

Probabilistic forecasting of photovoltaic power supply — A hybrid approach using D-vine copulas to model spatial dependencies

A. Schinke-Nendza^{a,*}, F. von Loeper^b, P. Osinski^a, P. Schaumann^b, V. Schmidt^b, C. Weber^a

^a House of Energy Markets and Finance, University of Duisburg-Essen, Universitätsstr. 12, D-45117 Essen, Germany

^b Institute of Stochastics, Ulm University, Helmholtzstr. 18, D-89069 Ulm, Germany

ARTICLE INFO

Keywords:

Solar power supply
Forecasting
Physical PV model
VARX model
Error distribution
D-vine copula
Spatial dependency

ABSTRACT

The fast growth of installed photovoltaic capacity is leading to an increasing impact of variable photovoltaic generation on the overall electricity industry, affecting all stakeholders in this sector. As a consequence, the importance of appropriate photovoltaic power forecasts for planning and decision support is rising, to cope with the resulting uncertainty. In particular, probabilistic forecasts are becoming increasingly important to assess the underlying risks, e.g., depicting the effect of adverse combinations. Whereas deterministic forecasts, while having the advantage of being more detailed, suffer from reflecting only an average expectation. Therefore, this paper proposes a comprehensive hybrid approach to generate deterministic and probabilistic photovoltaic power forecasts, while introducing several improvements for intra-day and day-ahead modelling and forecasting applications. In this context, several pre- and post-processing steps have been combined for the deterministic model, while the spatial interrelation of the forecasting errors is taken into account by applying D-vine copulas for the probabilistic forecasts. The reliability of the proposed hybrid approach is validated, using a comprehensive case study with high-resolution numerical weather predictions and real-world measurement data over several years for multiple photovoltaic units. Furthermore, the proposed model is benchmarked against various combinations of a photovoltaic power model (with and without statistical post-processing) and typical probabilistic models. As part of the evaluation the Energy score, Variogram-based score and Diebold–Mariano test are applied to evaluate the proposed model and highlight the strong performance of the proposed hybrid approach.

1. Introduction

The decarbonization of the energy sector requires an increasing use of variable renewable energies (RE). As a result, the generation structures are evolving from a centralized structure to a decentral energy system, raising various energy related issues. In this context, photovoltaic (PV) power has become one of the most important RE in Germany over the last decades, reaching an installed capacity of 49 GW in 2019 [1]. In particular, the characteristics of PV generation and other RE lead to a spatial and temporal decoupling of electricity generation and consumption, resulting notably in an increased need for improved modelling and forecasting techniques while providing more flexibility to the energy system. In the future, this need can be alleviated by several options, in particular the use of storage and sector coupling technologies (e.g. between electricity, heat and gas grids and with the mobility sector), but also by adapting energy grid infrastructures. Towards these options, an improved understanding of the applicability of uncertainty forecasts is of utmost importance for

the electric power industry. Hence, forecasting PV power generation becomes increasingly relevant for operational and long-term tasks of system operators and market participants, respectively. For instance, the forecasts are used for a secure network operation, maintenance scheduling and long-term portfolio management, see e.g. [2] where the application of probabilistic weather forecasts in decision making processes is discussed, and [3] considers the ambient temperature as input for an optimal power flow program. Especially situations in which high renewable infeed might cause congestion are often solved by curtailing RE units proactively without taking uncertainties into account. In this context, appropriate forecasts can help to reduce RE curtailment by supporting system operators and market participants in day-ahead and intraday planning. Instead of making decisions based on deterministic values, probabilistic forecasts can be more informative as a range of possible scenarios.

In terms of *deterministic forecasts* of PV power generation, both, statistical and physical models, are established methods. Physical PV

* Corresponding author.

E-mail address: aiko.schinke-nendza@uni-due.de (A. Schinke-Nendza).

<https://doi.org/10.1016/j.apenergy.2021.117599>

Received 2 April 2021; Received in revised form 4 August 2021; Accepted 9 August 2021

Available online 2 September 2021

0306-2619/© 2021 Elsevier Ltd. All rights reserved.

models typically use numerical weather predictions (NWP) and a model representation of the PV power plant to convert meteorological parameters into power output [4]. Most available meteorological measurements encompass global horizontal irradiance (GHI) and its decomposition into direct and diffuse irradiance as well as air temperature. However, in terms of NWP data, GHI might be provided exclusively, which introduces the necessity for a separate model for direct and diffuse irradiance decomposition. In literature, several approaches exist to derive the direct and diffuse irradiance from other meteorological sources. In this context, especially the Boland–Ridley–Laurent (BRL) model, following [5], performs reasonably well compared to other models by using GHI forecasts exclusively. Besides physical models, statistical models can be used to forecast PV generation. These models are typically based on historical time series, live measurement data and the previous forecasting performance, while sometimes taking into account exogenous explanatory variables (e.g. from NWP). A comprehensive overview is given by Antonanzas et al. [6], listing several publications which use different machine-learning techniques such as artificial neural networks (ANNs), support vector machines (SVMs) and random forests. Among the most common techniques are linear autoregressive methods (AR) and auto-regressive moving-average (ARMA) models which in addition to the time series of the power performance itself, take the historical time series of the model error into account. In this context, accurate deterministic forecasts based on statistical PV models are typically complex and computationally intensive, see [7]. However, combinations of statistical and physical models can be used to cope with the non-linear patterns in the residuals and to take advantage of the strengths of both techniques [8].

However, deterministic forecasts are not sufficient to cope with the uncertainty of forecasts which becomes increasingly relevant, e.g. for market participants to balance their portfolio of day-ahead hourly contracts by trading intraday quarter-hourly contracts or for system operators to quantify the risk of congestion. A review on *probabilistic PV power forecasting* has been provided by van der Meer et al. [9]. Common methods are e.g. quantile regression [10], quantile regression forests [11] and analog model chain forecasts [12]. More recently, a bivariate copula model has been proposed for the prediction of level-crossing probabilities of solar power supply [13]. The validation results highlight strong performance of the latter model. The bivariate copula model might be improved by including further influence factors. That topic was addressed in [14] by including influence factors such as further weather information and solar power supply of previous hours as explanatory variables using multivariate D-vine copulas. In addition, probabilistic forecasts of the output power of photovoltaic systems using ANNs seem to be promising, see [15], by improving the prediction accuracy for different seasons and with respect to different prediction horizons. Yet, ANNs are computationally quite expensive and time-consuming to train using conventional CPUs, so this factor still limits their potential applications. Therefore, the focus of the present study is to improve the prediction accuracy of conventional methods by introducing various steps of pre- and post-processing of the available data and by improving the probabilistic predictions by using copulas.

It can be observed that the *forecasting errors* of different PV units are correlated. Hence, the PV power injection at different network nodes might superpose in terms of transmission line loading affecting power system operation. Therefore, *spatial probabilistic forecasts* are needed, which describe the correlated uncertainties at all relevant network nodes, i.e., of those network nodes which are affected by uncertain PV net power injections. Higher-dimensional copulas can be applied to describe the dependency structure of the multidimensional random vectors [16]. In particular, D-vine copulas have been applied to model the spatial [17] and temporal [18] multivariate probability distribution of wind power supply. Moreover, Gaussian and R-vine copulas were utilized to estimate the multivariate probability distribution for the solar power supply of a small sample of individual PV units [19].

Consequently, this paper proposes a novel hybrid approach to generate probabilistic forecasts of PV power supply at different network nodes, by using a combination of physics-based methods for (deterministic) forecasting of the PV power supply together with a data-driven (statistical) pre- and post-processing of the input data and a data-driven (probabilistic) copula model to forecast the spatio-temporal uncertainty of the forecasting errors. Hereby, the physical PV model computes deterministic predictions of solar power supply at the corresponding network nodes which are then post-processed by applying an adaptive vector autoregressive model with exogenous variables (VARX model). Using the forecasting errors of the PV units, in a first step a one-dimensional marginal distribution is fitted for each of the network nodes. In a second step, a D-vine copula is fitted, which describes the spatial dependency of the forecasting errors at different network nodes. Subsequently, probabilistic forecasts are generated using the NWP-based deterministic forecasts of the physical PV model and the uncertainty sets drawn from the D-vine copula model. To evaluate the performance of the proposed approach, the deterministic forecasts are benchmarked to the physical model without statistical post-processing and an adaptive VARX model as a purely statistical approach. For the probabilistic forecasts, the D-vine copula model is compared to a multivariate normal distribution. Furthermore, to see the significance of taking spatial dependence into account, the D-vine copula model is furthermore benchmarked to a probabilistic model applying products of univariate normal distributions describing the forecasting errors at each node, thus assuming spatial independence.

The paper is structured as follows (cf. Fig. 1). In terms of *deterministic forecasts* in Section 2, the physical PV model is introduced briefly. Section 3 explains how the VARX model is applied to improve the forecasts. Regarding the *probabilistic forecasts*, Section 4 introduces D-vine copulas, which are applied to model the error distribution and subsequently combined with the deterministic predictions to generate probabilistic forecasts. Thereupon, the proposed approach is tried out in a real-world *application*. Therefore, an evaluation methodology is presented in Section 5. In Section 6 the hybrid approach is applied to the case study mentioned above, whereby the available data is separated into a training dataset for fitting the models and a test dataset for their evaluation. Thus, the performance of the hybrid approach on the test dataset is benchmarked to alternative approaches of deterministic and probabilistic modelling. Therefore, the *deterministic forecasts* and the *probabilistic forecasts* are investigated for both, intraday and day-ahead applications. Section 7 concludes, highlighting the key findings of this paper.

