



HAL
open science

A Monte Carlo Based Solar Radiation Forecastability Estimation

Cyril Voyant, Philippe Lauret, Gilles Notton, Jean-Laurent Duchaud, Alexis Fouilloy, Mathieu David, Zaher Mundher Yaseen, Ted Soubdhan

► **To cite this version:**

Cyril Voyant, Philippe Lauret, Gilles Notton, Jean-Laurent Duchaud, Alexis Fouilloy, et al.. A Monte Carlo Based Solar Radiation Forecastability Estimation. *Journal of Renewable and Sustainable Energy*, 2021. hal-03162966

HAL Id: hal-03162966

<https://hal.science/hal-03162966v1>

Submitted on 8 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Monte Carlo Based Solar Radiation Forecastability Estimation

Cyril Voyant,^{1, a)} Philippe Lauret,² Gilles Notton,¹ Jean-Laurent Duchaud,¹ Alexis Fouilloy,¹ Mathieu David,² Zaher Mundher Yaseen,³ and Ted Soubdhan⁴

¹⁾University of Corsica - SPE laboratory, Ajaccio, Corsica, France

²⁾University of Reunion - PIMENT laboratory, Saint-Denis, Reunion, France

³⁾Sustainable Developments in Civil Engineering Research Group, Faculty of Civil Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam

⁴⁾University of Antilles, LARGE, Pointe à Pitre, Guadeloupe, France

(Dated: 8 March 2021)

Based on the reported literature and commonly used metrics in the realm of solar forecasting, a new methodology is developed for estimating a metric called forecastability (F). It reveals the extent to which solar radiation time series can be forecasted and provides the crucial context for judging the inherent difficulty associated to a particular forecast situation. Unlike the score given by the standard smart Persistence model, the F metric which is bounded between 0% and 100% is easier to interpret hence making comparisons between forecasting studies more consistent. This approach uses the Monte Carlo method and estimates F from the standard error metric $RMSE$ and the Persistence predictor. Based on the time series of solar radiation measured at 6 very different locations (with optimized clear sky model) from a meteorological point of view, it is shown that F varies between 25.5% and 68.2% and that it exists a link between forecastability and errors obtained by machine learning prediction methods. The proposed methodology is validated for 3 parameters that may affect the F estimation (time horizon, temporal granularity and solar radiation components) and for 50 time series relative to McClear web service and to the central archive of Baseline Surface Radiation Network.

1. INTRODUCTION

The intermittent nature of the solar resource and consequently the difficulty of its prediction constitutes a limiting factor for a greater integration of solar power generation in the energy field¹. In the solar forecasting community, many researchers use different metrics in order to assess the difficulty in generating good forecasts for different climates². One first possibility is to estimate the solar variability embodied in the solar time series. Variability is devoted to the quantification of the lack of consistency and gives a way to describe to which extent data sets vary³. This type of metric is used to compare the data at hand to other sets of data. Some authors, like Marquez and Coimbra⁴, Perez and Hoff⁵ or Blaga and Paulescu⁶ have endeavoured to describe mathematically this variability while most of the others researchers considered this characteristic of the solar irradiance time series as a basic assumption and have made efforts to implement predictive strategies of increasing complexity⁷. As often in physics, the description and especially the understanding of phenomena that are causing a problem (i.e., the difficulty of predicting solar irradiance/irradiation or the solar radiation components) leads to a better characterization of the situation^{8,9}.

Variability of global horizontal solar irradiation (GHI) is due to two terms. The first one originates from the predictable geometric trajectory of the sun while the second unpredictable component is due to effects induced by the atmosphere and the clouds. These unpredictable effects are captured by the clear sky index k_t^* defined as the ratio of the global horizontal irradiation GHI to GHI_c (GHI under clear sky conditions). In Perez and Hoff⁵ and Marquez and Coimbra⁴, k_t^* is the key

parameter used to calculate the variability. In the first paper, for a specific time scale Δt of the time series, variability is given by the standard deviation of the changes in the clear sky index denoted by $\sigma(\Delta k_{t, \Delta t}^*)$ while the second one proposes to evaluate variability by computing the magnitude of the ramp rates (i.e. changes in the clear sky index). Other authors like Fouilloy *et al.*¹⁰ and Voyant *et al.*¹¹ estimated the variability of the solar irradiation time series by quantifying the *mean absolute log return*. Although the results were interesting, the lack of theoretical consistency (passing through the $L1$ norm) and the absence of normalization militate for a proposition of a new approach.

As mentioned above, the variability metrics computed at different time scales can be used as proxies to estimate the difficulty in forecasting in some particular sites. However, these kind of metrics are independent of the forecasting time horizon and consequently are not suited for a detailed evaluation of the intrinsic difficulty related to a specific forecasting context.

Furthermore, and as stated by Pedro and Coimbra², it would be interesting to have an idea of the expected performance of the prediction models prior to their implementation and performance evaluation. To the best of our knowledge, in the realm of solar forecasting, Pedro and Coimbra² were the first to propose a combination of two metrics in order to assess the forecasting performance one may expect before any forecasts are generated for a particular site. The authors defined such as an a priori assessment as the forecastability. The two proposed metrics (computed for each forecast horizon) are respectively the density of large irradiance ramps (i.e. the density of changes in k_t^*) and a statistical metric called the time series determinism. Unfortunately, the combination of the two metrics to assess forecastability seems quite complicated and makes interpretation of the results difficult.

Before going further, it must be emphasized that the term predictability instead of forecastability is also used by some

^{a)}Also at hospital of Castelluccio - Radiotherapy Unit, Ajaccio, Corsica, France.; Electronic mail: voyant_c@univ-corse.fr

76 authors like Yang¹². In this work, and based on a biblio-
 77 graphic survey, we opt for the notion of forecastability. Based
 78 on a rather general bibliographic survey, the next subsection
 79 tries to shed some light on the difference between *predictabil-*
 80 *ity* (P)¹³ and *forecastability* (F)¹⁴ⁱ.

81 A. Difference between forecastability and predictability

82 One of the first references about the forecastability and the
 83 time series formalism is the result of the work of the co-
 84 recipient of the Nobel Prize Clive W.J. Granger¹⁵. Authors
 85 define it as the variance of the optimal forecast divided by the
 86 unconditional variance of the time series. This definition and
 87 the resulting Q parameter (forecastability quotient) were ex-
 88 tensively studied in economics, gradually giving way to new
 89 kind tools as sample and approximate entropy¹⁶, correlation
 90 and mutual information metrics¹⁷.

91 These two notions (P and F) are conceptually very close: if
 92 the predictability (P)¹⁸ studies how trajectories of the true sys-
 93 tem diverge¹⁹, the forecastability (F) describes how a model
 94 trajectory diverges from a true system trajectory¹⁴. A common
 95 explanation is that a predictable process is able to be predicted
 96 while a forecastable one is able of being forecasted. With this
 97 last definition, the concept of modelling appears, thereby a
 98 forecastable system is necessarily predictable but the opposite
 99 is not trueⁱⁱ. The predictability term which is often used with
 100 dynamic processes, is closely related to notions like causality
 101 ²¹ or chaos (i.e. failure of predictability²²), found for example,
 102 in all weather series and where the typical predictable times
 103 (or barriers) concerns the prediction horizons smaller than 1
 104 or 5 days²³. In the context of the present study (nowcasting or
 105 very short term), the chaotic aspect is not directly studied, so,
 106 the term forecastability seems more suitable than predictabil-
 107 ity one even if, to the best of our knowledge, there isn't any
 108 consensus on that.

109 Note that other concepts could have been detailed here (ob-
 110 servability and detection reliability²⁴) but they do not have
 111 much sense in the study of solar radiation given the quantities
 112 involved are directly measurable and the associated uncertain-
 113 ties are very low.

114 B. Framework of the study

115 As has been done for wind speed forecasting²⁵, one may
 116 also use the performanceⁱⁱⁱ²⁶. of the Persistence (P) model
 117 which is, in this work, related to a persistent clear sky index

ⁱ In this paper, authors assume that "forecastability is not the same thing as predictability"

ⁱⁱ Kumar and Chen²⁰ highlight this difference stating "inherent predictability being the upper bound of forecastability"

ⁱⁱⁱ The performance of deterministic forecasting models are also usually reported by the root mean square error (RMSE) and its normalized counterpart ($nRMSE$) obtained by dividing the $RMSE$ by the mean of the irradiation values

118 k_t^* . More precisely, the $nRMSE$ obtained by this reference
 119 model can be used to gauge the forecastability of a particu-
 120 lar site. However, although this method is feasible, it would
 121 not provide the same type of information. Indeed, a Persis-
 122 tence forecast with an $nRMSE$ equal to 20 % is not sufficient
 123 to consider if the site is related to a high or a low forecastabil-
 124 ity. It depends on the kind of solar component (direct, diffuse,
 125 global), on the forecast horizon, on the latitude, on the to-
 126 pographic relief, on the nebulosity, etc. For instance, for GHI
 127 hourly data and 4-hour forecast horizon, a Persistence forecast
 128 with an $nRMSE$ equal to 20% could be an indication of a high
 129 forecastability, but for GHI data with a 5-min granularity and
 130 for a 5-min forecast horizon, certainly not. Put differently, the
 131 main drawback of using the score of the reference Persistence
 132 model is that it is not bounded by an upper limit.

133 In this work, the objective is to propose a forecastability
 134 metric (denoted hereafter F) which is easy to compute and
 135 easy to interpret. In other words, this new metric should pro-
 136 vide users a global reference in order to assess the inherent
 137 difficulty to issue forecasts for a specific site and consequently
 138 if it makes sense to add an extra effort to build more and more
 139 complex forecasting models. Additionally, it would be de-
 140 sirable to relate in a simple way this (ex-ante) forecastability
 141 metric with the (ex-post) skill score metric defined by Mar-
 142 quez and Coimbra⁴. Hence, the F metric should be helpful
 143 in comparing the numerous forecasting methods proposed by
 144 the community.

