Forecasting and optimization of ancillary services provision by renewable energy sources

Simon Camal

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à MINES ParisTech

Prévision et optimisation de l'offre de services système par des énergies renouvelables
Forecasting and optimization of ancillary services provision by renewable energy sources

Soutenue par
Simon Camal
Le 25 02 2020

École doctorale n°621
Ingénierie des Systèmes, Matériaux, Mécanique, Energétique

Spécialité
Energétique et procédés

Composition du jury :

Marc PETIT	*Président*
Prof., Centrale-Supélec	*Rapporteur*
Manuel MATOS	*Rapporteur*
Prof., INESC TEC	
Nicolaos Antonio CUTULULIS	*Rapporteur*
Prof., Technical University of Denmark	
Ivana KOCKAR	*Examinateur*
Dr., Reader, University of Strathclyde	
Vincent LEFIEUX	*Examinateur*
Dr., Ingénieur, RTE	
Georges KARINIOTAKIS	*Directeur de thèse*
HDR, Directeur de Recherche, MINES ParisTech	
Andrea MICHIORRI	*Maître de thèse*
Dr., Enseignant-Chercheur, MINES ParisTech	

Acknowledgements

Une seule certitude suffit à celui qui cherche. *Albert Camus, Le mythe de Sisyphe*

Contents

List of Tables

List of Figures

Nomenclature

Abbreviations

Power systems

AAP	Available Active Power
aFRR	automatic Frequency Restoration Reserve
AGC	Automatic Generation Control
BRP	Balancing Responsible Party
BSP	Balancing Supply Provider
FCR	Frequency Containment Reserve
GCT	Gate Closure Time
mFRR	manual Frequency Restoration Reserve
MOL	Merit-Order List
TSO	Transmission System Operator
VPP	Virtual Power Plant
VRE	Variable Renewable Energy

Forecasting, Statistics

AIC	Akaikes Information Criterion
ARIMA	Auto-Regressive Integration Moving Average
AV	Availability
BS	Brier Score
BMBR	Beta Mixture Bayesian Regression
CNN	Convolutional Neural Network
CvM	Cramer-von Mises Statistic
DG	Direct Gaussian
EVT	Extreme Value Theory
FCNN	Fully-Connected Neural Network
GBT	Gradient Boosting Trees
GMBR	Gaussian Mixture Bayesian Regression
IG	Indirect Gaussian
IV	Indirect Vine
KDE	Kernel Density Estimator
KKT	Karun-Kush Tucker
KL	Kullback-Leibler divergence
LSTM	Long Short Term Memory
MASE	Mean Average Scaled Error
MCMC	Monte Carlo Markov Chain
$\mathcal{MVN}$	Multivariate Normal
MI	Mutual Information
MRS	Markov Regime Switching

PI	Prediction Interval
PIT	Probability Integral Transform
PCA	Principal Component Analysis
PSD	Power Spectrum Density
QRF	Quantile Regression Forests
QL	Quantile Loss
QLR	Quantile Linear Regression
QS	Quantile Score
Relu	Rectified Linear Unit
RF	Random Forests
RMSSE	Root Mean Squared Scaled Error
RNN	Recurrent Neural Network
RUF	Rate of Under-Fulfilment
SF	Smoothing Factor
SVI	Stochastic Variational Inference
SVR	Support Vector Regression
VS	Variogram-based Score

Indices and Superscripts

$\uparrow,\downarrow$	Upward,downward regulation
agg	Level of aggregation in the VPP
c	channel in convolutional layer
da	Day-ahead variable
γ	Order of VS
h	Horizon of forecast
k	Component in a density mixture
l	Leaf of a tree in a QRF model
(l)	Layer index in a neural network
ω	Index of production scenarios
rt	Real-time variable
s	Index of energy source type in the VPP
$sources$	Relative to energy sources in the VPP
t	Runtime of forecast
t'	Timestep within validity period of reserve bid
$total$	Relative to the total production of the VPP
w	Width of convolutional filter

Variables

$\alpha_{\Delta R^-,.}$	Positive variable, used for identification of CVaR on reserve deficit
a_R	Reserve activation probability
b	Binary variable for ramp function
b	Bias in neural network
$\mathbf{c}$	Cost vector associated to energy and reserve bidding
$\mathbf{c}_{(t)}$	Long-term memory in LSTM
$\hat{c}$	Copula density function
$\hat{C}_{emp}$	Empirical copula cumulative density function
$\mathbf{B}$	Vector of decision variables for bidding
δ	Gradient indicator function for Brier Score
Δf	Frequency deviation
$\Delta \pi_{RE}$	Realization of spread price between reserve and energy
ΔP_{RE}	Random variable of spread price between reserve and energy
$\Delta R^{\uparrow,-}$	Deficit of upward reserve in real-time
$\Delta R^{\downarrow,-}$	Deficit of downward reserve in real-time
D_k	Distribution of component k in distribution mixture
$\epsilon_{MA,j}$	Moving average term of order j
$\epsilon_{\pi,.}$	Error of price forecasts
E	Bid on the energy market
$E^{rt,+}$	Energy surplus on the real-time market
$E^{rt,-}$	Energy deficit on the real-time market
$ELBO$	Evidence Lower Bound
ϕ	Generator function of Archimedean copulas
f	Filter of convolutional neural network
$\mathbf{fg}_{(t)}$	Forget gate in LSTM
g_t	Negative gradient loss in GBT model
$\mathbf{g}_{(t)}$	Main gate in LSTM
h	Hidden layer in CNN
$\mathbf{h}_{(t)}$	Short-term memory in LSTM
$\mathbf{ig}_{(t)}$	Input gate in LSTM
K_p	Proportional gain of FCR provision
K_u, K_v	Kernel of KDE Copula on u, v
λ	Lagrange multiplier of multi-product reserve optimization
L	Lagrangian of multi-product reserve optimization
l_E	Linear loss function on the energy market
l_R	Linear loss function on the reserve market
$l_{R,PI}$	Linear loss function on the reserve market with perfect information
$\mathbf{o}_{(t)}$	Output layer in LSTM
ψ	Positive variable for ramp function
p_ω	Probability of occurrence for scenario ω

$\pi_E^{rt,-}$	Real-time energy price
π_E^*	Net penalty price on the energy market
$\pi_{E_R}^{da}$	Day-ahead price for reserve balancing energy
$\hat{\pi}_{persist}$	Persistence price forecast
π_R^{da}	Day-ahead price for reserve capacity
$\pi_{\bar{R}}^{da}$	Day-ahead price for average reserve price
π_R^*	Net penalty price on the reserve market
ϖ_k	Optimal prediction at node k of a tree in GBT model
ρ_ω	Upper bound of distance between revenue at scenario ω and value-at-risk
R	Bid on reserve market
R_0	Total reserve capacity for multi-product reserve bidding
R_{demand}	Reserve demand of a TSO
Σ	Covariance matrix of multivariate copula
s_{aFRR}	Setpoint of aFRR
s_i	Resource coefficient of reserve product i in multi-product reserve bidding
s_t	State of RNN model at time index t
S_l	Subset of data in QRF model at leaf l
S_{multi}	Stock constraint in multi-product reserve bidding
$\tau(\tau_{min}, \tau_{max})$	Quantile nominal value (min, max)
τ_R	Quantile nominal value of forecasted distribution for reserve
τ_R^{opt}	Optimal quantile nominal value of forecasted distribution for reserve
τ_R^{marg}	Marginal quantile nominal value of forecasted distribution for a specific reserve product
τ_{rev}	Quantile nominal value of revenue
θ_{da}	Value-at-Risk of revenue in stochastic optimization
t_s	Threshold value in QRF model
u	PIT transform of VPP production
$\mathbf{u}$	Vector of uniform i.i.d. variables
U	Generalized cumulative distribution function
u_{ext}	Threshold of EVT model
v	PIT transform of spread price between reserve and energy
V	Total volume traded on market
w	Weight of observation-prediction pair in VS
w^T	Weights in LSTM layers
x	Realization of explanatory variables
X	Random variable modelling vector of explanatory variables
y	Realization of power production
$\hat{y}_{AAP}$	Estimated available active power
y_{ref}	Constant reference set-point of production
Y	Random variable associated with power production
Y^{agg}	Random variable associated with aggregated power production
$\mathbf{Z}$	Multivariate normal random variable
$\mathbf{z}$	Realization of multivariate normal random variable

Parameters

α_{CVaR}	Quantile value of CVaR revenue in stochastic optimization
α_k	Hyperparameter, concentration of probability of mixture component k
β_{bn}	Offset of batch normalization
$\hat{\alpha}_{\text{base}}$	Mean parameter in base regime of MRS model
$\hat{\beta}_{\text{base}}$	Mean-reverting parameter in base regime of MRS model
β	Aversion to risk on revenue
β_V	Coefficient of Volumes V in ARIMA model
B	Number of data batches
c_k	Hyperparameter, concentration of precision of component k in density mixture
$db_{\Delta f}$	Frequency deadband in FCR
$\mathcal{D}$	Decision space
ϵ	Confidence parameter of chance constraints
ϕ_i	Auto-regressive term of order i in ARIMA model
ϕ_k	Precision of component k in density mixture
γ	Extreme value Index of EVT model
γ_{bn}	Scaling factor of batch normalization
η	Shrinkage parameter in GBT model
h_u, h_v	Bandwidth of KDE on u, v
k	Threshold quantile value for EVT model
$k_{w/h/c}$	Size of convolutional kernel
K_p	Droop control coefficient of plant p
μ_{bn}	Mean of batch normalization
μ_k	Mean of component k in density mixture
μ_{μ_k}	Hyperparameter, mean of mean of component k in density mixture
M	Big-M parameter
ν	Hyperparameter set of an MDN
N	Number of observations
$N_{iterations}$	Number of iterations
ω	Distance between means in Beta mixture
Ω	Number of scenarios
O	Number of output nodes in neural network
P	Number of plants in VPP
π_k	Proportion of component k in density mixture
r_k	Hyperparameter, rate of precision of component k in density mixture
r_{BS}	Threshold value of ramp in BS for production ramps
$s_{w/h}$	Stride of convolutional kernel
σ	Scale of EVT model
σ_{bn}	Standard deviation of batch normalization
σ_{μ_k}	Hyperparameter variance of mean of component k in density mixture
S	Number of energy sources in the VPP

θ	Parameter vector of regression models
θ_j	Moving average coefficient of order j in ARIMA model
τ_{lim}	Nominal value of reference quantile for exponential model
$\sqcup$	Index of tree in decision-tree models
$\mathcal{T}$	Number of trees in decision-tree models
$T_{validity}$	Duration of reserve validity period
TS	Thresholds of regimes in MRS model
u_{Be}	Parameter of Beta mixture
v_{Be}	Parameter of Beta mixture
y_{max}	Installed power capacity

Operators and Functions

$a_y(.)$	Log-normalizer of response in exponential distribution family
$Be(.)$	Beta Distribution
$Dir(.)$	Dirichlet distribution
$f(.)$	Probability density function
$\bar{F}(.)$	Survival cumulative distribution function
$F(.)$	Cumulative distribution function
$F^{-1}(.)$	Quantile function
$\Phi(.)$	Probit function
$\Gamma(.)$	Gamma distribution
$g(.)$	Link function between features and response in mixture density model
$H(.)$	Heaviside function
$\mathcal{L}$	Likelihood
L	Lag operator
$m_{.}$	Marginal CDF of copula
$m_{\hat{Y}}(.)$	Regression model of aggregated production
$\mathcal{M}(.)$	Multinomial distribution
o	Objective function for risk-neutral stochastic optimization
o_{β}	Objective function for risk-averse stochastic optimization
$p(.)$	Prior distribution
$q(.)$	Posterior distribution
$r(.)$	Natural parameter of exponential distribution family
$\sigma(.)$	Sigmoid activation function
$t(.)$	Sufficient statistic of exponential distribution family
$T(.)$	Decision-tree in QRF or GBT model

Chapter 1

Introduction

1.1 Context

The integration of Variable Renewable Energy (VRE) in electrical networks is rising in many parts of the world. In Europe, where power generation is still dominated by non-renewable energy sources, a target of 30 % of gross final energy demand delivered by renewable sources in 2030 has been defined in the Clean Energy Package of the European Commission [1]. The transition towards renewable production is not only recorded in legislation, but also observed on the field: between 2015 and 2017, 54 GW of Wind and PV generation capacity have been installed in Europe [1]. This leads to a new situation of electrical networks, where the share of renewables in the electricity production is mostly driven by variable sources (mostly Wind, with small contributions of Photovoltaics and run-of-river Hydro) and tends progressively to compete with conventional non-renewable generation. In this new situation, the variability of VRE production challenges the operation of electrical networks.

An electrical network needs to operate at stable conditions to guarantee a continuous power supply with maximum reliability. The reliability of power provided by the network to its users is high in Europe: the average duration of disconnection of users on the French network over the decade 2008-2018 is inferior to 3 minutes per year [2]. This level of reliability requires however a permanent control effort: due to the fact that the transmission infrastructure is not designed to store large amounts of electrical energy, flows of active power produced by generators must match at every instant the global power consumption on the network. If an additional electric load connects to the network, while the number of connected synchronous generators remains constant, then the power produced by generators becomes inferior to the total power requested. By conservation of momentum, the frequency on the network drops, with a rate of change of frequency that depends on the inertia of the entire power system. If frequency falls outside of the standard frequency range around the nominal value ($50Hz \pm 50mHz$ in continental Europe), the deviation in frequency must be first contained and then restored in order to avoid damage to electrical devices or blackouts, which cause significant losses to the economy.

Frequency levels can be restored by adjusting power flows on the transmission system: when a Transmission System Operator (TSO) observes an imbalance between power production and

1

consumption over its perimeter or "control area", it requests the activation of power reserves from flexible assets to counter the effect of the imbalance, in the opposite direction to the frequency deviation. The flexible assets that are currently activated include:

- Dispatchable power plants such as nuclear, thermal or hydro power plants,

- Storage devices such as batteries, flywheels, supercapacitors or pumped-hydro,

- Flexible consumers

One notes that Variable Renewable Energy (VRE) sources, such as Wind, Photovoltaics (PV) or run-of-river Hydro power plants, are not present in the list above. We will see later that recent developments have shown that VRE, albeit the significant uncertainty on their production level, dispose of the technical capacity to regulate efficiently their active power. The grid-following, inverter-based production of VRE may become dominant in the total generation capacity at medium-term horizons [3]. As VRE tend to substitute conventional synchronous generators in the generation mix, they reduce the reserve capacity that can be traditionally activated without significant uncertainty on production [4]. The contribution of VRE to balancing AS appears therefore as a necessity for electrical networks. Beyond balancing, other AS are critical for electrical networks, e.g. inertia, fast frequency response or voltage support. Services concerned with response times at sub-second scale to ensure system stability are also heavily impacted by VRE integration. Developments on control systems for renewables improve their capacity to emulate inertia [5] and their grid-forming capabilities [3]. VRE are also capable of providing voltage support [6], [7]. However the present thesis will focus on balancing Ancillary Services (AS) for frequency control, provided by Balancing Service Providers (BSP) to TSOs.

1.1.1 Structure of the balancing AS markets

The activation of power reserve in European synchronous areas is structured following three successive levels, represented in Figure 1.1. The first level of power reserve is Frequency Containment Reserve (FCR), activated instantaneously over the whole synchronous area: all plants that take part in the various FCR mechanisms existing in Europe regulate a share of their power generation in a droop control mode. The droop control of a plant p modulates active power in (1.1) proportionally with a factor K_p to the frequency deviation from the nominal frequency Δf, if the absolute deviation is higher than a dead-band value $db_{\Delta f}$. This decentralized response activated in less than 30 seconds, enables to contain the frequency deviation to a maximum of 200 mHz.

$$FCR_p(\Delta f(t)) = -K_p \Delta f(t), \quad \forall \Delta f(t), |\Delta f(t)| \geq db_{\Delta f} \tag{1.1}$$

After frequency has been contained, a second reserve level defined as automatic Frequency Restoration Reserve (aFRR) is mobilized by TSOs in their respective control areas. TSOs send regulation set-points to BSPs in their control area to restore the nominal frequency within 15 min at most. A plant p bidding an energy quantity E_p^{da} on the day-ahead energy market and a reserve capacity $R_{aFRR,p}^{da}$ on a day-ahead aFRR market will see its production modulated in (1.2) as a function of the Automatic Generation Control (AGC) set-point $s_{aFRR}(t) \in [-1, 1]$ sent by the TSO. The
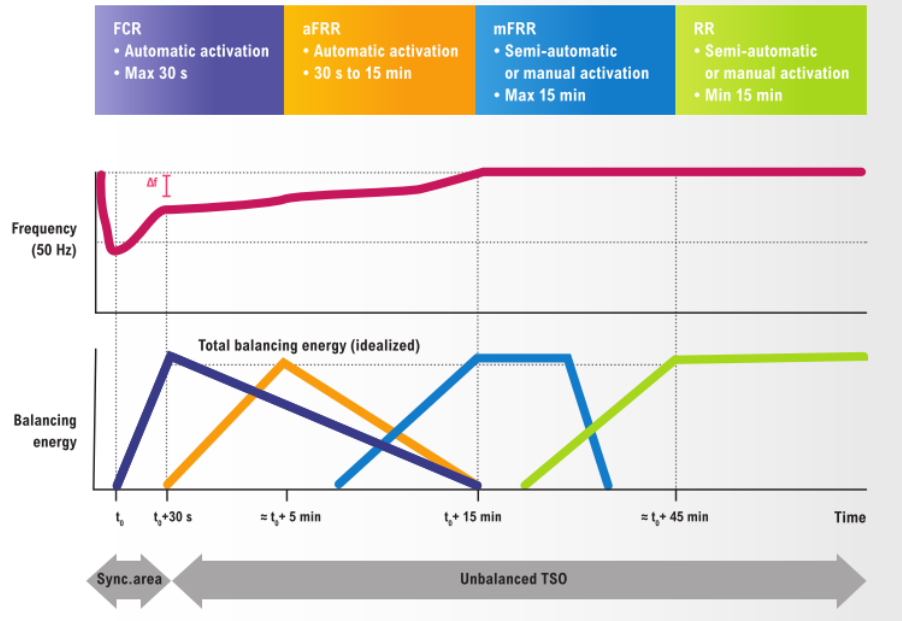
Figure 1.1:   Frequency control scheme implemented in Europe [8]

reserve capacity must be guaranteed during the whole delivery period $T_{validity}$. This duration is frequently denominated "product length" in the AS terminology.

$$Y_p(t) = E_p^{da}(t) + E_p^{rt}(t) + s_{aFRR}(t)R_{aFRR,p}^{da}, \quad \forall t \in T_{validity} \tag{1.2}$$

Finally, the nominal frequency is maintained by semi-automatic or manual reserve requests through manual Frequency Restoration Reserve (mFRR) and Replacement Reserve (RR), where new BSPs substitute BSPs implicated in aFRR, in order to preserve sufficient reserve capacity. In what follows, the markets concerned with imbalances and ancillary services for frequency control (FCR, aFRR, mFRR/RR) will be denominated "balancing reserve markets".

Sizing balancing reserve needs. Deviations between generation and consumption can originate either from incidents occurring on transmission elements, power generators, or load), or imbalances between expected and delivered quantities in power generation, storage or load. TSOs prepare to incidents by sizing reserves with diverse static approaches listed in Table 1.1.

Sizing methodology	Balancing AS	References
N-1 criterion: cover loss of largest plant in Europe	FCR	[9]
Deterministic model, function of network state in TSO's control area	aFRR	[9]
Convolution model on sources of uncertainty	aFRR, mFRR	[4]

Table 1.1: Static sizing approaches for several balancing AS

The increased share of uncertain production from VRE in the network challenges the sizing of balancing AS, both in volume and in methodology. German TSOs expect that an increase of 40 GW of VRE capacity will lead to an increase of balancing need of 1.5 GW [10]. Probabilistic AS

sizing methodologies have been proposed to take into account the uncertainty of VRE production [11]. Stochastic unit commitment models propose to optimize the scheduling of dispatchable plants in order to accommodate VRE penetration while minimizing costs [12]. A first scheduling can be issued in the day before delivery, and modified in intraday according to updates on the expected state of the network. The dynamical scheduling of balancing reserve is employed by several TSOs such as RTE in France. In the USA, the wholesale energy market and the AS markets are co-optimized in a joint unit commitment, initiated in the day-ahead and update in intraday.

The volume of imbalances also depends on how TSOs expect BSPs to anticipate their imbalances: in Germany, BRPs must declare in the day ahead a balance between their power supply (e.g. power production or power purchase) and their power demand (e.g. power consumption or power sales). This physical balance is not mandatory in France, resulting in less recourse to the intraday energy market (4 TWh in France, 40 TWh in Germany in 2010 [13]) and higher volumes of imbalances to be paid at the TSO after delivery. In any case, the volume of balancing markets is limited when compared to the wholesale short-term energy market: the yearly volume of imbalance in France has reached 1.7 % of the total electricity consumption [14]. Most of the balancing markets in Europe accept balancing offers from suppliers located in neighbouring control zones, e.g. 41 % of the upward volumes of balancing energy requested in France were from foreign countries [14]. The increased cooperation between control areas also led to standardization of procedures, illustrated by a decrease in AS volumes requested by the German TSOs between 2010 and 2011 [4]. However, the overall volume of balancing reserve in the European synchronous area increased of 21 % between 2016 and 2017, reaching a total contracted capacity of 4.56 TW [1].

Harmonization of balancing in Europe. The Electricity Balancing Guideline (EB) regulation (2017/2195) [15] established in 2017 by the European Commission profoundly reshapes the structure of balancing AS markets. The main objectives of the EB regulation are the standardization of the procurement of balancing energy throughout Europe, an improvement of exchanges across regions and countries, and a better integration of new players such as demand response and renewable production. The regulation is completed by proposals from TSOs where they expose their implementation objectives on specific AS markets. The most salient outcomes are:

- **Single-price mechanism for imbalance energy markets** (Article 52c in [15]). The previous dual price in imbalance markets penalized only imbalances that had worsen the total net imbalance, while remaining neutral if imbalances happened to be of the opposite sign of the total net imbalance. This was thought to incentivize BRPs to adjust their perimeter before delivery, but economic works have shown that the single price scheme, illustrated in Figure 1.2 is more efficient in terms of global cost, because it depends on the total net imbalance instead of individual states of BSPs [16].

- **Gate closure times close to delivery.** The Gate Closure Time (GCT) of balancing energy for aFRR, RR, mFRR is placed less than one hour before delivery, to ease the asset management of BSPs [17].

- **Pay-as-cleared auctions for balancing energy**. Auctions for balancing energy bids (Article 30a in [15]) should be cleared following the pay-as-cleared principle instead of the pay-

4

as-bid scheme implemented in some auctions such as the FCR platform before 2025. This clearing is thought to improve the overall economical efficiency of balancing markets.

- **Short validity periods.** Long validity periods, where balancing capacities must be ensured over several hours up to a entire week, have been blocking the integration of fluctuating assets such as VRE. Short validity periods of balancing AS located below the hour and compatible with the imbalance settlement period, such as 15 min, appear to form consensus among TSOs [18].

- **Allow for aggregations to provide balancing.** Aggregations of flexible assets should be allowed to provide balancing. Some TSOs underline that detailed information on the assets composing the aggregations is necessary to guarantee a secure operation of balancing AS [13].

- **Allow for separate bids of upward and downward reserve**. For FCR, offering simultaneously upward and downward capacity seams technically reasonable as the automatic containment of frequency deviations is equally likely to be needed in either upward or downward direction. In contrast, aFRR, mFRR and RR products are activated by TSOs as a function of the specific conditions of the imbalance to resolve: therefore bids on these markets can easily be separated into upward and downward capacities (Article 32(3) in [15]).
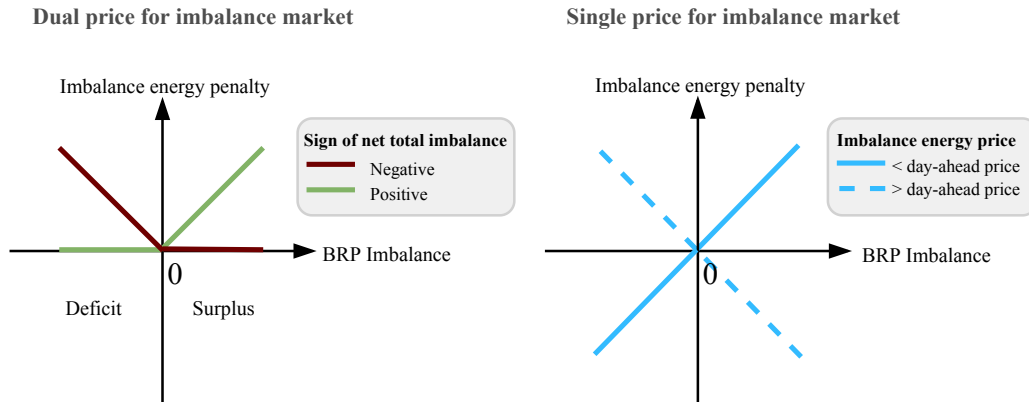
Figure 1.2: Pricing of imbalance energy markets, dual price versus single price

Payment of balancing capacity and energy. The FCR market bases its tenders on capacity prices, because the FCR activation can be considered symmetrical over validity periods, resulting in small net volumes of balancing energy. In contrast, activations in aFRR lead to significant net volumes of balancing energy in either upward or downward direction. Therefore there are two options for clearing aFRR tenders: (1) a double-price with a merit-order based on capacity price and a second merit-order for activation based on balancing energy price, (2) merit-order based on balancing energy price only. Current aFRR prices in Europe are mostly based on a double-price market, in order to secure enough capacity at medium-term (bilateral contracts) or at short-term (aFRR pool). Bidding downward reserve is a profitable action for VRE in a broad economical sense: pre-curtailment of production is not required, therefore renewable energy contribution is

maximized which is beneficial to global political targets and renewable producers do not miss the revenue opportunity on the wholesale energy by bidding all their expected volume.

Pre-qualification of BSPs. BSPs who wish to enter balancing AS markets must be able to provide the AS product according to frequency or AGC signal with minimal delay and deviation, according to the pre-qualification rules defined by each TSO. Following the EBGL regulation, a standardization of balancing AS products has started to harmonize the main temporal and volume characteristics of products (cf. Figure 1.3). The full activation time of the product, after which the regulation should have attained 100% of the bid volume, is an achievable constraint for VRE plants: individual plants react within seconds, but in case of aggregations of VRE plants inside a VPP efficient communication and control protocols must be implemented to minimize the reaction delay [19].
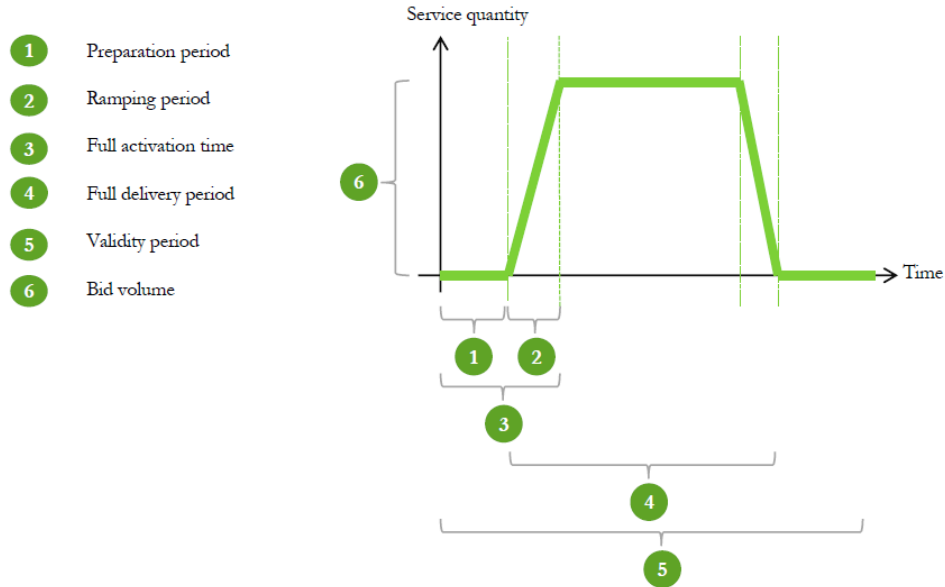
Figure 1.3: Standard balancing AS product, inspired from ENTSO-E [8]

In order to accommodate the variability of renewable production during the validity period, German TSOs integrate in their pre-qualification procedures that the response of the activated asset can deviate from the regulation signal within defined intervals or "corridors" [20]. These intervals are defined for the various operating states within the test period (idle, activation, full delivery, deactivation, as illustrated in Figure 1.4). Altogether requirements for qualification to balancing AS are still challenging for fluctuating producers:

- German TSOs tolerate underfulfillments under 10% of the requested reserve capacity during the full delivery period, and 20% during the ramping period [20].

- German TSOs limit the frequency of deviations (within tolerated range) to 5% of all measurements during pre-qualification tests [20].

- French TSOs impose a penalty equal to 5 times the reference capacity price in case of under-fulfillment on FCR or aFRR [14].

Lastly, these requirements are likely to evolve in the coming years, along with the implementation of new balancing markets in Europe.
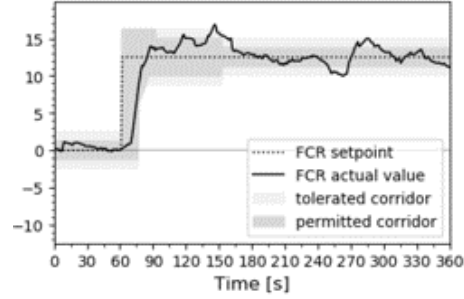
Figure 1.4: FCR pre-qualification test corridor, defined by German TSOs, with example response [20]

Balancing AS markets, impacted by the rising generation from VRE sources, show promising evolutions towards the integration of VRE as providers of balancing: balancing markets are re-organized at short term coherently with energy markets, and aggregations, which are necessary to attain similar capacities than large dispatchable plants, can participate into these markets. However challenges are identified: delivery periods, even if recently reduced, remain of several hours and constrain VRE to offer their minimal output over these periods. Furthermore, participation in upward regulation imposes curtailment of renewable production, which may be in contradiction with the societal objectives of maximizing low-carbon generation in the electricity mix.

1.1.2 Provision of balancing AS by renewables

Capacity of power regulation. Weather-dependent renewable power plants dispose of the adequate control technology to provide active power regulation within seconds and sustain the control during the whole AS balancing product length. More specifically, wind turbines can be controlled by their torque speed [21], Maximum Power Point tracking (MPPT) and pitch regulation in high wind conditions [22]. The variations in wake effect occurring on wind farms after turbines have been curtailed must be assessed [23] and the global uncertainty on available power estimation from real-time SCADA measurements effectively reduced by statistical methods based on real-time SCADA measurements [24]. PV inverters can incorporate a modified Maximum Power Point Tracker (MPPT) which enables subsecond active power control [25]. In Figure 1.5, the provision of aFRR from a large PV power plant is verified experimentally during tens of minutes during a clear morning. The green curve indicates the available power of the farm estimated in real-time, and the orange curve represents the available power reduced by the reserve capacity band (called headroom on the Figure). It is observed that production measured in yellow, initially curtailed at 100% of the reserve capacity, is regulated and follows with accuracy the power regulation signal in red. In less clear-sky conditions, the evaluation of reserve effectively provided is more challenging

7

due to higher uncertainties on the estimation of the available power. While research is still in early stages regarding AS at subsecond scale such as synthetic inertia or grid-forming inverters [3], the active power regulation of VRE has already reached development at the industrial scale.

Figure 1.5: Provision of aFRR by a PV power plant of 300 MW capacity, illustration by NREL [5]

Need for available active power estimation The provision of balancing from VRE raises an additional issue, which is the estimation in real-time of the effective volume of reserve provided by fluctuating plants during validity periods of their balancing bid. To this end, a reference production, had the plant not been activated for reserve, should be robustly available to TSOs. A possible method to evaluate balancing provision from VRE is the *constant set-point method*: production is capped to a reference set-point y_{ref} before the delivery, and is curtailed during the delivery period Δt by regulating power from y_{ref} to the activated reserve band in (1.3), as a function of reserve capacities $(R_\uparrow, R_\downarrow)$ and activation in either upward or downward direction $(a_\uparrow, a_\downarrow)$. In contrast, the *AAP method* regulates power in (1.4) with reference to a real-time estimation of the AAP $\hat{y}_{AAP}(t)$ based on a production model integrating down-regulation and real-time information on weather conditions. The constant reference set-point y_{ref} is not favourable for VRE production because it may lead to overestimations of the reserve capacity if the wind speed or radiation decreases during activation (cf. Figure 1.6), whereas the AAP adapts in real-time to observed weather conditions (cf. Figure 1.7).

$$y(t) = y_{ref} - (1 - a_\uparrow(t))R_\uparrow + a_\downarrow(t)R_\downarrow \tag{1.3}$$

$$y(t) = \hat{y}_{AAP}(t) - (1 - a_\uparrow(t))R_\uparrow + a_\downarrow(t)R_\downarrow \tag{1.4}$$

The error of the AAP estimation is significant in the case of wind farms: when turbines located at the front row of the farm start to down-regulate their production, they decrease wake losses, consequently the direct aggregation of power curves at turbine levels would overestimate the true AAP at the farm level. A better solution consists in a robust wake model adapted to down-regulation and integrated into the SCADA system of the wind farm [28]. The re-calibrated Larsen Model of [24] decreases the mean bias on the AAP estimation of a donw-regulated offshore wind farm to 1% compared to 8 % with direct sum of power-curve turbine-based AAP estimations. For large PV farms, the challenges of AAP estimation come from the possible heterogeneity in solar resource over the plant. This can be tackled by recalibrating estimations of I-P curve based on meteorological measurement by sensors (pyranometers, temperature, wind speed) dispersed over
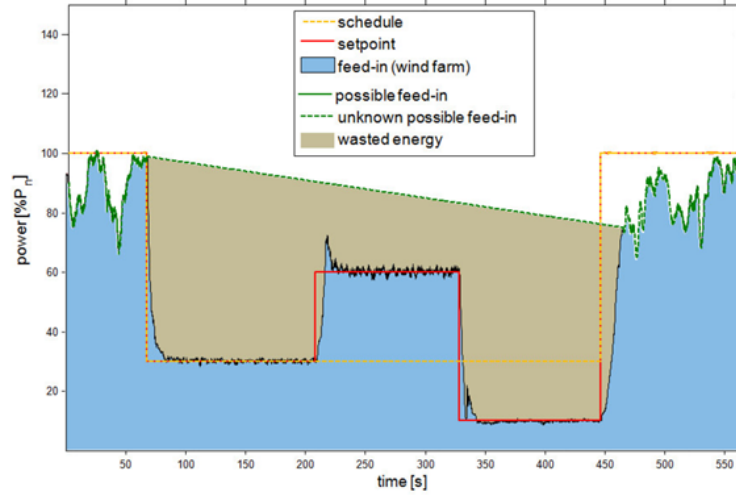
Figure 1.6: Activation of reserve on wind farm based on a constant reference setpoint [26]
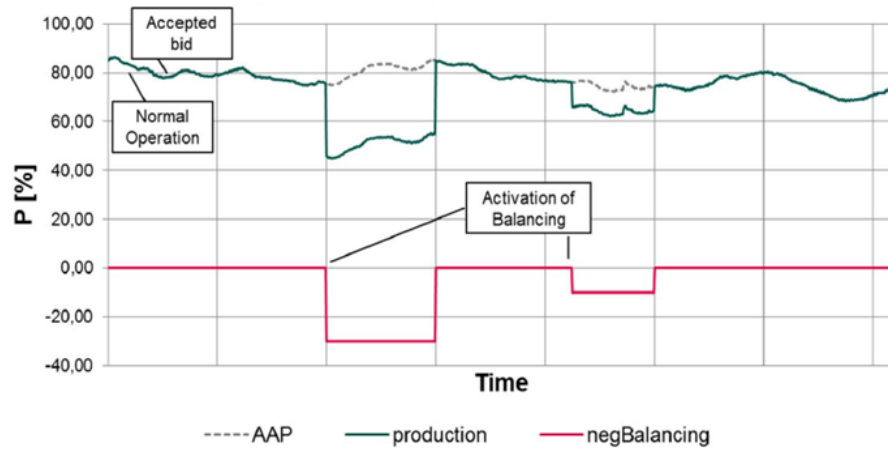
Figure 1.7: Activation of reserve based on the AAP of a Wind farm [27]

the farm. An alternative can be to dedicate reference arrays that won't be regulated, and fit a regression model enabling to estimate the production of regulated arrays given the observation of the unregulated reference arrays [19] (Figure 1.8).

In conclusion, variable renewables appear to have the technical capacity to provide active power reserve. However, it is clear that the available production must be accurately assessed in real-time to validate the effective volume of reserve provided. As it is known that the uncertainty of aggregated VRE production is lower than the uncertainty of single power plant [29], an aggregation combining Wind and PV plants is an interesting candidate for AS provision. The next section describes the available solutions to perform such an aggregation.
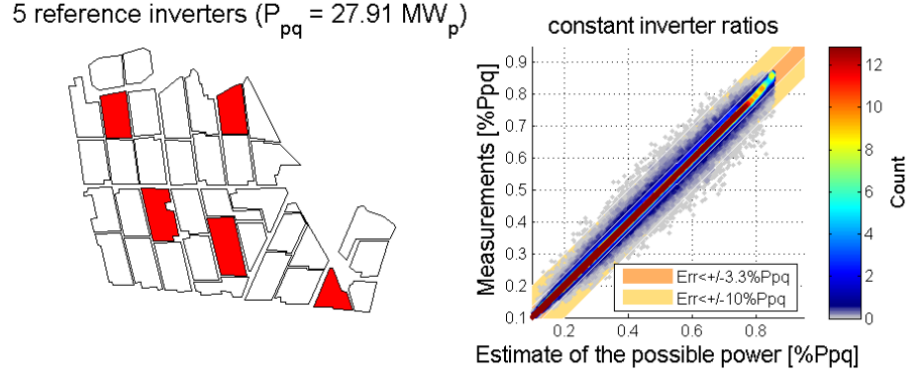
Figure 1.8: Estimation of AAP of PV farm by regression on reference inverters [19]

1.1.3 Industrial applications based on the Virtual Power Plant (VPP) technology

In order to supply balancing AS, the operation of a VRE aggregation must be allowed not only perform a financial aggregation of the assets, but also to dispatch set-points to the individual plants. The VPP technology offers a mature solution for the aggregation of power plants and the dispatch of control signals, possibly over large distances. Various projects have already employed VPP to provide AS, and in particular VRE-based balancing AS. The Kombikraftwerk project deployed an internet-based VPP connecting 80 MW capacity of onshore wind power plants, photovoltaics and biomass generation, all located in Germany to experimentally provide a balancing AS in conditions reproducing pre-qualification tests. The experimentations of Kombikraftwerk showed promising responses to regulation set-points, however results have proven to worsen when wind conditions were low, some turbines of a wind farm operating around 5 % of the installed capacity disconnected in the middle of the test in Figure 1.9. As a result, the overall quality of reserve provision was not satisfactory when compared to conventional power plants providing balancing AS.

The follow-up project ReWP [19] increased the effort on the estimation of AAP for PV parks, and achieved overall best results even if some tests were conducted on purpose during low wind conditions and mixed-sky conditions over PV farms. The project suggests that Wind and PV plants reacts within full activation times of aFRR without problems when proper communication protocols are established. Reserve allocations to Wind farms should reach at least 10 % of the farm capacity to secure a fast response (see reserve activation on a large onshore Wind farm in Figure 1.10).
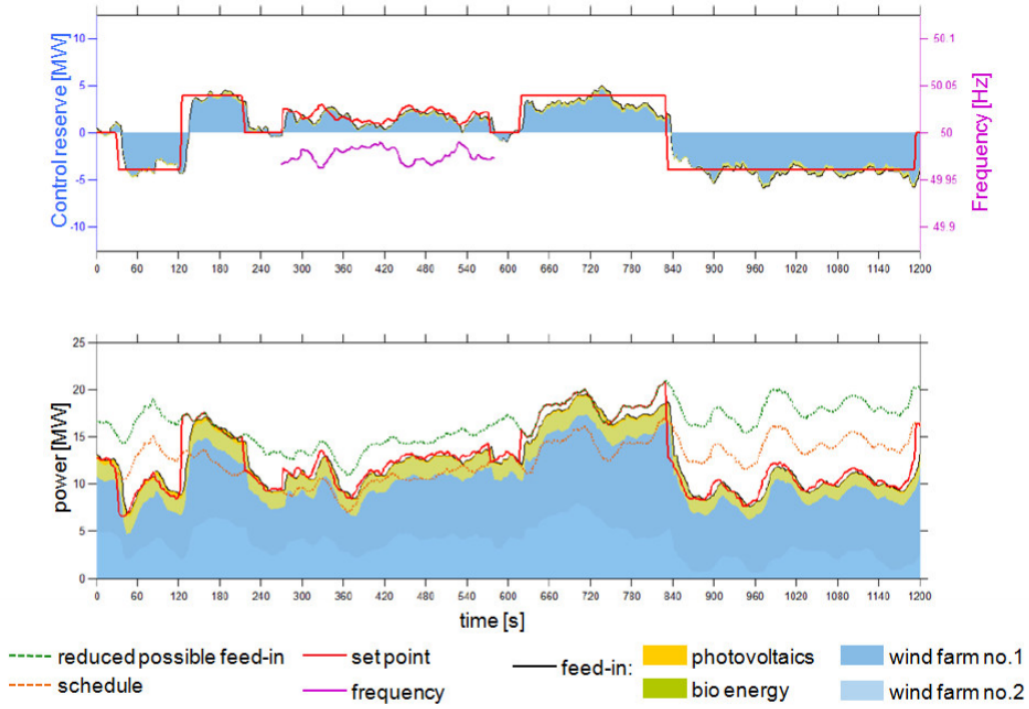
Figure 1.9: Experimentation of reserve delivery of a small Wind/PV/biomass VPP in Germany [26]
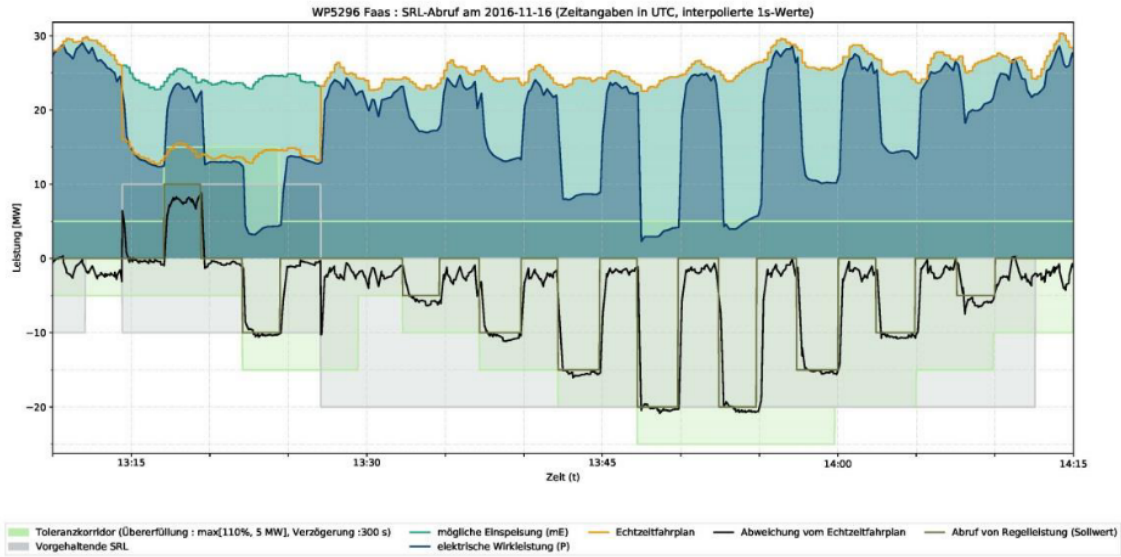
Figure 1.10: Balancing AS experimentation, inspired by AS pre-qualification tests, for a single Wind farm [19]. The AAP (in green) shows a slight deviation from the active power (in blue) when no curtailment is requested (e.g. between 13:17 and 13:20). Between 13:15 and 13:27, a symmetrical reserve provision with pre-curtailment of power. After 13:27, a downward-only provision is tested. Response to requested reserve is figured by the black curve of the negative power axis.

1.2 Motivation and Challenges

The integration of renewable power plants into electricity markets requires that their variable production should be properly anticipated before offering this production under the form of energy volumes and reserve capacities. Quantification of uncertainty on production is therefore needed to take informed decisions on volumes to bid on the energy and AS markets. But one may wonder how a production forecast, with quantification of production uncertainty, can help formulate offers on a market.

The uncertain nature of weather-dependent renewable production led to the development of probabilistic forecasting methods, producing a distribution of possible production values. The moments of the resulting distribution depend on explanatory variables relevant for the specific energy source, location and forecasting horizon. In contrast with deterministic forecasts, probabilistic forecasts enable to identify an optimal bid to hedge against penalties incurred when delivered production is distinct from the proposed bid. Bidding optimal quantiles of the forecast distribution has been applied to renewable participation in energy markets [30], and finds its theoretical ground in the portfolio theory developed by [31]. Extensive reviews [32], [33] show that the probabilistic forecasting of power production for Wind and PV (to a lesser extent) is a mature field of research. However, forecasts implemented in operational context have still significant error levels, with average absolute deviations comprised between 5 and 10 %. The penalties associated to these forecast errors have a significant impact on the profitability of renewable power plants [34], [35], [36].

Production is not the only source of uncertainty: before bidding on AS markets, a VRE aggregator is not aware of prices for reserve and its probability of activation. This echoes the more classical situation of bidding renewable production on the wholesale short-term energy market: the aggregator must adopt strategies that hedge against uncertainties on production and prices, by adapting their bidding volume on the various markets open before delivery (day-ahead, intraday...) in order to minimize the economical loss [37]. The uncertainties in balancing AS markets are expected to differ from uncertainties on the energy markets. Balancing AS markets have many characteristics that increase the complexity of the bidding strategy:

- bids must be highly reliable, because reserve is crucial for power systems,

- a reserve bid must be continuously available over the delivery period, which can last several hours,

- reserve induces production curtailment and therefore limits volumes allocated to the wholesale energy market, so potentially creating opportunity costs when prices for energy are more favourable than those of balancing AS,

- reserve activation is an uncertain variable which has a decisive impact on the revenue of producers.

Apart from the opportunity costs associated to the curtailment of production, one could add also that providing power regulation generates additional specific costs. The extra capital expenditure to upgrade control capacity and monitoring for balancing AS is estimated at less than 1% of capital costs by [38]. Concerning the operational costs associated to reserve provision, the frequent

regulations associated to power control are likely to induce supplementary costs in weartear, especially for wind turbines. However there is limited experience from the industry on this topic, more research is needed to integrate such costs in the global economical assessment of reserve provision by renewables.

> As a whole, the **overarching motivation of this thesis** is to facilitate the decisions taken by actors of VRE-penetrated grids who need to **optimize the value of VRE production, including the possible provision of ancillary services**. They must consider the numerous possibilities of aggregation based on several energy sources (solar, wind, hydro) and the entire range of markets, including the new prospect of ancillary service participation.

This motivation concerns obviously owners of renewable power plants who aim at maximizing their individual profit, but also operators of grids who need to accommodate the variable production at the best possible social welfare. In more details, two main challenges arise.

First, as aggregation appears necessary to reduce VRE production variability and provide balancing AS, forecasting methods are required:

Challenge 1: Quantify the uncertainty in aggregated VRE production.

This challenge can take a variety of forms. Several approaches of aggregated forecasting have been identified and are illustrated in Figure 1.11. A forecaster disposing of detailed information (weather forecasts, production data) on all plants within an aggregated portfolio may tackle the following problems:

1. **Forecast the total production of a VRE aggregation**, as the balancing bid will take place at the aggregated level operated by the VPP. This forecast must include a reliable quantification of uncertainty. An inaccurate forecast leading to under-fulfilments in reserve provision could jeopardize the participation of an aggregator to balancing AS markets. This forecast can be done directly on the total production (Figure 1.11, top right) or indirectly at the site level (Figure 1.11, top left), then aggregated taking into account correlations between sites. A third approach consists in reconciling the direct forecast of aggregation with forecasts at site level (Figure 1.11, bottom left), which is less useful in the present context as forecasts at sub-levels have no direct utility for AS provision.

2. **Increase the reliability of forecasts,** in case a TSO expects a maximum reliability, e.g. superior or equal to 99.9 %. This calls for specific models on extremely low quantiles of the forecast distribution (Figure 1.11, bottom right).

3. **Generate scenarios of VRE aggregated production**, in order to reproduce realistically temporal correlations in the aggregated production signal (Figure 1.11, bottom middle).

Second, balancing AS are procured in markets which bear their own uncertainties and impact the overall bidding strategy of VRE producers. The operator of a VRE-based VPP is therefore challenged to update his bidding strategies:

Challenge 2: Define a bidding strategy integrating balancing AS markets, considering uncertainties in aggregated production and market conditions.

1. **Ensure reserve capacity over the whole delivery period.** As reserve must be guaranteed up to several hours, a bidding strategy must integrate the expected variability of aggregated production during the delivery period, regardless of the status of reserve activation.

2. **Consider interactions between uncertainties on production and market conditions.** In a network with increasing renewable capacity, it is likely that the production pattern a VRE-VPP will have some degree of correlation with the total renewable production on the market. As renewables impact markets and will probably penetrate shortly balancing markets, the bidding strategy should take into account the impact of expected VPP production on balancing prices.
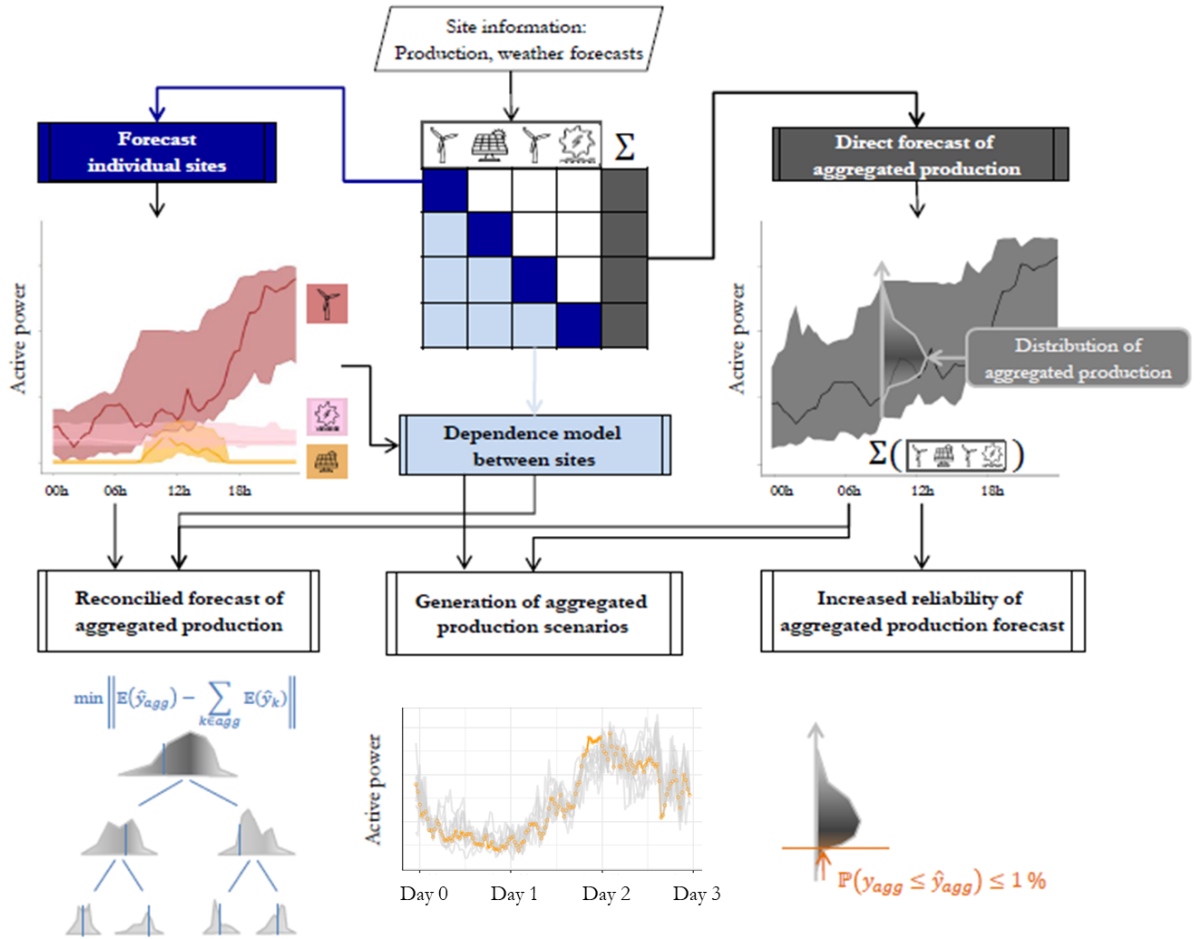
Figure 1.11: Identified challenges in quantifying the uncertainty of an aggregated VRE production

1.3 State of the art

This section investigates the most salient works tackling the afore-mentioned challenges on VRE aggregated forecasting and bidding integrating balancing AS markets. The aim of this analysis is to identify **research gaps**, which will be addressed by the contributions of this thesis.

