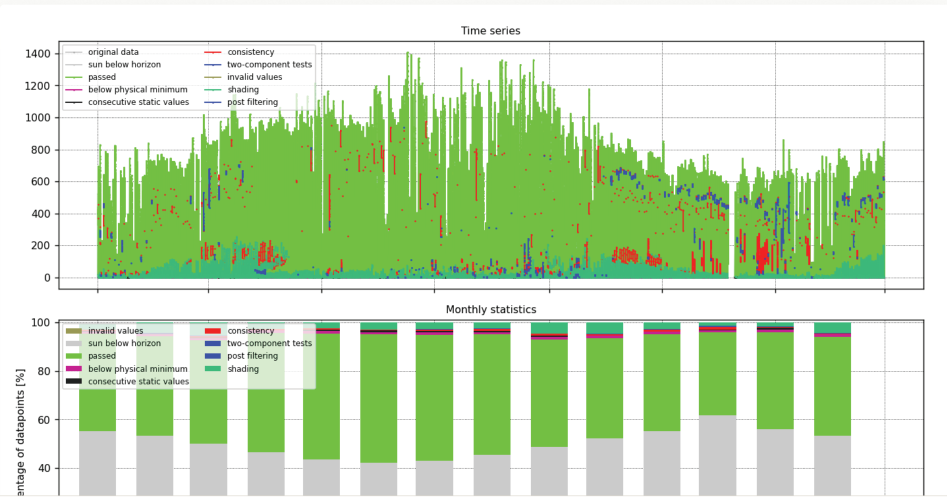
SOLARGIS

How to validate solar models using ground measurements



March 2025

Table of Contents

01

Data from satellite models or pyranometers: when to use what?

Understand how data is acquired

Satellite-based models, such as those from Solargis, use data continuously measured by remote sensing instruments installed onboard geostationary satellites as well as data from global meteorological models. Solar irradiance is calculated considering three key factors:

- Cloud properties derived from the satellite data
- Atmospheric properties, such as aerosol load and water content, derived from the meteorological models
- Terrain properties, such as elevation and shading, derived from the digital terrain models.

Typically, the models calculate solar irradiance as high-resolution grid map layers globally, more precisely between latitudes 65 degrees North and South. The models calculate solar irradiance by looking at sites "from the top".

To validate solar models, data from pyranometers installed at meteorological stations are used. The pyranometers are devices that measure irradiance "on the ground". For solar energy applications, they should be routinely maintained. Only Class A pyranometers (secondary standard) should be used to achieve the required accuracy.

