Deriving multivariate probabilistic solar generation forecasts based on hourly imbalanced data

HEMF Working Paper No. 07/2024

by

Yannik Pflugfelder,

Aiko Schinke-Nendza,

Jonathan Dumas

and

Christoph Weber

November 2024

UNIVERSITÄT
DUISBURG
ESSEN

Open-Minded

Deriving multivariate probabilistic solar generation forecasts based on hourly imbalanced data

Yannik Pflugfelder[a], Aiko Schinke-Nendza[a], Jonathan Dumas[b], Christoph Weber[a]

[a] *House of Energy Markets and Finance, University of Duisburg-Essen, Universitätsstr. 12, 45117, Essen, Germany*
[b] *RTE - Réseau de Transport d'Electricité, S.A., 7C place du Dôme, Paris La Défense, 92073, France*

Abstract

Accurate forecasting of solar PV generation is critical for integrating renewable energy into power systems. This paper presents a multivariate probabilistic forecasting model that addresses the challenges posed by imbalanced data resulting from day and night-time periods in solar photovoltaic (PV) generation. The proposed approach offers a robust and accurate method for predicting solar PV output by incorporating forecast updates and modeling the temporal interdependencies. The methodology is applied to a case study in France, demonstrating effectiveness across different spatial granularities and forecast horizons. The model uses advanced data handling methods combined with copula models, resulting in improved Energy Scores and Variogram-based Scores. These improvements underscore the importance of addressing imbalanced data and utilizing multivariate models with repeated updates to enhance solar forecasting accuracy. This work contributes to advancing forecasting techniques essential for integrating renewable energy into power grids, supporting the global transition to a sustainable energy future.

Keywords: Multivariate probabilistic forecasts, Forecast updates, Solar generation, Copula

Email addresses: yannik.pflugfelder@uni-due.de (Yannik Pflugfelder), aiko.schinke-nendza@uni-due.de (Aiko Schinke-Nendza), jonathan.dumas@rte-france.com (Jonathan Dumas), christoph.weber@uni-due.de (Christoph Weber)

Contents

1. Introduction

Climate change demands urgent global action, particularly through adopting renewable energy technologies. Solar photovoltaic (PV) systems play a critical role in this transition due to their scalability and potential for reducing carbon emissions (Fernández et al., 2023). However, the integration of large-scale solar PV presents significant challenges, primarily due to the variability and uncertainty inherent in solar PV generation (Sinsel et al., 2020). These challenges impact both grid stability and market dynamics, making decision-support instruments essential for energy systems (Hong et al., 2020). While network operators face challenges in maintaining grid stability due to the variability of solar PV generation (Ahmadi et al., 2018), market participants must manage increased price volatility and uncertainty in supply, requiring advanced trading and balancing strategies (Visser et al., 2024). In this context, particularly novel forecasting techniques are necessary to cope with the variability and uncertainty inherent to solar PV generation (Yang et al., 2022).

1.1. Background and motivation

With increasing shares of renewable energy sources (RES) in decarbonized energy systems (Fernández et al., 2023), accurate and reliable forecasts become a fundamental element in the decision-making processes of network operators and market participants (Sweeney et al., 2020). However, over the past decade, literature on forecasting focused mainly on wind and load predictions, yielding well-established methods (Hong et al., 2020). This trend might partly be due to the distinct challenges that solar PV introduces—particularly in managing periods of low or zero generation during winter and night-time hours (Yang et al., 2022). Addressing these issues requires the development of forecasting models capable of capturing the full range of variability and uncertainty associated with solar PV (Li and Zhang, 2020).

A particular challenge is introduced by *forecast updates*, regularly causing imbalances in the entire energy system (see, e.g., Zalzar et al., 2020; Boehnke et al., 2024), and individual trading portfolios (see, e.g., Riddervold et al., 2021; Visser et al., 2024). To cope with such repeated forecast updates, the limitations of traditional point forecasts or realization time series have to be overcome (Kolkmann et al., 2024). Optimal decisions have to be based on possible stochastic evolutions of infeed expectations (Glas et al., 2020), as summarized in so-called *forecast trajectories*, which capture the full uncertainty in renewable infeed (Kolkmann et al., 2024). These trajectories, however, are rarely subject to research for solar PV generation as their modeling poses a significant challenge given the intermittency of solar radiation—particularly when covering both day and night-time periods. Instead, solar PV forecasts commonly focus on infeed expectations for one given look-ahead

horizon, either resulting in realization trajectories (see, e.g., Golestaneh et al. (2016)) or individual timeseries data (see, e.g., Schinke-Nendza et al., 2021; Schinke-Nendza and Weber, 2024).

Therefore, we propose a multivariate probabilistic model to generate solar PV forecast trajectories based on repeated updates that are capable of coping with the stochastic evolutions of infeed expectations. We are hence extending forecast methodologies typically applied to wind power (Kolkmann et al., 2024), to imbalanced data of day and night-time forecasts for solar PV. This model generates probabilistic forecasts by providing continuous forecast trajectories, offering consistent scenarios particularly valuable in intraday markets and real-time decision-making (Zalzar et al., 2020). A key innovation of this approach is its ability to handle the inherent data imbalances caused by day and night cycles, improving both the accuracy and robustness of solar PV forecasts.

By capturing the complex dependencies and variability in solar generation, this model enhances the ability of system operators and market participants to manage the challenges of integrating intermittent renewable energy sources (Yang et al., 2022). This contributes to more efficient and reliable energy system operations (Boehnke et al., 2024), ultimately supporting the broader goals of decarbonization and energy transition (Fernández et al., 2023).

1.2. Contributions and structure of this paper

Decision-making processes in energy systems increasingly rely on accurate forecasts. Yet, most available techniques consist of predicting realization time series or point forecasts rather than forecast trajectories that are consistent with forecast updates. Addressing this gap, Kolkmann et al. (2024) developed a multivariate model for wind power, treating forecast updates as stochastic processes to generate continuous forecast trajectories. While this method, in general, improves the continuous decision-support based on energy forecasts (Boehnke et al., 2024), transferring it to solar PV forecasts presents unique challenges. Specifically, the imbalanced nature of PV data — caused by the binary distinction between day and night — limits the availability of forecast updates across all horizons, making stochastic modeling more complex. Consequently, this paper extends the work of Kolkmann et al. (2024) by addressing these imbalances in PV forecast trajectories through comprehensive data handling and algorithmic methods, thus enabling more accurate and consistent probabilistic forecasting. This drives the contributions of this paper:

1. We develop a dedicated approach for multivariate probabilistic forecasts for solar PV generation. This approach enables continuous decision support by

providing forecast trajectories with consistent forecast updates — crucial for the decision-making of traders and system operators.

2. In this context, we comprehensively address the issue of imbalanced data for continuous PV forecasts based on various — inter alia copula-based — modeling techniques for multivariate probabilistic forecasts.
3. Different methodical approaches are introduced on the data and algorithm levels to handle the data imbalance (i.e., due to nocturnal hours) in the training and test data. We systematically compare different methodical approaches using multivariate probabilistic forecast metrics and identify a benchmark for further investigation.
4. A case study for short-term solar PV forecasts is presented based on real-world data from the French TSO RTE. A comprehensive assessment of the different multivariate probabilistic modeling approaches with different underlying data handling methods is conducted for an electrical region in France. Furthermore, the model performance is assessed in detail for data at different spatial granularities in France.

The paper is structured as follows, see Figure 1: Section 2 presents the foundations of multivariate probabilistic forecasting, focusing on renewable energy generation. It emphasizes the modeling of forecast updates as stochastic processes. Additionally, this section explores the concept of forecast trajectories, which involves modeling a sequence of forecasts over different horizons to improve accuracy and reliability. Section 3 delves into the modeling of temporal interdependencies and methods for handling imbalanced data. It introduces techniques for capturing dependencies across forecast horizons. Section 4 introduces the evaluation metrics used to assess the accuracy and reliability of the forecast models. Specifically, it explains the Energy Score and Variogram-based Score and the statistical significance tests applied to compare different forecasting approaches. Section 5 applies the proposed methodology to a real-world case study involving short-term solar PV forecasts in a region of France. This section evaluates the performance of various models and data handling techniques, with an extended analysis that examines the impact of different spatial granularities and forecast horizons. Section 6 concludes the paper by summarizing the key findings, discussing the implications for both research and practical applications, and suggesting directions for future work in renewable energy forecasting.

