Applied Energy 302 (2021) 117498

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Probabilistic forecasts of the distribution grid state using data-driven forecasts and probabilistic power flow

Jorge Ángel González-Ordiano [a,b], Tillmann Mühlpfordt [a], Eric Braun [a], Jianlei Liu [a], Hüseyin Çakmak [a], Uwe Kühnapfel [a], Clemens Düpmeier [a], Simon Waczowicz [a], Timm Faulwasser [a,c], Ralf Mikut [a], Veit Hagenmeyer [a,*], Riccardo Remo Appino [a]

[a] Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, Germany
[b] Departamento de Estudios en Ingeniería para la Innovación, Universidad Iberoamericana Ciudad de México, Mexico
[c] Institute of Energy Systems, Energy Efficiency and Energy Economics, TU Dortmund University, Germany

ARTICLE INFO

Keywords:
Probabilistic forecasts
Probabilistic power flow
Distribution grid
Uncertainty quantification

ABSTRACT

The uncertainty associated with renewable energies creates challenges in the operation of distribution grids. One way for Distribution System Operators to deal with this is the computation of probabilistic forecasts of the full state of the grid. Recently, probabilistic forecasts have seen increased interest for quantifying the uncertainty of renewable generation and load. However, individual probabilistic forecasts of the state defining variables do not allow the prediction of the probability of joint events, for instance, the probability of two line flows exceeding their limits simultaneously. To overcome the issue of estimating the probability of joint events, we present an approach that combines data-driven probabilistic forecasts (obtained more specifically with quantile regressions) and probabilistic power flow. Moreover, we test the presented method using data from a real-world distribution grid that is part of the Energy Lab 2.0 of the Karlsruhe Institute of Technology and we implement it within a state-of-the-art computational framework.

1. Introduction

The number of distributed energy resources currently being installed in distribution grids is constantly growing [1]. This is particularly true for distributed generation systems based on renewable energy resources (RESs). While these systems increase the sustainability of the overall energy grid, they also complicate the operational duties of Distribution System Operators (DSOs). In particular, DSOs have to manage unprecedented bidirectional uncertain power flows, which are caused by the growing correlation between electric generation/consumption and uncertain phenomena, such as weather conditions. Traditional operation based on an "install-and-forget" strategy with low levels of monitoring may not be sufficient to guarantee a safe and reliable operation of the distribution grid: overvoltages can arise due to photovoltaic overgeneration at some nodes [2], the transient stability of the grid might be affected by the intermittent nature of some RESs [3], etc. Thankfully, distributed energy resources and energy storage systems may allow DSOs to more actively manage the distribution grid. For instance, they could use inverters and electric vehicles to regulate reactive power and active power injections, respectively [4,5]. However, in order to make use of these new tools in an efficient manner, DSOs need to know the grid state. In this context, the DSOs' operational tasks will profit from reliable estimates of the future state of the grid and the uncertainty margins around them. To be more specific, this means giving DSOs a probabilistic forecast of all the variables that define the state of a distribution grid; i.e. the current at every branch, and the voltage and active/reactive power injections at every node. The DSOs can then use this information to forecast operations close to the grid limits, which will allow them to act accordingly in order to avoid critical situations.

In the past couple of years, probabilistic forecasts [6] have been deemed useful at describing the uncertainty of both future load and renewable generation. Probabilistic forecasts are generally divided in two main categories; namely, parametric and non-parametric forecasts [7]. The first type assumes that future values follow a given parametric distribution, while the second type does not require this assumption. The absence of this critical assumption has driven non-parametric probabilistic forecasts to popularity, as shown, for instance, during the Global Energy Forecasting Competition of 2014 [8,9].

* Corresponding author.
E-mail address: veit.hagenmeyer@kit.edu (V. Hagenmeyer).

https://doi.org/10.1016/j.apenergy.2021.117498
Received 25 January 2021; Received in revised form 20 July 2021; Accepted 28 July 2021
Available online 9 August 2021

Unfortunately, the mere calculation of individual probabilistic forecasts of the state variables does not allow the estimation of the probability of joint events, such as the line flows surpassing their limit at the same time. This is due to the fact that these forecasts do not provide information about the correlation/dependence of the variables. In principle, one could solve this issue by using a copula function [10] to estimate the joint probability distribution of the forecast values. However, doing this is not trivial. For instance, parametric copulas require us to assume a copula type and to estimate some parameters, while empirical copulas are more time consuming to calculate [11]. Therefore, we propose in the present paper a new alternative that is not only able to obtain probabilistic forecasts of the grid state, but also allows the estimation of joint event probabilities, without using a joint probability distribution. This new alternative will allow DSOs to forecast operations close to grid limits more easily, which in turn can help them avoid future critical situations (such as an overload of the grid's branches).

The new method we present combines non-parametric probabilistic forecasts of the active power injection of all nodes with probabilistic power flow (PPF) [12]. We compute non-parametric probabilistic forecasts of the active power at each node using quantile regressions [13] obtained via the method described in [14]. Afterwards, we propagate the power uncertainty to the remaining variables by solving the PPF problem. Interested readers are referred to the review articles found in [15–17] for more information on probabilistic power flow and to [18,19] for examples of how quantile regressions have been used to forecast energy related values. Note that non-parametric probabilistic forecasts may lead to arbitrary probability distributions. Thus, the scheme used to solve the PPF problem has to be able to cope with general and time-varying distributions. Sampling-based approaches are straightforward methods to tackle this challenge—with the Monte Carlo Simulation being one of the most widely used, as in [20–22]. However, sampling-based approaches can be computationally demanding. Therefore, for the sake of computational tractability, we solve the PPF problem via the analytic approach described in [23]. This method is chosen since it allows for generic probability distributions without resorting to sampling. Finally, it is important to mention that the presented method can easily be expanded to include probabilistic forecasts of the reactive and active power as input for solving the PPF. This would only require a description of the correlation between the active and the reactive power, yet the workflow of the presented method would remain unchanged.

The main difference between the present article and other works that also apply PPF to grids with renewable energies (such as [24–28]) is not only that the presented method combines quantile regressions with an analytic approach for solving PPF, but also the description of a workflow that goes from data collection to the visualization of the results. To the best of the authors' knowledge, a work such as the one presented is not to be found in literature. To be more specific, the main contributions of the present article are twofold. First, we describe a novel algorithm to obtain probabilistic forecasts of the state of a distribution grid starting from a set of available data (e.g., past realizations of the active power injections, weather forecasts, etc.) and a model of the distribution grid (topology and line parameters). A crucial aspect of the algorithm is the data processing that is necessary to obtain the parameters of the tractable PPF problem from the outputs of the non-parametric probabilistic forecasting models. Second, we implement the presented method using data from the distribution grid at the Karlsruhe Institute of Technology (which forms part of the Energy Lab 2.0 [29]) and a state-of-the-art computational framework. This real-world use case not only allows the validation of the presented method and its implementation, but it also provides a benchmark for future works.

The remainder of the article is divided as follows: Section 2 describes the presented method and its implementation in more detail. Section 3 presents a validation experiment, while Section 4 shows and discusses the results. Finally, Section 5 offers the conclusion and outlook of the present work.

2. Method

2.1. Outline

As mentioned in the introduction, the goal is to estimate the future state of a distribution grid in a probabilistic manner. In other words, we want to compute probabilistic forecasts of:

1. the injected/adsorbed active and reactive power at each node;
2. the absolute value of the voltage at every node; and
3. the absolute value of the current at every branch.

Note that we model all previous values as random variables, as we consider them uncertain. In the present article, we use a sans-serif notation to distinguish between a (complex/real-valued) random variable y and its (complex/real) realization y.