After this introductory analysis, additional information on the state of the art relative to each contribution will be provided throughout the different Chapters.

1.3.1 Forecasting aggregated multi-source renewable production

Precise production forecasts taking into account the production uncertainty are a necessary input in reserve offering strategies. The proposals of European TSOs for balancing AS markets define a first gate closure time placed in the morning or the afternoon of the day before delivery [39], meaning that the horizon (or look-ahead time) of forecasts ranges typically from 12h to 48h. In this context, the forecast of Wind and PV production relies on Numerical Weather Predictions (NWP) [33]. An exhaustive review of the state of the art on existing forecasting methods for VRE production is given in [32]. Even if many authors propose models for either PV or Wind production, there is limited existing work on the probabilistic forecasting of a multi-source aggregated production (e.g. Wind and PV plants).

Regression models for VRE production. The probabilistic forecasting of single VRE power plants is an established field of research. Reviews are regularly published on the probabilistic forecasting of VRE production of various energy sources, in particular Wind ([40]) and PV ([41]) and some works give a more general view of the probabilistic forecasting of VRE production ([32], [42]).

A first candidate is the Quantile Regression Forests (QRF) model, which is a well-established probabilistic model for wind and PV power forecasting, and figures among the best performing models for both wind and PV forecasting [43]. Almeida et al. found that QRF was suitable for the prediction of PV power [44]. They also proposed to filter the training data by days showing an empirical distribution similar to the distribution of the NWP on the day to forecast, which improves performance. The gradient boosting algorithm, based on fitting iteratively shallow decision trees, is applicable to the quantile regression problem for both wind [45], [46] and PV [47]. Models based on decision trees are known to show better performance than linear models for day-ahead horizons, where the non-linearities between weather conditions and production levels are strong. An appealing answer to regression in large dimension is to deploy neural networks of several layers to approximate non-linearly the relationship between explanatory variables and the response variable of interest. Neural network models, with possibly deep architectures (large number of hidden layers) show generally better performance than standard machine learning models, but do not rank systemically first on the existing VRE forecasting competitions such as [43]. Neural networks are appealing because they can be seen theoretically as general learners, they have been frequently used in the forecasting community (from early works in [48] to more recent such as [49], [50]). The drawbacks associated with neural networks are long training times, recourse to calibration techniques to avoid exploding/vanishing gradients (e.g. batchnorm) or overfitting (e.g. dropout)

with little theoretical background, and difficult interpretation of the model [51]. Among possible neural network architectures for VRE production forecasting, we can list three architectures available in the state of the art of neural networks that can fit our context based on physical intuition or statistical observations: the standard Fully-Connected Neural Network (FCNN) or Multi-Layer Perceptron, which is a general solver but sensible to correlation in inputs and limited in depth due to its high number of connections between nodes (application to PV forecasting in [52]; the Convolutional Neural Network inspired by visual recognition (applications in [50] and [49] for Wind); recurrent neural networks modelling temporal dependencies between production sequences ([53] and [54] for Wind). It appears from the literature that quantile regression is seldom employed on neural networks to produce a probabilistic forecast of VRE production. The model of [53] issues a probabilistic forecast, but by learning the parameters of a mixture distribution instead of optimizing a quantile loss. In summary, several models existing for regression on VRE production, but they focus mostly on a unique source of energy.

Probabilistic forecast of aggregated VRE production. In contrast, the probabilistic forecast of aggregated solar or wind production, especially at the scale of several plants owned by a renewable producer, is an emerging topic in the literature. Bottom-up approaches, based on the fitting of copulas, propose a probabilistic forecast of the aggregated production of PV and Wind power plants [55]. Copulas are sampled to approximate the distribution of aggregated production. Alternatively to sampling methods, reconciliation methods such as in [56] build hierarchical partial sums of independent forecasts issued at the plant level to obtain an aggregated forecast that is coherent with a direct forecast of the aggregated production. The hierarchical model of [46] builds individual deterministic models at the level of wind turbines and obtains an aggregated deterministic forecast of the wind farm production with a penalized regression. A second stage consists in applying quantile regression to derive a probabilistic forecast of the aggregated production. The idea could potentially be rescaled to hierarchies of larger scales, i.e. aggregating wind farms located in different horizons. Spatio-temporal statistical models are known to improve the performance of probabilistic forecasts by exploiting the information available over a region where multiple VRE plants are located (see among many [57] for Wind, [58] for PV). Few spatio-temporal model focus on the prediction of an aggregated production, with the exception of [59], . A regional aggregation of VRE plants may be candidate for balancing As provision if it exploits favourable synergies between different energy sources, however the total aggregated variability is likely to remain higher than an aggregation of plants located in distinct climate zones. Therefore spatio-temporal models are of secondary importance in the context of balancing AS provision. To the author's knowledge, little has been published on the aggregated probabilistic forecasting of both wind and PV plants.

Models mentioned above are defined by parameters or hyper-parameters, which can be inferred by heuristic methods. As an alternative, Bayesian inference offers a framework to optimize the choice of these parameters: this can be achieved for linear regression models [60] as well as deep networks [61].

Reconciled hierarchical forecasts. If it is important for an aggregator to obtain coherent forecasts over all levels of the hierarchy defined in its aggregation, then optimization models can minimize the distance between forecasts at plant level and at aggregated level. The bottom-up ap-

proach of [62] introduces a Game-Theoretically Optimal reconciliation, where an optimal reconciled forecast at all hierarchical levels is obtained by minimizing a quadratic distance with the inconsistent forecasts, following a min-max zero-sum game constrained by the available information (e.g. prediction intervals). An advantage of reconcilied forecasting over direct aggregated forecasting is that the additional forecasts at lower hierarchical may improve the forecasting result [62]. The model ensures by construction that forecasts at bottom level do not harm the overall performance. The reconciliation framework has been applied to the context of aggregated wind production by [56], who introduce a constraint of privacy between individual producers participating to a common aggregation. The work of [63] extends the method to probabilistic forecasts: a sequence of permutations of the base density forecasts approximates the joint dependencies, and the resulting empirical copula is integrated to produce consistent aggregated density forecasts. While this approach is effective on aggregations which follow naturally a tree structure such as electricity networks, its 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.

Spatio-temporal models. Spatio-temporal statistical methods can be employed to generate distributions that take into account the spatial and temporal correlations among sites and instants within a delivery period of a given AS product. The boundaries of spatio-temporal distributions can then be used in robust optimisation frameworks. The latent Gaussian field approach proposed by [59] models spatio-temporal dependencies in aggregated wind production at a national level, yielding calibrated forecasts in both space and time. The spatio-temporal can also be thought in a nested architecture: the temporal feature extraction by LSTMs in [54] is encapsulated in a convolutional graph, which is dedicated to spatial dependencies.

Generate trajectories of aggregated VRE production. An important request in balancing AS markets is that capacities must be maintained during successive lead-times over the whole validity period of the balancing product. A validated solution for optimizing such bids with temporal constraints and uncertain variables is to model the problem as a stochastic optimization. This type of optimization requires scenarios, called also trajectories, that discretize the distribution of the uncertain variable and reproduce the temporal variability of the uncertain variable.

It is known that the quality of VRE production scenarios has a direct impact on the performance of optimization models under renewable uncertainty [64]. In the case of short-term horizons, scenarios should combine two properties: reproduce the interdependence 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 [65], and the possible correlation between power plants.

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 [65], or production of several PV plants at different lead times [66]). 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 [65]. Vine copulas, which form flexible trees of bivariate copulas, have been successfully used

for the probabilistic forecast of multiple Wind [29] and PV [67] 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 VRE 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 [63] and insurance exposure [68]. A specific challenge for scenarios of a multi-source VRE aggregation is to model the dependencies between energy sources, for simultaneous and successive lead-times, while preserving the conditional response to explanatory variables.

Alternative methods to generate scenarios exist. Time series analysis can derive spatio-temporal models for renewable power plants at multiple sites [69]. The machine learning approach of [70] 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, [12] develop a multi-stage scenario tree based on the distribution of wind forecasting errors. In [71], 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 [72] 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 [72] 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.

Research gaps: In summary, the following research gaps can be identified with respect to the probabilistic forecasting of aggregated VRE production:

1. (RG 1): There is a lack of direct probabilistic forecasting models for an aggregated VRE production composed of multiple energy sources, scalable to VPPs containing up to 100 individual plants.

2. (RG 2): Similarly, it is unclear how one should generate scenarios of aggregated VRE production with multiple energy sources, conditioned by short-term weather forecasts.

1.3.2 Forecasting highly reliable volumes of renewable production

Even if aggregating distant VRE plants reduces the variance of the total production, uncertainty remains significant. For balancing provision, the challenge is therefore to increase the reliability of the production forecasts on aggregation sizes coherent with the typical capacity of BSP portfolios implicated in balancing, reaching easily hundreds of MW. Relatively few publications have tackled the specific objective of highly reliable forecasting in the context of balancing AS. Existing solutions proposed in recent published research works can be classified as follows:

1. **Bottom-up dependence models.** After production has been forecast at plant level, a survival copula extracts the highest joint error on the aggregated plants to derive a reliable reserve capacity [73]. Other hierarchical models evaluate the risk of an aggregated portfolio: [68] models dependence through hierarchical sums of copulas. The work of [73] on reliable coalitions providing AS assumes that the deterministic production forecast of PV and Wind plants can be modelled as an homoskedastic Normal distribution. This assumption is an important limitation, because Normal errors fail to represent correctly the global uncertainty of VRE forecasts errors, especially the fat tails generally observed on low production levels [65]. Furthermore a Gaussian copula is chosen for the survival copula as standard practice, whereas non-Gaussian dependencies may be more adapted to describe non-linear relationships. The hierarchical copula model of [68] deals with financial risk aggregation. It identifies the structure of the hierarchical tree by rank correlation and bivariate copulas are selected based on tests for independence, symmetry, extremeness and are fitted by maximum likelihood.

2. **Distributional regression.** Instead of approximating the quantile function by minimizing a quantile loss in a quantile regression framework, the problem is formulated as a discrete approximation of the (cumulative) distribution function. A collection of binary loss is minimized from the empirical distribution function, over a finite grid of response values, conditionally on explanatory variables [19]. References indicate that distributional regression may have better convergence properties than quantile regression [74]. This has been applied for forecasting a reserve capacity in an industrial project [19] but no forecasting performance has been reported.

3. **Kalman Filtering.** Another family of models for tails relies on Kalman Filter (KF) or its non-linear versions (Unscented KF, Particle Filter). The conditional bias-penalized KF of [75] extends the classical KF formulation by penalizing the conditional bias of the state estimate in the objective function to be minimized. Adjusting the scale of this bias enables to enhance the reliability of the KF prediction on lower and upper tails of the distribution. The drawback of KF-based solutions is that they rely on a dynamical model, which can be challenging for problems in large dimension such as the production of a VPP.

4. **Simple parametric distribution for extremes.** A relatively direct parametric approach to model the tail of of a distribution consists in the interpolation of the Cumulative Distribution Function (CDF) of the response with an exponential functions. The thickness parameter of the function is tuned by weighting observations following clusters of forecast values. Having verified that quantile regression is reliable down to an empirically chosen percentile threshold (for instance 5 %), one can censor the distribution for quantiles below this threshold and replace it by a parametric distribution. A few publications have proposed empirical parametric solutions to model tails: [76] models the tails of wind power production with a parametric exponential function, where the coefficient is tuned by clusters of median production forecast. The rationale is that the shape of the tail is dependent on the wind production regime, and can be interpolated between 0 and the quantile forecast at the upper bound of the tail. An extension of this methodology is proposed by [77] in the context of Dynamic Line Rating: a

clustering of observations enables to fit exponential functions for the tails conditionally to the expected conditions. Exponential functions are also proposed to model extremal quantiles of a probabilistic day-ahead electricity price forecast in [78]. Rates of the exponential functions are inferred by maximum likelihood estimation on a rolling window without forgetting. The principal weakness of this approach is its lack of theoretical background, in particular its convergence properties.

5. **Advanced parametric distribution with Extreme Value Theory.** The Extreme Value Theory (EVT) proposes a more robust framework for the prediction of extremes. A predictive model is built on a subset of all available observations, retaining only a subset of extreme observations for defining the model, for instead blocks over a threshold. The EVT shows that the distribution of extreme observations can be modelled by a Pareto distribution [79]. Parameters of this distribution are inferred by methods based on maximum likelihood, statistical moments or bayesian approach. Few publications have dealt with EVT in the context of renewable production forecasting. One exception is the study of [80] on extremal quantiles of VRE production for the evaluation of transfer capacity under large renewable penetration. Similar reliability is found with EVT and exponential functions. In the context of short-term electricity price forecasting, a conditional EVT model is built on the identically and independent distributed (i.i.d) residuals of an auto-regressive forecasting model. The Peak-Over-Threshold (POT) retains the highest residuals over a threshold defined by maximum likelihood estimation. The reliability on quantiles above 99% is significantly improved compared to the auto-regressive model.

6. **Bayesian Inference.** Bayesian inference is known to be valuable when data is scarce [81] and when the response is modelled by a mixture of distributions [82]. Given that parametric statistical models are interesting candidates for forecasting distribution tails (as seen in EVT), bayesian inference is a promising framework for works on extremes, already explored in the financial sector (eg. [83]). Bayesian inference can capture rare behaviors by defining appropriate parametric distributions for the uncertainty on the response variable and valid assumptions on the possible values for the corresponding parameters. In [83], a Vector Auto-Regressive (VAR) model with Student distributed shock forecasts extreme economic events. It is shown that the Student distribution captures better rare events than a conventional Gaussian distribution. The Student distribution can also be skewed, as in [84]. The covariate-dependent copula model of [85] parametrizes the dependences between covariates to obtain an heteroskedastic conditional estimation of volatility. Both models yield a better tail forecasting score (here 1%-Value at Risk) than benchmark solutions. These publications are all based on a Bayesian framework, where the endogenous variables are assumed to follow parametric distributions, the parameters of which are in turn considered uncertain and generally estimated by Monte Carlo Markov Chain sampling. The Bayesian approach is appealing in the context of extreme forecasting because it integrates all the information available about endogenous and explanatory variables and works even with a limited sample size. The Bayesian inference approach can be applied easily to linear regression [81]. Knowing

that renewable power generation is non-linear with respect to weather conditions and past production levels, models able to capture this non-linear behaviour and project it on a linear space are natural candidates for the Bayesian regression of aggregated renewable production. The Extreme Learning Machine (ELM) is adapted to the non-linear regression of renewable production [86]. A Bayesian ELM is formulated by [87]. In this work, the uncertainty in the response variable is assumed to follow a Gaussian distribution, which is known to be inexact for renewable production [65]. For better statistical behaviour, an option is to infer a mixture of Gaussian distributions, relieving the strong assumption of gaussianity of the modelled process. The Bayesian inference of a Gaussian mixture model (GMM), enables to identify latent binary variables associating observations with components of the mixture, with more stability than frequentist approaches such as the Expectation-Maximization algorithm [82]. Still, working with Gaussian distribution is best suited for unbounded processes, which is not the case of renewable electricity generation. Conversely, the model adopted by [88] proposes a Bayesian regression framework which works directly on a bounded space. The specific mixture of Beta distributions ensures that the likelihood is tractable and is able to model a variety of tail behaviour.

Research gaps. Finally, the following research gaps appear on the forecasting of aggregated VRE production with maximum reliability:

- (RG 3): Quantile regression based models, e.g. based on neural networks, have not been adapted to the forecast of extremes in aggregated VRE production.

- (RG 4): The methodology to fit parametric models such as EVT or density mixtures should be optimized for the context of aggregated VRE production.

1.3.3 Optimization of ancillary service provision

The optimization of AS provision from flexible assets is an active field of research. The existing literature on this topic can be mapped in the following categories:

- **Optimal quantile.** Inspired by the literature on portfolio theory and its application to bidding on wholesale electricity markets [30], an optimal quantile of the expected production derived from probabilistic forecasts is dedicated to reserve [89], as a function of prices observed on markets for reserve and energy. This translates into allocating a share of the available power to reserve, and is proven to be operationally efficient when applied on wind-based reserve provision [90]. The existing works on optimal quantiles for reserve miss an important feature in the context of decision-making under uncertainty: they do not propose a methodology to forecast the various prices involved in the reserve bidding decision. Another limit, intrinsic to the optimal quantile approach, is that it can not integrate specific constraints (eg resulting from control capabilities or from market conditions).

- **Stochastic optimization.** The joint bidding of energy and reserve is formulated as a bilinear stochastic optimization in [89]. A Mc Cormick relaxation scheme is introduced to linearize the

bilinear constraints formed by the parallel participation on the two distinct markets of energy and reserve, when it is assumed that there is a constant ratio between energy and reserve bids (note however that transmission operators do not require such a ratio between bids of energy and reserve). This formulation is more flexible than the optimal quantile approach, because it can incorporate directly temporal constraints such as maintaining the reserve bid during the validity period of the offer.

- **Chance-constrained optimization.** Instead of classically balancing between average profit and worst-case losses or formulating a robust optimization based on extreme scenarios, a chance-constrained optimization can control the probability of undesired technical constraints imposed by fluctuating production. The reserve bidding model of [91] applies such a chance constraint to maintain constant the expected Loss of Load Probability (LOLP) of a micro-grid integrating flexible assets.

- **Robust optimization.** An alternative to stochastic optimizations consists in formulating a robust optimization problem, where the uncertainties are modelled by their boundaries instead of being realistically sampled with scenarios. Theses boundaries enable to hedge against worst-case realizations of production or other uncertain conditions such as prices. In [92], an adaptive robust optimization is proposed to optimize the cost of energy purchase of a residential aggregator of prosumers integrating batteries and PV production. The robust formulation enables to hedge against uncertainties in demand, PV production and energy prices. The limitation of robust methods is that they produce conservative bids, with sub-optimal revenues not extracting the full value of flexibility in a balancing AS market.

- **Linear Decision Rules.** Solving stochastic and chance-constrained optimizations requires generating scenarios and generally reducing them so that the resulting problem remains tractable [64]. A simpler alternative consists in approximating the bounds of the uncertainty set with piece-wise Linear Decision Rules (LDR). This robust estimation of bounds conserves most of the temporal information if enough pieces are used [89], and can serve as a base of a real-time management of power delivery taking into account market penalties [93].

- **Optimisation of reserve dispatch to aggregated resources in real-time.** Once reserve bids have been accepted by system operators, aggregated flexible resources must be managed in real-time to optimize the dispatch of activated reserve bids during the whole validity period. The optimisation of dispatch creates technical challenges and economical challenges. Technical challenges arise from the variability of flexible resources and their combined dynamic responses to the reserve setpoint. A simple approach to optimize reserve dispatch of aggregated resources consists in a merit-order activation. The merit-order can be based on economical criteria or technical criteria such as the quality of response by a power plant to prior activation [26]. However, methods of higher complexity can be suited for this task, for instance Model Predictive Control (MPC). In [94], observed activation reserve signals create a balancing energy constraint in the rolling optimal schedule which preserves an FCR bid offered by an aggregation of flexible consumers. The drawback of MPC methods is that they

require an accurate state-space model of the aggregation, which can be challenging in large-scale and diverse aggregation such as a VRE-based VPP with multiple energy sources. While the optimisation of reserve dispatch to aggregated resources is an interesting subject, it will not be covered in this thesis, which focuses on the decision-making relative to bidding reserve and energy ahead of delivery.

Research gaps. To summarize, the existing work in the bidding optimization of balancing AS show several gaps:

1. (RG 5): Uncertainties in balancing AS prices and their interaction with VRE production are not modelled.

2. (RG 6): The probability that a VRE-based VPP is activated by the TSO to provide reserve is not included, although it is a decisive parameter for bidding on aFRR, mFRR or RR markets.

3. (RG 7): A comparison of methods optimizing both revenues and risk of reserve under-fulfilment is missing.

4. (RG 8): The profitability of the bidding of multiple balancing AS products by a VRE-VPP has not been investigated.

1.4 Research Questions and Contributions

After identification of the research gaps emerging from the analysis of the available state of knowledge, the thesis is articulated on a series of research questions, stemming from two main challenges, as represented in Figure 1.12. Four research questions concern Challenge 1, i.e. forecasting the aggregated VRE production, while three research questions are connected to Challenge 2, i.e. the development of bidding strategies integrating balancing AS. Finally, research questions address all the research gaps identified in the analysis of the state of the art.

The 7 research questions are:

- RQ1: How to forecast an aggregated multi-source VRE production, with the best global probabilistic forecasting performance?

- RQ2: What is the optimal methodology to generate scenarios of aggregated multi-source VRE production, reproducing accurately the temporal variability in production?

- RQ3: What is the influence of the mix of energy sources (Wind, PV, Hydro) on the variability of the VPP production and on the performance of its forecast?

- RQ4: How to develop a quantile regression model for very low quantiles of aggregated VRE production, with reliability > 99 % ?

- RQ5: How to integrate market uncertainty into reserve bidding strategies?
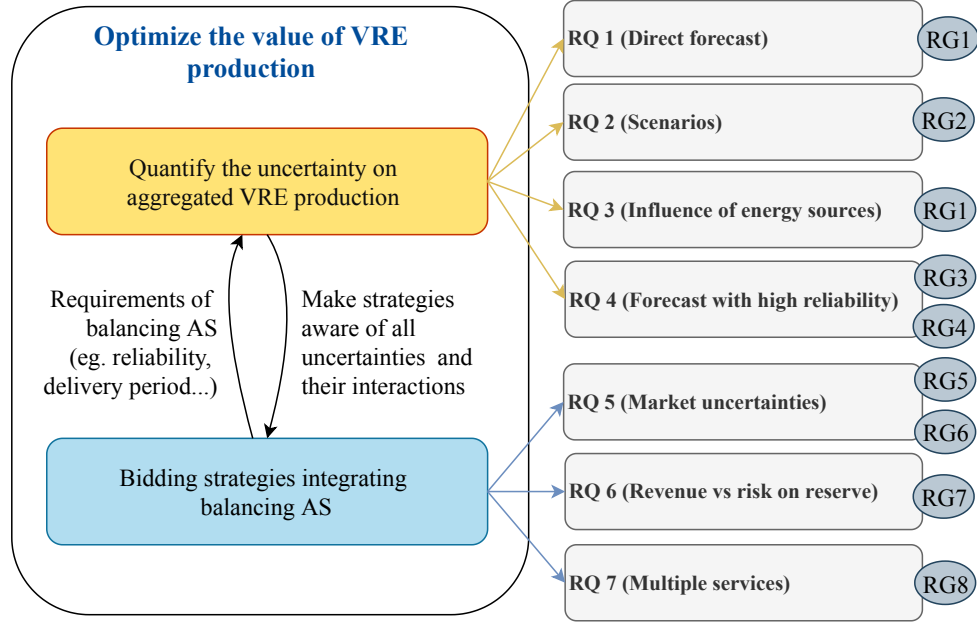
Figure 1.12: Research Questions of the thesis

- RQ6: Which type of bidding strategy provides best results in terms of maximized revenue and minimized risk of reserve uder-fulfilment?

- RQ7: How to optimize a bidding strategy integrating energy and multiple services?

The contributions of this work are:

C.1. The entire value chain of bidding a VRE-based balancing AS is covered, from the available data on VRE plants to the market bid: the production is forecasted with uncertainty, and this forecast is integrated into bidding strategies. Forecasts are developed in order to be useful for bidding, and bidding strategies are adapted to integrate the specific case of aggregated VRE forecasts with uncertainty.

C.2. Tailored probabilistic forecasting models for multi-source VRE aggregated production.

C.3. A comparative analysis of methods to generate trajectories of aggregated VRE production.

C.4. A set of forecasting models for the extremely low quantiles of aggregated VRE production.

C.5. An integration of probabilistic forecasts of production and price in bidding strategies for energy and reserve

C.6. A study of the influence of trajectories on the revenue of VRE producers offering energy and reserve

C.7. A simple methodology for bidding multiple ancillary services and energy

1.5 Structure of the thesis

The first chapter of this thesis has introduced the subject of balancing provision by a VRE-based VPP. In the following chapters, answers to the research questions listed above will be presented. A short summary in French ends every chapter. An overview of the structure of the thesis is given by Figure 1.13.
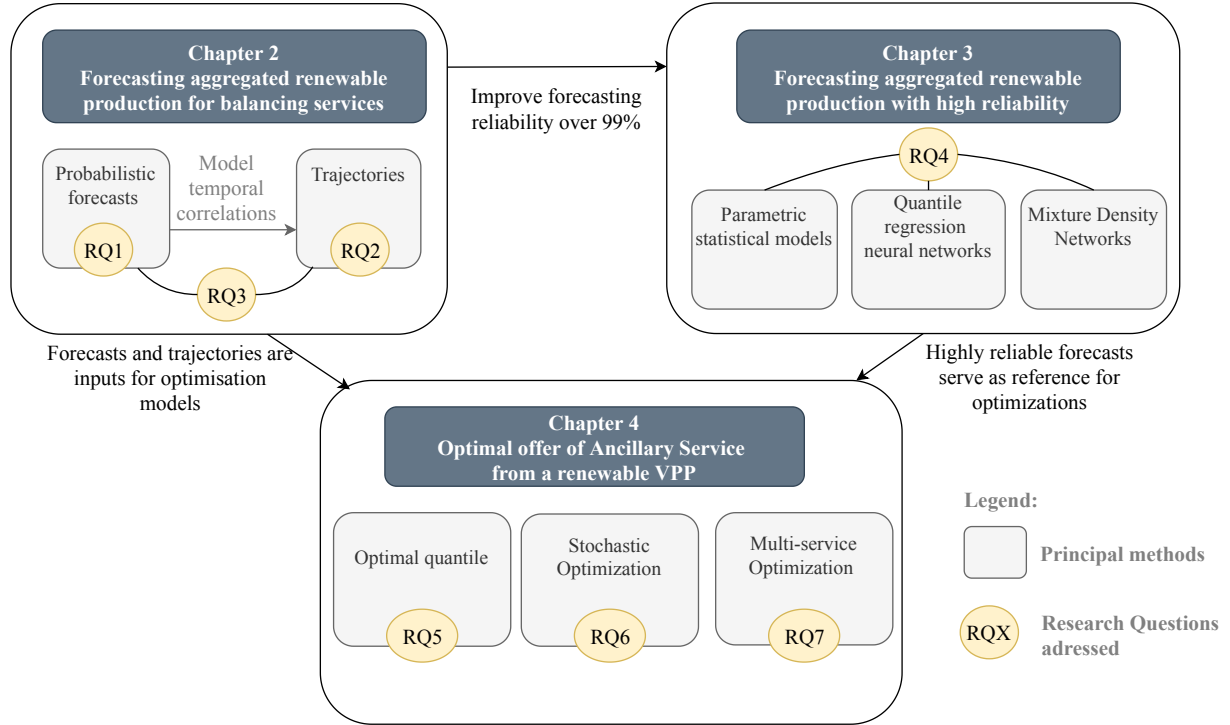
Figure 1.13: Overview of the structure of the thesis

The first proposition of this thesis is the development in Chapter 2 of methods for the probabilistic forecasting of an aggregated VRE production. After characterizing the uncertainty of aggregated production and analyzing suitable regression models, a direct approach is built for the forecasting of production at the aggregated level. Then the forecasting methodology is enhanced with a scenario generation model, which uses the direct forecasting model and considers alternatively separate forecast for each energy source of the aggregation.

The methodology developed in Chapter 2 is adapted to various VPP configurations and different shares of energy sources. However models are not tailored to the very high level of reliability needed to participate into balancing AS markets. In Chapter 3, several models are built to predict the most likely aggregated production levels, i.e. the lowest quantiles of the forecast distribution ($\leq 1\%$).

The operator of a VPP disposes now of detailed information about the expected production of its VPP and the associated uncertainty. In Chapter 4, solutions to optimize the bidding strategies of the operator in the balancing markets are proposed, taking into account the uncertainties on market conditions and allowing for minimal probabilities of reserve under-fulfillment.

1.6 List of publications

The present thesis led to the following publications:

Peer-reviewed journals:

1. **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, Nov. 2018, https://dx.doi.org/10.1109/TPWRS.2018.2847239, Postprint:https://hal-mines-paristech.archives-ouvertes.fr/hal-01816576

2. **S. Camal**, F. Teng, A. Michiorri, G. Kariniotakis, L. Badesa, "Scenario generation of aggregated Wind, Photovoltaics and small Hydro production for power systems applications", Applied Energy, Elsevier, 2019, Volume 242, pp.1396-1406. https://dx.doi.org/10.1016/j.apenergy.2019.03.112. Postprint:https://hal-mines-paristech.archives-ouvertes.fr/hal-02081282

3. **S. Camal**, A. Michiorri, G. Kariniotakis, "Highly Reliable Forecasts of aggregated multi-source Variable Renewable Energy Production", In preparation for IEEE Transactions on Smart Grids.

International Conferences:

1. **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", *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe* (ISGT-Europe), Torino, 2017, pp. 1-6. https://dx.doi.org/10.1109/ISGTEurope.2017.8260311, Open archive: https://hal-mines-paristech.archives-ouvertes.fr/hal-01615232

2. **S. Camal**, A. Michiorri, G. Kariniotakis. "Probabilistic forecasting and bidding strategy of ancillary services for aggregated renewable power plants". *6th International Conference Energy and Meteorology*, WEMC - World Energy and Meteorology Council, Jun 2019, Copenhaguen, Denmark. Open archive: https://hal-mines-paristech.archives-ouvertes.fr/hal-02177537

3. A. Michiorri, **S. Camal**, G. Kariniotakis. "Ancillary services from a renewable-sourced virtual power plant: the REstable project". *6th International Conference Energy and Meteorology*, WEMC - World Energy and Meteorology Council, Jun 2019, Copenhaguen, Denmark. Open archive: https://hal-mines-paristech.archives-ouvertes.fr/hal-02177556

4. **S. Camal**, A. Michiorri, G. Kariniotakis. "Forecasting Extremes of Aggregated Production from a RES Virtual Power Plant". *Proceedings of the Wind Energy Science Conference 2019*, EAWE - European Academy of Wind Energy, Jun 2019, Cork, Ireland. Open archive:https://hal-mines-paristech.archives-ouvertes.fr/hal-02158589

5. J. Strahlhoff, A. Liebelt, S. Siegl, **S. Camal**. "Development and Application of KPIs for the Evaluation of the Control Reserve Supply by a Cross-border Renewable Virtual Power Plant". *INFORMATIK 2019*. Bonn: Gesellschaft für Informatik e.V. (S. 517-530). https://dx.doi.org/10.18420/inf2019_69, Open archive: https://hal-mines-paristech.archives-ouvertes.fr/hal-02415649

26

6. **S. Camal**, A. Michiorri, and G. Kariniotakis. "Forecast trajectories for the production of a renewable virtual power plant able to provide ancillary services". *EGU General Assembly 2020*, Online, 4–8 May 2020, EGU2020-19081, https://doi.org/10.5194/egusphere-egu2020-19081

1.7 Résumé en Français

Contexte

La fiabilité actuelle de la fourniture d'électricité aux usagers des réseaux électriques est généralement élevée en Europe: à titre d'exemple, la durée moyenne de déconnexion en France sur la décade 2008-2018 est de 3 minutes par an [2]. Pour atteindre ce niveau de fiabilité, les gestionnaires de réseau adaptent en permanence les flux de puissance générées ou consommées en fonction des imprévus, tels qu'une indisponibilité imprévue d'une centrale, la perte de capacité de transport sur une ligne ou une augmentation de la consommation supérieure aux prévisions. Un déséquilibre de puissance active sur le réseau de transport se traduit immédiatement par une déviation de la fréquence au-dela de sa valeur nominale ($50Hz$ en Europe). Ces déviations de fréquence doivent être contenues puis supprimées afin d'éviter des dégâts majeurs aux réseaux électriques. Pour cela, les gestionnaires de réseau électrique font appel à des centrales pilotables (nucléaires, thermiques, hydraulique de barrage) capables de réguler une partie de leur puissance disponible avec précision. Des consommateurs flexibles peuvent également effectuer ce type de réglage. La fourniture de cette réserve de puissance est organisée au sein du mécanisme dit des *services système*, dont l'action coordonnée permet de retrouver la valeur nominale de fréquence à la suite d'un déséquilibre.

Le processus de contrôle de la fréquence est harmonisé en Europe et suit trois phases, représentées en Figure 1.14: la première phase consiste à mobiliser automatiquement en moins de 30 secondes, sur l'ensemble du réseau européen une réserve primaire servant à contenir la déviation de fréquence (*Frequency Containment Reserve, FCR*). La réserve secondaire (*automatic Frequency Restoration Reserve, aFRR*) prend le relai pour corriger progressivement, par envoi de consigne de réglage au niveau national, les écarts restants dans un délai de 15 minutes. Enfin la réserve tertiaire (*manual Frequency Restoration Reserve, mFRR* et *Replacement Reserve, mFRR*) permet de maintenir l'équilibre sur les 15 minutes suivantes, voire plus selon les pays, et de remplacer les centrales impliquées dans les réserves précédentes en prévision d'un futur déséquilibre.

Les renouvelables variables ne font pas partie des fournisseurs traditionnels de réserve du fait de l'incertitude liée à leur production. Pourtant ils disposent des capacités techniques de fournir de la réserve : les systèmes de contrôle des centrales variablees peuvent réguler leur puissance active à l'échelle de quelques secondes, en agissant sur l'angle de calage des pales et le couple du rotor des éoliennes [22], ou en déplaçant le point de fonctionnement du point de puissance maximum sur des centrales photovoltaïques (PV) [25]. Il est également possible de tenir compte de l'impact du sillage induit par les éoliennes adjacentes dans une ferme éolienne sur le productible de chaque éolienne, [23]. Ceci permet d'expérimenter la livraison de réserve à l'échelle d'une agrégation de centrales variablees contrôlées par une centrale virtuelle. La Figure 1.15 montre un test de fourniture de réserve par une centrale virtuelle éolien/PV/biomasse. On y voit que l'agrégation réagit à la consigne de réglage en rouge, qui suit des paliers à la hausse et à la baisse, ou est proportionnelle à
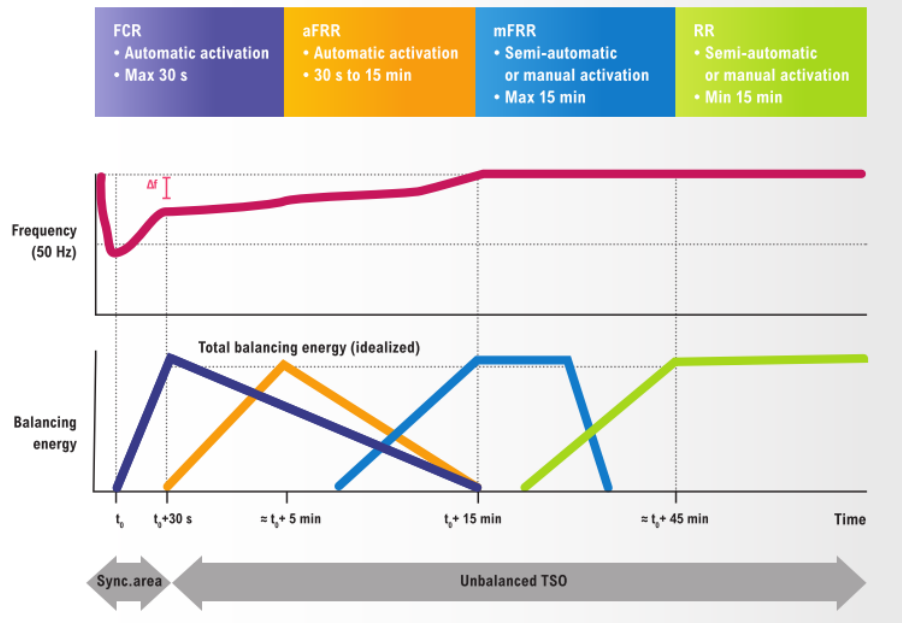
Figure 1.14: Schéma du contrôle de fréquence en Europe [8]

la variation de fréquence (intervalle de 300 à 600 secondes). La qualité de la fourniture de réserve peut être quantifiée en (2.1), après estimation du productible disponible (ou *Available Active Power, AAP*) $\hat{y}_{AAP}$ à partir des conditions météorologiques mesurées, et calcul de la réserve fournie comme étant l'écart entre le productible disponible et la puissance produite y. La réserve fournie dépend de l'activation à la hausse ou à la baisse $a_\uparrow$, $a_\downarrow$ requise par le réseau. L'estimation de la puissance disponible donne lieu à des erreurs, des méthodes de calibration avancées couplant statistiques et physique permettent de diminuer ces erreurs [24].

$$y(t) = \hat{y}_{AAP}(t) - ((1 - a_\uparrow(t))R_\uparrow + a_\downarrow(t)R_\downarrow)$$

Cette réserve donnant lieu à une réduction de l'énergie produite doit être valorisée à un niveau suffisant pour garantir l'équilibre économique de l'installation de production variablee. Les centrales variablees intègrent actuellement les marchés de l'électricité en Europe: au lieu de de bénéficier d'obligations d'achat à tarif constant, les producteurs renouvelables se rémunèrent directement sur les marchés, tout en disposent d'un soutien transitoire sous forme de complément de rémunération décidé par les régulateurs des marchés de l'énergie [96]. De la même façon, certains pays préparent la participation des renouvelables variables aux marchés de réserve (notamment au Danemark et en Espagne, voir [97]). Les opérateurs de centrales variablees seront donc appelés à formuler des offres d'énergie et de réserve en tenant compte de sources d'incertitude multiples:

- production agrégée,
- prix sur le marché de l'énergie,
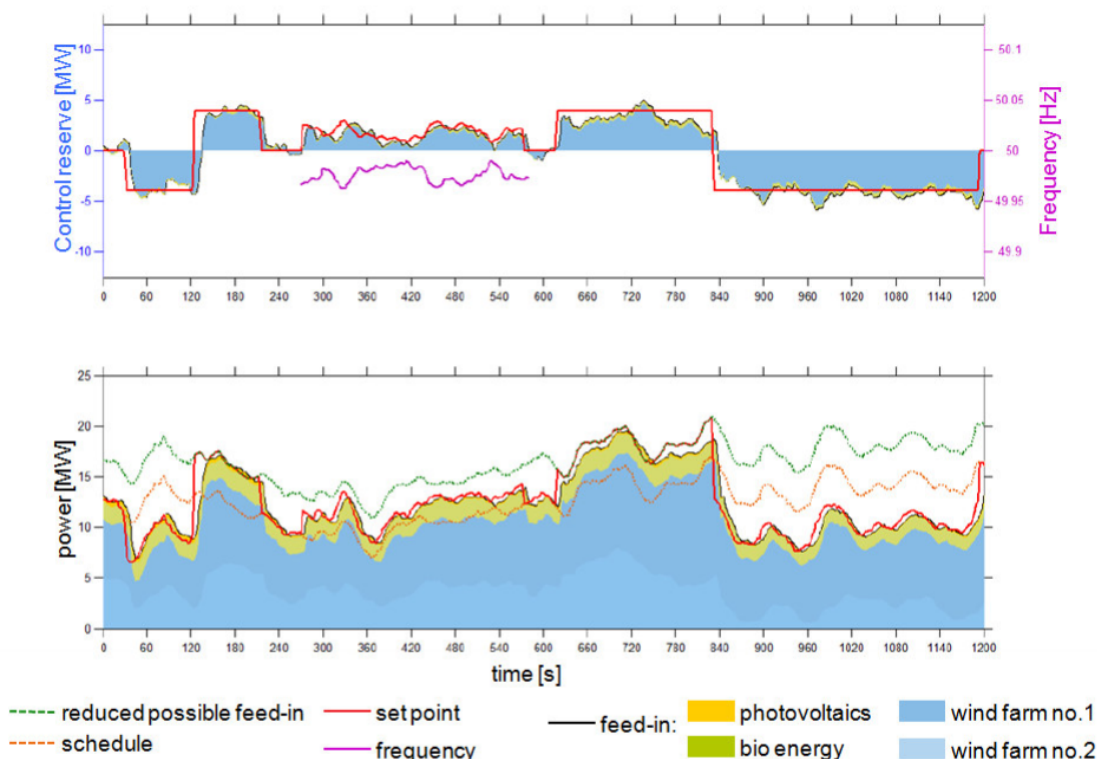
Figure 1.15: Fourniture de réserve par une centrale virtuelle éolien/PV/biomasse [95]

- prix sur les marchés de réserve,

- probabilité d'être activé pour fourniture de réserve,

- en cas de défaillance de fourniture de réserve, les pénalités imposées par le gestionnaire de réseau peuvent être très élevées, voire conduire à l'exclusion de la fourniture du service.

Enfin, les marchés de réserve sont en voie d'harmonisation en Europe, suite à la directive européenne sur l'équilibrage de l'électricité [98]. Il en ressort que la durée de validité de l'offre de réserve, qui peut aller aujourd'hui jusqu'à 1 semaine [18], sera portée progressivement à quelques heures puis à 15 minutes à horizon 2025. Les valeurs seront déterminées pour chaque type de réserve après concertation. Cette réduction de la durée de validité facilite naturellement l'intégration des renouvelable variables, l'offre de réserve pouvant être adaptée à la variabilité de production attendue. La disponibilité attendue de la réserve est toutefois théoriquement de 100% sur toute la durée de validité. Une disposition particulière proposée par les gestionnaires de réseau allemand précise que 99.73% des valeurs de réserve mesurées doivent être contenues dans une enveloppe de $\pm 10\%$ autour de la réserve attendue [99]. L'exigence de fiabilité de la réserve fournie est donc importante, tout en prenant en compte les déviations inhérentes aux incertitudes sur l'estimation de la puissance disponible.

Enjeux

A partir de cette vision du contexte, deux besoins majeurs sont identifiés:

1. **Prévoir la production agrégée d'une centrale virtuelle variablee**, afin de préparer les offres de réserve, et respecter les critères de disponibilité de la capacité de réserve posés par les gestionnaires de réseau.

2. **Définir une stratégie d'offre de services système**, qui puisse considérer les incertitudes de production et de prix. En effet, l'offre de réserve doit être assurée sur la totalité de la durée de validité, et doit se concentrer sur les périodes où un écart de prix favorable pour la réserve peut être anticipé.

Etat de l'art

L'avancée actuelle de la recherche sur la prévision de production agrégée variablee et sur les stratégies d'offre de réserve par les renouvelables a été analysée. Les points principaux identifiés concernant la prévision de production sont:

- Plusieurs modèles de prévision probabiliste de production variablee sont disponibles pour l'éolien et le PV. Parmi les modèles de prévision, la plupart effectuent des régressions quantiles. On identifie les modèles basés sur les arbres de décisions comme le Quantile Regression Forests (QRF) [44] et le Gradient Boosting Tree (GBT) [47]. Les réseaux de neurones, en particulier les réseaux convolutifs et les réseaux récurrents [53], déjà présents dans la discipline [48], font leur retour grâce à la démocratisation de leur implémentation. La prévision de production agrégée, en combinant des prévisions au niveau de chaque centrale à l'aide de copules, a été proposée dans le contexte de la production éolienne [29]. On remarque que peu d'auteurs ont traité de l'aggrégation variablee multi-source.

- L'offre de réserve doit être maintenue sur des intervalles de livraison durant lesquels la production variablee agrégée va varier. Si l'on souhaite modéliser les corrélations temporelles de la production dans un modèle d'opitmisation (par exemple une optimisation stochastique), il est nécessaire de générer des scénarios ou trajectoires de production agrégée. Des méthodes de génération de scénarios de production renouvelable proposent d'utiliser des prévisions probabilistes [100], voire de modéliser les corrélations entre centrales d'une même source d'énergie [66]. Le cas de scénarios de production agrégée multi-source semble avoir été peu traité.

- La prévision d'extrêmes de production variablee a fait l'objet de moins d'attention. Toutefois des modèles paramétriques, basés sur des distributions exponentielles ou sur la théorie des valeurs extrêmes [76] ont été proposés pour modéliser des extrêmes élevés d'incertitude variablee. L'approche par *distributional regression* approchant le problème de prévision des extrêmes comme une classification d'évènements dans la distribution possible des productions variablees a été proposée par [19] à partir des travaux de [74], sans toutefois détailler la qualité des résultats obtenus. Les réseaux récurrents de type Long Short Term Memory ont été utilisés pour la prévision d'extrêmes dans le contexte des transports [101]. Les approches de

type filtre de Kalman ont été rencontrées en économétrie [75], mais nécessitent une modélisation état-espace qui est difficile à implémenter pour une agrégation de plusieurs dizaines de centrales variablees. Enfin des méthodes basées sur des mélanges de densité, inférés par méthode bayésienne [88] peuvent modéliser des extrêmes avec flexibilité. En résumé, une gamme de solutions de prévision d'extrêmes existe, mais les extrêmes de production agrégée multi-source n'ont jusqu'ici pas été abordés.

Quant aux stratégies d'offre de réserve, les contributions intéressantes pour cette thèse sont les suivantes:

- Lorsque l'agrégateur gérant la centrale virtuelle souhaite offrir conjointement sur le marché de l'énergie et sur un marché de réserve spécifique (par exemple la réserve secondaire aFRR), il peut alors utiliser une méthode analytique dite du *quantile optimal*, qui affecte à la réserve un quantile de la production prévue. Cette méthode a été appliquée au marché de l'énergie [102] et à la fourniture de la réserve par une centrale éolienne [103]. Les travaux existants ne considèrent toutefois pas d'incertitude sur les prix.

- Des modèles d'optimisation stochastique permettent d'intégrer la variabilité de production variablee via des scénarios de production. Plusieurs formulations sont proposés par [89], sans toutefois traiter de la production agrégée. On peut également introduire une contrainte technique probabiliste dans le problème d'optimisation, afin de maîtriser le risque de défaillance, comme l'optimisation proposée par [91] pour le dimensionnement d'un micro-grid. L'optimisation robuste consiste enfin à remplacer les scénarios par des enveloppes modélisant les intervalles de variation des variables incertaines. Ceci permet de dimensionner le besoin de réserve d'un réseau comme dans [89], mais est difficilement applicable à l'offre jointe d'énergie et de réserve.

- Les modèles de type contrôle prédictifs permettent d'optimiser la fourniture de réserve en cours de journée. Ils sont bien adaptés aux problèmes intégrant des batteries, pour lesquels la gestion de charge doit être effectuée en fenêtre glissante en fonction des données techniques et économiques de la journée [94]. Ce type de méthode est moins utile lorsqu'il s'agit d'optimiser l'offre initiale, du jour pour le lendemain, d'énergie et de réserve.

Objectifs et contributions

L'objectif général de cette thèse est d'optimiser la capacité d'une centrale virtuelle variablee à fournir des services système de réglage en fréquence. Il s'agit en particulier de proposer des méthodologies d'offre de service intégrant les incertitudes sur les prix et sur la production agrégée.

Après étude des problématiques scientifiques et de l'état de l'art, les enjeux principaux relevés en Figure 1.16 orientent le travail de thèse. Le premier enjeux est de quantifier l'incertitude de la production variablee agrégée, et le second d' établir des stratégies d'offre sur les marchés intégrant la fournitre de services système d'équilibrage. Ces deux enjeux génèrent 7 questions de recherche:

RQ1: Comment développer une prévision probabiliste directe de la production agrégée multi-source?

RQ2: Quelle est la méthodologie optimale pour générer des scénarios ou trajectoires de production agrégée mulit-source ?

RQ3: QUelle est l'influence du mix des sources d'énergie sur la variabilité de la production agrégée et sur la performance de la prévision ?

RQ4: Comment développer des prévisions spécifiques pour les quantiles extrêmement faibles de production agrégéee variablee, avec une fiabilité > 99 % ?

RQ5: Comment intégrer les incertitudes de marché dans des stratégies d'ofre de réserve ?

RQ6: Quel type de stratégie d'offre dans les meilleurs résultats en termes de maximisation de revenus et de minimisation du risque de non-fourniture de réserve ?

RQ7: Comment optimiser une offre couplée d'énergie et de services système multiples ?
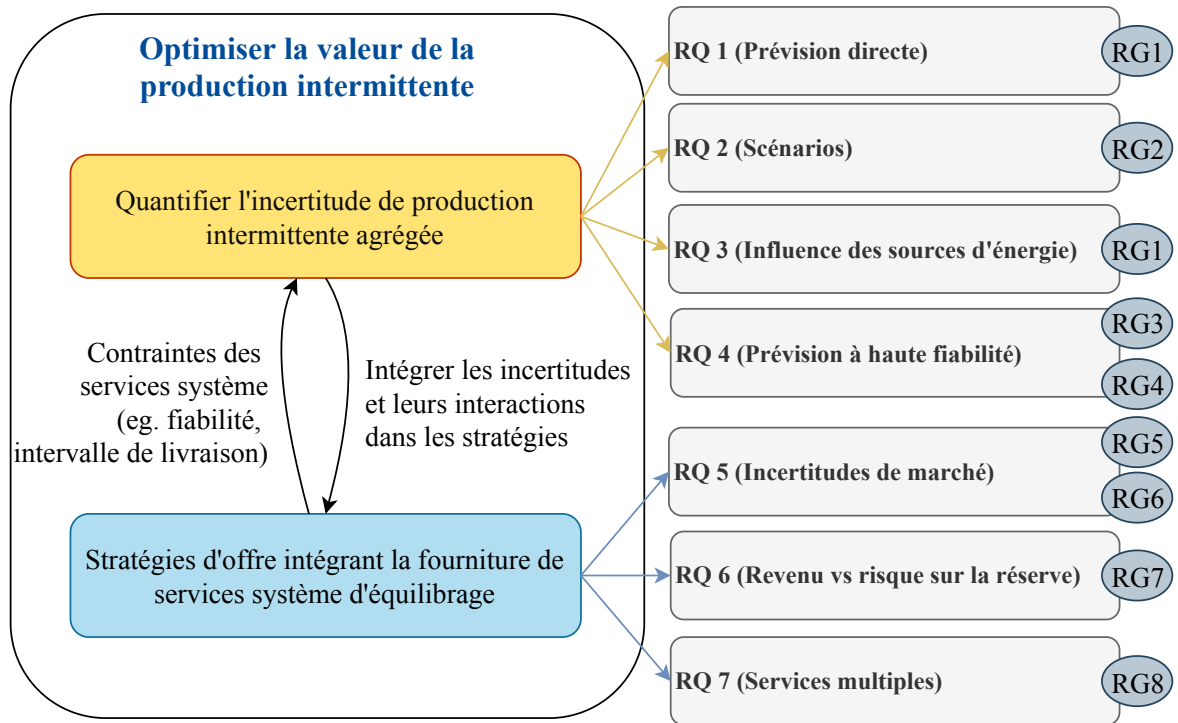
Figure 1.16: Organisation des questions de recherche abordées dans la thèse.

