Reliable Provision of Ancillary Services from Aggregated Variable Renewable Energy Sources through Forecasting of Extreme Quantiles

Simon Camal, Andrea Michiorri, George Kariniotakis

▶ **To cite this version:**

Simon Camal, Andrea Michiorri, George Kariniotakis. Reliable Provision of Ancillary Services from Aggregated Variable Renewable Energy Sources through Forecasting of Extreme Quantiles. IEEE Transactions on Power Systems, 2023, 38 (4), pp.3070-3084. ⟨10.1109/TPWRS.2022.3198839⟩. ⟨hal-03759187v2⟩

HAL Id: hal-03759187

https://hal.science/hal-03759187v2

Submitted on 2 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

HAL Authorization

Reliable Provision of Ancillary Services from Aggregated Variable Renewable Energy Sources through Forecasting of Extreme Quantiles

Simon Camal *Member, IEEE*, Andrea Michiorri, and Georges Kariniotakis, *Senior Member, IEEE*

***Abstract*—Virtual power plants aggregating multiple renewable energy sources such as Photovoltaics and Wind are promising candidates for the provision of balancing ancillary services. A requisite for the provision of these services is that forecasts of aggregated production need to be highly reliable in order to minimize the risk of not providing the service. Yet, a reliability greater than 99% is unattainable for standard forecasting models. This work proposes alternative models for the day-ahead prediction of the lowest quantiles (0.1% to 0.9 %) of renewable Virtual power plant production. The proposed approaches derive conditional quantile forecasts of aggregated Wind/PV/Hydro production, obtained from tailored parametric models and machine learning models, including a Convolutional Neural Network architecture for predicting extremes. Reliability deviation is reduced up to 50 % and probabilistic skill score up to 18% compared to Quantile Regression Forest. Forecasting models are subsequently applied to the provision of downward reserve capacity by a renewable Virtual power plant. Increased forecasting reliability leads to a higher reliability of the reserve capacity, but reduces the average reserve volume offered by the renewable aggregation.**

***Index Terms*—Aggregation, Extremes, Forecasting, Reliability, Renewables, Reserve, Virtual Power Plants**

I. Nomenclature

Variables

$\hat{.}$	Predicted variable
ϵ	Exceedances from reference quantile
$\mathcal{D}$	Generic term for distribution
$F_.$	CDF forecast of production
f	Filter of convolutional neural network
Φ	Feature map of convolutional neural network
$\mathcal{L}$	Likelihood function
R	Reserve bid
x	Features of aggregated production forecast
y	Aggregated renewable production
u	Threshold value of Extreme Value Theory
z	Indicator variable of components in mixture

Parameters and indices

α^0	Rate of Dirichlet prior distribution
c	Clustering index
c^0	Concentration of Gamma prior distribution
C	Number of conditioning partitions
d	Index of explanatory feature
D	Number of explanatory features
ΔT	Validity period of reserve bid
ϕ	Precision of marginal component in mixture
$\mathcal{F}_.$	Set of convolutional feature maps
γ	Shape of Generalized Pareto distribution
kn	Kernel size of convolutional filter or min-pooling
k	Threshold of Extreme Value Theory
h	Prediction horizon
$\mathcal{H}$	Set of prediction horizons
λ	Mean of component in Beta mixture
m	Index of marginal component in mixture
n_τ	Number of quantiles in the quantile interval
M	Number of marginal components in mixture
μ	Mean of marginal component in Gaussian mixture
μ^0	Mean of prior distribution
ω	Inter-mean distance of Beta mixture components
π	Probability of component in mixture
p	Index of plant in aggregation
$\mathcal{P}$	Number of plants in aggregation
ρ	Thickness of exponential distribution
r^0	Rate of Gamma prior distribution
s_u, s_v	Stride parameters of convolutional filter
$\mathcal{S}$	Number of energy sources in aggregation
σ	Scale of Generalized Pareto distribution
t	Forecasting runtime
T	Length of evaluation period
$\mathcal{T}$	Prediction set
θ	Parameter vector
τ	Quantile value of production forecasts
τ_R	Maximum accepted frequency of underfulfillments
v^0	Variance of prior distribution

The authors are with the Centre for Processes, Renewable Energies and Energy Systems (PERSEE), MINES Paris - PSL University, Sophia Antipolis, France, e-mail: (name.surname@minesparis.psl.eu).

The present research was carried out as part of the REstable Project (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, and Smart4RES Project (European Union's Horizon 2020, No. 864337). The sole responsibility of this publication lies with the author. The European Union and INEA are not responsible for any use that may be made of the information contained therein.

II. Introduction

A. Motivation

Power system operators expect Variable Renewable Energy (VRE) to contribute to the stability of power systems by providing frequency control through Ancillary Services (AS), which are increasingly procured by dedicated markets [1]. Wind farms have already demonstrated their technical

capacity to provide AS such as frequency control [2], while photovoltaics (PV) plants can also be equipped with suitable systems for active power regulation [3]. However, uncertainties regarding weather conditions make it difficult for a single Wind or PV plant to guarantee power reserve capacities. In contrast, aggregated Wind and PV plants dispersed over different climate zones and controlled by a Virtual Power Plant (VPP) are able to provide significant volumes of active power reserve [4].

In order to be accepted by TSOs, a reserve capacity offer of a VRE-based VPP must have a reliability level close to that observed for a single dispatchable plant (i.e. in the range of 99%-99.9%). Without the contribution of storage to compensate for deviations between the offered volume of reserve and the actual observed capacity, reserve offers rely on a highly reliable forecast of the total production of the VRE-VPP. This can be done by predicting very low quantiles of the expected distribution of the VRE-VPP production with nominal values ranging from 1% to 0.1%. It is known that state-of-the-art probabilistic forecasting models are accurate on non-extremal quantiles (e.g. between 5% and 95%, [5]) but are unreliable for extremal quantiles. This behaviour has been demonstrated on the case of quantile regression models (see for instance [6]) and explained by the challenges posed by the sparsity of the data useful for the modelling of extremes. The effect of sparsity is amplified when considering that conditioning the prediction model on contextual information e.g. expected weather variables is beneficial for the prediction of extremes. The present paper investigates alternative approaches to improve the forecasting performance on very low quantiles of aggregated VRE production. A correlated question addressed in this paper is to assess if an improved forecasting performance on low quantiles translates into a higher value for reserve provision from a VRE-VPP (i.e. here higher reliability of reserve capacities).

In the existing forecasting literature, two methods are generally put forward to improve the forecast of very low quantiles: either develop parametric models specifically adapted to extremes, or adjust non-parametric models so that they can more efficiently learn rare patterns. The next section presents related works on these two approaches in the context of VRE production forecasting.

B. Related Research

The issue of improving forecast reliability for extreme values is increasingly important in power systems. Recently solutions have been proposed for different applications such as setting transmission transfer capacity [5], forecasting overhead conductor ampacity [7] and wind production [6].

The first family of models for forecasting extremes consists in censoring the forecast Cumulative Density Function (CDF) using a parametric distribution specifically adapted to the tail of the distribution. A simple example is given by the exponential distribution. In [5], very high quantiles of VRE production are modelled by an exponential distribution. An extension of this methodology is proposed by [8] in the context of Dynamic Line Rating, who fit the distribution on clusters of quantile pairs. The principal weakness of this approach is its lack of theoretical background, in particular its convergence properties.

The Extreme Value Theory (EVT) proposes a more robust framework for the prediction of extremes, which are modelled with a Pareto distribution [9]. Few publications have dealt with EVT in the context of renewable production forecasting, with the exceptions of [5], [6] and [10] who propose EVT forecasts of extremal quantiles of VRE production. In [6], a gradient boosting model predicts non extremal quantiles that are further ranked by a similarity measure. Extreme levels of these predictions serve to fit the Pareto distribution modelling the distribution tail. In [10], extreme quantiles of net-load are predicted by integrating EVT into an additive model that is conditioned on a grid of weather prediction for the region of interest.

In the context of renewable power forecasting, it is established that the Gaussian distribution is not suitable to model the uncertainty in production [11]. For better statistical behaviour, an alternative parametric approach consists of the inference of a mixture of Gaussian distributions as done by the Mixture Density Network (MDN) of [12]. The MDN combines a mixture of Gaussian distributions with a non-linear relationship between explanatory variables and response. However the performance of the MDN on low quantiles was not in the scope of this study. A mixture better adapted to the prediction of extremes than the Gaussian Mixture is the Beta mixture proposed by [13] to predict distribution tails for processes bounded in the $[0, 1]$ interval. The regression case studies of this work have low dimensionality and do not relate to forecasting or power systems, so it appears difficult to directly apply this approach to the problem tackled here.

As an alternative to parametric models, non-parametric machine learning models have recently emerged that rely on neural networks specifically designed for an improved prediction of extremes. As an example, the approach proposed by [14] derives an auto-encoder stacking Long Short Term Memory (LSTM) layers, which improves the detection of extreme situations compared to a univariate time series model. In [15], neural networks model different wind production events that can be classified as extreme for power system operation, such as large production ramps, severe variability and NWP failures to anticipate the weather front. Both of the machine learning models presented above integrate extreme events, but do not specifically address the problem of probabilistic regression on very low quantiles.

Finally, convolutional Neural Networks (CNNs) constitute another type of neural network that have been successfully implemented in the context of PV [16] and Wind [17] production forecasting. They have also been used to detect extreme weather patterns [18], but to the authors' knowledge have not been applied to forecasting extremes of renewable production.

