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

A. Schinke-Nendza[1a], F. von Loeper[b], P. Osinski[a], 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*

Abstract

The fast growth of photovoltaic (PV) power generation leads to an increasing impact of decentralized PV units on the power system operation. As a consequence, the importance of appropriate PV power forecasts for planing and decision support is rising. This paper introduces a hybrid approach combining a statistically post-processed physical PV power model with a D-vine copula model to generate probabilistic forecasts on a network node level. In a first step, deterministic predictions are generated by a PV power model and statistically post-processed using a vector autoregressive model with exogenous variables. In a second step, the spatial error distribution of the post-processed deterministic predictions is taken into account by applying D-vine copulas. The model is benchmarked against various combinations of a PV power model (with and without statistical post-processing) and a probabilistic model. The Energy score, Variogram-based score and Diebold-

[1] Corresponding author
email: aiko.schinke-nendza@uni-due.de

Mariano test are applied to evaluate the proposed model and highlight the strong performance of the proposed hybrid approach.
Keywords: Solar power supply, forecasting, physical PV model, VARX model, error distribution, D-vine copula, spatial dependency

1. Introduction

Over the last decades, photovoltaic (PV) power has become one of the most important renewable energies (RE) in Germany, reaching an installed capacity of 49 GW in 2019 (BNetzA, 2020). 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 can be used for a secure network operation, maintenance scheduling and long-term portfolio management (Bessa et al., 2017; Schermeyer et al., 2018). 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 models typically use numerical weather predictions (NWP) and a model representation of the PV power plant to convert meteorological parameters into power output (Lorenz et al., 2011; Saint-Drenan et al., 2017). Most available mete-

orological 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. Most authors have identified the clearness index, i.e., the ratio of global irradiance to extraterrestrial irradiance, as the most influential variable (Erbs et al., 1982; Perez et al., 1986; Reindl et al., 1990), but several other variables such as temperature, wind speed and visibility have been considered as well (Lou et al., 2016; Skartveit and Olseth, 1987). In this context, especially the BRL model, following Ridley et al. (2010), 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. (2016), 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 auto-regressive 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 (Chu et al., 2015; Pedro and Coimbra, 2012). Furthermore, combinations of statistical and physical models can be used to cope with the

non-linear patterns in the residuals (Bouzerdoum et al., 2013).

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. (2017). Common methods are e.g. quantile regression (Bacher et al., 2009), quantile regression forests (Almeida et al., 2015) and analog model chain forecasts (Alessandrini et al., 2015). More recently, a bivariate copula model has been proposed for the prediction of level-crossing probabilities of solar power supply (von Loeper et al., 2020). 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 von Loeper et al. (2021) by including influence factors such as further weather information and solar power supply of previous hours as explanatory variables using multivariate D-vine 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 relvant 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 (Joe, 2014; Nelsen, 2006). In particular, Gaussian and D-vine copulas have been applied to model the multivari-

ate probability distribution of wind power supply at neighboring wind parks (Becker, 2017; Lu et al., 2014; Tastu et al., 2013). 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 (Golestaneh et al., 2016; Golestaneh and Gooi, 2017).

When considering predictive grid management systems, following Bessa et al. (2017), the question on the performance of previously presented methods arises, since most approaches either focus on individual PV units or aggregate them without paying attention to underlying network topologies. However, the aggregation of PV units for different purposes may diverge from the spatial dependency of the individual PV units. Most probabilistic forecasts are based on statistical PV models, coming along with rather high computational effort, see Gigoni et al. (2018), and facing some elementary drawbacks compared to physical models. For instance, statistical models of individual PV units are not capable to estimate exceptional situations and can be rather weak in out-of-sample applications, e.g. due to modifications (i.e., the extension or removal of defective panels) or with regard to local weather conditions. Furthermore, statistical PV models for clusters of PV units are suffering losses regarding their forecasting performance in case of renewable curtailment or disconnection of individual units due to maintenance.

Consequently, this paper proposes a hybrid approach to generate probabilistic forecasts of PV power supply at different network nodes, by combining a physical PV model with a statistical post-processing and applying a multivariate parametric D-vine copula model to cope with the uncertainty

of forecasting errors. 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. In Section 2, the physical PV model is introduced briefly. Section 3 explains how the VARX model is applied to improve the deterministic forecasts. In Section 4, D-vine copulas are applied to model the error distribution, which is subsequently combined with the deterministic predictions to generate probabilistic forecasts. The evaluation methodology is presented in Section 5. Finally, in Section 6 the hybrid approach is applied to a case study, 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. 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 Figure 1. The model is based on statistically post-processed NWP data which are combined with synoptic observations, see Baldauf et al. (2011), Hess (2020) and Schaumann et al. (2020). In addition, a processing of weather data is carried out to receive all required inputs for the physical PV model. In contrast to common probabilistic approaches, see Saint-Drenan et al. (2017), 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, this physical model (abbreviated as PM) computes the net power injection for individual PV units. In this context, four modeling components are applied subsequently using weather data, data of PV units provided by the system operator and representative parameters for panels and inverters:

1. Irradiance decomposition model

   The available global irradiance on a horizontal surface is decomposed into a direct and diffuse fraction by means of a logistic approximation

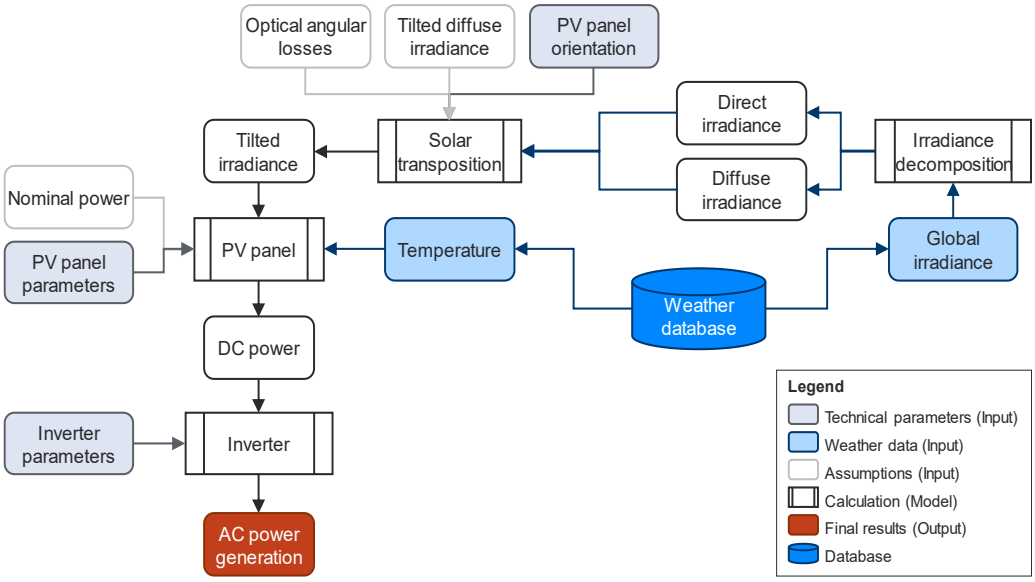


Figure 1: Physical model of photovoltaic power generation.

following Ridley et al. (2010). Moreover, the parameters of the model are recalibrated for the considered location, hence, increasing forecasting performance.

2. Solar transposition model

   The effective irradiance on the tilted surface of the PV panels is determined following Saint-Drenan et al. (2015). In this context, the diffuse irradiance on the tilted surface of the PV panels is determined by applying an isotropic approach, according to Liu and Jordan (1963), and taking the optical angular losses on the surface of the PV panels into account, following Martin and Ruiz (2001).

3. PV panel model

   The power output of the PV panels is obtained by using a simplified single diode solar cell model to determine the maximum power point

(MPP) voltage and current, see Chan et al. (1986) and Wagner and Bendel (2003). Thereby, a linear temperature dependency of the MPP voltage and current is assumed, based on commonly used voltage and current temperature coefficients as prosposed by Marion et al. (1999), together with a non-linear temperature model for the saturation current as described by Kou et al. (1998).

