Renewable and Sustainable Energy Reviews 202 (2024) 114655

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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

Original research article

# Spatial solar forecast verification with the neighborhood method and automatic threshold segmentation

Xiaomi Zhang [a], Dazhi Yang [a,*], Hao Zhang [a,*], Bai Liu [a], Mengying Li [b], Yinghao Chu [c], Jingnan Wang [d], Xiang'ao Xia [e]

[a] School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin, Heilongjiang, China
[b] Department of Mechanical Engineering & Research Institute for Smart Energy, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region
[c] Department of Advanced Design and Systems Engineering, City University of Hong Kong, Hong Kong Special Administrative Region
[d] State Grid Corporation of China, Harbin, Heilongjiang, China
[e] Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

ARTICLE INFO

*Keywords:*
Spatial forecast verification
Solar irradiance
Neighborhood method
Satellite-derived products
Automatic threshold segmentation

ABSTRACT

Numerical weather prediction (NWP) has hitherto been the default tool for providing day-ahead forecasting services to the solar energy industry. Rapid advancements in solar forecasting using NWP call for more appropriate forecast verification procedures. Current solar forecast verification is almost always carried out through ground-based radiometric data collected at point locations. Consequently, spatial features embedded in the gridded NWP forecasts cannot be verified. This study presents the spatial verification of solar irradiance forecasts using the neighborhood method, with the main goal of emphasizing the importance of such verification procedures. By applying spatial smoothing one establishes a way to directly compare the observed and forecast fields, and concurrently, mitigate verification errors that may arise from small-scale spatial displacements. Within this framework, two variants of the neighborhood-based verification, namely, the fraction-field method and the upscaling method, are examined with respect to two reanalysis products, namely, ERA5 and MERRA-2. The results suggest that, in comparison to the upscaling method, the fraction-field method can better quantify forecast performance by providing fractions skill scores. On top of the traditional neighborhood approach, which involves the subjective selection of threshold for dichotomization, an automatic threshold segmentation method based on the three-component skew-normal mixture model is proposed to resolve the issue, which can also lead to substantial time savings in data processing. Given the spatio-temporal attributes and benefits of visualization, spatial verification is anticipated to serve as a complementary practice to the current mainstream point-location forecast verification.

## 1. Introduction

Solar power generation, as a clean and renewable energy technology, holds great potential in reducing greenhouse gas emissions and meeting energy demands [1,2]. Nevertheless, the ability to manage the variable solar power generation, so as to better align the solar generation profile with the load profile, largely depends on the forecast quality of solar irradiance [3,4]. In practical applications, accurate solar irradiance forecasting plays a crucial role in enabling decision-makers to efficiently schedule power generation, strategically set flexible resources, and optimize energy supply [5,6]. This, in turn, opens up the possibility of large-scale and high-penetration integration of photovoltaic (PV) power into the grid [7,8].

Solar irradiance forecasting refers to the process of predicting the future solar irradiance at a specific location or over an area [9]. The amount of radiation reaching the surface is primarily influenced by clouds, aerosols, surface albedo, and water vapor, through complex radiative transfer processes. All these influencing factors vary in both space and time, thus making solar irradiance forecasting a challenging task. Given the significance of solar forecasting for the utilization of solar energy, the verification and assessment of forecasts are of particular importance. In recent years, research on forecast verification has emerged as a central focus in the field of solar energy meteorology [2,10]. The verification process typically involves comparing forecasts to observations—the correspondence between forecasts and observations gauges the forecast quality, which is one aspect of the goodness of forecasts, alongside consistency and value [11]. In most cases, observations come from ground-based radiometric stations, which provide the most reliable and accurate means of acquiring information pertaining to

* Corresponding authors.
*E-mail addresses:* yangdazhi.nus@gmail.com (D. Yang), zh_hit@hit.edu.cn (H. Zhang).

https://doi.org/10.1016/j.rser.2024.114655
Received 9 December 2023; Received in revised form 14 April 2024; Accepted 10 June 2024
Available online 21 June 2024

## Nomenclature

### Abbreviations

| | |
|---|---|
| CAMS | The ECMWF's Copernicus Atmosphere Monitoring Service |
| CAMS-Rad | CAMS radiation service |
| ECMWF | European Centre for Medium-Range Weather Forecasts |
| ERA5 | Fifth-generation ECMWF reanalysis |
| ETS | Equitable threat score |
| FBS | Fractions Brier score |
| FSS | Fractions skill score |
| MBE | Mean bias error |
| MERRA-2 | Modern-Era Retrospective Analysis for Research and Applications, version 2 |
| NAM | North America Mesoscale |
| NSRDB | National Solar Radiation Data Base |
| NWP | Numerical weather prediction |
| POD | Probability of detection |
| PV | Photovoltaic |
| RMSE | Root mean square error |
| SURFRAD | Surface Radiation Budget Network |

### Variables

| | |
|---|---|
| $\#H_{\mathrm{random}}$ | The random hits |
| $\eta$ | Threshold |
| $\kappa$ | Clear-sky index |
| $\kappa_f(i,j)$ | The clear-sky index for forecast fields at grid box position $(i,j)$ |
| $\kappa_x(i,j)$ | The clear-sky index for observed fields at grid box position $(i,j)$ |
| $\langle \kappa_f \rangle_m(i,j)$ | The clear-sky index for forecast fields after smoothing at grid box position $(i,j)$ |
| $\langle \kappa_x \rangle_m(i,j)$ | The clear-sky index for observed fields after smoothing at grid box position $(i,j)$ |
| $\langle I_f \rangle_m(i,j)$ | The forecast fraction at grid box position $(i,j)$ |
| $\langle I_x \rangle_m(i,j)$ | The observed fraction at grid box position $(i,j)$ |
| $\mathcal{D}'_s$ | The set containing all neighborhoods |
| $\mathcal{D}_s$ | The total verification domain |
| $\mathrm{FSS}_{\mathrm{uniform}}$ | Useful skill score |
| $\lvert\mathcal{D}_s\rvert$ | The total number of pixels in the verification domain |
| $f_0$ | The basic rate—the proportion of the area in the observation or forecast fields where the clear-sky indexes exceed a specified threshold |
| $g(z;\Theta)$ | A common form of finite mixture density |
| $I_f(i,j)$ | The forecast binary event at grid box position $(i,j)$ |
| $I_f^{\uparrow}(i,j)$ | The forecast binary event at grid box position $(i,j)$ under the *upscaling method* |
| $I_x(i,j)$ | The observed binary event at grid box position $(i,j)$ |
| $I_x^{\uparrow}(i,j)$ | The observed binary event at grid box position $(i,j)$ under the *upscaling method* |
| $K_m$ | The convolution kernel of the $m \times m$ mean filter |
| $x,\ f$ | Generic variable denoting observations and forecasts |

### Functions

| | |
|---|---|
| $\Phi$ | The cumulative distribution function of the standard normal distribution |
| $f(z;\theta)$ | The density of the skew-normal distribution |
| $f(z;\theta_t)$ | The probability density function (PDF) of the $t$th component in the mixture |

### Parameters

| | |
|---|---|
| #CR | The correct rejections, where $I_x^{\uparrow}(i,j) = 0$ and $I_f^{\uparrow}(i,j) = 0$ |
| #FA | The false alarms, where $I_x^{\uparrow}(i,j) = 0$ and $I_f^{\uparrow}(i,j) = 1$ |
| #H | The hits, where $I_x^{\uparrow}(i,j) = 1$ and $I_f^{\uparrow}(i,j) = 1$ |
| #M | The misses, where $I_x^{\uparrow}(i,j) = 1$ and $I_f^{\uparrow}(i,j) = 0$ |
| $\theta_t$ | The parameter $\theta$ of $t$th component |
| $\theta_t^{\top}$ | The transpose of $\theta_t$ |
| $\eta_1$ | Threshold intensity—the $\kappa$ value corresponding to the intersection of the curves for overcast conditions and other conditions |
| $\eta_2$ | Threshold intensity—the $\kappa$ value corresponding to the intersection of the curves for other conditions and clear-sky conditions |
| $\lambda$ | Skewness parameter |
| $\mu$ | Mean of PDF |
| $\phi$ | The PDF of a normal distribution with mean $\mu$ and variance $\sigma^2$ |
| $\sigma^2$ | Variance of PDF |
| $\Theta$ | The parameter vector of this mixture distribution, where $\Theta = (p_1,\dots,p_n,\theta_1^{\top},\dots,\theta_n^{\top})^{\top}$ |
| $n$ | The number of components in the finite mixture model |
| $N_x$ | The number of columns in the entire verification domain |
| $N_y$ | The number of rows in the entire verification domain |
| $p_t$ | The mixture weight of the $t$th component in the mixture |
| $z$ | Random variable |

### Indexes

| | |
|---|---|
| $i,j$ | Index for a spatial grid box position on forecast or observation fields, $i = 1,\dots,N_x$, $j = 1,\dots,N_y$ |
| $k$ | Index for the row of neighborhood window with a size of $m \times m$, $k = 1,\dots,m$ |
| $l$ | Index for the column of neighborhood window with a size$m \times m$, $l = 1,\dots,m$ |
| $m$ | Neighborhood scales, $m = 1,3,5,\dots$ |
| $t$ | Index for component of the three-components skew-normal mixtures model, $t = 1,\dots,n$ |

surface irradiance. In other situations where ground-based measurements are not available, one may also gauge forecasts against the irradiance retrieved from satellite images; this point is to be revisited shortly after. Since forecast quality depends upon time scale, season, location, and sky conditions, verification is often conducted separately for each condition [12,13].

From a statistical viewpoint, surface shortwave downward solar radiation (better known as the global horizontal irradiance in solar engineering) is a spatio-temporal process that is continuous in both space and time. When such a process is sampled, by either ground-based radiometry or remote sensing, it is discretized. Consequently, virtually all irradiance datasets are in the form of time series (point-location measurements), or time series of lattice processes (gridded products). In the case of the former, the absence of spatial information can potentially result in a misalignment of the forecast and observation time series, consequently yielding misleading forecast verification results. This represents a fundamental challenge encountered in traditional point-location forecast verification. As such, the most desirable verification procedure necessitates considering both temporal and spatial information, and the corresponding verification methods and evaluation metrics need to be established.

### 1.1. Best solar forecasts are generated spatio-temporally

Echoing the need for spatial forecast verification is the fact that the best kind of solar forecast is more often than not generated with physics-based methods, which all have a spatio-temporal appeal. On this point, one may question the validity of forecasting irradiance solely based on historical point-location measurements. The answer to this question is twofold. First, the accuracy of forecasts made using data collected at a point location is limited, because such forecasting systems do not consider any spatio-temporal information. The generation, advection, diffusion, and extinction of clouds cannot be totally captured, at least not very effectively, by point-location measurements. Hence, in most, if not all, scenarios, spatio-temporal forecasting methods can easily outperform point-location forecasting methods [14,15]. Second, it has been repeatedly emphasized that forecasting in physical sciences has to differentiate itself from other forecasting domains, such as econometrics or financial forecasting, where statistical forecasting methods dominate [16]. More specifically, a solar forecaster has to know what makes solar forecasting special—the physics and chemistry of the earth's atmosphere.

It has long been known that the images from sky camera [17], satellite imagery, and numerical weather prediction (NWP) output [18,19] are the indispensable exogenous data used in modern solar forecasting, for intra-hour, intra-day, and day-ahead horizons, respectively [20]. Clearly, all of them are spatio-temporal in nature. Stated differently, they can be used to produce forecasts on a spatial lattice, i.e., the forecasts are gridded. Whereas the physics-based solar forecasting methods are rapidly gaining attention from the energy meteorology community, the methods to verify such gridded forecasts, as mentioned earlier, seem to be underdeveloped.

### 1.2. Why do we need spatial forecast verification?

Forecast verification compares a set of forecasts, $f$, to a set of observations, $x$, and thus provides some ideas about the goodness of forecasts. Typically, measures of quality, such as mean bias error (MBE), root mean square error (RMSE), or RMSE skill score, are used. Alternatively, one can verify the joint distribution of $f$ and $x$, using graphical tools and summary statistics [21]. Since the ground-based observations are made at point locations, only the forecasts in the corresponding grid cells are verified. Such traditional verification is not useful on at least two aspects: (1) the performance of the gridded forecasts is unknown at those unobserved locations, and (2) the "near-miss" in the forecasts caused by small-scale variability is not represented.

It is well known that the performance of any solar forecasting model depends on climatic and weather conditions [22]. Although one may invoke Tobler's first law of geography—near things are more related than distant things—and argue that the performance of a model over a cluster of grid boxes is similar, this needs not to be the general case. For instance, the Big Island of Hawaii is famous for its ecological diversity; it has eight of the world's thirteen climate sub-zones according to the Köppen–Geiger climate classification [23]. Moreover, the area of the Big Island is 10,432 km$^2$, which is approximately the area covered by 100 grid boxes with a 10 km by 10 km resolution. This means that the accuracies of the gridded forecasts in adjacent pixels may in fact be quite different.

Furthermore, a near-miss in a weather forecasting context means that the forecast event occurs around the neighborhood of the observed event. For example, one can say the forecast is a near-miss, if the forecast cloud field is shifted two pixels to the left, and it then overlaps with the observed cloud field nicely. Such near-misses are mostly due to the small-scale variability in the spatio-temporal processes, e.g., parallax effects from sun–cloud shadow projections. To that end, having the gridded irradiance forecasts verified spatially is of great interest. If a forecast irradiance field is spatially similar to the observed field, some measures are needed to quantify such similarities. From a mathematical perspective, spatial forecast verification allows for the detection and quantification of deviations, rotations, deformations, and other phenomena within forecast fields. These aspects cannot be captured through point-location verification. However, to the best of the authors' knowledge, there is no published work on the spatial verification of solar forecasts [2]. One must therefore turn to the field of meteorology, which has a relatively rich literature on spatial verification.

### 1.3. A review of spatial forecast verification methods

Spatial forecast verification is often exemplified through precipitation forecasts, as the dichotomous variable corresponding to "rain" or "no rain" offers the most fundamental case that can facilitate the study of meteorological fields. To address the displacement errors in the forecast precipitation field, which cannot be examined by traditional point-location verification, Gilleland et al. [24,25] reviewed and analyzed four classes of spatial forecast verification methods, namely, the neighborhood methods, scale-separation methods, feature-based methods, and field deformation methods, which give a fairly complete typology of spatial forecast verification. Among the four classes of methods, the first two essentially perform a filtering procedure to the meteorological fields, such that the small, high-frequency noise from the forecast and observed fields can be removed, exposing the latent spatial processes for verification. As for the feature-based and field deformation methods, they seek to capture the displacement of the forecast field from the observed one, so as to gauge the quality of the forecast.