C. Key Contributions of the Present Work

The research question posed by this work is how to adapt candidate forecasting models for low quantile predictions, so that they reach the required performance for reserve provision

by a VRE-VPP operating without storage. The question is answered by an adaptation methodology that enables forecasting practitioners (either developers or end-users) to select the type of model as a function of the obtained performance in terms of forecasting score and value for reserve provision. The simplicity and interpretability potential of the different model types may be additional selection criteria. The main contributions of this paper are:

1) A **methodology to forecast low quantiles of VRE-based VPP production**, exploring alternative approaches based on statistical and machine learning models. Existing forecasting approaches for low quantiles [6], [10] address problems of low dimensionality (single VRE source, load). However they do not scale well with the size and heterogeneous type of explanatory variables required to predict the production of a VRE-VPP. Also, these papers present approaches that are centered on parametric approaches for low quantiles. The methodology proposed here encompasses both parametric and neural-network approaches, therefore covering a broader scope of forecasting models.
2) The methodology is validated on the complex case of a **VPP composed of three different types of renewable energy source**, which makes the forecasting exercise more challenging: the different contributions of energy sources to the aggregate VPP production vary over time, and the influence of available data at the site level on low quantiles of production is not well known. This VPP configuration corresponds to a real world problematic but it is not a standard forecasting case in the abundant literature on VRE forecasting.
3) **Specific developments are proposed to adapt three families of forecasting models** to produce accurate prediction of low quantiles of aggregated VRE production. **Conditional statistical models** based on exponential distributions and EVT extend existing approaches dealing with single renewable energy sources e.g. [6], by selecting features that are meaningful to model the impact of uncertain weather conditions on low quantiles of aggregated renewable production. A **CNN regression method** improves existing CNN approaches for RES forecasting on non-extremal quantiles [19] with the integration a specific ordering of VRE plants in the input data, a min-pooling layer and training on a skill score for tails. Finally the mixture of distributions in an **MDN** is adapted to the problem of finding low quantiles of an aggregation of VRE plants whereas the work of [12] considered a single VRE plant and did not evaluate the performance of the predicted distribution on low quantiles.
4) The proposed forecasting models are **applied as inputs to a reserve capacity offer** by a VRE-VPP, which is an important use case not covered by other works proposing extreme quantile predictions. Relationships between forecasting performance and reserve offer reliability are evaluated.

The present paper is organized as follows: the forecasting methodology is presented in Section III, and applied on a case study of day-ahead aggregated production forecast in Section IV. After presenting the evaluation metrics in Section V, the results are discussed in Section VI.

III. Methodology

Maximizing the reliability of an offered reserve volume R (active power) can be formalized in (1) as the problem of finding a time-dependent offer R, such that the frequency of events where the available production y is lower than R is at most τ_R during a testing period of length T. The frequency of such events is called hereafter the Rate of Under-Fulfillment (RUF) [20], and τ_R is in the range of $[0.1\% - 1\%]$.

$$RUF(R, y) := \frac{1}{T} \sum_{i \in [1,T]} \mathbf{1}_{y_i < R_i} \leq \tau_R \tag{1}$$

Consider now that the VPP is asked to formulate its offer on a short-term reserve market. The gate closure time of the market is assumed to be at a day-ahead horizon as implemented in the European Frequency Containment Reserve (FCR) cooperation [21] and in the German market for automatic Frequency Restoration Reserve (aFRR) [22]. The market procures reserve capacity for the 24 hours of the next day, which corresponds to the interval of prediction horizons $\mathcal{H}$. Due to the inherent uncertainty of the VRE production and the absence of storage in the VPP, the reserve offer problem expressed above requires the derivation of a probabilistic forecast of the VPP production $\hat{y}$ that aims in (2) at minimizing the reliability deviation of a quantile forecast $\hat{y}^{(\tau)}$, where τ is a low quantile value, $\tau \leq \tau_R$. This forecast will be used as a basis for the reserve capacity offer. Recall that the reliability of a probabilistic forecast corresponds to its property of being in line with the conditional relative frequencies of the observations [23]. In the long run, a reliable 0.1% quantile forecast would be superior to the actual VPP production during 0.1% of the time. Forecasts issued at the different prediction runtimes $t \in \mathcal{T}$ evolve conditionally to the values taken by the matrix $\mathbf{x}$ of explanatory variables (also called features) available for the $\mathcal{P}$ plants in the VPP.

$$\begin{aligned} &\min_{\hat{y}} \left| \frac{1}{T} \sum_{i \in [1,T]} \mathbf{1}_{y_i < \hat{y}_i^{(\tau)}} - \tau \right| \\ &\hat{y}_{t+h}^{(\tau)} := F_{\hat{y}}^{-1}(\tau | \mathbf{x}_{t+h}), \quad \forall t \in \mathcal{T}, h \in \mathcal{H} \end{aligned} \tag{2}$$

Finally, the predicted quantile of the VPP production can be converted into a reserve offer R in (3) by taking the minimum of the predicted production over the validity period ΔT of the reserve offer. By doing so, the capacity is guaranteed over the entire validity period [24] assuming that the observed reliability deviation of the forecast is within its expected bounds.

$$R_{t+h} = \min_{t' \in \Delta T(t+h)} \hat{y}_{t'}^{(\tau)} \quad \forall t \in \mathcal{T}, h \in \mathcal{H} \tag{3}$$

The methodology followed to develop forecasting models for low quantiles of VPP production following three alternative approaches is presented in the next subsections and

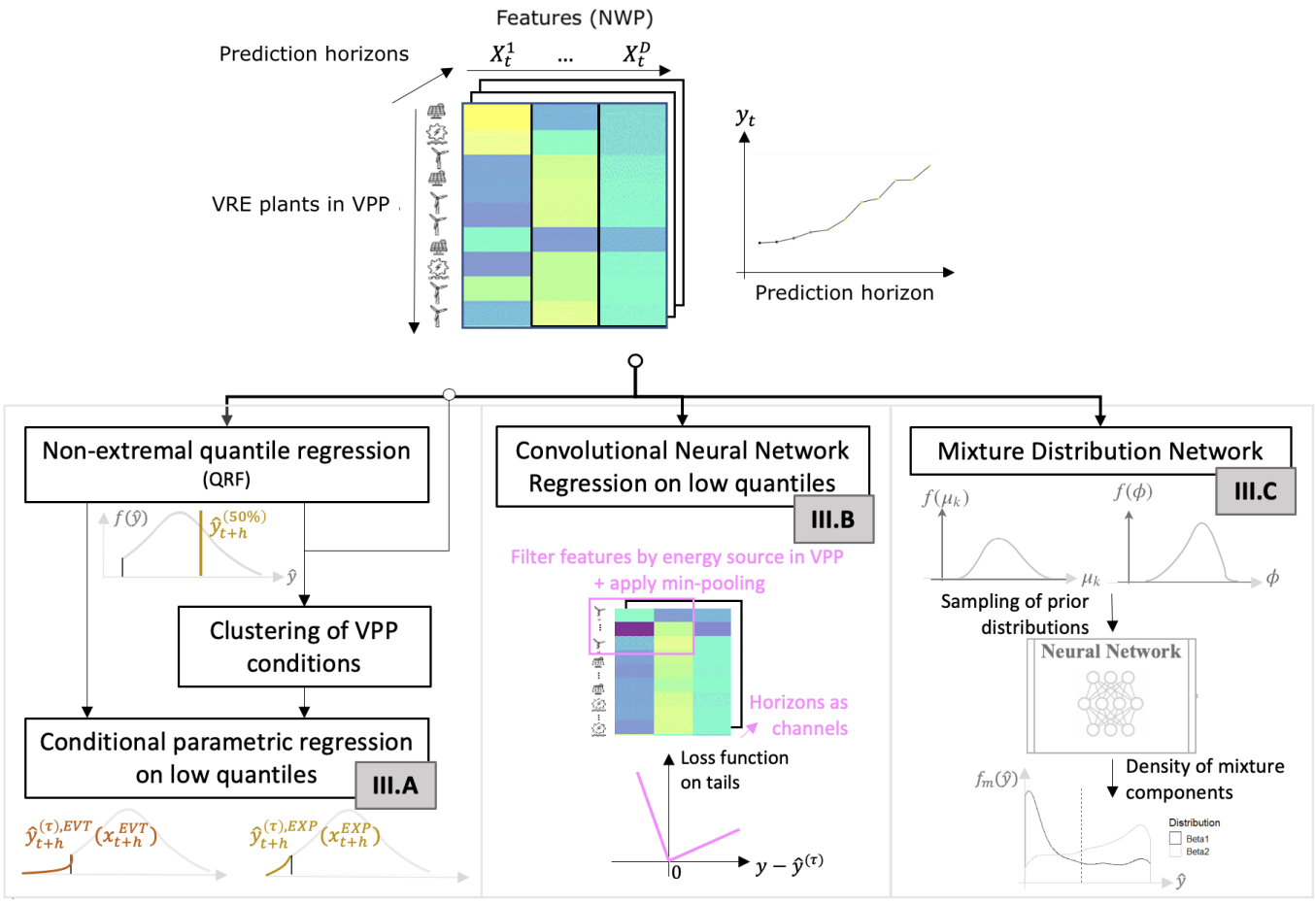

Fig. 1: Methodology workflow: Proposed approaches for forecasting low quantiles of VPP production. Each approach is developed in the designated Sections III.A to III.C

ilustrated in Fig. 1. In this work a day-ahead forecasting horizon is assumed in order to align on the reserve capacity market closure. At this horizon level, explanatory variables of VRE production are known to be principally Numerical Weather Predictions (NWP). The dataset describing the VPP corresponds to a collection of features X constituted by NWP variables for all VRE plants in the VPP and the entire range of prediction horizons, associated to the vector y of VPP production observations along the horizon. A first approach consists in parametric regression models of low quantiles conditioned by information on the VPP production, presented in Section III-A: an exponential distribution and a generalized Pareto distribution derived from EVT. The other two approaches are based on neural networks: a CNN regression presented in Section III-B and a Mixture Density Network approach in Section III-C.

