

# Seamless intra-day and day-ahead multivariate probabilistic forecasts at high temporal resolution

Dennis van Der Meer, Simon Camal, Georges Kariniotakis

▶ **To cite this version:**

Dennis van Der Meer, Simon Camal, Georges Kariniotakis. Seamless intra-day and day-ahead multivariate probabilistic forecasts at high temporal resolution. 17th International Conference on Probabilistic Methods Applied to Power Systems, PMAPS 2022, Jun 2022, Manchester - Online, United Kingdom. 10.1109/PMAPS53380.2022.9810606 . hal-03638925

**HAL Id: hal-03638925**

**https://minesparis-psl.hal.science/hal-03638925**

Submitted on 12 Apr 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Seamless intra-day and day-ahead multivariate probabilistic forecasts at high temporal resolution

Dennis van der Meer, Simon Camal, Georges Kariniotakis
*MINES Paris – PSL University – Centre for Processes, Renewable energies and Energy systems (PERSEE)*
Sophia Antipolis, France
{dennis.van_der_meer, simon.camal, georges.kariniotakis}@minesparis.psl.eu

***Abstract*—High temporal resolution intra-day and day-ahead photovoltaic (PV) power forecasts are important to maximize the value of PV systems because they enable stakeholders to participate in both the energy and ancillary services markets. Whereas most day-ahead electricity markets feature an hourly temporal resolution, intra-day markets may require forecasts at 5-minute resolution. In addition, battery integration can improve power system management in isolated grids with high PV power penetration, but battery control requires high temporal resolution forecasts. We propose an efficient method based on pattern matching to generate multivariate probabilistic forecasts, approximated by trajectories, at high temporal resolution and without the need to separately forecast the marginals and estimate the covariance matrix. We compare the proposed method against quantile regression forests in combination with copula theory and show that our method reduces the forecast time by approximately 98% and simplifies the modeling chain while incurring a minor performance penalty.**

***Index Terms*—Scenarios, operational forecasting, renewable energy sources, photovoltaic generation, virtual power plant**

## I. Introduction

Murphy divides the "nature of goodness in weather forecasting" into three aspects: (i) consistency, (ii) quality and (iii) value [1]. In itself, a forecast does not have value until it is used in a decision-making process [1]. In order to maximize the monetary value of a renewable energy source (RES) forecast, it is important that the system owner or system aggregator be able to participate in both the energy and ancillary service markets on the intra-day and day-ahead levels. Electricity markets feature various requirements for parties to place bids; of particular interest to this study are the requirements on temporal resolution and forecast horizon. For instance, the California Independent System Operator (CAISO) requires, every 5 minutes, forecasts with a lead time of 7.5 minutes, a temporal resolution of 5 minutes and a forecast horizon of 65 minutes for their real-time economic dispatch [2]. For its short-term unit commitment, CAISO requires forecasts with a lead time of 75 minutes, a temporal resolution of 15 minutes and a forecast horizon of 5 hours, updated every hour [2]. For its day-ahead market, CAISO requires hourly forecasts with a lead time of 14 hours and a forecast horizon of 24 hours [2]. Typically, research implementing RES forecasts in power system applications focuses on one of the aforementioned markets. For instance, [3] estimate the optimal automatic frequency restoration reserve (aFRR) offer that RESs can provide on the day-ahead market. Similarly, [4] focus on offers of aFRR and energy on the day-ahead market. In contrast, [5] concentrate on CAISO's short-term unit commitment market described above.

The value of a forecast also relates to the continuity of balance between supply and demand. Especially in isolated grids featuring high RES penetration, the lack of synchronous machines introduces challenges in power system management [6]. In such a scenario, high temporal resolution RES forecasts are important to inform the grid operator on future variability, as well as to enable battery control, which in turn allows for simultaneous provision of numerous services (e.g., peak shaving or frequency containment reserve) [7]. In the context of battery control, [8] generate 15-min resolution forecasts up to 24 h ahead that are updated every 15 mins and subsequently fed into a model predictive control algorithm.

Besides value, Murphy defines consistency as follows: "a forecast should always correspond to a forecaster's best judgment" [1]. Since the forecaster faces several uncertainties , his or her judgment is inherently uncertain and the forecast should reflect that [1]. Moreover, sequential optimization problems, e.g., model predictive control, require trajectory forecasts that account for dependencies in time and space[1] [9]. The combination of a multitude of trajectory forecasts represents an approximation of the multivariate probabilistic forecast and can readily be employed in Monte-Carlo-type decision-making processes [9].