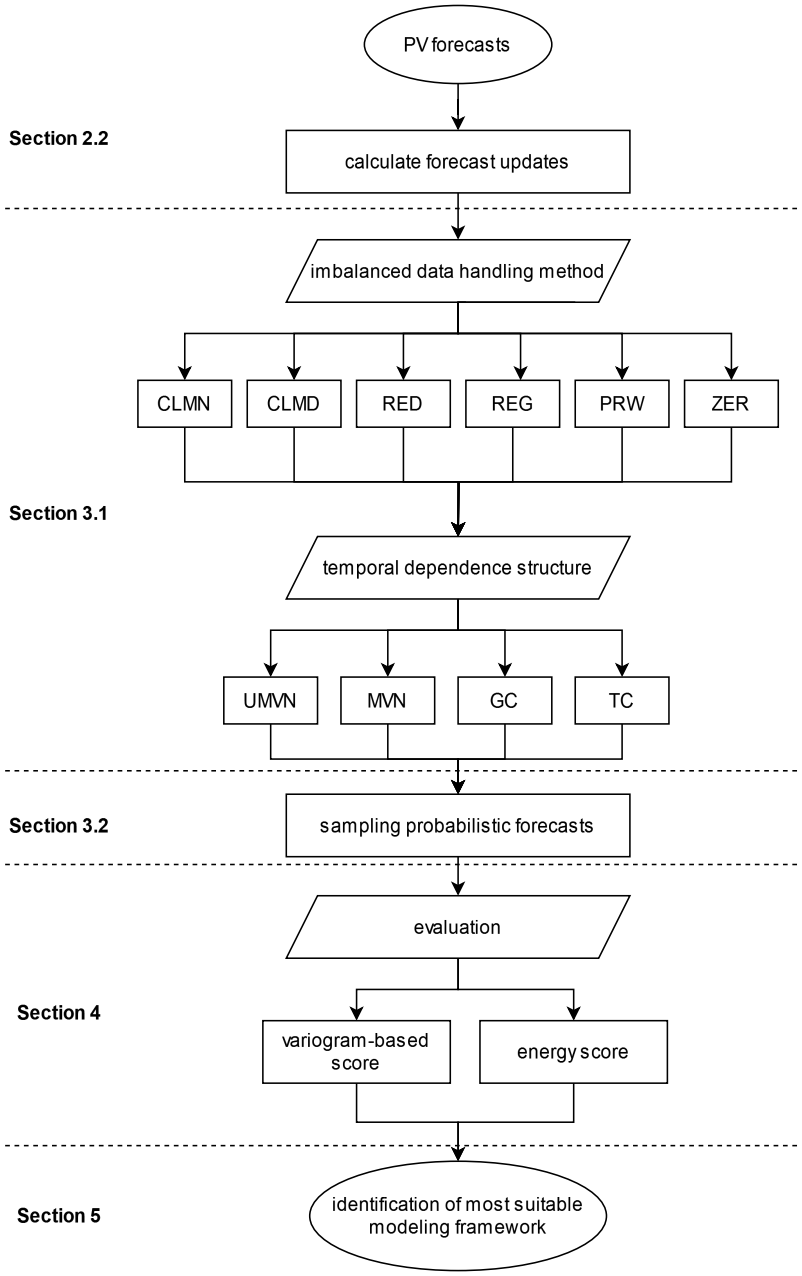

Figure 1: General methodology of this paper

2. Multivariate probabilistic forecasting and data imbalances

This Section explores multivariate probabilistic forecasting in renewable energy, focusing on refining predictions through forecast updates and handling uncertainties (cf. Section 2.1). Section 2.2 explains how forecast trajectories can be modeled as stochastic processes to improve decision-making, especially in energy trading and risk management. Section 2.3 also addresses challenges posed by imbalanced data in solar forecasts, proposing data-level and algorithm-level approaches to manage the day-night imbalance without traditional balancing techniques.

2.1. Overview on multivariate probabilistic forecasting

Over the past decades, numerous methods for forecasting renewable energy generation have been developed and studied (Sweeney et al., 2020). In this context, (multivariate) probabilistic forecasting sticks out, focusing on techniques to quantify the uncertainty in renewable energy generation predictions (Hong et al., 2020). Probabilistic forecasts provide a comprehensive view of uncertainty by offering a range of potential outcomes, unlike single-point deterministic forecasts (Gneiting and Katzfuss, 2014). This enhances decision-making, especially in volatile energy markets, by allowing for better risk management and reduced balancing costs. Probabilistic methods are adaptable to real-time updates and can incorporate spatial and temporal dependencies, improving the accuracy and robustness of predictions (Pinson et al., 2009). Two main perspectives can be distinguished within this forecasting framework: *forecast errors* and *forecast updates*. *Forecast errors* can be derived as the difference between deterministic point forecasts and actual observations. Probabilistic models can be trained to handle the resulting uncertainties from these imperfect forecasts (e.g. Golestaneh et al., 2016; Hong et al., 2020). Related methods help understand and improve the accuracy of initial predictions (Peña and Toth, 2014). *Forecast updates*, in contrast, consider the continuous flow of new information over time, with such updates corresponding to stochastic processes (Krzysztofowicz, 1987; Altendorfer and Felberbauer, 2023). Through a probabilistic treatment of forecast updates, prediction ranges, and multivariate forecast distributions can be continuously refined based on the latest available data, enabling consistent decision-making (Kolkmann et al., 2024).

For this reason, the remainder of the paper focuses on *forecast updates*. Understanding the interdependencies between these updates is crucial for improving risk assessment and management. As highlighted by Garnier and Madlener (2015), real-time forecasting updates significantly enhance decision-making processes, particularly for portfolio managers in power systems. These updates allow for the development of more effective trading strategies and better risk management. In intra-

day markets, rapidly incorporating new information is vital for optimizing trading positions, adjusting bids dynamically, and managing forecast errors — ultimately reducing costs related to energy imbalances (Koch, 2021). Gürtler and Paulsen (2018) further explore the influence of renewable energy forecasts on electricity prices in the German market, demonstrating that wind and solar PV generation typically lowers prices.

2.2. Modeling forecast updates and trajectories

In general, we consider *forecast trajectories* of solar generation, defined as a sequence of forecasts $p_{t,T}$ that are issued at different *forecast times* t (time of information update) for the same *delivery time* T (Kolkmann et al., 2024). The *forecast horizon* k can be defined as $k = T - t$. Forecast trajectories for two different T are shown in Figure 2 together with the updates resulting from advancing t. The forecast errors are also shown for clarification and differentiation.

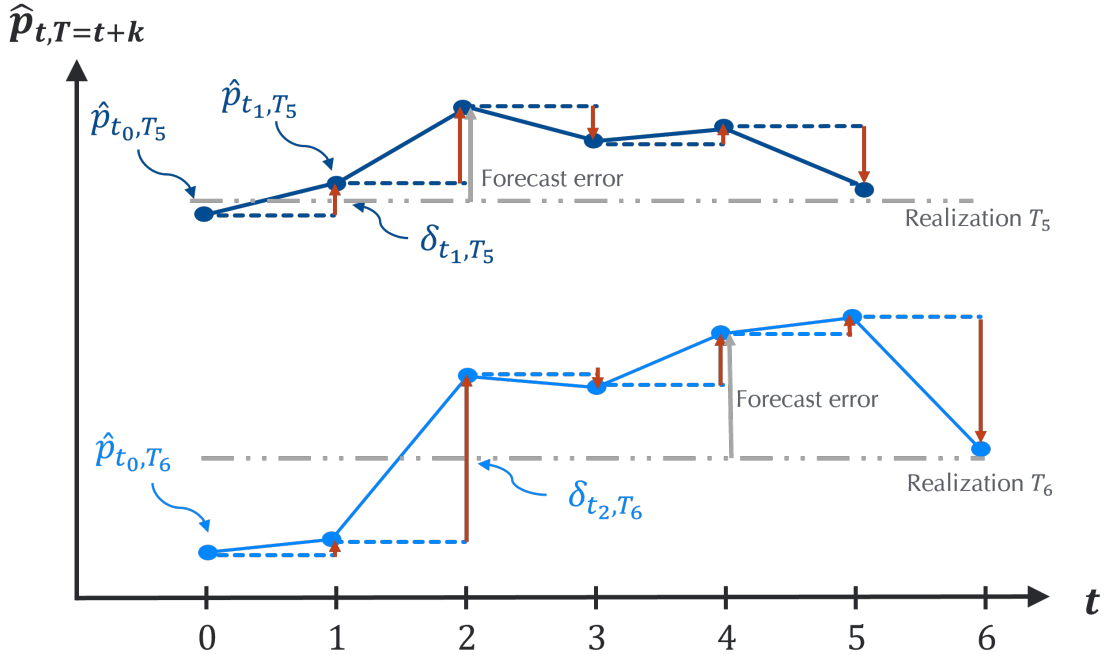

Figure 2: Forecast trajectories for two delivery times T. The updates are marked in red, and the errors are in grey arrows. The focus is on the correlation between updates across different forecast horizons k.

To obtain (normalized) *forecast updates*, we normalize the data using the clear sky production p_{CS}, i.e., the estimated maximum potential power output of the PV

system under ideal weather conditions without any atmospheric obstructions. Thereupon, we consider a matrix $\hat{\boldsymbol{p}}_{|\mathcal{T}|,|\mathcal{K}|} = (\hat{p}_{t,t+k})_{t\in\mathcal{T},k\in\mathcal{K}}$ of these normalized forecasts with elements $\hat{p}_{t,t+k} \in [0,1]$. The forecast updates can be calculated as follows

$$\delta_{t,t+k} = \hat{p}_{t,t+k} - \hat{p}_{t-1,t+k}, \qquad \forall t \in \mathcal{T}_{train}, k \in \mathcal{K}, \tag{1}$$

with $\delta_{t,t+k} \in [-1,1]$. The forecast trajectories for two different delivery times T and the corresponding forecast updates and errors are shown in Figure 2. We focus on the temporal interrelation between forecast trajectories, i.e., the *inter-correlation* of forecast trajectories. In contrast, the interrelation within one forecast trajectory, i.e., the *intra-correlation*, should be negligible (Samuelson, 1965; Kolkmann et al., 2024). Structuring the updates in a matrix $\boldsymbol{\Delta}_{|\mathcal{T}_{train}|,|\mathcal{K}|} = (\delta_{t,t+k})_{t\in\mathcal{T}_{train},k\in\mathcal{K}}$ analogous to Figure 3, we thus analyze the interrelation of the columns. In this illustration, the diagonal movement from the top right to the bottom left corresponds to a trajectory: For a constant T, the forecasts are updated as t progresses. We also have a vector $\hat{\boldsymbol{p}}_{|\mathcal{T}|} = (p_T)_{T\in\mathcal{T}}$ of normalized (pseudo-)observations.