4. Inverter model

   The efficiency of the PV inverter is modelled using a generic quadratic approach for the losses based on Driesse et al. (2008) and neglecting the voltage dependency to ensure the applicability of the model to all types of inverters, following Baumgartner et al. (2007).

The corresponding pre-processing of the NWP data is explained in Section 2.1 and the PV panel and inverter models are introduced in Section 2.2 briefly. More details on the PV panel and inverter models are provided in Schinke and Hirsch (2019).

2.1. Pre-processing of weather data

In case the raw NWP data do not contain the necessary information for the PV panel model introduced below, the question arises how to deal with this lack of information, notably if they do not provide information on the direct horizontal irradiance. In the following, we apply the so-called Boland–Ridley–Laurent (BRL) model, following Ridley et al. (2010), which provides the main advantage that it relies on the global irradiance exclusively and the direct horizontal irradiance can be expressed in a simple equation, see Equation 1 below. Furthermore, the BRL model has performed reasonably

well compared to other models (Torres et al., 2010). However, Laiti et al. (2018) report a substantial improvement of the BRL model performance by refitting the coefficients β_0 to β_5 based on regional measurements. Therefore, the coefficients of the BRL model will be recalculated in Section 6.3.1, whereby the improved results are benchmarked to the values given in literature. In this context, the BRL model introduces a proportional relation between the diffuse fraction of the direct horizontal irradiance $E_{\mathrm{dif,hor}}$ and the global irradiance $E_{\mathrm{G,hor}}$, on the one hand, meanwhile assuming a logistic relation to the clearness index k_t during time step t (i.e., using average values based on a certain time granularity, e.g. of one hour) and further variables, on the other hand:

$$E_{\mathrm{dif,hor},t} = E_{\mathrm{G,hor},t} \frac{1}{1+\mathrm{e}^{\beta_0+\beta_1 k_t+\beta_2 AST+\beta_3 \sin\gamma_{\mathrm{s},t}+\beta_4 K_T+\beta_5 \varphi_t}}, \tag{1}$$

where AST is the apparent solar time in decimal hours and γ_{s} is the solar elevation. Furthermore, K_T is the daily clearness index defined as the sum of the global irradiance divided by the sum of the extraterrestrial irradiance of all hours of the day, and φ_t serves as an indicator for the persistence of k_t defined as $\varphi_t = (k_{t-1} + k_{t+1})/2$ if the hours t-1 and t+1 are not before sunrise nor after sunset and $\varphi_t = k_t$ otherwise.

2.2. PV panel model and inverter model

To determine the power generation $\widehat{P}_{\mathrm{Phys},t}$ of an individual PV unit, the time-variant efficiency of the solar panels $\eta_{\mathrm{pnl},t}$ is considered in a first step. Thereupon, the power output of the panels $P_{\mathrm{pnl},t}$ can be expressed as

$$P_{\mathrm{pnl},t} = \eta_{\mathrm{pnl},t} A_{\mathrm{Tot}} E_{\mathrm{Eff,tilt},t}, \tag{2}$$

where A_{Tot} is the total surface area of the considered PV unit and $E_{\mathrm{Eff,tilt,t}}$ is the time-variant effective irradiance on the tilted surface as described by Saint-Drenan et al. (2015), hence, taking sun position, angular optical losses as well as PV panel coordinates, elevation and azimuth angles into account. In a second step the power infeed into the electricity grid requires a conversion to alternating current (AC) using an inverter with efficiency $\eta_{\mathrm{inv},t}$ resulting in the relationship:

$$\widehat{P}_{\mathrm{Phys},t} = \eta_{\mathrm{inv},t} P_{\mathrm{pnl},t} = \eta_{\mathrm{inv},t} \eta_{\mathrm{pnl},t} A_{\mathrm{Tot}} E_{\mathrm{Eff,tilt},t}. \tag{3}$$

The efficiency $\eta_{\mathrm{pnl},t}$ of the solar panels depends on the tilted irradiance and the panel temperature, which in turn depends on irradiance and ambient temperature. The time-variant efficiency $\eta_{\mathrm{inv},t}$ of the inverter solely depends on the direct current (DC) power input, neglecting the effect of voltage fluctuations and assuming an approximately constant temperature at the inverter site. Thereby the 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 Schinke and Hirsch (2019).

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 pure vector autoregressive (VAR) models, e.g.

following Bessa et al. (2015), the VARX model allows to take into account exogenous variables and to benchmark the results of the physical model. Furthermore, the VARX model is not only used for benchmarking but also for statistical post-processing of the forecasts.

In the following, the cumulative net power injection of all PV units connected to a certain network node $k \in \mathcal{K} = \{1, \ldots, K\}$ is considered as the relevant forecast variable $\widehat{P}_{\mathrm{Phys},t,k}$. Thereupon, the VARX model forecasts the net power injection $\hat{P}_{t,k}$ at each network node $k \in \mathcal{K} = \{1, \ldots, K\}$ 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 $m \in \mathcal{M} = \{1, \ldots, M\}$ is the index of the exogenous variables:

$$\frac{\hat{P}_{t,k}}{P_{0,k}} = \alpha_k + \sum_{l \in \mathcal{L}} \sum_{i \in \mathcal{K}} \beta_{l,k,i} \frac{P_{t-l,i}}{P_{0,i}} + \sum_{m \in \mathcal{M}} \sum_{i \in \mathcal{K}} \gamma_{m,k,i} X_{m,t,i}. \tag{4}$$

In Equation 4, $P_{0,k}$ is the cumulative nominal power of the PV units connected to node k. Following, e.g., Bessa et al. (2015), the parameters of the VARX model α_k, $\boldsymbol{\beta_{l,k}} = (\beta_{l,k,1}, \ldots, \beta_{l,k,K})$ with $l \in \mathcal{L}$ and $\boldsymbol{\gamma_{m,k}} = (\gamma_{m,k,1}, \ldots, \gamma_{m,k,K})$ with $m \in \mathcal{M}$ can be determined by K separate ordinary least square (OLS) minimizations of the error between predicted power $\widehat{P}_{\tau,k}$ and observed power $P_{\tau,k}$ of each network node $k \in \mathcal{K}$ using a moving-horizon of T_h hours preceding to the forecasting time step t:

$$\left\{\widehat{\alpha}_k, \widehat{\boldsymbol{\beta}}_{\boldsymbol{l,k}}, \widehat{\boldsymbol{\gamma}}_{\boldsymbol{m,k}}\right\} = \underset{\alpha_k, \boldsymbol{\beta_{l,k}}, \boldsymbol{\gamma_{m,k}}}{\operatorname{argmin}} \sum_{\tau=t-T_h}^{t-1} \left(P_{\tau,k} - \widehat{P}_{\tau,k}\right)^2. \tag{5}$$

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 Bessa et al. (2015). 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 $X_{1,t,k} = \overline{E}^{(k)}_{\mathrm{G,hor},t}$, $X_{2,t,k} = \overline{\vartheta}^{(k)}_{a,t}$ and $M = 2$,
2. PM-VARX: Set $X_{1,t,k} = \widehat{P}_{\mathrm{Phys},t,k}$ and $M = 1$,

where $\overline{E}^{(k)}_{\mathrm{G,hor},t}$ is the average global horizontal irradiance, $\overline{\vartheta}^{(k)}_{a,t}$ is the ambient temperature and $\widehat{P}_{\mathrm{Phys},t,k}$ 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 (McNeil et al., 2005). 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) $C : [0,1] \times [0,1] \to [0,1]$ 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 $R_1, \ldots, R_n$ for any fixed $n > 2$. Further details regarding general copula theory can be found e.g. in Aas et al. (2009), Joe (2014) and Nelsen (2006).