Structure de la thèse

Le premier Chapitre introduit le sujet de la fourniture de services système par une agrégation variablee. Les chapitres suivants répondent aux questions listées ci-dessus. La première proposition de cette thèse est une méthodologie de prévision de la production agrégée présentée en Chapitre 2. Après avoir caractérisé la variabilité de la production agrégée, plusieurs modèles de prévision probabiliste directe sont proposés, effectuant une régression quantile basée sur des arbres de décision

et des réseaux de neurones. Ensuite des scénarios de production agrégée sont générés afin de reproduire la variabilité temporelle de la production de la centrale virtuelle.

Ensuite, le Chapitre 3 présente une méthodologie de prévision fiable des quantiles bas de la production agrégée, afin d'obtenir une fiabilité supérieure à 99 %. Les modèles de régression quantile du chapitre précédent sont adaptés à la prévision des quantiles bas, et plusieurs approches paramétriques alternatives sont proposées. L'agrégateur de la centrale virtuelle dispose maintenant de prévisions de production fiables.

Le Chapitre 4 met à sa disposition plusieurs stratégies d'offre de réserve: une méthode du quantile optimal utilisant les prévisions probabilistes de production, une méthode d'optimisation stochastique sous contrainte probabiliste utilisant les scénarios de production agrégée, et enfin une méthode pour l'offre de services système multiple. Les conclusions générales sont présentées au Chapitre 5.

FR

Chapter 2

Forecasting of Aggregated Renewable Production for Reliable Balancing Services

2.1   Introduction

An aggregator operating a renewable Virtual Power Plant (VPP) needs a probabilistic forecasting model of the aggregated production to decide how much to bid on energy and reserve markets. Consider for illustration the density forecasts potentially issued by such a model in Figure 2.1 (specific models adapted to this task will be presented later in the Chapter). The VPP comprises multiple sources of variable energy: a first challenge of prediction models for VRE-based VPP production is therefore to account for the contributions of each energy source (PV, Wind, or run-of-river Hydro) in the aggregated production profile. On the three days represented in the Figure, the prediction intervals defined by the different quantiles of the forecasts appear to be in line with the production of the VPP, albeit errors of phase and magnitude can be readily observed before quantifying the forecasting error. Given that reserve offers from the VPP should be as reliable as possible, it is clear that an information of the uncertainty related to production is essential. Consequently, deterministic forecasts are not sufficient and probabilistic forecasts such as those depicted here at necessary to adapt reserve capacities as a function of this uncertainty.

At this point, a second challenge arises: how to optimize the precision of VPP production forecasts, that guarantees the effective availability of reserve offers and minimize the penalties on energy offers due to prediction errors? The dimension of this precision is essentially threefold:

- *reliability*: a reliable probabilistic forecast anticipates correctly the risk of underfulfilling the reserve bid,

- *sharpness*: a sharp forecasting model is able to minimize the range of prediction intervals conditioned by explanatory variables, therefore producing a concentrated information on the expected future behaviour.

- *realistic temporal correlations*: if the market imposes temporal constraints, or if bidding
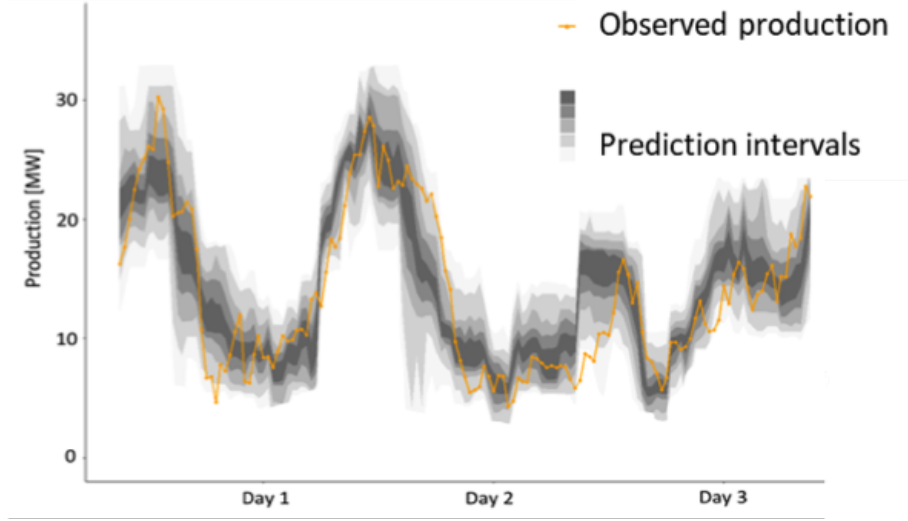
Figure 2.1: Illustration of a probabilistic density forecast of the production of a VRE-based VPP

strategies should be built on several consecutive horizons, then the forecasting model must reproduce accurately the temporal variability of the VPP.

The global **objective** of this Chapter is to identify the most precise probabilistic forecasting model for a multi-source VRE aggregated production.

In what follows, the context will be placed on day-ahead forecasting, i.e. from 12 to 48 hours ahead. Several challenges can be identified for the day-ahead forecasting model of a multi-source VRE aggregated production:

- devise the relative influence of individual energy sources on the aggregated behaviour,

- accommodate the deterministic trend of PV generation, for any share of PV capacity in the aggregation,

- reproduce the smoothing effect of aggregation,

- improve sharpness with respect to off-the-shelf implementations of available models, in order to avoid large prediction intervals, when information of expected weather on PV and Wind plants is associated with diverging levels of aggregated production in the past.

- detect invariant production patterns in the aggregation to improve the reliability of the forecast

- scale to VPPs containing a large number of plants

- generate trajectories with realistic temporal correlations.

36

2.2 Methodology

This Section describes briefly the methodology of the Chapter. The general workflow is illustrated in Figure 2.2. The research questions defined in Section 1.4 that are addressed by each method are explicitly underlined.

The aggregated production is first characterized in Section 2.3 to investigate how the proportions of the different energy sources and the level of aggregation influence the total uncertainty. This answers partially Research Question 3.

Then Section 2.4 proposes a direct probabilistic forecast of the aggregated production, using expected conditions on each site of the aggregation. This responds to Research Question 1. The reason for a direct forecast of the aggregated production is to provide insight to the decision-maker of a VRE-based VPP with a simple forecasting architecture.

If the decision-maker needs trajectories which represent correctly the temporal variability of aggregated production, then a probabilistic forecast is not sufficient. A scenario generation method is proposed in Section 2.5 to obtain trajectories that reproduce the aggregated behaviour, for various shares of energy sources in the aggregation. This answers Research Question 2.
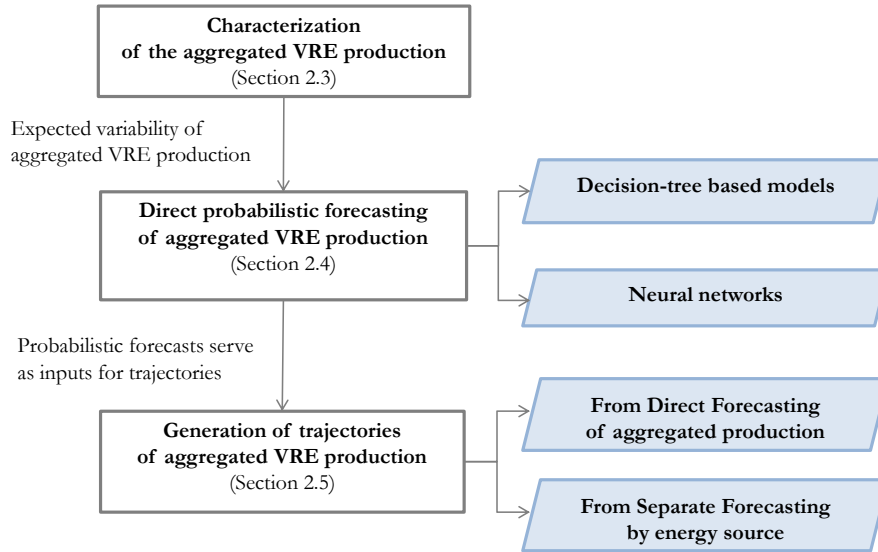
Figure 2.2: General workflow of Chapter

2.3 Characterization of the aggregated weather-dependent renewable production

Renewable aggregations bear a significant level of uncertainty in production that needs to be properly evaluated. If we consider the aggregated production as a random variable, its uncertainty can be separated in two distinct quantities: the *aleatoric* uncertainty, viewed as the intrinsic statistical

variation of the physical phenomenon, and the *epistemic* uncertainty, i.e. the uncertainty created when we approximate the phenomenon via an estimator [104]. Without loss of generality, if one observes the realizations of a random variable y over a period T, conditioned on the observation of explanatory variables x, then the variance of the conditional response $Var(p(y|x))$ writes:

$$Var(p(y|x)) = \underbrace{\frac{1}{T}\sum_{t=1}^{T} diag(y_t) - y_t.y_t^\intercal}_{\text{aleatoric uncertainty}} + \underbrace{\frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)((y_t - \hat{y}_t))^\intercal}_{\text{epistemic uncertainty}} \quad (2.1)$$

This section aims at quantifying the aleatoric uncertainty associated to an aggregation of weather-dependent renewable plants. The epistemic uncertainty will be treated in the next sections, where forecasting models will be proposed. The aleatoric uncertainty of the multi-source aggregated production can be analyzed spatially (how correlations between plants influence the total production) and temporally (auto-correlation of the VPP production, variability at different frequencies). The temporal variability of renewable production has a particular importance for the offer of AS, because variations at the day/hour scale impact the potential reserve volume, while sub-hourly variations impose an additional constraint on the reserve volume and challenge the effective power regulation to provide reserve.

More specifically, the following questions regarding the uncertainty of aggregated VRE production will be investigated:

- Consider one Wind plant and one PV plant that are included in the aggregation, installed in distant locations. Do they contribute significantly to the aggregated production? Are their productions independent?

- How does the variance of a multi-source aggregation compare to the variance of a single power plant in the aggregation?

- Is the variance of a multi-source aggregation lower than the variance of an aggregation of same cardinality with only one source of energy, and if so, what is the expected gain in variance?

2.3.1 Spatial correlation between plants in the aggregation

The relation between aggregated renewable production and production at the individual plant level has been studied in [105] in the context of regional wind power forecasting. More precisely, [105] observes a linear correlation between the wind production at a regional level (onshore Denmark) and a single wind farm or a subset of wind farms, with little influence of wind conditions on this correlation. In the context of a multi-source VRE VPP containing tens of power plants, correlations are expected to be diverse across regions and energy sources. The linear cross-correlation of between the production of all plants in the aggregation within a Franco-German Wind+PV VPP, measured during one year with a resolution of 10 minutes, is represented in Figure 2.3. A decrease in correlation is observed with increased distance as expected. High correlations between plant productions are observed on the plants closely located, which are here mostly wind power plants located in the same region. It is noticeable that on average, correlation between distant plants

(distance > 250 km) of same or of different energy sources, is low. This is a desired property of a VPP providing balancing services: low correlations are favourable to the compensation of variability between sites. This finding is coherent with those of [106], although in the cited work the correlation was established on weather variables (wind speed and irradiation) instead of production and on a different climate (nordic).
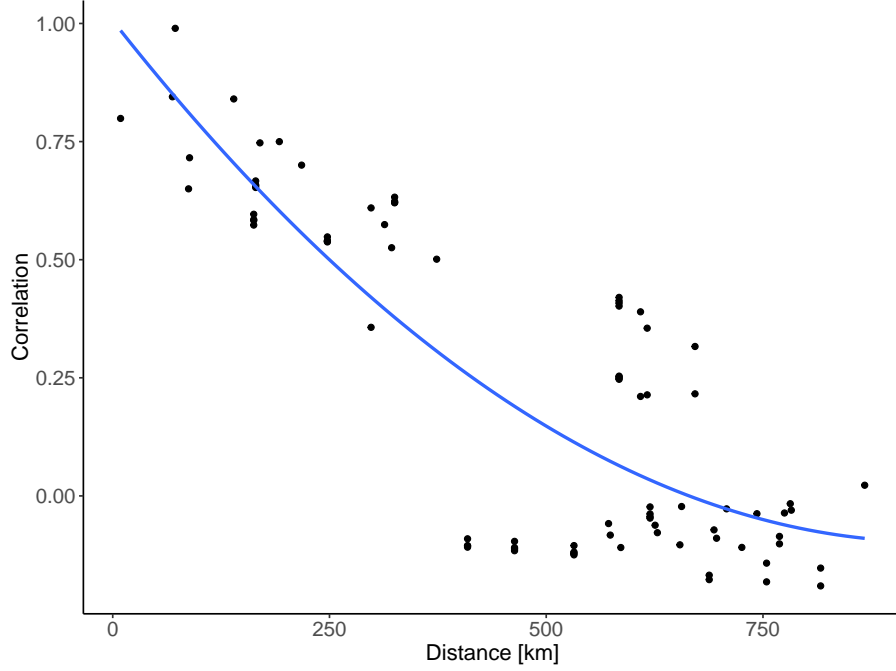
Figure 2.3: Scatterplot of cross correlation between all plants composing the VPP, as a function of their distance in km. The blue line indicates a polynomial regression.

It is known that the aggregated production of renewable plants exhibits a smoother profile compared to its individual plants, due to the diversity in production patterns [105]. A simple indicator of the smoothing effect is the Smoothing Factor (SF), defined in (2.2). The SF quantifies the reduction in the variance of the aggregated production of P plants compared to the average variance at the level of individual plants.

$$SF = 1 - \frac{\mathbb{V}(\sum_{p=1}^{P} y_p)}{\frac{1}{P} \sum_{p=1}^{P} \mathbb{V}(y_p)} = 1 - \frac{\mathbb{V}(y_{agg})}{\frac{1}{P} \sum_{p=1}^{P} \mathbb{V}(y_p)} \qquad (2.2)$$

The impact of aggregation with a sole source of energy, namely wind power is assessed first. The SF of the aggregated pool of wind power plants in the VPP presented above is equal to 45% at the level of a control area (3 plants in Germany), and 57% at the European level (3 plants in Germany + 4 plants in France). These values confirm that the smoothing effect of a Franco-German aggregated wind production is higher than for an aggregation of a few power plants at a regional level ([105] evaluates an average smoothing factor between 15% and 20% for 4 onshore wind power plants in Denmark). Now the SF of a multi-source VPP is evaluated. The VPP comprises 15 plants

among wich 3 Wind plants, 3 run-of-river hydro plants and 9 PV plants, located in different weather regions of France. The production data is 1 year with 10 minutes resolution. After computing all possible combinations of plants we obtain an SF for VPP configurations of varying cardinalities, from 2 plants to 15 plants, and varying shares of energy sources.
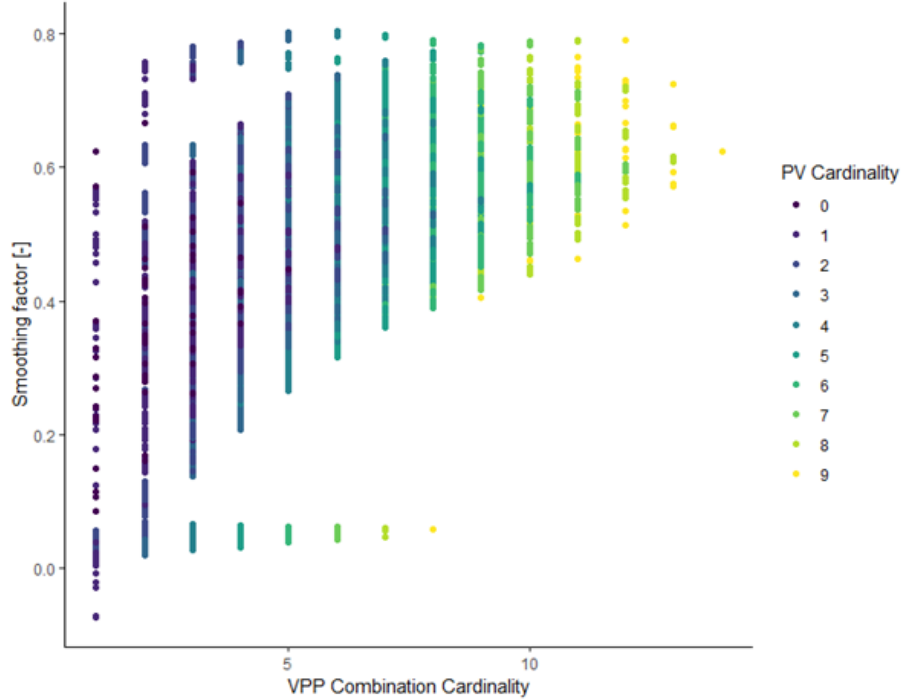
Figure 2.4: Smoothing Factor on all combinations between power plants in a Wind-PV-Hydro aggregation

One observes in Figure 2.4 a distinct zone with low smoothing with increasing cardinality: this corresponds to the aggregation of PV plants only. It is quite clear in this case, given that PV plant in this case are distant of 200 km at most, that the smoothing effect is reduced (inferior to 10%). A large variation of SF is observed at all cardinalities, with a tendency to decrease with greater cardinalities (analogous to the decrease of variance in $\mathcal{O}(\frac{1}{n})$ with increasing size of a Gaussian vector). For cardinalities above 5 plants, the average SF is mostly superior or equal to 50%. This high smoothing is thought to originate from the diversity of production profiles between the three energy sources PV, Wind and Hydro.

Is this complementarity effect of energy sources still valid when considering larger aggregations? A study of the optimal mixture of Wind and PV plants on Northern Europe by [107] minimizes the total aggregated variance of Wind+PV production, constrained by expected energy production levels and upper bounds of foreseen installation capacities. It is found that in a scenario where the annual production is doubled, the optimized standard deviation is 19% lower than merely doubling the capacity of each existing site. It is noted that negative correlations between Wind and PV are expected to be more pronounced with future Wind turbine technologies, reaching more yield by using higher wind hub heights.

In conclusion, the smoothing of variance of multi-source VPP is higher than the usual smoothing observed in aggregations of a single source of energy. Therefore, a multi-source VRE VPP has an intrinsic advantage over single source VRE VPP for the provision of balancing services, due to the fact that its production is expected to be less variable.

2.3.2 Temporal characteristics of aggregated production variability

In the previous paragraphs, the variability of aggregated production has been studied within the space defined by the plants of the aggregation. The temporal variability of production is another important feature for reserve provision: the supplier should know which variation in VPP production is expected between successive timesteps of the reserve delivery period.

In Figure 2.5, the auto-correlation in the production of the VPP presented beforehand is decomposed, starting from the level of single wind farms up to Wind/PV aggregations, for lags inferior or equal to a week. Blue lines represent aut-correlations obtained from Wind plants, and orange lines represent aggregations containing both Wind and PV plants. The thickness of lines is proportional to the number of aggregated plants. The auto-correlation of aggregated production is on average higher than individual plants, especially in the first 48 hours. This is a consequence of the diversity and complementarity between production signals. The addition of a modest share of PV in the aggregation (here up to 10% of the capacity) does not have a large impact on the auto-correlation.

Additional analysis is reported in Appendix A in order to understand better smoothing effect impacts production over the range of temporal scales (from seasonal variations to ramps with sub-hourly periods) and over the range of spatial scales (e.g. correlations between energy sources).

In conclusion to this Section, the aggregation of VRE plants with multiple energy sources has a lower variance and a higher auto-correlation than the plants composing the aggregation.
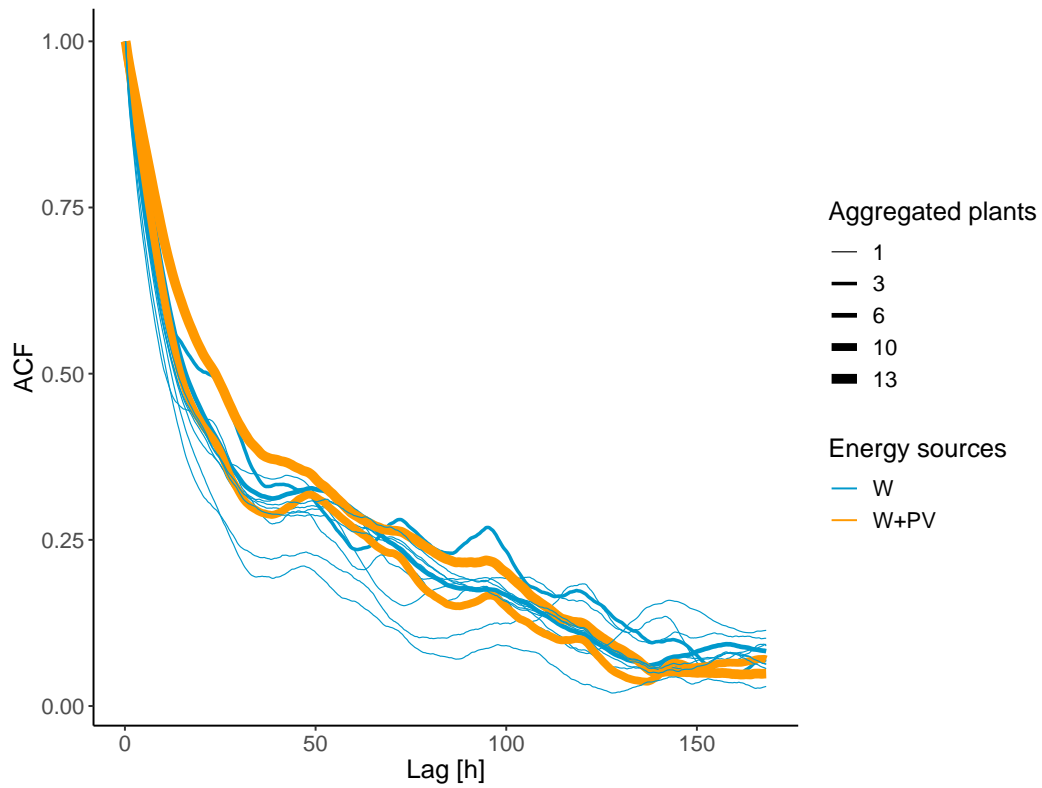
Figure 2.5: Auto-correlation diagram of productions, from Wind plants and aggregations (Wind 'W') to Wind/PV aggregations ('W+PV'). The line thickness is proportional to the number of aggregated plants.

2.4 Direct probabilistic forecasting of the aggegated VRE production

The total production of a VRE aggregation is of interest for the provision of ancillary services, because the smoothing effect brought by aggregation reduces the variability of the signal (cf. Section 2.3), which eases forecasting and consequently lowers risks of failing to deploy reserve capacities. The objective of this Section is to develop a regression model tailored to the direct forecast of aggregated VRE production, i.e. learning directly the aggregated production of a VRE-based VPP from the information available on all plants of the VPP.

More formally, considering $Y = \sum_{p=[1,P]} Y_p$ the random variable modelling the aggregated production of P plants, we look for a prediction model $m_{\hat{Y}}$ depending of the parameter set θ which maximizes the likelihood $\mathcal{L}$ of observations considering the conditions expressed by explanatory variables X.

$$\arg\max_{\theta} \mathcal{L}(m_{\hat{Y}}(.|X,\theta),Y) \tag{2.3}$$

In the context of renewable production forecasting with uncertainty, choosing for regression model $m_{\hat{Y}}$ a simple distribution of uncertainty such as a Gaussian distribution is known to be generally outperformed by distribution-free regression models [32]. An effective resolution of the distribution-free regression model consists in reformulating the problem as a quantile regression, minimizing the expected Quantile Loss (QL) for all quantiles $\tau \in [\tau_{min}, \tau_{max}]$ discretizing the Cumulative Distribution Function (CDF) of production:

$$\arg\min_{\theta} \sum_{\tau=\tau_{min}}^{\tau_{max}} \mathbb{E}[QL_\tau(m_{\hat{Y}}(\tau|X,\theta),Y)] \tag{2.4}$$

with

$$QL_\tau(\hat{y},y) = max(\tau.(\hat{y}^\tau - y), (1-\tau).(\hat{y}^\tau - y)) \tag{2.5}$$

The choice of directly forecasting the aggregated VRE production, instead of forecasting hierarchically starting from sub-aggregation levels (e.g. at the level of the individual plants), has the following advantages and drawbacks:

- *Advantages of direct forecasting*: The direct approach streamlines the forecasting process, avoids multiple forecasts at sub-levels of the VPP (plants or smaller aggregations) and an additional model of the dependencies between these sub-levels.

- *Drawbacks of direct forecasting*: A direct forecast ignores information provided by the production at sub-levels of the aggregation, it can therefore issue a biased response, not having observed for instance that one sub-level is saturating.

In the present Section, the following assumptions hold:

- Forecasting is restricted to the day-ahead horizon (12 to 48 hours), which corresponds to the main horizon of interest for bidding on energy and reserve capacity markets.

- At the day-ahead forecasting horizon, recent levels of production are not influential, they can be excluded from the set of features [32].

- Information about the production of any plant in the aggregation is available, without privacy constraints.

2.4.1 Quantile regression with decision-tree based models

Decision trees partition the space defined by explanatory variables using a succession of splits. Developed originally mostly for classifications, decision trees are also fit for regression purposes: regression models grow trees in parallel (Random Forests, RF [108]) or iteratively (Gradient Boosting Trees, GBT [109]) with randomized selection of the inputs to optimize splits at each node of the trees. In the following, two models based on decision trees and adapted to probabilistic forecasting of renewable production are presented.

Originality of the approach

The probabilistic forecasting of VRE production by means of decision-tree models such as QRF and GBT has been previously studied for both wind production (e.g. in [110]) and PV production (e.g. in [44]). What is new in the present work is the proposition of applying these state-of-the art regression models to the *direct prediction of an aggregated VRE production where multiple sources of energy are present.*

Quantile Regression Forests

Quantile Regression Forests (QRF) [111] offer a well-suited model for the forecast of aggregated VRE production, because they have the availablity to treat avariables of large dimension with diverse levels of correlation. The QRF model is an adaptation of Random Forests proposed by [108], in which all observations are partitioned in regression trees, and the number of observations contained in each node of the tree enable to approximate the quantile loss. The QRF evaluates in Algorithm 1 a conditional distribution of the power production weighted by the features, which are randomly selected by a parameter θ in each tree of index $\sqcup$ and tested within aggregated decision trees. Each split at a node of the tree is carried out to maximize the diversity of variance within the subspace S_l of the learning set defined by a leaf l. The number of trees grown $\mathcal{T}$ must be sufficiently large to fit several times each training point identified by the indicator function. The regression performance is rather insensitive to the number of variables randomly selected at each split, if it is at least equal to the recommended value for regression ($s \sim 1/3$ of the number of explanatory variables).

QRF has been chosen for its ability to perform regression on multivariate inputs of large dimensions [111] and for its proven performance for individual wind [110] and PV [44] day ahead

---

Algorithm 1

Quantile Regression Forests

1: **Initialize:** Define a random generator $\theta = \{\theta_\omega, \quad \omega \in [1, \Omega]\}$.

2: Grow regression trees $T(\theta_\omega), \quad \omega \in [1, \Omega]$:

  At each leaf l of a tree, a subset of features $X_{s,\theta_\omega}, s \sim \frac{1}{3}d, X \in \mathbb{R}^d$ is randomly selected.

  A regression tree $T(\theta_\omega)$ looks for the threshold values t_s that maximize the variance explanation after splitting the leaf subspace S_l into $S_{l'}$ [112].

  $T(\theta_\omega) = \{t_s^l : \arg\max_{t_s^l} (\sum_{l'} \mathbb{V}(S_{l'}) - \mathbb{V}(S_l)^2 \quad \forall l, l' \in \omega\}$

3: The conditional distribution of production over all N observations and all trees Ω is evaluated:

  $\hat{F}_{Y_{t+h}|\mathbf{X}_t}(y, x) = \sum_{i=1}^N \frac{1}{\Omega} \sum_{\omega=1}^\Omega \frac{\mathbf{1}_{\mathbf{x_{t,i}} \in S_l(x, \theta_\omega)}}{Card([j:\mathbf{X_{t,i}} \in S_l(x, \theta_\omega)]} \mathbf{1}_{Y_i \leq y}$

---

forecasting. Finally, the decorrelation of the explanatory variables obtained in the QRF by bagging and random variable selection is an interesting feature for differentiating various plants of the same energy source.

Gradient Boosting Trees

In contrast with QRF which operates on parallel trees, the Gradient Boosting Trees (GBT) model [109] builds iteratively shallow decision trees which maximize the improvement in the gradient of quantile loss (see Algorithm 2). It has also been implemented successfully for renewable production forecasting (PV [45], Wind [46]).

Algorithm 2

Gradient Boosting Trees

1: Fitted response $\hat{y}$ is initialized for all available observations in sample, $t \in [1, N]$ to a constant $\hat{y}^{(0)}$

2: **for** $t \in [1, N]$, for i in $[1, N_{iterations}]$ **do**

3:   compute the negative gradient of loss g_t associated to the decision tree T:

  $g_t = -\frac{\partial}{\partial \hat{y}_t} QL(y_t, \hat{y}_t), \hat{y}_t = \hat{y}_t^{(i-1)} \quad$ where $\quad \hat{y}_t = T(x_t)$

4:   sampling randomly a subset batch of data, and fit a regression tree explaining the gradient loss based on explanatory variables x_t

5:   find the optimal prediction ϖ_k for each terminal node n_k in $[1, K]$ of the tree:

  $\varpi_k = \arg\min_\varpi \sum_{x_t \in n_k} \mathcal{L}(T^{(i-1)}(x_t) + \varpi, y_t)$

6:   the model estimation is updated for $t \in [1, T]$:

  $T^{(i)}(x_t) = T^{(i-1)}(x_t) + \eta\varpi_t$, where η is a shrinkage parameter and ϖ_t is the optimal prediction associated to the observed response y_t.

Hyper-parameters exist to control the fit, but they have limited influence on the forecasting result in the present context. These parameters include the rate of subsampling, the number of trees, the number of hierarchical levels in the trees, and the minimal number of observations per node. The relative insensitivity to parameter selection is comparable with QRF. It is an advantage for easy implementation but it also implies limits in the learning capability of GBT: for instance,

increasing the depth of GBT will not yield a better performance in forecasting because as loss is minimized independently at each terminal node, nodes at the bottom of a profound decision tree, associated with few observations, will have a neglectible impact on the iterative loss reduction.

Pre-processing features for decision-tree based models

Features influencing the aggregated production at day-ahead horizons are constituted by day-ahead Numerical Weather Predictions (NWP), for each power plant of the aggregated portfolio. Decision-tree models such as QRF and GBT rely on input diversity. Fitting these models only on the weather conditions at a given horizon and a precise location restricts their learning capabilities. As decision-tree models are relatively immune to correlation between features [108], additional variables can be included into the feature vector following the classical "data augmentation" technique [51]: the explanatory variables X associated with a given plant of index p in (2.6) integrate, along with weather forecasts at the site of each plant at a horizon h, the same weather forecasts at several lags and leads (e.g. h-1 hour and h+1 hour), and statistical moments (e.g. standard deviation, skewness, kurtosis) of the distribution of weather variables over the grid points neighboring each plant.

$$X^p = [X_h^p, X_{h+1}^p, X_{h-1}^p, X_{h,sd}^p, X_{h,skew}^p, X_{h,kurto}^p] \tag{2.6}$$

2.4.2 Quantile regression with neural networks

The decision-tree based models presented above have limited tuning capacities. This leads to investigate versatile models able to establish more complex relationships between features and aggregated VRE production. More specifically, two types of regression models based on neural networks could be relevant for the present problem: *convolutional neural networks* (CNNs) filter spatial information in a hierarchical structure to build progressively abstract representations of the response, and *recurrent neural networks* (RNNs) model explicitly sequential dependencies which can be useful to reproduce the behaviour of aggregated VRE production viewed as a time series. Lastly, the standard Fully-Connected Neural Network (FCNN) or Multi-Layer Perceptron is implemented as a reference.

The three types of neural network architectures are shortly described below, with reference to existing works related to VRE production forecasting.

- Fully Connected Neural Network (FCNN): an FCNN, also called Multi-Layer Perceptron, is defined by several hidden layers of neurons, fully-connected to the preceding and succeeding layer, resulting in a complex non-linear approximation of the relationship between explanatory variables and response. Stacking several hidden layers and controlling back-propagation of gradients is a generic solution to problems of large dimension that can be easily applied to quantile regression problems. It is however difficult to implement deep FCNNs on problems of large dimension due to the high number of parameters necessary. This is why other network architectures have been proposed, inspired by findings on visual recognition and signal processing.

- Convolutional Neural Networks (CNN): a CNN filters the input data in its various dimensions, extracting information at various spatial and temporal scales. The CNN proposed by [49] for intraday PV forecasting shows improvement of probabilistic forecasting performance (evaluated by Continuous Ranked Probabilistic Score (CRPS)) compared to Support Vector Machines and conventional shallow neural networks.

- Recurrent Neural Networks (RNN): an appealing model for predicting time series is the Recurrent Neural Network model. The RNN consists in linking neural networks as cells, which share memory information. The Long-Short-Term Memory network (LSTM), first developed by [113] has been applied to short-term wind power forecasting with improvements compared to alternative statistical models [53].

Originality of the approach

Albeit neural network regression have already been proposed to generate probabilistic forecasts of VRE production, their application to the problem of muli-source VRE aggregated production has not been significantly investigated. A particular novelty of the approach presented here is the *adaptation of CNN regression to the feature space defined by the plants of various souces in the VPP.*

More specifically, the features available at plant level (in this work day-ahead horizons are considered, therefore features are weather predictions) are organized in volumes where plants are placed in rows, features in columns and horizons as channels. This creates a multi-horizon regression model. Furthermore, two alternative configuration of the feature volume are investigated considering the case of a VRE-based VPP: (1) rows are filled with plants of same energy source in order to filter first cross-plant information with similar production patterns, and (2) rows are filled with randomly shuffled plants in order to filter first cross-source information on VPP plants of diverse production patterns. A second contribution lies in the comparison of this CNN architecture with sate-of-the-art neural network regression models, namely LSTM and FCNN, for the probabilistic prediction of aggregated VRE production.

Convolutional Neural Networks

Convolutional neural networks are a mature solution for image recognition and other tasks involving data obeying to spatial organizations. One important characteristic of CNNs is their ability to detect invariant phenomenons: their hierarchical spatial filtering enables them to recognize phenomenons even if those are subject to variations in space. Instead, a standard fully-connected network can not detect a phenomenon which has been dilated or translated [51]. Principles of CNN have common roots with the Wavelet Transform [114], [115], which enables to locate variations both in time and space (space being defined according by the problem at hand). The pooling operation on the feature maps obtained by convolution layers shrinks progressively the information. This produces a model less sensible to locations of information, the resulting location invariance [61] is a possible advantage over fully-connected networks. The filtering and pooling leads also to a reduced

number of parameters compared to a fully-connected network of similar depth, CNNs are therefore less costly to train [51].

In the present problem, the space is defined by the ensemble of features available on all plants of the VPP. Assuming D features for P plants, then the available information is displayed as a volume of sequential $D.P$ matrices over the horizon range H. A tensor x is obtained by joining all matrices over the horizon range. Each slice in depth of the tensor is called a *channel*.

The CNN regression of the aggregated VRE production is summarized in Algorithm 3. It starts with the configuration of the inputs: how should the features of the different plants composing the aggregation be placed in the feature tensor? Two different configurations are considered and illustrated in Figure 2.6:

- **Source-ranked plants:** Plants sharing the same energy source can be positioned contiguously in rows of x to filter first by energy source then mix sources at a more abstract level.

- **Shuffled plants**: Plants are placed randomly in rows of x, then the model will start filtering mostly cross-source dependencies between features at the level of plants (e.g. one Wind plant with one PV plant) before learning the smoothing effect of aggregation due to diversity in production profiles.
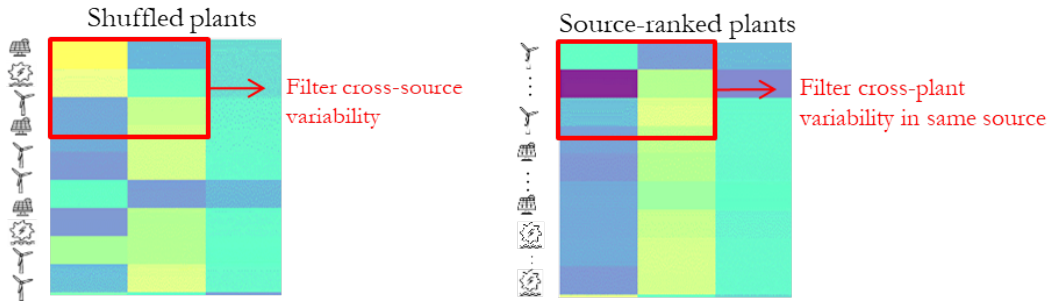
Figure 2.6: VPP configurations of feature space in CNN. Plants are placed in rows, either shuffled (plants) or placed contiguously by energy source (right). Features corresponding to the plants are placed in columns. Color levels symbolize centered and scaled values of explanatory variables. Red rectangles materialize a convolutional filter.

In this work a reduction of the dimensionality of features was implemented, in an attempt to decrease computational burden and ease the learning capacities of the network. To do so, a Principal Component Analysis (PCA) is applied at each batch and horizon. Applying PCA separately on each horizon preserve the temporal correlation between horizons, its aim is to reduce the number of features per plant in the VPP without loosing too much information. However, if computational burden is not an issue, this first step in Algorithm 3 can be discarded. In that case, as plants do not have necessarily the same number of features, columns not associated with a feature variable for a specific plant can be padded very zeros.

Then convolutional layers filter progressively the information contained in all channels. In classification problems, a subsampling step called *pooling* extracts synthetic results of feature maps,

by taking averages or maxima over a defined window. This is effective for detecting if an edge is present in a part of an image, and improves the invariance of learning, but will create information loss here where the objective is to have a full predictive capacity over the production distribution. Finally, the last layer is connected to multiple outputs through fully-connected layers. The outputs are nodes associated each with a specific quantile loss, with as many quantiles as necessary to cover the whole CDF. The full architecture of a CNN for aggregated VRE forecast is illustrated in Figure 2.8.

Algorithm 3

Convolutional Neural Network for regression of aggregated VRE production

---

1: A PCA reduces the number of features of each plant.

2: Features X are arranged in volumes with height h corresponding to all plants in the VPP, width w to features and channels c to horizons.

3: The first convolutional layer applies on X a filter $f^{(1)}$ with a *kernel* of size (k_w, k_h, k_c). A *stride* parameter (s_w, s_h) adds distance between kernels in order to reduce the dimensionality of the convolutional layer (see Figure 2.7). A bias $b^{(l)}$ is applied to all pixels in the obtained feature map $h^{(l)}$.

   $h^{(1)} = X \circledast f^{(1)} + b^{(1)}$, i.e.

   $h_{ijc}^{(1)} = \sum_{u=1}^{k_w} \sum_{v=1}^{k_h} \sum_{c'=1}^{k_c} x_{i'j'c'} w_{uvcc'} + b_c^{(1)}, \quad i' = u.s_w + k_w - 1, \quad j' = u.s_h + k_h - 1$

4: **for** l in L convolutional layers: **do**

5:    Filter $f^{(l)}$ is convolved with the previous layer $h^{(l-1)}$, corrected by a bias term $b^{(l)}$: $h^{(l)} = h^{(l-1)} \circledast f^{(l)} + b^{(l)}$

6:    No pooling on the obtained layer.

7:    Batch normalization on the feature axis (see Section 2.4.2)

8: **end for**

9: Flatten layer: all filters of the last convolutional layer L are flattened: $y_{flat} = [y_f^{(L)}, \forall f \in F)]$

10: Output layer: the predicted production at horizon h is associated to O nodes, where each node $i \in O$ is associated to a quantile loss of value τ_i.

11: Optimization of all quantile losses in the output through gradient backpropagation.
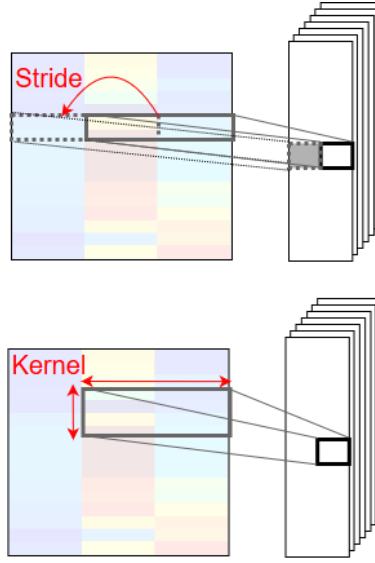
Figure 2.7: Illustration of stride and kernel of a convolutional layer in a CNN
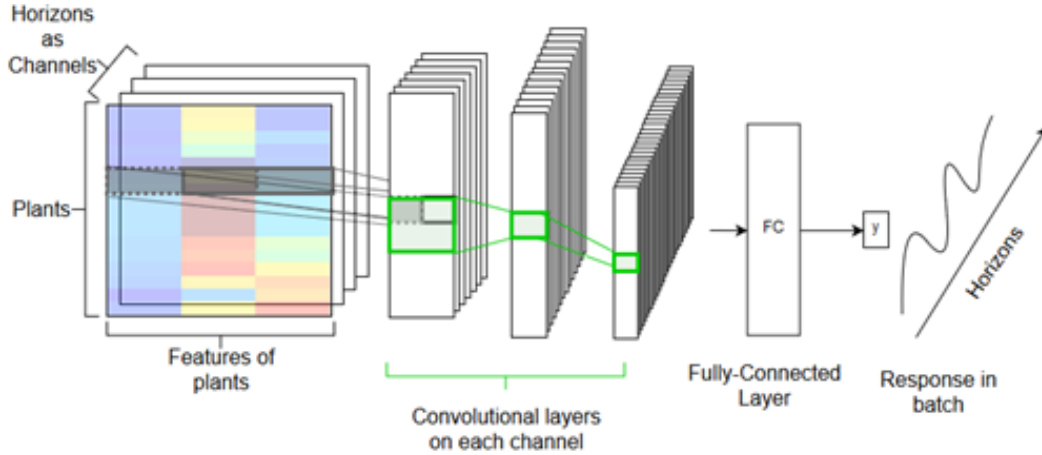
Figure 2.8: Overview of a CNN model for the multiple quantile regression of aggregated VRE production

Batch normalization on the feature axis

In a neural network with many layers, the information learnt can vanish or saturate quickly when the gradient of error is propagated back into the network. Considering that the CNN model is trained on batches of training samples, we remediate to the saturations or vanishing of gradient by applying *batch normalization* to every layer of the network. The normalization occurs by centering and scaling each batch before feeding to a layer [51]. However, by doing so network weights deviate from the optimum, so parameters of the normalization must be optimized themselves, after a first pass of backpropagation. A specificity of batch normalization on CNNs is that normalization is done on all dimensions except the channel depth, whereas in fully-connected networks normalization is done on all axis. This is done to preserve the ability to discriminate information across channels
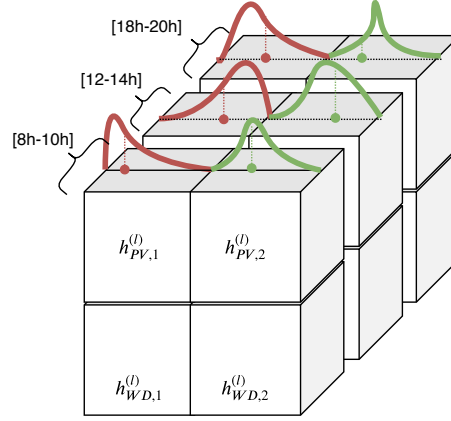
Figure 2.9: Illustrated distributions of samples in a batch over several channels (horizon intervals) of a CNN, for filtered features connected to PV plants (volumes in the top row). Let the two PV filtered features $h_{PV,1}^{(l)}$ and $h_{PV,2}^{(l)}$ be horizon-dependent as shown (e.g. representation of solar radiations or temperatures). Mean values over the batch (solid points) at each channel are different, so normalization must be done separately for each channel.

(see Figure 2.9). In summary, the tensor at layer (l) $x_b = h_{b,c}^{(l)}$ is normalized on every batch $b \in \mathbb{R}^B$ and channel c, with $\mu_{bn}, \sigma_{bn}, \gamma_{bn}, \beta_{bn}$ being learnable parameters representing the mean, standard deviation, scale and offset. A non-learnable parameter ϵ_{bn} avoids division by zero.

$$\mu_{bn} = \frac{1}{B} \sum_{i=1}^{B} x_{b,i} \tag{2.7}$$

$$\sigma_{bn}^2 = \frac{1}{B} \sum_{i=1}^{B} (x_{b,i} - \mu_{bn})^2 \tag{2.8}$$

$$\hat{x}_{b,i} = \gamma_{bn} \frac{x_{b,i} - \mu_{bn}}{\sqrt{\sigma_{bn}^2 + \epsilon_{bn}}} + \beta_{bn}, \quad \forall i \in \mathbb{R}^B \tag{2.9}$$

Recurrent Neural Networks

The Recurrent Neural Network (RNN) architecture is a natural candidate for time series prediction, which has been applied for wind and PV power forecasting as seen in the Introduction. The information can propagate forward as in a temporal process, and link to a sequence of output such as the successive observations of production over a forecasting horizon. There is however limited existing work on the properties of RNNs for aggregated VRE production with multiple energy sources.

RNNs are based on neural networks that treat sequences of data: for each timestep t of the training dataset, a neural network, usually called a cell treats the available information $\mathbf{x}_t$, outputs a response $\mathbf{y}_t$ and shares a state variable s_t with the following cell. The state is in turn used as inputs to the adjacent cell: in our example the state s_{t-1} transferred from the previous cell

passes a form of "sequential memory" to the present cell. As renewable production process can be considered causal, with known auto-correlation of production between neighbouring timesteps, we are primarily interested in causal, mono-directional networks. Note that bi-directional RNNs can share states forward and backward to the previous cell (this can be useful in language processing where words at the end of a sentence inform on the global meaning of the sentence). Cells can also be stacked in several layers to form deep RNNs.

An RNN cell can take a variety of forms, which differ in the modelling of state and in the treatment of information within the cell. The most popular forms of cells in the context of time-series forecasting are Long-Short Term Memory (LSTM, [113]) and Gated Recurrent Units (GRU). The LSTM decomposes the state in two components, long-term ($\mathbf{c}_{(t)}$ where c stands for "cell") and short term ($\mathbf{h}_{(t)}$ where h stands for "hidden"). A forget gate $\mathbf{fg}_{(t)}$ updates the previous long term memory by forget weights, obtained from short-term memory and features. The long term memory is subsequently further modified by a bias term produced by the input gate $\mathbf{ig}_{(t)}$, which controls how much of the main layer $\mathbf{g}_{(t)}$ will pass to the long-term memory.

$$\mathbf{c}_t = \mathbf{fg}_{(t)} \odot \mathbf{h}_{t-1} + \mathbf{i}_{(t)} \odot \mathbf{g}_{(t)} \tag{2.10}$$

$$\mathbf{y}_t = \mathbf{h}_{(t)} = \mathbf{o}_{(t)} \odot \tanh \mathbf{c}_t \tag{2.11}$$

$$\mathbf{g}_{(t)} = \tanh \left(w_{x,g}^{\mathsf{T}} \mathbf{x}_t + w_{h,g}^{\mathsf{T}} \mathbf{h}_{t-1} + b_g \right) \tag{2.12}$$

$$\mathbf{i}_{(t)} = \sigma (w_{x,i}^{\mathsf{T}} \mathbf{x}_t + w_{h,i}^{\mathsf{T}} \mathbf{h}_{t-1} + b_i) \tag{2.13}$$

$$\mathbf{o}_{(t)} = \sigma (w_{x,o}^{\mathsf{T}} \mathbf{x}_t + w_{h,o}^{\mathsf{T}} \mathbf{h}_{t-1} + b_o) \tag{2.14}$$

where $\odot$ denotes the element-wise Hadamard product of matrices. Interestingly, the state parameters can flow through the network units without being modified during gradient propagation, thanks to the possibility to discard / forget the information of previous steps. Therefore the structure of the LSTM offers a form of protection against vanishing or exploding gradients. Parameters of the LSTM can be optimized via state-of-the-art maximization of the likelihood by back-propagation the error gradient over sequences.

If one wishes to regularize the network and make it able to express uncertainty, he/she can turn to a Bayesian approach, where network weights are treated as random variables that can be sampled and optimized via probabilistic inference techniques. For instance the Bayes By BackProp inference scheme [116] minimizes the KL-divergence between the posterior and prior distributions of parameters, and can be implemented into Bayesian Recurrent Neural Network [117]. In this work for the sake of simplicity the Bayesian approach is not implemented, network parameters remain deterministic. Other state-of-the art techniques help control learning in the LSTM: gradients are stabilized with batch normalization (see Section on convolutional networks) and regularized for better handling of uncertainty by applying *dropout*. Dropout consists in masking some of the hidden neurons during training. It can be recurrently integrated into every cell, or applied only between layers (e.g. between two stacked LSTM layers, or between the last LSTM layer and the fully-connected layer associated to with the quantile loss function). In the present case it is found that non-recurrent dropout with moderate ratios ($\leq 30\%$), helps reduce overfitting.

Reducing the dimensionality of features for neural networks

Neural networks can treat inputs of large dimensions. However, nodes of CNNs and RNNs are connected to limited parts of the feature space (in the spatial direction for CNNs and in the temporal direction for RNNs). Hence, a reduction of the dimensionality of features helps both types of network find significant relationships with hidden layers of moderate dimensions. Note that, as stated in the presentation of the CNN model, networks can operate without reduction of feature dimensionality, but they may have more difficulties in learning efficiently relevant patterns for aggregated production if the amount of features is large, especially if those are highly correlated. A reduction strategy is implemented here to produce $D^{p,-}$ features for each plant p of the aggregation, with $D^{p,-} \leq D^p$. Features are reduced while preserving their variance with the help of a Principal Component Analysis (PCA), unrolled on every plant of the VPP simultaneously. PCA has proven to be effective at reducing dimensionality of weather features on multiple sites in the context of renewable power production [118]. PCA consists in reconstructing features via a compression by a layer of hidden neurons, linearly activated (cf. Figure 2.10). Encoded features of dimension $D^{p,-}$ are optimized by minimization of the squared loss between the original features and their reconstruction. Note that by applying non-linear activations and stacking more hidden layers, the model becomes an auto-encoder that can derive more complex reductions within NWPs [119].
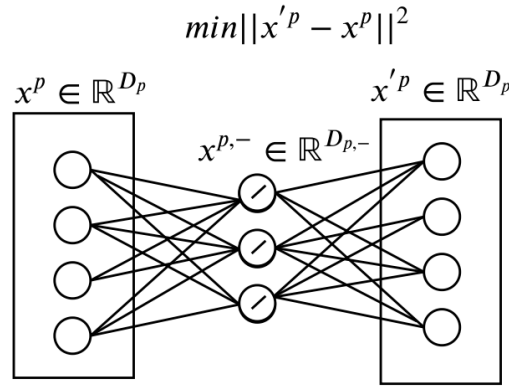
Figure 2.10: Principle of PCA applied for dimensionality reduction

2.4.3 Case study 1: Comparison of decision-tree models (QRF, GBT)

The purpose of this first Case Study is to compare the merits of both decision-tree based models, namely QRF and GBT, for the probabilistic forecast of aggregated multi-source VRE production. Decision-tree models are benchmarked against a simpler quantile linear regression model (QLR), fitted on weather variables without data augmentation to avoid singularities in the covariance matrix (weather variables and their lagged counterparts or statistical moments are highly correlated).

The forecasting models are evaluated on a VPP combining Wind and PV plants operating in France, with a total capacity of 42.3 MW and a 24% share of PV. The distance between any two power plants varies between 30 km and 700 km. These relative high distances imply that spatio-temporal correlations in the production are rather low. This is an interesting feature for

the present application, i.e. the more diverse the production profiles are within the VPP, the less variable its power output will be, and likely with higher minima. Real production data covering the period September to December 2015 are used to train the forecasting models, while data from January to March 2016 are employed to evaluate them. Probabilistic forecasts are generated using historic production data and NWPs from the European weather forecasting center ECMWF. The NWPs used are the predictions published at 00.00 UTC, in order to consider delivery delay and have sufficient time to process the forecast before the day-ahead gate closure of balancing market, assumed to take place in the morning between 08.00 and 12.00 UTC.

