Title

Predictability and forecast skill of solar irradiance over the contiguous United States

Authors

Liu, Bai

Yang, Dazhi

Mayer, Martin János

et al.

Publication Date

2023-08-01

DOI

10.1016/j.rser.2023.113359

Predictability and forecast skill of solar irradiance over the contiguous United States

Bai Liu[a], Dazhi Yang[a,*], Martin János Mayer[b], Carlos F. M. Coimbra[c], Jan Kleissl[c], Merlinde Kay[d], Wenting Wang[a], Jamie M. Bright[e], Xiang'ao Xia[f], Xin Lv[g], Dipti Srinivasan[h], Yan Wu[i], Hans Georg Bayer[j], Gokhan Mert Yagli[k], Yanbo Shen[l]

[a]*School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin, Heilongjiang, China*
[b]*Department of Energy Engineering, Faculty of Mechanical Engineering, Budapest University of Technology and Economics, Műegyetem rkp. 3, H-1111 Budapest, Hungary*
[c]*Department of Mechanical and Aerospace Engineering, University of California, San Diego, CA, USA*
[d]*School of Photovoltaic and Renewable Energy Engineering, University of New South Wales, Sydney, NSW, Australia*
[e]*UK Power Networks, London, UK*
[f]*Key Laboratory for Middle Atmosphere and Global Environment Observation (LAGEO), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China*
[g]*School of Science and Industrial Technology, Harbin Institute of Technology, Harbin, Heilongjiang, China*
[h]*Department of Electrical Engineering, National University of Singapore, Singapore*
[i]*State Power Investment Corporation Limited, Harbin, Heilongjiang, China*
[j]*Faculty of Science and Technology, University of the Faroe Islands, Tórshavn, Faroe Islands*
[k]*Solar Energy Research Institute of Singapore, National University of Singapore, Singapore*
[l]*China Meteorological Administration, Beijing, China*

Abstract

Current solar forecast verification processes place much attention on performance comparison of a group of completing methods. However, forecast verification ought to further answer how the best method within the group performs relative to the best-possible performance which one can attain under that forecasting situation, which makes the quantification of predictability and forecast skill immediately relevant. Unfortunately, the literature on the quantification of relative performance of solar irradiance has hitherto been lacking, and no study has ever reported the spatial distributions of predictability and forecast skill of solar irradiance. The predictability and forecast skill of an atmospheric process depend on two concepts: (1) the growth of initial error in unresolved scale of motion, and (2) the forecast performance of the standard of reference. Based upon this formalism, predictability and forecast skill of solar irradiance in the United States are quantified and mapped. Through this study, a couple of common misconceptions in regard to irradiance predictability are refuted, and the original formulation of skill score revived.

Highlights

- A formal discussion on predictability of solar irradiance is presented.
- A rigorous way of computing the forecast skill score is emphasized.
- Predictability of solar irradiance in the United States is estimated.
- Bounds of mean square error of ECMWF irradiance forecasts in the United States are derived.

Keywords: Predictability, Solar irradiance, Solar Forecasting, Forecast verification, NSRDB, ECMWF

Word count: 7606.

*Corresponding author. Tel.: +65 9159 0888.
Email address: yangdazhi.nus@gmail.com (Dazhi Yang)

Abbreviations: CLIPER, the optimal convex combination of climatology and persistence; CONUS, the contiguous United States; ECMWF,

Nomenclature

α	Smoothness parameter of the correlogram	$\mathcal{S}^*$	The reduced skill score
β	Shape parameter of the correlogram	ν	Nugget effect of the correlogram
$\boldsymbol{\theta}$	Correlogram parameters, $\boldsymbol{\theta} = (\nu, r, \alpha, \beta)^\top$	ρ_h	Lag-h autocorrelation of clear-sky index
κ	Clear-sky index	τ	Continuous time lag
$\mathbb{E}$	Expectation operator	A_f	Accuracy of forecasts of interest
$\mathbb{V}$	Variance operator	A_p	Accuracy of perfect forecasts
C	Correlogram	A_r	Accuracy of reference forecasts
$\mathcal{P}$	Predictability defined in this work	c	Clear-sky irradiance
$\mathcal{P}^*$	Predictability defined by Yang et al. (2021)	h	Discrete time lag
$\mathcal{P}^{**}$	Predictability defined by Anthes and Baumhefner (1984)	r	Scale parameter of the correlogram
		x	Generic variable denoting forecast
$\mathcal{S}$	The original skill score	y	Generic variable denoting observation

1. Introduction

In the field of solar energy meteorology, solar resource assessment and solar forecasting are of primary interest. Whereas the former deals with predicting the long-term (years to decades) availability of resources, which is vital for siting, sizing and bankability of solar projects, the latter finds relevance in predicting the short-term (seconds to days) fluctuations in irradiance, which is needed for operations management of solar power plants. It is forecasting with which this work is concerned.

The scientific theory on weather forecasting dates back to the dawn of the twentieth century, and has gone through a quiet revolution [1]. Whereas much more is known now than was known then, what constitutes a good forecast still attracts heated debates among weather forecasters. With an ever-increasing grid penetration level of solar energy, solar forecast verification has received unprecedented attention from not only weather forecasters, but also scientists and engineers from numerous disciplines. The biggest challenges in solar forecast verification are never about metrics and methods, instead, the ultimate question is perpetually: How to compare forecasts made for different locations and over different time periods [2, 3]? Without a proper answer to this question, anyone can, in principle, claim superiority of one forecasting method over another, since what is available in one forecasting study is rarely available to others—such superiority claims may appear to be scientifically sound but are not falsifiable, which have limited values [4].

In answering that question, two conceptions are essential, first of predictability, which suggests the extent of which one can issue meaningful forecasts, and second of skill score, which quantifies the skillfulness of a forecaster or a forecasting model with respect to its peers. Both conceptions carry a notion of normalization, in that, they place emphasis on utilizing some reference forecasts, such that the quantification is in relative terms. Nonetheless, predictability of solar irradiance is a topic which often suffers from the illusion of explanatory depth—many know the basic notion yet few could explain with technical precision [5]. Forecast skill is a familiar concept, but its original formulation and intention have hitherto been thrust into the background by the widespread uptake of an alternative form of skill score [6]. Both phenomena can be attributed to the lack of consensus on what constitutes a perfect forecast or a perfect forecasting model.

European Centre for Medium-Range Weather Forecasts; GHI, global horizontal irradiance; HRES, ECMWF's High Resolution model; MSE, mean square error; MSPEG, mean square predictability error growth; NWP, numerical weather prediction; NSRDB, National Solar Radiation Database; PEG, predictability error growth; RMSE, root mean square error; RMSPEG, root mean square predictability error growth.

One may regard a forecast as being perfect if it aligns with the materialized observation exactly. However, in such situations, the quantity or event does not warrant forecasting, since it is entirely mechanical, e.g., there is no need to forecast whether the Sun will rise tomorrow [7]. Uncertainty is an essential part of nature, and forecasting hence needs to anticipate that. One may also view a forecasting model as being perfect if it apprehends wholly the data-generating laws of nature. Notwithstanding, the limitation of induction, which suggests that no one can fully test a scientific theory over all possible scenarios [8], prevents us from identifying the data-generating laws of nature with absolute certainty. The last popular appeal of the notion of perfection is in economic terms, which is particularly relevant in energy forecasting. If a forecasting strategy can lead to decision making that meets the maximum reward expected by the forecast user, it might be called perfect. However, in a game theoretic viewpoint, where both the forecasting strategy and user expectation constantly change with those of peers, this definition is again problematic.

Indeed, there does not seem to be any universally acceptable way of defining a perfect forecast or a perfect forecasting model. But the pressing needs of reconciling forecasting practices and comparing forecasts must be attended to, insofar as the field of solar energy meteorology should advance. Combining both opinions, this work puts forward the first attempt in quantifying predictability and forecast skill of solar irradiance over large geographical regions, which is consistent with the needs of grid integration, that is, a continental- and country-wise unified framework for solar forecast generation, verification, and decision making.

The remaining part of the paper is organized as follows. Section 2 formally defines predictability, and reveals how this particular definition is related to the skill score. Data and method are presented in Section 3. In Sections 4 and 5, the predictability of solar irradiance and the forecast skill of a representative numerical weather prediction (NWP) model over the contiguous United States (CONUS) are discussed, and various maps are drawn to provide a visual aid to examining the spatial distribution of the measure of predictability and forecast skill. Section 6 compares the proposed definition of predictability with two other existing definitions, and shows their inter-relationship. In the same section, possible extensions of the proposed method are also elaborated. Conclusions follow at the end.

2. Defining predictability and forecast skill

Any attempt to quantify the predictability should proceed with definition, and there are many alternatives available in the literature [e.g., 9–11]. This work deals with the predictability of irradiance, which is an atmospheric process, hence, without loss of generality, the classic definition of Anthes and Baumhefner [12] is followed, which has the generic form:

$$\mathcal{P} = 1 - \frac{A_p}{A_r}, \tag{1}$$

where A_p and A_r are the performance (or accuracy) of perfect forecasts and that of reference forecasts, respectively. Stated differently, predictability is quantified through the relative performance of the best-possible and worst-possible forecasts. In this work, the accuracy measure, A, is taken to be negatively oriented, i.e., the smaller its value, the better. When $A_p = A_r$, $\mathcal{P} = 0$, which means that when the perfect forecasts are unable to outperform the reference ones, there is no predictability. When $A_p = 0$, $\mathcal{P} = 1$, which means that when the perfect forecasts are able to predict the exact outcome consistently, the process has full predictability. Accounting for both extremes, we arrive at the range $0 \leq \mathcal{P} \leq 1$. The mathematical form of Eq. (1) is, in the main, consistent with the common measures of predictability in the fields of statistics and economics [e.g., see 9, 13, 14].[1]

On the other hand, forecast skill is an aspect of quality, and the generic skill score, which is used to quantify skill, is formulated as:

$$\mathcal{S} = \frac{A_f - A_r}{A_p - A_r}, \tag{2}$$

where, in addition to A_p and A_r defined earlier, A_f is performance of some forecasts of interest, i.e., the forecasts from a model that needs to be verified [15–17]. When $A_f = A_r$, $\mathcal{S} = 0$, which suggests that when the forecasts of interest

[1] If one explores the formulations of predictability defined within these works, each is distinct, see for instance the discussion of Yang et al. [4]. However, in form, that is, one minus the ratio of some statistics of some forecasts that can be deemed optimal in some sense, over that of some reference forecasts, all aforementioned works resemble each other.

generated by a forecaster (or a forecasting model) have the same accuracy as the reference forecasts, that forecaster (or that model) possesses no skill. Moreover, since perfect forecasts should represent the best-possible performance, improvements made by the forecasts of interest over the reference forecasts can never exceed those made by the perfect forecasts, meaning $A_r - A_f \leq A_r - A_p$, and thus $\mathcal{S} \leq 1$. It should be noted that the upper bound of $\mathcal{S}$ is 1 but its lower bound is not 0. In practice, negative $\mathcal{S}$ values can occur, which correspond to the situations when the forecasts of interest are of lower performance than the reference forecasts. These very straightforward properties constitute the theoretical basis of $\mathcal{P}$ and $\mathcal{S}$.