To achieve our goal, the article presents an approach that combines data-driven non-parametric probabilistic forecasts of the active power injections of all nodes with probabilistic power flow. Compared to other uncertainty estimation approaches (such as interval forecasts), this method has two main advantages: (i) it estimates the cumulative distribution function (CDF) of the state defining variables instead of intervals, quantiles, etc. and (ii) it enables us to calculate the probability of joint events – for instance, the probability of the voltage of different nodes surpassing a certain value at the same time – without having to compute a joint probability distribution. Note that the presented method is based on the following assumptions:

1. historical measurements of the nodes' active power need to be available;
2. the network must be radial[1] (as in [30]) and its topology and line parameters must be known; and
3. the independence of the injected power values, but not of the electrical currents and voltages.

The third assumption stems from the fact that calculating the joint distribution of all injected power values would be out of the scope of the present article, as it is a non-trivial task. Nevertheless, an advantage of the proposed approach is that removing this assumption would not drastically change it, as the method would only require the dependence structure of the injected power values as additional information.

In the remainder of Section 2, the general requirements and steps of the proposed approach are presented. For the sake of clarity, Fig. 1 depicts the implemented workflow. This workflow includes not only the presented method, but also additional steps that are necessary for the collection of the input data and for the visualization of the results. Furthermore, every step of the presented approach is implemented within a computational framework [31] that simplifies automation and guarantees scalability. Note that in this paper, the obtainment of active power forecasts is considered the first step of the presented method. Nevertheless, the presented approach can easily be expanded to include the calculation of reactive power forecasts within the first step. Doing this would only require small changes within the individual steps (for instance within the compatibility step) and a description of the correlation between the active and reactive power, yet the workflow shown in Fig. 1 would, for the most part, remain unchanged.

Fig. 1 shows that the first step in the presented workflow consists of collecting all the necessary information that the presented method uses as input: active power measurements, weather data, etc. To collect this data we use the *e*ASIMOV framework [32]. More information on the data collection step can be found in Section 2.2.

Afterwards, the collected data is propagated to the presented method, which is divided into three steps, see Fig. 1. The first step uses the collected data to obtain probabilistic forecasts of the future active

[1] PPF can also be applied to meshed networks; the idea of the presented method is not affected.

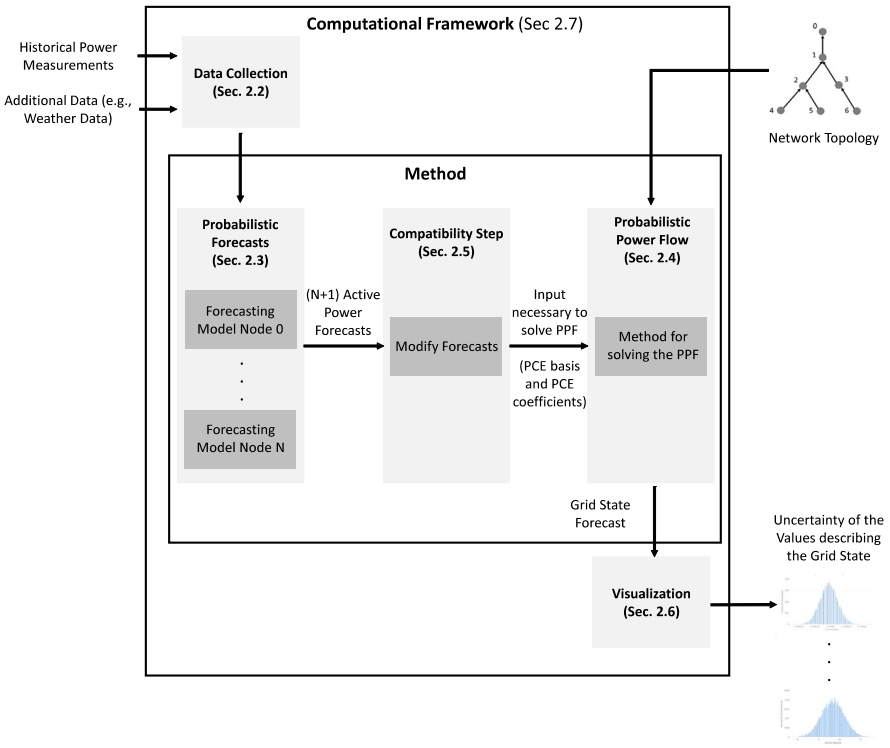

Fig. 1. Depiction of the workflow used to obtain probabilistic forecasts of the grid state showing the input and output of every step of the new method; PPF: probabilistic power flow; PCE: polynomial chaos expansion.

power at each node. These forecasts are obtained using a data-driven approach based on quantile regression and a method described in [14]. The second step ensures that the first and the last step of the presented method are compatible. This is important, since the output of the first step is not the input that the third step requires. Finally, the third step solves the probabilistic power flow (PPF) problem using the output of the second step, the network topology, and the approach described in [23]. The result of all of these steps is the desired probabilistic forecast of the full state of the distribution grid. Readers interested in the initial probabilistic forecasts and in the method for solving the PPF problem are referred to Sections 2.3 and 2.4, respectively. Furthermore, Section 2.5 offers the specifics on how the compatibility between the first and last step of the presented method is achieved.

As Fig. 1 shows, the grid state forecasts are propagated to a further step that allows for their visualization. The *e*ASIMOV framework is again used in this step. More information on the visualization can be found in Section 2.6.

Finally, Fig. 1 also shows the complexity of the workflow, as it consists of a combination of sequential steps that need to be automated and that are implemented using different software tools. Therefore, it is important for us to implement the workflow using a computational framework that not only simplifies its automation, but also helps in ensuring its scalability. The use of such a computational framework is also depicted in Fig. 1. Furthermore, readers are referred to Section 2.7 for more information on the computational framework used.

2.2. Data collection

As mentioned in Section 2.1, the *e*ASIMOV framework [32] is used to collect and access the necessary input data. To be more specific, the data can be extracted in the form of power consumption/generation time series through the framework's *e*POWWEB module. This module preprocesses the data to increase its quality and makes it accessible at various aggregation levels and time periods through a service featuring a REST interface.

2.3. Probabilistic forecasts

As discussed in the introduction, non-parametric probabilistic forecasts have become quite popular in the literature, since they avoid assumptions about the future value's distribution. Within the existing methods, quantile regression [13,18,19] is one of the most preferred and the one used in the present article. As the name suggests, quantile regression estimates a quantile of its output value given the regression's input [33]. A quantile regression can, for instance, be described as follows:

$$\hat{y}_\tau = f(x) \text{ with,}$$
$$\Pr(\mathrm{y} \leq y_\tau | x) = \tau; \tag{1}$$

where y is the desired output described as a random variable, y_τ is its $\tau-$quantile, $\tau \in (0,1)$ is a probability value, $\hat{y}_\tau$ is the estimate given by the regression model, and x is the model's input.

This type of regression can be used to create probabilistic forecasts by considering y as a representation of a future value of interest (for instance, load, renewable power, demand response load, etc.). Then, quantile regression models can be combined to estimate intervals with a probability of the future value lying within or to estimate the CDF of y [34].

In the present article, we use quantile regression to forecast the quantiles of the active power that is injected on every node.

2.4. Probabilistic power flow

Solving the power flow problem means to compute all nodal voltages and all branch currents given the complex power injections at each node and the characteristics of the grid (i.e. topology and line parameters). The relation among these variables is governed by the well-known power flow equations [35]. Probabilistic power flow extends the traditional power flow to the case in which the power injections are modeled as random variables. Thus, solving the probabilistic power flow problem means finding all nodal voltages and all branch currents

in terms of random variables such that the power flow equations are satisfied for each realization of the random variables. Solving the PPF poses two major challenges:

1. rendering the problem computationally tractable, and
2. solving the (deterministic) power flow equations—which in general do not have a closed analytic form.