A. Conditional Parametric Models

The parametric distributions proposed by [5] and [6] to predict extremes of VRE production adapt predictions by a conditioned inference of the parameters θ of a distribution $\mathcal{D}$. Forecasts on low quantiles are then retrieved by inverting in (4) the quantile function of the distribution conditioned by a pre-processed set of explanatory variables $\mathbf{x}^{\mathcal{D}}$ representative of the dynamic contributions of the various plants in the VPP. For better readability, temporal indices are omitted in what follows unless necessary for the understanding of the model structure.

$$\hat{y}^{(\tau)} = F_{\mathcal{D}}^{-1}(\tau|\theta(\mathbf{x}^{\mathcal{D}})) \tag{4}$$

1) Exponential distributions: This subsection presents the first conditioned parametric model, namely an exponential distribution. The choice of an exponential distribution to predict low quantiles is justified by the out-of-sample distribution of exceedances ϵ, formulated in (6) as the positive distance between a predicted reference low quantile $\hat{y}^{(\tau_{ref})}$ and production [25]. The reference quantile is empirically chosen as the lowest nominal quantile value for which a standard forecasting model such as QRF remains reliable (e.g. 1%). An advantage of this distribution is the simplicity of the inference: θ corresponds here to a unique parameter, the thickness ρ conditioned by the pre-processed set for the exponential distribution $\mathbf{x}^{\text{exp}}$. Finally the low quantile prediction can be retrieved in (7) by inverting the quantile function of the exponential distribution.

$$\epsilon := [\hat{y}^{(\tau_{ref})} - y]^+ \tag{5}$$

$$F_{\text{exp},\epsilon}(\tau|\mathbf{x}^{\text{exp}}) \sim \tau_{ref}.e^{-\rho(\mathbf{x}^{\text{exp}})\epsilon}, \quad \epsilon \geq 0 \tag{6}$$

$$\hat{y}^{(\tau)}_{t+h} = F^{-1}_{\text{exp},\epsilon}(\tau|\rho(\mathbf{x}^{\text{exp}}_{t+h}))), \quad \forall t \in \mathcal{T}, \forall h \in \mathcal{H} \tag{7}$$

The forecasting method is presented in more detail in Algorithm 1. It starts by deriving forecasts from a standard model such as QRF on the reference quantile and for the median. Then two choices of feature selection are proposed to condition the model based on a feature set $\mathbf{x}^{exp}$:

- following [5], equally-spaced intervals of median production forecast,
- as an original approach, clusters obtained by unsupervised learning based on median production forecasts and augmented with selected NWP features from different sites of the VPP.

The latter feature selection method is performed by k-means clustering on an empirically chosen set of variables. This feature set $\mathbf{x}^{\text{cluster}}$ is presented in (8), and includes the median VPP production forecast, augmented with the minimum, mean and maximum value of each NWP variable $d \in D^s$, where D^s is the number of distinct NWP variable per energy source s in the VPP. This set is thought to better characterize the potentially low production regimes of the VRE-VPP due to the integration of multivariate information from weather predictions on all plants of the VPP. Additionally, and of a more even population of cluster with k-means compared to equally-spaced intervals of production levels.

$$\mathbf{x}^{\text{cluster}} = \begin{pmatrix} \{\min_{p\in[1,P_s]} x^d_{p,t}, & t \in \mathcal{T}, d \in D_s, s \in S\} \\ \{\text{mean}_{p\in[1,P_s]} x^d_{p,t}, & t \in \mathcal{T}, d \in D_s, s \in S\} \\ \{\max_{p\in[1,P_s]} x^d_{p,t}, & t \in \mathcal{T}, d \in D_s, s \in S\} \\ \{\hat{y}^{(50\%)}_t, & t \in \mathcal{T}\} \end{pmatrix}$$

2) Extreme Value Theory: Distributions of extremes with higher versatility than the exponential model formulated above can be inferred with the EVT. This work follows a Peak-Over-Threshold (POT) approach, which collects observations over a chosen threshold u, taken in (14) as the k-highest value of the sorted available record. Since the present article focuses on minima of VPP production and the POT works on maxima, the POT approach is applied to $y^* = -y$.

$$u = y^*_{k,N}, \quad y^*_{1,N} \leq ... \leq y^*_{k,N} \leq y^*_{N,N} \tag{14}$$

Pickland's theorem [9] stipulates that the distribution of the maxima of the i.i.d variable of y^* above the threshold u converges towards a Generalized Pareto Distribution (GPD). The GPD is defined by its parameter vector $\theta := \{u, \sigma, \gamma\}$, where γ defines the overall shape of the distribution of extremes and σ quantifies the spread of extreme values. Similarly to the exponential model above, the present work proposes a conditional estimation of the GPD parameters as a function of selected features for the EVT $\mathbf{x}^{\text{EVT}}$. The prediction of the low quantile of VPP production is obtained directly in (15) from the quantile of the survivor function modelled by the GPD of $\bar{F}_{y^*,\text{GPD}}$, where $\bar{F} = 1 - F$.

ALGORITHM 1

Low quantile prediction by exponential distribution

Input:

- Explanatory variables $\mathbf{x}$
- Feature selection choice for the exponential model $\mathbf{x}^{\text{exp}}$
- Number of conditioning partitions C
- Reference quantile value, τ_{ref}

Output: Quantile forecast $\hat{y}^{(\tau)}$

1: VPP production forecast at reference quantile and median quantile, on validation and testing set

$$\hat{y}^{(\tau')}_{QRF} = F^{-1}_{\hat{y},QRF}(\tau'|\mathbf{x}), \quad \tau' = \{\tau_{\text{ref}}, 50\%\} \tag{8}$$

2: **switch** (Choice of feature selection for the exponential model $\mathbf{x}^{\text{exp}}$)

3: **case** $\mathbf{x}^{exp} := \hat{y}^{(50\%)}$**:**

4: Classify production levels in equally-spaced intervals $R_c, c \in C$, on validation set indexed by $\{j, j \in [1, N]\}$

$$[R_0, ..., R_c, ..., R_C] := \cup_{c\in\mathbb{N}^+\cap[0,C-1]}[\frac{c}{C} : \frac{c+1}{C}] \tag{9}$$

$$Card(\{\hat{y}^{(50\%)}_j \in I_c, j \in [1, N]\}) = N_c, \quad \forall c \in C \tag{10}$$

5: **case** $\mathbf{x}^{\text{exp}} := \mathbf{x}^{\text{cluster}}$**:**

6: Pre-process features for clustering $\mathbf{x}_{\text{cluster}}$

7: Classify features $\mathbf{x}^{exp}$ in clusters $R_c, c \in C$ via k-means

$$[R_0, ..., R_c, ..., R_C] := \text{k-means}(\mathbf{x}_{\text{cluster}}, C) \tag{11}$$

8: Compute exceedances ϵ, on validation set, cf. (6)

9: Estimate exponential thickness ρ_c by maximum likelihood for each interval or each cluster, on validation set

$$\rho_c = \frac{1}{\frac{1}{N_c}\sum_{i=1}^{N} \epsilon_i 1_{\mathbf{x}^{\text{exp}}_i \in R_c}} \quad \forall c \in C \tag{12}$$

10: Retrieve conditional thickness $\rho(\mathbf{x}^{\text{exp}}_{t+h})$ on testing set $\mathcal{T}_{te}$

$$\rho(\mathbf{x}^{\text{exp}}_{t+h}) = \rho_c : \mathbf{x}^{\text{exp}}_{t+h} \in R_c, \quad \forall t \in \mathcal{T}_{te}, \forall h \in \mathcal{H} \tag{13}$$

11: Conditional prediction on testing set $\mathcal{T}_{te}$, cf. (7)

$$\hat{y}^{(\tau)} = \bar{F}^{-1}_{y^*,\text{GPD}}(\tau|\theta(\mathbf{x}^{\text{EVT}})) \tag{15}$$

The EVT model of [6] is fitted on samples of non-extremal predictions that have the highest similarity with past observations. However in the present problem, it is likely that extremes of the total VPP production depend not only on the expected total production level but also at a significant extent on extreme realizations of the uncertain multivariate weather conditions experienced by each plant in the VPP. Therefore, the present work proposes to condition the EVT parameters on a clusterized set based on the set $\mathbf{x}^{\text{cluster}}$ presented above for the exponential model. The quantile function expressed in (15) is then derived in (16) for each cluster $c \in C$. The shape of the distribution of extremes γ_c is allowed to vary as a function of the different production regimes described by the clusters, but

should not be impacted by short-term variations of weather variables. Conversely, the scale factor σ, which quantifies the spread of extreme values, is assumed to be dependent also on the variations of explanatory values $\sigma_c(\mathbf{x}^{\text{EVT}})$.

$$\bar{F}^{-1}_{y^*,c}(\tau|\mathbf{x}^{\text{EVT}}) = u_c + \frac{\sigma_c(\mathbf{x}^{\text{EVT}})}{\gamma_c}[(\frac{\tau}{\bar{F}_c(u_c)})^{-\gamma_c} - 1], \quad \forall c \in C \tag{16}$$

Not all of the D^{EVT} features of $\mathbf{x}^{\text{EVT}}$ may be beneficial to the quantification of uncertainty. Adding too many parameters may add noise and ultimately prevent the model from detecting extreme conditions. We propose here to select features using a feature selection algorithm presented in the Algorithm 4 reported in Appendix A. Features are integrated into the conditional estimator of scale as long as the Akaike Information Criterion (AIC) improves, and the overall scale value is positive. After feature selection, the parameters are inferred by Maximum Likelihood Estimation. The entire forecasting model of low quantiles by EVT is reported in Algorithm 2.

ALGORITHM 2
Low quantile prediction by EVT

Input:
- Explanatory variables $\mathbf{x}$
- Number of conditioning partitions C
- Threshold value k

Output: Quantile forecast $\hat{y}^{(\tau)}$

1: Compute complementary VPP production $y^* = -y$
2: Classify features $\mathbf{x}^{\text{EVT}}$ in clusters $R_c, c \in C$ via k-means
3: Forward selection of features for the scale parameter $\sigma_c(\mathbf{x}^{\text{EVT}}), \forall c \in C$ (cf. Algorithm 4)
4: Identify the threshold $u_c, \forall c \in C$ on validation and testing sets
5: Estimate scale $\sigma_c(\mathbf{x}^{\text{EVT}})$ and shape $\gamma_c, \forall c \in C$ by likelihood maximization of the GPD on validation set
6: Retrieve conditional threshold and parameters on testing set $\mathcal{T}_{te}$
7: Conditional prediction on testing set $\mathcal{T}_{te}$, cf. (16)

B. Convolutional Neural Network Regression Model

A CNN quantile regression is proposed as a direct approach to forecast very low quantiles of aggregated VRE production. The complete algorithm of the prediction of low quantiles by this model is presented in Algorithm 3. The regression model consists of three adaptations of standard CNN regression to the present problem, detailed in the subsections below.

1) Configuration of the feature space: The configuration of the proposed CNN model is illustrated in the lower central image of Figure 1. The CNN network takes as input a 3D matrix of features gathering all the information on the various plants in the aggregation, that can be organized following the requested horizons of the prediction model.

