Generalizing Renewable Energy Forecasting Using Automatic Feature Selection and Combination

Dennis van Der Meer, Simon Camal, Georges Kariniotakis

Generalizing Renewable Energy Forecasting Using Automatic Feature Selection and Combination

Dennis van der Meer, Simon Camal, Georges Kariniotakis
Center for processes, renewable energies and energy systems (PERSEE)
MINES ParisTech, PSL University, Sophia Antipolis, France
{dennis.van_der_meer, simon.camal, georges.kariniotakis}@minesparis.psl.eu

***Abstract*—Spatially aggregating renewable power plants is beneficial when participating in electricity markets. In this context, a substantial number of features is available from various data sources. In machine learning, feature selection is common so as to relieve the curse of dimensionality and avoid overfitting. However, there is no guarantee that the selected features result in reliable forecasts and post-processing can therefore be valuable. In this study, we combine model agnostic feature selection with linear and nonlinear probabilistic forecast combination techniques. Moreover, the filters automatically compute the weights for our analog ensemble (AnEn) forecast model. We verify our model chain by generating intra-day forecasts of the aggregated output of 20 photovoltaic power plants using 831 input features in total. We show that the collection of filters selects a heterogeneous feature set but that each individual AnEn-filter combination results in underdispersed forecasts, which is efficiently remedied by the forecast combination techniques.**

***Index Terms*—Filtering, forecast combination, high-dimensional data, virtual power plant, probabilistic forecasts**

I. Introduction

The power generated by renewable energy sources (RESs) such as photovoltaic (PV) systems or wind turbines is highly variable and it is well-established that probabilistic forecasts contribute to optimising power system management when used in functions like unit commitment [1]. Unit commitment is usually performed on the day-ahead market (DAM) and real-time market (RTM), where penalties are incurred based on the forecast errors. Consequently, the accuracy of the probabilistic forecasts is important, which is challenged by the aforementioned variability. The variability can be relieved by aggregating spatially distributed power plants. The resulting virtual power plant (VPP) outputs a smoother power profile caused by the reduced spatial correlation. However, forecasting the VPP output requires more input features as the VPP likely covers large geographical areas reflected by multiple grid points of Numerical Weather Prediction (NWP) models or satellite images.

There are at least two challenges that arise when including vastly more input features: (i) features may be multicollinear; and (ii) the curse of dimensionality. In other words, it is important to select important features and discard redundant ones in order to retain accuracy and reduce time complexity. Often, feature selection is classified into three categories, namely: (i) filtering methods; (ii) wrapper methods; and (iii) embedded methods [2]. Filtering methods rank features based on a score, e.g., Pearson correlation, to subset the features, and is therefore model agnostic. Recently, an ultra-fast algorithm for similarity search was proposed to preselect relevant features [3]. Wrapper methods aim to find the optimal feature subset by subsetting the feature set, learning a predictor on this subset and comparing its performance to the same predictor on other feature subsets. Wrapper methods are not common due to their computational cost although a brute-force attempt to find a relevant subset of endogenous variables was made in [4]. Finally, the embedded methods embed the feature selection method, such as Lasso regression [5]. A multitude of studies in solar forecasting have applied embedded methods, e.g, Lasso combined with quantile regression [6].

As the penetration of RESs increases, it becomes more pertinent to aggregate these assets as VPPs in order to participate—with high reliability—in market bidding and power system regulation. For such large portfolio end-users dispose often multiple forecasts originating from different service providers or different NWPs etc. In this paper we combine feature selection methods with forecast combination techniques to derive a model that outperforms the state of the art. It is well-known that forecast combination can reduce modeling uncertainty although it is not commonly researched in solar forecasting [7]. Probabilistic forecasts can be combined using heuristic, linear and nonlinear combination methods; see [7] for an overview. For instance, the opinion linear pool (OLP) is a heuristic method that linearly combines predictive distributions by assigning equal weights to each expert. In solar forecasting, minimization of the continuous ranked probability score (CRPS) has been used to learn the weights to linearly combine predictive distributions in a traditional linear pool (TLP) [8]. Thorey et al. [9] developed an online optimization of the CRPS to linearly combine forecasts, which improved the calibration although underdispersion remained. Besides TLP, nonlinear combination methods have been proposed, i.e., the spread-adjust linear pool (SLP) and the beta-transformed linear pool (BLP) [10]. Möller and Groß used SLP to calibrate the ensemble prediction system of the European Center for Medium-range Weather Forecasts (ECMWF) [11], while BLP was used to combine two parametric predictive densities from regression models based solely on lagged observations [12].