In order to solve the PPF problem, the present article uses the method described in [23]. This method is a deterministic reformulation of the stochastic problem that can cope with non-Gaussian random variables under the condition of having a radial grid. The method utilizes polynomial chaos expansion (PCE) [36] to render the problem deterministic, and it employs the backward–forward-sweep (BFS) [37] scheme to solve the power flow equations.

Polynomial chaos is a Hilbert space method used in the field of uncertainty quantification [36]. It is to a random variable what a Fourier series is to a periodic signal: a method that allows to characterize a random variable by a finite number of Fourier coefficients. Mathematically, this reads

$$\mathrm{x} = \sum_{\ell=0}^{L} x_\ell \phi_\ell, \tag{2}$$

where ϕ_ℓ denotes the ℓth basis function, which is both a polynomial function and a random variable.[2] The main assumption of PCE is that all random variables are required to have a finite variance—a fairly mild assumption. The main advantages of using PCE in probabilistic power flow are:

1. applicability to Gaussian and/or non-Gaussian random variables or combinations thereof (e.g., mixed Gaussian distributions);
2. efficient uncertainty propagation through model equations by means of Galerkin projection; and
3. computation of moments of random variables without having to sample.

It is important to mention that Galerkin projection is only one of multiple methods to propagate uncertainties, other approaches such as non-intrusive methods are applicable too [36]. Using PCE, the probabilistic power flow problem can be redefined as follows: *determine all* PCE *coefficients of all random variables such that the power flow equations are satisfied in terms of random variables*.

BFS is an iterative method that exploits the networks' tree structure to obtain a solution of the power flow equations [37]. The BFS consists of the recursive application of Kirchhoff's voltage and current laws. As the name suggests, the approach is divided in a backward and a forward step. The backward step computes the nodal currents – starting from the power injections and the nodal admittance – using values of the nodal voltages from a previous iteration. Afterwards, the branch currents are calculated using the network's topology and branch parameters. In turn, the forward step uses the currents and admittance of a given branch to compute its voltage drop. Given the fixed and known voltage at the root node and the radial topology, the nodal voltages can be calculated from these voltage drops. These nodal voltages are then used to update the values of the nodal voltages which will be used in the next iteration. The algorithm is repeated until convergence is achieved up to a prescribed numerical tolerance[3]. The main advantage of this method is that, contrary to the Newton–Raphson method, it does not require the inversion of the power-flow equations' Jacobian, which may pose computational issues [38].

The algorithm to solve the probabilistic power flow combining the polynomial chaos expansion approach with a backward–forward-sweep scheme is summarized in Fig. 2. Just as with the deterministic BFS, each iteration begins with a backward step for the calculation of the branch currents. However, this backward step is subdivided in two: a first step for calculating the PCE coefficients of the (random) current injections at each node and a second one to compute all PCE coefficients of the (random) branch currents. Afterwards, the forward step is applied.

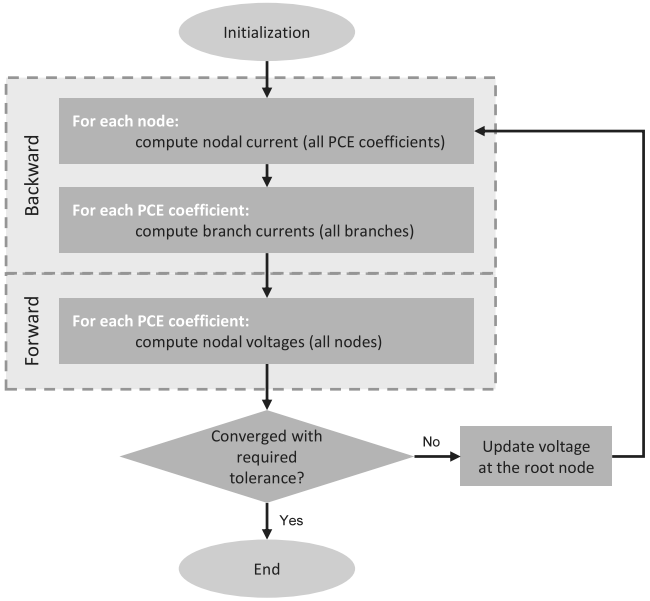

Fig. 2. Backward–forward-sweep-algorithm used to solve the probabilistic power flow (PPF).

Note that this algorithm requires the following inputs to work:

- the characteristics of the analyzed grid (network topology and line parameters);
- a PCE basis $\{\phi_0, \ldots \phi_L\}$; and
- the PCE coefficients for the random power injections (i.e. active and reactive power).

As depicted in Fig. 1, the characteristics of the analyzed grid are one of the initial inputs of the forecasting scheme, while the remaining inputs have to be deduced from the outputs of the forecasting models discussed in Section 2.3. To this end, some data processing is required, which is detailed in the next section.

2.5. Compatibility step

As mentioned in the previous sections, the output of the data-driven forecasting models are quantiles of the active power at every node, while the input of the PPF step are the PCE basis and the PCE coefficients of the power injections (i.e. active and reactive) at all nodes. Therefore, we propose an intermediate step (cf. Fig. 1) that renders the data-driven forecasts compatible with the PPF method. For the sake of clarity, Fig. 3 shows how the PCE basis and coefficients are obtained from the active power probabilistic forecasts.

As Fig. 3 shows, the compatibility step consists of four main elements: First, we estimate the analytic CDF of the future active power of every node. As depicted in Fig. 4a, the output of the nth node's quantile regressions can be used as a point-wise estimate of its active power CDF. Therefore, the analytic description of the CDF can be obtained by fitting that initial estimate to an analytic function. For example, Fig. 4b shows an analytic CDF based on the sum of two logistic functions fitted via a numerically solved least squares method. The sum of two logistic functions has been shown to deliver good estimates of the CDF [39] and thus is one of the functions we use herein. To be more specific, we fit the sum of two logistic functions, as well as a single logistic function

[2] For Fourier series, the ℓth basis function would be $\exp(\mathrm{j}\ell t)$.

[3] More information on the convergence of BFS can be found in [30].

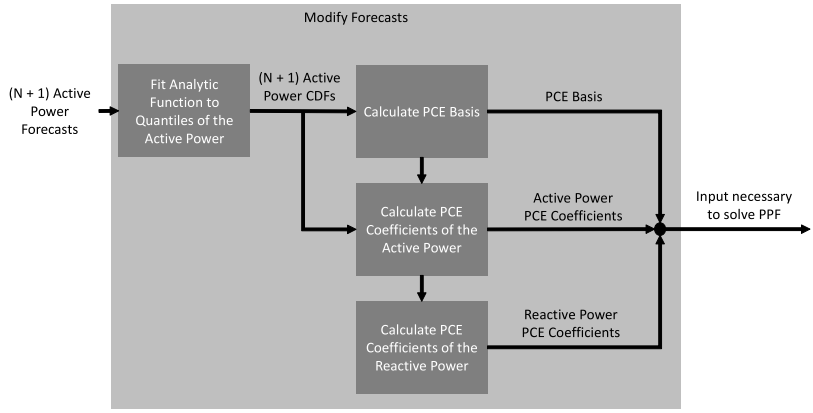

Fig. 3. Calculation of PCE basis and coefficients from the active power probabilistic forecasts.

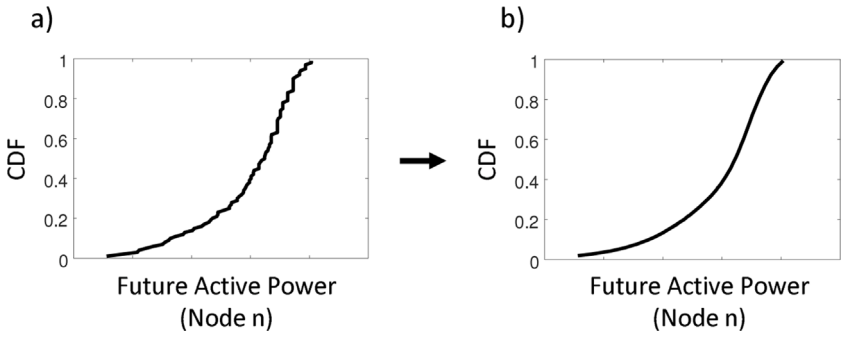

Fig. 4. Example of fitting the outputs of quantile regressions to an analytic cumulative distribution function (CDF).