In this model the different channels of the CNN correspond to the different prediction horizons $h \in \mathcal{H}$, plants in the VPP are organized in rows and the various explanatory variables or features for each plant are in the columns of each 'image' associated to a specific horizon h. Batches of such input matrices are obtained by slicing the available dataset in periods of length $\mathcal{H}$. Learning is done by applying a sequence of multiple 2D convolutional filters, followed by intermediate layers presented in Algorithm 3. The rationale of applying 2D convolutional filters on a feature space organized by prediction horizon is that the network will learn dependencies across plants and energy source for each prediction horizon independently. This will result in a sequence of abstract representations of the multivariate weather conditions experienced with the VPP over the entire horizon interval. At the final layer of the CNN, a direct relationship between this sequence of weather representations and the sequence of observed VPP production can be derived, therefore conserving the temporal correlation in the spaces of the features and of the response.

The first convolutional layer applies on $\mathbf{x}$ a filter $f^{(1)}$ with a 2D convolutional kernel kn of size (k_u, k_v) on rows and columns of the input matrix. A *stride* parameter $st = (s_u, s_v)$ adds distance between kernels in order to reduce the dimensionality of the convolutional layer. The filter learns a weight matrix w on all pixel indices $(i^{'}, j^{'})$ of the input matrix, and a constant bias term $b^{(1)}$. The pixels of the resulting feature map $\Phi^{(1)}_{ij}$ represent the filtered relationships between contiguous plants and features in the input matrix. A 3D convolutional layer would also filter the information across channels, which means here across prediction horizons. This might be beneficial in a different problem such as forecasting at intraday horizons where the temporal correlation between features is high, or the generation of trajectories. This goes beyond the scope of the present study.

$$\Phi^{(1)} = \mathbf{x}_h \circledast f^{(1)}(\text{kn}) + b^{(1)} \tag{17}$$

$$\Phi^{(1)}_{ij} = \sum_{u=1}^{k_i}\sum_{v=1}^{k_j} \mathbf{x}_{h,i^{'}j^{'}} w_{uv} + b^{(1)}, \tag{18}$$
$$i^{'} = u.s_u + k_i - 1,$$
$$j^{'} = v.s_v + k_j - 1$$

The ordering of different VRE plants and associated features in the VPP input matrix is important for the learning process of the CNN model. If VRE plants follow a random order in the input matrix (cf. right of Fig. 2), then learning first focuses on the dependency between features of individual plants (e.g. one Wind plant and one PV plant) before learning the smoothing effect of aggregation over all energy sources in the VPP. In contrast, VRE plants can be ranked contiguously by energy source as on the left of Fig. 2. Consider for example that the vectors $x^{s1,\cdot}$ in Fig. 3 correspond to the $\mathcal{P}_{s1}$ Wind farms in the VPP. Then, the $\mathcal{F}_1$ different feature maps of the first convolutional layer of the CNN $\Phi^{(1),\cdot}$ learn cross-plant information on the impact of weather diversity on the wind production process. The $\mathcal{F}_2$ feature maps of the second convolutional layer $\Phi^{(2),\cdot}$ combine the information from the first layer, and will learn the dependencies between Wind and the other energy sources (e.g. PV, Hydro). Given that

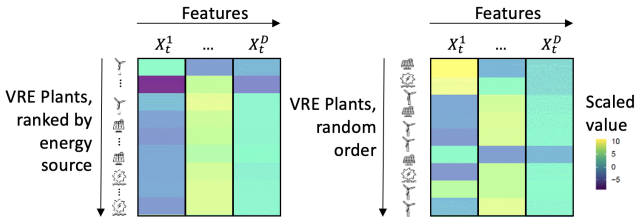

Fig. 2: Ordering of VRE plants in the VPP data input matrix

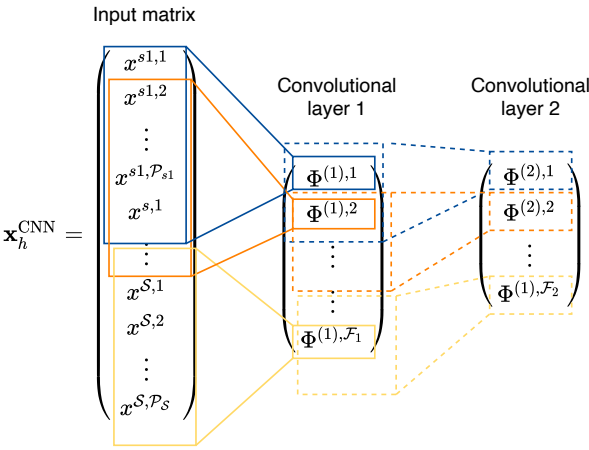

Fig. 3: Sequence of convolutional filters on the input matrix where VRE plants are ranked by energy sources $s1, ..., s, ..., \mathcal{S}$

the objective of the network is to detect minimal levels of production, these will be identified more easily among plants of the same energy source, which are likely to have conditions more correlated than with plants of a different energy source. Therefore, the former approach is implemented, where plants are ranked contiguously by energy source.

Note that in a VPP configured for AS provision, VRE plants are recruited in order to obtain a high diversity of climate conditions in the portfolio. Consequently, the spatio-temporal correlation between plants of the same energy source is low, and the placement in the input space based on geographical location is not of primary importance. If the problem is scaled to a large number of VRE plants where multiple plants are located in the same region, the configuration of the input space should integrate a placement rule that would preserve as much as possible spatial relations in all regions.

2) Min-pooling layer: In classification problems, a sub-sampling step called *pooling* extracts synthetic results from feature maps, by taking averages or maxima over a defined window. This is efficient for detecting whether an edge is present in a part of an image, and improves the invariance of learning. We propose instead here to integrate a *min-pooling* layer illustrated in Fig. 4. This layer extracts in (19) the minimum value obtained in a local area of the network by filtering the previous layer with a kernel $\text{kn}_{\text{minpool}}$. The rationale is that this layer detects the minimum value of features in a local area of the previous layer. Applied to the entire network, min-pooling helps detect low values of the features, e.g. in the

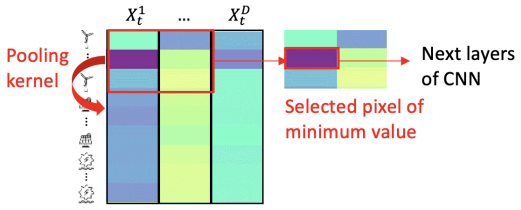

Fig. 4: Principle of min-pooling layer in the proposed CNN model

first layer of the network the minimum wind speed forecast across all wind farms in the VPP. The effect of such minimum values is then propagated through the CNN until it is evaluated on the observed aggregated production. Linking minimum values of filtered input weather variables with low quantiles of VRE production is justified by the fact that the most influential variables on VRE production at day-ahead horizons exhibit positive correlation with the production of the different energy sources (e.g. wind speed forecast and wind power, solar irradiance forecast and PV production, precipitation forecast and run-of-river Hydro production). Without the min-pooling layer, the gradient of the loss function with respect to the convolutional filter weights depends on all input variables, and therefore low values of expected weather conditions will be smoothed out during learning. Consequently, min-pooling facilitates the learning patterns of extremely low levels of aggregated VRE production.

$$\Phi^{(l)}_{minpool} = \min(\{\Phi^{(l-1)}_{i,j}, (i,j) \in \text{kn}_{\text{minpool}}\}) \qquad (19)$$

Finally, the last layer is connected to multiple outputs through fully-connected layers. The outputs are nodes associated with a specific loss for each quantile to be predicted in the interval of interest, i.e. [0.1%-0.9%].

3) Specific loss function for low quantiles: We propose to test two different loss functions for the prediction of low quantiles. A first loss function is the well-known pinball function used for quantile regression [26]. The Skill Score $SkSc$ is obtained in (20) by averaging the pinball losses over the entire interval of low quantiles $[\tau_{\min}, \tau_{\max}]$ of size n_τ. The score is negatively oriented here to minimize the CNN loss function during gradient descent.

$$SkSc(y, \hat{y}) = -\frac{1}{n_\tau} \sum_{\tau=\tau_{\min}}^{\tau_{\max}} (1_{y-\hat{y}^\tau} - \tau)(y - \hat{y}^\tau) \qquad (20)$$

The second loss function corresponds to a weighted version of the Quantile Score called hereafter wQS that has been proposed for the evaluation of the forecast of extremes, see for instance in [27] and [28]. Following [27], in the case of low quantiles (left tail of the distribution), the wQS weights in (21) the pinball losses of the different quantiles by a factor $(1-\tau)^2$. This factor enables a better balance between the

losses associated to lowest and highest quantiles in the quantile interval.

$$wQS(y,\hat{y}) = \frac{1}{n_\tau}\sum_{\tau=\tau_{\min}}^{\tau_{\max}}(1-\tau)^2(\tau - 1_{y-\hat{y}^\tau})(y-\hat{y}^\tau) \quad (21)$$

Both loss functions in (20) and (21) are convex but non-differentiable when the error $y-\hat{y}$ is zero due to the presence of the indicator function. In order to apply gradient descent, the pinball loss is reformulated as element-wise maximum of the penalized error $max_{1\leq i\leq n}(\tau-1)(y-\hat{y}), \tau(y-\hat{y}))$ (refer e.g. to the *pinball loss* function implemented in Tensorflow [29]). Then a smooth approximation can be applied to error levels close to zero via a Huber loss [30] in order to obtain a fully differentiable loss function. However, even without this approximation it is possible to perform gradient descent on this element-wise maximum function thanks to the automatic differentiation algorithms implemented in state-of-the-art libraries such as Tensorflow or PyTorch. At zero error, these librairies fix arbitrarily the value of the function error gradient to zero [31]. The impact of this choice has been shown to be neglectible in a simple setup with SGD presented in [31], especially if batch normalization is implemented as it is the case in Algorithm 3.