The neighborhood approach is also known as the fuzzy approach, which has been thoroughly reviewed by Ebert [26]. Its general idea is to make the forecast and observed fields less sharply defined, and the approximate agreement between the two fields is what the verification method seeks to assess. Neighborhood methods differ from one another mainly in terms of the decision model, the quantity being compared, and the error metric. The term "decision model" simply defines what is a good forecast. For instance, Yates et al. [27] and Zepeda-Arce et al. [28] considered a variant of the neighborhood approach, known as upscaling, that has a decision model of "good forecasts resembles the observations when upscaled to coarser scales". In another example, Roberts and Lean [29] used the fractions skill score (FSS), which is associated with a decision model of "good forecast has a similar frequency of forecast events as the observation". As for quantity being compared, the neighborhood methods can work with the original fields, averaged fields, or fraction fields, which gives rise to some flexibility in data preprocessing (e.g., smoothing after dichotomization or dichotomization after smoothing). Various error metrics may be employed to eventually gauge the goodness of forecasts. Depending on whether the quantity being compared is binary, multi-categorical, or continuous, metrics such as the Brier score, FSS, or RMSE may be opted for.

The underlying principle of the scale-separation methods is simple: The field is decomposed into several fields, each representing the variation on a particular spatial scale. For example, in the case of precipitation, physical features on a larger spatial scale may be associated with frontal systems, whereas those on a smaller spatial scale may be associated with convective showers. Scale separation is performed with bandpass filters such as the Fourier or wavelet decomposition [30]. This philosophy of examining a spatial field in its spectral domain is commonly employed in spatial statistics, and the relevant techniques are well elaborated in the book by Cressie and Wikle [31]. Insofar as the spectral representation of a field is concerned, it allows one to discern phase and amplitude errors, which are typically not possible with traditional statistical measures [32]. Moreover, since the latent scales of a field are singled out, it is now possible to scrutinize these scales independently and observe the scale at which the no-skill-to-skill transition occurs [33]. This is well-suited if the forecast is purposed to a specific scale. Scale-separation methods use the traditional statistical measures at each scale in terms of the decision model.

Generally speaking, feature-based methods seek to capture specific features within the observed and forecast fields, discover the best feature matches across both fields, and then compare these matched features according to various criteria. Feature-based methods are also known as object-oriented methods, for the goodness of a forecast field is evaluated in terms of the objects contained in it. On this point, object identification is needed. Ebert and McBride [34] considered the so-called contiguous rain areas, which are regions enclosed by a user-specified isopleth of precipitation. With the identified features in both the observed and forecast fields, their horizontal displacement error may be computed by moving the forecast feature toward the observed feature until their alignment is maximized. One drawback of the feature-based methods is that they do not facilitate any analysis depending on the spatial scale of the quantity of interest [24].

Last but not least, field-deformation methods spatially morph the forecast field to resemble as closely as possible the observed field. The "strength" or "energy" required for the deformation therefore gauges the similarity between the two fields. This class of methods was first introduced by Hoffman et al. [35] in the mid-1990s, by which time the various tools for mathematical morphology were already developed [e.g., 36]. Given any two images, their warping relationship can be described through a set of two thin-plate splines, and the bending energy can be calculated with an elegant algebra. This philosophy of quantifying the difference between two images brings out another major tool for field-deformation methods, that is, optical flow, which is a technique used to describe the motion of a field or an object. Optical flow can be categorized into dense and sparse optical flow, depending on whether the motion vector is computed at every single pixel within the frame. Since optical flow now is a well-known method in computer vision, ample references, and software packages can be found that can facilitate forecast verification through field deformation.

In summary, spatial forecast verification aims to assess four objectives: (1) scales at which a forecast has the skill, (2) location precision, (3) the intensity of the field, and (4) the structure of the field. The first objective is easy to comprehend, whereas the location precision refers to whether the event-occurring location can be sufficiently forecast, which is more relevant to discrete weather variables. Besides, the intensity of the field is more relevant to continuous random variables. As for the structure of the field, it is synonymous with spatial features or objects, which are again more suitable for assessing the forecasts of the rain areas. Table 1 gives a summary of the pros and cons of the four classes of spatial verification methods, in terms of the four objectives.

**Table 1**
A comparison of four classes of spatial verification methods [24].

| Class | Scale | Location | Intensity | Structure |
|---|---|---|---|---|
| Neighborhood | ✓ | ✗ | ✓ | ✗ |
| Scale-separation | ✓ | ✗ | ✓ | ✗ |
| Feature-based | ✗ | ✓ | ✓ | ✓ |
| Deformation | ✗ | ✓ | ✓ | ✗ |

### *1.4. How good are the latest satellite-derived irradiance products?*

To perform spatial verification, gridded observations are needed. Whereas forecasts of some meteorological variables can be gauged with high-resolution gridded reference datasets, for example, from mosaics of radar rainfall estimates, such an instrument that collects ground-based gridded irradiance is nonexistent (perhaps the sky-camera-based [37] radiometry technique proposed by Kurtz and Kleissl [38] has some potential, but at the moment, the accuracy of such measurements are still far from being satisfactory). In this regard, remote-sensed irradiance is naturally the next-best option, and most likely the only option [39].

The solar energy meteorology community is constantly debating whether or not the latest satellite-derived irradiance products are accurate enough. One school, led by Richard Perez, believes that satellite-derived irradiance has evolved to a stage where site adaptation is no longer required [40]. The remaining ones are more conservative and still view satellite-derived irradiance as a suboptimal source of irradiance data [41]. It is believed that all of these arguments are meaningless if the context is missing from the discussion. That is, one must reference an application—how the satellite-derived data is to be used—before interpreting the quantified accuracy.

Recently, Yang and Perez [42] put forward several case studies to answer the question "Can we gauge forecasts using satellite-derived solar irradiance?" Both the highest-accuracy ground-based measurements from the Surface Radiation Budget Network (SURFRAD) and the satellite-derived irradiance from the National Solar Radiation Data Base (NSRDB) were used to compute the accuracy of forecasts made by the North America Mesoscale Forecast System (NAM), at seven locations in the contiguous United States. It was found that the RMSEs of the NAM forecasts gauged using SURFRAD and NSRDB are almost identical. Furthermore, when the NAM forecasts are bias-corrected using simple polynomial regression, satellite-based irradiance is clearly able to detect the improvement in RMSE from the raw NAM forecasts [43]. Following that work, Yang and Boland [44] presented studies that compared the effectiveness of using satellite-derived irradiance and ground-based irradiance in two other solar energy applications, namely, solar radiation separation modeling [44] and model output statistics for NWP forecasts [45]. In the former, it was shown that by involving satellite-derived irradiance, the RMSE of the best separation model, the Engerer2 model, can be reduced substantially (25% reduction on average). In the latter, research indicates that the satellite-derived irradiance is almost as effective as the ground-based measurements in terms of its NWP post-processing capability (in most cases, the difference in their percentage RMSEs is within 1%). The conclusions of these previous works lay down a solid foundation for the current discussion—if the uncertainty in the satellite-derived irradiance is much smaller than that in the gridded forecasts, it can be justified to use satellite-derived irradiance to perform spatial verification [45]. Yagli et al. [46] utilized an error decomposition framework to assess ground-based measurement data and satellite data from 15 research-oriented monitoring stations in Europe, South America, and Africa. They found that forecasts generated using bias-corrected satellite data exhibited the same level of quality as forecasts generated using ground-based measurement data. Jiménez et al. [47] similarly demonstrated that satellite-derived solar irradiance can serve as a viable substitute for high-quality ground-based measurement data.

In regard to the timeliness and availability of satellite data as to serving the task of forecast verification, which some readers may raise concerns about, many latest satellite-derived irradiance databases, such as the Japan Aerospace Exploration Agency's Himawari-8 product [48, 49], are near real-time, and that is likely to become a standard in the near future. More importantly, it should be noted that forecast verification is mainly concerned with the long-term performance of a model (in solar forecasting, a year of data is common), therefore, the delay in the satellite-derived irradiance data is not of primary concern.

### 1.5. The novelty and outline of this work

Since this is the first work on spatial verification of solar irradiance forecasts while considering factors such as cloud deformation and drift, the neighborhood approach [50] is opted without loss of generality. In short, the neighborhood verification approach is used to evaluate the resemblance between forecasts and observations within a spatio-temporal neighborhood window. This is achieved by gradually enlarging the neighborhood window and applying a smoothing process to both the observational and forecast data. The specific implementation steps for spatial forecast verification are as follows. First, the observed and forecast images are converted into binary images based on a predefined threshold—this step is known as *dichotomization*. Subsequently, the binary images are smoothed using neighborhood windows of various spatial scales, traversing from the top-left corner and moving systematically to the right-bottom corner—this step is known as *smoothing*. Finally, error measurement statistics are computed for the smoothed data. Within the realm of probabilistic forecasting of continuous random variables, the introduction of error metrics for spatial prediction verification, such as the FSS, is utilized to compare forecast and observed score fields within spatial neighborhoods. A higher FSS value signifies a greater similarity between observations and forecasts, indicating higher forecast quality. If the order of the first two steps in the preceding workflow is reversed, it results in another variant of the neighborhood verification, with a notable difference in terms of the field under verification, which is binary in this case. Utilizing this alternative workflow, traditional error metrics for binary forecasts, such as the probability of detection (POD) or equitable threat score (ETS), are introduced. Similarly, higher POD or ETS values suggest better forecasts. In what follows, the two variants of the neighborhood verification should be referred to as the *fraction-field method* and *upscaling method*, respectively.

One of the essential steps in neighborhood-based forecast verification is the determination of one or more threshold values, by which the original continuous random variable is classified into a binary or categorical variable. In the usual case, the selection of the threshold is largely based on the forecaster's belief, which can be subjective. Therefore, an automatic threshold segmentation method is proposed in this work. The theory of the segmentation method is based on the fact that the clear-sky index has a bimodal distribution, and thus can be modeled using a mixture model. More specifically, the bimodal distribution is modeled as a semi-parametric distribution with a mixture of two- or three-component (skewed) normal. This proposal is innovative and general; it can be applied to all cases insofar as a clear-sky index field is available. In this work, a three-component skew-normal mixtures model is used, which classifies the clear-sky index into three bins, each corresponding to a sky condition, namely, clear-, overcast-, and other-sky conditions with the clear-sky index being located in the middle of these two extreme conditions.

The empirical part of the work verifies and compares two reanalysis datasets, against a satellite-derived irradiance dataset, over Europe. It should be noted that reanalyses are no different from NWP forecasts, except that the former is a form of hindcasting and is typically run over a short horizon with a "frozen" model. The two reanalysis products are the National Aeronautics and Space Administration's Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2), and the European Centre for Medium-Range Weather Forecasts (ECMWF's) fifth-generation reanalysis (ERA5), whereas the satellite-derived irradiance product is the ECMWF's Copernicus Atmosphere Monitoring Service (CAMS) Radiation Service (CAMS-Rad). Various accuracy metrics are applied to assess the forecasting skills of these datasets.

The remaining part of the work is organized as follows. Section 2 introduces the mathematical principles of the neighborhood approach, along with the specific implementation steps, and introduces various error metrics, including FSS and RMSE for the faction-field method, and POD and ETS for the upscaling method. Additionally, it introduces an automatic threshold segmentation method using a three-component skew-normal mixtures model that removes much subjectivity from the verification procedure. Subsequently, Section 3 presents a brief overview of the observation and forecast datasets, alongside the instructions on how to access and utilize them. Section 4 presents the numerical results of the verification. The verification is divided into absolute verification and comparative verification [51], with the former being concerned with the performance of an individual forecasting system and the latter with several systems. Absolute verification primarily concentrates on the forecasting performance within a single forecasting system, as detailed in Sections 4.1 and 4.2. Comparative verification involves comparing the forecasting performance between two distinct systems, which is put forth in Section 4.3. In Section 4.4, the efficacy of the automatic threshold segmentation method is tested. Finally, Section 5 provides a summary and analysis of the research content, as well as the main findings of this study.

## 2. Methodology

It should be first noted that solar irradiance is a sub-grid process, in that, irradiance values in adjacent grid pixels are susceptible to the influence of small-scale spatial displacements. On this point, a proper spatial verification method should allow certain displacement in spatial neighborhoods, which is gauged by the neighborhood scale [26]. Besides allowing some displacements, smoothing can also effectively mitigate forecast errors induced by scale mismatch. It evaluates the similarity between forecasts and observations within a spatio-temporal neighborhood rather than at the grid-box scale. Ebert [50] suggested that the incorporation of observational data from within the proximate grid boxes surrounding matching observational points lends greater rationality. For that reason, neighborhood spatial smoothing is adopted in this study to mitigate the forecast errors arising from spatial displacements, thereby surmounting the positional constraints inherent in conventional forecast verification at point locations.

### 2.1. The neighborhood approach for spatial forecast verification: The fraction-field method

The terminology and notations are first established to describe the neighborhood methods within a common verification framework. The symbols $x$ and $f$ are used to denote observations and forecasts, respectively. Suppose $\kappa$, the clear-sky index, is spatially indexed by $i$ and $j$, which indicates the position of a grid box on a two-dimensional lattice, a particular observation is then denoted with $\kappa_x(i,j)$ and the corresponding forecast is $\kappa_f(i,j)$. Both $\kappa_x(i,j)$ and $\kappa_f(i,j)$, $\forall i,j \in \mathcal{D}_s$, are images, or snapshots. Since the verification usually takes place over the entire lattice, for convenience, $\kappa_x$ and $\kappa_f$ are used to denote the observed and forecast clear-sky index fields.

To facilitate verification, the images are often converted to binary images, for which an event needs to be defined [26]. The observed and forecast binary images are denoted with $I_x$ and $I_f$, respectively. In meteorological forecasting of continuous random variables, an event is typically defined by the forecast variable (e.g., rainfall or wind speed) exceeding a certain threshold. In solar irradiance forecasting, the event can be defined as the clear or cloudy state of the sky. For instance, for an arbitrary location $(i,j)$, one can define a clear-sky situation as

$\kappa(i,j) \geq 0.9$, so that $I_x(i,j)$ and $I_f(i,j)$ take the value 1 if $\kappa(i,j) \geq 0.9$, and 0 otherwise. Hence, by comparing $I_f$ to $I_x$, the forecaster can obtain information relevant to verification, such as how often the clear-sky situations can be correctly forecast. Similarly, if the capability of forecasting the overcast situations is of interest, one can define the indicator functions to be 1 if $\kappa(i,j) < 0.2$, 0 otherwise.

Both $\kappa(i,j) \geq 0.9$ and $\kappa(i,j) < 0.2$ are examples of threshold rules, i.e., the rules that define the indicator functions. On this point, many operators, such as "≥", "≤", ">", or "<", are all suitable for defining the threshold rules. Stated differently, the user can decide how the event is defined, according to his particular verification needs. For instance, if the "≥" rule is used, and is given as follows:

$$I_x(i,j) = \begin{cases} 0, & \kappa_x(i,j) < \eta; \\ 1, & \kappa_x(i,j) \geq \eta, \end{cases} \tag{1}$$

$$I_f(i,j) = \begin{cases} 0, & \kappa_f(i,j) < \eta; \\ 1, & \kappa_f(i,j) \geq \eta, \end{cases} \tag{2}$$

where $\eta$ is the threshold value. Sometimes, instead of using $\kappa_x$ and $\kappa_f$, a smoothed version of the clear-sky index field can be used. This smoothed field is known as the *fraction field*.

