

# Forecasting Solar Radiation and Photovoltaic Power

Five (5) Continuing Education Hours  
Course #EE1510

Approved Continuing Education for Licensed Professional Engineers

*EZ-pdh.com*  
*Ezekiel Enterprises, LLC*  
*301 Mission Dr. Unit 571*  
*New Smyrna Beach, FL 32170*  
*800-433-1487*  
*support@ez-pdh.com*



### Course Description:

The Forecasting Solar Radiation and Photovoltaic Power course satisfies five (5) hours of professional development.

The course is designed as a distance learning course that overviews the basic concepts of solar irradiance and Photovoltaic (PV) power forecasting. Course based on *Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications*, 3rd edition, NREL, 2021.

### Objectives:

The primary objective of this course is to enable the student to understand solar irradiance forecasting and to understand the basic concepts and various methods of solar irradiance forecasting as well as forecasting the power production of Photovoltaic (PV) power plants.

### Grading:

Students must achieve a minimum score of 70% on the online quiz to pass this course. The quiz may be taken as many times as necessary to successfully pass and complete the course.

A copy of the quiz questions is attached to the last pages of this document.

# Table of Contents

## Forecasting Solar Radiation and Photovoltaic Power

**Forecasting Solar Radiation and Photovoltaic Power**

**1 Introduction ..... 1**

**2 Empirical and Physical Solar Irradiance Forecasting Methods ..... 5**

**3 Irradiance Forecasting Based on Irradiance Time Series and Post-Processing with Statistical and Machine Learning Methods ..... 16**

**4 PV Power Forecasting and Regional Upscaling..... 22**

**5 Evaluation of Irradiance and PV Power Forecasts ..... 28**

**6 Probabilistic Solar Forecasts ..... 38**

**7 Summary and Recommendations for Irradiance Forecasting..... 49**

**Quiz Questions..... 52**

# Forecasting Solar Radiation and Photovoltaic Power

## 1 Introduction

Solar resource forecasting is critical for the operation and management of solar power plants and electric grids. Solar radiation is highly variable because it is driven mainly by synoptic and local weather patterns. This high variability presents challenges to meeting power production and demand curves, notably in the case of solar photovoltaic (PV) power plants, which have little or no storage capacity. For concentrating solar power (CSP) plants, variability issues are partially mitigated by the thermal inertia of the plant, including its heat transfer fluid, heat exchangers, turbines, and potentially coupling with a heat storage facility; however, temporally and spatially varying irradiance introduces thermal stress in critical system components and plant management issues that can result in the degradation of the overall system's performance and reduction of the plant's lifetime. The variability can also result in lower plant efficiencies than those that occur under operation in stable conditions because optimally operating the plant under variable conditions is significantly more challenging. For PV power plants that have battery storage, forecasts are helpful for scheduling the charging process of the batteries at the most appropriate time, optimizing the fractions of electricity delivered and stored at any instant, and thus avoiding the loss of usable energy.

Solar radiation forecasting anticipates the solar radiation transients and the power production of solar energy systems, allowing for the setup of contingency mechanisms to mitigate any deviation from the required production.

With the expected integration of increasing shares of solar power into the electric grid, reliable predictions of solar power production are becoming increasingly important. PV power represents one of the main shares of the total renewable energy, along with wind power generation (IRENA 2019). High penetrations of PV power generation pose several challenges for the stability of the electric grid because of the stochastic variability of the residual electric load (i.e., the difference between the energy need—or load—and the distributed PV power generation, depending on meteorological conditions and sun position); therefore, accurate forecasting of PV power

generation is required for energy scheduling and for balancing demand and supply. This information is essential for distribution system operators and transmission system operators (TSOs) as well as for aggregators and energy traders (Pierro et al. 2017).

Today, PV power prediction systems are an essential part of electric grid management in countries that have substantial shares of solar power generation, among which Germany is a paradigmatic case. For example, in 2020, Germany had an installed PV power capacity of more than 50 GW<sub>peak</sub>, supplying more than 50% of the total load on sunny summer days at noon. In this context, and according to the German Renewable Energy Sources Act, TSOs are in charge of marketing and balancing the overall fluctuating PV power feed-in, which requires the use of regional forecasts for the designated control areas. Additionally, optional direct marketing of PV power is based on forecasts for the PV power plant's output. PV power is first offered on the day-ahead auction at the European Power Exchange. Subsequently, amendments based on updated forecasts can be made on the intraday market, when electricity can be traded until 30 minutes before delivery begins. Remaining deviations between scheduled and needed power are adjusted using balancing power. A similar procedure for California's electricity market was described by Mathiesen, Kleissl, and Collier (2013). Also, Kleissl (2013) described the stakeholder needs from the perspective of independent system operators and energy traders. Hence, accurate PV power forecasts at different spatial and temporal scales are extremely important for cost-efficient grid integration because large errors in the day-ahead forecast can cause either very high or negative prices on the intraday market, and intraday forecast errors determine the need for costly balancing power.

Several studies have evaluated the added value of solar irradiance forecasting for solar energy applications. For example, Dumortier (2009) gave a preliminary overview of such applications. Many other authors have detailed specific use cases and benefits of solar power forecasting. The following is a nonexhaustive list:

- In the realm of electric grids, Perez et al. (2007) evaluated the operational accuracy of end-use forecasts and their ability to predict the effective capacity of grid-connected PV power plants.
- Kaur et al. (2016) described the benefits of solar forecasting for energy imbalance markets.
- The specific needs of solar forecasting for the real-time electricity market and forecasting requirements from the California Independent System Operator have been examined by Yang, Wu, and Kleissl (2019), showing that hourly forecasts could be appropriately downscaled to the contemplated 15-minute resolution.
- Rikos et al. (2008), Diagné et al. (2013), and Simoglou et al. (2014) examined the solar power forecasting requirements to support microgrid and island systems with respect to stability and power quality. More specifically, Martinez-Anido et al. (2016) evaluated the value of solar forecast improvements for the Independent System Operator – New England.

- At the power plant level, Marcos et al. (2013) described the benefits of power prediction to optimize a storage system that attenuates power fluctuations in large PV power plants.
- Almeida, Perpiñán, and Narvarte (2015) explored the skill of a nonparametric method to predict the AC power output of PV power plants.
- Wittmann et al. (2008) and Kraas, Schroedter-Homscheidt, and Madlener et al. (2013) used case studies to show the economic benefit of supplying direct normal irradiance (DNI) forecasts for the optimized operation strategies of CSP plants.
- Schroedter-Homscheidt et al. (2013) evaluated the aerosol forecasting requirements for forecasts of concentrating solar electricity production.
- Law et al. (2014) reviewed different DNI forecasting methods and their applications to yield forecasting of CSP plants. In a later publication, Law, Kay, and Taylor (2016) reviewed the benefits of short-term DNI forecasts for the CSP technology.
- Hirsch et al. (2014) specifically evaluated the use of 6-hour forecasts (nowcasting) to operate CSP plants.

In a broader context, different solar radiation forecasting approaches, targeted at various time horizons, have been developed using different input data and data processing methods. In the IEA PVPS context, the state of the art of solar forecasting has been addressed in a report for Task 14 (Pelland et al. 2013). A nonexhaustive list includes methods based on:

- Statistical inference on ground-observed time-series (Huang et al. 2013; Lonij et al. 2013; Voyant et al. 2014; Boland and Soubdhan 2015; Graditi, Ferlito, and Adinolfi 2016)
- Use of cloud motion vectors (CMVs) and other cloud advection techniques on data from all-sky cameras and satellite imagery (Hammer et al. 1999; Perez et al. 2010; Chow et al. 2011; Marquez and Coimbra 2013; Quesada-Ruiz et al. 2014; Schmidt et al. 2016; Lee et al. 2017; Arbizu-Barrena et al. 2017; Miller et al. 2018)
- Forecasts based on numerical weather prediction (NWP) models (Mathiesen and Kleissl 2011; Lara-Fanego et al. 2012; Pelland, Galanis, and Kallos 2013; Ohtake et al. 2013; Perez et al. 2013; Jimenez et al. 2016a; Jimenez et al. 2016b) or even hybrid techniques (Marquez and Coimbra 2011; Marquez, Pedro, and Coimbra 2013; Perez et al. 2014; Dambreville et al. 2014; Wolff et al. 2016; Mazorra Aguiar et al. 2016).

This course provides an overview of basic concepts of solar irradiance forecasting by referring to selected examples and operational models rather than reviewing the state of the art because such reviews can be found elsewhere, including in Lorenz and Heinemann (2012); Inman, Pedro, and Coimbra (2013); Kleissl, Schroedter-Homscheidt, and Madlener (2013), and, for PV applications, in Antonanzas et al. (2016). The evaluations and comparisons of different irradiance forecasting approaches focus on global horizontal irradiance (GHI), with DNI being discussed in less detail. Nevertheless, forecasting and, in particular, evaluation methods apply to DNI to some extent. A focus on DNI forecasting can be found in Schroedter-Homscheidt and Wilbert (2017). The selected examples presented below have been investigated in the context of the International Energy Agency (IEA) Solar Heating and Cooling Programme (SHC) Task 36 and Task 46, and Photovoltaic Power Systems Programme (PVPS) Task 16.

Irradiance is a key driver for solar power output, but other environmental factors—including ambient temperature, air humidity, wind speed, and wind direction—have a nonnegligible impact on the final power yield of the plant. Ambient temperature, for instance, affects the PV efficiency and the thermal regime of CST plants. Humidity might also have some impact on CSP systems.

Similarly, wind speed and especially wind gust prediction are important for preventing strong mechanical loads in tracking systems; therefore, the forecasting of such other ancillary factors will provide tangible benefits for the effective operation of power plants. Forecasts of these ancillary variables, however, are not discussed here.

### 1.1 Overview of Solar Irradiance Forecasting Methods

Depending on the specific application and requirements regarding forecast horizon and spatiotemporal resolution, different forecasting methods are customarily used. From short to long forecasting horizons, the most important solar forecasting methods are the following:

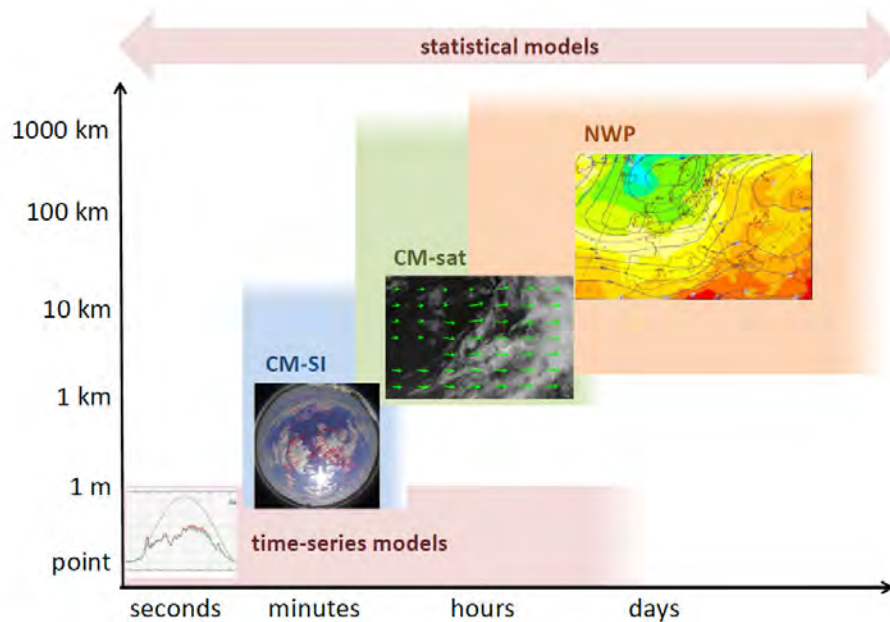
- **Intrahour forecasts with high spatial and temporal resolution:** These require on-site observations of irradiance and/or cloud conditions that are processed using statistical methods and, more recently, artificial intelligence and machine learning models, such as neural networks, as discussed in Section 3. Those that are based on solar irradiance measurements and, for instance, conventional autoregressive techniques might provide meaningful forecasts even up to a few hours ahead under relatively stable sky conditions; however, these methods rarely have good skill under variable sky conditions, given the chaotic behavior of the cloud system and the limited information contained in point-wise observations. In these cases, the local distribution of clouds, as gathered by one or more ground-based sky imagers, might enhance the forecast skill. This cloud-related information allows for the generation of solar irradiance forecasts with a temporal resolution on the order of a few minutes and a spatial resolution from 10–100 m covering a few square kilometers around the sky imagers. The typical forecast horizon of these systems is from 10–20 minutes, depending on the cloud height and speed.
- **Forecasts up to 4 hours ahead:** These are conventionally derived by extrapolating the cloud locations into the future using CMV techniques based on satellite imagery, and they are often referred to as *nowcasts*. The typical spatial resolution is from 1–5 km for the current generation of geostationary satellites, with forecast updates every 10–30 minutes.
- **Intraday and day-ahead forecasts:** These are based on NWP models, which typically offer higher performance for forecast horizons more than several hours and up to several days ahead. These models predict the evolution of the atmospheric system, including the formation, advection, diffusion, and dissipation of clouds. They are based on a physical description of the dynamic processes occurring in the atmosphere by solving and parameterizing the governing system of equations, and they depend on an observed set of initial conditions; see Section 2.2 for details. Current global NWP models cover the Earth with a spatial resolution from approximately 0.1–0.5° and a temporal resolution from 1–3 hours. Regional models, which are sometimes referred to as limited area models or mesoscale models, have a spatial resolution of a few kilometers and an intrahour temporal resolution in the area of interest.

An illustration of these different forecasting methods for various spatial and temporal scales is given in Figure 1.

In addition to this broad classification, when historical or near-real-time on-site solar irradiance or PV yield observations are available, these methods can be further improved by combination with machine learning (hybrid methods). For NWP-based methods, particularly, model output



statistics (MOS) techniques are often applied; see, e.g., Yang (2019a) and Yagli, Yang and Srinivasan (2020) These techniques are sometimes referred to as statistical downscaling techniques. These methods learn error patterns by comparing forecasts and observations and use them to reduce the error of the final prediction.



**Figure 1. Illustration of different forecasting methods for various spatial and temporal scales. The y-axis shows the spatial resolution, and the x-axis shows the forecast horizon intended for the different forecasting techniques. CM-SI: cloud motion forecast based on sky imagers; CM-sat: cloud motion forecast based on satellite imagery. Statistical models apply to all forecast horizons.**

## 2 Empirical and Physical Solar Irradiance Forecasting Methods

This section presents empirical and physical solar forecasting methods. Solar irradiance forecasting methods using statistical approaches and machine learning are described in Section 3. The empirical methods introduced here rely on the correlation between the cloud structures, atmospheric conditions, and solar irradiance. When using satellite data to calculate solar irradiance with radiative transfer models, wind fields from NWP models are used for cloud advection. For physical solar irradiance forecasting methods, various NWP models are discussed.

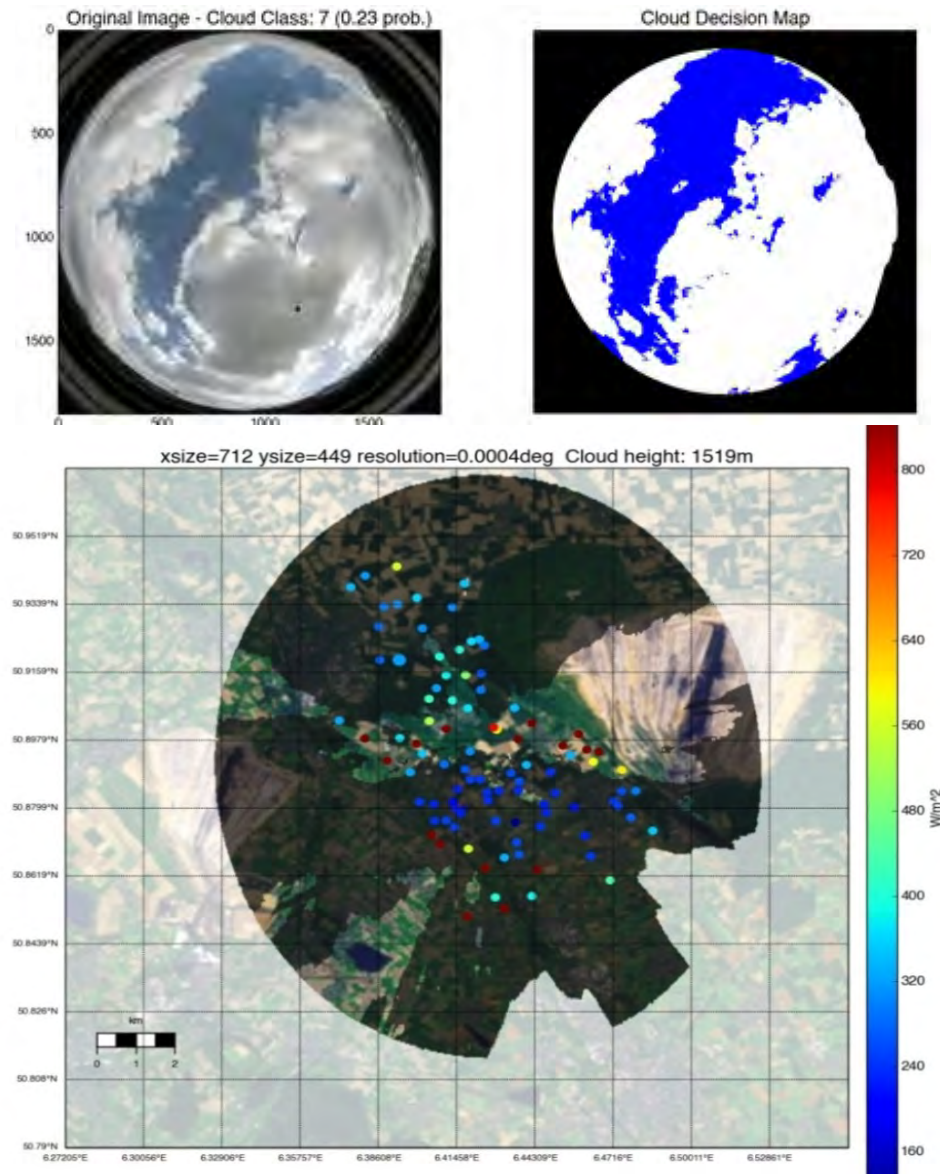
### 2.1 Irradiance Forecasting with Cloud Motion Vectors

At timescales from a few minutes to a few hours, horizontal advection has a strong influence on the temporal evolution of cloud patterns, with the shape of clouds often remaining quite stable. Here, the spatial scale is also extremely important because small-scale cloud structures change faster than larger structures. In these situations, techniques for detecting clouds and their motion trajectories, referred to as CMV techniques, are used to provide valuable information for irradiance forecasting. Obviously, the performance of these forecasting methods degrades as the importance of local processes of cloud formation and dissipation, such as strong thermally driven convection, increases.



CMV-based techniques consist of the following basic steps:

- Images with cloud information are derived from satellite- or ground-based sky-imager measurements.
- Assuming stable cloud structures and optical properties, the CMVs are determined by identifying matching cloud structures in consecutive cloud images.
- To predict future cloud conditions, the CMVs are applied to the latest available cloud image assuming cloud speed persistence.
- Solar irradiance forecasts are calculated from the predicted cloud structures.

