Solar Energy 258 (2023) 175–193

Contents lists available at ScienceDirect

Solar Energy

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

Improvement of satellite-derived surface solar irradiance estimations using spatio-temporal extrapolation with statistical learning

Hadrien Verbois [a,*], Yves-Marie Saint-Drenan [a], Quentin Libois [b], Yann Michel [b], Marie Cassas [b], Laurent Dubus [c], Philippe Blanc [a]

[a] Mines Paris, Université PSL, Centre Observation Impacts Energie (O.I.E.), 06904 Sophia Antipolis, France
[b] CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France
[c] Réseau de Transport d'électricité, Paris, France

ARTICLE INFO

Keywords:
Solar irradiance
Machine Learning
Satellite-derived solar radiation
Global adaptation
Modeled solar radiation data
Solar resource mapping

ABSTRACT

Estimations of solar surface irradiance (SSI) derived from meteorological satellites are widely used by various actors in the solar industry. However, even state-of-the-art empirical and physical SSI retrieval models exhibit significant errors; the estimations provided by these models are thus traditionally corrected using ground-based measurements of SSI as references. The literature is rich with such correction methods, often called adaptation techniques. Most of the proposed models, however, are local or site-specific, i.e., they do not extrapolate the correction in space and are only applicable to the location of the ground-based measurements.

In this work, we propose a novel global adaptation technique, that can extrapolate the correction in both space and time. To that end, we leverage (1) a dense network of measurement stations across France, (2) a relatively large number of predictors, and (3) a non-linear, sophisticated regression algorithm, the Extreme Gradient Boosting. The model is applied to the HelioClim3 database; its performance is benchmarked against raw HelioClim3 estimations, and alternative, simpler adaptation techniques.

Our analysis shows that this global model significantly improves satellite-derived SSI estimations from the HelioClim3 database, even when the evaluation is carried out on measurement stations that were not part of the training set of the algorithm. Our proposed model also outperforms all tested alternative global adaptation techniques.

These results suggest that global adaptation techniques leveraging advanced machine learning and high dimensionality have the potential to significantly improve satellite-derived SSI estimations, notably more than traditional adaptation approaches. There is certainly room for improvement, but the development of such techniques is a promising research topic.

1. Introduction

Accurate information about surface solar irradiance (SSI) is key in many areas of the photovoltaic and solar industry but also in other domains such as climatology, architecture, agriculture, etc. Early in a project life cycle, it is critical for the design and feasibility studies of a new solar power plant (da Graça et al., 2012); later on, in the operational phase, it is required to monitor the production of the plant (Marcos et al., 2011) and ameliorate energy market participation (Antonanzas et al., 2017; Alessandrini et al., 2014). On a larger scale, it is also paramount to e.g. optimize unit commitment (Brancucci Martinez-Anido et al., 2016) or manage regional transmission networks (Saint-Drenan et al., 2017). Depending on the application, different spatial resolutions and extents may be required; SSI may be needed in a single location or over a whole area and, in the latter case, with a spatial resolution ranging between sub-kilometer and countrywide scales. Furthermore, different time resolutions might be necessary, ranging from seconds to years or even decades (Blanc et al., 2017). Lastly, SSI may be required over a past period, for the present, or a specific horizon in the future.

To answer these various requirements, different data sources may be helpful. All come with strengths and weaknesses. Ground-based instruments, such as pyranometers, are likely the most accurate (Badosa et al., 2014). However, they only cover a very restricted area, and because they are expensive to install and maintain, they rarely are in place for long periods. On the other hand, Numerical weather models cover large areas — global models even cover the whole earth. Since, in

* Corresponding author.
E-mail address: hadrien.verbois@minesparis.psl.eu (H. Verbois).

https://doi.org/10.1016/j.solener.2023.04.037
Received 17 August 2022; Received in revised form 21 March 2023; Accepted 17 April 2023
Available online 10 May 2023

addition, they can be run retroactively, they can theoretically provide SSI estimations without restriction in space and time. Unfortunately, as they rely only sparsely on data – through data assimilation – and because of various numerical approximations, even state-of-the-art numerical models are often biased and can exhibit large errors (Perez et al., 2013, 2010; Jimenez et al., 2016). Because of the computational cost of running such models, they also have relatively coarse spatial and temporal scales (Verbois et al., 2020). Satellite-derived solar irradiance estimations rely on both observations – the satellite channels – and models to estimate the SSI from these channels (Rigollier et al., 2002; Tournadre et al., 2021). They can be seen as a compromise between ground-based sensors and numerical weather models: like the latter, they cover a large area and period, but their temporal and spatial resolution is usually finer (Ineichen, 2014). In Europe, it is possible to get SSI estimations at a resolution of 15 min and 3 km using e.g. HelioClim3 (Blanc et al., 2011b) or CAMS (Schroedter-Homscheidt et al., 2016). In the United States, the National Solar Radiation Database (NSRDB, Sengupta et al., 2014) provides SSI estimates with 30 min and 4 km resolution. Satellite estimations of SSI are nonetheless less accurate than ground-based measurements, and, like numerical models, often exhibit systematic errors (Qu et al., 2014; Ineichen, 2014). The inaccuracy of satellite estimations is partially due to modeling errors – similar to numerical weather models – and a lack of temporal and spatial resolutions, constrained by satellite instruments.

To leverage both data sources simultaneously, it is common to correct long-term satellite-derived estimations using short-term ground-based measurements (Polo et al., 2016, 2020). This approach is often referred to as *adaptation technique*. In the past few years, many models have been implemented and tested. For post-hoc correction – typically used for local resource assessment – quantile mapping (QM) is perhaps the most popular approach. QM corrects the Cumulative Distribution Function (CDF) of the long-term satellite estimations to match that of the short-term ground measurements. It is very efficient at removing bias and adjusting other higher-order moments of the distributions (e.g. standard deviation) (Yezeguelian et al., 2021). However, QM does not concern itself with the co-occurrence of estimations and measurements. It is thus improper for applications where the SSI estimations must not only be adjusted in terms of their distribution but also temporally accurate (Polo et al., 2016). In that context, regression-based adaptation techniques – sometimes referred to as Model Output Statistics (MOS) – are more relevant. The goal of MOS is to establish a relationship between the satellite estimated SSI – and possibly more observable predictors – and the measured SSI. In the literature, many regression models were tested. When SSI estimated by the satellite is the single predictor, simple algorithms show promising results: Davy et al. (2016) successfully implemented a simple generalized additive model, Mazorra Aguiar et al. (2019) combined a simple linear regression with k-mean clustering, and Vernay et al. (2013) used a simple linear correction in the Fourier domain. Several works proposed to use more complex algorithms, usually combined with a larger predictor set, including several components of SSI, solar geometry information, and sometimes basic weather data. Various classical machine learning algorithms were tested: neural networks (Şahin et al., 2014; Cornejo-Bueno et al., 2019), Random Forest (Babar et al., 2020), support vector machine, and Gaussian process (Cornejo-Bueno et al., 2019).

In most of these studies, the adaptation technique is local, i.e. it only corrects satellite estimations for locations where ground-based measurements are available. This is sufficient in some use cases, such as the sizing of a new photovoltaic power plant. However, for other applications, SSI estimations are required over an extended spatial coverage. For example, transmission system operators (TSO) need that spatially distributed information to estimate or forecast the power output of a whole region (Saint-Drenan et al., 2017). Even forecasting models focusing on a single location sometimes utilize irradiance data from the surroundings and therefore benefit from improved SSI estimations over multiple pixels (Sperati et al., 2017). Adaptation techniques that can extrapolate in space are sometimes referred to as regional or global adaptation techniques. Even though most of the satellite adaptation literature focuses on local models, few papers developed global alternatives. Davy et al. (2016) propose a global MOS for Australia: they train generalized additive models for each location with ground measurements and use a distance-weighted interpolation of these models for every other location in Australia. Babar et al. (2020) implement a Random Forest for the same purpose, but in Scandinavia; to help their model generalize spatially, they do not use interpolation, but directly give the latitude and longitude as input to their model.

In this work, we aim to contribute to the global adaptation technique literature by developing a novel MOS model and proposing a comprehensive performance analysis framework. Furthermore, we implement several benchmark models to challenge the main design choices of our model. Our proposed MOS model is somehow akin to that of Babar et al. (2020) in that it uses a single regression model and does not rely on any interpolation routine. However, our approach to predictor selection, particularly our consideration of past time steps, is fundamentally different. We intend to provide a large-enough predictor set so that the correction model can adapt its correction to a specific temporal context.

The study is conducted using a large network of ground-based instruments — more than 200 measurement stations, spread across France for up to 9 years. While our results are a priori specific to mid-latitudes, they demonstrate that our proposed approach is scalable to a country.

Our method is exposed in Section 2: we describe the data used and the associated quality checks (Section 2.1), detail the MOS models implemented (Section 2.2), and discuss our experimental setup (Section 2.3). The results are presented in Section 3 and divided into three parts: first, a multi-model comparison is conducted (Section 3.1), then we perform a condition-dependent analysis (Section 3.2), and finally, we evaluate the plausibility of our estimations (Section 3.3). Conclusions are given in Section 4, where we also discuss future perspectives.

2. Method

2.1. Data

2.1.1. Ground stations

This work uses ground-based irradiance measurements from 286 weather stations operated by Meteo France. The stations are equipped with Kipp&Zonen pyranometers[1] and measure 1-min Global Horizontal Irradiance (GHI). In this work, only hourly averages of GHI are used. As illustrated in Fig. 1, the stations are spread across metropolitan France relatively homogeneously. The data span 9 years, between 2010 and 2019, but not all stations were operational during the whole period.

2.1.2. Quality check

A thorough quality check (QC) procedure is applied to the ground measurements. First, the automated checks recommended by Espinar et al. (2013) are applied. As we only have access to GHI and because we take a conservative approach, they are reduced to the *extremely rare limits* (ERL) equation:

$$-2 < GHI < 1.2 I_{sc} cos^{1.2}(\theta_z) + 50 \text{ (in W m}^{-2}\text{)}$$

where I_{sc} is the solar constant adjusted for Earth–Sun distance, and θ_z the mean hourly solar zenith angle.

We then use a digital model of the horizon (Blanc et al., 2011a) to exclude all points below the horizon. Points with a mean hourly elevation angle below 10° are also discarded systematically.

[1] Several models are used depending on the station: CM11, CM6B, CMP6, CMP10, and CMP11. The details of the instrument at each station can be found at: https://donneespubliques.meteofrance.fr/?fond=contenu&id_contenu=37.

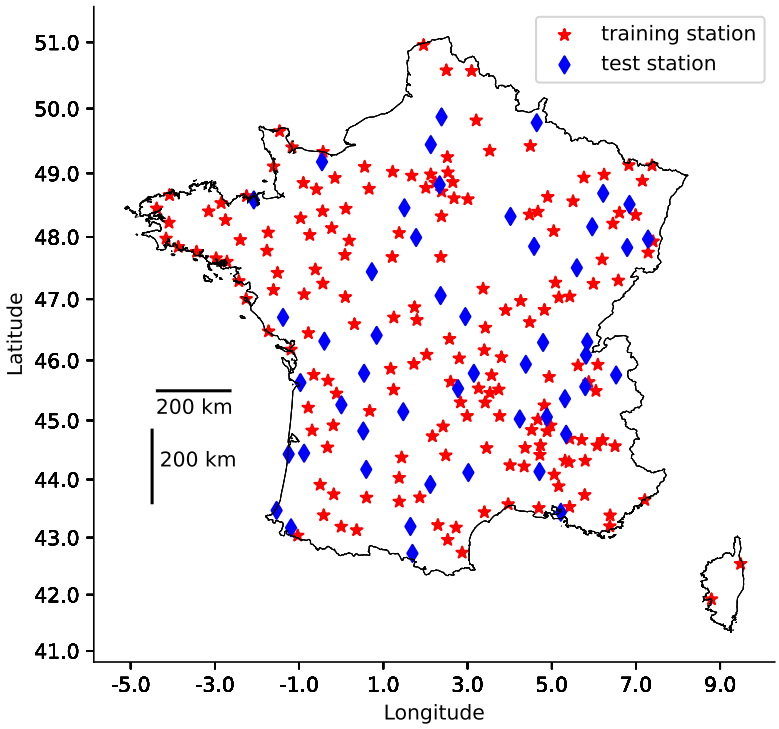


Fig. 1. Location of Meteo France stations used in this study, divided into training and test sets, as described in Section 2.3.2.

We subsequently conduct a visual check to detect possible calibration errors. The clear sky index k_c is computed as the ratio of the measured irradiance and the clear sky irradiance – estimated by the McClear model (Lefèvre et al., 2013). For each station, two time series are plotted:

1. the station k_c;
2. the ratio $R_{neighbors}$ of the station k_c to the k_c averaged over the neighboring stations (within a 50 km radius).

If we see a consistent drop in the maximal k_c and $R_{neighbors}$ for several months, the period is marked as failing QC. Fig. 2 illustrates the process.

Finally, a second visual QC is applied to detect possible additional shadowing, illustrated for one station in Fig. 3. For each station:

1. we select all instants T_{clear} for which HC3 clear sky index k_c^{HC3} is greater or equal to 0.75: $T_{clear} = \{t \text{ such that } k_c^{HC3}(t) \geq 0.75\}$;
2. we consider the corresponding ground-measured clear sky index k_c in the solar elevation — solar azimuth (α, ϕ) plane: $\{k_c(t, \alpha, \theta) \text{ for } t \in T_{clear}\}$.

As there are several years of data for each station and because the sun is in the exact same position twice a year, several k_c values fall in the same (α, θ) point in that plane.

3. For each given (α, ϕ) point, we select the highest value, $k_c^{max}(\alpha, \theta) = max_t\{k_c(t, \alpha, \theta) \text{ for } t \in T_{clear}\}$, and display it in on an azimuth/elevation graph.

If there is no object in a given elevation/azimuth line of sight, we expect k_c^{max} to be close to 1 - at a minimum greater than 0.75. On the contrary, a low k_c^{max} value is a strong indication that there might be a shadowing object in that direction. In particular, we are looking for low k_c^{max} region in the azimuth/elevation graph that originated from the ground ($\alpha = 0$).

As the data has a temporal resolution of one hour, we see a discontinuous pattern.

4. The image is thus convoluted with a median filter of size ($10° \times 10°$), F_{median}, to obtain a continuous picture.

