Scenario generation of aggregated Wind, Photovoltaics and small Hydro production for power systems applications

Simon Camal, Fei Teng, Andrea Michiorri, Georges Kariniotakis, L Badesa

Scenario generation of aggregated Wind, Photovoltaics and small Hydro production for power systems applications

S. Camal[a], F. Teng[b], A. Michiorri[a], G.Kariniotakis[a], L. Badesa[b]

[a] *MINES ParisTech - PSL University, Research Center PERSEE, Sophia Antipolis, France*
[b] *Imperial College London, Electrical and Electronic Engineering Department, London, SW7 2AZ, UK*

Abstract

This paper proposes a methodology for an efficient generation of correlated scenarios of Wind, Photovoltaics (PV) and small Hydro production considering the power system application at hand. The merits of scenarios obtained from a direct probabilistic forecast of the aggregated production are compared with those of scenarios arising from separate production forecasts for each energy source, the correlations of which are modeled in a later stage with a multivariate copula. It is found that scenarios generated from separate forecasts reproduce globally better the variability of a multi-source aggregated production. Aggregating renewable power plants can potentially mitigate their uncertainty and improve their reliability when they offer regulation services. In this context, the first application of scenarios consists in devising an optimal day-ahead reserve bid made by a Wind-PV-Hydro Virtual Power Plant (VPP). Scenarios are fed into a two-stage stochastic optimization model, with chance-constraints to minimize the probability of failing to deploy reserve in real-time. Results of a case study show that scenarios generated by separately forecasting the production of each energy source leads to a higher Conditional Value at Risk than scenarios from direct aggregated forecasting. The alternative forecasting methods can also significantly affect the scheduling of future power systems with high penetration of weather-dependent renewable plants. The generated scenarios have a second application here as the inputs of a two-stage stochastic unit commitment model. The case study demonstrates that the direct forecast of aggregated production can effectively reduce the system operational cost, mainly through better covering the extreme cases. The comprehensive application-based assessment of scenario generation methodologies in this paper informs the decision-makers on the optimal way to generate short-term scenarios of aggregated RES production according to their risk aversion and to the contribution of each source in the aggregation.

Keywords: Ancillary Services, Chance-constrained Optimization, Forecasting, Renewable Energy, Scenarios, Stochastic Scheduling

1. Nomenclature

1.1. Indices

agg	Level of aggregation in the VPP
s	Index of energy source type in the VPP
da	Day-ahead variable
g	Index of thermal generators
rt	Real-time variable
$\uparrow,\downarrow$	Upward,downward regulation
ω	Index of production scenarios

1.2. Sets

c	Coefficients of the objective function
$\mathcal{G}$	Set of thermal generators
$\mathcal{O}$	Subset of slow thermal generators
π	Prices for energy and reserve

1.3. Variables

a_R	Reserve activation probability
b	Binary variable for ramp function
$\mathbf{B}$	Vector of decision variables for bidding
δ	gradient indicator function for Brier Score
$\Delta R^{\uparrow,-}$	Deficit of upward reserve in real-time
$\Delta R^{\downarrow,-}$	Deficit of downward reserve in real-time
E^{da}	Energy on the day-ahead market
$E^{rt,+}$	Energy surplus on the real-time market
$E^{rt,-}$	Energy deficit on the real-time market

[☆] This paper was submitted for review on 03.07.2018. The research was carried as part of the European project REstable (Reference Number 77872), supported by the ERA-NET Smart Grids Plus program with financial contribution from the European Commission, ADEME, Juelich Research Center, Fundacao para a Ciencia e a Tecnologia

Email address: name.surname@mines-paristech.fr (S. Camal)

F	Cumulative distribution function
m	Marginal distribution function of a copula
o_T	Objective function for risk-neutral bidding
$o_{\beta,T}$	Objective function for risk-averse bidding
ψ	Positive variable for ramp function
P^{LS}	Load-shedding
P^g	Power produced by generator g
P^{RC}	RES curtailment
R^{da}	Reserve on the day-ahead market
R^{rt}	Reserve on the real-time market
Σ	Covariance matrix of Normal variables
$\mathbf{u}$	Vector of uniform i.i.d. variables
U	Generalized cumulative distribution function
$\mathbf{X}$	Vector of features for production forecasting
Y^{agg}	Power production at aggregation level agg
y^g	Binary variable, commitment decision for generator g
z^g	Startup state of generator g

1.4. Parameters

β	Aversion to risk on revenue
c^{LS}	Load-shedding cost
$\mathrm{c}^g_{\mathrm{m}}$	Marginal cost of generator g
$\mathrm{c}^g_{\mathrm{nl}}$	No-load cost of generator g
$\mathrm{c}^g_{\mathrm{st}}$	Startup cost of generator g
ϵ	Confidence parameter of chance constraints
k	Horizon of forecast
Ω	Number of scenarios
p_ω	Probability of occurrence for scenario ω
$\mathrm{P}^g_{\mathrm{max}}$	Rated power of unit g
$\mathrm{P}^g_{\mathrm{msg}}$	Minimum stable generation of unit g
$\mathrm{P}^g_{\mathrm{rd}}$	Maximum ramp-down capability of unit g
$\mathrm{P}^g_{\mathrm{ru}}$	Maximum ramp-up capability of unit g
$\mathrm{P}^{\mathrm{D}}_{\omega,t}$	Demand in scenario ω and time-step t
$\mathrm{P}^{\mathrm{R}}_{\omega,t}$	Power produced by RES in scenario ω and time-step t
t	Runtime of forecast
τ_t	Duration of time-step t in the two-stage SUC
T	Period of bidding optimization

2. Introduction

Weather-dependent Renewable Energy Sources (RES) constitute an increasing share of power generation, but the uncertainty of their production raises challenges for their interaction with power systems. In particular, the variability of aggregated RES power production must be accounted for in power system applications which exploit the complementarity between RES (e.g. bidding on electricity markets by an aggregator operating RES plants), or consider the systemic impact of aggregated RES plants (e.g. scheduling of power systems with high RES penetration).

The complementarity between RES is a subject of interest in applications related to the provision of Ancillary Services (AS). Historically, only conventional dispatchable power plants provided AS, but recent studies have shown that RES plants of various sources have the technical capacity to offer AS, namely Wind [1], run-of-river hydro [2], and Photovoltaics (PV) [3]. However the uncertainty associated with the production of a single RES plant is too large to provide AS with the high level of reliability required by system operators. A first solution to increase the reliability of AS delivered by RES is to combine RES plants with Energy Storage Systems (ESS) which can compensate forecast error, but with a high investment cost in the case of static storage [4]. A second solution is to aggregate RES plants blending energy sources and weather conditions, in order to reduce the uncertainty of the total production [5]. The aggregated plants are coordinated via a Virtual Power Plant (VPP) which ensures sufficient regulation capacity to comply with the stringent pre-qualification tests for the provision of AS [6]. Such a VPP can integrate other agents like flexible consumers or ESS from static storages and electric mobility [7],[8], [9]. Following the ongoing development of short-term markets for AS [10], methodologies for the optimal offer of energy and reserve have been proposed for wind farms [11],[12], microgrids [13], and aggregated flexible loads [14] [15]. The optimization models of these methodologies rely on probabilistic forecasts or trajectories of RES production. Adding chance constraints to the optimization makes it possible to balance revenue with the technical risk of not supplying reserve [14]. The realization of these constraints depends on scenarios of RES production, reproducing the expected uncertainty at short-term. In contrast with the reviewed works, which considered single energy sources or several sources independently, the present paper investigates how to forecast the expected interaction between multiple RES at short-term and quantify the associated impact on AS bidding.

The aggregated production of weather-dependent renewable power plants plays also a major role in the scheduling of electrical networks. When scheduling electrical systems, operators decide the commitment decision of generators to meet the demand with minimal cost. At present, AS are scheduled following deterministic rules that involve imposing pre-defined requirements to tackle the uncertainties associated with forecasting error and equipment outages. As the uncertainty introduced by the total generation of weather-dependent renewable power plants is much more significant than that of demand, scheduling processes performed under deterministic rules become inefficient [16] [17] [18] [19]. In contrast, stochastic unit commitment models employ scenarios to model the uncertainty of weather-dependent renewables [20] [21]. These models ensure economic efficiency and network security.

In both applications introduced above, short-term decisions are derived from a stochastic optimization which relies on scenarios of aggregated RES production as inputs. It is known that the quality of the scenarios has a direct impact on the performance of the optimization model [21]. In the case of short-term horizons, scenarios should combine two properties: reproduce the interde-

pendence in the aggregated renewable production process, and vary as a function of the influence of explanatory variables such as weather forecasts. The interdependence in RES production comprises classically the temporal dependencies between successive lead-times, which is generally not captured by production forecasts [22], and the possible correlation between power plants. A specific challenge for scenarios of a multi-source RES aggregation is to model the dependencies between energy sources, for simultaneous and successive lead-times, while preserving the conditional response to explanatory variables.

A popular method to generate scenarios is to build a multivariate Gaussian copula from probabilistic forecasts of the marginal distributions (e.g. production of a given wind farm at different lead times [22], or production of several PV plants at different lead times [23]). Copulas are flexible tools to model dependencies in uncertainties, although in problems focused on extreme regions of the marginal distributions, analytical models using exponential functions may be more appropriate [22]. Vine copulas, which form flexible trees of bivariate copulas, have been successfully used for the probabilistic forecast of multiple Wind [24] and PV [25] plants. However to the authors' knowledge, neither Gaussian copulas nor Vine copulas have been reported in the literature in the context of scenarios for a multi-source RES aggregation. Hierarchical copula models, which base on independent forecasts of each contributor of an aggregation at several hierarchical levels, have been proposed in the context of electricity demand [**?**] and insurance exposure [26]. While these approaches are effective on aggregations which follow naturally a tree structure such as electricity networks, their application on a multi-source RES aggregation which has no obvious hierarchical structure would require a large number of permutations to assess all possible sums between plants or sources.