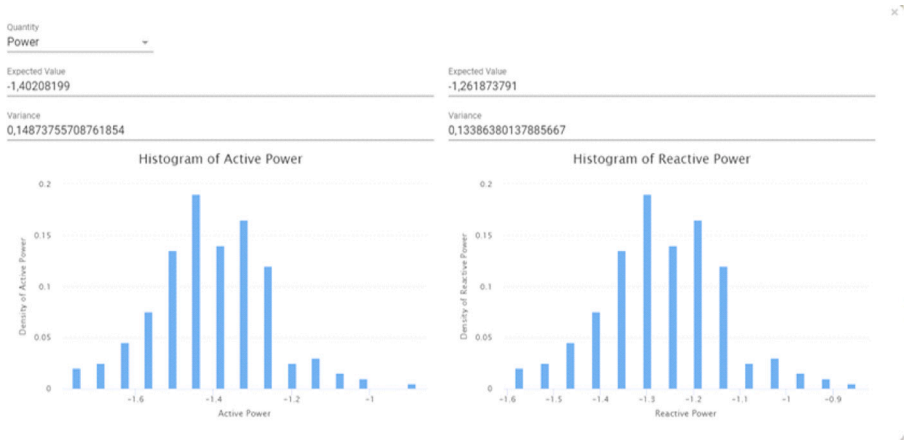

Fig. 5. Visualization of the probabilistic power flow results using *e*powvis. The histograms are formed using values sampled from the distributions obtained via the presented method.

to the quantiles and choose the best fit as the analytic CDF. The CDF described as a single logistic function is given as

$$\mathrm{F}(\mathrm{y}|x) = \frac{\theta_1}{1 + e^{-\theta_2(\mathrm{y}-\theta_3)}}\ , \tag{3}$$

while the CDF described as the sum of two logistic functions can be written as

$$\mathrm{F}(\mathrm{y}|x) = \frac{\theta_1}{1 + e^{-\theta_2(\mathrm{y}-\theta_3)}} + \frac{\theta_4}{1 + e^{-\theta_5(\mathrm{y}-\theta_6)}}\ . \tag{4}$$

In the previous equations, the values θ_i are the parameters that we fit via the least squares method, while $\mathrm{F}(\mathrm{y}|x)$ is the analytic CDF of the forecast value y (i.e. a node's active power), given the forecasting models input x. In addition, the algorithm models a variable as deterministic if it considers the spread of its CDF to be too small. To be more specific, if the difference between a variable's 0.95 and 0.05-quantile is smaller than 15%[4] of its median, the forecast uncertainty is ignored and the median becomes that variable's point (i.e. deterministic) forecast.

It is important to mention that estimating the analytic CDF does not require that we assume a specific parametric distribution (for instance, a normal distribution). This means that we retain the main advantage of using non-parametric probabilistic forecasts: no specific type of distribution (with finite variance) is assumed.

Based on an analytical expression of the CDF of the active power for each node, the next step is to compute an orthogonal polynomial basis—this is required for the PCE approach we employ to solve PPF.

[4] The 15% is chosen based on numerical studies.

There exist stable numerical routines to construct polynomials that are orthogonal to any given probability density function (which is the derivative of the CDF). The Julia package `PolyChaos.jl` provides exactly this functionality and is hence used [40]. `PolyChaos.jl` also facilitates to obtain the numerical values of the PCE coefficients (a crucial step in the presented method). After estimating the active power PCE coefficients using `PolyChaos.jl`, we calculate the coefficients of the reactive power by taking the active power coefficients and multiplying them with a scaling factor based on the power triangle.

2.6. Visualization

Similarly to the data collection step, the visualization step also makes use of one of the modules of the *e*ASIMOV framework [32], namely the *e*POWVIS module. This module is able to provide an interactive web-based visualization of a distribution grid. Additionally, the module also provides the possibility of clicking on the grid nodes and branches to open a new window showing the corresponding results of the probabilistic power flow step. For instance, Fig. 5 shows histograms that represent the distribution of both the active and reactive power at a given node. This histograms are obtained from values sampled from the distributions estimated using the presented method. Note that in this final step we can also implement transformations to go from rectangular to polar descriptions.

It is important to mention that *e*ASIMOV offers additional visualization options that can later be implemented within the presented method. For instance, grid visualizations that can help with the analysis of future grid states (including critical states). For the sake of illustration, Fig. 6 shows a visualization of the grid used for the present article's experiment that depicts the absolute value of the nodes' voltages.

Both Fig. 6a and b show the same grid state, but define the colors for the voltages using different scales. Fig. 6a uses a scale that shows that no voltage limit is being violated. At the same time, Fig. 6b is used to show how a visualization of a future critical state might look like and to demonstrate that *e*ASIMOV allows us to modify the color scaling quite easily, which might simplify the analysis of differences between the state defining variables no matter the magnitude of this differences.

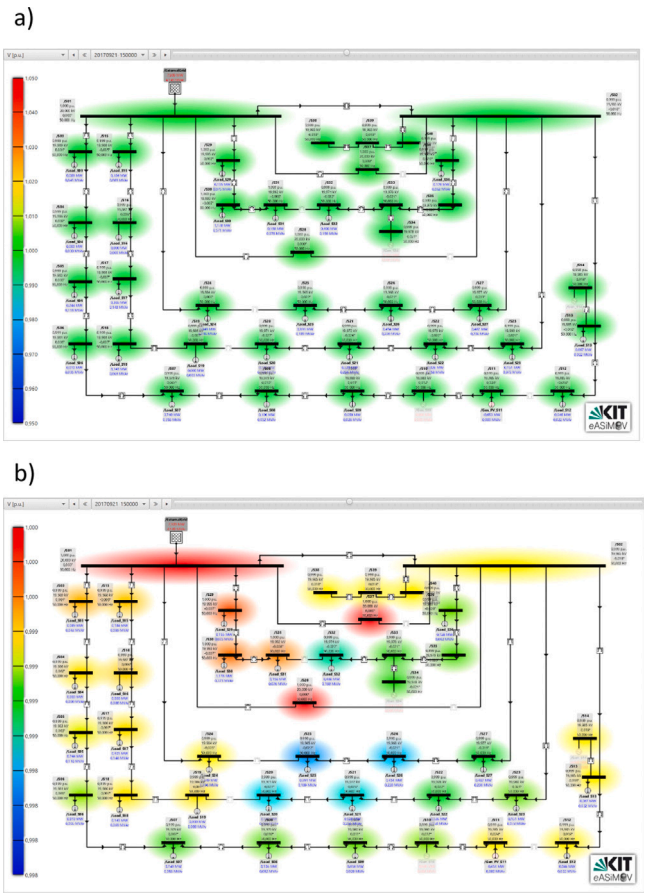

Fig. 6. Example of a grid visualization showing the absolute value of the nodes' voltages; Both images show the same grid state, but with different color scaling for the voltages.

2.7. Computational framework

The presented method is a complex workflow that consists of a combination of steps that need to be conducted in a certain order and that are implemented using a number of software tools.

The complexity of the workflow makes its execution tedious, time-consuming, and error-prone; especially if the workflow has to be executed repeatedly with different input data. Another argument against the manual execution of the workflow is the fact that the output data of one step typically needs to be modified, before it can be used by the following step. For example, information of the grid topology needs to be added to the measurement data that is coming from the database. As mentioned previously, doing this type of data transformation manually at each step is quite tedious and error-prone. Therefore, the computational framework described in [31] is used to automate the workflow, presented herein.