Mathematically, when $A_p = 0$, the skill score defined in Eq. (3) reduces to:

$$\mathcal{S}^* = 1 - \frac{A_f}{A_r}, \tag{3}$$

which is arguably a more common form of skill score used in the literature [6, 18, 19]. In the field of solar forecasting, Eq. (3) is known to fame through the series of works by Marquez and Coimbra [20, 21], Coimbra et al. [22] in the early 2010s, and has been used extensively when reporting the forecast skill of solar irradiance. The implicit assumption made in Eq. (3) is that *all* sources of error can be explained, such that it is theoretically possible to identify the Nature's model or data-generating process that leads to deterministic observations, i.e., no uncertainty. However, insofar as model is concerned, Wheatcroft [17] among others noted that all real-world models contain some degree of structural error, which echoes the famous aphorism of George Box that "*All models are wrong, but some are useful*." Particularly in the context of weather forecasting, even if the physical laws governing the atmospheric motion can be known with absolute certainty, all dynamical weather models suffer from inaccurate approximations of the physical laws and numerical approximations in solving the governing equations—perfect dynamical weather models do not exist. Insofar as observations are concerned, Lorenz [23] noted that one cannot expect in atmospheric science that the details of the small-scale features can be ever revealed on a global basis by a regular observational network. As observational uncertainty propagates into initial conditions and diagnostic analyses, no matter how small these errors may initially be, they will grow even under the perfect model. Thus, A_p can never be zero in reality, rendering this interpretation of Eq. (3) somewhat unrealistic.

When $A_p \neq 0$, $\mathcal{S}^*$ can no longer be regarded as a reduced form of $\mathcal{S}$. Yet the former can be viewed as an alternative (i.e., separate) definition of skill score.[2] Under this premise, if we are to examine the relationship between $\mathcal{S}^*$, $\mathcal{S}$, and $\mathcal{P}$ from a purely mathematical perspective, two things may be noticed. One of those, as stated by Anthes and Baumhefner [12], is that $\mathcal{P}$ is the highest possible $\mathcal{S}^*$, because $A_p \leq A_f$. Secondly, by combining Eqs. (1)–(3), we see that $\mathcal{S}^* = \mathcal{S} \times \mathcal{P}$. This implies that the sort of skill scores reported in the literature can be interpreted as the product of the forecast skill as defined in Eq. (2) and predictability as defined in Eq. (1). Murphy [15] noted that a main purpose of reporting the skill score is to allow comparison of forecasts made using different models, at different locations, and over different time periods, but in this regard, the coupling between $\mathcal{S}^*$ and $\mathcal{P}$ limits its functionality. Stated differently, even if $\mathcal{S}^*$ is computed and reported, it is *not* independent of predictability, which in itself varies in accord with the forecasting situation, and thus defies the original intention of reporting the skill score. That said, one should note that the skill score $\mathcal{S}^*$ is still relevant as a verification tool (especially when reported as one metric among other error metrics) if used to compare the performance of different forecast models produced for the same periods of time and for the same locality.

To the best of our knowledge, these important caveats on the interpretation of $\mathcal{S}^*$, or how to remedy them, has not been discussed in the field of solar forecasting. Hence, this work considers a physics-based method for estimating A_p, such that $\mathcal{S}$ and $\mathcal{P}$ can in turn be estimated. The approach for estimating A_p is based on quantifying the growth of difference between a control and a perturbed forecast. The method was proposed by Anthes and Baumhefner [12], who considered the same problem, but in the context of the synoptic-scale 500-mb height. In order to adapt their approach to solar irradiance, several modifications are made.

First and foremost, instead of using the error variance and climatology, which were used by Anthes and Baumhefner [12] as the accuracy measure and the standard of reference, respectively, we follow the recommendations for deterministic solar forecast verification [2, 3, 6, 20], and use the root mean square error (RMSE) as the accuracy measure, and

[2] There are many other alternative definitions besides $\mathcal{S}^*$ and $\mathcal{S}$, see Jolliffe and Stephenson [16] for a list of skill score definitions available in the atmospheric science literature.

the optimal convex combination of climatology and persistence (CLIPER) as the standard of reference. The advantage of of CLIPER over climatology is that the former can produce reference forecasts of much higher accuracy over short horizons [see Fig. 1 of Ref. 24]. The second modification pertains to the variable of concern. Anthes and Baumhefner [12] estimated A_p and A_r directly on the synoptic-scale 500-mb height, but for solar irradiance, these mathematical operations must be performed in clear-sky index terms, and then back-transformed into irradiance terms. This is because solar irradiance exhibits double-seasonal variation due to yearly and diurnal cycles in the relative position of the sun and the earth, which implies that the sizes of A_p and A_r would change with the amplitude of the double-seasonal pattern. Solar forecasters are accustomed to working with the clear-sky index [25, 26], so this point needs no further discussion. Thirdly, the method for estimating A_p and A_r is empirical (i.e., based on samples), which suggests that the estimated values are functions of the discrete horizon h. To allow retrieval of values of A_p and A_r for any horizon, two curves are fitted to the empirically computed A_p and A_r values, such that they become functions of τ, a random variable denoting the continuous horizon. With these modifications in mind, the data and method for estimating A_p and A_r are introduced next.

3. Data and method

The main goal of this work is to obtain predictability maps for CONUS. For that, two datasets are needed. One of those consists of satellite-derived irradiance, which is needed to compute the mean square error (MSE) of CLIPER, or A_r^2. Here, the National Solar Radiation Database (NSRDB) [27] is utilized. The other dataset is the NWP forecast irradiance, which is needed to compute the MSE of the perfect forecasts, or A_p^2. Since perfect forecasts are unattainable, A_p^2 has to be approximated as the inherent error growth of a dynamical system starting from some initial error distribution [28]. Here, the European Centre for Medium-Range Weather Forecasts' (ECMWF's) High Resolution (HRES) model[3] is selected as the dynamical system of concern. It should be noted that HRES, like all other NWP models, is not perfect, however, its control and perturbed forecasts can be used to estimate the inherent nonlinear error growth (see below). Both datasets are downloaded over a period of two years (2019–2020), at a $0.5^\circ \times 0.5^\circ$ spatial resolution and an hourly temporal resolution over the CONUS. It should be noted that ECMWF issues forecasts at four different initialization times each day; in this work, only the forecasts from the 00Z runs are used. Because both gridded datasets are already quality controlled before they are disseminated, thus, no pre-processing is required.

The main steps of obtaining the predictability and forecast skill estimates are as follows:

1. Obtain the MSE of CLIPER via Eq. (4) for a discrete set of forecast horizons $h = 1, 2, \ldots, H$.
2. Fit a parametric correlation function to the empirical lag-h correlations, and use Eq. (5) to estimate the RMSE of CLIPER at any arbitrary continuous forecast horizon τ; this is the upper bound of forecast RMSE.
3. Obtain the MSE between the control and perturbed GHI forecasts via Eqs. (6) and (7) for a discrete set of forecast horizons $h = 1, 2, \ldots, H$.
4. Fit a linear line to the MSE between the control and perturbed clear-sky index forecasts via Eq. (8), and use Eq. (9) to estimate the RMSE between the control and perturbed GHI at any arbitrary continuous forecast horizon τ; this is the lower bound of forecast RMSE.
5. Use Eqs. (1) and (2) to compute predictability and forecast skill, respectively.

While the above steps are to be elaborate shortly, it should be noted that the upper and lower bounds of RMSE are estimated separately. Moreover, one should be aware of that the RMSE computation from NSRDB or ECMWF data is only possible at discrete time lags, as multiples of data resolution (i.e., steps 1 and 3 above). Therefore, to acquire the upper and lower bounds of RMSE at arbitrary forecast horizon, fitting is needed (i.e., steps 2 and 4 above).

3.1. MSE of CLIPER

Throughout this work, y, c and κ are used to denote all-sky irradiance, clear-sky irradiance, and clear-sky index, respectively. It follows that $\kappa = y/c$. For simplicity, we do not explicitly distinguish in notation random variable

[3] HRES is the name of the ECMWF model that issues deterministic forecasts. However, it is also the model that enables ECMWF's ensemble forecasts, which are generated by the Ensemble Prediction System, which runs HRES with slightly different initial conditions.

and particular realization, but the latter is accompanied by a subscript indexing time, so which is which is implied. Expectation and variance operators are written as $\mathbb{E}$ and $\mathbb{V}$, and lag-h autocorrelation of the clear-sky index time series is expressed as ρ_h.[4] To distinguish estimated quantities from the true ones, the former is annotated with a hat symbol.

With these notational conventions, the MSE of CLIPER, or A_r^2 in the present context, following the findings of previous works [2, 3, 29], is given as:

$$A_r^2 = \mathrm{MSE}_y \approx \mathbb{E}\left(c^2\right) \cdot \mathrm{MSE}_\kappa = \left(1 - \rho_h^2\right) \mathbb{V}(\kappa) \mathbb{E}\left(c^2\right). \tag{4}$$

In this equation, $\mathrm{MSE}_\kappa = \left(1 - \rho_h^2\right) \mathbb{V}(\kappa)$ which is the MSE of CLIPER forecasts of clear-sky index. Since it is irradiance that is of our final interest, MSE_κ is scaled to MSE_y, which is the MSE of clear-sky CLIPER,[5] in irradiance terms; this is known as *MSE scaling* [29], which states that if stationarity can be assumed in clear-sky index time series, the MSE in irradiance terms is equal to the MSE in clear-sky index terms multiply by the mean square clear-sky irradiance. MSE scaling is an approximation of which the degree of validity is sufficient for the current purpose.

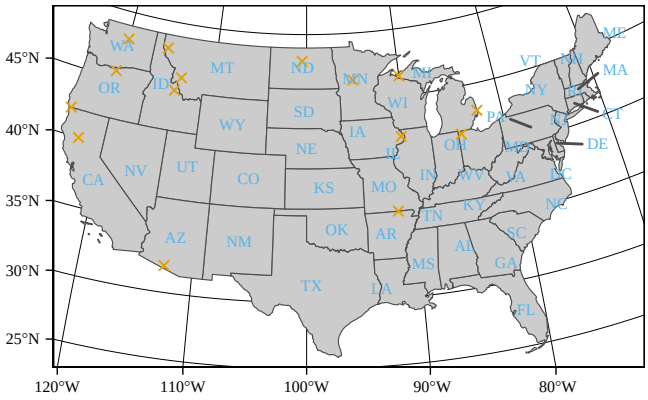

Figure 1: State abbreviations and the 15 randomly chosen locations.

Examining Eq. (4) reveals that the dependency of A_r on horizon is solely up to the autocorrelation ρ_h, which can be computed empirically from data. Notwithstanding, time series data with a certain resolution are only able to support discrete ρ_h, whereas it is more convenient if A_r can be expressed as a function of a continuous horizon τ. In this regard, a popular concept in spatial statistics called the correlation function (also known as correlogram) [30] can facilitate the estimation of $A_r(\tau)$. A correlogram is herein denoted as $C(\tau; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the unknown vector of parameters that describes the shape of the function. Fitting a parametric correlogram is a mature procedure, but requires some foundation in spatial statistics; the detailed procedure used in this work has been outlined by Yang [24] and reiterated in Appendix A.