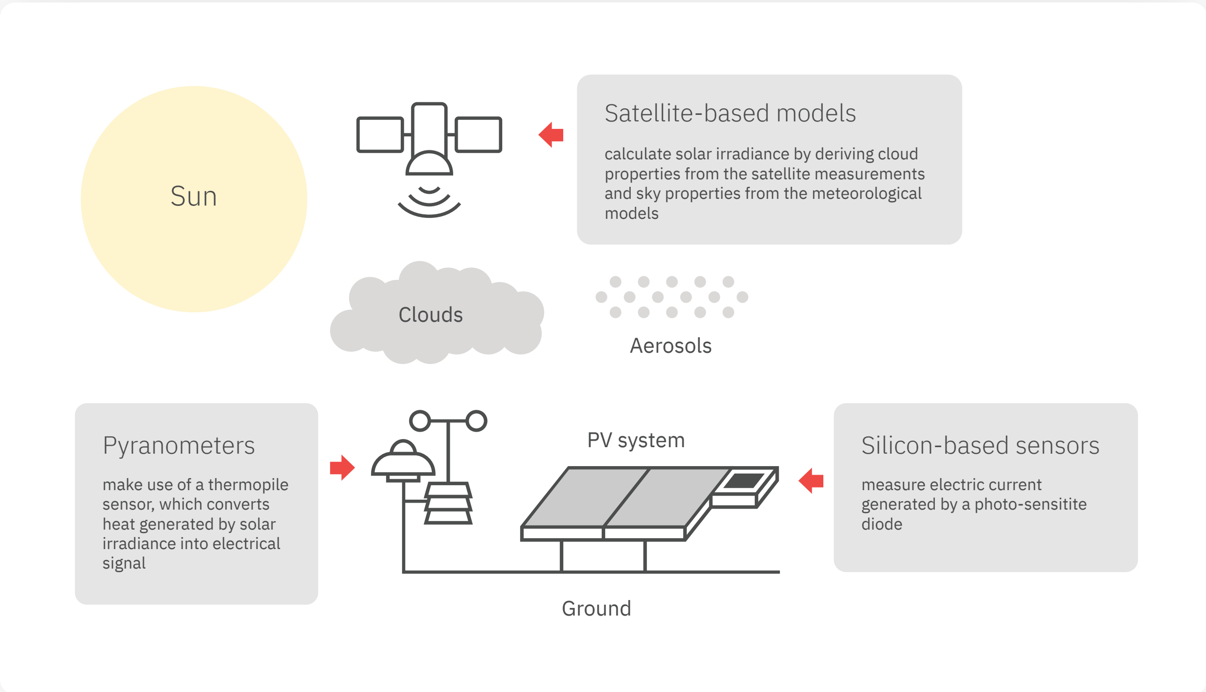


Fig. 1. Diagram showing the working principle of solar measurements and solar models. As for ground measurements, in the absolute majority of solar energy applications, the use of pyranometers is preferred to silicon-cell sensors.

Tab. 1. Main features for solar measurements and satellite-model data.

	Satellite-based model	Pyranometer	Silicon-based sensor
Measured period	18–30 YEARS APPROX. SINCE SATELLITE MISSION STARTED (1994/1999/2007)	UNDETERMINED SINCE DATE OF INSTALLATION	UNDETERMINED SINCE DATE OF INSTALLATION
Spatial coverage	GLOBAL AND CONTINUOUS LATITUDE -65° TO 65° RESOLUTION 4–60 KM² ENHANCED TO < 1 KM²	POINT-BASED ONLY COVERS THE SITE OF THE INSTALLATION	POINT-BASED ONLY COVERS THE SITE OF THE INSTALLATION
Time step	EVERY 10 OR 15 MINUTES	EVERY FEW SECONDS AGGREGATED TYPICALLY INTO 1, 5, 10 MINUTES	EVERY FEW SECONDS
Suitable for solar assets	YES	YES	NO

The satellite-model data represent a long history (for the case of Solargis, starting in the years 1994, 1999, and 2007, depending on the region) with 10- or 15-minute time resolution.

The satellite model outputs characterize an average solar irradiance over an area, typically the size of 4 to 60 square kilometers, or more, depending on latitude. The spatial resolution of the satellite-model time series can be enhanced by a digital terrain model to achieve sub-kilometer grid resolution. A long history of consistent and gap-free satellite-model time series representing all components (global, direct, diffuse) can be calculated for any location.

Outputs from modern satellite-based models are quite accurate. The temporal mismatch observed when comparing satellite-model outputs and measurements is given by three factors (in the order of their importance):

- Mismatch between cloud properties measured by a ground sensor at a given location and their average grid representation seen by the satellite sensors
- Imperfections of the atmospheric model data (mainly aerosols) used as inputs to solar models.
- Imperfections of the model algorithms.

The pyranometers installed at well-designed and rigorously maintained solar meteorological stations offer accurate site-specific measurements in high-frequency time steps. They are typically aggregated into 1-, 5- or 10-minute time series before being stored in a database. The data is available from the date when the device was installed and data recording commenced. Solar measurements represent the specific microclimate of a site where a solar meteorological station is installed. To monitor all three solar components, a combination of different instruments is used.

Another type of instrument that is sometimes offered for measuring irradiance is silicon-based solar sensors. The lower price of these devices made them popular in the industry, yet their limited accuracy does not qualify them for standard measuring solutions in solar industry applications.

The nature of the measuring and modeling techniques and the characteristics of the time series data influence the way they are used across different tasks related to the development and operation of solar power plants.

Satellite-model data are the ideal source for **long-term energy yield assessments**. For such cases, data from pyranometers are mainly useful for reducing the uncertainty of the solar models (performed by site adaptation methods).

Pyranometers would be the recommendable option for **real-time PV power plant operation** so that instant changes in performance can be detected and reported.

For **regular performance assessment** of large solar power plants or entire portfolios, the exclusive use of pyranometers shows serious limitations. This is due to the difficulty of keeping the same level of data accuracy and calibration stability under different site conditions and maintenance routines. The use of satellite-model data helps to control the stability and quality of the ground measurements and build harmonized and verified time series data for bankable performance analysis.