2.3. Imbalanced data in solar forecast trajectories

In this paper, we address the challenges posed by imbalanced data in multivariate probabilistic forecasts of PV generation, focusing on predicting consistent forecast trajectories. The following distinctions are pertinent: *Imbalanced data* refers to any dataset with an unequal distribution of observations over classes, with some classes having significantly more instances than others (see, e.g. He and Garcia, 2009; Haixiang et al., 2017). This disparity can bias machine learning algorithms, leading them to favor most classes during training. *Unbalanced data* refers to a distribution over classes that is not balanced due to incomplete data or changes in the statistical population (see, e.g. Baltagi and Song, 2006; Baltagi and Liu, 2020). Unlike imbalanced data, which are inherently unequally distributed, unbalanced data imply a certain degree of incompleteness or correctable class distribution. Meanwhile, *missing data* occurs when some observations or entries are absent due to reasons such as errors in data collection, data loss, or non-responses in surveys. Proper handling of missing data is crucial, as it can lead to biased estimates and reduce the statistical power of analyses if not addressed appropriately (see, e.g. Little and Rubin, 2019; Tawn et al., 2020). The primary distinction between imbalanced, unbalanced, and missing data lies in their impact on dataset integrity: imbalanced data refer to an inherently unequal representation of classes, affecting classification (Haixiang et al., 2017), whereas unbalanced and missing data involve the absence of values, potentially compromising the accuracy and reliability of data analysis (see Baltagi and Liu, 2020; Tawn et al., 2020).

In the case of PV, there are two classes of observations: periods (hours) during daytime and periods during nighttime. The considered normalized forecast updates $\delta_{t,t+k}$ are limited to the domain $[-1, 1]$ during day-time and become zero during nighttime. In this context, Figure 3 illustrates the twofold structure of imbalanced training data. This imbalanced data serves as a basis for creating multi-variate probabilistic PV forecasts (i.e., scenarios of forecast trajectories) by considering the correlation across multiple forecast horizons. The training data is structured as follows: First, the row-wise consideration per forecast time t defines a forecast issued at time t for k forecast horizons. Each issued forecast possibly covers multiple day-time and night-time periods—or may exclusively consist of day-time and night-time periods. Hence, the imbalance ratio, measuring this disparity, varies significantly, depending on the forecast time t, the season (sunrise and sunset), and the number of forecast horizons k. Second, the column-wise consideration per forecast horizon k yields the distributions of the two classes. In the case of night-time periods, the forecast updates become zero per definition. During daytimes $\delta_{t,t+k} \in [-1, 1]$ holds.

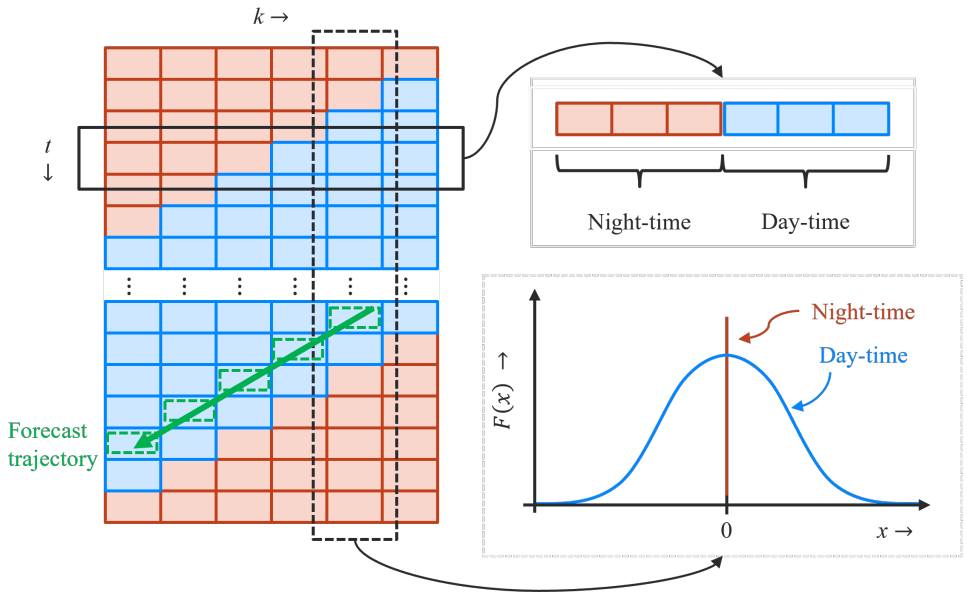

Figure 3: Twofold structure of imbalanced data due to daytimes and nighttimes: Imbalance per forecast issuing time t and per forecast horizon k.

Consequently, handling imbalanced data is crucial, especially when modeling consistent forecast trajectories. Simply eliminating issued forecasts with imbalances from the dataset will at least cause a loss of information. This loss may be significant, or it may not be that important — depending on whether the maximum forecast horizon

K is larger than the number of consecutive daytime periods. A common approach is to limit investigations of (multivariate) probabilistic PV forecasts to specific trajectories (e.g. Golestaneh et al., 2016), or timeseries data (e.g. Schinke-Nendza et al., 2021; Schinke-Nendza and Weber, 2024), that exclude nocturnal hours—thus avoiding the issue of data imbalances. However, these approaches lack the capability to provide continuous and consistent forecast trajectories that are pivotal for the decision-support of traders and system operators. A joint consideration of night and day periods would heavily distort column-wise distributions in these approaches. Consequently, this necessitates adequate techniques for dealing with the previously raised issue of imbalanced datasets.

However, in contrast to the common viewpoint on imbalanced data, cf. He and Garcia (2009), we are not interested in balancing data, e.g., by oversampling or undersampling, but in obtaining unary-classified data (i.e., of daytime periods) exclusively. Similar to the techniques of treating imbalanced data, cf. He and Garcia (2009), two approaches can be distinguished:

- First, addressing imbalances at the *data level*. Night-time instances in the forecast are replaced to create a unary-classified (daytime) dataset. Similar to oversampling techniques for imbalanced data (He and Garcia, 2009), this generates synthetic data, potentially leading to over-fitting. Nevertheless, this data-level adjustment facilitates the application of standard methods for training multivariate probabilistic forecasting models.

- Second, addressing imbalances at the *algorithm level*. Existing training methods for multivariate probabilistic forecasting are tailored to handle imbalanced data directly—allowing the usage of imbalanced data. This approach ensures that the models are adapted to reflect inherent data imbalances during training and sampling, i.e., Monte Carlo simulations.

Therefore, this paper aims to explore the extent to which data-based and algorithm-based techniques improve the quality of multivariate probabilistic forecasts across various models, with a particular focus on achieving consistent forecast trajectories, following Kolkmann et al. (2024).

3. Modeling temporal interdependencies for imbalanced data

This Section outlines the methodology for developing multivariate probabilistic forecasts for solar PV power, focusing on data preprocessing, modeling temporal interdependencies, and handling imbalanced data. In Subsection 3.1, methods for

modeling temporal interdependencies are introduced, including separating training and test datasets and modeling marginal distributions of forecast updates. Additionally, techniques to address data imbalances at both the algorithmic and data levels are discussed. In Subsection 3.2, the generation of forecast trajectories using probabilistic models and Monte Carlo simulations is covered, ensuring consistent and accurate forecasts over time.

3.1. Modeling interdependencies

For modeling temporal interdependencies, we can separate both matrices, i.e., forecast updates and observed forecast trajectories, into two sets $\mathcal{T}_{train}$ and $\mathcal{T}_{test}$. Here, $\mathcal{T}_{train}$ is the training dataset that serves as a basis for training probabilistic models introduced below. Thereupon, multivariate probabilistic forecasts can be obtained, e.g., using a Monte Carlo simulation, and the performance of different models can then be tested for the test dataset $\mathcal{T}_{test}$.

3.1.1. Marginal distributions

The marginal distribution of forecast updates $\delta_{t,t+k}$ can be modeled using empirical cumulative distribution functions (CDF), see McNeil et al. (2015). The marginal distribution F_k for $k \in \mathcal{K}$ can then be approximated by the empirical CDF as

$$F_k(\delta_k) \approx \widetilde{F}_k(\delta_k) = \frac{1}{|\mathcal{T}_{train}| + 1} \sum_{\tau \in \mathcal{T}_{train}} I_{\left\{\widetilde{\delta}_{k,\tau} \leq \delta_k\right\}} \tag{2}$$

where $I_{\left\{\widetilde{\delta}_{k,\tau} \leq \delta_k\right\}}$ is the indicator[1] and $\left(\widetilde{\delta}_{k,\tau}\right)_{k \in \mathcal{K}, \tau \in \mathcal{T}_{train}}$ describes the training dataset.[2] Furthermore, $u_k = F_k(\delta_k)$ describes the probability of δ_k for $k \in \mathcal{K}$.

In extreme value theory, these empirical CDFs are known to be poor estimators of the underlying distribution in the tails (McNeil et al., 2015). However, for our applications, where forecast updates are strictly limited to a given interval, i.e., $\delta_{t,t+k} \in [-1, 1]$, the boundedness of the distribution may justify the use of empirical CDFs.

3.1.2. Handling data imbalances

As discussed in Section 2.3, we need to develop adequate techniques to deal with imbalanced data before capturing the temporal dependencies between the updates of

[1] Returns 1 if the condition in the subscript $\{\cdot\}$ holds and 0 if not.

[2] Note that the denominator $|\mathcal{T}_{train}| + 1$ in Eq. (2) ensures that the data strictly lies within the unit cube.

the different forecast horizons in Section 3.1.3. At the *algorithmic level*, we introduce the new pairwise methodological approach PRW. It is characterized by the pairwise handling of relationships between forecast horizons, which can be modeled using various approaches. For each pair of forecast horizons i and j within u_k, the PRW method models the joint behavior of $u_i = F_i(\delta_i)$ and $u_j = F_j(\delta_j)$. For each pair of horizons, the relevant parameters — depending on the chosen temporal dependence structure (cf. Section 3.1.3) — are estimated using the available data. After estimating the pairwise relationships for all horizon pairs, these are combined into a complete $k \times k$ matrix, encapsulating the dependencies across all horizons. It is a flexible and advanced technique designed to manage datasets imbalances, preserving the day-time data's integrity.