Alternative methods to generate scenarios exist. Time series analysis can derive spatio-temporal models for renewable power plants at multiple sites [27]. The machine learning approach of [28] builds an iterative neural network that outputs scenarios by random generation of errors and step-ahead forecasts. For stochastic scheduling of power systems with high penetration by wind power, [20] develop a multi-stage scenario tree based on the distribution of wind forecasting errors. In [21], scenarios of wind forecasting errors are generated by a Levy α-stable distribution. These approaches are based on deterministic forecasts of production, which is valid for a single energy source but would miss important interdependences of uncertainty between energy sources. The Generative Adversarial Networks of [29] produce trajectories with adequate diversity and similar statistical properties to historical data thanks to the ability of this unsupervised deep learning to learn complex non-linearities and classify large inputs. The minimization of the Wasserstein distance between the generator and discriminator provides good climatological properties of the scenarios, and could be applied to multi-source RES aggregations. However, the solution of [29] proposes only a classification of typical conditions (e.g. scenarios for a sunny winter day). This is suitable for long-term scenarios but not directly applicable for short-term scenarios where scenarios should reflect the expected conditions such as weather forecasts.

In the presented literature few efforts have been made to assess how different scenario generation approaches impact the results of the optimization. More specifically, the use of aggregated, multi-source RES trajectories in stochastic optimization has not been frequently investigated. Uncertainty in production is of particular interest when aggregating power plants of different energy sources, possibly subject to different climates. The present paper addresses the challenge of generating scenarios of the output of a mix of power plants harvesting different energy sources. The dimension of the forecasting problem is significantly more complex than forecasting single energy sources. Furthermore, we investigate which scenario generation method is most suitable considering the framework of the applications, with examples for AS bidding and power system scheduling.

The framework of the paper, illustrated in Figure 1, is structured as follows: we start by forecasting the production of a portfolio of PV, Wind and small Hydro plants, separately for each source and directly at the aggregated level. From these probabilistic forecasts we generate scenarios by applying multivariate copulae on the available marginal distributions over the range of lead-times, and over the range of sources in the case of separate forecasts. The Gaussian copula is first applied as a standard spatio-temporal dependence model, and a Vine copula is proposed to model the dependencies between separate forecasts with more flexibility. Then, these scenarios are applied to two relevant case studies, in which the solution of the stochastic optimization is highly dependent on the characterization of the uncertainty. The first case study is the day-ahead bidding of energy and automatic Frequency Restoration Reserve (aFRR) by a renewable VPP using a two-stage stochastic optimization. In the second case study, we schedule a network with significant penetration of Wind, PV and small hydro.

The key contributions of this work are identified as follows:

1. Firstly, this paper addresses the problem of generating probabilistic forecasts for the aggregated power of a set of renewable power plants harvesting different energy sources. Despite abundant works in the literature on models that separately forecast PV, Wind or Hydro, the problem has not been tackled jointly.
2. Based on probabilistic forecasts of either the individual energy sources or the aggregation, this paper proposes to generate scenarios of the aggregated production, with a multivariate Gaussian copula taking into account correlations.

3. In addition, to evaluate the statistical performance of the forecasting methods, this paper compares the methods in two particular applications, reserve bidding and system scheduling. The problem of reserve bidding is formulated as a chance-constrained optimization to include a maximum frequency of reserve under-fulfillment. A two-stage stochastic unit commitment model is applied in this paper to evaluate the impact on the system operational costs.
4. The comprehensive application-based assessment of scenario generation methodologies informs the decision makers on the optimal way to generate short-term scenarios of aggregated RES production according to their risk aversion and to the contribution of each source in the aggregation.

The methodology developed for the probabilistic forecasting of weather-dependent RES production is presented in Section 3. Section 4 develops the methodology for generating scenarios. Then, we develop the evaluation framework for two applications: a reserve bidding optimization problem in Section 5, and a stochastic unit commitment problem in Section 6. We present the results in Section 7 and conclude in Section 8.

3. Probabilistic Forecasting of Weather-dependent Production

The production of weather-dependent renewable power plants is forecast on a day-ahead horizon with two possible levels of aggregation: (1) plants of same energy source and (2) all plants of possibly different energy sources located under a common perimeter (e.g. the portfolio of an aggregator, or a control area). Among the various existing approaches for the density forecast of renewable production, we choose the Quantile Regression Forest (QRF) model. This is a validated nonparametric model, which can achieve adequate performance in terms of reliability [30] and global forecasting skill [31].

The production forecast of an aggregated portfolio agg at runtime t and horizon k is given in (1) from the explanatory variables of all plants under the aggregation level chosen X_t^{agg}. Features of Wind and PV plants comprise day-ahead Numerical Weather Predictions (NWP) retrieved at each power plant site. In the case of run-of-river hydro, we consider NWP and past production values. Hydro production levels in recent weeks inform about features that are independent from day-ahead weather forecasts, such as the hydrological conditions of the river or the availability of the plant. Learning in the QRF model is mainly influenced by three hyper-parameters: the number of variables randomly selected $mtry$, the number of trees δ and the minimum number of observations in leaves $minobs$. The number of selected variables $mtry$ is kept to 1/3 of the available variables. The number of trees δ must be several hundreds (at least 500) to make sure that important

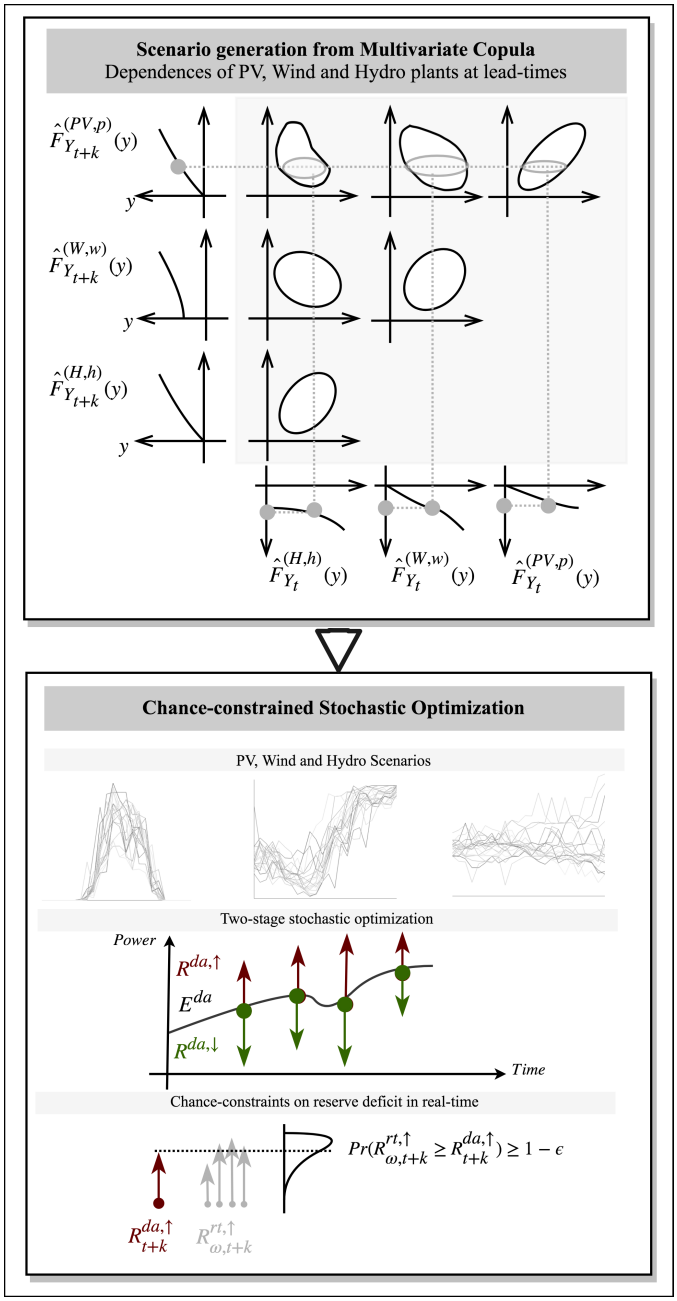

Figure 1: Methodology framework: aggregated production scenarios are generated and evaluated on a stochastic optimization problem (here chance-constrained reserve bidding)

variables are selected a few times. The minimum number of observations per leaf $minobs$ is kept to 10 to avoid overfitting. Finally, the aggregated production behaves differently according to the horizon (e.g. diurnal presence of PV, intra-day wind patterns...), so the QRF model is trained on multiple horizon intervals.

$$\hat{f}_{Y_{t+k}^{agg}|\mathbf{X}_t^{agg}} = QRF(Y_{t+k}^{agg}, X_t^{agg}, mtry, \delta, minobs) \quad (1)$$

In the case of direct aggregated forecast, the reliability is improved by adding variables to the marginal features

corresponding to each plant of the aggregation: (1) minimum and maximum values of variables across plants sharing the same source (e.g. minimum and maximum wind speed) help the QRF model explore more robustly the variance of the aggregated production process, (2) lagged values of marginal features add information on the weather patterns around the time of prediction or on persistent behaviour of production for a given energy source. In the case of separate forecasting by energy source, a preprocessing treatment of the production variable Y^{agg}_{t+k} helps reduce the impact of deterministic trends on learning and obtain conditional variances independent of conditional means [22]. The treatment is specific to each energy source and is carried out before summing up the power production data: for Wind we apply the logit transform [22], while for PV we normalize production with a simple analytical model for the top of atmosphere global horizontal irradiance. The logit transform could also been applied to Hydro (by analogy with Wind as a double-bounded production that is non-linearly dependent on an inflow). However, in this case the hydro production is not explained by a water flow forecast but rather by features mixing production and weather variables, meaning there is no homogeneity in the non-linear relationships between features and production, and therefore logit transform is not employed.