ALGORITHM 3
Prediction of multiple low quantiles by CNN

Input:

- Explanatory variables $\mathbf{x}$
- Stride and window parameters of kernels $\text{kn}, \text{kn}_{\text{minpool}}$
- Number of convolutional layers L

Output: Quantile forecast $\hat{y}^{(\tau)}$

1: Features $\mathbf{x}$ are arranged in volumes with height corresponding to all plants in the VPP, width to features and channels to prediction horizons.
2: VRE plants are ranked contiguously by energy source.
3: **for** l in L convolutional layers: **do**
4: Filter $\Phi^{(l)}$ is convolved with the previous layer $\Phi^{(l-1)}$, corrected by a bias term $b^{(l)}$: $\Phi^{(l)} = \Phi^{(l-1)} \circledast f^{(l)} + b^{(l)}$
5: Min-pooling, cf (19)
6: Exponential Linear Unit activation
7: Batch normalization on the feature axis to ensure efficient backpropagation
8: **end for**
9: Flatten layer: The F_L feature maps in the last convolutional layer $\Phi^{(L)}$ are flattened: $y_{flat} = [\Phi^{(L),j}, \forall j \in F_L)]$
10: Output layer: the predicted production at all horizons $\hat{y}^{(\tau)}_{h,i} \in \mathcal{H}$ is associated to an observed batch of production values $y_{h,i}$, where i is the index of a batch in the training set
11: Training on the Skill Score cf. (20) through Stochastic Gradient Descent (SGD).
12: Conditional prediction on batches of testing set $\mathcal{T}_{te}$

C. Mixture Density Network

The Mixture Density Network (MDN) combines the flexibility of mixtures to capture extremes with the capacity of a neural network $g(.)$ to approximate non-linear relationships between features and response variables. Three classical architectures of neural networks are tested as regression function in this paper: the Fully-Connected Neural Network (FCNN) called also Multi Layer Perceptron, a CNN without min-pooling layer, and a Long Short Term Memory (LSTM) network. In order to predict the low quantiles of the expected distribution of the VPP production, it is proposed here to formulate a specific mixture distribution that is (1) adapted to a VRE-VPP and (2) has good forecasting performance on its lowest quantiles. Consider that the random variable associated to the VPP production y follows in (22) a mixture of M components, where each component m is associated with a probability π_m, such that $\sum_{m=1}^{M}\pi_m(x) = 1$.

$$\begin{aligned} y &\sim \sum_{m=1}^{M}\pi_m(x) f_m(\theta_m(x)) \\ g(x) &= (\pi_m(x), \theta_m(x) : m \in M) \end{aligned} \quad (22)$$

The density of the response variable is obtained in (23) by integrating the Bayes chain rule over the range of parameter values.

$$p(y_t|x_t) = \int_\theta \sum_{m=1}^{M}\pi_m(x_t) f_m(y_t|x_t,\theta_m) p(\theta|y_t,x_t) p(\theta) d\theta \quad (23)$$

A standard approach consists in proposing a mixture of Gaussian components. The parameters θ_m of the components are generally estimated by Expectation-Minimization without assuming problem-specific values for the parameters, i.e. the mean and variance of each component. In contrast, we introduce here *a priori* knowledge on the mixture to model that the distribution should reflect the aggregated behaviour of multiple plants in the VPP. In particular, the means and variances of the different components are inferred via a bayesian approach detailed in Appendix B where the hyper-parameters correspond to the means and variances of the production of each plant in the VPP. This formulation of the Gaussian mixture is adapted to the context of aggregated VRE production. The limit of this formulation is that it is not guaranteed to perform well on low quantiles: its performance will be evaluated in Section VI.

A second alternative mixture distribution is a special Beta mixture proposed by [13] that is well suited to the modelling of extremes. The parameters of the mixture components presented in Appendix B are inferred to model the shape of the distribution, without direct link to the plants composing the VPP. The adaptation of this mixture to the present problem consists in integrating it into a MDN.

Finally, the mixture parameters of both Gaussian MDN and Beta MDN are inferred via the Stochastic Variational Inference (SVI) method developed by [32], chosen for its ability to work on mixture distributions of the exponential family (comprising Gaussian and Beta) and its gradient-descent approach compatible with MDNs.

Energy source s	PV	Wind	Hydro
Number of plants $\mathcal{P}_s$	9	3	3
Installed capacity [MW]	9	32	12
Solar radiation GHI [W/m2]	x		x
Total cloud cover [-]	x		
Air temperature 2 m [K]			x
Hourly rainfall [mm/h]	x		
Daily cumulated rainfall [mm/day]			x
Zonal, meridional wind speed at 100m [m/s]		x	

TABLE I: VPP plants description and NWP by energy source

Parameter	Gaussian Mixture	Beta Mixture
π	$\alpha_\pi^0 = (\frac{1}{\mathcal{P}}, ..., \frac{1}{\mathcal{P}})$	$\alpha_\pi^0 = (0.5, 0.5)$
μ	$\mu_\mu^0 = \{\mu_p, p \in [1, \mathcal{P}]\}$ $v_\mu^0 = \{3, p \in [1, \mathcal{P}]\}$	$\alpha_\mu^0 = 1$ $\beta_\mu^0 = 1$
ϕ	$c^0 = \{v_p^{-1}, p \in [1, \mathcal{P}]\}$ $r^0 = \{3, p \in [1, \mathcal{P}]\}$	$c^0 = 1$ $r^0 = 0.1$
ω	-	$\alpha_\omega^0 = 1$ $\beta_\omega^0 = 1$

TABLE II: Hyperparameters of distribution mixtures

Parameter	Model	Value Range
Number of clusters	Exponential	[2,4,...,32]
Number of clusters	EVT	[2,3,4,5]
Reference quantile	Exponential	[1%, 2%, ..., 5%]
Threshold k	EVT	[90%, 91%, ..., 99%]

TABLE III: Parameter ranges for parametric models

IV. Case Study

A. Data description

The merits of the models developed for extremes are compared with those of QRF on a case study concerning a VPP with 3 different energy sources: PV, Wind and run-of-river Hydro. The number of plants P in the VPP equals 15 plants located in several climates in France (moderate warm continental, moderate cold continental, atlantic, see Fig. 5). The production data comprises 9 months of measurements at 30 minutes resolution between July 2015 and February 2016, constituting a total dataset of 9600 samples.The training, validation and testing of models is done by a 7-fold cross-validation on weekdays. Additional details on the installed capacities and retrieved NWP per energy source are given in Table I.

The NWPs are retrieved from the ECMWF HRES service, which is a state-of-the art solution for short-term weather forecast. Note that forecasts at higher resolution such as AROME from MeteoFrance are expected to have lower weather forecasting errors, especially in specific climate zones where the higher spatio-temporal resolution of the NWP model improves forecasting performance. The impact of such improved NWP on the proposed methods in this study is an interesting perspective for future work. NWPs issued at 00h00 on the previous day are used in order to participate in a day-ahead reserve offer mechanism. The resulting forecasting horizon is thus comprised between 24h and 48h.

B. Hyperparameters

The hyperparameters values of the Gaussian and Beta Mixture are reported in Table II. The mean and precision parameters of the Gaussian Mixture are fitted on the past observed mean μ_p and variance v_p of each plant p in the aggregation. The hyperparameters of the Beta Mixture reflect a low level of knowledge a priori, e.g. the concentration rates of the two Beta components α_π^0 are assumed to be equal, and the parameters of the precision ϕ describe a low precision a priori. These hyperparameter values let therefore the Bayesian inference optimize the final posterior values of the Beta mixture parameters in a large range of possible values. The sensitivity of the inference to different choices of hyperparameters is beyond the scope of this work.

The characteristics of the CNN regression models (number of layers, dimensions of filters) are established via grid search. Note however that a more systematic tuning of the depth and filter characteristics of the CNN is out of the scope of the present paper. As the CNN model is run on CPU only and with limited memory capacity, a Principal Component Analysis (PCA) is applied to the complete feature set in order to reduce the computational burden. A smaller subset of 3 features per plant in the VPP is retrieved, explaining 90% of the original variance. Note that the PCA step can be skipped if the model is run on GPU or memory capacities are not heavily constrained, without any impact on the rest of the CNN implementation.

EVT and exponential models have been fitted on a grid of parameter values reported in Table III that include the numbers of clusters, the reference quantile nominal value for the exponential model and the threshold k for the EVT. The sensitivity of the performance of all variants of these models with respect to the number of clusters has been analyzed (not shown here for brevity) and only the best performing models are presented in the Result section. The number of clusters for the EVT is smaller than for the exponential model because the number of available points above the threshold k is small. The threshold is evaluated as a function of the maximization of the likelihood and of the stationarity of the series of peaks (optimal value is $k = 0.97$).

C. Application to Reserve provision

Finally, forecasts obtained by each method are converted into offers of reserve capacity by the VRE-VPP. For simplicity, only downward reserve activation is considered and two validity periods for the offer are tested, 1h, 2h 4h. The latter value corresponds to the current rules in the aFRR German market at the time of writing, while 1 h is a potential future standard value that can be expected from the ongoing effort to harmonize European short-term markets for balancing capacity.

V. Evaluation Metrics and implementation

As this paper focuses on forecasted extremes, the evaluation metrics are specifically chosen for the assessment of probabilistic performance on the lowest quantiles of the distribution and include:

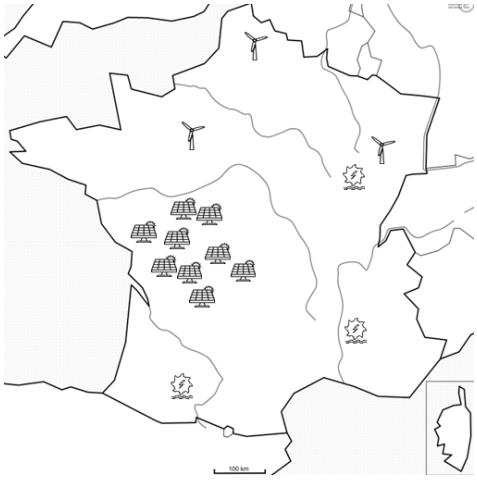

Fig. 5: Localization of Wind, PV, and Hydro plants in the VPP.

