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

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 Soubdhan[4]

[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.

I. 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 field[1]. In the solar forecasting community, many researchers use different metrics in order to assess the difficulty in generating good forecasts for different climates[2]. 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 vary[3]. This type of metric is used to compare the data at hand to other sets of data. Some authors, like Marquez and Coimbra[4], Perez and Hoff[5] or Blaga and Paulescu[6] 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 complexity[7]. 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 Hoff[5] and Marquez and Coimbra[4], 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 proxys 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 Coimbra[2], 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 Coimbra[2] 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

authors like Yang [12]. In this work, and based on a bibliographic survey, we opt for the notion of forecastability. Based on a rather general bibliographic survey, the next subsection tries to shed some light on the difference between *predictability* (P)[13] and *forecastability* (F)[14i].

A. Difference between forecastability and predictability

One of the first references about the forecastability and the time series formalism is the result of the work of the co-recipient of the Nobel Prize Clive W.J. Granger[15]. Authors define it as the variance of the optimal forecast divided by the unconditional variance of the time series. This definition and the resulting Q parameter (forecastability quotient) were extensively studied in economics, gradually giving way to new kind tools as sample and approximate entropy[16], correlation and mutual information metrics[17].

These two notions (P and F) are conceptually very close: if the predictability (P)[18] studies how trajectories of the true system diverge[19], the forecastability (F) describes how a model trajectory diverges from a true system trajectory[14]. A common explanation is that a predictable process is able to be predicted while a forecastable one is able of being forecasted. With this last definition, the concept of modelling appears, thereby a forecastable system is necessarily predictable but the opposite is not true [ii]. The predictability term which is often used with dynamic processes, is closely related to notions like causality [21] or chaos (i.e. failure of predictability[22]), found for example, in all weather series and where the typical predictable times (or barriers) concerns the prediction horizons smaller than 1 or 5 days[23]. In the context of the present study (nowcasting or very short term), the chaotic aspect is not directly studied, so, the term forecastability seems more suitable than predictability one even if, to the best of our knowledge, there isn't any consensus on that.

Note that other concepts could have been detailed here (observability and detection reliability[24]) but they do not have much sense in the study of solar radiation given the quantities involved are directly measurable and the associated uncertainties are very low.

B. Framework of the study

As has been done for wind speed forecasting[25], one may also use the performance[iii][26]. of the Persistence (P) model which is, in this work, related to a persistent clear sky index k_t^*. More precisely, the *nRMSE* obtained by this reference model can be used to gauge the forecastability of a particular site. However, although this method is feasible, it would not provide the same type of information. Indeed, a Persistence forecast with an *nRMSE* equal to 20 % is not sufficient to consider if the site is related to a high or a low forecastability. It depends on the kind of solar component (direct, diffuse, global), on the forecast horizon, on the latitude, on the topographic relief, on the nebulosity, etc. For instance, for *GHI* hourly data and 4-hour forecast horizon, a Persistence forecast with an *nRMSE* equal to 20% could be an indication of a high forecastability, but for *GHI* data with a 5-min granularity and for a 5-min forecast horizon, certainly not. Put differently, the main drawback of using the score of the reference Persistence model is that it is not bounded by an upper limit.

In this work, the objective is to propose a forecastability metric (denoted hereafter F) which is easy to compute and easy to interpret. In other words, this new metric should provide users a global reference in order to assess the inherent difficulty to issue forecasts for a specific site and consequently if it makes sense to add an extra effort to build more and more complex forecasting models. Additionally, it would be desirable to relate in a simple way this (ex-ante) forecastability metric with the (ex-post) skill score metric defined by Marquez and Coimbra [4]. Hence, the F metric should be helpful in comparing the numerous forecasting methods proposed by the community.

To compute the F metric, the user will need to compute the *RMSE* of the Persistence model [iv]. Moreover, as it will be discussed in section III C, this approach will allow adapting the forecastability with the forecast horizon and with the different components of the solar radiation.

The rest of the manuscript is structured in four sections. The first one (section II) presents the mathematical formulation of the forecastability F. In section III, the sensitivity of F to various parameters is evaluated while in section IV is proposed an estimate of F concerning 50 time series of *GHI*. Finally, Section V gives some conclusions and some perspectives.