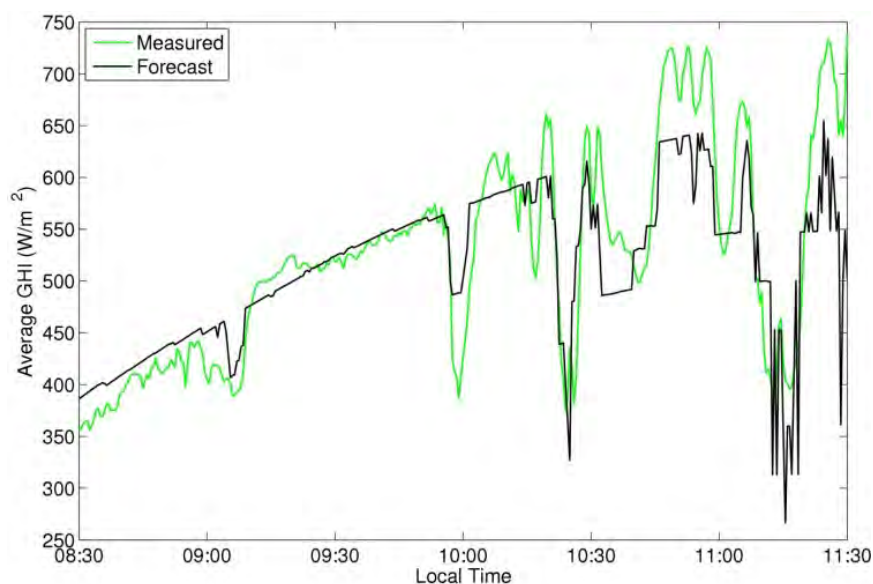


**Figure 2. Cloud information from ASI: (upper left) original image; (upper right), cloud decision map; and (bottom), shadow map with irradiance measurements. Sky image and irradiance measurements were taken in Jülich, Germany, on April 9, 2013, at 12:59 UTC in the framework of the HD(CP)<sup>2</sup> Observational Prototype Experiment (HOPE) campaign (Macke and HOPE-Team 2014). Images from the University of Oldenburg**

### 2.1.1 Forecasting Using Ground-Based All-Sky Imagers

Solar irradiance forecasts at subhourly scales with high temporal and spatial resolutions can be derived from ground-based all-sky imagers (ASIs). Such cameras are installed horizontally and photograph the whole sky above them (see Figure 2, upper left image). ASIs are at times also called whole-sky imagers or sky imagers. The word *imager* is sometimes replaced by the term *camera*, even though they are not strictly identical. In the IEA PVPS Task 16, the term *ASI* is normally used.

ASIs can capture sudden changes in irradiance, which are often referred to as *ramps*, at temporal scales from seconds to minutes (Figure 3). Cloud fields sensed from ASIs or from an assembly of ASIs can be resolved in high detail (e.g., 10 x 10-m resolution), allowing the partial cloud cover on large PV installations to be modeled and forecasted (see Figure 2). The maximum predictable horizon strongly depends on cloud conditions (i.e., cloud height and velocity), and it is constrained by the cloud speed and the field of view of the ASIs. This forecast horizon is typically in the range of 10 minutes, but it can reach 30 minutes in some cases.



**Figure 3. Example of 5-minute-ahead GHI forecast using a sky imager. Location: University of California, San Diego, November 14, 2012. Image from the University of California, San Diego, Center for Energy Research**

Currently, there is no defined standard for sky-imaging hardware, camera calibration, or image-processing techniques. Systems in use include commercially available, low-cost, webcams or surveillance cameras and systems developed specifically for sky imaging; e.g., Urquhart et al. (2015). Most systems use digital RGB (red-green-blue) cameras with fish-eye lenses and therefore consider visible radiation, although some systems rather work with infrared cameras, which are more expensive. In particular, older RGB systems and some infrared cameras use a downward-looking camera that takes photos of an image of the sky that appears on a roughly spherical upward-looking mirror. (This is where the term *imager* comes from.) This concept—unlike the smaller lens or dome of fish-eye cameras—has the disadvantage that the *whole* mirror must be cleaned. Moreover, some older systems use sun-tracked “shadowbands” to prevent the direct sunlight from reaching the camera. This can reduce lens flare-induced saturated areas in

the photos, but the shadowband also covers a noticeable part of the image. Because the required tracking of the shadowband entails higher costs and can lead to system failures, shaded devices have been uncommon in recent years.

The operation of all ASI-based forecasts typically involves these five steps:

1. Take images of the sky and detect which pixels show clouds and which do not.
2. Detect the cloud motion in image series.
3. Geolocate the clouds, including cloud height (if irradiance maps are forecasted).
4. Project the shadow on the ground or determine whether a shadow is at the location of the sky imager (for the current and future cloud positions).
5. Estimate the radiative effect of the clouds on the ideal cloudless DNI, GHI, or GTI (global tilted irradiance).

Machine learning methods are used in some ASI systems for these individual tasks. Some ASI systems do not follow these steps and instead use machine learning methods to directly connect image series to the current and future GHI at the site of the camera (Pierer and Remund 2019).

Cloud detection (which is often also referred to as cloud segmentation) from ASI observations is performed by evaluating different image properties. The red-to-blue ratio (RBR) or multicolor criteria (Kazantzidis et al. 2012) have been used as a main indicator for clouds because of their different spectral-scattering properties (high RBR) compared to clear-sky (low RBR) conditions (Shields, Johnson, and Koehler 1993; Long and DeLuisi 1998). Binary cloud decision maps (Figure 2, top right) can be derived based on threshold procedures. Evaluating the RBR in relation to a clear-sky library (Chow et al. 2011; Shaffery et al. 2020) has proven helpful to account for a nonuniform clear-sky signal over the sky hemisphere that depends on the position of the sun and the turbidity of the atmosphere (Ghonima et al. 2012). Cloud detection is particularly difficult in the circumsolar and solar disk regions because of saturated pixel information that have high RBR values for not only cloudy but also clear-sky conditions. High potential has been seen in machine learning-based segmentation (Hasenbalg et al. 2020).

Detecting cloud motion is the next step to derive irradiance forecasts. For instance, Chow et al. (2011) identified cloud motion based on a normalized cross-correlation procedure—in other words, by maximizing the cross-correlation between shifted areas in two consecutive images. In contrast, Quesada-Ruiz et al. (2014) proposed a discretization method (the sector method) of the cloud image that helps to derive both the direction and speed of clouds. Alternatively, cloud movement can be analyzed by applying optical flow techniques to subsequent images (Lucas and Kanade 1981; Wood-Bradley, Zapata, and Pye 2012). The derived CMVs are then used to cast the observed cloud scenes into the future. For point-wise forecasts at the sky-imager location, information about cloud height is not required because the cloud movement can be parameterized in terms of “pixels per second.” In contrast, for applications requiring mapping cloud shadows, the cloud speed derived using CMVs needs to be expressed in meters per second; this requires knowing the cloud height, which cannot be derived using a single ASI.

The multiple options to determine cloud height include the application of two or more ASIs, ceilometers, distributed radiometers, satellite methods, and NWP data. In particular, the most accurate information on cloud-base height directly above the instrument is currently obtained using ceilometers (Arbizu-Barrena et al. 2015), which are typically employed at airport weather

stations; however, the different clouds seen in a sky image can have different cloud heights, and the ceilometer measures only the cloud height directly above it. Thus, the applicability of ceilometers for this purpose strongly depends on the particular cloud arrangement. Retrieving the cloud-top height from satellite images gives spatially continuous information but shows large uncertainties. Different methods to determine cloud height using combined information from more than one ASI are described by Nguyen and Kleissl (2014) and Wang, Kurtz, and Kleissl (2016). Some of these methods allow deriving different cloud heights for the individual clouds seen in the sky image (Peng et al. 2015). Additionally, the combination of one ASI's CMV in pixels per second with another device's absolute CMV in meters per second can be used to determine the cloud height. Spatially distributed radiometers can be used to derive CMVs in meters per second, as described by Wang, Kurtz, and Kleissl (2016) and Kuhn et al. (2017a). A system using two ASIs can achieve higher accuracy than the application of NWP and distributed radiometers (Kuhn et al. 2018).

Cloud shadow maps at the surface (see Figure 2, bottom) are produced by projecting the forecasted cloud scenes with their assigned height using information about the position of the sun and a digital elevation model. The impact of the projection method on solar forecast accuracy can be large. Local irradiance or PV power measurements can be used to estimate the effect of the clouds on irradiance or PV power. Urquhart et al. (2013) analyzed the frequency distributions of PV power normalized to clear-sky conditions to determine a clear and a cloudy mode and to assign them to shaded and unshaded cells, respectively. Schmidt et al. (2016) used the clear-sky index derived from pyranometer measurements to determine the forecasted all-sky GHI. Similarly, Blanc et al. (2017) used the beam clear-sky index determined from the last 30 minutes of pyrhemometer measurements to derive the cloud transmittance. Additional information on cloud type in the monitored scene indicates cloud optical thickness and cloud height and can be obtained with cloud classification algorithms or by using infrared and thermographic sky imagers. Ghonima et al. (2012) proposed a method to differentiate thin and thick clouds for various atmospheric conditions using a clear-sky library. Gauchet et al. (2012) proposed using a regression model in combination with a clear-sky model to estimate the surface solar irradiance from segmented sky images with information about clear-sky areas; bright and dark clouds; circumsolar area; and solar disk. This specific segmentation is made to optimally accommodate various luminance thresholds.

Instead of using only one or a few ASI systems, networks of approximately 10 or more ASIs can be created to increase the spatial coverage, the forecast horizon, and the accuracy of observations. The combination of several ASIs can provide a more accurate 3D reconstruction of the cloud field (Mejia et al. 2018). Also, the combination of several ASI-derived irradiance maps or intermediate results (e.g., segmentation and cloud height) can be used to improve the nowcasts (Blum et al. 2019). In addition to irradiance nowcasting, ASIs have many other applications that are relevant to meteorology and solar energy. Deriving GHI and/or DNI from sky images is discussed by Schmidt et al. (2016), Chauvin et al. (2018), Kurtz and Kleissl (2017), and Gauchet et al. (2012). Estimating the sky radiance distribution is also possible (Chauvin et al. 2015). Further, the aerosol optical depth (AOD) can be retrieved from ASIs as well (e.g., Olmo et al. 2008 and Kazantzidis et al. 2017).

Another camera-based nowcasting method uses so-called shadow cameras installed at elevated positions, such as on towers or mountains, that take images of the ground around their position

(Kuhn et al. 2017a). In such photos, cloud shadows can be detected, and the brightness of the pixels contains information on the irradiance at the pixel of interest. Unlike ASIs, these systems have the advantage that no modeling of the clouds is required to obtain irradiance maps; however, the development of shadow-camera systems is still in an early phase compared to ASI systems.

### 2.1.2 Satellite-Based Forecasts

Forecasts up to approximately 6 hours ahead require wide-area observations of cloud fields. For example, assuming a maximum cloud velocity of 160 km/h, a region of approximately 2,000 km by 2,000 km would need to be monitored if the goal is to track any arriving cloud 6 hours ahead. Satellite data with broad coverage are appropriate sources for these horizons.

Cloud and irradiance information from satellite images can be derived by a variety of methods. In principle, all of them can be applied to cloud predictions using CMVs to obtain forecasts of solar irradiance. In addition, multiple methods exist to derive CMVs, as described in Section 8.2.1.1 for ASIs. Methods to calculate CMVs from satellite images have also been developed and are routinely used in operational weather forecasting, where CMSs are used to describe wind fields at upper levels in the atmosphere in NWP models.

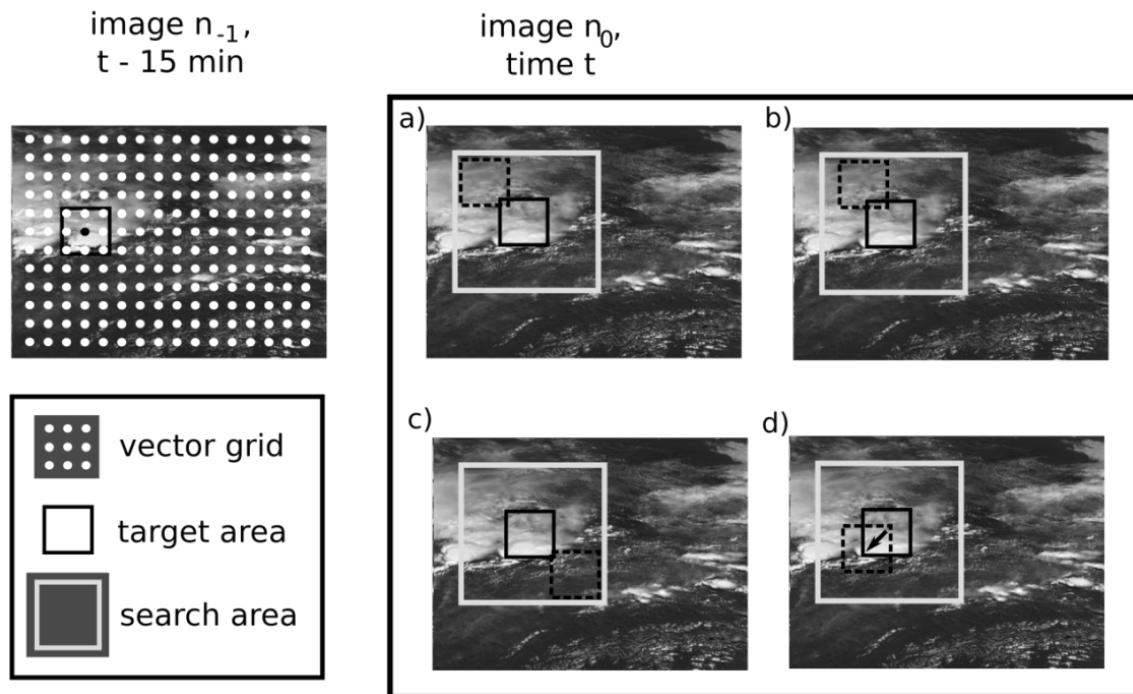
Satellite-based nowcasting schemes for solar irradiance forecasts have been developed mostly in the past decade based on CMVs or sectoral cloud tracking (Hammer et al. 2003; Schroedter-Homscheidt and Pulvermüller 2011). The satellite-based forecasting scheme from the University of Oldenburg in Germany (Lorenz, Hammer, and Heinemann 2004; Kühnert, Lorenz, and Heinemann 2013) is introduced here as an example of such a system. It uses images of the geostationary Meteosat Second Generation (MSG) satellites. The semiempirical HELIOSAT method (Hammer et al. 2003) is applied to obtain information about clouds and irradiance. A characteristic feature of the method is the dimensionless cloud index, which provides information about cloud transmittance.

CMVs are derived by identifying corresponding cloud patterns in two consecutive images (Figure 4). Rectangular areas—the “target areas”—are defined with an approximate size of 90 km by 90 km. This is large enough to contain information about temporally stable cloud structures and small enough that cloud motion for this area can be described by a single vector. Mean square pixel differences among target areas in consecutive images ( $n_0$  and  $n_{-1}$ ) are calculated for displacements in all directions (Figure 4, a–c). The maximum possible displacement (“search area”) is determined by the maximum wind speeds at typical cloud heights. The displacement that yields the minimum mean square pixel difference for a given target area is assigned as a motion vector (Figure 4, d). The derived motion vectors are applied to the cloud index image,  $n_0$ , to predict future cloud conditions. A smoothing filter is applied to the predicted cloud index image to eliminate randomly varying small-scale structures that are hardly predictable. Finally, solar irradiance is derived from the predicted cloud index images using the HELIOSAT method.

The SolarAnywhere short-term forecasting scheme (Perez and Hoff 2013) for the United States is based on Geostationary Operational Environmental Satellite (GOES) imagery and follows a similar approach to detect cloud motion. It is also based on a semiempirical cloud index method. In parallel, Solargis has developed a CMV short-term forecasting scheme that is run under the principles just described but incorporates a multiresolution

treatment of cloud structures. Another method—presented by Schroedter-Homscheidt and Pulvermüller (2011)—discriminates between tracking optically thin cirrus and tracking optically thick cumulus or stratus with respect to the need for increased accuracy in direct irradiance nowcasting aimed at concentrating technologies.

Müller and Remund (2014) proposed a method that combines cloud index values retrieved from MSG satellites with wind fields from a NWP model. The wind fields are predicted with the Weather Research and Forecasting (WRF) model (Skamarock et al. 2005) with hourly resolution and are applied to the forward propagation of the observed cloud patterns from the satellite imagery. Information about the height of the monitored clouds is needed to determine the corresponding NWP model's pressure level. Müller and Remund (2014) assumed fixed cloud heights for this purpose. An advantage of the application of NWP wind fields over satellite-derived CMVs is the potential to describe changes in the direction and speed of cloud movement during the extrapolation process.



**Figure 4. Schematic representation of the CMV derivation using satellite images. Images reproduced from Kühnert et al. (2013)**

A method for satellite-based short-term forecasting using a physical cloud and irradiance retrieval scheme was introduced by Miller, Heidinger, and Sengupta (2013) and Miller et al. (2018). The method processes GOES satellite observations with the National Oceanic and Atmospheric Administration (NOAA) Pathfinder Atmospheres Extended (PATMOS-x) retrieval package (Heidinger et al. 2014), which is a stand-alone radiative transfer code, and combines them with wind field data from the Global Forecast System (GFS) model. Cloud properties are retrieved with PATMOS-x in a first step. Next, the cloud fields are advected using GFS winds at the vertical level matching the cloud-top height as retrieved from PATMOS-x. Finally, solar irradiance at the surface is calculated with radiative transfer calculations using predicted cloud properties and additional atmospheric parameters.



Another satellite-based irradiance scheme that is based on cloud physical properties (CPPs) is the Spinning Enhanced Visible and InfraRed Imager (SEVIRI) CPP method, which uses the SEVIRI instrument onboard the MSG satellites. The method is based on advecting the cloud properties and is used to forecast both GHI and DNI. Details about the CPP surface solar forecasting algorithm and its evaluation are presented by Wang et al. (2019). The forecast horizon of this method is from 0–4 hours at a 15-minute temporal resolution. The forecast has been tested over the Netherlands at a spatial resolution of approximately 4 km by 6 km. CMVs are derived from three cloud properties (cloud top height, cloud optical thickness, and particle effective radius) with a method adapted from the weather radar precipitation forecast system of the Royal Netherlands Meteorological Institute (KNMI). The advected cloud properties are used as input for the CPP-SICCS (Surface Insolation under Clear and Cloudy skies derived from SEVIRI imagery) algorithm to calculate the surface solar irradiance.

## **2.2 Irradiance Forecasting with Numerical Weather Prediction**

NWP models are routinely operated by weather services to forecast the state of the atmosphere. Starting from initial conditions that are derived from routine Earth observations from worldwide networks of ground, airborne, and spaceborne sensors, the temporal evolution of the atmosphere is simulated by solving the equations that describe the physical processes occurring in the atmosphere. Such physical modeling is the only feasible approach when there is little correlation between the actual observations and the forecasted values, which is typically the case for time horizons longer than approximately 5 hours ahead. A comprehensive overview of NWP modeling was given by Kalnay (2003).

Global NWP models predict the future state of the atmosphere worldwide. To determine the initial state from which an NWP model is run, data assimilation techniques are applied to make efficient use of worldwide meteorological observations (Jones and Fletcher 2013). These include observations from ground-based weather stations, buoys, and spaceborne sensors (i.e., satellites). The simulation with NWP models involves spatial and temporal discretization, and the resolution of this discretization determines the computational cost of the simulation. In addition, many physical processes occur on spatial scales much smaller than the grid size—including, for example, condensation, convection, turbulence, as well as scattering and absorption of shortwave and longwave radiation. The effect of these unresolved processes on the mean flow at the model's grid size is evaluated with the so-called parameterizations of atmospheric physics. They include interactions of the land and ocean with the atmosphere, vertical and temporal development of the planetary boundary layer, cumulus triggering and cloud microphysics, as well as shortwave and longwave radiation. The physical parameterizations are a key component of the prediction with NWP models. They bridge the small-scale and large-scale processes, and they make possible the convergence of the numerical routines that solve the physical equations. Today, global NWP models are run by approximately 15 national and international weather services, and their resolutions range from approximately 10 km to approximately 50 km. The temporal resolution of the global model outputs is typically 1 or 3 hours, and their forecasts are normally updated every 6 or 12 hours.

Mesoscale or regional models cover only a limited area of the Earth. They take the initial and lateral boundary conditions from a previous global NWP model run and bring the spatial and temporal grid of the global NWP model down to a finer resolution. Weather services typically operate mesoscale models with a spatial resolution ranging from 1–10 km, and they provide



hourly forecasts, though higher resolutions are feasible. Compared to global models, the higher spatial resolution of mesoscale models allows for explicit modeling of small-scale atmospheric phenomena.