To give perspective of the result, the empirical and fitted autocorrelations, over horizons up to 102 h, at 15 randomly chosen locations (as marked on the map in Fig. 1) are depicted in Fig. 2. The black dots and orange curves in Fig. 2 represent the empirically calculated lag-h autocorrelation $\hat{\rho}_h$ and fitted correlogram $C\left(\tau; \hat{\boldsymbol{\theta}}\right)$, for each location. Here, the Cauchy correlogram with a nugget effect is used. Stated simply, the nugget effect is where the correlogram (almost) cuts the ordinate.[6] With the estimated correlogram, A_r as a function of τ can be obtained by replacing ρ_h in

[4] All programming language offers standard computations of these statistics. However, it should be noted that variances and expectations are calculated only with the data corresponding to solar zenith angles $Z > 85°$, whereas autocorrelation should be computed with care—after the nighttime data are removed, the "holes" (or data gaps) should be replaced with NAs, as to preserve the original temporal spacing of observations.

[5] In view of the well-known terminology of "clear-sky persistence" in the field of solar forecasting, this method should be called "clear-sky CLIPER."

[6] As τ decreases, the value of correlogram approaches the size of the nugget effect, but when τ reaches 0, the value of correlogram jumps to 1, because the autocorrelation with no lag is always 1.

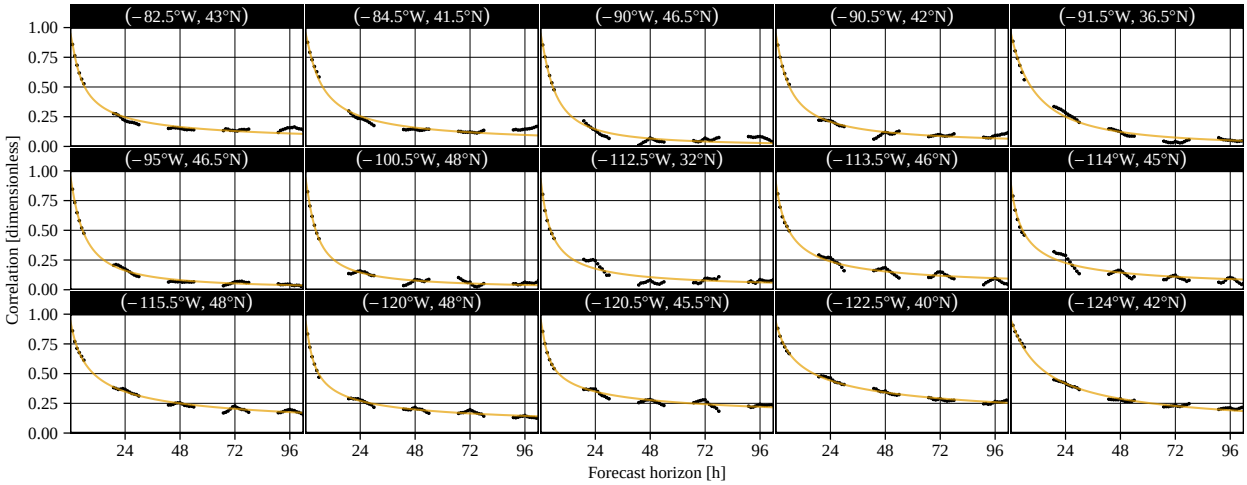

Figure 2: Empirical autocorrelations $\hat{\rho}_h$ (black dots) and fitted Cauchy correlograms $C\left(\tau;\hat{\boldsymbol{\theta}}\right)$ (orange lines) at 15 randomly chosen locations, using hourly NSRDB GHI data from 2019 to 2020. The REST2 model is used to derive the clear-sky index from GHI.

Eq. (4) by $C\left(\tau;\hat{\boldsymbol{\theta}}\right)$, which yields:

$$\hat{A}_r(\tau) = \left\{\left[1 - C^2\left(\tau;\hat{\boldsymbol{\theta}}\right)\right]\mathbb{V}(\kappa)\mathbb{E}\left(c^2\right)\right\}^{\frac{1}{2}}. \tag{5}$$

Figure 3 shows the $\hat{A}_r(\tau)$ estimated at the 15 locations of Fig. 2. Additionally, sample RMSEs of clear-sky CLIPER are marked as dots—they are computed by substituting $\hat{\rho}_h$ into Eq. (4). The number near the y-intercept in each subplot denotes the size of nugget effect in RMSE terms. The estimated $\hat{A}_r(\tau)$ shown in Fig. 3 represents the upper bound of RMSE at each location—if some forecasts of interest receive an RMSE that exceeds the upper bound, the forecaster is better off by using just the clear-sky CLIPER forecasts.

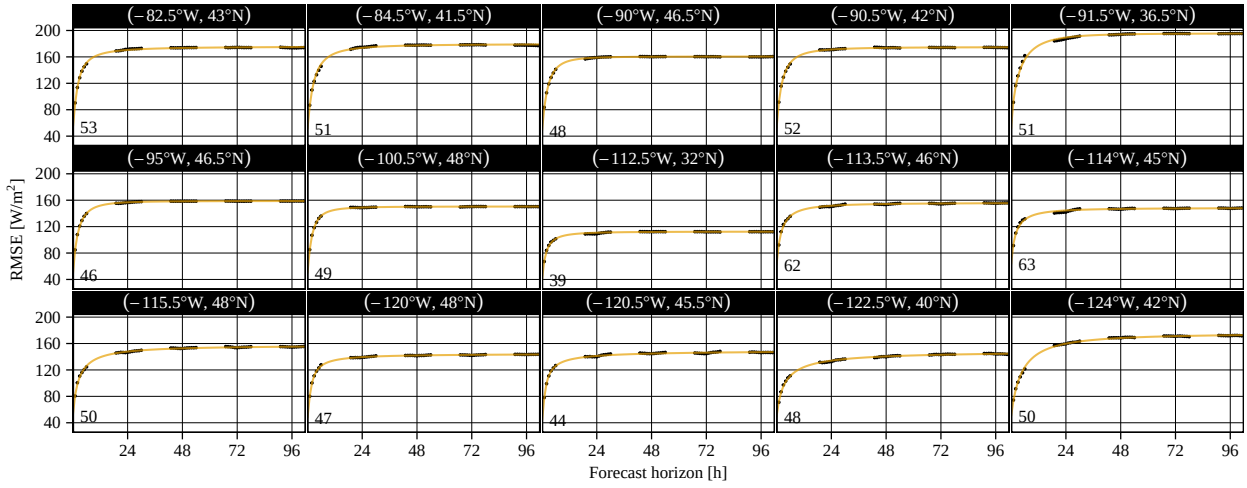

Figure 3: Sample RMSE of clear-sky CLIPER via Eq. (4) (dots) and $\hat{A}_r(\tau)$ estimated via Eq. (5) (lines) at 15 randomly chosen locations, as appeared in Fig. 2. Point (−112.5°W, 32°N) is located in Sonoran Desert, Arizona (AZ), and thus sees a much lower RMSE than the other locations, cf. Fig. 1.

3.2. *Predictability error growth*

In weather forecasting, the error growth of a dynamical system is called the *predictability error growth* (PEG)—not to be confused with predictability $\mathcal{P}$ itself. PEG describes how the initial error in the analysis, which is due to both the limited observations entering assimilation and the deficiencies in assimilation method itself, evolves with time. Conceptually, PEG is similar to forecast error. More specifically, for a forecast x_t and its corresponding observation y_t at time t, the forecast error $e_t = x_t - y_t$ is defined as the difference between the forecast and the observation. But for PEG, it is defined as the difference between the control forecast $x_t^{(c)}$ and a perturbed forecast $x_t^{(p)}$, i.e., $\mathrm{peg}_t = x_t^{(c)} - x_t^{(p)}$. On this point, the usual aggregation statistics that are applicable to e_t would also function for peg_t. For instance, given n samples of forecast error, the MSE is calculated as $n^{-1}\sum_{t=1}^{n} e_t^2$, in that, given n samples of PEG, the mean square PEG (MSPEG) can be defined and calculated analogously as $n^{-1}\sum_{t=1}^{n} \mathrm{peg}_t^2$. The original formulation of PEG of Baumhefner [28] considers only one perturbed forecast, however, most modern NWP systems, such as the ECMWF Ensemble Prediction System, issue multiple perturbed forecasts. Therefore, the formula for MSPEG can be modified naturally as:

$$\frac{1}{n}\sum_{t=1}^{n}\left(\frac{1}{m}\sum_{i=1}^{m}\mathrm{peg}_{t,i}^2\right) = \frac{1}{n}\sum_{t=1}^{n}\left[\frac{1}{m}\sum_{i=1}^{m}\left(x_t^{(c)} - x_{t,i}^{(p)}\right)^2\right], \tag{6}$$

where m denotes the number of ensemble members, and an additional subscript i is used to index these members.

The rule-of-thumb of solar forecasting [see 5, 31]—performing model training in clear-sky index terms, and back-transforming the clear-sky index forecasts to irradiance for verification—is also applicable to the estimation of MSPEG of irradiance. In this regard, denoting the MSPEG of clear-sky index as MSPEG_κ, and that of irradiance as MSPEG_y, the MSE scaling is again useful:

$$A_p^2 = \mathrm{MSPEG}_y \approx \mathbb{E}\left(c^2\right) \cdot \mathrm{MSPEG}_\kappa, \tag{7}$$

the idea of which is identical in form to the first part of Eq. (4).

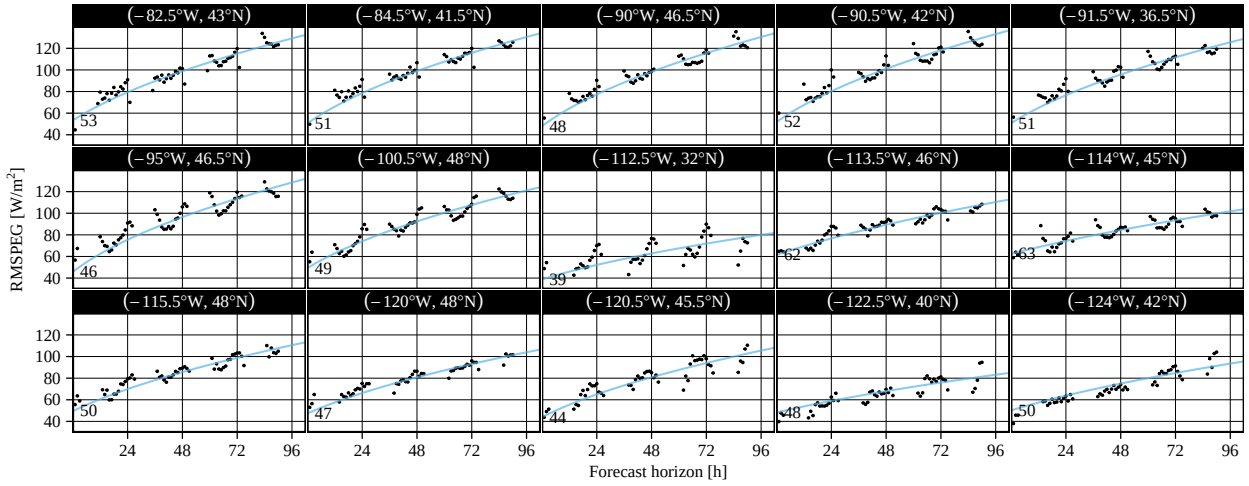

Figure 4: Sample RMSPEGs (dots) and linearly fitted RMSPEGs (lines) at 20 randomly chosen locations, using hourly ECMWF control and perturbed forecasts from 2019 to 2020.