The computational framework uses state-of-the-art technologies for the run-time automation and coordination of sub-tasks, such as container virtualization and distributed message oriented middleware. In other words, it defines an architecture for distributed process execution and coordination. The framework automatically distributes software applications that perform different tasks to different nodes within a cluster. These applications can access necessary data and communicate with other components using the framework's communication adaptors and high-performance messaging channel infrastructure. Additionally, the framework provides an easy-to-use web user interface based on Apache NIFI that allows users to create, run, and control workflows regardless of their understanding of the underlying computing infrastructure. For the sake of illustration, Fig. 7 shows how the method has been implemented within the computational framework described in the present paper.

Fig. 7 shows that the workflow implemented within the computational framework includes an additional step between the compatibility and the PPF steps. This step has the task of waiting on the forecast of several nodes to communicate them to the PPF step as a batch. This is needed by the algorithm solving the PPF. Note that an advantage of the framework used is that – through a simple configuration – certain tasks can be executed in parallel. Hence, an optimal performance is achieved when using a computing cluster. The parallelization is realized by the framework's coordination service, which allows for parallel data processing and step execution using separate Docker containers with different input values.

Readers interested in the technical details of the computational framework are referred to [31].

3. Experiment

3.1. Test case

In the present article, we use a physically existing distribution grid as a use case of the implementation described in Section 2. The grid in question is the grid of the north campus of the Karlsruhe Institute of Technology (KIT), which is operated as a radial grid and thus fulfills the assumptions of the proposed approach. The goal of this experiment is to forecast the KIT's hourly grid state one, two, and three hours in

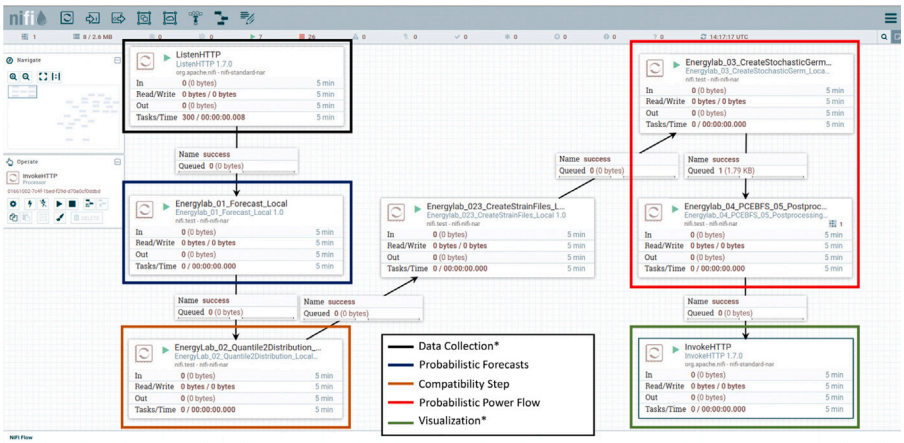

Fig. 7. Implementation of the present article's method within the computational framework used; ∗ Both the data collection and visualization steps are represented by the blocks that connect to *e*asiMov to extract data or to save data for visualization.

advance at 12 PM (i.e. the grid state at 1 PM, 2 PM, and 3 PM) on three different days. To be more specific, the days for which we conduct the forecasts are: August 10th, 2017, September 10th, 2017, and October 10th, 2017.

The grid consists of 38 individual nodes that are divided in two main groups: 11 nodes are classified as *certain* and 27 as *uncertain*. The so-called certain nodes are loads without historical measurements, generators, and slack buses that are considered to have deterministic power injections. In contrast, the so-called uncertain nodes are either loads or renewable generators for which we need to obtain models that forecast the uncertainty of their future active power. Note that these models are trained using the nodes' 2016 hourly active power measurements. We use hourly measurements, because we want to obtain hourly forecasts. However, the presented method is not limited to hourly forecasts, we can change the resolution (if required for scheduling/control reasons) without changing the structure of the method. The only thing we need to do is to train the quantile regression models (cf. Section 2.3) using data that matches the new desired resolution.

After inspecting the historical data of all uncertain nodes, we divide them in two additional groups. The first contains 6 nodes with great amounts of missing historical data, while the second contains 21 nodes for which the historical data is sufficient to train quantile regressions (cf. Section 2.3). Even though we do not train quantile regressions for the 6 nodes of the first group we still quantify the uncertainty of their hourly active power by calculating 99 empirical quantiles ($\tau = 0.01, 0.02, \ldots, 0.99$) of their historical measurements. For instance, the uncertainty of a node's active power at 1 PM is represented by the empirical quantiles of historical values measured at 1 PM. For the nodes within the second group we train a series of quantile regressions that estimate 99 different quantiles ($\tau = 0.01, 0.02, \ldots, 0.99$) of the nodes' active power one, two, and three hours in advance using as input the active power measurements of the past two days. These regressions are all polynomials of maximum degree two trained using the MATLAB open source toolbox SCIXMINER [41] and the method described in [14]. As previously mentioned, the PPF uses as input the PCE coefficients and bases of the future active and reactive power. Therefore, we use the approaches described in Section 2.5 to transform the active power quantiles into the input that is necessary to solve PPF. Note that the scaling factor used to estimate the reactive power is 0.4843 (cf. Section 2.5).

After calculating the necessary input, the method for solving the PPF can be applied. In the present article, we use the Julia package `PolyChaos.jl` [40]. After solving the PPF, we obtain – in addition to the injected power forecasts – probabilistic forecasts of the absolute value of the voltage and current at every node and branch, respectively. Since we are estimating these forecasts based on power values of the next 3 h (i.e. the forecast horizons of the quantile regression models used), the forecasts we obtain are a probabilistic description of the grid state one, two, and three hours ahead.

3.2. Forecast evaluation

To visualize (as in Fig. 5) and evaluate the forecasts, we draw 200 random samples of the grid voltages, currents, and injected power values. The forecast evaluation is based on the commonly used pinball-loss metric—a metric used to evaluate the accuracy of quantile forecasts. To use this metric we calculate 99 quantiles ($\tau = 0.01, 0.02, \ldots, 0.99$) of the random samples of each value. Afterwards, we use the following equation to calculate the pinball-loss:

$$Q_\tau = \frac{1}{\#\{\mathcal{G}\}\ N_\mathrm{S}} \sum_{n\in\mathcal{G}} \sum_{i=1}^{N_\mathrm{S}} \frac{(y_{ni} - \hat{y}_{ni,\tau})}{f_{\mathrm{Norm},n}} (\tau - I(y_{ni} < \hat{y}_{ni,\tau})) \ , \tag{5}$$

where N_S is the number of days for which we obtain forecasts (i.e. 3) and $\mathcal{G}$ is a set of the grid's nodes when calculating the error for the voltage and power forecasts or the set of branches when evaluating the forecasts of the electrical current. At the same time, y_{ni} represents the true power, voltage, or current, $\hat{y}_{ni,\tau}$ is the values' corresponding τ-quantile forecast, $\#\{\cdot\}$ is the cardinality operator, and $I(\cdot)$ is an indicator function that equals 1 if its condition is fulfilled and 0 otherwise. Finally, $f_{\mathrm{Norm},n}$ is a value used to make the errors of the different buses and branches comparable. To explain why this value is important, consider the following example. Imagine two nodes, one with power always around 100 kW and another with power always around 20 kW. If we forecast the power of both nodes and make a mistake of 10 kW, the magnitude of the error – relative to their nominal power – will be completely different for each node. Therefore, to evaluate the accuracy of the grid state forecast, we need to re-scale the errors to make them comparable. In the present article, we set $f_{\mathrm{Norm},n}$ for the active/reactive power and current equal to the difference between the maximal and minimal measurements of the values during the period used as training data (i.e. 2016). Note that for the voltage $f_{\mathrm{Norm},n}$ is set equal to 1, as all voltages are measured and forecast in per-unit values; values that in our case are always around 1. For this reason, the errors of the nodes' voltage forecasts are already comparable and thus do not need to be re-scaled. The previous information, i.e. the values that the variables in Eq. (5) take based on the forecasts being evaluated (i.e. power, voltage, or current), is summarized in Table 1.