For irradiance forecasting, the parameterizations of radiation transfer and cloud properties are of special importance. Larson (2013) compared the respective model configurations with respect to GHI for four operational NWP models, including the Integrated Forecast System (IFS) of the European Center for Medium-Range Weather Forecasts (ECMWF) and the Global Forecast System (GFS) run by NOAA. In particular, Larson (2013) discussed deep and shallow cumulus parameterizations, turbulent transport, stratiform microphysics and prognosed hydrometeors, cloud fraction and overlap assumptions, aerosols, and the shortwave radiative transfer schemes. But Larson (2013) emphasized that “because of the strong feedback and interactions of physical processes in the atmosphere,” other processes might have a significant impact on irradiance forecasting.

Today, most NWP models offer GHI as direct model output, and some also provide forecasts of direct and diffuse irradiance. Although in principle direct model output can be used for solar energy applications, in practice additional post-processing is customarily applied to improve forecast accuracy.

### *2.2.1 Examples of Operational Numerical Weather Prediction Models*

This section describes some examples of NWP models enumerated together with their spatial resolutions and output time intervals. In particular, it highlights cloud fraction parameterizations and radiation schemes. Additionally, specific references are provided with respect to the application and evaluation of irradiance forecasts in the context of solar energy applications. A comparison of GHI forecasts based on these models was described by Lorenz et al. (2016) for Europe and by Perez et al. (2013) for the United States, Canada, and Europe. It should be emphasized that the sample of operational models and applications given here is non-exhaustive; it simply summarizes the research experience and lessons learned from some research completed within the frameworks of the IEA SHC Task 36 and Task 46 and the IEA PVPS Task 16.

The IFS of the ECMWF is a global model currently being operated with a horizontal grid spacing of approximately 12 km and 137 vertical levels for high-resolution deterministic forecasts. Operational output is available with a temporal resolution of 3 hours and up to 6 days ahead, with a higher resolution of 1 hour being accessible in the framework of research projects. The model is cycled every 12 hours. The radiation code is based on a version of the Rapid Radiation Transfer Model for General Circulation Models (RRTMG) that has been specially developed for use in NWP models (Mlawer et al. 1997; Iacono et al. 2008). Cloud-radiation interactions are taken into account in detail by using the values of cloud fraction and liquid, ice, and snow water content from the cloud scheme using the Monte Carlo Independent Column Approximation (McICA) method (Pincus, Barker, and Morcrette 2003; Morcrette et al. 2008). McICA uses a stochastic approach to infer the cloud extinction of shortwave and longwave solar radiation from only a random selection of calculations. The prognostic scheme for clouds and large-scale precipitation is based on Tiedtke (1993). The ECMWF irradiance forecasts were analyzed by Lorenz et al. (2009) with respect to different relevant properties for PV power prediction applications. In addition, Lorenz et al. (2011) proposed and evaluated an approach based on the ECMWF forecasts for regional PV power prediction for improved electric grid integration.

NOAA's GFS is currently being operated at a spatial resolution of approximately 13 km and 64 vertical levels; however, the outputs are provided in a regular latitude/longitude grid with a resolution of 0.25° and 46 levels, with hourly resolution up to 120 hours ahead and 3-hour resolution up to 240 hours ahead. The model is cycled every 6 hours. Model physics related to clouds and radiation were summarized by Larson (2013). Note that cloud fraction is a diagnostic variable in the GFS model in contrast to the IFS model. Mathiesen and Kleissl (2011) compared intraday GHI forecasts of the GFS and IFS forecasts from the ECMWF and the North American Model.

Environment Canada's Canadian Meteorological Centre operates the Global Environmental Multiscale (GEM) model. It is run in different configurations, including a regional deterministic configuration (Mailhot et al. 2006) that generates forecasts up to 48 hours ahead at a 7.5-minute time step and with a spatial resolution of approximately 15 km centered at the grid. Pelland, Galanis, and Kallos (2013) investigated solar irradiance and PV power forecasting with post-processing applied to the high-resolution GEM forecasts.

The mesoscale (or regional) WRF model (Skamarock et al. 2005) was developed in the framework of a long-term collaborative effort of several institutes led by the National Center for Atmospheric Research (NCAR) in the United States. Now it is a community model, meaning it is publicly and freely available and can receive contributions from all participants. The WRF model is nonhydrostatic, has multiple nesting capabilities, and offers several schemes for each different parameterization of the atmospheric physical processes. This allows the WRF model to be adapted to widely different climate conditions and different applications over virtually any region of interest. The shortwave radiation parameterization usually runs the Dudhia (1989) scheme; however, the latest version of the WRF model includes up to eight different shortwave parameterization schemes (v. 3.9, 2017). This includes the RRTMG radiative scheme already mentioned for the ECMWF's IFS model but also other advanced and research-class radiative models, such as the New Goddard shortwave radiation scheme of the WRF model (Chou and Suarez 1999), the NCAR Community Atmosphere Model (Collins et al. 2004), the Fu-Liou-Gu model (Gu et al. 2011), and the Fast All-sky Radiation Model for Solar applications (FARMS) model (Xie, Sengupta, and Dudhia 2016). The user can select any of these schemes. The current WRF model's cloud fraction schemes are diagnostic. The impact of the resolved topography on the downward solar radiation can be optionally included in the computations. The direct aerosol impact can also be modeled using built-in climatologies or inputs from the user.

The ability of the WRF model and its precursor, the fifth generation of the Penn State University/NCAR Mesoscale Model (MM5), to produce solar radiation forecasts have been evaluated (Guichard et al. 2003; Zamora et al. 2003; Zamora et al. 2005; Ruiz-Arias et al. 2008; Wen et al. 2011). More recently, and mostly toward solar energy applications, the WRF model has been extensively evaluated. For instance, within the framework of the IEA SHC Task 36 and Task 46, Lara-Fanego et al. (2012) evaluated 3-day-ahead hourly and 10-minute WRF model forecasts of GHI and DNI in Spain. Perez and Hoff (2013) conducted a benchmarking study of multiple NWP models, including the WRF model, at a number of European and North American radiometric sites. Lorenz et al. (2016) compared the GHI predictions of multiple models, including the WRF model, and various model configurations against irradiance measurements in Europe. Many other studies from the last few years have evaluated the model over different worldwide regions, including Isvoranu and Badescu (2015) in Romania; Zempila et al. (2016) in

Greece; Aryaputera, Yang, and Walsh (2015) in Singapore; He, Yuan, and Yang (2016) in China; Lima et al. (2016) in Brazil; Gueymard and Jimenez (2018) in Kuwait; and Sosa-Tinoco et al. (2016) in Mexico.

Other studies have analyzed the causes of model errors, and some studies have proposed improvements. For instance, Mathiesen, Collier, and Kleissl (2013) proposed a direct cloud assimilation technique tailored for the WRF model to improve its representation of clouds along the California coastline for improved solar radiation forecasts. Ruiz-Arias et al. (2013) performed surface clear-sky shortwave radiative closure intercomparisons of various shortwave radiation schemes, including the RRTMG, Goddard, and Dudhia models, in which RRTMG showed the highest performance, whereas some deficiencies were found in the the Goddard radiative scheme. A correction for these deficiencies was proposed by Zhong, Ruiz-Arias, and Kleissl (2016). Ruiz-Arias, Dudhia, and Gueymard (2014) proposed a parameterization of the shortwave aerosol optical properties for surface direct and diffuse irradiances assessment. And Ruiz-Arias et al. (2015) described problems with WRF when simulating convective clouds in the Iberian Peninsula and highlighted the need for a dedicated shallow cumulus scheme to reduce model biases.

An important milestone in the use of the WRF model for solar radiation applications has been the recent development of WRF-Solar, a dedicated suite of WRF model parameterizations for solar radiation forecasting (Deng et al. 2014; Ruiz-Arias, Dudhia, and Gueymard 2014; Thompson and Eidhammer 2014) within the U.S. Department of Energy’s Sun4Cast project (Haupt et al. 2016). Some of these improvements, and others, have been summarized by Jimenez et al. (2016b). Moreover, the Sun4Cast project has contributed to the development of the Multisensor Advection Diffusion nowCast (MADCast) system (Descombes et al. 2014), which is a particular configuration of the WRF model for fast assimilation of satellite reflectance images. That configuration can be used to obtain a proxy field to cloud fraction that can be subsequently advected in WRF and used to compute solar radiation nowcasts. Lee et al. (2017) presented a comparative evaluation of WRF-Solar, MADCast, and satellite-based forecasts and found that WRF-Solar performed generally well at predicting GHI under challenging situations in California.

The WRF model is operated for solar irradiance forecasting at several public and private entities, including Solargis (Slovakia); Meteotest (Switzerland); GL-Garrad Hassan (Mathiesen, Kleissl, and Collier 2013); the Atmospheric Sciences Research Center of the University of Albany as part of the operational air quality forecasting program; and AWS Truepower, a UL company in the United States.

The High Resolution Limited Area Model (HIRLAM) is a hydrostatic regional NWP model operated by several national meteorological services in Europe, including the Spanish National Weather Service and the Danish Meteorological Institute. The Spanish National Weather Service runs HIRLAM four times daily in three spatial configurations (one covering Europe at a resolution of 16 km and two covering Spain and the Canary Islands at a resolution of 5 km) with 40 vertical levels. The Danish Meteorological Institute runs its highest-resolution HIRLAM model, “SKA,” for an area covering Northwestern Europe with a grid size of  $0.03^\circ$  ( $\approx 3$  km) and 65 vertical levels. HIRLAM uses the clear-sky irradiance scheme of Savijärvi (1990) and the cloud scheme of Wyser, Rontu, and Savijärvi (1999). The nonhydrostatic HIRLAM ALADIN

Regional Mesoscale Operational NWP In Europe (HARMONIE) regional model is being run experimentally by the Spanish National Weather Service daily over Spain and the Balearic Islands at a resolution of 2.5 km, with 65 vertical levels. Closely related to HARMONIE and HIRLAM is the STRÅNG mesoscale model, which provides hourly nowcasts of GHI, DNI, erythema ultraviolet, photosynthetically active radiation, and sunshine over Scandinavia at 2.5-km resolution.

The German Weather Service (Deutscher Wetterdienst, or DWD) has an operational model chain consisting of the global model ICON (ICOsahedral Nonhydrostatic); the ICON-EU, which is the nested European regional model; and the regional model for Germany called COSMO-D2. The horizontal resolution of the ICON model is 13 km, and it has 90 model layers extending to 75 km (Zängl et al. 2014). The ICON-EU is nested via a two-way interaction. It has a horizontal model resolution of approximately 7 km and 60 model layers up to a height of 22.5 km. The third model is the regional nonhydrostatic COSMO-D2 model, which has a horizontal resolution of 2.2 km and 65 vertical levels. The COSMO model was developed by the Consortium for Small-scale Modeling (COSMO), which consists of the national meteorological services of Germany, Greece, Italy, Poland, Romania, Russia, and Switzerland. The COSMO-D2 (Baldauf et al. 2011) is a model for short-term forecasts of +27 hours, except for the 03:00 UTC run, which has a forecast length of +45 hours. The radiation scheme in the ICON and ICON-EU is currently called once per hour; however, the radiative transfer scheme of Ritter and Geleyn (1992) is called every 15 minutes within COSMO-D2. Thus, the direct and diffuse radiation predictions are available every 15 minutes via direct model output. Using the regional COSMO model, DWD performed a statistical analysis to detect those days with the highest forecast error in Germany, and they identified that NWP forecasts have frequent errors in the presence of low stratus. To address those situations, they proposed a low stratus-detection method that operationally uses post-production.

Another example is NOAA's High-Resolution Rapid Refresh (HRRR) model for the United States, which provides forecasts at 3-km by 3-km resolution with hourly updates.

Finally, the "SKIRON" regional weather forecasting system (Kallos 1997) is operated for solar energy applications at the National Renewable Energy Centre of Spain (Gastón et al. 2009).

### **3 Irradiance Forecasting Based on Irradiance Time Series and Post-Processing with Statistical and Machine Learning Methods**

Statistical learning models are widely used for solar irradiance and power forecasting. The dependence between input variables (predictors) and forecast values (predictands) is established in a training phase by learning from historical data, assuming that patterns in the historical data sets are repeated in the future and thus might be exploited for forecasting. Statistical methods include classical regression methods, such as autoregressive and autoregressive integrated moving average models as well as machine learning or artificial intelligence techniques, such as artificial neural networks (ANNs), k-nearest neighbors, or support vector regression. Coimbra and Pedro (2013a), Diagné et al. (2013), and Yang et al. (2018) provided an overview of different statistical approaches used for solar irradiance forecasting. Voyant et al. (2017) and Sobri

et al. (2018) reviewed the topic with a heavy focus on the use of machine learning methods for solar radiation or power forecasting as well as for post-processing.

Statistical and machine learning models are applied for different purposes in irradiance and PV power forecasting. Pure time-series approaches aim to derive solar irradiance or power forecasts based solely on local measurements without involving any physical modeling (i.e., *time-series approaches with no exogenous input*). They are suitable for forecast horizons from several minutes to several hours ahead.

Statistical and machine learning methods also play an important role in enhancing the output of NWP models and CMV forecasts. Regardless of the physics-based forecasting model used, errors that are partly stochastic and partly systematic will always remain. These errors can be reduced with statistical learning based on historical data sets of predicted and measured irradiance or PV power. Further, statistical learning methods can be employed to derive quantities not included in the native model output. Different terminology is used for this combination of statistical and physical forecasting methods, depending on the perspective of the researchers. The community of statistical modeling and artificial intelligence refers to these models as *statistical models with exogenous input*. Meteorologists commonly use the terms *statistical post-processing* or, more specifically, model output statistics (MOS) in the context of NWP, which is the terminology adopted here.

Section 8.3.1 provides an overview of selected machine learning models, and Section 8.3.2 addresses pure time-series models based on irradiance measurements. Finally, Section 8.3.3 describes the application of statistical and machine learning models for post-processing.

### **3.1 Examples of Machine Learning Models Applied for Forecasting**

The use of state-of-the-art machine learning models is popular in irradiance as well as in PV power forecasting. This section lists several approaches discussed by Winter et al. (2019).

#### **3.1.1 Artificial Neural Networks**

ANNs constitute one of the most versatile machine learning methods and are known for their use in complex tasks such as image or speech recognition (LeCun et al. 1989; Sak and Beaufays 2014).

As described by Bishop (1995), an ANN consists of a fixed number of nodes, called units, that can take on numerical values and are arranged in several layers. The input layer contains one unit for each feature of the data set, whereas the output layer, in case of a single regression problem, is only one unit. The layers between the input and output layers are referred to as hidden layers. The key task is to establish a connection between the nodes by assigning to each unit in one layer the weighted sum of the previous layer's units, and to then apply a nonlinear activation function. In the case of a regression problem, a linear activation function is applied to the weighted sum of the output unit.

By training an ANN on a given set of input and output data, all its weights are adjusted to minimize an error function, typically the mean square error (MSE). This is usually done by back-propagation, an iterative process for calculating the gradient of the error function with respect to each weight (Rumelhart, Hinton, and Williams 1986). In each step, the weights get updated by using a gradient descent optimization algorithm. An alternative option is the method of adaptive

moment estimation, or “Adam,” as described by Kingma and Ba (2015). Instead of calculating the gradient of the error function with respect to the full data set, in each step the weights can be updated only with respect to a subset of the data set (see Bottou 1998; Ruder 2016). The weights can be initialized using a common heuristic described by Glorot and Bengio (2010).

To enable an ANN to learn nonlinear relationships between input and output, a nonlinear activation function must be chosen. For example, the leaky rectified linear unit activation function can be used (Maas, Hannun, and Ng 2013).

### **3.1.2 Extreme Learning Machines**

An extreme learning machine, as proposed by Huang, Zhu, and Siew (2006), is an ANN with a single hidden layer between the input and output layers. Its learning method does not rely on gradient descent. Instead, the weights between the input and hidden layers are chosen randomly. In this way, only the weights between the hidden and output layers need to be determined. Because this is just a linear regression problem, an analytic solution exists, which can be calculated directly without an iterative optimization algorithm. Hence, training the model is considerably faster while maintaining good performance of the model.

### **3.1.3 Gradient Boosted Regression Trees**

Gradient boosted regression trees are an ensemble technique using multiple classification and regression trees (CART) (Breiman et al. 1984). The CART algorithm creates binary decision trees, which means that at each new node the data is split into two parts according to a threshold value. Starting with a root node, which in general contains all training data, the tree grows until some stop condition is reached. The last nodes form the tree’s leaves. Each splitting leads to either another node or a leaf. The leaf contains the class to be predicted. In the case of regression, a leaf returns the mean value of the training samples it contains.

The principle of boosting is described by Friedman (2001). Starting with a single CART tree that is fit to minimize the MSE on the training data, the following trees are trained consecutively so that each new tree predicts the residual error. This residual error is proportional to the gradient of the MSE. By scaling the new tree’s prediction with a step size between 0 and 1 and by adding it to the current ensemble, every new tree aims to further reduce the MSE of the ensemble’s prediction.

### **3.1.4 Random Forest**

A random forest is another technique based on ensembles of CARTs and is presented by Breiman (2001). The ensemble’s prediction is the average over all single tree predictions. Each tree is trained on a bootstrap data set generated by randomly drawn samples with replacement from the original data set (Efron 1979). Further, for each node split, only a random subset of features is considered. By omitting data randomly, the resulting trees become less correlated. This lowered correlation of single trees has been observed to reduce the model error.

## **3.2 Time-Series Models Based on Measurements**

Intrahour or hours-ahead solar irradiance and PV power forecasting with time-series models use recent measurements of irradiance or PV power as a basic input, possibly complemented by measurements of other variables. Examples are the application of a coupled autoregressive and dynamic system model for forecasting solar radiation on an hourly timescale, as described by

Huang et al. (2013); the comparison of ANN and classical time-series models. as by Reikard (2009); and the short-term PV power prediction approach of Bacher, Madsen, and Nielsen (2009). Through their review of machine learning methods, Voyant et al. (2017) concluded that although ANN and autoregression-style methods still dominate statistical forecasting, other methods (e.g., support vector regression, regression tree, random forest, and gradient boosting) are increasingly being used. Although the ranking of such methods is complicated by many factors, it generally holds that a multi-model approach results in improvement in forecasting performance (Zemouri, Bouzgou, and Gueymard 2019).

For any statistical model, the selection and availability of appropriate input variables as well as the optimized preprocessing of the data is of critical importance for good forecast performance. Also, the choice of the model configuration (e.g., the ANN architecture or the selection of hyperparameters in machine learning models) is essential. Finally, the setup of the training sample (e.g., the number of days and sites used for the training) has a noteworthy influence on the forecast accuracy. Coimbra and Pedro (2013a) showed the benefits of the application of a generic algorithm to identify the most suitable ANN architecture, preprocessing scheme, and training data.

The advantages and limits of purely statistical approaches are discussed next. High-quality measurements of the actual surface solar irradiance or PV power are the best possible starting point for any forecast. In comparison, the assessment of the initial irradiance conditions (i.e., the irradiance analysis) with an empirical or physical forecasting model shows considerably higher uncertainties. Any physics-based forecasting model has an inherent uncertainty, regardless of the forecast horizon, that is caused by limits in spatial and temporal resolution, uncertainty in input parameters, and simplifying assumptions within the model. Time-series models exploit the autocorrelation in time series of solar irradiance, cloud cover and, possibly, other explanatory variables. For very short-term forecast horizons, forecasts based on accurate on-site measurements and statistical methods reach forecast errors that are smaller than even the NWP analysis errors or the initial errors of irradiances derived from satellite images.

Given the inherent chaotic nature of weather phenomena, any existing autocorrelation decreases as the time lag between time-series instances increases. Hence, the performance of these models is (1) strongly determined by the underlying autocorrelation of each particular weather condition and (2) decreases as forecast lead time increases. For longer forecast horizons, wide-area observations (e.g., those from satellites) or physical models (e.g., NWP models) are required to meet the forecast skill requirements.