For each grid point in the clear-sky index field, its smoothed value is calculated based on the values of the surrounding grid points within a neighboring window with a size of $m \times m$, where $m$ is called the spatial resolution of verification. To denote a size-$m \times m$ neighborhood, the operator $\langle \cdot \rangle_m$ is used. The results are shown in Eqs. (3) and (4), where $\langle I_x \rangle_m$ denotes the observed fractional field obtained from the binary field $I_x$ in size of $m \times m$ neighborhood, and $\langle I_f \rangle_m$ denotes the forecast fractional field obtained from the binary field $I_f$ in size of $m \times m$ neighborhood [29].

$$\langle I_x \rangle_m(i,j) = \frac{1}{m^2} \sum_{k=1}^{m} \sum_{l=1}^{m} I_x \left( i+k-1-\frac{(m-1)}{2}, j+l-1-\frac{(m-1)}{2} \right), \tag{3}$$

$$\langle I_f \rangle_m(i,j) = \frac{1}{m^2} \sum_{k=1}^{m} \sum_{l=1}^{m} I_f \left( i+k-1-\frac{(m-1)}{2}, j+l-1-\frac{(m-1)}{2} \right). \tag{4}$$

In this context, $(i,j)$ denotes the desired grid box position, where $i$ ranges from 1 to $N_x$, with $N_x$ representing the number of columns in the entire verification domain, and $j$ ranges from 1 to $N_y$, with $N_y$ representing the number of rows in the whole verification domain. Since the neighborhood of a grid box should be centered on it, $m$ has to take odd integer values, that is, $m \in \{1,3,5,\dots\}$. Clearly then, if some scores such as FSS are used to evaluate forecasts, their values often depend upon the value of $m$. $k$ and $l$ represent the row and column indices, respectively, of a neighborhood window with a size-$m \times m$, both ranging from 1 to $m$.

To facilitate understanding, Fig. 1 shows a toy example of an observed binary field and the corresponding forecast binary field. The pixel values within individual grid boxes signify the occurrences of the events. The forecast data is assumed to entail a displacement to the right by one grid box with respect to the observed data. At the center point of the neighborhood window with side $m = 5$, the observed binary field value is 1 and the forecast binary field value is 0. Hence, when comparing the forecast event value and the observed event value at this point location, the forecast is incorrect. Nonetheless, when employing Eqs. (3) and (4), one may find that $\langle I_x \rangle_5 = \langle I_f \rangle_5 = 10/25$ at that point location, indicating quite a good forecast. In short, spatial neighborhood smoothing can effectively mitigate forecast errors induced by minor spatial displacements.

Since the goal is to create a smoothed field, may it be observed or forecast, one may extend the preceding calculation by invoking kernels. Stated differently, instead of averaging binary values in space, adding kernels allows a smoother transition among values in the neighborhood. Various choices of kernels can be employed for that purpose, such as the mean and Gaussian kernels. In this study, a mean filter

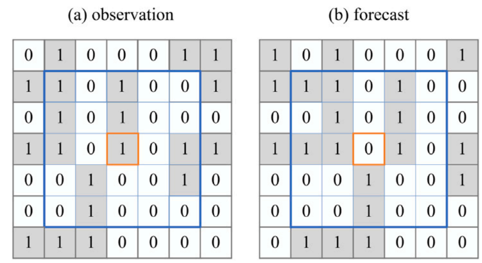


**Fig. 1.** A toy image of an observed binary field and forecast binary field at the same spatio-temporal location are presented, with blue squares representing neighborhood windows with a scale of $m = 5$.

convolution kernel is applied to the binary domain, and Eqs. (3) and (4) can be rewritten as Eqs. (5) and (6):

$$\langle I_x \rangle_m(i,j) = \sum_{k=1}^{m} \sum_{l=1}^{m} I_x \left( i+k-1-\frac{(m-1)}{2}, j+l-1-\frac{(m-1)}{2} \right) K_m(k,l), \tag{5}$$

$$\langle I_f \rangle_m(i,j) = \sum_{k=1}^{m} \sum_{l=1}^{m} I_f \left( i+k-1-\frac{(m-1)}{2}, j+l-1-\frac{(m-1)}{2} \right) K_m(k,l), \tag{6}$$

where $K_m(k,l)$ is the convolution kernel of the $m \times m$ mean filter, which can be expressed as

$$K_m = \frac{1}{m^2} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}. \tag{7}$$

When kernel-based smoothing is applied to a binary image, in a rolling manner from top left to bottom right, a smoothed image can be obtained. Fig. 2 illustrates the procedure with the same forecast field as Fig. 1(b), but with a window size of 3 to facilitate understanding. In summary, the neighborhood verification workflow consists of the following three steps:

1. Convert the clear-sky index fields to binary images according to a definition of the event, i.e., $\kappa_x \to I_x$, and $\kappa_f \to I_f$.
2. Define a series of spatial resolutions and compute the corresponding smoothed images, i.e., $I_x \to \langle I_x \rangle_m$, and $I_f \to \langle I_f \rangle_m$, $m \in \{1,3,5,\dots\}$. It is common to have multiple resolutions defined so that the forecast verification can reveal the performance over different spatial scales.
3. Compute error statistics using $\langle I_x \rangle_m$, $\langle I_f \rangle_m$.

This procedure is exemplified in Figs. 3 and 4, using data from a single time slice (at 12:00 UTC on 4 August 2016) over Europe. Fig. 3(a) shows the observed clear-sky index field from CAMS-Rad (i.e., the AGATE volume, see Section 3 for more description), as well as the forecast fields from MERRA-2 and ERA5. The continuous fields are converted to binary fields based on a decision rule of $I_{\kappa>0.5} = 1$, where pixels are set to 1 if the condition is met and 0 otherwise; the results are shown in Fig. 3(b). The deep-blue patches represent $\kappa \leq 0.5$, indicating a higher cloud cover over those patches; it can be seen that the areas of the deep-blue patches in MERRA-2 are smaller. Comparing the two binary forecast fields, the one from ERA5 resembles the observation more closely. Fig. 4 displays the results following the smoothing process. Smoothed fields at two neighborhood scales, namely, $m = 7$ and $m = 15$. As the neighborhood scale increases, the clear-sky index field is smoothed within the neighborhood window, resulting in reduced sharpness. Based on the visualization of the clear-sky index, it is subjectively evident that ERA5, when compared to MERRA-2, demonstrates a closer alignment with observation in terms of cloud coverage and clear-sky areas.

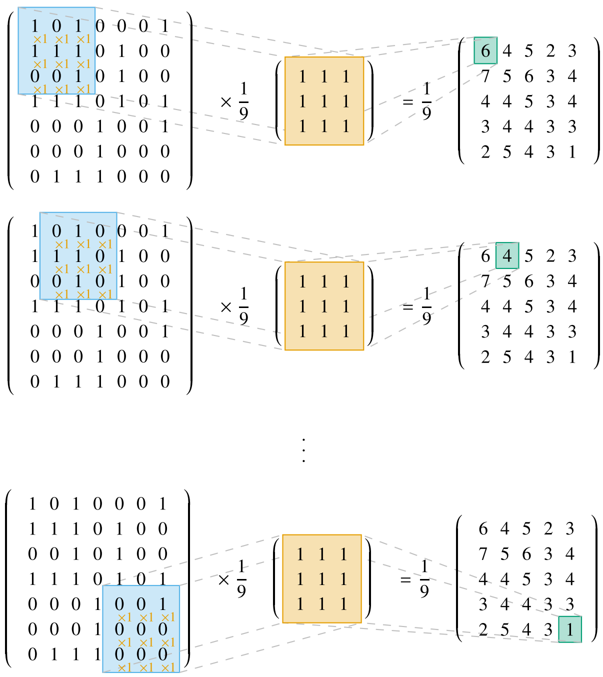

**Fig. 2.** Mean kernel-based smoothing is applied to a binary image.

Several error metrics (i.e., performance measures) can gauge the similarity between the observed and forecast images. RMSE is one of the most intuitive options, as it is a performance measure that gauges the accuracy of forecasts. It is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{|\mathcal{D}_s|} \sum_{(i,j)\in\mathcal{D}_s} \left[\langle I_f\rangle_m(i,j) - \langle I_x\rangle_m(i,j)\right]^2}, \tag{8}$$

where $\mathcal{D}_s$ is the total verification domain, and $|\mathcal{D}_s|$ is the total number of pixels in the domain, i.e., the cardinality of $\mathcal{D}_s$. Since both $\langle I_f\rangle_m(i,j)$ and $\langle I_x\rangle_m(i,j)$ are fraction fields, Ebert [50] calls the square of RMSE (i.e., mean square error), in this context, the fractions Brier score (FBS):

$$\text{FBS} = \frac{1}{|\mathcal{D}_s'|} \sum_{(i,j)\in\mathcal{D}_s'} \left[\langle I_f\rangle_m(i,j) - \langle I_x\rangle_m(i,j)\right]^2. \tag{9}$$

Eqs. (8) and (9) are almost identical. Besides the square root, another minor difference is with the domain of aggregation: RMSE is computed over the entire verification domain $\mathcal{D}_s$, whereas FBS is only computed over $\mathcal{D}_s'$, which is the set containing all neighborhoods. Stated simply, $|\mathcal{D}_s'|$ is the total number of pixels in the verification domain minus those boundary pixels without a complete neighborhood—in the case of Fig. 1, $|\mathcal{D}_s'| = 9$; in the case of Fig. 2, $|\mathcal{D}_s'| = 25$. Both RMSE and FBS are *negatively oriented* performance measures (i.e., the smaller the better).

However, neither RMSE nor FBS alone is entirely useful, as their numerical values highly depend on the frequency of the event itself. As such, Roberts and Lean [29] developed an error metric for spatial forecast verification, namely, FSS, which compares the forecast and observed fractional fields within spatial neighborhoods. FSS, being a skill score, takes the same form as the forecast skill score that is familiar to most solar forecasters, that is:

$$\text{FSS} = 1 - \frac{\text{FBS}}{\frac{1}{|\mathcal{D}_s'|}\left[\sum_{(i,j)\in\mathcal{D}_s'} \langle I_f\rangle_m^2(i,j) + \sum_{(i,j)\in\mathcal{D}_s'} \langle I_x\rangle_m^2(i,j)\right]}. \tag{10}$$

In other words, FSS is one minus the ratio of the performance of forecasts of interest and that of a set of reference forecasts. It is also worth noting that $\langle I_x\rangle_m(i,j)$ represents the observed probability that the event is true within the neighbor of $(i,j)$. Similarly, $\langle I_f\rangle_m(i,j)$ represents the forecast probability that the event is true within the neighbor of $(i,j)$. As such, FSS is often classified into the knowledge domain of probabilistic forecasting. The denominator in Eq. (10) represents the FBS of some low-performance reference forecasts. More specifically, the denominator represents the largest possible FBS obtained in the observation and forecast domains. It signifies the worst possible forecast scenario where there is no overlap at all between the forecast and the observed event.

In neighborhood verification, a forecast is said to be useful if the forecast frequency of the event is similar to the observed frequency of the event [26]. In that, the smaller the FBS value is, the larger the FSS would be, and the better the forecast is perceived to be. The range of FSS is between 0 and 1, which differs from the regular skill score that can reach negative values. This can be seen if the numerator in Eq. (10) is expanded:

$$\begin{aligned}
&\sum_{(i,j)\in\mathcal{D}_s'} \left[\langle I_f\rangle_m(i,j) - \langle I_x\rangle_m(i,j)\right]^2 \\
&= \sum_{(i,j)\in\mathcal{D}_s'} \langle I_f\rangle_m^2(i,j) + \sum_{(i,j)\in\mathcal{D}_s'} \langle I_x\rangle_m^2(i,j) - 2\sum_{(i,j)\in\mathcal{D}_s'} \langle I_f\rangle_m(i,j)\langle I_x\rangle_m(i,j) \\
&\leq \sum_{(i,j)\in\mathcal{D}_s'} \langle I_f\rangle_m^2(i,j) + \sum_{(i,j)\in\mathcal{D}_s'} \langle I_x\rangle_m^2(i,j),
\end{aligned} \tag{11}$$

because $\langle I_x\rangle_m(i,j)$ and $\langle I_f\rangle_m(i,j)$ are strictly non-negative. The magnitude of the FSS value indicates how well the forecast matches the observation, with 0 indicating a complete mismatch and 1 indicating a complete match. When FSS $= 0$, $\sum_{(i,j)\in\mathcal{D}_s'} \langle I_f\rangle_m(i,j)\langle I_x\rangle_m(i,j)$, necessitating either $\langle I_f\rangle_m(i,j)$ or $\langle I_x\rangle_m(i,j)$ being 0, $\forall (i,j) \in \mathcal{D}_s'$, i.e., a complete mismatch. When FSS $= 1$, $\sum_{(i,j)\in\mathcal{D}_s'} \left[\langle I_f\rangle_m(i,j) - \langle I_x\rangle_m(i,j)\right]^2 = 0$, necessitating $\langle I_f\rangle_m(i,j) = \langle I_x\rangle_m(i,j)$, $\forall (i,j) \in \mathcal{D}_s'$, i.e., a complete match.

Roberts and Lean [29] defined the "random skill score" and the "useful skill score". The random skill score is obtained from a random forecast and is equivalent to the ratio of the area with clear and cloudless conditions. The proportion of the area in the observation or forecast fields where the clear-sky indexes exceed a specified threshold, often referred to as the base rate, is denoted as $f_0$. At the neighborhood scale of $m = 1$, falling between perfect prediction and random prediction, the formula for the "useful skill score" is as follows:

$$\text{FSS}_{\text{uniform}} = 0.5 + f_0/2, \tag{12}$$

where the base rate $f_0$ is defined as

$$f_0 = \frac{\sum_{i=1}^{N_x}\sum_{j=1}^{N_y} I_x(i,j)}{N_x N_y}, \tag{13}$$

with $(i,j)$ denoting the desired grid box position, where $i$ ranges from 1 to $N_x$, with $N_x$ representing the number of columns in the entire verification domain, and $j$ ranges from 1 to $N_y$, with $N_y$ representing the number of rows in the whole verification domain. It is generally considered that when the FSS value exceeds $\text{FSS}_{\text{uniform}}$, the forecast contains valuable information and is regarded as skillful. Conversely, when the FSS value computed for the forecast dataset falls under $\text{FSS}_{\text{uniform}}$, it is deemed that the forecast dataset lacks valuable information or contains only a minimal amount of useful information [29,52].

### 2.2. An alternative workflow for the neighborhood approach: The upscaling method

The first two steps in the preceding verification workflow can be exchanged, that is, the original clear-sky images $\kappa_x$ and $\kappa_f$ are first smoothed, and then converted to binary images according to a preset threshold [27]. This variant of the neighborhood approach is known as the upscaling method, as smoothing is essentially a form of upscaling (i.e., making the field "blurry," therefore representing features on a larger scale).