2. Physical model

A physical PV model is introduced to generate deterministic forecasts of the PV net power injection of individual PV units, see Fig. 2. The model is based on statistically post-processed NWP data which are combined with synoptic observations, see [20,21] and [22,23]. In addition, the available weather data is pre-processed to receive all required inputs for the physical PV model. In contrast to common probabilistic approaches, see [24], the net power injection at a certain node of the underlying power system is then determined by aggregating the individual forecasts. This posterior aggregation enables the system operator to get a more detailed view on the individual PV units.

Using hourly weather data for the global horizontal irradiance and the ambient temperature as input variables (indicated as Boxes 0.a and 0.b in Fig. 2), this physical model (abbreviated as PM) computes the net power injection for individual PV units. In this context, four modelling components (indicated as Boxes 1 to 4 in Fig. 2) are applied subsequently using weather data, data of PV units provided by the system operator and representative parameters for panels and inverters:

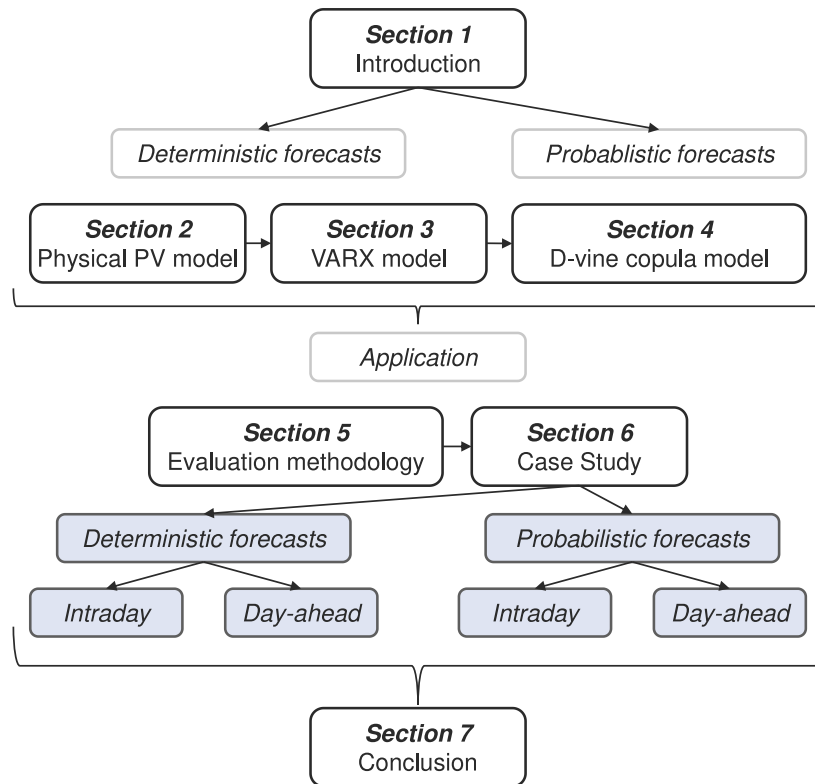


Fig. 1. Flowchart describing the structure of this paper and the content of each section.

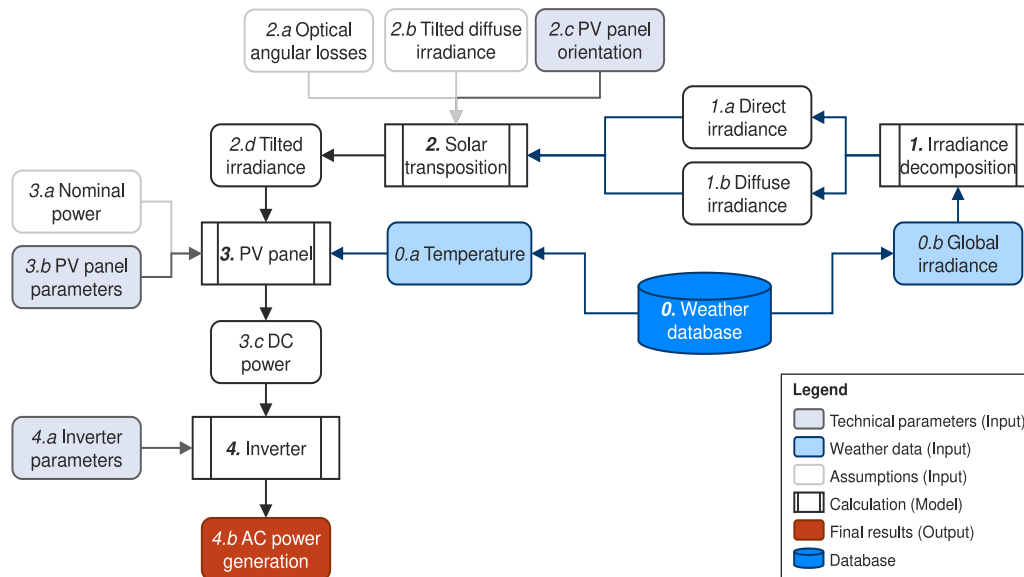


Fig. 2. Outline of the different inputs, model components, parameters and the output of the physical model for photovoltaic power generation.

1. Irradiance decomposition model

The available global irradiance on a horizontal surface is decomposed into a direct and diffuse fraction (cf. Boxes 1.a and 1.b in Fig. 2) by means of a logistic approximation following [5]. Moreover, the parameters of the model are recalibrated for the considered location, hence, increasing forecasting performance, as proposed by Laiti et al. [25].

2. Solar transposition model

The effective irradiance on the tilted surface (referred to as the tilted irradiance, cf. Box 2.d in Fig. 2) of the PV panels is

determined following [26]. In this context, the diffuse irradiance (cf. Box 2.b) on the tilted surface of the PV panels is determined by applying an isotropic approach, as proposed by Kamphuis et al. [27]. Furthermore, the optical angular losses (cf. Box 2.a) on the surface of the PV panels are determined following [28]. In this context, the data of the PV unit's panel orientation, provided by the system operator, are used (cf. Box 2.c), meanwhile typical values of silicon PV panels are assumed with respect to the angular loss coefficient and the fitting parameters of the model, as proposed by Martín and Ruiz [29].

3. PV panel model

The DC power output of the PV panels (cf. Box 3.c in Fig. 2) is obtained by using a simplified single diode solar cell model to determine the maximum power point (MPP) voltage and current, see [30] and [31]. Thereby, a linear temperature dependency of the MPP voltage and current is assumed, based on commonly used voltage and current temperature coefficients as proposed by Marion et al. [32], together with a non-linear temperature model for the saturation current as described by Kou et al. [33]. In terms of the PV panel parameters (cf. Box 3.b), it is assumed that all PV units of the case study use customary monocrystalline silicon PV panels, without considering the individual configuration and the installed components in detail. The nominal power of PV units (cf. Box 3.a) is provided by the system operator.

4. Inverter model

The efficiency of the PV inverter is modelled using a generic quadratic approach for the losses based on [34] and neglecting the voltage dependency to ensure the applicability of the model to all types of inverters, following [35]. Analogous to the previous assumptions, the use of customary components with standard inverter parameters (cf. Box 4.a in Fig. 2) is assumed for the inverter model, without considering the individual configuration and the installed components in detail. Finally, the AC power generation (cf. Box 4.b) describes the feed-in of the PV unit into the power system.

Considering the alignment, orientation and dimensioning of the individual PV units assumptions are made with reference to findings from a comprehensive survey for the application area of the case study introduced below, see [24]. As already stated above, the relevant input parameters for the PV panel and inverter modelling may be estimated using data available in the technical specifications and data sheets of representative poly-crystalline and mono-crystalline modules. More details on the applied PV panel and inverter models are provided in [36].

3. VARX model for benchmarking and statistical post-processing

To validate and possibly improve the performance of the physical model considered in this paper, an adaptive vector autoregressive model with exogenous variables (VARX model) is used. In contrast to previous publications on PV power forecasts using VARX models, e.g. following [37], a statistical post-processing of the previously introduced physical model (PM) is carried out using the forecast of the PM as exogenous variables. In addition, the VARX model allows to benchmark the results of the proposed hybrid model chain and the physical model. Especially in terms of neighbouring PV units, the VARX model seems to be the most promising approach compared to other models, cf. [37]. For more information on the advantages and disadvantages of VARX models, see [6].

In the following, the cumulative net power injection of all PV units connected to a certain network node $k \in \mathcal{K} = \{1, \dots, K\}$ is considered as the relevant forecast variable $\hat{P}_{\text{phys},t,k}$. Thereupon, the VARX model forecasts the net power injection $\hat{P}_{t,k}$ at each network node $k \in \mathcal{K} = \{1, \dots, K\}$ during time step t based on past observations $P_{t-l,i}$ of all other network nodes $i \in \mathcal{K}$ (including themselves). Hereby, \mathcal{L} denotes the set of lags used (any combination of natural numbers greater than zero) and the exogenous variables $X_{m,t,i}$ and $m \in \mathcal{M} = \{1, \dots, M\}$ is the index of the exogenous variables:

$$\frac{\hat{P}_{t,k}}{P_{0,k}} = \alpha_k + \sum_{l \in \mathcal{L}} \sum_{i \in \mathcal{K}} \beta_{l,k,i} \frac{P_{t-l,i}}{P_{0,i}} + \sum_{m \in \mathcal{M}} \sum_{i \in \mathcal{K}} \gamma_{m,k,i} X_{m,t,i}. \quad (1)$$

In Eq. (1), $P_{0,k}$ is the cumulative nominal power of the PV units connected to node k . Following, e.g., [37], the parameters of the VARX model $\alpha_k, \beta_{l,k} = (\beta_{l,k,1}, \dots, \beta_{l,k,K})$ with $l \in \mathcal{L}$ and $\gamma_{m,k} =$