Pure time-series approaches are typically applied to forecast horizons ranging from several minutes to a few hours ahead. Evidently, their performance in comparison to other methods strongly depends on the prevailing climate and weather conditions (e.g., the stability of the sky situation), the spatiotemporal resolution of the forecasts, and the models to which they are compared.

In this context, Bacher, Madsen, and Nielsen (2009) compared an autoregressive model for hourly solar power forecasting combined with and without exogenous inputs from a diverse origin. The study was based on PV plants in Denmark, and the authors found that ground-observed data are the most important class of inputs up to approximately 2 hours ahead, whereas the NWP forecast parameters are adequate for next-day horizons. A comparison of pure time-series models with satellite-based CMV forecasts was given by Wolff et al. (2016) for PV



systems in Germany. The authors found that CMV forecasts outperformed the time-series approach for forecast horizons more than 30 minutes ahead for single sites and for forecast horizons of more than 2 hours ahead for the German average.

Further, sky camera imagery-based forecasting methods were demonstrated to be valuable for short-term high-resolution forecasting. Pedro et al. (2018) and Huang et al. (2019) assessed intrahour hybrid forecasting models that combine statistical (or machine learning) methods with information extracted from sky imagery and found substantial improvements. Huang et al. (2019) proposed the conditional autoregressive method of clear-sky index, which can separate and model characteristic weather events through the identification of key condition variables. Based on high-frequency data measured in Australia, it was shown that by adding exogenous forecasts derived from sky imagery, their hybrid model could produce accurate forecasts seamlessly across timescales from 10 seconds to 10 minutes ahead.

### 3.3 Statistical Post-Processing Methods

Statistical post-processing (or machine learning with exogenous input) plays an important role in irradiance and PV power forecasting. Post-processing methods are applied to:

- Reduce model errors by considering unaccounted or partially accounted local and regional effects (e.g., topography and aerosols)
- Combine the outputs of different models
- Derive quantities that are not direct model outputs.

In what follows, various statistical post-processing methods are summarized for the possible applications enumerated.

#### 3.3.1 Model Output Statistics to Reduce Forecast Errors

MOS are widely used to refine the output of NWP models, primarily to account for local variations in weather and surface conditions (Glahn and Lowry 1972). They use measurements and/or climatology for specific locations as a basis to adapt the forecasts. For example, MOS techniques constitute a powerful tool to adapt the results from NWP or satellite-based models to site-specific conditions (Gueymard et al. 2012). For solar irradiance forecasting, satellite-derived values might be used in lieu of ground measurements. The set of predictors consists of NWP output and might be extended by including any relevant information—for example, prior observations or climatological values.

Originally, the term *model output statistics* was associated with the use of regression equations; however, a generalization of this concept now involves other statistical approaches. Lorenz et al. (2009) applied a bias correction MOS based on solar elevation and clear-sky index to ECMWF irradiance forecasts. Kalman filters have also been proposed by Pelland, Galanis, and Kallos (2013) to improve irradiance forecasts of the Canadian GEM model and by Diagné et al. (2014) in the case of WRF model solar irradiance forecasts. Marquez and Coimbra (2011) investigated the application of ANNs to predicted variables from a weather forecasting database, and Gastón et al. (2009) used a machine learning algorithm to enhance SKIRON solar irradiance forecasts. Pierro et al. (2015) proposed a MOS technique to correct WRF-based GHI forecasts by coupling two intermediate MOS consisting of correlations with relative humidity and ANNs, respectively. Other powerful post-processing approaches have been thoroughly reviewed by Yang and van der Meer (2021).

### 3.3.2 Combination of Forecast Model Outputs

Combining the output of different models can considerably increase the forecast accuracy. First, simple averaging is beneficial for models with similar accuracy, exploiting the fact that forecast errors of different models are usually not perfectly correlated (Perez et al. 2013; Lorenz et al. 2016).

Combining methods using more advanced techniques might also account for strengths and weaknesses of the different models for certain situations—for example, by adapting the contribution of each model depending on the weather situation. In particular, they might be applied to establish a forecast consensus covering horizons from several minutes to several days ahead by integrating measurements, climate monitoring, and NWP forecasts. Various approaches to this aim have been proposed. For instance, Lorenz and Heinemann (2012) used a weighted average with weights optimized for each forecast horizon. Sanfilippo et al. (2016) applied a multi-model approach to solar forecasting that uses supervised classification of forecasting evaluation results to select the best predictions from persistence, support vector regression, and diverse stochastic models. Wolff et al. (2016) and Mazonra Aguiar et al. (2016) combined forecasts based on support vector regression machines and ANNs respectively. Yang et al. (2017) used a hierarchical scheme and minimization of the trace of the forecast error covariance matrix. Within the context of the Sun4Cast project, NCAR's DICAST system (Myers et al. 2011, 2012) has been applied to blend multiple solar radiation forecasts. This system—which has already been applied in other forecasting areas, such as transportation, agriculture, and wind energy—consists of a two-step process: (1) a statistical bias correction process using a dynamic MOS and (2) optimization of the model blending weights for each lead time (Haupt et al. 2016).

### 3.3.3 Post-Processing to Derive Additional Quantities

Not all quantities of interest in the context of irradiance forecasting (i.e., GTI, DNI or PV power) are always available as direct NWP output or as a result of CMV forecasts. Post-processing can be applied to derive these quantities. To that aim, statistical or machine learning methods are typically employed, but empirical or physical models are also frequently used to derive the desired quantity from the direct output of the forecasting model.

Although GHI has become a standard output of most NWP models, this was not the case when the field of solar forecasting started to emerge. For example, Perez et al. (2007) proposed an empirical solar radiation forecast model relating sky-cover predictions from the National Digital Forecast Database to the clear-sky index to derive GHI forecasts.

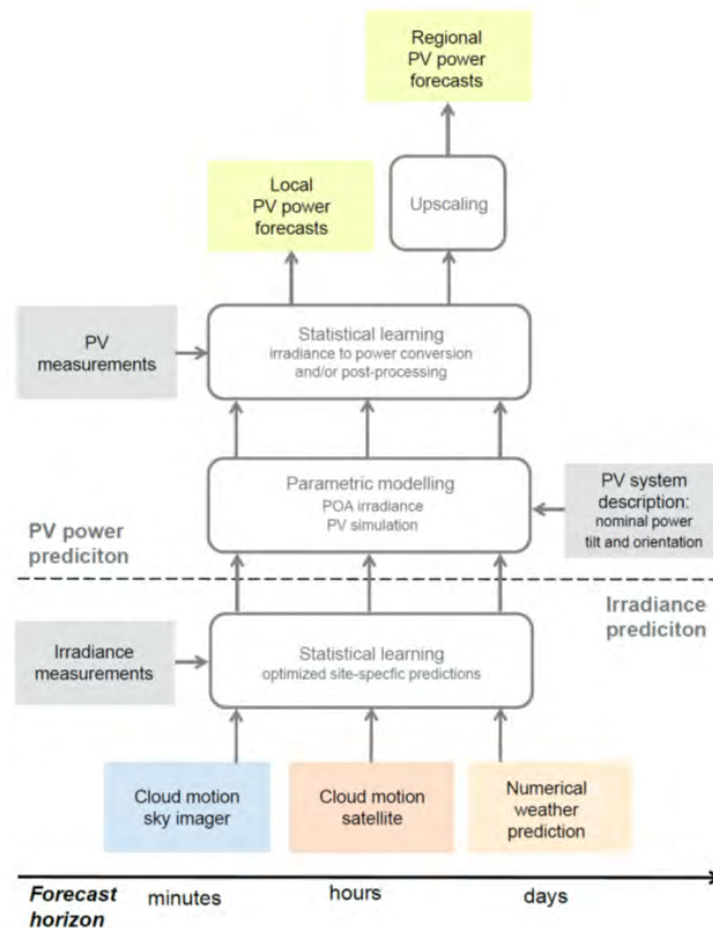
The irradiance components (DHI and DNI) are still not provided as direct output from many irradiance forecasting systems. To derive them from GHI forecasts, several empirical diffuse or direct fraction models can be used, many of which were originally developed for application to measurements or satellite data. These models are also being used in DNI forecasting systems that are based on a GHI forecast (e.g., Schroedter-Homscheidt, Benedetti, and Killius 2016). For DNI forecasts, several physical post-processing approaches have also been proposed, specifically for better consideration of aerosols. Breitzkreuz et al. (2009) proposed a forecasting approach for direct and diffuse irradiance based on the combination of a chemistry transport model and an NWP model in which forecasts of AOD are directly collected from the chemistry transport model outputs. Similarly, Gueymard and Jimenez (2018) used WRF-Solar with hourly inputs of aerosol forecasts from NASA's Goddard Earth Observing System Model 5 (GEOS-5) atmospheric analysis model. Such aerosol forecasts, together with

other remote sensing data (ground albedo and ozone) and NWP parameters (water vapor and clouds), are used as input to radiation transfer calculations to derive the irradiance forecasts. A similar approach was used by Lara-Fanego et al. (2012) to derive DNI from WRF output using aerosol observations from the Moderate Resolution Imaging Spectroradiometer (MODIS) onboard the Terra satellite.

In the context of PV applications, deriving GTI (or plane-of-array [POA]) forecasts is also of interest (see Section 4.1.1).

## 4 PV Power Forecasting and Regional Upscaling

PV power forecasts for a given location or region are important for plant operators, grid operators, and the marketing of produced energy. They are derived from irradiance predictions with physics-based or statistical methods or a combination of both (see Figure 5). The exceptions are time-series approaches for very short-term forecast horizons that are solely based on PV power measurements.



**Figure 5. Overview of basic modeling steps in PV power prediction.**

**Irradiance prediction:** Different forecasting models for different forecast horizons (e.g., cloud-motion sky imager and satellite data, NWP) and combination with statistical learning approaches for optimized site-specific predictions.

**PV power prediction:** Conversion of irradiance to PV power with parametric PV simulation models and/or statistical learning approaches; regional PV power prediction requires upscaling as a last step. *Image reproduced from Lorenz (2018)*

Physics-based parametric modeling involves transposing GHI to GTI (POA) irradiance (Section 4.1.1) and then applying a PV simulation model (Section 4.1.2). For this, information on the characteristics of the PV system configuration is required in addition to the meteorological input data; this includes information on nominal power, tilt, and orientation of a PV system as well as characterization of the module efficiency as a function of irradiance and temperature.

Alternatively, the relationship between PV power output and irradiance forecasts and other input variables can be directly established with statistical or machine learning on the basis of historical data sets including measured PV power. In practice, the approaches are often combined, and statistical post-processing using measured PV power data is applied to improve predictions with parametric simulation models (Section 4.1.3). PV power prediction for grid operators requires forecasts of the aggregated PV power generation for a specified area (i.e., regional forecasts instead of single-site forecasts). These regional predictions are typically obtained by upscaling methods (see Section 4.2).

#### **4.1 Simulation of PV Power Plant Production**

The simplest way to forecast the production of a PV power plant is to apply a PV power simulation model to the forecast of the relevant predicting variables (primarily irradiance, but also environmental temperature and wind speed). Physics-based models explicitly require specific inputs. ANNs or other machine learning models might be more flexible and benefit from a more extensive set of input variables.

##### **4.1.1 Estimating Plane-of-Array or PV Power from Irradiance Forecast**

Because empirical PV simulation models use irradiance or the POA as key inputs, the transposition of GHI into GTI to obtain the POA irradiance is the first modeling step. For example, the PV power forecasting approaches presented by Lorenz et al. (2011) and Pelland, Galanis, and Kallos (2013) involve empirical models to derive the POA irradiance as input for PV simulation models. Unless DNI and DHI are explicitly provided by the forecast model, this first step requires splitting GHI into its direct and diffuse irradiance components. For that purpose, many empirical diffuse or direct fraction models that were originally developed for application to measurements or satellite data can be used (see also Section 4.2). Gueymard and Ruiz-Arias (2015) and Aler et al. (2017) presented an unprecedented worldwide evaluation of 140 of these separation models proposed during the last 60 years.

Next, the direct and diffuse components are projected or “transposed” to the POA. The transposition of the direct irradiance is straightforward, subject only to geometric considerations. The transposition of the diffuse irradiance requires, again, an empirical model for the directional distribution of radiance over the sky, describing anisotropic effects such as horizon brightening and circumsolar irradiance (Perez et al. 1987; Gueymard 1987; Hay 1979). Validation studies of these transposition models are provided by Behr (1997); David, Lauret, and Boland (2013); Gueymard (2009); Ineichen (2011); and Kambezidis et al. (1994). The validation of combined separation and transposition models has been undertaken by Gueymard (2009); Orehounig, Dervishi, and Mahdavi (2014); Lave et al. (2015); and Yang (2016).

#### 4.1.2 PV Power Simulation

In the next step, the POA irradiance is converted to PV output power. Most simple PV simulation models use only the global tilted irradiance on the POA as input and scale it with the PV module array area and efficiency:

$$P_{PV} [kW] = I_{POA} \left[ \frac{kW}{m^2} \right] * A_{PV} [m^2] * \frac{\eta_{mod}[\%]}{100} * f_{Loss} \quad (1)$$

with:

- $I_{POA} \left[ \frac{kW}{m^2} \right]$ , irradiance on the module's POA
- $A_{PV} [m^2]$ , the PV array module area
- $\eta_{mod} [\%]$ , the efficiency of the PV modules
- $f_{Loss}$ , a factor accounting for additional deviations (e.g., deviations that are due to electrical, optical, or thermal losses).

State-of-the-art PV simulation models consider additional influencing factors. Because of optical losses on the module surface, the effective irradiance is lower than the incoming POA irradiance (e.g., Martin and Ruiz 2001). The DC module efficiency depends on the POA irradiance and decreases with increasing temperature. It is secondarily also affected by wind speed and direction (e.g., Beyer et al. 2004). The spectral distribution of irradiance is another influencing factor. Moreover, the conversion efficiency of DC-to-AC inverters is not constant and should be also modeled (e.g., Schmidt and Sauer 1996).

A deeper insight into the modeling of PV power and corresponding variables can be achieved with the tools provided by *pvlb*, a software package for modeling PV systems (Andrews et al. 2014). The choice of input parameters is an issue for using such a model. A natural approach is to use the metadata available for the PV system (e.g., module and inverter specifications, orientation, and peak power); however, this information is frequently missing or erroneous, especially for smaller PV systems. An alternative is learning the parameters from historical data. Here, it is emphasized that measurement issues, plant outages, or shading can impact the estimation of these parameters. Failures of technical components are likely to have large impacts, but the same is true with grid codes, consumption, and curtailments, which are caused by grid operation or electricity market price. To overcome these issues, a robust training method has been proposed by Saint-Drenan et al. (2015).

#### 4.1.3 Statistical and Machine Learning Methods for PV Power Forecasting Based on PV Power or Irradiance Measurements

When PV power measurements are available, the output of a forecast derived with PV power simulation is often adapted to PV power measurements with statistical or machine learning methods to improve the predictions (e.g., Kühnert 2015). When aiming to fit a time series of measured PV power plant feed-in, one needs to account for external effects reducing the production, as described in Section 4.2.

Recently, the direct simulation of PV power with statistical or machine learning models (see Section 3.1) has also gained popularity. This forecasting technique is based on historical data, either by means of a statistical analysis of the different input variables (e.g., autoregressive



moving average or autoregressive integrated moving average) or by using machine learning algorithms that can also handle nonlinear and nonstationary data patterns (Das, Tey and Seydmahmoudia 2018; Ulbricht 2013).

## **4.2 Estimation and Forecasting of Regional PV Power Feed-In**

A very large number of PV systems contribute to the overall PV power generation in a control area or a country. TSOs and utilities require forecasts and estimates of this overall PV power (e.g., as a basis for energy trading).

For many small PV systems, which contribute a large share of the overall feed-in, PV power is not measured with sufficient resolution (e.g., 15-minute or hourly) in many countries; only annual energy totals are available. Consequently, the actual overall PV power feed-in must be estimated using other available data. These estimates of regionally aggregated PV power are important as a starting point for the shortest-term forecasting in real time and as a reference for the statistical training of regional forecasts as well as for evaluations.

Regional PV power feed-in can be estimated using the following data:

- Measurements of the output of PV plants
- Meteorological data (irradiance and temperature)
- Information on the fleet of PV system: coordinates and installed capacity (along with tilt and orientation if available).

An approach frequently applied for both the estimation and the forecasting of regional PV power feed-in is upscaling from a representative set of PV systems in combination with information on the PV fleet. Another approach combines meteorological information (e.g., real-time irradiance from satellite data or irradiance forecasts) with PV simulation using information of the characteristics of the PV fleet. This information is available at different levels in various countries but remains difficult to obtain on a regular basis. Killinger et al. (2018) addressed this issue by collecting the data and applying the method at several thousands of PV system characteristics. As a first step, this approach does not rely on PV measurements, but in practice some-post processing is often applied.

With respect to forecasting of regional PV power, there are additional options. In principle, it would be possible to predict the PV power output for each PV plant in a region (even if PV power measurements were unavailable), and subsequently aggregating the predictions for the whole area (i.e., a “bottom-up” or accumulation approach). This method, however, is characterized by a high computational burden and requires a detailed knowledge of every plant in the area; therefore, it is difficult to achieve, especially for large areas. Nonetheless, two examples of PV-system-based forecasts have been published (Vaz et al. 2016; Carillo et al. 2020).

Finally, if an estimated time series of regional PV power time series is given, forecast providers can also train their models directly to this PV power time series without requiring detailed information of the PV fleet (i.e., a “models input average” approach).

These approaches are introduced in sections 4.2.1, 4.2.2, and 4.2.3.

#### 4.2.1 Upscaling Based on Representative PV Systems

One option for upscaling is to rescale the output of the reference plants to the overall installed capacity in a given region (Lorenz et al. 2011; Kühnert 2015). This approach exploits the strong correlation of the power output of nearby PV systems and allows for the estimation and the prediction of PV power with good accuracy given a sufficient number of reference plants and given that the representative set reflects the basic properties of the total data set (Kühnert 2015; Saint-Drenan et al. 2016). Representation of the spatial distribution of the nominal power and of the PV systems' tilt and orientation is most important in this approach. In operational PV power prediction systems based on this approach, the upscaling is typically performed for small subregions (e.g., down to postal codes) and potentially also for different classes of PV system size in a first step. Then, the estimates for the different subregions and classes are aggregated to the region of interest.

Another approach for upscaling based on representative PV systems uses spatial interpolation methods. Starting with a set of reference power plants, one can interpolate the (measured or predicted) power to any other power plant, assuming that at least the exact coordinates of reference as well as target locations are known. Geosciences offer several methods to conduct that task; see the review by Li and Heap (2014). The most popular of these methods are the simple but robust inverse distance weighting (IDW), and some authors apply the more complex kriging method, at least to interpolate irradiation data (Jamaly and Kleissl 2017; Yang et al. 2013). Because irradiance is the most important variable for PV power production, methods are likely to apply to both measured and predicted power.

One key difference of the two mentioned methods is the feature of convexity: the IDW method is convex and therefore generates interpolated values only in the range from the min to the max input values; on the other hand, kriging is a nonconvex method that can produce results outside the range of input measurements. Considering the aim to estimate many local PV power production values based on a set of measurements from different locations, it seems reasonable to rely on a set of references that is as large as possible when using the IDW method.