II. ESTIMATION PROPOSAL FOR F

Before proposing a new method to estimate the forecastability F, it is important to understand why in the existing formalism of the solar radiation time series, some elements could be improved. As discussed in the introduction, in the time series domain, the variability (V) is often estimated by computing the variance[6]. In the context of solar radiation, the time series data are very particular in the sense that the underlying trend is periodic and easily quantifiable using robust models validated by the community.

These models are denoted clear sky models[27] and they correspond to the solar irradiation estimated under clear sky conditions (GHI_c) and computed from the sun position and vari-

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

[ii] Kumar and Chen [20] 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

[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).

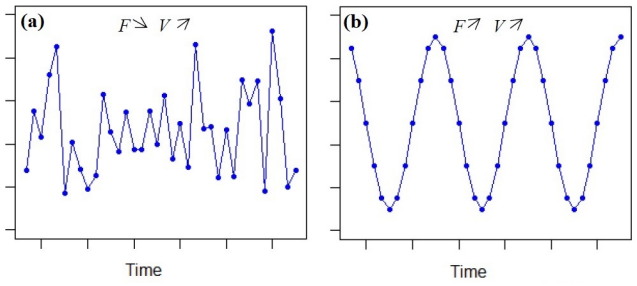

FIG. 1. Intuitive relationship between forecastability (F) and variability (V)

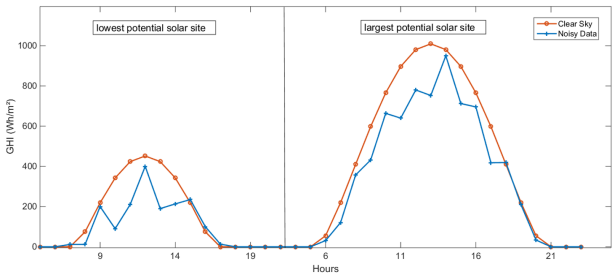

FIG. 2. Clear sky modeling vs noisy data for two sites with different solar potential

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.

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

Figure 1 permits to intuitively introduce the notion of forecastability. 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 forecastability 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 horizontal 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.

B. Interest of the normalization in computing F

In statistics, normalization is a very frequently used technique which refers to the creation of shifted and scaled versions of certain variables (see normalization of scores proposed by Dobson [30]). The primary intent of this change is that these normalized values promote the comparison of a certain 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 normalization in the calculation of F, Figure 2 shows one day of modelled GHI_c 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* related 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 problematic prompts us to propose a *normalized* formulation of F that takes into account the studied site and its characteristics (including the data filtering process) as we will see in the following subsection.

Again, it may seem tempting to quantify the forecastability relying solely on the interpretation of the *nRMSE* of the Persistence forecast. But, as already stated in the introduction, 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.

C. Mathematical formalism for calculating F

In what follows, we detail the mathematical formalism using 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 forecasting 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 data[31]. Note that recent studies have highlighted the use of "improved" Persistence model such as an optimal convex optimization of Persistence on k_t^* and climatology[12]. 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 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 over-irradiance[33] (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 unsignificant 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 method[35], is based on a computational algorithm that relies on repeated random sampling to obtain numerical results[36].

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 $GHI(t) = k_t^*(t).GHI_c(t)$. 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.