Note that the current, voltage, and reactive power values used to evaluate the forecasts in Eq. (5) are not measurements. They are, instead, values obtained using the active power measurements and a grid simulation.

In addition, to compare the accuracy of the presented method to a more classical approach (i.e. one that directly estimates the uncertainty of the state defining variables), we train 99 quantile regressions ($\tau = 0.01, 0.02, \ldots, 0.99$) that directly estimate the quantiles of the currents one, two, and three hours in advance at every branch without taking

Table 1
Definitions of the values used in Eq. (5) to evaluate the power, voltage, and current forecasts.

y_{ni}	$\mathcal{G}$	$f_{\text{Norm},n}$
Voltage absolute value	Grid nodes	1
Current absolute value	Grid branches	Difference between the maximal and minimal current of the nth branch in 2016
Active power	Grid nodes	Difference between the maximal and minimal active power of the nth node in 2016
Reactive power	Grid nodes	Difference between the maximal and minimal reactive power of the nth node in 2016

into consideration the variables' correlation. Afterwards, we use these regressions to forecast the quantiles of the current absolute values for the same three days and hours on which we tested the presented method. We then calculate the pinball-loss of the quantiles (cf. Eq. (5)) and compare the results to those obtained by the presented method's current absolute value forecasts. Note that these quantile regressions have the same structure as the ones used for the power. They are polynomials of maximal degree two that take the current measurements of the past two days as input and that are trained using the MATLAB open source toolbox scixminer [41] and the method described in [14].

The presented approach also allows us to obtain correlated samples of the state defining variables that we can then use to forecast the probability of joint events, without having to estimate the joint distribution. For this reason as a final part of the experiment, we evaluate how likely it is that the current of two branches connected to the same node surpass a given threshold in the next hour. To do so, we calculate the probability of the two currents surpassing the 1 h ahead 0.9 quantile (estimated from the currents' random samples given by the presented approach) when we consider the currents' correlation/dependence using the presented method, when we assume them to be statistically independent, and when we assume "full dependence". "Full dependence" meaning in this context that if the current on one branch surpasses the threshold, the other will too. This experiment is conducted on two separate pairs of branches.

4. Results and discussion

The results of the one, two, and three hours ahead grid state forecasts (cf. Section 3) are shown in Fig. 8. The pinball-loss is given in percentages of the range of values observed in 2016. The only exception is the voltage pinball-loss, which is depicted in per unit. As explained above, this is due to the fact that the voltages in per unit are always around 1 on every bus, thus making the comparison of the forecast errors possible without the need of rescaling.

As we can observe, the errors of the active and reactive power are all below the 3% mark. At the same time, the pinball-loss curves show a form that is commonly observed, i.e. an inverted parabola that has is maximum value around the 0.5 quantile. The curves for the three hour ahead forecasts seem to be the only exception: they are shifted slightly towards the right. This points at a slight underestimation of some of the larger quantiles. Furthermore, as the reactive power is calculated through a scaling of the active power (cf. Section 2.5), its pinball-loss curves should be the same as the active power ones. However, this is not completely the case as the active and reactive power curves show small differences. The reason for this is that within the KIT grid there is a bus that represents a PV power plant that we assume always has a reactive power of 0. Therefore, we remove this bus when calculating the reactive power pinball-loss curves, causing the small differences that we observe.

Moreover, Fig. 8 shows that the pinball-loss curves of the voltage forecasts have a completely different behavior than all the others. To be more specific, the maximum of the pinball-loss curves is completely shifted to the left. This means that the distribution of the bus voltage uncertainty is being overestimated. However, we can also observe that the error is small, lower than $3 \cdot 10^{-4}$ p.u. In other words, even though there seems to be a systematic difference between the true voltage and its probabilistic forecasts, the difference is so small that we can still consider the forecast to be acceptable. Still, this systematic overestimation is something that needs to be addressed in future related works.

Fig. 8 shows that the pinball-loss values of the current forecast are always under 5%. At the same time we can also observe that the curves are slightly shifted to the left, which is something that might be caused by an overestimation of some of the lower quantiles. Regardless, the results appear to be acceptable as the magnitude of the errors remains small.

To summarize, the previous results show that the presented method is able to provide acceptable probabilistic forecasts of the values that define the grid state (i.e. active/reactive power and the absolute values of the voltages and currents). Nevertheless, we can also observe aspects that still need to be slightly improved in future related works, like the overestimation of the current and voltage forecasts; with the latter being the most extreme case.

Fig. 9 shows the comparison between the current absolute value forecasts obtained with the presented method and those obtained by quantile regressions that are trained to estimate the quantiles of the current absolute values directly. Since these regressions do not take into consideration the correlation between the currents, we refer to them as independent forecasts.

As we can observe in Fig. 9, the independent forecasts are more accurate than those obtained with our approach. The largest pinball-loss of the independent forecasts is around 2.2%, while the largest pinball-loss of the presented method is around 4.7%. This result is not surprising considering that independent forecasts consist of quantile regressions that are specifically trained to estimate the uncertainty of the current absolute values directly, while the presented method estimates that uncertainty indirectly by using the PPF to propagate the uncertainty of the initial active power forecasts to all other state defining variables. Nevertheless, the independent forecasts do not take into consideration the correlation between the variables, in other words, they cannot be used to estimate the probability of joint events, which is something that the presented method can be used for. Therefore, though the presented method delivers accurate forecasts it does show a trade-off between forecast accuracy and the ability to estimate the probability of joint events without the need of a joint distribution. If the latter is important in a given use case, the presented method has a clear advantage over independent forecasts. The ability to estimate the probability of joint events is further discussed in the remainder of this section.

Fig. 10 shows the probability of the current of two branches surpassing a threshold (i.e. the 0.9 quantile of the one hour ahead forecast) at the same time when we consider their correlation/dependence using the presented method, as well as when we assume their independence or their "full dependence". Note that Fig. 10 shows two separate plots, since we conduct this experiment on two separate pairs of branches.

Since we define the threshold as the currents' 0.9 quantile, the probability of two currents surpassing the threshold is 1% if we consider them to be independent and 10% if we assume "full dependence". In contrast, we forecast a probability of 5 and 6% for the first and second pair of branches, respectively, if we estimate the probability of this joint event using the presented approach. This mean that when we consider the variables' correlation/dependence in our example, the future joint events become more likely than when we assume statistical

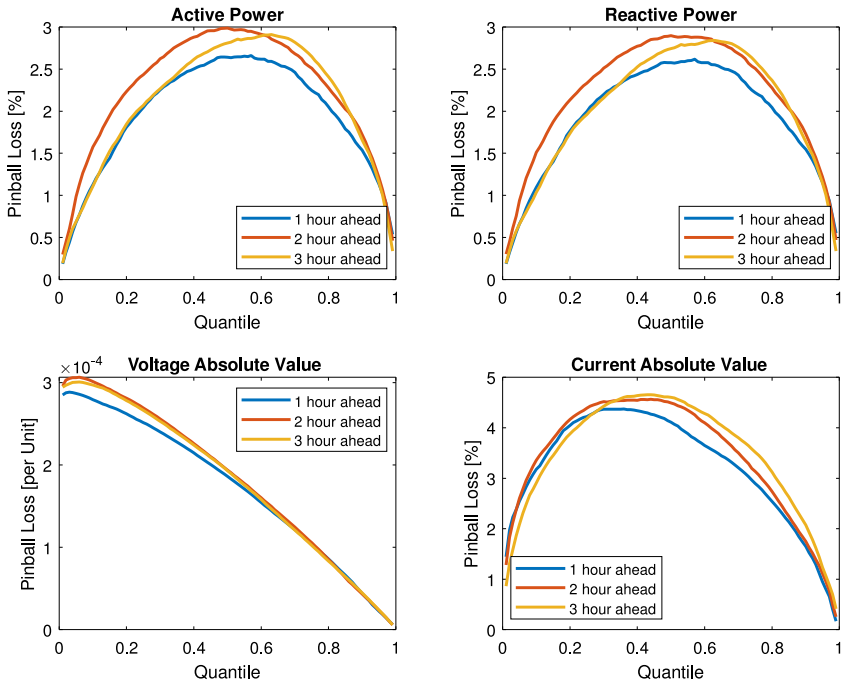

Fig. 8. Pinball-loss of the forecasts of the state defining variables, i.e.: active power; reactive power; voltage absolute value; current absolute value.