Because this robust method is one of the most commonly implemented methods, as exemplified by Saint-Drenan (2011) and Bright et al. (2018), it is briefly described here. The PV power,  $P_j$ , for power plant,  $j$ , is the weighted sum of  $n$  surrounding power plants,  $i$ , where the weights,  $w_{ij}$ , are calculated based on the inverse of distance,  $d$ , between  $j$  and  $i$ , so that the sum of all weights equals one. The exponent,  $u$ , is typically optimized and found to be approximately 1.7–2.0.

$$P_j(t) = \sum_{i=1}^n w_{ij} P_i(t) \quad (2)$$

$$w_{ij} = \frac{d(i,j)^{-u}}{\sum_{k=1}^n d(k,j)^{-u}}, u > 0 \quad (3)$$

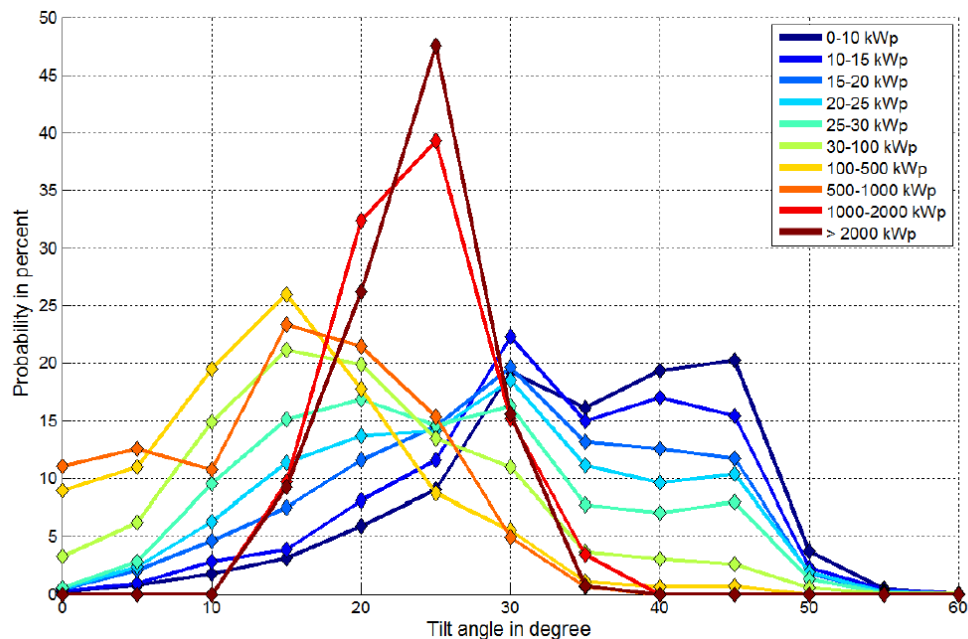
The targets,  $j$ , can alternatively be seen as the center of all installed capacity of an area of interest; however, assuming similar characteristics (e.g., orientation angles) between references and targets is needed for this. Improvements have been observed by accounting for known orientations (e.g., Killinger et al. 2016). Last, an important step is the aggregation of all relevant targets (power plants or areas) for a region of interest. When it comes to unknown or dynamically changing electric grid connections, a new source of uncertainty becomes important. An extensive investigation of more general uncertainties can be found in Saint-Drenan et al. (2016).



#### 4.2.2 Regional Model Based on Statistical Analysis of the Fleet of PV Systems

As mentioned, simulating each single plant installed in a region is not realistic because of the very high computational costs. As shown by Saint-Drenan, Good, and Braun (2017), a realistic alternative is to group all plants with similar characteristics (i.e., orientation angles) and simulate this limited number of groups. It can easily be shown that this computational technique allows for speeding up the calculation without losing information, making this approach viable. This type of implementation requires determining the share of each group of plants—in other words, the share of installed capacity for each class of orientation must be assessed. Several approaches can be followed to obtain this information.

The first approach, which is described by Saint-Drenan et al. (2017), involves conducting a statistical analysis on a subset of the installed PV plants (see Figure 6). The risk here is that the selected samples are not representative of the actual PV system. To address this issue, it is possible to train the distribution of the different classes of PV plants; this option was demonstrated by Saint-Drenan et al. (2019), where a Bayesian approach was used to regularize the training approach. This statistical approach has also been employed in the Copernicus Climate Change Service to generate regional PV power for each region in Europe using the ERA5 reanalysis data set. As described by Saint-Drenan et al. (2018), the parameters of the statistical regional model have been derived from the optimal tilt angle, allowing this model to be implemented anywhere in Europe without the need for a training data set. A comparison of the model output with the European Network of TSOs for Electricity data showed that this approach is well accepted. Note that estimates of the regional power production for each European Union country is calculated operationally each month with this method. The data can be found on the Copernicus Data Store.



**Figure 6. Distribution of the tilt angle of German PV systems for different classes of peak power.**  
*Image from Saint-Drenan et al. (2017)*

### 4.2.3 Forecasting Regional PV Power Based on Averaged Model Inputs

The direct approach, called “model inputs average,” is based on the spatial smoothing of the input features. In this case, the PV power generation of the area is considered to be a virtual power plant, and prediction is made directly at the regional level by using a historical data set of regional PV power (measured or estimated) and input meteorological forecasts aggregated at a smaller spatial scale than the region of interest. The main advantage of this approach is the possibility of obtaining a reliable power forecast without additional details about the installations beyond the total installed capacity of the whole area. The relationship between meteorological forecasts and PV power time series is typically established with machine learning approaches. An overview of the relationship is given by Betti et al. (2020). A prerequisite of this approach is the availability of a time series of actual PV power feed-in, which can be estimated by the upscaling models described in Section 4.2.1. In many countries, estimates of the regional PV power feed-in are published, for example, by grid operators, and they are available to forecast data providers as a basis for training.

## 5 Evaluation of Irradiance and PV Power Forecasts

The evaluation of solar irradiance forecasts provides users with the necessary information about forecast accuracy and helps them choose different forecasting products or assess the risk when using a particular forecast as a basis for decisions. An extensive overview of forecast verification methods was given by Jolliffe and Stephenson (2011). This section addresses the evaluation of deterministic irradiance forecasts that provides an overall indication of the uncertainty of a specific forecast model. Probabilistic solar forecasts assigning uncertainty estimates to each individual forecast value are described in Section 6.

The quality of forecasts is evaluated by assessing their similarity to reference data. Most often, irradiance measurements are used as reference data, which are commonly referred to as ground truth data. Nevertheless, reference data are always affected by a certain degree of uncertainty. Alternatively, satellite-retrieved irradiance values or the output of a detailed physical model might serve as a reference.

The choice of appropriate metrics and concepts for the evaluation of solar irradiance and power forecasts is the subject of ongoing discussions within the solar forecasting community; see, e.g., Hoff et al. (2012a) and Marquez and Coimbra (2013). Recently, Yang et al. (2020) proposed applying the well-established Murphy-Winkler framework for distribution-oriented forecast verification as a standard practice to analyze and compare solar forecasts.

Here, the most standard evaluation methods are outlined, including (1) statistical error measures (Section 5.1.1); (2) comparison to reference models using the skill score parameter (Section 5.1.2); and (3) other important considerations, such as the representation of the observed frequency distribution and the forecast “goodness” as a function of solar position, hour of the day, cloud variability, or even spatiotemporal averaging (Sections 5.2–5.4). These concepts are introduced using examples from an observational data set of hourly pyranometer measurements from 18 weather stations of the German Weather Service from March 2013 to February 2014 (Lorenz et al. 2016) and fo

- High-resolution deterministic global IFS model, operated at the ECMWF, with spatial resolution of 0.125°, 3-hourly outputs, and forecast horizon of 24 hours issued every day at 00:00 UTC
- High-resolution regional HIRLAM SKA model, operated at the Danish Meteorological Institute, with spatial resolution of 3 km, hourly outputs, and forecast horizons from 4–9 hours ahead, issued daily at 00:00 UTC, 06:00 UTC, 12:00 UTC, and 18:00 UTC.

In addition, Section 5.5 addresses the evaluation of regional PV power forecasts in a case study in Italy. Finally, Section 5.6 introduces the concept of “firm power forecasts” as an effective model validation metric to account for the economic value of solar forecasts.

## 5.1 Error Measures

Statistical error measures and skill scores are applied for quantitative forecast evaluation.

### 5.1.1 Statistical Error Measures

Here, the most commonly used error measures based on first-order statistics are presented. The error of a single measurement is given as:

$$\varepsilon_i = I_{pred,i} - I_{meas,i}, \quad (4)$$

where  $I_{pred,i}$  denotes a predicted irradiance value (GHI or DNI), and  $I_{meas,i}$  is the corresponding measured value.

To evaluate the forecast accuracy of the solar power predictions, the root mean square error (RMSE) is commonly used:

$$RMSE = \frac{1}{\sqrt{N}} \sqrt{\sum_{i=1}^N \varepsilon_i^2} \quad (5)$$

where  $N$  is the number of data pairs. The mean square error,  $MSE = RMSE^2$ , is also commonly used. Typically, only daytime values are considered for the evaluation. Relative errors for the irradiance forecast are generally derived by normalization with respect to mean measured irradiance of a given time interval. In contrast, relative errors of PV power forecasts for utility applications are often normalized to the installed power rather than the mean measured value (e.g., Lorenz et al. 2011).

The RMSE can be split into two components: systematic (1) or bias error and (2) stochastic error or standard deviation. The bias is the difference between the mean of the predicted and measured values (systematic error):

$$bias = \bar{\varepsilon} = \frac{1}{N} \sum_{i=1}^N \varepsilon_i \quad (6)$$

A positive bias means the predicted values exceed the measurements on average.

The standard deviation of the errors,  $stderr$ , is defined as:

$$stderr = \sigma(\varepsilon) = \frac{1}{\sqrt{N}} \sqrt{\sum_{i=1}^N (\varepsilon_i - \bar{\varepsilon})^2} \quad (7)$$

The  $stderr$  provides information on the spread of the errors around their mean value and might be further decomposed into one part related to the error amplitude  $[\sigma(I_{pred}) - \sigma(I_{meas})]$  and another part related to the correlation coefficient,  $r$ , of the time series, which is defined as:

$$r = \frac{\sum_{i=1}^n (I_{pred,i} - \overline{I_{pred}}) \cdot (I_{meas,i} - \overline{I_{meas}})}{\sigma(I_{pred}) \sigma(I_{meas})}. \quad (8)$$

Overall, the complete decomposition of RMSE yields:

$$RMSE^2 = bias^2 + (\sigma(I_{pred}) - \sigma(I_{meas}))^2 + 2\sigma(I_{meas}) \sigma(I_{pred}) (1 - r) \quad (9a)$$

or, equivalently, and more simply:

$$RMSE^2 = bias^2 + stderr^2. \quad (9b)$$

Another common measure to assess forecast accuracy is the mean absolute error (MAE):

$$MAE = \frac{1}{N} \sum_{i=1}^N |\varepsilon_i| \quad (10)$$

which is recommended by Hoff et al. (2012a) as a preferred measure, in particular for reporting relative errors.

From a user's point of view, the choice of the most suitable error measure will be based on the impact of forecast errors on their application. MAE is appropriate for applications with linear cost functions (i.e., when the costs caused by inaccurate forecast are proportional to the forecast error). The RMSE is more sensitive to large forecast errors and hence suitable when small errors are more tolerable and larger errors cause disproportionately high costs, which is the case for many applications in the energy market and for grid management issues.

In addition to the computation of these error measures, at least some basic visual analysis is strongly recommended. A direct comparison of measurements and forecasts in scatter plots or 2D histograms and time series is helpful to develop a better understanding of forecast performance.

### 5.1.2 Skill Score and Persistence Forecast Model

Skill score (also referred to as forecast skill) is used to quantify the forecast performance relative to a reference model. The RMSE is normally used for this comparison; other scores such as MAE or MSE are also often used. The skill score is defined as the difference between the score of the reference model and the forecast model divided by the difference between the score of the

reference model and a perfect model; note that a perfect model yields zero RMSE. For RMSE, the skill score,  $SS_{RMSE}$ , is calculated as:

$$SS_{RMSE} = \frac{RMSE_{ref} - RMSE}{RMSE_{ref}}, \quad (11)$$

where  $RMSE_{ref}$  refers to the reference model, and RMSE refers to the investigated forecasting algorithm (Coimbra and Pedro 2013a). The skill score's value of 1 hence indicates a perfect forecast, and a skill score of 0 means that the investigated algorithm has the same RMSE as the reference forecast. A negative value indicates performance that is worse than the reference. Skill scores might be applied for comparisons to a simple reference model and also for intercomparisons of different forecasting approaches (i.e., improvement scores).

In solar radiation forecasting, *persistence* is the simplest and most widely used forecast reference model. The persistence model is a trivial model that assumes that the current situation does not change during the forecasted lead time. Several definitions of persistence exist, including *simple* persistence; *scaled* persistence, which accounts for solar geometry changes; and more-advanced concepts, such as *smart* persistence. The most widely used definitions are presented next.

For day-ahead forecasts, the simplest approach is to assume that irradiance,  $I$  (GHI or DNI), persists during a period of 24 hours, that is:

$$I_{per,24h}(t) = I_{meas}(t - 24h). \quad (12)$$

A more elaborate option for GHI, which produces higher accuracy forecasts, is to separate the clear and cloudy contributions to solar radiation and assume that only the cloudy component (i.e., the random component of GHI) persists during the forecast lead time. The clear component is strongly influenced by the deterministic solar geometry and can be described with reasonable accuracy using a clear-sky radiation model. In such a modeling approach, the persisting magnitude is the clear-sky index,  $K_c$ , calculated from the measured GHI. For forecast horizons of several hours ( $\Delta t$ ) ahead, persistence GHI<sub>per, $\Delta t$</sub>  for time  $t$  is then defined as:

$$GHI_{per K_c, \Delta t}(t) = GHI_{clear}(t) K_c(t - \Delta t). \quad (13)$$

For DNI, a similar approach can be used based on the beam clear-sky index or the Linke turbidity factor (Kuhn et al. 2017b).

In the context of the IEA Task 46 (IEA 2015), the so-called “smart persistence” has been proposed. It consists of increasing the integration time that defines the current conditions commensurately to the forecast time horizon  $\Delta t$ :

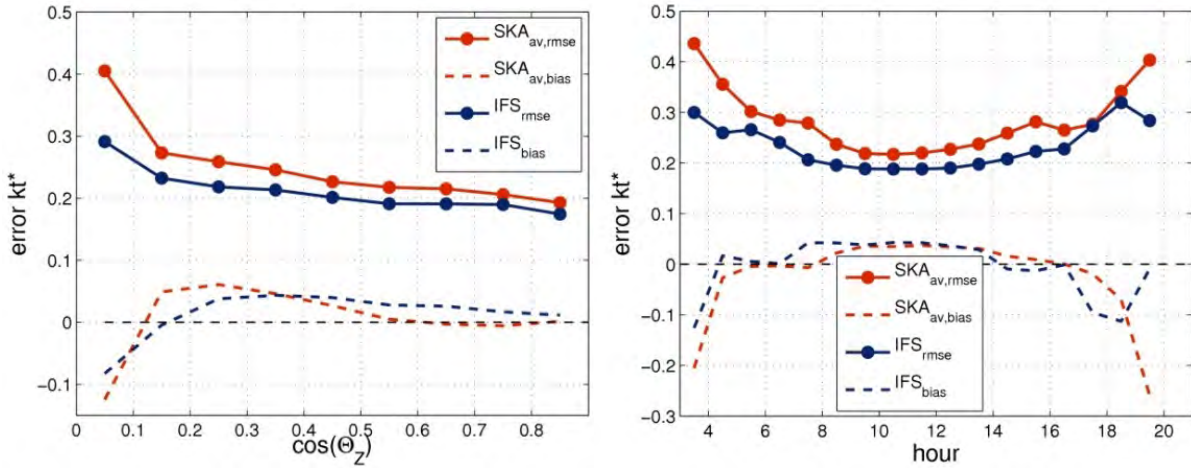
$$GHI_{per smart, \Delta t}(t) = GHI_{clear}(t) \frac{1}{\Delta t} \int_{t-2*\Delta t}^{t-\Delta t} K_c(t') dt' \quad (14)$$

Or, for measurements available in discrete time interval  $\Delta t_{meas}$  :

$$GHI_{per\ smart, \Delta t}(t) = GHI_{clear}(t) \frac{1}{N} \sum_{i=1}^N K_c(t - \Delta t \left(1 + \frac{i}{N}\right)), \quad (15)$$

with  $N = \frac{\Delta t}{\Delta t_{meas}}$ .

Another less-used reference model is based on climatological mean values. Alternatively, combinations of climatology and persistence can be applied as a reference, as recommended by Yang et al. (2020). Further discussion on forecast benchmarking using the skill score and clear-sky persistence is provided by Yang (2019b).



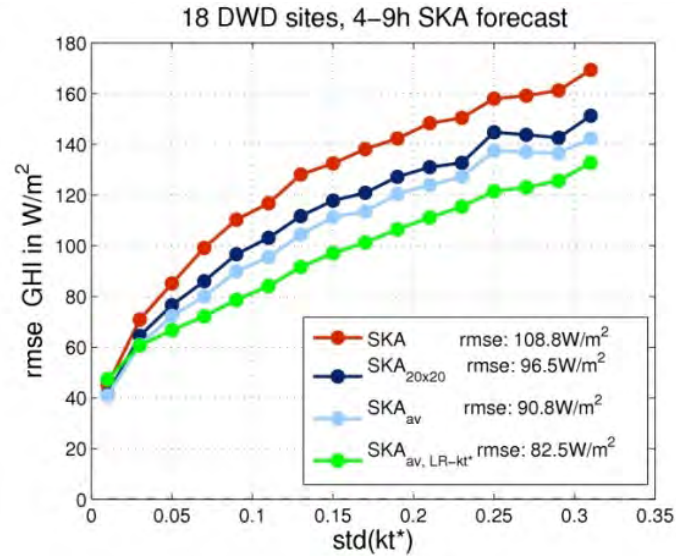
**Figure 7. Clear-sky index (here noted as  $kt^*$ ) forecast error as a function of (left) cosine of solar zenith angle and (right) hour of the day for the forecasts issued by the IFS and SKA NWP models (blue and red lines, respectively). Solid lines show RMSE, and dashed lines show mean bias error. The evaluated period is from March 1, 2013–February 28, 2014.**

## 5.2 Analysis of Forecast Error with Respect to Solar Elevation

A special feature of solar irradiance is its very strong deterministic component, which results from the daily and seasonal course of the sun. This deterministic signal strongly influences the forecast error signal. Hence, to investigate the solar irradiance forecast errors, it is sometimes advisable to evaluate only the nondeterministic part of solar radiation, which is primarily caused by errors in the representation of clouds. To this aim, the analyzed variable is often the clear-sky index forecast error instead of GHI forecast errors.

Figure 7 shows the RMSE and bias of the clear-sky index,  $K_c$ , as a function of the cosine of the solar zenith angle (Figure 7, left) and the time of day (Figure 7, right) for two different NWP model forecasts (IFS and SKA). The two models show similar behavior: RMSE increases with low SZA or, equivalently, during morning and evening hours, as is also the case with the magnitude of bias. This error pattern is very often caused by deficient modeling of the atmospheric transport of radiation for low solar altitudes. This limitation is a well-known flaw of the two-stream schemes used in most NWP models. Other model limitations also exist, such as 3D effects and atmospheric refraction issues whose impact is enhanced for low solar altitudes.





**Figure 8. RMSE of various versions of the SKA forecasts as a function of the standard deviation of measurement-based clear-sky index (here noted as  $kt^*$ ) (red) SKA.**

**Dark blue: Nearest grid point, SKA20 x 20 averaged throughout 20 by 20 grid points.**

**Light blue: SKAav 5-hour sliding mean of the clear-sky index of the forecasts of the average throughout 20 by 20 grid points. Green: SKAav, LR. $kt^*$ : linear regression of the clear-sky index of the forecasts applied to SKAav. The evaluated period is from April 3, 2013–February 28, 2014.**

**Training set: Last 30 days, all 18 DWD sites**

### 5.3 Analysis of Forecast Error with Respect to Cloud Variability and Spatiotemporal Averaging

Forecasts generally show good agreement with measurements for clear-sky periods or even completely overcast days, which basically have a quasi-constant clear-sky index; however, cloud variability strongly impacts solar forecasting accuracy. Hence, considerable deviations from the measurements are typically observed for days with variable cloudiness. An evaluation of the SKA forecast errors as a function of the measurement-derived  $K_c$  variability, here represented by the standard deviation of  $K_c$  throughout 5 hours, is shown in Figure 8. The evaluation also shows this dependence for multiple spatial and temporal averaging configurations of the SKA forecasts. Overall, Figure 8 shows:

1. The forecast error increases with enhanced cloud variability.
2. Spatial and temporal forecast averages result in reduced RMSE values, going from negligible reductions for very stable conditions to large reductions for highly variable conditions.