The present research was carried as part of the Smart4RES Project (European Union's Horizon 2020, No. 864337).

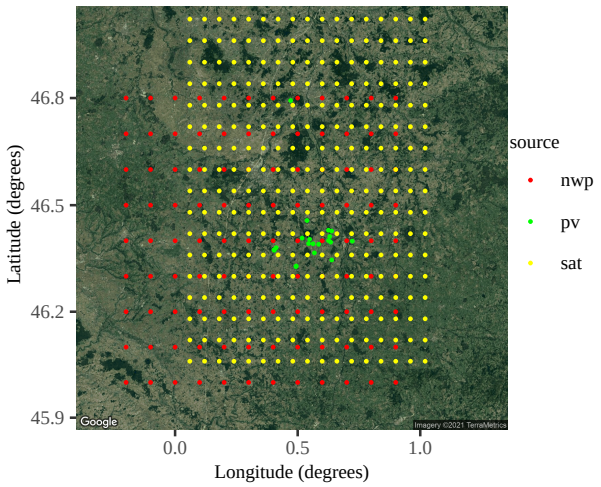

Fig. 1. Overview of the satellite and NWP grid points and the PV systems.

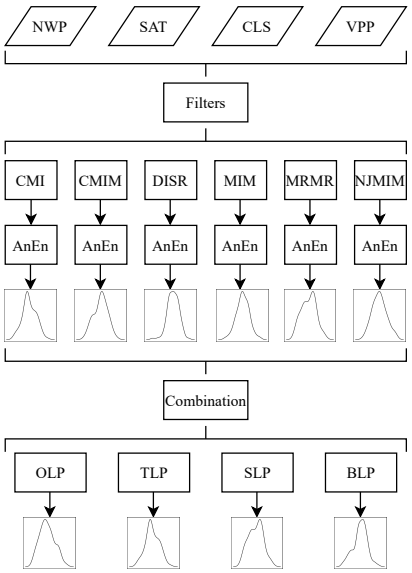

Fig. 2. A flowchart of the pre-process, forecast, and post-process steps.

Consequently, there is need for additional study into nonlinear combination methods applied to solar forecasting.

In this study, we focus on operational forecasts of the intra-day aggregated output of a VPP that consists of 20 PV plants with a combined nominal capacity of 4005 kW, located in mid-west France. The total number of features is 831: 289 satellite pixels and 108 NWP grid points with 5 variables and 2 astronomical variables (cf Fig. 1). We generate the probabilistic forecasts with an Analog Ensemble (AnEn), which is an established forecasting method that compares the current weather state to historical weather states and outputs the observed power generation accompanying the most similar historical states as an ensemble forecast [13]. We use 6 filtering methods based on mutual information to select the input features for our AnEn model and automatically compute their weights. Since a recent benchmark study of 22 filter methods concludes that "there is no subset of filter methods that outperforms all other filter methods" [14], we combine the predictive distributions using 1 linear and 2 nonlinear combination methods. The contributions of this study can therefore be summarized as follows: (i) we employ 6 filtering methods based on mutual information to account for the fact that no filter works best in all circumstances [14] and to account for the nonlinearity between PV power generation and the features; (ii) we apply linear and nonlinear probabilistic combination methods to minimize model uncertainty, which is an underserved area of research in solar forecasting [7]; and (iii) these methods combined generalize existing probabilistic forecasting approaches by automating filtering and model combination.

The remainder of this paper is organized as follows. Section II details the methods used in this study. Then, Section III presents the results and Section IV concludes this paper.

II. Methodology

This section describes the methodology of the paper and starts with describing the filtering methods and the forecast combination methods. Then, the forecast evaluation tools used in this study are described. Figure 2 presents an overview of the methodology.