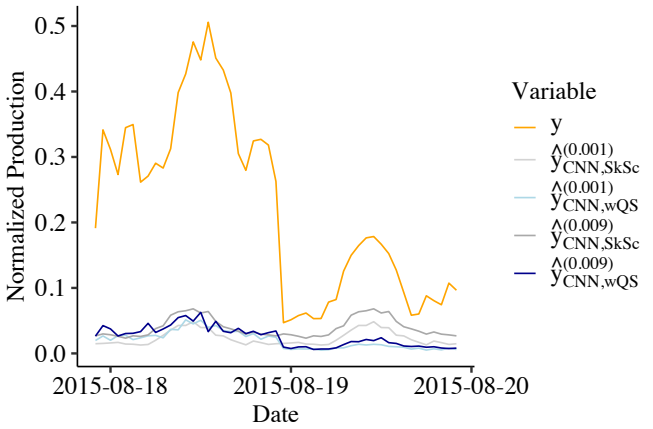

Fig. 6: Illustration of the interval of predicted quantiles from the CNN trained on SkSc and wQS. The orange line is the actual VPP production. 0.001 and 0.009 stands for 0.1% quantile and 0.9% quantile respectively.

- the weighted version of the Quantile Score (wQS) adapted to tails, converted in logarithm form for better readibility;
- Reliability on low quantiles, including uncertainty bars accounting for sampling effect [33];
- Sharpness of prediction on low quantiles.

The sharpness Sh on the quantile interval $[\tau_{min}, \tau_{max}]$ writes as in (24) for an evaluation period T. The quantile interval [0.1%-0.9%] predicted by the CNN trained either on the SkSc (grey lines) or on the wQS (blue lines) are illustrated on two days in Fig. 6. The interval of the CNN with wQS is much narrower than the interval of the CNN with SkSc. The width of this interval varies during the day, for instance in this summer period it is larger around noon where the PV contribution is maximum, as visible in the second day with sunshine and no wind.

$$Sh(\hat{y}) = \frac{1}{T}\sum_{t=1}^{T}(\hat{y}_t^{(\tau_{\max})} - \hat{y}_t^{(\tau_{\min})}) \tag{24}$$

The performance of reserve offers is evaluated via metrics that quantify the technical reliability of the reserve capacity and characterize the offered volumes:

- Frequency of reserve under-fulfillment quantified by the RUF (1)
- Maximum Reserve deficit $\max_{i\in[1,T]}(R_i - y_i)1_{y_i - R_i}$

Finally, the code implementation is done in R for the parametric models and in Python for the CNN and MDN. The following packages have been essential in the implementation: `quantregForest` for the QRF [34], `extRemes` for the EVT [35], `keras` and `tensorflow` for the CNN and MDN. Code is run using only CPU (no GPU) on an Intel Xeon 2.40 GHz, 128 Go 32 cores. Training the machine learning models on the entire set of training batches takes the following approximate computational time: 40 minutes for the MDN, 35 minutes for the CNN, 10 minutes for the QRF, and 2 minutes for the k-means clustering.

VI. Results

The performance of models for forecasting very low quantiles of VRE-VPP production is summarized in Table IV and compared with QRF. Models are compared on three results: the average reliability deviation on the quantile range [0.1% - 0.9 %], the weighted QS (in log form for a more compact presentation of the low scores obtained), and the sharpness between the 0.9% and 0.1 % quantiles.

The most reliable model is the exponential distribution based on k-means clustering, which halves the reliability deviation of the QRF. The reliability of the exponential distributions improves with a lower reference quantile. The number of k-means clusters has also a significant impact on reliability: low numbers of clusters impede discrimination and high numbers of clusters limit the capacity to generalize observations. The reliability diagram in Fig. 7 shows that the reliability of exponential distributions is best in this case for medium-sized clusters (16 clusters). Note that sharpness does not improve compared to QRF.

The EVT model depends on the combined influence of the threshold value k, which determines the range of production values considered as peaks, and of the number of clusters characterizing the production regime. The most reliable EVT configuration combines a high threshold of 97% with a small number of clusters (2). The global forecasting score has improved compared to QRF (-7.04 vs -7.01 in log wQS), but not the reliability (-0.12% vs -0.10%). It seems that the gain in sharpness due to clusterization leads to biased predictions of extremes.

The MDNs with Beta Mixture rank in the top 3 methods in terms of forecasting score, with a slight improvement in sharpness compared to QRF (2.0% - 2.2% vs 2.5%). The reliability deviation shows that Beta MDN forecasts are too conservative, which is on the safe side for the present application (reserve underfulfillments are minimized) but leads to limited value (reserve volumes based on these forecasts will be small). The MDN with Gaussian Mixture, albeit applied to

logit-transformed production series, constantly predicts values close to zero for all quantiles, resulting in reliability deviation around 2% and sharpness equal to zero.

Finally, Table IV indicates that a CNN quantile regression model configured for a VRE-VPP and trained for very low quantiles as explained in Section III-B has the best forecasting score (-7.21) and the best sharpness (1.6%) of all models. Beyond the CNN architecture, the two specific characteristics namely *min-pooling* and training on *Skill-Score* are key to improving reliability. Fig. 7 shows that:

- Training the CNN on the SkSc instead of the wQS significantly improves the reliability. In the quantile regression with wQS observed forecast levels are too low, suggesting that learning on this score is not discriminative enough for very low quantiles.
- For quantiles below 0.4%, the CNN integrating the min-pooling layer (red curve) shows adequate reliability whereas the CNN without min-pooling (violet curve) generate uncalibrated forecasts with deviations outside the uncertainty bars. We observe the opposite on quantiles above 0.6% (i.e. min-pooling degrades reliability). As it is at the lowest quantiles that the QRF model is not reliable, we can conclude that min-pooling helps improve forecasts of extremes of VPP production compared to the reference QRF model.

The methods with best forecasting performance are used to simulate the offering of downward reserve capacity by the VRE-VPP. The reserve capacity is derived from forecasts at quantiles between 0.1% and 0.9% and length of reserve validity period equal to 1h, 2h and 4h. Offers based on forecasts from the exponential model (with 16 clusters from k-means and 3% reference quantile) and from the CNN (with min-pooling and Skill-Score loss function) are compared to offers based on QRF forecasts. Table V summarizes the technical performance of the reserve offer based on a 0.1% quantile forecast and different lengths of the reserve validity period. The exponential model produces the most reliable forecasts and this translates into the most reliable reserve offer with lower RUF than the QRF on the different validity periods considered. In contrast, as forecasts from the exponential are lower on average than the QRF forecasts, the volume of reserve capacity is smaller for the exponential-based offer compared to the QRF-based offer. Interestingly, the maximum observed reserve deficit is smaller for the exponential-based offer when the period length equals to 1h, i.e. close to the temporal resolution of the VPP quantile forecast. At a higher period length of 4h, the RUF reduces for all forecasting models, and the QRF which is less sharp and more variable in time than the CNN as a smaller RUF than the CNN. This result suggests that retaining the minimum forecast value over a validity period longer than the forecast temporal resolution may increase the technical reliability of a reserve offer when the low quantile forecast shows a significant temporal variability. CNN-based reserve offers limit significantly the risk of large values of reserve deficits. This is because the CNN forecasts are lower on average, which also causes lower average reserve volumes and standard deviation.

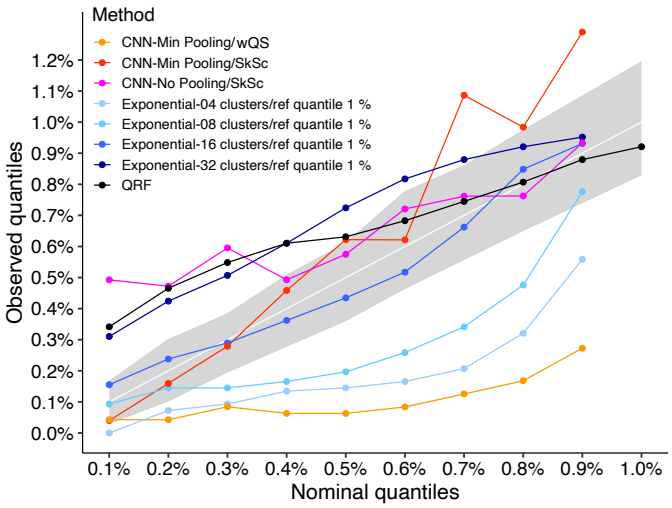

Fig. 7: Reliability diagram of QRF, exponential distributions with k-means clusters and 1% reference quantile, CNN regression with different configurations. The shaded area represents the 5%-95% cosistency interval of acceptable reliability deviations due to the sampling effect [33]

Lastly, Fig. 8 presents the average reserve capacity and RUF obtained as a function of the different validity period lengths and forecast quantile values between 0.1% and 0.9%. As expected, a shorter length of the validity period increases the average reserve volume offered at a constant quantile value of the forecast. Interestingly, offers based on QRF forecast at quantiles 0.1% - 0.3 % obtain results close to those of offers based on exponential forecast at higher quantiles 0.4% - 0.6 %. The reasons for this behaviour should be further studied, e.g. the impact of the parametrization of the exponential model based on the QRF output, or an approximate exponential learning process of extremes within the QRF albeit the estimated quantiles are unreliable. The CNN achieves the best RUF at quantiles below 0.3%, but this result is achieved due to small volumes of reserve offer on average. The non-dominated solutions in terms of maximizing the average reserve capacity and minimizing RUF are obtained from the exponential model for quantiles below 0.5 %, where the QRF starts to be unreliable. This result is valid for all validity period lengths tested, which indicates robustness of the reserve offer methodology based on reliable forecasts of low quantiles.

VII. Conclusion

This paper proposes specific models for forecasting extremes of a VRE-based VPP production with improved reliability compared to decision-tree based approaches on very low quantiles (i.e. below 1%). The first type of models consists of parametric models, namely an exponential distribution and an EVT model, conditioned by information on VPP production and weather forecasts. The case study shows that conditioning by k-means clustering on production forecasts and weather conditions improves reliability of the exponential model based on an upper bound reference quantile obtained from QRF, whereas the EVT has a higher sharpness but worse reliability.