For comparison purposes, we also implement the methods on the *data level* given in Table 1. Each method addresses the issue of missing or incomplete night-time data differently, impacting the overall analysis of temporal dependencies in solar PV forecasting.

Method	Description
CLMN	For the column-wise mean method (CLMN), every night-hour value is replaced with the mean of the whole period. This approach simplifies the dataset but may introduce bias by artificially inflating night-hour values.
CLMD	The column-wise median method (CLMD) is similar to CLMN, but instead of the mean, the median value of the entire period is used for night-hour entries. This offers more robustness against outliers compared to the mean-based approach.
RED	The reduced approach (RED) retains only rows with day-time entries in every column, eliminating rows with missing data for night hours. However, as the forecast horizon increases, the data available for calculating temporal dependencies decreases, potentially reducing model accuracy.
REG	A regression-based imputation (REG) is applied to fill in nocturnal data, using existing day-time data to predict missing values. This method enhances the completeness of the data matrix, enabling a more accurate analysis of temporal dependencies. The model is trained using neighboring columns with day-time data, and predicted values are inserted back into the dataset.
ZER	This method sets all values for night hours to zero (ZER), offering a straightforward and intuitive approach. While simple, it is a useful benchmark for comparing more complex imputation techniques.
PRW	New methodological approach with pairwise (PRW) estimation of relationships between all pairs of forecast horizons (cf. 3.1.2).

Table 1: Overview of methods for handling imbalanced data at the data and algorithmic level

3.1.3. Temporal dependence structure

The distribution and temporal interdependence of forecast updates $\delta_{t,t+k}$ can be described using various statistical models. Among these, the multivariate normal MVN and uncorrelated multivariate normal UMVN distributions rely on standard statistical assumptions, thus serving as benchmarks.

In the MVN framework, the forecast updates $\delta_{t,t+k}$ are assumed to follow a normal distribution with a mean vector $\boldsymbol{\mu}$ and a covariance matrix $\boldsymbol{\Sigma}$. The joint density

function is given by

$$f(\boldsymbol{\delta}) = \frac{1}{(2\pi)^{K/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{\delta}-\boldsymbol{\mu})^{\top}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\delta}-\boldsymbol{\mu})\right), \tag{3}$$

where K is the number of forecast horizons, and $\boldsymbol{\delta}$ is the vector of forecast updates.

In the UMVN approach, the forecast updates are assumed to be uncorrelated, meaning the covariance matrix $\boldsymbol{\Sigma}$ is diagonal. The joint density function simplifies to

$$f(\boldsymbol{\delta}) = \prod_{k=1}^{K} \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{1}{2}\frac{(\delta_k-\mu_k)^2}{\sigma_k^2}\right), \tag{4}$$

where μ_k and σ_k^2 are the mean and variance of the forecast update for horizon k, respectively.

By employing these methods, the temporal interdependence of forecast updates and errors may be captured based on an established and robust framework for multivariate modeling. However, more sophisticated approaches such as copulas are commonly used in forecasting applications to capture better the dependence structure of multiple forecasts (see, e.g. Golestaneh et al., 2016). Copulas can describe the temporal interdependence of forecast updates $\delta_{t,t+k}$. Recalling that δ_k for $k \in \mathcal{K}$ denotes the random variable of updates, the multivariate distribution F (with marginal distributions F_k for $k \in \mathcal{K}$) can be represented by the copula C as

$$F\left(\delta_1, \ldots, \delta_{|\mathcal{K}|}\right) = C\left(F_1\left(\delta_1\right), \ldots, F_{|\mathcal{K}|}\left(\delta_{|\mathcal{K}|}\right)\right), \tag{5}$$

based on the theorem of Sklar (1959).

Although multiple copulas are defined for bivariate distributions, only the Gaussian copula GC and the t-copula TC may be directly generalized to higher dimensional multivariate cases. Gaussian copulas are commonly used in forecasting applications. The stochastic dependence structure of $\boldsymbol{u}$ – the probabilities of forecasting deviations – is thereby modeled using a copula C_G based on the multivariate normal distribution according to

$$C_G\left(\boldsymbol{u}|\boldsymbol{R}_G\right) = \frac{1}{\sqrt{\det \boldsymbol{R}_G}} \exp\left(-\frac{\boldsymbol{x}^{\top}\left(\boldsymbol{R}_G^{-1}-\boldsymbol{I}\right)\boldsymbol{x}}{2}\right), \tag{6}$$

where $\boldsymbol{R}_G$ is the Gaussian copula's correlation matrix, $\boldsymbol{x} = \left(\Phi^{-1}\left(u_1\right), \ldots, \Phi^{-1}\left(u_{|\mathcal{N}|}\right)\right)^{\top}$ is the vector of inverse cumulative standard normal distribution and I the identity matrix (McNeil et al., 2015; Bouyé et al., 2000).

Despite being used frequently in forecasting applications, Gaussian copulas cannot describe the tail-dependence of continuous marginal distributions, i.e., the probability of co-occurrence of relatively rare events (McNeil et al., 2015). Therefore, we propose a t-copula to model the temporal interrelation of forecasting updates for multiple forecast horizons. In this context, the density of the t-copula is given by

$$C_t\left(\boldsymbol{u}|\boldsymbol{R}_T,\nu\right) = \frac{\Gamma\left(\frac{\nu}{2}\right)^{|\mathcal{N}|-1}\Gamma\left(\frac{\nu+|\mathcal{N}|}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)^{|\mathcal{N}|}\sqrt{\det\boldsymbol{R}_t}}\cdot \\ \cdot\left(1+\frac{\boldsymbol{x}^\top\boldsymbol{R}_t^{-1}\boldsymbol{x}}{\nu}\right)^{-\frac{\nu+|\mathcal{N}|}{2}}\prod_{n\in\mathcal{N}}\left(1+\frac{x_n^2}{\nu}\right)^{\frac{\nu+1}{2}}, \tag{7}$$

with $\Gamma(\cdot)$ the gamma function, ν the degrees of freedom, and $\boldsymbol{R}_T$ the correlation matrix of the t-copula (McNeil et al., 2015; Bouyé et al., 2000).

3.2. Obtaining forecast trajectories

We are interested in *forecast trajectories*, which involve simulating the evolution of forecasts over the intraday market period based on a sequence of forecast updates. To generate these trajectories, we fit multivariate probabilistic models to the data of forecast updates $\boldsymbol{\delta} \in \mathbb{R}^{|\mathcal{T}_{train}|\times|\mathcal{K}|}$. Thereupon, we can draw samples from the multivariate probabilistic distributions, with $s \in \mathcal{S}$ scenarios to generate a multivariate probabilistic forecast. We simulate forward from the first known forecast for T, constructing anteceding consistent forecasts by adding all subsequent updates from the sample.

Combining the sample $\delta_{t,t+k}^{(s)}$ of forecast updates with the normalized forecasts of the test dataset $\hat{\boldsymbol{p}}_{|\mathcal{T}_{test}|,|\mathcal{K}|}$ we can derive a probabilistic forecast as follows

$$\widetilde{p}_{t,t+k}^{(s)} = \widetilde{p}_{t-1,t+k}^{(s)} + \delta_{t,t+k}^{(s)}, \qquad \forall s\in\mathcal{S}, t\in\mathcal{T}_{test}, k\in\mathcal{K}, \tag{8}$$

and

$$\widetilde{p}_{t-1,t+k}^{(s)} = \widetilde{p}_{t-2,t+k}^{(s)} + \delta_{t-1,t+k}^{(s)}, \qquad \forall s\in\mathcal{S}, t\in\mathcal{T}_{test}, k\in\mathcal{K}, \tag{9}$$

thus

$$\widetilde{p}_{t,t+k}^{(s)} = \hat{p}_{t+k-|\mathcal{K}|,t+k} + \sum_{\kappa=0}^{|\mathcal{K}|-k-1}\delta_{t-\kappa,t+k}^{(s)}, \qquad \forall s\in\mathcal{S}, t\in\mathcal{T}_{test}, k\in\mathcal{K}, \tag{10}$$

where $s \in \mathcal{S}$ defines the forecast trajectories (scenarios).

Reversing the normalization ensures that energy-based metrics like the Energy and Variogram-based Score accurately reflect actual power values. This process

involves multiplying the normalized forecast values by the clear sky production and dividing by the installed capacity. This step is crucial for the correct interpretation and meaningful comparison of forecast performance in real-world terms.

4. Evaluation metrics

To assess and compare the forecast performance of the different approaches and models for handling imbalanced data, we use two standard evaluation metrics, the Energy Score (Section 4.1) and the Variogram-based Score (Section 4.2).