The de facto method to issue trajectory forecasts is to forecast the marginal cumulative distribution functions (CDFs) over a forecast horizon and compute the probability integral transform (PIT) variables to estimate a covariance matrix that describes the dependence structure, e.g., [3], [4], [8], [9]. However, as both the temporal resolution and the forecast horizon increase, the number of required marginal CDFs increases rapidly, as does the computational burden. For instance, a 48 h ahead forecast at a 5 min resolution requires 576 marginal CDFs, after which the covariance matrix needs to be estimated. To alleviate the computational burden and to simplify the model chain, we propose to generate trajectory

Parts of this research were carried in the frame of the Smart4RES project (No. 864337), supported by the Horizon 2020 Framework Program.

[1]Note that the spatial dependency is disregarded in the present study because it focuses on aggregated power production.

forecasts using pattern matching in order to downscale hourly numerical weather prediction (NWP) ensemble forecasts to power forecasts at a 5 min resolution, similar to [5]. However, unlike [5]—who use Mueen's Algorithm for Similarity Search [10]—we employ k-dimensional tree (kd-tree, [11]) to allow for multivariate input features as recommended in [12]. This approach constitutes a seamless method for the generation of RES trajectories at intraday and day-ahead horizons, in contrast with the seamless RES forecasting method of [13] that generates probabilistic density forecasts without temporal correlation.

The contributions of this paper can be summed up as follows:

- We use a pattern matching algorithm to downscale NWP ensemble forecasts to high temporal resolution multivariate probabilistic PV power forecasts, which has not been attempted before as far as we are aware.
- The proposed method significantly reduces the computational burden and greatly simplifies the forecast model chain.
- The use of kd-tree means that the proposed method can be applied to any RES power forecast model chain using any number of input features.

The remainder of this paper is organized as follows: Section II provides an overview of the state-of-the-art and describes the data, methods and scoring rules. Section III presents and discusses the results while Section IV concludes our study.

## II. Background and Methodology

### A. Problem formulation

In this study, we concern ourselves with temporal trajectory forecasts only and what follows is a description of how such trajectories are often generated. We follow [14] to estimate the covariance matrix that describes the dependencies across time. This method requires that the marginal predictive CDFs be calibrated, which is to say that the observations $Y$ are statistically similar to samples from the CDFs $F$, i.e., $Y \sim F$.

Mathematically, a series of forecasts with horizon $k$ is said to be calibrated when $P_k$ is uniformly distributed, i.e., $P_k \sim \mathcal{U}[0,1]$, where $p_{k,i}$ can be computed using the probability integral transform (PIT)

$$p_{k,i} = F_{i+k|i}(y_{i+k|i}), \quad \forall i \tag{1}$$

where $i$ is the index of the training set that runs from $1,\dots,N$. Likewise, $t$ is the index of the testing set that runs from $1,\dots,T$.

Subsequently, it is possible to transform $P_k$ to a standard normal random variable $Z_k \sim \mathcal{N}(0,1)$ using the inverse Gaussian CDF

$$z_{k,i} = \Phi^{-1}(p_{k,i}). \quad \forall i \tag{2}$$

In case of $k = 1,\dots,K$ forecast horizons, the random vector $\boldsymbol{Z} = (Z_1,\dots,Z_K)^\top$ follows a multivariate normal distribution, i.e., $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu}_0$ is a vector of zeros and $\boldsymbol{\Sigma}$ a covariance matrix that describes the dependencies between forecast horizons with ones on its diagonal [14]. An unbiased estimate of $\boldsymbol{\Sigma}$ is [14]:

$$\boldsymbol{\Sigma} = \frac{1}{N-1}\sum_{i=1}^{N} \boldsymbol{Z}_i \boldsymbol{Z}_i^\top. \tag{3}$$

The trajectory forecasts can then be generated by sampling $S$ $K$-dimensional vectors using a multivariate normal random number generator with $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}$. Subsequently, the sampled vectors can be transformed back to the uniform variable $P_k$ using the standard normal CDF $\Phi$:

$$p_{s,k} = \Phi(z_{s,k}). \quad \forall s,k \tag{4}$$

Once the marginal inverse predictive CDFs $F^{-1}_{t+k|t}$ have been generated on the testing set, the autocorrelated standard uniform samples $p_{s,k}$ can be used to generate trajectory forecasts $f_{s,t+k|t}$:

$$f_{s,t+k|t} = F^{-1}_{t+k|t}(p_{s,k}). \quad \forall s,k \tag{5}$$

The above-described method becomes computationally very demanding when the temporal resolution and the forecast horizon increase. For instance, a forecast horizon of 48 h at a temporal resolution of 5 min requires 576 marginal predictive CDFs, whereas a forecast horizon of 24 h at a temporal resolution of 1 min requires 1,440 marginal predictive CDFs. Clearly, an alternative method that performs similarly but significantly faster is preferred, which is what we describe in the next section.