A. Filtering

Filtering is a category of feature selection in which a score is computed to rank the importance of features. The goal is to find a feature subset $\mathbb{S}$ from the complete feature set $\mathbb{X}$ that represents the target variable Y with high accuracy and minimal residual uncertainty [14]. In this study, the filtering methods are based on mutual information, which originates from information theory and describes the amount of information that can be known about random variable Y when knowing random variable X [14].

For the remainder, it is helpful to introduce the entropy of a discretized random variable Y with marginal probability mass function p [14]:

$$H(Y) = -\sum_{y} p(y)\log_2\left(p(y)\right), \tag{1}$$

which measures the uncertainty of Y. In addition, the conditional entropy of Y given X is defined as [14]:

$$H(Y|X) = \sum_{x} p(x)\left(-\sum_{y} p(y|x)\log_2\left(p(y|x)\right)\right). \tag{2}$$

The mutual information between Y and X is then defined as $I(Y;X) = H(Y) - H(Y|X)$ and describes how the uncertainty of Y is lowered by knowing X [14]. In this study, we use the `praznik` package [15] in the statistical software R. The `praznik` toolbox discretizes the range of the continuous features into $\max\{\min\{\frac{n}{3}, 10\}, 2\}$ intervals that are equally spaced, where n is the number of training samples [15].

As mentioned above, we use 6 different filtering methods based on mutual information. The first filter, mutual information maximization (MIM), computes the mutual information between feature k and target variable Y as:

$$J_{\text{MIM}}(X^{(k)}) = I(Y; X^{(k)}), \tag{3}$$

and returns a predetermined number of features that maximizes J [15].

The remaining filters greedily add a feature to $\mathbb{S}$ that at each iteration maximizes the score $J(X^{(k)})$. The first of those is minimal conditional mutual information (CMIM), which uses the score

$$J_{\text{CMIM}}(X^{(k)}) = \min\left\{ I(Y; X^{(k)}), \min_{X^{(j)} \in \mathbb{S}} I(Y; X^{(k)} | X^{(j)}) \right\}, \tag{4}$$

where $I(Y; X^{(k)}|X^{(j)}) = H(Y|X^{(j)}) - H(Y|X^{(k)}, X^{(j)})$ [15]. Similar to MI, CMIM describes how knowing $X^{(k)}$ lowers the uncertainty of Y given that we already know $X^{(j)}$.

The third filter is conditional mutual information (CMI), which ranks features according to the score:

$$J_{\text{CMI}}(X^{(k)}) = I(Y; X^{(k)} | \mathbb{S}), \tag{5}$$

and essentially evaluates the added value of feature $X^{(k)}$ given the already selected features.

The fourth filter is double input symmetrical relevance (DISR), which uses the score [15]:

$$J_{\text{DISR}}(X^{(k)}) = \sum_{X^{(j)} \in \mathbb{S}} \frac{I(Y; X^{(k)}, X^{(j)})}{H(Y, X^{(k)}, X^{(j)})}, \tag{6}$$

where $I(Y; X^{(k)}, X^{(j)})$ evaluates the complementary information that $X^{(k)}$ and $X^{(j)}$ provide for Y, whereas $H(Y, X^{(k)}, X^{(j)})$ reduces the possibility to choose highly variable features [14].

Maximum relevance and minimum redundancy (MRMR) is the fifth filter and uses the following score to rank the features:

$$J_{\text{MRMR}}(X^{(k)}) = I(Y; X^{(k)}) - \frac{1}{|\mathbb{S}|} \sum_{X^{(j)} \in \mathbb{S}} I(X^{(k)}; X^{(j)}), \tag{7}$$

which is designed to ensure that there is as little redundancy as possible between $X^{(k)}$ and $\mathbb{S}$ (second term) and providing maximal information about Y [15].

The sixth and final filter is minimal normalized joint mutual information (NJMIM), which scores the features according to [15]:

$$J_{\text{NJMIM}}(X^{(k)}) = \min_{X^{(j)} \in \mathbb{S}} \left\{ \frac{I(Y; X^{(k)}, X^{(j)})}{H(Y, X^{(k)}, X^{(j)})} \right\}, \tag{8}$$

which is a modification of filter DISR and evaluates the minimal relative information between Y, $X^{(k)}$ and already selected features instead of the sum [14].