4.1. Energy Score

The Energy Score (ES) is, according to Gneiting et al. (2007), defined as

$$\begin{aligned} ES_t = &\frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \left\| \hat{\boldsymbol{p}}_{\boldsymbol{t}} - \widetilde{\boldsymbol{p}}_t^{(s)} \right\|_2 \\ &- \frac{1}{2\,|\mathcal{S}|^2} \sum_{s_1, s_2 \in \mathcal{S}} \left\| \widetilde{\boldsymbol{p}}_t^{(s_1)} - \widetilde{\boldsymbol{p}}_t^{(s_2)} \right\|_2, \end{aligned} \tag{11}$$

where $\widetilde{\boldsymbol{p}}_t^{(s)} = \left(\widetilde{p}_{t,t+k}^{(s)}\right)_{k \in \mathcal{K}} = \left(\widetilde{p}_{t,t+1}^{(s)}, \ldots, \widetilde{p}_{t,t+|\mathcal{K}|}^{(s)}\right)^\top$ is the vector of forecasts in scenario s issued at time-step t and $\hat{\boldsymbol{p}}_t$ the vector of the observed forecasts at the same time-step. Here, $\widetilde{\boldsymbol{p}}_t^{(s)}$ can be obtained from the forecast update process via Eq. (10). Following Pinson and Girard (2012), we can approximate the Energy Score as

$$\begin{aligned} ES_t = &\frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \left\| \hat{\boldsymbol{p}}_{\boldsymbol{t}} - \widetilde{\boldsymbol{p}}_t^{(s)} \right\|_2 \\ &- \frac{1}{2\,(|\mathcal{S}| - 1)} \sum_{s \in \mathcal{S} \setminus \{|\mathcal{S}|\}} \left\| \widetilde{\boldsymbol{p}}_t^{(s)} - \widetilde{\boldsymbol{p}}_t^{(s+1)} \right\|_2, \end{aligned} \tag{12}$$

The Energy Score is frequently used to evaluate probabilistic forecasts (Bjerregård et al., 2021), but Scheuerer and Hamill (2015) showed that the ES lacks sensitivity to incorrectly specified correlations in multivariate forecasts. While it effectively detects mean-biased forecasts, it is unreliable when it comes to identifying incorrect dependency structures or inaccurate variance predictions. That is why the Variogram-based Score (VS) is employed as an additional proper scoring rule to evaluate the temporal dependence structure of multivariate probabilistic forecasts.

4.2. Variogram-based Score

Given the explanations in Section 3.2, we construct forecast trajectories by simulating from the first known forecast for a fixed delivery time T. These forecasts are compared to each other for all horizons $k \in \mathcal{K}$ and the Variogram-based Score (VS) follows as

$$VS_t = \sum_{k_1,k_2 \in \mathcal{K}} w_{k_1,k_2} \left(\frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \left| \widetilde{p}^{(s)}_{T-k_1,T} - \widetilde{p}^{(s)}_{T-k_2,T} \right|^{\gamma} - \left| \hat{p}_{T-k_1,T} - \hat{p}_{T-k_2,T} \right|^{\gamma} \right)^2, \tag{13}$$

where γ is the order of the VS (whereby Scheuerer and Hamill (2015) propose 0.5 for accurate results) and w_{k_1,k_2} is a weighting factor. Like the ES, this scoring rule is also negatively oriented: lower values indicate a higher forecasting accuracy.

4.3. Statistical significance test

To evaluate the comparative performance of the two models, we employ the Diebold-Mariano (DM) test, following Gneiting and Katzfuss (2014). It is designed to test the null hypothesis that the predictive accuracy of the models is statistically indistinguishable. The test statistic is calculated based on the mean differences of the performance scores ES and VS. For a given significance level, the DM test statistic can indicate which model shows superior predictive performance (Gneiting and Katzfuss, 2014). This approach ensures a robust assessment of model efficacy in probabilistic forecasting contexts.

5. Application

In this section, we apply the methodology to a case study in France, evaluating the models' performance across different spatial scales and forecast horizons to validate the approach using real-world forecast data. In this context, Section 5.1 describes the dataset that serves as a case study in this work. Thereupon, in Section 5.2 the results are analyzed—assessing the probabilistic forecast approaches as outlined in Section 3.2 based on the scoring methods detailed in Section 4.

5.1. Case study

The case study used in this work is based on a forecasting dataset provided by RTE, the electricity transmission system operator of France. In this context, data

is available on three different levels of spatial granularity: 1) substation level, 2) regional level, and 3) country level. On the *substation level*, the dataset includes aggregated data from five substations in the Centre-Val de Loire region. On the *regional level*, the dataset includes seven electrical areas R1 to R7, which do not necessarily correspond to geographical regions, with regional dispatching in the cities of Lille, Nancy, Lyon, Marseille, Toulouse, Nantes, and St Quentin. On the *country level*, the data covers the entirety of France. Figure 4 depicts the substations and regions.

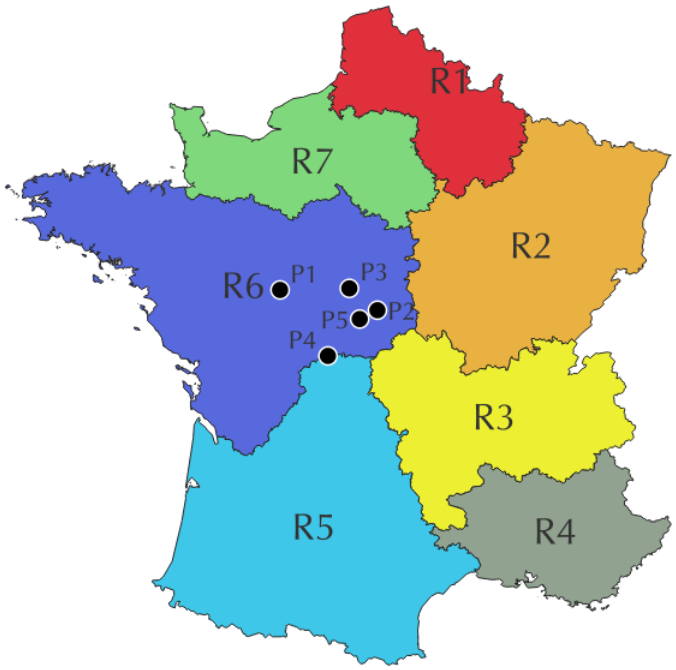

Figure 4: Overview of regions and selected substations in France

Our evaluation process involves generating 1,000 Monte Carlo simulations for various imputation methods, based on a month-wise train-test split of the entire dataset of 2021, i.e., separately using each month as test data, while the remainder of the data serves as training data for the model.

We structure our analysis of the results twofold: First, in Section 5.2.1, insights for the different probabilistic forecasting approaches are given on a regional level—using one exemplary region, with a long forecast horizon of 12 hours, i.e., where data

is highly imbalanced. Second, in Section 5.2.2 follows with an in-depth analysis for different levels of spatial granularity and forecast horizons.

5.2. Results

In the following, the results of different probabilistic forecasting approaches are assessed to provide insights into the effectiveness of different data handling methods and temporal dependence structures, as well as their impact on forecast accuracy. By systematically comparing these methods, we aim to identify the most robust approach for capturing temporal correlations of PV forecast updates.

5.2.1. Regional-level with a 12 hour forecast horizon

For our first analyses, we exemplarily consider region R6 around Nantes, see Figure 4, which corresponds to the regions Bretagne, Pays de Loire, Centre-Val de Loire and the former region Poitou-Charentes, and a forecast horizon of 12 hours. A visualization for the simulations using PRW and TC is given in Figure 5 as a quantile plot of simulations and observations. They all start at the same point,

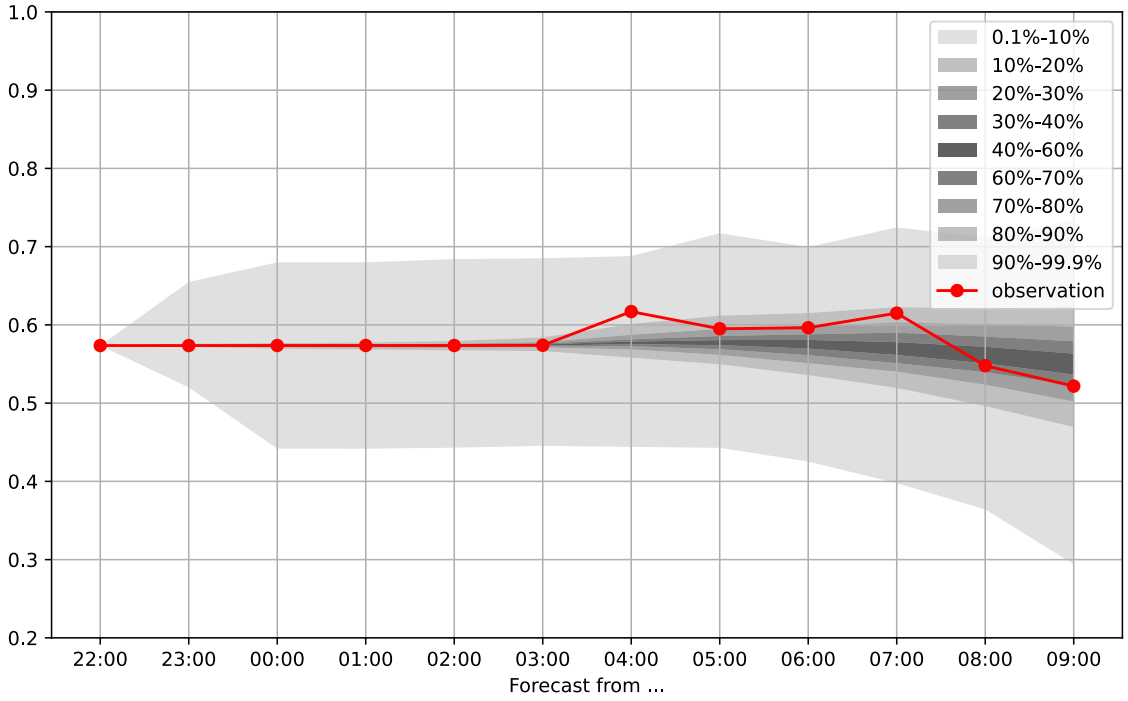

Figure 5: Quantile plot of 1000 PRW-TC simulations for T = 2nd July 2021 10:00 am with $k = 12$ in R6 region

which is the first available forecast, and then deviate since they use different samples for the forecast trajectories. Table 2 depicts – as an example – the results of the ES for different distribution models as monthly means, given the data handling method PRW. The same does Table 3 for the VS.