### B. Pattern matching

The pattern matching model (PMM) proposed here is conceptually straightforward and is based on the analog ensemble [15]. However, rather than searching for analogs for a single forecast horizon to approximate $F_{t+k|t}$, we aim to search for $S$ analog trajectories to approximate the multivariate predictive CDF $\boldsymbol{F}_t$. In addition, we intend to leverage the computational efficiency of kd-tree, which can improve the computational performance by three orders of magnitude compared to brute-force search [16]. To that end, we modify the original similarity metric in [15] such that $\boldsymbol{\mathcal{X}}_t$ contains the query NWP forecast issued at testing time $t$ organized as one vector. Similarly, $\boldsymbol{\mathcal{A}}_i$ is a single vector containing the analog NWP forecast issued at training time $i$. The similarity metric accepts any number of variables and is defined as [12]

$$d(\boldsymbol{\mathcal{X}}_t, \boldsymbol{\mathcal{A}}_i) = \sqrt{\sum_{j=1}^{J} w_j \left(x_t^{(j)} - x_i^{(j)}\right)^2}, \tag{6}$$

where for simplicity $J$ is the total dimension and $w_j = 1 \; \forall \; j$.

The result is the approximate predictive distribution $\boldsymbol{F}_t$ comprising detrended 5-min PV power measurements observed from time $i+1$ up to time $i+K$, such that $\boldsymbol{F}_t \in \mathbb{R}^{S\times K}$ and the corresponding observations are $\boldsymbol{y}_t \in \mathbb{R}^K$. In addition, we compute the clear-sky global horizontal irradiance (GHI) vector $\boldsymbol{g}_t \in \mathbb{R}^K$ with which $\boldsymbol{F}_t$ is multiplied to attain PV power measurements, see Section II-D. In order to value each feature similarly, $\boldsymbol{\mathcal{A}}_i$ is centered and scaled and these factors are used to center and scale $\boldsymbol{\mathcal{X}}_i$.

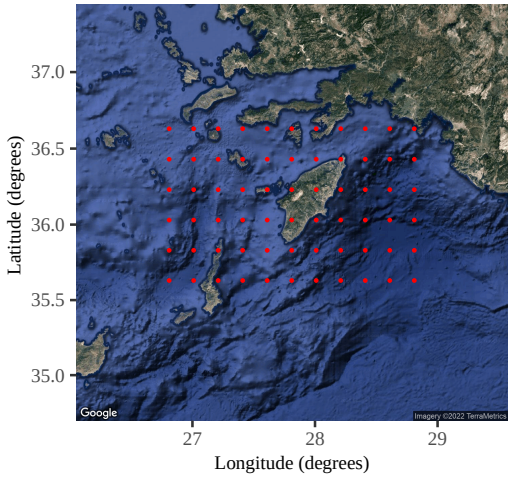



Fig. 1. Map of the area of interest where the red points represent the grid points at which we have collected NWP forecasts.

### C. Benchmarks

*1) Quantile regression forests (QRF):* In random forests, the prediction is the weighted average of the observed response variables. In contrast, in QRF, the output is the weighted distribution of the observed response variables [17]. During training, each tree is grown on a random sample of the training data, thus reducing the correlation between the trees. To further decrease this correlation, a random subset of the features is selected at each candidate split [17]. Note that QRF is combined with the Gaussian copula described above to generate state-of-the-art trajectory forecasts.

*2) Multivariate probabilistic ensemble:* The second benchmark is the naive multivariate probabilistic ensemble (MuPEn) of which the marginal predictive CDFs are identical to those of the complete-history persistence ensemble [18]. It can be constructed by gathering all $N$ $K$-length vectors from the historical observations that start at the same time ("HH:MM") as the current forecast issue time $t$. The result is an $N \times K$ matrix from which we randomly sample—without replacement—$S$ trajectories such that the result is a multivariate predictive distribution $\boldsymbol{F}_t \in \mathbb{R}^{S \times K}$.

### D. Data

In this study we use two data sources, the first of which is the aggregated PV power measurements at a 1-min resolution from Rhodes, Greece, which we average to a 5-min resolution. The only available metadata is the installed capacity, which is 18,164 kW. The PV power measurements are subsequently detrended using clear-sky GHI generated by the McClear model [19] for a tilt of 25° and azimuth of 180° due south. These values constitute the optimal orientation on Rhodes and we assume that a large share of the PV systems are installed with this orientation [20]. However, we notice that the tails of the PV power measurements at sunrise and sunset do not coincide with the clear-sky GHI, which is likely caused by a number of systems orientated towards the east and west, respectively. The detrended PV power measurements are therefore set to 0 when the zenith angle is larger than 90° or when the clear-sky GHI is lower than 1 W/m$^2$ as these values would otherwise be significantly higher than can be reasonably assumed. The PV power is normalized with the installed capacity so that the errors can be presented as dimensionless numbers or as percentages.