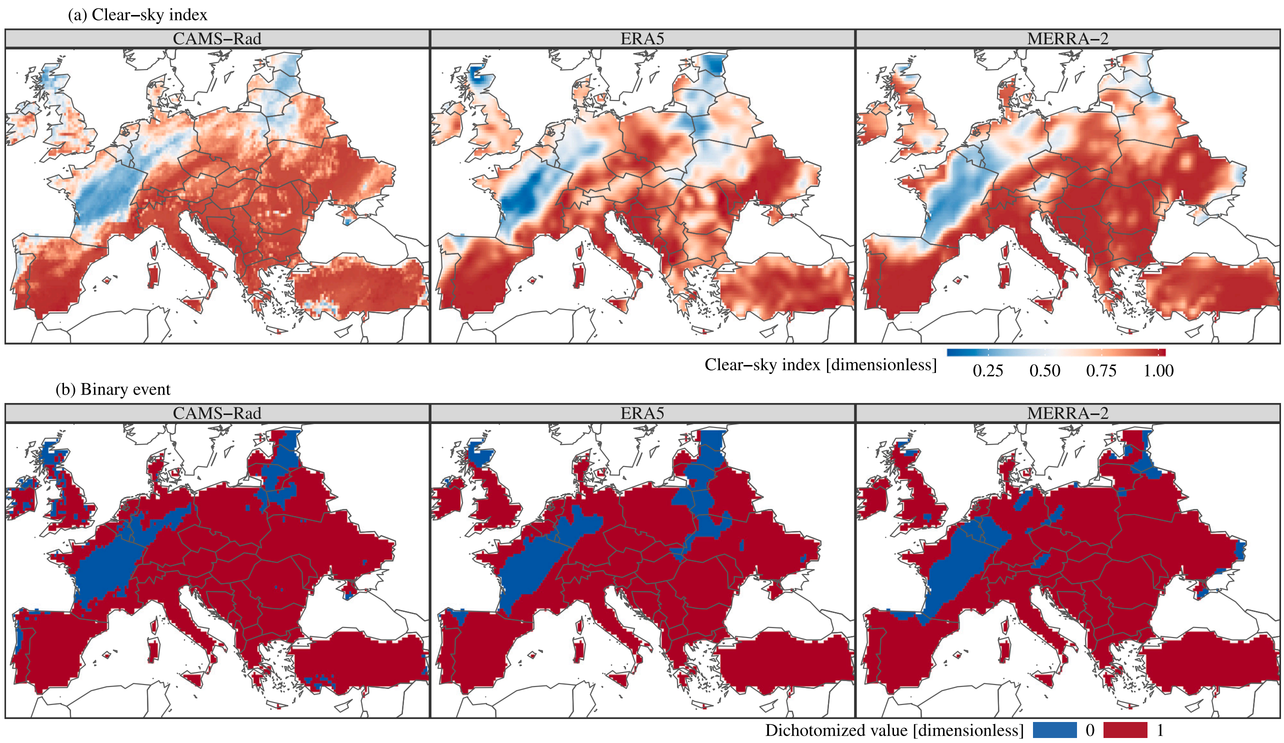


**Fig. 3.** Example of various fields involved in neighborhood approaches. Data over Europe at 12:00 UTC on 4 August 2016 from CAMS-Rad satellite-derived clear-sky index, MERRA-2 reanalysis forecasts, and ERA5 reanalysis forecasts are displayed. Subplot (a) shows the clear-sky index, $\kappa$, ranging from 0 to 1.05; (b) shows the binary event field with a threshold of 0.5 (i.e., a pixel takes the value 1, if $\kappa > 0.5$, 0 otherwise).

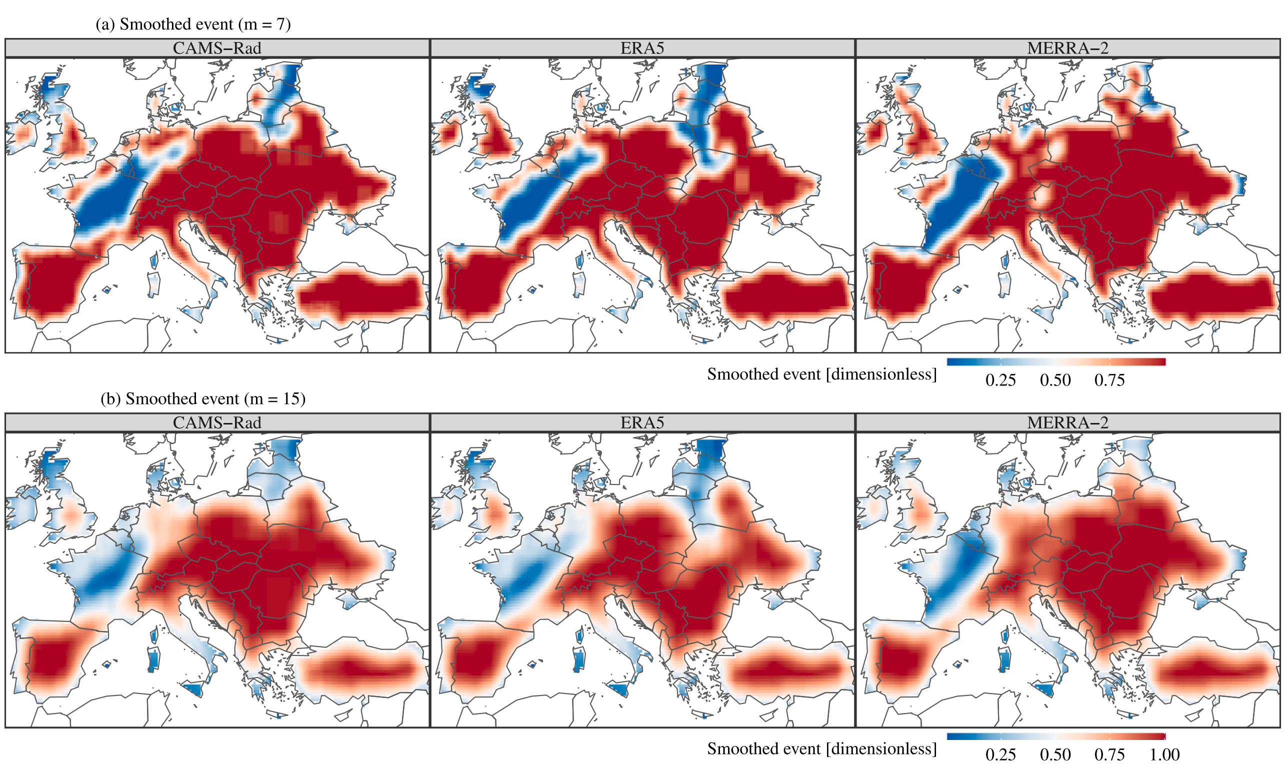


**Fig. 4.** The smoothed fields at two neighborhood scales: (a) $m = 7$ and (b) $m = 15$.

**Table 2**
Categorical contingency table defining the possible situations during forecast verification of a binary.

| Observed | Forecast | Category |
|---|---|---|
| $I_x^{\uparrow}(i,j) = 1$ | $I_f^{\uparrow}(i,j) = 1$ | Hits |
| $I_x^{\uparrow}(i,j) = 0$ | $I_f^{\uparrow}(i,j) = 1$ | False alarms |
| $I_x^{\uparrow}(i,j) = 1$ | $I_f^{\uparrow}(i,j) = 0$ | Misses |
| $I_x^{\uparrow}(i,j) = 0$ | $I_f^{\uparrow}(i,j) = 0$ | Correct rejections |

For instance, if the smoothed $\kappa$ images are denoted as $\langle\kappa_x\rangle_m$ and $\langle\kappa_f\rangle_m$, the quantities being verified are

$$I_x^{\uparrow}(i,j) = \begin{cases} 0, & \langle\kappa_x\rangle_m(i,j) < \eta; \\ 1, & \langle\kappa_x\rangle_m(i,j) \geq \eta, \end{cases} \tag{14}$$

$$I_f^{\uparrow}(i,j) = \begin{cases} 0, & \langle\kappa_f\rangle_m(i,j) < \eta; \\ 1, & \langle\kappa_f\rangle_m(i,j) \geq \eta, \end{cases} \tag{15}$$

where $I_x^{\uparrow}$ and $I_f^{\uparrow}$ are the observed and forecast binary fields, converted from the upscaled clear-sky images.

Since $I_x^{\uparrow}$ and $I_f^{\uparrow}$ are binary, the performance of the forecast can be gauged using various accuracy measures for binary variables. One can also use FSS to quantify the difference between $I_x^{\uparrow}$ and $I_f^{\uparrow}$, but traditional metrics designed specifically for binary fields are certainly more apt. Table 2 shows the categorical contingency table defining the possible situations during forecast verification of a binary event at location $(i,j)$.

Based on the number of hits (#H), false alarms (#FA), misses (#M), and correct rejections (#CR), several statistics can be computed over the verification lattice. For instance, the probability of detection (POD) is given as:

$$\text{POD} = \frac{\#\text{H}}{\#\text{H} + \#\text{M}}, \tag{16}$$

which measures the fraction of observed events that are correctly forecast. Another choice is the equitable threat score (ETS), which is given as:

$$\text{ETS} = \frac{\#\text{H} - \#\text{H}_{\text{random}}}{\#\text{H} + \#\text{M} + \#\text{FA} - \#\text{H}_{\text{random}}}, \tag{17}$$

where

$$\#\text{H}_{\text{random}} = \frac{(\#\text{H} + \#\text{M}) \times (\#\text{H} + \#\text{FA})}{\#\text{H} + \#\text{M} + \#\text{FA} + \#\text{CR}}. \tag{18}$$

ETS is a measure of the fraction of all events forecast and/or observed that are correctly diagnosed, and adjusted for those hits that could be attributed to random chance. In short, all measures that describe the relationship between the four counts (hits, false alarms, misses, and correct rejections) can be used for upscaling-based spatial verification.

### 2.3. Automatic segmentation of thresholds using three-component skew-normal mixtures model

Up to this stage, the neighborhood approach for spatial forecast verification has been introduced. However, a remaining issue is to determine the neighborhood size and threshold value, of which the choice would impact the verification. Since the neighborhood size is related *a priori* to the scale of spatial features, the $m$ that leads to the highest FSS is often interpreted as the preferred choice that can maximize the utility of the neighborhood approach. As for the threshold value, its determination can be problematic for continuous random variables, such as solar irradiance or clear-sky index. In the usual case, one has to test many different threshold values, in order to examine the impact of the choice on FSS. This process results in extensive computations, lengthy program execution times, high memory usage, and intricate subsequent data processing, and it often requires experienced forecasters to subjectively identify the optimal combination values through the visualized images. To that end, an automatic threshold selection algorithm is proposed in this study, to segment the range of the clear-sky index and convert it into a binary (or multi-categorical) variable.

Extensive research has been conducted on the statistical distributions of the clearness index and clear-sky index at various time scales [53–55]. These studies indicate that the probability density of the clearness index often exhibits a bimodal distribution, which is what the theory of segmentation is based upon. The bimodal distribution may be expressed as a weighted sum of two or more unimodal parametric distributions. The weighted sum of multiple parametric distributions of the same family is known as a finite mixture model in statistics. A common form of finite mixture density is

$$g(z;\Theta) = \sum_{t=1}^{n} p_t f(z;\theta_t), \tag{19}$$

where $p_t \geq 0$, $\sum_{t=1}^{n} p_t = 1$, for $t = 1, \ldots, n$, are the mixture weight; $f(z;\theta_t)$ is the probability density function (PDF) of the $t$th component in the mixture, parameterized by parameter $\theta_t$; $\Theta = (p_1, \ldots, p_n, \theta_1^{\top}, \ldots, \theta_n^{\top})^{\top}$; and $n$ is the number of components in the finite mixture model.

In choosing $n$ and $f(\cdot)$, different modelers hold diverse opinions, but the vast majority of the literature suggests using a two- or three-component normal mixture, of which the latter is slightly more accurate and thus more popular than the former [54,55]. Fig. 5(a) and (b) illustrate the PDFs of the $\kappa$ over Europe at 12:00 UTC on 20 August 2016 and 20 June 2016, respectively—the data is to be introduced later. From this figure, both PDFs show clear bimodality, corresponding to the clear and cloudy states of the sky, with the left mode being broader, while the right mode tends to be sharper, which agrees with the observation reported in the literature [e.g., 55]. Overlaid on the PDFs are the fitted three-component normal mixture densities (drawn as the gray dashed lines). It is evident that mixture density is able to explain the empirical density to a large extent. However, as also can be seen from the figure, the two modes of the empirical density are skewed—it exhibits positive skewness at low $\kappa$ values and negative skewness at high $\kappa$ values—which renders the normal component density somewhat inappropriate. On this point, this work proposes using the skewed-normal mixture models, which are also overlaid in Fig. 5. The advantage is immediately obvious, from just eye-balling.

The skew-normal distribution is an extension of the normal distribution that introduces a skewness parameter [56]. A random variable $Z$ follows a skew-normal distribution with location parameter $\mu$, scale parameter $\sigma^2$, and skewness parameter $\lambda$ [57]. The density formula is expressed as

$$f(z;\theta) = 2\phi(z;\mu,\sigma^2)\Phi\left(\frac{\lambda(z-\mu)}{\sigma}\right), \tag{20}$$

where $\theta = (\mu, \sigma^2, \lambda)^{\top}$; $\phi$ is the PDF of a normal distribution with mean $\mu$ and variance $\sigma^2$, whereas $\Phi$ is the cumulative distribution function of the standard normal distribution. In this study, the PDF of the $\kappa$ is fitted using the three-component skew-normal mixtures. The parameter vector of this mixture distribution, denoted as $\Theta = (p_1, p_2, p_3, \theta_1^{\top}, \theta_2^{\top}, \theta_3^{\top})^{\top}$, is estimated through maximum likelihood using an expectation–maximization algorithm [56].