$(\gamma_{m,k,1}, \dots, \gamma_{m,k,K})$ with $m \in \mathcal{M}$ can be determined by K separate ordinary least square (OLS) minimizations of the error between predicted power $\hat{P}_{t,k}$ and observed power $P_{t,k}$ of each network node $k \in \mathcal{K}$ using a moving-horizon of T_h hours preceding to the forecasting time step t :

$$\{\hat{\alpha}_k, \hat{\beta}_{l,k}, \hat{\gamma}_{m,k}\} = \underset{\alpha_k, \beta_{l,k}, \gamma_{m,k}}{\operatorname{argmin}} \sum_{\tau=t-T_h}^{t-1} (P_{\tau,k} - \hat{P}_{\tau,k})^2. \quad (2)$$

Moreover, to determine the optimal specifications of the VARX model the auto-correlation function can be considered in a first step to identify the most promising sets of lags \mathcal{L} , following [37]. Meanwhile in a second step, different variations of those lags can then be combined with different lengths of the moving horizon T_h . In this context, we apply the VARX model in two ways: For post-processing of the forecasts obtained from the physical PV model (PM-VARX) and for benchmarking (VARX-B). For that purpose, different exogenous variables are used for both applications:

1. VARX-B: Set $X_{1,t,k} = \overline{E}_{G,\text{hor},t}^{(k)}$, $X_{2,t,k} = \overline{\vartheta}_{a,t}^{(k)}$ and $M = 2$,
2. PM-VARX: Set $X_{1,t,k} = \hat{P}_{\text{phys},t,k}$ and $M = 1$,

where $\overline{E}_{G,\text{hor},t}^{(k)}$ is the average global horizontal irradiance, $\overline{\vartheta}_{a,t}^{(k)}$ is the ambient temperature and $\hat{P}_{\text{phys},t,k}$ is the deterministic forecast of the physical model at network node $k \in \mathcal{K}$ and time step t . To compare the performance of the deterministic forecasts obtained by the proposed PM-VARX approach, the solely physical model (PM-B) is used for benchmarking as well.

4. Hybrid model using D-vine copulas

Since deterministic predictions usually differ from actual measurements of PV net power injection, the resulting forecasting errors must be taken into consideration. For that purpose, there are well-known tools for modelling uncertainties such as the multivariate normal distribution. However, these tools are facing elementary limitations regarding marginal distributions and tail dependency modelling to address joint extreme values adequately [38]. In contrast, D-vine copulas are well suited to take tail dependencies into account and flexible regarding the number of input variables.

4.1. D-vine copula

A bivariate copula is defined as the joint cumulative distribution function (CDF) $C : [0, 1] \times [0, 1] \rightarrow [0, 1]$ of a 2-dimensional random vector with uniform marginals on the unit interval $[0, 1]$. Based on bivariate copulas, so-called vine copulas or pair-copula architectures can be constructed. Each type of vine copula corresponds to a different kind of decomposition of an n -dimensional density into a product of marginal densities and conditional bivariate copula densities with conditional CDFs as their arguments. An n -dimensional type of vine copula can be applied to estimate the joint multivariate distribution of arbitrary random variables R_1, \dots, R_n for any fixed $n > 2$. Further details regarding general copula theory can be found e.g. in [16,39] and [40].

There exist different types of vine copulas. However, D-vine copulas are chosen in this paper, because they are the only type of vine copulas where each node of their graph-based architecture is connected to exactly two edges, thus giving equal importance to all input variables, see also [16]. The D-vine copula corresponds to the decomposition of an n -dimensional joint density $f_{1,\dots,n}$ of the random variables R_1, \dots, R_n given by

$$f_{1,\dots,n} = \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{i,i+j|i+1,\dots,i+j-1} (F_{i|i+1:i+j-1}, F_{i+j|i+1:i+j-1}) \cdot \prod_{i=1}^n f_i, \quad (3)$$

where f_i denotes the density of the random variable R_i and $c_{i,i+j|i+1,\dots,i+j-1}$ is some bivariate copula density. The bivariate copula

densities $c_{i,i+j|i+1,\dots,i+j-1}$ are applied to the conditional CDFs $F_{i|i+1:i+j-1}$ and $F_{i+j|i+1:i+j-1}$ of R_i and R_{i+j} given the random variables $R_{i+1}, \dots, R_{i+j-1}$. A mathematical derivation of Eq. (3) can be found in [16].

4.2. Multivariate distribution model

First, a univariate distribution has to be fitted separately for each of the random variables R_1, \dots, R_n . For that purpose, histograms are computed providing a visual impression which types of marginal distributions might be most suitable for describing the realizations of R_1, \dots, R_n . Since all candidates of one-dimensional marginal distributions considered in the present paper have exactly two parameters penalizing the number of parameters as done in the Akaike or Bayesian information criterion becomes unnecessary [41]. Thus, the maximum likelihood is computed to assess the goodness of fit and choose the distribution type with the highest maximum likelihood as most suitable distribution type. Then, in a next step, the data is transformed by applying the CDFs F_{R_i} of the fitted marginal distributions.

To fit the D-vine copula, the bivariate conditional copulas $c_{i,i+j|i+1,\dots,i+j-1}$ in the D-vine copula structure are estimated sequentially based on the transformed data [16]. For that purpose, the Archimedean copula types Joe, Frank, Gumbel and Clayton are considered [40]. Each conditional copula density is fitted to the outputs of the CDFs, which are given as arguments in Eq. (3). The copula parameters are estimated for each considered copula type using the inference function for margins method as proposed in [42]. To conclude fitting the multivariate distribution model, the copula type with the highest maximum likelihood is chosen for each bivariate conditional copula.

We apply the multivariate distribution model to generate samples of the random vector (R_1, \dots, R_n) . Using the sampling algorithm for vine copulas explained by Aas et al. [39], n -dimensional samples can be drawn from the fitted D-vine copula. These samples have uniform marginal distributions and describe the interdependence of the fitted D-vine copula. Thus, the inverse of the fitted marginal CDFs $F_{R_i}^{-1}$ has to be applied on the drawn samples, to get corresponding samples of the random vector (R_1, \dots, R_n) .

4.3. Model combination

For any given forecast horizon, each deterministic model (PM-VARX, VARX-B and PM-B as introduced in Sections 2 and 3 generates deterministic forecasts of power injection for all network nodes at the considered time steps t . Now we want to combine these deterministic models with D-vine copulas. For that purpose, we use the following modelling steps:

1. For any given forecast horizon, compute deterministic forecasts by means of the considered deterministic model.
2. Compute and normalize forecasting errors for each network node as described in Eq. (4) below.
3. Fit marginal distributions and a D-vine copula to the forecasting errors at all considered network nodes as described in Section 4.2.
4. Sample the fitted error distribution as discussed in Section 4.2 and reverse the normalization.
5. Generate probabilistic forecasts by adding renormalized samples of error vectors to the deterministic forecasts.

Note that step 2 is performed as follows. For each deterministic model δ the forecasting errors $\epsilon_{t,k}^{(\delta)} = P_{t,k} - \hat{P}_{t,k}^{(\delta)}$ are computed by comparing the forecasts to actual measurements subsequently, these are normalized as follows:

$$\epsilon_{t,k}^{(\delta)} = \frac{\epsilon_{t,k}^{(\delta)} - \frac{1}{|T_{\text{Train},t}|} \sum_{\tau \in T_{\text{Train},t}} \epsilon_{\tau,k}^{(\delta)}}{P_{\text{Ex},t,k}}, \quad (4)$$

where $T_{\text{Train},t}$ contains all time steps of the training dataset (cf. Section 6.1) with the same daily hour as time step t (with $|T_{\text{Train},t}|$ the size of the dataset), while $P_{\text{Ex},t,k} = E_{\text{Ex},\text{hor},t,k} P_{0,k} / E_{\text{Ex},0}$ represents the theoretical extraterrestrial PV power generation at time step t . Here $E_{\text{Ex},\text{hor},t,k}$ is the extraterrestrial irradiance which is computed based on the physical model of Inman et al. [43], $E_{\text{Ex},0}$ is the extraterrestrial solar constant and $P_{0,k}$ is the cumulative nominal power of the PV units connected to node k .

The resulting normalized forecasting errors $\epsilon_{t,k}^{(\delta)}$ at the network node $k \in \mathcal{K}$ are interpreted as realizations of a random variable R_k , see Section 4.2. For each deterministic model and considered forecast horizon, a multivariate error distribution is thus fitted and sampled as explained in Section 4.2. Finally, we renormalize the sampled forecasting errors by reversing the procedure described in Eq. (4) and add the renormalized samples of error vectors to the deterministic forecasts.

We propose the combination of the statistically post-processed physical PM-VARX model with the D-vine copula model as the hybrid model to generate probabilistic forecasts. While PM-VARX is benchmarked to the deterministic models VARX-B and PM-B, the D-vine copula model is compared to a multivariate normal distribution (MVN) fitted to the forecasting errors at all nodes. Moreover, to highlight the importance of modelling spatial dependence, univariate normal distributions (UVN), fitted to the forecasting errors at each node, are sampled separately and the samples are added to the corresponding deterministic forecasts at the corresponding node. The resulting probabilistic prediction model is compared in Section 6 to the probabilistic forecasts based on the D-vine copula and the MVN.

Thus, all combinations of the deterministic models PM-VARX, VARX-B and PM-B and the probabilistic models based on D-vine copula, MVN and UVN are considered, leading to a total of 9 model combinations. In the following these model combinations are referred to as *model chains*. Each model chain is fitted to a training dataset and validated based on a test dataset (cf. Section 6.1).

5. Evaluation methodology

To quantify the performance of the proposed hybrid model in Section 4, a three-stage evaluation is applied: In a first step, the improvements resulting from the pre-processing of meteorological data are considered, whereas in the second step the PM-VARX model is benchmarked to the PM-B and VARX-B models using deterministic performance metrics. Thereupon, the probabilistic forecasts computed by the D-vine copula model are benchmarked to the probabilistic forecasts based on MVN and UVN in a third step, see Section 4.3.