The second data source is the hourly NWP forecasts issued by the European Centre for Medium-Range Weather Forecasts (ECMWF), issued daily at 00:00 UTC. Figure 1 presents the grid points in red at which we have collected NWP forecasts, i.e., 66 grid points in total to capture the large scale weather pattern. The variables used in this study are surface solar downward radiation (SSRD) converted to GHI and then clear-sky index, total cloud cover (TCC), 10 meter u- and v-wind components (U10 and V10) and 2-meter temperature (T2M). In order to have a sense of the variability present in the forecasts, we use the ensemble prediction system featuring 50 ensemble members [21]. Given the 48 h forecast horizon of interest, query $\boldsymbol{\mathcal{X}}_t$ has a length of $48 \times 66 \times 50 \times 5 = 792,000$. To reduce the vector length, we summarize the grid points and ensemble members in two ways: (i) the mean and standard deviation of the ensemble members at each grid point, resulting in 31,680 features (referred to as case study MS); and (ii) the $0.01, 0.02, \ldots, 0.99$ quantiles over all ensemble members and grid points, resulting in 23,760 features (referred to as case study QS).

Given the relatively short length of the data set (1 year), we test on a monthly basis while the remaining months serve as training data. For instance, when forecasting January, we use February up to and including December as training data. In addition, it is important to note that forecasts are issued on a rolling basis with hourly intervals. This means there are approximately 700 forecasts per month, although this can be fewer in the case where the PMM is unable to find sufficient quality analogs. Finally, we train 48 QRF models instead of 576 and use each model to forecast the 12 horizons within the hour to ease the computational burden.

### E. Forecast verification

Since consistency is one of the three aspects of a "good" forecast, it is encouraged in expectation by proper scoring rules [1]. Such scoring rules can be used to evaluate the quality of forecasts, which is Murphy's second aspect [1]. In this study, we use the continuous ranked probability score (CRPS) to evaluate the sharpness and calibration of the marginal predictive CDFs $F_{t+k|t}$, which is defined as [22]

$$\mathrm{CRPS}(F_{t+k|t}, y_{t+k|t}) = \int_0^\alpha \left(F_{t+k|t}(x) - \mathbb{1}\{y_{t+k|t} \leq x\}\right)^2 \mathrm{d}x, \tag{7}$$

where $\mathbb{1}$ is the indicator function and $\alpha$ is the maximum observed power. Subsequently, we average the CRPS over all forecast valid times where the zenith angle is smaller than 85°.

To evaluate the multivariate predictive CDFs $\boldsymbol{F}_t$, we use the energy score (ES) and variogram score (VS). ES is a generalization of CRPS and can be defined as [22]

$$\mathrm{ES}\left(\boldsymbol{F}_t, \boldsymbol{y}_t\right) = \mathbb{E}_{\boldsymbol{F}}\|\boldsymbol{X}_t - \boldsymbol{y}_t\| - \frac{1}{2}\mathbb{E}_{\boldsymbol{F}}\|\boldsymbol{X}_t - \boldsymbol{X}_t'\|, \quad (8)$$

where $\boldsymbol{X}_t$ and $\boldsymbol{X}_t'$ are independent random vectors sampled from $\boldsymbol{F}_t$ and $\|\cdot\|$ represents the Euclidean norm. A disadvantage of ES is that it is unable to discriminate between poorly or correctly specified dependence structures [23]. Therefore, we include VS to compare the quality of the dependence structure specification, which is defined as [24]:

$$\mathrm{VS}_p\left(\boldsymbol{F}_t, \boldsymbol{y}_t\right) = \sum_{i,j=1}^{K} w_{ij}\left(|y_i - y_j|^p - \mathbb{E}_{\boldsymbol{F}}|\boldsymbol{x}_i - \boldsymbol{x}_j|^p\right)^2, \quad (9)$$

where $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are components $i$ and $j$ of random vector $\boldsymbol{X}_t$, which is distributed according to $\boldsymbol{F}_t$. In this study, $\boldsymbol{F}_t$ is approximated by $S$ $K$-dimensional trajectories $\begin{pmatrix}\boldsymbol{x}^{(1)} & \boldsymbol{x}^{(2)} & \cdots & \boldsymbol{x}^{(S)}\end{pmatrix}^\top$ and $\mathbb{E}_{\boldsymbol{F}}|\boldsymbol{x}_i - \boldsymbol{x}_j|^p$ can be approximated by [24]:

$$\mathbb{E}_{\boldsymbol{F}}|\boldsymbol{x}_i - \boldsymbol{x}_j|^p \approx \frac{1}{S}\sum_{s=1}^{S}|x_i^{(s)} - x_j^{(s)}|^p, \quad i,j = 1,\ldots,K, \quad (10)$$

where $x_i^{(s)}$ and $x_i^{(s)}$ are elements $i$ and $j$ of the $s^{\text{th}}$ trajectory forecast. Although weights $w_{ij}$ can be used to add or reduce importance between certain forecast horizons, we use identity weights and set $p = 0.5$ as recommended in [24]. Finally, CRPS, ES and VS are averaged over the testing samples after which skill scores relative to MuPEn are computed.

## III. Results and Discussion

### A. Experimental setup

Preliminary experiments reveal that the zenith angle along the entire forecast horizon provides valuable information to accurately determine sunrise and sunset timing. Similarly, U10, V10 and T2M do not contribute positively to the forecast accuracy and are therefore omitted for the PMM. In addition, we find that a larger number of trajectories $S$ positively affects VS, likely because it is more challenging to approximate the prevailing dependencies with fewer trajectories. We therefore set $S = 50$ in our experiments. Furthermore, recall that the forecast horizon is 48 h and that the temporal resolution is 5 min, which means that $K = 576$. Finally, recall that the two ways of representing the spatio-temporal information from the NWP forecasts are abbreviated by MS and QS, resulting in four case studies (e.g., PMM+MS or QRF+QS).

### B. Univariate forecasts

Figure 2 presents histograms of the PIT variables computed over the testing set using (1). In addition, Fig. 3 presents the CRPS in percent of the nominal capacity as a function of the forecast horizon. The mean and standard deviations are computed across the 12 testing months, as described above. Several interesting observations can be made from these

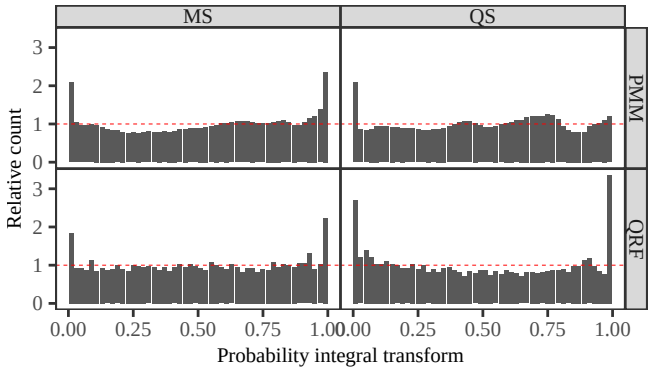



Fig. 2. Histograms of the marginal PIT variables combined over all forecast horizons and testing instances where the zenith angle is smaller than 85°.

figures. First, Fig. 2 shows that the PMM+MS and QRF+QS forecasts tend to be underdispersed. In contrast, forecasts generated by PMM+QS and QRF+MS are calibrated better in the main distribution although the lowest quantile deviates in case of the former and the most extreme quantiles deviate in case of the latter.

Second, Fig. 3 shows that PMM+MS does not perform well and that is mainly caused by poor forecasts in December (not shown here). The poor performance during December is likely due to the inability of the PMM to find high quality analogs, although PMM+QS did not suffer from this problem. The poor performance in December could therefore be caused by the way the NWP ensemble information is summarized.

Third, CRPS of PMM+QS is constant over the forecast horizon, whereas CRPS of QRF+MS and QRF+QS increases sharply over the first hours and then stabilizes. In control problems, stability in the forecast error variance over the entire forecast horizon is preferable because it reduces the bullwhip effect [5]. For instance, underdispersive forecasts could result in an optimistic strategy by the control algorithm that could lead to the overutilization of available storage to correct for the forecast errors.

Finally, QRF+MS performs best in terms of CRPS. However, it is likely that more data would favor the PMM as it requires a long history to find quality analogs. In contrast, QRF has been proven to be robust against relatively small data sets. In terms of computation time, it is important to note that the PMM is approximately 98% faster than QRF.

### C. Multivariate forecasts

Figure 4 presents ES and VS as monthly averages in addition to the total averaged scores. The figure clearly shows that PMM+MS performs poorly in December. Besides December, the figure shows that QRF+MS performs substantially better in February than the other models. Overall, QRF+MS outperforms the others but given the limited number of forecast-verification pairs—recall that there are about 700 forecast-verification pairs per month—it is worthwhile to test the significance of the difference between the forecasts.