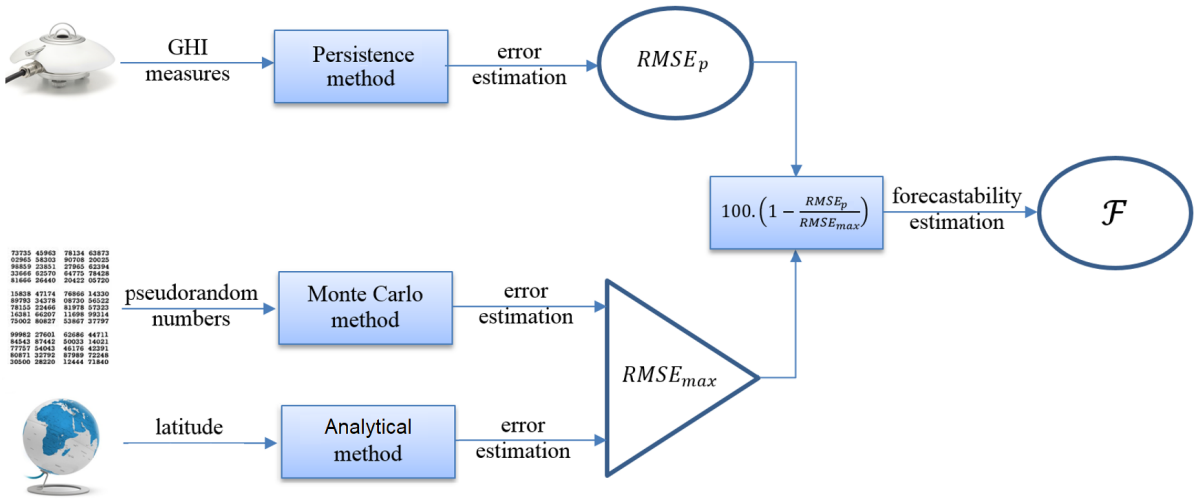

FIG. 3. Methodology to estimate the forecastability (F)

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 $RMSE_{max}^{mcarlo}$), 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 $RMSE_{max}^{mcarlo}$

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 $RMSE_{max}^{mcarlo} = RMSE(GHI - \widehat{GHI})$

[a] See Eq 5

With the Solis model[11], 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 $R^2 = 0.999$):

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

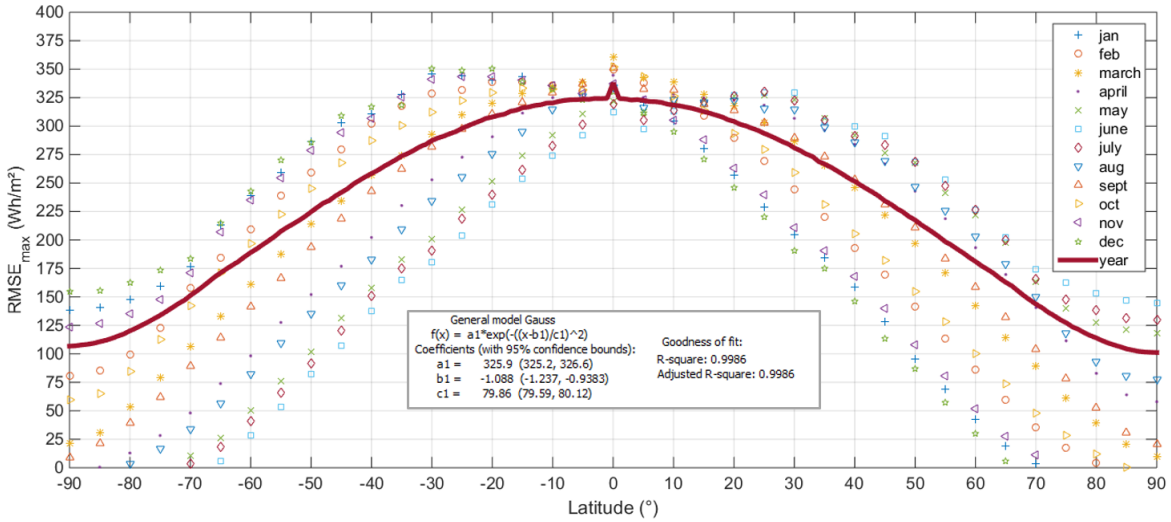

FIG. 4. The magnitudes of the $RMSE_{max}^{analytic}$ versus the latitude. Monthly and yearly values are given.

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

Currently, most of the forecasting studies evaluate the performance of the models in terms of forecast skill. As mentioned in the introduction, it would be desirable to simply relate the forecastability metric to a skill score. As the forecastability is computed from the Persistence forecasts, it is straightforward to establish the link between the skill score ($SS = 1 - \frac{RMSE}{RMSE_P}$) and F. Considering Equations 4 and Algorithm 1, Equation 7 related to the link between F and SS can be established as:

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

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

III. FACTORS INFLUENCING F

Some parameters influence the factor F such as the site location, the solar radiation component (global, beam and diffuse), the prediction horizon and the time granularity (this list is not intended to be exhaustive).