145 To compute the F metric, the user will need to compute the
 146 $RMSE$ of the Persistence model^{iv}. Moreover, as it will be dis-
 147 cussed in section III C, this approach will allow adapting the
 148 forecastability with the forecast horizon and with the different
 149 components of the solar radiation.

150 The rest of the manuscript is structured in four sections.
 151 The first one (section II) presents the mathematical formula-
 152 tion of the forecastability F . In section III, the sensitivity of
 153 F to various parameters is evaluated while in section IV is
 154 proposed an estimate of F concerning 50 time series of GHI .
 155 Finally, Section V gives some conclusions and some perspec-
 156 tives.

157 II. ESTIMATION PROPOSAL FOR F

158 Before proposing a new method to estimate the forecasta-
 159 bility F , it is important to understand why in the existing for-
 160 malism of the solar radiation time series, some elements could
 161 be improved. As discussed in the introduction, in the time se-
 162 ries domain, the variability (V) is often estimated by comput-
 163 ing the variance⁶. In the context of solar radiation, the time
 164 series data are very particular in the sense that the underlying
 165 trend is periodic and easily quantifiable using robust models
 166 validated by the community.

167 These models are denoted clear sky models²⁷ and they cor-
 168 respond to the solar irradiation estimated under clear sky con-
 169 ditions (GHI_c) and computed from the sun position and vari-

^{iv} Which does not require any additional effort and calculation since the $RMSE$ of the reference Persistence model is systematically computed in the solar forecasting studies (e.g., for skill score calculation).

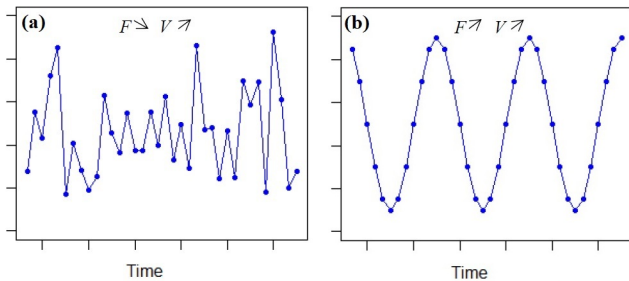


FIG. 1. Intuitive relationship between forecastability (F) and variability (V)

ous meteorological parameters characterizing the atmosphere
state. Among others, one can cite the *Kasten, Bird, ESRA,*
Solis, McClear, SMARTS, etc. models. The clear sky models
are extensively used in the literature^{28,29} to build increasingly
powerful predictive models and thus it is essential to use this
specificity in our turn to quantify the forecastability.

176 A. The use of the specifics of solar radiation time series

Figure 1 permits to intuitively introduce the notion of fore-
castability. Figure 1(a) presents a Gaussian noise signal. In
this non-periodic case, one can assume that the forecastability
and the variability are strongly linked ($F \simeq 1 - V$). In other
words, when the variance of the signal increases, the fore-
castability decreases. Conversely, for a periodic time series
data, the conclusion is different as shown in Figure 1(b) with
a *cosine* function. Even if the variability is important, one can
not conclude that forecastability is low ($F \not\simeq 1 - V$) if there is
a way to estimate the trend and seasonality present in the time
series. In the present case, the prediction becomes easy using
a simple trigonometric function.

In the case of the global solar irradiation, if one attempts to
estimate the forecastability of the GHI time series, it must
certainly rely on a clear sky modeling of the global hori-
zontal irradiation. With this assumption, a methodology to
estimate F based on a detrending of the GHI time series
or rather on a seasonal adjustment like the clear sky index
($k_t^*(t) = GHI(t)/GHI_c(t)$) is proposed hereafter.

196 B. Interest of the normalization in computing F

In statistics, normalization is a very frequently used tech-
nique which refers to the creation of shifted and scaled ver-
sions of certain variables (see normalization of scores pro-
posed by Dobson³⁰). The primary intent of this change is
that these normalized values promote the comparison of a cer-
tain effect for different data sets in a way that eliminates the
effects of certain gross influences.

In order to exemplify the interest to make use of the nor-
malization in the calculation of F , Figure 2 shows one day
of modelled GHI_c for two sites with different solar potential.

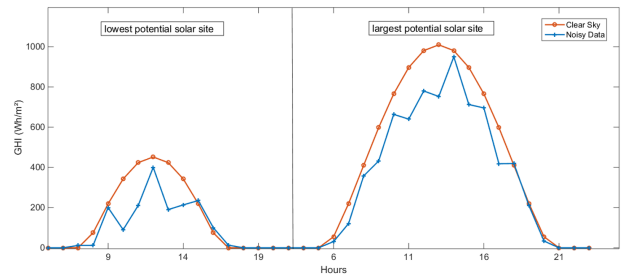


FIG. 2. Clear sky modeling vs noisy data for two sites with different solar potential

Figure 2 plots also a synthetic time series (denoted here noisy
data) built from the sum of an uniform white noise (same level
of noise for the left and right parts of the figure) plus 80% of
the GHI_c values. Although this example covers only one day,
it highlights phenomena that occur annually for two sites with
high and low solar energy potential or for one site when it
is studied during two different periods of the year (summer
and winter for instance). When a prediction is performed and
when an usual relative error metric like the normalized root
mean squared error ($nRMSE$) is computed, the impact of the
noise is more detrimental in the left case. The $nRMSE$ re-
lated to the left part of the Figure 2 is 65.6% (49.2% with
nighttime filtering) while for the right part it is 45.9% (39.8%
with nighttime filtering). With an estimate of F that does not
take into account the solar potential of the site, these two GHI
curves could be characterized by an identical F given that the
two signals are constructed in the same way. This problem-
atic prompts us to propose a *normalized* formulation of F that
takes into account the studied site and its characteristics (in-
cluding the data filtering process) as we will see in the follow-
ing subsection.

Again, it may seem tempting to quantify the forecastabil-
ity relying solely on the interpretation of the $nRMSE$ of the
Persistence forecast. But, as already stated in the introduc-
tion, this way of proceeding does not make it possible to rule
on the forecastability regarding particular forecast horizons or
time scales present in the data. Finally, we explain in Annex
A why the normalization can't be operated by a simple use of
the mean of GHI_c .

236 C. Mathematical formalism for calculating F

In what follows, we detail the mathematical formalism us-
ing the GHI . This formalism can be derived for the other solar
radiation components but will not be given here.

The most used model as naive or reference predictor in
the solar forecasting community is the so-called Persistence
model^v defined in this work by the Persistence of the clear

^v Naive forecasts serve as a placebo in evaluating the performance of fore-
casting processes

sky index k_t^* . It has the advantage to be quickly implemented without resorting to a learning phase nor having access to a large set of historical data³¹. Note that recent studies have highlighted the use of “improved” Persistence model such as an optimal convex optimization of Persistence on k_t^* and climatology¹². As currently this type of model is not yet a common used model in the solar forecasting studies, it will be not tested in this study. In addition, using this kind of “smart” Persistence instead of the classical persistence would marginally impact the evaluation of the inherent difficulty of a forecast situation provided by the F metric.

The Persistence model is also considered as the reference for calculating the forecast skill score $SS = 1 - RMSE/RMSE_P$ ³² where $RMSE$ ²⁶ is the root mean square error related to the studied model and $RMSE_P$ is the score of the Persistence model. The error metric $RMSE$ computed for the time index set T reads as (circumflex for the prediction):

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n \left(GHI(t) - \widehat{GHI}(t) \right)^2}, \forall n \in T. \quad (1)$$

The Persistence forecast estimates at the time $t + 1$ (i.e. $\widehat{GHI}(t + 1)$) uses the clear sky modelling value (GHI_c) and is given by Equation 2.

$$\widehat{GHI}(t + 1) = GHI(t) \times \frac{GHI_c(t + 1)}{GHI_c(t)} \quad (2)$$

Given the $RMSE$ score obtained by the Persistence predictor ($RMSE_P$), the formulation of F is based on a normalization involving an estimation of the maximum and minimum values that the $RMSE$ can reach (respectively $RMSE_{max}$ and $RMSE_{min}$) for a particular site. Hence, Equation 3 defines F .

$$F = 100\% \times \frac{RMSE_{max} - RMSE_P}{RMSE_{max} - RMSE_{min}} \quad (3)$$

With this metric, when $RMSE_P = RMSE_{min}$ (the lowest acceptable value) $F = 100\%$ and when $RMSE_P = RMSE_{max}$ (the highest allowable value) $F = 0\%$. The F score is therefore between 0% and 100%. $RMSE_{min}$ is reached when the forecastability is maximum i.e. cloudless sky and clear sky conditions with $GHI(t) = GHI_c(t)$. In this case, k_t^* is always equal to 1 and the Persistence model is an ideal predictor inducing $RMSE_{min} \simeq 0$ (not really observable in practice but it is a necessary theoretical condition for the presented methodology)^{vi}. According to this simplification, the updated version of F is as follows:

$$F \approx 100\% \times \left(1 - \frac{RMSE_P}{RMSE_{max}} \right) \quad (4)$$

^{vi} inducing 2 assumptions: there are no clouds and the clear sky is perfectly defined. Unfortunately, neither is physically realistic, if the first hypothesis is rarely established (except in desert regions and not yet, during the whole year), the second one is dependent on modeling, and as it will be visible in Section III B, the errors of the clear sky can vary between 1% and 5%

$RMSE_{max}$ (purely theoretical parameter defined over the desired prediction period: months, years, etc.) is reached when the prediction is not possible thus when the predictability becomes too low for a model to be relevant. For example, on Ajaccio within the framework of an hourly study of the GHI , whatever the method used, its $RMSE$ will be between $RMSE_{min}$ ($= 0Wh/m^2$) and $RMSE_{max}$ ($= 249.7Wh/m^2$). This last limit case occurs when there is no link (in the sense of statistical dependence) between the future elements of the time series and the present ones. We artificially generated a k_t^* time series with elements governed by a random process and an uniform probability distribution ($\varepsilon(t)$, between 0 and 1 - See Equation 5 below). This is equivalent to uniformly distribute $GHI(t)$ between 0 and $GHI_c(t)$.