At the end of this first step of the methodology, we obtain at runtime t, two sets of forecasts for all horizons $k \in [1, K]$:

1. The total aggregated production for all plants, $\hat{F}_{Y^{total}_{t+k}}$;
2. The production of plants sharing the same energy source, namely Wind, PV and run-of-river Hydro, $\hat{F}_{Y^{Wind}_{t+k}}$, $\hat{F}_{Y^{PV}_{t+k}}$, $\hat{F}_{Y^{Hydro}_{t+k}}$.

4. Generation of Production Scenarios

The probabilistic forecasts obtained in the previous section inform us about the predicted levels of production and their relative uncertainty. However simple Monte Carlo sampling on the predictive densities obtained does not consider correlations resulting thus in non realistic alternative forecast scenarios with close horizons or spatial correlations between plants or energy sources. In contrast, adding a model of the correlations between the predictive densities, such as a multivariate copula [**?**], produces scenarios with more realistic behavior with respect to the real production patterns.

4.1. Multivariate Copula based on Probabilistic Forecasts

The multivariate copula is a multivariate distribution function, the marginals of which should be distributed uniformly in the rank domain [23]. The marginals are predictive densities of the production for specific dimensions of the problem, for instance the various horizons or the different energy sources of the aggregation. Due to the boundedness of intermittent production, the forecasted distribution function $\hat{F}_Y$ of an intermittent production Y is not strictly monotonous, hence a given power observation y_{obs} can be associated with several quantile values $\hat{F}_Y(y_{obs})$, for instance if y_{obs} is a wind power observation occurring at wind speeds over the nominal speed value. In order to obtain marginals uniformly distributed from intermittent production forecasts, we apply to each forecast the distributional transform developed by Ruschendorf [32], which generalizes in (2) the property of uniform distribution to discontinuous Cumulative Distribution Functions (CDFs).

$$U(y) = \hat{F}_{Y^-}(y) + V(\hat{F}_Y(y) - \hat{F}_{Y^-}(y)) \tag{2}$$

where V is a random variable following the uniform distribution and $\hat{F}_{Y^-}$ is the left-hand CDF of the production variable Y. Note that $U(y) = \hat{F}_Y(y)$ if the CDF is continuous. We can now construct our multivariate copula from the transformed marginal distributions.

At this point, we propose two distinct methods. The first method, entitled "Direct Gaussian" (DG), constructs a multivariate temporal dependence model between the variables Y^{total}_k, which represent the values taken by the total aggregated production at the successive horizons $k \in [1, K]$. We start by collecting a series of observed aggregated productions y^{total}_{t+k}, not included in the training set of the forecasting model, for all horizons k. The position of these observed productions in the forecast distribution $\hat{F}_{Y^{total}_{t+k}}$ at each horizon k is evaluated according to (2), and constitutes a realization u^{total}_{t+k} of the uniformly distributed marginal U^{total}_k.

$$u^{total}_{t+k} = U^{total}_k(y^{total}_{t+k}), \quad \forall t, \forall k \tag{3}$$

All marginals are then converted in (4) into normally distributed variables Z^{total}_k using the probit function Φ [22], forming a multivariate variable $\mathbf{Z^{total}}$ normally distributed with a zero mean vector and covariance Σ^{total} of dimension $K \times K$.

$$z^{total}_{t+k} = Z^{total}_k(y^{total}_{t+k}) = \Phi(u^{total}_{t+k}), \quad \forall t, \forall k \tag{4}$$

$$\mathbf{Z^{total}} = (Z^{total}_1, Z^{total}_2, ..., Z^{total}_k, ..., Z^{total}_K) \tag{5}$$

$$\mathbf{Z^{total}} \sim \mathcal{MVN}(0, \Sigma^{total}) \tag{6}$$

The density of $\mathbf{Z^{total}}$ forms a Gaussian copula parametrized by the mean vector and the covariance matrix, which is computed here as the empirical covariance matrix on the observed normally transformed marginals. A final step consists in drawing samples from the copula to generate trajectories which reproduce the temporal correlation between horizons. To generate Ω distinct scenarios of total aggregated production for a given period of interest $[t+1, t+K]$ we:

1. Draw Ω i.i.d.random vectors $\mathbf{s}_\omega$ following the uniform distribution $U(0,1)^K$, where $\omega \in [1, \Omega]$;
2. Convert them into realizations $\mathbf{z}^{total}_\omega$ of $\mathbf{Z^{total}}$;
3. Generate trajectories $\hat{y}^{total,DG}_{\omega,t+k}$ for the period of interest by applying in (7) the quantile values given by each $\mathbf{z}_\omega$ to the marginal forecasts $\hat{F}_{Y^{total}_{t+k}}$:

$$\hat{y}^{total,DG}_{\omega,t+k} = \hat{F}^{-1}_{Y^{total}_{t+k}}(z^{total}_{\omega,t+k}) \quad \forall\omega, \forall t, \forall k \tag{7}$$

A second method is based on the separate forecasting by energy source instead of the direct forecasting of the total aggregated production. This approach, entitled "Indirect Gaussian" (IG), models the dependencies between productions of the different energy sources, over all horizons. The observed productions are now collected and aggregated for each energy source separately, resulting in an observation vector $\mathbf{y}^{sources}_{t+k}$ of dimension S, S being the number of sources (in the present case S=3 with Wind, PV and small Hydro). The multivariate copula of the variable $\mathbf{Z}^{sources}$ is constructed in the same way as for the DG method, giving a covariance matrix $\Sigma^{sources}$ of dimension $SK \times SK$. Scenarios are generated by sampling the covariance matrix and affecting the resulting quantiles $\mathbf{z}^{sources}_{\omega}$ of dimension SK to the probabilistic forecasts of the respective energy sources for the period of interest. Lastly the obtained equiprobable trajectories are summed across all sources in (8) to form Ω trajectories of total aggregated production.

$$\hat{y}^{total,IG}_{\omega,t+k} = \sum_{s=[1,S]} \hat{F}^{-1}_{Y^{s}_{t+k}}(\mathbf{z}^{sources,s}_{\omega,t+k}) \quad \forall\omega \in [1,\Omega], \forall t, \forall k \tag{8}$$

A variant of this method, entitled "Indirect Vine" (IV), consists in replacing the Gaussian copula by a regular Vine copula to model non-Gaussian dependencies between horizons and energy sources. A regular Vine copula is formed sequentially by joining bivariate copulas into trees. The selected tree among all possible combinations is the tree that maximizes the sum of empirical rank correlations over the possible pairs (Maximum Spanning Tree algorithm, see [24]). To generate a number of scenarios Ω from the Vine copula at horizon $t+k$ (omitted in notations below for the sake of simplicity) we:

1. Draw Ω i.i.d random vectors $\mathbf{s}_{\omega}$ following the uniform distribution $U(0,1)^{SK}$
2. Retrieve the uniform marginal CDF value $m_{\omega,d}, d \in SK$ of the production variable Y_d, conditioned by the other variables, by inverting the h-function of the Vine copula [24]

$$m_{\omega,d} = \hat{F}^{-1}_{d|d-1,...,1}(\mathbf{s}_{\omega,d}|m_{\omega,d-1}, ..., m_{\omega,1}) \tag{9}$$

3. Invert the CDF of the marginal production variable $\hat{F}^{(d)}_{Y}$ to obtain the production trajectory.

$$y_{\omega,d} = \hat{F}^{-1}_{Y^d}(z_{\omega,d}) \tag{10}$$

In the next section, we evaluate the quality of the trajectories of total aggregated production obtained by direct aggregated forecasting and separate forecasting by energy source.

4.2. Evaluation of Trajectories

The generated trajectories must reproduce correlations between horizons, locations and energy sources. We assess the quality of trajectories by a proper score, the Variogram-based score (VS) [33], to determine whether trajectories reproduce correctly the main moments of the original production [23]. The VS of order γ can be expressed in (11) as the quadratic difference between the Variogram of the original production data y and the Variogram of the forecast trajectories $\hat{y}_{\omega t}$. The latter is approximated by the mean of the score over the scenarios. Here, pairs of points are equally weighted, $w_{ij} = 1$. Points with a low correlation and thus a low signal-to-noise ratio are therefore not penalized [33]. The discrimination ability of the score could be lower than with a correlation model fitted on data, but we choose to use equal weights to investigate the whole range of correlations including long intervals (production gradients over several hours are an important input for reserve bidding).

$$VS^{(\gamma)}_{t} = \sum_{i,j \in M} w_{ij}(|y_{t,i} - y_{t,j}|^{\gamma} - \frac{1}{\Omega}\sum_{\omega \in [1,\Omega]} (|\hat{y}_{\omega t,i} - \hat{y}_{\omega t,j}|^{\gamma}))^2 \tag{11}$$

Beyond similarity in the trajectory distributions, trajectories should also exhibit characteristic events of the original time series, such as gradients. Gradients up to a few hours are of particular interest when offering reserve capacities: the validity period (contracted duration of capacity) of secondary reserve (aFRR) goes from 15 min in the Netherlands to 1h in Portugal and 4h in Germany [34]. We evaluate the similarity of gradients observed in the original time series with gradients in scenarios by means of a Brier Score (BS) defined in (13). The events considered in the score are production gradients δ_t over an interval Δt, which are higher than a threshold value r, taken as the average observed gradient over the interval.

$$\delta_t(y; \Delta t) = \mathbf{1}(|y_{t+\Delta t} - y_t| \geq r) \tag{12}$$

$$BS = \frac{1}{T}\sum_{t=1}^{T}(\frac{1}{\Omega}\sum_{\omega \in [1,\Omega]} \delta_t(\hat{y}_{\omega}; \Delta t) - \delta_t(y; \Delta t))^2 \tag{13}$$

The generated scenarios are now applied to two stochastic optimization problems, reserve bidding and network scheduling.