Our algorithm runs over testing time instances $t = n + 1, \cdots, n'$. The data that the filters are applied to comprise the d previous days recorded during the same time stamp as t (expressed as "HH:MM"). This setup allows us to include the most recent data in our filtering method. In order to account for temporal dependencies, we also include one time instant prior to "HH:MM" and one time instant after "HH:MM" (except for $t + 1$). Consequently, the filters have a total of $3 \cdot d - 1$ time instances at their disposal.

These filters have been selected because (i) they can uncover nonlinear relationships; (ii) do not require the random variables to be on the same scale; (iii) performed decently as [14] concludes; (iv) tend to select a different order of features [14], which could benefit forecast diversity; and (v) except for MIM, take into account interactions between features.

B. Forecast Combination

As mentioned above, OLP is a heuristic method in which m predictive distributions F_j are linearly combined as $G = \sum_{j=1}^{m} w_j F_j$ where $w_j \geq 0$ and $\sum_{j=1}^{m} w_j = 1$ with equal weights $w_j = 1/m$. In this work, however, we determine the weights w_j based on a scoring rule; specifically, the logarithmic score (IGN) as proposed by Gneiting and Ranjan [10]. The following subsections describe the three combination methods in detail. In addition, OLP is used as a reference method to compare the combination methods below with.

1) Linear Combination: The traditional linear pool (TLP) is similar to the opinion pool, except that we determine the weights by optimizing the logarithmic score. The weights can be determined by evaluating $g_i(y) = \sum_{j=1}^{m} w_j f_{ij}(y)$ and minimizing over n fitting samples:

$$\text{IGN} = -\frac{1}{n} \sum_{i=1}^{n} \log\left(\sum_{j=1}^{m} w_j f_{ij}(y_i) \right), \tag{9}$$

which is equivalent to maximizing the log-likelihood [7]. The main challenge of TLP is that it increases dispersion (cf. Theorem 3.1 in [10]) and is therefore not suited when the forecast members are probabilistically calibrated or overdispersed.

2) Nonlinear combination: In this study, we consider two nonlinear combination methods. The first is the spread-adjusted linear pool (SLP) that includes a strictly positive parameter c which adjusts the spread of the predictive distributions. The combined predictive distribution is defined as follows [10]:

$$G_i^c(y) = \sum_{j=1}^{m} w_j F_{ij}^0\left(\frac{y - \mu_{ij}}{c} \right), \tag{10}$$

where $F_{ij} = F_{ij}^0(y - \mu_{ij})$ and μ_{ij} is the median of predictive distribution of F_{ij}. Equation 10 shows that, in case $c = 1$, SLP reduces to TLP. Setting $c < 1$ tends to improve calibration when the individual predictive distributions are overdispersed or neutrally dispersed and setting $c \geq 1$ may benefit underdispersed predictive distributions [10]. The flexibility of SLP is limited but it can be effective when the component forecasts are underdispersed or neutrally dispersed [10].

Therefore, we additionally consider the beta-transformed linear pool. Generalized by Gneiting and Ranjan [10], the

predictive cumulative distribution function (CDF) of the beta-transformed linear pool is defined as

$$G_i^{\alpha,\beta}(y) = B_{\alpha,\beta}\left(\sum_{j=1}^{m} w_j F_{ij}(y)\right), \tag{11}$$

where $B_{\alpha,\beta}$ represents the beta CDF with parameters $\alpha > 0$ and $\beta > 0$, and BLP reduces to TLP when $\alpha = \beta = 1$. Gneiting and Ranjan [10] prove that when the weights $w_1, \ldots, w_m$ are fixed, the variance of the probability integral transform random variable can attain any value in the open interval $\left(0, \frac{1}{4}\right)$, with $\frac{1}{12}$ indicating neutrally dispersed probabilistic forecasts.