There exist many different types of vine copulas. However, D-vine copulas are choosen 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 Joe (2014). The D-vine copula corresponds to the decomposition of an n-dimensional joint density $f_{1,\ldots,n}$ of the random variables $R_1, \ldots, R_n$ given by

$$f_{1,\ldots,n} = \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{i,i+j|i+1,\ldots,\, i+j-1} \left(F_{i|i+1:i+j-1}, F_{i+j|i+1:i+j-1} \right) \cdot \prod_{l=1}^{n} f_l, \quad (6)$$

where f_i denotes the density of the random variable R_i and $c_{i,i+j|i+1,\ldots,\, i+j-1}$ is some bivariate copula density. The bivariate copula densities $c_{i,i+j|i+1,\ldots,\, i+j-1}$ 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 $R_{i+1}, \ldots, R_{i+j-1}$. A mathematical derivation of Equation 6 can be found in Joe (2014).

4.2. Multivariate distribution model

First, an univariate distribution has to be fitted separately for each of the random variables $R_1, \ldots, R_n$. For that purpose, histograms are computed providing a visual impression which types of marginal distributions might be most suitable for describing the realizations of $R_1, \ldots, R_n$. 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 (Konishi and Kitagawa, 2008). 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 $c_{i,i+j|i+1,\ldots,\ i+j-1}$ in the D-vine copula structure are estimated sequentially based on the transformed data (Joe, 2014). For that purpose, the Archimedean copula types Joe, Frank, Gumbel and Clayton are considered (Nelsen, 2006). Each conditional copula density is fitted to the outputs of the CDFs, which are given as arguments in Equation 6. The copula parameters are estimated for each considered copula type using the inference function for margins method as proposed in Joe and Xu (1996). 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 $(R_1, \ldots, R_n)$. Using the sampling algorithm for vine copulas

explained by Aas et al. (2009), 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 $(R_1, \ \ldots, R_n)$.

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 modeling 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 Equation 7 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 $\varepsilon_{t,k}^{(\delta)} = P_{t,k} - \widehat{P}_{t,k}^{(\delta)}$ are computed by comparing the forecasts

to actual measurements subsequently, these are normalized as follows:

$$e_{t,k}^{(\delta)} = \frac{\varepsilon_{t,k}^{(\delta)} - \frac{1}{|T_{\text{Train},t}|} \sum_{\tau \in T_{\text{Train},t}} \varepsilon_{\tau,k}^{(\delta)}}{P_{\text{Ex},t,k}}, \tag{7}$$

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 $P_{\text{Ex},t,k} = E_{\text{Ex,hor},t,k} P_{0,k} / E_{\text{Ex},0}$ represents the theoretical extraterrestrial PV power generation at time step t. Here $E_{\text{Ex,hor},t,k}$ is the extraterrestrial irradiance which is computed based on the physical model of Inman et al. (2013), $E_{\text{Ex},0}$ 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 $e_{t,k}^{(\delta)}$ 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 Equation 7 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 cor-

responding 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. van der Meer et al. (2017):

$$NRMSE_k = \sqrt{\frac{1}{|T_{\text{Test}}|} \sum_{t \in T_{\text{Test}}} \left(\frac{P_{t,k} - \hat{P}_{t,k}}{P_{0,k}} \right)^2}, \tag{8}$$

where T_{Test} is the test dataset (cf. Section 6.1). Furthermore, the quantity $NRMSE_k$ given in Equation 8 can be used to determine the forecast skill S_k, following Coimbra et al. (2013), 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 $E_{\text{Ex,hor},t,k}$. Thus, the benchmarking persistence forecasts are computed using $\hat{P}_{t,k} = E_{\text{Ex,hor},t,k} \left(P_{t,k-1}/E_{\text{Ex,hor},t,k-1}\right)$. The forecast skill is then defined as

$$S_k = 1 - \frac{NRMSE_k}{NRMSE_{\text{pers},k}}, \tag{9}$$

with

$$NRMSE_{\text{pers,k}} = \sqrt{\frac{1}{|T_{\text{Test}}|} \sum_{t \in T_{\text{Test}}} \left(\frac{P_{t,k}}{P_{0,k}} - \frac{P_{t,k-1}}{P_{0,k}} \frac{E_{\text{Ex,hor},t,k}}{E_{\text{Ex,hor},t,k-1}} \right)^2}. \tag{10}$$

The forecast skill S_k given in Equation 9 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. Testing the distribution of forecasting errors

To evaluate the benefits of the proposed D-vine copula model a test for univariate normality of the forecasting errors is applied. Furthermore, the representativity of the training dataset for application on the test dataset

shall be investigated in advance to preclude systematic deviations of the forecasting errors e.g. resulting from seasonal or annual weather phenomena.

In this context, the *Shapiro-Wilk test* is applied to test the null hypothesis that the univariate distribution of the forecasting errors at each network node follows a normal distribution (Shapiro and Wilk, 1965). The *Shapiro-Francia test* (as an extension of the Shapiro-Wilk test) provides better results in case of leptokurtic samples (Royston, 1993), yet the *Shapiro-Wilk test* should be preferred for platykurtic samples. Therefore, we evaluate the kurtosis of the considered sample of forecasting errors in a first step. Thereupon, the best fitting test (i.e., either the original Shapiro-Wilk test or its extension, the Shapiro-Francia test) is selected automatically in a second step. Furthermore, the acquired test statistic is normalized according to Royston (1993) in a third step, providing an asymptotic standard normal distribution. For simplicity reasons we reference this procedure as the *extended Shapiro-Wilk test*. In case the Shapiro-Wilk test provides ambiguous results in terms of univariate normality, the spatial distributions of the forecasting errors can be furthermore tested for multivariate normality using *Royston's H test* (Royston, 1983).

Besides testing for normality of the univariate distributions, the two-sample *Kolmogorov-Smirnov test* can be carried out to preclude systematic deviations of the forecasting errors (Klotz, 1967). In this context, the null hypothesis ist tested that the distributions of the forecasting errors of the training dataset are from the same statistical population as the distributions of the test dataset. In case the null hypothesis is rejected, both datasets can be assumed to differ substantially, indicating systematic changes between

both datasets, e.g. caused by weather phenomena or dismantling of defective PV panels. The test statistic is given as

$$D_{\text{Test,Train}} = \sup_{x \in \mathbb{R}} \left| F^{(\delta)}_{\text{Test}}(x) - F^{(\delta)}_{\text{Train}}(x) \right| \tag{11}$$

for the empirical cumulative distribution functions $F^{(\delta)}_{\text{Test}}$ and $F^{(\delta)}_{\text{Train}}$ obtained from the forecasting errors of the deterministic model δ in the test and training datasets, respectively. In this context, under the null hypothesis, the test statistic follows a distribution which can be described by look-up tables, see e.g. Miller (1956).

5.3. 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 $\widehat{\boldsymbol{P}}^{(\xi)}_t$ of model chain ξ a sample of size m of the forecasting errors $\boldsymbol{\varepsilon}^{(\xi)}_t$ is drawn from the corresponding probabilistic model (cf. Section 4.3). Thereupon, the probabilistic forecast is defined as $\widetilde{\boldsymbol{P}}^{(\xi)}_t = \widehat{\boldsymbol{P}}^{(\xi)}_t + \boldsymbol{\varepsilon}^{(\xi)}_t$, 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 (2007). In this context, the energy score (ES) during time step t can be written as

$$ES_t = \frac{1}{m} \sum_{j=1}^{m} \sqrt{\sum_{k \in \mathcal{K}} (p_{t,k} - \widetilde{p}_{t,k,j})^2} - \frac{1}{2m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} \sqrt{\sum_{k \in \mathcal{K}} (\widetilde{p}_{t,k,i} - \widetilde{p}_{t,k,j})^2}, \tag{12}$$

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

However, Pinson and Tastu (2013) show that, while the energy score is useful to detect differences between the observed and predicted power sup-

ply, its discriminative ability in terms of variance and correlation is poor. Therefore, we follow Golestaneh et al. (2016) 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 (Scheuerer and Hamill, 2015; Golestaneh et al., 2016). In case of a K-variate probabilistic forecast represented by a sample of size m, the variogram-based score (VS) is defined as

$$VS_t = \sum_{i \in \mathcal{K}} \sum_{k \in \mathcal{K}} w_{i,j} \left(|p_{i,t} - p_{t,k}|^{0.5} - \frac{1}{m} \sum_{j=1}^{m} |\widetilde{p}_{t,i,j} - \widetilde{p}_{t,k,j}|^{0.5} \right)^2, \qquad (13)$$

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 Scheuerer and Hamill (2015). 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., $ES = \frac{1}{|T_{\text{Test}}|} \sum_{t \in T_{\text{Test}}} ES_t$ and $VS = \frac{1}{|T_{\text{Test}}|} \sum_{t \in T_{\text{Test}}} VS_t$.