As shown in Fig. 5, the three component densities from left to right represent three distinct atmospheric states: overcast conditions, other-sky conditions, and clear-sky conditions [54]. In accordance with the fitted density of the clear-sky index, the three distinct weather states can be determined based on the intersections between the component densities, which will subsequently be used as thresholds during forecast verification. The classification rule corresponding to thresholds of the clear-sky index is

$$I_x(i,j) = \begin{cases} 0, & \kappa_x(i,j) < \dfrac{\eta_1 + \eta_2}{2}; \\ 1, & \kappa_x(i,j) \geq \dfrac{\eta_1 + \eta_2}{2}, \end{cases} \tag{21}$$

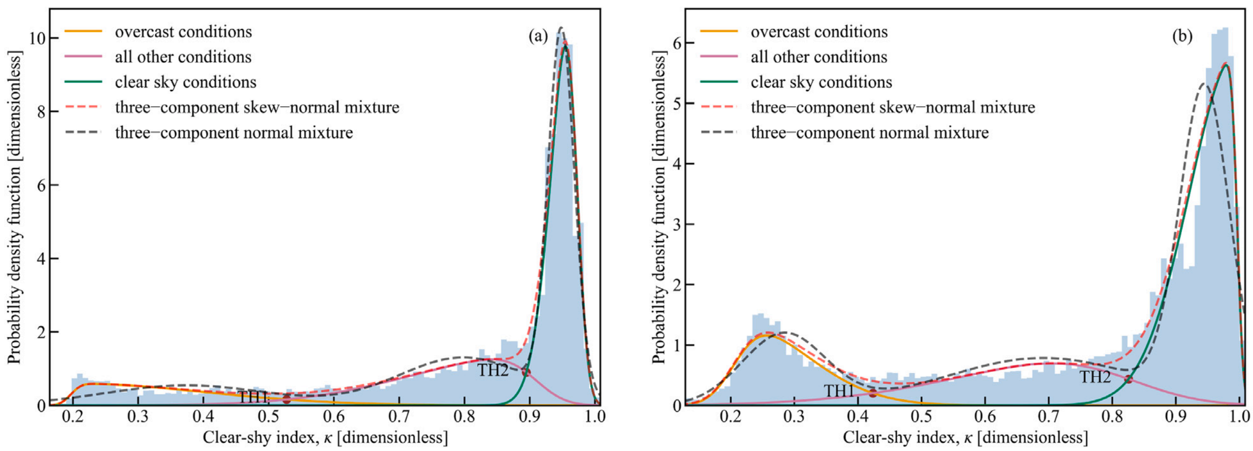


**Fig. 5.** Subplot(a) and (b) introduce the PDF of the $\kappa$ over Europe for satellite-derived observational data at 12:00 UTC on 20 August 2016 and 20 June 2016, respectively. The PDF is fitted using a three-component normal mixture (drawn as the gray dashed lines) or a three-component skew-normal mixture (drawn as the red dashed lines). The three components of a three-component skew-normal mixture model are depicted as solid lines, corresponding from left to right to the three atmospheric states: overcast conditions, other conditions, and clear-sky conditions.

where $\eta_1$ represents the $\kappa$ value corresponding to the intersection of the component PDFs for overcast conditions and other-sky conditions, and $\eta_2$ corresponds to the $\kappa$ value at the intersection of the component PDFs for other-sky conditions and clear-sky conditions. In accordance with the results discussed in Section 4.1, low thresholds tend to produce excessively high skill scores, giving rise to an unreasonable overestimation phenomenon. Conversely, excessively high thresholds often result in lower skill scores, contributing to a noticeable underestimation. Hence, in this context, the automatic threshold segmentation method continues to convert the clear-sky index range into a binary variable, with the threshold set at $(\eta_1 + \eta_2)/2$. When the $\kappa$ value is lower than $(\eta_1 + \eta_2)/2$, $I_x(i, j)$ is assigned the value of 0. When the $\kappa$ value surpasses $(\eta_1+\eta_2)/2$, $I_x(i, j)$ is assigned the value of 1. This classification rule is closely linked to sky conditions, which makes the threshold segmentation algorithm logical. One should note that the automatic threshold segmentation method does not impact the calculation of the statistics for FSS and binary variables (e.g., POD or ETS).

## 3. Data description

CAMS-Rad is a joint initiative managed by the European Commission and the European Space Agency, as part of the earth meteorological observation system [58]. CAMS-Rad provides three commonly used solar irradiance components, namely, the global horizontal irradiance, diffuse horizontal irradiance, and beam normal irradiance. The solar irradiance provided by CAMS-Rad is typically available for download in a location-by-location fashion (http://www.soda-pro.com). However, for spatial forecast verification, irradiance over wide geographical areas is needed. On this point, the Europe and Africa volumes of the data, which are named AGATE and JADE, can be downloaded as HDF5 files, from the same website, upon request. The observation data selected for this study is the AGATE volume over Europe, over a six-month period from June to December 2016. The temporal resolution of AGATE is 1 h.

The forecasts are obtained from the ERA5 and MERRA-2 reanalyses. ERA5 is the fifth-generation atmospheric reanalysis dataset by the ECMWF, covering a period from 1940 to the present [59,60]. ERA5 assimilates historical observation data and uses a legacy version of the ECMWF's High-Resolution model to make forecasts over a 12-h period in each run, so as to create a globally comprehensive atmospheric and surface meteorological dataset with high temporal and spatial resolution. It is the latest reanalysis product following the third-generation ERA-Interim dataset [59]. ERA5 features a higher spatial resolution ($0.25° \times 0.25°$) and temporal resolution (hourly), accessible through the Climate Data Store (CDS), which is available at https://cds.climate.copernicus.eu. The variable "surface downward solar radiation," abbreviated as SSRD, provides the global horizontal irradiance in units of $J/m^2$. To convert the SSRD values in $W/m^2$, simply divide them by 3600.

MERRA-2 introduces the assimilation of aerosol information into reanalysis, for the first time, to improve the simulation of meteorological data changes [61,62]. It is a reanalysis product developed by the Global Modeling and Assimilation Office of the National Aeronautics and Space Administration using the Goddard Earth Observing System Model. This dataset covers a long-term time series from 1980 to the present day, and provides global coverage with a spatial resolution of $0.5° \times 0.625°$ and a temporal resolution of 1 h (https://disc.gsfc.nasa.gov). Additionally, Bright et al. [63] made a Python library for downloading MERRA-2 data. The metadata of the three datasets is provided in Table 3.

## 4. Result and discussion

The dimensionality and complexity of the spatial forecast verification exercises involved in this work are rather high. More specifically, verification is to be conducted through the two variants of the neighborhood methods (recall Sections 2.1 and 2.2), which further contain several accuracy measures. For each variant, different choices of the neighboring window $m$ and threshold $\eta$ should be considered. Since an automatic threshold segmentation method has been proposed, its efficacy has to be gauged with respect to the traditional ad hoc threshold determination method. Another important mission is to conclude whether spatial forecast verification holds advantages over point forecast verification; this also needs to be discussed. Additionally, it is noted that spatial forecast verification is commonly conducted for a single time instance, in that, the procedure has to be repeated as many times as the number of time instances in the verification dataset, which typically spans months if not years. Last but not least, two reanalysis products are to be examined and compared in this work, which again contributes to the dimensionality of verification. It is clearly not efficient to tabulate and display all results. Hence, a verification workflow is carefully designed, as depicted in Fig. 6, to facilitate all discussions that need to be made here.

Most generally, forecast verification can be divided into two kinds, namely, absolute verification and comparative verification [51]. The former focuses on evaluating the forecast performance of a single forecasting system, whereas the latter involves the comparison of performance among two or more forecasting systems, generated under either identical or different conditions [51]. Since absolute verification

**Table 3**
Basic information of grid irradiance products.

| Dataset | Domain | Spatial resolution (lat ×lon) | Temporal resolution | Spatial range (N, E, S, W) |
|---|---|---|---|---|
| CAMS-Rad | $275 \times 126$ pixels | $0.25° \times 0.25°$ | 1 h | $60°, 44.6°, 35°, -10.2°$ |
| ERA5 | $275 \times 126$ pixels | $0.25° \times 0.25°$ | 1 h | $60°, 44.6°, 35°, -10.2°$ |
| MERRA-2 | $219 \times 100$ pixels | $0.50° \times 0.625°$ | 1 h | $60°, 44.5°, 35°, -10.0°$ |

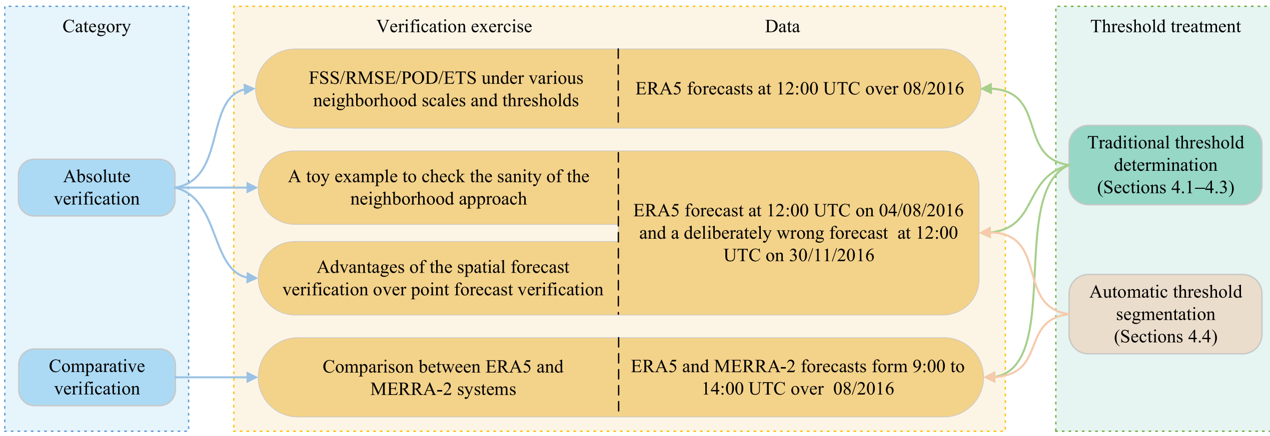


**Fig. 6.** A forecast verification workflow for the neighborhood approach.

is concerned with only a single forecasting system, ERA5 is used without loss of generality. A total of three verification exercises are devised for absolute verification. First, utilizing the ERA5 forecasts for 12:00 UTC timestamps in August 2016 (i.e., a total of 31 instances), the effects of neighborhood scale and threshold selection on performance measures are inquired in Section 4.1. Stated more precisely, FSS, RMSE, POD, and ETS are computed under a combination of different neighborhood scales and threshold values. Subsequently, in Section 4.2, a toy example is presented to examine the neighborhood method's ability to discern good forecasts from bad ones. For this purpose, the CAMS-Rad observation for 12:00 UTC on 4 August 2016, is used and the corresponding ERA5 forecast is taken as the "good" forecast, whereas the "bad" forecast is deliberately selected from another day (12:00 UTC on 30 November 2016). On this point, if the neighborhood method is rational, it should yield a much higher score/performance for the "good" forecast than the "bad" one. Using the same two forecasts, the advantage of spatial forecast verification over point forecast verification is investigated. Through this exercise, it is demonstrated that point forecast verification may assign a high score even if the underlying forecast is wrong, whereas the spatial verification method is not liable for such deficiencies. The subsequent stage entails comparative verification. In Section 4.3, a comparative analysis utilizing the FSS method is carried out to evaluate the performance of two forecasting systems, namely, ERA5 and MERRA-2, in predicting solar irradiance. ERA5 and MERRA-2 forecast datasets come from all 9:00 to 14:00 UTC timestamps in August 2016 (i.e., a total of 186 instances), over Europe. Through comparative verification, it is concluded that the overall forecast performance of solar irradiance in the ERA5 system surpasses that of the MERRA-2 system. Thus far, all exercises concern just traditional spatial forecast verification. In Section 4.1, the efficacy of the proposed automatic threshold segmentation is demonstrated.

### *4.1. Effects of neighborhood scale and threshold on performance measures*

The quantitative results of spatial forecast verification based on the neighborhood approach depend on the neighborhood scales and threshold values. To investigate the effect of neighborhood scales and threshold values on verification, the neighborhood scales are set to vary from $m = 1$ to $m = 15$, whereas the threshold values are selected to range from 0.5 to 0.9. Using ERA5 forecasts from 12:00 UTC on days in August 2016 (i.e., a total of 31 instances) and the corresponding CAMS-Rad observations over Europe, the FSSs and RMSEs are computed with respect to the fraction-field method, whereas the PODs and ETSs are computed with respect to the upscaling method. This results in four tables, which are plotted in the neighborhood-scale-threshold diagrams as shown in Fig. 7. Each entry in this plot denotes the average value of the performance metric over the 31 instances.

Fig. 7(a) depicts the FSS values of ERA5 forecasts under the fraction-field method. The FSS values are enhanced for visibility by overlaying them with varying shades of colors, where brighter colors indicate better forecasting skills—recall that a higher FSS indicates better forecasts, whereas a score of 1 represents perfect forecasts, and a score of 0 signifies no forecasting skill. As the neighborhood scale increases, FSS increases, indicating an improvement in the forecasting skill. This is anticipated because, with a larger neighborhood window, it is more probable for the forecast fraction to resemble that of the observation. On the contrary, when adopting a higher threshold, the FSS value decreases, indicating a decrease in the apparent forecasting performance. (The word "apparent" is used because the forecasting performance realized here depends on the selected threshold and is less related to the intrinsic performance of forecasts.) In neighborhood-based verification, one usually selects the threshold that gives the highest FSS as the final choice—in this case, the choice of $\eta = 0.5$ would be selected. This is however problematic, for the value 0.5 is unable to physically dichotomize the sky condition into clear and cloudy, cf. TH2 values from Fig. 5; this echoes the need for automatic threshold segmentation. Fig. 7(b) depicts the RMSE values when different neighborhood scales and threshold values are used. As the neighborhood scale increases, the decrease in RMSE indicates that enlarging the neighborhood window reduces the discrepancies between observations and forecasts, thereby improving forecasting skills. The RMSE is lowest when $m = 15$ and $\eta = 0.5$. This observation aligns with the performance of FSS in forecast verification.

Fig. 7(c) and (d) respectively depict the monthly averaged POD and ETS of ERA5 forecasts at 12:00 UTC for each day in August 2016, under different combinations of neighborhood scales and threshold values. Similar to FSS, higher values of POD and ETS indicate better forecasting performance. Interestingly, the figure reveals that the POD and ETS values are relatively insensitive to changes in neighborhood scale, with variations primarily linked to the choice of threshold. The results of

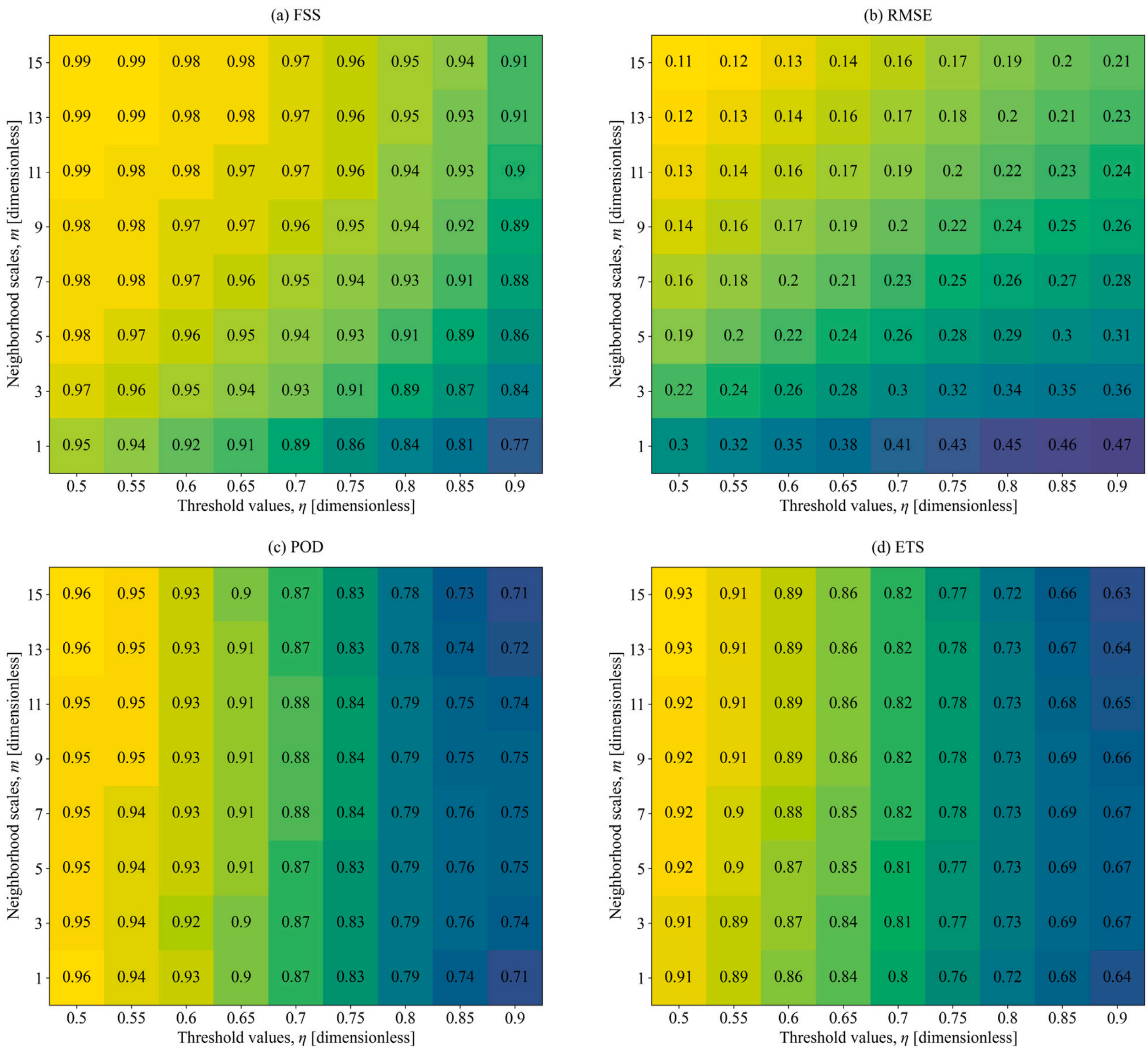


**Fig. 7.** Performance of ERA5 clear-sky index forecasts at 12:00 UTC timestamps in August 2016 over Europe against CAMS-Rad observations, under varying neighborhood scales and threshold values using the four evaluation metrics. Whereas FSSs and RMSEs are computed under the fraction-field method, PODs and ETSs are computed under the upscaling method. Better results are coded with brighter colors.

the upscaling method are in agreement with that of the fraction-field method, in that, higher neighborhood scales and low threshold values correspond to better apparent performance.