5. Case Study 1: Day-ahead Bidding of Energy and Automatic Frequency Restoration Reserve

A VPP aggregating Wind, PV and small Hydro power plants jointly offers energy and symmetrical automatic Frequency Restoration Reserve (aFRR) on a day-ahead market. It earns revenues for reserve capacities and energy activated for upward reserve (which could not be sold in the energy market), while it pays the grid operator for the activated energy during downward reserve (to compensate for the energy not delivered). We consider no uncertainty

of market conditions, in order to individuate the impact of production scenarios on the result.

A specific market condition of this problem is the activation of the VPP: considering that the TSO activates the aFRR under a merit-order scheme, what is the probability of the VPP being activated and at what intensity? In a simplified approach we assume that the VPP has similar marginal costs to its competitors. The activation is then modeled by an activation probability, denoted below a_R: the probability of activation equals the effective aFRR demand divided by the tendered demand [5].

5.1. Mathematical Formulation

This bidding problem can be formulated as a two-stage stochastic linear optimization. In a first stage, the VPP offers volumes of energy and reserve in their respective day-ahead (da) markets for each market time unit t of the optimization period T. Then in a recourse stage occuring in real time (rt), the VPP compensates its imbalances in the energy market, delivers the requested reserve and any penalties if it fails to do so. At this stage the decisions of the VPP are computed for each scenario of index ω Assuming that the VPP is price-taker and risk-neutral, the objective function writes:

$$\min o_T = \mathbb{E}(f(\mathbf{B},\omega)) = \sum_{\omega\in[1,\Omega]} p_\omega f(\mathbf{B},\omega) \tag{14}$$

where the bidding net penalty for scenario ω is

$$f(\mathbf{B},\omega) = \sum_{t\in[1,T]} [c^{\mathcal{T}}_{da,t}.\mathbf{B}^{da}_t + c^{\mathcal{T}}_{rt,t}.\mathbf{B}^{rt}_{\omega t}] \tag{15}$$

with the following decision variables:

$$\mathbf{B}^{da}_t = (E^{da}_t, R^{\uparrow,da}_t, R^{\downarrow,da}_t) \tag{16}$$

$$\mathbf{B}^{rt}_{\omega t} = (E^{rt,-}_{\omega t}, E^{rt,+}_{\omega t}, R^{\uparrow,rt}_{\omega t}, R^{\downarrow,rt}_{\omega t}, \Delta R^{\downarrow,-}_{\omega t}, \Delta R^{\uparrow,-}_{\omega t}) \tag{17}$$

and their corresponding costs:

$$c^{\mathcal{T}}_{da,t} = (-\pi^{da}_{E,t}, -\pi^{da}_{R\uparrow,t}, -\pi^{da}_{R\downarrow,t})$$

$$c^{\mathcal{T}}_{rt,t} = (\pi^{rt,-}_{E,t}, -\pi^{rt,+}_{E,t}, a^{rt}_{R\uparrow,t}.\pi^{rt}_{R\uparrow,t}, - a^{rt}_{R\downarrow,t}.\pi^{rt}_{R\downarrow,t}, -\pi^{rt,-}_{R\uparrow,t}, -\pi^{rt,-}_{R\downarrow,t}) \tag{18}$$

This problem is subject to the following constraints:

1. The simulated production of the VPP must match the sum of energy and reserve, considering day-ahead and real-time deviations.

$$E^{da}_t + E^{rt,+}_{\omega t} - E^{rt,-}_{\omega t} + a^{rt}_{R\uparrow,t}.R^{\uparrow,rt}_{\omega t} - a^{rt}_{R\downarrow,t}.R^{\downarrow,rt}_{\omega t} = Y^{agg}_{\omega t} \tag{19}$$

2. We add the possibility for the operator to offer less reserve capacity than contracted: this reserve deficit equals the deviation between day-ahead reserve offer and real-time reserve offer

$$\Delta R^{\uparrow,-}_{\omega t} = R^{\uparrow,da}_t - R^{\uparrow,rt}_{\omega t} \quad \forall\omega,\text{t} \tag{20}$$

$$\Delta R^{\downarrow,-}_{\omega t} = R^{\downarrow,da}_t - R^{\downarrow,rt}_{\omega t} \quad \forall\omega,\text{t} \tag{21}$$

3. The offer is symmetrical: the upward day-ahead reserve equals the downward day-ahead reserve.
4. The total power offered on energy and reserve markets can not exceed the installed capacity of the VPP.
5. The downward reserve can not exceed the energy offered.

This problem is easily generalized into a risk-averse formulation inserting an economic Conditional Value-at-Risk (CVaR), where the Value-at-Risk θ_{da} is the upper bound of revenues r having the probability $1-\alpha$ to be exceeded for all scenarios. The CVaR is linear with respect to the variables, so the problem remains linear. In this formulation we add in (25) a non-anticipativity constraint to ensure that the day-ahead decisions are independent from the outcomes of the production scenarios.

$$\max o_{\beta,T} = (1-\beta).\mathbb{E}(r(\mathbf{B},\omega)) + \beta.(\theta_{da} - \frac{1}{1-\alpha}\sum_{\omega\in[1,\Omega]} p_\omega\rho_\omega) \tag{22}$$

s.t.

$$\theta_{da} - r(\mathbf{B},\omega) \le \rho_\omega, \quad \forall\omega\in\Omega \tag{23}$$

$$\rho_\omega \ge 0 \quad \forall\omega\in\Omega \tag{24}$$

$$\mathbf{B}^{da}_{\omega,t} = \mathbf{B}^{da}_{\omega',t} \quad \forall\omega,\omega'\in\Omega \tag{25}$$

$$\text{(13)-(15)} \tag{26}$$

5.2. Chance-constrained Stochastic Optimization

A reserve offer that is unfulfilled too frequently might be discarded by network operators. Although reserve penalties may be in force, these are mainly considered to be a dissuasive signal rather than an arbitrage opportunity. The purely economic approach described above tends to minimize the volume of reserve offered to hedge against high penalties. A more balanced behavior between revenue and risk of underfulfillment can be obtained by adding chance constraints to the optimization problem. A solution is deemed feasible if the constraints representing the underfulfillment have a very low probability of occurrence over the scenario set. We add to the previous model a chance constraint on upward reserve (27) and downward reserve (28) to ensure that the reserve in the real-time is at least equal to the day-ahead reserve volume (i.e. no reserve deficit) with a probability of $1-\epsilon$.

$$Pr(\Delta R^{\uparrow,-}_{\omega t} \le 0) \ge 1-\epsilon \quad \forall\omega,\text{t} \tag{27}$$

$$Pr(\Delta R^{\downarrow,-}_{\omega t} \le 0) \ge 1-\epsilon \quad \forall\omega,\text{t} \tag{28}$$

A chance constraint is difficult to solve in its general form because it is not convex. The uncertain parameters are here the productions of each plant. These parameters are not normally distributed, so we can not easily convert it into a second-order cone constraint by inverting the Gaussian distribution function [13]. One option is to

derive the ϕ-divergence between the distribution of production and the normal distribution, then apply a distributionally robust chance-constrained programming model using this ϕ divergence [14]. As the distributions of renewable production show significant divergences with the Gaussian distribution (right skews and fat tails), we opt for an alternative approach which is distribution-free and scenario-oriented, i.e. the constraint is approximated by a non-decreasing convex function [35]. The constraint is conservatively approximated by a technical CVaR function on the distribution of reserve deficit $\mathrm{CVaR}_{\Delta R^{\uparrow,-}_{\omega t}}$ such that:

$$\mathrm{CVaR}_{\Delta R^{\uparrow,-}_{\omega t}}(1-\epsilon) = \inf_{\alpha}\left(\frac{\mathbb{E}([\Delta R^{\uparrow,-}_{\omega t} + \alpha]^{+})}{\epsilon}\right) - \alpha \quad (29)$$

Then the chance constraint (27) can be conservatively linearized as in (30). The expected value of the ramp function is obtained in (31) by averaging its value over the number of scenarios.

$$\mathbb{E}([\Delta R^{\uparrow,-}_{\omega t} + \alpha^{\uparrow}_{t}]^{+}) \leq \alpha^{\uparrow}_{t}.\epsilon \quad \forall\ \mathrm{t} \quad (30)$$

$$\mathbb{E}([\Delta R^{\uparrow,-}_{\omega t} + \alpha^{\uparrow}_{t}]^{+}) = \sum_{\omega\in[1,\Omega]} p_{\omega}[\Delta R^{\uparrow,-}_{\omega t} + \alpha^{\uparrow}_{t}]^{+} \quad (31)$$

The positive ramp function appearing in (30) is approached using the big M constraint technique: we insert binary variables $b_{\omega t}$ and positive variables $\psi_{\omega t}$ for each scenario ω such that :

$$-M.(1-b^{\uparrow}_{\omega t}) \leq \Delta R^{\uparrow,-}_{\omega t} + \alpha^{\uparrow}_{t} \leq M.b^{\uparrow}_{\omega t} \quad (32)$$

$$\psi^{\uparrow}_{\omega t} = b^{\uparrow}_{\omega t}.(\Delta R^{\uparrow,-}_{\omega t} + \alpha^{\uparrow}_{t}) \quad (33)$$

$$-M.(1-b^{\uparrow}_{\omega t}) \leq \psi^{\uparrow}_{\omega t} - (\Delta R^{\uparrow,-}_{\omega t} + \alpha^{\uparrow}_{t}) \leq M.(1-b^{\uparrow}_{\omega t}) \quad (34)$$

$$-M.b^{\uparrow}_{\omega t} \leq \psi^{\uparrow}_{\omega t} \leq M.b^{\uparrow}_{\omega t} \quad (35)$$

6. Case Study 2: Stochastic Unit Commitment

The impact that the two forecasting methods have on the power system operation is studied in this section. The system operation is simulated by a two-stage Stochastic Unit Commitment (SUC) framework. This SUC obtains the optimal scheduling and dispatch for the system generators in a 24h-horizon, using the forecasted RES generation scenarios as an input. For each time-step in the two-stage SUC, the first stage corresponds to the Day-Ahead (DA) decisions, while the second stage is the Real-Time (RT) decision. In the DA, the commitment decision for slow generators (which cannot suddenly start generating in RT) is fixed for all scenarios. The RT decision corresponds to the dispatch of online slow units, the commitment state and dispatch of fast generators, the RES curtailment and the load shedding. The computational efficiency of the SUC model is a key element for its practical applications. High-quality scenario generation may reduce the number of scenarios that are needed to describe the uncertainty, leading to reduced computational time for SUC.