5.4. 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 Diebold and Mariano (1995), and Gneiting and Katzfuss (2014). 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 $\varepsilon_{1,k}^{(\delta)}, \varepsilon_{2,k}^{(\delta)}, \ldots$ introduced in Section 4.3

as realizations of a certain random variable $\varepsilon_k^{(\delta)}$. 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.,

$$\mathbb{E}\left(g\left(\varepsilon_k^{(\delta_i)}\right) - g\left(\varepsilon_k^{(\delta_j)}\right)\right) = 0 \tag{14}$$

for some function $g : \mathbb{R} \to \mathbb{R}$. Possible choices for g are $g(\varepsilon) = |\varepsilon|$ or $g(\varepsilon) = \varepsilon^2$, 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 $t_{\mathrm{DM},k}$, given by

$$t_{\mathrm{DM},k} = \frac{\bar{d}_k \sqrt{|T_{\mathrm{Test}}|}}{\sqrt{\nu_k}}, \tag{15}$$

where $|T_{\mathrm{Test}}|$ is the number of forecasts in the test dataset and $\bar{d}_k$ is the sample mean of the differences $d_{t,k} = g(\varepsilon_{t,k}^{(\delta_i)}) - g(\varepsilon_{t,k}^{(\delta_j)})$. Furthermore, the normalizing factor ν_k in Equation 15 is given by

$$\nu_k = \varsigma_0 + 2 \sum_{i=1}^{h-1} \frac{(|T_{\mathrm{Test}}| - i)}{|T_{\mathrm{Test}}|} \varsigma_i, \tag{16}$$

where ς_i is the i-th sample autocovariance of $d_{1,k}, d_{2,k}, \ldots$ and h is the forecast horizon in hours, see Harvey et al. (1997). Under the null hypothesis, the test statistic $t_{\mathrm{DM},k}$ is asymptotically standard normal distributed, i.e. $t_{\mathrm{DM},k} \sim \mathcal{N}(0, 1)$.

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 Gneiting and Katzfuss (2014). Hence, the time series of two scores $S(\widetilde{\boldsymbol{P}}_t^{(\xi_i)}, \boldsymbol{p}_t)$ and $S(\widetilde{\boldsymbol{P}}_t^{(\xi_j)}, \boldsymbol{p}_t)$ are used,

which can be either the energy score (ES) or the variogramm-based score (VS) introduced in Section 5.3, with $\boldsymbol{p}_t$ 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 $\bar{S}^{(\xi)} = \frac{1}{|T_{\text{Test}}|} \sum_{t \in T_{\text{Test}}} S(\widetilde{\boldsymbol{P}}_t^{(\xi)}, \boldsymbol{p}_t)$, the test statistic can be accessed as

$$t_{\text{DM}}^{(\xi_i,\xi_j)} = \sqrt{|T_{\text{Test}}|} \frac{\bar{S}^{(\xi_i)} - \bar{S}^{(\xi_j)}}{\widehat{\sigma}^{(\xi_i,\xi_j)}} \tag{17}$$

with

$$\widehat{\sigma}^{(\xi_i,\xi_j)} = \sqrt{\frac{1}{|T_{\text{Test}}|} \sum_{t \in T_{\text{Test}}} \left(S(\widetilde{\boldsymbol{P}}_t^{(\xi_i)}, \boldsymbol{p}_t) - S(\widetilde{\boldsymbol{P}}_t^{(\xi_j)}, \boldsymbol{p}_t) \right)^2}. \tag{18}$$

Referring to Gneiting and Katzfuss (2014), the test statistic $t_{\text{DM}}^{(\xi_i,\xi_j)}$ given in Equation 17 is asymptotically standard normal distributed under the null hypothesis mentioned above, i.e., $t_{\text{DM}}^{(\xi_i,\xi_j)} \sim \mathcal{N}(0,1)$. When rejecting the null hypothesis, the model chain ξ_i is preferred if $t_{\text{DM}}^{(\xi_i,\xi_j)}$ is negative and ξ_j is preferred if $t_{\text{DM}}^{(\xi_i,\xi_j)}$ is positive.

6. Example of application to real power systems

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 von Loeper et al. (2020). 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, Figure 2 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.

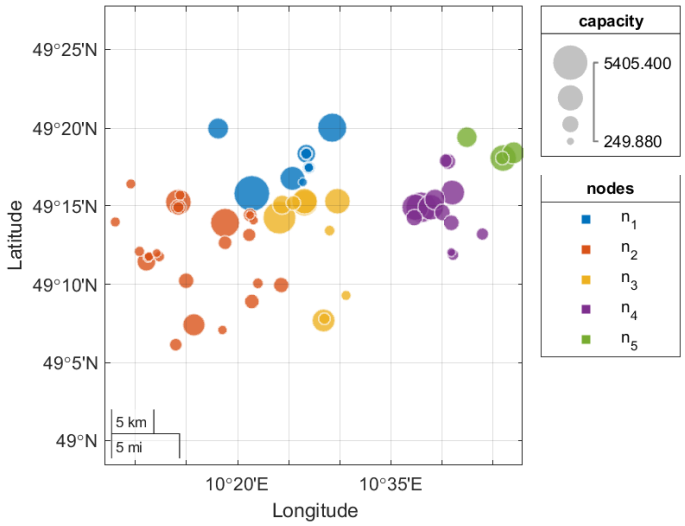


Figure 2: PV units considered in this case study, where the size of the circles corresponds to the nominal power (in kW) and the color indicates the associated network node

For deterministic forecasting of PV power supply, statistical forecasts of global horizontal irradiance and ambient temperature (2m above ground) of Ensemble-MOS of DWD are used, see also Baldauf et al. (2011), Hess (2020) and Schaumann et al. (2020). 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 hours. The forecasts are updated every 3 hours. 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 hours in advance) hereinafter referenced as the *intraday forecast* and on the earliest available forecast (i.e., with a forecast horizon of 17-19 hours 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).
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 hours 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.

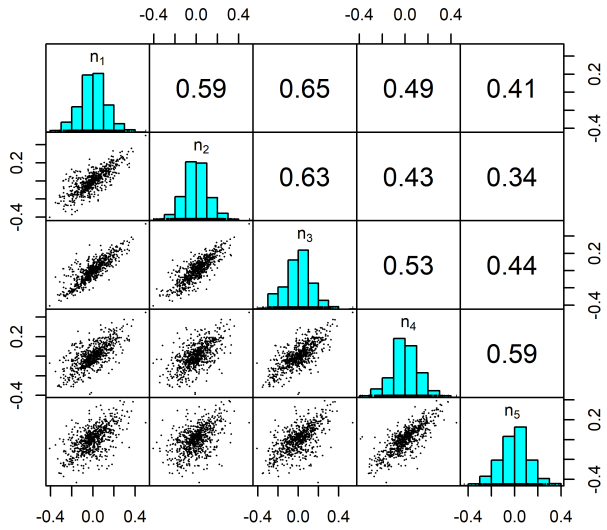


Figure 3: Histograms, pairwise scatter plots and pairwise Kendall's rank correlation coefficients of the forecasting errors (for intraday forecasts of the PM-B approach) in May, June and July of year 2016 evaluated for the five network nodes $n_1, \ldots, n_5$.