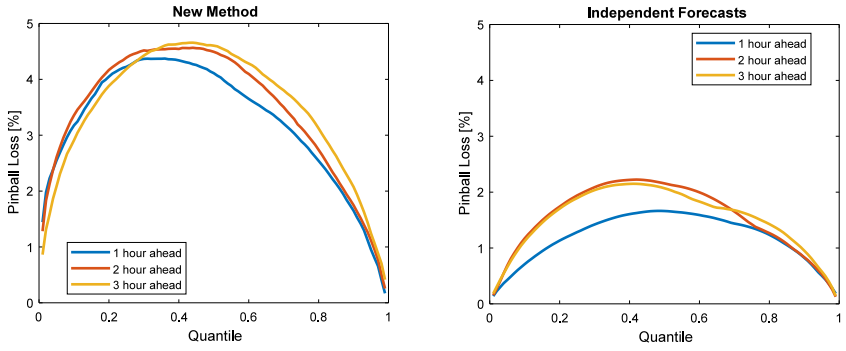

Fig. 9. Pinball-loss of the forecasts of the current absolute values that are obtained by the presented method and by the independent forecasts (i.e. quantile regressions that directly estimate the quantiles of the current absolute values).

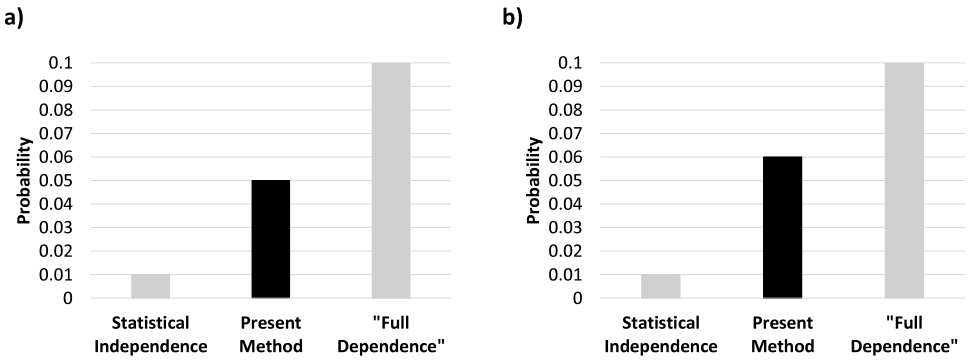

Fig. 10. Probability of pairs of branches surpassing the 0.9 quantile threshold of their one hour ahead current forecast at the same time; (a) pair 1; (b) pair 2.

independence. Being able to correctly estimate the probability of these joint events is of great importance in cases in which those events are critical situations within the grid. For instance, if the current's threshold of the above example represents an overload of the corresponding branches. At the same time, the presented method estimates joint event probabilities that are lower than those of the "full dependence" case. In other words, we are able to obtain (in our example) less conservative estimates of the probability of possible critical situations.

To summarize, the previous results show that the presented method offers information that will allow the DSOs to estimate more accurately the future behavior of the grid. At the same time, this will allow the

DSOs to avoid critical situations, as they will be able forecast operations close to the grid limits more accurately.

5. Conclusion and outlook

We present a novel approach for obtaining probabilistic forecasts of the state of a distribution grid. The presented method is able to forecast the uncertainty of the active/reactive power at every bus, as well as absolute values of the voltage and current at every bus and branch, respectively. This is a non-trivial task as the values might be correlated and thus the forecast of their uncertainty might require the estimation of a multivariate distribution. In order to avoid this, we estimate the full state of the grid using the combination of data-driven probabilistic forecasts of the active power at all buses and an approach for solving the probabilistic power flow problem. Since the presented method considers the correlation/dependence of the state defining variables, it allows an estimation of the probability of joint events, such as the current of two branches surpassing a given value at the same time. In addition, we present an implementation of the presented method using data from the distribution grid of the Karlsruhe Institute of Technology and an experiment to validate the forecasts. The results show that the presented method is able to obtain accurate probabilistic forecasts of the grid state. All computations have been run within a state-of-the-art computational framework. Currently, the presented method is being implemented within the Energy Lab 2.0.

In the future, additional studies to further improve and evaluate the presented method shall be conducted, for instance: (i) a test on real-time data of the grid of the Karlsruhe Institute of Technology; (ii) the research of models with greater forecast horizons and of methods able to remove the slight overestimation of the voltages; (iii) the study of the economic and technical impact that the forecasting errors have on the DSOs and the distribution grid; and (iv) the investigation on how the results of the presented method can be used within a power system planning context. Additionally, since the power generated by renewable sources might show correlation within a given region, an analysis without the assumption of the injected power being independent must also be undertaken. Moreover, the possibility of visualizing future critical states must be implemented within the presented method. Finally, the use of the presented method to forecast the uncertainty in high voltage grids (for instance, to estimate critical situations in northern Germany, due to offshore wind power) shall also be investigated in future related works.

CRediT authorship contribution statement

Jorge Ángel González-Ordiano: Formal analysis, Investigation, Methodology, Software, Validation, Writing – original draft. **Tillmann Mühlpfordt:** Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Writing – review & editing. **Eric Braun:** Data curation, Investigation, Methodology, Software, Validation, Writing review & editing. **Jianlei Liu:** Data curation, Investigation, Methodology, Software, Validation, Writing review & editing. **Hüseyin Çakmak:** Data curation, Investigation, Methodology, Software, Validation, Visualization, Writing – review & editing. **Uwe Kühnapfel:** Writing – review & editing. **Clemens Düpmeier:** Writing – review & editing. **Simon Waczowicz:** Writing – review & editing. **Timm Faulwasser:** Supervision, Writing – review & editing. **Ralf Mikut:** Supervision, Writing – review & editing. **Veit Hagenmeyer:** Funding acquisition, Resources, Supervision, Writing – review & editing. **Riccardo Remo Appino:** Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Writing – review & editing.

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.

Acknowledgments

The present contribution is supported by the Helmholtz Association under the Joint Initiative “Energy System 2050, Germany – A Contribution of the Research Field Energy”. The authors would like to thank Prof. Steven Simske for proofreading the article.