The shadowed area can then be identified manually (a pole and some trees are recognizable in Fig. 3). All the time-steps falling in the shadowed area are flagged as falling QC.

Note that shadows caused by the topography surrounding a station (mountain for example) were already automatically detected thanks to the digital model of the horizon.

Summary. Out of 286 stations, 46 stations failed QC for more than 40% of their time-steps and were totally removed from the data-set. The distribution of days passing QC for each station is illustrated in Fig. 4.

2.1.3. HelioClim3

Satellite-based irradiance estimations are taken from the HelioClim3 version 5 database (HC3, Espinar et al., 2012). HC3 uses the Heliosat2 method (Rigollier et al., 2004) to derive surface irradiance from MeteoSat second generation (MSG) data (Rigollier et al., 2002). Since version 5, HC3 also relies on CAMS aerosol data (Schroedter-Homscheidt et al., 2016) and the McClear clear-sky model (Lefèvre et al., 2013) to estimate clear-sky irradiance.

HC3 estimations natively come with a 15 min time resolution and the same spatial resolution as MSG, 3 km at the nadir. Because MSG observation time is not the same for each pixel of its sensor, interpolation is done so that all HC3 SSI estimations are given for the same time-step.

In this work, we use 1-hour averages to match the time resolution of the measurements (see Section 2.1.1) and a regular projection above France, with a spatial resolution of 0.04 degrees.

Note that for HC3 as for ground data, the time intervals are *right-labeled*: hourly SSI at e.g. 11:00 refers to the average SSI between 10:01 and 11:00.

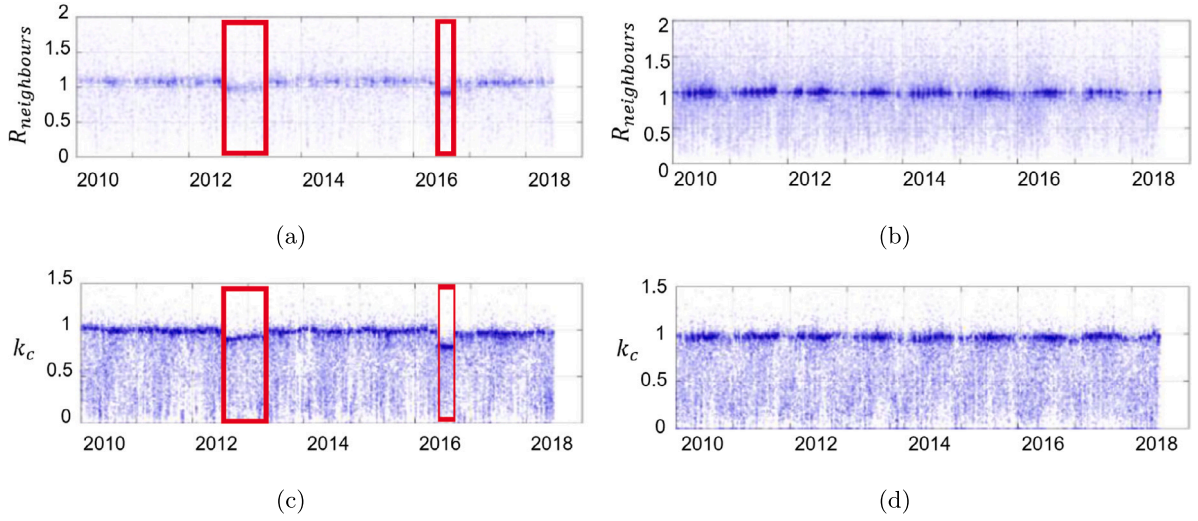


Fig. 2. Example of visual QC to find calibration errors based on the station k_c and $R_{neighbors}$. On the left plots (a, c), we see concurrent drops in k_c and $R_{neighbors}$ (in red) - they are flagged as not passing QC. The data on the right plots (b, d) passes this QC test. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

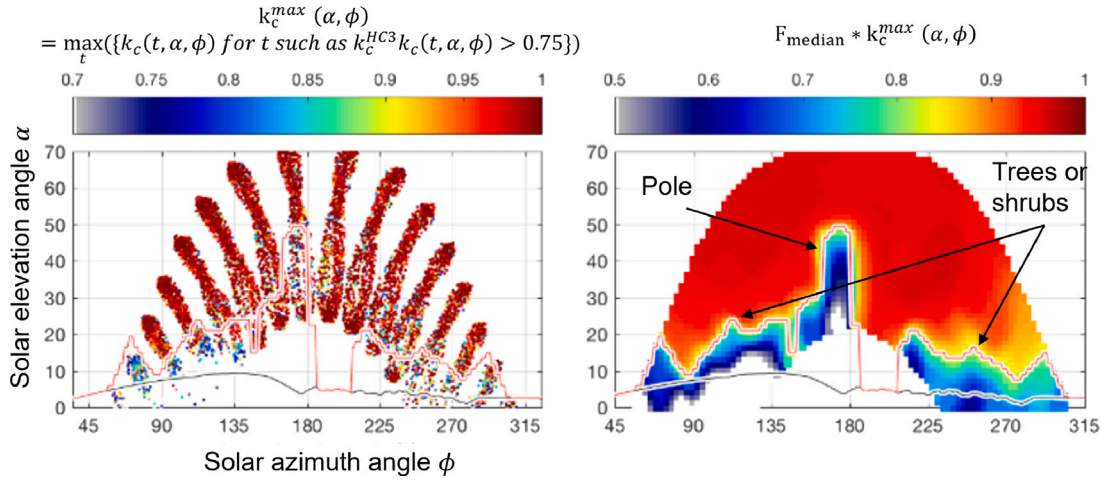


Fig. 3. Example of visual QC to detect shadowing. Hourly k_c^{max} are shown on the left plot, and after passing a median filter on the right plot. Shadows from what we suspect are a pole and neighboring trees are clearly visible. The horizon from the digital model of the horizon is also shown in both plots by a dark line. All time-steps for which the station is shaded were flagged as failing QC.

2.2. Adaptation techniques

2.2.1. XGB_global: our proposed model

Our proposed model, XGB_global, relies on a non-linear regression algorithm popular in Kaggle competitions (Kaggle, 2022): Extreme Gradient Boosting (XGB) (Chen and Guestrin, 2016). XGB is an improved implementation of Tree Gradient Boosting (GB), known to handle relatively high dimensionality well, and based on regression trees. At training, regression tree partitions the predictor space into a set of complementary rectangles; the target variable is then estimated by the mean response in each partition. Regressions trees have low bias but high variance. XGB is designed to take advantage of this feature: the algorithm combines a large number of regression trees, that are built to be as uncorrelated as possible with each other. This reduces the variance while keeping the low bias. XGB algorithm has several so-called *hyper-parameters* that are not learned during training but must be tuned to the problem at hand. These hyper-parameters either control how each individual tree is built or how the different trees are combined:

1. The maximum depth of each tree can be specified. A small maximum depth reduces the size of trees; it can thus be seen as a regularization term.
2. The depth of each tree can also be constrained using the minimum number of samples in each leave of each tree.
3. The total number of trees used must be set by the user. A higher number of trees implies a more complex algorithm (and thus more risk of overfitting).
4. To reduce the correlation between two regression trees, it is possible to train each tree with a subset of the data. The ratio of train data used to fit each tree is a hyper-parameter.
5. Finally, the contribution of each tree is constrained by a learning rate, which must also be set by the user.

The hyper-parameters selection process is further described in Section 2.3.3. For more details on regression trees and gradient boosting, the reader may refer to Hastie et al. (2009); Section 4.2.2.6 of Verbois (2019) may provide a solar forecasting perspective.

To give the algorithm as much context as possible, we use a set of 18 predictors. From HC3 and McClear, we use the 5 previous time steps estimations of hourly GHI and hourly k_c, as well as the last clear-sky irradiance (CSI) estimation. The sun elevation and azimuth angles

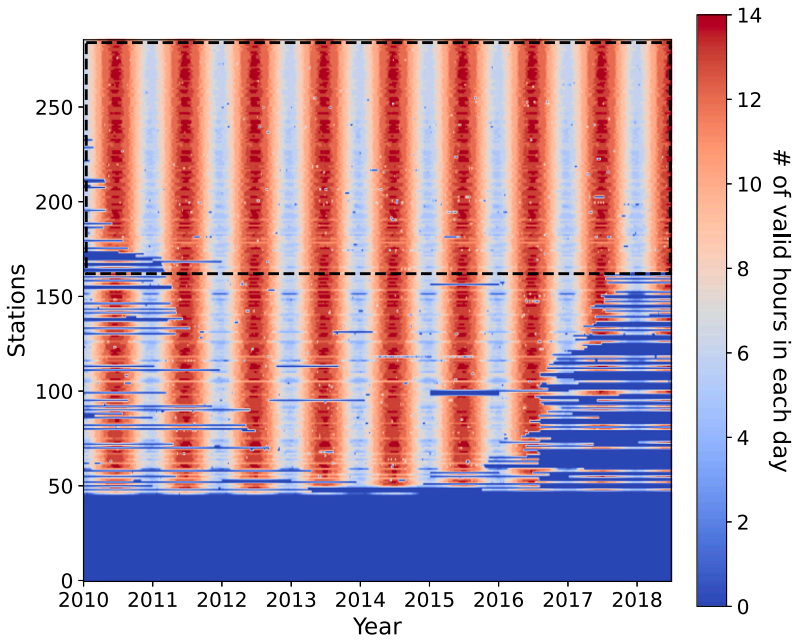


Fig. 4. Quality check of stations data: number of hours per day passing the QC. The set of stations from which 50 training stations are randomly drawn (see Section 2.3.2) is indicated by a black dashed rectangle.

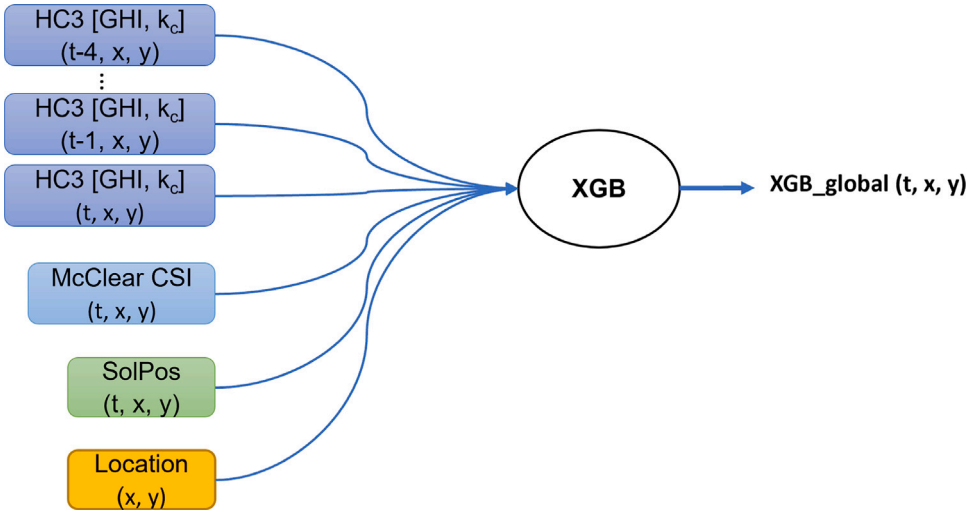


Fig. 5. Overview of the XGB_global model. SolPos contains solar elevation and azimuth angle, as well as the time of day and day of the year ; location contains latitude, longitude and altitude.

— computed using the sg2 library (Blanc and Wald, 2012), the day of the year and the time of day are also added. Finally, the latitude, longitude, and altitude of the HC3 pixel are used as extra predictors. This predictor set is notably redundant: the k_c, GHI, and CSI values are strongly correlated; solar elevation, azimuth, time of day, and day of year are not independent either. We thus rely on the ability of XGB to handle correlated inputs and to leverage weak predictors (Hastie et al., 2009). The model is summarized in Fig. 5, and the predictors are given in Table 1.

It should be emphasized that this model can operate in real time. Indeed, while the training of the model is time-consuming, applying the already trained model to new data is very fast.[2] Every time a new HC3 map is available, all the data necessary to correct this map are available and XGB_global can thus be obtained immediately.

2.2.2. Reference models

The primary objective of an adaptation technique is to improve HC3 estimations. We will thus thoroughly compare the performances and properties of XGB_global and HC3. However, our proposed XGB_global model is fairly complicated; we, therefore, implement several alternative adaptation models to justify our design.

BRR_global: a linear model. To isolate the effect of the XGB_global regression algorithm and of its non-linearity in particular, we implement

[2] Training a single XGB model on our training set takes between one and three hours on a single processor, depending on the hyper-parameters. Because we use grid-search and cross-validation to tune the hyper-parameters, several days of computation were necessary for the whole cross-validation process. In contrast, applying the trained model to one HC3 time-step and for all pixels in France (110,000 pixels) only takes a few seconds on the same processor.

an alternative global model based on linear regression, BRR_global. We choose a Bayesian Ridge Regression (BRR, MacKay, 1992), because (1) it implements both L1 and L2 regularization, which is useful to best benefit from multiple predictors and because (2) the L1 and L2 coefficients are determined with a Bayesian framework, which does not require any cross-validation which is advantageous considering the size of our data-set.

We emphasize that BRR_global uses the same predictors as XGB_global; it only differs in the choice of the regression algorithm.

XGB_interpol: a local model with interpolation. Another specificity of XGB_global is that a single model is needed for all of France. That is because the spatial variations of the required corrections are directly handled by the XGB algorithm, which gets the location (latitude, longitude, and altitude) as predictors. In contrast, other approaches in the literature train separate adaptation models for each training station and interpolate their predictions for test stations. We thus implement a model that relies on interpolation from training to test stations, XGB_interpol:

1. We first implement local adaptation models based on the XGB algorithm for each training station: $\mathrm{XGB}^{i}_{\mathrm{local}}$, where $i \in$ {Training Stations} indicates the station and $x_i = (Lat_i, Lon_i)$ its location.
2. A global model, XGB_interpol is built to correct HC3 estimations in any x_j, where $j \in$ {test stations} by interpolating XGB^{i}_{local} corrections for x_i in the neighborhood of x_j. We chose to use the 4 nearest training stations and modulate their contribution by the squared inverse of the geodesic distance $d(x_i, x_j)$ between x_i and x_j:

$$\mathrm{XGB\_interpol}(x_j) = \frac{1}{\sum_{N(x_j)} \frac{1}{d(x_i,x_j)^2}} \sum_{N(x_j)} \frac{1}{d(x_i,x_j)^2} \mathrm{XGB}^{i}_{\mathrm{local}}(x_j); \tag{1}$$

where $N(x_j)$ is the set of 4 nearest neighbors.