Regarding the first point, the solar radiation forecast error shows a clear dependency with respect to cloud variability and, more generally, with respect to the cloud conditions. Combining the error trend in the dependence of cloud conditions and the solar elevation has been proposed as an efficient method to reduce the systematic error in NWP model forecasts using a post-processing MOS. In particular, Lorenz et al. (2009) used a polynomial function with  $\cos(\text{SZA})$  and  $K_c$  as independent variables to parameterize the forecast bias error from historical forecasts relative to observations and ultimately to subtract the parameterized error from operational forecasts. This approach has also been adapted and evaluated for other NWP models and different climates. Mathiesen and Kleissl (2011) found improved accuracies when applying that approach to three



different NWP models—GFS, North American Model, and IFS—for stations in the continental United States. Pelland, Galanis, and Kallos (2013) did the same for the Canadian GEM model, and Müller and Remund (2010) for the WRF model forecasts in Switzerland.

Regarding the second point, the rationale of the RMSE decreases when an averaging scheme is applied, and this is explained by the existence of small correlations among the pixels over which the averaging scheme is applied. This leads to random error cancellations during the averaging process. In contrast, for stable conditions, when the correlation among neighboring pixels is very high, the cancellation of random errors is small.

The optimal region size and time interval for RMSE reduction using averaging depends on the correlation structure among neighboring forecasts, both in time and space. Multiple studies have been conducted on this topic. For instance, a detailed evaluation of irradiance forecasts from the Canadian GEM model resulted in a reduction of forecast errors in the range from 10% to 15% when the model outputs were averaged throughout several hundred kilometers (Pelland, Galanis, and Kallos 2013). A similar improvement was achieved with WRF forecasts provided by Meteotest using averages over an area of 50 km by 50 km (Müller and Remund 2010). In parallel, Mathiesen and Kleissl (2011) reported an averaging area of 100 km by 100 km as suitable for irradiance forecasts using the GFS and North American Mesoscale forecast system models. The benefit of horizon-dependent smoothing filters for CMV forecasts was also shown by Lorenz, Hammer, and Heinemann (2004) and by Kühnert, Lorenz, and Heinemann (2013).

The reduction of RMSE by spatial and temporal averaging can be extrapolated to the particular case in which the forecasting model performance is evaluated throughout multiple sites across a wide region (also referred to as *regional* forecast) or for coarser temporal granularities, such as monthly or yearly. In these cases, there is a reduction of random errors with respect to point-wise evaluations that make regional forecasting more accurate than point-wise forecasting. Again, the extent of the reduction depends on the particular correlation levels among the aggregated values in each case. An analysis of regional forecast errors for different region sizes and different forecast models was given by Lorenz et al. (2009); Kühnert, Lorenz, and Heinemann (2013); and Lorenz and Heinemann (2012).

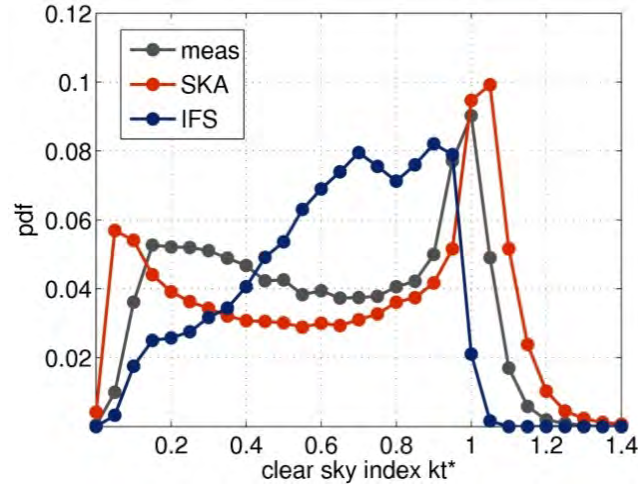
Temporal and spatial averaging can be also considered for ASI-based forecasts. It has been found that in a nowcasting system with four sky imagers during days with many transient clouds, the DNI RMSE for forecasts that are 10 minutes ahead is reduced from 13.0% to 6.5% using averages of 4 km<sup>2</sup> and 15 minutes with respect to pixel-wise forecasts (Kuhn et al. 2017c).

Despite the positive impact of spatiotemporal averaging on reducing the RMSE of a forecast, there is a negative effect that adversely impacts the frequency distribution of forecasted data because averaging reduces extreme forecasted values and distorts the original frequency distribution of the forecast data. Consequently, forecast averages should be used only when the forecast frequency distribution is not critical.

#### **5.4 Analysis of the Frequency Distributions of Forecasted Values**

The ability of a model to reproduce the observed frequency distribution of both solar irradiance and clear-sky index is a required property for some applications. In addition, it can provide insights about potential problems in the forecast model.

Figure 9 shows the probability density function (PDF) of the clear-sky index,  $K_c$ , for forecasts issued by the SKA and IFS NWP models, as in Figure 7, and the actual PDF obtained from observations. These plots show that the SKA model systematically overpredicts clear-sky situations and underpredicts overcast conditions. Consequently, intermediate situations are underrepresented. On the other hand, the IFS model underrepresents very clear and very cloudy conditions and overrepresents intermediate situations. Although this gives insightful information about the forecast performance, the similarity of the distribution functions of measurements and forecasts does not guarantee an accurate forecast because it does not include information regarding the correct timing of the modeled events.



**Figure 9. PDF of the clear-sky index (here noted as  $kt^*$ ) derived from measurements (gray), SKA model forecasts (red), and IFS model forecasts (blue). The evaluated period is from March 1, 2013–February 28, 2014;  $\cos(SZA) > 0.1$ .**

A quantitative evaluation of the agreement between the observed and forecasted distribution functions can be done using the Kolmogorov-Smirnov integral (Espinar et al. 2008; Gueymard 2014), which is usually applied to distribution functions of GHI or DNI rather than to  $K_c$  (Beyer et al. 2009; Perez and Hoff 2013).

### 5.5 Analysis of Regional Forecasts

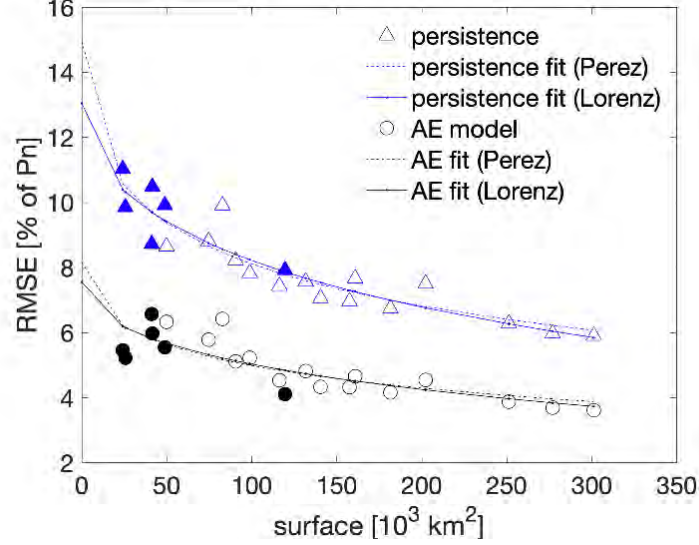
Regionally aggregated forecasts of PV power, typically derived through upscaling (see Section 4.2), are required by grid operators. Regional forecasts show much lower uncertainties than single-site forecasts. By enlarging the footprint of the forecast region of interest, forecast errors are reduced (Hoff and Perez 2012; Lorenz et al. 2009, 2011; Fonseca et al. 2014; David et al. 2016; Saint-Drenan et al. 2016). This phenomenon, which is called the smoothing effect, is related to the correlation between the forecast errors at different locations. The larger the region, the more locations with different irradiance variability are included, and thus solar forecast errors of the different sites are less correlated. This subsequently leads to a higher accuracy of the regional PV power forecasts.

An example is shown in Figure 10, which depicts the day-ahead forecast accuracy that can be reached in Italy by predicting the PV generation of different control areas: The adopted model corresponds to an upscaling method using averaged model inputs and forecasting the power generation at market zone level directly (Betti et al. 2020).

In addition, a measure of PV power variability is displayed in Figure 10. With  $P(t)$  denoting the PV power output at time  $t$ , the change in PV power for a given time step,  $\Delta t$ , is defined as:

$$\Delta P_{\Delta t} = P(t) - P(t - \Delta t). \quad (16)$$

Hourly values and a time step  $\Delta t$  of 24 hours are specifically considered in Figure 10.



**Figure 10. Smoothing effect over Italy: RMSE of regional forecasts (circles) and persistence (triangles) as a function of the area size for the market zones in Italy (full circles/triangles) and for areas merging several adjacent market zones (empty circles/triangles).**

PV power variability in each zone is defined as the standard deviation,  $\sigma(\Delta P_{\Delta t})$ , as proposed by Perez et al. (2016), which is equivalent to the RMSE of the persistence of PV power (see also equations 8.5 and 8.12)

$$RMSE_{per} = \frac{1}{\sqrt{N}} \sqrt{\sum_{i=1}^N (\Delta P_{\Delta t})^2} = \sigma(\Delta P_{\Delta t}). \quad (17)$$

Here, it is commonly assumed that the average of  $\Delta P_{\Delta t}$  should be zero.

Both the variability and the forecast errors decrease with an increase in the size of the region and the number of PV systems considered. They can be well fitted either by a hyperbolic function, similar to one proposed by Perez et al. (2016), or by an exponential function, similar to the one proposed by Lorenz et al. (2009). As shown in Figure 10, by enlarging the footprint of the forecast region from the prediction of the PV generation in each market zone in Italy to the prediction of the PV generation over all of Italy, the RMSE can decrease from 5.5% (market zones average) to 3.6% (countrywide).

To summarize, expanding the transmission grid to manage the power generation in large areas (e.g., entire countries instead of market zones) not only reduces congestion and constraints on production capacity but also increases the forecast accuracy, as shown with the Italy example.

## 5.6 Effective Model Validation Benchmarking: Operationally Firm Solar Forecasts

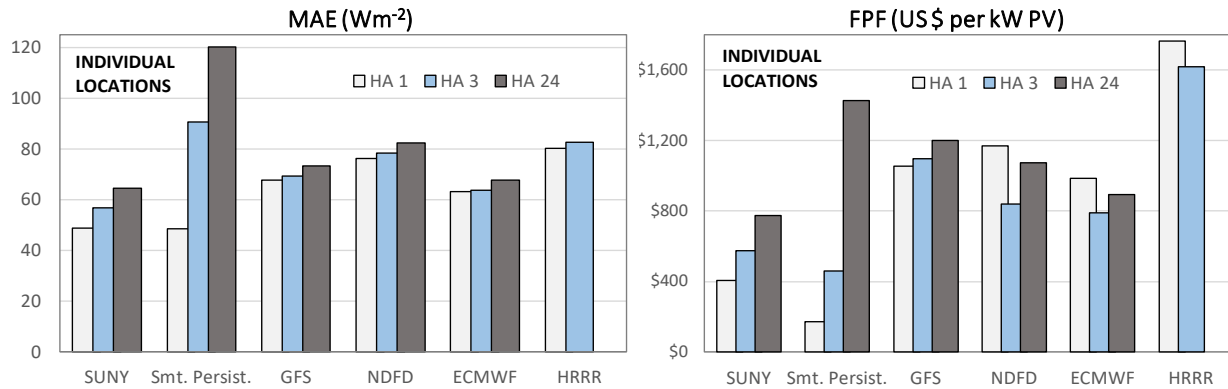
When validating solar forecasts, the classical error metrics (e.g., RMSE and MAE) introduced in Section 5.1 are commonly used. With continuous development in technology and changes in the energy market, the need arises for new validation measures that account for the economic value of solar forecasts. Thus, the firm power forecast (FPF or “perfect forecast”) concept was developed and introduced in a recent series of publications by Perez et al. (2019a, 2019b) and Pierro et al. (2020a). This forecast is both an effective model validation metric and an operational strategy to integrate increasing amounts of variable solar generation on electric grids. The costs incurred in transforming imperfect forecasts into firm predictions define the new metric: These include the costs of energy storage and output curtailment needed to make up for any over- or underprediction situations. It was shown by Perez et al. (2019a, 2019b) and Pierro et al. (2020a) that delivering firm predictions (i.e., fully eliminating grid operator renewable supply-side uncertainty and, incidentally, the need to characterize forecasts probabilistically) could be achieved at a modest operational cost.

Although other recent publications have focused on standardizing error metrics and forecast model validation practice (e.g., Hoff et al. 2012a; Yang et al. 2020), the choice of possible references as well as the possible definitions of persistence constitute sources of ambiguity when evaluating results from different studies, particularly reports emanating from the industry. In addition, the standard metrics, however “standardized,” are not directly exploitable by grid operators for estimating operational costs incurred by forecast errors.

The FPF metric is defined as the optimum (i.e., minimum possible) capital cost of PV plant oversizing and storage that is sufficient to make up for all instances of over- and underforecasts. This minimum is a function of the assumed capital costs for PV and storage and, of course, of the quality of the forecast. This metric bypasses both the ambiguity of standard metrics and their exploitability by grid operators because the metric is (1) a tangible hardware cost and (2) an indirect measure of operational costs resulting from solar supply-side uncertainty, because applying firm forecasts would entirely eliminate the said uncertainty.

Figure 11 compares the MAE and FPF metrics for GHI forecasts at 1-, 3-, and 24-hour forecast horizons. Results stem from the analysis of time series at the seven NOAA Surface Radiation Budget Network (SURFRAD) sites for a period of 1 year. The forecast models include smart persistence (Section 5.1.2), SUNY (also known as SolarAnywhere; SolarAnywhere 2019), and its underlying NWP components: IFS, GFS, National Digital Forecast Database, and the HRRR. In this example, the hardware costs quantifying the FPF metric amount to \$1,200 per kW for PV oversizing and \$200 per kWh for storage.

A noticeable difference between the two metrics is the performance of smart persistence relative to the other models, especially for short-term horizons. Persistence turns out to be operationally more robust than generally assumed because, whereas its dispersion can be large (i.e., large MAE and RMSE), this dispersion tends to be well balanced around the 1:1 diagonal, with fewer instances of the prolonged over/underestimations that are operationally costly.



**Figure 11. Comparison of forecast model performance at 1-, 3-, and 24-hour horizons for all SURFRAD stations using (left) the MAE metric and (right) the FPF metric.**

## 6 Probabilistic Solar Forecasts

A forecast is inherently uncertain and a proper assessment of its associated uncertainty offers the grid operator a more informed decision-making framework. For example, a deterministic forecast that includes predictions intervals is of genuine added value and, if appropriately incorporated in grid operations, might permit an increase in the value of solar power generation (Morales et al. 2014).

This section is restricted to the univariate context that corresponds to probabilistic forecasts that do not consider the spatiotemporal dependencies generated by stochastic processes such as solar power generation. Two types of solar probabilistic forecasts are considered here: *quantile forecasts* and *ensemble forecasts* (i.e., those using the Ensemble Prediction System, or EPS). Quantile forecasts are quite versatile probabilistic models and as such might address different forecasting time horizons, whereas EPS forecasts generally provide day-ahead probabilistic forecasts. Further, a verification framework is considered for evaluating the quality of solar probabilistic forecasts. The evaluation framework is based on visual diagnostic tools and a set of scores mostly originating from the weather forecast verification community (Wilks 2014). What follows provides an overview of the basic concepts related to solar probabilistic forecasting methods with an emphasis on the specific associated verification metrics. Comprehensive overviews regarding forecasting methods and the verification of solar probabilistic forecasts metrics can be found in van der Meer, Widén, and Munkhammar (2018); Antonanzas et al. (2016); and Lauret, David, and Pinson (2019).

### 6.1 Nature of Probabilistic Forecasts of Continuous Variables

In contrast to deterministic forecasts, probabilistic forecasts provide additional information about the inherent uncertainty embodied in NWP. The probabilistic forecast of a continuous variable, such as solar power generation or solar irradiance, takes the form of either a cumulative distribution function (CDF),  $F(Y)$ , or a PDF,  $f(Y)$ , of the random variable of interest,  $Y$  (e.g., GHI).

The CDF of a random variable  $Y$  is given as:

$$F(y) = P(Y \leq y) \quad (18)$$

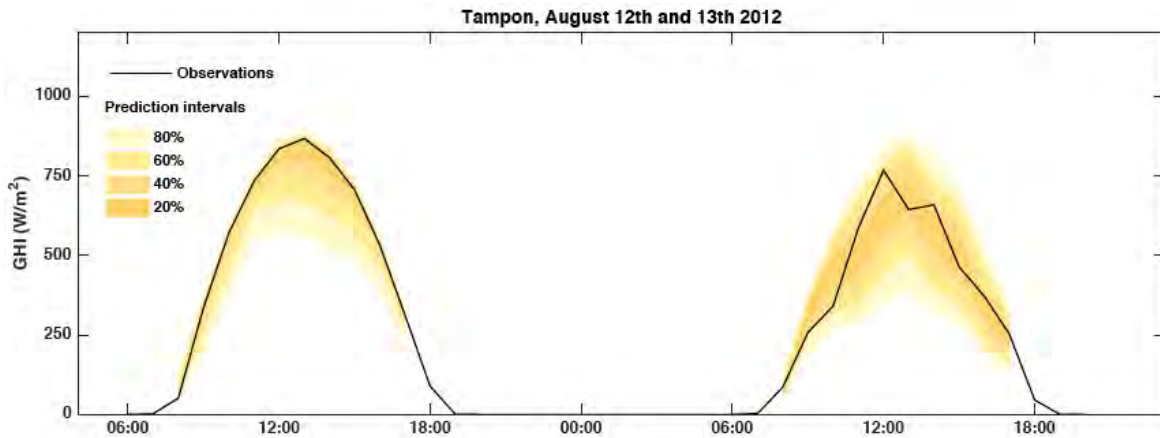
where  $P(Y \leq y)$  represents the probability that this random variable is less or equal to  $y$ .

The predictive distribution can be summarized by a set of quantiles. The quantile,  $q_\tau$ , at probability level  $\tau \in [0,1]$  is defined as follows:

$$q_\tau = F^{-1}(\tau), \quad (19)$$

where  $F^{-1}$  is the so-called *quantile function*. A quantile,  $q_\tau$ , corresponds to the threshold value below which an event,  $y$ , materializes with a probability level,  $\tau$ .

Prediction intervals (also called interval forecasts) can be inferred from the set of quantiles. Prediction intervals define the range of values within which the observation is expected to be with a certain probability (i.e., its nominal coverage rate) (Pinson et al. 2007). For example, a central prediction interval with a coverage rate of 95% is estimated by using the quantile  $q_{\tau=0.025}$  as the lower bound and  $q_{\tau=0.975}$  as the upper bound. Figure 12 shows an example of the probabilistic forecasts of solar irradiance where prediction intervals have been computed for nominal coverage rates ranging from 20% to 80%.



**Figure 12. Example of probabilistic solar irradiance forecasts: two days of measured GHI at Le Tampoon, France, and associated 1-hour ahead forecasts with prediction intervals (yellow) generated with the quantile regression forest model.**

## 6.2 Quantile Forecasts

Two approaches are commonly used in the community to generate quantile forecasts (see Figure 13) addressing different forecast horizons. As input, they use past ground observations and satellite images for intraday forecasting or NWP deterministic forecasts that are more effective for day-ahead forecasting. The first approach (Bacher, Madsen, and Nielsen 2009; Pedro et al. 2018) involves directly generating the quantiles of the predictive distribution of the variable of interest (e.g., GHI, DNI, or PV power). The second approach (Lorenz et al. 2009; David et al. 2016; Grantham, Gel, and Boland 2016; Pierro et al. 2020b) produces the interval forecasts from the combination of a deterministic (point) forecast and quantiles of the prediction error.