5.1. Deterministic performance metrics

To evaluate the quality of the generated deterministic forecasts for a certain network node $k \in \mathcal{K}$, the normalized root mean square error (NRMSE) can be used as performance metric, following e.g. [9]:

$$NRMSE_k = \sqrt{\frac{1}{|T_{\text{Test}}|} \sum_{t \in T_{\text{Test}}} \left(\frac{P_{t,k} - \hat{P}_{t,k}}{P_{0,k}} \right)^2}, \quad (5)$$

where T_{Test} is the test dataset (cf. Section 6.1). Furthermore, the quantity $NRMSE_k$ given in Eq. (5) can be used to determine the forecast skill S_k , following [44], which benchmarks the quality of the forecast of the considered PV model to a naive persistence method. Hereinafter, the benchmarking persistence model relies on the assumption that the net power injection of the PV units at node $k \in \mathcal{K}$ is constant relative to the extraterrestrial irradiance $E_{\text{Ex},\text{hor},t,k}$. Thus, the benchmarking persistence forecasts are computed using $\hat{P}_{t,k} = E_{\text{Ex},\text{hor},t,k} (P_{t-1,k} / E_{\text{Ex},\text{hor},t-1,k})$. The forecast skill is then defined as

$$S_k = 1 - \frac{NRMSE_k}{NRMSE_{\text{pers},k}}, \quad (6)$$

with

$$NRMSE_{pers,k} = \sqrt{\frac{1}{|T_{Test}|} \sum_{t \in T_{Test}} \left(\frac{P_{t,k}}{P_{0,k}} - \frac{P_{t-1,k}}{P_{0,k}} \frac{E_{Ex,hor,t,k}}{E_{Ex,hor,t-1,k}} \right)^2}. \quad (7)$$

The forecast skill S_k given in Eq. (6) does not measure the quality of the physical model exclusively, but rather the performance of the weather forecast together with the physical model. However, when benchmarking different PV models, the forecast skill S_k indicates the improvement of the considered model compared to the persistence method. Therefore, the best deterministic model achieves the highest forecast skill S_k which is corresponding to the lowest $NRMSE$.

5.2. Probabilistic performance metrics

In this section, the performance metrics for evaluating the probabilistic forecasts generated by different model chains, see Section 4.3, are introduced. Based on the deterministic forecasts $\hat{P}_t^{(\xi)}$ of model chain ξ a sample of size m of the forecasting errors $\epsilon_t^{(\xi)}$ is drawn from the corresponding probabilistic model (cf. Section 4.3). Thereupon, the probabilistic forecast is defined as $\tilde{P}_t^{(\xi)} = \hat{P}_t^{(\xi)} + \epsilon_t^{(\xi)}$, constituting a $|\mathcal{K}| \times m$ dimensional matrix.

A common performance metric for multivariate probabilistic forecasts is the *energy score* introduced by Gneiting and Raftery [45]. In this context, the energy score (ES) during time step t can be written as

$$ES_t = \frac{1}{m} \sum_{j=1}^m \sqrt{\sum_{k \in \mathcal{K}} (p_{t,k} - \tilde{p}_{t,k,j})^2} - \frac{1}{2m^2} \sum_{i=1}^m \sum_{j=1}^m \sqrt{\sum_{k \in \mathcal{K}} (\tilde{p}_{t,k,i} - \tilde{p}_{t,k,j})^2}, \quad (8)$$

where $p_{t,k}$ is the observed power supply at network node $k \in \mathcal{K}$ and $\tilde{p}_{t,k,j}$ is the j th of m realizations of the predicted power supply.

However, Pinson and Tastu [46] show that, while the energy score is useful to detect differences between the observed and predicted power supply, its discriminative ability in terms of variance and correlation is poor. Therefore, we follow [47] using a *variogram-based score* as an additional scoring rule, which is reported to have a higher discriminative power regarding the variance and the interdependence structure of multivariate distributions [47,48]. In case of a K -variate probabilistic forecast represented by a sample of size m , the variogram-based score (VS) is defined as

$$VS_t = \sum_{i \in \mathcal{K}} \sum_{k \in \mathcal{K}} w_{i,j} \left(|p_{t,i} - p_{t,k}|^{0.5} - \frac{1}{m} \sum_{j=1}^m |\tilde{p}_{t,i,j} - \tilde{p}_{t,k,j}|^{0.5} \right)^2, \quad (9)$$

where $w_{i,j}$ are arbitrarily selectable weighting factors for all $i, j \in \mathcal{K}$, e.g. representing the historical correlation of the spatio-temporal forecasting errors or power injections at different locations, see [48]. To evaluate our results given in Section 6 below, we usually refer to the average energy and variogram-based scores over the full validation period, i.e., $ES = \frac{1}{|T_{Test}|} \sum_{t \in T_{Test}} ES_t$ and $VS = \frac{1}{|T_{Test}|} \sum_{t \in T_{Test}} VS_t$.

5.3. Statistical significance

To investigate if the observed differences of the results obtained by the deterministic and probabilistic models are statistically significant, the *Diebold–Mariano test* can be used in two ways, following [49], and [50]. Note that the notation changes when applying the test to deterministic or probabilistic forecasts. Thus, the test will be used for both cases separately.

In terms of *deterministic forecasts*, for any network node $k \in \mathcal{K}$ and model δ , we interpret the forecasting errors $\epsilon_{1,k}^{(\delta)}, \epsilon_{2,k}^{(\delta)}, \dots$ introduced in Section 4.3 as realizations of a certain random variable $\epsilon_k^{(\delta)}$. Then, this test checks the null hypothesis that the difference of the expected forecasting errors of two models, say δ_i and δ_j for $i \neq j$, is equal to zero for a certain network node $k \in \mathcal{K}$ (and a certain forecast horizon h), i.e.,

$$\mathbb{E} \left(g \left(\epsilon_k^{(\delta_i)} \right) - g \left(\epsilon_k^{(\delta_j)} \right) \right) = 0, \quad (10)$$

for some function $g : \mathbb{R} \rightarrow \mathbb{R}$. Possible choices for g are $g(\epsilon) = |\epsilon|$ or $g(\epsilon) = \epsilon^2$, hence, representing either the mean absolute error or the mean squared error. Hereinafter, we only consider the mean squared error. The null hypothesis of the Diebold–Mariano test can be evaluated using the test statistic $t_{DM,k}$, given by

$$t_{DM,k} = \frac{\bar{d}_k \sqrt{|T_{Test}|}}{\sqrt{v_k}}, \quad (11)$$

where $|T_{Test}|$ is the number of forecasts in the test dataset and \bar{d}_k is the sample mean of the differences $d_{t,k} = g(\epsilon_{t,k}^{(\delta_i)}) - g(\epsilon_{t,k}^{(\delta_j)})$. Furthermore, the normalizing factor v_k in Eq. (11) is given by

$$v_k = \varsigma_0 + 2 \sum_{i=1}^{h-1} \frac{(|T_{Test}| - i)}{|T_{Test}|} \varsigma_i, \quad (12)$$

where ς_i is the i th sample autocovariance of $d_{1,k}, d_{2,k}, \dots$ and h is the forecast horizon in hours, see [51]. Under the null hypothesis, the test statistic $t_{DM,k}$ is asymptotically standard normal distributed, i.e., $t_{DM,k} \sim \mathcal{N}(0, 1)$.

In terms of *probabilistic forecasts*, for two model chains ξ_i and ξ_j with $i \neq j$, the Diebold–Mariano test can be used to compare the differences of two probabilistic performance metrics, see [50]. Hence, the time series of two scores $S(\tilde{P}_t^{(\xi_i)}, p_t)$ and $S(\tilde{P}_t^{(\xi_j)}, p_t)$ are used, which can be either the energy score (ES) or the variogram-based score (VS) introduced in Section 5.2, with p_t the actual realization of the net power injections at all network nodes from \mathcal{K} . Therefore, the null hypothesis can be tested that the expected difference of both scores is equal to zero, i.e., on average, both model chains ξ_i and ξ_j have the same forecast performance. Based on the sample mean $\bar{S}^{(\xi)} = \frac{1}{|T_{Test}|} \sum_{t \in T_{Test}} S(\tilde{P}_t^{(\xi)}, p_t)$, the test statistic can be accessed as

$$t_{DM}^{(\xi_i, \xi_j)} = \sqrt{|T_{Test}|} \frac{\bar{S}^{(\xi_i)} - \bar{S}^{(\xi_j)}}{\hat{\sigma}^{(\xi_i, \xi_j)}}, \quad (13)$$

with

$$\hat{\sigma}^{(\xi_i, \xi_j)} = \sqrt{\frac{1}{|T_{Test}|} \sum_{t \in T_{Test}} \left(S(\tilde{P}_t^{(\xi_i)}, p_t) - S(\tilde{P}_t^{(\xi_j)}, p_t) \right)^2}. \quad (14)$$

Referring to Gneiting and Katzfuss [50], the test statistic $t_{DM}^{(\xi_i, \xi_j)}$ given in Eq. (13) is asymptotically standard normal distributed under the null hypothesis mentioned above, i.e., $t_{DM}^{(\xi_i, \xi_j)} \sim \mathcal{N}(0, 1)$. When rejecting the null hypothesis, the model chain ξ_j is preferred if $t_{DM}^{(\xi_i, \xi_j)}$ is negative and ξ_i is preferred if $t_{DM}^{(\xi_i, \xi_j)}$ is positive.