Additionally, spatial smoothing could be selectively employed to retain the displacement-induced errors that are of reference significance. However, in this study, when the neighborhood scale is set to 1 (i.e., without spatial smoothing), the FSSs tend to be lower, as depicted in Fig. 7 (a). These distinctions are more pronounced in high-resolution grid datasets [26]. Simultaneously, the FSS values obtained without spatial smoothing are very close to the useful skill scores ($\text{FSS}_{\text{uniform}}$, see Section 4.2 for details), making it difficult to establish clear criteria to quantify the quality gap between the relevant forecasts and the useful forecasts. Accordingly, smoothing is needed for spatial solar forecast verification.

### 4.2. A toy example to check the sanity of the neighborhood approach

One of the essential capabilities of performance metrics is to discern good forecasts from bad ones. Insofar as absolute verification is concerned, one has to be able to conclude whether the forecasts of interest are acceptable solely based on the forecasts themselves. Clearly, then, RMSE, POD, and ETS offer only limited information for such inquiries. For instance, it is known that a perfect forecast would correspond to a POD value of 1 (i.e., all hits no misses). However, since perfect forecasts are unattainable, the verification would return a POD value of less than 1. In this case, without a notion of the difficulty of the forecasting situation, it would be quite difficult to interpret any POD value obtained thereof. The same can be said for RMSE and ETS. The situation for FSS is, nevertheless, different. The useful skill score as defined in Eq. (12) provides just the insights needed.

The blue curve in Fig. 8(a) indicates the FSS values under $\eta = 0.5$ and $m = 1, 3, \ldots, 15$, using the ERA5 forecast at 12:00 UTC on 4 August 2016; this is referred to as ERA5_case1. The green dashed line in the figure shows the $\text{FSS}_{\text{uniform}}$, which lies beneath the blue curve, suggesting that the ERA5 forecast is useful. In the same subfigure, the orange curve marks the FSS values evaluated using a deliberately selected wrong forecast, which is in fact a forecast issued at 12:00 UTC on 30 November 2016; this is referred to as ERA5_case2. It can be seen that the yellow bars, despite being quite high numerically (FSS $\approx 0.8$), are lower than $\text{FSS}_{\text{uniform}}$, suggesting that the forecast is worse

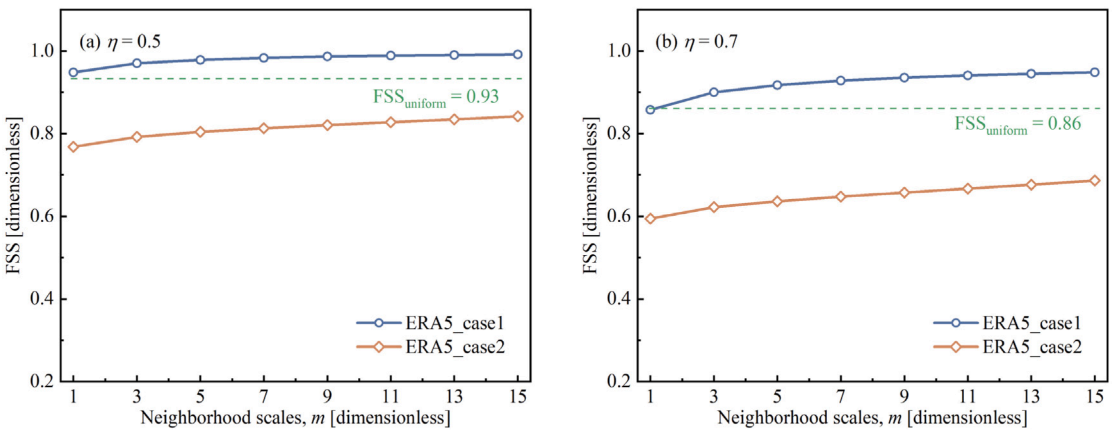


**Fig. 8.** Subplot(a) and (b) show the FSS values for ERA5_case1 and ERA5_case2 with $\eta = 0.5$ and $\eta = 0.7$, respectively.

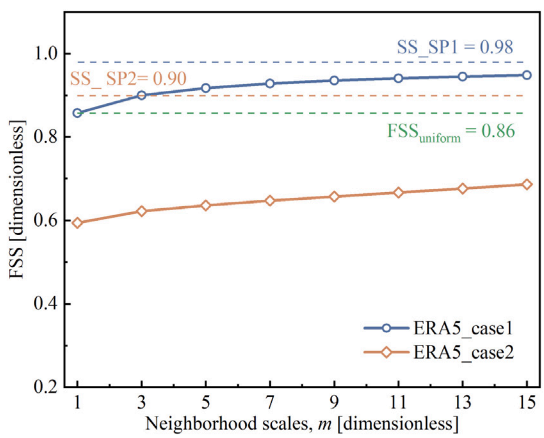


**Fig. 9.** Same as Fig. 8(b), but with two additional dashed lines indicating the point-location skill scores of the two forecasts.

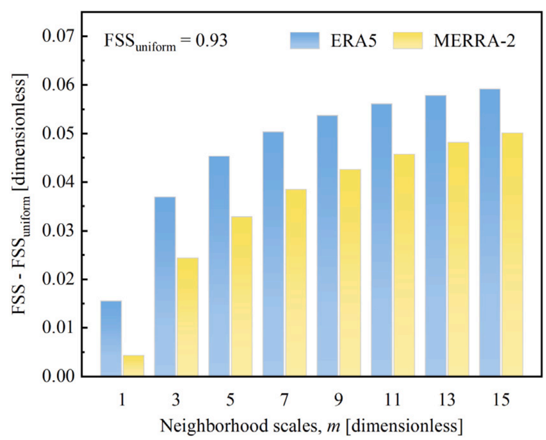


**Fig. 10.** The differences between the monthly average FSS and monthly average $\mathrm{FSS}_{\mathrm{uniform}}$ of all 9:00 to 14:00 timestamps in August 2016, with varying neighborhood scales, evaluated using ERA5 and MERRA-2 forecasts with $\eta = 0.5$.

than a random guess, and therefore is poor. In Fig. 8(b), another case with $\eta = 0.7$ is shown, and the same conclusion follows. After this toy example, the FSS should be favored over the other performance measures, due to its capacity to discern good forecasts from bad ones through $\mathrm{FSS}_{\mathrm{uniform}}$.

Subsequently, the advantage of spatial forecast verification over point-location forecast verification is elaborated with the same toy example. The point-location skill score, in the style of FSS, for the original forecast field (i.e., without dichotomization and smoothing) is given by:

$$\mathrm{SS} = 1 - \frac{\frac{1}{|D_s|}\sum_{(i,j)\in D_s}\left[\kappa_f(i,j) - \kappa_x(i,j)\right]^2}{\frac{1}{|D_s|}\left[\sum_{(i,j)\in D_s}\kappa_f^2(i,j) + \sum_{(i,j)\in D_s}\kappa_x^2(i,j)\right]}. \tag{22}$$

According to this definition, the point-location skill score of the ERA5 forecast at 12:00 UTC on 4 August 2016, and that of the deliberately selected wrong forecast are marked as the blue and orange dashed lines, respectively, in Fig. 9; this figure is identical to Fig. 8(b), except with the two additional dashed lines. Interestingly, the skill scores computed under the point-location forecast verification framework of the two cases are quite close (with skill scores of 0.98 and 0.90), albeit one of the forecasts is entirely nonsense. The limitation of the point-location verification is immediately obvious. More specifically, evaluating spatial forecasts with point-location verification metrics can be inappropriate, as they may issue superficially high scores when the forecasts are in fact poor. In conclusion, for applications of regional forecasting that permit small displacement errors, spatial forecast verification becomes highly necessary for high spatial resolution irradiance products. Conversely, if users are more interested in assessing forecast accuracy at a specific location, traditional point forecast verification could be more appropriate.

### *4.3. Comparison between ERA5 and MERRA-2*

To further examine the developed method under the comparative verification framework, the two sets of forecasts, namely, ERA5 and MERRA-2, are assessed and compared using FSS. Forecasts from those 9:00 to 14:00 UTC timestamps in August 2016 are used (i.e., a total of 186 instances). Note that ERA5 has a spatial resolution of $0.25° \times 0.25°$, whereas MERRA-2 has a spatial resolution of $0.5° \times 0.625°$. For comparison purposes, the two sets of forecasts must be first adjusted to the same spatial resolution. To that end, the bilinear interpolation method is utilized to adjust the spatial resolution of ERA5 to match that of MERRA-2. Using the fraction-field method, FSS values for ERA5 and MERRA-2 are shown in Fig. 10.

Fig. 10 shows the bar chart depicting the differences between the monthly average FSS and monthly average $\mathrm{FSS}_{\mathrm{uniform}}$ with varying neighborhood scales for ERA5 and MERRA-2. As the neighborhood scale increases, the monthly average FSS values also increase accordingly. The FSS values of ERA5 and MERRA-2 forecasts consistently exceed $\mathrm{FSS}_{\mathrm{uniform}}$, indicating that both sets of forecasts contain valuable information. However, the monthly average FSS values of ERA5 are higher than those of MERRA-2, suggesting superior forecasting performance for ERA5. The FSS verification results echo the observation made

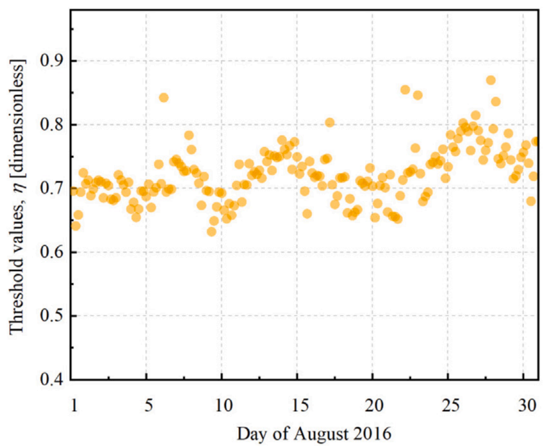


**Fig. 11.** The time series of automatic thresholds of CAMS-Rad satellite observations through the automatic threshold segmentation method for all 9:00 to 14:00 UTC timestamps in August 2016.

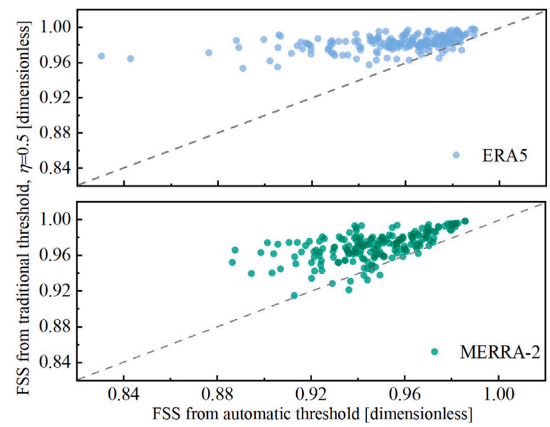


**Fig. 13.** Scatter plots of the FSS calculated through the automatic threshold segmentation method versus the FSS calculated using the traditional threshold determination method (with $\eta = 0.5$), with $m = 7$. The ERA5 and MERRA-2 forecasts are from 9:00 to 14:00 UTC timestamps in August 2016, with a total of 186 instances.

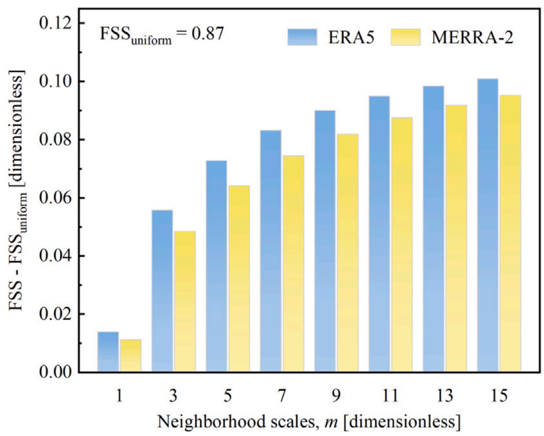


**Fig. 12.** The differences between average FSS and average $FSS_{uniform}$ of all 9:00 to 14:00 UTC timestamps in August 2016, with varying neighborhood scales for ERA5 and MERRA-2 forecast verification using automatic threshold segmentation method.

with the binary image of the $\kappa$ in Fig. 3(b), i.e., the cloud coverage area of the ERA5 forecast (at least for that instance) better resembles the observation. In conclusion, under the fraction-field method, both ERA5 and MERRA-2 are meaningful forecasts, although the former should be favored over the latter due to its higher skill.

### 4.4. Verification results of automatic thresholds segmentation

In traditional neighborhood-based spatial forecast verification, the selection of thresholds has often been subjective, and determining suitable values has typically involved the computation of multiple sets of thresholds, as seen in Section 4.1. To eliminate such subjectivity and improve verification efficiency, a three-component skew-normal mixtures model is employed to fit the probability densities of observational data; recall Section 2.3. The thresholds are determined as the mean $\kappa$ values corresponding to the intersections of the component PDFs. Fig. 11 shows the time series of automatic thresholds of CAMS-Rad satellite observations through the automatic threshold segmentation method for all 9:00 to 14:00 UTC timestamps in August 2016. This method enables the automatic calculation of thresholds based on the sky conditions, with the threshold values concentrated in the range of 0.7 to 0.8.