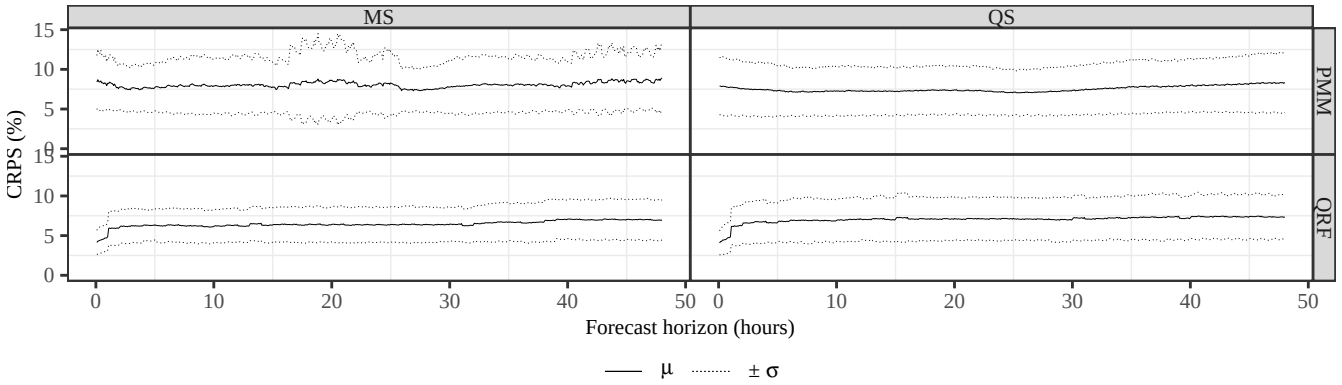



Fig. 3. CRPS in percent of nominal capacity as a function of forecast horizon. The mean and standard deviation are computed across the 12 testing months.

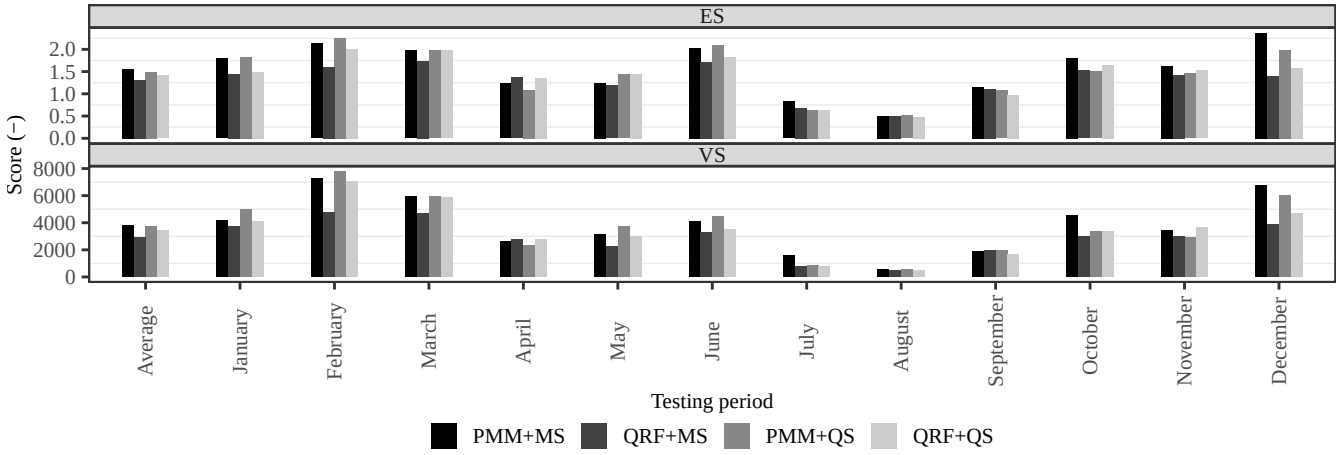



Fig. 4. ES and VS averaged over the entire testing set and on a monthly basis. Note that the scores are dimensionless.

In hypothesis testing it is common to set the null hypothesis ($H_0$) such that there is no difference between the forecasts whereas the alternative hypothesis ($H_a$) is that there is a difference. The Diebold-Mariano (DM) test is commonly used to test $H_0$. The test requires computation of the forecast error loss differential $d_t = \ell(\boldsymbol{F}_{1,t}, \boldsymbol{y}_t) - \ell(\boldsymbol{F}_{2,t}, \boldsymbol{y}_t)$. However, the DM test is designed specifically for a forecast horizon and consequently, we are uncertain whether it applies here. Furthermore, the paired $t$-test is not recommended when temporal dependence and contemporaneous correlation are present [25]. Instead, [25] recommend the Hering-Genton (HG) test and circular block bootstrapping. We choose to use block bootstrapping because the HG test, like the DM test, assumes that the series $\boldsymbol{d}$ is covariance stationary, whereas circular block bootstrapping is relevant for small testing sets [25]. The bootstrap is repeated 10,000 times with a block length of $\sqrt{T}$ and the results are presented in Table I as the mean plus-minus the standard deviation and the confidence interval (CI, significance level $\alpha = 5\%$) in parentheses. As the table shows, 0 always lies within the CI and therefore we cannot reject $H_0$.