Similar to the process used for estimating A_r, we wish to express A_p as a function of τ, such that $\mathcal{P}$ can be estimated for arbitrary forecast horizons. Based on the empirical evidence (see below), it is concluded that a linear fit of sample MSPEG against h seems adequate. In other words, the data suggests that the MSPEG of irradiance increases linearly as forecast horizon increases. A linear regression of sample MSPEG on h is parameterized by a slope and an intercept, which could be estimated simultaneously, e.g., via least squares. However, as MSPEG at $\tau = 0$ should represent the initial error size, which is preferred to be consistent with the size of the nugget effect acquired from the procedure for

estimating A_r, it can be used as the intercept of the linear regression. Then, one just needs to find the slope, which is the approach used by Yang [24] as well as in this work.[7] Denoting the nugget of the correlogram as ν, the size of the estimated nugget effect, written in the terms of MSE of clear-sky index, is $\left[1-(1-\hat{\nu})^2\right]\mathbb{V}(\kappa)$, and suppose the fitted slope is $\hat{a}$, the fitted line has the equation:

$$\mathrm{MSPEG}_\kappa(\tau;\hat{a},\hat{\nu}) = \hat{a}\tau + \left[1-(1-\hat{\nu})^2\right]\mathbb{V}(\kappa), \tag{8}$$

where $\hat{a}$ is the fitted slope. Consequently, A_p can be estimated as:

$$\hat{A}_p(\tau) = \left\{\hat{a}\tau\mathbb{E}\left(c^2\right) + \left[1-(1-\hat{\nu})^2\right]\mathbb{V}(\kappa)\mathbb{E}\left(c^2\right)\right\}^{\frac{1}{2}}. \tag{9}$$

To give perspective, the sample and fitted root mean square PEGs (RMSPEGs) for the same 20 locations are shown in Fig. 4. The numbers near the y-intercept again denote where the linear fits cut the ordinate, which, following the above discussion, are identical to those in Fig. 3. It should be clarified that the linearly fitted line exhibit some (concave) curvature, which is because MSPEG is linear to τ, and when it takes square root, its relationship with τ is no longer linear. The estimated $\hat{A}_p(\tau)$ shown in Fig. 4 represents the lower bound of RMSE at each location—theoretically, no forecasts are able to receive an RMSE that falls below the lower bound.

4. Predictability of solar irradiance in the United States

Using hourly NSRDB data and the method outlined in Section 3.1, six maps of $\hat{A}_r(\tau)$ are shown in Fig. 5, which correspond to forecast horizons $\tau = 1, 3, 6, 24, 48, 72$ h. The results are highly consistent with what have long been known by solar forecasters. For instance, the first conspicuous feature of the maps is that the western United States, particularly California (CA), Arizona (AZ), and Nevada (NV), sees much lower RMSE than the eastern United States. This is because the cloudiness of western United States is much lower than that of the eastern half—if one compares Fig. 5 with Fig. 2 of Yang et al. [32], which depicts the annual mean cloud cover over the CONUS in 2019, the correspondence between the RMSE of clear-sky CLIPER and cloudiness would be obvious at once. The second generalization which one can induce with high certainty is that the RMSE of clear-sky CLIPER, regardless of location, increases with forecast horizon, but generally becomes saturated within 24 h—comparing the maps for $\tau = 24, 48, 72$, the color differences are barely noticeable by the naked eye. The saturation is also clearly visible from Fig. 3, in that the "elbows" of the RMSE curves take place before $\tau = 24$, beyond which the growth of $\hat{A}_r(\tau)$ saturates.

Using the control and the 50-member perturbed ECMWF forecasts, as well as the method outlined in Section 3.2, six maps of $\hat{A}_p(\tau)$ are shown in Fig. 6, which correspond to the same horizons used in Fig. 5. Several observations made from these map need to be discussed. Foremost, the maps for $\tau = 1, 3, 6$ are highly similar in terms of spatial features, but more importantly, only marginal increase in RMSPEG over the first six hours is seen. This forms a clear contrast to the case of CLIPER, of which the RMSE increases substantially during the first six hours. This indicates the fact that the error growth over short horizons is small as compared to the size of error in analysis, which has long been noted by Lorenz [23]. Immediately, one has reason to believe that using additional observations that are not assimilated in global models, as in the case of dynamical downscaling with regional NWP models, has beneficial effects for short-range NWP forecasts, because better analysis reduces the initial error. The second observation which one can make is that PEG is location-dependent, e.g., PEGs in California, west Arizona, and south Nevada, are significantly lower than the other parts of the CONUS. This means that the NWP model of interest is able to discern forecasting situations. That said, there does not seem to be a strong correspondence between the spatial distribution of PEG and that of Köppen–Geiger climate classification (KGC), of which a map is shown in Fig. 7. The reason is that KGC is mostly separated by using temperature (afterwards, by precipitation) which has little to do with (cloud modulated) solar irradiance. This opens up a new issue: Whether climate-based solar forecast verification is justified? Verifying forecasts and reporting forecast performance based on climates has been a very well-estimated practice in the field of

[7] Here, the procedure of estimating the predictability proceeds from estimating A_r, which results in a nugget value that can be subsequently used during the estimation of A_p. Alternatively, if A_p is estimated first, which implies that the intercept of the linear fit needs to be first determined, and the determined value can be taken as the size of nugget effect during A_r estimation. The difference in results of these two alternative approaches is not discussed here for brevity, and should be minor anyway.

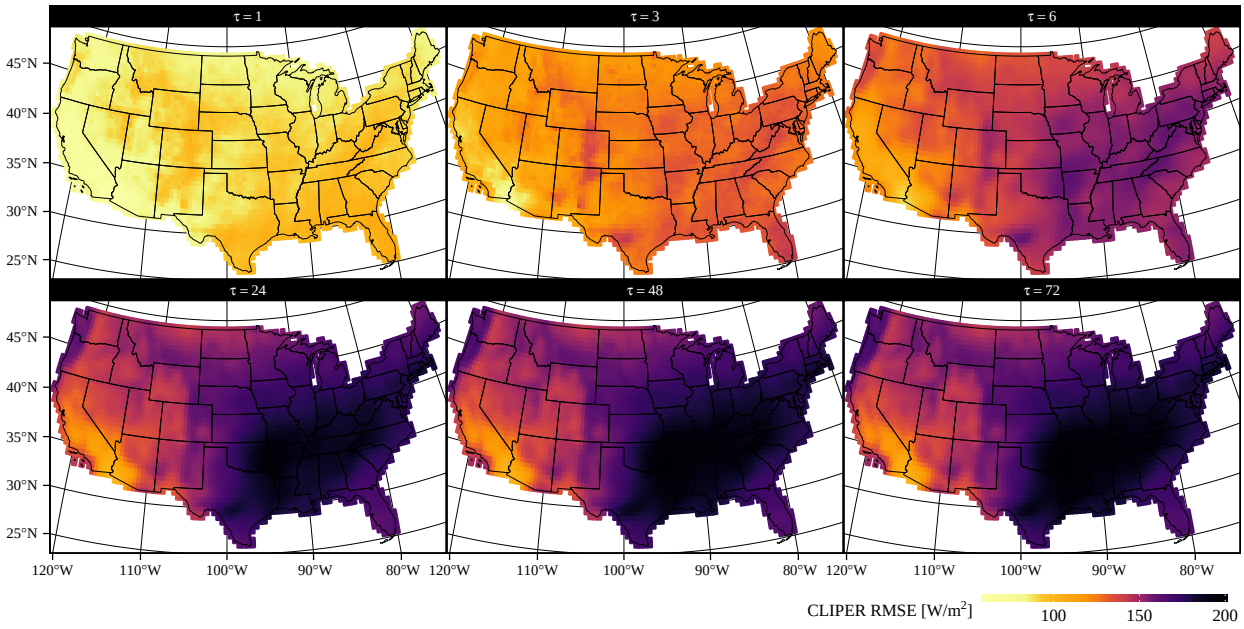

Figure 5: Maps of $\hat{A}_r(\tau)$, i.e., estimated RMSE of clear-sky CLIPER from NSRDB irradiance, over six forecast horizons, $\tau = 1, 3, 6, 24, 48, 72$.

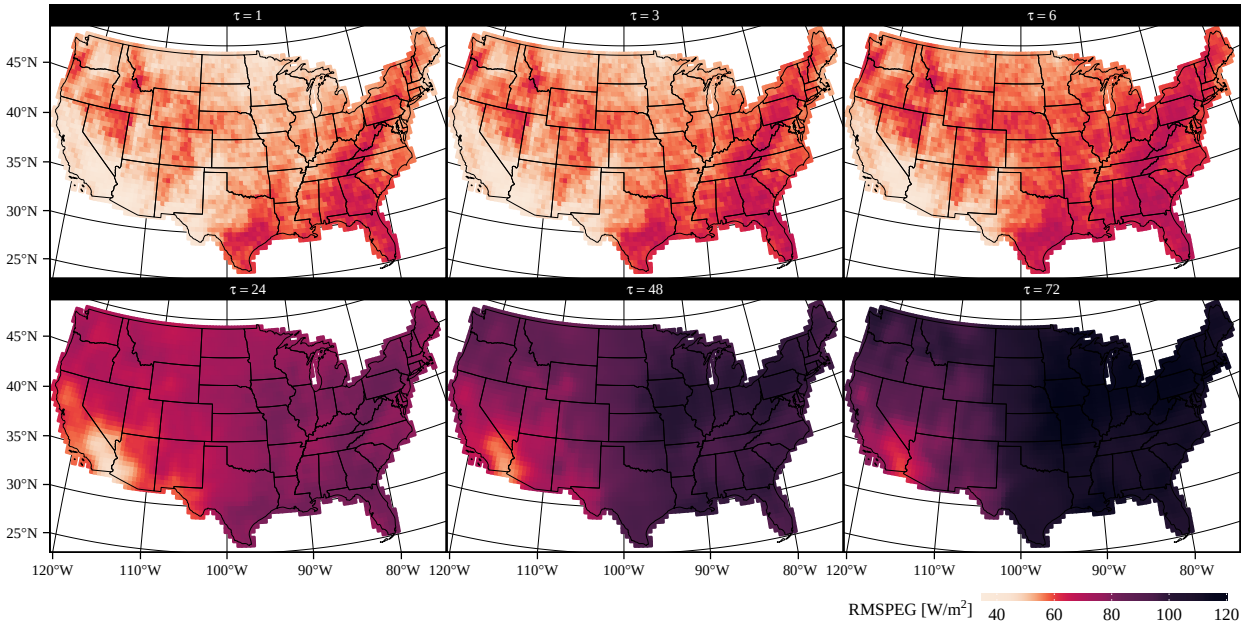

Figure 6: Maps of $\hat{A}_p(\tau)$, i.e., RMSPEG estimated from ECMWF control and perturbed irradiance forecasts, over six forecast horizons, $\tau = 1, 3, 6, 24, 48, 72$.

solar forecasting. The field seems to have believed since quite an early date that forecast performance is a function of climate class. However, besides irradiance forecasts produced for arid climates, which have lower RMSE in general, the accuracies of forecasts for all other climates are rather indistinguishable [e.g., see 33–35]. Therefore, the present results suggest, to a large extent, the divorce between climate class and solar forecast performance. (Coincidentally, it

is thought that climatology of cloud occurrence should be a dominant factor determining PEG distribution.)