We consider the five network nodes $n_1, \ldots, n_5$, see Figure 2. 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 Figure 3. Based on visual inspection of the histograms, the forecasting errors at the network nodes $n_1, \ldots, n_5$ 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 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

$$\widehat{\tau} = \frac{2}{n(n-1)} \sum_{i<j} \operatorname{sgn}(x_i - x_j) \operatorname{sgn}(y_i - y_j) \tag{19}$$

for given realizations $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ of the random variables X and Y. The pairwise Kendall's rank correlation coefficients in Figure 3 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 Figures 2 and 3 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. That results in 30 fitted marginal distributions for each univariate distribution type. In Figure 4 the maximum values of the loglikelihood function of the fitted marginal distributions are visualized for the gamma, logistic, normal and Weibull distribution respectively. The

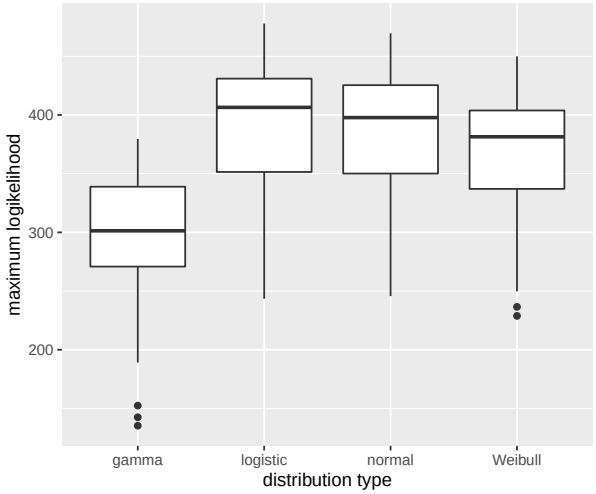


Figure 4: Boxplots of the maximum loglikelihoods for different univariate distribution types.

boxplot shows that the logistic distribution is slightly better than the normal distribution, whereas the Weibull and gamma distribution are clearly worse.

For intraday forecasts of the PM-B approach, we fitted a copula model as described in Section 4.2. The structure of the fitted D-vine copula and its (conditional) copula densities are visualized in Figure 5. Each panel entitled with $i, i+j|i+1, \ldots, i+j-1$ shows the fitted (conditional) copula density $c_{i,i+j|i+1,\ldots,i+j-1}$ in Equation 6. Note that the vine copula structure builds on vine copula structures of lower dimensions. For example, the copula densities $c_{1,2}, c_{2,3}$ and $c_{1,3|2}$ are part of the 3-dimensional D-vine copula corresponding to the error distribution of the nodes n_1, n_2 and n_3. To get the 4-dimensional D-vine copula for the error distribution of the nodes n_1, n_2, n_3 and n_4, the copula densities $c_{3,4}, c_{2,4|3}$ and $c_{1,4|2,3}$ have to be computed and for the 5-dimensional the rest of the copula densities in Figure 5 is

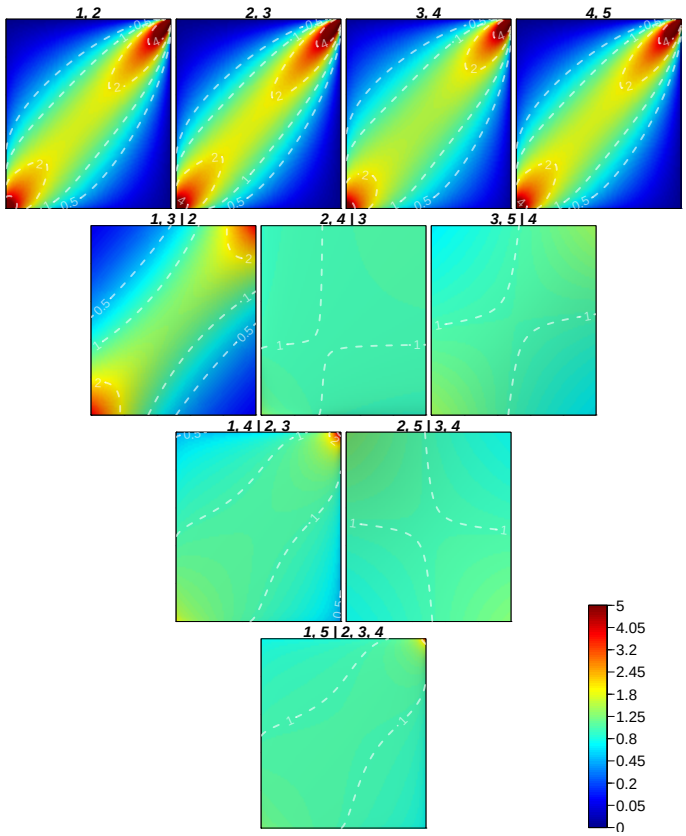


Figure 5: Graph structure of the fitted D-vine copula and its (conditional) copula densities for PM-B approach and intraday forecasts.

needed. Furthermore, the fitted copula densities reflect the interdependence between the considered forecast errors. One aspect is that the more uneven the conditional copula density, the greater is the interdependence. For instance, if the copula densities $c_{3,5|4}, c_{2,5|3,4}$ and $c_{1,5|2,3,4}$ would be all equal to one, the 5-dimensional joint density in Equation 6 would be the product of a 4-dimensional joint density for the nodes n_1, n_2, n_3 and n_4 and the two-dimensional joint density for the nodes n_4 and n_5. In that case, the joint density takes the dependence between the forecast errors at nodes n_4 and n_5

instead of the dependence between n_5 and other nodes such as n_1, n_2 and n_3 into account. Although in our case the conditional densities are $c_{3,5|4}, c_{2,5|3,4}$ and $c_{1,5|2,3,4}$ are not completely equal to one, we notice that the impact of errors at n_5 is more important for errors at n_4 than for the other nodes. That matches the fact that the distance between n_5 and n_1, n_2 and n_3 is larger than between n_5 and n_4, see Figure 2.

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 2.1) 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. Weather data pre-processing

As already mentioned in Section 2.1, the coefficients of the BRL model have been recalculated based on two central European measurements. Therefore, global and diffuse irradiance data for 2015 and 2016 from the weather stations in Lindenberg, Germany and Payerne, Switzerland (World Radiation Monitoring Center, 2019) are used. Data from both stations is available at a temporal resolution of one minute. Obviously erroneous data were filtered, mean values of global and diffuse irradiance per hour were calculated and all

required input variables for the BRL model were derived. For both stations, the coefficients β_i with $i \in \{1, \ldots, 5\}$ appearing in Equation 1 were recalculated while the calibrated models were cross-validated with the dataset from the other station. The resulting NRMSEs (see Equation 8) of the predicted diffuse irradiance compared to the measured diffuse irradiance are shown in Table 1. In this context, BRL stands for the original BRL model, meanwhile Fit PAY and Fit LIN stand for the calibrated models with coefficients fitted by means of the datasets of Payerne and Lindenberg.

Test dataset	**NRMSE**		
	BRL	Fit PAY	Fit LIN
Payerne	0.3316	0.2700	0.2840
Lindenberg	0.3180	0.2800	0.2590

Table 1: NRMSE for different combinations of utilized models and datasets used for testing.

Both recalibrations of the model produce a substantially better fit to the datasets than the original BRL model, see Table 1. Although comparisons between different weather stations are difficult because climatic conditions influence the predictive performance of the model, it can be noted that the achieved errors are lower than those of all tested models in Laiti et al. (2018). A recalibration of the coefficients seems hence appropriate. To make best use of the available European data, the final calibriation is based on the combined dataset for both weather stations (Payerne and Lindenberg). The coefficients are determined as $\beta_0 = -5.64$, $\beta_1 = 7.80$, $\beta_2 = 0.01$, $\beta_3 = -0.75$, $\beta_4 = 2.23$ and $\beta_5 = 0.20$.

The calibrated BRL model is applied to the ModelMix forecast data of DWD, described in Section 6.1. 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.