This approach obviously excludes the phenomenon of overirradiance³³ (i.e. $k_t^* > 1$). However, when the granularity of the GHI time series is greater than a few minutes, we assume that these phenomena are rare enough to consider values of k_t^* greater than 1 as insignificant events. Moreover, for example in Ajaccio, the values of measured k_t^* maximum (time granularity of 1h and horizon 1h) fluctuate considering the year used as a basis for the calculation (between 1.76 in 2000 and 2.08 in 1999). So considering the available years, the value of F (which depends indirectly on the value of the k_t^* maximum if the latter is not taken equal to 1) would fluctuate a lot. Consequently, and in order to keep the methodology as simple as possible, we limit the generation of random k_t^* between 0 and 1.

Figure 3 shows the methodology used to estimate F for a specific site using pseudo-random numbers generator (Matlab®)³⁴. The methodology, although different from what is usually the goal of a Monte Carlo method³⁵, is based on a computational algorithm that relies on repeated random sampling to obtain numerical results³⁶.

From the definitions of the $RMSE$ (Equation 1) and the Persistence model (Equation 2), it is possible to calculate the $RMSE_{max}$ score in relation with the $GHI_c(t)$ value and pseudo-random numbers series denoted by $\varepsilon(t)$. Posing $k_t^*(t) = \varepsilon(t)$ and $k_t^*(t - 1) = \varepsilon(t - 1)$, it is possible to integrate these values in Equation 1 considering that $GHI(t) = k_t^*(t) \cdot GHI_c(t)$. Therefore, $RMSE_{max}$ reads as:

$$RMSE_{max} = \sqrt{\frac{1}{n} \sum_{t=1}^n \left(GHI_c(t) \times \left(\varepsilon(t) - \varepsilon(t - 1) \right) \right)^2} \quad (5)$$

Obviously, the approach set out here is only possible if the number n of random samples in Equation 5 is large enough. In practice, more than 1000 samples^{vii} are necessary to obtain a fairly estimate of F . As a consequence, more than 1000 pairs of forecast/measurement are necessary to compute the score of the Persistence model in Equation 4.

It is easy to notice in Equation 5 that $RMSE_{max}$ depends on a variable with a physical meaning namely $GHI_c(t)$. As a

^{vii} but the higher the number n , the better the estimate of F .

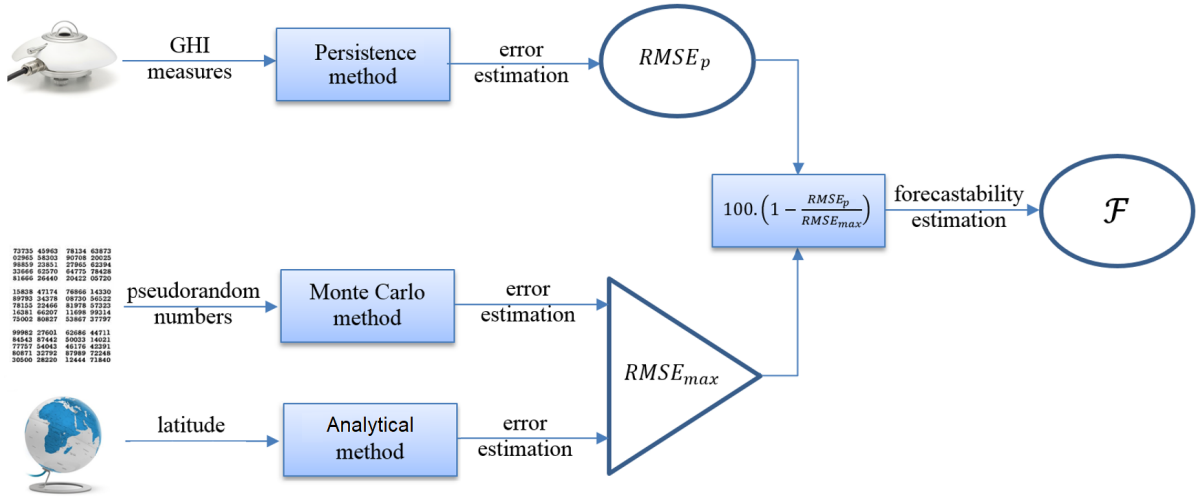


FIG. 3. Methodology to estimate the forecastability (F)

consequence, two sites with the same clear sky characteristics will exhibit the same $RMSE_{max}$ score. Two possibilities exist (Figure 3), either to use random numbers (the most efficient method, see Algorithm 1 for $RMSE_{max}^{mc}$), or to make hypotheses (which we will detail in the following $RMSE_{max}^{analytic}$) and use an analytical version of the parameter F . For the second way, as a first approximation, it is not necessary to take into account all the parameters needed to compute the clear sky model.

Algorithm 1: Calculate $RMSE_{max}^{mc}$

Require: clear sky time series GHI_c
for all $t \in T$ **do**
 generate $\varepsilon(t)^a \in (0, 1)$
 $GHI(t) \leftarrow GHI_c(t) \times \varepsilon(t)$
return $\widehat{GHI}(t) \leftarrow GHI(t - 1)$
end for
return $RMSE_{max}^{mc} = RMSE(GHI - \widehat{GHI})$

^a See Eq 5

With the Solis model¹¹, the main parameter influencing the $RMSE_{max}^{analytic}$ value is the latitude which is linked to the solar position in the sky. The $RMSE_{max}^{analytic}$ (which is a purely theoretical and statistical parameter without real physical sense) is, thereby, assumed equivalent for Sahara desert, Caribbean, coastal Mexico, south east Asia, ... i.e. countries or regions along the same latitude band. This is of course a strong assumption but which is necessary if one wants to use the clear sky modeling (operated by the most of the researchers in the solar forecasting community) in simulations and not the extraterrestrial irradiation or the solar elevation. In this case, other setting parameters induce only slight uncertainties in the observed results (only verified for the tested sites in this study but certainly quite simply transposable to most sites that do not have extreme characteristics).

Even if all existing clear sky models were not checked, this conclusion seems to be generalizable. Anyway, this assumption can be easily verified for all clear sky models used by the researchers. In the current case, an uncertainty of $\pm 3.7\%$ was found when the Aerosol Optical Depth (AOD monthly updated in the study) varies between 0.1 and 0.2 (these latter values are very small compared to exceptional observations in desert climates) or $\pm 0.5\%$ when the altitude fluctuates between 0 and 1000m. Note that in Ajaccio, the arrivals of ferries boats add a systematic pollution and the measurement of “ AOD ” exhibits strong intraday fluctuations which are not taken into account in the simulations. The estimated $RMSE_{max}^{analytic}$ versus the latitude computed using the Solis model is plotted in Figure 4. For each latitude, monthly and yearly values are given. They allow to know the magnitudes of the approximate $RMSE_{max}^{analytic}$ according to the latitude and the periods during which solar radiation measurements are available (month, season or overall year). A filtering process was applied, all data with solar elevation $h < 1^\circ$ (nocturnal values) are deleted because during these periods, the k_t^* values are not defined. With another limit of filtering, the curves are no longer usable; hence, it is considered that the validity interval of the solar elevation filtering for the proposed curves of Figure 4, is between 0° and 5° (corresponding to a $RMSE_{max}^{analytic}$ difference of $\pm 1\%$).

In Figure 4, it must be noted that $RMSE_{max}^{analytic}$ values used to estimate the forecastability of the considered site has been computed with an uncertainty close to $\pm 5\%$. In the annual case (red line), a Gaussian fit (see Equation 6) with a slight difference related to latitudes close to 0° can be used (goodness-of-fit of analytical method close to $R^2 = 0.999$):

$$RMSE_{max}^{analytic} = 325.9e^{-\left(\frac{lat+1.088}{79.86}\right)^2}$$

$$\Rightarrow F = 100\% \times \left(1 - \frac{RMSE_p}{325.9e^{-\left(\frac{lat+1.088}{79.86}\right)^2}}\right) \quad (6)$$

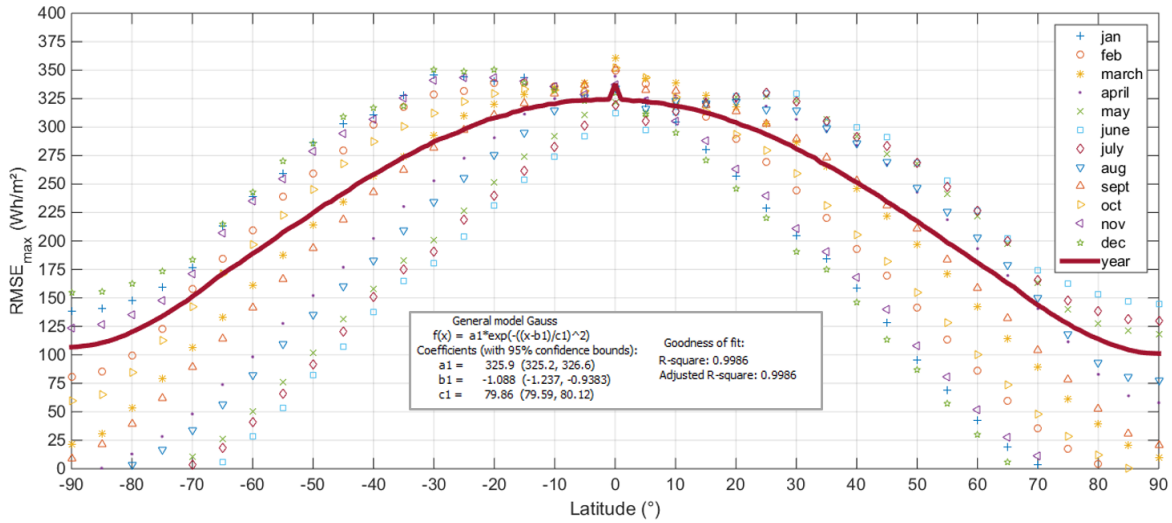


FIG. 4. The magnitudes of the $RMSE_{max}^{analytic}$ versus the latitude. Monthly and yearly values are given.