6.1. Mathematical Formulation

The SUC minimizes the expected operational cost over all scenarios and time-steps:

$$\min \sum_{t\in[1,\mathrm{T}]} \sum_{\omega\in[1,\Omega]} p_{\omega} \left(\sum_{g\in\mathcal{G}} C^{g}_{\omega,t} + C^{\mathrm{LS}}_{\omega,t} \right) \quad (36)$$

Where the operating cost for each generator $C^{g}_{\omega,t}$ and the load-shedding cost $C^{\mathrm{LS}}_{\omega,t}$ are defined as:

$$C^{g}_{\omega,t} = \mathrm{c}^{g}_{\mathrm{st}} \cdot z^{g}_{\omega,t} + \tau_t \left(\mathrm{c}^{g}_{\mathrm{nl}} \cdot y^{g}_{\omega,t} + \mathrm{c}^{g}_{\mathrm{m}} \cdot P^{g}_{\omega,t} \right) \quad (37)$$

$$C^{\mathrm{LS}}_{\omega,t} = \tau_t \cdot \mathrm{c}^{\mathrm{LS}} \cdot P^{\mathrm{LS}}_{\omega,t} \quad (38)$$

The problem is subject to the following constraints:

$$\sum_{g\in\mathcal{G}} P^{g}_{\omega,t} + \mathrm{P}^{\mathrm{R}}_{\omega,t} - P^{\mathrm{RC}}_{\omega,t} = \mathrm{P}^{\mathrm{D}}_{\omega,t} - P^{\mathrm{LS}}_{\omega,t} \quad \forall\omega,t \quad (39)$$

$$y^{g}_{\omega,t} \cdot \mathrm{P}^{g}_{\mathrm{msg}} \leq P^{g}_{\omega,t} \leq y^{g}_{\omega,t} \cdot \mathrm{P}^{g}_{\mathrm{max}} \quad \forall g,\omega,t \quad (40)$$

$$-\tau_{t-1} \cdot \mathrm{P}^{g}_{\mathrm{rd}} \leq P^{g}_{\omega,t} - P^{g}_{\omega,t-1} \leq \tau_{t-1} \cdot \mathrm{P}^{g}_{\mathrm{ru}} \quad \forall g,\omega,t \quad (41)$$

$$z^{g}_{\omega,t} \geq y^{g}_{\omega,t} - y^{g}_{\omega,t-1} \quad \forall g,\omega,t \quad (42a)$$

$$z^{g}_{\omega,t} \geq 0 \quad \forall g,\omega,t \quad (42b)$$

$$y^{g}_{\omega,t} = y^{g}_{1,t} \quad \forall g \in \mathcal{O}, \forall\omega,t \quad (43)$$

Constraint (39) enforces the power balance, (40) enforces generation limits, (41) enforces ramp limits, (42a)-(42b) define the startup state of generators and (43) is the non-anticipativity condition, which fixes the commitment decision for slow generators in the DA. Frequency response modeling is not considered in this model, as the key objective is to analyze the quality of the generated scenario by alternative methods. Nevertheless, it is straightforward to include the various forms of frequency response constraints.

7. Evaluation of the methodology

In this section, we evaluate the proposed methodology, starting with the forecasting of the aggregated production and the generation of scenarios. The obtained scenarios are then applied to the two case studies, namely reserve bidding and stochastic scheduling.

7.1. Forecasting of Aggregated Production

We forecast the production of a following VPP for the day ahead comprising 3 Wind farms, 3 small Hydro plants and 9 PV farms, all located in France within the same control area but with different climates. To assess the sensitivity of our generation method to the relative proportion of each energy source (Wind, PV, Hydro), we scale the installed capacities of the farms to obtain two different VPP configurations in Table 1. The first VPP (VPP1) is dominated by Wind, whereas the second VPP (VPP2) is

Configuration	Wind	PV	Hydro
VPP1	32	9	12
VPP2	12.4	36	4.6

Table 1: Installed capacities of VPP configurations in MW

dominated by PV. VPP2 shows the same capacity ratio between Wind and Hydro as for VPP1.

The following NWP are retrieved from the ECMWF forecasting center: for the geographical location of each PV plant, surface solar radiation oriented downwards, total cloud coverage, hourly rainfall; for each Wind plant, zonal and meridional wind speeds at 10 m; and for each Hydro plant, daily cumulated rainfall, surface solar radiation downwards, air temperature at 2 m. The production forecast is assumed to be done before noon when the day-ahead energy market closes, so we use NWP issued at 00h00 of the previous day. The resulting forecasting horizon is thus comprised between 24h and 48h. We forecast the production of all plants of the same source (PV, Wind and Hydro) with the QRF model. Then, we forecast the aggregated production (53 MW capacity) directly, this time without preprocessing of the production variables to preserve coherence between energy sources.

The QRF model is trained on 6 horizon intervals of 4 hours each, and evaluated using a 7-fold cross-validation: one day of the week is chosen alternatively as the test set on which the forecasting is evaluated and the model is trained on the remaining days of the week. The available dataset consists in 200 days of power production between June 2015 and March 2016, with a 30-min resolution for each plant (9600 points). The reliability diagram in Figure 2 indicates that direct forecasting of the aggregated production of VPP1 shows adequate reliability. The separate forecasting is also reliable: the observed frequencies of the forecast distribution transform for Hydro (see Figure 3) are distributed almost uniformly (a similar performance is obtained for PV and Wind, not shown).

The Root Mean Square Error (RMSE) and the Continuous Ranked Probability Score (CRPS) in Table 2 indicate that the QRF model has comparable performance with respect to the state of the art in forecasting of Wind and PV production [36]. Note that the direct forecast of the aggregated production shows slightly lower errors than the forecasts for each energy source, because the aggregated production has a smoother profile and because the QRF model is able to learn from explanatory variables of large dimension, even if they reflect different energy sources. The VPP dominated by PV (VPP2) shows lower forecasting error at night because the total available capacity is lower (no PV production).

7.2. Generation of Scenarios

The covariance matrix obtained from separate forecasts on each energy source show in Figure 4 that the correlations between energy sources are low but existing: positive correlations between 0.2 and 0.5 can be observed between Wind and Hydro around noon, while PV and Wind alternate between low positive and negative correlations. Negative correlations are interesting for the present application: a source ramping up (eg: Wind speed rising at a given Wind farm) can potentially substitute a drop in production of another source (eg: cloud passing upon a PV plant).

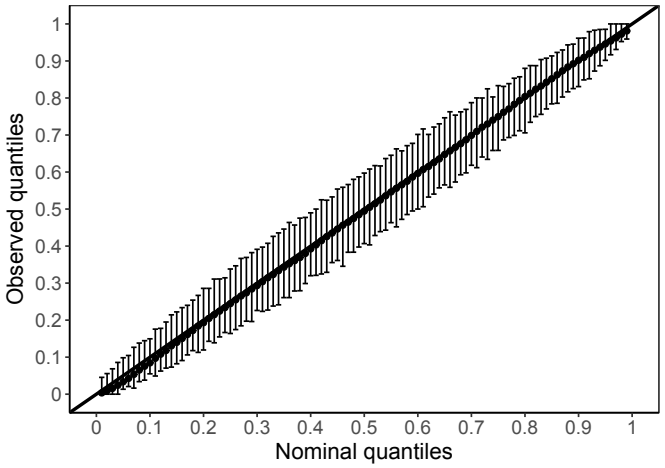

Figure 2: Reliability diagram for the direct forecast of the total aggregated production in VPP1. Consistency bars are 5%-quantile and 95%-quantile of resampled forecasts

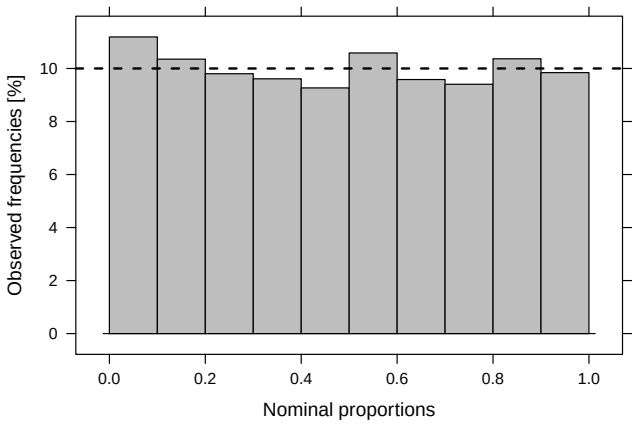

Figure 3: Observed frequencies of distribution transform for Hydro plants in VPP1. The dashed line represents an ideal uniform distribution

VPP	Level	NRMSE	NCRPS			
-	-	(0-24h)	(0-6h)	(6-12h)	(12-18h)	(18-24h)
VPP1	PV	0.045	–	0.046	0.048	–
VPP1	Wind	0.095	0.044	0.045	0.048	0.051
VPP1	Hydro	0.081	0.032	0.047	0.044	0.044
VPP1	Aggreg.	0.065	0.030	0.036	0.036	0.035
VPP2	Aggreg.	0.046	0.012	0.036	0.037	0.014

Table 2: Forecasting results, normalized by installed capacity. NRMSE average on horizon, NCRPS by intervals of day

Properties and scores of scenarios for VPP1 and VPP2 obtained by the scenarios from DG and IG methods, with Gaussian copula, are displayed as a function of the hourly horizon in Figure 5 (results of DG scenarios are represented with solid lines and denoted as "aggregated", results of IG scenarios are represented with dashed lines and denoted as "separate"). The mean values of scenarios are close for both methods on VPP1 and VPP2. This proves that the methods DG and IG are coherent, as expected: the summed expectation of energy sources is equal to the expectation of the sum. The amplitude of the scenario set (difference between minimum and maximum values of scenarios at each time step) is lower for the IG method, for both VPPs. This is due to the fact that extreme aggregated production levels observed during training have been considered directly in the dependence model of the DG method, whereas they are only reconstructed a posteriori by the IG method. In terms of bias, the IG method is more biased around noon when PV production is maximum, which is probably related to a higher bias in the separate PV forecast.