6.3.2. Deterministic intraday forecasts

For each forecasting model with a forecast horizon of 1 to 3 hours, the NRMSE and forecast skill S, as introduced in Section 5.1, are determined at the five considered network nodes. The results depicted in Figure 6 show the results for the proposed PM-VARX model as well as for both benchmark models PM-B and VARX-B (cf. Section 3).

Figure 6: NRMSE (a) and forecast skill S (b) of the five considered network nodes for intraday forecasts.

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 NMRSE values at nodes n_2 to n_5, and provides an overall better forecasting performance than the PM-B approach.

Network	**PM-VARX vs. VARX-B**		**VARX-B vs. PM-B**	
node	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

Table 2: Test statistic of the Diebold-Mariano test computed for deterministic intraday and day-ahead forecasts comparing all three deterministic models.

To test for significance of the differences in the NRMSE the Diebold-Mariano test, introduced in Section 5.4, is carried out for the intraday and day-ahead forecast horizons. The corresponding results are depicted in Table 2. 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

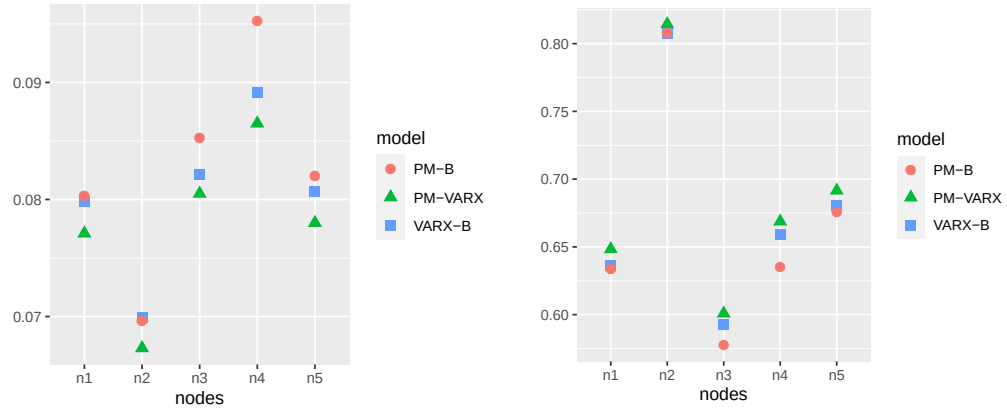
is not rejected, plus sign (+) confidence = 90%, one-asterisk (*) confidence = 95%, two-asterisks (**) confidence = 99% and three-asterisks (***) confidence = 99.9%.

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 6.3.3.

6.3.3. Deterministic day-ahead forecasts

The respective results for NRMSE and forecast skill for a forecast horizon of 17 to 19 hours are depicted in Figure 7. In this context, the NRMSEs for day-ahead forecasts at the network node level increase to a range from 6.7% to 8.9%.

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 2. Unlike to the intraday forecasts, the performance differences between VARX (benchmark) and the physical model



(a) (b)

Figure 7: NRMSE (a) and forecast skill S (b) of the five considered network nodes for day-ahead forecasts.

are not statistically significant, except for the network node n_4.

6.3.4. Testing the distributions of forecasting errors

As proposed in Section 5.2, a test for univariate normality of the forecasting errors is applied before fitting individual marginal distributions and D-vine copulas to model the multivariate marginal and spatial distributions of the forecasting error. In addition, the representativity of the training dataset for the test dataset is tested in advance to exclude exogenous effects. When considering the univariate distributions of the forecasting errors for fitting individual marginal distributions for each network node, the extended Shapiro-Wilk test is applied to test the null hypothesis that the univariate distribution is following a normal distribution. Table 3 provides the corresponding normalized test statistics for the training and test dataset, with asterisks indicating a rejection of the null hypothesis (cf. Section 6.3.2).

Training dataset

Network	**PM-VARX**		**VARX-B**		**PM-B**	
node	Intraday	Day-ahead	Intraday	Day-ahead	Intraday	Day-ahead
n_1	10.707^{***}	9.844^{***}	10.668^{***}	9.778^{***}	5.337^{***}	4.755^{***}
n_2	9.792^{***}	8.899^{***}	9.69^{***}	8.754^{***}	5.049^{***}	4.845^{***}
n_3	10.636^{***}	10.036^{***}	10.602^{***}	9.973^{***}	3.503^{***}	5.024^{***}
n_4	10.711^{***}	9.959^{***}	10.645^{***}	9.877^{***}	3.329^{***}	3.909^{***}
n_5	10.871^{***}	10.269^{***}	10.855^{***}	10.213^{***}	4.812^{***}	5.483^{***}

Test dataset

Network	**PM-VARX**		**VARX-B**		**PM-B**	
node	Intraday	Day-ahead	Intraday	Day-ahead	Intraday	Day-ahead
n_1	4.927^{***}	4.774^{***}	4.436^{***}	3.909^{***}	4.575^{***}	4.931^{***}
n_2	3.511^{***}	3.811^{***}	2.994^{**}	2.548^{**}	5.919^{***}	3.635^{***}
n_3	5.65^{***}	6.043^{***}	4.500^{***}	4.763^{***}	6.391^{***}	6.797^{***}
n_4	3.795^{***}	4.460^{***}	3.253^{***}	3.515^{***}	3.413^{***}	4.392^{***}
n_5	4.363^{***}	5.175^{***}	3.793^{***}	4.486^{***}	4.752^{***}	6.045^{***}

Table 3: Normalized test statistic of the extended Shapiro-Wilk test computed for the training and test datasets for the deterministic intraday and day-ahead forecasts.

The results in Table 3 clearly indicate that the univariate distributions of the forecasting errors at each network node are not normally distributed, based on a confidence level of at least 99%. Solely for the VARX-B model at node n_2 the null hypothesis cannot be rejected with 99.9% of confidence, but only with 99%. For all other models and nodes, the null hypothesis can be

rejected even at a confidence level of 99.9%. Since the extended Shapiro-Wilk test clearly rejects the hypothesis of univariate normality by means of the forecasting errors at each individual network node the multivariate normality can be precluded as well.

Furthermore, to preclude systematical deviations of the unbiased and normalized forecasting errors, e.g. resulting from seasonal or annual weather phenomena, between the training and test dataset, the results of the two-sample Kolmogorov-Smirnov test (cf. Section 5.2) are depicted in Table 4, whereby the notation corresponds to Table 3. In this context, the null hypothesis that both datasets, test and training, of forecasting errors corresponding to a certain deterministic model are drawn from the same statistical population is not rejected at a significance level of 95% or higher for the proposed PM-VARX approach. Only for one node in the intraday forecast and three nodes in the day ahead forecast, the null hypothesis is rejected to a confidence level of 90%. In terms of the PM-VARX approach, the training dataset is hence found to be at least partwise representative for the test dataset. In opposite to these results, both datasets are indicated to belong to the same population without any exceptions for the VARX-B approach. Regarding the PM-B approach, training and test data do finally not seem to be drawn from the same population, especially in terms of the day-ahead forecasts. Here the null hypothesis is not rejected only for two nodes in the intraday forecast. Therefore, the Kolmogorov-Smirnov test indicates the VARX-based approaches to eliminate systematical deviations, hence, improving the comparability of forecasting errors in the training and test dataset.

Network	**PM-VARX**		**VARX-B**		**PM-B**	
node	Intraday	Day-ahead	Intraday	Day-ahead	Intraday	Day-ahead
n_1	0.04	0.04	0.04	0.05	0.06	0.12^{***}
n_2	0.06	0.07^{+}	0.06	0.05	0.05	0.07^{*}
n_3	0.06^{+}	0.06^{+}	0.04	0.05	0.08^{*}	0.1^{***}
n_4	0.04	0.06^{+}	0.04	0.05	0.08^{**}	0.12^{***}
n_5	0.04	0.05	0.03	0.05	0.07^{*}	0.11^{***}

Table 4: Test statistic of the Kolmogorov-Smirnov test computed for the unbiased and normalized forecasting errors for all deterministic models of the training and test dataset.