XGB_station: a local model without spatial interpolation. To further explore the implication of using a single global XGB regressor in XGB_global, we also evaluate the performance of XGB^{j}_{local} for j in test stations. It should be noted that this model, named XGB_station, is a local model; it has access to more information than the other MOS models tested in this paper and only extrapolates the correction in time, not in space. XGB_station is thus not a fair challenger but constitutes an insightful reference.

XGB_spatial: a spatial model without temporal interpolation. Furthermore, we implement a model that does not extrapolate in time: XGB_spatial. The configuration of this model is identical to that of XGB_global, but while a strict spatio-temporal separation between training and test sets is enforced for XGB_global, only spatial separation is imposed on XGB_spatial (See Section 2.3.3 for more details on the design of training and test sets.). XGB_spatial, similarly to XGB_station, has thus access to more information than the other MOS models tested in this paper; it only extrapolates in space, not in time.

SLR_global: a simple global model. Finally, we implement a simple global model based on a simple linear regression between HC3 k_c and measured k_c: SLR_global. The model, summarized in Fig. 6, serves as reference and constitutes a lower bound in terms of complexity and performance.

2.2.3. Summary

The 6 correction models implemented in this work are listed in Table 1, with their respective predictors. SLR_global, BRR_global, XGB_global, and XGB_interpol are global models, while XGB_station is a local model and XGB_spatial a spatial model.

2.3. Experimental setup

2.3.1. Metrics

The root mean square error, RMSE, is arguably the most popular metric for the forecasting and estimation of solar irradiance. Given the error ϵ such as $\epsilon_t = \hat{I}_{t,forecast} - I_{t,observed}$, where $I_{t,observed}$ is the observed irradiance at time t, $\hat{I}_{t,forecast}$ the forecasted irradiance, the RMSE is defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\hat{I}_{t,forecast} - I_{t,observed}\right)^2}; \tag{2}$$

RMSE measures both the precision and accuracy of the estimation. To separate these two aspects of the error, RMSE can be decomposed into the Mean Bias Error, MBE, and the standard deviation of the error, SDE[3]:

$$RMSE^2 = MBE^2 + SDE^2 \tag{3}$$

where:

$$MBE = \frac{1}{n}\sum_{t=1}^{n}\hat{I}_{t,forecast} - I_{t,observed}. \tag{4}$$

and

$$SDE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{I}_{t,forecast} - I_{t,observed} - MBE)^2}. \tag{5}$$

MBE measures the precision – or bias – of the estimations, while SDE measures their accuracy.

To compare two models A and B in terms of SDE or RMSE, it is sometimes useful to define the change in SDE or RMSE, ΔSDE^{B}_{A} and $\Delta RMSE^{B}_{A}$ (Verbois et al., 2022):

$$\Delta SDE^{B}_{A} = SDE_B - SDE_A. \tag{6}$$

$$\Delta RMSE^{B}_{A} = RMSE_B - RMSE_A. \tag{7}$$

Because MBE can be negative, we use the absolute change in MBE, $\Delta|MBE|^{B}_{A}$:

$$\Delta|MBE|^{B}_{A} = |MBE|_B - |MBE|_A. \tag{8}$$

ΔSDE^{B}_{A} ($\Delta|MBE|^{B}_{A}$ and $\Delta RMSE^{B}_{A}$ resp.) is positive when model A outperforms model B in terms of SDE (MBE and RMSE resp.).

When two models A and B have similar performance in terms of a metric $\mathcal{M}$, for example $\mathcal{M}_A \lesssim \mathcal{M}_B$, it is useful to conduct statistical testing. Following Verbois et al. (2022), we use bootstrapping to reject the null hypothesis $\mathcal{M}_A = \mathcal{M}_B$. Here, however, we use blockwise bootstrap (Kunsch, 1989) to account for the dynamics of the signal; we use days as block.

2.3.2. Train-test split

How one splits the data into a training and testing set is critical. The split must indeed ensure that we test the ability of our global models (SLR_global, BRR_global, XGB_global, and XGB_interpol) to generalize in time and space. We, therefore, reserve 50 stations and 3 years for testing (2017–2019). Some of the data is thus discarded: (1) these 50 stations' data for years before 2017 and (2) the other stations' data for the years 2017 to 2019. The process is illustrated in Fig. 7(a)

For the local model, XGB_station, we use a different setup: for each of the 50 test stations, the years 2010 to 2016 are used for training, and the years 2017 to 2019 for testing, as illustrated in Fig. 7(b).

Finally, for the spatial model, XGB_spatial, the years 2017 to 2019 are used for both testing and training, as illustrated in Fig. 7(c)

Testing and training stations must fulfill several requirements:

[3] This decomposition can be easily derived if one remembers that RMSE is an estimate of $\left(\mathbb{E}[\epsilon^2]\right)^{1/2}$, SDE of $\left(\mathbb{E}[(\epsilon - \mathbb{E}[\epsilon])^2]\right)^{1/2}$ and MBE of $\mathbb{E}[\epsilon]$, where ϵ is the error.

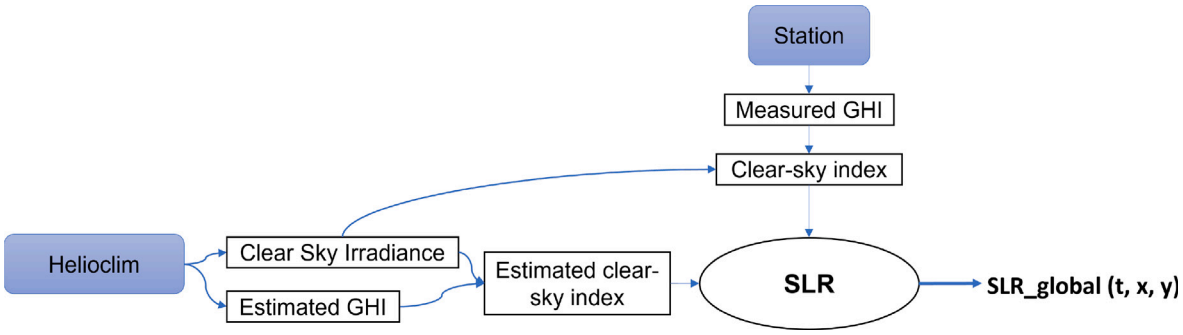


Fig. 6. Overview of the SLR_global model.

Table 1
Predictors used by each model. When passed time steps are used for a variable, the number of time steps used is indicated in brackets.

Predictors ↓ \ Models →	SLR_global	BRR_global	XGB_global	XGB_interpol	XGB_station	XGB_spatial
ghi from HC3		X (5)	X (5)	X (5)	X (5)	X (5)
k_c from HC3	X	X (5)	X (5)	X (5)	X (5)	X (5)
CSI from HC3		X	X	X	X	X
solar elevation angle		X	X	X	X	X
solar azimuth angle		X	X	X	X	X
time of day		X	X	X	X	X
day of year		X	X	X	X	X
latitude		X	X			X
longitude		X	X			X
altitude		X	X			X

Table 2
Hyper-parameters tuned for XGB models, with the range of values tested during cross-validation. The hyper-parameters selected for XGB_global by the cross-validation procedure are shown in bold.

Parameter name	Values tested	Description
max_depth	[5, **10**, 15]	Maximum depth of each regression tree
min_child_weight	[1, 5, **10**]	Minimum number of sample(s) in each leaf
subsample	[.3, .5, **1**]	Ratio of train data randomly picked to train each tree
learning_rate	[.05, .1, .5]	Weight of the contribution of each regression tree
n_estimators	[300, **500**]	Total number of regression trees used

1. Each test station must have enough data during the testing period so that station-wise performance analysis is relevant.
2. Each sub-model (i.e. developed for each individual station) used by the local model, XGB_station, must have enough training data.
3. Each sub-model (i.e. developed for each individual station) used by the global interpolated model, XGB_interpol, must have enough training data.

To guarantee 1., we require that each *test* station passes QC for more than 30% of the hours over the *testing* period. To guarantee 2., since XGB_station sub-models are trained on test stations (see Fig. 7(b)), we require that a *test* station passes QC for more than 30% of the hours over the *training* period. To guarantee 3., since XGB_interpol sub-models are trained on training stations, we require that each *training* station passes QC for more than 30% of the hours over the *training* period. The 50 test stations are chosen randomly among the stations fulfilling requirements 1 and 2. All the remaining stations fulfilling requirement 3 are taken as training stations.

Considering that nighttime data are flagged as failing QC (see Section 2.1.2), the threshold of 30% corresponds to an average of more than 2628 h of available daylight data, i.e. it can be considered a rather strict criterion.[4]

Of the 286 original measurement stations, 46 are removed by QC (see Section 2.1.2), 50 are reserved for testing and, out of the 190 remaining, 173 pass requirement 3. and are used as training stations. The locations of test and training stations are shown in Fig. 1 - we see that train and test stations are evenly spread across France.

It should be emphasized that while the training data may be different for different models, the test performances are evaluated on the exact same data set for all models. As discussed in Verbois et al. (2022), this is paramount for a proper benchmark.

2.3.3. Cross-validation

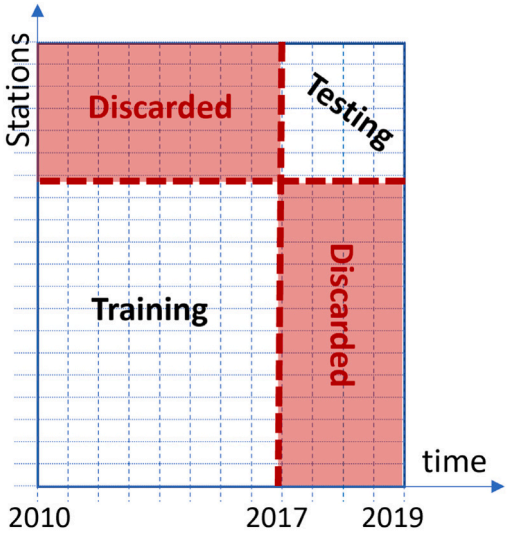
XGB_global, XGB_interpol and XGB_station have several hyper-parameters that need to be optimized, detailed in Table 2. This optimization must be done on the training set, using here cross-validation techniques (Hastie et al., 2009). In this work, the training stations are randomly grouped into 4 validation sets, each containing data from 43 or 44 stations for 2010–2017. These 4 sets are used to conduct 4-fold cross-validation.

BRR_global has two hyper-parameters (the weights of the L1 and L2 regularization terms), but uses a Bayesian framework to optimize them and thus does not require any validation. SLR_global does not have any hyper-parameters.

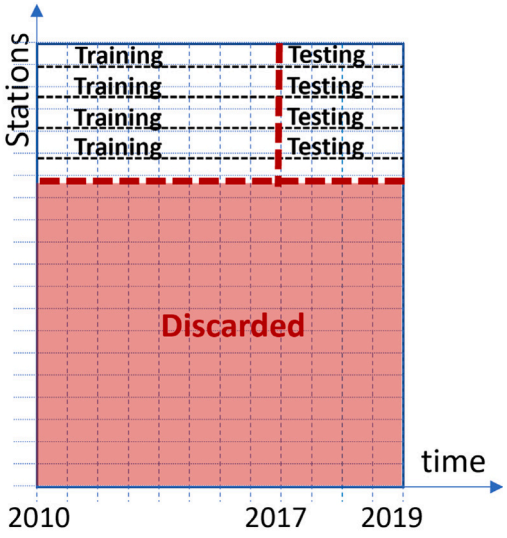
3. Results

This section is organized into three parts. In Section 3.1, we compare the overall and station-wise performance of our proposed MOS model, XGB_global, with that of HC3 and of the 5 other MOS models. In Section 3.2, we go further and perform a condition-dependent analysis of XGB_global and HC3 performances. Finally, in Section 3.3, we go beyond the MBE and SDE metrics and study the effect of our MOS model on the plausibility of the GHI estimations.

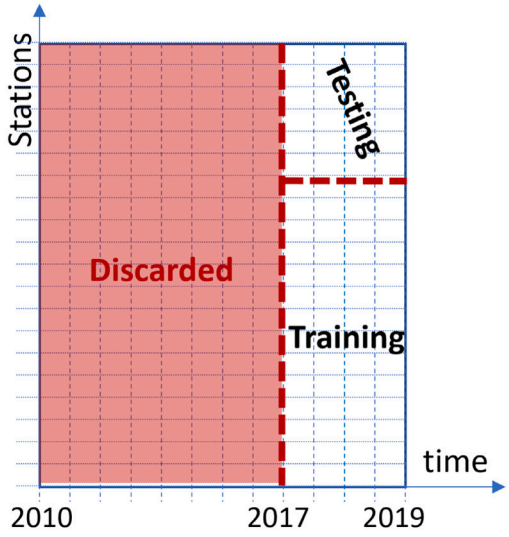
[4] Since Metropolitan France latitude is between 42° and 51°, hours with a mean hourly elevation angle above 10° constitute between 39% and 43% of all hours.



(a) Split for global models



(b) Split for local model



(c) Split for spatial model

Fig. 7. Illustration of the training/testing split setup, for the global models (XGB_global, BRR_global, SLR_global and XGB_interpol) in (a), for the local model (XGB_station) in (b) and for the spatial model (XGB_spatial) in (c).

3.1. Multi-model comparison

3.1.1. Overall performances

We first look at the overall performances, computed over the 50 test stations and 2 test years. Table 3 shows the overall MBE of HC3 and the 6 MOS models. HC3 has a bias of 13.2 W m^{-2} and all MOS models have similar MBEs, between −2 and 2 W m^{-2}, which is a significant reduction of HC3 bias. SLR_global MBE is statistically significantly lower than that of all other models, but the difference with other MOS models is likely negligible in practice as they are comparable to the measurements noise.[5]

Table 4 shows the overall SDE of HC3 and the 6 MOS models. The improvements over HC3 are not as important as for MBE, but there is a stronger differentiation between MOS models. In fact, the SDE seems to decrease with the increasing complexity of the algorithm used. With an SDE of 64.8 W m^{-2}, SLR_global, the simplest model, slightly deteriorate HC3 SDE — 64.6 W m^{-2} ; BRR_global performs somewhat better with a reduction of 3% of HC3 SDE; XGB_global, the most sophisticated model, reaches an overall SDE of 59.8 W m^{-2}, 8% less than HC3. The local model, XGB_station, and the spatial model, XGB_spatial, have an even lower SDE. This was expected since they benefit from additional information; the difference with XGB_global is, nonetheless, small.