Satellite-model data are used also for filling the gaps that are common in ground measurements, on the condition that the process of combining both data sources is done properly. In the case of smaller solar power plants (indicatively 5 MW and smaller), the use of satellite-model data is usually sufficient.

Finally, it is to be noted that the non-uniform spectral response and the thermal dependence of silicon cells, limiting their accuracy, do not make them suitable for performance evaluation of solar power plants.

The pyranometers described so far are specifically designed to measure the global component of solar irradiance, horizontal or plane of array (tilted plane). To measure the direct and diffuse components, instruments such as pyrheliometers and shaded pyranometers (mounted on solar trackers), SPN1, or RSR are used.

In the specific case of albedo calculation, measurements of reflected irradiance are obtained by a combination of two horizontally mounted pyranometers, one facing the sky, and the other the ground.

Tab. 2. Common solar energy applications for solar measurements and satellite-model data.

	Pyranometer	Satellite-based model	Silicon-based sensor
Real-time power plant operation	YES	LIMITED	NO
Portfolio performance assessment	LIMITED	YES	NO
Long-term energy yield assessment	NO	YES	NO
Satellite model validation	YES	–	NO
Satellite model uncertainty reduction	YES	–	NO
Gap-filling of ground measurements	–	YES	NO

02

Before validation: how to filter incorrect values?

Be methodical with the reference data

Satellite-model time series data are based on the combined use of remote sensing technologies and global meteorological models. The nature of issues found in the model is different from those we can encounter in ground measurements.

Outcomes of satellite-model data rely on the frequency and resolution of satellite data acquisition instruments. Certain types of cloud situations, mainly smaller size, high variability clouds or a combination of snow/ice and cloud scenes are difficult to detect at full accuracy.

Data coming from the models can be validated only with the best quality pyranometer measurements. However, measurements by ground sensors are affected by environmental conditions (dust, humidity, dew, snow) and technical errors caused, for example, by misalignment and miscalibration of instruments or data logger issues. Therefore, before the measurements can be considered a trustable “ground reference” the values have to be quality controlled and errors filtered out.

A full suite of ground-measured quality control procedures is required to achieve a complete data quality analysis. The solar data quality assessment from Solargis consists of the following steps:

1. **Identification and correction of time shifts**, time drifts, and other time-related issues based on testing diurnal symmetry (morning vs. afternoon). The time reference is critical for all subsequent quality tests. Data analysts can manually identify the shifts or run auto-detection algorithms, which automatically suggest the location and value of shifts and drifts in the data.

2. **Detection of invalid values**, nighttime/daytime, artificial static values, breaking physical limits, and consistency of solar irradiance components. Each record that failed in these tests is flagged by a predefined flag value. Other situations like possible tracker malfunction can be flagged, showing conditions when the instrument for tracking sun position is not working properly.

3. **Visual inspection of flagged data values**. At the end of the process, flag values are saved with the data and the quality status is updated.

4. **Post-processing to remove valid data flags** left in the dataset. Some individual flags often don't provide much value, and it is recommended to remove the flags before going ahead with the analysis.