6.3.5. 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. The result is then the $|\mathcal{K}| \times m$-dimensional matrix $\widetilde{\boldsymbol{P}}_t^{(\xi)} = \widehat{\boldsymbol{P}}_t^{(\xi)} + \boldsymbol{\varepsilon}_t^{(\xi)}$, to which the energy score (ES) as well as the variogram-based score (VS) described in Section 5.3 are applied, in order to compare the probabilistic forecast $\widetilde{\boldsymbol{P}}_t^{(\xi)}$ to the $|\mathcal{K}|$-dimensional vector of the actually measured PV power supplies $\boldsymbol{P}_t$. 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.

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	2.11	3.09	2.78
VARX-B	MVN	2.18	3.12	2.81
VARX-B	UVN	2.23	3.62	3.28
PM-B	D-vine	2.23	3.61	3.27
PM-B	MVN	2.38	3.86	3.49
PM-B	UVN	2.39	4.01	3.62

Table 5: Energy score (ES) and variogram-based scores (VS1 and VS2) of the intraday forecasts for different model chains.

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 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 statistical significant for intraday and day-ahead forecasts (cf. Sections 6.3.2 and 6.3.3). Thus, the Diebold-Mariano test is carried out for a pairwise comparison of the different probabilistic models only, see Table 6.

Deterministic	**D-vine vs. MVN**			**MVN vs. UVN**		
model	ES	VS1	VS2	ES	VS1	VS2
PM-VARX	-5.77^{***}	-1.79^{+}	-1.8^{+}	-6.2^{***}	-7.61^{***}	-7.66^{***}
VARX-B	-7.25^{***}	-1.42	-1.44	-5.89^{***}	-8.25^{***}	-8.29^{***}
PM-B	-13.09^{***}	-6.86^{***}	-6.95^{***}	-0.67	-2.02^{*}	-1.99^{*}

Table 6: Test statistic of the Diebold-Mariano test for pairwise comparison of the forecast performance based on ES, VS1 and VS2 of the intraday forecasts.

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 5 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 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.

6.3.6. Probabilistic day-ahead forecasts

For the day-ahead forecasts, i.e., for a forecast horizon of 17 to 19 hours, the energy score (ES) and the variogram-based scores (VS1 and VS2) are depicted in Table 7 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

Deterministic model	Probabilistic model	ES	VS1	VS2
PM-VARX	D-vine	2.52	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	2.63	4.01	3.61
VARX-B	MVN	2.82	4.15	3.75
VARX-B	UVN	2.81	4.68	4.24
PM-B	D-vine	2.68	4.25	3.84
PM-B	MVN	2.92	4.71	4.25
PM-B	UVN	2.91	4.71	4.26

Table 7: Energy score (ES) and variogram-based scores (VS1 and VS2) of the intraday forecasts for different model chains.

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 8.

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

Deterministic	**D-vine vs. MVN**			**MVN vs. UVN**		
model	ES	VS1	VS2	ES	VS1	VS2
PM-VARX	-8.5^{***}	-3.97^{***}	-3.99^{***}	-0.17	-6.75^{***}	-6.8^{***}
VARX-B	-10.58^{***}	-4.32^{***}	-4.35^{***}	0.14	-6.24^{***}	-6.29^{***}
PM-B	-14.09^{***}	-8.85^{***}	-8.99^{***}	0.99	-0.04	-0.02

Table 8: 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.

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 PV 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 PV forecasts and carrying out a comprehensive analysis using well-established models as benchmarks. On the one hand, the hybrid approach is based on a physical PV model with statistical pre-processing of the weather data and (VARX-based) post-processing of the deterministic forecasts. On the other hand, univariate marginal distributions and a D-vine copula are fitted to the normalised forecasting errors of the deterministic forecasting model to model the multivariate error distribution.

A case study on a power system in the south of Germany is carried out in order to demonstrate the advantages of the proposed hybrid approach. In this context, the available weather and measurement data is split up into a training and test dataset to fit the different models of the proposed approach and the benchmark models. Thereby, the results of intraday and day-ahead forecasts are evaluated separately for the deterministic forecasts (using NRMSE and forecast skill) and probabilistic forecasts (based on the energy score and variogram-based scores). The statistical significance of both forecasts is investigated by applying a pairwise Diebold-Mariano test. In order to justify the application of the proposed D-vine copula model, tests for univariate normality have been carried out. For 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 D-vine copula model, a multivariate normal

distribution and, in particular, a product of univariate normal distributions. In addition, even though the VARX models are already taking spatial dependencies into account, there is still a statistically significant interdependence in the forecasting errors, hence, justifying the application of D-vine copulas. Furthermore, one of the key findings in this case study appears to be that probabilistic PV forecasts in general are more sensitive to the deterministic models than to the probabilistic models for both, intraday and day-ahead forecasts, respectively.

Further research could take the time-coupling of spatially interdependent PV forecasting errors into account or extend the proposed D-vine copula model to other sources of uncertainty, e.g., wind or hydro power infeed.

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 Bäsmann of N-Energie GmbH for their support by providing the meteorological forecasting data and power-supply measurement data of our case study.

References

Aas, K., Czado, C., Frigessi, A., Bakken, H., 2009. Pair-copula constructions of multiple dependence. Insurance: Mathematics and Economics 44, 182–198.

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

Almeida, M.P., Perpiñán, O., Narvarte, L., 2015. PV power forecast using a nonparametric PV model. Solar Energy 115, 354–368.

Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-de Pison, F., Antonanzas-Torres, F., 2016. Review of photovoltaic power forecasting. Solar Energy 136, 78–111.

Bacher, P., Madsen, H., Nielsen, H., 2009. Online short-term solar power forecasting. Solar Energy 83, 1772–1783.

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M., Reinhardt, T., 2011. Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities. Monthly Weather Review 139, 3887–3905.

Baumgartner, F.P., Schmidt, H., Burger, B., Bruendlinger, R., Haeberlin, H., Zehner, M., 2007. Status and relevance of the DC voltage dependency of the inverter efficiency, in: Proceedings of the 22nd European Photovoltaic Solar Energy Conference and Exhibition, pp. 2499–2506.

Becker, R., 2017. Generation of time-coupled wind power infeed scenarios using pair-copula construction. IEEE Transactions on Sustainable Energy 9, 1298–1306.

Bessa, R., Möhrlen, C., Fundel, V., Siefert, M., Browell, J., Haglund El Gaidi,

S., 2017. Towards improved understanding of the applicability of uncertainty forecasts in the electric power industry. Energies 10, 1402.

Bessa, R.J., Trindade, A., Silva, C.S., Miranda, V., 2015. Probabilistic solar power forecasting in smart grids using distributed information. International Journal of Electrical Power & Energy Systems 72, 16–23.

BNetzA, 2020. Marktstammdatenregister. URL: `http://www.marktstammdatenregister.de/MaStR`. accessed: 2020-04-28.

Bouzerdoum, M., Mellit, A., Pavan, A.M., 2013. A hybrid model (SARIMA–SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant. Solar Energy 98, 226–235.

Chan, D., Phillips, J., Phang, J., 1986. A comparative study of extraction methods for solar cell model parameters. Solid-State Electronics 29, 329–337.

Chu, Y., Urquhart, B., Gohari, S.M., Pedro, H.T., Kleissl, J., Coimbra, C.F., 2015. Short-term reforecasting of power output from a 48 MWe solar PV plant. Solar Energy 112, 68–77.

Coimbra, C.F., Kleissl, J., Marquez, R., 2013. Solar Energy Forecasting and Resource Assessment. Elsevier / Academic Press.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business & Economic Statistics 20, 134–144.

Driesse, A., Jain, P., Harrison, S., 2008. Beyond the curves: Modeling the

electrical efficiency of photovoltaic inverters, in: Proceedings of the 33rd IEEE Photovoltaic Specialists Conference, IEEE. pp. 1–6.