[5] It is difficult to quantify the measurement noise precisely, because not all stations use the same pyranometer model and because pyranometers expected uncertainties are typically condition-dependent (Habte et al., 2016). Nonetheless, if we consider for example the Kipp&Zonen CMP11 – installed in several of the stations considered in this study – Strobel et al. (2009) estimate a higher bound of ±3.10%±1.10 W m^{-2} for the combined standard uncertainty. With irradiance values above 100 W m^{-2}, that is larger than the MOS MBEs.

Table 3
MBE for HC3 and MOS models, computed once for all test stations and test years. For each pair of models M_1, M_2, the statistical significance of $MBE_1 \neq MBE_2$ is tested; the resulting p-values ($\mathcal{P}(MBE_1 = MBE_2)$) are shown in the table. MBE statistically significantly non null are indicated by a *.

	MBE	HC3	SLR_global	BRR_global	XGB_global	XGB_interpol	XGB_station	XGB_spatial
	W m^{-2}	p-values						
HC3	13.2*	NA	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
SLR_global	1.5*	<0.01	NA	<0.01	<0.01	<0.01	<0.01	<0.01
BRR_global	1.9*	<0.01	<0.01	NA	0.02	<0.01	0.03	<0.01
XGB_global	2.0*	<0.01	<0.01	0.02	NA	0.07	0.25	<0.01
XGB_interpol	2.1*	<0.01	<0.01	<0.01	0.07	NA	0.32	<0.01
XGB_station	1.9*	<0.01	<0.01	0.03	0.25	0.32	NA	<0.01
XGB_spatial	−2.1*	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	NA

Table 4
SDE for HC3 and MOS models, computed once for all test stations and test years. For each pair of models M_1, M_2, the statistical significance of $SDE_1 \neq SDE_2$ is tested; the resulting p-values ($\mathcal{P}(SDE_1 = SDE_2)$) are shown in the table.

	SDE W m^{-2}	p-values						
		HC3	SLR_global	BRR_global	XGB_global	XGB_interpol	XGB_station	XGB_spatial
HC3	64.6	NA	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
SLR_global	64.8	<0.01	NA	<0.01	<0.01	<0.01	<0.01	<0.01
BRR_global	62.6	<0.01	<0.01	NA	<0.01	<0.01	<0.01	<0.01
XGB_global	59.8	<0.01	<0.01	<0.01	NA	<0.01	<0.01	<0.01
XGB_interpol	61.7	<0.01	<0.01	<0.01	<0.01	NA	<0.01	<0.01
XGB_station	59.0	<0.01	<0.01	<0.01	<0.01	<0.01	NA	0.01
XGB_spatial	59.1	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	NA

3.1.2. Station-wise performances

To better discriminate between the models and gain a deeper understanding of their (relative) performances, we break down the MBE and SDE per station. Fig. 8 shows the distribution of these metrics across the 50 test stations, using boxplots. We first notice that even though the overall MBE are ≤2.1 W m^{-2} (Table 3) for all MOS models, a relatively large bias remains for some stations. Furthermore, even though XGB_interpol has almost the same MBE as XGB_global on average, it is significantly higher for 3 stations. Lastly, the superiority in terms of bias reduction of the local model, XGB_station, is clearer: not only the mean MBE is the lowest (Table 3), but the distribution of MBE is also more tightly wrapped around 0.

Generally, however, the cross-model comparison is not easy in Fig. 8 because the cross-station variations of performances are more important than the cross-model variations. To properly evaluate the benefit of using XGB_global, we need to compare its MBE and SDE with every other model, independently for each station. To that end, the distributions of $\Delta|MBE|^{model}_{\text{XGB\_global}}$ and $\Delta SDE^{model}_{\text{XGB\_global}}$ across test stations are shown in Fig. 9, for $model \in$ {HC3, SLR_global, BRR_global, XGB_interpol, XGB_station, XGB_spatial}. XGB_global decreases HC3 MBE in 80% of the stations. In addition, while the reduction of MBE is as good as 47 W m^{-2} in one location, the deterioration never exceeds 13 W m^{-2}, as can be seen from the extreme points of the boxplots. In terms of SDE, the improvement is even more systematic, as HC3 sees its SDE increased by XGB_global in only one station (or 2%) and by less than 2 W m^{-2}.

XGB_global also shows improved performance over other global MOS models. Its MBE is indeed lower than that of SLR_global and BRR_global in more than 60% of the stations. However, it only has a lower MBE than XGB_interpol in 50% of the stations. Its superiority in terms of SDE is more significant as it outperforms the two linear models (SLR and BRR) in all stations but one and does better than XGB_interpol in 90% of them.

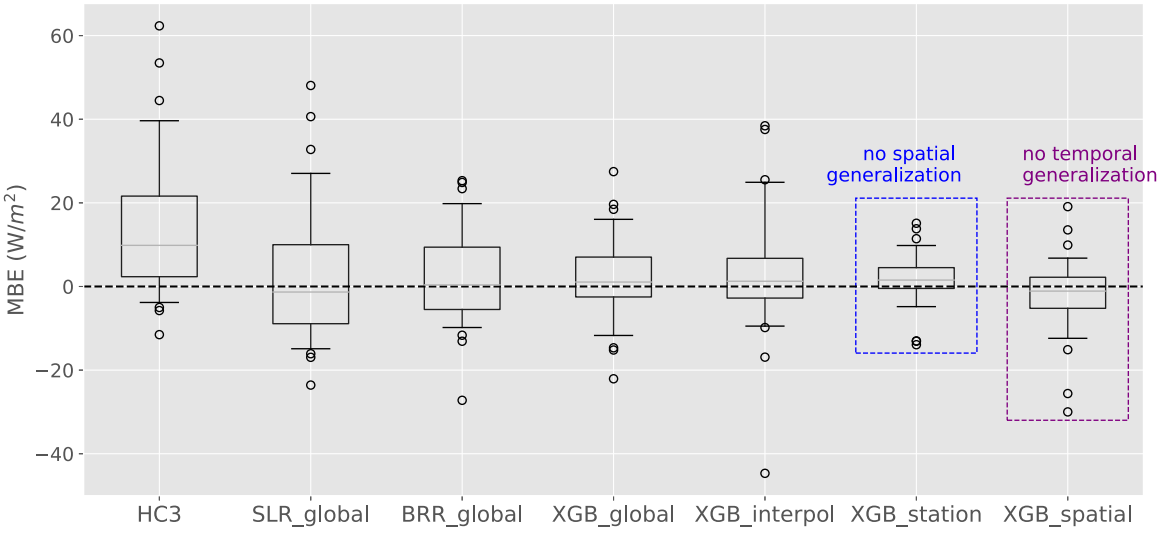
The comparison with the local MOS model, XGB_station, and the spatial model, XGB_spatial, is also insightful. Firstly, although the local model has a lower MBE and SDE on average (Tables 3 and 4), it has a higher SDE in 56% of the stations. This suggests that the global model is more robust than XGB_station, even though it does not perform quite as well on average. XGB_spatial, on the other hand, while it has only a slightly lower SDE than XGB_global overall (59.1 W m^{-2} vs 59.8 W m^{-2}), has a lower SDE in 84% of the test stations. The improvement, albeit small, is thus nearly systematic. The fact that the improvement of XGB_spatial over XGB_global is significantly more systematic than that of XGB_station over XGB_global suggests that, in the current setup, interpolation in time is more challenging than interpolation in space. One must keep in mind, however, that training and test stations are tightly interlaced (see Fig. 1); had we use a geographical split that enforce for example a minimum distance between training and test stations, the conclusion may have been different.

3.1.3. Monthly station-wise performances

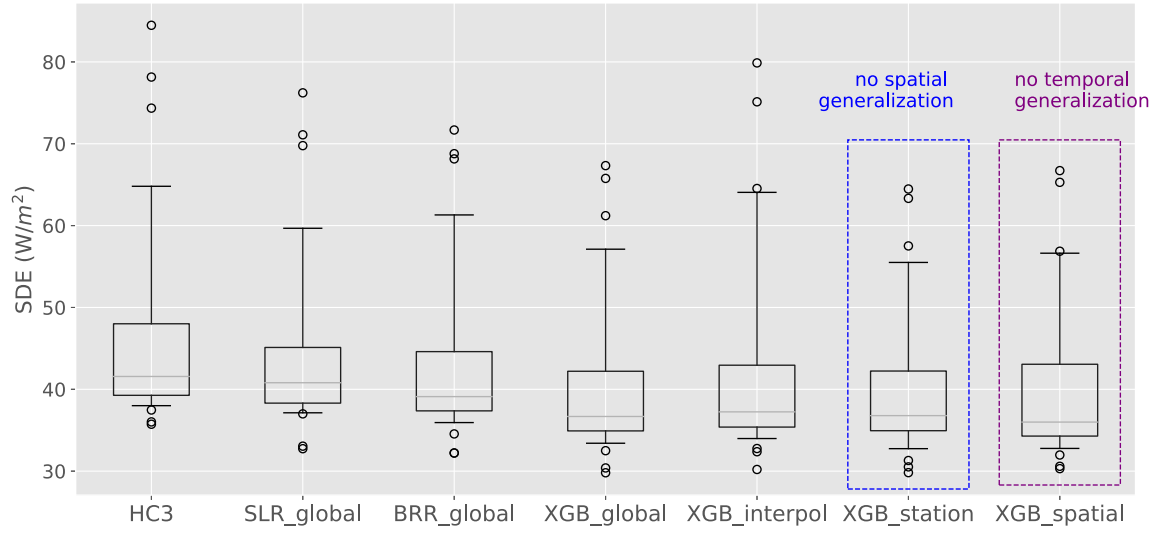
We can further analyze the systematicity of XGB_global's superiority over the other MOS models and HC3. To this end, we look at the distribution of the monthly station-wise performances. Similar to the station-wise analysis (Figs. 8 and 9), we are interested in the station and month-wise difference in SDE and MBE, rather than in absolute values. Fig. 10 thus shows the distribution of $\Delta|MBE|^{model}_{\text{XGB\_global}}$ and $\Delta SDE^{model}_{\text{XGB\_global}}$ for each month and station. Each sample used to build the boxplots in Fig. 10(b) is one value of ΔSDE for a given month in a given station.

These results confirm that XGB_global outperforms HC3 and other global MOS models in most cases, especially in terms of SDE. In particular, it improves HC3 SDE in 93.5% of months and stations. The local MOS model, XGB_stations, has a lower MBE than XGB_global in 61% of the cases, but a lower SDE in only 48.7% of the cases. This confirms that XGB_global is slightly more robust than XGB_station. The spatial model, XGB_local, on the other hand, has a lower MBE nearly half the time (49.3% of the cases), but a lower SDE in more than 70% of the cases. This confirms the conclusions drawn from Section 3.1.2: in the current setup, interpolation in time is more challenging than interpolation in space.

At this point in the analysis, we are confident that XGB_global is a better choice than other tested global MOS models, at least in



(a) MBE



(b) SDE

Fig. 8. Boxplots of stations' MBE and SDE. Each sample used to build a given boxplot is the MBE (resp. SDE) for a given model in a single test station during the test period. The whiskers of the boxplot indicate the 5 and 95% percentiles; the remaining samples are represented by points outside the box and its whiskers.

terms of MBE and SDE. Admittedly, the difference with other models – XGB_interpol in particular – is relatively small, but it is systematic. Furthermore, XGB_global compares well with two sophisticated local and spatial MOS models, XGB_station and XGB_global, even though they have additional information (in space or time, respectively).

In the rest of this paper, we will focus on the comparison between XGB_global and HC3, to understand whether the solar irradiance estimations of our proposed MOS model can be substituted to those of HC3.

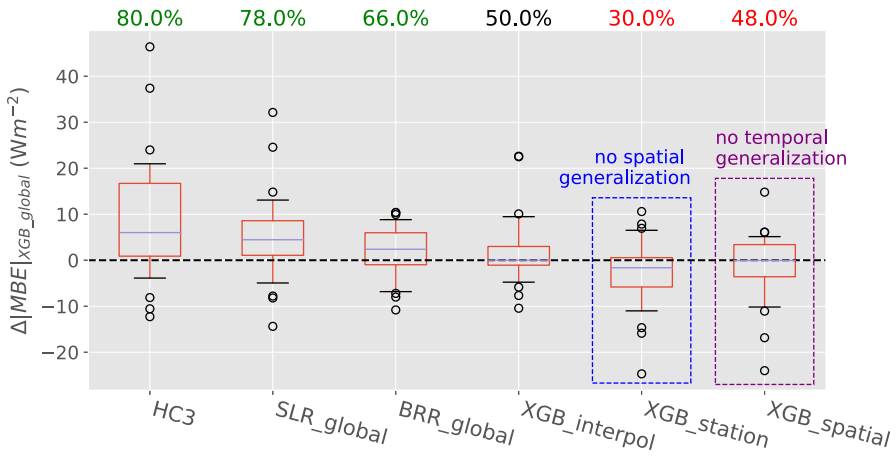
3.2. Condition-dependent analysis

In this section, we propose to look in more detail into the performances of XGB_global and HC3 to better understand the added value of the former. We first study the dependence of their performances on the sky conditions, and then verify if their error has a daily or seasonal dependence.