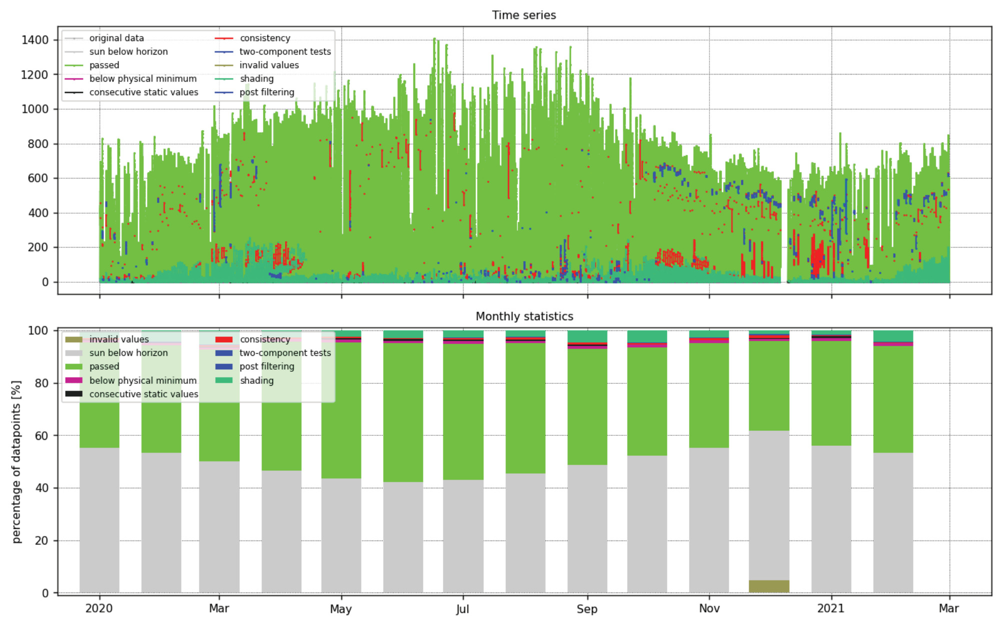


Fig. 2. A diagram showing graphical results of solar data quality control procedures.
Image captured from Solargis Analyst, the solar data analysis software.

At the end of the process, the issues are identified and the erroneous data values are flagged and discarded from further processing, namely:

- Invalid records
- Records of sun below horizon
- Values below minimum physical limit
- Consecutive static values
- Consistency issues (three components of irradiance: GHI, DIF, DNI)
- Failed two-component tests (comparing GHI vs DIF)
- Records affected by shading

03

1-to-1 comparisons: what key metrics shall I calculate?

Do not use just one variable in data comparisons

Whenever a model is compared with a ground-based data reference, we assess the similarity between the two sets of records for the same site and the same period. A specific set of statistical indicators should be used in the model-measurement comparisons.

When running model-measurements solar irradiance data comparisons **Bias** calculations give results close to zero when the average is calculated. However, it might be the case of positive and negative errors canceling out when the sum is calculated, and that is why an average is calculated over the absolute error values (**Mean Absolute Deviation** or MAD).

Calculating the mean over the absolute values can still hide something: the big errors remain undetected when they balance out with the smaller ones. To give a higher weight to bigger errors, an average over the squared deviations is calculated (**Root Mean Square Deviation** or RMSD). This indicator can be calculated for different aggregation levels (dispersion of sub-hourly, hourly, daily values, etc).

To be able to compare RMSD, we use normalization of values (calculating the percentage using a reference value). We can also benefit from other unit-invariant metrics like the **Coefficient of determination and correlation** (R2 and R).

To complete the set of indicators, to estimate the model's ability to represent various solar radiation conditions we calculate the **Kolmogorov-Smirnov Index** (KSI).

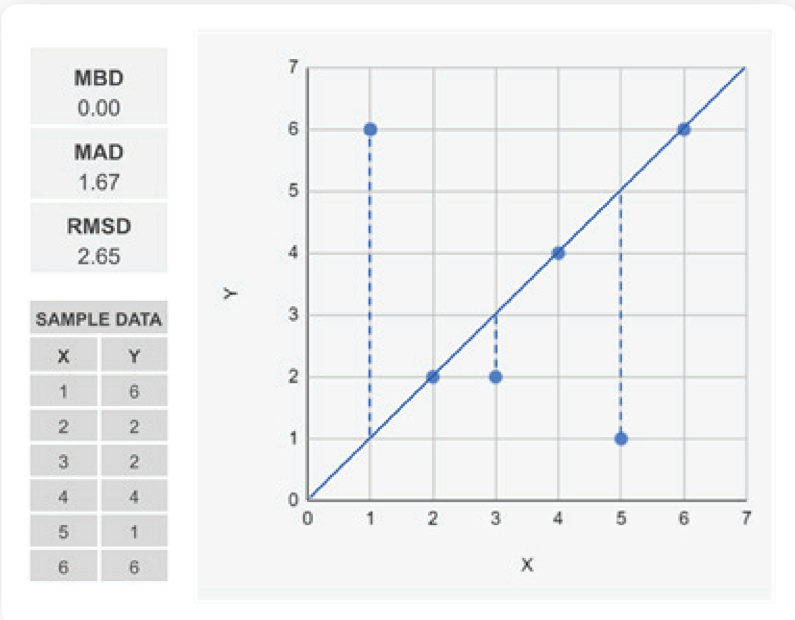


Fig. 3. Sample calculation of MBD, MAD, and RMSD showing the "data cancellation effect" and the need of having various data accuracy indicators.