A. F variation with location

The F value quantifies the difficulty to predict the solar radiation components. Table I lists the F values for 6 locations (see also Figure 5): Ajaccio (Corsica island, France), Nancy (East of France), Odeillo (mountainous site, Pyrenées Orientales, France), Tilos (Greek island), Saint Pierre (Reunion island, France), Le Raizet (Guadeloupe archipelago, France). Those stations are not enough to prove the generalization of the forecastability. In section IV, it will be tested on 50 stations spread all over the globe. Here the solar radiation component of interest is the GHI and the forecast horizon and the time granularity is 1 hour for all sites.

FIG. 5. The 6 locations used to estimate F

As shown by Table I, the higher the forecastability (F), the lower the error metric ($nRMSE$) related to predictions generated by a multilayer perceptron (MLP[10,38,39]). The relationship between these two metrics is not obvious to estimate. However, it must be stressed that factors such as the quality of the measurements as well as the concordance of the measurements with the clear sky model may complicate the interpretation. This table validates what has been widely demonstrated in the literature i.e. mountainous regions (Odeillo) and continental climates (Nancy) are more difficult to apprehend than coastal areas (Ajaccio and Tilos). With this table, it is interesting to realize that the SS alone does not allow to judge the

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* predictions

TABLE 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

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

B. F evolution with the solar radiation components

The methodology used to compute F can be applied to the 3 components of solar radiation (global, beam and diffuse)[40]. The only prerequisite is to estimate the clear sky solar value corresponding to each component. Table II gives the F values and the prediction errors for Odeillo (the site with the most important variability, for a forecast horizon and a time granularity of 1 hour[10,39]).

As shown by Table II, the other solar radiation components (beam and diffuse) exhibit lower F values than the one estimated for the *GHI*. Conversely to global irradiation, the clear sky modelling for the beam and diffuse components is less efficient. As an illustration, the values obtained with the simplified Solis version are within 1% for the global component, 2% for the beam component and 5% for the diffuse component (read Ineichen[42] for details). In this subsection, the conclusion is similar than the one stated in subsection III A that is $nRMSE$ trend is a bijective function of F but the relationship between F and $nRMSE$ is non-linear. However, the difference in forecastability between Beam and Diffuse components is close to 1% while the $nRMSE$ fluctuates by more than 2 percentage points. Consequently, it must be stressed that the proposed methodology for estimating F will not be relevant if the clear sky model related to the solar radiation component of interest lacks precision.

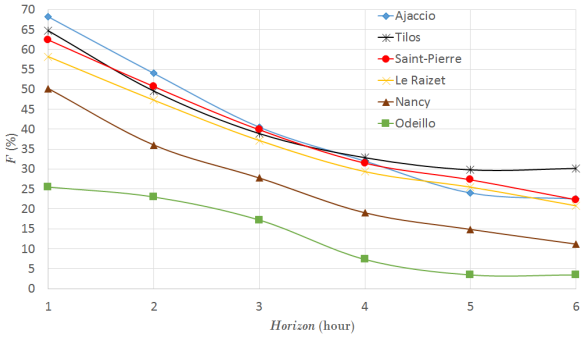

FIG. 6. Forecastability and prediction horizons (F computed from $RMSE_P$ found in[10,38,41])

C. F evolution with the forecast time horizon

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

D. F variation with time granularity

In Table III, it can be seen that for Tilos (the only site where 10 minutes data were available), when we realized global irradiation forecasts (MLP[39]) for horizons of respectively, 1 hour, 15 minutes and 10 minutes[10], the conclusion follows the logic observed so far that is F and $nRMSE$ are strongly statistically dependent and when one increases the other decreases.