Table 2: Energy Scores of probabilistic forecast update models based on different distribution assumptions and the PRW data handling method for the forecast horizon $k = 12$ in Region R6

Month	UMVN	MVN	GC	TC
1	0.0396	**0.0396**	**0.0395**	0.0397
2	**0.0525**	**0.0523**	0.0526	0.0529
3	0.0488	0.0488	**0.0483**	**0.0485**
4	0.0493	0.0493	**0.0481**	**0.0481**
5	0.0569	0.0569	**0.0564**	**0.0566**
6	0.0565	0.0566	**0.0563**	**0.0563**
7	0.0535	0.0537	**0.0534**	**0.0532**
8	0.0612	0.0612	**0.0609**	**0.0611**
9	0.0482	0.0484	**0.0480**	**0.0480**
10	0.0452	0.0450	**0.0450**	**0.0448**
11	**0.0466**	**0.0464**	0.0467	0.0468
12	**0.0478**	**0.0479**	0.0481	0.0481

Ranking: **1st**, **2nd**, 3rd

Building on the discussions from Section 4, it is essential to reiterate that the ES and the VS differ in their structural approaches. The ES primarily evaluates the probability distribution for a given combination of forecast time and horizon, making it suitable for assessing overall distributional similarity. In contrast, the VS is more specific as it focuses on reproducing differences for updates across various forecast horizons, which is crucial for accurately capturing the temporal dynamics. For our application, a preferred model should perform well regarding the Energy Score, positioning itself as a strong aspirant. However, the main evaluation comes through the Variogram-based Score, where the best model should outperform others by effectively capturing the temporal correlations and differences in forecast updates.

An analysis of the ranking of the scores suggests that the copula models exhibit the best performance for the PRW data handling method. For the ES (cf. Table 2), copula models perform best in most cases. In terms of the VS (cf. Table 3), TC, and GC consistently rank first and second best. We employ the Diebold-Mariano test to determine the statistical significance of the observed differences in scores (see

Table 3: Variogram-based Scores of probabilistic forecast update models based on different distribution assumptions and the PRW data handling method for the forecast horizon $k = 12$ in Region R6

Month	UMVN	MVN	GC	TC
1	0.4354	0.4338	**0.4020**	**0.4037**
2	0.5750	0.5737	**0.5426**	**0.5460**
3	0.5880	0.5858	**0.5018**	**0.5037**
4	0.7564	0.7582	**0.6240**	**0.6236**
5	0.6617	0.6613	**0.5675**	**0.5678**
6	0.5744	0.5751	**0.4991**	**0.4984**
7	0.6477	0.6497	**0.5486**	**0.5490**
8	0.7329	0.7347	**0.6451**	**0.6433**
9	0.5243	0.5242	**0.4437**	**0.4440**
10	0.4921	0.4909	**0.4251**	**0.4241**
11	0.4599	0.4591	**0.4518**	**0.4498**
12	0.7229	0.7230	**0.7062**	**0.7075**

Ranking: **1st**, **2nd**, 3rd

Section 4.3) at a 5 percent significance level. This test enables pairwise comparisons across all model variants, and the overall results collected in Table 4 for the Energy Score and Table 5 for the Variogram-based Score facilitate the selection of the most suitable model configuration. The DM test is applied to the underlying hourly time series, with the results aggregated into scores: each model listed in a row earns one point per month when it outperforms the model specified in the columns. The last column provides the total number of "beats" each model achieved. The last row gives information about how often, in terms of months, the model is beaten by another one. Consequently, a well-performing model should exhibit high row entries and achieve a high sum in the last column. Additionally, the selected model variant should have as many zero entries as possible in the corresponding column, indicating that it is not outperformed by any other model variant in any month.

For the ES (Table 4), the models RED-TC with a beat sum of 169, PRW-GC (166), PRW-TC (164), and RED-GC (164) stand out, having also the lowest numbers of defeats in the column sums. In the decisive VS (Table 5), the PRW method emerges as the favorite option with 214 beats and 5 losses with TC and 212 beats and 2 losses with GC. Both RED-TC and RED-GC reach 211 beats and 3 or 4 losses. Overall, it can be observed that PRW and RED, combined with copulas, perform best and are most suitable for simulating update-based forecast trajectories. Both

models differ only slightly in performance. RED is a consistent method for large and complete datasets and limited forecast horizons. As described in Section 3.1.2, using RED with large forecast horizons could remove too much data due to the twofold data structure in the case of PV. However, in specific applications, this approach may still be sufficient. The new methodological approach PRW includes more of the real-world dataset, making it slightly more accurate, although computationally more demanding.

Table 4: Energy Score: Diebold-Mariano test beats

		CLMN				**CLMD**				**PRW**				**REG**				**RED**				**ZER**				**Sum Beats**
		UMVN	MVN	GC	TC	UMVN	MVN	GC	TC	UMVN	MVN	GC	TC	UMVN	MVN	GC	TC	UMVN	MVN	GC	TC	UMVN	MVN	GC	TC	
CLMN	**UMVN**	-	0	12	12	5	5	8	9	1	1	0	0	0	0	0	0	1	1	0	0	6	6	9	8	84
	MVN	2	-	12	12	6	5	9	9	1	1	0	0	0	0	0	0	1	1	0	0	6	6	9	8	88
	GC	0	0	-	0	0	0	4	4	0	0	0	0	0	0	0	0	0	0	0	0	4	4	6	5	27
	TC	0	0	0	-	0	0	4	4	0	0	0	0	0	0	0	0	0	0	0	0	4	4	6	5	27
CLMD	**UMVN**	4	3	11	11	-	1	11	11	1	1	0	0	0	1	2	3	0	1	0	0	6	7	10	10	94
	MVN	4	4	11	11	2	-	11	11	0	1	0	0	0	1	3	3	1	1	1	0	6	6	10	10	97
	GC	0	0	4	4	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	6	6	22
	TC	0	0	5	4	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	1	1	6	6	23
PRW	**UMVN**	6	8	11	11	8	8	8	8	-	1	1	2	5	5	6	6	1	0	1	2	8	8	8	8	130
	MVN	6	7	11	11	7	8	8	8	0	-	3	3	5	5	5	6	0	2	2	3	8	8	8	8	132
	GC	10	9	11	11	8	8	9	9	3	6	-	4	7	7	8	8	5	5	2	1	9	9	9	8	166
	TC	11	10	11	11	8	8	9	9	5	4	1	-	6	6	8	9	5	6	1	1	9	9	9	8	164
REG	**UMVN**	11	11	12	12	9	8	12	11	2	2	0	0	-	1	10	11	1	3	0	0	7	7	10	10	150
	MVN	11	11	12	12	8	8	10	10	1	1	0	0	2	-	9	9	1	2	0	0	7	8	10	10	142
	GC	8	9	12	12	6	7	10	10	1	1	0	0	1	1	-	0	1	1	1	1	6	7	9	8	112
	TC	9	8	12	12	6	6	10	10	1	1	0	0	1	1	0	-	1	1	1	1	6	6	9	8	110
RED	**UMVN**	6	6	11	11	8	8	8	8	0	1	3	3	5	5	6	6	-	0	2	3	8	8	8	8	132
	MVN	6	6	11	11	8	8	8	8	2	3	3	4	5	4	6	6	2	-	2	4	8	8	8	8	139
	GC	10	10	11	11	8	8	9	9	5	5	1	1	8	6	8	8	6	4	-	1	9	9	9	8	164
	TC	11	10	11	11	8	8	9	9	4	7	1	1	7	7	9	9	5	6	2	-	9	9	8	8	169
ZER	**UMVN**	3	3	6	6	2	2	8	8	2	2	2	2	2	2	3	3	2	2	2	2	-	2	11	11	88
	MVN	3	3	6	6	2	2	8	8	2	2	2	2	2	2	2	3	2	2	2	2	3	-	11	11	88
	GC	2	2	3	4	0	0	3	2	0	0	0	0	0	0	1	2	0	1	0	0	0	0	-	1	21
	TC	2	2	4	4	1	1	3	3	1	1	1	0	1	1	1	2	1	1	1	0	0	0	4	-	35
Beaten by		125	122	210	210	110	109	179	178	32	41	18	22	57	55	87	94	36	40	20	21	131	133	193	181	