Finally, we report the skill scores relative to MuPEn in Table II, computed as $1 - L_{\text{model}}/L_{\text{MuPEn}}$ and where $L$ is the loss averaged over the testing set. Evidently, QRF+MS performs best on all scores. However, given that the marginal predictive CDFs of PMM+QS are calibrated slightly better, that no learning step but only proper data organization is required, and that computation time is reduced by approximately 98%, we argue that the PMM is at the very least a highly efficient and interpretable asset in a forecaster's toolbox.

## IV. Conclusions

We proposed a forecast model based on a pattern matching model (PMM) and used it to generate multivariate probabilistic forecasts of aggregated photovoltaic (PV) power production, approximated by trajectories. We compared the PMM against the de facto method of generating trajectories, namely by separately forecasting the marginal predictive distributions and estimating the covariance matrix that describes the dependencies between forecast horizons. Quantile regression forests (QRF) were used to forecast the marginals and mostly outperformed the PMM, although block bootstrapping indicated that we could not reject the null hypothesis of an accuracy difference between the forecasts with $\alpha = 5\%$. We argue that because of its simplicity and efficiency, PMM can be a valuable asset

TABLE I
BLOCK BOOTSTRAPPED LOSS DIFFERENTIAL PRESENTED AS $\mu \pm \sigma$ $(2.5\% - 97.5\%)$.

| | ES | | | |
|---|---|---|---|---|
| | PMM+QS | QRF+QS | PMM+MS | QRF+MS |
| PMM+QS | 0±0 (0—0) | 0.06±0.31 (-0.5—0.83) | -0.08±0.21 (-0.67—0.31) | 0.13±0.32 (-0.4—0.82) |
| QRF+QS | (-) | 0±0 (0—0) | -0.11±0.29 (-0.78—0.46) | 0.11±0.22 (-0.2—0.68) |
| PMM+MS | (-) | (-) | 0±0 (0—0) | 0.15±0.27 (-0.31—0.74) |
| QRF+MS | (-) | (-) | (-) | 0±0 (0—0) |
| | VS | | | |
| | PMM+QS | QRF+QS | PMM+MS | QRF+MS |
| PMM+QS | 0±0 (0—0) | 397.02±1116.4 (-1557.52—3187.29) | -246.11±794.92 (-1996.46—1632.91) | 722.43±1199.35 (-977.84—3611.67) |
| QRF+QS | (-) | 0±0 (0—0) | -533.09±1108.93 (-3073.65—1598.25) | 534.45±1119.49 (-657.42—4011.29) |
| PMM+MS | (-) | (-) | 0±0 (0—0) | 776.92±1091.41 (-770.52—3274.18) |
| QRF+MS | (-) | (-) | (-) | 0±0 (0—0) |

TABLE II
CRPS, ES AND VS SKILL SCORES, RELATIVE TO MUPEN. NOTE THAT WE COMPUTE THE MEAN AND STANDARD DEVIATION ($\mu \pm \sigma$) OVER ALL FORECAST HORIZONS FOR CRPS.

| Model | CRPS | ES | VS |
|---|---|---|---|
| PMM+MS | 0.188 ± 0.250 | 0.237 | 0.384 |
| PMM+QS | 0.241 ± 0.234 | 0.272 | 0.398 |
| QRF+QS | 0.289 ± 0.221 | 0.308 | 0.452 |
| QRF+MS | 0.348 ± 0.203 | 0.361 | 0.536 |

in a forecaster's toolbox. Lastly, it is important to highlight that the limited amount of available data may have had a negative effect on the quality of the analogs found by the PMM, whereas QRF is known to be robust against small sample sizes. Furthermore, extensive feature selection could improve the PMM's performance further. These aspects will be investigated in future work.

## ACKNOWLEDGMENT

The authors wish to thank ECMWF for the access to Numerical Weather Predictions and HEDNO for the provision of renewable production data.

## REFERENCES

[1] A. H. Murphy, What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting, Weather and Forecasting 8 (2) (1993) 281–293.

[2] Y. V. Makarov, P. V. Etingov, J. Ma, Z. Huang, K. Subbarao, Incorporating uncertainty of wind power generation forecast into power system operation, dispatch, and unit commitment procedures, IEEE Transactions on Sustainable Energy 2 (4) (2011) 433–442.