References

[1] ENTSOG, ENTSO-E. TYNDP 2018 scenario report. Tech. Rep., 2018, URL: https://tyndp.entsoe.eu/tyndp2018/scenario-report/; [accessed August, 2018].
[2] Hafiz F, de Queiroz AR, Husain I. Solar generation, storage, and electric vehicles in power grids: Challenges and solutions with coordinated control at the residential level. IEEE Electrif Mag 2018;6(4):83–90.
[3] Selvam MM, Gnanadass R, Padhy N. Initiatives and technical challenges in smart distribution grid. Renew Sustain Energy Rev 2016;58:911–7.
[4] Vlahinić S, Franković D, Komen V, Antonić A. Reactive power compensation with PV inverters for system loss reduction. Energies 2019;12(21):4062.
[5] Wang M, Mu Y, Jia H, Wu J, Yu X, Qi Y. Active power regulation for large-scale wind farms through an efficient power plant model of electric vehicles. Appl Energy 2017;185:1673–83.
[6] Gneiting T, Katzfuss M. Probabilistic forecasting. Annu Rev Stat Appl 2014;1:125–51.
[7] Zhang Y, Wang J, Wang X. Review on probabilistic forecasting of wind power generation. Renew Sustain Energy Rev 2014;32:255–70.
[8] Hong T, Pinson P, Fan S, Zareipour H, Troccoli A, Hyndman RJ. Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. Int J Forecast 2016;32(3):896–913. http://dx.doi.org/10.1016/j.ijforecast.2016.02.001.
[9] Juban R, Ohlsson H, Maasoumy M, Poirier L, Kolter JZ. A multiple quantile regression approach to the wind, solar, and price tracks of GEFCom2014. Int J Forecast 2016;32(3):1094–102.
[10] Jaworski P, Durante F, Hardle WK, Rychlik T. Copula theory and its applications. Berlin, Heidelberg, Germany: Springer Berlin Heidelberg; 2010.
[11] Panamtash H, Zhou Q, Hong T, Qu Z, Davis KO. A copula-based Bayesian method for probabilistic solar power forecasting. Sol Energy 2020;196:336–45.
[12] Borkowska B. Probabilistic load flow. IEEE Trans Power Appar Syst 1974;(3):752–9.
[13] Koenker R. Quantile regression. Cambridge University Press; 2005.
[14] González Ordiano JA, Gröll L, Mikut R, Hagenmeyer V. Probabilistic energy forecasting using the nearest neighbors quantile filter and quantile regression. Int J Forecast 2020;36(2):310–23. http://dx.doi.org/10.1016/j.ijforecast.2019.06.003.
[15] Chen P, Chen Z, Bak-Jensen B. Probabilistic load flow: A review. In: Proc., 2008 third international conference on electric utility deregulation and restructuring and power technologies. Piscataway, NJ, USA: IEEE; 2008, p. 1586–91.
[16] Prusty BR, Jena D. A critical review on probabilistic load flow studies in uncertainty constrained power systems with photovoltaic generation and a new approach. Renew Sustain Energy Rev 2017;69:1286–302.
[17] Ramadhani UH, Shepero M, Munkhammar J, Widén J, Etherden N. Review of probabilistic load flow approaches for power distribution systems with photovoltaic generation and electric vehicle charging. Int J Electr Power Energy Syst 2020;120:106003.
[18] He Y, Li H. Probability density forecasting of wind power using quantile regression neural network and kernel density estimation. Energy Convers Manage 2018;164:374–84.
[19] Taieb SB, Huser R, Hyndman RJ, Genton MG. Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression. IEEE Trans Smart Grid 2016;7(5):2448–55.
[20] Carpinelli G, Caramia P, Varilone P. Multi-linear Monte Carlo simulation method for probabilistic load flow of distribution systems with wind and photovoltaic generation systems. Renew Energy 2015;76:283–95.
[21] Constante-Flores GE, Illindala MS. Data-driven probabilistic power flow analysis for a distribution system with renewable energy sources using Monte Carlo simulation. IEEE Trans Ind Appl 2018;55(1):174–81.
[22] Hajian M, Rosehart WD, Zareipour H. Probabilistic power flow by Monte Carlo simulation with latin supercube sampling. IEEE Trans Power Syst 2012;28(2):1550–9.
[23] Appino RR, Mühlpfordt T, Faulwasser T, Hagenmeyer V. On solving probabilistic load flow for radial grids using polynomial chaos. In: Proc., 2017 IEEE Manchester powertech, Manchester, UK. Piscataway, NJ, USA: IEEE; 2017, p. 1–6.
[24] Morales JM, Baringo L, Conejo AJ, Mínguez R. Probabilistic power flow with correlated wind sources. IET Gener Transm Distrib 2010;4(5):641–51.
[25] Fan M, Vittal V, Heydt GT, Ayyanar R. Probabilistic power flow studies for transmission systems with photovoltaic generation using cumulants. IEEE Trans Power Syst 2012;27(4):2251–61.

[26] Nikmehr N, Ravadanegh SN. Heuristic probabilistic power flow algorithm for microgrids operation and planning. IET Gener Transm Distrib 2015;9(11):985–95.
[27] Villanueva D, Pazos JL, Feijoo A. Probabilistic load flow including wind power generation. IEEE Trans Power Syst 2011;26(3):1659–67.
[28] Widén J, Shepero M, Munkhammar J. Probabilistic load flow for power grids with high PV penetrations using copula-based modeling of spatially correlated solar irradiance. IEEE J Photovolt 2017;7(6):1740–5.
[29] Hagenmeyer V, Cakmak HK, Düpmeier C, Faulwasser T, Isele J, Keller HB, Kohlhepp P, Kühnapfel U, Stucky U, Waczowicz S, Mikut R. Information and communication technology in Energy Lab 2.0: Smart energies system simulation and control center with an open-street-map-based power flow simulation example. Energy Technol 2016;4:145–62. http://dx.doi.org/10.1002/ente.201500304.
[30] Bompard E, Carpaneto E, Chicco G, Napoli R. Convergence of the backward/forward sweep method for the load-flow analysis of radial distribution systems. Int J Electr Power Energy Syst 2000;22(7):521–30.
[31] Liu J, Braun E, Düpmeier C, Kuckertz P, Ryberg DS, Robinius M, Stolten D, Hagenmeyer V. Architectural concept and evaluation of a framework for the efficient automation of computational scientific workflows: An energy systems analysis example. Appl Sci 2019;9(4):728.
[32] Kyesswa M, Çakmak H, Kühnapfel U, Hagenmeyer V. A matlab-based dynamic simulation module for power system transients analysis in the eASiMOV framework. In: 2017 European modelling symposium (EMS). IEEE; 2017, p. 157–62.
[33] Fahrmeir L, Kneib T, Lang S, Marx B. Regression: Models, methods and applications. Berlin, Germany: Springer; 2013.
[34] Gaglianone WP, Lima LR. Constructing density forecasts from quantile regressions. J Money Credit Bank 2012;44(8):1589–607.
[35] Grainger J, Stevenson WD. Power system analysis. McGraw-Hill; 1994.
[36] Xiu D. Numerical methods for stochastic computations: A spectral method approach. Princeton, NJ, USA: Princeton University Press; 2010.
[37] Eminoglu U, Hocaoglu MH. Distribution systems forward/backward sweep-based power flow algorithms: A review and comparison study. Electr Power Compon Syst 2008;37(1):91–110.
[38] Adusumilii BS, Kumar BK. Backward/forward sweep based power flow analysis of distribution systems under uncertainty using new affine arithmetic division. In: 2020 IEEE power & energy society innovative smart grid technologies conference (ISGT). IEEE; 2020, p. 1–5.
[39] Appino RR, González Ordiano JA, Mikut R, Hagenmeyer V, Faulwasser T. Storage scheduling with stochastic uncertainties: Feasibility and cost of imbalances. In: Proc., 2018 power systems computation conference (PSCC), Dublin, Ireland. Piscataway, NJ, USA: IEEE; 2018, p. 1–7. http://dx.doi.org/10.23919/PSCC.2018.8442529.
[40] Mühlpfordt T, Zahn F, Hagenmeyer V, Faulwasser T. PolyChaos. jl–A Julia package for polynomial chaos in systems and control. 2020, arXiv preprint arXiv:2004.03970.
[41] González Ordiano JA, Mikut R. The forecasting toolbox of the MATLAB-toolbox SciXMiner short manual. Tech. Rep., 2020, http://dx.doi.org/10.13140/RG.2.2.31678.84801/1.