E. Conclusion about the F estimation

This section showed the link between the F factor computed from Equation 4 and some parameters frequently used in studies related to the prediction of global irradiation or solar radiation components through the time series formalism. In Figure 7, this link can be estimated by comparison between the forecastability value and the prediction error related to *MLP* forecasts. For each kind of parameters (respectively

TABLE III. Value of F according to the time granularity and $nRMSE$ realated to MLP prediction[10] 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

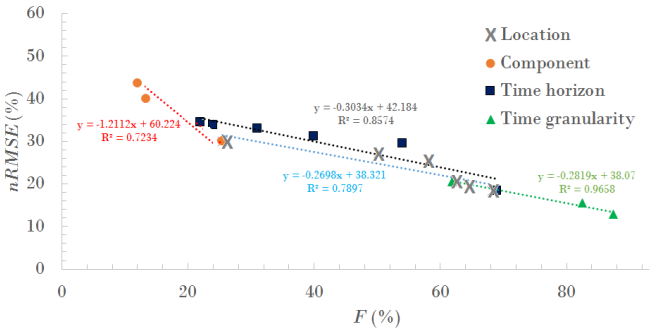

FIG. 7. Relationship between F and prediction error $nRMSE$. Linear fits are provided in order to better highlight the monotonic relationship between F and $nRMSE$.

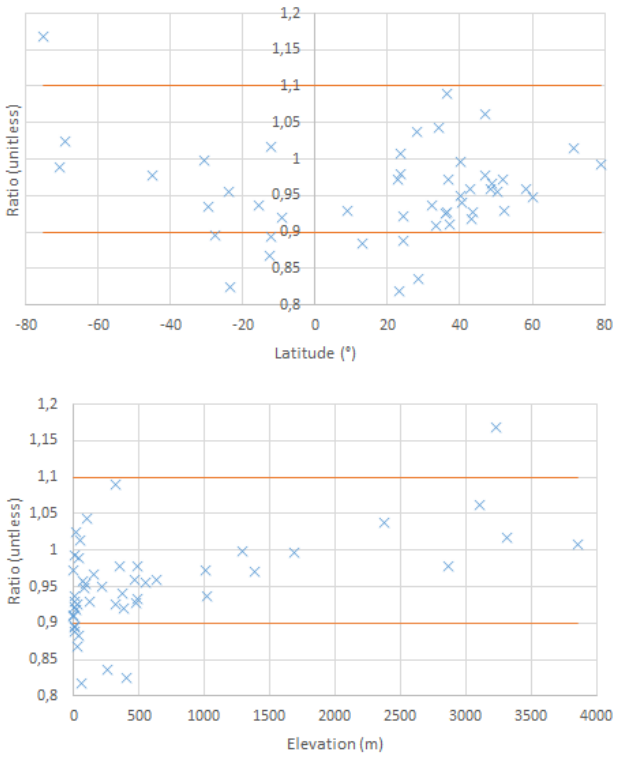

FIG. 8. $Ratio = RMSE_{max}^{mcarlo}/RMSE_{max}^{analytic}$ considering elevation and latitude. 82% of ratio value are comprised between 0.9 and 1.1.

the location, radiation component, the time granularity and the lead time) the link between $nRMSE$ and F is monotonic and when F increases the error metrics decreases. The relationship is not linear and certainly depends on the predictor used (MLP in our case). What is verified is that when the forecastability is good (in the sense of a high value of F) the forecast becomes easier.

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

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

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

In this section, we propose to estimate $RMSE_{max}$ from McClear web service series. The calculation methodologies previously developed and based from random number generation (Cf $RMSE_{max}^{mcarlo}$ in section II C) and based solely on the latitude ($RMSE_{max}^{analytic}$ computed from Eq 6) are compared. The characteristics of the time series are available in Appendix B (Table IV). The way of calculating the forecastability from random numbers (Figure 3) is synthesized by Algorithm 1, which can be applied from anywhere on the surface of the globe, as long as a clear sky radiation estimate (GHI_c) is known but above all optimized to be as reliable as possible.

Figure 8 makes it possible to quantify the difference between a direct calculation ($RMSE_{max}^{mcarlo}$ described by Algorithm 1) and the use of Equation 6 ($RMSE_{max}^{analytic}$). We can appreciate a good match between these two ways of operating. However, stations abbreviated as (see Table IV) COC, DOM, FLO, GAN, GOB, GUR, HOW, ISH, TIR are the stations with more than 10% of difference between the two methods (the mean of ratio is 0.95). There are several ways for understanding these discrepancies, such as considering only the latitude in Equation 6 or the fact that no post-processing was performed with the McClear model. This model is undoubtedly one of the best performing model at the present, but there are some uncertainties relating to certain locations that have been reported in the literature. For example, Laguarda *et al.*[43] reported errors related to the use of McClear model close to 5% in average and which can reach more than 10% some particular periods.