For both approaches, quantiles can be estimated either by assuming a parametric law for the predictive distribution or by nonparametric methods, which make no assumptions about the shape of the predictive distribution.

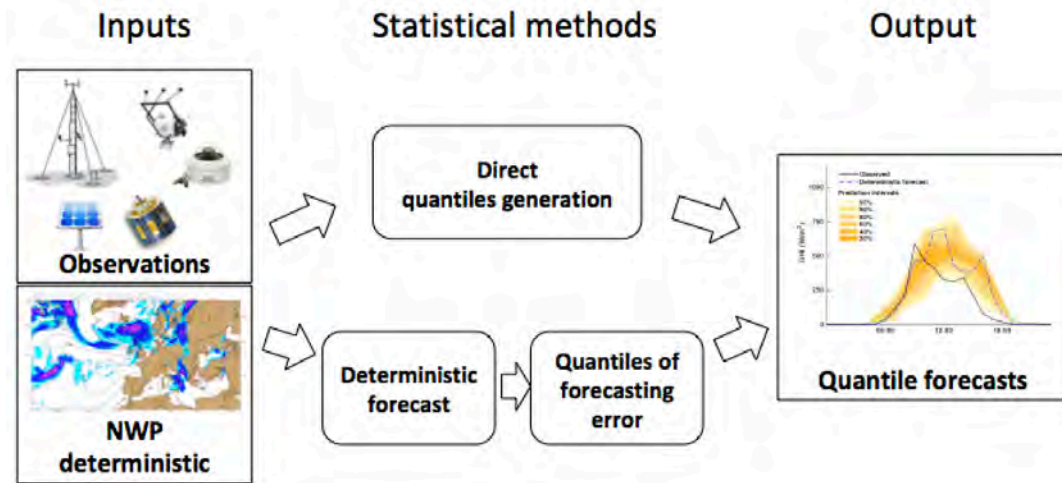


Figure 13. The two typical workflows used to generate quantile forecasts from recent past observations and/or deterministic NWP forecasts

### 6.2.1 Parametric Methods

Parametric models assume that the variable of interest or the prediction error follows a known law of distribution (e.g., a doubly truncated Gaussian for GHI or a Gaussian for the error distribution). Only a few parameters (e.g., mean and variance) are needed to fully characterize the predictive distribution. Consequently, this approach is particularly interesting in an operational context because it requires a low computational effort.

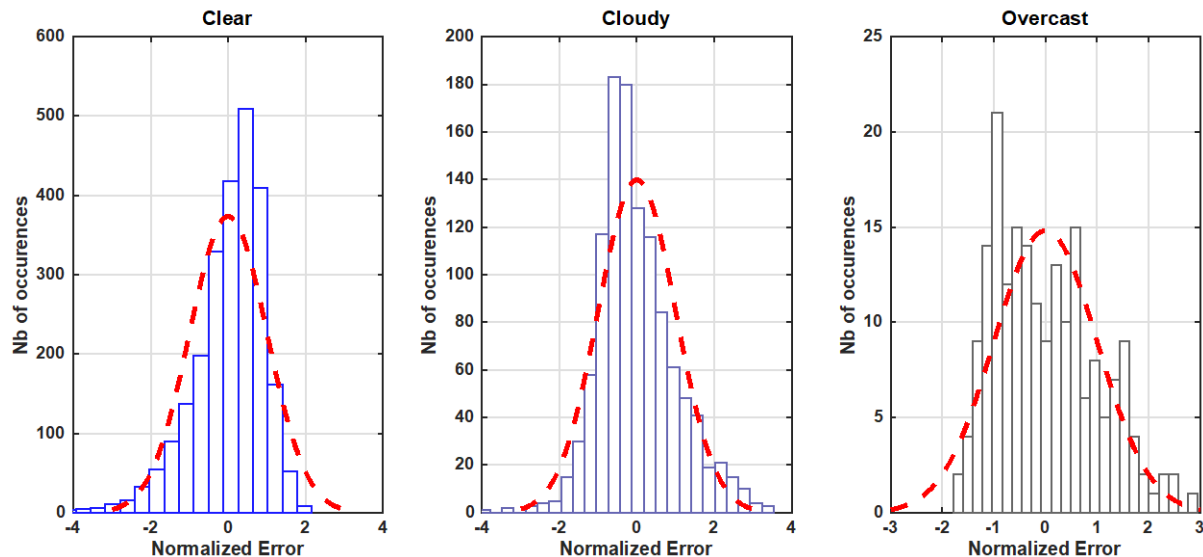


Figure 14. PDF of the normalized error (zero mean and unit variance) of the hourly profile of day-ahead forecasts of the clear-sky index provided by ECMWF-HRES for three different sky conditions and for the site of Saint-Pierre (21.34°S, 55.49°E), Reunion, France, in 2012. The red dashed line represents the fitted standard normal PDF. *Image from David and Lauret (2018)*

In the solar forecasting community, it is very common to fit a Gaussian distribution to the errors even though errors derived from deterministic forecasts of solar irradiance and of clear-sky index do not follow a Gaussian distribution (see Figure 14). For instance, Lorenz et al. (2009) developed the first operational PV forecasting model in Germany by assuming a Gaussian distribution of the error of the deterministic GHI forecasts generated by the IFS. More precisely, the predictive CDF was a Gaussian distribution with a mean corresponding to the point forecast and a standard deviation derived from a fourth-degree polynomial function for different classes of cloud index and solar elevation. For intrahour and intraday solar irradiance probabilistic forecasts, David et al. (2016) assumed a Gaussian error distribution of the deterministic forecast to generate a predictive CDF with a Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) model. Instead of fitting a parametric PDF to the error distribution, Fatemi et al. (2018) proposed a framework for parametric probabilistic forecast of solar irradiance using the beta distribution and standard two-sided power distribution.

### 6.2.2 Nonparametric Methods

To avoid making assumptions about the shape of the predictive distribution, numerous nonparametric methods have been proposed in the literature (van der Meer, Widén, and Munkhammar 2018). Examples are techniques such as bootstrapping (Efron 1979; Grantham, Gel, and Boland 2016), kernel density estimation (Parzen 1962), or k-nearest neighbor (Pedro et al. 2018). Here, two prominent and simple nonparametric methods are discussed briefly: the quantile regression and analog ensemble techniques.

Quantile regression models relate quantiles of the variable of interest (predictand) to a set of explanatory variables (predictors). Statistical or machine learning techniques—such as linear quantile regression, quantile regression forest, or gradient boosting (David and Lauret 2018; van der Meer, Widén, and Munkhammar 2018)—are commonly used to produce the set of quantiles with probability levels spanning the unit interval.

The following summarizes the linear quantile regression method first proposed by Koenker and Bassett (1978). See David, Luis, and Lauret (2018) for details about the implementation of other regression methods, including other variants of the linear quantile regression, quantile regression forest, quantile regression neural network, or boosting.

The linear quantile regression technique estimates a set of quantiles of the cumulative distribution function,  $F$ , of some response variable,  $Y$  (the predictand), by assuming a linear relationship between the quantiles of  $Y$  ( $q_\tau$ ) and a set of explanatory variables,  $X$  (the predictors):

$$q_\tau = \beta_\tau X + \epsilon, \quad (20)$$

where  $\beta_\tau$  is a vector of the parameters to be optimized at each probability level,  $\tau$ , and  $\epsilon$  represents a random error term (Koenker and Bassett 1978).

Numerous implementations of the linear quantile regression technique (and of its related variants) have been proposed in the literature to generate quantile forecasts for different forecast horizons and using different types of predictors,  $X$ . See, e.g., Bacher, Madsen, and Nielsen (2009); Zamo et al. (2014); Lauret, David, and Pedro (2017); van der Meer, Widén, and Munkhammar 2018; and Bakker et al. (2019).

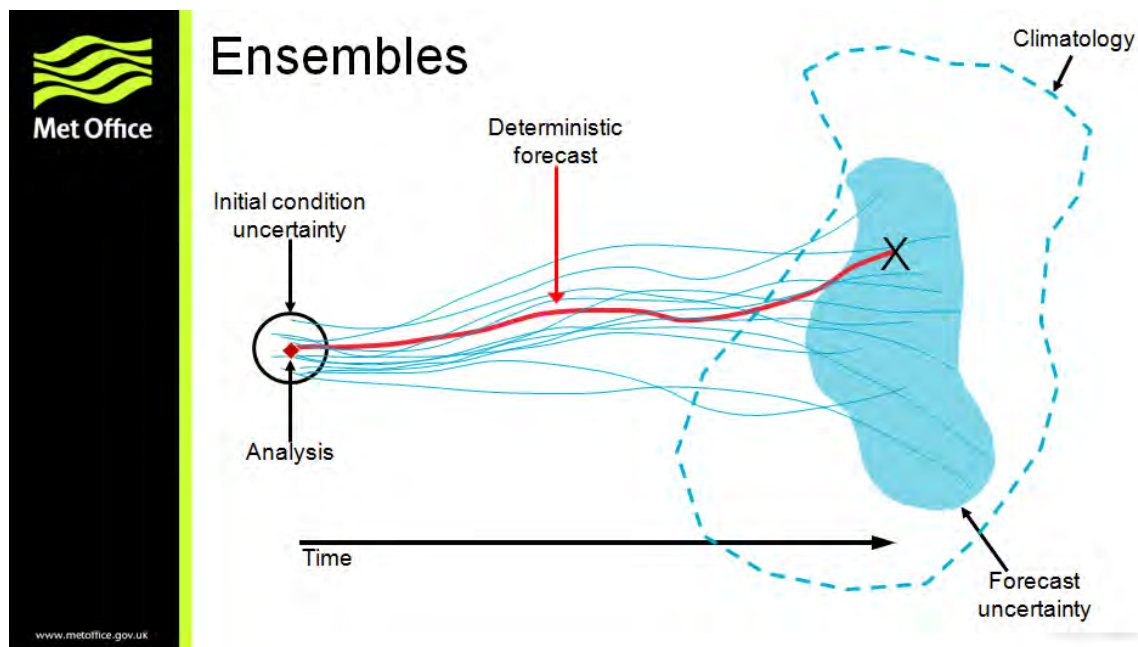
The analog ensemble (AnEn) method (Delle Monache et al. 2013) is a simple nonparametric technique used to build the predictive distributions. The aim is to search for similar forecasted conditions in the historical data and to create a probability distribution with the corresponding observations. Alessandrini et al. (2015) applied an AnEn approach to a set of predicted meteorological variables (e.g., GHI, cloud cover, and air temperature) generated by the Regional Atmospheric Modeling System (RAMS). Note that the AnEn technique is mostly employed for day-ahead forecasts and generates the predictive distribution using NWP deterministic forecasts.

### 6.3 Ensemble Prediction System

#### 6.3.1 Definition

The EPS corresponds to a perturbed set of forecasts generated by slightly changing the initial conditions of the control run and of the modeling of unresolved phenomena (Leutbecher and Palmer 2008). Figure 15 shows a schematic representation of an ensemble forecast generated by an NWP model. The trajectories of the perturbed forecasts (blue lines) can differ strongly from the control run (red line). The spread of the resulting members (blue stain) represents the forecast uncertainty. For example, the ECMWF provides an ensemble forecast from the IFS model. It consists of 1 control run and 50 “perturbed” members.

Though members of the ensemble are not directly linked to the notion of quantiles, they can be seen as discrete estimates of a CDF when they are sorted in ascending order. Lauret, David, and Pinson (2019) proposed different ways to associate these sorted members to a CDF.



**Figure 15. A schematic illustration of an ensemble forecast generated with an NWP model.**  
*Image from Met Office, © British Crown copyright (2021)*

#### 6.3.2 Post-Processing of the Ensemble Prediction System

Global and regional NWP models are designed to forecast a large variety of meteorological variables (but mainly precipitation and temperature) and have not previously focused on the accurate

generation of the different components of solar radiation. Consequently, raw ensembles provided by meteorological centers suffer from a lack of accuracy, a lack of calibration, or both (Leutbecher and Palmer 2008). Additionally, see, e.g., Yang et al. (2020) for definitions and discussions about the specific meaning of accuracy, calibration, and other specialized terms in the field of forecasting—some of which appearing in the following. Overall, raw ensemble forecasts are systematically refined by post-processing techniques (also called calibration techniques) to further improve their quality.

The aim of post-processing is to apply a statistical calibration to the PDF drawn by the raw initial ensemble forecasts to optimize a specific metric (e.g., the continuous ranked probability score [CRPS] described in Section 6.4) used to assess the quality of probabilistic forecasts. Indeed, as well as having a coarse spatial resolution, the ensemble forecasts from NWP are known to be underdispersive; in other words, they exhibit a lack of spread (Leutbecher and Palmer 2008). To address this, Sperati, Alessandrini, and Delle Monache (2016) proposed two different methods already used in the realm of wind forecasting: The variance deficit method designed by Buizza et al. (2003) and the ensemble model output statistic (MOS) method proposed by Gneiting et al. (2005). Even if these methods cannot be considered to be parametric, they are based on the characteristics of a normal distribution. Indeed, such a distribution is appealing because it can be assessed with only two parameters: the mean and the standard deviation, which are related to the average bias and the spread of the ensemble, respectively.

Another method of calibration is based on the rank histogram (see Section 6.4); it was initially proposed by Hamill and Colucci (1997) for precipitation forecasts. Zamo et al. (2014) applied this method to the Météo France’s EPS, called PEARP, to generate probabilistic solar forecasts. The aim of this method is to build a calibrated CDF from the rank histogram derived from past forecasts and observations. Other techniques of EPS calibration exist, but they have not been used for solar forecasting. For example, Pinson (2012) and Pinson and Madsen (2009) suggested a framework for the calibration of wind ensemble forecasts. Junk, Delle Monache, and Alessandrini (2015) proposed an original calibration model for wind-speed forecasting applied to ECMWF-EPS based on the combination of nonhomogeneous Gaussian regression and AnEn models. Likewise, Hamill and Whitaker (2006) suggested an adaptation of the AnEn technique for the calibration of ensemble precipitation forecasts using the statistical moments of the distribution, such as the mean and spread of the members as predictors. See Wilks (2018) for a thorough review of univariate ensemble postprocessing methods.

## 6.4 Verification of Solar Probabilistic Forecasts

### 6.4.1 Properties Required for a Skillful Probabilistic System

Several attributes characterize the quality of probabilistic forecasts (Jolliffe and Stephenson 2011; Wilks 2014), but two main properties (reliability and resolution) are used to assess the quality of the forecasts.

The *reliability* or calibration refers to the statistical consistency between forecasts and observations; in other words, a forecast system has a high reliability if the forecast probability and observed frequency agree. The reliability property is an important prerequisite because

nonreliable forecasts would lead to a systematic bias in subsequent decision-making processes (Pinson et al. 2007).

The *resolution* measures the ability of a forecasting model to generate predictive distributions that depend on forecast conditions. Put differently, the more distinct the observed frequency distributions for various forecast situations are from the full climatological distribution, the more resolution the forecast system has. Climatological forecasts are perfectly reliable but have no resolution. Consequently, a skillful probabilistic forecasting system should issue reliable forecasts and should exhibit high resolution.

*Sharpness*, which refers to the concentration of predictive distributions, can be measured by the average width of the prediction intervals. Unlike reliability and resolution, sharpness is a function of only the forecasts and does not depend on the observations. Consequently, a forecasting system can produce sharp forecasts yet be useless if the probabilistic forecasts are unreliable.

#### 6.4.2 Probabilistic Verification Tools

The evaluation framework is based on visual diagnostic tools and numerical scores.

##### 6.4.2.1 Visual Diagnostic Tools

Table 1 lists the diagnostic tools for which Lauret, David, and Pinson (2019) provided pros and cons as well as detailed information about their implementation. Note that some tools were initially designed for a specific type of forecast (i.e., an ensemble or quantile forecast) and that there is apparently no visual diagnostic tool to assess the resolution property.

**Table 1. Visual Diagnostic Tools**

Diagnostic Tool	Remarks
Reliability diagram	Initially designed for the reliability assessment of quantile forecasts Can be used for ensemble forecasts if members are assigned specific probability levels (see Lauret, David, and Pinson [2019])
Rank histogram	Initially designed for the reliability assessment of ensemble forecasts Can be extended to quantile forecasts if quantiles are evenly spaced
Probability integral transform histogram	Represents a reliability assessment of quantile forecasts
Sharpness diagram	Plots the average width of the prediction intervals for different nominal coverage rates Sharpness can only contribute to a qualitative evaluation of the probabilistic forecasts. Even if narrow prediction intervals are preferred, sharpness cannot be seen as a property to verify the quality of probabilistic forecasts but is more like the consequence of a high resolution.

**Table 2. Forecast Metrics**

Forecast Metric	Remarks
Continuous ranked probability score (CRPS)	Can be normalized to define a skill score (CRPS skill score) Can be further partitioned into the two main attributes: reliability and resolution
Ignorance score	Local score (i.e., the score depends only on the value of the predictive distribution at the observation) Cannot be normalized
Interval score	Specifically designed for interval forecasts
Quantile score	Forecast performance of specific quantiles

### 6.4.2.2 Numerical Scores

Numerical scores provide summary measures for the evaluation of the quality of probabilistic forecasts. Table 2 enumerates the main scoring rules for evaluating the quality of probabilistic forecasts of a continuous variable. All the scores listed in the table are proper scoring rules (Gneiting and Raftery 2007), hence ensuring that perfect forecasts are given the best score value. Lauret, David, and Pinson (2019) gave a detailed definition of each score.

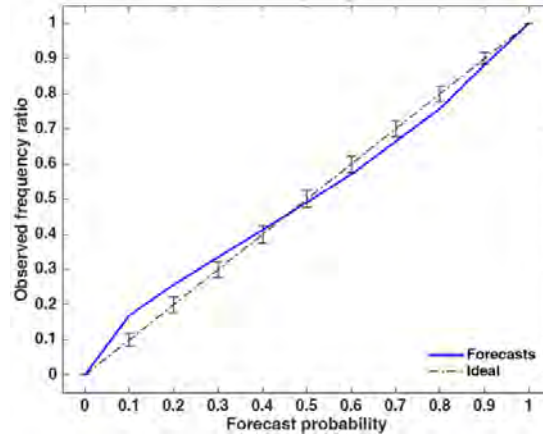
### 6.4.3 Presentation of Some Tools and Scores

This section describes in detail some diagnostic tools and numerical scores. See Lauret, David, and Pinson (2019) and Yang et al. (2020) for descriptions of other metrics.

#### 6.4.3.1 Reliability Diagram

A reliability diagram is a graphical verification display used to assess the reliability attribute of quantile forecasts. Quantile forecasts are evaluated one by one, and their observed frequencies are reported versus their forecast probabilities (see Figure 16). Such a representation is appealing because the deviations from perfect reliability (the diagonal) can be visually assessed (Pinson, McSharry, and Madsen 2010); however, because of both the finite number of pairs of observation/forecast and also possible serial correlation in the sequence of forecast-verification pairs, observed proportions are not expected to lie exactly along the diagonal, even if the density forecasts are perfectly reliable. Pinson, McSharry, and Madsen (2010) proposed a method to add consistency bars to the reliability diagram. This adding of consistency bars to the reliability diagrams can help users gain more confidence in their (possibly subjective) judgment regarding the reliability of the different models. Figure 16 shows an example of reliability diagram with consistency bars. In this example, the forecasts cannot be considered reliable because the line corresponding to the forecasts falls outside the consistency bars. More elaborate reliability diagrams are proposed by Yang (2019a, 2019c).