6. Application

To demonstrate the advantages of the proposed hybrid forecasting model, see Section 4, a case study based on data observed for an existing power system is introduced in Section 6.1. The hybrid forecasting model is fitted and some model characteristics are discussed in Section 6.2. In Section 6.3, short-term and long-term forecasts of PV power supply are evaluated, based on all model chains considered in this paper.

6.1. Case study

In our case study we consider the high voltage power system of N-ERGIE Netz GmbH in the south of Germany and select 53 PV units which are connected to five nodes of the considered power system, see also [13]. Furthermore, the underlying distribution network (i.e. the interconnections between the considered nodes and the individual PV units) is neglected to focus on the probabilistic forecasts exclusively. In this context, Fig. 3 visualizes the spatial distribution of the considered PV units, their installed capacity and the corresponding network node. The nominal power of the individual PV units (i.e., the installed capacity) varies between 250 kW and 5.4 MW and the maximum distance between the two outermost PV units is less than 50 km.

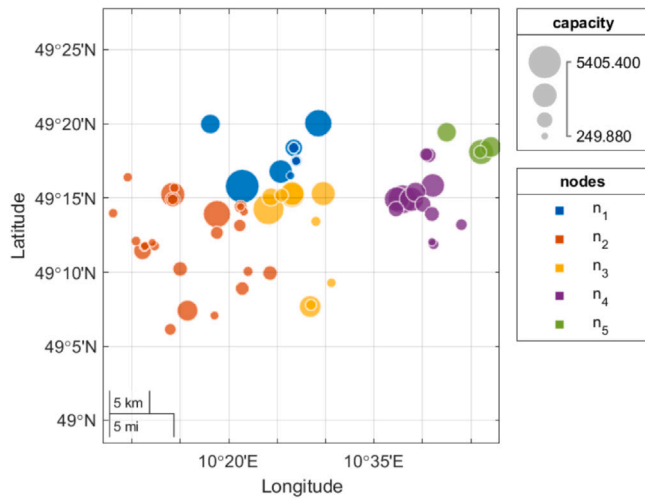


Fig. 3. Locations of PV units considered in this case study, where the size of the circles corresponds to the nominal power (in kW) and the colour indicates the associated network node.

For deterministic forecasting of PV power supply, statistical forecasts of global horizontal irradiance and ambient temperature (2 m above ground) of Ensemble-MOS of DWD are used, see also [20,21] and [22,23]. The weather forecasts are used in a spatial resolution of 20×20 km, with a temporal resolution of one hour and a forecast horizon of 19 h. In this context, the parameters of the irradiance decomposition model (cf. Section 2) have been recalibrated for the considered location, whereby the forecasting performance has been improved significantly. Subsequently, direct and diffuse irradiance at the location of each PV unit is calculated for each hour via inverse distance weighting of the four nearest available grid points of the weather dataset. The same interpolation procedure is applied to the temperature, used as an input variable of PV panel efficiency in the following section. The forecasts of irradiance and temperature are updated every 3 h. The updates improve the forecast quality for a given point in time in general, whereas the improvement is most significant for very short-term forecasts (i.e., of the next hour) but rather negligible at the end of the forecast horizon. Here, we focus on the latest available forecast for each hourly time step (i.e., with a forecast horizon of 1–3 h in advance) hereinafter referenced as the *intraday forecast* and on the earliest available forecast (i.e., with a forecast horizon of 17–19 h in advance) subsequently referenced as the *day-ahead forecast*. Furthermore, since focusing on PV generation and applications to power systems, all models are applied for the hours between 9 a.m. and 18 p.m. exclusively, hence, neglecting the morning and evening hours while focusing on those hours of the day with the highest impact on the power system.

For empirical validation, measurement data of the net power supply for May to July is available for all PV units considered in this paper and the years 2015, 2016 and 2017. In this context, we split up the available data into two sets: The *training dataset*, denoted by T_{Train} , consists of N_{Train} hourly data between May and July for the years 2015 and 2016, and the *test dataset*, denoted by T_{Test} , consisting of $|T_{\text{Test}}|$ hourly data between May and July of the year 2017. This hourly data contains measurement time series for individual PV units as well as forecasting time series of the weather data obtained from ModelMix of DWD. As these time series of weather forecasts and observed PV power injections are in some cases subject to missing data, noise or minor calibration errors, the following filtering steps are conducted:

1. In case of detecting unreasonable offsets, the time series is recalibrated by superposition with a time invariant constant (e.g. indicated by a certain power in-feed during night times).

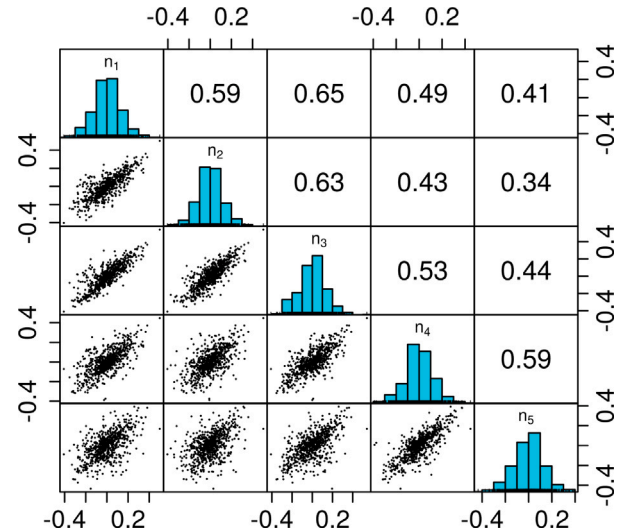


Fig. 4. Histograms (diagonal), pairwise scatter plots (bottom left) and pairwise Kendall's rank correlation coefficients (top right) of the forecasting errors (for intraday forecasts of the PM-B approach) in May, June and July of year 2016 evaluated for the five network nodes n_1, \dots, n_5 .

2. All hours of days with observations of zero power supply between 9 and 18 pm (UTC+2) are treated as missing data.
3. For some hours, data gaps occur in either the weather dataset or in the measurement data of the considered PV units. Days, with such data gaps occurring in at least one hour, are excluded entirely (in both training and evaluation). Removed days are 2015-05-01, 2015-07-31, 2016-05-01, 2016-06-20, 2016-07-31, 2017-05-01, 2017-06-01, 2017-07-17 and 2017-07-31.

In this context, the PM-B model has been calibrated on the training dataset once and applied subsequently to the test dataset. Furthermore, the PV power injections have been aggregated for each network node and subsequently used as an input variable regarding the post-processing for the PM-VARX model. In terms of the VARX-based models, the best results were obtained using lags of 1 and 2 h for intraday and day-ahead forecasts, representing the highest autocorrelations of the considered time series in the test dataset. In addition, the VARX-based models are taking a moving window of the last four weeks into account. For that reason and furthermore to eliminate the effect of seasonal or annual weather phenomena, the probabilistic models (D-vine copula, MVN and UVN) are fitted on a subset of the training dataset, hence, taking the time series of 2016 into account exclusively. Since the models are calibrated and trained using (subsets of) the training dataset, the performance metrics and their statistical significance are evaluated typically for the test dataset.

6.2. Application of the D-vine copula model

Now we apply the D-vine copula model described in Section 4 to the case study presented in Section 6.1.

We consider the five network nodes n_1, \dots, n_5 , see Fig. 3. For the intraday forecasts of the PM-B approach, normalized forecasting errors in May, June and July of year 2016 are visualized using histograms, pairwise scatterplots and pairwise Kendall's rank correlation coefficients in Fig. 4. Based on visual inspection of the histograms, the forecasting errors at the network nodes n_1, \dots, n_5 correspond to unimodal distribution types such as the normal, Weibull, gamma and logistic distribution. We selected these four distribution types as candidates for the marginal distributions. Moreover, the pairwise scatterplots provide detailed information about the interdependence of any pairs of forecasting errors. Based on the scatterplots we conclude that the

Table 1

Test statistic of the Diebold–Mariano test computed for deterministic intraday and day-ahead forecasts comparing all three deterministic models. Confidence levels: No symbol = H_0 not rejected, (+) = 90%, (*) = 95%, (**) = 99%, (***) = 99.9%.

Network node	PM-VARX vs. VARX-B		VARX-B vs. PM-B	
	Intraday	Day-ahead	Intraday	Day-ahead
n_1	-2.679***	-3.848***	0.476	-0.343
n_2	-4.586***	-3.551***	-2.189*	0.201
n_3	-4.446***	-2.20*	-2.210*	-1.479
n_4	-2.273*	-3.101***	-4.378***	-3.128***
n_5	-4.399***	-4.034***	-1.218	-0.645

errors at two network nodes gather near a line through the origin. To quantify the interdependence of two random variables X and Y , we apply Kendall's rank correlation coefficient defined by

$$\hat{\tau} = \frac{2}{n(n-1)} \sum_{i < j} \text{sgn}(x_i - x_j) \text{sgn}(y_i - y_j), \quad (15)$$

for given realizations (x_1, \dots, x_n) and (y_1, \dots, y_n) of the random variables X and Y . The pairwise Kendall's rank correlation coefficients in Fig. 4 show that the forecasting errors are strongly correlated and a multivariate distribution model such as the D-vine copula model is suitable. Furthermore, by comparing Figs. 3 and 4 we notice that the forecasting errors of pairs of nodes are stronger correlated if the distance between the nodes is smaller. In Step 3 of Section 4.3 we fit marginal distributions to the normalized forecasting errors for each considered network node, deterministic model and forecast horizon. Comparing the loglikelihood function of the fitted marginal distributions it seems that the logistic distribution is slightly better than the normal distribution, whereas the Weibull and gamma distribution are clearly worse.