For VPP1 (wind-dominated), the average VS of scenarios is of similar level for both DG and IG methods. For VPP2 (PV-dominated), the VS of the IG approach is significantly lower, which indicates a more realistic variability of scenarios from separate forecasts by source. The aggregated production profile depends here more on the horizon than for the wind-dominated case. Despite the lower performance of separate probabilistic forecasts (cf. 2), the IG method compensates with a high number of possible combinations between sources. In contrast, the covariance of the DG method summarizes the variability with less versatility. In addition, the DG method ignores the saturations occuring for each source, which creates an underestimation of the smoothing effect.

Concerning gradients of aggregated production, the corresponding BS are reported in Figure 5 for intervals of 1h to 4h. The events evaluated by the BS are gradients that exceed the average gradient values for each interval. For VPP1, we observe that IG scores are slightly better than DG except for the 1-hour gradient at night. The 1-h auto-correlation is higher for Hydro than for Wind during this period, and the IG method seems to overestimate the weight of Hydro in this case. The auto-correlations are lower at further lags, so this effect disappears for BS at intervals superior to 1 h. For VPP2, where dependence on the horizon is more pronounced, the DG method has a better BS when production is high because it considers more extreme production, and worse BS at night when production is more stable.

Finally, the distribution of errors for VPP2 in Figure 6 shows that the variance of the error is slightly reduced for the IG method compared to the DG method. In summary, scenarios from separate forecasts show better average properties than scenarios from direct aggregated forecast for stochastic optimization, especially when the aggregated production profile depends on the horizon (e.g. high PV share in the aggregation).

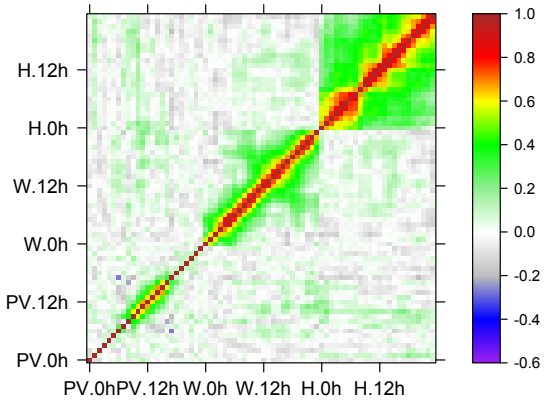

Figure 4: Covariance matrix of the multivariate Gaussian Copula, in the rank domain of Wind (W), Hydro (H) and PV (PV) for each hour of the day

7.3. Case Study 1: Reserve Bidding

The optimization is computed for 100 days using price data from Portugal, with 100 trajectories of aggregated production of VPP1, generated from the Direct Aggregated production forecast and Gaussian Copula (DG), the Indirect aggregated production forecast and Gaussian Copula (IG) and Vine Copula (IV). The M coefficient is set to a number close to the aggregated installed capacity (60 MW vs 53 MW) and ϵ is set to 1%. We consider a moderately risk-averse VPP, $\beta = 0.5$. The penalties for not supplying reserve are taken equal to two times the price of restoration reserve (instead of 1.5 times in current Portuguese rules). These high penalties lead the optimization model to face higher penalties on the imbalance energy market in order to be able to supply reserve.

The frequency of reserve deficit simulated by the model is null for all scenarios, which was not the case for a model without chance-constraints. Economic and technical results are reported in Table 3. The use of the direct aggregated forecast, which generates scenarios with higher amplitudes, reduces the amount of reserve offered to avoid penalties for reserve underfulfillment. Considering that in the case study the day-ahead price for energy is higher than the day-ahead price for reserve, the average revenue obtained in the objective function increases up to 6% when comparing with the scenarios from separated forecasts. In contrast, scenarios from separate forecasting give a higher 5%-Conditional Value at Risk (CVaR), up to 18%. This is associated with a more conservative bidding in the day-ahead market: more reserve capacity and less energy is offered. The CVaR increases as less penalties are to be paid for energy deficit in real-time. This is due to the lower Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) of scenarios based on separate forecasting.

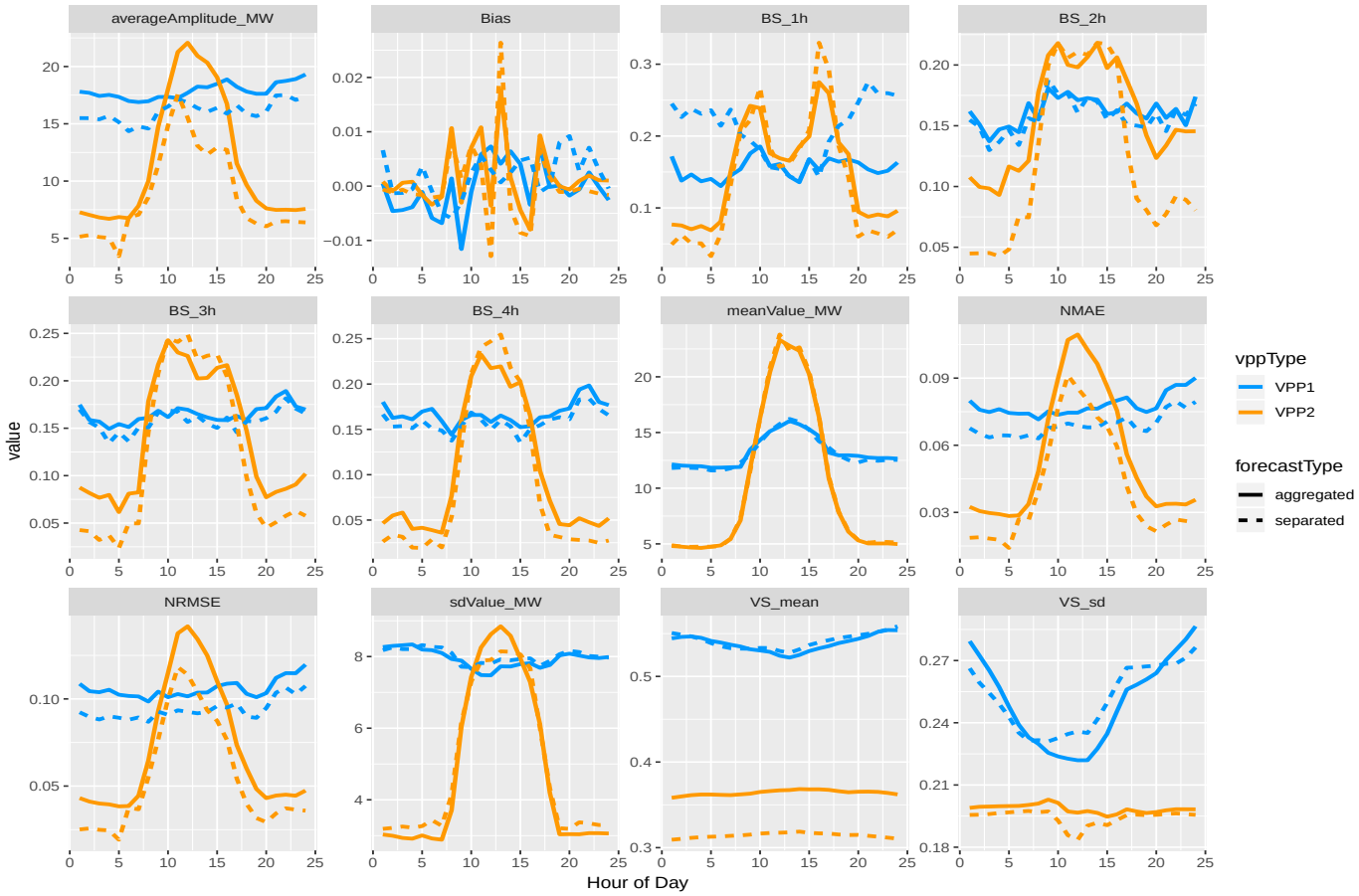

Figure 5: Properties and scores of scenarios generated with Gaussian Copula, directly aggregated from the DG method ("aggregated", solid line) and indirectly aggregated by separate production forecast from the IG method ("separated", dashed-line). Starting at top-left: average Amplitude of the scenario set, in MW; average bias of scenarios, scaled by installed capacity; Brier Score for ramps on several intervals, from 1 hour to 4 hours; mean value of scenarios, in MW; Normalized Mean Average Error of scenarios (NMAE), scaled by installed capacity; Normalized Root Mean Square Error of scenarios (NMRSE), scaled by installed capacity; standard deviation of the scenarios, in MW; mean value of the Variogram Score with moment = 0.5; standard deviation of the Variogram Score with moment = 0.5

Lastly, the flexible dependence model of the Vine Copula generates more extreme values of aggregated production from the separate forecast than the Gaussian Copula. In return, the results for the method IV is intermediary between the results of DG and IG: higher average revenue than IG but lower CVar.