Table 5: Variogram-based Score: Diebold-Mariano test beats

		CLMN				CLMD				PRW				REG				RED				ZER				Sum Beats
		UMVN	MVN	GC	TC	UMVN	MVN	GC	TC	UMVN	MVN	GC	TC	UMVN	MVN	GC	TC	UMVN	MVN	GC	TC	UMVN	MVN	GC	TC	
CLMN	**UMVN**	-	2	11	11	9	8	11	11	4	4	0	0	1	1	0	0	4	4	0	0	9	9	11	11	121
	MVN	1	-	10	10	9	8	11	11	4	4	0	0	1	1	0	0	4	4	0	0	9	9	11	11	118
	GC	1	1	-	1	1	1	12	12	1	1	0	0	1	1	0	0	1	1	0	0	1	1	12	12	61
	TC	1	1	2	-	1	1	12	12	1	1	0	0	1	1	0	0	1	1	0	0	1	1	12	12	62
CLMD	**UMVN**	0	0	9	9	-	1	11	11	4	4	0	0	1	1	0	0	4	4	0	0	8	8	11	11	97
	MVN	0	0	9	9	2	-	11	11	4	4	0	0	1	1	0	0	4	4	0	0	9	9	11	11	100
	GC	0	0	0	0	0	0	-	0	1	1	0	0	1	1	0	0	1	1	0	0	0	0	12	12	30
	TC	0	0	0	0	0	0	1	-	1	1	0	0	1	1	0	0	1	1	0	0	0	0	12	12	31
PRW	**UMVN**	5	5	5	5	5	5	8	8	-	1	0	0	3	3	4	3	0	1	0	0	5	5	8	8	87
	MVN	5	5	5	5	5	5	7	7	0	-	0	0	3	3	4	4	1	2	0	0	5	5	8	8	87
	GC	10	10	11	11	10	10	11	11	11	11	-	1	12	12	8	8	11	11	0	1	10	10	11	11	212
	TC	10	10	11	11	10	11	11	11	11	11	0	-	12	12	8	8	11	11	2	0	10	10	11	11	214
REG	**UMVN**	8	8	8	8	8	8	11	11	7	7	0	0	-	0	4	4	7	7	0	0	8	8	11	11	144
	MVN	8	8	8	8	8	8	11	11	7	7	0	0	2	-	4	4	7	7	0	0	8	8	11	11	146
	GC	11	11	11	11	11	11	12	12	7	7	1	1	7	7	-	1	7	7	1	1	12	12	12	12	185
	TC	11	11	11	11	11	11	12	12	7	7	1	1	7	7	0	-	7	7	1	1	12	12	12	12	184
RED	**UMVN**	5	5	5	5	5	5	7	8	0	0	0	0	3	3	4	4	-	0	0	0	5	5	8	8	85
	MVN	5	5	5	5	5	5	8	8	2	2	0	0	3	3	4	4	0	-	0	0	5	5	8	8	90
	GC	10	10	11	11	10	10	11	11	11	11	0	1	12	12	8	8	11	11	-	0	10	10	11	11	211
	TC	10	10	11	11	10	10	11	11	11	11	0	1	12	12	8	8	11	11	0	-	10	10	11	11	211
ZER	**UMVN**	0	0	7	7	0	0	11	11	4	4	0	0	1	1	0	0	4	4	0	0	-	1	11	11	77
	MVN	0	0	7	8	0	0	11	11	4	4	0	0	1	1	0	0	4	4	0	0	0	-	11	11	77
	GC	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1	0	0	0	0	-	0	4
	TC	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1	0	0	0	0	2	-	6
Beaten by		101	103	157	157	120	118	211	211	104	105	2	5	86	84	56	56	103	105	4	3	137	138	238	236	

5.2.2. Extended analysis using different granularities and horizons

We now compare the results obtained across different forecast horizons and various levels of spatial granularity. As highlighted in Section 5.1, the dataset includes three different levels of spatial granularity: 1) substation level, 2) regional level, and 3) country level.

The values from the respective spatial granularity levels are averaged to provide a comprehensive summary. Tables 6 and 7 depict the results of Energy Score and Variogram-based Score for different probabilistic forecast approaches with different spatial granularity and two forecast horizons (i.e., 6 and 12 hours). The results showcase the mean scores and the deviations of various data handling methods relative to the PRW method—as it achieves the lowest scores in most cases for the decisive Variogram-based score.

Table 6: Benchmark of mean *Energy Scores* and deviations for different spatial granularit and forecast horizons.

Scope		**CLMN**	**CLMD**	**PRW**	**REG**	**RED**	**ZER**
Substation 6h	Mean Score	0.0624	0.0628	0.0613	0.0614	**0.0613**	0.0628
	Deviation	*1.74%*	*2.44%*	-	*0.09%*	*-0.04%*	*2.53%*
Substation 12h	Mean Score	0.0772	0.0776	0.0763	0.0765	**0.0763**	0.0776
	Deviation	*1.14%*	*1.63%*	-	*0.25%*	*-0.07%*	*1.65%*
Region 6h	Mean Score	0.0426	0.0429	0.0416	0.0417	**0.0416**	0.0431
	Deviation	*2.50%*	*3.26%*	-	*0.17%*	*-0.03%*	*3.67%*
Region 12h	Mean Score	0.0538	0.0540	0.0528	0.0532	**0.0528**	0.0542
	Deviation	*1.83%*	*2.21%*	-	*0.68%*	*-0.05%*	*2.64%*
France 6h	Mean Score	0.0266	0.0267	0.0256	0.0258	**0.0256**	0.0272
	Deviation	*3.90%*	*4.10%*	-	*0.44%*	*-0.02%*	*5.92%*
France 12h	Mean Score	0.0335	0.0335	0.0327	0.0329	**0.0326**	0.0339
	Deviation	*2.63%*	*2.66%*	-	*0.83%*	*-0.08%*	*3.79%*

The RED method demonstrates very similar performance to PRW, sometimes even delivering slightly lower scores in specific cases, such as the substation-level ES for a 6-hour forecast horizon, where RED scores 0.0613 compared to PRW's identical score, though RED's deviation is marginally better at -0.04% versus PRW's baseline. This alignment suggests that RED may be a viable alternative under certain conditions, especially when computational or data handling constraints favor its application. The distinction between PRW, RED, and the other methods becomes particularly evident when examining the deviations. Methods like CLMN and CLMD

Table 7: Benchmark of mean *Variogram-based Scores* and deviations for different spatial granularit and forecast horizons.

Scope		**CLMN**	**CLMD**	**PRW**	**REG**	**RED**	**ZER**
Substation 6h	Mean Score	0.2729	0.2746	0.2699	0.2743	**0.2697**	0.2746
	Deviation	*1.10%*	*1.74%*	-	*1.61%*	*-0.07%*	*1.74%*
Substation 12h	Mean Score	1.1209	1.1287	1.0538	1.0913	**1.0536**	1.1288
	Deviation	*6.37%*	*7.10%*	-	*3.55%*	*-0.02%*	*7.12%*
Region 6h	Mean Score	0.1453	0.1479	**0.1381**	0.1389	0.1382	0.1561
	Deviation	*5.15%*	*7.08%*	-	*0.56%*	*0.02%*	*12.96%*
Region 12h	Mean Score	0.6389	0.6449	**0.5741**	0.6074	0.5741	0.6475
	Deviation	*11.30%*	*12.33%*	-	*5.80%*	*0.00%*	*12.80%*
France 6h	Mean Score	0.0818	0.0836	**0.0737**	0.0745	0.0738	0.0868
	Deviation	*10.97%*	*13.38%*	-	*1.07%*	*0.05%*	*17.74%*
France 12h	Mean Score	0.3779	0.3817	**0.3256**	0.3541	0.3257	0.3888
	Deviation	*16.08%*	*17.24%*	-	*8.77%*	*0.03%*	*19.43%*

consistently show higher scores, with deviations of up to 17.24% for VS at the country level with a 12-hour horizon, as compared to PRW. This difference underscores the robust handling of spatial and temporal data by PRW and RED, largely attributable to their incorporation of t-copula (TC). When comparing the VS with the ES, it is evident that the relative differences are more pronounced for the VS. For example, at the regional level with a 12-hour forecast horizon, the deviation for CLMD reaches 12.33% for VS, compared to just 2.21% for ES, highlighting the higher sensitivity of VS to variations in forecast quality. This trend emphasizes the importance of selecting the right evaluation metric based on the specific requirements of the forecast application. Another general trend observable across both scores is the reduction of scores with coarser spatial granularity. As an example, the VS mean score for PRW and 12 horizons decreases from 1.0538 at the substation level to 0.3256 at the country level. This can be explained by the averaging across larger geographic areas, where localized short-term phenomena, such as variations in cloud coverage, have a diminished impact. As a result, forecasts at coarser spatial granularity benefit from a smoothing effect, leading to better overall performance and lower scores. In summary, PRW and RED consistently rank at the top across all levels of spatial granularity and forecast horizons, with PRW generally offering a slightly more robust performance for VS. However, RED's comparable results and slightly better deviations in some instances make it a strong alternative, depending on specific operational needs.

6. Conclusion

In this paper, we developed a multivariate probabilistic forecasting model for solar PV generation, addressing the significant challenge of imbalanced data resulting from day and night-time periods. By leveraging forecast updates and modeling the complex temporal interdependencies in solar PV generation, our approach offers a more accurate and consistent method for predicting renewable energy output.

The application of our methodology to real-world data from a region in France demonstrated its practical effectiveness. The case study results highlighted the differences in performance of various methods in generating reliable forecast trajectories across different spatial granularities and forecast horizons. Here, the introduced pairwise new methodology and reduced methods consistently outperformed other approaches—specifically when combined with copula models—in terms of both evaluation metrics, i.e., the Energy and Variogram-based Score. The results on a regional level indicate that the proposed methods accurately capture capture key probabilistic properties of forecast updates, including the distributional properties and temporal interdependence structures of solar PV generation. Furthermore, the extended case study showed that with a coarser spatial granularity, the forecast accuracy improves, highlighting the adaptability of the proposed methodology to varying spatial scales and its potential for effective application in aggregated forecasting applications. The pairwise method, in particular, proved to be a reliable option for handling real-world data, providing consistent performance. Our findings indicate that the reduced method can deliver competitive performance. Specifically, for the Energy Score — which has weaknesses compared to the Variogram-based Score in reliably penalizing incorrect dependency structures — the reduced method occasionally outperforms the pairwise approach slightly. This suggests that the computationally simpler reduced method may be a viable alternative under certain conditions. However, the limitations of the reduced method must be carefully considered. For higher forecast horizons, the dataset becomes overly constrained, leading to data elimination, which can compromise prediction quality and render the method infeasible. In such scenarios, the pairwise approach provides a more robust and reliable framework, ensuring a sufficient data basis. Consequently, for applications with stringent data requirements, the pairwise approach remains the preferred method.

Our findings underscore the critical importance of addressing data imbalances in solar forecasting and demonstrate the advantages of using multivariate models with repeated updates. These models provide decision-makers with more reliable and actionable insights, enhancing the integration of renewable energy into power systems and markets. Our work contributes to developing more sophisticated forecasting tools for successfully integrating renewable energy sources into the grid, supporting the

global transition to a sustainable energy future.