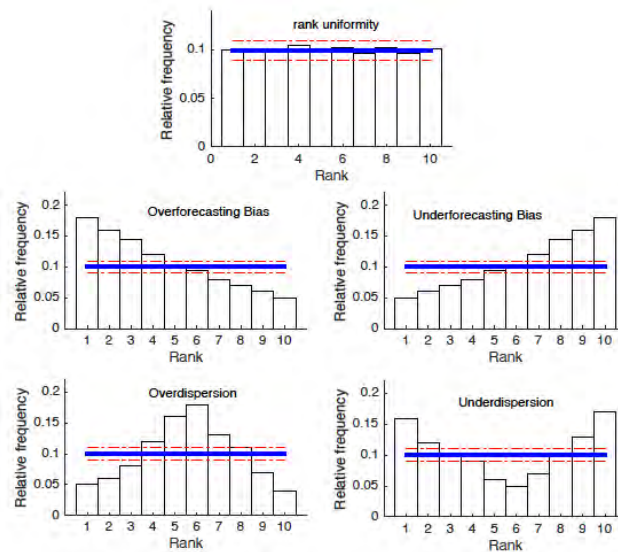


**Figure 16. Example of a reliability diagram. Consistency bars for a 90% confidence level around the ideal line are individually computed for each nominal forecast probability.**

### 6.4.3.2 Rank Histogram

A rank histogram is a graphical display initially designed for assessing the reliability of ensemble forecasts (Wilks 2014). Rank histograms permit users to visually assess the statistical consistency of the ensemble—that is, if the observation can be seen statistically like another member of the ensemble (Wilks 2014). A flat rank histogram is a necessary condition for ensemble consistency and shows an appropriate degree of dispersion of the ensemble. An underdispersed or overdispersed ensemble leads to U-shaped or hump-shaped rank histogram; see Figure 17.

In addition, some unconditional biases can be revealed by asymmetric (triangle-shaped) rank histograms. It must be stressed that one should be cautious when analyzing rank histograms. As shown by Hamill (2001), a perfectly flat rank histogram does not mean that the corresponding forecast is reliable. Further, when the number of observations is limited, consistency bars can also be calculated with the procedure proposed by Bröcker and Smith (2007).



**Figure 17. Illustrative examples of rank histograms for an ensemble of  $M = 9$  members. The horizontal solid blue line denotes the statistical consistency of the ensemble. The dashed-dotted lines represent the consistency bars. *Figure inspired from Wilks (2014)***

### 6.4.3.3 Overall Skill Assessment with the Continuous Ranked Probability Score

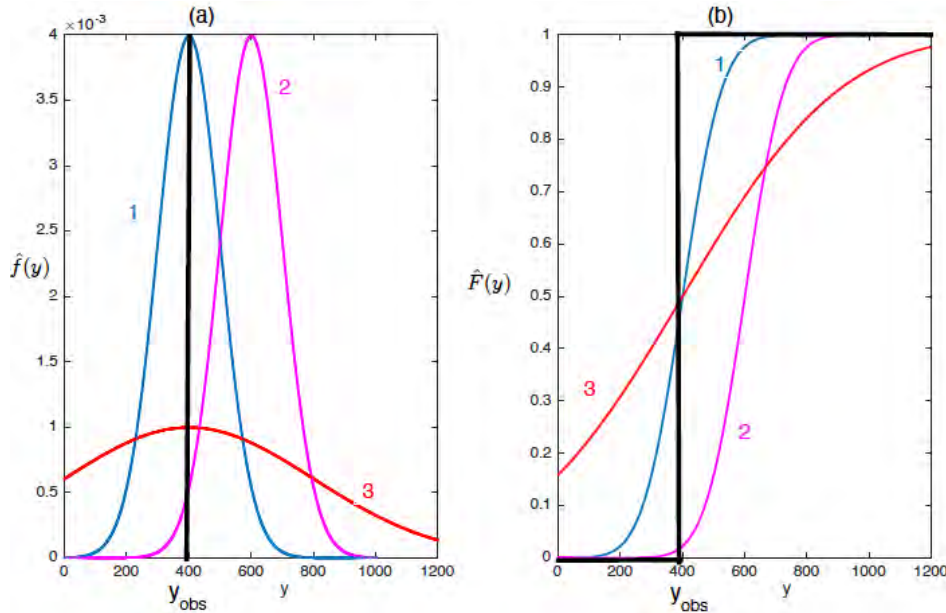
The most common skill score for evaluating the quality of predictive densities of continuous variable is the CRPS, whose formulation is:

$$CRPS = \frac{1}{N} \sum_{i=1}^N \int_{-\infty}^{+\infty} [\hat{F}_{fcst}^i(y) - F_{y_{obs}}^i(y)]^2 dy, \quad (21)$$

where  $\hat{F}_{fcst}^i(y)$  is the predictive CDF of the variable of interest,  $x$  (e.g., GHI), and  $F_{y_{obs}}^i(y)$  is a CDF of the observation (i.e., a step function that jumps from 0 to 1 at the point where the forecast variable,  $y$ , equals the observation,  $y_{obs}$ ). The squared difference between the two CDFs is averaged over the  $N$  forecast/observation pairs. Note that the CRPS is negatively oriented (smaller values are better) and has the same dimension as the forecasted variable.

Figure 18(a) shows three hypothetical predictive PDFs, and Figure 18(b) plots the corresponding predictive CDFs. The black thick line in Figure 18(b) represents the CDF of the observation,  $F(y)$ .

Because CRPS represents the integrated squared difference between the two CDFs, the pair of observation/forecast indicated by 1 will be assigned the best score. Conversely, forecasts indicated by 2 and 3 will lead to a higher CRPS. Indeed, although it has the same degree of sharpness as Forecast 1, Forecast 2 is not centered on the observation (i.e., this is a biased forecast). Regarding Forecast 3, even though it is centered on the observation, it is less sharp than forecasts 1 and 2. In summary, CRPS rewards the concentration of probability around the step function located at the observed value (Hersbach 2000).



**Figure 18. Schematic of the CRPS skill score.** Three forecast PDFs are shown in relation to the observed variable in (a). The corresponding CDFs are shown in (b), together with the step function CDF for the observation (black heavy line). Forecast PDF 1 would produce a small (i.e., good) CRPS. This would not be the case for Forecast 2 or Forecast 3. *Illustration inspired from Wilks (2014)*

The CRPS can be further partitioned into the two main attributes of probabilistic forecasts described: reliability and resolution. The decomposition of the CRPS leads to:

$$CRPS = \text{RELIABILITY} - \text{RESOLUTION} + \text{UNCERTAINTY}. \quad (22)$$

The uncertainty term cannot be modified by the forecast system and depends only on the observation's variability (Wilks 2014). Because the CRPS is negatively oriented, the goal of a forecast system is to minimize the reliability term and maximize the resolution term as much as possible. Hersbach (2000) and Lauret, David, and Pinson (2019) detail the procedures for calculating the different terms (reliability and resolution, respectively) for ensemble and quantile forecasts.

It must be stressed that the decomposition of the CRPS provides quantitative overall measures of reliability and resolution, hence providing additional and valuable insight into the performance of a forecasting system.

Similarly, to obtain skill scores used for evaluating deterministic forecasts (Coimbra et al. 2013b), a CRPS skill score (CRPSS) can be derived to quantify the improvement brought by a new method over a reference easy-to-implement (or “baseline”) model, such as:

$$CRPSS = 1 - \frac{CRPS_{new\ method}}{CRPS_{Reference}}. \quad (23)$$

Negative values of CRPSS indicate that the new proposed method fails to outperform the reference baseline model, and positive values of CRPSS mean that the new method outperforms the reference model. Further, the higher the CRPSS, the better the improvement. Note that the uncertainty part of the decomposition of the CRPS (which corresponds to the score of the climatology) can be used as a reference baseline model. CRPSS and a mean-normalized CRPS are also discussed by Yang (2020).

#### 8.6.4.3.4 Interval Score

The interval score (IS) specifically assesses the quality of interval forecasts. As shown by Eq. 24 the interval score rewards narrow prediction intervals but penalizes (with a penalty term that increases with increasing nominal coverage rate) the forecasts for which the observation,  $x_{obs}$ , is outside the interval. For a  $(1 - \alpha) \times 100\%$  nominal coverage rate, the interval score reads as:

$$IS_{\alpha} = \frac{1}{N} \sum_{i=1}^N (U^i - L^i) + \frac{2}{\alpha} (L^i - x_{obs}^i) I_{x_{obs}^i < L^i} + \frac{2}{\alpha} (x_{obs}^i - U^i) I_{x_{obs}^i > U^i} \quad (24)$$

where  $I_u$  is the indicator function ( $I_u=1$  if  $U$  is true and 0 otherwise).  $U^i$  and  $L^i$  represent the upper  $\left(\tau = 1 - \frac{\alpha}{2}\right)$  and lower  $\left(\tau = \frac{\alpha}{2}\right)$  quantiles, respectively.

A plot of interval scores for different nominal coverage rates might offer a consistent evaluation of the quality of interval forecasts. Consequently, such a plot could advantageously replace the sharpness diagram.

#### 6.4.4 Benchmark Probabilistic Models

This section describes the benchmark probabilistic models used to gauge the performance of new proposed probabilistic methods using skill scores such as the CRPSS. By analogy with the deterministic approach, persistence ensemble (PeEn) models based on GHI (Alessandrini et al. 2015) and on clear-sky index (David et al. 2016) have been proposed. The empirical CDF of a PeEn forecast is simply built with the most recent  $k$  past measurements of solar irradiance. Considering an infinite number of past measurements, the PeEn turns to be the climatology. In numerous other fields of meteorology, climatology is often considered to be a reference that can be used to test the performance of probabilistic models (Wilks 2014). Indeed, the climatology is perfectly reliable, but it has no resolution. Also, an advanced climatology, called the complete-history persistence ensemble, was proposed by Yang (2019b); this reference model corresponds to a conditional climatology where the time of the day is used as a predictor. In addition, for ensemble forecasts, the CRPS of the raw ensemble can serve as a benchmark.

## 7 Summary and Recommendations for Irradiance Forecasting

Solar power forecasting is essential for the reliable and cost-effective system integration of solar energy. It is used in for a variety of applications with their specific requirements with respect to forecast horizon and spatiotemporal resolution. To meet these needs, different solar irradiance and power forecasting methods have been developed, including physical and empirical models as well as statistical and machine learning approaches. Based on these developments, a number of forecasting services of good quality is available for users today. In the following, a summary of different forecasting methods and their applicability for different tasks is given, along with criteria that determine model performance as well as recommendations for forecast evaluation.

- Different empirical and physical models are suitable for different forecast horizons. Generally, the spatiotemporal resolution of irradiance forecasts decreases with increasing forecast horizon.
- Short-term irradiance forecasts from 10–20 minutes ahead resolving irradiance ramps with a temporal resolution of minutes or even less can be derived from ASIs using cloud motion-based methodologies.
- Irradiance forecasts up to several hours ahead with typical resolutions from 10–15 minutes can be derived from satellite data covering large areas, also using cloud motion-based methodologies.
- Irradiance forecasting from several hours to days ahead essentially relies on NWP models, with their capability to describe complex atmospheric dynamics, including advection as well as the formation and dissipation of clouds.

The performance of the different forecast models depends on multiple factors that have different impacts depending on the forecast horizons:

- The capability of the models to predict changes in clouds and irradiance
- The performance of the models for irradiance retrieval/analysis

- The capability of the models to predict AOD, especially for DNI forecasting in arid regions
- Input data to the model (parameters as well as the area covered by the input data)
- Time for a model run (run-time determines a lower limit for the delay of observations at the time of forecast delivery)
- Spatiotemporal resolution.

Complementing empirical and physical models, statistical and machine learning methods are widely used in solar irradiance and power forecasting:

- For time-series models and post-processing, which require reference values for training, the availability of irradiance and/or PV power measurements is crucial, as is proper quality control of the data. Satellite-derived data can also be used for training on irradiance.
- Short-term forecasting up to approximately 1 hour ahead benefits greatly from the use of local online irradiance or PV power measurements as input; however, pure time-series approaches based on local measurements only are outperformed by approaches integrating empirical and/or physical model forecasts from a few minutes to hours onward, depending on the spatiotemporal scale of the forecasts and the climatic conditions of the forecast location.
- Statistical and machine learning approaches are applied effectively for improving forecasts with empirical or physical models (post-processing). They can reduce systematic meteorological forecast errors. Training to PV power measurements in addition allows adaptation to the specifics of a given PV plant or to replace PV simulation models.
- Machine learning models are increasingly applied to replace parts of empirical models, e.g., algorithms to compute optical flow in cloud motion approaches.

State-of-the-art solar irradiance or PV power forecasting services do not rely on a single forecasting model but integrate and optimize different tools and data, with prominent examples given here:

- High-resolution intrahour forecasting systems combine the use of local measurements and ASI data with empirical and machine learning approaches.
- Forecasting systems for the intraday energy market up to several hours ahead integrate online measurements, satellite-based forecasts, and NWP model-based forecasts with statistical and/or machine learning approaches.
- Forecasting systems from several hours to several days ahead use different NWP models as input in combination with statistical and/or machine learning approaches.

Besides forecasting for single PV power plants, the estimation and forecasting of regionally aggregated PV power is important for grid operators. It involves the same modeling approaches described. Here, an additional challenge is that information on all the PV power plants contributing to the overall feed-in is often incomplete. Also, PV power is not measured at a sufficient resolution for most plants in many countries; therefore, upscaling approaches have been developed and are applied effectively to derive and forecast regionally aggregated PV power. Because of spatial smoothing effects, forecast errors of regionally aggregated PV power (normalized to the installed power) are much smaller than for single PV plants, depending on the size of the region and the set of PV plants contributing.

Forecast evaluations provide users with necessary information on forecast accuracy, assisting them in choosing between different forecasting services or assessing the risk when a forecast is used as a basis for decisions. The assessment of forecast accuracy should combine visual diagnostics (e.g., scatter plots or 2D histograms of forecasts and observations) and quantitative error measures (e.g., RMSE and skill in comparison to persistence). In addition to the model used, forecast accuracy depends on different factors, including the climatic conditions and the spatiotemporal scale; therefore, specific evaluation for a given application considering these factors is recommended—i.e., an evaluation for sites in a similar climate and with similar spatiotemporal resolutions.

Beyond general information on the overall accuracy of a deterministic forecast, probabilistic forecasts provide specific uncertainty information for each forecast value, depending on the weather conditions. Probabilistic forecasts take the form of CDFs or PDFs. They are summarized by quantiles from which prediction intervals can be inferred. Quantiles can be estimated using either a parametric or a nonparametric approach. In the latter case, statistical or machine learning techniques can be used to estimate the quantiles. Although NWP ensemble members are not directly linked to the notion of quantiles, different propositions exist to infer a CDF from an ensemble. As an example, for deterministic forecasting, the assessment of the quality of the probabilistic forecasts is based on visual diagnostic and proper scoring rules. In particular, the CRPS seems to have all the features needed to evaluate a probabilistic forecasting system and, as such, could become a standard for verifying probabilistic forecasts of solar irradiance and power.

Finally, forecasting solar irradiance should be evaluated in the context of strategies for the system integration of solar power, which aim to provide the necessary power to cover demand at any time. These strategies include spatial smoothing for grid-integrated PV and increasingly also the use of storage (batteries) and curtailment as well as in combination with other variable renewable energy sources, especially wind power. Applying these strategies reduces the variability of solar power as well as forecast errors.



1. **True or False? High penetrations of PV power generation pose several challenges for the stability of the electric grid because of the stochastic variability of the residual electric load (i.e., the difference between the energy need—or load—and the distributed PV power generation, depending on meteorological conditions and sun position).**
  - ☐ True
  - ☐ False
2. **Which forecasting method is based on numerical weather prediction (NWP) models?**
  - ☐ Intraday and day-ahead forecasts
  - ☐ Intrahour forecasts with high spatial and temporal resolution
  - ☐ Forecasts up to 4 hours ahead
  - ☐ Irradiance forecasting with cloud motion vectors
3. **This forecasting method uses horizontally installed cameras which photograph the whole sky above them?**
  - ☐ Irradiance Forecasting with Cloud Motion Vectors
  - ☐ Satellite-Based Forecasts
  - ☐ Forecasting Using Ground-Based All-Sky Imagers
  - ☐ Irradiance Forecasting with Numerical Weather Prediction
4. **True or False? Instead of using only one or a few ASI systems, networks of approximately 10 or more ASIs can be created to increase the spatial coverage, the forecast horizon, and the accuracy of observations.**
  - ☐ True
  - ☐ False
5. **Methods to calculate Cloud Motion Vectors (CMV)s from satellite images \_\_\_\_\_ in operational weather forecasting.**
  - ☐ are not accurate enough to be used in
  - ☐ is new and maybe upcoming but not yet proven with a high degree of accuracy for
  - ☐ have been developed and are routinely used
  - ☐ is antiqued method no longer used for

6. **Today, most NWP models offer \_\_\_\_\_ as direct model output.**
- ☐ Direct normal irradiance (DNI)
  - ☐ Global horizontal irradiance (GHI)
  - ☐ Diffuse horizontal irradiance (DHI)
  - ☐ None of the above
7. **NOAA's High-Resolution Rapid Refresh (HRRR) model is an example of which type of forecast model?**
- ☐ Operational Numerical Weather Prediction Models
  - ☐ High-Spectral-Resolution Models
  - ☐ Water Clouds Climate Models
  - ☐ Semiempirical BRDF Models
8. **Which the of following are Examples of machine learning models applied for forecasting?**
- ☐ Artificial Neural Networks
  - ☐ Gradient Boosted Regression Trees
  - ☐ Extreme Learning Machines
  - ☐ All of the above
9. **Why does statistical post-processing (or machine learning with exogenous input) play an important role in irradiance and PV power forecasting?**
- ☐ Combines the outputs of different models
  - ☐ Reduces model errors by considering unaccounted or partially accounted local and regional effects (e.g., topography and aerosols)
  - ☐ Derives quantities that are not direct model outputs.
  - ☐ All of the above
10. **True or False? The simplest way to forecast the production of a PV power plant is to apply a PV power simulation model to the forecast of the relevant predicting variables (primarily irradiance, but also environmental temperature and wind speed).**
- ☐ True
  - ☐ False

**11. Regional PV power feed-in can be estimated using which of the following data?**

- ☐ Meteorological data (irradiance and temperature)
- ☐ Measurements of the output of PV plants
- ☐ Information on the fleet of PV system: coordinates and installed capacity (along with tilt and orientation if available).
- ☐ All of the above

**12. True or False? The evaluation of solar irradiance forecasts provides users with the necessary information about forecast accuracy and helps them choose different forecasting products or assess the risk when using a particular forecast as a basis for decisions.**

- ☐ True
- ☐ False

**13. In regard to analysis of forecast error with respect to cloud variability and spatiotemporal averaging, forecast error \_\_\_\_\_ with enhanced cloud variability.**

- ☐ is typically unchanged even
- ☐ decreases
- ☐ increases
- ☐ variables must be adjusted

**14. By enlarging the footprint of the forecast region of interest, forecast errors are \_\_\_\_\_.**

- ☐ reduced
- ☐ not changed
- ☐ increased
- ☐ exponentially increased

**15. Expanding the transmission grid to manage the power generation in large areas (e.g., entire countries instead of market zones has which benefit?**

- ☐ Reduces congestion
- ☐ Reduced constraints on production capacity
- ☐ Increases the forecast accuracy
- ☐ All of the above

16. **Which type of forecast provides additional information about the inherent uncertainty embodied in numerical weather prediction (NWP)?**
- ☐ Deterministic forecasts
  - ☐ Physical forecasts
  - ☐ Probabilistic forecasts
  - ☐ Empirical forecasts
17. **What model assumes that the variable of interest or the prediction error follows a known law of distribution?**
- ☐ Parametric
  - ☐ Nonparametric
  - ☐ Ensemble Prediction System
  - ☐ Probabilistic
18. **The \_\_\_\_\_ measures the ability of a forecasting model to generate predictive distributions that depend on forecast conditions.**
- ☐ sharpness
  - ☐ resolution
  - ☐ reliability
  - ☐ quantitative
19. **True or False? Solar power forecasting is essential for the reliable and cost-effective system integration of solar energy.**
- ☐ True
  - ☐ False
20. **True or False? State-of-the-art solar irradiance or PV power forecasting services do not rely on a single forecasting model but integrate and optimize different tools and data.**
- ☐ True
  - ☐ False