385 D. Link between F and forecast skill (skill score SS)

386 Currently, most of the forecasting studies evaluate the per-
 387 formance of the models in terms of forecast skill. As men-
 388 tioned in the introduction, it would be desirable to simply re-
 389 late the forecastability metric to a skill score. As the fore-
 390 castability is computed from the Persistence forecasts, it is
 391 straightforward to establish the link between the skill score
 392 ($SS = 1 - \frac{RMSE}{RMSE_P}$) and F . Considering Equations 4 and Algo-
 393 rithm 1, Equation 7 related to the link between F and SS
 394 can be established as:

$$SS = 1 - \frac{RMSE}{(1 - \frac{F}{100}) \times RMSE_{max}^{carlo}} \quad (7)$$

395 For example, in Ajaccio (Corsica island) with Latitude =
 396 41.9° , $F = 68.2\%$ and $RMSE_{max}^{carlo} = 249.7 Wh/m^2$, Equation
 397 7 becomes a simple relationship between the skill score SS
 398 and the $RMSE$ of the studied forecasting method i.e. $SS \simeq$
 399 $1 - RMSE/79$.

400 III. FACTORS INFLUENCING F

401 Some parameters influence the factor F such as the site lo-
 402 cation, the solar radiation component (global, beam and dif-
 403 fuse), the prediction horizon and the time granularity (this list
 404 is not intended to be exhaustive).

405 A. F variation with location

406 The F value quantifies the difficulty to predict the solar ra-
 407 diation components. Table I lists the F values for 6 locations
 408 (see also Figure 5): Ajaccio (Corsica island, France), Nancy
 409 (East of France), Odeillo (mountainous site, Pyrenées Orien-
 410 tales, France), Tilos (Greek island), Saint Pierre (Reunion is-



FIG. 5. The 6 locations used to estimate F

411 land, France), Le Raizet (Guadeloupe archipelago, France).
 412 Those stations are not enough to prove the generalization of
 413 the forecastability. In section IV, it will be tested on 50 sta-
 414 tions spread all over the globe. Here the solar radiation com-
 415 ponent of interest is the GHI and the forecast horizon and the
 416 time granularity is 1 hour for all sites.

417 As shown by Table I, the higher the forecastability (F), the
 418 lower the error metric ($nRMSE$) related to predictions gener-
 419 ated by a multilayer perceptron ($MLP^{10,38,39}$). The relation-
 420 ship between these two metrics is not obvious to estimate.
 421 However, it must be stressed that factors such as the quality of
 422 the measurements as well as the concordance of the measure-
 423 ments with the clear sky model may complicate the interpreta-
 424 tion. This table validates what has been widely demonstrated
 425 in the literature i.e. mountainous regions (Odeillo) and conti-
 426 nental climates (Nancy) are more difficult to apprehend than
 427 coastal areas (Ajaccio and Tilos). With this table, it is inter-
 428 esting to realize that the SS alone does not allow to judge the
 429

TABLE I. Values of F and $nRMSE$ for 6 locations

Site	Köppen class ³⁷	Latitude	F ^a	$nRMSE$ ^b	SS ^b
Ajaccio	Csa	41°55'N	68.2%	18.3%	0.04
Tilos	Csa	36°25'N	64.7%	19.6%	0.02
St Pierre	Aw	21°20'S	62.4%	21.1%	0.01
Le Raizet	Af	16°26'N	58.2%	25.9%	0.07
Nancy	Cfb	48°41'N	50.2%	27.4%	0.05
Odeillo	Cfb	42°30'N	25.5%	29.9%	0.19

^a See Equation 4^b Related to the *MLP* predictionsTABLE II. Value of F for the 3 radiation components and $nRMSE$ ⁴¹ in Odeillo site

Radiation Components	F ^a	$nRMSE$ ^b	SS ^b
Global	25.5%	29.9%	0.19
Beam	13.4%	38.2%	0.02
Diffuse	12.1%	40.9%	0.35

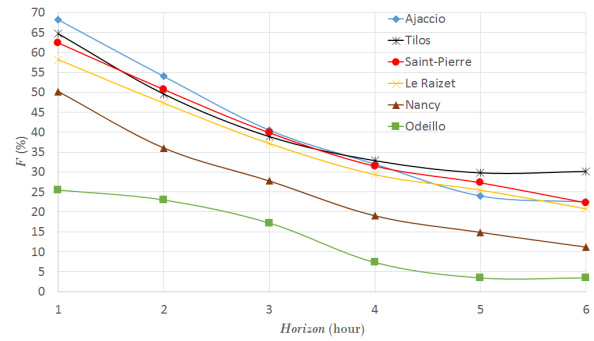
^a See Equation 4^b Related to the *MLP* predictions

430 forecastability of a site. Indeed, it only indicates the degree of
 431 improvement that the predictive methodology generates (*MLP*
 432 in this case) with respect to persistence. By reasoning simply,
 433 two totally different phenomena can be characterized by the
 434 same SS . It is that one observes with a constant (thus persist-
 435 ent) phenomenon, $F = 100\%$ logically but $SS = 0$ also be-
 436 cause one cannot do better than a forecast by persistence. In
 437 the same way, a white noise with $F = 0\%$ is also characterized
 438 by $SS = 0$ because here again it will not be possible to do bet-
 439 ter than a simple persistence. Therefore, SS cannot be a good
 440 indicator of forecastability.

441 B. F evolution with the solar radiation components

442 The methodology used to compute F can be applied to the
 443 3 components of solar radiation (global, beam and diffuse)⁴⁰.
 444 The only prerequisite is to estimate the clear sky solar value
 445 corresponding to each component. Table II gives the F values
 446 and the prediction errors for Odeillo (the site with the most
 447 important variability, for a forecast horizon and a time granu-
 448 larity of 1 hour^{10,39}).

449 As shown by Table II, the other solar radiation components
 450 (beam and diffuse) exhibit lower F values than the one esti-
 451 mated for the *GHI*. Conversely to global irradiation, the clear
 452 sky modelling for the beam and diffuse components is less
 453 efficient. As an illustration, the values obtained with the sim-
 454 plified Solis version are within 1% for the global component,
 455 2% for the beam component and 5% for the diffuse component
 456 (read Ineichen⁴² for details). In this subsection, the conclu-
 457 sion is similar than the one stated in subsection III A that is
 458 $nRMSE$ trend is a bijective function of F but the relationship

FIG. 6. Forecastability and prediction horizons (F computed from $RMSE_P$ found in^{10,38,41})

459 between F and $nRMSE$ is non-linear. However, the difference
 460 in forecastability between Beam and Diffuse components is
 461 close to 1% while the $nRMSE$ fluctuates by more than 2 per-
 462 centage points. Consequently, it must be stressed that the pro-
 463 posed methodology for estimating F will not be relevant if
 464 the clear sky model related to the solar radiation component
 465 of interest lacks precision.

466 C. F evolution with the forecast time horizon

467 Intuitively, it makes sense to think that the longer the fore-
 468 cast horizon, the lower the forecastability. This fact is verified
 469 for all the locations depicted in Figure 6 and particularly for
 470 Ajaccio with *GHI* time series of 1 hour time granularity. The
 471 F factor is halved in value when the lead time goes from 1h to
 472 6h for all the locations. The related prediction errors $nRMSE$
 473 generated by the *MLP*³⁹ predictor in Ajaccio are respectively
 474 18.3%, 29.5%, 31.2%, 33.0%, 33.8% and 34.5%. It can be
 475 noted that between the 5 and 6 hours horizons, the estimates of
 476 F are not significantly different for most of the studied cities.

478 D. F variation with time granularity

479 In Table III, it can be seen that for Tilos (the only site where
 480 10 minutes data were available), when we realized global irra-
 481 diation forecasts (*MLP*³⁹) for horizons of respectively, 1 hour,
 482 15 minutes and 10 minutes¹⁰, the conclusion follows the logic
 483 observed so far that is F and $nRMSE$ are strongly statistically
 484 dependent and when one increases the other decreases.