Scenarios	$o_{\beta,T}$	$CVar_{5\%}$	E^{da}	$E^{rt,-}$	$E^{rt,+}$	R^{da}
-	[€/MWh]	[€/MWh]	[MWh]	[MWh]	[MWh]	[% Pn]
DG	**73**	48	**9.2**	1.9	**5.2**	6.6
IG	71	**58**	7.9	**1.4**	6.1	**7.5**
IV	69	54	7.5	**1.4**	6.2	6.8

Table 3: Average profits and volumes of energy and reserve for the optimized bidding, depending on the scenario generation method

7.4. Case Study 2: Stochastic scheduling

To evaluate the impact of different forecasting methods on power system operation, a two-step approach is applied:

1. First, the two-stage SUC is used to obtain the optimal scheduling of the system for the forecasted RES-generation scenarios. After solving the optimization, the DA decisions are recorded.
2. Then, a real time dispatch programme is carried out over the real RES production data, with the DA decisions obtained in the first part fixed.

This two-step methodology is applied to evaluate the two scenario generation methods, namely the scenarios issued from aggregated wind, PV and hydro forecast proposed in this paper and the scenarios from separate production forecasts for each energy source. We use here the Gaussian copula in both cases to generate scenarios from the probabilistic forecasts.

A test case is designed to evaluate the two methods: Directly aggregated with Gaussian Copula (DG) and Indirectly aggregated with Gaussian Copula (IG). The characteristics of thermal generators are included in Table 4, where Combined Cycle Gas Turbines (CCGT) are treated as slow generators and Open Cycle Gas Turbines (OCGT)

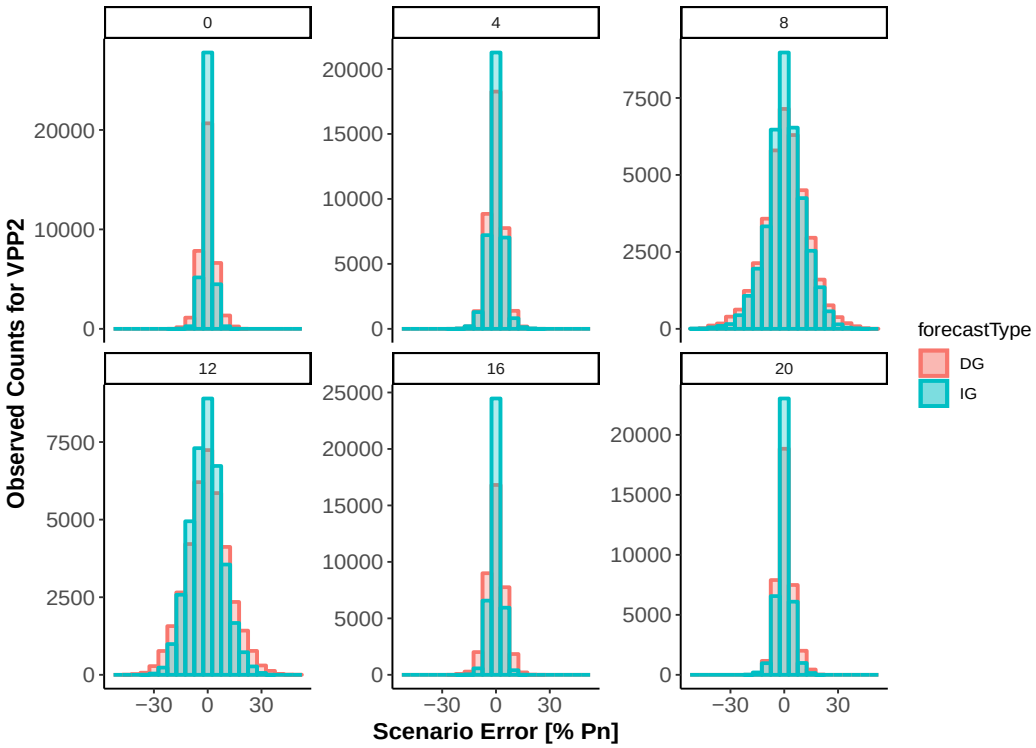

Figure 6: Distribution of scenario error in MW for VPP2, for selected horizons between 0h and 20h

	CCGT	OCGT
Number of Units	10	5
Rated Power (MW)	50	10
Min Stable Generation (MW)	30	5
No-Load Cost c_g^{nl} (£/h)	540	300
Marginal Cost c_g^{m} (£/MWh)	47	200
Startup Cost c_g^{st} (£)	1000	0

Table 4: Characteristics of thermal generators

as fast generators. The two VPPs in Table 1 are considered, while the same 100 RES-generation trajectories as in section 7.3 are used but scaled up by a factor of 10, corresponding then to installed capacities of 530MW. A scaled demand profile corresponding to Great Britain's equivalent consumption is used, with minimum and maximum demand of 200MW and 600MW, respectively. An MIP gap of 0.1% is considered for the optimizations and c^{LS} is set to 30k£/MWh.

A simulation spanning 60 consecutive days was carried out, for which the average daily net-demand (demand minus RES generation) was of roughly 290MWh for both VPPs. The average daily savings due to using DG when compared to IG for the stochastic scheduling are of 1.6% and 5.1%, for VPP1 and VPP2 respectively. Interestingly, IG achieves slightly lower operating costs for most of the days, but the overall advantage of DG is driven by a few incidents that high cost load-shedding happens when using IG. This is due to the fact that the lower forecasting variance in IG tends to lead less CCGTs to be started in the day-ahead. Although the bias is slight higher in IG, it is already covered, for most of the cases, by the online CCGTs. On the other hand, DG schedules more CCGTs due to the higher variance, and therefore typically induces higher part-loading and higher operation cost. However, there are few incidents with particularly low variance for IG, in which very few CCGTs are started up in the day ahead scheduling, and when it is combined with large positive error (much lower RES generation realised than forecasted), even to start up all OCGTs in the real-time operation is not sufficient to maintain the system operation without costly load-shedding. In summary, for the point of system scheduling, the main advantage of DG is from its capability to reduce costs while at the same time cover the extreme errors in the generated scenarios.

8. Conclusion

A QRF model directly derives a short-term probabilistic forecast of the aggregated production of dispersed Wind, PV and small hydro plants. The same model applied to a separate forecasting of the energy sources composing the aggregation results in higher forecast errors than the direct forecast of the aggregated production. This is due to the efficient learning capacity of the QRF in high dimension (many explanatory variables from multiple production sites), but also to the smoothing effect of aggregation which reduces the variability and intrinsically reduces the

uncertainty in production. An extension of this work could consist in forecasting the multi-source aggregated production with a deep quantile regression approach, with the objective to outperform the performance of the machine learning model used here and scale to potentially larger aggregations.

Scenarios of aggregated production have been generated from these probabilistic forecasts. The impact of choosing either a direct forecast of the aggregated production or a separate forecast for each energy source is assessed on two power system applications requiring such scenarios. A multivariate copula, of type either Gaussian or Vine, is applied to obtain scenarios from the day-ahead production forecasts. The generation methods are evaluated on two different aggregations with different capacity shares for the different energy sources.

It is found on one hand that scenarios from the aggregated forecast have a higher amplitude, because the scenario generation method incorporates directly past extreme levels of aggregated production, meanwhile scenarios from separate forecasts reconstruct extreme aggregated levels from the marginal contributions of each source in the aggregation. On the other hand, scenarios issued from the separate forecasting of each source reproduce more accurately the variability of the aggregated production when the production is highly dependent on the horizon, for instance when PV is dominant in the total production. This is quantified by a sensible gain on the Variogram Score for the PV-dominated aggregation, whereas the gain is close to zero for a Wind-dominated aggregation. This is due to a better reproduction of the various conditions leading to the observed smoothing effect in aggregated production. In this case the available information about the variability for each source is preserved, whereas a given level of total production can arise from multiple distinct production patterns at the level of each source, hence loosing information on the diversity of production profiles.

When used in both stochastic applications of reserve bidding and unit commitment, scenarios from direct aggregated production forecast generate more average value (increased profit or reduced costs) than those from separate forecasts, because these applications are sensible to the amplitude of the scenario set. Extreme levels of aggregated production are more present in the scenarios from direct aggregated production forecast, which can secure highly risk-averse decisions (e.g. unit commitment under extreme RES aggregated production) but also hinders decisions that could be valuable for the agent (e.g. scenarios of high amplitude will limit the offer of reserve of an aggregator, with a possible opportunity cost if activated reserve would have increased his revenue). Finally, a moderately risk-averse decision maker will observe that scenarios generated from separate forecasts create less penalties due to their sharper distribution and more realistic variability: an aggregator bidding AS and energy will increase his Conditional Value -at-Risk, and a system operator will decrease his lower operational costs for most days of the year. In conclusion, the assessment of scenario generation methodologies in this paper informs the decision-makers on the optimal way to generate short-term scenarios of aggregated RES production according to their risk aversion and to the contribution of each source in the aggregation.

References

[1] M. Yin, Y. Xu, C. Shen, J. Liu, Z. Y. Dong, Y. Zou, Turbine Stability-Constrained Available Wind Power of Variable Speed Wind Turbines for Active Power Control, IEEE Transactions on Power Systems 32 (3) (2017) 2487–2488. doi:10.1109/TPWRS.2016.2605012.

[2] L. Spitalny, D. Unger, J. M. A. Myrzik, Potential of small hydro power plants for delivering control energy in Germany, in: 2012 IEEE Energytech, IEEE, 2012, pp. 1–6. doi:10.1109/EnergyTech.2012.6304700.
URL http://ieeexplore.ieee.org/document/6304700/

[3] A. F. Hoke, M. Shirazi, S. Chakraborty, E. Muljadi, D. Maksimovic, Rapid Active Power Control of Photovoltaic Systems for Grid Frequency Support, IEEE Journal of Emerging and Selected Topics in Power Electronics 5 (3) (2017) 1154–1163. doi:10.1109/JESTPE.2017.2669299.

[4] A. Gonzalez-Garrido, A. Saez-de Ibarra, H. Gaztanaga, A. Milo, P. Eguia, Annual Optimized Bidding and Operation Strategy in Energy and Secondary Reserve Markets for Solar Plants with Storage Systems, IEEE Transactions on Power Systems 8950 (c) (2018) 1–10. doi:10.1109/TPWRS.2018.2869626.