6.3. Forecasting results

At first, the improvements of the pre-processing for the weather data are evaluated, whereby additionally the BRL model parameters (see Section PhysicalModelSec) are recalibrated. Secondly, the deterministic forecasts of the PM-VARX approach are benchmarked to the PM-B and VARX-B approaches, which are introduced in Section 3, for different forecast horizons (intra-day and day-ahead). Then, the marginal and spatial distributions of the forecasting errors resulting from different deterministic models are tested for normality and representativity. Finally, the probabilistic forecasting is carried out and evaluated for the different forecast horizons (intra-day and day-ahead).

6.3.1. Deterministic intraday forecasts

For each forecasting model with a forecast horizon of 1 to 3 h, the NRMSE and forecast skill S , as introduced in Section 5.1, are determined at the five considered network nodes. The results depicted in Fig. 5 show the results for the proposed PM-VARX model as well as for both benchmark models PM-B and VARX-B (cf. Section 3).

As expected, all deterministic models show substantially lower forecast errors than the persistence benchmark indicated by a consistently positive forecast skill. Hence, the NRMSE is ranging from 5.7% to 6.8% for PM-VARX, from 5.9% to 6.9% for VARX-B and from 6.2% to 7.5% for PM-B. This leads to average forecast skills of 57.9% for the PM-VARX model, 56.5% for the VARX-B and 54.1% for the PM-B. In this context, the results are clearly indicating the proposed PM-VARX approach to outperform both benchmarks (VARX-B and PM-B). Furthermore, the VARX-B model is preferable in comparison with the PM-B for most of the network nodes, due to the lower NRMSE values at nodes n_2 to n_5 , and provides an overall better forecasting performance than the PM-B approach.

To test for significance of the differences in the NRMSE the Diebold–Mariano test, introduced in Section 5.3, is carried out for the intraday

and day-ahead forecast horizons. The corresponding results are depicted in Table 1. To establish consistency to following tables, we introduce a notation which indicates the confidence level on which a null hypothesis H_0 of equally performing models is rejected: No asterisk means that the null hypothesis is not rejected, plus sign (+) confidence = 90%, one-asterisk (*) confidence = 95%, two-asterisks (**) confidence = 99% and three-asterisks (***) confidence = 99.9%.

When comparing the proposed PM-VARX approach to the VARX-B the test statistics indicate that the null hypothesis of equally performing models can be rejected for most of the network nodes, so the advantages of the VARX (post-processing) are – albeit not very large – highly significant. Comparing the VARX-B and PM-B approaches, the performance differences are in favour of the VARX-B approach but not as clear as above: The significance levels for the intraday-forecasts are reaching confidence above 95% only for three network nodes, meanwhile for day-ahead forecast the test indicates equal forecast performance for almost all network nodes, see also Section 6.3.2.

6.3.2. Deterministic day-ahead forecasts

The respective results for NRMSE and forecast skill for a forecast horizon of 17 to 19 h are depicted in Fig. 6. In this context, the NRMSEs for day-ahead forecasts at the network node level increase to a range from 6.7% to 8.9%.

When comparing the forecasting performance of the proposed model to the PV forecasts of whole regions, see e.g. see [24] for Germany, there are two outcomes: On the one hand, the day-ahead forecasting performance on a regional level seems to exceed the performance when forecasting PV production on a network node level. On the other hand, since the variance of the forecast decreases with an increasing number of PV units, the observed forecasting performance on a regional level is not a good benchmark regarding forecasts on a network node level. As the performance of the persistence model deteriorates considerably compared to the intraday forecasts, the forecasting skill rises to a range between 57% and 81%. Again, the proposed PM-VARX approach outperforms both benchmarks (VARX-B and PM-B). However, although the differences in NRMSE are small, the performance of PM-VARX is still significantly better compared to the other models according to the test statistic of the Diebold–Mariano test given in Table 1. Unlike to the intraday forecasts, the performance differences between VARX (benchmark) and the physical model are not statistically significant, except for the network node n_4 .

6.3.3. Probabilistic intraday forecasts

When considering the previously discussed results, the question arises how sensitive the probabilistic forecasts are with respect to the performance of the deterministic models and which benefits can be achieved by using D-vine copula models instead of well-established approaches such as MVN or UVN. In this context, and as already mentioned above, the probabilistic forecast for a certain model chain ξ is carried out by applying Monte Carlo simulation, i.e., for each network node $k \in \mathcal{K}$ and each time step t , a sample of size m is drawn for the model chain (with $m=3,000$). The result is then the $|\mathcal{K}| \times m$ -dimensional matrix $\tilde{\mathbf{P}}_t^{(\xi)} = \hat{\mathbf{P}}_t^{(\xi)} + \epsilon_t^{(\xi)}$, to which the energy score (ES) as well as the variogram-based score (VS) described in Section 5.2 are applied, in order to compare the probabilistic forecast $\tilde{\mathbf{P}}_t^{(\xi)}$ to the $|\mathcal{K}|$ -dimensional vector of the actually measured PV power supplies \mathbf{P}_t . Furthermore, VS can be determined without weighting (VS1: $w_{ij} = 1$ for all $i, j \in \mathcal{K}$) and with weighting, e.g., according to the correlation ρ_{ij}^{Train} of the power injection for each pair $i, j \in \mathcal{K}$ of network nodes in the test dataset (VS2: $w_{ij} = \rho_{ij}^{\text{Train}}$ for all $i, j \in \mathcal{K}$). The results of the evaluation are depicted in Table 11 for different model chains, i.e. combinations of the deterministic and probabilistic models.

The results of energy and variogram-based scores indicate the proposed hybrid model chain, incorporating a PM-VARX approach together with a D-vine copula model (with individually fitted marginal distributions and pre-processing of the forecasting errors) to outperform

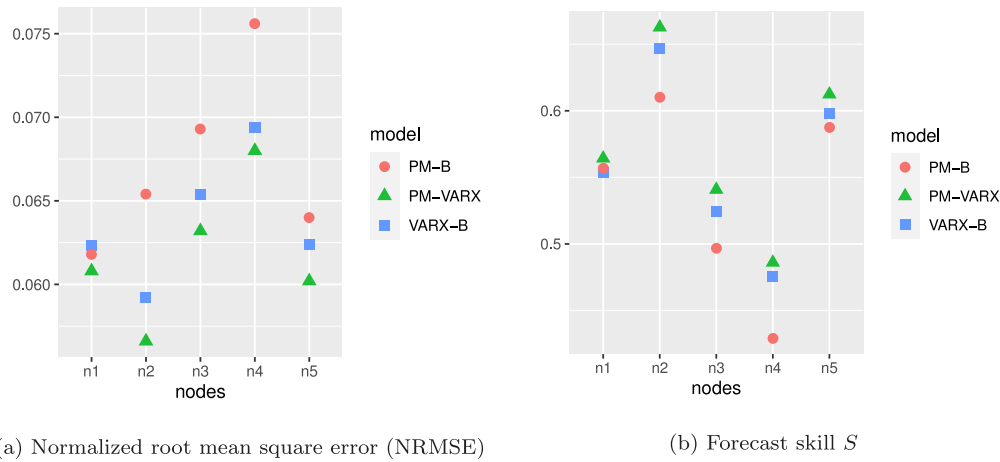


Fig. 5. Comparison of validation scores for 3 models at the five considered network nodes for **intraday forecasts**. Note that a better model performance is indicated in (a) by a lower score and in (b) by a higher score.

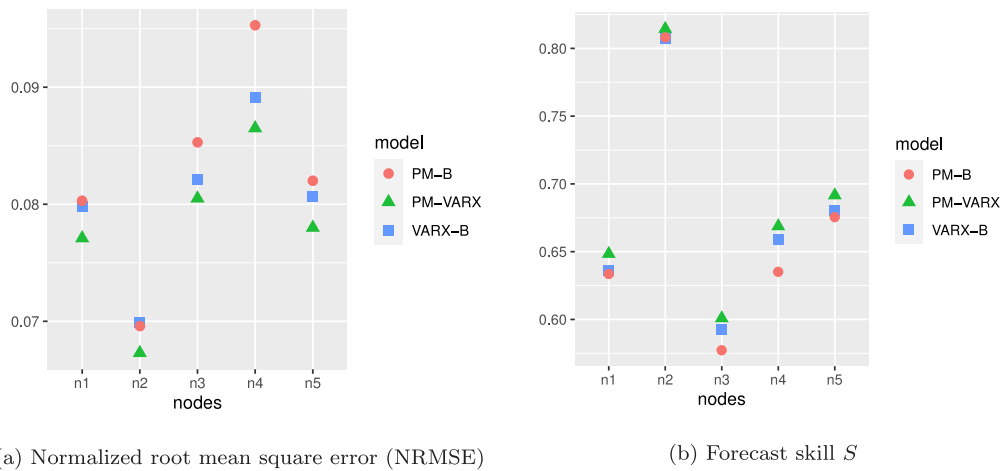


Fig. 6. Comparison of validation scores for 3 models at the five considered network nodes for **day-ahead forecasts**. Note that a better model performance is indicated in (a) by a lower score and in (b) by a higher score.

Table 2

Energy score (ES) and variogram-based scores (VS1 and VS2) of the intraday forecasts for different model chains. Better model performance is indicated by lower scores.