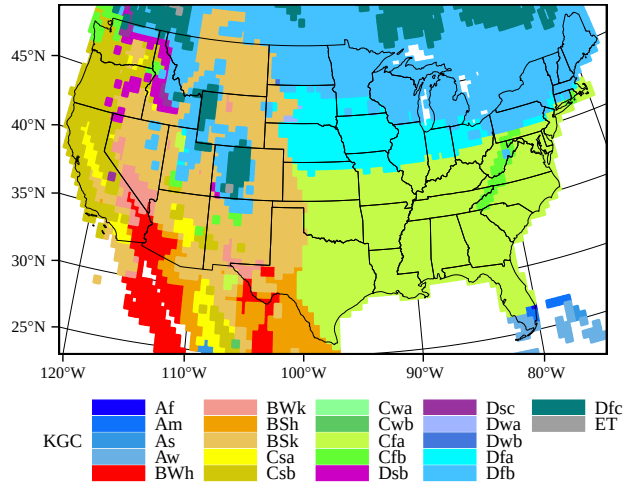

Figure 7: Köppen–Geiger climate classification over the CONUS and its surrounding regions.

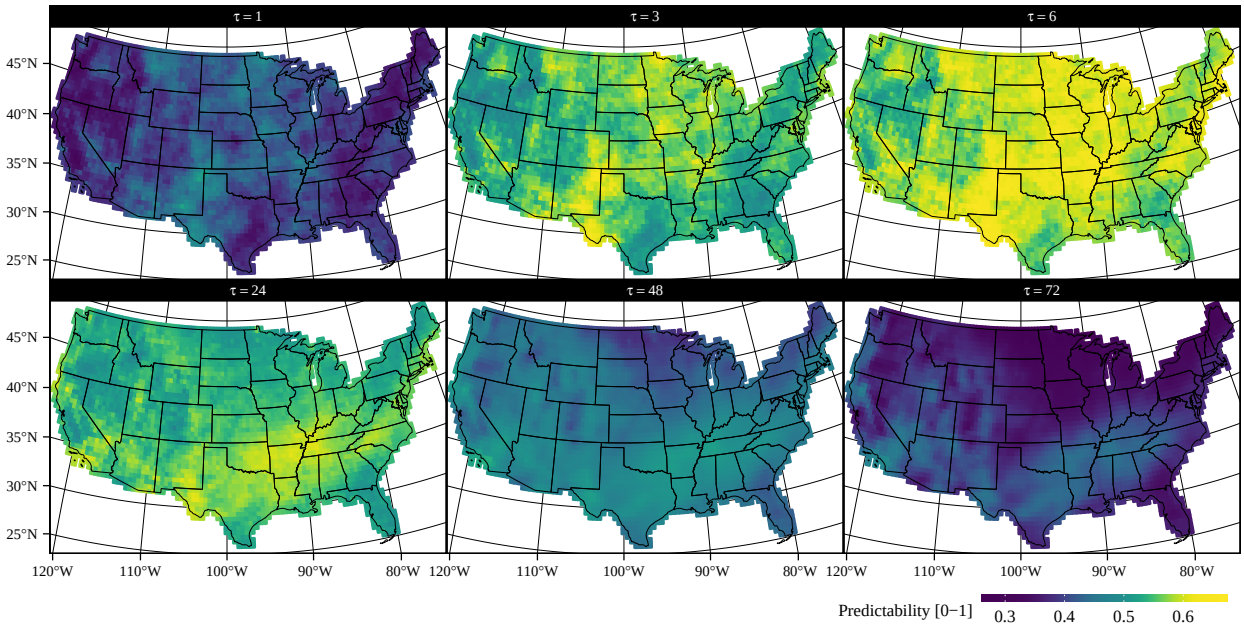

Figure 8: Maps of the estimated predictability over the CONUS for six forecast horizons, $\tau = 1, 3, 6, 24, 48, 72$.

Having obtained the maps for $\hat{A}_r(\tau)$ and $\hat{A}_p(\tau)$, the maps of predictability can be readily obtained by applying Eq. (1) for each location and for each forecast horizon. Figure 8 displays these maps, and the results should appear rather surprising to those who hold traditional suppositions. One of the traditional suppositions is that predictability should reduce as forecast horizon increases; the rationale is that longer horizons are harder to forecast, and thus have lower predictability. This supposition is refuted by the maps shown in Fig. 8, since the predictability for $\tau = 6$ is evidently higher than that for shorter and longer horizons. On this point, $\mathcal{P}$, if it is regarded as a function of τ, is clearly not monotone. More specifically, predictability neither just increase nor just decrease, but depends on the

relative rate of growth of $A_r(\tau)$ and $A_p(\tau)$. From Figs. 3 and 5, it can be concluded that the rate of growth of MSE of clear-sky CLIPER is fast over the short horizon but saturates after several hours. Therefore, if $A_p(\tau)$ grows linearly, as suggested by the present model, one should expect an increasing predictability before the saturation of MSE of clear-sky CLIPER. On the other hand, after the saturation, $A_r(\tau)$ no longer changes much—its value approaches the RMSE of climatology asymptotically—one should expect an decreasing predictability, for $A_p(\tau)$ continues to grow with respect to a saturated $A_r(\tau)$ value. Another traditional supposition which has often been made is that predictability is related to the sky condition; the logic is that clear skies are easier to forecast than cloudy ones, and thus have higher predictability. The maps shown in Fig. 8 again reject this supposition, since sunnier regions like California do not always have a higher predictability than other regions. The reason can again be understood from the relativity between $A_r(\tau)$ and $A_p(\tau)$. Sunnier regions receive smaller $A_r(\tau)$, which is certain, but at the same time, $A_p(\tau)$ may also be small, causing the ratio $A_p(\tau)/A_r(\tau)$ to be big and predictability to be low. Analogously, cloudier regions receive higher $A_r(\tau)$, which is certain, but $A_p(\tau)$ is not necessarily high, which may result in good predictability.

In conclusion, both suppositions mentioned above can be attributed to the same confusion, that is, one often associates predictability with forecast error, which is incorrect, and has been pointed out by Yang et al. [4] but through another means. Note also that the solar forecasting skill score (Eq. 3) proposed by Marquez and Coimbra [20] does not aggregate value to the score during clear sky or fully overcast periods since the RMSEs of even excellent forecasts are not smaller than the near-zero RMSEs of reference persistence forecasts in these predictable situations. In those cases, the skill score in Eq. (3) is either zero or slightly negative because $\mathcal{S}$ is near zero even if $\mathcal{P}$ may approach unity.

5. Forecast skill of ECMWF HRES solar irradiance in the United States

Having obtained $\hat{A}_r(\tau)$ and $\hat{A}_p(\tau)$, one is able to computed forecast skill according to Eq. (2), if the accuracy of some forecasts of interest $A_f(\tau)$ is also known. Hence, in this section, we should wish to analyze the skill score of ECMWF HRES forecasts over the CONUS. There is however a practical difficulty: ECMWF issues forecasts four times a day, which implies that each time stamp can only receive forecasts over a few but not all horizons. For this reason, NWP-based solar forecasts, or weather forecasts in general, are often verified over daily horizons, i.e., all forecasts within a day are verified jointly [35].

Using the 1–72-h-ahead forecasts from the 00Z runs of ECMWF HRES, the forecasts are separated into three daily blocks, namely, the 1–24-h block, the 25–48-h block, and the 49–72-h block. For each of these blocks of horizons, forecasts from consecutive days can be appended into a complete time series, with which the RMSE can be calculated with respect to the observation time series, namely, the NSRDB irradiance. This block-based verification leads to three RMSE values for each location, which are denoted as $A_f(\text{Day 1})$, $A_f(\text{Day 2})$, and $A_f(\text{Day 3})$. Echoing this aggregation approach, the $\hat{A}_r(\tau)$ and $\hat{A}_p(\tau)$ values are aggregated in the same fashion, e.g., $\hat{A}_r(\text{Day 1}) = \sum_{h=1}^{24} \hat{A}_r(h)/24$ or $\hat{A}_p(\text{Day 3}) = \sum_{h=49}^{72} \hat{A}_p(h)/24$.

Figure 9 shows the maps of forecast skill of ECMWF HRES solar irradiance, for three daily horizons. The skill scores are computed in two ways. The ones shown in the upper row of Fig. 9 follows Eq. (2), in that, the estimated A_p values are used. The second approach follows Eq. (3), which is what the current practice in the field of solar forecasting, and the results are displayed in the second row of Fig. 9. The contrast between $\mathcal{S}$ and $\mathcal{S}^*$ is strong, for all locations and for all horizons. More particularly, the relationship $\mathcal{S}^* = \mathcal{S} \times \mathcal{P}$ has been made clear in the introduction, and since $0 \leq \mathcal{P} \leq 1$, the inequality $\mathcal{S}^* \leq \mathcal{S}$ follows. In words, the skill scores reported in the current literature are lower than the actual skill scores. This shortfall of skill score is undesirable, since it could exaggerate the perceived room for improvement. Forecasts with a small $\mathcal{S}^*$ may appear to be less attractive to forecast users, but the same forecasts may be quite good in fact if they are presented in terms of $\mathcal{S}$. Figure 10 shows what we call *skill score deficit*, which is simply defined as the difference between $\mathcal{S}^*$ and $\mathcal{S}$. Spatial inhomogeneity is evident, which implies that the skill scores reported in the literature cannot be converted to the actual skill score by any easy means.