2.4.4 Case Study 2: Performance of QRF in an operational context for reserve capacity forecast

An operational forecast of the European VPP of the REstable project, presented in the Introduction, has been deployed based on training data for 17 power plants. The training data covers a period starting from September 2017 and ending in May 2018, for a testing period comprised between November 2018 and March 2019. Two alternative sources of NWP have been used: ARPEGE from Meteofrance and GFS. The direct probabilistic forecast was done with a QRF model for quantiles between 1% and 99% at several aggregation levels: power plant, BSP, TSO and Europe.

The probabilistic forecast serve as the basis of a reserve capacity forecast, which is taken in (2.15) as the minimum 1%-quantile of the aggregated production forecast over the validity period $T_{validity}$, here taken at 1 hour, diminished by the largest 1%-quantile forecast of plants under the aggregation to increase the robustness of the forecast.

$$\hat{R}_t = min_{t' \in T_{validity}(t)} \hat{y}_{agg,t'}^{(1\%)} - max_{p \in VPP} \hat{y}_{p,t}^{(1\%)} \tag{2.15}$$

2.4.5 Case Study 3: Comparison of decision-tree and neural networks

In this case study, the merits of CNN and LSTM are compared to those of QRF, on a VPP with high diversity in production profiles due to the presence of 3 different energy sources: PV, Wind and run-of-river hydro. The VPP comprises 15 plants (3 Wind, 3 run-of-river Hydro, 9 PV) all located in several climates in France (Moderate warm continental, Moderate cold continental, Atlantic). The production data comprises 10^5 points of measurement at resolution 30 minutes, between July 2015 and February 2016.

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 NWP issued at 00h00 of the previous day are used. The resulting forecasting horizon is thus comprised between 24h and 48h. To assess the sensitivity of the methodology to the relative proportion of each energy source (Wind, PV, Hydro), the installed capacities of the farms are scaled to obtain two different VPP configurations in Table 2.1. The first VPP (VPP1) is dominated by Wind, whereas the second

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

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

Table 2.1: Installed capacities of VPP configurations in MW

Models have been trained and tested in a cross-validation framework, organized as follows:

- **Aggregated response**: The VRE production is forecasted at the aggregated level of the VPP and by energy source (Wind, PV, Hydro). For the forecast of production by energy source, a pre-processing of response variables is implemented: in the case of PV production, production variables and radiation NWPs are normalized by an analytical Top of Atmosphere model; in the case of Wind and Hydro, a logit transform enables to perform regression on an unbounded space $[-\infty; +\infty]$ instead of the $[0 - 1]$ interval.

- **Cross-validation**: Models are trained on 6 weekdays and tested on the remaining weekday. This does not correspond to a real-life industrial deployment, where models would be trained and tested on a rolling-window approach, but this permits to obtain a cross-seasonal assessment of the forecasting performance. The forecast horizon is 24h-48h, so it is considered that the infringement of temporality due to cross-validation has little impact on results.

- **QRF setup**: Separate QRF models have been trained on 6 horizon intervals of 4 hours each, in order to individuate better the different regimes of aggregated VRE production. CRPS is evaluated based on the interpolation of the CDF constructed by percentile forecasts [1%, 2%,...,99%].

- **Pre-processing** for CNN and LSTM: Features for CNN and LSTM are pre-processed via the PCA presented in Section 2.4.2, which generates 3 features per plant.

- **Loss functions** for CNN and LSTM: models are trained on 10 quantile losses from 10% to 90%, and CRPS is later interpolated from the decile forecasts.

- **Optimizer and batch dimensions for CNN:** The CNN is trained with the Stochastic Gradient Descent (SGD) optimizer on a mini-batch, which performs noisy but sensitive gradient backpropagation. The dimension of the mini-batch is a trade-off between model variance (higher with smaller batches) and bias (higher with larger batches, because it is more difficult to optimize the link between features and response on large batches). A batch size of 16 gives satisfactory results (ie a batch comprises 16 days, with a full horizon interval of 24 hours).

- **Optimizer and batch dimensions for LSTM:** The LSTM, having fewer nodes than the CNN, is found to train more efficiently on batches of minimal size. Furthermore, the stateful property (cf. Section 2.4.2) enables to train it even in *online learning* mode, i.e. with a batch

size of 1. The optimal batch size found empirically in the present case is however higher, in the range of 12 to 24.

Optimal parameters of CNN and LSTM are investigated via grid search. It should be stressed that the aim here is not to find the best possible network architectures, which would require large computational effort on GPU and specific approaches to tackle the challenge of very deep models with tens of layers. The aim is rather to illustrate the potentiality of those networks compared to standard machine learning models. Networks of moderate depth, inferior to 10 layers, prove to be sufficient for this task.

2.4.6 Evaluation metrics

A large number of evaluation metrics for forecasting exist in the literature. The metrics used in this thesis to evaluate forecasts of VRE production are of common use in the VRE forecasting community. The reader may refer to [30] and [120] for more in-depth insights about the evaluation of production forecasting, in its deterministic and probabilistic form. Error metrics are scaled by the installed capacity Pn of the production of interest, in general an aggregated portfolio of plants within a VPP. The following scores are used to quantify the *deterministic* performance of forecasts $\hat{y}$, for N observations of production y:

- Normalized Mean Absolute Error (NMAE)

$$NMAE(y, \hat{y}) = \frac{1}{Pn} \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{2.16}$$

- Normalized Root Mean Squared Error (NRMSE)

$$NRMSE(y, \hat{y}) = \frac{1}{Pn} \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \tag{2.17}$$

Two scores are used to quantify the *probabilistic* performance of forecasts. These scores are proper scores, i.e. an improvement in these scores can be trusted by the forecaster as meaning a higher skill of the forecast [30]. With $\hat{F}_Y$ denoting the predicted CDF of production, these scores are:

- Quantile Score (QS): The QS is the average of the Quantile Losses (QL, see (2.5)) over the range of predicted quantiles $[\tau_{min}, \tau_{max}]$:

$$QS(y, \hat{y}) = \sum_{\tau_{min}}^{\tau_{max}} QL_\tau(y, \hat{y}) \tag{2.18}$$

- Continuous Ranked Probability Score (CRPS) The CPRS quantifies the distance between the predicted CDF of production and a theoretical perfect forecast aligned on the observed production level without uncertainty, which can be formalized by a Heaviside function H placed at the level of observed production [121]:

$$CRPS(y, \hat{y}) = \frac{1}{Pn} \frac{1}{N} \sum_{i=1}^{N} \int_{-\infty}^{+\infty} (\hat{F}_{y,i} - H(y_i))^2 dy \tag{2.19}$$

In addition to the above scores, which give a global view on the probabilistic performance, specific scores are used to evaluate particular aspects of the forecast. The first aspect is the calibration or reliability of probabilistic forecasts, which is evaluated in two ways:

- Reliability diagram:

  The reliability diagram plots the nominal quantiles τ against the observed quantiles $\hat{\tau}$. Observed quantiles are computed by recording the number of observations below the associated predicted quantile $\hat{y}^\tau$:

$$\hat{\tau} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(y_i - \hat{y}_i^\tau) \tag{2.20}$$

  A reliable forecast has a deviation between observed and nominal quantiles close to zero, meaning that the reliability diagram of a reliable forecast is aligned on the diagonal.

- Probability Integral Transform (PIT) :

  The PIT ranks the position of an observation y within the quantiles associated to a predicted distribution F_Y [122]. With a perfectly reliable forecast, the values taken by the PIT over a sample of large size should follow a Uniform distribution.

$$PIT(y) = F_Y(y) \tag{2.21}$$

  Considering that the CDF issued by probabilistic forecasts of renewable production is not always strictly monotonous, ambiguities may arise with the above formula (eg when different forecast quantiles are associated to the same production level, which occurs at minimum or maximum expected production levels). This is why the Rüschendorf distribution transform U which generalizes the PIT [123], is used here instead:

$$U(y) = \hat{F}_{Y^-}(y) + V(\hat{F}_Y(y) - \hat{F}_{Y^-}(y)) \tag{2.22}$$

  with V a uniformly distributed random variable and $\hat{F}_{Y^-}(y)$ the closest quantile below the observation y.

The second aspect is sharpness, which is defined in [30] as the ability to concentrate the probabilistic information of the forecast. More precisely, it quantifies the distance between upper and lower quantiles of the predicted distribution of production. The sharpness indicator with nominal coverage rate β writes:

$$\delta^\beta = \frac{1}{N} \sum_{i=1}^{N} \hat{y}_i^{1-\beta/2} - \hat{y}_i^{\beta/2} \tag{2.23}$$

Finally, the reserve capacity forecast in Case Study 2 is evaluated in terms of availability AV, which is simply defined as the frequency of observed production y laying below the proposed reserve capacity R over an evaluation period T:

$$AV = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(y_i \leq R_i)$$

2.4.7 Results of Case Study 1

The first case study aims at comparing the performances of two alternative regression models based
on decision trees, namely QRF and GBT, for the direct forecast of aggregated VRE production.
Before comparing their forecasting scores, it is interesting to assess how such a type of model
weights the different explanatory variables (here NWP). The explicative value of NWP variables is
evaluated by the QRF model via an *importance factor*, which quantifies the increase in regression
error when values of the selected variable are randomly permuted and tested against out-of-bag
samples. The variables with higher importance are listed in Table 2.11. We note that weather
variables have similar importance levels among plants of the same energy source. The spatial
distribution coefficients were found to improve the aggregated forecast for PV plants only.

Variables	PV	Wind
NWP	SSRD, Temp	U100,V100,W100
Lagged NWP	SSRD.k-1/k+1 Temp.k-1	U100.k-1, V100/W100.k-1/.k+1
Spatial distribution NWP	Temp.sd, Temp.skewness	-

Legend:
- SSRD: solar surface radiation downwards
- U100/V100/W100: zonal/meridional/absolute wind speed at 100 m
- Temp: air temperature at 2 m
- k-1: previous hour

Figure 2.11: Features with most importance in the QRF model

The reliability of the QRF model is evaluated by means of the Probability Integral Transform in
Figure 2.12. The nearly uniform distribution indicates that the probabilistic calibration is adequate,
even if the moderate U-shape of the diagram shows that the model is narrow [124], which means
that the prediction intervals does not cover well the lowest and highest production levels.

The global quantile score on the entire quantile range, reported by monthly averages in Table
2.13, shows that QRF is slightly more efficient than GBT on average. This is confirmed by the
analysis of the Continuous Ranked Probability Score in Figure 2.14. One can observe here that
the QRF model seems to be more flexible in its capacity to learn that a Wind/PV aggregation
has a dual operation, with no PV production at night. Even though the iterative behaviour of
GBT ensures a minimized bias of the forecasting model, the major depth of QRF appears to be an
advantage for the multi-source context.
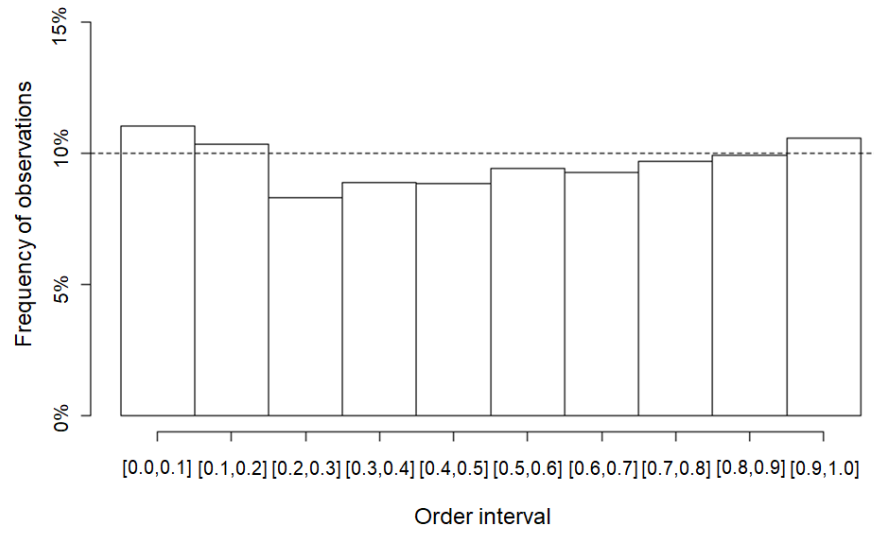
Figure 2.12:   PIT rank histogram of QRF aggregated forecast

Model	January	February	March
QRF	0.023	0.026	0.024
GBT	0.031	0.033	0.032
QLR	0.049	0.057	0.054

Figure 2.13: Monthly-averaged Quantile Score on 5% to 95% quantiles
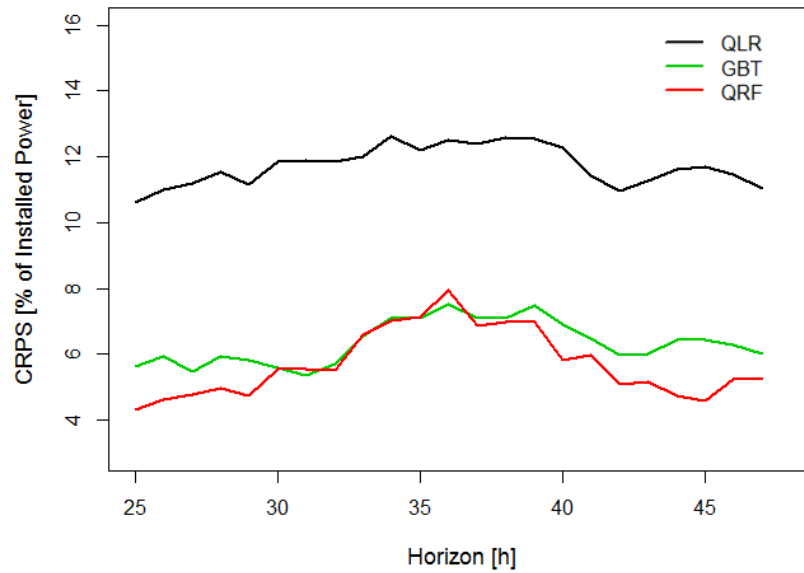
Figure 2.14:   Continuous Ranking Probability Score (CRPS) as a function of the forecast horizon for the QRF and the benchmarks QLR and GBT

2.4.8 Results of Case Study 2

After having identified the QRF as a suitable model for forecasting multi-source VRE production in an offline context, this Case Study evaluates whether QRF is also valid in an operational context, where forecasts are issued on a dedicated server every morning during 3 months. Furthermore, the 1% quantile forecast of the aggregated production is used as a base for the operational reserve capacity forecast of a VPP.

The QRF model is run for various aggregation levels, from single plants to 13 plants. Figure 2.15 presents the NMAE (red curve), NRMSE (green curve) and Quantile Scores at 1% and 5% (blue and violet curves respectively) obtained by QRF for forecasts at various levels of installed capacity, from a single plant to the full VPP capacity. Scores are scaled with the performance of the worst performing plant (leftmost point). As the aggregation grows in size and diversity, a tendency of decreasing scores (deterministic and probabilistic) with increasing capacity is observed.
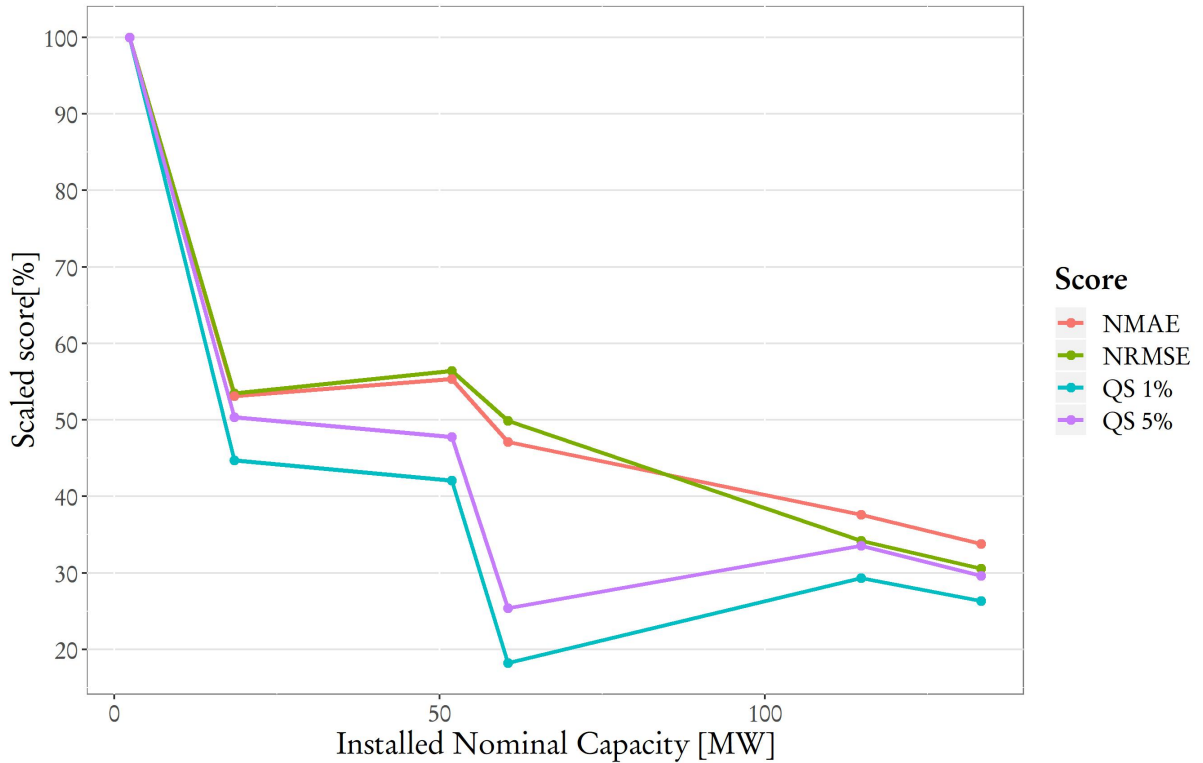
Figure 2.15: Evolution of forecasting score with increasing installed capacity

The distribution of reserve capacity forecasts is reported in Figure 2.16 for 4 sub-aggregations of the VPP, and compared to the distribution of the observed production. It is are concentrated on low shares of the installed aggregated capacity, regardless of the composition of the aggregations, which are here either Wind or PV based.

Finally, the use of NWP with higher resolution and quality (ARPEGE instead of GFS) significantly improves the availability of the reserve forecast, defined in Section 2.4.6. It is observed in Figure 2.17 that the availability of reserve, quantified on the y-axis, increases (round points associ-
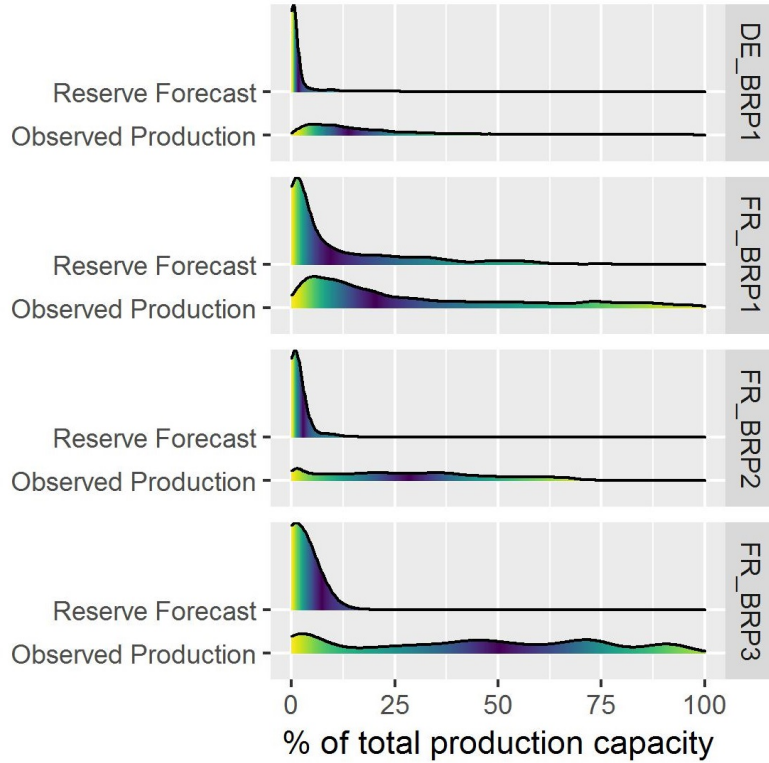
Figure 2.16: Distribution of reserve forecast capacities and observed production, for two wind aggregations (DE BRP1, FR BRP1) and two PV aggregations (FR BRP2, FR BRP3)

ated to ARPEGE lay above triangles associated with GFS). This is done at the cost of decreasing the mean reserve capacity (see x-axis).

In summary, this Case Study indicates that the QRF model may be used in an operational context for probabilistic forecasts. Moreover, the VPP reserve capacity forecast based on 1%-quantile QRF forecasts show an availability of reserve capacity at least equal to 98% on 4 out of 5 different aggregation levels, if NWPs of high quality are used. This is an encouraging although imperfect result (TSOs are likely to expect at least 99% reserve availability).

2.4.9 Results of Case Study 3

Performance of QRF forecasts on various energy sources

The reliability diagram in Figure 2.18 indicates that direct forecasting of the aggregated production of VPP1 with QRF shows adequate reliability. The reliability is achieved for aggregated production as well as forecasts by energy source: the PIT (in its Rüschendorf version) of Wind and PV production shown in Figure 2.19 is rather uniformly distributed, similarly to the aggregated production (cf. Figure 2.19a).

Table 2.2 reports the Root Mean Square Error (RMSE) and the Continuous Ranked Probability Score (CRPS) associated with the forecasts of the aggregated production and of the production of each energy source (PV, Wind, Hydro) for both VPPs (VPP1, wind-dominated, and VPP2, PV-
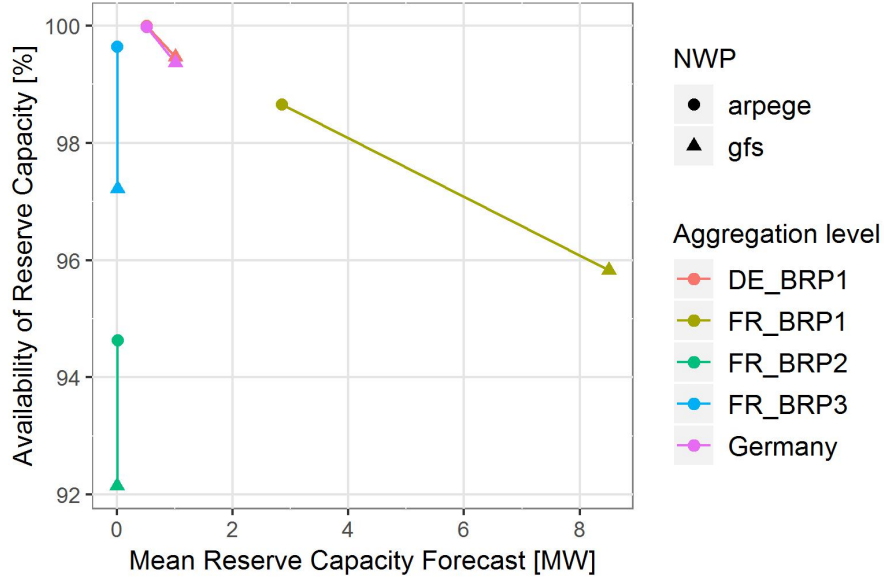
Figure 2.17: Availability of Forecasted Reserve Capacity depending on NWP sources ('arpege' NWP model from Meteo France represented with points, 'gfs' NWP model with triangles), for 5 aggregation levels in the VPP (1 BRP in Germany, 3 different BRPs in France, and all plants in Germany)
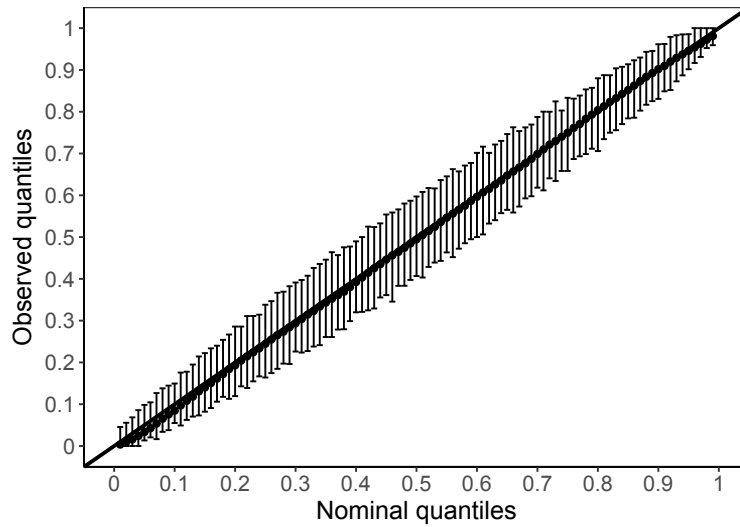
Figure 2.18: Reliability diagram for the direct forecast of the total aggregated production in VPP1. Bars indicate for each quantile the confidence interval associated to the 5%-quantile and 95%-quantile obtained by consistency resampling [125]

dominated). Both RMSE and CRPS indicate that the QRF model has comparable performance with respect to the state of the art in forecasting of Wind and PV production [32]. 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).

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

The sharpness of QRF predictions is presented in Figure 2.20 for several horizons between 10h and 18h and aggregation levels. We see that sharpness of the forecast is higher at the aggregated level (total VPP production, in red) than at the level of each energy source (Hydro, PV or Wind in green, blue and violet respectively). The increase in sharpness at the aggregated level is coherent with the increased smoothing effect observed in Section 2.3. The sharpness of forecasts on PV production (all PV plants) is obviously horizon-dependent, and is performing worst around noon when production levels and uncertainties are highest. The 80% interval is comprised between 30% and 35% of *Pmax* for wind aggregated plants, and 25% for the total aggregation. In light of these elements, the QRF model presents a challenging performance for its neural network-based counterparts.

Figure 2.19: Observed frequencies of PIT for forecasts of aggregated production and for Wind and PV production in VPP1. The dashed line represents an ideal uniform distribution
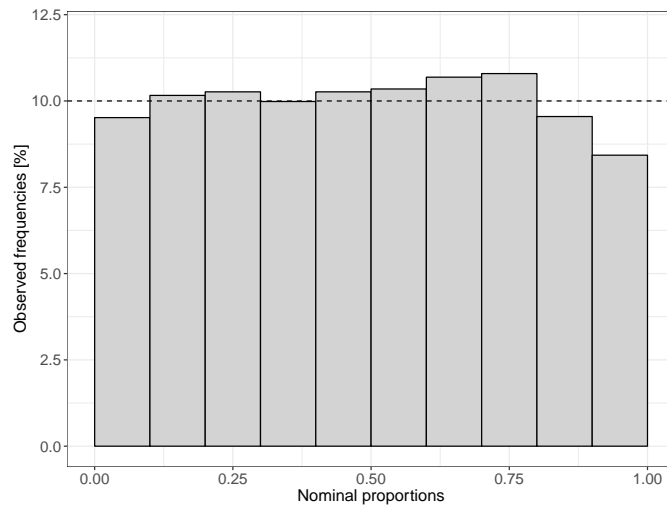
(a) PIT of aggregated production forecast in VPP
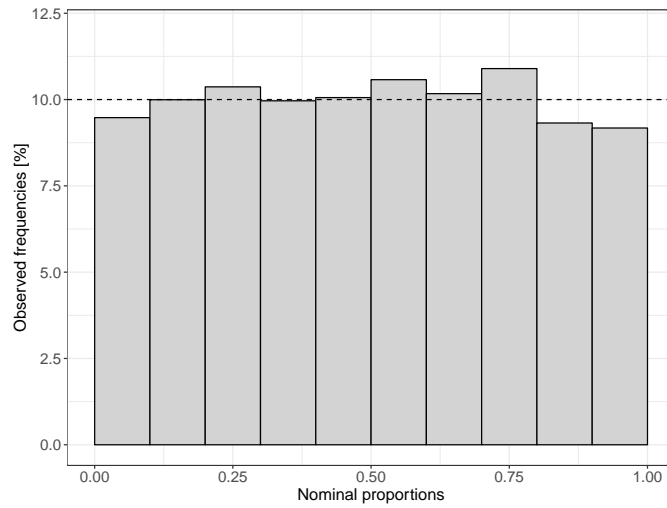
(b) PIT of Wind production forecast
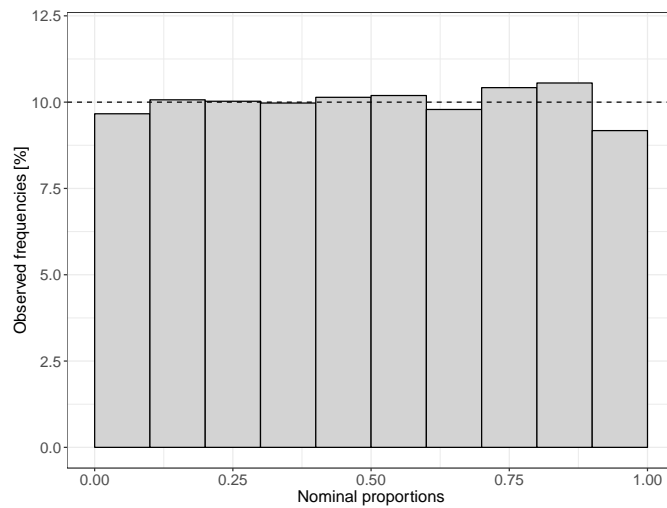
(c) PIT of PV production forecast
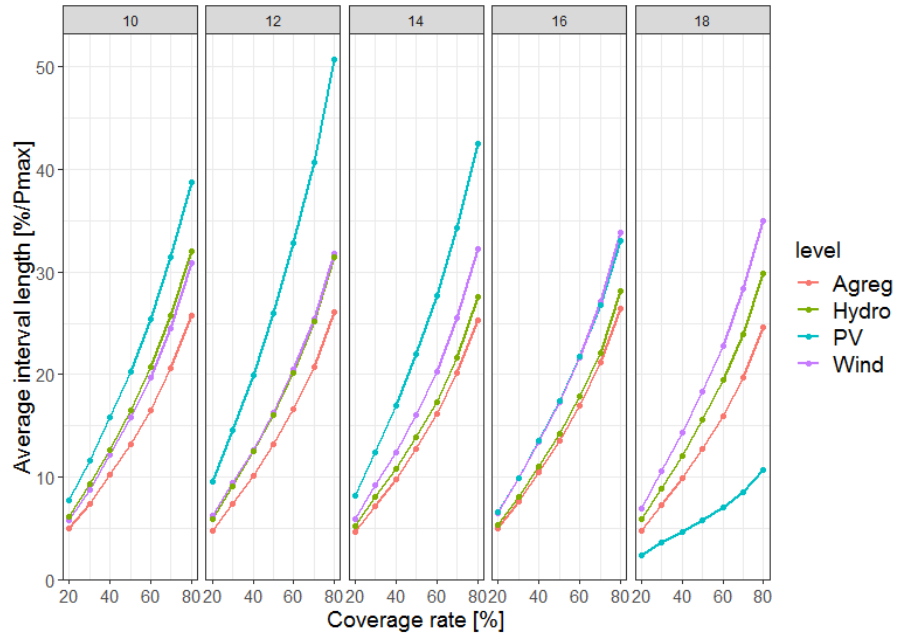
Figure 2.20: Sharpness of QRF forecast for horizons between 10 and 18 hours and different aggregation levels, VPP1.

Performance of neural networks compared to QRF

Before comparing QRF to its alternatives, it is important to assess whether the neural networks proposed are actually able to learn steadily on various configurations. It is verified in Figure 2.21 that the training of CNN is robust to the variety of data configurations encountered in a cross-validation framework: training losses are similar on the 7 cross-validation datasets.
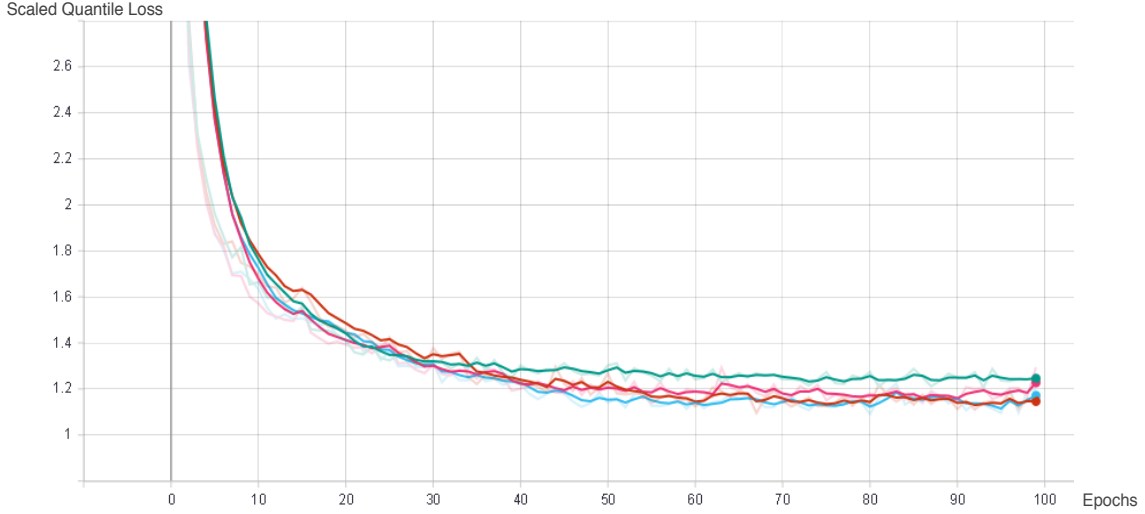
Figure 2.21: Training loss of CNN model for VPP1. Colors represent the different cross-validation datasets.

The main scores obtained on the forecast of the total aggregated production of VPP1 (PV+Wind+Hydro), comparing the results of QRF, CNN and LSTM models are reported in Table 2.3.

Model	parameters	CRPS	RMSE	Sharpness 80% interval
-	-	[-]	[-]	[-]
QRF	500 trees	0.033	0.065	0.250
FCNN	10 layers, relu, bn	0.035	0.055	0.244
CNN	Source-ranked plants, Conv8-16-32-64x4	0.033	0.058	0.159
CNN	Source-ranked plants, Conv16-16-32-64x4	0.031	0.052	0.152
CNN	Shuffled plants, Conv16-16-32-64x4	0.028	0.048	0.148
LSTM	2 stacked stateful layers + FC48x2	0.037	0.064	0.080

Table 2.3: Comparison of aggregated forecast models on VPP1, over all horizons. CNN: kernel size = (4,4), batch size = 16. LSTM: dropout = 0.2, batch size = 16

It is observed that neural networks exhibit a higher sharpness on the average of horizons compared to QRF, with the exception of the FCNN. The CNN reaches a sharpness of 14.8% to 15.9% depending on the selected configurations. However, the first configuration of CNN with 8 convolutional layers and a first layer comprising 8 filters, is outperforming QRF in terms of RMSE but has a similar CRPS. It is thought to stem from a better reliability of QRF in this case, which compen-

sates for its lower sharpness. However, increasing the number of filters in the first layer of CNN is sufficient to improve CRPS and overcome QRF. A grid search on the optimal size of convolutional filters for the CNN forecast of aggregated production restricted to wind plants is reported in Figure 2.22. The model is found to perform best with an average size for the first kernel (4 pixels x 4 pixels). One notes also that the number of filters of the first convolutional layer giving the best result is 8, whereas it is equal to 16 in the forecast of the total VPP aggregation. This may indicate the need to increase filter dimensions according to the problem size.

The probabilistic performance of the CNN is ensured over the whole horizon range, as shown in Figure 2.23. The CRPS of CNN models is lower than the CRPS of both QRF and FCNN, which have similar performances on VPP1 (see left panel). The configuration of inputs in the CNN model impacts significantly the probabilistic score: a CNN trained on shuffled plants reaches a better score than a CNN trained on source-ranked plants, on both the total production of VPP1 (left panel) and a reduced aggregation of VPP1 limited to PV and Wind plants (right panel).
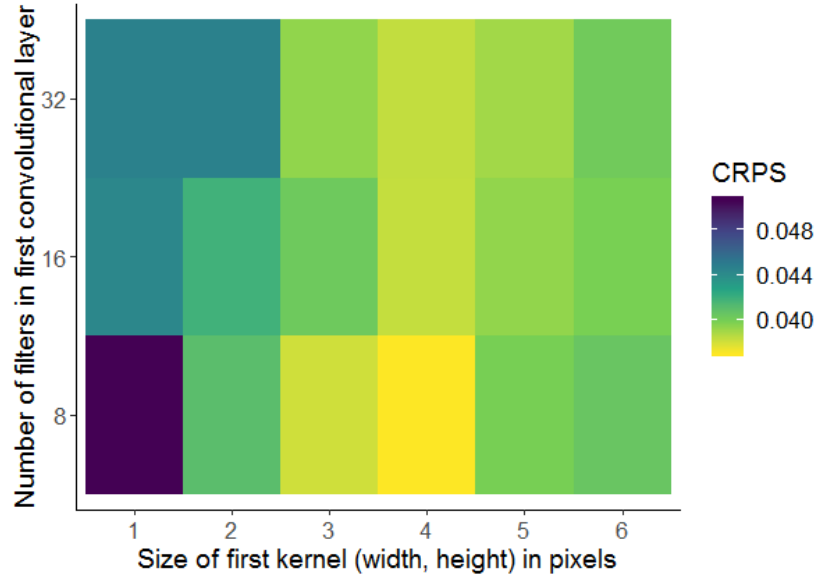
Figure 2.22: Grid search of CNN parameters: CRPS on wind aggregated production as a function of number of filters in first convolutional layer and size of first kernel. For reference, QRF has CRPS = 0.045
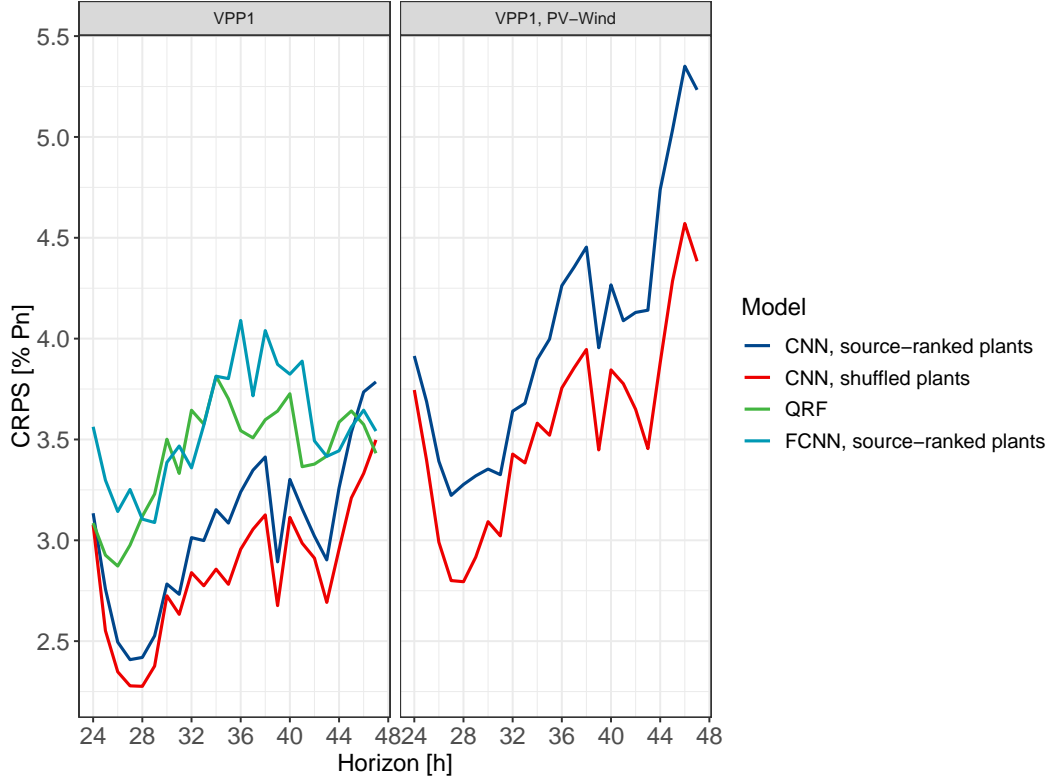
Figure 2.23: CRPS of aggregated production forecasts on VPP1 (PV+Wind+Hydro, left) and VPP1 restricted to PV and Wind plants (right)

2.5 Generate trajectories of aggregated VRE production

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 [100], produces scenarios with more realistic behavior with respect to the real production patterns.

Originality of the approach

Generating trajectories of renewable production on multiple sites has been proposed by [66] in the context of PV production. This work proposed to fit a multivariate copula on density probabilistic forecasts of each PV site. In the present work, a similar approach will be taken but with two main novel contributions: (1) marginal forecasts will be done on different energy sources within the VPP, instead of a single energy source, and (2) the obtained trajectories will be summed to obtain trajectories of aggregated production. Furthermore, these trajectories will be compared with trajectories obtained from direct probabilistic forecast of aggregated VRE production. The resulting approach is, to the best of the author's knowledge, innovative as it addresses directly the

challenge of aggregated multi-source VRE production trajectories.

2.5.1   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 [66]. 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 variable production, the forecasted distribution function $\hat{F}_Y$ of an variable 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 variable production forecasts, we apply to each forecast the distributional transform developed by Ruschendorf [123], which generalizes in (2.24) the property of uniform distribution to discontinuous CDFs.

$$U(y) = \hat{F}_{Y^-}(y) + V(\hat{F}_Y(y) - \hat{F}_{Y^-}(y)) \tag{2.24}$$

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 the multivariate copula from the transformed marginal distributions.

At this point, two distinct methods are proposed. The first method, entitled "Direct Gaussian" (DG), constructs a multivariate temporal dependence model between the variables Y_h^{total}, which represent the values taken by the total aggregated production at the successive horizons $h \in [1, H]$. First, a series of observed aggregated productions y_{t+h}^{total} is collected, not included in the training set of the forecasting model, for all horizons h. Second, the position of these observed productions in the forecast distribution $\hat{F}_{Y_{t+h}^{total}}$ at each horizon h is evaluated according to (2.24), and constitutes a realization u_{t+h}^{total} of the uniformly distributed marginal U_h^{total}.

$$u_{t+h}^{total} = U_h^{total}(y_{t+h}^{total}), \quad \forall t, \forall h \tag{2.25}$$

All marginals are then converted in (2.26) into normally distributed variables Z_h^{total} using the probit function Φ [65], forming a multivariate variable $\mathbf{Z^{total}}$ normally distributed with a zero mean vector and covariance Σ^{total} of dimension $H \times H$.

$$z_{t+h}^{total} = Z_h^{total}(y_{t+h}^{total}) = \Phi(u_{t+h}^{total}), \quad \forall t, \forall h \tag{2.26}$$

$$\mathbf{Z^{total}} = (Z_1^{total}, Z_2^{total}, ..., Z_h^{total}, ..., Z_H^{total}) \tag{2.27}$$

$$\mathbf{Z^{total}} \sim \mathcal{MVN}(0, \Sigma^{total}) \tag{2.28}$$

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. The generation of Ω distinct scenarios of total aggregated production for a given period of interest $[t + 1, t + H]$ follows the steps below:

1. Draw Ω i.i.d.random vectors $\mathbf{s}_\omega$ following the uniform distribution $U(0, 1)^H$, 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 (2.42) the quantile values given by each $\mathbf{z}_\omega$ to the marginal forecasts $\hat{F}_{Y^{total}_{t+h}}$:

$$\hat{y}^{total,DG}_{\omega,t+h} = \hat{F}^{-1}_{Y^{total}_{t+h}}(z^{total}_{\omega,t+h}) \quad \forall\omega, \forall t, \forall h \tag{2.29}$$

A second method is based on the separate forecasting by energy source instead of the direct forecasting of the total aggregated production. This approach models the dependencies between productions of the different energy sources, over all horizons. This is an extension of the methodology proposed by [66] to generate multivariate scenarios for multiple plants of the same energy source. The observed productions are now collected and aggregated for each energy source separately, resulting in an observation vector $\mathbf{y}^{sources}$ of dimension S, S being the number of sources (in the present case S=3 with Wind, PV and Hydro). In Figure 2.24, the principle of the method is represented. The position of production observations y into the forecast CDFs of each energy source $\hat{F}^{PV}_Y, \hat{F}^W_Y, \hat{F}^H_Y$ are recorded at all horizons (here two horizons t and $t+h$ are represented for simplicity). This forms a multivariate copula, with a specific density for each source-horizon pair, represented in the Figure by white-filled regions. Lastly, trajectories are generated by drawing randomly Note that this method can be extended straightforwardly to a multivariate copula between all plants of the aggregation (in this case the indices p, w, h in Figure 2.24 are associated to plants of the different sources).
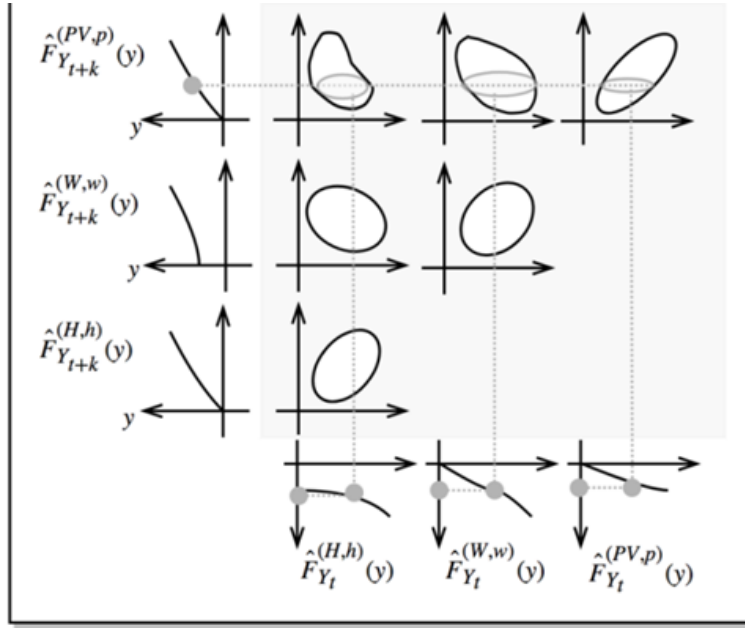
Figure 2.24: Principle of scenario generation from separate probabilistic forecasts by energy sources ('W' stands for Wind, 'H' stands for Hydro)

If the multivariate copula is considered to be Gaussian, then this method is called 'Indirect Gaussian (IG)'. The multivariate copula of the variable $\mathbf{Z}^{sources}$ is constructed from forecasts and

observed production following the same process as for the DF method, giving a covariance matrix $\Sigma^{sources}$ of dimension $SH \times SH$.

Scenarios are generated by sampling the covariance matrix and affecting the resulting quantiles $\mathbf{z}_\omega^{sources}$ of dimension SH 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 (2.30) to form Ω trajectories of total aggregated production.

$$\hat{y}_{\omega,t+h}^{total,IG} = \sum_{s=[1,S]} \hat{F}_{Y_{t+h}^s}^{-1}(\mathbf{z}_{\omega,t+h}^{sources,s}) \quad \forall \omega \in [1,\Omega], \forall t, \forall h \tag{2.30}$$

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 [55]). The generation of Ω scenarios from the Vine copula at horizon $t + h$ (omitted in notations below for the sake of simplicity) is executed as follows:

1. Draw Ω i.i.d random vectors $\mathbf{s}_\omega$ following the uniform distribution $U(0,1)^{SH}$

2. Retrieve the uniform marginal CDF value $m_{\omega,d}, d \in SH$ of the production variable Y_d, conditioned by the other variables, by inverting the h-function of the Vine copula [55]

$$m_{\omega,d} = \hat{F}_{d|d-1,...,1}^{-1}(\mathbf{s}_{\omega,d}|m_{\omega,d-1},...,m_{\omega,1}) \tag{2.31}$$

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

$$y_{\omega,d} = \hat{F}_{Y^d}^{-1}(z_{\omega,d}) \tag{2.32}$$

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.

2.5.2 Evaluation metrics

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) [126], to determine whether trajectories reproduce correctly the main moments of the original production [66]. The VS of order γ can be expressed in (2.33) 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 [126]. 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).

$$VS_t^{(\gamma)} = \sum_{i,j \in M} w_{ij}(|y_{t,i} - y_{t,j}|^{\gamma} - \frac{1}{\Omega} \sum_{\omega \in [1,\Omega]} (|\hat{y}_{\omega t,i} - \hat{y}_{\omega t,j}|^{\gamma}))^2 \tag{2.33}$$

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 [99]. The similarity of gradients observed in the original time series with gradients in scenarios is quantified by means of a *Brier Score* (BS) defined in (2.35). The events considered in the score are production gradients δ_t over an interval Δt, which are higher than a threshold value r_{BS}, taken as the average observed gradient over the interval.

$$\delta_t(y; \Delta t) = \mathbf{1}(|y_{t+\Delta t} - y_t| \geq r_{BS}) \tag{2.34}$$

$$BS = \frac{1}{N} \sum_{t=1}^{N} (\frac{1}{\Omega} \sum_{\omega \in [1,\Omega]} \delta_t(\hat{y}_\omega; \Delta t) - \delta_t(y; \Delta t))^2 \tag{2.35}$$

2.5.3 Case Study

The proposed methodology for the generation of scenarios is evaluated on the basis of the day-ahead VPP production forecasts presented in Section 2.4.5. Recall that the original VPP (VPP1) is composed of 3 Wind farms, 3 small Hydro plants and 9 PV farms, all located in France within the same control area but with different climates. A variant (VPP2) has modified installed capacities to be dominated by PV instead of Wind in the case of VPP1. The evaluation covers 9 months at a 30-min resolution in a cross-validation by weekdays.

2.5.4 Results

The covariance matrix obtained from separate forecasts on each energy source shows in Figure **??** 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 between Wind and PV production are in accordance. The fact that negative correlations occur is interesting for the provision of balancing services: 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), thereby potentially maintaining the capacity of the VPP to provide reserve. These negative correlations between PV and Wind are in accordance with the findings of [107], albeit those stand for aggregated production on a much larger size (Northern Europe).

An example of the trajectories obtained with the direct and indirect approaches is shown for VPP1 (cf. 2.4.5) during 3 winter days in Figure 2.26. The trajectories of VPP production obtained by the direct approach DG (cf. top graph in Figure 2.26a) correctly reproduce the production pattern, but display higher amplitudes than trajectories obtained with the indirect approach IG
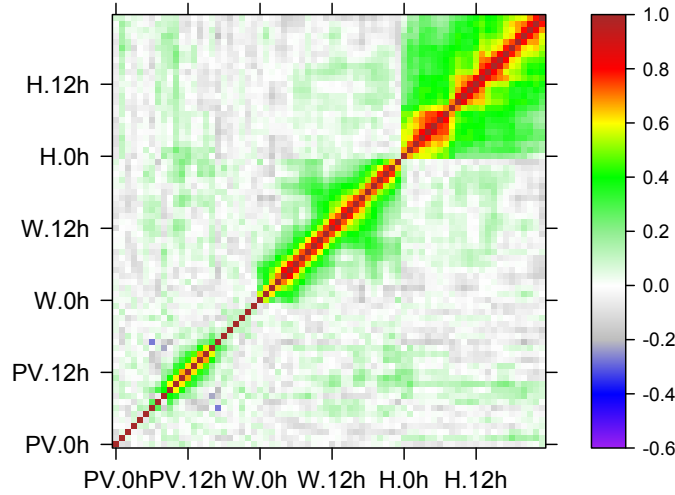
Figure 2.25: Covariance matrix of the multivariate Gaussian Copula, in the rank domain of Wind (W), Hydro (H) and PV (PV) for each hour of the day

(cf. bottom graph in Figure 2.26a). Figure 2.26b shows the breakdown by energy source of the trajectories obtained by the IG approach. In those 3 days, the VPP production is mostly covered by Wind, with overcast conditions on PV sites and low production levels for Hydro (installed capacity of Hydro is 12 MW). Trajectories of Wind production evolve with relatively narrow levels of amplitude. This suggests that the forecasting model of Wind production is able to issue a sharper prediction than the models of PV and Hydro. Additionally, the frequent ramps observed in PV and Hydro indicate that the scenario generation model outputs an expected high temporal variability. Wind trajectories show a bias at high production regimes in those particular days, this is thought to come from plants not operating at full capacity (unvailability of some turbines). The high PV trajectories relative to the PV production observed in the third day originate from a significant overestimation of PV production by the QRF forecasting model on those dates. Finally, the trajectories of Hydro production seem noisy compared to the smoothed observed production level. The variations mimic the observed frequent ramps in some of the hydro plants installed in narrow rivers near mountains, and the range of power levels is inline with the observed daily variation of production. Overall, the sharpness of Wind forecasts appears to dominate over the other factors in the resulting trajectories of VPP production.