Declaration of generative AI and AI-assisted technologies in the writing process

While preparing this work, the authors used ChatGPT and Grammarly to improve the language and readability of selected parts of the manuscript. After using these tools, the authors carefully reviewed and edited the content as needed, taking full responsibility for the content of the publication.

Funding source

This work results from a collaborative research effort between the *House of Energy Markets and Finance, University of Duisburg-Essen* and *RTE Réseau de Transport d'Électricité S.A.*. The collaboration was supported by funding from RTE, emphasizing a joint commitment to advancing renewable energy forecasting methodologies in a cooperative academic-industrial partnership.

References

Ahmadi, M.H., Ghazvini, M., Sadeghzadeh, M., Alhuyi Nazari, M., Kumar, R., Naeimi, A., Ming, T., 2018. Solar power technology for electricity generation: A critical review. Energy Science & Engineering 6, 340–361. doi:10.1002/ese3.238.

Altendorfer, K., Felberbauer, T., 2023. Forecast and production order accuracy for stochastic forecast updates with demand shifting and forecast bias correction. Simulation Modelling Practice and Theory 125, 102740. doi:10.1016/j.simpat.2023.102740.

Baltagi, B.H., Liu, L., 2020. Forecasting with unbalanced panel data. Journal of Forecasting 39, 709–724. doi:10.1002/for.2646.

Baltagi, B.H., Song, S.H., 2006. Unbalanced panel data: A survey. Statistical Papers 47, 493–523. doi:10.1007/s00362-006-0304-0.

Bjerregård, M.B., Møller, J.K., Madsen, H., 2021. An introduction to multivariate probabilistic forecast evaluation. Energy and AI 4, 100058. doi:10.1016/j.egyai.2021.100058.

Boehnke, F., Kolkmann, S., Leisen, R., Weber, C., 2024. How wind forecast updates are balanced in coupled european intraday markets. IET Renewable Power Generation doi:10.1049/rpg2.13065.

Bouyé, E., Durrleman, V., Nikeghbali, A., Riboulet, G., Roncalli, T., 2000. Copulas for finance-a reading guide and some applications. Available at SSRN 1032533 doi:10.2139/ssrn.1032533.

Fernández, A., Spencer, T., Bouckaert, S., McGlade, C., Remme, U., Wanner, B., D'Ambrosio, D., 2023. Net zero by 2050: A roadmap for the global energy sector doi:10.1787/c8328405-en.

Garnier, E., Madlener, R., 2015. Balancing forecast errors in continuous-trade intraday markets. Energy Systems 6, 361–388. URL: http://dx.doi.org/10.1007/s12667-015-0143-y, doi:10.1007/s12667-015-0143-y.

Glas, S., Kiesel, R., Kolkmann, S., Kremer, M., Graf von Luckner, N., Ostmeier, L., Urban, K., Weber, C., 2020. Intraday renewable electricity trading: advanced modeling and numerical optimal control. Journal of Mathematics in Industry 10. doi:10.1186/s13362-020-0071-x.

Gneiting, T., Balabdaoui, F., Raftery, A.E., 2007. Probabilistic Forecasts, Calibration and Sharpness. Journal of the Royal Statistical Society Series B: Statistical Methodology 69, 243–268. doi:10.1111/j.1467-9868.2007.00587.x.

Gneiting, T., Katzfuss, M., 2014. Probabilistic forecasting. Annual Review of Statistics and Its Application 1, 125–151. URL: http://dx.doi.org/10.1146/annurev-statistics-062713-085831, doi:10.1146/annurev-statistics-062713-085831.

Golestaneh, F., Gooi, H.B., Pinson, P., 2016. Generation and evaluation of space–time trajectories of photovoltaic power. Applied Energy 176, 80–91. doi:10.1016/j.apenergy.2016.05.025.

Gürtler, M., Paulsen, T., 2018. The effect of wind and solar power forecasts on day-ahead and intraday electricity prices in germany. Energy Economics 75, 150–162. doi:10.1016/j.eneco.2018.07.006.

Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., Bing, G., 2017. Learning from class-imbalanced data: Review of methods and applications. Expert systems with applications 73, 220–239. doi:10.1016/j.eswa.2016.12.035.

He, H., Garcia, E.A., 2009. Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering 21, 1263–1284. doi:10.1109/TKDE.2008.239.

Hong, T., Pinson, P., Wang, Y., Weron, R., Yang, D., Zareipour, H., 2020. Energy forecasting: A review and outlook. IEEE Open Access Journal of Power and Energy 7, 376–388. doi:10.1109/OAJPE.2020.3029979.

Koch, C., 2021. Intraday imbalance optimization: incentives and impact of strategic intraday bidding behavior. Energy Systems 13, 409–435. doi:10.1007/s12667-021-00445-9.

Kolkmann, S., Ostmeier, L., Weber, C., 2024. Modeling multivariate intraday forecast update processes for wind power. Energy Economics 139, 107890. doi:https://doi.org/10.1016/j.eneco.2024.107890.

Krzysztofowicz, R., 1987. Markovian forecast processes. Journal of the American Statistical Association 82, 31–37. doi:10.1080/01621459.1987.10478387.

Li, B., Zhang, J., 2020. A review on the integration of probabilistic solar forecasting in power systems. Solar Energy 210, 68–86. doi:10.1016/j.solener.2020.07.066.

Little, R.J.A., Rubin, D.B., 2019. Statistical analysis with missing data. volume 793. John Wiley & Sons. doi:10.1002/9781119482260.

McNeil, A.J., Frey, R., Embrechts, P., 2015. Quantitative risk management: concepts, techniques and tools-revised edition. Princeton university press.

Peña, M., Toth, Z., 2014. Estimation of analysis and forecast error variances. Tellus A: Dynamic Meteorology and Oceanography 66, 21767. doi:10.3402/tellusa.v66.21767.

Pinson, P., Girard, R., 2012. Evaluating the quality of scenarios of short-term wind power generation. Applied Energy 96, 12–20. doi:10.1016/j.apenergy.2011.11.004.

Pinson, P., Madsen, H., Nielsen, H.A., Papaefthymiou, G., Klöckl, B., 2009. From probabilistic forecasts to statistical scenarios of short-term wind power production. Wind Energy 12, 51–62. doi:10.1002/we.284.

Riddervold, H.O., Aasgård, E.K., Haukaas, L., Korpås, M., 2021. Internal hydro- and wind portfolio optimisation in real-time market operations. Renewable Energy 173, 675–687. doi:10.1016/j.renene.2021.04.001.

Samuelson, P.A., 1965. Proof that properly anticipated prices fluctuate randomly. Industrial Management Review 6, 41.

Scheuerer, M., Hamill, T.M., 2015. Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. Monthly Weather Review 143, 1321–1334. doi:10.1175/MWR-D-14-00269.1.

Schinke-Nendza, A., von Loeper, F., Osinski, P., Schaumann, P., Schmidt, V., Weber, C., 2021. Probabilistic forecasting of photovoltaic power supply—a hybrid approach using d-vine copulas to model spatial dependencies. Applied Energy 304, 117599. doi:10.1016/j.apenergy.2021.117599.

Schinke-Nendza, A., Weber, C., 2024. Event-based evaluation of congestions in grids with high shares of renewable generation using multivariate probabilistic forecasts. To be published .

Sinsel, S.R., Riemke, R.L., Hoffmann, V.H., 2020. Challenges and solution technologies for the integration of variable renewable energy sources—a review. Renewable Energy 145, 2271–2285. doi:10.1016/j.renene.2019.06.147.

Sklar, M., 1959. Fonctions de repartition an dimensions et leurs marges. Publ. inst. statist. univ. Paris 8, 229–231.

Sweeney, C., Bessa, R.J., Browell, J., Pinson, P., 2020. The future of forecasting for renewable energy. Wiley Interdisciplinary Reviews: Energy and Environment 9, e365. doi:10.1002/wene.365.

Tawn, R., Browell, J., Dinwoodie, I., 2020. Missing data in wind farm time series: Properties and effect on forecasts. Electric Power Systems Research 189, 106640. doi:10.1016/j.epsr.2020.106640.

Visser, L., AlSkaif, T., Khurram, A., Kleissl, J., van Sark, W., 2024. Probabilistic solar power forecasting: An economic and technical evaluation of an optimal market bidding strategy. Applied Energy 370, 123573. doi:10.1016/j.apenergy.2024.123573.

Yang, D., Wang, W., Gueymard, C.A., Hong, T., Kleissl, J., Huang, J., Perez, M.J., Perez, R., Bright, J.M., Xia, X., et al., 2022. A review of solar forecasting, its dependence on atmospheric sciences and implications for grid integration: Towards carbon neutrality. Renewable and Sustainable Energy Reviews 161, 112348. doi:10.1016/j.rser.2022.112348.

Zalzar, S., Bompard, E., Purvins, A., Masera, M., 2020. The impacts of an integrated european adjustment market for electricity under high share of renewables. Energy Policy 136, 111055. doi:10.1016/j.enpol.2019.111055.

Correspondence

M.Sc. Yannik Pflugfelder[1]
(Corresponding author)
Tel. +49 201 183-6458
E-Mail Yannik.Pflugfelder@uni-due.de

M.Sc. Aiko Schinke-Nendza[1]
E-Mail Aiko.Schinke-Nendza@uni-due.de

Jonathan Dumas[2]
E-Mail Jonthan.Dumas@rte-france.com

Prof. Dr. Christoph Weber[1]
Chair for Management Science and Energy Economics
Tel. +49 201 183-2966
E-Mail Christoph.Weber@uni-due.de

[1]House of Energy Markets and Finance
University of Duisburg-Essen, Germany
Universitätsstr. 12, 45117 Essen
www.hemf.net

[2]RTE – Réseau de Transport d'Electricité, S.A.
7C place du Dôme, Paris La Défense, 92073, France