Deterministic model	Probabilistic model	ES	VS1	VS2
PM-VARX	D-vine	2.04	3.06	2.76
PM-VARX	MVN	2.09	3.11	2.80
PM-VARX	UVN	2.15	3.56	3.22
VARX-B	D-vine	2.10	3.08	2.78
VARX-B	MVN	2.18	3.12	2.81
VARX-B	UVN	2.23	3.62	3.28
PM-B	D-vine	2.22	3.59	3.25
PM-B	MVN	2.38	3.86	3.49
PM-B	UVN	2.39	4.01	3.62

all other model chains. Not only the PM-VARX approach surpasses the VARX-B and PM-B benchmarks when comparing the same probabilistic models, but also the D-vine copula surpasses the MVN and UVN when comparing the same deterministic models. We notice that the differences in the scores seem to be rather small (especially when comparing the D-vine copula and MVN), which rises the question whether the differences in the forecast performance of competing model chains are statistically significant. However, the outstanding forecasting performance of the PM-VARX approach has already been confirmed to be statistically significant for intraday and day-ahead forecasts (cf. Sections Section 6.3.1 and 6.3.2). Thus, the Diebold–Mariano test is carried out

Table 3

Test statistic of the Diebold–Mariano test for pairwise comparison of the forecast performance based on ES, VS1 and VS2 of the intraday forecasts. Confidence levels: No symbol = H_0 not rejected, (+) = 90%, (*) = 95%, (**) = 99%, (***) = 99.9%.

Deterministic model	D-vine vs. MVN			MVN vs. UVN		
	ES	VS1	VS2	ES	VS1	VS2
PM-VARX	−8.67***	−3.96***	−3.98***	−0.17	−6.75***	−6.8***
VARX-B	−10.62***	−4.42***	−4.45***	0.14	−6.24***	−6.29***
PM-B	−13.96***	−8.63***	−8.79***	0.99	−0.04	−0.02

for a pairwise comparison of the different probabilistic models only, see Table 3.

When considering the results, the highly negative values of the test statistic based on the ES clearly indicate the D-vine copula model to outperform the MVN and UVN approach for all deterministic models with confidence level above 99.9%. Only VS1 and VS2 do not confirm these results when comparing MVN and the D-vine copula model, but indicate a confidence level of 90% for intraday forecasts incorporating the PM-VARX approach instead. As a consequence, even though the differences of the scores depicted in Table 2 are small, a statistical significance can be observed for most of the model chains in pairwise comparisons.

Therefore, two results can be concluded for intraday forecasts: On the one hand, the VARX-based models seem to improve the spatial

Table 4

Energy score (ES) and variogram-based scores (VS1 and VS2) of the day-ahead forecasts for different model chains. Better model performance is indicated by lower scores.

Deterministic model	Probabilistic model	ES	VS1	VS2
PM-VARX	D-vine	2.53	3.93	3.54
PM-VARX	MVN	2.66	4.06	3.66
PM-VARX	UVN	2.66	4.60	4.17
VARX-B	D-vine	2.64	4.01	3.62
VARX-B	MVN	2.82	4.15	3.75
VARX-B	UVN	2.81	4.68	4.24
PM-B	D-vine	2.70	4.23	3.82
PM-B	MVN	2.92	4.71	4.25
PM-B	UVN	2.91	4.71	4.26

Table 5

Test statistic of the Diebold–Mariano test for pairwise comparison of the forecast performance based on ES, VS1 and VS2 of the day-ahead forecasts. Confidence levels: No symbol = H_0 not rejected, (+) = 90%, (*) = 95%, (**) = 99%, (***) = 99.9%.

Deterministic model	D-vine vs. MVN			MVN vs. UVN		
	ES	VS1	VS2	ES	VS1	VS2
PM-VARX	−5.37***	−1.73+	−1.74+	−6.2***	−7.61***	−7.66***
VARX-B	−7.48***	−1.61	−1.62	−5.89***	−8.25***	−8.29***
PM-B	−13.16***	−6.92***	−7.02***	−0.67	−2.02*	−1.99*

modelling, hence, taking spatial dependencies already into account during the deterministic forecasting process. On the other hand, the forecasting errors can be approximated appropriately using a MVN in a first step. Nevertheless, the D-vine copula model is capable to improve the probabilistic intraday forecasts, especially for physical models.

In addition, the results indicate a higher sensitivity of the probabilistic forecasts to the deterministic models than to the probabilistic models, i.e., the PM-VARX and VARX-B with UVNs outperform or perform comparable to the PM-B approach with a D-vine copula. Hence, when generating intraday probabilistic forecasts, the initial focus should be laid on an appropriate deterministic model as a basis for the probabilistic modelling (with marginal distributions and D-vine copulas) to build on.

6.3.4. Probabilistic day-ahead forecasts

For the day-ahead forecasts, i.e., for a forecast horizon of 17 to 19 h, the energy score (ES) and the variogram-based scores (VS1 and VS2) are depicted in Table 4 for different model chains.

In this context, the findings for the intraday forecasts can be confirmed partly: The results of energy and variogram-based scores indicate the proposed hybrid model chain, incorporating a PM-VARX approach together with a D-vine copula model, to outperform all other model chains. However, in contrast to the intraday forecasts the differences in the ES between MVN and UVN are vanishing for all three deterministic models. These results are confirmed by VS1 and VS2 for the PM-B approach only, meanwhile the variogram-based scores indicate unequal forecasting performance for the VARX-based approaches. Even though the differences in the scores for the day-ahead forecasts are higher than for intraday forecasts, the question on the statistical significance in the forecast performance of competing model chains remains. Therefore, the Diebold–Mariano test is carried out again for a pairwise comparison of the different probabilistic models, see Table 5.

In contrast to the intraday forecasts, the results of the VARX-based models for all three scoring rules are in accordance with each other, namely, the D-vine copula outperforms the MVN and UVN benchmarks. Hence, when comparing the D-vine copula model to the MVN the highly negative test statistics clearly indicate the D-vine copula model to provide the better forecast performance with a confidence level above 99.9%. Furthermore, even though there are some positive differences when comparing MVN and UVN for the deterministic benchmark models (i.e., the UVN surpasses the MVN approach in terms of the PM-B),

the D-vine copula surpasses both approaches clearly since these differences are statistically not significant. Finally, three main findings of the probabilistic intraday forecasts may be retained for the day-ahead forecasts:

1. The VARX-based models enable to take into account spatial dependencies during deterministic modelling.
2. For day-ahead forecasts, the multivariate spatial dependency modelling using D-vine copulas outperforms the MVN and UVN benchmarks. Even though a more detailed modelling of the marginal distributions (e.g. using kernel density estimators) may improve the forecasting performance further, the spatial dependency modelling can be assumed to be more relevant for the performance of day-ahead forecasts.
3. Furthermore, the findings suggest that the probabilistic forecasts are more sensitive to the deterministic models than to the probabilistic models.

7. Conclusion

Probabilistic photovoltaic forecasts are mostly based either on completely statistical approaches or neglect the underlying spatial structure, e.g., of electricity networks. This paper goes further by introducing a hybrid approach for modelling spatially correlated probabilistic photovoltaic forecasts and carrying out a comprehensive analysis using well-established models as benchmarks. In opposite to previous research in this field a comprehensive hybrid model chain with several improvements for intra-day and day-ahead deterministic and probabilistic forecasts has been proposed. In this context, several pre- and post-processing steps have been combined and benchmarked using a case study for a certain location in the south of Germany. Therefore, high-resolution numerical weather predictions and real-world measurement data over several years have been used to demonstrate the advantages of the proposed hybrid approach. The pre- and post-processing steps include especially an improved irradiance decomposition using the Boland–Ridley–Laurent model with individual calibration of the model parameters, an additional post-processing using a vector autoregressive model with exogenous variables, i.e., a VARX model, and introducing a novel normalization of forecasting errors. Furthermore parametric univariate marginal distributions and a vine copula have been applied to cope with the remaining uncertainty of the probabilistic forecasts. The results of the case study have been evaluated by applying diverse deterministic and probabilistic performance measures while considering statistical significance of the outcome as well.

As one of the key findings of this case study, the proposed hybrid approach achieves the best results compared to different alternative model chains, i.e., combinations of either a solely physical model or a solely VARX model together with either a vine copula model, a multivariate normal distribution and, in particular, a product of univariate normal distributions. Another key finding of this work is that although VARX models already account for spatial dependencies, there is still a statistically significant interdependence in the forecasting errors, justifying the use of vine copulas. Furthermore, one of the key findings in this case study appears to be that probabilistic photovoltaic forecasts in general are more sensitive to the deterministic models than to the probabilistic models for both, intraday and day-ahead forecasts, respectively.

Therefore, the proposed hybrid approach offers advantages for all applications where high accuracy of the results and computational performance of the models are crucial. For instance, in terms of online predictive grid management systems most probabilistic forecasts are based on statistical photovoltaic models, coming along with rather high computational effort, and facing some elementary drawbacks compared to physical models. Moreover, these approaches are not able to estimate exceptional situations and may be rather weak in out-of-sample

applications, e.g., due to modifications (i.e., expansion or removal of defective panels).

Further research could take the time-coupling of spatially interdependent photovoltaic forecasting errors into account or extend the proposed vine copula model to other sources of uncertainty, e.g., wind or hydro power infeed. As part of the future work, the authors are currently investigating spatio-temporal probabilistic photovoltaic forecasts based on artificial neural networks. Furthermore, the key findings of this work can be used to investigate probabilistic forecasts of network congestions and form the basis for optimization approaches to reduce renewable curtailment.

CRedit authorship contribution statement

A. Schinke-Nendza: Conceptualization, Methodology, Data curation, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **F. von Loeper:** Software, Validation, Investigation, Writing – original draft, Writing – review & editing, Visualization. **P. Osinski:** Data Curation, Software, Validation, Formal analysis, Investigation, Writing – original draft. **P. Schaumann:** Software, Validation. **V. Schmidt:** Conceptualization, Writing – review & editing, Supervision, Project administration. **C. Weber:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

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

Acknowledgements

This work has been carried out as part of the research project ‘Verteilnetze’ which is funded by the Federal Ministry for Education and Research under grant agreement numbers 05M18PGA and 05M18VUB. In addition, we would like to thank our project partners Reinhold Hess of DWD and Rainer Bäsman of N-Energie GmbH for their support by providing the meteorological forecasting data and power-supply measurement data of our case study.