Another vizualisation of obtained trajectories is given in Figure 2.27 over a sequence of 5 days. In this Figure, 3 main observations can be made:

1. On the first day, the VPP production pattern is dominated by a sharp decrease of wind production. This situation is correctly anticipated by the prediction models in both approaches.

2. Between the second and the third day, trajectories from separate forecasts exhibit a variability sensibly higher than the observed VPP production. It is an evidence that trajectories from

separate forecasts ignore true VPP ramps and therefore tend to be less realistic regarding ramps than trajectories from direct forecast of the aggregated VPP production.

3. On the fourth day, PV becomes the dominant source in the VPP because of low wind conditions. In this situation, the DG approach is unaware of possible saturation of PV production, therefore DG trajectories overestimate the contribution of PV to the real VPP production pattern.

In the next paragraphs, properties and scores of the trajectories enable to distinguish the outcomes over the whole evaluation period. Before analyzing scores, we evaluate the capacity of trajectories to model correctly the temporal correlations in the VPP production signal. The auto-correlation of 10 trajectories, reduced from the 100 generated trajectories by fast-forward selection on the cumulated absolute deviation to production on all horizons [127], is represented for both DG and IG methods in Figure 2.28. Trajectories obtained with the IG method (dark blue) are closer to the auto-correlation of the aggregated production of the VPP (yellow) than trajectories from the DG method (light blue), especially on the first 48 hours.
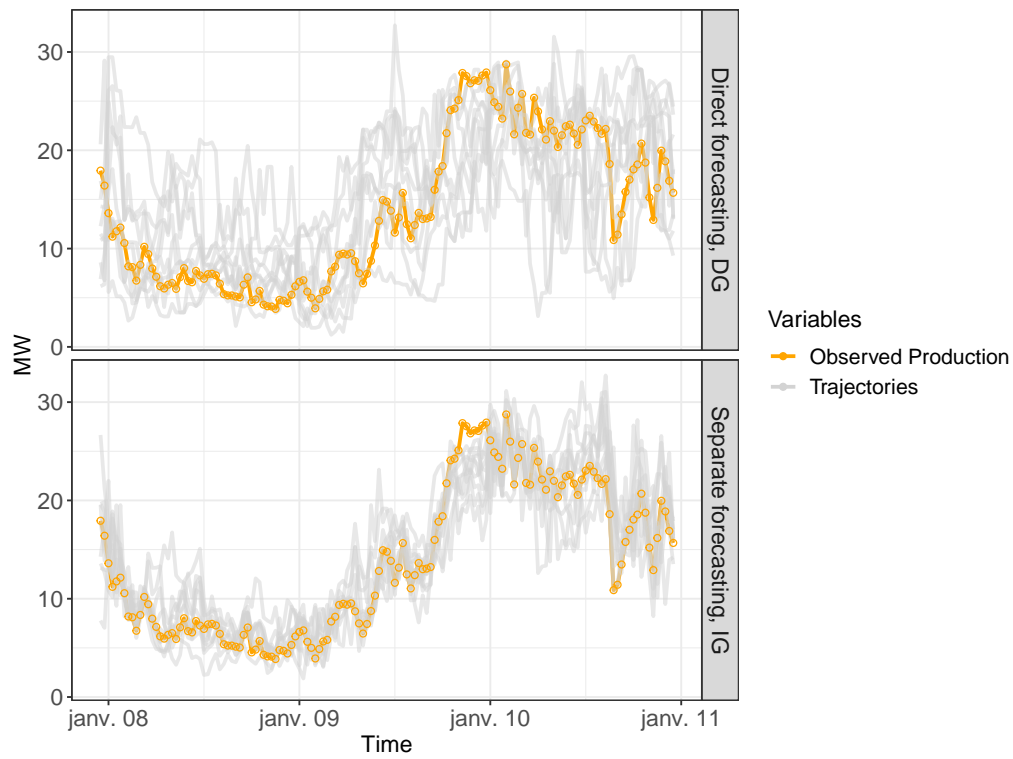
(a) Trajectories of VPP production obtained by approach DG (top) and approach IG (bottom), VPP1.
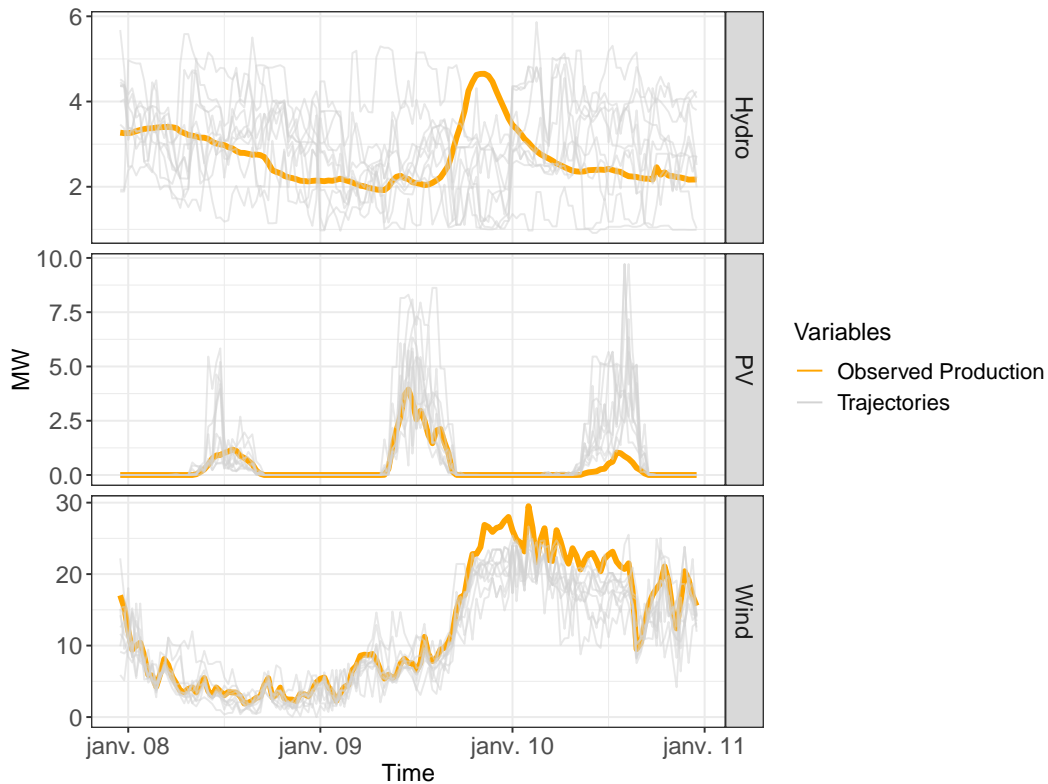
(b) Trajectories of production by energy source obtained by the approach IG

Figure 2.26: Trajectories of aggregated production and production by energy source (20 trajectories randomly selected out of 100), along with observed production of VPP1, on 3 winter days.
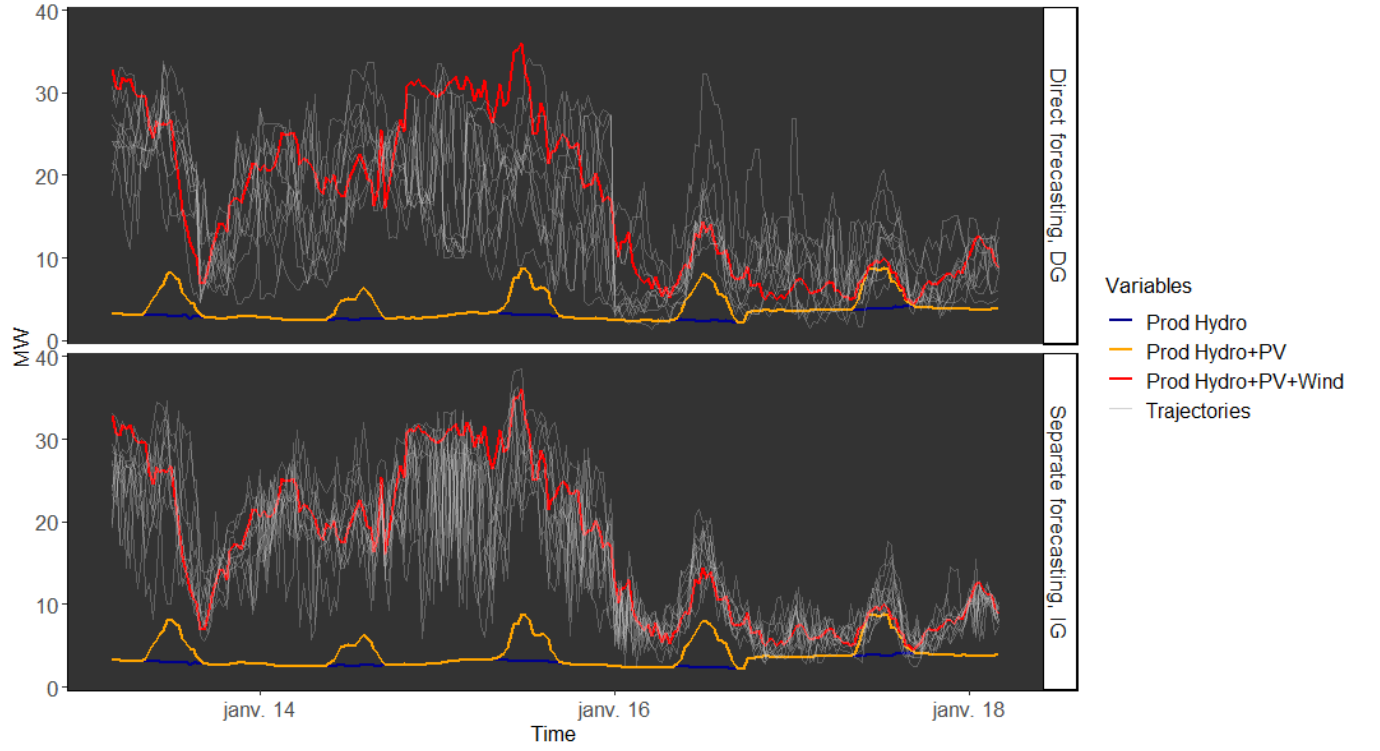
Figure 2.27: Trajectories of VPP production with approach DG (top) and IG (bottom) on a second period of 5 days
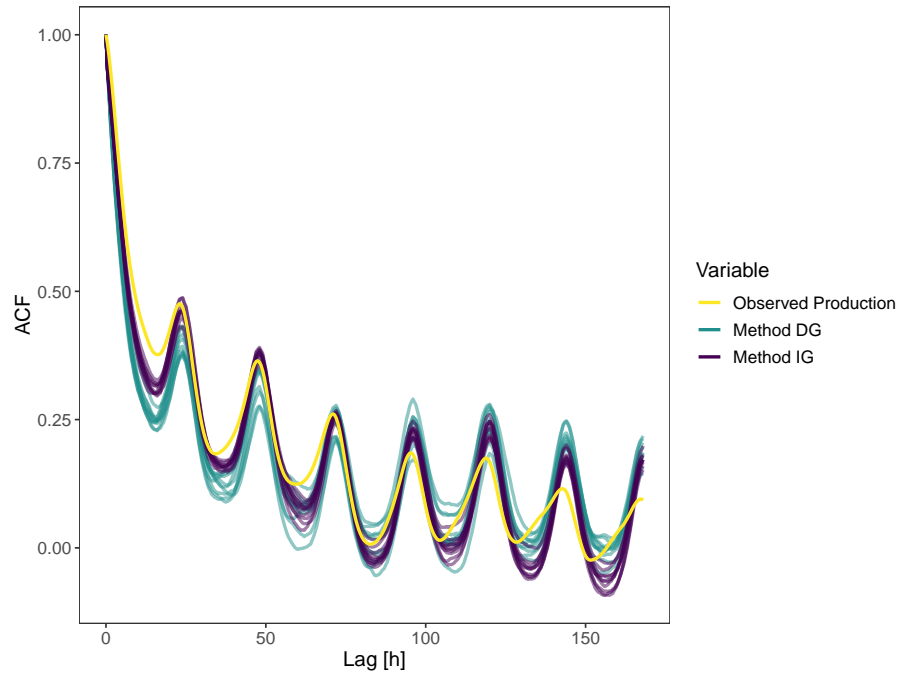
Figure 2.28: ACF of production and reduced trajectories generated with approaches DG and IG, VPP1

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 2.29 (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. 2.2), 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 occurring for each source, which creates an underestimation of the smoothing effect.

Concerning gradients of aggregated production, the corresponding BS are reported in Figure 2.29 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 2.30 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).
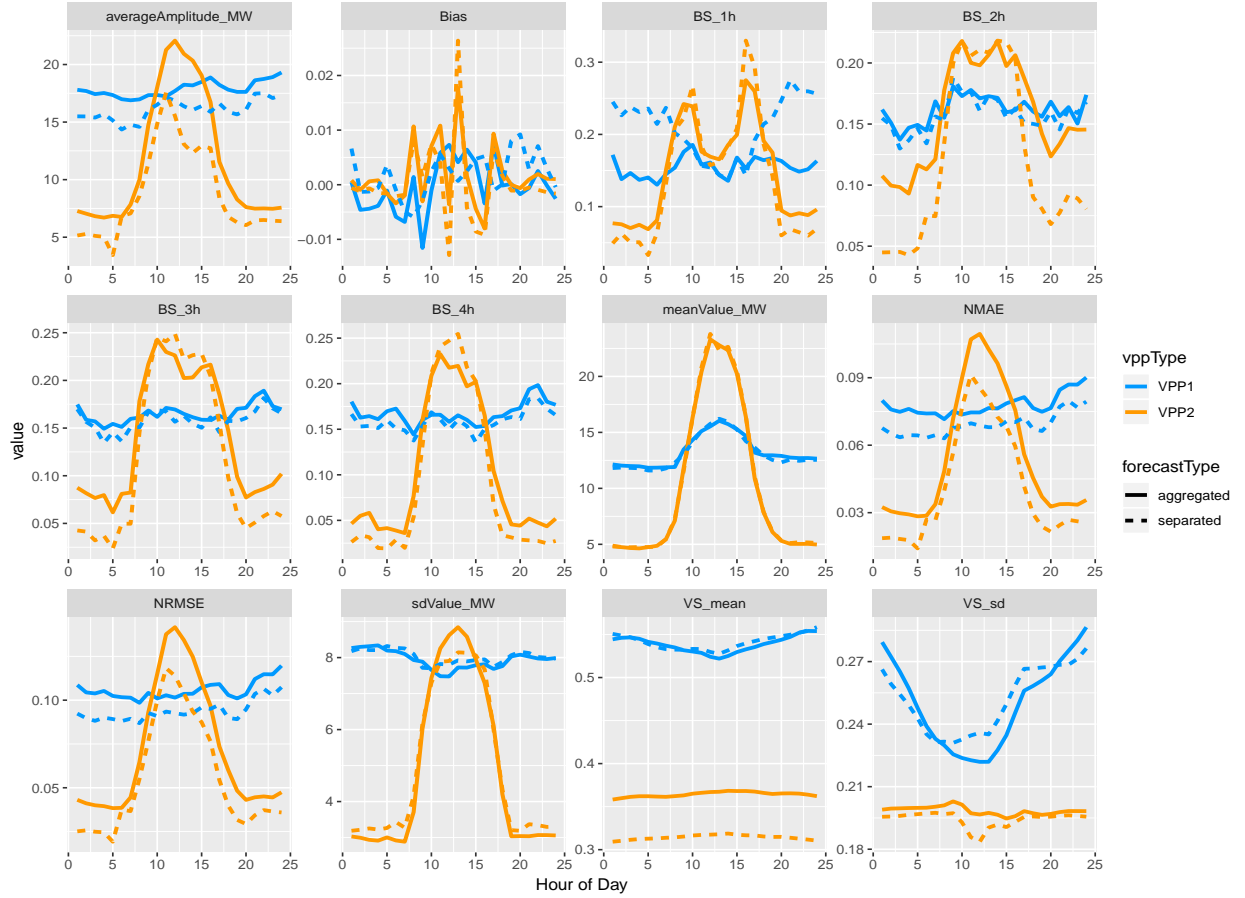
Figure 2.29: Properties and scores of scenarios generated with Gaussian Copula, directly aggregated from the DG method ("aggregated", solid line) and indirectly aggregated by separate production forecast from the IG method ("separated", dashed-line). Starting at top-left: average Amplitude of the scenario 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
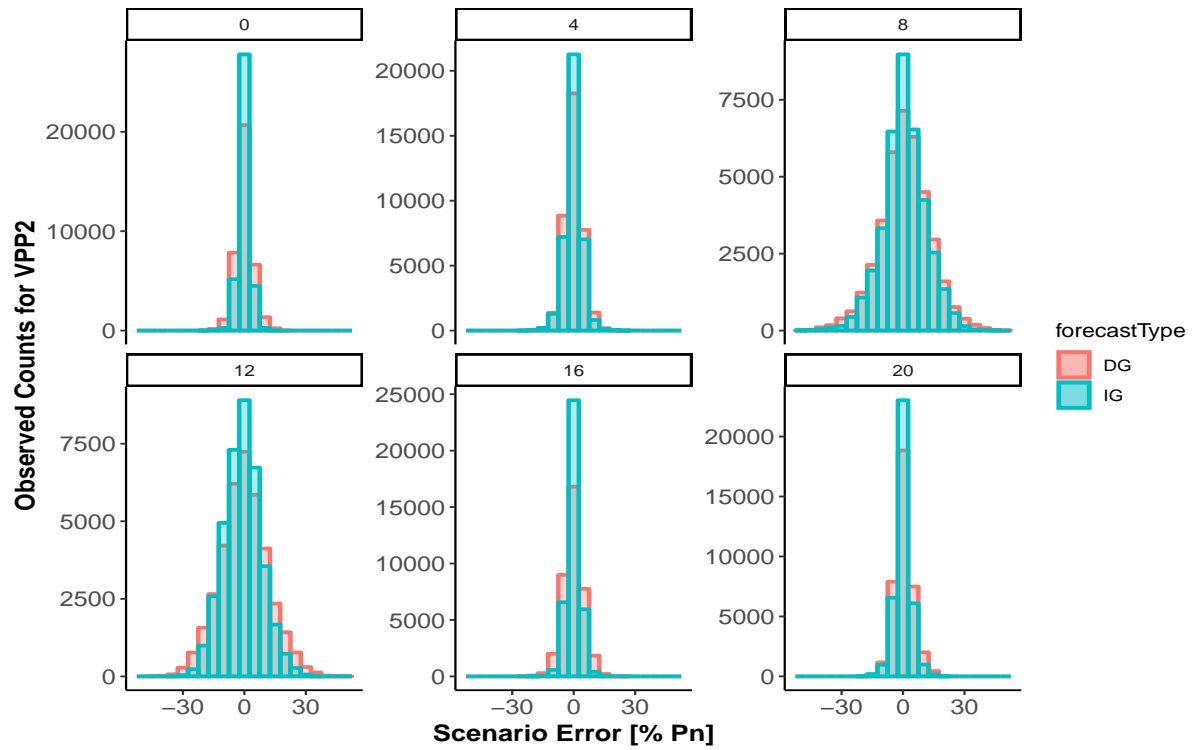
Figure 2.30: Distribution of scenario error in MW for VPP2, for selected horizons between 0h and 20h. *DG*: Direct forecast with Gaussian copula; *IG*: Indirect forecast separate by energy sources with Gaussian Copula

2.6 Conclusion

A methodology for the direct probabilistic forecast of aggregated VRE production has been proposed in Section 2.4. It appears from the case studies that decision-tree based such as Quantile Regression Forests (QRF) provide adequate reliability over a large quantile range [1%-99%], even in an operational context. Data augmentation technique, where features are supplied to the model with temporal lags and points on the neighbouring geographical area, improves performance. When trained over the whole horizon at once, QRF outperforms Gradient Boosting Trees (GBT) in terms of global probabilistic forecasting score (quantile score), because it can build relatively profound trees which accomodate well with a multi-source context, whereas GBT will learn on a reduced set of most influential features.

Considering decision-tree models as a reference, this work proposes regression models based on neural networks as an alternative. Besides the conventional fully-connected neural network (FCNN), The convolutional network is built by stacking features belonging to each power plant into an image, where plants are either randomly placed or ranked contiguously by energy source. Horizons form the channels of the network. Recurrent networks are based on the Long Short Term Memory (LSTM) cell, with stacked layers and stateful property, meaning that memory states learned on a previous batch are fed as the initial memory state of the next batch. The main results are shown in Figure 2.31: the obtained CRPS, RMSE and sharpness at 80% interval are scaled with results of the QRF as a reference. The three CNN variants in blue perfom better than QRF with up to 15% reduction of CRPS, due to a reduction in errors visible in the RMSE but also an improvement of around 40% in sharpness. The FCNN has a better RMSE than QRF but a worse CRPS and a similar sharpness, this illustrates the difficulty of deriving precise probabilistic forecasts with FCNN in the context of multi-source aggregated VRE production. Lastly, the LSTM is competitive in terms of CRPS but its very high sharpness and worst CRPS score indicate a lack of ability in capturing correctly the uncertainty of aggregated production.

In conclusion on the direct forecasting of aggregated VRE production, the recommended model is the CNN regression, if the computational and tuning effort can be afforded. QRF remains an easier solution with competitive results for practical applications. An alternative formulation of the loss function, adding for instance an auxiliary loss for reliability in addition to the quantile regression loss, may remediate to the excessive sharpness of LSTM predictions observed in the case study.

Scenarios of aggregated VRE production are generated from probabilistic forecasts and a multivariate copula dependence model in Section 2.5. The proposed methodology compares two alternative methods for generating scenarios. The first method relies on the direct univariate forecast of aggregated production, which is combined with a temporal dependency model.The second method builds independent separate forecasts at lower aggregation levels, which are combined in a further step with a dependency model by energy source and horizon and summed in a last step to form trajectories of aggregated production.

The ability of trajectories to reproduce the temporal variability observed in the aggregated production signal has been evaluated by means of the Variogram Score for a hindsight on global performance and of the Brier Score for the reproduction of ramps. Forecasts have been issued
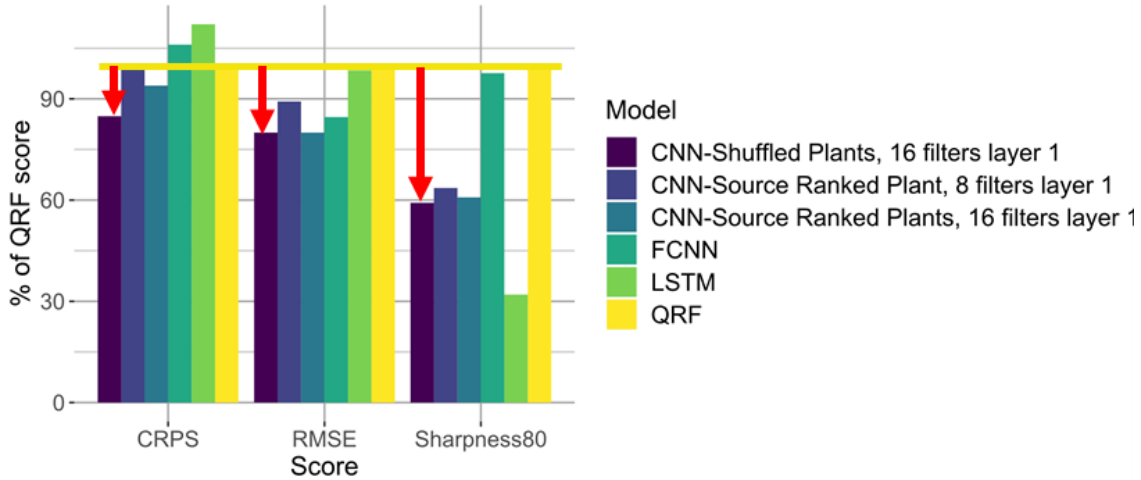
Figure 2.31: Summary of results for the direct forecast of aggregated VRE production, in terms of CRPS, RMSE and Sharpness at 80% interval, scaled by the QRF scores as a reference. VPP1.

with QRF, which is not the best performing model but is competitive and easy to deploy in an industrial context. Regardless of the capacity shares for the energy sources in the VPP, we observe that scenarios from separate forecasts have lower instantaneous errors (NMAE, NRMSE) than scenarios from direct aggregated forecasts, although forecasting at source level performs worst than forecasting at aggregated level. This is due to the higher amount of information on the structure of temporal correlation available in the covariance built from separate forecasts. In addition, the direct forecast ignores the marginal saturation of each plant in high wind/sun conditions, and therefore tends to underestimate the smoothing effect of aggregation. In the case of a VPP with a highly horizon-dependent production profile, such as a VPP with high PV capacity, the global variability is better reproduced with scenarios issued from forecasts by source because they distinguish more efficiently the different production regimes. In contrast, during the day the direct forecast models better multi-source ramps which are ignored by the separate forecasts by source. This comparative assessment is summarized in Table 2.4. In summary, the method with separate forecasts for each energy source produces more realistic trajectories, and should therefore have priority for stochastic applications.

VPP type	Instantaneous errors	Ramps	Global variability
high horizon dependence	separate	separate or direct	separate or direct
low horizon dependence	separate	separate (at night)	separate

Table 2.4: Summary of best method (direct aggregated forecast / separate forecast) for the generation of aggregated VRE production scenarios according to VPP type and scenarios properties

2.7 Résumé en français

Introduction

L'opérateur d'une agrégation de centrales variablees doit disposer de prévisions de production agrégée qui modélisent l'incertitude de cette production, afin de décider quelle part du productible affecter à la réserve. Cette thèse s'intéresse aux marchés d'énergie et de réserve dont la clôture s'effectue le jour précédant la livraison (aussi appelés marchés day-ahead). Ainsi l'horizon des prévisions utiles aux prises de position sur ces marchés est compris entre 12h et 48h. Les caractéristiques attendues pour une prévision performante sont les suivantes:

- fiabilité : une prévision fiable quantifie avec une erreur minimale la probabilité d'observer une production inférieure à un quantile prévu.

- finesse : une prévision fine minimise l'écart entre les quantiles prévus, de façon à concentrer l'information utile à la décision.

- corrélations temporelles réalistes : le processus de production agrégée variablee multi-sources démontre un effet de lissage atténuant les variations temporelles propres à chaque centrale prise individuellement, et mélange les caractéristiques propres à chaque source d'énergie.

L'objectif de ce chapitre est de proposer la prévision probabiliste la plus performante vis-à-vis de ces trois critères. Une première proposition de ce travail consiste à prévoir directement la production agrégée à partir des informations disponibles sur l'ensemble des centrales de l'agrégation, afin de limiter les modélisations intermédiaires entre centrales et agrégation, dont les approximations successives peuvent faire perdre de la précision. Ceci répond à la Question de Recherche 1. Les trois paragraphes suivants synthétisent les modèles de prévision directe proposés: modèles de *machine learning* basés sur les arbres de décision, puis réseaux de neurones convolutifs et récurrents, dont les architectures paraissent adaptées à la régression d'une production variablee multi-source.

Il reste ensuite à proposer une méthode pour reproduire les corrélations temporelles de la production agrégée. Ceci permet d'anticiper les variations de production et ainsi d'offrir de la réserve en se prémunissant contre les pénalités liées à la réserve non fournie en cas de productible insuffisant. Une méthode de génération de trajectoires de production agrégée multi-source est présentée, et l'évaluation de cette méthode clôt le chapitre. Cette méthode et son évaluation répondent à la Question de Recherche 3.

Prévision directe de production variablee agrégée par modèles basés sur les arbres de décision

La première proposition de modèle de prévision agrégée consiste à effectuer la régression directe de la production de l'ensemble des centrales du VPP à partir des informations de chaque centrale. Les modèles basés sur les arbres de décision, Quantile Regression Forests (QRF) et Gradient Boosting Trees (GBT), sont adaptés au problème car ils permettent de traiter des données X pouvant rassembler plusieurs centaines de variables, sans être trop impactés par le niveau de corrélation entre ces variables. Toutefois ces modèles n'incorporent pas de structure spatiale ou temporelle,

ils gagnent donc à disposer d'une vision étendue sur le processus de production. Ces modèles sont approvisionnés en données *augmentées* en temps et en espace, c'est-à-dire que nous ajoutons aux données initiales de chaque centrale p de la centrale virtuelle à l'horizon h les données relatives à des horizons voisins $h-1$ et $h+1$, et nous ajoutons des informations sur la distribution des conditions météorologiques prévues sur la grille entourant le site de chaque centrale (ces informations sont synthétisées par l'écart-type et la kurtosis).

$$\hat{X}^p = [X_h^p, X_{h+1}^p, X_{h-1}^p, X_{h,sd}^p, X_{h,skew}^p, X_{h,kurto}^p] \tag{2.36}$$

Prévision directe de production variablee agrégée par réseaux convolutifs

Les réseaux de neurones convolutifs ou *Convolutional Neural Networks* (CNN), sont capables de filter hiérarchiquement une information spatiale (par exemple des images), et ce sur plusieurs canaux portant chacun une information propre (par exemple la décomposition d'une image en niveaux de rouge, vert, bleu). Les neurones du canal c au sein d'une couche $h^{(l)}$ sont obtenus par convolution du filtre $f^{(l)}$ avec la couche inférieure $h^{(l-1)}$, corrigé par un biais $b^{(l)}$.

$$h^{(l)} = h^{(l-1)} \circledast f^{(l)} + b^{(l)} \tag{2.37}$$

On paramétrise le réseau en définissant la taille de la fenêtre vue par le filtre, le nombre de filtres et éventuellement des couches intermédiaires (comme la réduction de dimension par *pooling*, plus utile en classification). Ce type de modèle apprend avec moins de neurones qu'un réseau multi-couches standard et est capable de repérer des motifs invariants (par exemple par translation).

Cette structure est adaptée au problème de prévision agrégée multi-source en traitant les données par horizon (les horizons sont considérés comme autant de canaux), et en ordonnant les variables explicatives selon la centrale correspondante. Les données concernant les mêmes sources d'énergie sont placées soit côte à côte de façon à filtrer d'abord les centrales ayant le même type de profil de production avant d'apprendre la composition des sources dans l'agrégation, soit de façon aléatoire afin de filtrer d'abord les corrélations croisées entre sources avant d'apprendre l'effet de foisonnement par la taille de l'agrégation. Ces deux approches de présentation des données sont comparées dans un cas d'étude. La structure globale du réseau convolutif est représentée en Figure 2.32.

Prévision directe de production variablee agrégée par réseaux récurrents

Les réseaux récurrents permettent de mémoriser de l'information à court et moyen terme à partir de séquences de données. Nous implémentons ici le modèle Long Short Term Memory (LSTM), que nous entraînons sur des séquences de 24 heures. L'utilisation de la technique *dropout*, masquant un certain nombre de noeuds pendant l'apprentissage, permet d'éviter le sur-apprentissage. Les cellules récurrentes peuvent être empilées pour augmenter le pouvoir d'abstraction. Enfin le réseau est configuré pour être *stateful*, c'est-à-dire que le dernier état mémorisé sur un batch de données peut être réutilisé pour le batch suivant. L'architecture du réseau LSTM est illustrée en Figure 2.33.
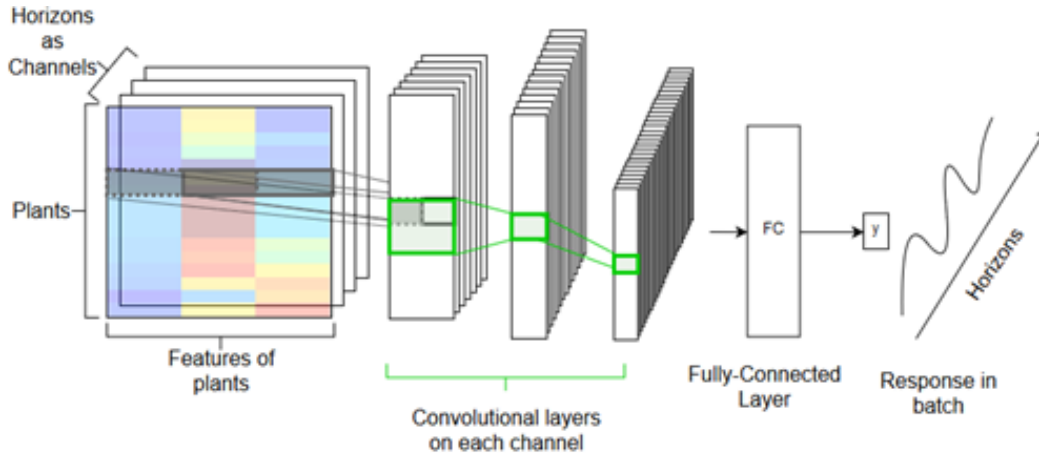
Figure 2.32: Structure d'un modèle CNN pour la prévision de production agrégée
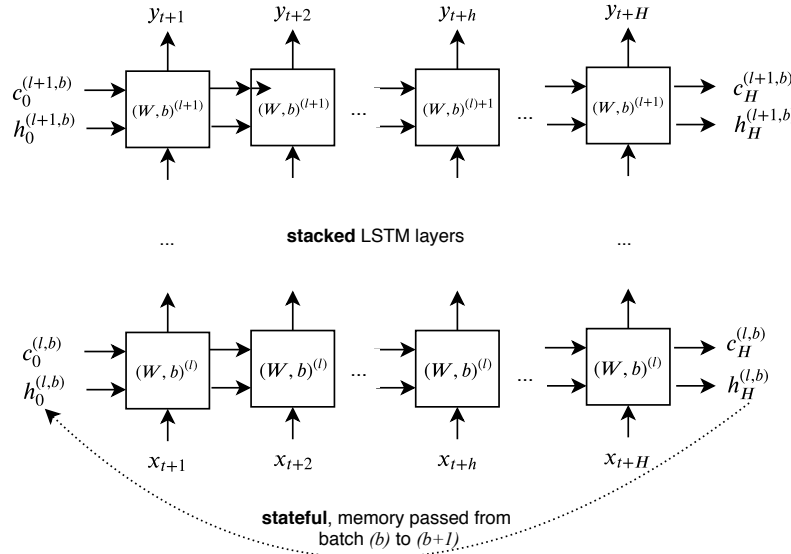
Figure 2.33: Architecture du modèle LSTM. b est l'indice de batch et l l'indice de couche

Cas d'étude pour la prévision directe de production variablee agrégée

Les performances des modèles QRF et GBT sont comparées sur un premier cas d'étude avec un VPP constitué de centrales PV et éoliennes situées en France métropolitaine, sous les climats atlantique et continental. La part de capacité PV est de 24%. L'apprentissage s'effectue de Septembre à Décembre 2015 et est évalué de Janvier à Mars 2016, avec une résolution temporelle de 30 minutes. Les prévisions météorologiques utilisées sont issues de la publication ECMWF en J-1 à 00h00. Les modèles sont entraînés sur l'ensemble des horizons, et le Continuous Ranked Probability Score (CRPS) montré en Figure 2.34 indique que le QRF surpasse le GBT ainsi qu'une régression linéaire (QRL). Le QRF est notamment meilleur pendant la nuit, ce qui démontre une plus grande
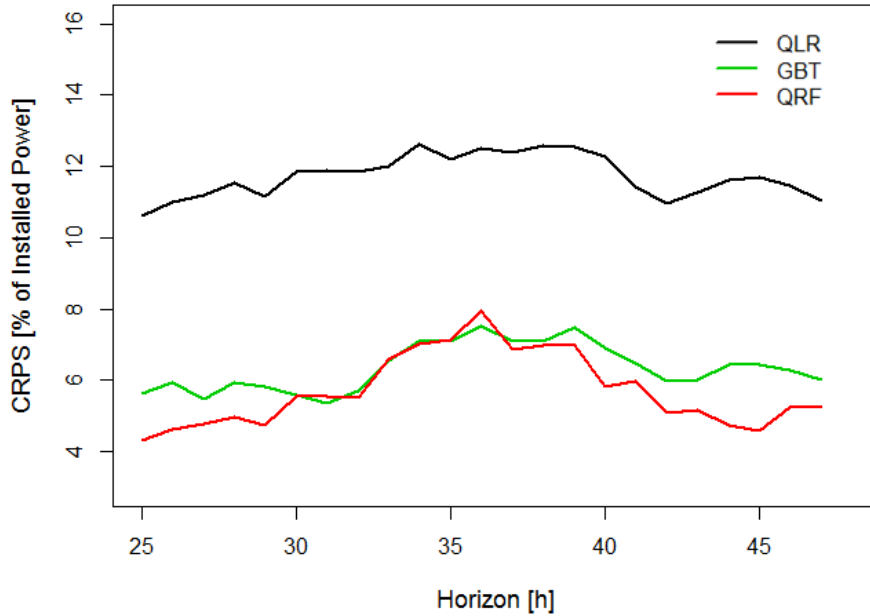
Figure 2.34: CRPS de la prévision agrégée en fonction de l'horizon sur un VPP éolien+PV obtenue par QRF, GBT et QLR

adaptabilité à la transition entre production éolienne seul et production multi-source éolien+PV. On remarque que l'erreur de prévision est plus grande lorsque le PV produit, et que son influence sur l'erreur est supérieure à son poids dans l'agrégation.

Un second cas d'étude évalue les capacités des réseaux convolutifs et récurrents à dépasser les performances du QRF. Les prévisions s'effectuent ici sur 9600 points de production mesurée entre Septembre 2015 et Mars 2016 à résolution 30 minutes, sur un VPP comprenant 15 centrales situées en France dont 3 centrales éoliennes, 3 centrales hydrauliques au fil de l'eau et 9 centrales PV.

La prévision s'effectue en validation croisée selon les jours de la semaine: les modèles sont entraînés sur 6 jours et évalués sur le 7ème. La moyenne des résultats obtenus par permutation du jour de test donne le score définitif de la prévision. Le QRF est entrainé sur des intervalles d'horizon de 4 heures afin de mieux séparer les différents régimes de production, et en particulier l'influence du PV. La prévision est également effectuée sur la production agrégée par source d'énergie. Pour ce faire, la production PV et les prévisions de rayonnement solaire sont normalisées par un profil de ciel clair analytique simple de type Top-of-Atmosphere.

La performance d'une prévision probabiliste peut se résumer à la combinaison de la fiabilité et de la finesse. On vérifie sur le diagramme de fiabilité en Figure 2.35 que les prévisions du modèle QRF sont fiables sur l'ensemble de la distribution prévue: les quantiles observés sont très proches des quantiles nominaux sur l'ensemble de l'intervalle 1%-99%.

Quant à la finesse, elle est comparée à celle des réseaux convolutifs et neuronaux en Table 2.5. Pour ces deux réseaux, le nombre de variables explicatives a été réduit par Analyse en Composantes
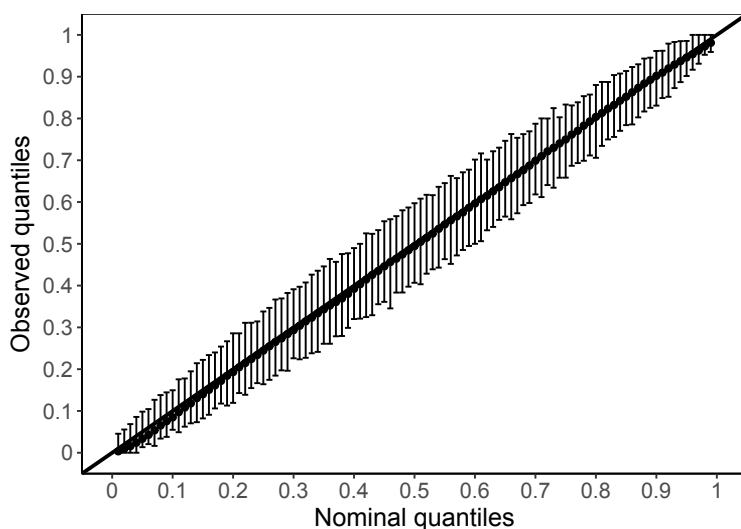
Figure 2.35: Diagramme de fiabilité de prévision agrégée par QRF sur VPP PV/Wind/Hydro. Les barres d'erreur indiquent l'intervalle de confiance à 90% d'une prévision parfaitement fiable obtenue par ré-échantillonnage [125]

Principales afin de faciliter l'apprentissage. En ce qui concerne la configuration des données des modèles CNN, le placement aléatoire des centrales dans le volume d'entrée donne de meilleurs résultats que le placement des centrales triées par source d'énergie, en termes de score global (CRPS de 2.8 % contre 3.1 %) et de score déterministe (RMSE de 4.8 % contre 5.2%).

A la lecture de la Table 2.5, il apparaît que les réseaux convolutifs, même de dimension modeste (8 couches cachées) dépassent le QRF en termes de finesse (de 15% à 25%) et de prévision deterministe (4.8 % contre 6.4 %). L'amélioration de CRPS est due à l'amélioration de la finesse. Le réseau récurrent simple testé (2 couches récurrentes, séquences de 24 h) atteint lui un RMSE similaire à celui du QRF, une finesse supérieure au QRF et au CNN mais un CRPS moins bon du fait d'une plus mauvaise fiabilité. Un meilleur paramétrage du réseau devrait néanmoins permettre de résoudre ce problème de fiabilité.

Modèle	Paramètres	CRPS	RMSE	Finesse
		[% Pmax]	[% Pmax]	[%]
QRF	500 arbres	3.3	6.5	25.0
CNN	16-32x2-64x4, k=4, FC200x2	2.8	4.8	14.8
LSTM	2 couches stateful empilées + FC48x2	3.7	6.4	8.1

Table 2.5: Résultats de la prévision directe de production agrégée multi-source. FC: couches fully-connected. knl: taille du noyau (kernel).

Génération de trajectoires de production variablee agrégée

Un agrégateur variable souhaitant participer à un marché de réserve doit pouvoir assurer son offre pendant toute la durée d'un intervalle de livraison. Cette durée varie selon les réglementations en vigueur (de 15 minutes à 1 semaine), elle est en tous les cas suffisante pour y observer une variabilité forte de la production variablee. Une formulation du problème en tant qu'optimisation stochastique permet de maximiser l'espérance de gain en tenant compte de la contrainte temporelle énoncée plus haut.

La résolution d'un problème d'optimisation stochastique nécessite la génération de scénarios caractérisant l'incertitude sur la production agrégée et reproduisant ses caractéristiques temporelles.

Les prévisions probabilistes effectuées par les réseaux récurrents reproduisent partiellement la corrélation entre les horizons, et l'on peut imaginer une architecture basée sur des réseaux profonds qui puisse générer un ensemble de trajectoires. Toutefois il est proposé ici d'améliorer la prévision probabiliste agrégée avec une méthode spécifiquement dédiée à la modélisation des dépendances temporelles entre les densités de probabilité produites à chaque horizon. Cette méthode consiste à définir une copule modélisant la probabilité jointe de la production agrégée sur l'ensemble des horizons à prévoir.

Deux variantes de cette méthode sont proposées, l'une basée sur la *prévision directe* de la production agrégée, l'autre sur la *prévision séparée* par niveaux inférieurs à l'agrégé, par exemple par source d'énergie. Les scénarios issus de prévisions séparées sont sommés afin d'obtenir des scénarios de production agrégée.

La méthode basée sur la prévision directe commence par placer les observations de production agrégée totale y^{total} dans l'espace des probabilités défini par la fonction de répartition prévue pour l'horizon h, U_h^{total}.

$$u_{t+h}^{total} = U_h^{total}(y_{t+h}^{total}), \quad \forall t, \forall h \tag{2.38}$$

Ces réalisations peuvent être converties en variables normalement distribuées, sur toute la plage d'horizon $h \in [1, H]$. On peut alors inférer une loi de probabilité jointe $\mathbf{Z^{total}}$ sur l'ensemble des horizons dont la fonction de répartition est décrite par une copule. En faisant l'hypothèse que les dépendances entre horizons suivent des lois gaussiennes, on peut générer des trajectoires en échantillonnant la copule définie par la covariance $\Sigma^{total} \in \mathbb{R}^{H.H}$.

$$\mathbf{Z^{total}} = (Z_1^{total}, Z_2^{total}, ..., Z_h^{total}, ..., Z_H^{total}) \tag{2.39}$$

$$\mathbf{Z^{total}} \sim \mathcal{MVN}(0, \Sigma^{total}) \tag{2.40}$$

Les réalisations dans l'espace des probabilités $z_{\omega,t+h}^{total}$ sont ensuite converties en trajectoires de production en utilisant la fonction quantile des prévisions de production agrégée $\hat{F}_{Y^{total}}^{-1}$.

$$\hat{y}_{\omega,t+h}^{total,DG} = \hat{F}_{Y_{t+h}^{total}}^{-1}(z_{\omega,t+h}^{total}) \quad \forall \omega, \forall t, \forall h \tag{2.41}$$

La méthode de prévision séparée se base sur le même principe, la matrice de covariance s'agrandit et devient de dimension $SH.SH$ où S est le nombre de sources d'énergie. Les trajectoires équiprobables obtenues par sources sont ensuite sommées sur l'espace des sources afin d'obtenir des trajectoires de production agrégée.

$$\hat{y}_{\omega,t+h}^{total,DG} = \hat{F}_{Y_{t+h}^{total}}^{-1}(z_{\omega,t+h}^{total}) \quad \forall\omega, \forall t, \forall h \tag{2.42}$$

Cas d'étude pour la génération de scénarios de production agrégée

La méthodologie de génération de scénarios est appliquée sur le même cas d'étude que celui pour la comparaison de prévision directe entre QRF, CNN et RNN. Les trajectoires générées sont équiprobables, et nous avons vérifié que 100 trajectoires étaient suffisantes pour une évaluation comparative robuste entre scénarios issus de la prévision directe et scénarios issus de la prévision séparée par sources. Les puissances installées sont modifiées afin d'obtenir un VPP où le PV domine en capacité.
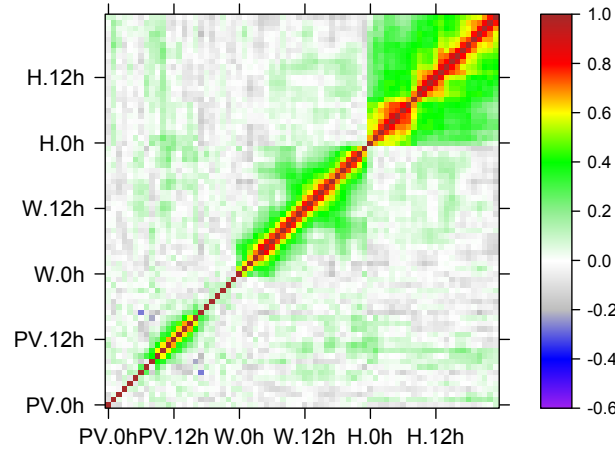
Figure 2.36: Matrice de covariance de la copule Gaussienne utilisée pour la génération de scénarios par prévisions séparées, dans le domaine de rang de l'éolien (W), de l'hydraulique (H) et du PV, pour chaque horizon

L'évaluation de la variabilité globale s'effectue à l'aide du Variogram Score et celle des rampes (durées de 1h à 4h) à l'aide du score de Brier.

Les résultats obtenus sont synthétisés en Table 2.6: les scénarios issus de prévision séparée sont plus performants que les scénarios issus de prévision directe de l'agrégé en termes d'erreurs instantanées (NRMSE, NMAE), de rampes et de variabilité globale. Ceci est dû à une plus grande variété de situations possiblement mobilisables par la copule, mais aussi au fait que l'approche directe les saturations marginales de chacune des sources. Les écarts sont plus marqués lorsque le VPP est dépendant de l'horizon (VPP à majorité PV): en effet dans ce cas la prévision séparée permet de modéliser plus précisément la variabilité de chacun des régimes.

Conclusion

La prévision directe de la production agrégée d'un VPP multi-source est réalisable à l'aide du modèle QRF avec des performances compétitives par rapport à d'autres modèles standards de

Type de VPP	Erreurs instantanées	Rampes	Variabilité globale
majorité éolien	séparée	directe ou séparée	directe
majorité PV	séparée	séparée	séparée

Table 2.6: Méthodes de génération de scénarios (prévision directe/prévision séparée) présentant les meilleurs scores selon le type de VPP et le type de propriétés recherchées

machine learning comme le GBT ou des approches statistiques classiques comme la régression quantile linéaire. Les modèles CNN et RNN permettent d'augmenter la finesse des prévisions par rapport au QRF. Le modèle RNN présenté n'est toutefois pas assez fiable pour être utilisé en l'état, sa fiabilité pourrait être améliorée à l'aide d'un objectif spécifique à intégrer à l'optimisation du modèle. La régression par CNN en revanche offre une formulation naturelle au problème de l'agrégation et bat le QRF sur l'ensemble des critères, au prix d'un effort de calcul plus important afin d'optimiser ses hyper-paramètres.

Une méthodologie de génération de trajectoires de production agrégée est proposée, qui se base sur la prévision probabiliste de la production. Pour ce faire, nous avons comparé deux approches:

- *Prévision directe de l'agrégé*: Prévision de la production agrégée totale, puis définition d'une copule modélisant la dépendance temporelle entre les prévisions à chaque horizon. L'échantillonnage de la copule produit des trajectoires cohérentes temporellement.

- *Prévision séparée par sources d'énergie*: Prévision de la production à plusieurs niveaux inférieurs de l'agrégation, par centrale ou par source d'énergie. Ensuite une copule modélise les dépendances spatiales (sources d'énergie) et temporelles (horizons). Les trajectoires obtenues pour chaque niveau inférieur (par exemple chaque source d'énergie) sont sommées afin d'obtenir des trajectoires de production agrégée.

Les scénarios issus de prévisions séparées par source reproduisent mieux la variabilité globale du signal de production agrégée, et ont des erreurs instantanées plus faibles. Ils devraient donc être utilisés en priorité dans des problèmes d'optimisation stochastique, comme l'offre de réserve formulée par un VPP variable qui sera présenté dans le Chapitre 4.

Chapter 3

Forecasting aggregated renewable production with high reliability

3.1 Introduction

Considering the vital importance of ancillary services for electrical networks, VRE-based reserve must be offered with a maximum level of reliability. In light of recent evolutions of AS pre-qualification tests in countries such as Germany where the VRE penetration is large, TSOs are likely to accept VRE-based reserve in a nearby future if and only if they can follow reserve setpoints with limited underfulfilments in terms of frequency and amplitude. Using the probabilistic forecasting methods presented in the previous Chapter, we are now able to propose reserve offers based on the lowest quantiles available through the direct regression model. Consider the illustrative case presented in Figure 3.1. Here, a reserve bid is chosen as the minimum value of the forecasted 1%-quantile given by the QRF prediction of the VPP production, over a validity period of 4 hours. Assuming that the model is reliable, which has been demonstrated in the previous Chapter, then the reliability of the reserve offer should be equal at least to 99%. But is this sufficient for TSOs?

In fact, the requisites formulated in [99] stipulate that the reserve capacity must be available during 100% of the validity period, and that 95% of the actual delivered reserve in response to TSO setpoints must be within an accepted deviation corridor. Even though these two rules may seem reasonable for a conventional power plant, they are less relevant for VRE plants for which deviations from setpoints may result of a lack of available active power instead of a poor power regulation. In this thesis, we choose to target a reliability of reserve as close as possible to the 100% availability requirement, so that it may be equivalent to the average availability of reserve provided by conventional dispatchable power plants. Therefore, the capacity of VPP forecasting models to provide reliability close to 100 % should be investigated.

A forecast that is at the same time useful and 100 % reliable is naturally unachievable. The **objective** of this Chapter consists more pragmatically in going beyond the reliability of previous forecasting models, and develop forecasting models of aggregated VRE production able to achieve more than 99% reliability. The Chapter as a whole answers to Research Question 4.