Tab. 3. Summary table with main indicators to be calculated for each validation site.

Indicator			What does it represent	Expected value
MBD	$\frac{\sum_{i=1}^{n}(x_i - y_i)}{n}$	**Bias or Mean Bias Deviation**	Characterizes systematic model deviation at a given size, i. e. systematic over- or underestimation.	A bias closer to zero means lower deviation
MAD	$\frac{\sum_{i=1}^{n}\lvert x_i - y_i \rvert}{n}$	**Mean Absolute Deviation**	Shows an indication of the spread of error. It is used to avoid deviations with opposite signs cancelling each other out.	A MAD closer to zero means a lower error
RMSD	$\sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}}$	**Root: Mean Square Deviation**	Shows an indication of the spread of error, giving higher weight to larger errors. It should be calculated at different granularities (sub-hourly, hourly, daily, monthly).	A lower RMSD means a lower spread of error
R	$\frac{\sum_{i=1}^{n}[(x_i - \underline{x}) * (y_i - \underline{y})]}{\sqrt{[\sum_{i=1}^{n}(x_i - \underline{x})^2 * \sum_{i=1}^{n}(y_i - \underline{y})^2]}}$	**Correlation coefficient**	Represents how close data points are to the regression line.	The closer to 1, the higher the similarity between the measurements and model
R^2	R^2	**Coefficient of determination**	Represents how well the model fits the data.	The closer to 1, the higher the similarity between the measurements and model
KSI	$sup\lvert F_x(X) - F_y(X)\rvert$	**Kolmogorov-Smirnoff Index**	Indicates a model's ability to represent various solar irradiance conditions and characterizes the representativeness of distribution of values.	A lower KSI means a higher similarity of distribution between datasets.

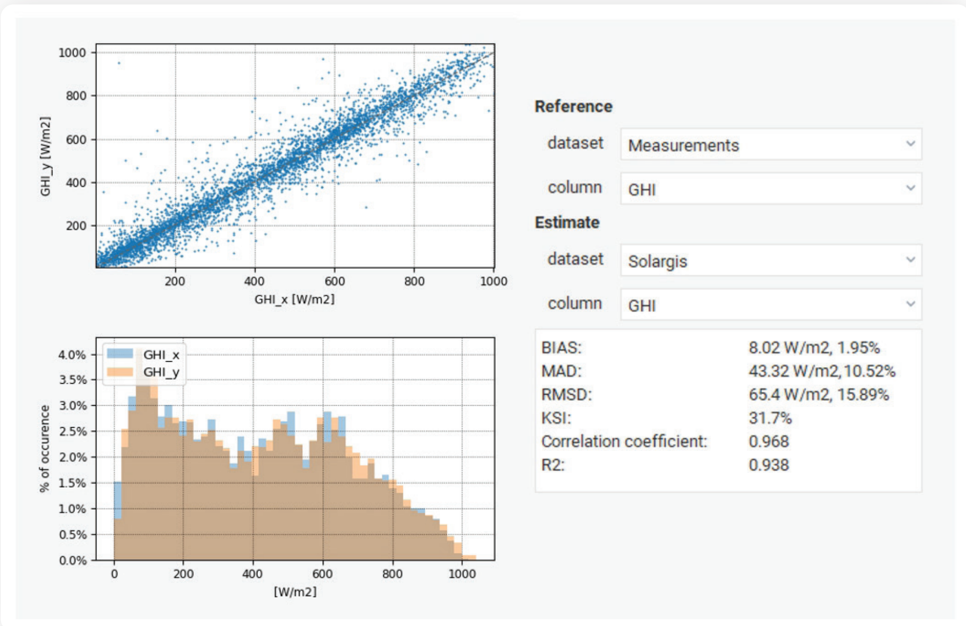


Fig. 4. Comparison charts (scatterplot and histogram) with calculated key indicators for a sample site. Image captured from Solargis Analyst.

04

Interpreting results: what factors influence satellite-model data uncertainty?