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

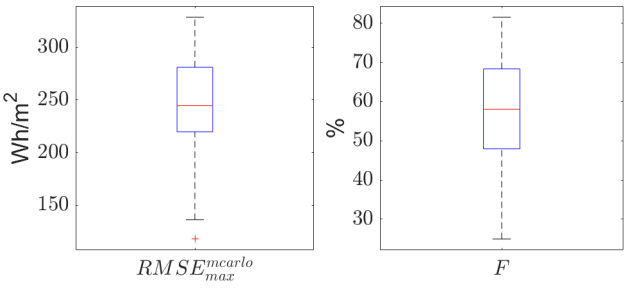

FIG. 9. Boxplots concerning $RMSE_{max}^{mcarlo}$ and F computed from McClear and BSRN data

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 I A). 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 previous 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 literature[10]. 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 Voyant[44], 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).

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

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

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

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 that[45]:

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

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 zero[45].

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 $\overline{\varepsilon} = 1/2$) $nRMSE_{max}^{CSI} = 2/\sqrt{6}$ 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 $41^o56'$) the $RMSE_{max}$ read in the Figure 4 is $249.7 Wh.m^{-2}$. For this same site $\overline{GHI_c}$ is $467.3 Wh.m^{-2}$ inducing a $RMSE_{max}$ equal to $381.2 Wh.m^{-2}$ (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 globe[46]. Developed by the Centre O.I.E. - MINES ParisTech/ARMINES, it uses the results of the numerical meteorological model of chemistry - transport of the European MACC projects[47]. *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 sites[48].

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.

[1] D. Yang, E. Wu, and J. Kleissl, "Operational solar forecasting for the real-time market," International Journal of Forecasting **35**, 1499–1519 (2019).
[2] H. T. Pedro and C. F. Coimbra, "Short-term irradiance forecastability for various solar micro-climates," Solar Energy **122**, 587 – 602 (2015).
[3] D. 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 McCLear[47] and BSRN[49] time series (1h time granularity and horizon; 2015). NA when less than 1000 data are available

Station	**Köppen**[37]	**Location**	**Lat (°)**	**Long (°)**	**Elev (m)**	$\mathbf{RMSE}_{\mathbf{max}}^{\mathbf{mcarlo}}(\mathbf{Wh/m^2})$	**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

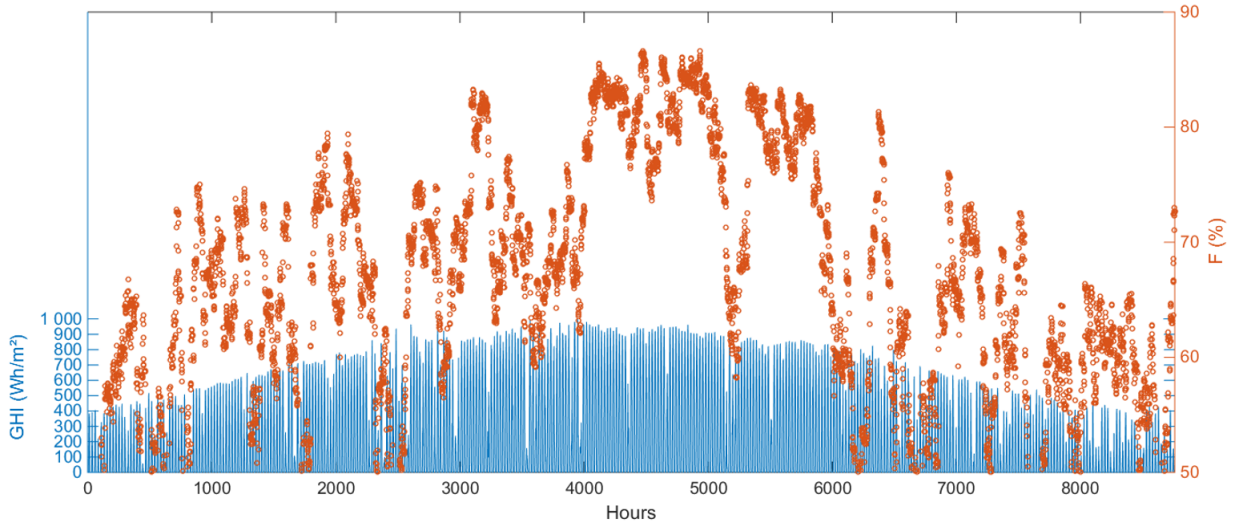

FIG. 10. Intra-annual variation of F. F is estimated from hourly *GHI* time series acquired at Ajaccio and for a 1h forecast horizon.