Existing studies on probabilistic forecast demonstrate a lack of performance of standard statis-
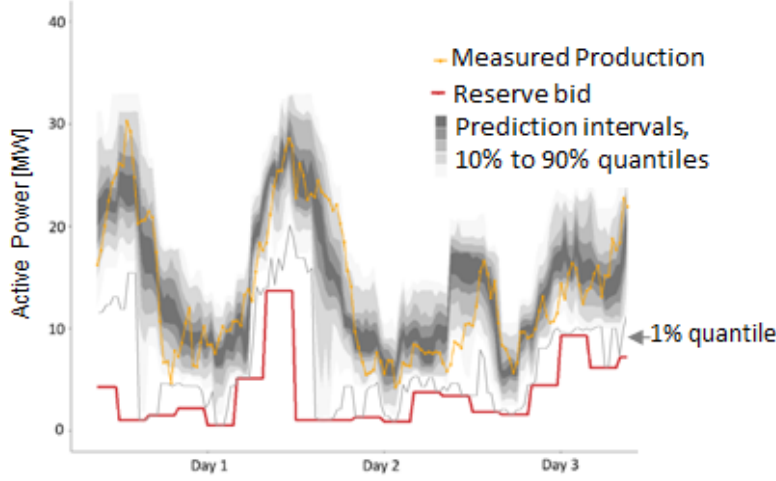
Figure 3.1: Reserve offer (in red) based on the 1%-quantile of a QRF prediction of total VPP production. The validity period of reserve is assumed to be 4 hours

tical models on low quantiles, e.g. quantiles below 1% (probability of occurence ≥ 99 %) (renewable production forecast [80], Dynamic Line Rating forecast [77], electricity price forecast [128]). This is due to the fact that minimizing a quantile loss with quantile nominal values close to zero leads to a poorly discriminative regression with state-of-the art regression models: the lowest observations are selected regardless of the values taken by the explanatory variables. Models such as QRF are indeed based on a shallow architecture: their inner hidden variables have few interactions, which results in a rather simple non-linear response. Consider that the depth of a tree in QRF is of the order of $\mathcal{O}(D)$ where D is the number of explanatory variables. In contrast, a Convolutional Neural Network, albeit a deep model with limited number of parameters, with L hidden layers each with F filters of dimension (m, n), and approximating the density of response with Q quantiles, has a total number of parameters of the order of $\mathcal{O}(LFmnQ)$, which easily exceeds $\mathcal{O}(D)$. Moreover, standard data-driven machine learning models can have difficulty in extrapolating extreme behaviors, which are per definition scarcely present in the learning data. This behaviour is verified with the QRF forecast presented in Section 2.4.5. The reliability diagram of forecasts obtained by QRF on quantiles comprised between 0.1% and 1%, presented in Figure 3.2, shows that QRF predictions are too narrow on the leftmost quantiles (inferior to 0.5 %). The observed frequencies are 2 to 3 times larger than the expected ideal behaviour.

Models developed for forecasting extremes of distributions, also frequently denominated *tails*, have been reviewed in 1.3.2. We observe that many of these models are based on the inference of parametric distributions. However there is a lack of references on how to model extremely reliable production levels for a multi-source aggregated VRE production. Given that the forecasting models developed in Chapter 2 on non-extremal quantiles do not guarantee reliability for quantiles below 1%, we propose a methodology for forecasting the left tail of aggregated VRE production, with enhanced attention to the selection of adequate parametric distributions. Neural networks that
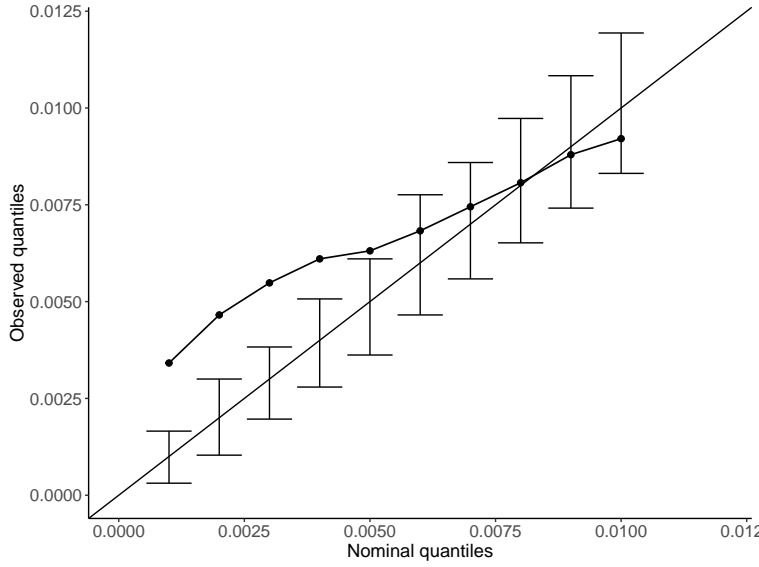
Figure 3.2: Reliability diagram of QRF forecast on low quantiles. The uncertainty bars represent the standard deviation of 1000 bootstraps of the forecast to quantify the uncertainty due to the sampling effect (see Section 3.8.1)

.

explicitly capture dependencies at various horizons such as Convolutional and Recurrent Neural Networks are interesting candidates for regression at extremely low quantiles. We will see that it can be valuable to couple non-parametric, non-linear regression approaches such as neural networks with distributions to propose a hybrid non-parametric/parametric model of extreme events.

An initial issue when forecasting extremely low quantiles is to know the minimal quantity of observations that is necessary to evaluate the forecasting reliability at low quantile levels with sufficient confidence. Let the natural quantile estimator $\hat{y}^\tau$ approximate the predicted τ-quantile of a random variable $Y \in \mathbb{R}^N$. The estimator converges in distribution towards a Normal distribution of asymptotic variance $\mathbb{V}(\hat{y}^\tau) = \frac{\tau(1-\tau)}{N.f^2(y^\tau)}$ [124]. The asymptotic prediction interval of order $1 - \alpha$ of the natural quantile estimator, defined in (3.1) depends on the standard deviation of a Binomial distribution with N trials and τ probability of success, scaled by the average density of the true underlying process.

$$PI_{natural}^\tau \sim \frac{\sqrt{\tau(1-\tau)}}{\sqrt{N}f(y^\tau)} \qquad (3.1)$$
$$\frac{PI_{natural}^\tau}{PI_{binomial}^\tau} = \frac{1}{f(y^\tau)}$$

If the prediction interval defined by the natural quantile estimator deviates significantly from the prediction interval defined by the Binomial distribution, then the evaluation of reliability is dependent on the distribution of the true process, which is unknown in practice. Figure 3.3 shows the ratio of prediction intervals obtained with the natural estimator and with Binomial variance only, simulating 50 bootstraps of the Normal distribution, as a function of th its number of observations. One can note that the size of prediction intervals differs visibly for quantiles below 10 observations

per quantile interval (e.g. quantiles below 10^{-3} for 10^4 observations). Considering that the usual amount of data available for evaluating forecasts is comprised between 10^4 points (1 year of yearly data) and 10^5 points at 5-min resolution), and that 10 observations per quantile interval are needed to reduce the influence of the effective distribution on reliability, the smallest quantile that can be evaluated with confidence that the outcome is representative of the underlying process is 0.1%.
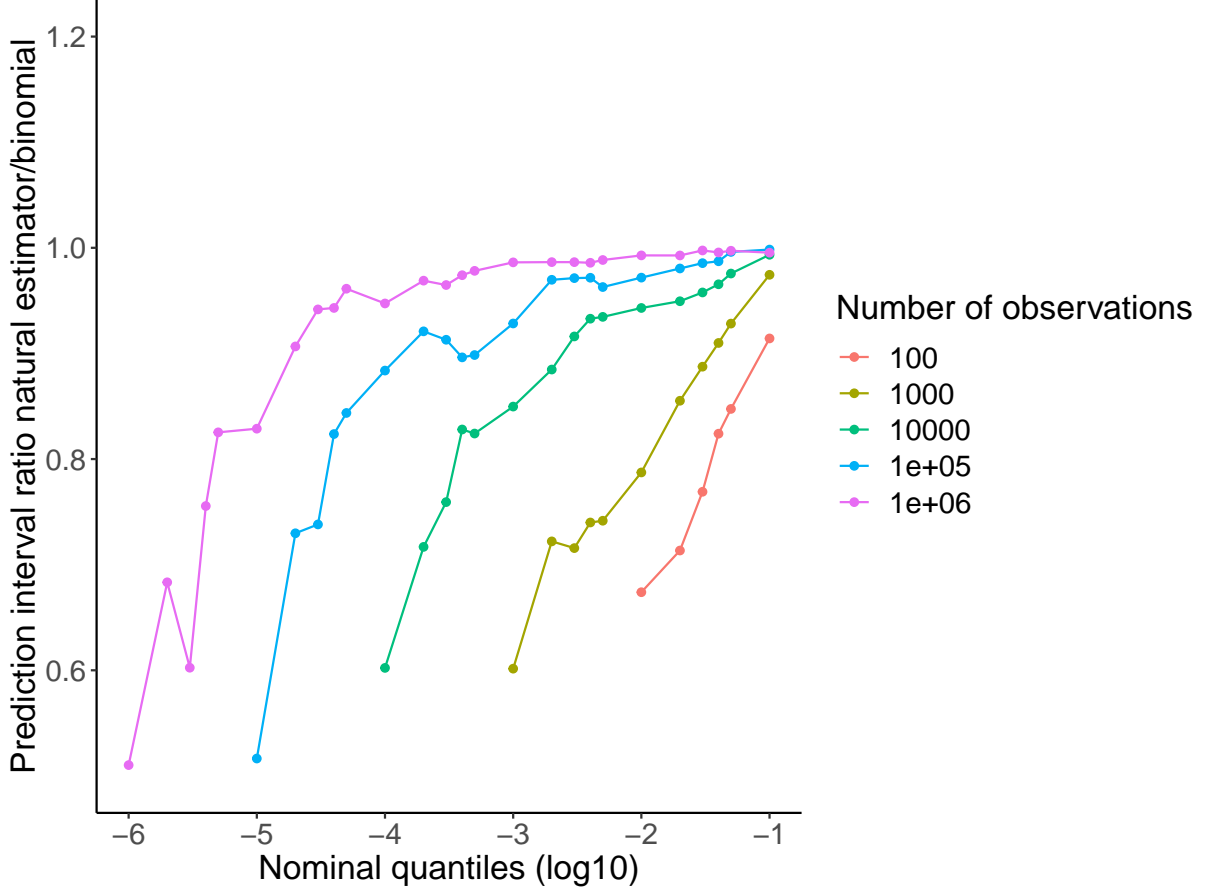
Figure 3.3: Ratio of prediction intervals obtained with natural quantile estimator and with natural quantile estimator

3.2   Methodology

In this Chapter, we propose several forecasting models for extremes of aggregated production covering extremely low quantiles within a [0.1%-0.9%] interval. These models aim at building an efficient estimator of the aggregated VRE production.
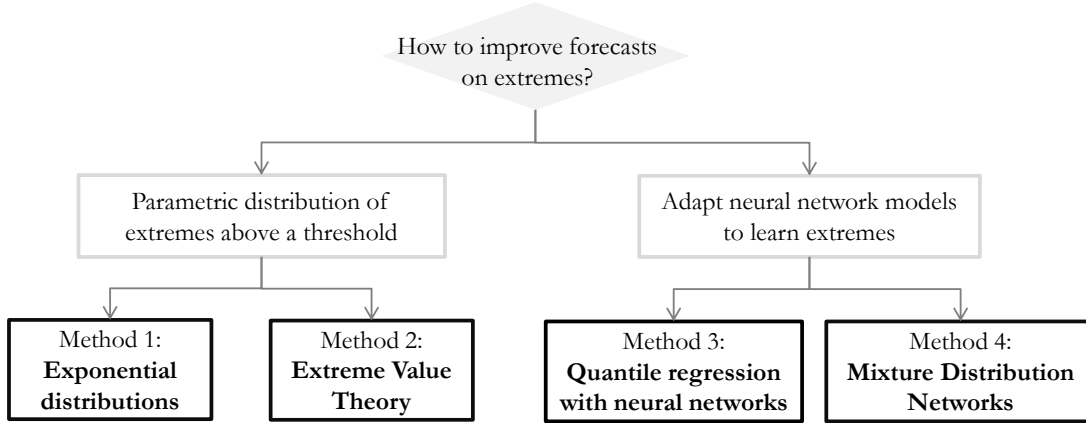
Figure 3.4: Overview of proposed models for forecasting extremely low quantiles

The proposed models are summarised in Figure 3.4. In Subsection 3.3, a first model is based on applying an exponential distribution on extremely low quantiles. This model is calibrated by the lowest reliable quantile forecast given by the models derived in Chapter 2. In addition, we formulate and verify the assumption that the distribution of extremely low production levels is dependent on the expected production regime. Therefore the parameter of the distribution is inferred on clusters of median aggregated production forecast.

Exponential distributions, albeit simple due to their definition through a unique rate parameter, can not represent a complex distribution of extremes, with *fat tails* (i.e. large intervals of low production levels with non-zero probability) or conversely sharper decays. In Section 3.4, we propose to improve the parametric model of extremes by replacing exponentials with Pareto distributions inferred from Extreme Value Theory (EVT). Instead of calibrating the model on the low quantile forecasts of a previous model, the EVT analyzes production levels exceeding a threshold, which is conditioned by median production forecasts, similar to the exponential model.

At day-ahead horizon, the distribution of extreme levels of a multi-source aggregated production is influenced by non-linear combinations of weather conditions over dispersed locations. The former presented models are able to conservatively anticipate extremes, but lack resolution due to their shallow structure. In Section 3.5, we propose neural networks architectures to perform directly quantile regression for extremes. Neural networks can also be used to fit a parametric distribution. It will be shown in Section 3.6 that mixtures of distributions are adequate to capture the extremely low quantiles of aggregated VRE production.

Originality of the approach

To the best of the author's knowledge, the approach consisting in tailoring several forecasting models for very low quantiles of a VRE-based VPP production is new. The comparison between relatively simple statistical models and more complex architectures containing mixtures and neural networks is also claimed to constitute an innovative contribution.

3.3 Method 1: Exponential distributions

A first approach consists in analyzing the deviation of production z from a quantile forecast of nominal value τ_{lim}, which is empirically given as the lowest bound of quantiles where a classical forecasting model such as QRF remains reliable.

$$z_t = [\hat{y}_{agg,t}^{(\tau_{lim})} - y_{agg,t}]^+, \quad \forall t$$

The analysis of deviations, illustrated in Figure 3.5, shows that an exponential distribution of parameter λ, defined below, may approximate the empirical distribution of deviations.

$$f(z, \lambda) = \lambda e^{-\lambda z}, z \geq 0$$
$$F(z, \lambda) = 1 - e^{-\lambda z}, y \geq 0$$
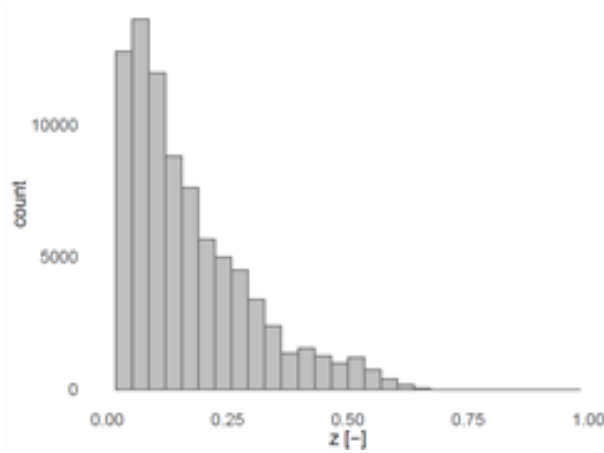
Figure 3.5: Deviations of production under a 1% quantile QRF forecast on VPP1

The rate parameter λ is then inferred by maximization of the likelihood. After straightforward analytical derivation, we obtain an estimator of the rate equal to the inverse of the average deviation.

$$\hat{\lambda} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{z_i} \tag{3.2}$$

It is known that forecasting errors can be characterized by the classification of production regimes, as demonstrated in [129] for wind power. The same idea led [80] to the classification of

VRE forecasting errors following clusters of median forecasts. The clustering is applied here as an extension of our previous methodology in (3.33): median production forecasts are partitioned into C quantile intervals, and the exceedances associated to each cluster give the rate λ_c corresponding to the VPP production regime.

$$\hat{\lambda}_c = \frac{1}{\frac{1}{N_c} \sum_{i=1}^{N} z_i \mathbb{1}_{\hat{y}_i^{(50\%)} \in I_c}} \forall c \in C \tag{3.3}$$

The CDF of QRF forecasts clustered by 10 equally-spaced production intervals on [0,1] are displayed in Figure 3.6. We observe that the dispersion of production levels associated to extremely low quantiles is narrow when the VPP forecast is low (cluster 1), and high when the VPP forecast is high (cluster 10). This fat-tail behaviour advocates for a model which individuates highly reliable production levels when the expected VPP production is above average.
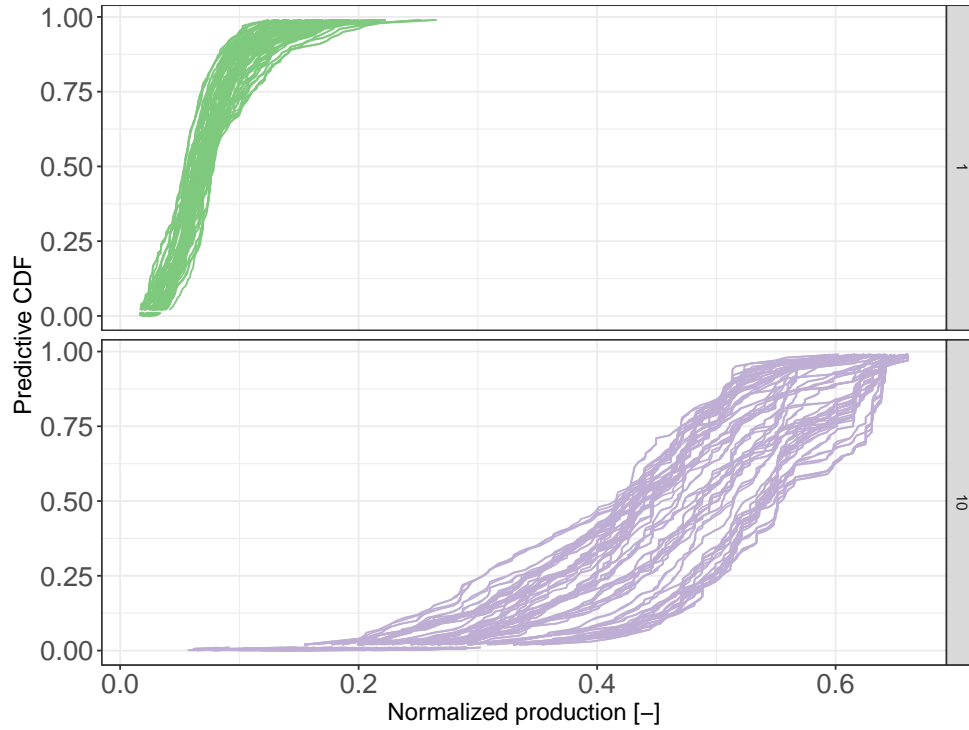
Figure 3.6: CDF of median forecasts for two clusters out of 10 interval-based clusters: lowest cluster (lowest production levels, top figure) and highest cluster (highest production levels, bottom figure), for VPP2.

Originality of the method

The idea of applying exponential models to tail forecasts is well spread in the domain of forecasting on renewables and electricity markets. The developments presented above for a VRE-based VPP apply developments made by [76] in the context of total RES production at national scale. The best configuration with respect to reference quantile will be investigated later, similarly to the work

of [77] on Dynamic Line Rating. However, Method 1 includes innovative content in its own right. Specifically, it *integrates a variant where the rate parameter is clusterized following a reduced set of weather conditions and median production forecasts of VPP production.* The augmentation of data used for tuning rates is thought to increase the selectivity of the model and its reliability *in fine.*

The choice of equally-spaced intervals of median production forecast for defining clusters of VPP production is not optimizing any criterion of similarity or density within clusters. Bad clustering leads to the assignment of false positives within clusters, and the overall result is a loss of information penalizing model inference. An unsupervised clustering algorithm such as k-means can partition expected conditions more robustly. A challenge of k-means is that it is not scalable to problems in high dimension, therefore we must classify observations according to a subset of features. The subset is determined empirically so as to characterize a VPP production regime and its low production levels: the min, mean, and max of weather variables across plants sharing the same energy source, and median VPP production forecast for a global insight on production regime.

$$x_{cluster} = ((min_{p \in [1,P]} x_{s,p}, mean_{p \in [1,P]} x_{s,p}, max_{p \in [1,P]} x_{s,p}, \quad \forall s \in \mathcal{S}), \hat{y}_{agg}^{(50\%)}) \tag{3.4}$$

The distribution of observed production values within clusters obtained by median forecast and k-means clustering are presented in Figure 3.7. The clustering by median forecasts (cf. Figure 3.7a) clearly partitions observations by production regime, while clustering by k-means

(a) Clusters by quantile intervals of median forecast

(b) Clusters by k-means
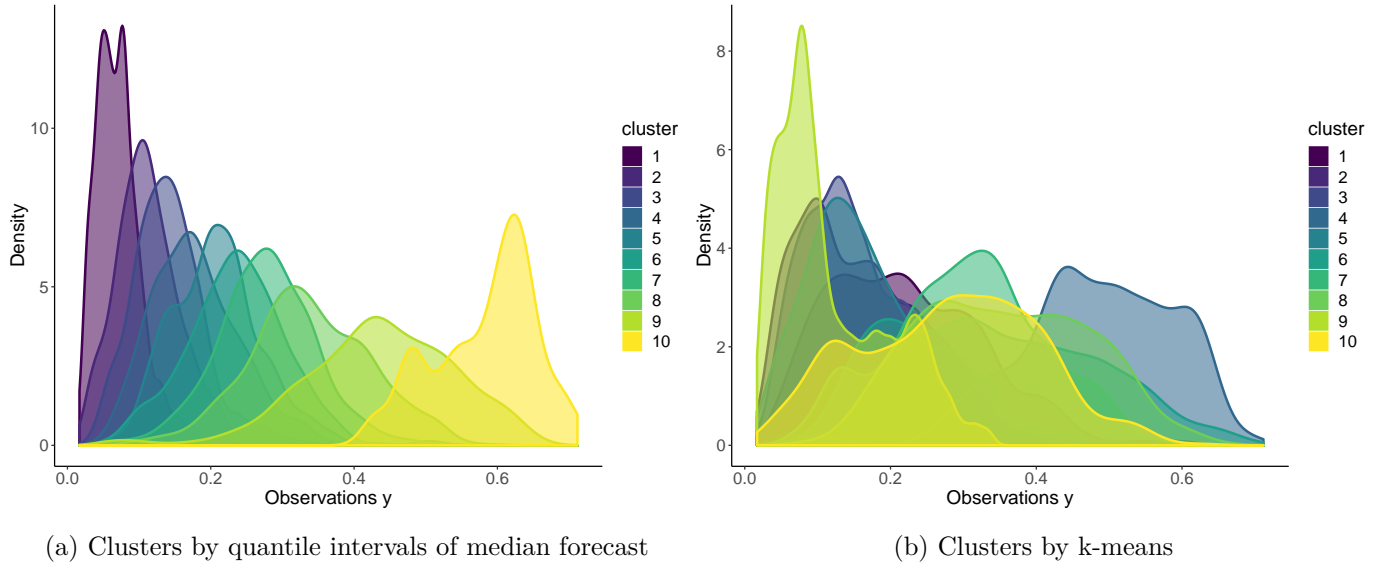
Figure 3.7: Distribution of clusterized observations for two clusterization methods, for 10 clusters in each methods.

3.4   Method 2: Extreme Value Theory

Distributions of extremes with higher versatility than the exponential model formulated above can be inferred using the Extreme Value Theory. The Fisher-Tippet Gnedenko theorem states that given a sequence of iid real random variables (Y_n), sorted in ascending order $(Y_{1,n} \leq Y_{2,n} < ... \leq Y_{n,n})$, the
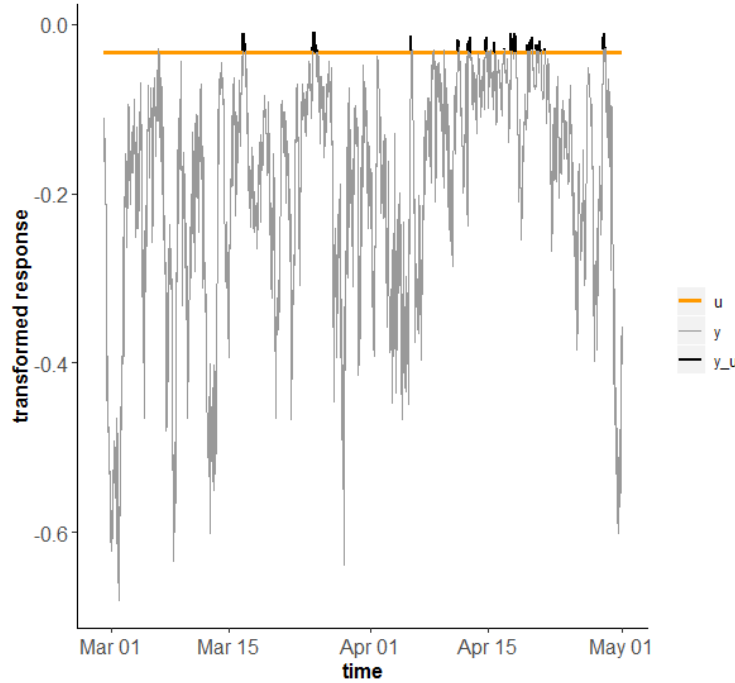
Figure 3.8: Principle of POT method. Logit transformed VPP production y, threshold u_{ext}.

maxima $Y_{n,n}$ centered by constant $b_n \in \mathbb{R}$ and scaled by constant $a_n > 0$, converge in distribution to a Generalized Extreme Value (GEV) distribution. This distribution is parametrized by an extreme value index γ, with $1 + \gamma x > 0$. Specific values of the index correspond to classic distributions: Gumbel for $\gamma = 0$, Fréchet for $\gamma = 1$, Reversed Weibull for $\gamma = -1$. The index γ is often denominated *shape*, because it determines if the distribution accepts a finite minimum ($\gamma < 0$) or has a long tail with infinite minimum (*gamma* > 0). Note that the theory developed here on maxima can be readily applied on minima by focusing on $-max(-x) = min(x)$.

$$\lim_{n \to \infty} p(\frac{Y_{n,n} - b_n}{a_n} \leq y) = exp(-(1 + \gamma y)^{-1/\gamma}) \tag{3.5}$$

Identifying extreme data. If we desire to forecast an extreme event, we need to extrapolate the distribution beyond the available data. The extrapolation is usually performed by the block maxima approach or the Peak Over Threshold (POT) approach. Block maxima are useful when observations of the phenomenon are scarce. In the power generation context, the POT approach is more valuable, as the parameters of the distribution for the tails are inferred from a collection of observations over a threshold u_{ext} . It is common practice to select u_{ext} as the k-highest value of the series, $u_{ext} = Y_{n-k+1,n}$.

A parametric distribution for extremes. Pickland's theorem [79] states that the survivor CDF $\bar{F} = 1 - F$ of values over the threshold can be approximated by a Generalized Pareto Distribution (GPD) defined by the extreme value index γ, the scale factor σ and the rank k defining the threshold u_{ext}. An extremal quantile can be obtained by inverting the CDF of the GPD defined in

(3.34).

$$\bar{F}(\tau) \sim u_{ext} + \frac{\sigma}{\gamma}[(\frac{\tau}{\bar{F}(u_{ext})})^{-\gamma} - 1], \tau \to 0 \tag{3.6}$$

Originality of the Method

An EVT model has been developed by [76] for estimating high quantiles of total RES production at a national level. This model was conditioned to median production forecasts. The conditioning of the EVT model will be discussed in the next paragraph. The novelty in the present Method consists in *merging conditioning by expected weather and VPP production forecasts with the idea of clustering by power regimes*. This follows the principle that the shape of the distribution tail will depend on power regimes, while the uncertainty on the distribution of extremes following this shape will evolve as a function of weather conditions. Lastly the selection of influential features for the conditioning is proposed to be based on a forward selection in order to avoid that adding too many parameters will add noise and ultimately prevent the model from detecting extreme conditions.

The distribution evaluated in (3.34) is an unconditional approximation of the extreme quantiles. However, it can be conjectured that the distribution of extremely low quantiles may be better anticipated with the help of the information contained in weather forecasts. Given that γ defines the overall shape of the distribution of extremes, we can safely assume that it is not influenced by short-term weather forecasts but is rather a constant characteristic of the production process. Conversely, the scale factor σ, which quantifies the spread of extreme values, is likely to be dependent on estimated weather conditions. The previous generic model is improved by conditioning the scale to weather features $\sigma(\mathbf{x})$. The extreme data is additionally clusterized in C clusters, which are identical to the clusters derived previously for exponential distributions: they originate either from intervals of median forecast or from k-means clustering.

$$\bar{F}_c(\tau) \sim u_c + \frac{\sigma_c(x)}{\gamma_c}[(\frac{\tau}{\bar{F}_c(u_c)})^{-\gamma_c} - 1], \tau \to 0, \quad \forall c \in [1, C] \tag{3.7}$$

Not all features may be beneficial to the quantification of uncertainty. We propose here to select features in a forward selection presented in Algorithm 4. Features are integrated into the conditional estimator of scale as long as the Akaikes Information Criterion (AIC) improves, and the overall scale value is positive.

Choice of threshold

The sensitivity of the EVT to the choice of threshold k is assessed with the help of the Akaikes Information Criterion (AIC), which penalizes the model likelihood with the number of parameters in the model. We present an example in Figure 3.9 for an unclusterized [1] EVT model on VPP1, the evolution of the AIC depending on k. One can observe that the goodness of fit improves with $k < 0.95$. This is expected as the model gains more information as k decreases. On the other hand, the stationarity of extremes degrades when the threshold decreases. The stationarity of the distribution of extremes qualitatively by a partial auto-correlation plot, and then quantitatively by a KPSS test. The pacf in Figure 3.10 shows that $k = 0.95$ is the upper bound of thresholds where

---

[1]Similar results are found for clusterized EVT models

Algorithm 4

Forward selection of features for conditional EVT forecasting of extremes

---

1: **Initialize**

   Compute maximum likelihood of unconditional model $\mathcal{L}^{(0)}(\gamma, \sigma^{(0)})$

   Evaluate $AIC_0 = 2k - 2ln\mathcal{L}_0$

2: **while** j < number of features in x **do**

   Add feature x_j to the scale $\sigma^{(j)}(x) = \sigma^{(j-1)} + \sigma_j.x_j$

   Compute maximum likelihood $\mathcal{L}^{(j)}(\gamma, \sigma^{(j)}(x))$

3:    **if** $AIC_j < AIC_{j-1}$ and $\sigma^{(j)}(x_i) > 0, \quad \forall x_i$ **then**

4:       $\sigma^{(j)} = (\sigma^{(j-1)}, \sigma_j)$

5:    **else**

6:       $\sigma^{(j)} = \sigma^{(j-1)}$

7:    **end if**

8: **end while**

---

the stationarity is deemed acceptable. In conclusion, $k = 0.95$ appears to compromise well between goodness of fit and stationarity in this case, and will be used throughout the rest of the study.

Parameters inference. The estimation of parameters γ, σ, k with Maximum Likelihood Estimation (MLE) is possible for $\gamma \geq -1$. Other inference models exist: the Hill and moment estimators derive γ as a function of k, and is known to converge in probability only for $\gamma \geq 0$ [79], while bayesian estimators perform inference at the cost of defining prior distributions. We use the MLE as it is a reasonable method in our context: it is applicable to most situations ($\gamma \geq -1$) and is less complex than bayesian estimators.
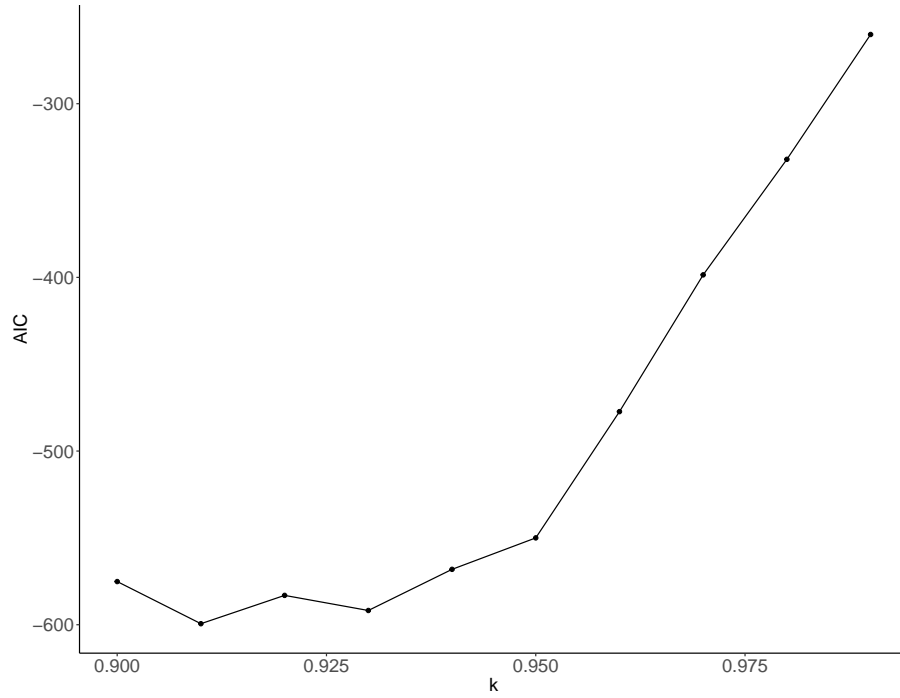
Figure 3.9: AIC of EVT model for several values of threshold quantile k
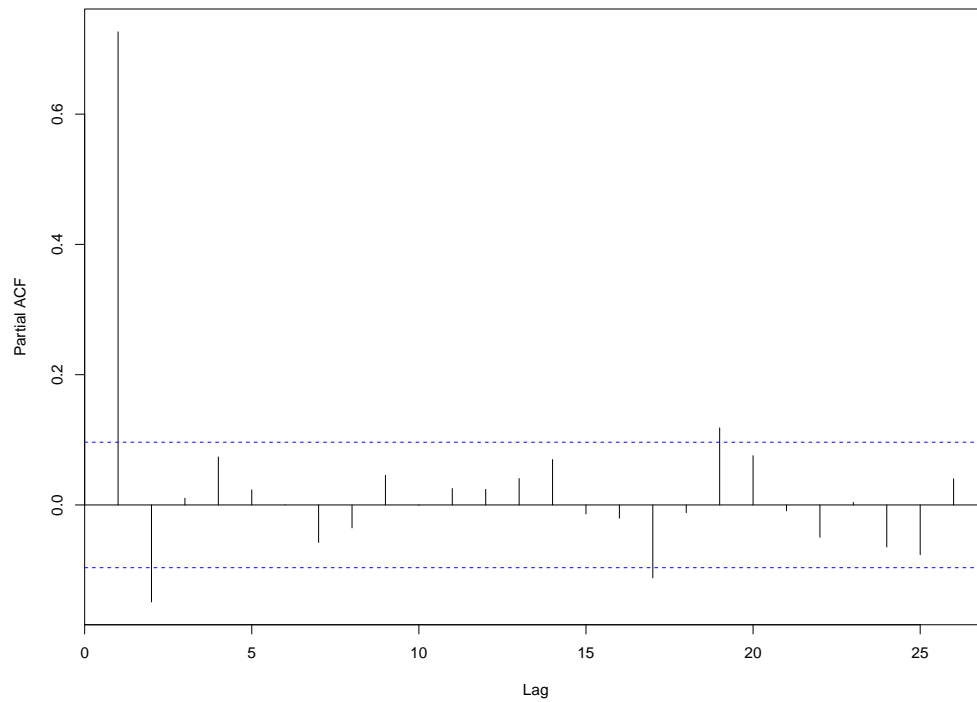
Figure 3.10: Partial auto-correlation plot of EVT model for k=0.95. Temporal resolution of forecast and production is 10 minutes.

3.5 Method 3: Quantile regression with neural networks

The previous Chapter has shown that quantile regression based on convolutional and recurrent neural networks are appealing for the prediction of aggregated production. However, due to the low number of observations susceptible to be penalized by the standard quantile loss function, it is proposed here to adapt the regression from neural networks to the context of predicting extremes of aggregated production.

Originality of the method

While neural networks have been intensively used in the past for deterministic forecasting of VRE production, starting from seminal works such as [48] on wind power, and recently developing to probabilistic forecasting e.g. in [130], there has been little attention devoted to the forecasting of extremes with such networks. The neural network proposed by [131] identifies extreme wind power events by learning to recognize several regimes of error. However the aim is here to reduce the global deterministic forecasting error. The recurrent neural network architecture of [101] is claimed to be adapted to the forecast of extreme events, due to the inclusion of an auto-encoder that helps the network detect extreme conditions by condensing the large amount of information. Nonetheless, there is little evidence that this approach could readily apply to the forecast of very low quantiles of aggregated VRE production. In contrast, the Method proposed in this Section consists in adapting the quantile regression networks with two innovations:

- Formulation of a *training loss function that is specific to extreme forecasting.*

- Adapt the *CNN architecture* to detect more easily the conditions leading to extreme production levels.

Regarding the first adaptation mentioned above, it has been observed in the Introduction that the main issue on forecasting very low quantiles is the loss of reliability. Therefore instead of training on the quantile loss, models are trained on the Skill Score *SkSc* [86] defined in (3.8) for a quantile τ, which offers a good balance between reliability (probabilistic deviation on left operand), and resolution and sharpness (production deviation on right operand).

$$SkSc^\tau(y, \hat{y}^\tau) = -(1_{y-\hat{y}^\tau} - \tau)(y - \hat{y}^\tau) \tag{3.8}$$

Results of Chapter 2 have shown that convolutional neural networks are effective in the context of multi-source aggregated production. If the configuration of CNN presented earlier for non-extremal parts is applied to the forecast of very low quantiles, the most influencing variables will be diluted by the successive filtering layers. Instead, if a specific pooling layer is introduced after convolutional filters to retain the minimal value on each kernel, then the network can focus more easily on the behaviour of aggregated production under extreme conditions. The pooling layer introduced is denominated here as *min pooling*, by opposition to the *max pooling* broadly used in CNNs for classification [51], [132]. The min-pooling layer is illustrated in Figure 3.11: it selects as
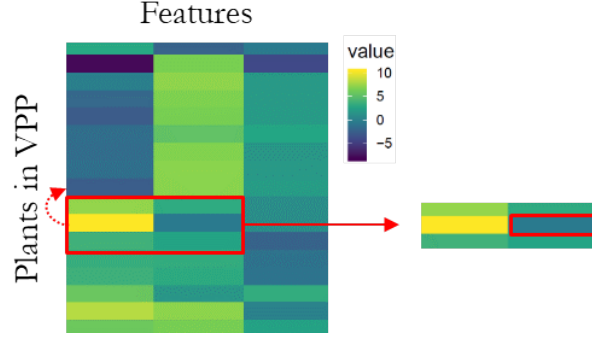
Figure 3.11: Illustration of the min-pooling layer

in (3.9) minimal values in window w of the inferior layer $l-1$, for a filter of stride s and kernel size k.

$$h^{(l)}_{minpool,k,s} = min(\{h^{(l-1)}_{i,j}, (i,j) \in w(s,k)\}) \tag{3.9}$$

Finally, the recurrent networks are equipped with the stateful property presented in Section 2.4.2 in order to memorize at which horizons extreme events can occur between batches.

3.6 Method 4: Mixture Density Networks

It is known that a plain Gaussian distribution can not model accurately the forecast error of renewable production [133]. If one wishes to describe the uncertainty of VRE production in a parametric framework, more flexible representations of the uncertainty are required. A first state-of-the-art approach consists in applying a mixture of distributions, because it allows more diversity on tails and asymmetries than a single distribution while remaining relatively tractable.

A mixture density model combines K marginal distributions $D_k, k \in [1, K]$, each associated with a probability of occurrence π_k. We assume at this point that marginal distributions originate from the same exponential family (the exponential family contains Gaussian, Gamma, Beta distributions, see definition in [81]). Distributions from the exponential family are convenient for modelling mixtures because the likelihood of their mixture is generally tractable [82]. For now, let us define distributions by the following parameters: mean μ_k and precision ϕ_k , where the precision is the inverse of the standard deviation. Distribution parameters and the probabilities of occurrence form a parameter set θ. The objective of this method is to find the best model that fits in (3.10) the aggregated production as a mixture of components conditioned by features x. The link between features and parameters is established with a link function $g(.)$. In the following we propose a neural network for g, to build non-linear relationships between θ and x. The corresponding prediction model is denominated following the literature as a Mixture Density Network (MDN).

$$y|x \sim \sum_{k \in [1,K]} \pi_k(x) D_k(\mu_k(x), \phi_k(x))$$

$$\sum_{k \in [1,K]} \pi_k(x) = 1, \quad \forall x$$

$$\theta(x) := (\pi_k(x), \mu_k(x), \phi_k(x)) = g(x) \tag{3.10}$$

Originality of the Method

An MDN has been recently proposed by [53] to issue probabilistic forecasts of wind power. This work focused on the generic probabilistic performance and not on the forecast of extremes. We assume here that mixture models can be effective at predicting rare events, as shown by [88] in the context of econometry, and we combine it with several simple neural network architectures. The novel contributions of the present Method are as follows:

- A first proposition to *apply and fit MDNs to the forecast of extremes of aggregated VRE production*

- A *comparison of a Beta mixture and a Gaussian mixture specifically designed for the problem.* The Beta mixture is adapted to the bounded nature of VRE production, and the Gaussian mixture is parametrized by the installed capacities of plants in the VPP. The inference of these mixtures is performed with Bayesian approaches that have proven to be effective in the context of limited data points [134] and neural networks [135].

The parameters θ of a mixture distribution can be inferred in a frequentist approach: the latent parameters associating observations to components can be estimated with an Expectation-Maximization (EM) algorithm. However, the EM does not quantify the uncertainty associated to the parameters, is sensible to initialization [82] and converges to local optima distant from the global optimum when the amount of samples is small [134]. This last point is important when inferring extremes, as few samples are effective in for learning extremes.

Alternatively, we can formulate the mixture regression following the Bayesian framework, where the parameters of the mixture are assumed to follow prior distributions, and the posterior distributions of parameters are inferred as a function of observed training data. More precisely, we infer the posterior distribution of parameters in (3.11) by invoking the Bayes rule: the posterior writes as the product of the likelihood $p(y, x|\theta)$ and of the prior distribution of parameters, normalized by the evidence of observations $p(y|x)$. The prior distribution of parameters models an existing belief on the possible values taken by parameters, considering the problem at hand. Then the density of the response variable can be obtained in (3.12) by integrating over the range of parameter values. In practice the integration is approximated thanks to sampling the parameter space. But beforehand we must obtain the likelihood of the mixture, from which we will later derive posteriors and simulate densities of all components f_k.

$$p(\theta|y, x) = \frac{p(y|x, \theta)p(\theta)}{p(y|x)} \tag{3.11}$$

$$p(y = y_t | x = x_t) = \int_\theta \sum_{k=1}^{K} \pi_k(x_t) f_k(y_t | x_t, \theta_k) p(\theta | y_t, x_t) p(\theta) d\theta \tag{3.12}$$

Alternative solutions such as *re-parametrization* or *non-exchangeable priors* solve the identifiability problem with better properties. The latter is used here: each marginal Gaussian distribution is affected to each of the P plants in the aggregation. Its prior is the observed mean and variance of each plant production over the training batches B. The mixture probability is parametrized in (3.13) by the share of installed capacity $y_{max,k}$ of plant k in the total aggregation.

$$\mu_{\mu_k}, \sigma_{\mu_k} = \arg\max_{\mu,\sigma} \prod_{b=1}^{B} f_{\mathcal{N}}(\mathbb{E}_b(y_k) | \mu, \sigma, .), \quad \forall k \in [1, P] \tag{3.13}$$

$$c_k, r_k = \arg\max_{c,r} \prod_{b=1}^{B} f_\Gamma(\frac{1}{\mathbb{V}_b(y_k)} | c, r, .), \quad \forall k \in [1, P] \tag{3.14}$$

$$\pi_k \sim Dir(\frac{y_{max,k}}{\sum_{k=1}^{P} y_{max,k}})$$

The next section presents a mixture of Beta distributions, which is well suited to the forecast of extremes due to its flexibility and remains in the bounded space of renewable production.

3.6.1 Inference of the mixture as an incomplete problem

The inference of the mixture presented above is an incomplete problem, because it is not possible to associate directly from data the global parameters θ to the local context formed by observations and the component they should be associated with. We can complete the inference by adding a missing/latent variable z_{ik} [81], linking any of the N available observations $y_1, \ldots, y_i, \ldots, y_N$ to a specific component of the mixture k:

$$z_{ik} = \begin{cases} 1 & \text{if } y_i \sim f_k(\mu_k(x_i), \phi_k(x_i)), \\ 0 & \text{else} \end{cases} \tag{3.15}$$

The random variable $\mathbf{z}$ is naturally modelled by a Categorical/Multinomial distribution, whose conjugate prior is the Dirichlet distribution: each observation y_i can be associated to the component distribution k proportionally to a weight/probability π_k following the Dirichlet distribution. The hyperparameters of the Dirichlet distribution $\alpha \in \mathbb{R}$, usually called concentration rates, can be viewed as the occurrences of observations in each component of the mixture. In this work, the number of components is assumed to be known (i.e. chosen empirically before inference), but it can be treated as an unknown random variable in the bayesian inference framework [49].

$$\mathbf{z}_i |. \sim \mathcal{M}_k(1; , \pi_1, ..., \pi_k) \tag{3.16}$$

The likelihood writes then as the product of marginal densities with parameters indicated by the value contained in $\mathbf{z}$.

$$p(y | \theta, \mathbf{z}) \sim \prod_{i=1}^{N} \prod_{k=1}^{K} f_k(\mu_k^{z_{ik}}, \phi_k^{z_{ik}}) \tag{3.17}$$

In this work, we focus on distributions pertaining to the exponential family. The density of a distribution in the exponential family writes [82]

$$p(y|\theta) = h(y)exp(r(\theta)^T t(y) - a_y(r(\theta))) \tag{3.18}$$

where h is a scalar function called *base measure* defined on $\mathbb{R}^+$. Distributions of the exponential family are assured to be sufficiently described, with constant dimension, by the *sufficient statistic* t. The *natural parameter* r links the distribution to the *natural* exponential form $exp(\theta x)$, and the *log-normalizer* a_y scales the projection $r(\theta)^T t(y)$ into the space of y.

If the conditional distributions of z_k given z_{-k} ,θ and y follow a distribution of the exponential family, then the prior distributions of θ belong also to a similar exponential distribution (priors are said to be conjugate, see [81]). In this exponential family framework, the posterior distributions of mixture parameters can be easily derived. Each component is associated with a conjugate prior $p(\theta_k|\alpha_k, \beta_k)$ where $\alpha_k \in \mathbb{R}$ and $\beta_k > 0$.

$$p(\theta_k|\alpha_k, \beta_k) \propto exp(r(\theta)^T \alpha_k - \beta_k a_y(r(\theta))) \tag{3.19}$$

The posterior distribution of θ depending on y and z writes [82] :

$$p(\theta|z, y) \propto \prod_{k=1}^K exp[r(\theta_k)^T (\alpha + \sum_{i=1}^n \mathbb{I}_{z_i=k} t(y_i)) - a_y(r(\theta_k))(n_k + \beta)] \tag{3.20}$$

where $n_k = \sum_{i=1}^n \mathbb{I}_{z_i=k}$.

We apply this scheme on a Gaussian Mixture to form a Gaussian Mixture Bayesian Regression dependent on features x. Generic terms in (3.18) become:

$$r(\theta) = \begin{bmatrix} \phi\mu \\ -\phi/2 \end{bmatrix}$$
$$t(y) = \begin{bmatrix} y \\ y^2 \end{bmatrix}$$
$$a_y(x) = \begin{bmatrix} 0 \\ log2\pi\frac{1}{x} \end{bmatrix}$$

The prior distribution for the variance of one component of the mixture follows a conjugate Gaussian distribution. As our mixture formulation involves a precision instead of the variance, we use a Gamma prior for precision. Priors distributions of the model are summarized in (3.21).

$$\pi_k \sim Dir(\alpha_\mathbf{k}) \tag{3.21}$$
$$\mu k \sim \mathcal{N}(\mu_{\mu k}, \sigma_{\mu k})$$
$$\phi k \sim \Gamma(c_k, r_k)$$

An issue for the Gaussian mixture is that it suffers from a lack of identifiability: posteriors of parameters of the mixture are invariant with permutations of mixture component indices,

$\sigma(1,...,K) \rightarrow (\sigma(1),...,\sigma(K))$. The invariance is easily demonstrated in (3.22) after a change of variables:

$$p(\pi, \theta) \sim p(\pi), p(\theta). \sum_{i=1}^{N} \sum_{k=1}^{K} \pi_k f(y|x, \theta_{\mathbf{k}})$$

$$p(\pi_\sigma, \theta_\sigma) \sim p(\pi_\sigma), p(\theta_\sigma). \sum_{i=1}^{N} \sum_{j=1}^{K} \pi_{\sigma(j)} f(y|x, \theta_{\sigma(\mathbf{j})}) = p(\pi), p(\theta). \sum_{i=1}^{N} \sum_{k'=1}^{K} \pi_{k'} f(y|x, \theta_{\mathbf{k'}}) \tag{3.22}$$

Adding a constraint such as ordering the means, which translates into truncating the original prior distribution, remediates to the identifiability problem [82]. This truncation is however not consistent with the likelihood, and may contradict the assumptions on priors, hereby leading to a lack of performance. Alternative solutions such as *re-parametrization* or *non-exchangeable priors* solve the identifiability problem with better properties. The latter is used here: by imposing distinct means and variances equal to the observed past mean and variance of each power plant in the aggregation.

3.6.2 Derivation of Bayesian Beta Mixture Regression (BBMR)

The limit of the Gaussian Mixture is that it is not specifically developed to model fat tails, and can not directly be applied to a bounded process such as renewable generation. Following the flexible Bayesian Beta Regression proposed by [88], the mixture is defined by two Beta densities which are re-parametrized to share a common precision parameter. This re-parametrization avoids the identifiability issue already mentioned. The Beta distribution is a flexible distribution, but is not appropriate to model heavy-tailed distributions. A generic mixture of Beta distributions can overcome these limitations, but its tractability is limited as its likelihood is unbounded and it has been proven sensible to initialization [88]. A flexible Beta mixture, where two Beta distributions share a common precision ϕ , has a bounded likelihood and can accommodate various types of tails and asymmetries [88]. Furthermore we have seen in the Gaussian mixture case that non-identifiability of mixture is detrimental to the performance of inference. The re-parametrization affects distinct parameters to each component of the mixture, therefore avoiding the identifiability problem [82]. The univariate random variable associated with the aggregated production y_{agg} follows (3.23),

$$y_{agg} \sim pBe(\lambda_1, \phi) + (1-p)Be(\lambda_2, \phi) \tag{3.23}$$

where the probability density function of a bounded random variable $z \in [0,1]$ following a Beta distribution is:

$$f_{Be}(z, \lambda, \phi) = \frac{\Gamma(\phi)}{\Gamma(\lambda\phi)\Gamma((1-\lambda)(\phi))} z^{\lambda\phi-1}(1-z)^{(1-\lambda)\phi-1}$$

The mixture is now re-parametrized by introducing in (3.24) the common mean μ and in (3.25), the distance ω between the means of the component distributions.

$$\mu = p\lambda_1 + (1-p)\lambda_2 \tag{3.24}$$

$$\omega = \lambda_1 - \lambda_2 \tag{3.25}$$

Afterwards, the parameters $(\lambda_i, \phi), i \in [1, 2]$ of marginal components are easily converted back into Beta form parameters (α_i, β_i):

$$
\begin{cases} \phi & = \alpha_i + \beta_i \\ \lambda_i & = \frac{\alpha_i}{\alpha_i + \beta_i} \end{cases} \Longleftrightarrow \begin{cases} \alpha & = \lambda_i . \phi \\ \beta & = \phi(1 - \lambda_i) \end{cases}
$$

Parameters are inferred by regression on features x through a link function, similarly to the Gaussian case. The choice of prior distributions is tailored to the behaviour of Beta distributions. We define Beta priors on the common mean μ and the distance between marginal means ω, and a Gamma prior on the common precision ϕ because low precisions are assumed to be probable compared to high precisions. The mixing probabilities are assumed to be distributed following a Dirichlet distribution, parametrized by the two concentration rates α_1, α_2 of the two mixture components.

$$
\begin{aligned}
\mu &\sim Be(\alpha_{mu}^0, \beta_{mu}^0) \\
\omega &\sim Be(\alpha_{omega}^0, \beta_{omega}^0) \\
\phi &\sim \Gamma(c_{phi}^0, r_{phi}^0) \\
\pi &\sim Dir(0.5, 0.5)
\end{aligned}
$$

3.6.3 Neural networks learning parameters of the mixture

It is assumed that the forecasting performance on extremes depends on the choice of the regression function $g(.)$ which links the features x to the parameters of the mixture, similarly to what happens on non-extremal quantiles. The regression function is therefore tested with the neural networks implemented previously for the non-extremal forecasting (FCNN, CNN and LSTM, see Section 2.4).