Using the same data as Section 4.3, but with automatic threshold segmentation, the skill scores are presented in Fig. 12. The conclusion that ERA5 is superior to MERRA-2 remains unchanged. However, as compared to the $\eta = 0.5$ case in Section 4.3, the forecasts now attain reduced values for both FSS and $FSS_{uniform}$. Fig. 13 shows the scatter plots of the FSS obtained using the automatic threshold segmentation method and the FSS calculated by the traditional threshold determination method (with $\eta = 0.5$), with a neighborhood-scale of $m = 7$. The forecasts are derived from the ERA5 and MERRA-2 systems for a total of 186 instances, spanning from 9:00 to 14:00 UTC timestamps in August 2016. It can be seen that, in comparison to the traditional threshold method, the FSS values obtained through automatic threshold segmentation are generally lower. The traditional fraction-field method posits that higher FSS values indicate better forecasting performance [26]. However, excessively high FSS values may not be suitable for the verification of solar irradiance forecast. Similarly, as observed in Fig. 7, under the traditional threshold method, very high FSS values obtained with low thresholds or large neighborhood scales may be misleading, indicating a notable overestimation in forecasting verification. At this point, it becomes crucial to identify an appropriate combination of neighborhood scale and threshold value that accurately reflects forecasting performance. The automatic threshold segmentation method effectively addresses the overestimation that occurs when subjective threshold selection is too low. Simultaneously, selecting a neighborhood scale of $m = 7$ mitigates the risk of excessively high FSS values resulting from an excessively large neighborhood scale. As illustrated in Fig. 13, the scatter plots provide a visual representation supporting the discussion mentioned earlier.

Using the same data as Section 4.2, but computed through the automatic threshold segmentation using the three-component skew-normal mixtures model, the skill scores for ERA5_case1 and ERA5_case2 are illustrated in Fig. 14. The conclusions align closely with the results obtained using the traditional threshold $\eta = 0.7$. The ERA5_case1 forecast exhibits high accuracy, while the ERA5_case2 forecast performs worse than random guessing. It is also noteworthy that point-location verification is not suitable for spatial forecast verification of solar irradiance.

Therefore, applying the automatic segmentation of thresholds using the three-component skew-normal mixtures model to neighborhood-based spatial forecast verification is more rational. It allows for a more accurate evaluation of the forecasting performance. The automated threshold segmentation method serves to eliminate subjective threshold selection. Thresholds can be categorized based on clear sky and cloudy conditions, with distinct threshold allocations for observational data across different time series. Furthermore, rational thresholds ensure a more reasonable FSS value, preventing the overestimation of forecasting skill due to the subjective choice of excessively low thresholds

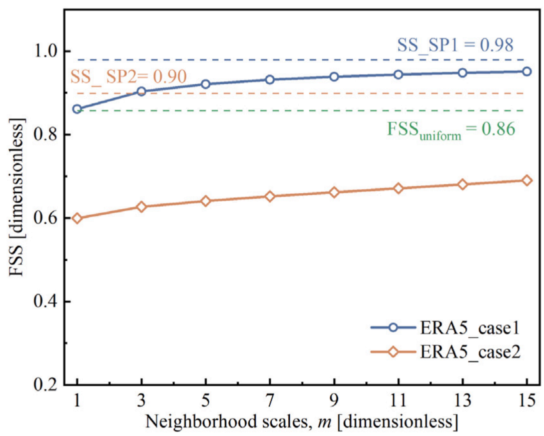


**Fig. 14.** Applying the automatic threshold segmentation using the three-component skew-normal mixture model to calculate the FSS curves for ERA5_case1 and ERA5_case2 data, along with the corresponding point-location verification curves.

or the underestimation of forecasting capability when thresholds are subjectively set too high.

## 5. Conclusion

Traditional ground-based forecast verification faces limitations due to observation location constraints, thereby lacking spatial feature information. This constraint often results in verification errors caused by small-scale spatial displacements. This study introduces a neighborhood-based spatial forecast verification approach to address such limitations. This approach involves smoothing observational and forecast data separately using neighborhood windows, relaxing the requirement for precise matching between observations and forecasts in a grid scale.

Within the spatial neighborhood smoothing framework, two workflows—the fraction-field method and the upscaling method—are employed to assess the forecast performance. Both methods demonstrated effectiveness in assessing forecast performance under conditions of moderate to low clear-sky index conditions but displayed a tendency for underestimation in forecasts with high clear-sky indices. In contrast, the accuracy measure FSS is better suited for spatial forecast verification, as it accurately quantifies forecast performance through explicit numerical values. Additionally, it encompasses its own useful skill scores ($FSS_{uniform}$) for various observations. When the FSS value surpasses $FSS_{uniform}$, it indicates the presence of useful information in the forecast. A higher FSS value corresponds to better forecast performance, with $FSS = 1$ representing a perfect match between forecast and observation. A lower FSS value corresponds to poorer forecast performance, and $FSS = 0$ signifies a complete mismatch between forecast and observation. This study reveals that the FSS values associated with incorrectly forecast information consistently remain under $FSS_{uniform}$. As forecast skill diminishes, there is a corresponding decrease in FSS values.

Furthermore, experiments indicate that, compared to point-location forecast verification, neighborhood-based spatial forecast verification can accurately reflect forecast performance. Skill scores in point forecast verification consistently remain high, making it challenging to distinguish the quality of forecast skills. To further examine the developed spatial forecast verification methodology, the forecast performance of two reanalyzed systems, namely ERA5 and MERRA-2, is subsequently compared. The results indicate that, in comparison to MERRA-2, ERA5 exhibits superior forecasting performance. This conclusion is consistent with the actual forecasting situations depicted in the visualization images for the two reanalysis systems. The application of the automatic threshold segmentation method based on the three-component skew-normal mixtures model results in a more rational threshold allocation, effectively avoiding subjective threshold selection or extensive computations to determine suitable threshold intensities. Simultaneously, it prevents the overestimation of forecasting skills resulting from the subjective choice of excessively low thresholds or the underestimation of forecasting skills associated with the subjective selection of excessively high thresholds. Additionally, to facilitate the practical application of the proposed neighborhood-based spatial forecast verification method, all parameter fitting processes can be encapsulated into a system program, requiring users only the input of the irradiance data to be verified and relevant parameters. The threshold is automatically computed by the automatic threshold segmentation method, thereby obviating the necessity for users to locally modify threshold parameters. The neighborhood scale used for spatial smoothing is determined based on the specific application. Typically, for irradiance data with the same high spatial resolution, a generic neighborhood scale value can be chosen. This technological approach makes it easier to construct a unified framework structure, facilitating the comparison of forecasting performance across a large number of forecast datasets. As such, the neighborhood-based spatial forecast verification becomes more reasonable and widely applicable.

Besides, neighborhood-based spatial forecast verification can also be used to monitor the monthly or quarterly forecasting performance in specific regions and evaluate the degree of forecasting performance improvement after upgrading forecasting models. Related topics could be a focal point for future exploration. Moreover, in view of the recent hype in machine-learning-based weather forecasting, which moves beyond training at a single location, but rather over an area. It should be highlighted that the recent advances in machine-learning-based weather forecasting provide important future applications for the proposal verification method. Unlike traditional machine-learning methods, which train the models with just local data, the recent deep-learning models almost always leverage areal data. In that sense, the present verification metric can be used as the loss function during such training, though such an avenue is not part of this work.

## CRediT authorship contribution statement

**Xiaomi Zhang:** Conceptualization, Methodology, Software, Formal analysis, Validation, Investigation, Resources, Writing – original draft, Visualization, Project administration. **Dazhi Yang:** Methodology, Software, Formal analysis, Investigation, Writing – original draft, Visualization, Data curation, Funding acquisition. **Hao Zhang:** Methodology, Validation, Investigation, Writing – review & editing, Funding acquisition. **Bai Liu:** Methodology, Validation, Investigation, Data curation. **Mengying Li:** Visualization, Writing – original draft, Writing – review & editing. **Yinghao Chu:** Methodology, Writing – review & editing. **Jingnan Wang:** Writing – review & editing. **Xiang'ao Xia:** Writing – review & editing.

## Declaration of competing interest

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

## Data availability

Data will be made available on request.

## Acknowledgments

Dazhi Yang is supported by the National Natural Science Foundation of China (project no. 42375192) and the China Meteorological Administration Climate Change Special Program (CMA-CCSP; project no. QBZ202315).

Hao Zhang is supported by the China Postdoctoral Science Foundation [2023M740908] and Heilongjiang Postdoctoral Foundation, China [LBH-Z22120].

## References


[1] Rahman A, Farrok O, Haque MM. Environmental impact of renewable energy source based electrical power plants: Solar, wind, hydroelectric, biomass, geothermal, tidal, ocean, and osmotic. Renew Sustain Energy Rev 2022;161:112279. http://dx.doi.org/10.1016/j.rser.2022.112279, URL https://www.sciencedirect.com/science/article/pii/S136403212200199X.

[2] Yang D, Wang W, Gueymard CA, Hong T, Kleissl J, Huang J, Perez MJ, Perez R, Bright JM, Xia X, van der Meer D, Peters IM. A review of solar forecasting, its dependence on atmospheric sciences and implications for grid integration: Towards carbon neutrality. Renew Sustain Energy Rev 2022;161:112348. http://dx.doi.org/10.1016/j.rser.2022.112348, URL https://www.sciencedirect.com/science/article/pii/S1364032122002593.

[3] Bollipo RB, Mikkili S, Bonthagorla PK. Hybrid, optimal, intelligent and classical PV MPPT techniques: A review. CSEE J Power Energy Syst 2021;7(1):9–33. http://dx.doi.org/10.17775/CSEEJPES.2019.02720.

[4] Purohit I, Purohit P. Inter-comparability of solar radiation databases in Indian context. Renew Sustain Energy Rev 2015;50:735–47. http://dx.doi.org/10.1016/j.rser.2015.05.020, URL https://www.sciencedirect.com/science/article/pii/S1364032115004797.

[5] Manoharan M, Mohamed L, Brahim B. Statistical analysis of novel ensemble recursive radial basis function neural network performance on global solar irradiance forecasting. J Electr Comput Eng 2023;2023:2554355. http://dx.doi.org/10.1155/2023/2554355.

[6] Yang G, Zhang H, Wang W, Liu B, Lyu C, Yang D. Capacity optimization and economic analysis of PV–hydrogen hybrid systems with physical solar power curve modeling. Energy Convers Manage 2023;288:117128. http://dx.doi.org/10.1016/j.enconman.2023.117128, URL https://www.sciencedirect.com/science/article/pii/S0196890423004740.

[7] de O. Santos DS, de Mattos Neto PSG, de Oliveira JFL, Siqueira HV, Barchi TM, Lima ARFM, Dantas DAP, Converti A, Pereira AC, de Melo Filho JB, Marinho MHN. Solar irradiance forecasting using dynamic ensemble selection. Appl Sci 2022;12(7):3510. http://dx.doi.org/10.3390/app12073510, URL https://www.mdpi.com/2076-3417/12/7/3510.

[8] Chinnavornrungsee P, Kittisontirak S, Chollacoop N, Songtrai S, Sriprapha K, Uthong P, et al. Solar irradiance prediction in the tropics using a weather forecasting model. Japan J Appl Phys 2023;62:SK1050. http://dx.doi.org/10.35848/1347-4065/acd4c8.

[9] Yang D, Kleissl J. Solar Irradiance and Photovoltaic Power Forecasting. CRC Press; 2024, http://dx.doi.org/10.1201/9781003203971.

[10] Yang D, Wang W, Xia X. A concise overview on solar resource assessment and forecasting. Adv Atmosph Sci 2022;39:1239–51. http://dx.doi.org/10.1007/s00376-021-1372-8.

[11] Murphy AH. What is a good forecast—An essay on the nature of goodness in weather forecasting. Weather Forecast 1993;8:281–93. http://dx.doi.org/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2.

[12] Yang D, Alessandrini S, Antonanzas J, Antonanzas-Torres F, Badescu V, Beyer HG, Blaga R, Boland J, Bright JM, Coimbra CFM, David M, Frimane Â, Gueymard CA, Hong T, Kay MJ, Killinger S, Kleissl J, Lauret P, Lorenz E, van der Meer D, Paulescu M, Perez R, Perpiñán-Lamigueiro O, Peters IM, Reikard G, Renné D, Saint-Drenan Y-M, Shuai Y, Urraca R, Verbois H, Vignola F, Voyant C, Zhang J. Verification of deterministic solar forecasts. Sol Energy 2020;210:20–37. http://dx.doi.org/10.1016/j.solener.2020.04.019, URL https://www.sciencedirect.com/science/article/pii/S0038092X20303947, Special Issue on Grid Integration.

[13] Liu B, Yang D, Mayer MJ, Coimbra CFM, Kleissl J, Kay M, Wang W, Bright JM, Xia X, Lv X, Srinivasan D, Wu Y, Beyer HG, Yagli GM, Shen Y. Predictability and forecast skill of solar irradiance over the contiguous United States. Renew Sustain Energy Rev 2023;182:113359. http://dx.doi.org/10.1016/j.rser.2023.113359, URL https://www.sciencedirect.com/science/article/pii/S1364032123002162.

[14] Yang D, Kleissl J, Gueymard CA, Pedro HTC, Coimbra CFM. History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. Sol Energy 2018;168:60–101. http://dx.doi.org/10.1016/j.solener.2017.11.023, URL https://www.sciencedirect.com/science/article/pii/S0038092X17310022.

[15] Yang D. A guideline to solar forecasting research practice: Reproducible, operational, probabilistic or physically-based, ensemble, and skill (ROPES). J Renew Sustain Energy 2019;11:022701. http://dx.doi.org/10.1063/1.5087462.

[16] Makridakis S, Hyndman RJ, Petropoulos F. Forecasting in social settings: The state of the art. Int J Forecast 2020;36(1):15–28. http://dx.doi.org/10.1016/j.ijforecast.2019.05.011, URL https://www.sciencedirect.com/science/article/pii/S0169207019301876.

[17] Rajagukguk RA, Kamil R, Lee H-J. A deep learning model to forecast solar irradiance using a sky camera. Appl Sci 2021;11:5049. http://dx.doi.org/10.3390/app11115049, URL https://www.mdpi.com/2076-3417/11/11/5049.

[18] Yang D, Yang G, Liu B. Combining quantiles of calibrated solar forecasts from ensemble numerical weather prediction. Renew Energy 2023;215:118993. http://dx.doi.org/10.1016/j.renene.2023.118993, URL https://www.sciencedirect.com/science/article/pii/S0960148123008996.

[19] Wang W, Yang D, Hong T, Kleissl J. An archived dataset from the ECMWF Ensemble Prediction System for probabilistic solar power forecasting. Sol Energy 2022;248:64–75. http://dx.doi.org/10.1016/j.solener.2022.10.062, URL https://www.sciencedirect.com/science/article/pii/S0038092X22008015.