[4] R. Marquez and C. F. M. Coimbra, "Proposed Metric for Evaluation of Solar Forecasting Models," Journal of Solar Energy Engineering **135** (2013), 10.1115/1.4007496.

[5] R. Perez and T. E. Hoff, "Chapter 6 - Solar Resource Variability," in *Solar Energy Forecasting and Resource Assessment*, edited by J. Kleissl (Academic Press, Boston, 2013) pp. 133–148.

[6] R. Blaga and M. Paulescu, "Quantifiers for the solar irradiance variability: A new perspective," Solar Energy **174**, 606–616 (2018).

[7] C. Pan, C. Wang, Z. Zhao, J. Wang, and Z. Bie, "A Copula Function Based Monte Carlo Simulation Method of Multivariate Wind Speed and PV Power Spatio-Temporal Series," Energy Procedia Renewable Energy Integration with Mini/Microgrid, **159**, 213–218 (2019).

[8] D. Yang, "A universal benchmarking method for probabilistic solar irradiance forecasting," Solar Energy **184**, 410–416 (2019).

[9] F. J. Rodríguez-Benítez, C. Arbizu-Barrena, F. J. Santos-Alamillos, J. Tovar-Pescador, and D. Pozo-Vázquez, "Analysis of the intra-day solar resource variability in the iberian peninsula," Solar Energy **171**, 374 – 387 (2018).

[10] A. Fouilloy, C. Voyant, G. Notton, F. Motte, C. Paoli, M.-L. Nivet, E. Guillot, and J.-L. Duchaud, "Solar irradiation prediction with machine learning: Forecasting models selection method depending on weather variability," Energy **165**, 620–629 (2018).

[11] C. Voyant, T. Soubdhan, P. Lauret, M. David, and M. Muselli, "Statistical parameters as a means to a priori assess the accuracy of solar forecasting models," Energy **90**, 671–679 (2015).

[12] D. Yang, "Making reference solar forecasts with climatology, persistence, and their optimal convex combination," Solar Energy **193**, 981 – 985 (2019).

[13] F. X. Diebold and L. Kilian, "Measuring predictability: theory and macroeconomic applications," Journal of Applied Econometrics **16**, 657–669 (2001), https://onlinelibrary.wiley.com/doi/pdf/10.1002/jae.619.

[14] C. Amat, T. Michalski, and G. Stoltz, "Fundamentals and exchange rate forecastability with simple machine learning methods," Journal of International Money and Finance **88**, 1–24 (2018).

[15] C. W. J. Granger and P. Newbold, "Forecasting transformed series," Journal of the Royal Statistical Society: Series B (Methodological) **38**, 189–203 (1976), https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1976.tb01585.x.

[16] J. Boylan, "Toward a More Precise Definition of Forecastability," Foresight: The International Journal of Applied Forecasting , 34–40 (2009).

[17] B. Weghenkel, A. Fischer, and L. Wiskott, "Graph-based predictable feature analysis," CoRR **abs/1602.00554** (2016), arXiv:1602.00554.

[18] T. DelSole, "Predictability and Information Theory. Part I: Measures of Predictability," Journal of the Atmospheric Sciences **61**, 2425–2440 (2004).

[19] A. Donnellan, P. Mora, M. Matsu'ura, and X.-c. Yin, *Computational earthquake science. 1* (Springer Science & Business Media, 2004) google-Books-ID: Zd1Lhzbb8vsC.

[20] A. Kumar and M. Chen, "Inherent Predictability, Requirements on the Ensemble Size, and Complementarity," Monthly Weather Review **143**, 3192–3203 (2015), https://journals.ametsoc.org/mwr/article-pdf/143/8/3192/4314575/mwr-d-15-0022_1.pdf.