Avoid simplification of uncertainty factors

Validation statistics provide useful metrics to evaluate the performance of a solar radiation model. Model performance analysis is not easy and requires expert-level knowledge of the model, its internal algorithm, and its inputs. On the other side, well-executed validation assumes that quality-controlled measurements from high-accuracy and rigorously maintained instruments are used.

There are simplified approaches to analyzing uncertainty that convert validation statistics to expected uncertainty for all the covered territories, or they simply calculate metrics by continent. It is a fact that satellite-based models perform better in more stable weather conditions. However, the climate is not the only factor affecting solar model performance. Deriving expected uncertainty by looking solely at climate does not yield a complete interpretation.

In practice, the expected uncertainty needs to be estimated on a case-by-case basis taking into account all factors including climate, geography, environment, and satellite technology (see table).

Only an indicative range of uncertainty can be determined without looking at the specific site conditions. Fast generalizations, such as "This model works better here, and the other one works better there" are to be avoided. Uncertainty estimates need to be studied carefully.

Tab. 4. Overview of identified factors affecting uncertainty of satellite-based irradiance data

Climatology	Geographic	Environmental	Satellite technology
Clouds persistence Clouds variability Aerosol optical depth Total water vapour Snow coverage	Terrain variability Distance to water surface High albedo surface	Anthropogenic pollution Extreme weather events	Frequency of data acquisition Data spatial resolution Pixel distortion Data acquisition issues

05

Collecting stats: how many solar meteorological stations are needed to validate a satellite-model data?

Validate in as many locations as possible

Solar developers have questions about what solar database would give them more accurate results. Since there is effectively no updated independent study that compares all databases globally, validation statistics are one of the most important tools when comparing models from different providers.

Comparing models based on validation results from an uneven number of sites leads to incorrect comparison results. In statistics, the size of the sample always matters and the same as happens with surveys, knowing the answers of just a few people does not help in understanding everyone's opinion.

To follow the analogy, saying that a model with one validation site is more accurate than a model with 320 validation sites, just because a lower bias was calculated for this 1 site, is akin to guessing the response of 320 people by asking one of them. For solar power projects, where every percentage point of accuracy matters, estimating accuracy using measurements from a few sites is inexcusable when more reference data is available.

So, for the question of how many sites with high-quality solar measurements are needed to objectively validate a solar model, the answer is clear: the higher the number of sites, the better. The complexity of factors affecting the solar irradiance model uncertainty requires having as many validation sites as possible.

To sum up, when looking at validations statistics from different models there are three additional factors to consider:

- **Quantity.** Large differences in the number of validation sites lead to unreliable model comparisons.
- **Quality.** Uneven data quality leads to unreliable statistics when datasets are combined together.
- **Distribution.** Having geographic conditions without validation makes any comparison less representative.