486 E. Conclusion about the F estimation

487 This section showed the link between the F factor com-
 488 puted from Equation 4 and some parameters frequently used
 489 in studies related to the prediction of global irradiation or so-
 490 lar radiation components through the time series formalism.
 491 In Figure 7, this link can be estimated by comparison be-
 492 tween the forecastability value and the prediction error related
 493 to *MLP* forecasts. For each kind of parameters (respectively

TABLE III. Value of F according to the time granularity and $nRMSE$ related to MLP prediction¹⁰ in Tilos island

Time Granularity	F ^a	$nRMSE$ ^b	SS ^b
1 hour	64.7%	19.6%	0.02
15 minutes	82.5%	15.6%	0.06
10 minutes	87.4%	12.8%	0.04

^a See Equation 4

^b Related to the MLP predictions

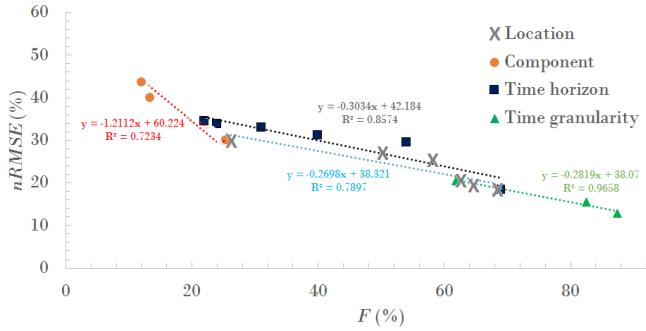


FIG. 7. Relationship between F and prediction error $nRMSE$. Linear fits are provided in order to better highlight the monotonic relationship between F and $nRMSE$.

494 the location, radiation component, the time granularity and the
 495 lead time) the link between $nRMSE$ and F is monotonic and
 496 when F increases the error metrics decreases. The relation-
 497 ship is not linear and certainly depends on the predictor used
 498 (MLP in our case). What is verified is that when the forecasta-
 499 bility is good (in the sense of a high value of F) the forecast
 500 becomes easier.

501 IV. $RMSE_{max}^{mcarlo}$ AND F CALCULATED FROM 50 TIME 502 SERIES OF GHI

503 In order to not limit the conclusions of this study to the only
 504 time series studied in the previous simulations, we propose
 505 here to estimate $RMSE_{max}^{analytic}$, $RMSE_{max}^{mcarlo}$ and F from data
 506 directly provided by large consortia working in the field of
 507 acquisition, processing and modeling of solar radiation (year
 508 2015, 1h time granularity and horizon).

509 A. Comparison between $RMSE_{max}^{analytic}$ and $RMSE_{max}^{mcarlo}$ 510 using McClear time series

511 In this section, we propose to estimate $RMSE_{max}$ from Mc-
 512 Clear web service series. The calculation methodologies pre-
 513 viously developed and based from random number generation
 514 (Cf $RMSE_{max}^{mcarlo}$ in section IIC) and based solely on the lati-
 515 tude ($RMSE_{max}^{analytic}$ computed from Eq 6) are compared. The
 516 characteristics of the time series are available in Appendix B

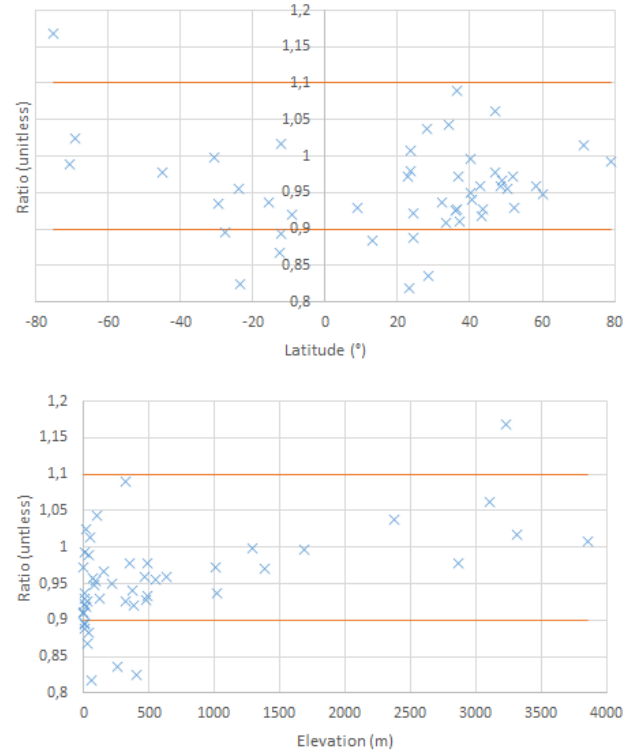


FIG. 8. $Ratio = RMSE_{max}^{mcarlo} / RMSE_{max}^{analytic}$ considering elevation and latitude. 82% of ratio value are comprised between 0.9 and 1.1.

517 (Table IV). The way of calculating the forecastability from
 518 random numbers (Figure 3) is synthesized by Algorithm 1,
 519 which can be applied from anywhere on the surface of the
 520 globe, as long as a clear sky radiation estimate (GHI_c) is
 521 known but above all optimized to be as reliable as possible.

522 Figure 8 makes it possible to quantify the difference be-
 523 tween a direct calculation ($RMSE_{max}^{mcarlo}$ described by Algo-
 524 rithm 1) and the use of Equation 6 ($RMSE_{max}^{analytic}$). We can
 525 appreciate a good match between these two ways of operat-
 526 ing. However, stations abbreviated as (see Table IV) COC,
 527 DOM, FLO, GAN, GOB, GUR, HOW, ISH, TIR are the sta-
 528 tions with more than 10% of difference between the two meth-
 529 ods (the mean of ratio is 0.95). There are several ways for
 530 understanding these discrepancies, such as considering only
 531 the latitude in Equation 6 or the fact that no post-processing
 532 was performed with the McClear model. This model is un-
 533 doubtedly one of the best performing model at the present, but
 534 there are some uncertainties relating to certain locations that
 535 have been reported in the literature. For example, Laguarda
 536 *et al.*⁴³ reported errors related to the use of McClear model
 537 close to 5% in average and which can reach more than 10%
 538 some particular periods.

539 As a synthesis, we cannot therefore make an objective de-
 540 cision as to the quality of the forecastability calculation based
 541 solely on the latitude of the site. However, we can think that
 542 this way of proceeding is not in total contradiction with the
 543 direct calculation (Algorithm 1), which is relatively simple to
 544 implement and it is certainly preferable to undertake it if one

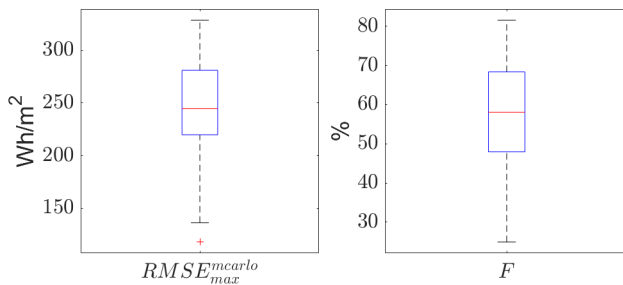


FIG. 9. Boxplots concerning $RMSE_{max}^{mcarlo}$ and F computed from McClear and BSRN data

wants precise results.

B. F calculation concerning 50 BSRN stations

In this section we propose F estimates (Equation 4) from $RMSE_{max}^{mcarlo}$ computed with McClear model and Algorithm 1. The measurements (BSRN data) are kindly provided by the renowned World Radiation Monitoring Center (Annex B and Table IV). The distributions of the F and $RMSE_{max}^{mcarlo}$ values can be seen in Figure 9 (average respectively close to 60% and 250 Wh/m²). The interpretation of the results should be taken lightly, because, as we have seen before, the determination of the $RMSE_{max}^{mcarlo}$ depends strongly on the clear sky model used. During these simulations, there was no post-processing to improve the reliability of the model contrary to what was done during the simulations relating to the sections III A to III D. Moreover, sometimes long acquisition periods were unusable making the calculation of forecastability certainly irrelevant for certain series. As could be expected, it is nevertheless possible to notice a large dispersion of the values of F (from 25% to 82%). The site with the lowest variability is in Japan (SAP) and the site with the highest, in Antarctica (DOM). Note that latitude is not a sufficient factor to judge forecastability. Indeed, two sites with the same latitude like IZA and GUR may have the same F value while others like DRA and E13 may have totally different F values. With this study we must valid that forecastability does not allow to pose whether a meteorological phenomenon will be easily or totally predictable (the role of predictability; see introduction part IA). The forecastability can be seen as the efforts to be implemented to carry out a forecast not of a meteorological phenomenon (such as cloudiness because it is indirectly what is done when modelling the GHI) but of the measurement of the latter under certain constraints (clear sky quality, horizon, time step, component, inclination, etc.).

V. CONCLUSIONS AND PERSPECTIVES

In the solar energy forecasting community, it is common to read more and more papers proposing increasingly complex forecasting methodologies and whose conclusions are limited to state that the proposed new method outperforms the previ-

ous ones without any consideration of forecastability or predictability. However, the inherent difficulty related to a particular forecasting situation should be studied prior the implementation and testing of a forecasting model. Put differently, statements regarding the quality of the generated forecasts must account for the forecastability of the variable of interest.

In this paper, a simple new methodology based on the $RMSE$ metric and the Persistence model was presented to estimate the forecastability F of the solar radiation components. This F metric is defined like a percentage between 0 and 100% and is very easy to interpret. The formalism used is reminiscent of what has been proposed in the literature over the past 10 years but with some small modifications and a normalization process based on a Monte Carlo approach ($RMSE_{max}$). Two ways of doing so were proposed: calculate $RMSE_{max}^{mcarlo}$ or make an analytical approximation of the latter by calculating $RMSE_{max}^{analytic}$. Even if the latter method gives good results, the first one, while being very easy to implement, is preferable and we recommend it. The results of the simulations validate the proposed theoretical framework and it appears that it is quite simple to quantify the forecastability (see Equation 6 and Algorithm 1) regardless the studied site. The real difficulty in using this methodology is that that the clear sky model must be reliable and carefully tuned. Otherwise, the methodology presented here becomes inappropriate. It is this prerequisite which led to less convincing results when we estimated F for the diffuse and solar radiation components. This point is important, because besides being used for estimating F , the clear sky modelling is becoming increasingly important for the derivation of the forecasting methods found in the literature¹⁰. The prediction results (and especially those related to the skill scores) can strongly diverge in case of incorrect tuning of the clear sky model. This phenomenon is not so highlighted in the other forecasting disciplines because, contrary to the solar radiation time series, it is not possible to estimate the underlying trend with the help of a physical-based parametric model like a clear sky model.