6. Discussion

Although predictability is a natural conception, its quantification must be carried out with human conventions. In that, all predictability studies proceed with definitions, which set the extent to which the analysis is valid. Since no two forecasters would agree completely on what is a perfect forecast or what is a perfect forecasting model, we should

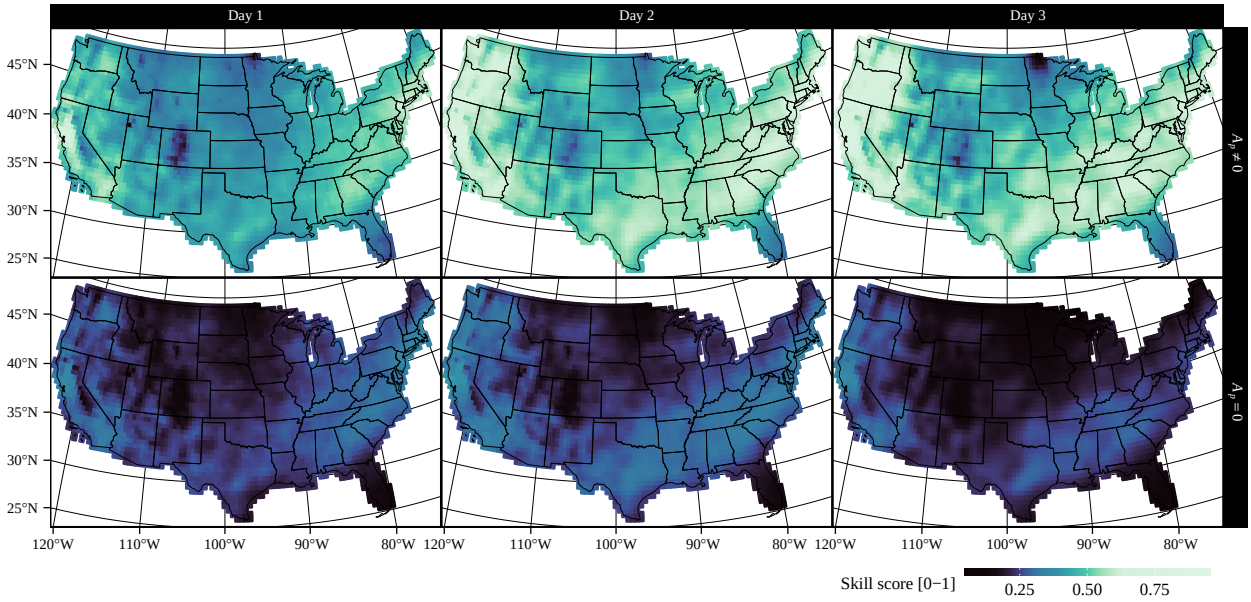

Figure 9: Maps of the forecast skill of ECMWF HRES solar irradiance, for horizons up to 3 days ahead. Two different formulas for skill score are used: (upper row) assuming $A_p \neq 0$, or $\mathcal{S} = (A_f - A_r)(A_p - A_r)$; and (lower row) assuming $A_p = 0$, or $\mathcal{S}^* = 1 - A_f/A_r$.

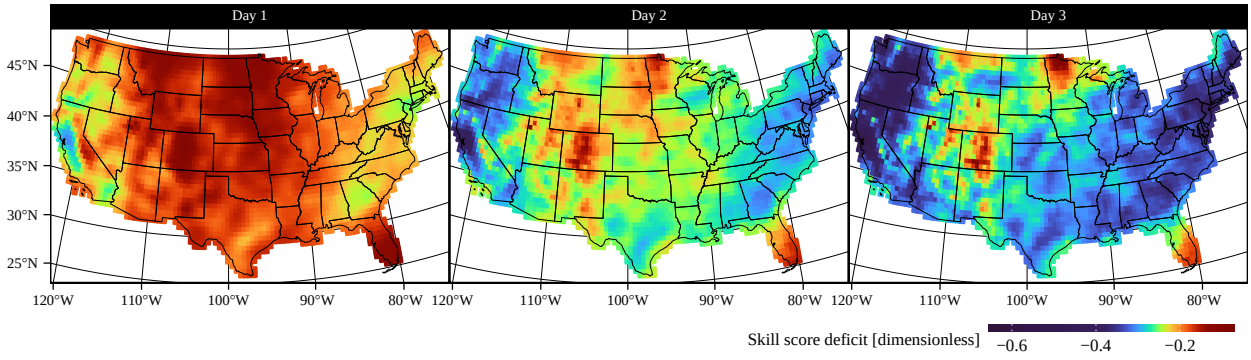

Figure 10: Maps of skill score deficit computed based on ECMWF HRES solar irradiance, for horizons up to 3 days ahead. The deficit is quantified as $\mathcal{S}^* - \mathcal{S}$. One can note the similarities in spatial features here and those in Fig. 9, which can be attributed to the interaction between that $\mathcal{S}$ and $\mathcal{P}$.

wish to first examine some alternative definitions of predictability in the literature, and how they are related to the present one. The next point of discussion is on the choice of fitting function, which could also affects the quantification of predictability. Last but not least, this section also imagines several possible extensions of the method, and set the tone for future works.

6.1. On alternative predictability definitions

Predictability, as noted by Anthes and Baumhefner [12], can be viewed as the upper limit of the forecast skill. Insofar as the mathematical expression is concerned, predictability is exactly the reduced skill score $\mathcal{S}^*$ of the perfect forecast, and as such, it is a measure of the relative performance of two forecasting methods. This notion of relative performance of two methods, first of a high-performing one and second of a baseline, gives rises to other alternative formulations of predictability, which can lead to vastly different numerical results. Yang et al. [4] calculated

predictability by taking CLIPER as the high-performing method, and climatology as the baseline, i.e.,

$$\mathcal{P}^* = 1 - \frac{\mathrm{RMSE}_{\mathrm{CLIPER}}}{\mathrm{RMSE}_{\mathrm{CLIM}}}, \tag{10}$$

and their results suggested that predictability is the highest at the shortest time horizon, and diminishes to almost zero in less than 24 hours, which clearly represents the added value of persistence over the climatology. In contrast, in the present work the error of the high-performing method is approximated by RMSPEG, a metric derived from NWP control and perturbed forecasts, and CLIPER is used as the reference, and therefore, the predictability:

$$\mathcal{P} = 1 - \frac{\mathrm{RMSPEG}_{\mathrm{NWP}}}{\mathrm{RMSE}_{\mathrm{CLIPER}}} \tag{11}$$

reflects the added value of the NWP, which is the highest on the 3–24-h time horizon.

The fundamental contradiction between the results of these two predictability quantification is evident, which must be resolved in order to make this work answer more questions than it raises. Anthes and Baumhefner [12] separated the contribution of the observations and the models to the forecast skill. The contribution of the observations can be calculated by the error reduction of the persistence compared to the climatology, which resembles the predictability formulation of Yang et al. [4]. The contribution of the model is the error reduction of the perfect forecast compared to the CLIPER, as presented in this work. To that end, both formulations of predictability are valid, but they have different meanings: the one presented by Yang et al. [4] is the predictability that can be owed to the integration of the new observations, while the one presented here is the predictability of the NWP model. The changes of these predictability scores over the time horizon are also in line with Fig. 1 of Anthes and Baumhefner [12], that depicts the contribution of the observations and models to the forecast accuracy. Finally, it is also possible to quantify predictability by combining these two approaches, i.e., using PEG as a measure of the perfect forecasts, and climatology as the reference,

$$\mathcal{P}^{**} = 1 - \frac{\mathrm{RMSPEG}_{\mathrm{NWP}}}{\mathrm{RMSE}_{\mathrm{CLIM}}}, \tag{12}$$

which would show a slowly monotonic decreasing predictability with the highest value at the time of analysis. And it is easy to show:

$$\mathcal{P}^{**} = \mathcal{P} + \mathcal{P}^* - \mathcal{P}\mathcal{P}^*. \tag{13}$$

Updating the forecasts with new observations is clearly important in practice, however, in the academia, the main purpose of research is to develop more advanced models to get as close to the perfect forecasts as possible—this is why the model-related predictability is the one presented in detail in this work. Furthermore, since $\mathcal{P}^*$ is solely dependent upon the observations, the predictability $\mathcal{P}$ advocated in this work can be converted to that of Anthes and Baumhefner [12] with ease via Eq. (13), or vice versa.

6.2. On choice of fitting functions

As noted by Dalcher and Kalnay [36], owing to the implicit assumption of a perfect model, the estimated MSPEG is only a lower bound. Correspondingly, the estimated predictability $\mathcal{P}$ is only an upper bound of the real predictability. Since this implicit assumption must be made, and all quantifications of predictability of this sort proceed from postulates by a model of perfection, one cannot but suppose this "take it or leave it" approach admissible. For more conservative forecasters, it is possible retreat to an even lower estimate of MSPEG. This can be done by fitting the infimum of the empirical MSPEG sequence. Mathematically, for a sequence x_n, the infimum of x_n is $\inf_{m \geq n} x_m$, and graphically, using the data of Fig. 4, the infimum fitting is demonstrated in Fig. 11. Analogously, for a forecaster optimistic about predictability, the supremum of the empirical MSPEG sequence may be used instead.

Besides using the infimum or supremum, other options of fitting are possible. For example, Lorenz [23] observed that the rate of change of RMSPEG for 500-mb height from ECMWF is quadratic to RMSPEG itself, which, after some algebraic derivation, suggests that RMSPEG is a hyperbolic tangent function of the forecast horizon, which is also concave for $h > 0$. Indeed, the shape of the RMSPEG curves shown in Fig. 4 have some resemblance with the curves shown in Fig. 1 of Lorenz [23]. Regardless, the choice of fitting curve should follow the empirical evidence, which necessitates exploratory analysis on data, and this rule also applies to the selection of correlogram.

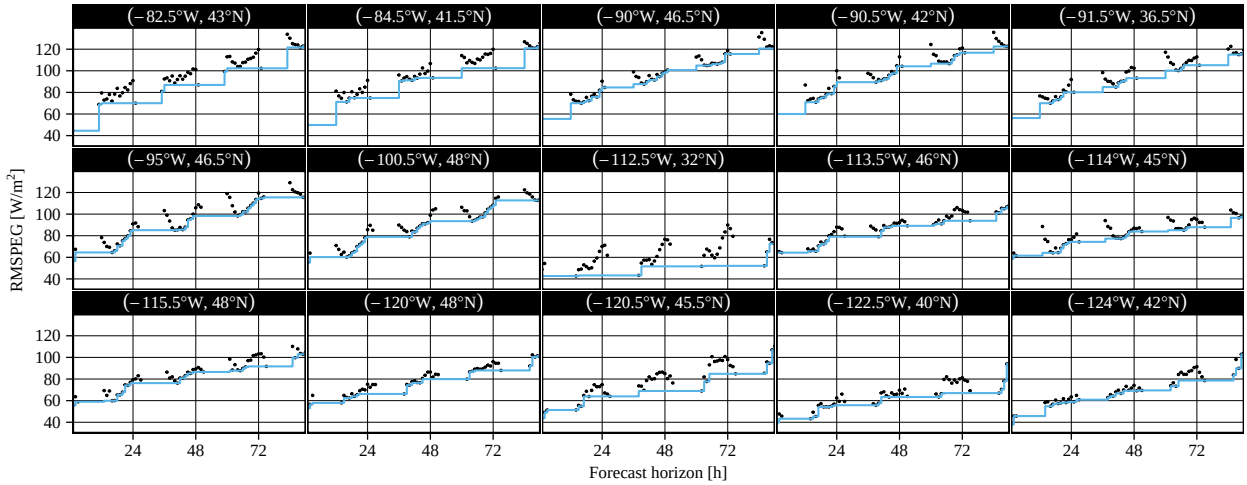

Figure 11: The infimum of the empirical RMSPEG sequence, using data of Fig. 4. This bound may be used as a more conservative estimate of the MSPEG.