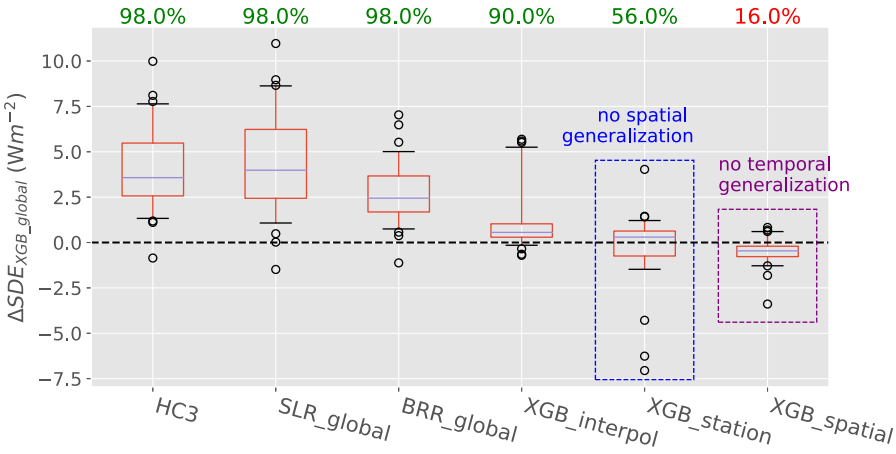
3.2.1. Dependence on the sky conditions

We first look at the performances of HC3 and XGB_global as a function of the sky conditions, for which we use the clear-sky index k_c as a proxy. Figs. 11(a) and 11(b) show the distribution of $\mathrm{MBE}_{\mathrm{HC3}}$, $\mathrm{MBE}_{\mathrm{XGB\_global}}$ and $\Delta|MBE|^{\mathrm{HC3}}_{\mathrm{XGB\_global}}$ as a function of observed k_c. HC3 generally overestimates the solar irradiance (i.e. $\mathrm{MBE}_{\mathrm{HC3}} > 0$) for $k_c \leq 0.8$ – in particular for $k_c \leq 0.2$ – and has a slightly negative bias ($\mathrm{MBE}_{\mathrm{HC3}} \lessapprox 0$) for $k_c > 0.8$. Because k_c is bounded by 0 and 1, we expect some underestimation for the highest k_c class and overestimation for the lowest k_c class.[6] The positive bias for $0.2 \leq k_c \leq 0.8$, however, cannot be attributed to mathematical artifacts and indicates that HC3 underestimates clouds. A known issue with empirical satellite retrieval methods such as the one used for HC3 is that when a higher cloud casts a shadow on a lower cloud, the apparent reflectance of the latter decreases and it is thus classified as a clear-sky. This could perhaps

[6] Since k_c is bounded by 0 and 1, estimations can never fall to the right of $[0.8, 1]$ or to the left of $[0, 0.1]$.



(a) $\Delta|MBE|_{\mathrm{XGB\_global}}$



(b) $\Delta SDE_{\mathrm{XGB\_global}}$

Fig. 9. Boxplot of $\Delta|MBE|_{\mathrm{XGB\_global}}$ and $\Delta SDE_{\mathrm{XGB\_global}}$ for HC3 and benchmark MOS models. Each sample used to build a given boxplot is the $\Delta|MBE|_{\mathrm{XGB\_global}}$ (resp. $\Delta SDE_{\mathrm{XGB\_global}}$) for a given model in a single test station during the test period. The percentage of stations for which XGB_global outperforms other models is shown above the boxplot.

explain some of the positive bias of HC3, but further analysis – beyond the scope of this paper – is needed to fully understand the causes.

XGB_global can significantly reduce the positive bias of HC3: it is beneficial to 100% of the test stations for very cloudy skies ($k_c \leq 0.2$) and to most of them for $k_c < 0.8$. For $k_c > 0.8$, however, $\mathrm{MBE}_{\mathrm{HC3}}$ is slightly negative, and XGB_global further decreases it, i.e. deteriorates the estimations, for 96% of the test stations. We suspect that this deterioration stems from the tendency of least-square regression models to group their prediction around the mean: in order to reduce the mean squared error, it is indeed beneficial to avoid extreme values – in this case, k_c close do 1. Typically, a similar pattern would be expected for small values of k_c. However, in the present scenario, HC3 performs extremely poorly for k_c values approaching zero, so XGB_global is still able to mitigate its bias.

The distributions of $\mathrm{SDE}_{\mathrm{HC3}}$, $\mathrm{SDE}_{\mathrm{XGB\_global}}$ and $\Delta SDE^{\mathrm{HC3}}_{\mathrm{XGB\_global}}$ as a function of the k_c are given in Figs. 11(c) and 11(d). Somewhat surprisingly, the improvement pattern is almost the opposite of that of MBE: XGB_global reduces $\mathrm{SDE}_{\mathrm{HC3}}$ most significantly for $0.4 \leq k_c \leq 0.8$ and generally for all $k_c \geq 0.4$. For overcast skies ($k_c \leq 0.4$), on the other hand, XGB_global slightly deteriorates the SDE of HC3. The contrast between MBE and SDE may represent the well-known bias–variance trade-off that statistical learning approaches must manage.

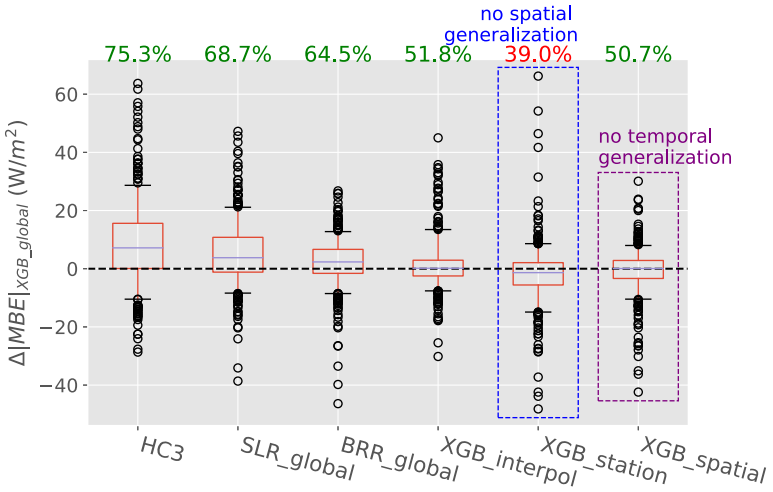
Because XGB_global has an opposite impact on MBE and SDE depending on the k_c class, we also look at the RMSE (that combines both MBE and SDE), in Figs. 11(e) and 11(f). We see that overall, XGB_global consistently and significantly improves HC3 for all $k_c \leq 0.8$ and only slightly deteriorates it for $k_c \geq 0.8$.

3.2.2. Dependence on the time of day and year

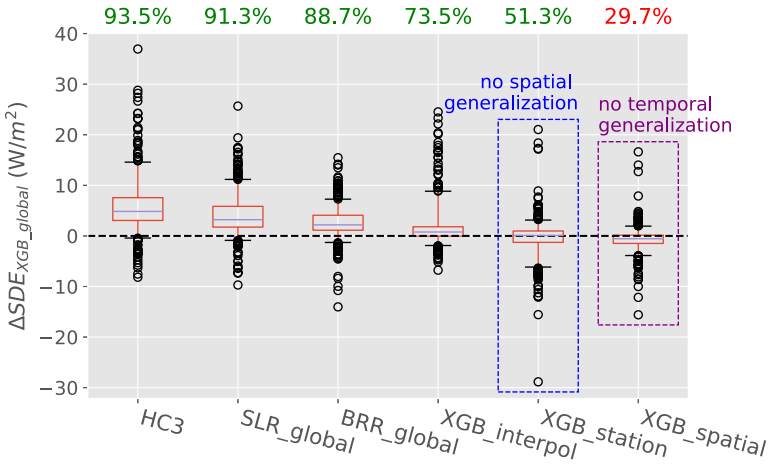
Because of the daily cycle of solar irradiance, it is interesting to look at the daily patterns of MBE and SDE. As the longitudes of the stations range between −4° and 8°, we use the true solar time (TST)[7] to normalize across stations. Figs. 12(a) and 12(b) show the distribution of $\mathrm{MBE}_{\mathrm{HC3}}$, $\mathrm{MBE}_{\mathrm{XGB\_global}}$ and $\Delta|MBE|^{\mathrm{HC3}}_{\mathrm{XGB\_global}}$ as a function of the true solar time. Both HC3 and XGB_global have a relatively constant MBE across the day. As this is an absolute value, the relative bias is significantly higher in the morning and evening. XGB_global reduces HC3 bias consistently during the day, for most stations. We note, however, that it decreases the MBE of more stations in this afternoon. This phenomenon may be attributed to the fact that XGB_global depends on past estimates of HC3 (see Section 2.2.1), which are not informative during the morning hours – integrating outputs of a night cloud-detection algorithm in the MOS inputs could help alleviate this drawback.

Figs. 12(c) and 12(d) show the distribution of $\mathrm{SDE}_{\mathrm{HC3}}$, $\mathrm{SDE}_{\mathrm{XGB\_global}}$, and $\Delta SDE^{\mathrm{HC3}}_{\mathrm{XGB\_global}}$ across stations as a function of the time of day.

[7] The true solar time is based on the apparent position of the sun at a given location. For example, it is noon TST when the sun is at its zenith at that location.



(a) Monthly variations of $\Delta|MBE|_{\mathrm{XGB\_global}}$



(b) Monthly variations of $\Delta SDE_{\mathrm{XGB\_global}}$.

Fig. 10. Boxplot of change in monthly station MBE and SDE between XGB_global and HC3 and the other MOS models. Each sample used to build the boxplots is the value of ΔSDE or ΔMBE for a given month in a given station. The percentage of stations for which XGB_global outperforms other models is shown above the boxplot.

Contrary to MBE, $\mathrm{SDE}_{\mathrm{HC3}}$ is not constant, and peaks around noon. XGB_global follows the same pattern and, as for MBE, consistently reduces the SDE of HC3 across the day. Unlike for MBE, however, the number of stations that see their SDE decreased by the statistical correction is almost constant across the day.

It can be insightful to also check for yearly patterns in the (relative) performances of HC3 and XGB_global. The distributions of $\mathrm{MBE}_{\mathrm{HC3}}$, $\mathrm{MBE}_{\mathrm{XGB\_global}}$, and $\Delta|MBE|^{\mathrm{HC3}}_{\mathrm{XGB\_global}}$ as a function of the month are shown in Figs. 13(a) and 13(b). Their SDE counterparts are shown in Figs. 13(c) and 13(d). HC3 displays a small seasonal bias, with a slightly lower MBE in summer, and a more significant seasonality in SDE, with larger values in summer.

XGB_global manages to remove some of the seasonality of $\mathrm{MBE}_{\mathrm{HC3}}$, but cannot completely erase that of $\mathrm{SDE}_{\mathrm{HC3}}$: a substantial seasonal variability remains. $\Delta SDE^{\mathrm{HC3}}_{\mathrm{XGB\_global}}$ is nonetheless larger in summer, when $\mathrm{SDE}_{\mathrm{HC3}}$ is largest; that is a desirable behavior as it contributes to even out the estimations error.

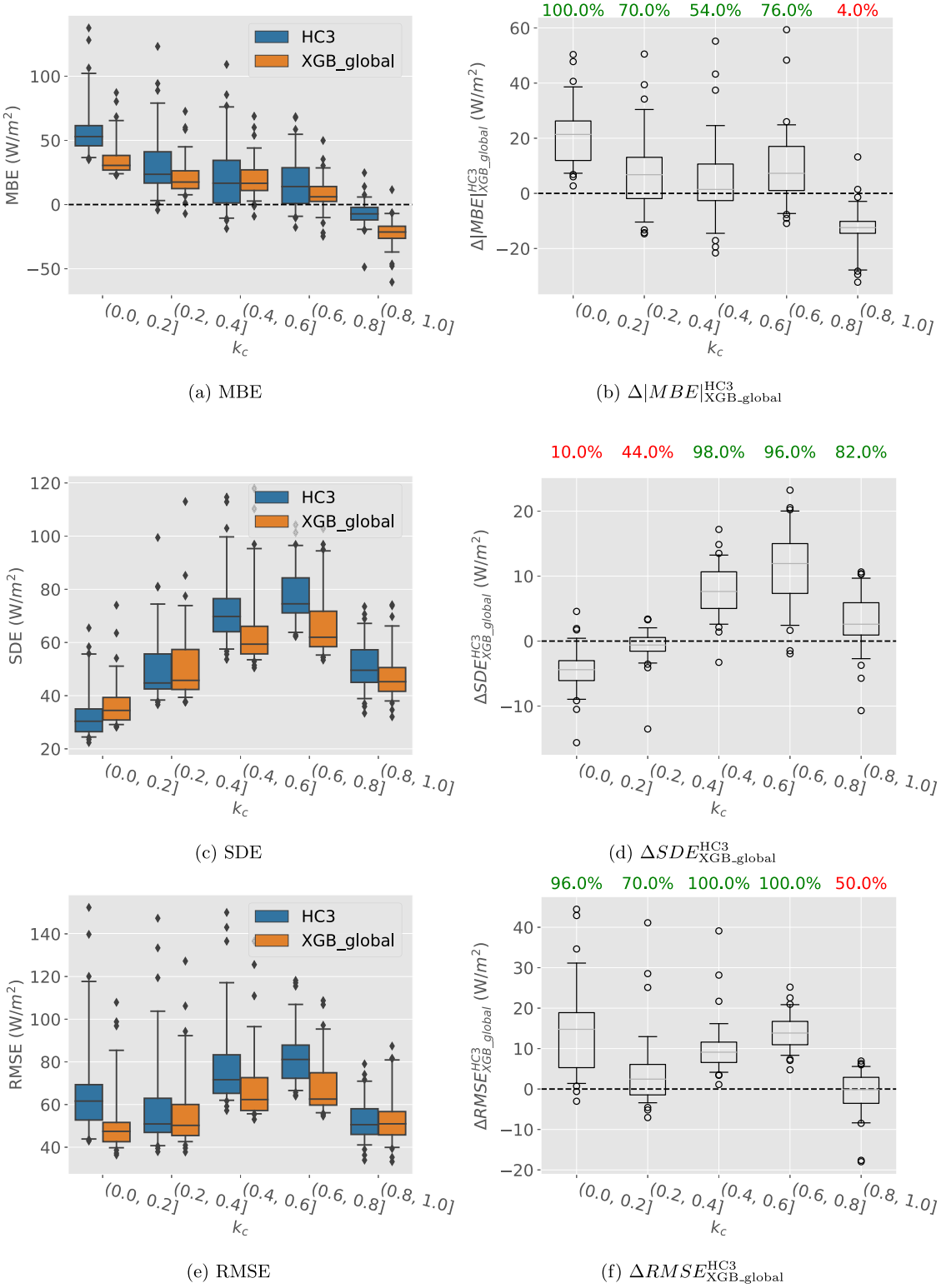
3.3. Plausibility

In the two previous result subsections, we thoroughly evaluated the impact of our proposed model, XGB_global, on the precision and accuracy of the irradiance estimations. These two metrics have some limitations (Vallance et al., 2017). In particular, they do not evaluate the plausibility of the models, i.e. how realistic the SSI estimations are Verbois et al. (2020). In this subsection, we thus propose to analyze the impact of the statistical correction on the plausibility of the irradiance estimations.