C. Analog Ensemble

The work in [13] was one of the first studies that applied AnEn to solar forecasting. Therein, the authors compare the most recent NWP forecast to historical NWP forecasts and select the most similar NWP historical forecasts (the analogs), of which the corresponding PV power measurements form the ensemble forecast. In this study, we modify the original similarity metric to include additional data sources, similar to [16]. Furthermore, we leverage the algorithmic efficiency of k-dimensional tree (kd-tree), as suggested by [17]. For this, we further modify the original similarity metric such that the lags of the features are distinct features [7]:

$$d(\boldsymbol{\mathcal{X}}_t^h, \boldsymbol{\mathcal{A}}_i^h) = \sqrt{\sum_{j=1}^{J} w_j^h \left(x_t^{(j)} - x_i^{(j)}\right)^2}, \tag{12}$$

where $J = N_v \cdot (2\tilde{t} + 1)$ is the total dimension, w_j^h is the j^{th} weight for the h^{th} forecast horizon and $x^{(j)}$ is the j^{th} feature. Vector $\boldsymbol{\mathcal{X}}_t^h$ contains the most recent filtered features, which comprises the NWP forecasts and clear-sky data (CLS) valid at time t, as well as the satellite derived clear-sky index and PV power measurement observed h time steps before. Vector $\boldsymbol{\mathcal{A}}_i^h$ contains the same filtered features as $\boldsymbol{\mathcal{X}}_t^h$ except historical ones. Note that the features in $\boldsymbol{\mathcal{A}}_i^h$ are scaled and centered and these scaling factors are used to scale and center $\boldsymbol{\mathcal{X}}_t^h$.

Herein, the weights w_j^h in eq. 12 are taken directly from the scores that the filtering methods assign to the features. The weights are subsequently normalized by the sum of all feature scores so that the weights sum to 1. This allows us to dynamically update the feature weights instead of employing a costly wrapper.

D. Complete History Persistence Ensemble

We employ the complete history persistence ensemble (CH-PeEn) to evaluate the forecast methods described above [18]. CH-PeEn is an external multiple-valued climatology, is independent of forecast horizon and its output is calibrated by design [18]. Note that the CH-PeEn is constructed using detrended PV power measurements, as Section II-F details.

E. Forecast Verification

An important aspect of forecasts is their quality, which in the case of probabilistic forecasts consists of calibration, resolution and sharpness. To evaluate the probabilistic calibration of our forecasts, we compute the probability integral transform (PIT) $z_t = F_t(y_t)$ and plot these as a histogram for visual verification [10]. In case of probabilistic calibration, the resulting histogram is close to flat and has a variance close to $1/12$.

In addition, we evaluate the forecasts numerically and normalize the results using the nominal capacity of the VPP. The first score is the CRPS, which is a strictly proper and a negatively-oriented score. Interestingly, it is possible to decompose the average CRPS into reliability, uncertainty and resolution terms as CRPS $=$ *Reliability* $+$ *Uncertainty* $-$ *Resolution* [19]. Additionally, we use the CRPS to compute a skill score, i.e., the relative improvement over CH-PeEn, that has a maximum value of 100%. Finally, we use the square root of the average of the variance (RMV) of the predictive distributions as a measure of the sharpness [10].

Although this study focuses on probabilistic forecasts, the root mean squared error (RMSE) is arguably one of the most common verification metrics and is therefore included in our analysis.

F. Data

The available PV power measurements extend from 2019-01-01 until 2020-09-30 and measurements recorded at a zenith angle larger than 85° are removed. The period from 2019-01-01 until 2019-09-30 is used for training, while 2019-10-01 until 2019-12-31 is used to optimize the combination weights. The remainder of the data is used for testing. The PV plants have the same orientation and tilt, therefore we compute the clear-sky global tilted irradiance (GTI_{cs}) at the averaged coordinates. We detrend the PV power measurements by dividing them with the GTI_{cs} and note that more elegant solutions exist but they require additional system information that is unavailable to us. We include the last PV power observation as feature (referred to as "VPP"), as well as the zenith angle and solar azimuth as features (referred to as "CLS").

The NWP forecasts come from the ECMWF, issued daily at 00:00 UTC with a spatial resolution of $0.1° \times 0.1°$. The 5 features that we extract are surface solar radiation downwards (SSRD), 100 m U- and V-wind speed (U100 and V100), 2 m temperature (2T) and total cloud cover (TCC). The SSRD is converted to global horizontal irradiance (GHI) and detrended to the clear-sky index using the McClear clear-sky model [20].

Finally, we include satellite derived GHI from Meteosat Second Generation (MSG) satellite images (referred to as "SAT"). The GHI is derived using an improvement over the Heliosat-2 method [21] and stored as a GHI time series for each pixel. Similar to NWP GHI forecasts, the satellite derived GHI is converted to the clear-sky index. Figure 1 indicates where the satellite and NWP grid points and PV systems are located.