[21] E. Bianco-Martinez and M. S. Baptista, "Space-time nature of causality," Chaos: An Interdisciplinary Journal of Nonlinear Science **28**, 075509 (2018).

[22] D. Ruelle, "Chaos, predictability, and idealization in physics," Complexity **3**, 26–28 (1997).

[23] L. Fortuna, G. Nunnari, and S. Nunnari, "Analysis of Solar Radiation Time Series," in *Nonlinear Modeling of Solar Radiation and Wind Speed Time Series*, edited by L. Fortuna, G. Nunnari, and S. Nunnari (Springer International Publishing, Cham, 2016) pp. 17–27.

[24] J. Ostrometzky, A. Bernstein, and G. Zussman, "Irradiance field reconstruction from partial observability of solar radiation," IEEE Geoscience and Remote Sensing Letters **16**, 1698–1702 (2019).

[25] J. Zhang and B. M. Hodge, "Forecastability as a design criterion in wind resource assessment: Preprint," **34** (2014), 10.1016/B978-0-444-63433-7.50095-X.

[26] D. A. Ahlburg, "Error measures and the choice of a forecast method," International Journal of Forecasting **8**, 99–100 (1992).

[27] S. Cros, O. Liandrat, N. Sébastien, N. Schmutz, and C. Voyant, "Clear sky models assessment for an operational pv production forecasting solution," in *28th European Photovoltaic Solar Energy Conference and Exhibition* (France, 2013) p. 5BV.4.69.

[28] C. 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.

[29] X. Sun, J. M. Bright, C. A. Gueymard, B. Acord, P. Wang, and N. A. Engerer, "Worldwide performance assessment of 75 global clear-sky irradiance models using principal component analysis," Renewable and Sustainable Energy Reviews **111**, 550 – 570 (2019).

[30] A. Dobson, "1. the oxford dictionary of statistical terms. yadolah dodge (ed.), oxford university press, oxford, 2003,hardcover. no. of pages: 506. price: Gbp 25.00. isbn 0-19-850994-4," Statistics in Medicine **23**, 1824–1825 (2004).

[31] C. Voyant and G. Notton, "Solar irradiation nowcasting by stochastic persistence: A new parsimonious, simple and efficient forecasting tool," Renewable and Sustainable Energy Reviews **92**, 343 – 352 (2018).

[32] J. G. Fortin, F. Anctil, L. Étienne Parent, and M. A. Bolinder, "Comparison

of empirical daily surface incoming solar radiation models," Agricultural and Forest Meteorology **148**, 1332 – 1340 (2008).
[33]L. R. do Nascimento, T. de Souza Viana, R. A. Campos, and R. Rüther, "Extreme solar overirradiance events: Occurrence and impacts on utility-scale photovoltaic power plants in brazil," Solar Energy **186**, 370 – 381 (2019).
[34]F. 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).
[35]C. 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).
[36]T. 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).
[37]J. 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).
[38]P. 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).
[39]A. 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).
[40]J. Kleissl, *Solar Energy Forecasting and Resource Assessment* (Academic Press, 2013) google-Books-ID: 94KI0_SPwW8C.
[41]L. 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).
[42]P. Ineichen, "A broadband simplified version of the solis clear sky model," Solar Energy **82**, 758 – 762 (2008).
[43]A. 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).
[44]M. Fliess, C. Join, and C. Voyant, "Prediction bands for solar energy: New short-term time series forecasting techniques," Solar Energy **166**, 519–528 (2018).
[45]M. Dekking, *A modern introduction to probability and statistics : understanding why and how* (London : Springer, 2005).
[46]D. 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.
[47]M. 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).
[48]A. 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).
[49]D. 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.
[50]C. 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).
[51]P. 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).
[52]R. 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).
[53]M. Lave and J. Kleissl, "Solar variability of four sites across the state of Colorado," Renewable Energy **35**, 2867–2873 (2010).
[54]T. E. Hoff and R. Perez, "Quantifying PV power Output Variability," Solar Energy **84**, 1782–1793 (2010).
[55]C. 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.
[56]M. 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).
[57]G. Lohmann, "Irradiance variability quantification and small-scale averaging in space and time: A short review," Atmosphere **9**, 264 (2018).