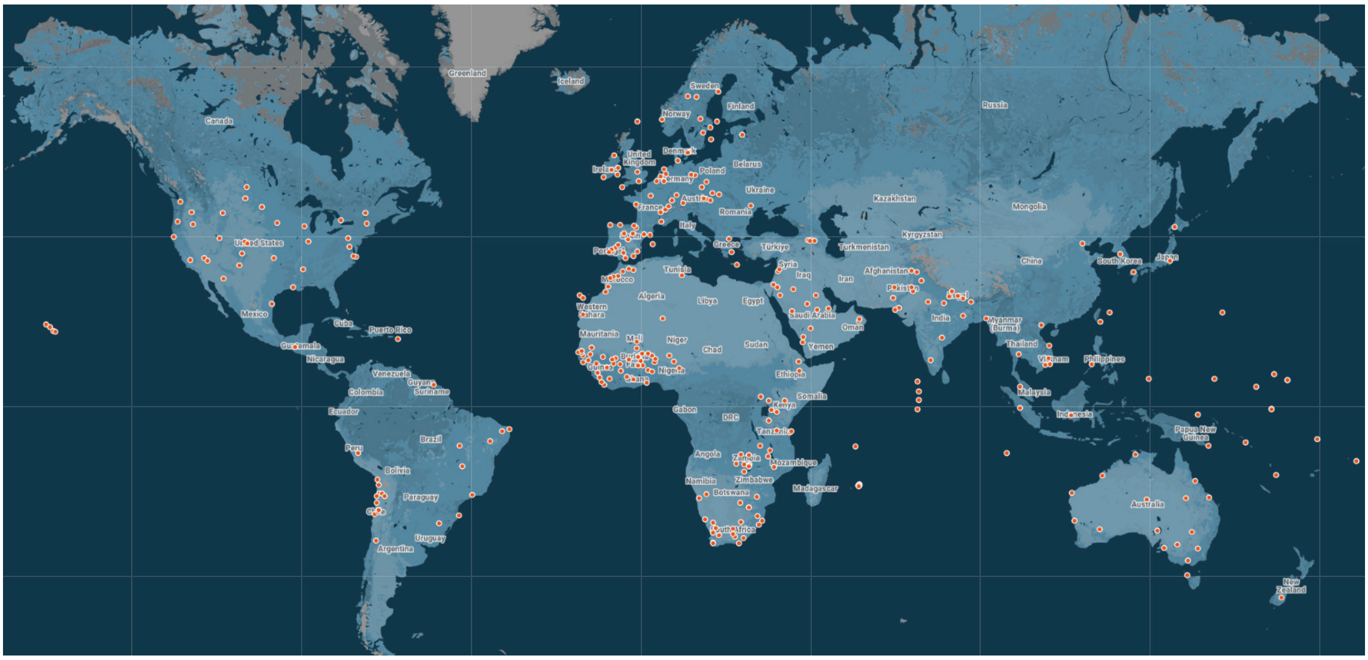


Fig. 5. High-quality public reference solar meteorological stations used by Solargis (320 sites as of March 2025).

06

Summary: four main steps to achieve useful satellite data validations using ground data

The process we take for validating satellite-model solar time series using high accuracy quality-controlled solar measurements is summarized as follows:

1. **Filter ground measurements**
   using high standard quality control procedures and identify non-valid data records

2. **Calculate key indicators**
   including all necessary metrics to characterize differences between the measured and modeled data

3. **Analyze specific site conditions**
   that may be affecting the satellite-based model performance.

4. **Keep collecting sites**
   and repeat the above steps with as many high-quality ground measurements as possible

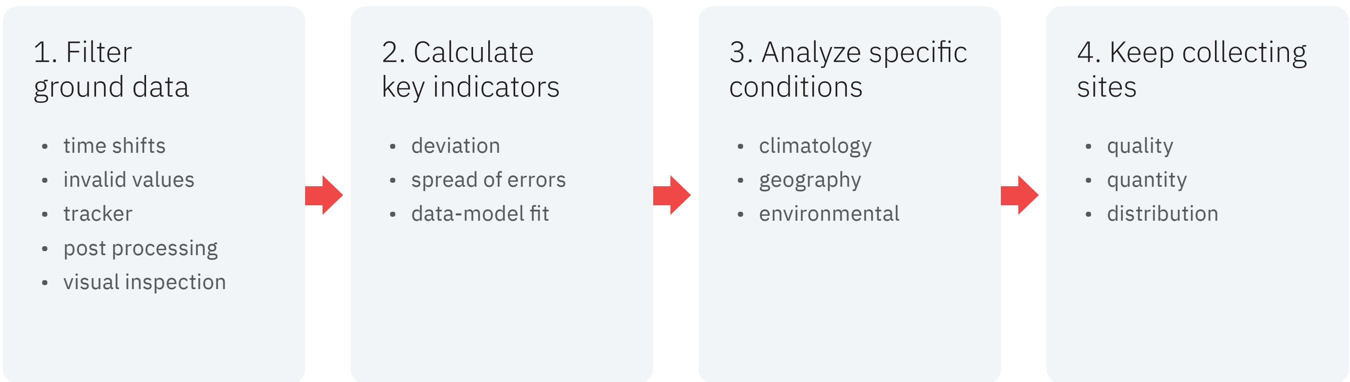


Fig.6. Main steps and key recommendations for validation of satellite-based solar irradiance models for solar energy applications

SOLARGIS

Headquarters
Solargis s.r.o.
Bottova 2A
811 09 Bratislava
+421 243 191 708
contact@solargis.com
solargis.com

Americas
Solargis Americas Inc.
150 King St. W, Suite #200
Toronto, ON M5H 1J9
Canada
+1 647 472 1588

APAC
Solargis APAC Pte. Ltd.
6 Battery Road, #716,
Singapore 049909
+65 9396 7410

Our partners
See our regional
partners at
solargis.com/partners