3.6.4 Choice of approximation method of the posterior distributions

Even when the posterior distribution has a closed-form, it is difficult to numerically minimize the loss of the Bayesian estimator based on this posterior because of the large dimensions of the parameter space $\Theta, \theta \in \Theta$ and the decision space $\mathcal{D}, (x, y) \in \mathcal{D}$.

Two main families of solutions are available when looking for posterior distributions that minimize the Bayesian estimator loss: Variational Inference (VI), and sampling under the Markov Chain Monte Carlo (MCMC) paradigm. VI approximates the posterior distribution by a surrogate model that is tractable in a classical optimization framework. It has been implemented by [49] for a wind power prediction model based on a Gaussian Mixture. In contrast, sampling methods discretize the exact distribution model by correlated samples [82]. MCMC sampling is asymptotically correct [136] but it is difficult to obtain satisfactory convergence in practice. Concerning VI, it is difficult to analyze the quality of the posterior approximation, even if some advanced evaluation tools have emerged recently [137]. In this work, VI is chosen for its adequate compatibility with gradient-descent optimization.

The likelihood $\mathcal{L}(y|x, \theta, \mathbf{z})$ of a mixture distribution with latent variables $\mathbf{z}$ can be decomposed in two parts: the Evidence Lower Bound (ELBO), and the Kullback-Leibler divergence between the

approximate distribution of the posterior distribution of parameters and latent variables $q(\theta, \mathbf{z}|.)$ and the prior distribution $p(\theta, \mathbf{z})$. With the assumption that parameters are independent, the likelihood can be approximated (even derived analytically in some simple cases as in a Gaussian Mixture), the KL term is easily computed and acts as a regularizing term.

$$\mathcal{L}(y|x, \theta, \mathbf{z}) = \mathbb{E}_{q(\theta, \mathbf{z})} log \frac{p(\mathbf{y}, \theta, \mathbf{z})}{q(\theta, \mathbf{z})} + KL(q(\theta, \mathbf{z}|y, x), p(\theta, \mathbf{z})) \tag{3.26}$$

$$max \quad \mathcal{L}(y|x, \theta, \mathbf{z}) = ELBO + KL(q(\theta, \mathbf{z}|y, x), p(\theta, \mathbf{z})) \tag{3.27}$$

$$\iff min - ELBO = -\mathcal{L}(y|x, \theta, \mathbf{z}) + \quad KL(q(\theta, \mathbf{z}|y, x), p(\theta, \mathbf{z})) \tag{3.28}$$

3.6.5 Variational inference of mixtures

Ma and Leijon [134] propose a VI scheme for Beta Mixtures. The intractable conjugate posterior of the Beta distribution is decomposed following the mean-field approach into independent densities relative to parameters u_{Be} and v_{Be}.

$$f(u_{Be}, v_{Be}|y, x) \simeq f(u_{Be}|y, x) f(v_{Be}|y, x) \tag{3.29}$$

After approximating the conjugate prior with the Gamma distribution, the VI algorithm derives the expectations of variational distributions as detailed in [134]. In this work, we choose a more general VI scheme that can be applied to any mixture of distributions that belong to the exponential family (among which Gaussian and Beta), namely the Stochastic Variational Inference (SVI) developed by [135].

The SVI relies on the inference of posterior variational distributions $q(\theta, z)$ over sampled observations. The inference is stochastic, consequently it needs an important number of iterations to converge. However, with the assumption of independent posteriors in the mean-field approximation, posteriors of all parameters can be derived in parallel and in mini-batches. SVI is then readily applicable to gradient-descent models, where neural networks perform the inference of parameters (see illustration in Figure 3.12).

Steps of SVI are summarized in Algorithm 5 for a Beta Mixture of parameters (μ, ω, ϕ, π). After having defined prior distributions on all parameters, a neural network infers the distribution of each parameter. A pass of data sample through the network approximates the posterior distribution of the parameter. In parallel, the network computes the likelihood of the mixture. The sum of the KL-divergence between priors and posteriors and the negative likelihood is minimized by gradient descent.
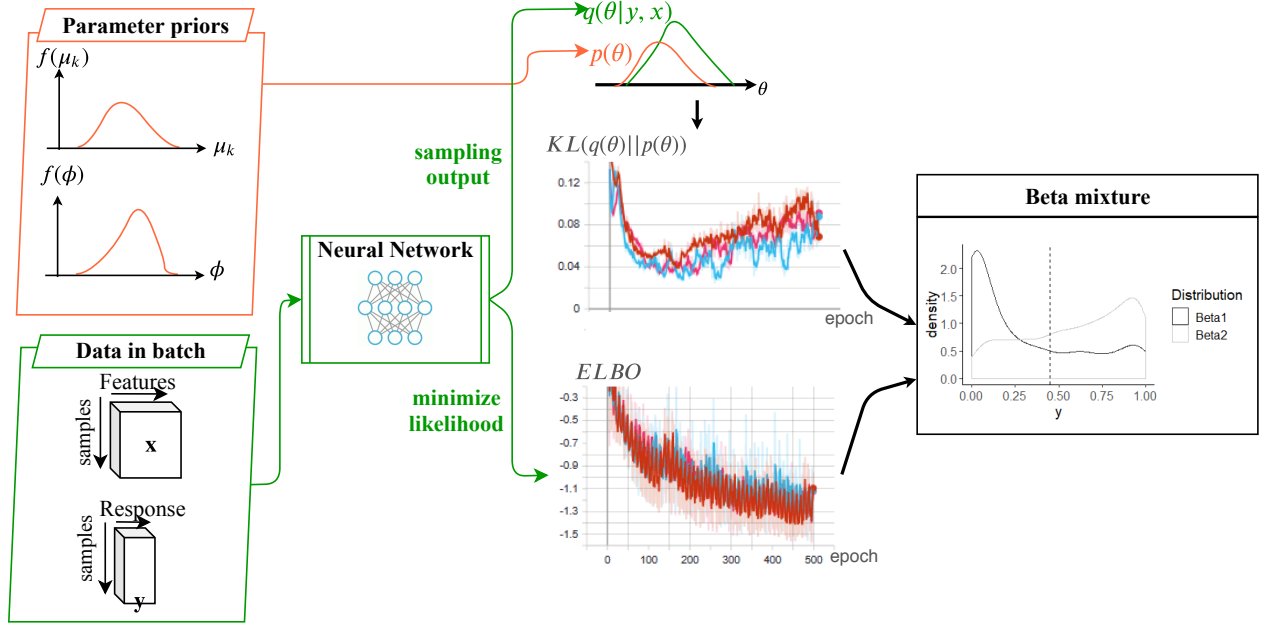
Figure 3.12: Principle of SVI on a Beta Mixture, with neural networks inferring parameters. The different colors in the KL and ELBO diagrams represent the various cross-validation datasets.

Algorithm 5 Stochastic VI algorithm for a Mixture Density Network

---

1: **Initialize** randomly parameters $\theta = (\mu, \omega, \phi, \pi)$ and hidden variables $\mathbf{z}$ of mixture model

2: **Prior distributions.** Each parameter θ_i of the mixture is associated with a prior distribution $\mathcal{D}_i$ and hyperparameters ν_i^0.
$p(\theta_i) \sim \mathcal{D}_i(\nu_i^0) \quad \forall i \in card(\theta)$

3: **while** epoch in 1...epochs **do**

4:    **Inference of hyperparameters by neural network.** A neural network g infers the hyperparameters of each parameter i over a mini-batch $b = (x, y) \in (\mathbb{R}^D, \mathbb{R}^1)$.
$\nu_i^{NN} = g(\nu_i|x, y)$

5:    **Approximate variational distribution.** The approximate variational posterior of each parameter i is obtained by taking the mean of the distribution $\mathcal{D}_i$ over the mini-batch b:
$q(\theta_i) = \mathbb{E}_{(b)}\mathcal{D}_i(\nu_i^{NN}|x, y \in b)$

6:    **Approximation of likelihood.** The variational distribution of parameters over the mini-batch enable to compute the negative likelihood of the mixture.
$-\mathcal{L}(y, x) = \prod_{t\in b} \prod_{k\in[1,K]} \pi_k(x_t)^{z_{t,k}} Be(y_t; \alpha_k(x_t), \beta_k(x_t)^{z_{t,k}}$

7:    **Loss minimization.** The ELBO loss is minimized by gradient descent.

8: **end while**

3.6.6 Limits of the approach

The present approach is readily applicable to a univariate regression, where we forecast directly the extremes of the total aggregated production. However, CNN, LSTM and Gaussian Mixture

forecasts can be easily extended to a multivariate regression framework, where each dimension corresponds for instance to one particular plant of the aggregation. Multivariate Extreme Value Theory could also be considered [138]. The multivariate extension is more challenging for the Beta mixture: the multivariate version of the Beta mixture is a Dirichlet mixture, and its application to the context of extremes is promising in a Bayesian inference framework, but works found in literature are limited to dimension five [139], which is insufficient for the present problem.

Once a multivariate forecast has been obtained, the ellipsoidal uncertainty regions proposed by [140] produce a prediction interval over all dimensions of the response. The reliability of such an approach for extremes of aggregated multi-source VRE production remains to be studied. Furthermore, an additional step would be needed to derive a reliable level of total production from multivariate prediction intervals.

3.7 Case Study

The presented methodology is evaluated in two different configurations, in order to address the different challenges arising concerning the forecast of extremes. We evaluate here the forecasting models on a cross-seasonal test, similarly to the previous Chapter, we evaluate our models by cross-validation on weekdays with production data at a 30-min resolution. This enables to assess if models can detect extreme production regimes in various weather conditions, with a minimal impact of the temporal auto-correlation on the result of the analysis. The test is applied to the Wind-PV-Hydro aggregation already presented in the case study of Section 2.5.

Regarding the practical implementation of models, the exponential distribution model is coded with user-defined functions executed on the R platform, while the EVT model is inferred using the extRemes package of the R platform [141] by maximization of the likelihood. The quantile regression on CNN and LSTM, as well as the Mixture Density Networks, are realized by means of Keras and Python. Gradient descent in networks is operated with the ADAM optimizer, which adapts its learning as a function of exponential decays of past gradients [51].The learning rate is kept at a low value 1e-3 to avoid large oscillations during learning.

3.8 Evaluation metrics

When evaluating forecasts in the context of extremes, one faces two issues:

- For extremely low quantiles, we expect to verify very few observations laying below forecast values. This challenges the evaluation of reliability: what is the degree of confidence one should have regarding a specific reliability score?

- Standard metrics for global forecasting score put more weight on central quantiles than extremal quantiles. Global forecasting scores must therefore be adapted for the prediction of extremes.

We present briefly below the methods employed to evaluate reliability and global score on tails.

3.8.1 Confidence interval of reliability considering sampling effect

As seen in the Introduction, the evaluation of forecasting models on samples of limited size is influenced by the distribution of observations within the evaluation set. Therefore even a perfectly reliable forecasting model will show deviations from the nominal quantile value. The sampling effect due to the size of the evaluation set is incorporated into the reliability diagram by adding consistency bars. Consistency bars indicate the confidence intervals where the observed reliability may be found without rejecting the assumption of reliability. In this work, bars are computed following [125] by resampling the available forecast to obtain a reliable surrogate forecast.

3.8.2 Quantile score on tails

The Quantile Score (QS) metric, introduced in Section 2.4, though a proper score, tends to understate the capacity of a forecasting model to forecast extremely low quantiles. Following [142] the QS turns into a weighted Quantile Score (wQS) which balances the QS with the quantile value:

$$wQS_{[0.1\%,0.9\%]}(\hat{y}, y) = \int_{\tau \in [0.1\%,0.9\%]} (1-\tau)^2 QS^{(\tau)}(\hat{y}^{(\tau)}, y)d\tau \qquad (3.30)$$

The interest of the wQS is that all quantiles in the quantile range contribute more evenly to the total score compared to the QS, where lowest quantiles have almost no impact.

3.9 Results

The performance of all models for forecasting extremely low quantiles of aggregated VRE production is discussed first for each model separately, the best configuration of each model is identified. Lastly, the best configurations for each model are compared and discussed.

Exponential distribution. The reliability of the exponential distribution model is displayed in Figure 3.13, as a function of the clustering of observations (the 10 clusters obtained by median forecast intervals are in black, the clusters obtained by k-means are in blue levels), and of the reference quantile (one diagram represents one reference quantile, from 1% on top to 5% at the bottom).
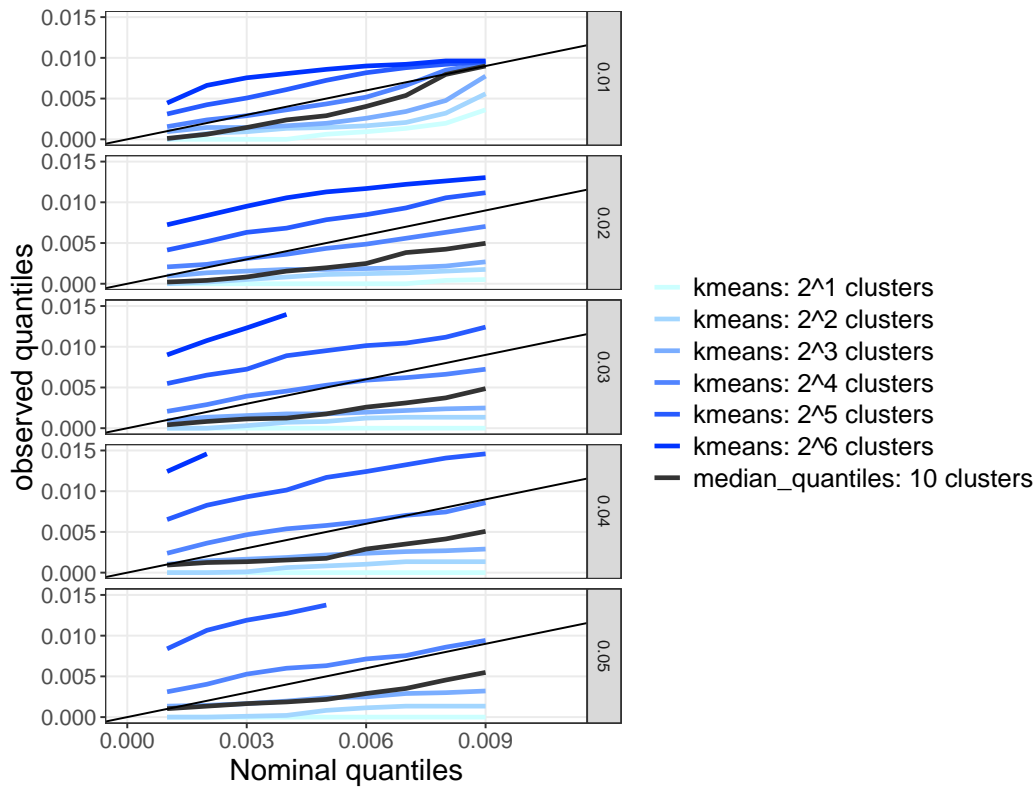
Figure 3.13: Reliability diagram for exponential distributions with different number of k-means clusters, as a function of the reference quantile (from 0.01 to 0.05). Results on VPP1 (wind-dominated)

In general terms, reliability improves with a lower reference quantile. The clustering by median forecasts shows acceptable reliability for reference quantile 1% only (deviations limited at 0.1 % at most). It is observed that the model with clusters of median quantiles is negatively biased, for all reference quantiles (observed quantiles are below nominal quantiles). K-means clustering improves reliability compared to clusters of median forecasts: a satisfactory reliability (deviations lower than 0.1 %) can be obtained at various reference quantiles, arguably because clusters with k-means are less dependent on the value of aggregated forecast. Reliability is best in this case for medium-sized clusters (16 clusters). This is expected as low numbers of clusters impede discrimination and high

numbers of clusters limit the capacity to classify observations.

The sensitivity to the cluster size is also observed in Figure 3.14, where the weighted QS finds its minimum at 16 clusters. The less reliable forecasts with median clusters, in yellow, show a slight advantage in terms of log weighted QS (-7.07 compared to -7.02 with 3% reference quantile).
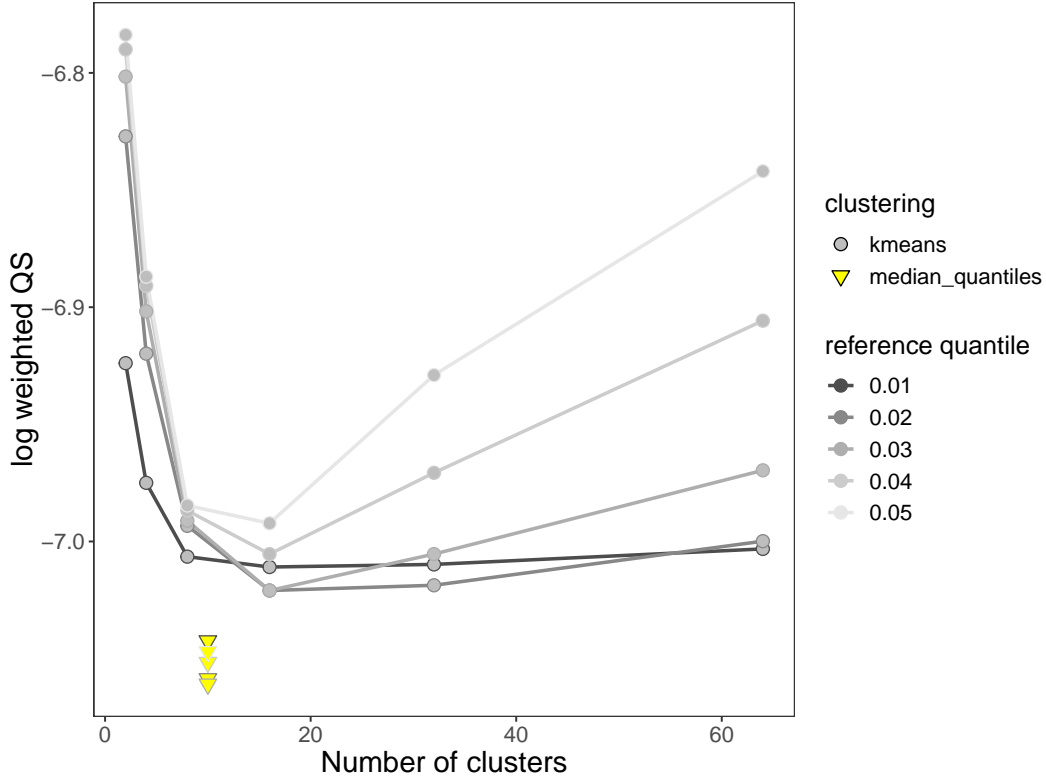
Figure 3.14: Logarithm of weighted QS for exponential distributions with different number of k-means clusters, as a function of the reference quantile (from 0.01 to 0.05)

Extreme Value Theory. 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. Clusters are limited in size compared to the case of exponential distributions, because thresholds already filter most of the available production data. Figure 3.15 presents the forecasting scores of EVT, as a function of the threshold and of the number of clusters. Two tendencies are observed: the global forecasting score (weighted QS, upper panel) and the average reliability over the extremal quantile range [0.1%-0.9%] (center panel) improve with an increasing value of the threshold. EVT models have better average reliability than the QRF only for threshold values above 95%, expect for the clustering with 5 clusters, the highest number, which is constantly too narrow. This may indicate a lack of diversity in peaks. Clusters of medium-size (3-4) compare marginally better with the cluster of small size (2 clusters) because they characterize the production process well enough to adapt their resolution to the amount of data available (with increasing thresholds above 95 %, reliability remains constant and sharpness decreases). The reliability as a whole is however not improving the situation when compared to

QRF, whereas the global forecasting score improves (QRF has a log wQS of -7.01). It seems that the gain in sharpness with clusterization leads to biased predictions of extremes.
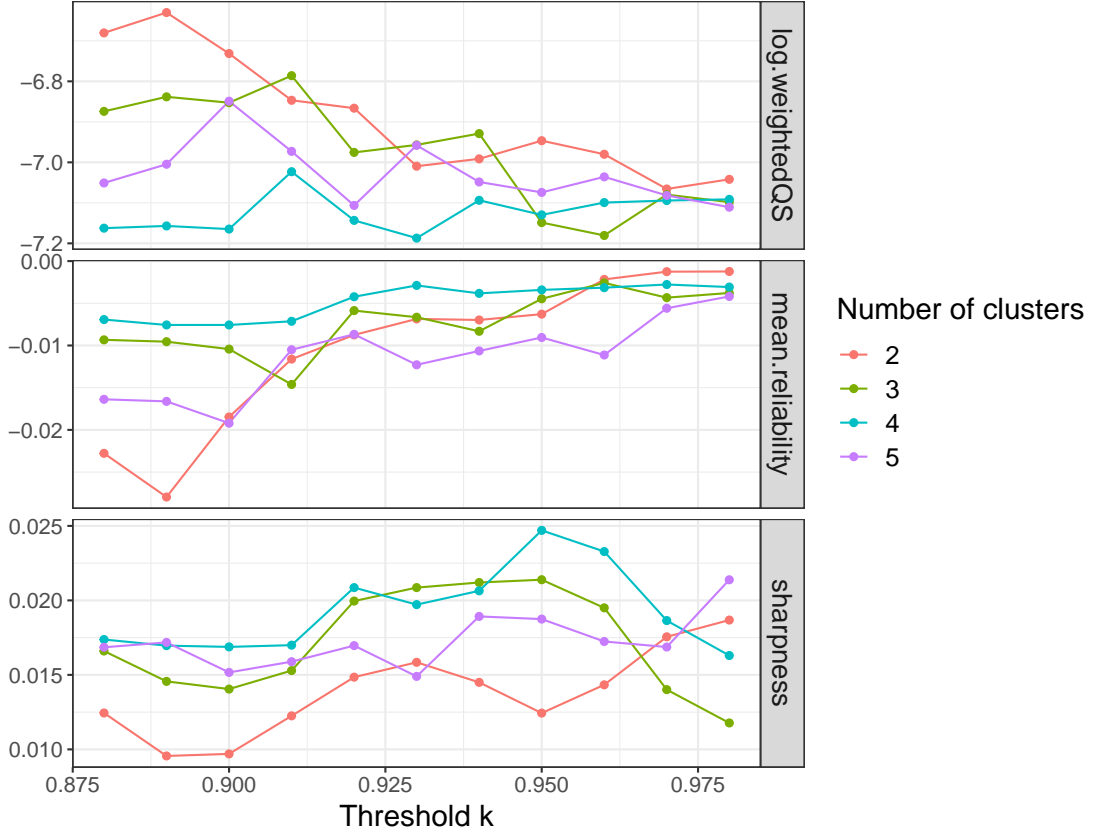
Figure 3.15: Results of the EVT model as a function of threshold k and reference quantile, for VPP1

Quantile regression with neural networks. Similarly to the non-extremal forecast, the regression based on LSTM is too sharp and biased to provide satisfactory results. Better results may be found with adapted weights initialization or kernel regularization. In contrast, the CNN regression with min-pooling offers a good balance between average reliability (second best model) and global score (best model). The effect of min-pooling on CNN regression forecasts is illustrated in Figure 3.16, which represents the forecasts issued for a day with medium-high levels of production (the observed production in orange lies within 30%-60% of the maximum capacity). The forecasts obtained by CNN regression with min-pooling are represented on the left panel, and compared to the same model without pooling layer. It can be seen that the min-pooling layer effectively dampens the forecast, especially in middle hours when the CNN model anticipates the highest production levels of the day.

This observation is corroborated by an analysis of the forecast reliability. The absence of the min-pooling layer in the CNN model creates uncalibrated forecasts on the lowest quantiles: the reliability diagram shown in Figure 3.17 indicates a positive deviation for quantiles below 0.4 % (red curve), whereas the CNN model with min-pooling forecasts within the uncertainty bars (orange
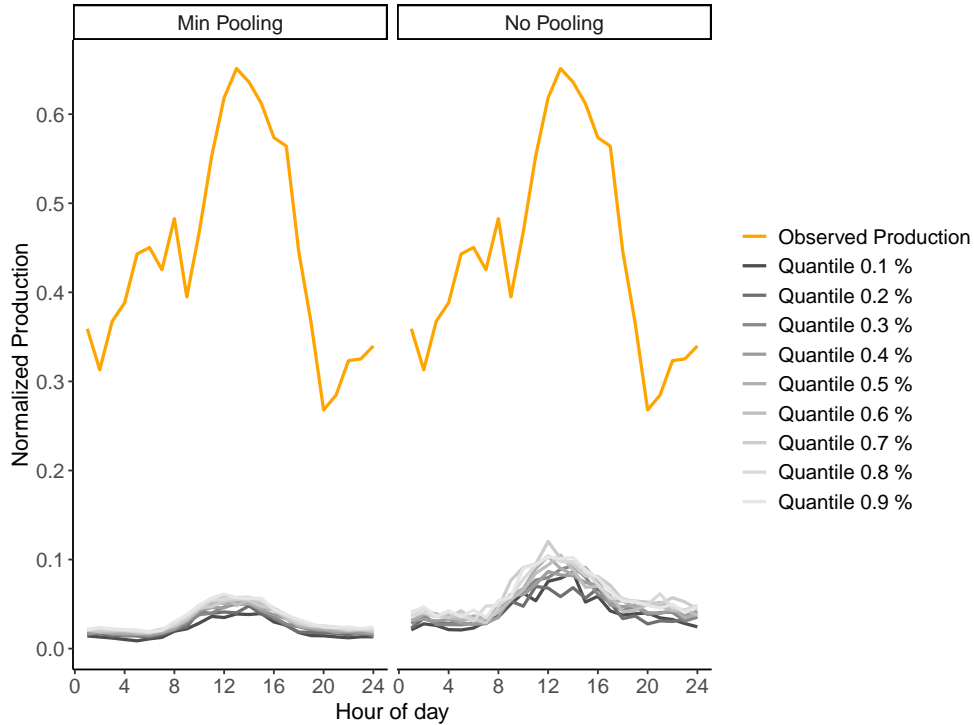
Figure 3.16: Example of forecasts obtained by the CNN regression model on VPP1, with min-pooling (left) and without pooling (right).

curve). Both models were trained by minimizing the Skill Score ('SkSc'). A variant of the CNN with min-pooling has been trained by minimizing the Quantile Score instead ('QS', purple curve). The observed forecast levels are too low, suggesting that a standard quantile regression on the QS is too conservative in this context.

Mixture density networks. The performance of mixture density networks is heavily dependent on the choice of the distributions. The Gaussian Mixture performs well in terms of CRPS and RMSE on non extremal quantiles, but lacks selectivity on tails. Quantiles below 1% are on average associated to 0 production levels, regardless of the type of network inferring parameters. It was not possible to obtain competitive results with a CNN (without min-pooling) inferring parameters, whether of Gaussian or Beta mixtures, because training was ineffective with saturating responses on the output layers. Reasons for this lack of performance should be further studied. Possible hints for adapted configurations may be found in bayesian CNNs [61], where weights follow themselves distributions obtained on batches. Results for Beta Mixture are encouraging. A Beta Mixture trained with a fully-connected network or a stateful LSTM improves the weighted Quantile Score and sharpness with respect to QRF. Similarly to the EVT, reliability is not improved. However deviations are on the safe side for the present problem: predictions are too wide, i.e. predicted levels are too low, which does not generate a risk of reserve under-fulfilment.

Results of the best configurations for each type of model are reported in Table 3.1, and compared to the QRF forecast as a reference. Models are compared on three results: the average reliability on the quantile range [0.1% - 0.9 %], the weighted QS (in log form for a more compact presentation
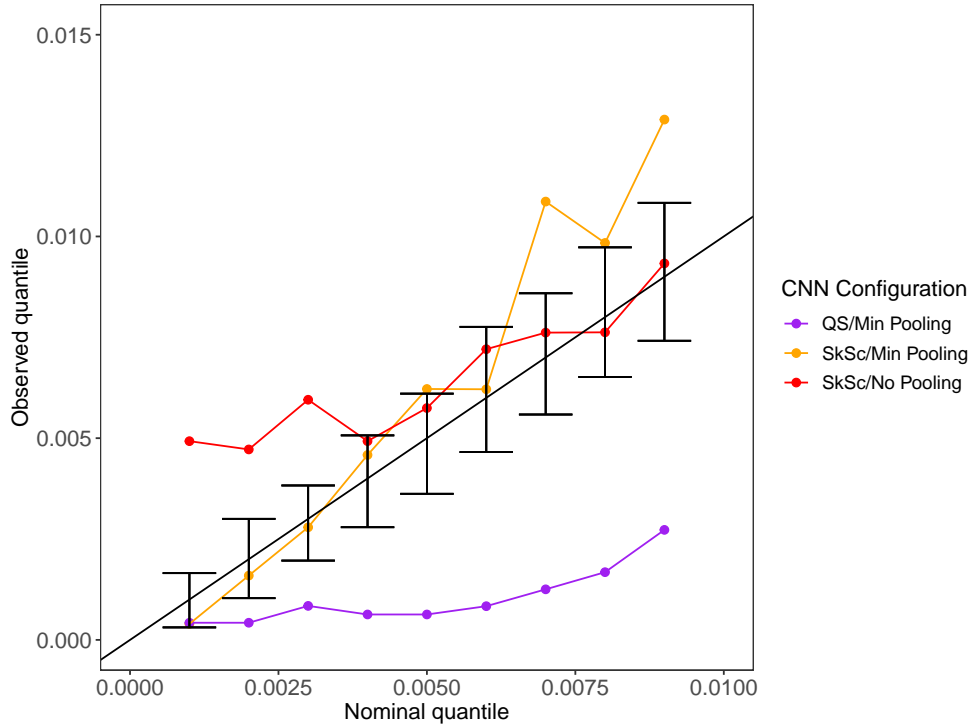
Figure 3.17: Reliability diagram for different CNN configurations

of the small scores obtained), and the sharpness between the 0.9% and 0.1 % quantiles. The two most reliable solutions of this case study are the quantile regression with CNN and the exponential distribution based on k-means clustering. The reliability diagram of QRF and the best reliable models is shown in Figure 3.18. The CNN forecast has a similar average reliability than QRF over the [0.1%-0.9%] range, but at closer scrutiny of the reliability diagram CNN forecasts have an interesting characteristic: they are the most reliable on low quantiles (below 0.5 %) where the QRF starts deviating significantly. In terms of global score, quantified in weighted QS, three models reach the best scores, with close values: EVT, quantile regression with CNN, and Beta Mixture Density Networks. The lack of reliability of the EVT model is however detrimental for the present application as it is not enough conservative (positive deviation of +0.2 %, meaning an increased risk of reserve underfulfilment when used for reserve provision). In contrast, the reliability deviation of Beta Mixtures is on the safe side (negative deviations). It would be interesting to investigate whether this behaviour is induced by the Mixture Density Networks, which tends to forecast at the bounds of the identified components of the mixture. As a final note, the forecaster disposes of two methods which are adapted to a reliable and precise forecast of lowest quantiles of aggregated VRE production:

- a simple parametric method (exponential distribution with k-means clustering), which is reliable but lacks sharpness.

- a neural network method (quantile regression with CNN), which offers good compromise between reliability and global performance, but requires more tuning effort.

Model	Best configuration	Average reliability	Log wQS	Sharpness
QRF	500 trees	-0.10 %	-7.01	2.5%
Exponential, forecast clusters	1% reference quantile, 10 clusters	-0.15 %	-7.07	3.0 %
Exponential, k-means clusters	3% reference quantile, 16 clusters	**-0.05** %	-7.02	2.5 %
EVT	4 clusters, k = 0.96	-0.20 %	-7.19	2.0 %
Quantile regression CNN	minpooling, f=16-32-64x5,k=2,s=1	-0.10%	**-7.21**	**1.6** %
Beta mixture FCNN	7 layers, 60 nodes	+0.16 %	-7.19	2.2 %
Beta mixture LSTM	4 layers, horizon 48 stateful	+0.20%	-7.17	2.0 %
Gaussian Mixture LSTM	4 layers, horizon 48 stateful	+2.0%	-6.89	0.0 %

Table 3.1: Summary of forecasting scores for the best of configurations of all models proposed. Results on VPP1.
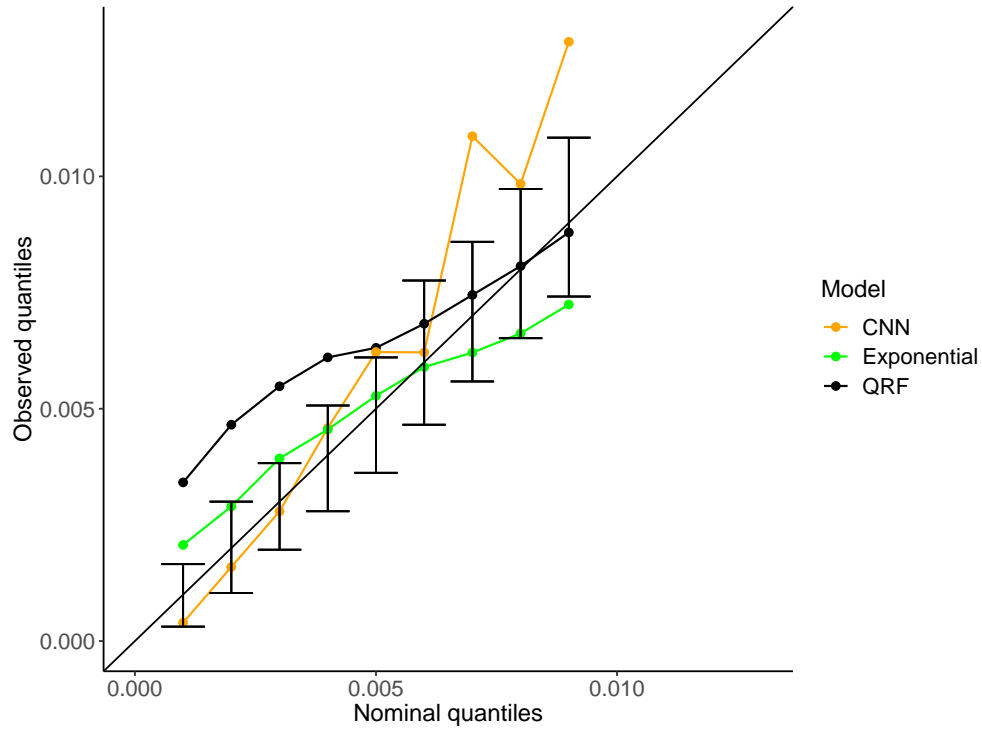
Figure 3.18: Reliability diagram presenting most reliable models compared to QRF
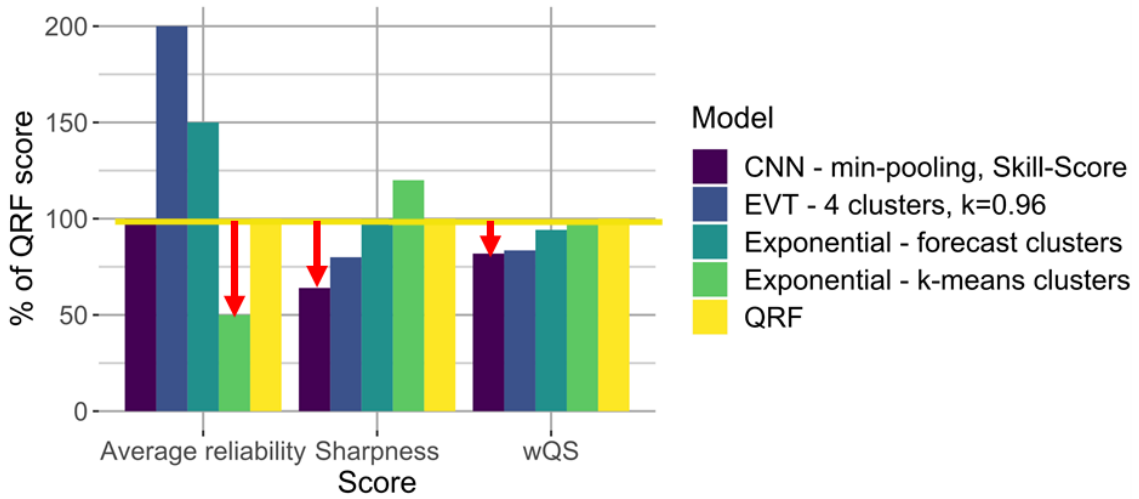
Figure 3.19: Summary of results for best models for forecasting extremes of aggregated VRE production, in terms of average reliability, sharpness on the [0.1%-0.9%] interval, and wQS. Scores are scaled with reference to the QRF. VPP1.

3.10   Conclusion

In this chapter, specific models have been developed for forecasting extremes of aggregated VRE production with improved reliability compared to decision-tree based approaches on quantiles below 1%. The series of models and the comparison of their performance answers Research Question 4. The performance of the best performing models is summarized in Figure 3.19, where scores are scaled by the scores obtained by the QRF as a reference.

The first type of solutions investigated is to develop parametric models specifically designed for forecasting extremes. An exponential distribution model is fitted on a reference quantile of the QRF prediction, which is chosen as the lowest reliable quantile in the distribution. Rates of the exponential distribution are fitted on clusterized median forecasts, following the assumption that the distribution of extremely extreme production events depends on the production regimes. The methodology is tested at several low quantile levels between 1% and 5%. An alternative clustering by k-means on production forecasts and weather conditions reaches adequate reliability for any reference quantile. The EVT model increases the flexibility of the parametric approach. Clustering of median forecasts is applied to improve the selectivity of the model. The EVT has a better global score than exponential distribution due to its higher sharpness, but is less calibrated than exponential distributions and QRF. The best parameteric approach in the context of forecasts for reserve is therefore the exponential distribution with k-means clustering, with 50% improvement in terms of average reliabity compared to QRF.

The second type of solutions improve the quantile regression proposed in the previous Chapter for the context of extremes. Quantile regression with CNN, configured for extremes with a specific min-pooling layer which captures minimal values of features, enables to reach a good balance between increase of reliability on the lowest quantiles and global score with respect to QRF, reaching

the best improvement of nearly 20% on wQS. Finally, mixture density networks are effective when based on Beta distributions, which are coherent with the bounded process of aggregated production. Gaussian distributions work fine for non-extremal quantiles, but should be avoided for the forecast of very low quantiles because their lowest marginal densities are too low and create mostly constant forecasts at 0 production level. A Beta Mixture trained by a neural network, fully-connected or LSTM, reaches the best performance in global score but is too conservative.

In conclusion, the simple model of exponential distributions may suffice for practical implementation of highly reliable forecasts for reserve bidding. But the CNN regression and Beta Mixture Density Network, although more complex to set up, offer a promising alternative. Being based on neural networks, the presented results can be potentially improved with optimized architectures of networks in terms of sizing, initialization or regularization.

3.11 Résumé en français

Besoins et enjeux de la prévision des extrêmes de production agrégée

Les services système étant vitaux pour le fonctionnement des réseaux, une offre de réserve doit être extrêmement fiable, ce qui est difficile à assurer pour un fournisseur de réserve variable. Toutefois les règles de pré-qualification pour la fourniture de services système en fréquence sont en évolution, notamment pour permettre la participation d'entités dont la réponse peut dévier de la consigne demandée, comme les renouvelables variables ou les consommateurs flexibles. A titre d'exemple, les gestionnaires de réseau allemands autorisent la fourniture de réserve si les responsables de réserve démontrent que leurs taux de défaillance sont inférieurs à des seuils de fréquence et d'amplitude définis [99]. Plus précisément, une disponibilité de 100% du volume de réserve est exigée durant toute la durée de livraison, et 95% des pas de temps pour lesquels la réserve fournie sont contenus dans un corridor autour de la consigne requise par le gestionnaire de réseau. Dans ce chapitre, on cherche à approcher l'objectif le plus contraignant, à savoir une disponibilité de 100%, similaire à celle d'une centrale pilotable conventionnelle. L'enjeu pratique pour une agrégtion variablee consiste à prévoir un niveau de production agrégée qui maximise sa disponibilité, tout en étant significativement supérieur à 0.

Cette disponibilité est directement issue de la fiabilité du modèle de prévision probabiliste de production: l'offre de réserve est basée sur la sélection d'un quantile bas de la distribution de production prévue. Si l'on veut évaluer la fiabilité, quel est le niveau de fiabilité maximum vérifiable en conditions opérationnelles? Compte tenu de la dimension usuelle des ensembles de données de production disponibles pour l'évaluation de la prévision (par exemple 1 an de mesure contient de l'ordre de 10^4 points à résolution horaire, présentant une auto-corrélation suffisamment faible pour être considérés comme des réalisations indépendantes du processus de production), et de la contrainte de disposer d'au moins 10 observations pour effectuer une statistique sur des quantiles extrêmement bas (inférieurs à 1%), la fiabilité maximum atteignable dans ces conditions est limitée à 99.9%, soit un quantile minimum de 0.1%.

Plusieurs travaux ont également observé une fiabilité moindre des modèles de prévisions probabilistes standards lorsqu'on les applique aux extrêmes, et ce dans plusieurs domaines (production éolienne [76], Dynamic Line Rating [77], prix de l'électricité [128]). Cette limite de fiabilité est due principalement à deux facteurs distincts. Premièrement, les modèles de régression quantile présentés au Chapitre 2 sont des modèles généralistes, qui reproduisent correctement l'incertitude globale de production mais ne sont pas spécifiquement adaptés à capter des évènements extrêmes. Deuxièmement, appliquer une fonction de perte quantile sur un niveau de quantile proche de 0 ne génère pas de différence significative sur des modèles régressifs peu profonds, qui comme le QRF, ont une réponse non-linéaire limitée à quelques interactions simples entre leurs variables internes [2]. Les modèles paramétriques présentés en Introduction (cf. Section 1.3.2) permettent, sans devoir recourir à une architecture d'apprentissage complexe, d'inférer une distribution probable

---

[2]la profondeur d'arborescence d'un arbre de QRF est de l'ordre de $\mathcal{O}(D)$ où D est le nombre de variables explicatives, alors qu'un réseau de neurones convolutif à L couches cachées, disposant chacune de F filtres de dimension (m, n) et approximant la densité de probabilité résultante avec Q quantiles, rassemble un ensemble de paramètres de l'ordre de $\mathcal{O}(LFmnQ)$, ce qui dépasse facilement $\mathcal{O}(D)$

des extrêmes, à partir d'observations extrêmes peu nombreuses. Pour rappel, parmi les modèles paramétriques adaptés aux extrêmes on trouve les distributions exponentielles [76], les distributions Pareto établies par théorie des valeurs extrêmes [79], ou encore les mélanges de densité [88]. Les réseaux de neurones sont quant à eux capables d'évaluer des interactions complexes entre les variables explicatives afin de reproduire les phénomènes extrêmes. L'enjeu principal consiste ici à adapter les modèles mentionnés ci-dessus au problème de la production agrégée variablee multi-sources.

Méthodologie

L'objectif global de ce Chapitre est donc de développer des modèles de prévision agrégée variablee multi-source ayant une fiabilité supérieure à 99% pour les quantiles bas. Ceci répond à la Question de Recherche 4. Ce chapitre propose ainsi deux approches: des modèles statistiques paramétriques spécifiquement dédiés aux quantiles extrêmement bas, et une régression quantile par réseaux profonds. On verra que les deux approches peuvent être couplées dans des réseaux de neurones à mélange de densité (l'apprentissage des paramètres du modèle statistique de mélange est effectué sur des réseaux de neurones). L'enjeu consiste ici à identifier les configurations adéquates de ces modèles compte tenu des caractéristiques de la production agrégée variablee:, à savoir un phénomène borné avec saturations des différentes sources d'énergie, et une dépendance plus ou moins marquée à l'horizon selon le mix des sources.

Les développements particuliers pour chacune de ces approches sont les suivants :

- **Modèles statistiques paramétriques:** Le premier modèle de prévision paramétrique proposé est basé sur la calibration d'une *distribution exponentielle* sur la partie basse de la distribution. Reprenant l'approche définie dans [76], cette distribution est d'abord établie sur des clusters de prévision médiane. Une *première proposition originale* vis-à-vis de l'état de l'art consiste à effectuer un clustering non-supervisé par l'algorithme k-means, afin de partitionner les observations en fonction des conditions climatiques et des prévisions de production disponibles. Un second modèle paramétrique basé sur la théorie des valeurs extrêmes apporte plus de garanties théoriques et une modélisation plus flexible de la queue de distribution. Enfin le mélange de densités de probabilité offre une alternative au clustering, chaque distribution composant le mélange pouvant capter une partie du phénomène de production agrégée. La deuxième proposition originale revient ici à *configurer le mélange de densité en adéquation avec l'objectif de détection des extrêmes de production agrégée*, notamment avec un mélange de densités Beta adapté aux phénomènes bornés comme la production renouvelable.

- **Régression quantile par réseaux de neurones:** Les modèles basés sur les réseaux convolutifs et récurrents présentés dans le Chapitre précédent sont adaptés à la régression sur les quantiles extrêmement faibles, avec l'incorporation de deux propositions nouvelles: (1) Une couche de filtrage spécifique est intégrée aux réseaux convolutifs et (2) une fonction de perte spécifique à l'apprentissage des extrêmes. Les réseaux récurrents sont paramétrés pour transmettre l'information mémorisée entre les batchs de données, permettant de mieux conserver l'observation d' extrêmes au sein d'une séquence. Enfin ces réseaux sont utilisés,

après comparaison avec des réseaux multicouches classiques, pour apprendre les paramètres des mélanges de densités évoqués plus haut.

Prévision des extrêmes par distributions exponentielles

Ce premier modèle consiste à calibrer une distribution exponentielle sur les quantiles inférieurs à un quantile de référence α_{lim}, dont la fiabilité est jugée suffisante (déviation inférieure aux incertitudes liées à l'effet d'échantillonnage induit par la sélection d'une période d'évaluation). Les dépassements z de puissance agrégée observés sous le quantile de référence pendant la période d'entraînement $\mathcal{T}$ sont supposés suivre la densité de probabilité définie en (3.32), dépendant d'un paramètre λ.

$$z_t = [\hat{y}_{agg,t}^{(\tau_{lim})} - y_{agg,t}]^+, \forall t \in \mathcal{T} \tag{3.31}$$

$$f(z, \lambda) = \lambda e^{-\lambda z}, z \geq 0 \tag{3.32}$$

Toutefois la distribution des extrêmes est conditionnée par le régime de production: il est connu que l'incertitude de production PV est plus grande lorsque le régime de production PV est élevé, et que l'incertitude de production éolienne est maximale dans des régimes de vent moyens [30]. Nous proposons alors, suivant l'exemple de [76], de classifier les régimes de production suivant la prévision médiane agrégée $\hat{y}_i^{(50\%)}$. Cette classification s'effectue en décomposant la prévision médiane selon C clusters, constitués par des intervalles de même effectif.

$$\hat{\lambda}_c = \frac{1}{\frac{1}{N_c} \sum_{i=1}^{N} z_i 1_{\hat{y}_i^{(50\%)} \in I_c}} \forall c \in C \tag{3.33}$$

Une dernière amélioration de ce modèle consiste à remplacer le clustering naïf précédent par un clustering permettant d'améliorer la séparabilité et la densité des clusters. Parmi les modèles de clustering possibles, nous implémentons le k-means clustering qui permet d'obtenir simplement des clusters de conditions décrivant la production agrégée. Cette méthode n'étant réalisable qu'avec un nombre limité de variables explicatives, on sélectionne empiriquement les variables explicatives qui permettent de décrire les régimes de production agrégée et leurs extrêmes: minimum des prévisions météorologiques au sein des centrales de même source et prévision médiane de production agrégée.

Théorie des valeurs extrêmes

La modélisation paramétrique de la distribution des extrêmes peut être améliorée en remplaçant la distribution exponentielle par une distribution plus flexible, et en inférant ses paramètre à partir de la théorie des valeurs extrêmes, qui bénéficie de garanties théoriques. En effet, le théorème de Fischer-Tippet-Gnedenko montre qu'une suite croissante de variables aléatoires indépendantes converge en loi vers une distribution de type valeur extrême généralisée. On en déduit ensuite que la fonction de survie pour un faible quantile τ d'observations supérieures à un seuil u suit une distribution de Pareto du type:

$$\bar{F}(\tau) \sim u_{ext} + \frac{\sigma}{\gamma}[(\frac{\tau}{\bar{F}(u_{ext})})^{-\gamma} - 1], \tau \to 0 \tag{3.34}$$

Dans le cas présent on s'intéresse aux quantiles bas, on utilise donc la théorie en minimisant l'opposé de la production de telle sorte que $max(-y) = -min(y)$. Les paramètres σ, γ sont estimés à partir des données clusterisées comme présenté pus haut.

Régression quantile par réseaux de neurones

Les architectures de régression utilisées sont similaires à celles présentées dans le Chapitre 2, à savoir une régression sur des quantiles multiples, reliés à une dernière couche de neurones denses, auxquels on applique pour chacun la fonction de perte quantile correspondante. Toutefois, compte tenu du faible niveau de perte associé à des prévisions conservatives lorsque le quantile est inférieur au 1%, et que les modèles de prévisions standard souffrent d'un manque de fiabilité aux quantiles très faibles, les réseaux sont entraînés avec le *Skill Score* défini en (3.35) , permettant de pénaliser à la fois le manque de finesse et le manque de fiabilité de la prévision[86]:

$$SkSc^\tau(y, \hat{y}^\tau) = -(1_{y-\hat{y}^\tau} - \tau)(y - \hat{y}^\tau) \tag{3.35}$$

Ensuite, les réseaux eux-mêmes sont modifiés pour s'adapter au problème de la prévision des quantiles bas. Ainsi, les réseaux convolutifs sont équipés d'une couche de filtrage supplémentaire permettant de retenir uniquement les valeurs les plus faibles observés par les filtres convolutifs. Cette étape, appelée ici *min pooling* (par opposition au *max pooling* couramment utilisé en classification [51]) sélectionne les valeurs minimales de la couche inférieure $l - 1$ au sein de la fenêtre w du filtre paramétrée par son pas ou *stride s* et sa taille en pixel k.

$$h^{(l)}_{minpool,k,s} = min(\{h^{(l-1)}_{i,j}, (i,j) \in w(s,k)\}) \tag{3.36}$$

Les réseaux récurrents LSTM sont eux équipés d'une fonction de transmission de la mémoire entre batchs, aussi appelée *stateful* dans la littérature, afin de mieux mémoriser l'apparition d'extrêmes dans les séquences observées.

Régression par mélange de densités

Nous pouvons combiner la capacité des réseaux de neurones à traiter des problèmes non-linéaires avec une distribution paramétrique représentant la production agrégée par un mélange de densités D_k. On peut prendre en compte l'influence de variables explicatives x sur les paramètres $\theta = (\pi, \sigma, \mu)$ du mélange (proportions de chaque composant π_k, paramètres de chaque composant μ_k, σ_k).

$$y_{agg}|x \sim \sum_{k \in [1,K]} \pi_k(x) D_k(\mu_k(x), \phi_k(x)) \tag{3.37}$$

Un tel mélange permet de représenter la distribution des extrêmes avec flexibilité et d'utiliser l'ensemble de l'information contenue dans les variables explicatives, là où une distribution de type

Pareto n'utilise qu'un sous-ensemble de points pour lesquels la production est au-delà du seuil défini.