The methodology proposed in this work is based on the clear sky index. Nonetheless, in order to overcome the uncertainties related to the modeling of the different interactions in the atmosphere (which are the basis of the clear sky models elaboration)⁶, a future work will be devoted to the derivation of the F metric based on the clearness index (ratio of the irradiation or irradiance to the extraterrestrial irradiation).

However, it should be noted that the uncertainties related to time-stamping and poor measurement-model synchronization can still affect the F estimation. Consequently, even with these other possible methodologies, it would not be possible to propose a "true" objective F estimation. This is a critical issue that deserves careful attention.

In addition to this essential point, the next objective of this study will be to propose a reliability index based on the variation of the intra-annual forecastability. Indeed, it is possible to compute F from a sliding window (100 hours taken here as example) and thus to simply estimate the reliability of a prediction from the computed forecastability. As shown by Figure 10, in summer (center of the hours axis), predictions are

reliable with $F \simeq 85\%$ which is not the case at the extremities (winter; $F \simeq 60\%$). Following the methodology developed by Fliess, Join, and Voyant⁴⁴, the next step in this research will be to enrich deterministic forecasts with predictions intervals that will take into account both the forecastability and the volatility of the solar radiation time series.

Finally, the idea of characterizing F over a large area (or a whole country) will quickly become essential. While the estimation of F for meteorologically homogeneous or regular areas should be straightforward, it will be not possibly the case for inhomogenous regions such as for instance Corsica island or the Pyrénées mountains. In these areas, forecastability can vary up to 20% under a distance of 2 or 3 km. To characterize such inhomogenous areas, many time series would be necessary and it may be necessary to use satellite-derived irradiance such as HelioClim-3 solar radiation database in real time.

Appendix A: Why not normalize F with the mean of GHI_c as we do with error metrics ($RMSE$ vs $nRMSE$)?

At a first glance, it would seem attractive to propose a normalization based on the average value of GHI_c rather than using a more complicated Monte Carlo type approach. In this appendix, we show that these two types of normalization are roughly equivalent. More precisely, we demonstrate that the first one is an approximate of the second one (under certain assumptions).

Considering a time series $\varepsilon(t)$ built with a sampling from an uniform distribution (between 0 and 1 equivalent to an artificial clear sky index (CSI), see section II C), it is possible to compute the $RMSE_{max}$ related to Persistence forecasting. In the next, we denote this parameter $RMSE_{max}^{CSI}$ so as not to be mistaken with the $RMSE_{max}$ relating to the GHI . Indeed, the MSE_{max}^{CSI} is given by the well known relation A1 (in which the mean of the signal $\bar{\varepsilon}$ has been added and then subtracted).

$$MSE_{max}^{CSI} = \frac{1}{n} \sum_{t=1}^n \left(\varepsilon(t) - \bar{\varepsilon} + \bar{\varepsilon} - \varepsilon(t-1) \right)^2. \quad (A1)$$

This algebraic identity can be simplified as shown in Equation A2:

$$MSE_{max}^{CSI} = \frac{1}{n} \sum_{t=1}^n \left(\varepsilon(t) - \bar{\varepsilon} \right)^2 + \frac{1}{n} \sum_{t=1}^n \left(\varepsilon(t-1) - \bar{\varepsilon} \right)^2 - \frac{2}{n} \sum_{t=1}^n \left(\varepsilon(t) - \bar{\varepsilon} \right) \left(\varepsilon(t-1) - \bar{\varepsilon} \right). \quad (A2)$$

The first two terms of the right member of this equation correspond to the variance of the signal (σ_{ε}^2), and in the case of an uniform law with values between 0 and 1 (and with n large enough), we know that⁴⁵:

$$\sum_{t=1}^n \left(\varepsilon(t) - \bar{\varepsilon} \right)^2 = \sum_{t=1}^n \left(\varepsilon(t-1) - \bar{\varepsilon} \right)^2 = n\sigma_{\varepsilon}^2 = \frac{n}{12}. \quad (A3)$$

Regarding the last term of the Equation A2, as the elements

constituting the series ε are completely independent ($\varepsilon(t)$ and $\varepsilon(t-1)$ are independent random variables), their covariance is thereby zero⁴⁵.

The above considerations allow rewriting Equation A1 as $MSE_{max}^{CSI} = 1/6$ and consequently $RMSE_{max}^{CSI} = 1/\sqrt{6}$.

If, in the case of random number between 0 and 1, the theoretical approach is possible, it is not the case for the MSE_{max} related to the GHI (between 0 and GHI_c) given that the latter fluctuates. In this case, the Monte Carlo approach is the only one really effective but an approximation can be made.

Assuming that the $nRMSE_{max}^{CSI}$ related to CSI gives the same error estimation than the $nRMSE_{max}$ related to the GHI , we obtain (posing $\bar{\varepsilon} = 1/2$) $nRMSE_{max}^{CSI} = 2/\sqrt{6}$ leading to $RMSE_{max} = (2/\sqrt{6})\overline{GHI}_c$ with \overline{GHI}_c the mean value of GHI_c . This approximate is valid only if a filtering is operated and if the data related to the night are not taken into account. So $RMSE_{max}$ could be computed only from \overline{GHI}_c but under certain conditions and accepting some uncertainty. For instance, in Ajaccio (latitude of $41^{\circ}56'$) the $RMSE_{max}$ read in the Figure 4 is $249.7Wh.m^{-2}$. For this same site \overline{GHI}_c is $467.3Wh.m^{-2}$ inducing a $RMSE_{max}$ equal to $381.2Wh.m^{-2}$ (using $RMSE_{max} = (2/\sqrt{6})\overline{GHI}_c$). The difference between the two methodologies is greater than 40% (381.2 versus 467.3). There are certainly special cases for which this simple approach gives good results but in all cases it is preferable to use the methodology on the generation of Monte-Carlo type random numbers.

Appendix B: McClear and BSRN studied sites

One can see the studied sites in the table IV. The interesting McClear model estimates clear sky radiation for any point on the globe⁴⁶. Developed by the Centre O.I.E. - MINES Paris Tech/ARMINES, it uses the results of the numerical meteorological model of chemistry - transport of the European MACC projects⁴⁷. *BSRN* is a project of the Panel on Data and Assessments of the Global Energy and Water Cycle Experiment (*GEWEX*) under the World Climate Research Programme (*WCRP*) and, as such, aims to detect significant changes in the radiation field at the Earth's surface that may be related to climate change. This group offers free quality GHI series at a wide range of sites⁴⁸.

Data availability

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due privacy restrictions.

¹D. Yang, E. Wu, and J. Kleissl, "Operational solar forecasting for the real-time market," *International Journal of Forecasting* **35**, 1499–1519 (2019).

²H. T. Pedro and C. F. Coimbra, "Short-term irradiance forecastability for various solar micro-climates," *Solar Energy* **122**, 587 – 602 (2015).

³D. Kumar, "Hyper-temporal variability analysis of solar insolation with respect to local seasons," *Remote Sensing Applications: Society and Environment* **15**, 100241 (2019).

TABLE IV. Characteristics of the locations related to the 50 McCLea⁴⁷ and BSRN⁴⁹ time series (1h time granularity and horizon; 2015). NA when less than 1000 data are available