[3] S. Camal, A. Michiorri, G. Kariniotakis, Optimal offer of automatic frequency restoration reserve from a combined pv/wind virtual power plant, IEEE Transactions on Power Systems 33 (6) (2018) 6155–6170.

[4] S. Camal, F. Teng, A. Michiorri, G. Kariniotakis, L. Badesa, Scenario generation of aggregated wind, photovoltaics and small hydro production for power systems applications, Applied Energy 242 (2019) 1396–1406.

[5] D. Yang, E. Wu, J. Kleissl, Operational solar forecasting for the real-time market, International Journal of Forecasting 35 (4) (2019) 1499–1519.

[6] H. Vasconcelos, C. Moreira, A. Madureira, J. P. Lopes, V. Miranda, Advanced control solutions for operating isolated power systems: Examining the portuguese islands., IEEE Electrification Magazine 3 (1) (2015) 25–35.

[7] P. Pinson, A. Pierrot, L. Han, R. Bessa, J. Gouveia, C. Moreira, S. Camal, Identification of new and enhanced forecasting products, Tech. rep., Smart4RES Deliverable 1.2 (2021).

[8] D. van der Meer, G. C. Wang, J. Munkhammar, An alternative optimal strategy for stochastic model predictive control of a residential battery energy management system with solar photovoltaic, Applied Energy 283 (2021) 116289.

[9] P. Pinson, Wind energy: Forecasting challenges for its operational management, Statistical Science 28 (4) (2013) 564–585.

[10] A. Mueen, Y. Zhu, M. Yeh, K. Kamgar, K. Viswanathan, C. Gupta, E. Keogh, The fastest similarity search algorithm for time series subsequences under euclidean distance (August 2017).

[11] J. L. Bentley, Multidimensional binary search trees used for associative searching, Commun. ACM 18 (9) (1975) 509–517.

[12] D. Yang, D. van der Meer, Post-processing in solar forecasting: Ten overarching thinking tools, Renewable and Sustainable Energy Reviews 140 (2021) 110735.

[13] T. Carriere, C. Vernay, S. Pitaval, G. Kariniotakis, A Novel Approach for Seamless Probabilistic Photovoltaic Power Forecasting Covering Multiple Time Frames, IEEE Transactions on Smart Grid (2019) 1–1.

[14] P. Pinson, H. Madsen, H. A. Nielsen, G. Papaefthymiou, B. Klöckl, From probabilistic forecasts to statistical scenarios of short-term wind power production, Wind Energy 12 (1) (2009) 51–62.

[15] S. Alessandrini, L. Delle Monache, S. Sperati, G. Cervone, An analog ensemble for short-term probabilistic solar power forecast, Applied Energy 157 (2015) 95–110.

[16] D. Yang, Ultra-fast analog ensemble using kd-tree, Journal of Renewable and Sustainable Energy 11 (5) (2019) 053703.

[17] N. Meinshausen, Quantile regression forests, Journal of Machine Learning Research 7 (2006) 983–999.

[18] D. van der Meer, A benchmark for multivariate probabilistic solar irradiance forecasts, Solar Energy 225 (2021) 286–296.

[19] M. Lefèvre, A. Oumbe, P. Blanc, B. Espinar, B. Gschwind, Z. Qu, L. Wald, M. Schroedter-Homscheidt, C. Hoyer-Klick, A. Arola, A. Benedetti, J. W. Kaiser, J.-J. Morcrette, McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions, Atmospheric Measurement Techniques 6 (9) (2013) 2403–2418.

[20] H. D. Kambezidis, B. E. Psiloglou, Estimation of the optimum energy received by solar energy flat-plate convertors in greece using typical meteorological years. part i: South-oriented tilt angles, Applied Sciences 11 (4) (2021).

[21] M. Leutbecher, T. Palmer, Ensemble forecasting, Journal of Computational Physics 227 (7) (2008) 3515–3539, predicting weather, climate and extreme events.

[22] T. Gneiting, A. E. Raftery, Strictly proper scoring rules, prediction, and estimation, Journal of the American Statistical Association 102 (477) (2007) 359–378.

[23] P. Pinson, R. Girard, Evaluating the quality of scenarios of short-term wind power generation, Applied Energy 96 (2012) 12–20, smart Grids.

[24] M. Scheuerer, T. M. Hamill, Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities, Monthly Weather Review 143 (4) (2015) 1321 – 1334.

[25] E. Gilleland, A. S. Hering, T. L. Fowler, B. G. Brown, Testing the tests: What are the impacts of incorrect assumptions when applying confidence intervals or hypothesis tests to compare competing forecasts?, Monthly Weather Review 146 (6) (2018) 1685 – 1703.