Model	Best configuration	Average reliability deviation	Log wQS	Sharpness [% Pn]
QRF	500 trees	-0.10%	-7.01	2.5%
Conditional exponential, forecast clusters	1% reference quantile, 10 clusters	-0.15%	-7.07	3.0%
Conditional exponential, k-means clusters	3% reference quantile, 16 clusters	**-0.05**%	-7.03	2.5%
EVT	2 clusters, k = 0.97	-0.12%	-7.04	1.8%
Quantile regression CNN	minpooling, SkSc f=16-32-64x5,kn=(2,2),st=(1,1)	-0.10%	**-7.21**	1.6%
Quantile regression CNN	minpooling, wQS f=16-32-64x5,kn=(2,2),st=(1,1)	-0.11%	-7.08	**0.5**%
Beta mixture FCNN	7 layers, 60 nodes	+0.16%	-7.19	2.2%
Beta mixture LSTM	4 layers, 1 cell per horizon	+0.20%	-7.17	2.0%

TABLE IV: Summary of forecasting scores for the best of configurations of all models proposed

Method	Product length	RUF [%]	Max Reserve Deficit [% Pn]	Reserve average [% Pn]	Reserve s.d. [% Pn]
QRF	1h	0.064	11.0	**8.2**	5.8
exponential	1h	**0.047**	8.7	7.0	5.6
CNN	1h	0.058	**1.1**	3.5	1.6
QRF	4h	0.030	7.5	**6.7**	4.8
exponential	4h	**0.026**	8.1	5.5	4.5
CNN	4h	0.035	**0.8**	2.9	1.3

TABLE V: Summary of results of downward reserve capacity offer with selected best forecasting methods compared to QRF benchmark, for $\tau = 0.1\%$. [% Pn] means results are given w.r.t to the normalized production of the VPP

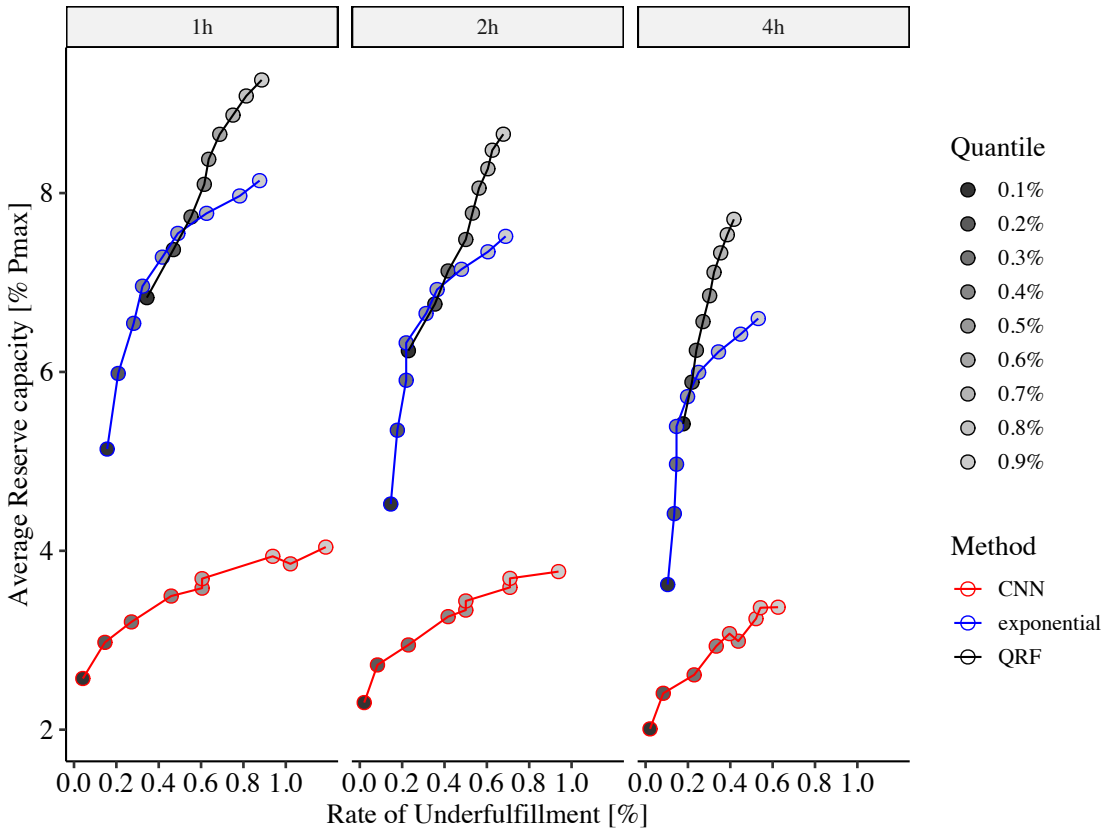

Fig. 8: Average reserve capacity and RUF obtained by the proposed forecasting models for reserve validity period lengths of 1h, 2h, 4h. Each point corresponds to a quantile value in the interval [0.1% - 0.9%], going from darkest grey for 0.1 % to lighter gray for 0.9 %.

The second type of model necessitates neural networks. Quantile regression with CNN, configured for extremes with a specific min-pooling layer which captures minimal values of features, strikes a good balance between increased reliability on the lowest quantiles and improvement of global score with respect to QRF. Min-pooling may lead to discard information conveyed by features that are negatively correlated to the aggregated VRE production, therefore other filtering or pooling techniques should be studied to obtain better performance over the entire interval of low quantiles. Mixture density networks are effective when based on Beta distributions, which are coherent with the bounded process of aggregated production. A Beta Mixture trained by a neural network, either fully-connected or LSTM, reaches one of the best performances in terms of global score but is too conservative to achieve acceptable reliability. The direct estimation of a Beta Mixture conditioned on the features of the VPP, without the complex integration of a neural network would be an interesting topic of further research. The computational complexity of the machine-learning models proposed here grows linearly with the number of features associated. So for a large VPP containing hundreds of plants, computationally efficient approaches would be needed to ensure that the proposed methods remain scalable. These include Extremely Randomized Trees instead of QRF and solving the CNN with GPU. For even larger dimensions of the VPP, transfer learning or specific statistical methods adapted to high dimension should be studied.

In conclusion, the simple model of conditional exponential distributions may suffice for the implementation of highly reliable forecasts of VRE-VPP production in practical applications such as reserve offer or unit commitment under extreme VRE scenarios. But the CNN regression constitutes an alternative because it provides sharp predictions and an optimization of its architecture could lead to further improvements in applications implicating VPPs. Both forecasting models can be used to derive reserve offers for VRE-based VPPs that minimize the probability of reserve under-fulfillments. In order to better prepare reserve dispatch within the VPP or interaction with a storage system, future work is needed to model temporal dependencies and predict very low quantiles at intraday horizons when VPP dispatch decisions need to be taken.

ACKNOWLEDGMENT

The authors wish to thank ECMWF for the provision of the Numerical Weather Predictions.

APPENDIX A
COMPLEMENTARY MATERIAL ON THE EVT MODEL

This appendix contains Algorithm 4, which selects features relevant for the conditioning of the scale parameter of the EVT model.

Algorithm 4 Feature selection algorithm for conditional scale of EVT model

Initialize
  Compute maximum likelihood of unconditional model $\mathcal{L}^{(0)}(\gamma, \sigma^{(0)})$
  Evaluate $AIC^{(0)} = 2D^{\text{EVT}} - 2ln\mathcal{L}^{(0)}$
while $j \geq D^{\text{EVT}}$ **do**
  Add feature x_j to the scale $\sigma^{(j)}(x) = \sigma^{(j-1)} + \sigma_j.x_j$
  Compute maximum likelihood $\mathcal{L}^{(j)}(\gamma, \sigma^{(j)}(x))$
  if $AIC^{(j)} < AIC^{(j-1)}$ and $\sigma^{(j)}(x_i) > 0, \quad \forall x_i$ **then**
    $\sigma^{(j)} = (\sigma^{(j-1)}, \sigma_j)$
  else
    $\sigma^{(j)} = \sigma^{(j-1)}$
  end if
end while

APPENDIX B
COMPLEMENTARY MATERIAL ON THE MDN MODEL

1) Inference of Gaussian Mixture: In a Bayesian inference of a mixture problem, the mixture proportion is associated in (25) with a prior Dirichlet distribution, with the hyperparameter rate vector α_m^0. The prior distributions for the means of mixture components follow as in (26) a conjugate Gaussian distribution with hyper-parameters $\mu_{\mu m}^0, \sigma_{\mu m}^0$. The conjugate priors for precisions are taken in (27) as a Gamma distribution with hyperparameters c_m^0, r_m^0.

$$\pi_m \sim Dir(\alpha_{\pi m}^0) \tag{25}$$

$$\mu_m \sim \mathcal{N}(\mu_{\mu m}^0, \sigma_{\mu m}^0) \tag{26}$$

$$\phi_m \sim \Gamma(c_m^0, r_m^0) \tag{27}$$

Then each component is associated with one of the VRE plants in the VPP. The hyper-parameters of the component are taken as the mean and variance of production of the corresponding plant. The prior distribution of the proportions is fitted on the shares of installed capacity of the aggregated plants. Finally, the response variable y is transformed from the bounded space $[0, 1]$ to $\mathbb{R}$ by applying the logit transform, by doing so the Gaussian mixture, unbounded in nature, can operate in an unbounded space.

2) Inference of Beta Mixture: Gaussian Mixtures are not able to capture fat tails and can not be directly applied to a bounded process such as renewable generation. Instead, the flexible Bayesian Beta Regression proposed by [13] has a bounded likelihood and can accommodate various types of tails and asymmetries [13]. This mixture is defined by two Beta densities which are re-parametrized to share a common precision parameter ϕ.

The random variable associated with the VPP production y now follows (28),

$$y \sim \sum_{m=1}^{2} \pi_m Be(\lambda_m, \phi) \tag{28}$$

The choice of prior distributions is tailored to the behavior of Beta distributions. We define Beta priors on the common mean $\mu = \sum_{m=1}^{2} \pi_m \lambda_m$ and the distance between marginal means ω, and a Gamma prior on the common precision ϕ because a low common precision is assumed to be probable compared to a high common precision. The mixing proportions are assumed to follow a Dirichlet distribution, parametrized

by the two concentration rates $\alpha^0_{\pi 1}, \alpha^0_{\pi 2}$ of the two mixture components.

$$\pi \sim Dir(\alpha^0_{\pi 1}, \alpha^0_{\pi 2}) \tag{29}$$
$$\mu \sim Be(\alpha^0_{\mu}, \beta^0_{\mu}) \tag{30}$$
$$\omega \sim Be(\alpha^0_{\omega}, \beta^0_{\omega}) \tag{31}$$
$$\phi \sim \Gamma(c^0_{\phi}, r^0_{\phi}) \tag{32}$$

REFERENCES

[1] E. Commission, "COMMISSION REGULATION (EU) 2017/ 2195 - of 23 November 2017 - establishing a guideline on electricity balancing," 2017.