6.3. Extensions of the method

The method introduced in this work is general, and thus applies to other contexts. First of all, since both NSRDB and ECMWF are available for worldwide locations (between ±60° latitudes), the exercises depicted in this work can be repeated for other regions of the world. What should be further investigated, however, is how instrumentation could affect the data quality of the retrieved irradiance and thus the estimates of A_r. Note that the low- to mid-latitude regions of the world is jointly covered by several geostationary satellites, but each uses a different radiometer/imager [37]. For instance, the Geostationary Operational Environmental Satellites (GOES) satellites, positioned to cover most of the Americas, carry the Advanced Baseline Imager (ABI), which has different spectral bands, spatial resolution, scan rate, among other specifications, from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) onboard Meteosat Second Generation (MSG) satellites over Europe, Africa and Indian Ocean—see `https://space.oscar.wmo.int/instruments` for a detailed comparison of these and other radiometers/imagers. When the Physical Solar Model [27, 38], which is the retrieval algorithm for producing NSRDB, is applied to the raw data acquired by different instruments, there must be some effects due to instrumentation, which are currently still largely unknown.

As for A_p estimation, the current method relies on dynamical ensemble NWP forecasts, which are less common as compared to deterministic (or control) NWP forecasts. In fact ECMWF only started to issue ensemble irradiance forecasts in hourly resolution since Cycle 43r1, which commenced on November 22, 2016. This explains, at least in part, why predictability study has hitherto been lacking in the field of solar forecasting. Whereas one may expect dynamical ensemble NWP forecasts to become more and more popular moving into the future, it is still desirable if PEG can be estimated solely from deterministic forecasts, such that the A_p estimates from several alternative NWP models can be compared. The method outlined by Lorenz [23] offers such a possibility. Instead of computing PEG from control and perturbed forecasts, Lorenz [23] proposed to estimate PEG from pairs of consecutive forecasts from the same run, for the error growth tends to amplify as the forecast horizon gets far. Since this work is not concerned with that alternative, the reader is referred to the original paper of Lorenz [23] if the detailed method is to be understood.

Lastly, the content of this work is readily transferable to other analyses of a like nature with other atmospheric variables, in particular, wind speed and temperature, which are essential in wind and load forecasting. What needs to be taken care of before applying the procedure of estimating predictability and compute skill score is however the salient features of the atmospheric variable under scrutiny. For instance, wind speed is well known to possess diurnal and seasonal nonstationarity, conditional heteroscedasticity, among other properties, which call for the identification of those nonstationary components, square root transform, and other possible treatments. Similarly, aerosol optical depth is log-normal distributed and subject to frequent retrieval failures due to clouds and high-albedo surfaces, in

that, log transform and gap fill seem almost always necessary. That said, we should leave these as exercises for the future, and to those who are more qualified and more relevant to study the predictability and forecast skill of those aforementioned atmospheric variables.

7. Conclusion

This paper studies, for the first time, the predictability of solar irradiance over the contiguous United States. Based upon the well-known meteorology theory of estimating predictability, several modifications have been made to adapt the theory to the salient features of irradiance. The current proposal takes predictability as one minus the ratio of the root mean square predictability error growth (RMSPEG) and root mean square error (RMSE) of the optimal combination of climatology and persistence (CLIPER), of which the former quantifies how fast the initial errors in the analysis grow, and the latter provides a reference on how naïve forecasts performance varies with the forecast situation of concern. The quantification of RMSPEG has a universal appeal, for one just needs to have a set of control and a set of perturbed forecasts. On the other hand, the RMSE of CLIPER can be computed with several statistics (mean, variance, autocorrelation) from satellite-derived irradiance. Additional to these, two curves are fitted to empirical RMSPEG and RMSE of CLIPER, in that, if this experiment is to be repeated using ECMWF and the other satellite-derived irradiance databases, predictability can be estimated for any location on earth and for any (non-discrete) forecast horizon.

The estimated predictability maps provide new insights to solar forecast verification. Firstly, the results concretely refute two common suppositions in the literature: (1) longer horizons are with lower predictability, and (2) clearer skies correspond to higher predictability. Both suppositions confuse the notion of predictability with forecast error. In reality, predictability should be quantified in relation to the performance of a set of perfect forecasts and that of a reference set, see Eq. (1). The present results show that predictability, as a function of horizon, is not monotone, and as a function of geographical location, has little to do with the climate and sky conditions.

Secondly, the estimated RMSPEG and RMSE of CLIPER can revive the skill score to its original formulation, which requires not only the accuracy of forecasts of interest and that of reference forecasts, but also the accuracy of perfect forecasts, which is commonly approximated by RMSPEG in the weather community, see Eq. (2). In solar forecasting, the reduced skill score, i.e., one minus the ratio of the accuracy of forecasts of interest and that of reference forecasts, has survived and flourished, owing to the hardship of materializing the perfect forecasts. But the reduced skill score is herein shown to be less valuable as a tool to compare forecasts made for different periods and regions. This is because the forecast skill computed that way is not independent of predictability, which defies the intention of reporting the forecast skill if not when comparing forecasting models for the same conditions (time and place). Of course, reporting additional error metrics (e.g., RMSE, MAE, MBE) for both reference and model forecasts in addition to the reduced skill score help alleviate some of these deficiencies when attempting to use such metrics for comparisons of forecasts obtained across different periods and/or locations. Nonetheless, the limitations of these metrics should be clearly understood for those using them.

The limitation of the present work resides in the fact that only predictability of irradiance is studied, which does not necessarily translate to predictability of solar power. Irradiance-to-power conversion must not be regarded as a trivial task, for the procedure requires not only meteorological knowledge, but also system specification and design. Particularly relevant are systems with a "memory" (i.e., systems containing storage elements), in that the accuracy of the accumulated forecast errors over consecutive time steps gains importance. On the other hand, the present discussion has only been focusing on the predictability of individual forecast horizons, which is unable to cater for optimizing systems that do not response instantly in time. Consequently, this calls for further studies on the temporal correlation of the forecast errors spanning over several time steps.

Acknowledgment

European Centre for Medium-range Weather Forecast (ECMWF) operational forecast and analysis data used in this study were downloaded from the ECMWF Meteorological Archival and Retrieval System (MARS). Access to the ECMWF archived data was provided by ECMWF's Data Services. Emma Pidduck, Ruth Coughlan, and Ilaria Parodi from the ECMWF's Data Services team are warmly thanked for their swift communication regarding data access.

Appendix A. Correlogram fitting procedure

For given time series $\{x_t : t = 1, \ldots, n\}$ and $\{c_t : t = 1, \ldots, n\}$, both $\mathbb{V}(\kappa)$ and $\mathbb{E}\left(c^2\right)$ in Eq. (4) are constant, and thus MSE_y is solely a function of ρ_h, which can be computed empirically from $\{\kappa_t : t = 1, \ldots, n\}$; the empirical lag-$h$ autocorrelation is denoted as $\hat{\rho}_h$. As mentioned in the introduction, we wish A_r to be attainable for any continuous horizon τ, a correlation function (also known as correlogram) is fitted using $\hat{\rho}_h$ and discrete h. There are many parametric correlograms available, but the Cauchy correlogram has been previously shown to be effective and flexible enough to capture the temporal correlation in irradiance-related quantities [39, 40]. For this reason, the Cauchy correlogram with a nugget effect is herein used, which has the form:

$$C(\tau; \boldsymbol{\theta}) = (1 - \nu) \cdot \left[1 + \left(\frac{\tau}{r}\right)^{\alpha}\right]^{-\beta} + \nu \boldsymbol{I}_{\tau=0}, \tag{A.1}$$

where r is the scale parameter, α is the smoothness parameter, β constitutes the shape parameter, ν represents the size of the nugget effect, and $\boldsymbol{I}_{\tau=0}$ is an indicator function which takes the value of 1 when $\tau = 0$, 0 otherwise. The statistics preliminary in regard to correlation functions can be found in abundance, but reading Gneiting [41, 42] is sufficient for the present content.

Unknown vector of parameters $\boldsymbol{\theta} = \{\nu, r, \alpha, \beta\}$ is estimated in two stages, as to acknowledge that some parameters may be deemed more important than others with respect to the features of the phenomenon being modeled [43]. Considering the current application, the nugget effect ν is evidently more important than the other parameters, for it can be regarded, in a statistical sense, as the irreducible initial error in the temporal process due to measurement uncertainty, or in a physical sense, effects on analysis due to unresolved microscale atmospheric processes (e.g., local flows and turbulent eddies) and insufficient observations. Therefore, following the suggestion of Montero et al. [43], the nugget effect is first estimated by extrapolating the first few empirical autocorrelations until they cut the ordinate. With the estimated nugget effect, the remaining parameters are estimated via a weighted least squares (WLS) estimator that is similar to the one proposed by Cressie and Huang [44]:

$$\underset{r,\alpha,\beta}{\operatorname{argmin}} \sum_h n_h \left[\left(\frac{\hat{\rho}_h - C(h; r, \alpha, \beta)}{1 - C(h; r, \alpha, \beta)}\right)^2\right], \tag{A.2}$$

where n_h is the number of data points used to compute $\hat{\rho}_h$. This two-stage parameter estimation procedure is fairly well known in the field of spatial statistics, but unfamiliar readers are referred to Montero et al. [43], Cressie and Huang [44] for its theoretical basis.

References

[1] P. Bauer, A. Thorpe, G. Brunet, The quiet revolution of numerical weather prediction, Nature 525 (2015) 47–55. URL: http://www.nature.com/articles/nature14956. doi:10.1038/nature14956.

[2] D. Yang, Standard of reference in operational day-ahead deterministic solar forecasting, Journal of Renewable and Sustainable Energy 11 (2019) 053702. doi:10.1063/1.5114985.

[3] D. Yang, Making reference solar forecasts with climatology, persistence, and their optimal convex combination, Solar Energy 193 (2019) 981–985. URL: http://www.sciencedirect.com/science/article/pii/S0038092X19309880. doi:10.1016/j.solener.2019.10.006.

[4] X. Yang, D. Yang, J. M. Bright, G. M. Yagli, P. Wang, On predictability of solar irradiance, Journal of Renewable and Sustainable Energy 13 (2021) 056501. doi:10.1063/5.0056918.

[5] D. Yang, W. Wang, C. A. Gueymard, T. Hong, J. Kleissl, J. Huang, M. J. Perez, R. Perez, J. M. Bright, X. Xia, D. van der Meer, I. M. Peters, A review of solar forecasting, its dependence on atmospheric sciences and implications for grid integration: Towards carbon neutrality, Renewable and Sustainable Energy Reviews 161 (2022) 112348. URL: https://www.sciencedirect.com/science/article/pii/S1364032122002593. doi:10.1016/j.rser.2022.112348.