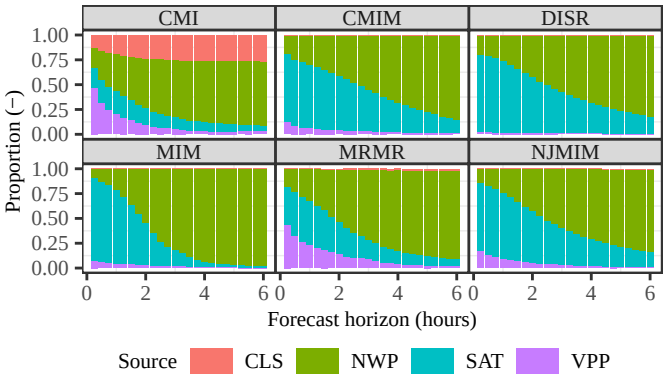

Fig. 3. Normalized scores of the filters as a function of forecast horizon, used as dynamic weights in the AnEn model. Recall that VPP represents PV power at $t-h$, SAT satellite derived clear-sky index at $t-h$ and CLS the zenith angle and solar azimuth at t.

III. Results

The first contribution of this study pertains to feature selection by means of the filters described in Section II-A. Figure 3 presents the normalized weights assigned by each filter to the feature groups as a function of the forecast horizon and for each forecast issue time. As can be seen, there are some notable differences between the filters. For instance, filter CMI includes feature group CLS even though the input and output features have been detrended. Note also that filters DISR and MIM to a large extent ignore the most recent power measurement. This can be expected of MIM since it does not consider interaction between input features and is therefore likely to select many similar features if they are relevant to Y. Since DISR evaluates the complementary information between input features, it does not specifically promote feature diversity. As mentioned above, filter NJMIM is a modification of DISR by evaluating the minimal relative information between features, which likely helps it to select more diverse features. As a final note, CMI and MRMR select diverse features as they are designed to reduce redundancy.

The second contribution concerning forecast combination aims at improving the calibration of the component forecasts. Table I presents the combination weights and SLP and BLP parameters, expressed as "mean $\pm$ standard deviation" over all forecast horizons. The differences between the combination methods are not drastic although it is interesting to note that BLP favors MIM and DISR, whereas TLP and SLP assign less importance to these filters. Interestingly, SLP parameter c is lower than 1, which may suggest overdispersed component forecasts [10]. In this case, we reason that the combination introduces overdispersion and parameter c tries to compensate.

Figure 4 presents the PIT histograms of the AnEn forecasts generated with the inputs provided by the filters and their combinations. The value in each subfigure corresponds to the variance, which should be close to $1/12$ (≈ 0.083) for neutral dispersion [10]. The figure shows that the predictive distributions are calibrated in the main distribution but deviate at the extreme quantiles, which could be attributed to the selection of too few or redundant features. Linear forecast combination increases the dispersion [10], and Fig. 4 confirms this; OLP and TLP are both much closer to uniformity and their variance is closer to $1/12$. In our setting, SLP combines the component forecasts most effectively, i.e., results in the most desirable calibration. Interestingly, BLP introduces negative bias, which could be due to the fact that the component forecasts are underdispersed, due to limited training instances or due to a discrepancy between the validation and verification sets.

The numerical results are presented in Fig. 5. It can be seen that the combination methods, especially OLP, TLP and SLP, improve upon the component forecasts across all scores and forecast horizons. Interesting to note is that OLP tends to perform only slightly worse when compared to TLP and SLP, which indicates that linear forecast combination can be highly effective with minimal computational effort as long as the component forecasts are underdispersed. Furthermore, it can be seen that the forecasts resulting from the filters are only slightly less sharp than the combination methods, which indicates that the PIT histogram shape is caused by the lowest (highest) quantile frequently being above (below) the target.