Erbs, D., Klein, S., Duffie, J., 1982. Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation. Solar Energy 28, 293–302.

Gigoni, L., Betti, A., Crisostomi, E., Franco, A., Tucci, M., Bizzarri, F., 2018. Day-ahead hourly forecasting of power generation from photovoltaic plants. IEEE Transactions on Sustainable Energy 9, 831–842.

Gneiting, T., Katzfuss, M., 2014. Probabilistic forecasting. Annual Review of Statistics and Its Application 1, 125–151.

Gneiting, T., Raftery, A., 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102, 359–378.

Golestaneh, F., Gooi, H.B., 2017. Multivariate prediction intervals for photovoltaic power generation, in: IEEE Innovative Smart Grid Technologies-Asia (ISGT-Asia), IEEE. pp. 1–5.

Golestaneh, F., Gooi, H.B., Pinson, P., 2016. Generation and evaluation of space–time trajectories of photovoltaic power. Applied Energy 176, 80–91.

Harvey, D., Leybourne, S., Newbold, P., 1997. Testing the equality of prediction mean squared errors. International Journal of Forecasting 13, 281–291.

Hess, R., 2020. Statistical postprocessing of ensemble forecasts for severe weather at Deutscher Wetterdienst. Nonlinear Processes in Geophysics 27, 473–487.

Inman, R.H., Pedro, H.T., Coimbra, C.F., 2013. Solar forecasting methods for renewable energy integration. Progress in Energy and Combustion Science 39, 535–576.

Joe, H., 2014. Dependence Modeling with Copulas. Chapman and Hall/CRC.

Joe, H., Xu, J., 1996. The estimation method of inference functions for margins for multivariate models. URL: https://open.library.ubc.ca/collections/facultyresearchandpublications/52383/items/1.0225985.

Klotz, J., 1967. Asymptotic efficiency of the two sample kolmogorov-smirnov test. Journal of the American Statistical Association 62, 932–938.

Konishi, S., Kitagawa, G., 2008. Information Criteria and Statistical Modeling. Springer.

Kou, Q., Klein, S., Beckman, W., 1998. A method for estimating the long-term performance of direct-coupled PV pumping systems. Solar Energy 64, 33–40.

Laiti, L., Giovannini, L., Zardi, D., Belluardo, G., Moser, D., 2018. Estimating hourly beam and diffuse solar radiation in an alpine valley: A critical assessment of decomposition models. Atmosphere 9, 117.

Liu, B.Y., Jordan, R.C., 1963. The long-term average performance of flat-plate solar-energy collectors: With design data for the U.S., its outlying possessions and Canada. Solar Energy 7, 53–74.

von Loeper, F., Kirstein, T., Idlbi, B., Ruf, H., Heilscher, G., Schmidt, V., 2021. Probabilistic analysis of solar power supply using D-vine copulas

based on meteorological variables. In: S. Goettlich, M. Herty and A. Milde (eds.) Mathematical MSO for Power Engineering and Management, Springer (in print) .

von Loeper, F., Schaumann, P., de Langlard, M., Hess, R., Bäsmann, R., Schmidt, V., 2020. Probabilistic prediction of solar power supply to distribution networks, using forecasts of global horizontal irradiation. Solar Energy 203, 145–156.

Lorenz, E., Scheidsteger, T., Hurka, J., Heinemann, D., Kurz, C., 2011. Regional PV power prediction for improved grid integration. Progress in Photovoltaics: Research and Applications 19, 757–771.

Lou, S., Li, D.H., Lam, J.C., Chan, W.W., 2016. Prediction of diffuse solar irradiance using machine learning and multivariable regression. Applied Energy 181, 367–374.

Lu, Q., Hu, W., Min, Y., Yuan, F., Gao, Z., 2014. Wind power uncertainty modeling considering spatial dependence based on pair-copula theory, in: Proceedings of the IEEE PES General Meeting— Conference & Exposition, IEEE. pp. 1–5.

Marion, B., Kroposki, B., Emery, K., Cueto, J., Myers, D., Osterwald, C., 1999. Validation of a photovoltaic module energy ratings procedure at NREL. Technical Report. National Renewable Energy Laboratory. Golden, Colorado, USA.

Martin, N., Ruiz, J., 2001. Calculation of the PV modules angular losses

under field conditions by means of an analytical model. Solar Energy Materials and Solar Cells 70, 25–38.

McNeil, A.J., Frey, R., Embrechts, P., 2005. Quantitative Risk Management: Concepts, Techniques and Tools. Princeton University Press.

van der Meer, D., Widén, J., Munkhammar, J., 2017. Review on probabilistic forecasting of photovoltaic power production and electricity consumption. Renewable and Sustainable Energy Reviews , 1484–1512.

Miller, L.H., 1956. Table of percentage points of kolmogorov statistics. Journal of the American Statistical Association 51, 111—121.

Nelsen, R., 2006. An Introduction to Copulas. Springer.

Pedro, H.T., Coimbra, C.F., 2012. Assessment of forecasting techniques for solar power production with no exogenous inputs. Solar Energy 86, 2017–2028.

Perez, R., Stewart, R., Arbogast, C., Seals, R., Scott, J., 1986. An anisotropic hourly diffuse radiation model for sloping surfaces: Description, performance validation, site dependency evaluation. Solar Energy 36, 481–497.

Pinson, P., Tastu, J., 2013. Discrimination ability of the energy score. Technical Report 15. Technical University of Denmark.

Reindl, D., Beckman, W., Duffie, J., 1990. Diffuse fraction correlations. Solar Energy 45, 1–7.

Ridley, B., Boland, J., Lauret, P., 2010. Modelling of diffuse solar fraction with multiple predictors. Renewable Energy 35, 478–483.

Royston, J.P., 1983. Some techniques for assessing multivariate normality based on the Shapiro-Wilk W. Journal of the Royal Statistical Society. Series C (Applied Statistics) 32, 121–133.

Royston, J.P., 1993. A pocket-calculator algorithm for the Shapiro-Francia test for non-normality: An application to medicine. Statistics in Medicine 12, 181–184.

Saint-Drenan, Y., Good, G., Braun, M., 2017. A probabilistic approach to the estimation of regional photovoltaic power production. Solar Energy 147, 257–276.

Saint-Drenan, Y.M., Bofinger, S., Fritz, R., Vogt, S., Good, G., Dobschinski, J., 2015. An empirical approach to parameterizing photovoltaic plants for power forecasting and simulation. Solar Energy 120, 479–493.

Schaumann, P., de Langlard, M., Hess, R., James, P., Schmidt, V., 2020. A calibrated combination of probabilistic precipitation forecasts to achieve a seamless transition from nowcasting to very short-range forecasting. Weather and Forecasting 35, 773–791.

Schermeyer, H., Vergara, C., Fichtner, W., 2018. Renewable energy curtailment: A case study on today's and tomorrow's congestion management. Energy Policy 112, 427–436.

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

Schinke, A., Hirsch, H., 2019. 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. pp. 1–5.

Shapiro, S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). Biometrika 52, 591–611.

Skartveit, A., Olseth, J.A., 1987. A model for the diffuse fraction of hourly global radiation. Solar Energy 38, 271–274.

Tastu, J., Pinson, P., Madsen, H., 2013. Space-time scenarios of wind power generation produced using a Gaussian copula with parametrized precision matrix. Technical Report 14. Technical University of Denmark.

Torres, J., De Blas, M., García, A., De Francisco, A., 2010. Comparative study of various models in estimating hourly diffuse solar irradiance. Renewable Energy 35, 1325–1332.

Wagner, A., Bendel, C., 2003. Photovoltaic measurement relevant to the energy yield, in: Proceedings of the WCPEC-3 World Conference on Photovoltaic Energy Conversion, Osaka, Japan.

World Radiation Monitoring Center, 2019. Data warehouse. URL: `www.bsrn.awi.de/data/data-retrieval-via-pangaea/data-warehouse/`. accessed: 2020-04-28.