[6] D. Yang, S. Alessandrini, J. Antonanzas, F. Antonanzas-Torres, V. Badescu, H. G. Beyer, R. Blaga, J. Boland, J. M. Bright, C. F. M. Coimbra, M. David, Â. Frimane, C. A. Gueymard, T. Hong, M. J. Kay, S. Killinger, J. Kleissl, P. Lauret, E. Lorenz, D. van der Meer, M. Paulescu, R. Perez, O. Perpiñán-Lamigueiro, I. M. Peters, G. Reikard, D. Renné, Y.-M. Saint-Drenan, Y. Shuai, R. Urraca, H. Verbois, F. Vignola, C. Voyant, J. Zhang, Verification of deterministic solar forecasts, Solar Energy 210 (2020) 20–37. URL: http://www.sciencedirect.com/science/article/pii/S0038092X20303947. doi:10.1016/j.solener.2020.04.019.

[7] J. S. Armstrong, Principles of Forecasting: A Handbook for Researchers and Practitioners, Springer, 2001.

[8] B. Russell, The Scientific Outlook, Routledge, 2017.

[9] F. X. Diebold, L. Kilian, Measuring predictability: Theory and macroeconomic applications, Journal of Applied Econometrics 16 (2001) 657–669. URL: http://www.jstor.org/stable/2678520. doi:10.2307/2678520.
[10] Z. Toth, Y. Zhu, T. Marchok, The use of ensembles to identify forecasts with small and large uncertainty, Weather and Forecasting 16 (2001) 463–477. URL: https://journals.ametsoc.org/view/journals/wefo/16/4/1520-0434_2001_016_0463_tuoeti_2_0_co_2.xml. doi:10.1175/1520-0434(2001)016<0463:TUOETI>2.0.CO;2.
[11] T. Schneider, S. M. Griffies, A conceptual framework for predictability studies, Journal of Climate 12 (1999) 3133–3155. URL: https://journals.ametsoc.org/view/journals/clim/12/10/1520-0442_1999_012_3133_acffps_2.0.co_2.xml. doi:10.1175/1520-0442(1999)012<3133:ACFFPS>2.0.CO;2.
[12] R. A. Anthes, D. P. Baumhefner, A diagram depicting forecast skill and predictability, Bulletin of the American Meteorological Society 65 (1984) 701–703. doi:10.1175/1520-0477-65.7.701.
[13] H. Kauppi, T. Virtanen, Boosting nonlinear predictability of macroeconomic time series, International Journal of Forecasting 37 (2021) 151–170. URL: https://www.sciencedirect.com/science/article/pii/S0169207020300571. doi:10.1016/j.ijforecast.2020.03.008.
[14] X. Li, V. Zakamulin, The term structure of volatility predictability, International Journal of Forecasting 36 (2020) 723–737. URL: https://www.sciencedirect.com/science/article/pii/S0169207019302341. doi:10.1016/j.ijforecast.2019.08.010.
[15] A. H. Murphy, Skill scores based on the mean square error and their relationships to the correlation coefficient, Monthly Weather Review 116 (1988) 2417–2424. doi:10.1175/1520-0493(1988)116<2417:SSBOTM>2.0.CO;2.
[16] I. T. Jolliffe, D. B. Stephenson, Forecast verification: a practitioner's guide in atmospheric science, John Wiley & Sons, 2012.
[17] E. Wheatcroft, Interpreting the skill score form of forecast performance metrics, International Journal of Forecasting 35 (2019) 573–579. URL: https://www.sciencedirect.com/science/article/pii/S0169207019300093. doi:10.1016/j.ijforecast.2018.11.010.
[18] A. H. Murphy, E. S. Epstein, Skill scores and correlation coefficients in model verification, Monthly Weather Review 117 (1989) 572–582. URL: https://journals.ametsoc.org/view/journals/mwre/117/3/1520-0493_1989_117_0572_ssacci_2_0_co_2.xml. doi:10.1175/1520-0493(1989)117<0572:SSACCI>2.0.CO;2.
[19] A. H. Murphy, Climatology, persistence, and their linear combination as standards of reference in skill scores, Weather and Forecasting 7 (1992) 692–698. doi:10.1175/1520-0434(1992)007<0692:CPATLC>2.0.CO;2.
[20] R. Marquez, C. F. M. Coimbra, A novel metric for evaluation of solar forecasting models, in: ASME 2011 5th International Conference on Energy Sustainability, ASME, 2011, pp. 1459–1467. doi:10.1115/ES2011-54519.
[21] R. Marquez, C. F. M. Coimbra, Proposed metric for evaluation of solar forecasting models, Journal of solar energy engineering 135 (2013) 011016. doi:10.1115/1.4007496.
[22] C. F. M. Coimbra, J. Kleissl, R. Marquez, Chapter 8 - Overview of solar-forecasting methods and a metric for accuracy evaluation, in: J. Kleissl (Ed.), Solar Energy Forecasting and Resource Assessment, Academic Press, Boston, 2013, pp. 171–194. URL: http://www.sciencedirect.com/science/article/pii/B9780123971777000085. doi:10.1016/B978-0-12-397177-7.00008-5.
[23] E. N. Lorenz, Atmospheric predictability experiments with a large numerical model, Tellus 34 (1982) 505–513. doi:10.1111/j.2153-3490.1982.tb01839.x.
[24] D. Yang, Correlogram, predictability error growth, and bounds of mean square error of solar irradiance forecasts, Renewable and Sustainable Energy Reviews 167 (2022) 112736. URL: https://www.sciencedirect.com/science/article/pii/S1364032122006244. doi:10.1016/j.rser.2022.112736.
[25] D. Yang, D. van der Meer, Post-processing in solar forecasting: Ten overarching thinking tools, Renewable and Sustainable Energy Reviews 140 (2021) 110735. URL: https://www.sciencedirect.com/science/article/pii/S1364032121000307. doi:10.1016/j.rser.2021.110735.
[26] D. van der Meer, D. Yang, J. Widén, J. Munkhammar, Clear-sky index space-time trajectories from probabilistic solar forecasts: Comparing promising copulas, Journal of Renewable and Sustainable Energy 12 (2020) 026102. doi:10.1063/1.5140604.
[27] M. Sengupta, Y. Xie, A. Lopez, A. Habte, G. Maclaurin, J. Shelby, The National Solar Radiation Data Base (NSRDB), Renewable and Sustainable Energy Reviews 89 (2018) 51–60. doi:https://doi.org/10.1016/j.rser.2018.03.003.
[28] D. P. Baumhefner, The relationship between present large-scale forecast skill and new estimates of predictability error growth, AIP Conference Proceedings 106 (1984) 169–180. doi:https://doi.org/10.1063/1.34270.
[29] D. Yang, Choice of clear-sky model in solar forecasting, Journal of Renewable and Sustainable Energy 12 (2020) 026101. doi:10.1063/5.0003495.
[30] N. Cressie, Statistics for Spatial Data, John Wiley & Sons, 2015.
[31] D. Yang, W. Wang, X. Xia, A concise overview on solar resource assessment and forecasting, Advances in Atmospheric Sciences 39 (2022) 1239–1251. doi:10.1007/s00376-021-1372-8.
[32] D. Yang, W. Wang, J. M. Bright, C. Voyant, G. Notton, G. Zhang, C. Lyu, Verifying operational intra-day solar forecasts from ECMWF and NOAA, Solar Energy 236 (2022) 743–755. URL: https://www.sciencedirect.com/science/article/pii/S0038092X22001645. doi:10.1016/j.solener.2022.03.004.
[33] D. Yang, R. Perez, Can we gauge forecasts using satellite-derived solar irradiance?, Journal of Renewable and Sustainable Energy 11 (2019) 023704. doi:10.1063/1.5087588.
[34] D. Yang, E. Wu, J. Kleissl, Operational solar forecasting for the real-time market, International Journal of Forecasting 35 (2019) 1499–1519. URL: https://www.sciencedirect.com/science/article/pii/S0169207019300755. doi:10.1016/j.ijforecast.2019.03.009.
[35] R. Perez, E. Lorenz, S. Pelland, M. Beauharnois, G. Van Knowe, K. Hemker, D. Heinemann, J. Remund, S. C. Müller, W. Traunmüller, G. Steinmauer, D. Pozo, J. A. Ruiz-Arias, V. Lara-Fanego, L. Ramirez-Santigosa, M. Gaston-Romero, L. M. Pomares, Comparison of numerical weather prediction solar irradiance forecasts in the US, Canada and Europe, Solar Energy 94 (2013) 305–326. URL: https://www.sciencedirect.com/science/article/pii/S0038092X13001886. doi:10.1016/j.solener.2013.05.005.
[36] A. Dalcher, E. Kalnay, Error growth and predictability in operational ECMWF forecasts, Tellus A: Dynamic Meteorology and Oceanography

39 (1987) 474–491. doi:10.3402/tellusa.v39i5.11774.
[37] D. Yang, 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 210 (2020) 3–19. URL: https://www.sciencedirect.com/science/article/pii/S0038092X20303893. doi:10.1016/j.solener.2020.04.016.
[38] Y. Xie, M. Sengupta, J. Dudhia, A Fast All-sky Radiation Model for Solar applications (FARMS): Algorithm and performance evaluation, Solar Energy 135 (2016) 435–445. URL: https://www.sciencedirect.com/science/article/pii/S0038092X16301827. doi:10.1016/j.solener.2016.06.003.
[39] D. Yang, C. Gu, Z. Dong, P. Jirutitijaroen, N. Chen, W. M. Walsh, Solar irradiance forecasting using spatial-temporal covariance structures and time-forward kriging, Renewable Energy 60 (2013) 235–245. URL: http://www.sciencedirect.com/science/article/pii/S0960148113002759. doi:10.1016/j.renene.2013.05.030.
[40] A. W. Aryaputera, D. Yang, L. Zhao, W. M. Walsh, Very short-term irradiance forecasting at unobserved locations using spatio-temporal kriging, Solar Energy 122 (2015) 1266–1278. URL: http://www.sciencedirect.com/science/article/pii/S0038092X15005745. doi:10.1016/j.solener.2015.10.023.
[41] T. Gneiting, Strictly and non-strictly positive definite functions on spheres, Bernoulli 19 (2013) 1327–1349. doi:10.3150/12-BEJSP06.
[42] T. Gneiting, Nonseparable, stationary covariance functions for space–time data, Journal of the American Statistical Association 97 (2002) 590–600. doi:10.1198/016214502760047113.
[43] J.-M. Montero, G. Fernández-Avilés, J. Mateu, Spatial and Spatio-Temporal Geostatistical Modeling and Kriging, volume 998, John Wiley & Sons, 2015.
[44] N. Cressie, H.-C. Huang, Classes of nonseparable, spatio-temporal stationary covariance functions, Journal of the American Statistical Association 94 (1999) 1330–1339. doi:10.1080/01621459.1999.10473885.