[20] Pedro HTC, Larson DP, Coimbra CFM. A comprehensive dataset for the accelerated development and benchmarking of solar forecasting methods. J Renew Sustain Energy 2019;11:036102. http://dx.doi.org/10.1063/1.5094494.

[21] Murphy AH, Winkler RL. A general framework for forecast verification. Mon Weather Rev 1987;115:1330–8. http://dx.doi.org/10.1175/1520-0493(1987)115<1330:AGFFFV>2.0.CO;2, URL https://journals.ametsoc.org/view/journals/mwre/115/7/1520-0493_1987_115_1330_agfffv_2_0_co_2.xml.

[22] Yagli GM, Yang D, Srinivasan D. Automatic hourly solar forecasting using machine learning models. Renew Sustain Energy Rev 2019;105:487–98. http://dx.doi.org/10.1016/j.rser.2019.02.006, URL https://www.sciencedirect.com/science/article/pii/S1364032119300905.

[23] Beck HE, Zimmermann NE, McVicar TR, Vergopolan N, Berg A, Wood EF. Present and future Köppen-Geiger climate classification maps at 1-km resolution. Sci Data 2018;5:180214. http://dx.doi.org/10.1038/sdata.2018.214.

[24] Gilleland E, Ahijevych DA, Brown BG, Ebert EE. Verifying forecasts spatially. Bull Am Meteorol Soc 2010;91:1365–76. http://dx.doi.org/10.1175/2010BAMS2819.1, URL https://journals.ametsoc.org/view/journals/bams/91/10/2010bams2819_1.xml.

[25] Gilleland E, Ahijevych D, Brown BG, Casati B, Ebert EE. Intercomparison of spatial forecast verification methods. Weather Forecast 2009;24:1416–30. http://dx.doi.org/10.1175/2009WAF2222269.1, URL https://journals.ametsoc.org/view/journals/wefo/24/5/2009waf2222269_1.xml.

[26] Ebert EE. Fuzzy verification of high-resolution gridded forecasts: A review and proposed framework. Meteorol Appl 2008;15:51–64. http://dx.doi.org/10.1002/met.25, URL https://api.semanticscholar.org/CorpusID:62140586.

[27] Yates E, Anquetin S, Ducrocq V, Creutin J-D, Ricard D, Chancibault K. Point and areal validation of forecast precipitation fields. Meteorol Appl 2006;13:1–20. http://dx.doi.org/10.5194/nhess-5-741-2005.

[28] Zepeda-Arce J, Foufoula-Georgiou E, Droegemeier KK. Space-time rainfall organization and its role in validating quantitative precipitation forecasts. J Geophys Res: Atmos 2000;105(D8):10129–46. http://dx.doi.org/10.1029/1999JD901087, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/1999JD901087.

[29] Roberts NM, Lean HW. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. Mon Weather Rev 2008;136:78–97. http://dx.doi.org/10.1175/2007MWR2123.1, URL https://journals.ametsoc.org/view/journals/mwre/136/1/2007mwr2123.1.xml.

[30] Lack SA, Limpert GL, Fox NI. An object-oriented multiscale verification scheme. Weather Forecast 2010;25(1):79–92. http://dx.doi.org/10.1175/2009WAF2222245.1, URL https://journals.ametsoc.org/view/journals/wefo/25/1/2009waf2222245_1.xml.

[31] Cressie N, Wikle CK. Statistics for Spatio-Temporal Data. John Wiley & Sons; 2015.

[32] Stamus PA, Carr FH, Baumhefner DP. Application of a scale-separation verification technique to regional forecast models. Mon Weather Rev 1992;120(1):149–63. http://dx.doi.org/10.1175/1520-0493(1992)120<0149:AOASSV>2.0.CO;2, URL https://journals.ametsoc.org/view/journals/mwre/120/1/1520-0493_1992_120_0149_aoassv_2_0_co_2.xml.

[33] Casati B. New developments of the intensity-scale technique within the spatial verification methods intercomparison project. Weather Forecast 2010;25(1):113–43. http://dx.doi.org/10.1175/2009WAF2222257.1, URL https://journals.ametsoc.org/view/journals/wefo/25/1/2009waf2222257_1.xml.

[34] Ebert EE, McBride JL. Verification of precipitation in weather systems: Determination of systematic errors. J Hydrol 2000;239(1):179–202. http://dx.doi.org/10.1016/S0022-1694(00)00343-7, URL https://www.sciencedirect.com/science/article/pii/S0022169400003437.

[35] Hoffman RN, Liu Z, Louis J-F, Grassoti C. Distortion representation of forecast errors. Mon Weather Rev 1995;123(9):2758–70. http://dx.doi.org/10.1175/1520-0493(1995)123<2758:DROFE>2.0.CO;2, URL https://journals.ametsoc.org/view/journals/mwre/123/9/1520-0493_1995_123_2758_drofe_2_0_co_2.xml.

[36] Bookstein FL. Principal warps: Thin-plate splines and the decomposition of deformations. IEEE Trans Pattern Anal Mach Intell 1989;11(6):567–85. http://dx.doi.org/10.1109/34.24792.

[37] Kumar DS, Yagli GM, Kashyap M, Srinivasan D. Solar irradiance resource and forecasting: A comprehensive review. IET Renew Power Gener 2020;14(10):1641–56. http://dx.doi.org/10.1049/iet-rpg.2019.1227, URL https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-rpg.2019.1227.

[38] Kurtz B, Kleissl J. Measuring diffuse, direct, and global irradiance using a sky imager. Sol Energy 2017;141:311–22. http://dx.doi.org/10.1016/j.solener.2016.11.032, URL https://www.sciencedirect.com/science/article/pii/S0038092X16305722.

[39] Bright JM. Solcast: Validation of a satellite-derived solar irradiance dataset. Sol Energy 2019;189:435–49. http://dx.doi.org/10.1016/j.solener.2019.07.086, URL https://www.sciencedirect.com/science/article/pii/S0038092X19307571.

[40] André M, Perez R, Soubdhan T, Schlemmer J, Calif R, Monjoly S. Preliminary assessment of two spatio-temporal forecasting technics for hourly satellite-derived irradiance in a complex meteorological context. Sol Energy 2019;177:703–12. http://dx.doi.org/10.1016/j.solener.2018.11.010, URL https://www.sciencedirect.com/science/article/pii/S0038092X18311101.

[41] Polo J, Wilbert S, Ruiz-Arias JA, Meyer R, Gueymard CA, Súri M, Martín L, Mieslinger T, Blanc P, Grant I, Boland J, Ineichen P, Remund J, Escobar R, Troccoli A, Sengupta M, Nielsen KP, Renne D, Geuder N, Cebecauer T. Preliminary survey on site-adaptation techniques for satellite-derived and reanalysis solar radiation datasets. Sol Energy 2016;132:25–37. http://dx.doi.org/10.1016/j.solener.2016.03.001, URL https://www.sciencedirect.com/science/article/pii/S0038092X16001754.

[42] Yang D, Perez R. Can we gauge forecasts using satellite-derived solar irradiance? J Renew Sustain Energy 2019;11:023704. http://dx.doi.org/10.1063/1.5087588.

[43] Lorenz E, Hurka J, Heinemann D, Beyer HG. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. IEEE J Sel Top Appl Earth Obs Remote Sens 2009;2(1):2–10. http://dx.doi.org/10.1109/JSTARS.2009.2020300.

[44] Yang D, Boland J. Satellite-augmented diffuse solar radiation separation models. J Renew Sustain Energy 2019;11:023705. http://dx.doi.org/10.1063/1.5087463.

[45] Yang D. Post-processing of NWP forecasts using ground or satellite-derived data through kernel conditional density estimation. J Renew Sustain Energy 2019;11:026101. http://dx.doi.org/10.1063/1.5088721.

[46] Yagli GM, Yang D, Gandhi O, Srinivasan D. Can we justify producing univariate machine-learning forecasts with satellite-derived solar irradiance? Appl Energy 2020;259:114122. http://dx.doi.org/10.1016/j.apenergy.2019.114122, URL https://www.sciencedirect.com/science/article/pii/S0306261919318094.

[47] Jiménez PA, Yang J, Kim J-H, Sengupta M, Dudhia J. Assessing the WRF-solar model performance using satellite-derived irradiance from the national solar radiation database. J Appl Meteorol Climatol 2022;61:129–42. http://dx.doi.org/10.1175/JAMC-D-21-0090.1, URL https://journals.ametsoc.org/view/journals/apme/61/2/JAMC-D-21-0090.1.xml.

[48] Bessho K, Date K, Hayashi M, Ikeda A, Imai T, Inoue H, Kumagai Y, Miyakawa T, Murata H, Ohno T, Okuyama A, Oyama R, Sasaki Y, Shimazu Y, Shimoji K, Sumida Y, Suzuki M, Taniguchi H, Tsuchiyama H, Uesawa D, Yokoda H, Yoshida R. An introduction to Himawari-8/9—Japan's new-generation geostationary meteorological satellites. Jo Meteorol Soc Japan Ser II 2016;94:151–83. http://dx.doi.org/10.2151/jmsj.2016-009.

[49] Zhang Q, Zhu J, Huang Y, Yuan Q, Zhang L. Beyond being wise after the event: Combining spatial, temporal and spectral information for Himawari-8 early-stage wildfire detection. Int J Appl Earth Obs Geoinf 2023;124:103506. http://dx.doi.org/10.1016/j.jag.2023.103506, URL https://www.sciencedirect.com/science/article/pii/S1569843223003308.

[50] Ebert EE. Neighborhood verification: A strategy for rewarding close forecasts. Weather Forecast 2009;24:1498–510. http://dx.doi.org/10.1175/2009WAF2222251.1, URL https://journals.ametsoc.org/view/journals/wefo/24/6/2009waf2222251_1.xml.

[51] Murphy AH. Forecast verification: Its complexity and dimensionality. Mon Weather Rev 1991;119(7):1590–601. http://dx.doi.org/10.1175/1520-0493(1991)119<1590:FVICAD>2.0.CO;2, URL https://journals.ametsoc.org/view/journals/mwre/119/7/1520-0493_1991_119_1590_fvicad_2_0_co_2.xml.

[52] Mittermaier M, Roberts N. Intercomparison of spatial forecast verification methods: Identifying skillful spatial scales using the fractions skill score. Weather Forecast 2010;25:343–54. http://dx.doi.org/10.1175/2009WAF2222260.1, URL https://journals.ametsoc.org/view/journals/wefo/25/1/2009waf2222260_1.xml.

[53] Voskrebenzev A, Riechelmann S, Bais A, Slaper H, Seckmeyer G. Estimating probability distributions of solar irradiance. Theor Appl Climatol 2015;119:465–79. http://dx.doi.org/10.1007/s00704-014-1189-9, URL https://link.springer.com/article/10.1007/s00704-014-1189-9.

[54] Hollands KGT, Suehrcke H. A three-state model for the probability distribution of instantaneous solar radiation, with applications. Sol Energy 2013;96:103–12. http://dx.doi.org/10.1016/j.solener.2013.07.007, URL https://www.sciencedirect.com/science/article/pii/S0038092X13002727.

[55] Jurado M, Caridad JM, Ruiz V. Statistical distribution of the clearness index with radiation data integrated over five minute intervals. Sol Energy 1995;55(6):469–73. http://dx.doi.org/10.1016/0038-092X(95)00067-2, URL https://www.sciencedirect.com/science/article/pii/0038092X95000672.

[56] Prates MO, Lachos VH, Barbosa Cabral CR. Mixsmsn: Fitting finite mixture of scale mixture of skew-normal distributions. J Stat Softw 2013;54(12):1—-20. http://dx.doi.org/10.18637/jss.v054.i12, URL https://www.jstatsoft.org/index.php/jss/article/view/v054i12.

[57] Yang D, Dong Z, Lim LHI, Liu L. Analyzing big time series data in solar engineering using features and PCA. Sol Energy 2017;153:317–28. http://dx.doi.org/10.1016/j.solener.2017.05.072, URL https://www.sciencedirect.com/science/article/pii/S0038092X17304796.

[58] Yang D, Bright JM. Worldwide validation of 8 satellite-derived and reanalysis solar radiation products: A preliminary evaluation and overall metrics for hourly data over 27 years. Sol Energy 2020;210:3–19. http://dx.doi.org/10.1016/j.solener.2020.04.016, URL https://www.sciencedirect.com/science/article/pii/S0038092X20303893.

[59] Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, Nicolas J, Peubey C, Radu R, Schepers D, Simmons A, Soci C, Abdalla S, Abellan X, Balsamo G, Bechtold P, Biavati G, Bidlot J, Bonavita M, De Chiara G, Dahlgren P, Dee D, Diamantakis M, Dragani R, Flemming J, Forbes R, Fuentes M, Geer A, Haimberger L, Healy S, Hogan RJ, Hólm E, Janisková M, Keeley S, Laloyaux P, Lopez P, Lupu C, Radnoti G, de Rosnay P, Rozum I, Vamborg F, Villaume S, Thépaut J-N. The ERA5 global reanalysis. Q J R Meteorol Soc 2020;146(730):1999–2049. http://dx.doi.org/10.1002/qj.3803, URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3803.

[60] Siles Soria G. ERA5 climatological reanalysis: A review of its use in calculating atmospheric attenuation in satellite communications systems. Investig Desarrollo 2022;22:145–59. http://dx.doi.org/10.23881/idupbo.022.1-12i.

[61] Gelaro R, McCarty W, Suárez MJ, Todling R, Molod A, Randles LTCA, Darmenov A, Bosilovich MG, Reichle R, Wargan K, Coy L, Cullather R, Draper C, Akella S, Buchard V, Conaty A, da Silva AM, Gu W, Kim G-K, Koster R, Lucchesi R, Merkova D, Nielsen JE, Partyka G, Pawson S, Putman W, Rienecker M, Schubert SD, Sienkiewicz M, Zhao B. The Modern-Era Retrospective Analysis for Research and Applications. J Clim 2017;30:5419–54. http://dx.doi.org/10.1175/JCLI-D-16-0758.1, URL https://journals.ametsoc.org/view/journals/clim/30/14/jcli-d-16-0758.1.xml.

[62] Randles CA, da Silva AM, Buchard V, Colarco PR, Darmenov A, Govindaraju R, Smirnov A, Holben B, Ferrare R, Hair J, Shinozuka Y, Flynn CJ. MERRA-2 aerosol reanalysis, 1980 onward. Part I: System description and data assimilation evaluation. J Clim 2017;30:6823–50. http://dx.doi.org/10.1175/JCLI-D-16-0609.1, URL https://journals.ametsoc.org/view/journals/clim/30/17/jcli-d-16-0609.1.xml.

[63] Bright JM, Bai X, Zhang Y, Sun X, Acord B, Wang P. Irradpy: Python package for MERRA-2 download, extraction and usage for clear-sky irradiance modelling. Sol Energy 2020;199:685–93. http://dx.doi.org/10.1016/j.solener.2020.02.061, URL https://www.sciencedirect.com/science/article/pii/S0038092X20301894.