IV. Concluding Discussion

In this study, we generated intra-day forecasts of the aggregated output of a virual power plant with a nominal capacity of 4005 kW. We applied the analog ensemble (AnEn) as a base model as it requires no training. Given the large number of inputs comprising numerical weather prediction, satellite, endogenous and astronomical data, we proposed to use a variety of filters to (i) reduce the problem dimensionality and (ii) automatically and dynamically learn the AnEn feature weights. The results showed that the collection of filters selects a rather heterogeneous feature set but that each individual AnEn-filter combination suffers calibration issues.

We employed linear and nonlinear forecast combination to improve the component forecasts, since that is an underserved area of research in solar forecasting. We showed that even a naive linear combination significantly improves the calibration, which is supported by theoretical results [10]. In our case, the spread-adjusted linear pool proved to be most effective

TABLE I
The combination weights and SLP and BLP parameters expressed as "mean $\pm$ standard deviation" over all forecast horizons.

Model	CMI	CMIM	DISR	MIM	MRMR	NJMIM	c	α	β
TLP	0.267 $\pm$ 0.07	0.214 $\pm$ 0.097	0.076 $\pm$ 0.056	0.06 $\pm$ 0.084	0.274 $\pm$ 0.06	0.108 $\pm$ 0.033	—	—	—
SLP	0.277 $\pm$ 0.065	0.206 $\pm$ 0.099	0.087 $\pm$ 0.065	0.068 $\pm$ 0.078	0.253 $\pm$ 0.057	0.109 $\pm$ 0.046	0.931 $\pm$ 0.028	—	—
BLP	0.227 $\pm$ 0.032	0.206 $\pm$ 0.066	0.11 $\pm$ 0.063	0.131 $\pm$ 0.069	0.224 $\pm$ 0.05	0.102 $\pm$ 0.026	—	0.855 $\pm$ 0.039	1.388 $\pm$ 0.069

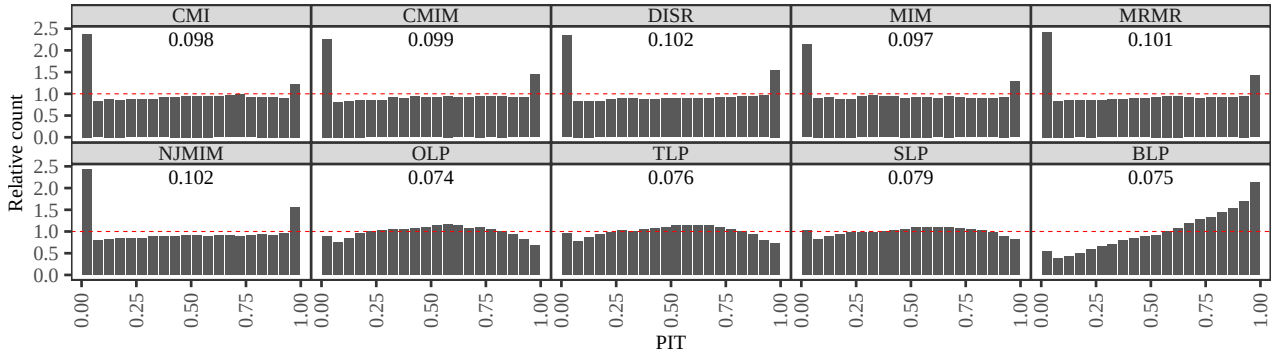

Fig. 4. Histograms of the PIT variables. The value in the subfigure corresponds to the variance, which should be close to $1/12$ for neutral dispersion.