Station	Köppen ³⁷	Location	Lat (°)	Long (°)	Elev (m)	RMSE _{max} ^{carlo} (Wh/m ²)	F (%)
Alice Springs (ASP)	Bwh	Australia	-23,8	133,9	547	291.1	48.7
Bermuda (BER)	Af	USA	32,3	-64,7	8	253.1	NA
Billings (BIL)	Dfa	USA	36,6	-97,5	317	242.7	38.1
Bondville (BON)	Cfa	USA	40,1	-88,4	213	236.4	52.8
Boulder (BOS)	BSk	USA	40,1	-105,2	1689	252.3	55.7
Brasilia (BRB)	Aw	Brazil	-15,6	-47,7	1023	290.9	43.9
Cabauw (CAB)	Cfb	Netherlands	52,0	4,9	0	200.4	64.5
Camborne (CAM)	Cfb	United Kingdom	50,2	-5,3	88	208.1	61.9
Cener (CNR)	Cfb	Spain	42,8	-1,6	471	229.1	NA
Cocos Island (COC)	Aw	Australia	-12,2	96,8	6	282.4	NA
De Aar (DAA)	BSk	South Africa	-30,7	24,0	1287	279.0	NA
Concordia Station (DOM)	EF	Antarctica	-75,1	123,4	3233	164.0	81.6
Desert (DRA)	BWh	USA	36,6	-116,0	1007	257.9	29.1
Darwin Met Office (DWN)	Aw	Australia	-12,4	130,9	32	285.9	79.9
Southern Great Plains (E13)	BSk	USA	36,6	-97,5	318	285.4	69.2
Florianopolis (FLO)	Cfa	Brazil,	-27,6	-48,5	11	267.3	50.8
Fort Peck (FPE)	BSk	USA	48,3	-105,1	634	216.2	46.1
Fukuoka (FUA)	Cwa	Japan	33,6	130,4	3	247.3	68.0
Gandhinagar (GAN)	BSh	India	23,1	72,6	65	240.8	39.9
Goodwin Creek (GCR)	Cfa	USA	34,3	-89,9	98	281.0	58.1
Gobabeb (GOB)	Csb	Namibia	-23,6	15,0	407	247.3	70.1
Gurgaon (GUR)	BSh	India	28,4	77,2	259	238.9	78.3
George von Neumayer (GVN)	EFs	Antarctica	-70,6	-8,2	42	151.3	69.8
Howrah (HOW)	Aw	India	22,5	88,3	51	243.1	53.9
Ishigaki jima (ISH)	Cfa	Japan	24,3	124,2	5,7	259.9	36.3
Izaña (IZA)	BWh	Spain	28,3	-16,5	2373	300.7	78.4
Kwajalein (KWA)	Af	Marshall Islands	8,7	167,7	10	294.6	54.1
Lauder (LAU)	Cfb	New Zealand	-45,0	169,7	350	236.2	NA
Lerwick (LER)	Cfb	United Kingdom	60,1	-1,2	80	172.8	48.5
Lindenberg (LIN)	Cfb	Germany	52,2	14,1	125	199.9	62.6
Lulin (LLN)	Cfa	Taiwan	23,5	120,9	2862	293.3	NA
Langley (LRC)	Cfb	USA	37,1	-76,4	3	239.1	68.2
Minamitorishima (MNM)	Cfa	Japan	24,3	154,0	7	267.3	68.5
Ny-Ålesund (NYA)	ET	Norway	79,0	11,9	11	117.9	64.8
Huancayo (OHY)	Cwb	Peru	-12,0	-75,3	3314	328.7	NA
Palaiseau (PAL)	Cfb	France	48,7	2,2	156	205.7	33.5
Payerne (PAY)	Cfb	Switzerland	46,8	6,9	491	219.8	67.2
Rock Springs (PSU)	BSk	USA	40,7	-77,9	376	237.5	62.4
Petrolina (PTR)	BSh	Brazil	-9,1	-40,3	387	295.2	31.6
Sapporo (SAP)	Dfb	Japan	43,1	141,3	17	227.3	24.9
São Martinho da Serra (SMS)	Cfa	Brazil	-29,4	-53,8	489	259.6	48.5
Sonnblick (SON)	Dfb	Austria	47,1	13,0	3109	246.1	49.4
Sioux Falls (SXF)	Dfa	USA	43,7	-96,6	473	231.8	53.4
Syowa (SYO)	EF	Antarctica	-69,0	39,6	18	161.2	73.1
Tamanrasset (TAM)	Csa	Algeria	22,8	5,5	1385	286.2	70.6
Tateno (TAT)	Dfb	Japan	36,1	140,1	25	236.4	66.3
Tiksi (TIK)	Dfd	Russia	71,6	128,9	48	136.3	64.5
Tiruvallur (TIR)	Aw	India	13,1	80,0	36	277.9	46.0
Toravere (TOR)	Dfb	Estonia	58,2	26,5	70	175.7	NA
Yushan Station (YUS)	Cwa	Taiwan	23,5	121,0	3858	298.5	NA

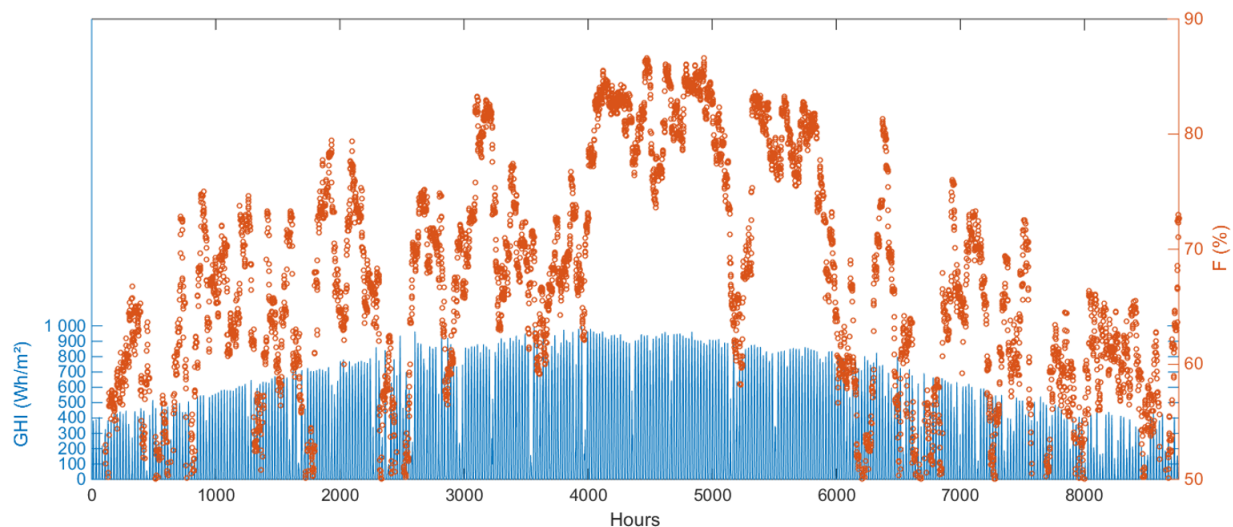


FIG. 10. Intra-annual variation of F . F is estimated from hourly GHI time series acquired at Ajaccio and for a 1h forecast horizon.

- 732 ⁴R. Marquez and C. F. M. Coimbra, “Proposed Metric for Evaluation of Solar
733 Forecasting Models,” *Journal of Solar Energy Engineering* **135** (2013),
734 10.1115/1.4007496.
- 735 ⁵R. Perez and T. E. Hoff, “Chapter 6 - Solar Resource Variability,” in *Solar*
736 *Energy Forecasting and Resource Assessment*, edited by J. Kleissl (Academ-
737 ic Press, Boston, 2013) pp. 133–148.
- 738 ⁶R. Blaga and M. Paulescu, “Quantifiers for the solar irradiance variability:
739 A new perspective,” *Solar Energy* **174**, 606–616 (2018).
- 740 ⁷C. Pan, C. Wang, Z. Zhao, J. Wang, and Z. Bie, “A Copula Function Based
741 Monte Carlo Simulation Method of Multivariate Wind Speed and PV Power
742 Spatio-Temporal Series,” *Energy Procedia Renewable Energy Integration*
743 with Mini/Microgrid, **159**, 213–218 (2019).
- 744 ⁸D. Yang, “A universal benchmarking method for probabilistic solar irradiance
745 forecasting,” *Solar Energy* **184**, 410–416 (2019).
- 746 ⁹F. J. Rodríguez-Benítez, C. Arbizu-Barrena, F. J. Santos-Alamillos,
747 J. Tovar-Pescador, and D. Pozo-Vázquez, “Analysis of the intra-day solar
748 resource variability in the iberian peninsula,” *Solar Energy* **171**, 374 –
749 387 (2018).
- 750 ¹⁰A. Foulloy, C. Voyant, G. Notton, F. Motte, C. Paoli, M.-L. Nivet, E. Guil-
751 lot, and J.-L. Duchaud, “Solar irradiation prediction with machine learn-
752 ing: Forecasting models selection method depending on weather variabil-
753 ity,” *Energy* **165**, 620–629 (2018).
- 754 ¹¹C. Voyant, T. Soubdhan, P. Lauret, M. David, and M. Muselli, “Statistical
755 parameters as a means to a priori assess the accuracy of solar forecasting
756 models,” *Energy* **90**, 671–679 (2015).
- 757 ¹²D. Yang, “Making reference solar forecasts with climatology, persistence,
758 and their optimal convex combination,” *Solar Energy* **193**, 981 – 985
759 (2019).
- 760 ¹³F. X. Diebold and L. Kilian, “Measuring predictability: theory and macroe-
761 conomic applications,” *Journal of Applied Econometrics* **16**, 657–669
762 (2001), <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jae.619>.
- 763 ¹⁴C. Amat, T. Michalski, and G. Stoltz, “Fundamentals and exchange rate
764 forecastability with simple machine learning methods,” *Journal of Interna-
765 tional Money and Finance* **88**, 1–24 (2018).
- 766 ¹⁵C. W. J. Granger and P. Newbold, “Forecasting transformed series,” *Journal*
767 *of the Royal Statistical Society: Series B (Methodological)* **38**, 189–
768 203 (1976), [https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-
769 6161.1976.tb01585.x](https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1976.tb01585.x).
- 770 ¹⁶J. Boylan, “Toward a More Precise Definition of Forecastability,” *Foresight:*
771 *The International Journal of Applied Forecasting*, 34–40 (2009).
- 772 ¹⁷B. Weghenkel, A. Fischer, and L. Wiskott, “Graph-based predictable fea-
773 ture analysis,” *CoRR* **abs/1602.00554** (2016), arXiv:1602.00554.
- 774 ¹⁸T. DelSole, “Predictability and Information Theory. Part I: Measures of Pre-
775 dictability,” *Journal of the Atmospheric Sciences* **61**, 2425–2440 (2004).
- 776 ¹⁹A. Donnellan, P. Mora, M. Matsu’ura, and X.-c. Yin, *Computational*
777 *earthquake science. 1* (Springer Science & Business Media, 2004) google-
778 Books-ID: Zd1Lhzbb8vsC.
- 779 ²⁰A. Kumar and M. Chen, “Inherent Predictability, Requirements on
780 the Ensemble Size, and Complementarity,” *Monthly Weather Re-
781 view* **143**, 3192–3203 (2015), [https://journals.ametsoc.org/mwr/article-
782 pdf/143/8/3192/4314575/mwr-d-15-0022_1.pdf](https://journals.ametsoc.org/mwr/article-pdf/143/8/3192/4314575/mwr-d-15-0022_1.pdf).
- 783 ²¹E. Bianco-Martinez and M. S. Baptista, “Space-time nature of causal-
784 ity,” *Chaos: An Interdisciplinary Journal of Nonlinear Science* **28**, 075509
785 (2018).
- 786 ²²D. Ruelle, “Chaos, predictability, and idealization in physics,” *Complexity*
787 **3**, 26–28 (1997).
- 788 ²³L. Fortuna, G. Nunnari, and S. Nunnari, “Analysis of Solar Radiation Time
789 Series,” in *Nonlinear Modeling of Solar Radiation and Wind Speed Time*
790 *Series*, edited by L. Fortuna, G. Nunnari, and S. Nunnari (Springer Inter-
791 national Publishing, Cham, 2016) pp. 17–27.
- 792 ²⁴J. Ostrometzky, A. Bernstein, and G. Zussman, “Irradiance field recon-
793 struction from partial observability of solar radiation,” *IEEE Geoscience*
794 *and Remote Sensing Letters* **16**, 1698–1702 (2019).
- 795 ²⁵J. Zhang and B. M. Hodge, “Forecastability as a design criterion in wind
796 resource assessment: Preprint,” **34** (2014), 10.1016/B978-0-444-63433-
797 7.50095-X.
- 798 ²⁶D. A. Ahlburg, “Error measures and the choice of a forecast method,” *Inter-
799 national Journal of Forecasting* **8**, 99–100 (1992).
- 800 ²⁷S. Cros, O. Liandrat, N. Sébastien, N. Schmutz, and C. Voyant, “Clear sky
801 models assessment for an operational pv production forecasting solution,”
802 in *28th European Photovoltaic Solar Energy Conference and Exhibition*
803 (France, 2013) p. 5BV.4.69.
- 804 ²⁸C. A. Gueymard, “Clear-Sky Radiation Models and Aerosol Effects,” in *Solar*
805 *Resources Mapping: Fundamentals and Applications*, edited by J. Polo,
806 L. Martín-Pomares, and A. Sanfilippo (Springer International Publishing,
807 Cham, 2019) pp. 137–182.
- 808 ²⁹X. Sun, J. M. Bright, C. A. Gueymard, B. Acord, P. Wang, and N. A.
809 Engerer, “Worldwide performance assessment of 75 global clear-sky irra-
810 diance models using principal component analysis,” *Renewable and Sus-
811 tainable Energy Reviews* **111**, 550 – 570 (2019).
- 812 ³⁰A. Dobson, “1. the oxford dictionary of statistical terms. yadolah dodge
813 (ed.), oxford university press, oxford, 2003,hardcover. no. of pages: 506.
814 price: Gbp 25.00. isbn 0-19-850994-4,” *Statistics in Medicine* **23**, 1824–
815 1825 (2004).
- 816 ³¹C. Voyant and G. Notton, “Solar irradiation nowcasting by stochastic per-
817 sistence: A new parsimonious, simple and efficient forecasting tool,” *Re-
818 newable and Sustainable Energy Reviews* **92**, 343 – 352 (2018).
- 819 ³²J. G. Fortin, F. Ancil, L. Étienne Parent, and M. A. Bolinder, “Comparison