[2] T. Göçmen *et al.*, "Uncertainty quantification of the real-time reserves for offshore wind power plants," 2016, 15th Wind Integration Workshop, Vienna, Austria.

[3] E. I. Batzelis, G. E. Kampitsis, and S. A. Papathanassiou, "Power Reserves Control for PV Systems With Real-Time MPP Estimation via Curve Fitting," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 3, pp. 1269–1280, jul 2017, doi: 10.1109/TSTE.2017.2674693.

[4] K. Knorr *et al.*, "Kombikraftwerk 2 Abschlussbericht," Fraunhofer IWES, Tech. Rep., 2014. [Online]. Available: http://www.kombikraftwerk.de/fileadmin/Kombikraftwerk_2/Abschlussbericht/Abschlussbericht_Kombikraftwerk2_aug14.pdf

[5] M. A. Matos, R. J. Bessa, C. Gonçalves, L. Cavalcante, V. Miranda, N. Machado, P. Marques, and F. Matos, "Setting the maximum import net transfer capacity under extreme RES integration scenarios," in *2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 2016, pp. 1–7, doi: 0.1109/PMAPS.2016.7764145.

[6] C. Gonçalves, L. Cavalcante, M. Brito, R. J. Bessa, and J. Gama, "Forecasting conditional extreme quantiles for wind energy," *Electric Power Systems Research*, vol. 190, 2021, doi: 10.1016/j.epsr.2020.106636.

[7] R. Dupin, L. Cavalcante, R. Bessa, G. Kariniotakis, and A. Michiorri, "Extreme Quantiles Dynamic Line Rating Forecasts and Application on Network Operation," *Energies*, vol. 13, no. 12, p. 3090, Jun. 2020, doi: 10.3390/en13123090.

[8] R. Dupin, "Prévision du dynamic line rating et impact sur la gestion du système électrique," Ph.D. dissertation, 2018, energétique et Procédés, PSL University.

[9] M. Garrido and P. Lezaud, "Extreme Value Analysis : an Introduction," *Journal de la Societe Française de Statistique*, vol. 154, no. 2, pp. pp 66–97, 2013.

[10] J. Browell and M. Fasiolo, "Probabilistic forecasting of regional net-load with conditional extremes and gridded nwp," *IEEE Transactions on Smart Grid*, vol. 12, no. 6, pp. 5011–5019, 2021, doi: 10.1109/TSG.2021.3107159.

[11] P. Pinson, "Very-short-term probabilistic forecasting of wind power with generalized logit-normal distributions," *Journal of the Royal Statistical Society. Series C: Applied Statistics*, vol. 61, no. 4, pp. 555–576, 2012, doi: 10.1111/j.1467-9876.2011.01026.x.

[12] J. Zhang, J. Yan, D. Infield, Y. Liu, and F. sang Lien, "Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model," *Applied Energy*, vol. 241, no. January, pp. 229–244, 2019, doi: 10.1016/j.apenergy.2019.03.044.

[13] S. Migliorati, A. M. Di Brisco, and A. Ongaro, "A new regression model for bounded responses," *Bayesian Analysis*, vol. 13, no. 3, pp. 845–872, 2018, doi: 10.1214/17-BA1079.

[14] N. Laptev *et al.*, "Time-series Extreme Event Forecasting with Neural Networks at Uber," *International Conference on Machine Learning - Time Series Workshop*, pp. 1–5, 2017.

[15] G. Sideratos and N. D. Hatziargyriou, "Wind power forecasting focused on extreme power system events," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 3, pp. 445–454, 2012, doi: 10.1109/TSTE.2012.2189442.

[16] H.-z. Wang *et al.*, "Deep learning based ensemble approach for probabilistic wind power forecasting," *Applied Energy*, vol. 188, pp. 56–70, feb 2017, doi: 10.1016/J.APENERGY.2016.11.111.

[17] Y. Chen, Y. Wang, D. Kirschen, and B. Zhang, "Model-Free Renewable Scenario Generation Using Generative Adversarial Networks," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 3265–3275, 2018, doi: 10.1109/TPWRS.2018.2794541.

[18] Y. Liu *et al.*, "Application of deep convolutional neural networks for detecting extreme weather in climate datasets," 2016, doi: 10.48550/ARXIV.1605.01156.

[19] M. Solas, N. Cepeda, and J. L. Viegas, "Convolutional neural network for short-term wind power forecasting," in *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, 2019, pp. 1–5.

[20] S. Camal, A. Michiorri, G. Kariniotakis, and A. Liebelt, "Short-term forecast of automatic frequency restoration reserve from a renewable energy based virtual power plant," in *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, 2017, pp. 1–6, doi: 10.1109/ISGTEurope.2017.8260311.

[21] ENTSO-E, "TSO's proposal for the establishment of common and harmonised rules and processes for the exchange and procurement of Balancing Capacity for Frequency Containment Reserves (FCR) TSOs' proposal for the establishment of common and harmonised rules and procurement," Oct. 2018.

[22] M. Merten, F. Rücker, I. Schoeneberger, and D. U. Sauer, "Automatic frequency restoration reserve market prediction: Methodology and comparison of various approaches," *Applied Energy*, vol. 268, p. 114978, 2020, doi: https://doi.org/10.1016/j.apenergy.2020.114978.

[23] J. W. Messner, P. Pinson, J. Browell, M. B. Bjerregård, and I. Schicker, "Evaluation of wind power forecasts—an up-to-date view," *Wind Energy*, vol. 23, no. 6, pp. 1461–1481, 2020, doi: https://doi.org/10.1002/we.2497.

[24] S. Camal, A. Michiorri, and G. Kariniotakis, "Optimal offer of automatic frequency restoration reserve from a combined pv/wind virtual power plant," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6155–6170, 2018, doi: 10.1109/TPWRS.2018.2847239.

[25] T. Jónsson, "Forecasting and Decision-Making in Electricity Markets with Focus on Wind Energy," Ph.D. dissertation, 2012.

[26] C. Wan, J. Lin, J. Wang, Y. Song, and Z. Y. Dong, "Direct quantile regression for nonparametric probabilistic forecasting of wind power generation," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2767–2778, 2017, doi: 10.1109/TPWRS.2016.2625101.

[27] T. Gneiting and R. Ranjan, "Comparing density forecasts using thresholdand quantile-weighted scoring rules," *Journal of Business and Economic Statistics*, vol. 29, no. 3, pp. 411–422, 2011, doi: 10.1198/jbes.2010.08110.

[28] C. Diks, V. Panchenko, and D. Van Dijk, "Likelihood-based scoring rules for comparing density forecasts in tails," *Journal of Econometrics*, vol. 163, no. 2, pp. 215–230, 2011, doi: 10.1016/j.jeconom.2011.04.001.

[29] TensorFlow Developers, "Tensorflow," May 2022, doi: 10.5281/zenodo.6574269.

[30] A. J. Cannon, "Quantile regression neural networks: Implementation in R and application to precipitation downscaling," *Computers and Geosciences*, vol. 37, no. 9, pp. 1277–1284, 2011, doi: 10.1016/j.cageo.2010.07.005.

[31] D. Bertoin, J. Bolte, S. Gerchinovitz, and E. Pauwels, "Numerical influence of ReLU'(0) on backpropagation," in *Advances in Neural Information Processing Systems*, ser. Advances in Neural Information Processing Systems 34 (NeurIPS 2021), Paris, France, Dec. 2021.

[32] M. D. Hoffman, D. M. Blei, C. Wang, and J. Paisley, "Stochastic variational inference," *Journal of Machine Learning Research*, vol. 14, no. 40, pp. 1303–1347, 2013.

[33] J. Bröcker and L. A. Smith, "Increasing the reliability of reliability diagrams," *Weather and Forecasting*, vol. 22, no. 3, pp. 651 – 661, 2007, doi: 10.1175/WAF993.1.

[34] N. Meinshausen, "Quantile regression forests," *Journal of Machine Learning Research*, vol. 7, no. 35, pp. 983–999, 2006.

[35] E. Gilleland and R. W. Katz, "extRemes 2.0: An extreme value analysis package in R," *Journal of Statistical Software*, vol. 72, no. 8, p. 1–39, 2016. [Online]. Available: https://www.jstatsoft.org/index.php/jss/article/view/v072i08

Simon Camal received his Eng. degree in Energy and Environmental Engineering from Mines Nancy, France in 2010 and his European Master of Science in Renewable Energy from Loughborough University, UK in 2011. He obtained his PhD at MINES ParisTech - PSL University in 2020 on forecasting and optimization of ancillary service provision by renewable energy power plants. He is currently Project Manager of the Horizon2020 Smart4RES Project, working at MINES ParisTech Center for Processes, Renewable Energies and Energy Systems (PERSEE) in Sophia Antipolis, France.

Andrea Michiorri received his Eng. degree in Mechanical Engineering from the University of Rome 'La Sapienza' in 2005. He then obtained a PhD from the University of Durham in 2010 with a dissertation on the thermal state estimation of power system components. He is currently an associate professor at the MINES ParisTech PERSEE Centre, working on the integration of renewable resources and distributed generators into the power system, with a focus on decision making under uncertainty.

Georges Kariniotakis (S'95-M'02-SM'11) was born in Athens, Greece. He received his Eng. and M.Sc. degrees from Greece in 1990 and 1992 respectively, and his PhD from Ecole des Mines de Paris in 1996. Currently he is Professor at MINES ParisTech. He is head of the Renewable Energies and Smart Grids Group at PERSEE Centre. He has authored more than 250 scientific publications in journals and conferences. He has been involved as participant or coordinator in more than 45 R&D projects in the fields of renewable energies and distributed generation. Among them, he was the coordinator of some major EU projects in the field of wind power forecasting such as Anemos, Anemos.plus and SafeWind projects. Currently he is the coordinator of the H2020 Smart4RES project. His scientific interests include among others timeseries forecasting, decision making under uncertainty, modeling, management and planning of power systems.