3.3.1. Empirical distributions of estimated k_c

We first look at the plausibility of the estimations with a "static" perspective, looking only at the empirical distributions of the clear-sky index k_c. The distribution of k_c is shown in Fig. 14 for HC3, XGB_global, and as measured at the stations; a single distribution is shown for the whole test set – 50 stations and 3 years.

XGB_global distribution of k_c is closer to the observations than HC3 for $k_c < 0.6$. Nonetheless, the MOS model still under-predicts overcast skies ($k_c < 0.2$), even if it also improves over HC3 in that range. On the contrary, for clearer skies ($k_c > 0.8$), the plausibility of irradiance estimations is slightly degraded by XGB_global. All three histograms exhibit a peak for $k_c > 0.8$, but the peak is too narrow for XGB_global. This is in agreement with the k_c-dependant results presented in Section 3.2, which show that XGB_global has a slightly higher RMSE and MBE than HC3 for $k_c > 0.8$.



(a) MBE (b) $\Delta|MBE|^{\text{HC3}}_{\text{XGB\_global}}$

(c) SDE (d) $\Delta SDE^{\text{HC3}}_{\text{XGB\_global}}$

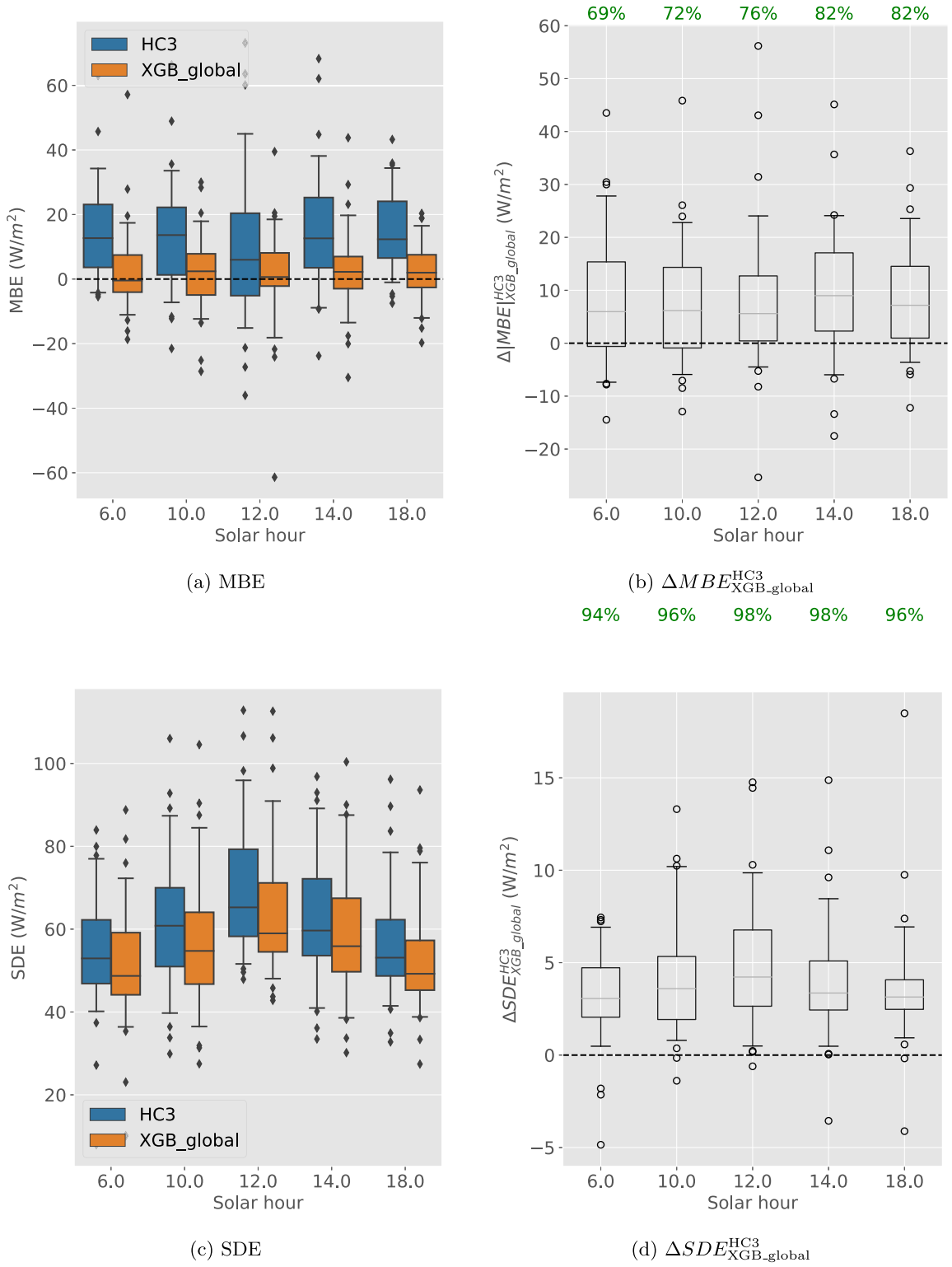
(e) RMSE (f) $\Delta RMSE^{\text{HC3}}_{\text{XGB\_global}}$

Fig. 11. Boxplot of SDE, $\Delta SDE^{\text{HC3}}_{\text{XGB\_global}}$, MBE, $\Delta|MBE|^{\text{HC3}}_{\text{XGB\_global}}$, RMSE, and $\Delta RMSE^{\text{HC3}}_{\text{XGB\_global}}$ for all stations, as a function of k_c. The percentage of stations for which XGB_global outperforms HC3 is indicated above each boxplot.

3.3.2. *Empirical distributions of ramp rate*

We then consider the dynamics of the signal, through the ramp rate of k_c, $r_{k_c}(t) = k_c(t) - k_c(t-1)$. The ramp-rate is a simple proxy for the variability of a k_c time series; it is also important for technical considerations, as a high ramp rate can impact the operation of an electrical grid or a photovoltaic power plant.

The distribution of the k_c ramp rates is shown for HC3, XGB_global and the measurements in Fig. 15. XGB_global positive ramp rates are almost always closer to the measurements than HC3. In particular, XGB_global partially compensates for the over-occurrence of positive ramp rates observed with HC3 and almost fully corrects the under-occurrence of HC3 negative ramp rates.



(a) MBE

(b) $\Delta MBE^{\mathrm{HC3}}_{\mathrm{XGB\_global}}$

(c) SDE

(d) $\Delta SDE^{\mathrm{HC3}}_{\mathrm{XGB\_global}}$

Fig. 12. Boxplot of MBE, $\Delta|MBE|^{\mathrm{HC3}}_{\mathrm{XGB\_global}}$, SDE and $\Delta SDE^{\mathrm{HC3}}_{\mathrm{XGB\_global}}$ across the 50 test stations, as a function of the true solar time. The percentage of stations for which XGB_global outperforms HC3 is indicated above each boxplot.

3.3.3. Spatial patterns

Finally, we want to probe the spatial behavior of the signals. To that end, we compute the correlation between every pair of test stations and group the results in 10 bins, depending on the pairwise geodesic distance. The average pairwise correlation for each bin is estimated using the sample mean estimator, whose standard deviation can also be estimated (see Appendix). The results are shown in Fig. 16, where the 90% confidence intervals are also represented. We see that HC3 and XGB_global pairwise correlation is generally higher than in the observations. Furthermore, we notice that for distances above 100 or 200 km, XGB_global pairwise correlation is slightly higher than that of HC3. The difference is not large, but it is statistically significant, and we can conclude that XGB_global slightly deteriorates the spatial behavior of HC3.

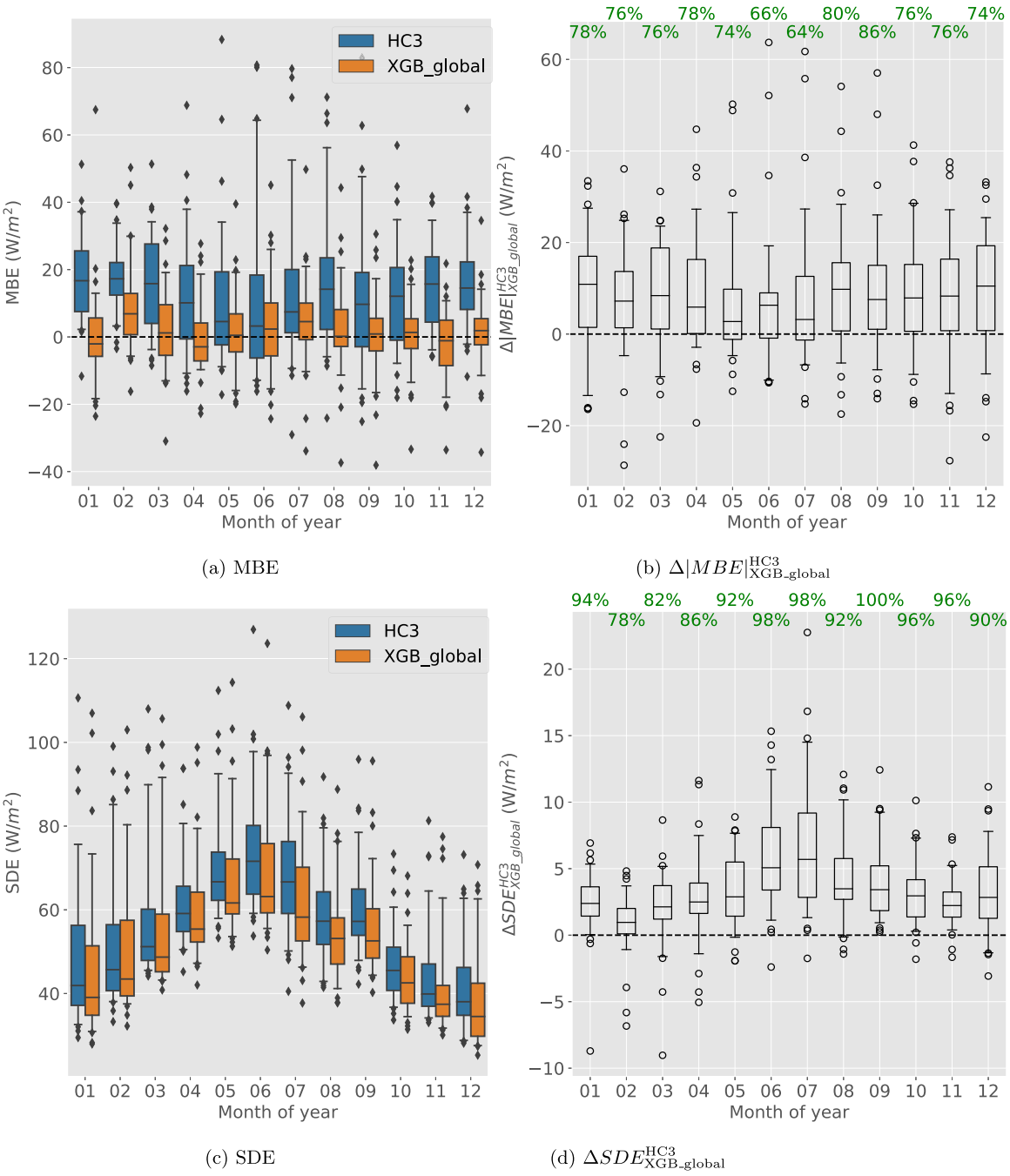


Fig. 13. Boxplot of station MBE, SDE, $\Delta|MBE|^{\text{HC3}}_{\text{XGB\_global}}$, and $\Delta SDE^{\text{HC3}}_{\text{XGB\_global}}$ as a function of the month of the year. The percentage of month × stations for which XGB_global outperforms HC3 is indicated for each month in green, above the boxplot.

A known pitfall of statistical post-processing in energy meteorology is that post-processing algorithms sometimes produce overly smooth estimations (Vallance et al., 2017). We saw in Section 3.3.2 that XGB_global does not deteriorate the *temporal* dynamics of the estimation. However, a possible explanation for its higher-than-expected *spatial* correlation is that it does smooth the signal in space.

While it may not impact the precision or accuracy of the GHI estimations, spatial smoothing might cause an underestimation of the variability and variety of GHI in a given region. This can impact resource management strategies at several time scales. Fully understanding and precisely describing the consequences of spatial smoothing on, e.g., a TSO, is however out of the scope of this paper.

To further understand the effect of XGB_global on the spatial characteristics of the SSI, we inspect the whole irradiance field produced by XGB_global, i.e. not only at the location of the measurement stations. Fig. 17 shows the average GHI over France for HC3 (a) and XGB_global (b), as well as the yearly average of the difference between HC3 and XGB_global (c). Spatial artifacts – latitude–longitude discontinuities – are clearly visible in Fig. 17(c) and can even be distinguished in Fig. 17(b).

They are likely caused by the regression trees underlying the extreme gradient boosting algorithm (XGB). As explained in Section 2.2.1, XGB prediction is based on a large number of regression trees, that partition the predictor space in rectangles. Since there are relatively few values of latitude and longitude (compared to e.g., values of k_c and GHI), it is possible to get steps in XGB estimation function where the latitude or longitude matches some of the training stations. A similar issue was described in Konstantinov et al. (2021).