[5] S. Camal, A. Michiorri, G. Kariniotakis, Optimal offer of automatic frequency restoration reserve from a combined PV/Wind Virtual Power Plant, IEEE Transactions on Power Systems 33 (6) (2018) 6155–6170. doi:10.1109/TPWRS.2018.2847239.

[6] K. Knorr, B. Zimmermann, M. Speckmann, M. Wunderlich, D. Kirchner, F. Steinke, P. Wolfrum, T. Leveringhaus, T. Lager, L. Hofmann, D. Filzek, T. Göbel, B. Kusserow, L. Nicklaus, P. Ritter, Kombikraftwerk 2 Abschlussbericht (August) (2014) 218.
URL http://www.kombikraftwerk.de/fileadmin/Kombikraftwerk_2/Abschlussbericht/Abschlussbericht_Kombikraftwerk2_aug14.pdf

[7] L. Rubino, C. Capasso, O. Veneri, Review on plug-in electric vehicle charging architectures integrated with distributed energy sources for sustainable mobility, Applied Energy 207 (2017) 438–464. doi:10.1016/J.APENERGY.2017.06.097.
URL https://www.sciencedirect.com/science/article/pii/S0306261917308358

[8] D. Dallinger, M. Wietschel, Grid integration of intermittent renewable energy sources using price-responsive plug-in electric vehicles (6 2012). doi:10.1016/j.rser.2012.02.019.
URL https://www.sciencedirect.com/science/article/pii/S136403211200113X?via%3Dihub

[9] O. Veneri, Technologies and Applications for Smart Charging of Electric and Plug-in Hybrid Vehicles, Springer, 2017.

[10] L. Hirth, I. Ziegenhagen, Control power and variable renewables, International Conference on the European Energy Market, EEM 50 (2013) 1035–1051. doi:10.1109/EEM.2013.6607359.
URL http://dx.doi.org/10.1016/j.rser.2015.04.180

[11] T. Soares, P. Pinson, T. V. Jensen, H. Morais, Optimal Offering Strategies for Wind Power in Energy and Primary Reserve Markets, IEEE Transactions on Sustainable Energy 7 (3) (2016) 1036–1045. doi:10.1109/TSTE.2016.2516767.

[12] T. Soares, P. Pinson, Renewable energy sources offering flexibility through electricity markets, Ph.D. thesis, Technical University of Denmark (2017).

[13] H. Fu, Z. Wu, X.-P. Zhang, J. Brandt, Contributing to DSOs Energy-Reserve Pool: A Chance-Constrained Two-Stage μVPP Bidding Strategy, IEEE Power and Energy Technology Systems Journal 4 (4) (2017) 1–1. doi:10.1109/JPETS.2017.2749256.
URL http://ieeexplore.ieee.org/document/8058436/

[14] H. Zhang, Z. Hu, E. Munsing, S. J. Moura, Y. Song, Data-driven Chance-constrained Regulation Capacity Offering for Distributed Energy Resources, IEEE Transactions on Smart Griddoi:10.1109/TSG.2018.2809046.
URL https://arxiv.org/pdf/1708.05114.pdf

[15] R. Pinto, R. J. Bessa, M. A. Matos, Multi-period flexibility forecast for low voltage prosumers, Energy 141 (2017) 2251–2263. doi:10.1016/j.energy.2017.11.142.
URL https://repositorio.inesctec.pt/bitstream/123456789/4767/3/P-00N-9Q4.pdf

[16] P. Meibom, R. Barth, B. Hasche, H. Brand, C. Weber, M. O'Malley, Stochastic Optimization Model to Study the Operational Impacts of High Wind Penetrations in Ireland, IEEE Transactions on Power Systems 26 (3) (2011) 1367–1379. doi:10.1109/TPWRS.2010.2070848.
URL http://ieeexplore.ieee.org/document/5587912/

[17] F. Bouffard, F. Galiana, Stochastic Security for Operations Planning With Significant Wind Power Generation, IEEE Transactions on Power Systems 23 (2) (2008) 306–316. doi:10.1109/TPWRS.2008.919318.
URL http://ieeexplore.ieee.org/document/4470561/

[18] J. M. Morales, A. J. Conejo, J. Perez-Ruiz, Economic valuation of reserves in power systems with high penetration of wind power, in: 2009 IEEE Power & Energy Society General Meeting, IEEE, 2009, pp. 1–1. doi:10.1109/PES.2009.5260229.
URL http://ieeexplore.ieee.org/document/5260229/

[19] A. Tuohy, P. Meibom, E. Denny, M. O'Malley, Unit Commitment for Systems With Significant Wind Penetration, IEEE Transactions on Power Systems 24 (2) (2009) 592–601. doi:10.1109/TPWRS.2009.2016470.

[20] F. Teng, G. Strbac, Full Stochastic Scheduling for Low-Carbon Electricity Systems, IEEE Transactions on Automation Science and Engineering 14 (2) (2017) 461–470. doi:10.1109/TASE.2016.2629479.
URL http://ieeexplore.ieee.org/document/7833096/

[21] K. Bruninx, Improved modeling of unit commitment decisions under uncertainty, Ph.D. thesis (2016).
URL https://www.mech.kuleuven.be/en/tme/research/energy_environment/Pdf/doctoraat-kenneth-bruninx.pdf

[22] P. Pinson, R. Girard, Evaluating the quality of scenarios of short-term wind power generation, Applied Energy 96 (2012) 12–20. doi:10.1016/j.apenergy.2011.11.004.

[23] F. Golestaneh, H. B. Gooi, P. Pinson, Generation and evaluation of spacetime trajectories of photovoltaic power, Applied Energy 176 (2016) 80–91. doi:10.1016/J.APENERGY.2016.05.025.
URL https://www.sciencedirect.com/science/article/pii/S0306261916306079

[24] Z. Wang, W. Wang, C. Liu, Z. Wang, Y. Hou, Probabilistic Forecast for Multiple Wind Farms Based on Regular Vine Copulas, IEEE Transactions on Power Systems 33 (1) (2018) 578–589. doi:10.1109/TPWRS.2017.2690297.
URL http://ieeexplore.ieee.org/document/7909035/

[25] W. Wu, K. Wang, B. Han, G. Li, X. Jiang, M. L. Crow, A Versatile Probability Model of Photovoltaic Generation Using Pair Copula Construction, IEEE Transactions on Sustainable Energy 6 (4) (2015) 1337–1345. doi:10.1109/TSTE.2015.2434934.

[26] M. P. Côté, C. Genest, A copula-based risk aggregation model, Canadian Journal of Statistics 43 (1) (2015) 60–81. doi:10.1002/cjs.11238.
URL http://doi.wiley.com/10.1002/cjs.11238

[27] J. Dong, T. Kuruganti, S. M. Djouadi, Very Short-term Photovoltaic Power Forecasting using Uncertain Basis Function, Information Sciences and Systems (CISS), 2017 51st Annual Conference on (2017) 1–6.

[28] S. I. Vagropoulos, E. G. Kardakos, C. K. Simoglou, A. G. Bakirtzis, J. P. S. Catalão, ANN-based scenario generation methodology for stochastic variables of electric power systems, Electric Power Systems Research 134 (2016) 9–18. doi:10.1016/j.epsr.2015.12.020.
URL https://ac.els-cdn.com/S0378779615003971/1-s2.0-S0378779615003971-main.pdf?_tid=28d90c9a-f6b3-11e7-b494-00000aacb361&acdnat=1515663681_181743d500e10436507caa9e7c9cf4be

[29] Y. Chen, Y. Wang, D. Kirschen, B. Zhang, Model-Free Renewable Scenario Generation Using Generative Adversarial Networks, IEEE Transactions on Power Systems 33 (3) (2018) 3265–3275. doi:10.1109/TPWRS.2018.2794541.
URL http://ieeexplore.ieee.org/document/8260947/http://arxiv.org/abs/1707.09676

[30] N. T. Tung, J. Z. Huang, Thuy Thi Nguyen, I. Khan, Bias-corrected Quantile Regression Forests for high-dimensional data, in: 2014 International Conference on Machine Learning and Cybernetics, IEEE, 2014, pp. 1–6. doi:10.1109/ICMLC.2014.7009082.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7009082

[31] Y. Zhang, J. Wang, X. Wang, Review on probabilistic forecasting of wind power generation, Renewable and Sustainable Energy Reviews 32 (2014) 255–270. doi:10.1016/j.rser.2014.01.033.
URL http://dx.doi.org/10.1016/j.rser.2014.01.033

[32] L. Rüschendorf, Mathematical Risk Analysis, Vol. 2013, 2013. doi:10.1007/978-3-642-33590-7.
URL http://link.springer.com/10.1007/978-3-642-33590-7

[33] M. Scheuerer, T. M. Hamill, Variogram-Based Proper Scoring Rules for Probabilistic Forecasts of Multivariate Quantities*, Monthly Weather Review 143 (4) (2015) 1321–1334. doi:10.1175/MWR-D-14-00269.1.
URL https://www.esrl.noaa.gov/psd/people/michael.scheuerer/variogram-score.pdfhttp://journals.ametsoc.org/doi/10.1175/MWR-D-14-00269.1

[34] A. E. R. T. T. B. 50hz, Amprion, Consultation on the design of the platform for automatic Frequency Restoration Reserve (aFRR) of PICASSO region.
URL https://www.entsoe.eu/Documents/Networkcodesdocuments/Implementation/picasso/PICASSO-Consultation_document.pdf?Web=1

[35] C. Liu, X. Wang, Y. Zou, H. Zhang, W. Zhang, A probabilistic chance-constrained day-ahead scheduling model for grid-connected microgrid, in: 2017 North American Power Symposium, NAPS 2017, no. 51577146, 2017. doi:10.1109/NAPS.2017.8107180.

[36] G. Kariniotakis, Renewable Energy Forecasting - From Models to Applications, Woodhead Publishing Series in Energy, Woodhead Publishing Series in Energy - Elsevier, 2017.