- of empirical daily surface incoming solar radiation models,” *Agricultural and Forest Meteorology* **148**, 1332 – 1340 (2008).
- ³³L. R. do Nascimento, T. de Souza Viana, R. A. Campos, and R. Rütger, “Extreme solar overirradiance events: Occurrence and impacts on utility-scale photovoltaic power plants in Brazil,” *Solar Energy* **186**, 370 – 381 (2019).
- ³⁴F. Monforti, T. Huld, K. Bódis, L. Vitali, M. D’Isidoro, and R. Lacal-Arántegui, “Assessing complementarity of wind and solar resources for energy production in Italy. A Monte Carlo approach,” *Renewable Energy* **63**, 576–586 (2014).
- ³⁵C. K. Kim, H.-G. Kim, Y.-H. Kang, C.-Y. Yun, and S. Y. Kim, “Probabilistic prediction of direct normal irradiance derived from global horizontal irradiance over the Korean Peninsula by using Monte-Carlo simulation,” *Solar Energy* **180**, 63–74 (2019).
- ³⁶T. Hou, D. Nuyens, S. Roels, and H. Janssen, “Quasi-Monte Carlo based uncertainty analysis: Sampling efficiency and error estimation in engineering applications,” *Reliability Engineering & System Safety* **191**, 106549 (2019).
- ³⁷J. Ascencio-Vásquez, K. Brecl, and M. Topič, “Methodology of Köppen-Geiger-photovoltaic climate classification and implications to worldwide mapping of PV system performance,” *Solar Energy* **191**, 672 – 685 (2019).
- ³⁸P. Lauret, C. Voyant, T. Soubdhan, M. David, and P. Poggi, “A benchmarking of machine learning techniques for solar radiation forecasting in an insular context,” *Solar Energy* **112**, 446 – 457 (2015).
- ³⁹A. Fouilloy, *Comparaison de méthodes d’apprentissage automatique de prévision de la ressource solaire pour une application à une gestion optimisée des réseaux intelligents*, Theses, Université de CORSE - Pascal PAOLI (2019).
- ⁴⁰J. Kleissl, *Solar Energy Forecasting and Resource Assessment* (Academic Press, 2013) google-Books-ID: 94KIO_SPwW8C.
- ⁴¹L. Benali, G. Notton, A. Fouilloy, C. Voyant, and R. Dizene, “Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components,” *Renewable Energy* **132**, 871–884 (2019).
- ⁴²P. Ineichen, “A broadband simplified version of the solis clear sky model,” *Solar Energy* **82**, 758 – 762 (2008).
- ⁴³A. Laguarda, G. Giacosa, R. Alonso-Suárez, and G. Abal, “Performance of the site-adapted cams database and locally adjusted cloud index models for estimating global solar horizontal irradiation over the pampa húmeda,” *Solar Energy* **199**, 295 – 307 (2020).
- ⁴⁴M. Fliess, C. Join, and C. Voyant, “Prediction bands for solar energy: New short-term time series forecasting techniques,” *Solar Energy* **166**, 519–528 (2018).
- ⁴⁵M. Dekking, *A modern introduction to probability and statistics : understanding why and how* (London : Springer, 2005).
- ⁴⁶D. Yang, “Choice of clear-sky model in solar forecasting,” *Journal of Renewable and Sustainable Energy* **12**, 026101 (2020), <https://doi.org/10.1063/5.0003495>.
- ⁴⁷M. Lefèvre, A. Oumbe, P. Blanc, B. Espinar, B. Gschwind, Z. Qu, L. Wald, M. Schroedter-Homscheidt, C. Hoyer-Klick, A. Arola, A. Benedetti, J. W. Kaiser, and J.-J. Morcrette, “McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions,” *Atmospheric Measurement Techniques* **6**, 2403–2418 (2013).
- ⁴⁸A. Driemel, J. Augustine, K. Behrens, S. Colle, C. Cox, E. Cuevas-Agulló, F. M. Denn, T. Duprat, M. Fukuda, H. Grobe, M. Haeffelin, G. Hodges, N. Hyett, O. Ijima, A. Kallis, W. Knap, V. Kustov, C. N. Long, D. Longenecker, A. Lupi, M. Maturilli, M. Mimouni, L. Ntsangwane, H. Ogihara, X. Olano, M. Olefs, M. Omori, L. Passamani, E. B. Pereira, H. Schmithüsen, S. Schumacher, R. Sieger, J. Tamlyn, R. Vogt, L. Vuilleumier, X. Xia, A. Ohmura, and G. König-Langlo, “Baseline surface radiation network (bsrn): structure and data description (1992–2017),” *Earth System Science Data* **10**, 1491–1501 (2018).
- ⁴⁹D. Yang and J. M. Bright, “Worldwide validation of 8 satellite-derived and reanalysis solar radiation products: A preliminary evaluation and overall metrics for hourly data over 27 years,” *Solar Energy* (2020), <https://doi.org/10.1016/j.solener.2020.04.016>.
- ⁵⁰C. Gueymard *et al.*, *SMARTS2: a simple model of the atmospheric radiative transfer of sunshine: algorithms and performance assessment* (Florida Solar Energy Center Cocoa, FL, 1995).
- ⁵¹P. Lauret, R. Perez, L. Mazorra Aguiar, E. Tapachès, H. M. Diagne, and M. David, “Characterization of the intraday variability regime of solar irradiation of climatically distinct locations,” *Solar Energy* **125**, 99–110 (2016).
- ⁵²R. Perez, S. Kivalov, J. Schlemmer, K. Hemker, and T. E. Hoff, “Short-term irradiance variability: Preliminary estimation of station pair correlation as a function of distance,” *Solar Energy Progress in Solar Energy* **3**, 86, 2170–2176 (2012).
- ⁵³M. Lave and J. Kleissl, “Solar variability of four sites across the state of Colorado,” *Renewable Energy* **35**, 2867–2873 (2010).
- ⁵⁴T. E. Hoff and R. Perez, “Quantifying PV power Output Variability,” *Solar Energy* **84**, 1782–1793 (2010).
- ⁵⁵C. A. Gueymard, “Clear-sky radiation models and aerosol effects,” in *Solar Resources Mapping: Fundamentals and Applications*, edited by J. Polo, L. Martín-Pomares, and A. Sanfilippo (Springer International Publishing, Cham, 2019) pp. 137–182.
- ⁵⁶M. Lave, R. J. Broderick, and M. J. Reno, “Solar variability zones: Satellite-derived zones that represent high-frequency ground variability,” *Solar Energy* **151**, 119 – 128 (2017).
- ⁵⁷G. Lohmann, “Irradiance variability quantification and small-scale averaging in space and time: A short review,” *Atmosphere* **9**, 264 (2018).