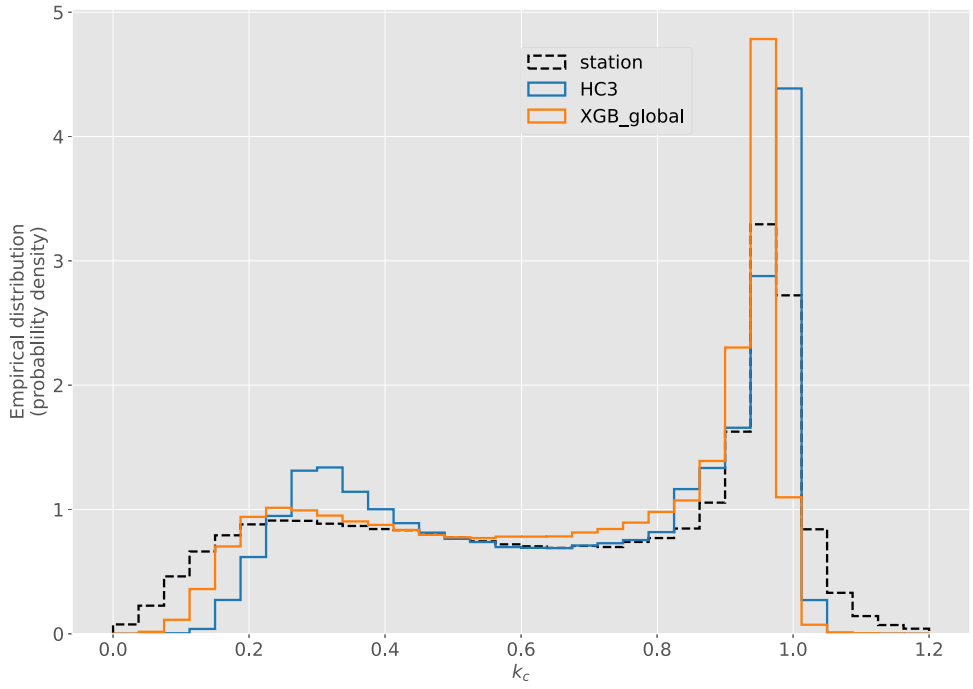


Fig. 14. Distribution of k_c over the 50 test stations and 2 test years for HC3 (solid orange), XGB_global (solid green) and as measured (dash black) for all stations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

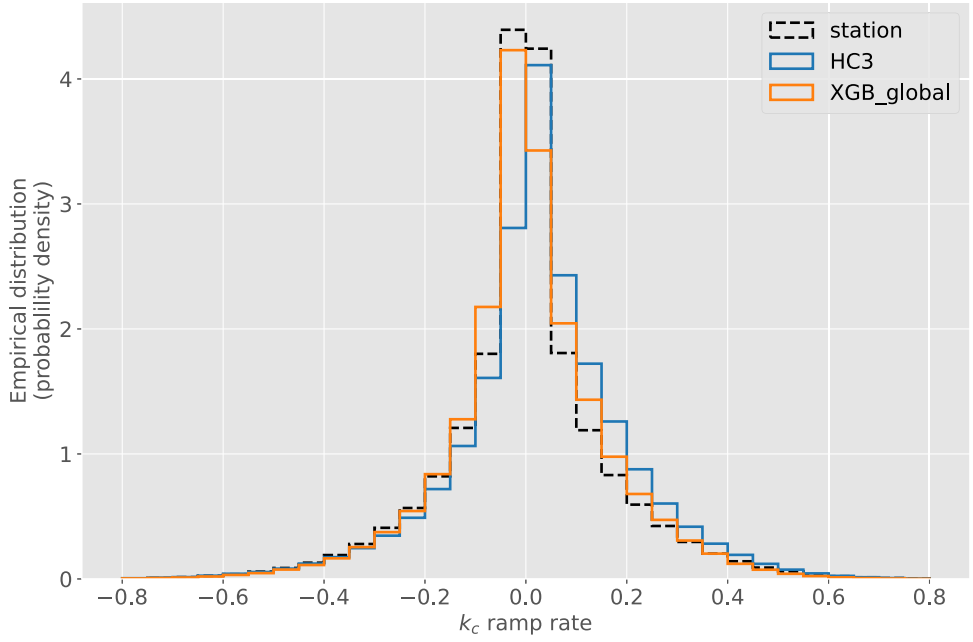


Fig. 15. Distribution of k_c ramp rates over the 50 test stations and 2 test years for HC3 (solid orange), XGB_global (solid green) and as measured (dash black) for all stations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The impact of such artifacts on the *quality* of the estimations is not clear, and requires further investigation.

4. Conclusions and perspectives

In this work we presented XGB_global, a new global adaptation technique of satellite-based estimations of SSI that extrapolates both in space and time from a set of pyranometers. The technique was designed so that it can be used operationally to improve hourly HelioClim3 maps of irradiance on-the-fly.

We benchmarked this correction model against several alternative statistical approaches. In particular, we showed that our proposed single location-aware model outperforms a combination of multiple local models with interpolation. We demonstrated the benefit of using a non-linear regression model as well. XGB_global was also compared to a local correction model – or site adaptation technique – and to a spatial correction model, based on the same regression algorithm. As expected, our global model does not perform as well as the local and spatial model on average. However, XGB_global turned out to be slightly more robust that the local model; this is perhaps because the global model benefits from a much larger training set.

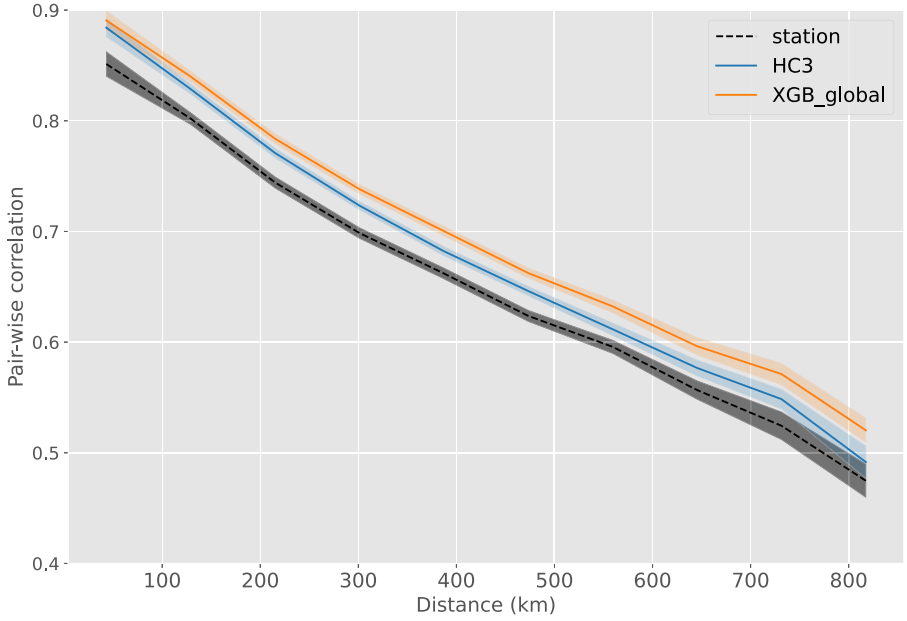


Fig. 16. Mean pairwise correlation as a function of the distance between pairs of stations for HC3, XGB_global and measured irradiance. The 90% confidence intervals are represented for each curve by a shaded area.

The performances of our proposed model, XGB_global, were then thoroughly compared with those of HelioClim3 (HC3). We found the improvement in both MBE, SDE, and, thus, RMSE, to be quite systematic and to hold for various conditions, in particular for most sky conditions. Admittedly, XGB_global does not improve HC3 for clear-sky conditions and even slightly increase the estimations' bias under such conditions. We hypothesize that the model benefits from a relatively large number of predictors and thereby can adapt its correction to the conditions. The plausibility of XGB_global and HC3 were also compared: corrected estimations were at least as plausible as HC3 estimations as long as we looked at time series.

When we considered the spatial properties of the estimations, however, XGB_global appeared to slightly deteriorate the plausibility of HC3. Perhaps more critically, it also introduced spatial discontinuities in the irradiance map. While it is not obvious how such artifacts impact the quality of the estimations, they would likely affect the confidence that potential users would put in them. This result also underlines the importance to go beyond classical metrics such as RMSE, SDE, and MBE when analyzing the performance of a model.

This work is a proof of concept. We showed that, under certain conditions, machine-learning-based global adaptation technique can significantly outperform traditional approaches. We also demonstrated the benefit of both non-linearity and large predictor space. An extreme gradient boosting algorithm (XGB) was used, but it is the authors opinion that other sophisticated non-linear regression algorithms could perform similarly well with the same predictors.

Admittedly, the conditions of the experiment were favorable. Firstly, the model benefited from a relatively dense network of measurement stations. Not all regions of the world enjoy such a network, and future work should investigate the sensitivity of the model to the network density. Secondly, the model training and testing was restricted to metropolitan France, which covers less than 1000 by 1000 km. It is not sure that a single global model could work for significantly larger areas. Extending our proposed approach to a continent or beyond might require some sort of partitioning based on geographical or climatic zones. This aspect should also be explored in future studies.

Finally, the spatial nature of the estimations require more work and should be investigated in further studies. In particular, developing an objective metric to measure spatial discontinuities would be useful, and could help better understand the impact they may have on downstream usages. Other algorithms should be tested, to try and get rid of the spatial artifacts. An alternative gradient boosting model, proposed by Konstantinov et al. (2021) uses partially randomized regression trees to avoid discontinuities in the predictor space, and would thus be a good candidate. Neural networks are also less likely to introduce discontinuities, by design, and should be tested — convolutional neural networks could be particularly appropriate in this context, as they are designed to handle spatial information. Lastly, it would be interesting to explore the potential of geo-statistical regression techniques such as Kriging.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank Augustin Touron from RTE for thoroughly reviewing this manuscript.

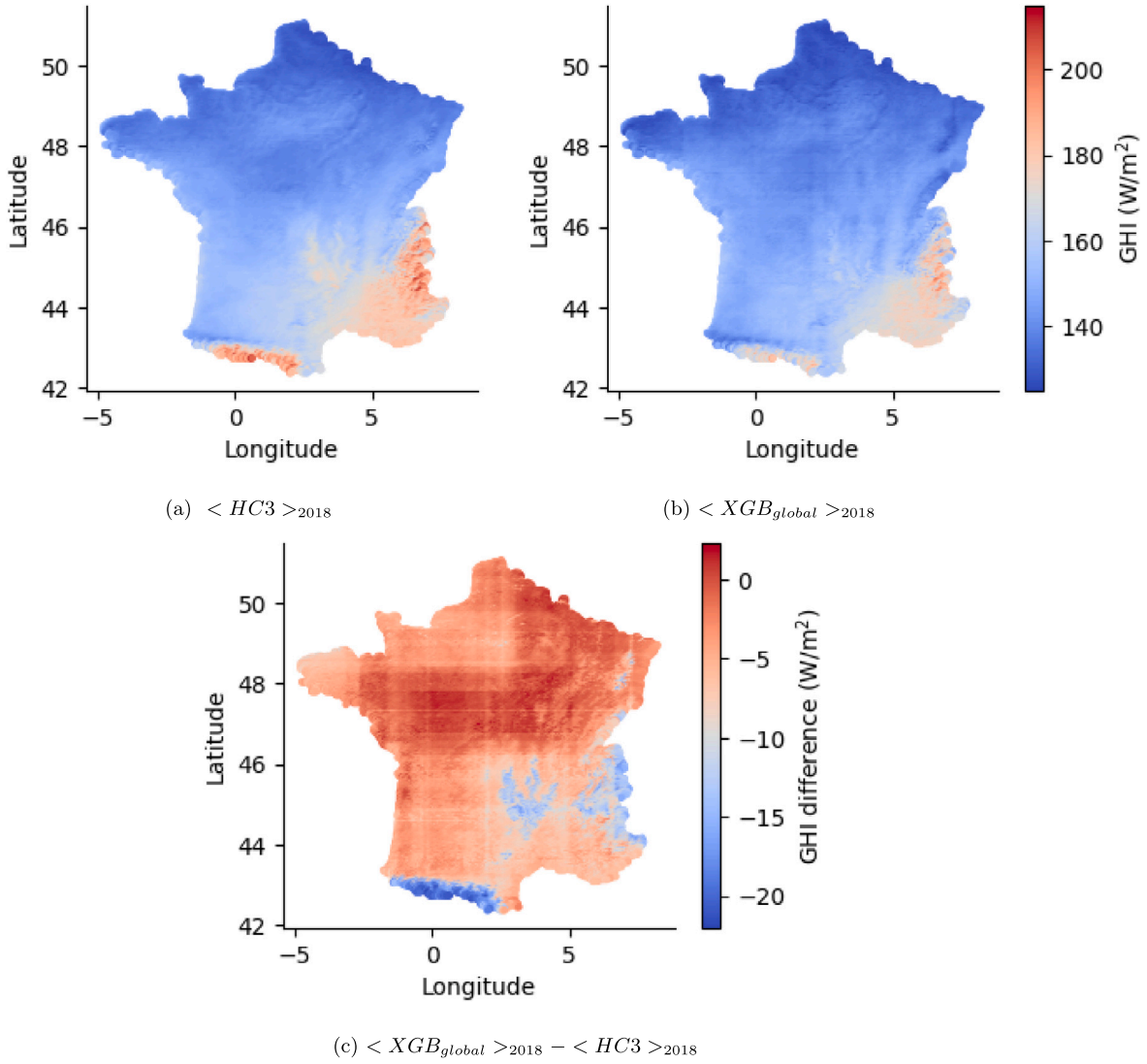
The work of Hadrien Verbois, Yves-Marie Saint-Drenan, and Philippe Blanc was supported by the SciDoSol chair.

The contribution of Météo-France to this work is being carried out as a part of the Smart4RES project (European Union's Horizon 2020, No. 864 337). The sole responsibility of this publication lies with the authors. The European Union is not responsible for any use that may be made of the information contained therein. All authors approved the version of the manuscript to be published.

Appendix. Estimating the standard deviation of the sample mean

Given a random variable x, with average μ and standard deviation σ, and a set of samples $\{x_i\}$, $i \in \{1..N\}$, the sample mean $\bar{x}$ define in Eq. (A.1) is an unbiased estimator of $\mathbb{E}[x]$.

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{A.1}$$



(a) $< HC3 >_{2018}$ (b) $< XGB_{global} >_{2018}$

(c) $< XGB_{global} >_{2018} - < HC3 >_{2018}$

Fig. 17. Mean GHI estimated by HC3 (a), XGB_global (b) for the test year 2018, and the yearly difference (c).

The variance of $\bar{x}$ is given by:

$$var(\bar{x}) = \frac{\sigma^2}{N} \tag{A.2}$$

N is known and σ^2 can be estimated by the sample variance $\widehat{\sigma^2}$:

$$\widehat{\sigma^2} = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2 \tag{A.3}$$

Eqs. (A.1) and (A.3) are valid for all distributions. To estimate the 90% confidence interval for $\bar{x}$, however, we assume that this sample mean estimator is Gaussian, so that we can use corresponding look-up tables between variance and percentiles.

References

Alessandrini, Stefano, Davò, F., Sperati, S., Benini, M., Delle Monache, Luca, 2014. Comparison of the economic impact of different wind power forecast systems for producers. Adv. Sci. Res. (ISSN: 1992-0636) 11 (1), 49–53. http://dx.doi.org/10.5194/asr-11-49-2014.