References

- [1] BNetzA. Marktstammdatenregister. 2020, <http://www.marktstammdatenregister.de/MaStR>, Accessed: 2020-04-28.
- [2] Bessa R, Möhrlein C, Fundel V, Siefert M, Browell J, Haglund El Gaidi S. Towards improved understanding of the applicability of uncertainty forecasts in the electric power industry. *Energies* 2017;10:1402.
- [3] Schermeyer H, Vergara C, Fichtner W. Renewable energy curtailment: A case study on today's and tomorrow's congestion management. *Energy Policy* 2018;112:427–36.
- [4] Lorenz E, Scheidsteiger T, Hurka J, Heinemann D, Kurz C. Regional PV power prediction for improved grid integration. *Prog Photovolt, Res Appl* 2011;19(7):757–71.
- [5] Ridley B, Boland J, Lauret P. Modelling of diffuse solar fraction with multiple predictors. *Renew Energy* 2010;35:478–83.
- [6] Antonanzas J, Osorio N, Escobar R, Urraca R, Martinez-de Pison F, Antonanzas-Torres F. Review of photovoltaic power forecasting. *Sol Energy* 2016;136:78–111.
- [7] Gigoni L, Betti A, Crisostomi E, Franco A, Tucci M, Bizzarri F. Day-ahead hourly forecasting of power generation from photovoltaic plants. *IEEE Trans Sustain Energy* 2018;9(2):831–42.
- [8] Bouzerdoum M, Mellit A, Pavan AM. A hybrid model (SARIMA–SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant. *Sol Energy* 2013;98:226–35.
- [9] van der Meer D, Widén J, Munkhammar J. Review on probabilistic forecasting of photovoltaic power production and electricity consumption. *Renew Sustain Energy Rev* 2017;1484–512.
- [10] Bacher P, Madsen H, Nielsen H. Online short-term solar power forecasting. *Sol Energy* 2009;83(10):1772–83.
- [11] Almeida MP, Perpiñán O, Narvarte L. PV Power forecast using a nonparametric PV model. *Sol Energy* 2015;115:354–68.
- [12] Alessandrini S, Delle Monache L, Sperati S, Cervone G. An analog ensemble for short-term probabilistic solar power forecast. *Appl Energy* 2015;157:95–110.
- [13] von Loeper F, Schaumann P, de Langlard M, Hess R, Bäsman R, Schmidt V. Probabilistic prediction of solar power supply to distribution networks, using forecasts of global horizontal irradiation. *Sol Energy* 2020;203:145–56.
- [14] von Loeper F, Kirstein T, Idlbi B, Ruf H, Heilscher G, Schmidt V. Probabilistic analysis of solar power supply using d-vine copulas based on meteorological variables. In: S. Goettlich MH, Milde A, editors. *Mathematical modeling, simulation and optimization for power engineering and management, mathematics in industry*, Vol. 34. Springer; 2021, p. 51–68.
- [15] Wang H, Yi H, Peng J, Wang G, Liu Y, Jiang H, et al. Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network. *Energy Convers Manage* 2017;153:409–22.
- [16] Joe H. *Dependence Modeling with Copulas*. Chapman and Hall/CRC; 2014.
- [17] Lu Q, Hu W, Min Y, Yuan F, Gao Z. Wind power uncertainty modeling considering spatial dependence based on pair-copula theory. In: *Proceedings of the IEEE PES General Meeting| Conference & Exposition. IEEE*; 2014, p. 1–5.
- [18] Becker R. Generation of time-coupled wind power infeed scenarios using pair-copula construction. *IEEE Trans Sustain Energy* 2017;9(3):1298–306.
- [19] Golestaneh F, Gooi HB. Multivariate prediction intervals for photovoltaic power generation. In: *IEEE Innovative Smart Grid Technologies-Asia (ISGT-Asia)*. IEEE; 2017, p. 1–5.
- [20] Baldauf M, Seifert A, Förstner J, Majewski D, Raschendorfer M, Reinhardt T. Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities. *Mon Weather Rev* 2011;139(12):3887–905.
- [21] Hess R. Statistical postprocessing of ensemble forecasts for severe weather at Deutscher Wetterdienst. *Nonlinear Process Geophys* 2020;27(4):473–87.
- [22] Schaumann P, de Langlard M, Hess R, James P, Schmidt V. A calibrated combination of probabilistic precipitation forecasts to achieve a seamless transition from nowcasting to very short-range forecasting. *Weather Forecast* 2020;35(3):773–91.
- [23] Schaumann P, Hess R, Rempel M, Blahak U, Schmidt V. A calibrated and consistent combination of probabilistic forecasts for the exceedance of several precipitation thresholds using neural networks. *Weather Forecast* 2021;36(3):1079–96.
- [24] Saint-Drenan Y, Good G, Braun M. A probabilistic approach to the estimation of regional photovoltaic power production. *Sol Energy* 2017;147:257–76.
- [25] Laiti L, Giovannini L, Zardi D, Belluardo G, Moser D. Estimating hourly beam and diffuse solar radiation in an alpine valley: A critical assessment of decomposition models. *Atmosphere* 2018;9(4):117.
- [26] Saint-Drenan Y-M, Bofinger S, Fritz R, Vogt S, Good G, Dobschinski J. An empirical approach to parameterizing photovoltaic plants for power forecasting and simulation. *Sol Energy* 2015;120:479–93.
- [27] Kamphuis N, Gueymard C, Holtzapfel M, Duggleby A, Annamalai K. Perspectives on the origin, derivation, meaning, and significance of the isotropic sky model. *Sol Energy* 2020;201:8–12.
- [28] Martin N, Ruiz J. Calculation of the PV modules angular losses under field conditions by means of an analytical model. *Sol Energy Mater Sol Cells* 2001;70(1):25–38.
- [29] Martin N, Ruiz J. Annual angular reflection losses in PV modules. *Prog Photovolt, Res Appl* 2005;13(1):75–84.
- [30] Wagner A, Bendel C. Photovoltaic measurement relevant to the energy yield. In: *Proceedings of the WCPEC-3 world conference on photovoltaic energy conversion*, Osaka, Japan; 2003.
- [31] Humada AM, Darweesh SY, Mohammed KG, Kamil M, Mohammed SF, Kasim NK, et al. Modeling of PV system and parameter extraction based on experimental data: Review and investigation. *Sol Energy* 2020;199:742–60.
- [32] Marion B, Kroposki B, Emery K, Cueto J, Myers D, Osterwald C. Validation of a photovoltaic module energy ratings procedure at NREL. Technical Report, Golden, Colorado, USA: National Renewable Energy Laboratory; 1999.
- [33] Kou Q, Klein S, Beckman W. A method for estimating the long-term performance of direct-coupled pv pumping systems. *Sol Energy* 1998;64(1–3):33–40.
- [34] Driesse A, Jain P, Harrison S. Beyond the curves: Modeling the electrical efficiency of photovoltaic inverters. In: *Proceedings of the 33rd IEEE Photovoltaic Specialists Conference. IEEE*; 2008, p. 1–6.
- [35] Baumgartner FP, Schmidt H, Burger B, Bruendlinger R, Haeblerlin H, Zehner M. Status and relevance of the DC voltage dependency of the inverter efficiency. In: *Proceedings of the 22nd European photovoltaic solar energy conference and exhibition*; 2007. p. 2499–2506.
- [36] Schinke A, Hirsch H. Impact of electric vehicle charging and photovoltaic generation on distribution system voltage volatility. In: *Proceedings of the IEEE power energy society general meeting (PESGM)*. IEEE; 2019, p. 1–5.
- [37] Bessa RJ, Trindade A, Silva CS, Miranda V. Probabilistic solar power forecasting in smart grids using distributed information. *Int J Electr Power Energy Syst* 2015;72:16–23.
- [38] McNeil AJ, Frey R, Embrechts P. *Quantitative risk management: Concepts, techniques and tools*. Princeton University Press; 2005.
- [39] Aas K, Czado C, Frigessi A, Bakken H. Pair-copula constructions of multiple dependence. *Insurance Math Econom* 2009;44(2):182–98.

- [40] Nelsen R. An introduction to copulas. Springer; 2006.
- [41] Konishi S, Kitagawa G. Information Criteria and Statistical Modeling. Springer; 2008.
- [42] Joe H, Xu J. The estimation method of inference functions for margins for multivariate models. University of British Columbia; 1996, <https://open.library.ubc.ca/collections/facultyresearchandpublications/52383/items/1.0225985>.
- [43] Inman RH, Pedro HT, Coimbra CF. Solar forecasting methods for renewable energy integration. Prog Energy Combust Sci 2013;39(6):535–76.
- [44] Coimbra CF, Kleissl J, Marquez R. Solar Energy Forecasting and Resource Assessment. Elsevier / Academic Press; 2013, p. 171–94.
- [45] Gneiting T, Raftery A. Strictly proper scoring rules, prediction, and estimation. J Amer Statist Assoc 2007;102(477):359–78.
- [46] Pinson P, Tastu J. Discrimination ability of the energy score. Technical Report 15, Technical University of Denmark; 2013.
- [47] Golestaneh F, Gooi HB, Pinson P. Generation and evaluation of space–time trajectories of photovoltaic power. Appl Energy 2016;176:80–91.
- [48] Scheuerer M, Hamill TM. Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. Mon Weather Rev 2015;143:1321–34.
- [49] Diebold FX, Mariano RS. Comparing predictive accuracy. J Bus Econom Statist 1995;20(1):134–44.
- [50] Gneiting T, Katzfuss M. Probabilistic forecasting. Annu Rev Stat Appl 2014;1:125–51.
- [51] Harvey D, Leybourne S, Newbold P. Testing the equality of prediction mean squared errors. Int J Forecast 1997;13(2):281–91.