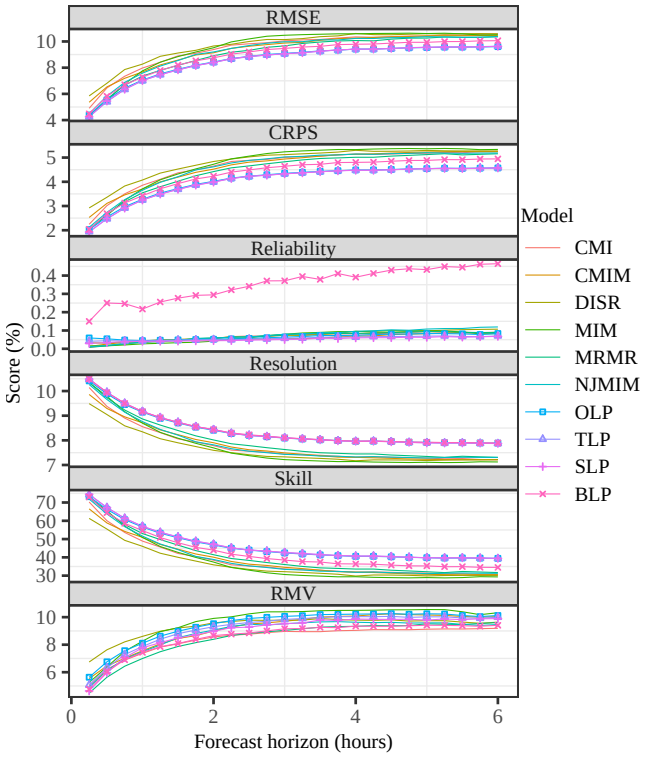

Fig. 5. The normalized numerical scores as a function of forecast horizon.

at calibrating component forecasts as well as maximizing sharpness. Regardless, we recommend forecast practitioners to at least include naive linear forecast combination.

An interesting direction for future research on feature selection would be to augment the filters with a time-dependent factor to weigh recent data more. Regarding forecast combination, interesting directions for future research include (i) combining a wider variety of forecast models and (ii) developing methods to compute the combination weights online.

References

[1] J. Wang, et al., Wind power forecasting uncertainty and unit commitment, Applied Energy 88 (11) (2011) 4014–4023.

[2] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, Journal of Machine Learning Research 3 (2003) 1157–1182.

[3] D. Yang, Ultra-fast preselection in lasso-type spatio-temporal solar forecasting problems, Solar Energy 176 (2018) 788 – 796.

[4] D. van der Meer, M. Shepero, A. Svensson, J. Widén, J. Munkhammar, Probabilistic forecasting of electricity consumption, photovoltaic power generation and net demand of an individual building using gaussian processes, Applied Energy 213 (2018) 195–207.

[5] R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society. Series B (Methodological) 58 (1) (1996) 267–288.

[6] X. G. Agoua, R. Girard, G. Kariniotakis, Probabilistic Model for Spatio-Temporal Photovoltaic Power Forecasting, IEEE Transactions on Sustainaible Energy - (-) (2018) 1–9.

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

[8] A. Bracale, G. Carpinelli, P. De Falco, A Probabilistic Competitive Ensemble Method for Short-Term Photovoltaic Power Forecasting, IEEE Transactions on Sustainable Energy 8 (2) (2017) 551–560.

[9] J. Thorey, C. Chaussin, V. Mallet, Ensemble forecast of photovoltaic power with online CRPS learning, International Journal of Forecasting 34 (4) (2018) 762–773.

[10] T. Gneiting, R. Ranjan, Combining predictive distributions, Electronic Journal of Statistics 7 (2013) 1747–1782.

[11] A. Möller, J. Groß, Probabilistic temperature forecasting with a heteroscedastic autoregressive ensemble postprocessing model, Quarterly Journal of the Royal Meteorological Society 146 (726) (2020) 211–224.

[12] S. A. Fatemi, A. Kuh, M. Fripp, Parametric methods for probabilistic forecasting of solar irradiance, Renewable Energy 129 (2018) 666–676.

[13] 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.

[14] A. Bommert, X. Sun, B. Bischl, J. Rahnenführer, M. Lang, Benchmark for filter methods for feature selection in high-dimensional classification data, Computational Statistics & Data Analysis 143 (2020) 106839.

[15] M. B. Kursa, praznik: Tools for Information-Based Feature Selection, r package version 8.0.0 (2020).

[16] 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 11 (3) (2020) 2281–2292.

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

[18] D. Yang, A universal benchmarking method for probabilistic solar irradiance forecasting, Solar Energy 184 (2019) 410–416.

[19] H. Hersbach, Decomposition of the continuous ranked probability score for ensemble prediction systems, Weather Forecasting 15 (5) (2000) 559–570.

[20] M. Lefèvre, et al., McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions, Atmospheric Measurement Techniques 6 (9) (2013) 2403–2418.

[21] L. F. Zarzalejo, J. Polo, L. Martín, L. Ramírez, B. Espinar, A new statistical approach for deriving global solar radiation from satellite images, Solar Energy 83 (4) (2009) 480–484.