Antonanzas, J., Pozo-Vázquez, D., Fernandez-Jimenez, L.A., Martinez-de Pison, F.J., 2017. The value of day-ahead forecasting for photovoltaics in the Spanish electricity market. Sol. Energy 158, 140–146.

Babar, Bilal, Luppino, Luigi Tommaso, Boström, Tobias, Anfinsen, Stian Normann, 2020. Random forest regression for improved mapping of solar irradiance at high latitudes. Sol. Energy 198, 81–92.

Badosa, J., Wood, J., Blanc, P., Long, C.N., Vuilleumier, L., Demengel, D., Haeffelin, M., 2014. Solar irradiances measured using SPN1 radiometers: Uncertainties and clues for development. Atmos. Meas. Tech. 7 (12), 4267–4283. http://dx.doi.org/10.5194/amt-7-4267-2014.

Blanc, Philippe, Espinar, Bella, Gschwind, Benoˆıt, Ménard, Lionel, Thomas, Claire, Wald, Lucien, 2011a. High spatial resolution solar atlas in provence-alpes-cote d'azur. In: ISES Solar World Congress 2011. pp. paper–34552.

Blanc, Philippe, Gschwind, Benoît, Lefèvre, Mireille, Wald, Lucien, 2011b. The HelioClim project: Surface solar irradiance data for climate applications. Remote Sens. (ISSN: 20724292) 3 (2), 343–361. http://dx.doi.org/10.3390/rs3020343.

Blanc, Philippe, Remund, Jan, Vallance, Loïc, 2017. Short-term solar power forecasting based on satellite images. Renewable Energy Forecasting: from Models to Applications. Elsevier Ltd, pp. 179–198. http://dx.doi.org/10.1016/B978-0-08-100504-0.00006-8.

Blanc, Philippe, Wald, Lucien, 2012. The SG2 algorithm for a fast and accurate computation of the position of the Sun for multi-decadal time period. Sol. Energy 86 (10), 3072–3083.

Brancucci Martinez-Anido, Carlo, Botor, Benjamin, Florita, Anthony R., Draxl, Caroline, Lu, Siyuan, Hamann, Hendrik F., Hodge, Bri-Mathias Mathias, 2016. The value of day-ahead solar power forecasting improvement. Sol. Energy (ISSN: 0038092X) 129, 192–203. http://dx.doi.org/10.1016/j.solener.2016.01.049.

Chen, Tianqi, Guestrin, Carlos, 2016. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining. pp. 785–794.

Cornejo-Bueno, L., Casanova-Mateo, C., Sanz-Justo, J., Salcedo-Sanz, S., 2019. Machine learning regressors for solar radiation estimation from satellite data. Sol. Energy (ISSN: 0038092X) 183 (March), 768–775. http://dx.doi.org/10.1016/j.solener.2019.03.079.

da Graça, Guilherme Carrilho, Augusto, André, Lerer, Maria M, 2012. Solar powered net zero energy houses for southern Europe: Feasibility study. Sol. Energy 86 (1), 634–646.

Davy, Robert J., Huang, Jing R., Troccoli, Alberto, 2016. Improving the accuracy of hourly satellite-derived solar irradiance by combining with dynamically downscaled estimates using generalised additive models. Sol. Energy (ISSN: 0038092X) 135, 854–863. http://dx.doi.org/10.1016/j.solener.2016.06.052.

Espinar, Bella, Blanc, Philippe, Wald, Lucien, Gschwind, Benoît, Ménard, Lionel, Wey, Etienne, Thomas, Claire, Saboret, Laurent, 2012. HelioClim-3: A near-real time and long-term surface solar irradiance database. In: Workshop on "Remote Sensing Measurements for Renewable Energy".

Espinar, B., Blanc, Philippe, Wald, Lucien, Hoyer-Klick, Carsten, Marion Schroedter Homscheidt, Thomas Wanderer, 2013. On quality control procedures for solar radiation and meteorological measures, from subhourly to montly average time periods. In: EGU General Assembly 2012, Apr 2012, Vienne, Austria, no. 305. 262892.

Habte, Aron, Sengupta, Manajit, Andreas, Afshin, Wilcox, Stephen, Stoffel, Thomas, 2016. Intercomparison of 51 radiometers for determining global horizontal irradiance and direct normal irradiance measurements. Sol. Energy (ISSN: 0038-092X) 133, 372–393. http://dx.doi.org/10.1016/j.solener.2016.03.065.

Hastie, Trevor, Tibshirani, Robert, Friedman, Jerome, 2009. The Elements of Statistical Learning, second ed. In: Springer Series in Statistics, Springer New York, New York, NY, ISBN: 978-0-387-84858-7, p. 282. http://dx.doi.org/10.1007/978-0-387-84858-7.

Ineichen, Pierre, 2014. Long term satellite global, beam and diffuse irradiance validation. Energy Procedia (ISSN: 18766102) 48, 1586–1596. http://dx.doi.org/10.1016/j.egypro.2014.02.179.

Jimenez, Pedro A., Hacker, Joshua P., Dudhia, Jimy, Haupt, Sue Ellen, Ruiz-Arias, Jose A., Gueymard, Chris A., Thompson, Gregory, Eidhammer, Trude, Deng, Aijun, 2016. WRF-solar: Description and clear-sky assessment of an augmented NWP model for solar power prediction. Bull. Am. Meteorol. Soc. (ISSN: 0003-0007) 97 (7), 1249–1264. http://dx.doi.org/10.1175/BAMS-D-14-00279.1.

Kaggle, 2022. Kaggle competition. https://www.kaggle.com/. (Online; Accessed 03 June 2022).

Konstantinov, Andrei, Utkin, Lev, Muliukha, Vladimir, 2021. Gradient boosting machine with partially randomized decision trees. In: 2021 28th Conference of Open Innovations Association. FRUCT, pp. 167–173. http://dx.doi.org/10.23919/FRUCT50888.2021.9347631.

Kunsch, Hans R., 1989. The Jackknife and the Bootstrap for General Stationary Observations. Ann. Statist. 17 (3), 1217–1241. http://dx.doi.org/10.1214/aos/1176347265.

Lefèvre, M., Oumbe, A., Blanc, P., Espinar, B., Gschwind, B., Qu, Z., Wald, L., Schroedter-Homscheidt, M., Hoyer-Klick, C., Arola, A., Benedetti, A., Kaiser, J.W., Morcrette, J.J., 2013. McClear: A new model estimating downwelling solar radiation at ground level in clear-sky conditions. Atmos. Meas. Tech. (ISSN: 18678548) 6 (9), 2403–2418. http://dx.doi.org/10.5194/amt-6-2403-2013.

MacKay, David J.C., 1992. Bayesian interpolation. Neural Comput. 4 (3), 415–447.

Marcos, Javier, Marroyo, Luis, Lorenzo, Eduardo, Alvira, David, Izco, Eloisa, 2011. From irradiance to output power fluctuations: the PV plant as a low pass filter. Prog. Photovolt., Res. Appl. 19 (5), 505–510.

Mazorra Aguiar, L., Polo, Jesus, Vindel, J.M., Oliver, A., 2019. Analysis of satellite derived solar irradiance in islands with site adaptation techniques for improving the uncertainty. Renew. Energy (ISSN: 18790682) 135, 98–107. http://dx.doi.org/10.1016/j.renene.2018.11.099.

Perez, Richard, Kivalov, Sergey, Schlemmer, James, Hemker, Karl, Renné, David, Hoff, Thomas E., 2010. Validation of short and medium term operational solar radiation forecasts in the US. Sol. Energy (ISSN: 0038092X) 84 (12), 2161–2172. http://dx.doi.org/10.1016/j.solener.2010.08.014.

Perez, Richard, Lorenz, Elke, Pelland, Sophie, Beauharnois, Mark, Van Knowe, Glenn, Hemker, Karl, Heinemann, Detlev, Remund, Jan, Müller, Stefan C., Traunmüller, Wolfgang, Steinmauer, Gerald, Pozo, David, Ruiz-Arias, Jose a., Lara-Fanego, Vicente, Ramirez-Santigosa, Lourdes, Gaston-Romero, Martin, Pomares, Luis M., 2013. Comparison of numerical weather prediction solar irradiance forecasts in the US, Canada and Europe. Sol. Energy (ISSN: 0038092X) 94, 305–326. http://dx.doi.org/10.1016/j.solener.2013.05.005.

Polo, Jesus, Fernández-Peruchena, Carlos, Salamalikis, Vasileios, Mazorra-Aguiar, Luis, Turpin, Mathieu, Martín-Pomares, Luis, Kazantzidis, Andreas, Blanc, Philippe, Remund, Jan, 2020. Benchmarking on improvement and site-adaptation techniques for modeled solar radiation datasets. Sol. Energy (ISSN: 0038092X) 201 (October 2019), 469–479. http://dx.doi.org/10.1016/j.solener.2020.03.040.

Polo, Jesus, Wilbert, S., Ruiz-Arias, J.A., Meyer, R., Gueymard, C., Súri, M., Martín, L., Mieslinger, T., Blanc, P., Grant, I., Boland, J., Ineichen, P., Remund, J., Escobar, R., Troccoli, A., Sengupta, M., Nielsen, K.P., Renne, D., Geuder, N., Cebecauer, T., 2016. Preliminary survey on site-adaptation techniques for satellite-derived and reanalysis solar radiation datasets. Sol. Energy (ISSN: 0038092X) 132, 25–37. http://dx.doi.org/10.1016/j.solener.2016.03.001.

Qu, Z., Gschwind, B., Lefevre, M., Wald, L., 2014. Improving HelioClim-3 estimates of surface solar irradiance using the McClear clear-sky model and recent advances in atmosphere composition. Atmos. Meas. Tech. (ISSN: 18678548) 7 (11), 3927–3933. http://dx.doi.org/10.5194/amt-7-3927-2014.

Rigollier, Christelle, Lef, M., Wald, L., 2004. The method heliosat-2 for deriving shortwave solar radiation from satellite images, 77. pp. 159–169. http://dx.doi.org/10.1016/j.solener.2004.04.017.

Rigollier, Christelle, Lefèvre, Mireille, Blanc, Philippe, Wald, Lucien, 2002. The operational calibration of images taken in the visible channel of the meteosat series of satellites. J. Atmos. Ocean. Technol. 19 (9), 1285–1293.

Şahin, Mehmet, Kaya, Yılmaz, Uyar, Murat, Yıldırım, Selçuk, 2014. Application of extreme learning machine for estimating solar radiation from satellite data. Int. J. Energy Res. 38 (2), 205–212.

Saint-Drenan, Yves-Marie, Good, G.H., Braun, M., 2017. A probabilistic approach to the estimation of regional photovoltaic power production. Sol. Energy (ISSN: 0038092X) 147, 257–276. http://dx.doi.org/10.1016/j.solener.2017.03.007.

Schroedter-Homscheidt, Marion, Arola, Antti, Killius, Niels, Lefèvre, Mireille, Saboret, Laurent, Wandji, William, Wald, Lucien, Wey, Etienne, 2016. The copernicus atmosphere monitoring service (CAMS) radiation service in a nutshell. In: Proc. SolarPACES16. pp. 11–14.

Sengupta, Manajit, Habte, Aron, Gotseff, Peter, Weekley, Andrew, Lopez, Anthony, Anderberg, Mary, Molling, Christine, Heidinger, Andrew, 2014. Physics-based GOES product for use in NREL's national solar radiation database: Preprint. URL https://www.osti.gov/biblio/1351563.

Sperati, Simone, Alessandrini, Stefano, Delle Monache, Luca, 2017. Gridded probabilistic weather forecasts with an analog ensemble. Q. J. R. Meteorol. Soc. (ISSN: 00359009) http://dx.doi.org/10.1002/qj.3137.

Strobel, M.B., Betts, T.R., Friesen, G., Beyer, H.G., Gottschalg, R., 2009. Uncertainty in photovoltaic performance parameters – Dependence on location and material. Sol. Energy Mater. Sol. Cells (ISSN: 0927-0248) 93 (6), 1124–1128. http://dx.doi.org/10.1016/j.solmat.2009.02.003.

Tournadre, Benoit, Gschwind, Benoit, Saint-Drenan, Yves-Marie, Blanc, Philippe, 2021. An improved cloud index for estimating downwelling surface solar irradiance from various satellite imagers in the framework of a heliosat-v method. Atmos. Measur. Tech. Discussions 2021, 1–30. http://dx.doi.org/10.5194/amt-2020-480.

Vallance, Loïc, Charbonnier, Bruno, Paul, Nicolas, Dubost, Stéphanie, Blanc, Philippe, 2017. Towards a standardized procedure to assess solar forecast accuracy: A new ramp and time alignment metric. Sol. Energy (ISSN: 0038092X) 150, 408–422. http://dx.doi.org/10.1016/j.solener.2017.04.064.

Verbois, Hadrien, 2019. Solar Irradiance Forecasting in the Tropics Using Numerical Weather Prediction and Statistical Learning (Ph.D. thesis).

Verbois, Hadrien, Blanc, Philippe, Huva, Robert, Saint-Drenan, Yves Marie, Rusydi, Andrivo, Thiery, Alexandre, 2020. Beyond quadratic error: Case-study of a multiple criteria approach to the performance assessment of numerical forecasts of solar irradiance in the tropics. Renew. Sustain. Energy Rev. (ISSN: 18790690) 117 (October 2019), 109471. http://dx.doi.org/10.1016/j.rser.2019.109471.

Verbois, Hadrien, Saint-Drenan, Yves-Marie, Thiery, Alexandre, Blanc, Philippe, 2022. Statistical learning for NWP post-processing: A benchmark for solar irradiance forecasting. Sol. Energy (ISSN: 0038092X) 238 (February), 132–149. http://dx.doi.org/10.1016/j.solener.2022.03.017.

Vernay, Christophe, Blanc, Philippe, Pitaval, Sébastien, 2013. Characterizing measurements campaigns for an innovative calibration approach of the global horizontal irradiation estimated by HelioClim-3. Renew. Energy (ISSN: 09601481) 57, 339–347. http://dx.doi.org/10.1016/j.renene.2013.01.049.

Yezeguelian, Loïc, Vernay, Christophe, Carrière, Thomas, Blanc, Philippe, 2021. Characterizing the convergence and robustness of the kernel density mapping method for site-adaptation of global horizontal irradiation in western Europe. In: EU PVSEC 2021.