Les mélanges de densité dépendent de variables latentes associant les observations aux différentes distributions composant le mélange. Il est possible d'inférer les variables latentes du mélange grâce à un algorithme de type *Expectation-Maximization*, mais qui est ici problématique: il est sensible à l'initialisation, et ne donne pas de garanties sur la convergence du mélange obtenu. Une autre méthode consiste à recourir à une approche *bayésienne*, qui modélise les paramètres comme des variables aléatoires. La probabilité des valeurs prises par les paramètres évolue a posteriori en fonction des données observées, à partir de l'estimation d'une vraisemblance obtenue par un modèle régressif et du choix de distributions fait a priori en fonction de caractéristiques attendues du problème.

$$p(\theta|y,x) = \frac{p(y|x,\theta)p(\theta)}{p(y|x)} \tag{3.38}$$

Parmi les méthodes bayésiennes disponibles pour traiter de l'inférence des mélanges, la méthode dite d'*inférence variationnelle stochastique* est retenue. La distribution postérieure $p(\theta|y,x)$ est décomposée en la somme d'une vraisemblance (approchée ici par réseau de neurones) et d'une divergence entre les distributions marginales des paramètres et leurs distributions a priori. Ceci repose sur l'hypothèse que les paramètres de θ sont indépendants. Nous appliquons cette méthode à un mélange de P lois Gaussiennes, avec P le nombre de centrales dans l'agrégation. Chaque loi marginale est paramétrée par la moyenne et la variance de chaque centrale. Toutefois la production agrégée étant bornée, nous évaluons également un mélange de deux lois Beta ayant même précision (le mélange de P lois Beta quelconque est difficile à traiter car peut diverger facilement).

Cas d'étude

Nous évaluons les modèles de prévisions proposés par la méthode de validation croisée sur les jours de la semaine effectuée au Chapitre 1. L'objectif est d'évaluer la performance des modèles sur plusieurs saisons. Le VPP étudié est identique à celui présenté en Section 2.4.5.

L'évaluation des modèles repose sur des métriques spécifiques à l'analyse des prévisions extrêmes. Ces méthodes sont présentées synthétiquement dans la section suivante.

Métriques d'évaluation

La taille limitée des échantillons de test crée des biais dans l'évaluation de la fiabilité. On intègre l'influence de l'échantillon dans le diagramme de fiabilité des prévisions, en quantifiant l'intervalle de confiance associé à chaque quantile à prévoir selon le *consistency resampling* proposé par [125].

Le score global utilisé pour évaluer la prévision est similaire à la perte quantile pondérée utilisée pour l'entraînement des modèles de régression. Il s'agit du score quantile évalué sur N observations, et pondéré par l'écart quadratique de la prévision quantile $\hat{y}^{(\tau)}$ à 1. Ainsi, les prévisions associées à des quantiles très faibles ne sont pas négligées dans le score global.

$$wQS_\tau(\hat{y}, y) = (1-\tau)^2 QS_\tau(\hat{y}^{(\tau)}, y) = (1-\tau)^2 \sum_{i=1}^{N} max(\tau.(y_i - \hat{y}_i^{(\tau)}), (1-\tau).(y_i - \hat{y}_i^{(\tau)})) \quad (3.39)$$

Résultats

La Table 3.2 rassemble les résultats obtenus par les meilleures configurations des différents modèles. Plusieurs modèles obtiennent des scores globaux proches (réseaux à mélange de densités Beta, CNN, EVT). Parmi les modèles paramétriques, la distribution exponentielle dispose d'une meilleure fiabilité si le clustering s'effectue par k-means plutôt que par clusters de prévision médiane. Le modèle par théorie des valeurs extrêmes obtient un des meilleurs scores globaux, mais une mauvaise fiabilité et ne devrait donc pas être utilisé dans le contexte d'une prévision pour offre de réserve. De même, les réseaux à mélange de densités Gaussiennes sont déconseillés, car si celles-ci conviennent pour une prévision probabiliste usuelle, elles manquent de sélectivité aux extrêmes et donnent des prévisions en moyenne nulles sur les quantiles bas. Par contre, le mélange de densités Beta, adapté au caractère borné de la production agrégé, obtient de bons scores globaux pour différents types de réseaux de neurones effectuant l'inférence (perceptron multi-couche fully-connected FCNN ou LSTM). La fiabilité est mauvaise mais du sens opposé à celle de l'EVT, à savoir que les prévisions sont trop conservatives (trop basses). Ceci n'est pas un obstacle à l'offre de réserve, car cela n'induit pas de risque de défaillance sur la réserve. Enfin la régression quantile par CNN, équipé de la couche de min-pooling, donne une fiabilité acceptable et obtient un des meilleurs scores globaux.

Model	Best configuration	Average reliability	Log wQS	Sharpness
QRF	500 trees	-0.10 %	-7.01	2.5%
Exponential, forecast clusters	1% reference quantile, 10 clusters	-0.15 %	-7.07	3.0 %
Exponential, k-means clusters	3% reference quantile, 16 clusters	-0.05 %	-7.02	2.5 %
EVT	4 clusters, k = 0.96	-0.20 %	-7.19	2.0 %
Quantile regression CNN	minpooling, elu f=16-32-64x5,k=2,s=1	-0.10%	**-7.21**	**1.6** %
Beta mixture FCNN	7 layers, 60 nodes, Relu	+0.16 %	-7.19	2.2 %
Beta mixture LSTM	4 layers, horizon 48 stateful	+0.20%	-7.17	2.0 %
Gaussian Mixture LSTM	4 layers, horizon 48 stateful	+2.0%	-6.89	0.0 %

Table 3.2: Scores de prévision pour l'ensemble des modèles, meilleure configuration pour chaque modèle.

FR

Conclusion

En résumé, le modèle par distribution exponentielle convient pour une prévision des extrêmes appliquée à la fourniture de réserve. Ce modèle est relativement simple à établir, si ce n'est qu'il nécessite une prévision probabiliste préalable sur les quantiles non-extrêmaux, par exemple via QRF. Ensuite, les réseaux à mélange Beta et la régression quantile par CNN min-pooling représentent les solutions les plus performantes, en termes de score global ou de fiabilité. Ces solutions devraient être privilégiées par des opérateurs souhaitant optimiser cette prévision, car les architectures des réseaux de neurones sous-jacents à ces deux méthodes ont un potentiel d'apprentissage supérieur à celui des distributions exponentielles.

FR

Chapter 4

Optimal offer of ancillary services from a renewable VPP

4.1 Introduction

The operator of a VRE-based VPP bidding on balancing AS markets needs to decide the quantity of active power reserve to offer. Informed on the global level of uncertainty by the direct production forecast of Chapter 2, and on balancing capacities with maximum reliability by the forecasting of extremes in Chapter 3, the operator wants to maximize earnings and minimize the probability of facing large financial losses or not being able to provide the offered reserve capacity.

Optimal decisions for VRE producers selling on the wholesale energy market have been extensively studied, the reader is referred to [143] for a comprehensive introduction on the matter. However bidding aggregated VRE production in balancing markets raises specific challenges that have not been addressed sufficiently in the literature:

1. The difference in price between reserve markets and the wholesale energy market will drive the decisions of the VPP operator. But prices on energy and reserve markets are unknown when the VPP operator define bids. The price difference and its uncertainty should be forecasted.

2. The utility of probabilistic forecasts and scenarios of aggregated VRE production for a bidding strategy in AS markets should be demonstrated.

3. In balancing capacity markets sized by static methods, the average activated reserve volume is lower than the contracted volume. Consequently the activation of the VPP is largely uncertain, and the probability of activation must be anticipated.

4. It is not clear how to optimize revenue on energy and reserve markets while controlling the risk of not providing reserve.

The main **objective** of this Chapter is to develop a strategy for optimizing the offer of the aggregated production of a VRE-based VPP, which has the capacity to provide one or more balancing AS on top of energy. The concept is illustrated in Figure 4.1. Reserve bids in red on the left figure are concentrated in periods identified as profitable. The optimization of reserve and energy

bids integrates the information on production uncertainty conveyed by the probabilistic forecasts. The right figure summarizes the problem: given a prediction of the prediction uncertainty (here depicted as a density forecast), and considering the possible market outcomes, the bidding optimization should find the reserve and energy bids that maximize revenue. These bids are constrained by the expected production level. The optimal value will depend not only on the production, but also on the uncertain price levels on both energy and reserve markets.
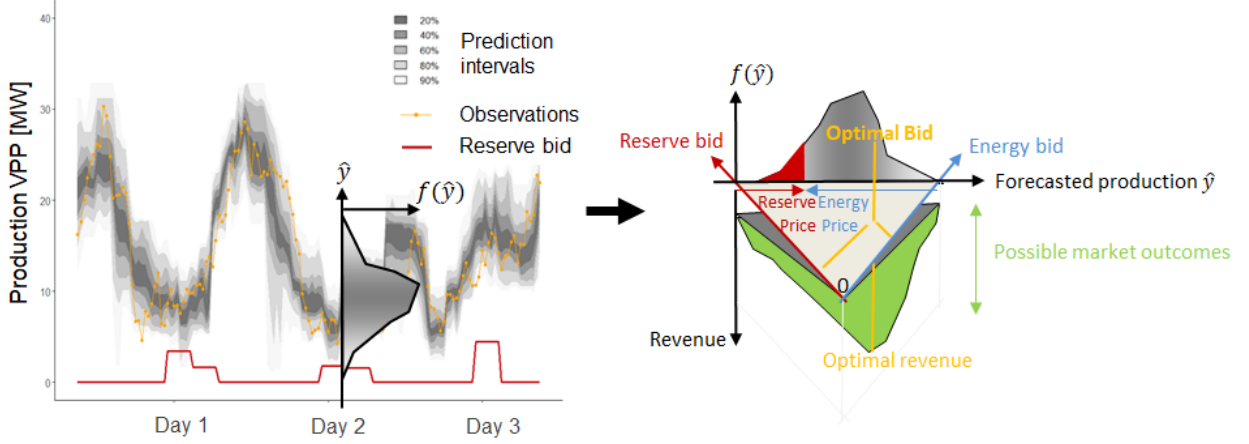
Figure 4.1: Illustration of joint optimization of energy and reserve.

4.2 Methodology

The strategy for the optimal offer of balancing AS is deployed in a gradual scale of complexity. Figure 4.2 summarizes the methods presented in this Chapter. The **Method 1** in Section 4.3 discards market conditions and bases entirely its offer on production forecasts. Highly reliable forecasts from models proposed in Chapter 3 are implemented to minimize reserve under-fulfillments. If the VPP operator instead wishes to prioritize the global profitability of the renewable production, then three models integrating market conditions are proposed:

- **Method 2**: An *optimal quantile for reserve* is derived in Section 4.4 from probabilistic forecasts of production, generated by the models in Section 2.4, and from forecasts of prices (deterministic and probabilistic). The integration of market uncertainties answers to Research Question 5.

- **Method 3**: If the VPP operator wishes to target a maximum rate of frequency under-fulfillment, then the *chance-constrained optimization* proposed in Section 4.5 enables to maximize revenue while controlling the technical risk on reserve provision. The effective variability of aggregated production is also integrated via the scenarios of aggregated production generated in Section 2.5. The consideration of both revenue and technical risk on reserve is directed towards Research Question 6.

- **Method 4**: A method for the bidding of multiple reserve products is proposed in Section 4.6. The stochastic optimization of Method 3 could be easily adapted to a multiple product context. What is proposed here is an alternative based only on probabilistic forecasts of production, without the need to produce scenarios or uncertainty regions as for robust optimization. It is based on an extension of the optimal quantile proposed in Method 2. This is directly addressing Research Question 7.
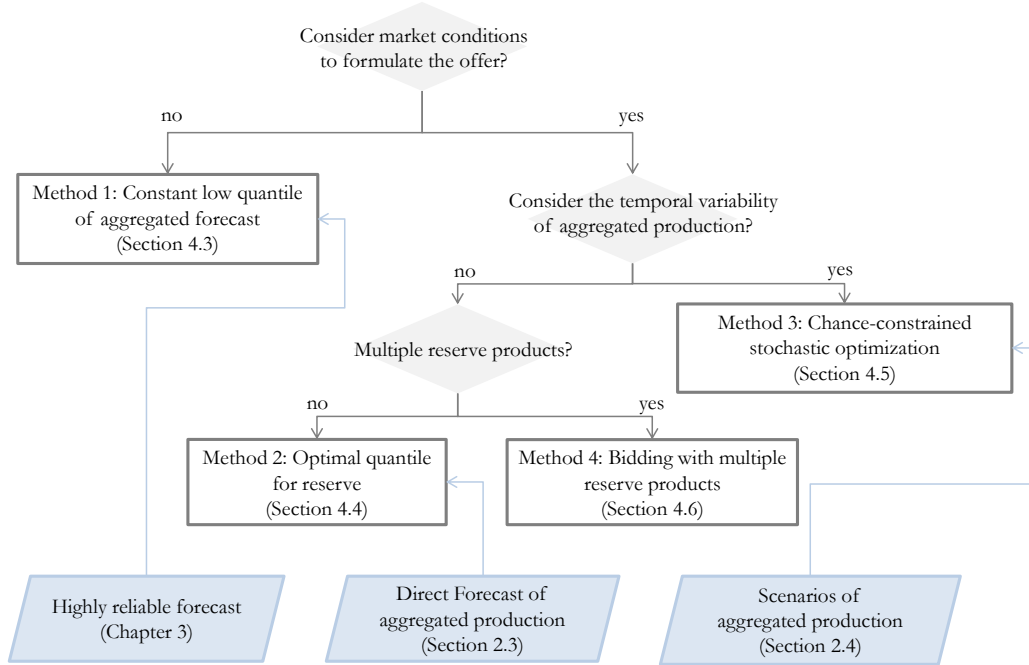
Figure 4.2: Workflow of the methodology for the optimal offer of balancing AS

The following **assumptions** hold for the market modelling:

- It is assumed that the VPP is a price-taker in the energy and reserve market. This does not imply that the power quantity offered must be independent from the prices, but rather that the VPP's possible offers would have no influence on the clearing prices.

- The decisions taken by the VPP are risk-neutral, if not stated otherwise.

- The observation of market values (prices, quantities) and their relationships with VRE production over a limited training period is assumed to inform on the future market conditions at short-term.

- TSOs allow under-fulfillments of reserve provision, and have fixed a pricing scheme for balancing penalties.

- Reserve markets are cleared in the morning of the day ahead, and the energy market is cleared before noon of the day ahead. Under this assumption it is safe to consider that the VPP operator will decide simultaneously of its bids on energy and reserve markets. Please note that this assumption is valid for bidding balancing capacities (available balancing power in MW) in most countries, but clearing for balancing energy (activated balancing energy in MWh) is expected to take place shortly before delivery (down to 15 minutes) in the coming years.

- The energy opportunity cost for upward activated reserve is remunerated by the TSOs.

- The VPP pays the TSOs for the energy not produced during activation of downward reserve.

4.3 Method 1: Constant low quantile of aggregated forecast

In this first method, which serves as a baseline, the offer strategy is independent from expected market conditions on energy and balancing AS. Its unique aim is to minimize the frequency of reserve under-fulfillment. The reserve capacity is chosen in (4.2) as the minimum of the aggregated production forecast at a low quantile over a validity period $T_{validity}$, similarly to the work of [144]. The low quantile forecast can be obtained from the quantile regression models presented in Chapter 2, or from the models dedicated to extremely low quantiles in Chapter 3.

$$R_t^{da} = min_{t' \in T_{validity}(t)} \hat{y}_{t'}^{\tau_R} \tag{4.1}$$

4.4 Method 2: Optimal quantile from forecasts of production and price

The second method aims at devising an optimal share of the uncertain production to dedicate to reserve, considering prices on the energy and reserve markets. This problem finds its roots in the theory of the newsvendor problem [31], where the newsvendor must optimize its sales considering

uncertain demand and costs associated with deviations between offer and demand. Transferred into the context of this thesis, uncertain demand becomes the uncertain production of the VPP, and the offer becomes the bid on electricity markets. Now the theory states that it is possible to derive an optimal quantile of the predictive distribution of production that should be offered on electricity markets, as a function of the expected revenues and penalties associated to the VPP bids.

Originality of this method

Optimal quantile models for the bidding of renewable production on the wholesale energy markets have been proposed by [36], and adapted to the reserve bidding for Wind power plants in [103]. The method presented here is an extension of the aforementioned works, integrating two innovations:

- The present method integrates for the first time, at the best of the author's knowledge, *realistic forecasts of all market quantities for energy and reserve*, instead of simple assumptions such as perfect or constant prices.

- A parametric dependence model between energy prices and renewable production has been proposed by [145]. In this method, a non-parametric Kernel Density Estimation (KDE) copula makes a direct use of probabilistic forecasts of renewable production and prices, enabling to *integrate uncertainties in both production and prices into the bidding strategies.*

The optimal quantile framework necessitates the assumption that the decisions of the VPP have no temporal dependence over a sequence of time steps. This assumption is justified here by the lack of decisions impacting the available production, such as trading in the intraday market or use of storage, between the day-ahead offer and the deployment of energy and reserve. This simplifies the problem of maximizing the summed daily revenue as maximizing the expected revenue over each time unit [146].

The workflow of the present method is illustrated in Figure 4.3. It starts with addressing the following issue: given that the integration of renewables in electricity markets becomes significant, should the model consider a possible dependence between the VPP production and observed market prices? In a first approach, the dependence can be ignored. The optimal quantile proposed in Section 4.4.1 is based on deterministic forecasts of prices on energy and reserve markets.
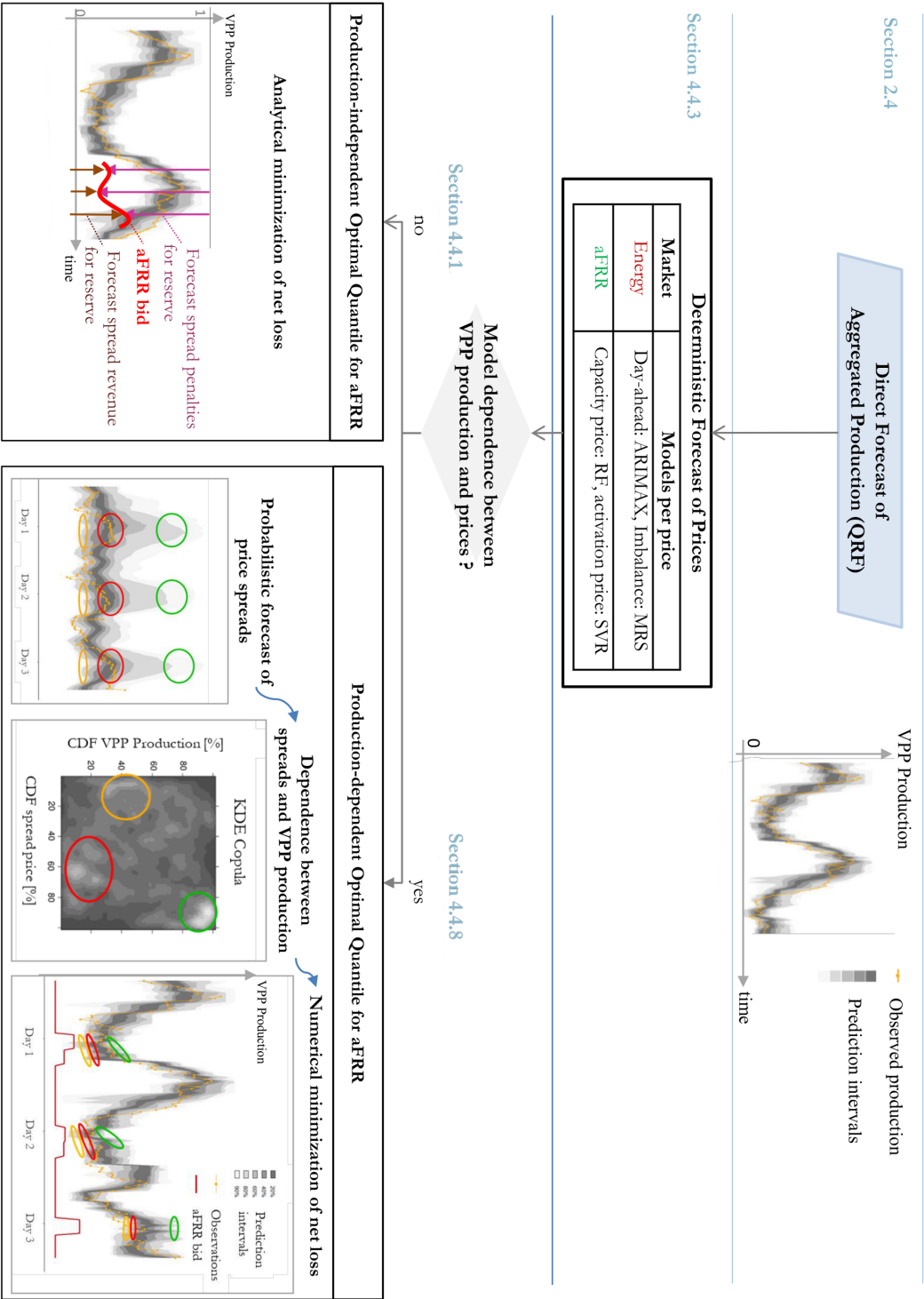
Figure 4.3: Methodology of optimal quantile based combined offer of reserve and energy

4.4.1 Production-independent Optimal Quantile for reserve

If the VPP operator integrates expected levels of prices into its strategy, then the reserve bid is based on the minimum of a quantile over a validity period $T_{validity}$ as in (4.2), but the quantile $\tau_R(t)$ is now variable, and depends on market conditions at each timestep.

$$R_t^{da} = min_{t' \in T_{validity}(t)} \hat{y}_{t'}^{\tau_R(t)} \tag{4.2}$$

After formulating losses associated to opportunity costs (energy was more profitable than reserve, curtailment has reduced profit) or missed sales (reserve was more profitable than energy, reserve bid too low), the derivative of the total loss with respect to reserve gives the optimal quantile for reserve. The derivation is given below.

The net penalty $l_{E,R}$ associated to the offer of energy and reserve is composed in (4.3) of three terminal linear loss functions: l_E accounts for penalties on the energy market, $l_{R,PI}$ corresponds to the value of perfect information on the reserve bid, and l_R quantifies penalties on the reserve market. The loss on the energy market is proportional to the net penalty price π_E^*, which refers to the price difference between the price paid for imbalances and the price of bids sold on the day-ahead market.

$$l_{E,R} = l_E(y, E) + l_{R,PI}(y) + l_R(y, R) \tag{4.3}$$
$$l_E(y, E) = \pi_E^*(E - y)$$
$$l_{R,PI}(y) = \pi_R^{\uparrow\downarrow}(y)$$
$$l_R(y, R) = \begin{array}{ll} (\pi_R^* - \pi_E^*)(R - y) & y \leq R \\ (\pi_R^{\uparrow\downarrow} - \pi_E)(y - R) & y > R \end{array}$$

In order to find the optimal reserve bid, the loss of interest is the third loss associated to penalties on the reserve market, because it is the only loss influenced by the day-ahead decision on reserve. As per the certainty equivalent theory [147], it is sufficient to search for the minimum of the loss expectation rather than its entire distribution.

$$\frac{\partial}{\partial R}\mathbb{E}(l_R) = 0 \tag{4.4}$$
$$\frac{\partial}{\partial R}\int_{y=0}^{y_{max}} l_R(y, R) f_Y(y) dy = 0 \tag{4.5}$$

After derivation with respect to reserve with the help of the Leibniz rule, the optimal quantile for reserve writes:

$$R^{opt} = F_y^{-1}\left(\frac{\Delta\hat{\pi}_{RE}^{da}}{\Delta\hat{\pi}_{RE}^{da} + \Delta\hat{\pi}_{RE}^{rt}}\right) \tag{4.6}$$

Assuming symmetrical reserve, the optimal quantile writes in (4.7) as a function of expected spread prices between energy and reserve. The forecast of these prices is explained in the next subsection.

$$\tau_R = \frac{\Delta\hat{\pi}_{RE}^{da}}{\Delta\hat{\pi}_{RE}^{da} + \Delta\hat{\pi}_{RE}^*} \tag{4.7}$$

The day-ahead spread price depends on the reserve capacity prices $\hat{\pi}_R$, balancing energy prices $\hat{\pi}_{E_R}$ and energy price $\hat{\pi_E}$:

$$\Delta\hat{\pi}_{RE}^{da} = \hat{\pi}_R^{\uparrow,da} + \hat{\pi}_R^{\downarrow,da} + \hat{a}_R^{\uparrow}.\hat{\pi}_{E_R}^{\uparrow,da} + \hat{a}_R^{\downarrow}.\hat{\pi}_{E_R}^{\downarrow,da} - \hat{\pi}_E^{da} \tag{4.8}$$

The real-time spread price quantifies the difference in net price of imbalances on both markets:

$$\Delta\hat{\pi}_{RE}^{rt,*} = \hat{\pi}_R^{rt,*} - \hat{\pi}_E^{rt,*} = \underbrace{\hat{\pi}_R^{\uparrow,-} - \hat{\pi}_R^{\uparrow,da} + \hat{\pi}_R^{\downarrow,-} - \hat{\pi}_R^{\downarrow,da}}_{\text{reserve imbalance price}} - \underbrace{\hat{\pi}_E^{rt,-} - \hat{\pi}_E^{da}}_{\text{energy imbalance price}} \tag{4.9}$$

In the equation above R stands for reserve, E for energy, E_R for energy associated with reserve activation. $\hat{a}$ is the expected probability of the VPP being activated among bidders present in the merit-order list, for upward and downward aFRR. In summary, the optimal quantile represents the balance between the lost gain opportunity (proportional to the day-ahead price spread $\Delta\hat{\pi}_{RE}^{da}$) and the increase in penalty losses (proportional to the penalty price spread in real-time $\Delta\hat{\pi}_{RE}^{rt,*}$) when reserve is more expensive than energy. One can see that estimating the optimal quantile involves forecasting two types of quantity: (1) the probability that the VPP will be activated, and (2) prices. The proposed methodology to derive and forecast these quantities is presented in the following two subsections.

The energy bid E_t^{da} is given in (4.10) considering that the balancing market is operated following a single-price paradigm. In this case the theoretical optimal bid consists in offering the installed capacity if the forecasted day-ahead price is higher than the price for imbalance, and zero otherwise to maximize the arbitrage opportunity between the day-ahead market and the balancing market [148]. Considering the high uncertainties regarding the imbalance price for the day ahead, we choose here a more conservative bid which hedges the offer against imbalance penalties: the VPP will bid the mean of its production forecast, limited by the installed capacity and the reserve offer, which has priority over energy. This approach is in line with the risk-constrained energy offer proposed by [148] in a single-price balancing market, which is also set with reference to the mean of the forecasted production.

$$E_t^{da} = min(y_{max} - R_t^{da}, \mathbb{E}(Y_t)) \tag{4.10}$$

4.4.2 Estimation of activation probability

The probability of activation of the VPP arises in the derivation of the optimal quantile in (4.7). Here, we present how this probability is estimated based on historical data from aFRR auction settlements and activations, while in the next sub-section we present how it is forecasted. It is assumed that the position of the VPP on the merit-order list (MOL) has a uniform probability. Although the low marginal costs of the renewable plants within the VPP would probably induce a

low marginal price from the VPP, it is hazardous to assume that the VPP will be systematically selected among the cheapest offers, at least in aFRR markets that are penetrated by hydro or that integrate large amounts of wind and PV. Therefore, the VPP is considered as equally likely to be located anywhere in the merit-order. The activation probability is then estimated based on (4.11) and is valid for both upward and downward activation. At each activation time step, the activation probability equals the sum of infra-marginal reserve capacities that are situated below R_{demand} the reserve activated by the TSO in the merit-order list (for these capacities, the indicator function returns 1, 0 otherwise), divided by the sum of the allocated reserve as a result of the aFRR auction.

$$a_R = \frac{\sum_{i \in MOL} R_i 1_{R_i \leq R_{demand}}}{\sum_{j \in MOL} R_j} \tag{4.11}$$

The model proposed here is based on the available history of reserve activations for all reserve suppliers. However, we can tune the model to the historical activation records of the VPP itself, if such data are available, in order to predict the VPP's actual probability of activation. This is expected to be the case in the future, when variable renewable VPPs enter the aFRR market. An alternative model for activation probability proposed by [127] is based on the density of the wind forecast error. This was devised for sizing reserve and does not include tendered reserve volumes. It is therefore less adapted to the forecast of aFRR market conditions than our model in (4.11). The activation probability obtained from this approach regarding the tendered and activated capacities on the aFRR market in Germany shows a high variability and discrepancies between upward and downward activation. In Figure 4.4 this activation probability is averaged on a rolling daily mean to show tendencies more clearly. The probability of a reserve being activated has an impact on the potential revenue of the VPP. A downward reserve paid for capacity (after being selected in the tender) that is rarely activated (no need to curtail the production and day-ahead energy offer, no payment to TSO for downward energy activation), is the best-case for the VPP profitability. Conversely, an upward reserve paid for capacity that is rarely activated is likely to constitute the worst-case for profitability, as production is curtailed and the day-ahead energy offer has to be reduced.

Forecast errors on load and renewable production are correlated to the probability of reserve activation: in a scenario of high penetration of renewable energy, an evaluation of the activation probability should consider renewable production forecast errors as an important factor. The diagram on the left of Figure 4.5 shows that the aFRR activation probabilities (January 2015-December 2016, 30-min resolution) are correlated with load forecast errors. Large negative load forecast errors, where grid operators require less reserve power than forecasted to ensure balance, are usually associated with significant reserve activation probability comprised between 0% and 40%. We observe that large positive load forecast errors, where load has been largely underestimated, are relatively rare. As expected in this case, only upward reserve is activated to supply more power to the grid. Low activation probabilities for large deviations could indicate that balance was mainly attained via services other than aFRR.
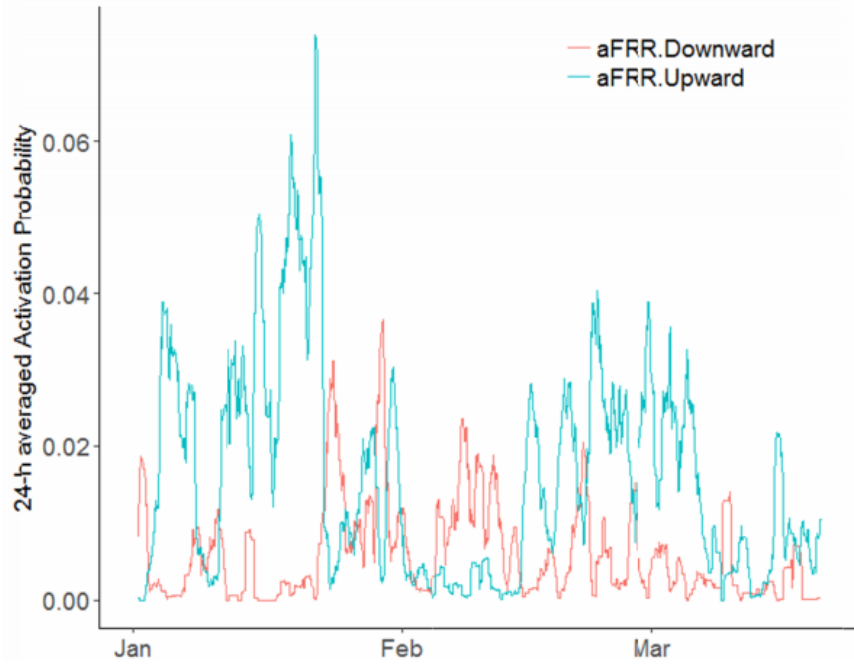
Figure 4.4: aFRR Activation probability (Germany , January-March 2016)
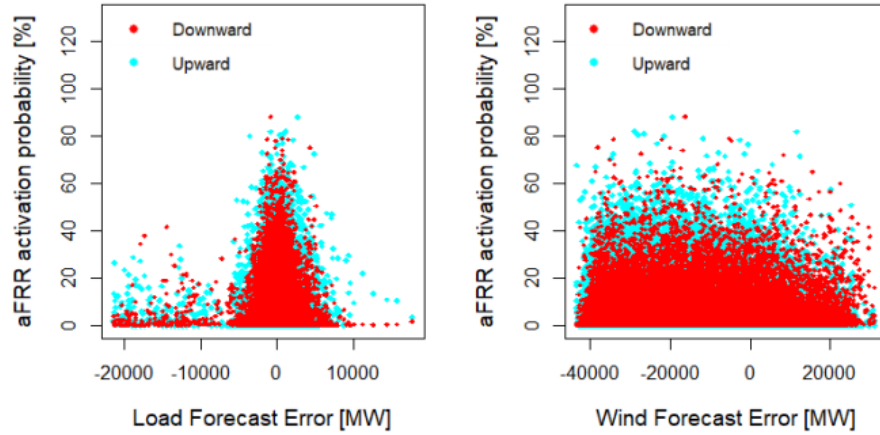
Figure 4.5: Observed aFRR activation probabilities and forecast errors on load and Wind forecasting errors, (Germany, 2015-2016)

4.4.3 Price forecasting on energy and balancing AS markets

As explained previously, the optimal quantile for aFRR defined by (4.7) calls for a forecast of the price spreads between energy and reserve. One could choose to forecast directly those spreads, however they stem from different markets so their complexity is difficult to apprehend with standard statistical models. In this section, all market conditions influencing price spreads are forecast separately in order to capture the dynamics of each market more easily.

This involves forecasting the following expected market conditions which compose the price spreads:

- day-ahead energy price and real-time imbalance energy price π_E^{da}, $\pi_E^{rt,-}$,

- aFRR average price (day-ahead capacity price + day-ahead price for balancing energy, considering activation probability) $\pi_{\bar{R}} = \pi_R^{da} + a_R \pi_{E_R}^{da}$,

- aFRR activation probability a_R.

A state-of-the art forecasting technique is implemented for each price, when available. The scope is not to optimize price forecasts, but rather to investigate how the information of all prices of energy and reserve markets may help the VPP operator in its reserve bidding. Price forecasting models are presented in details in the next sections. Reserve prices are forecasted separately for upward and downward components. Table 4.1 below summarizes the models employed:

Variable	Model	Explanatory variables
Energy price π_E^{da}	ARIMAX	lagged prices and volumes
Reserve capacity price π_R^{da}	Random Forests	lagged prices
Reserve average price $\pi_{\bar{R}}^{da}$	Support Vector Regression	reserve demand, national forecasts of load and renewable production
Reserve activation probability a_R	Support Vector Regression + PCA	past prices, reserve demand, national forecasts of load and renewable production
Energy imbalance price $\pi_E^{rt,-}$	Markov Regime Switching	past prices

Table 4.1: Summary of deterministic price forecasting models

Price forecasts are later combined to form deterministic forecasts of both spread prices.

$$\Delta \hat{\pi}_{RE}^{da} = \pi_R^{\hat{da},\uparrow} + \pi_R^{\hat{da},\downarrow} + \hat{a}_{R,\uparrow} . \hat{\pi}_{E_R}^{da,\uparrow} - \hat{a}_{R,\downarrow} . \hat{\pi}_{E_R}^{da,\downarrow} \tag{4.12}$$

$$\hat{\pi}_{E_R}^{da,s} = \frac{1}{\hat{a}_{R,s}} \hat{\pi}_{\bar{R}}^{da,s} - \pi_R^{\hat{da},s}, \quad \forall s \in (\uparrow,\downarrow) \tag{4.13}$$

$$\Delta \hat{\pi}_{RE}^{rt,*} = \hat{\pi}_R^{\uparrow,da} - \hat{\pi}_R^{\uparrow,-} + \hat{\pi}_R^{\downarrow,da} - \hat{\pi}_R^{\downarrow,-} + \hat{\pi}_E^{rt,-} - \hat{\pi}_E^{da} \tag{4.14}$$

4.4.4   Forecast of day-ahead energy price

As mentioned in the Introduction, renewable electricity producers progressively enter the electricity short-term markets. In these markets, producers offer volumes of energy on the day before delivery, with a gate closure time usually placed in the beginning of the afternoon, and a uniform pricing scheme [149]. Markets are operated at a national (FR) or multi-national level (DE/AT), supply and demand curves are harmonized in a later step by a market coupling algorithm [150].

The day-ahead energy price $\pi_{E,t+h}^{da}$ at horizon h is forecast in (4.15) at runtime t and horizon k with a seasonal Auto-Regressive Integrated Moving Average with eXogenous variables (ARIMAX) model of order (p,q,d). The ARIMAX is a validated model for short-term prediction of electricity prices [151], [78] . It can be constructed through simple assessment of the autocorrelation and partial-autocorrelation diagrams of the time series, and minimization of the Akaike Information
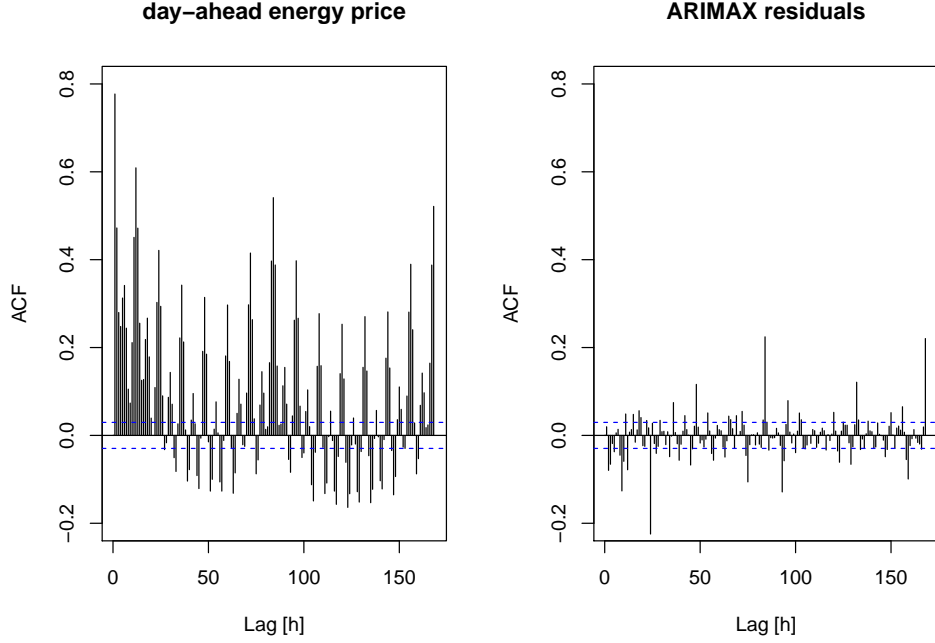
Figure 4.6: ACF of original day-ahead energy price series and ARIMA residuals

Criterion. The auto-regressive coefficients ϕ integrate a seasonal lag with the lag operator L^d of order d. A seasonal part without differentiation is added to reflect the daily periodicities observed in prices. Finally, volumes exchanged on the market during the previous day V_{t+h-24} are integrated as an exogenous variable, which helps measure the effect of demand on price. The model is fitted by means of the auto.arima function from the R forecast package. Figure 4.6 shows the ACF of the original series (left) and of the residuals (right). Residuals show remaining auto-correlation, especially on seasonal terms, so they can not be assimilated with white noise. This means that the obtained ARIMA model fails at capturing fully the dynamics in the day-ahead energy price. Notwithstanding this observation, the ARIMA model augmented with exogenous variables (ARIMAX) is retained sufficient because the aim is to cover as realistically as possible all sources of uncertainty and not optimizing each individual prediction models. An example of forecasts is displayed in Figure 4.7.

$$
\begin{aligned}
\hat{\pi}_{E,t+h}^{da} = {} & (1 - \sum_{i=[1,p]} \phi_i L_i).(1 - \sum_{i=24.[1,p]} \phi_i L_i) L^d \hat{\pi}_{E,t+h}^{da} \\
& + (1 - \sum_{j=[1,q]} \theta_j L_j).(1 - \sum_{j=24.[1,q]} \theta_j L_j) \epsilon_{MA,h} \\
& + \beta_V.V_{t+h-24}
\end{aligned}
\tag{4.15}
$$
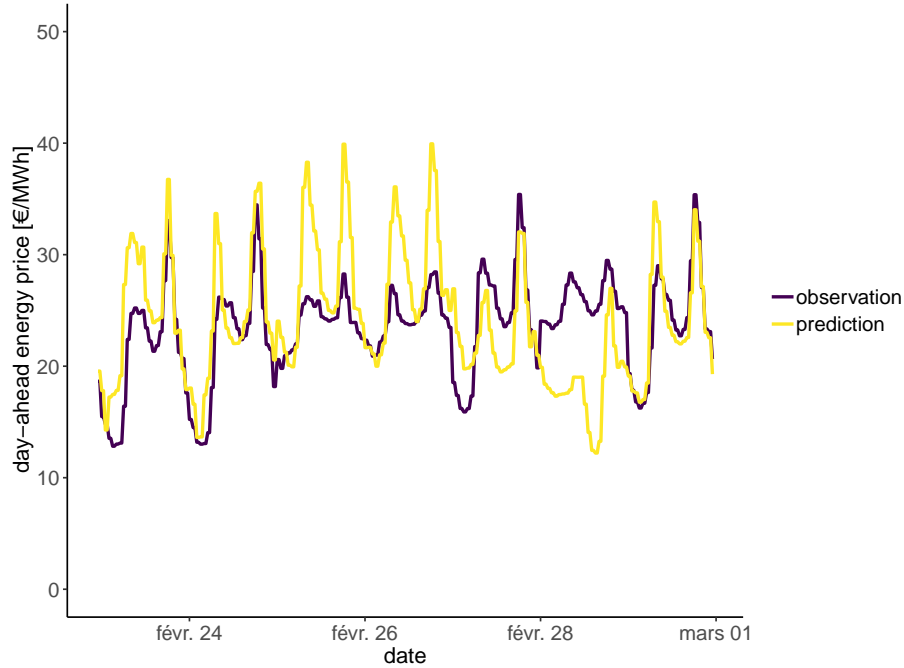
Figure 4.7: Illustration of day-ahead price forecasting

4.4.5 Forecast of imbalance energy price

Balance responsible parties, who buy and sell energy volumes on the day-ahead markets, are liable to TSOs for any imbalance observed between the volumes of their day-ahead decisions (sales or purchases) and their power injection or consumption in real-time. The imbalance is linearly penalized. As seen in the Introduction, most European countries are now adopting a single-price scheme for imbalances. These prices are meant to reflect the cost incurred by the TSO to remediate to imbalances. The real-time imbalance energy price $\pi_{E,t+h}^{rt,-}$ is therefore volatile and dependent on the amount of balancing capacity available. It shows frequent jumps which are difficult to forecast with ARIMA models [152]. We choose to forecast the imbalance energy price with a Markov Regime Switching (MRS) to reproduce the apparent price regimes with two main advantages [153]: it integrates the mean-reverting behavior observed in regulation markets, and it allows for consecutive spikes or drops.

Considering the occurrence of largely negative values in this market, we infer parameters and transition probabilities for 3 price regimes: base, spikes and drops. The day-ahead forecast is issued by binding 24 distinct MRS models calibrated on each hour hr of the day using the Expectation-Maximization algorithm. The base regime includes a mean parameter $\hat{\alpha}_{base,t+h}^{hr}$ and a mean-reverting parameter $\hat{\beta}_{base,t+h}$, while drops and spikes obey a log-normal distribution around manually-defined thresholds $TS_{drop,t+h}^{hr}, TS_{spike,t+h}^{hr}$. The forecast imbalance price is obtained in (4.16) by combining the forecasts of each regime weighted by their respective probability. Figure 4.8 represents imbalance forecasts. The price value is scaled to the maximum observed imbalance price over the period. In the illustrated period, large peaks and drops are observed. For instance, the negative price drops at the beginning of the sequence are associated to an excess of production
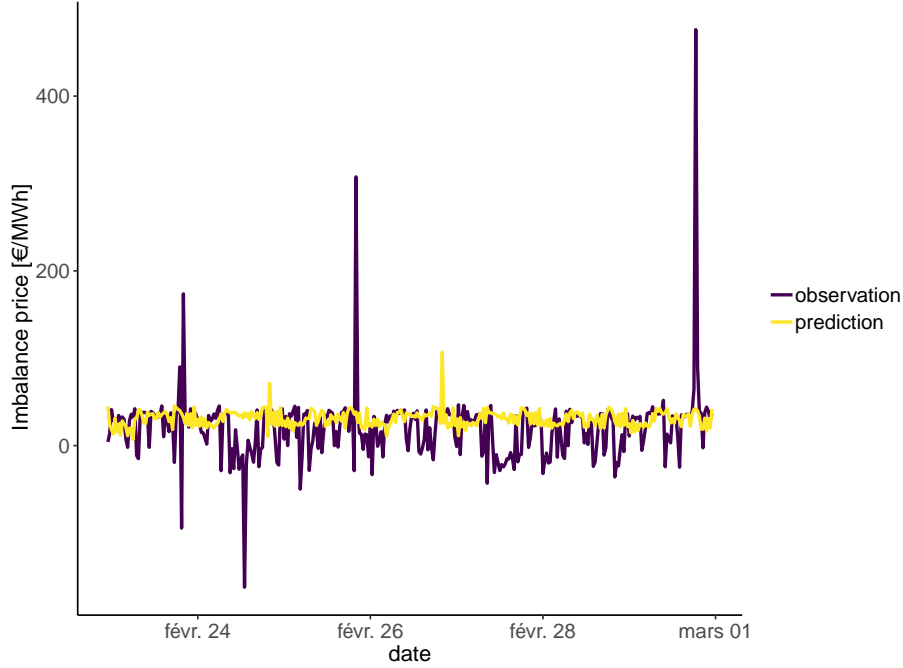
Figure 4.8: Forecasts of the energy imbalance price

and very low demand, requesting nearly full activation of downward reserve. Such events are difficult to predict with the chosen statistical model. However the model captures the average price level.

$$\hat{\pi}^{rt}_{E,t+h} = \hat{p}_{hr,base,h}(\hat{\alpha}_{hr,base,t+h} + (1 - \hat{\beta}_{base,t+h})\hat{\pi}^{rt}_{E,t}) +$$

$$\hat{p}_{hr,drop,t+h}(TS^{hr}_{drop,t+h} - e^{\hat{\alpha}_{hr,drop,t+h} + \frac{\hat{\sigma}^2_{drop,t+h}}{2}}) +$$

$$\hat{p}_{hr,spike,t+h}(TS_{hr,spike,t+h} - e^{\hat{\alpha}_{hr,spike,t+h} + \frac{\hat{\sigma}^2_{hr,spike}}{2}}) \qquad (4.16)$$

4.4.6 Forecast of aFRR price

Prices relative to aFRR are not frequently forecasted in the literature. The capacity prices for downward and upward aFRR are forecasted with a Random Forest regression model using lagged values of capacity prices as features.

Forecasts of aFRR balancing energy prices $\pi^{da}_{E_R}$ are challenging and not commonplace to the author's knowledge. As aFRR balancing energy is currently not integrated in the merit-order calculation in most countries in Europe [4], the few suppliers on the market may be tempted to bid low capacity prices and high balancing energy prices. The resulting price is highly volatile. Instead, the aFRR average price (capacity and balancing energy), $\pi^{da}_{\bar{R}} = \pi^{da}_R + a_R.\pi^{da}_{E_R}$ and aFRR activation probability a_R are forecasted with a Support Vector Regression (SVR) model (see application of SVR for day-ahead electricity price forecast in [154]). Another approach could be to directly

forecast the aFRR activation energy price if it were publicly available. The SVR approach is chosen for its ability to generalize non-linear relationships between response variable and features. These relationships are of interest in the context of aFRR activation which results from a sum of complex decisions on the energy and reserve markets. The feature vector $X_{\bar{R}}$ of dimension d gathers in (4.17) variables influencing the aFRR market, i.e. lagged aFRR demand, lagged aFRR average price, forecasted wind power $\hat{W}_t$, forecasted load $\hat{L}_t$ and forecasted system margin $\hat{M}_t$. System margin forecast helps detect imbalances in the energy market, which are likely to impact the activation price. The system margin forecast is defined as the ratio between load forecast and generation forecast [151].

$$X_{\bar{R}} = \{R_{demand,t-lags}, \pi_{\bar{R},t-l}, \hat{W}_t, \hat{L}_t, \hat{M}_t\} \tag{4.17}$$

The forecast of the average price is obtained in (4.19) by maximizing the dual objective function of an epsilon-SVR in the feature space through a functional f built with an ANOVA Kernel k [155]. The value of the forecast conditioned by a feature $x \in \mathbb{R}^N$ is obtained by fitting the multipliers $(\alpha_i, \alpha_i *), \forall i \in \mathbb{R}^N$ associated with the margin constraints and the bias b. The kernel k takes here the form of a multivariate Gaussian Kernel to capture separately the proximity in each of the features.

$$\pi_{\bar{R}} = f(X_{\bar{R}} = x) \tag{4.18}$$

$$f(x) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)k(x_i, x) + b \tag{4.19}$$

The cost coefficient of the SVR [156] is tuned to enlarge the margin and reach support vectors that explain moderate spikes, while deliberately ignoring large spikes, mostly induced by peaks of reserve activation. These peaks are generally shorter than the aFRR product length. Such as behaviour is observed in Figure 4.9, where the forecast of the average price for downward aFRR is presented. The usual aFRR price levels are correctly reproduced, but the large price drops mentioned earlier are not anticipated by the model.

4.4.7 Forecast of aFRR activation probability

The activation probability, as derived in Section 4.4.2, is done with the SVR model presented above. Regression is found to improve with building one model per hour of the day and with applying Principal Component Analysis (PCA) on the input variables. Results of the forecast illustrated in Figure 4.10, show that the model is able to individuate periods with higher activation probability, but constantly underestimates the magnitude of activations.
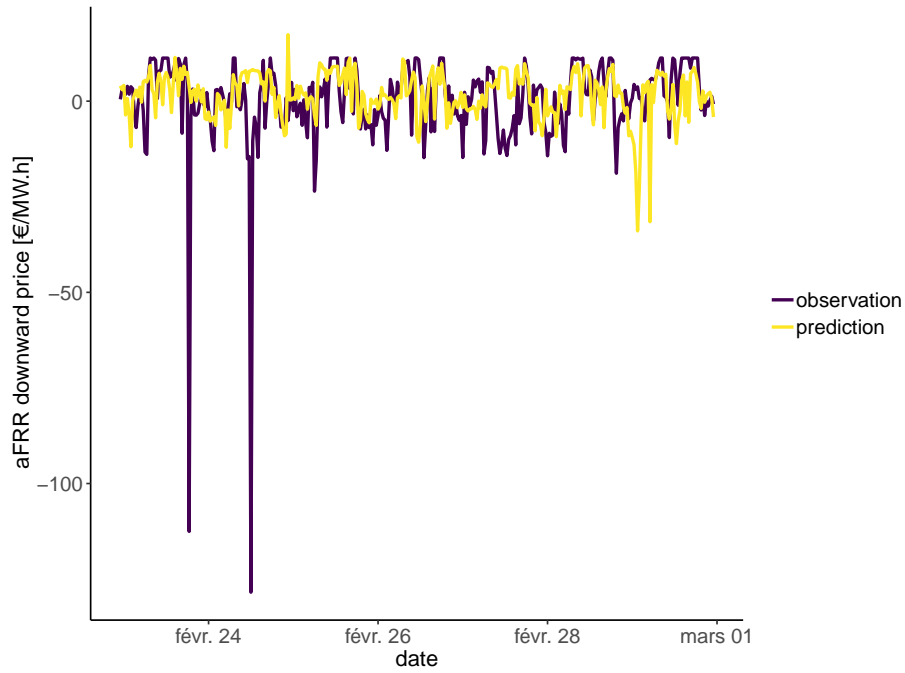
Figure 4.9: Forecasts of the aFRR average price, example for downward reserve
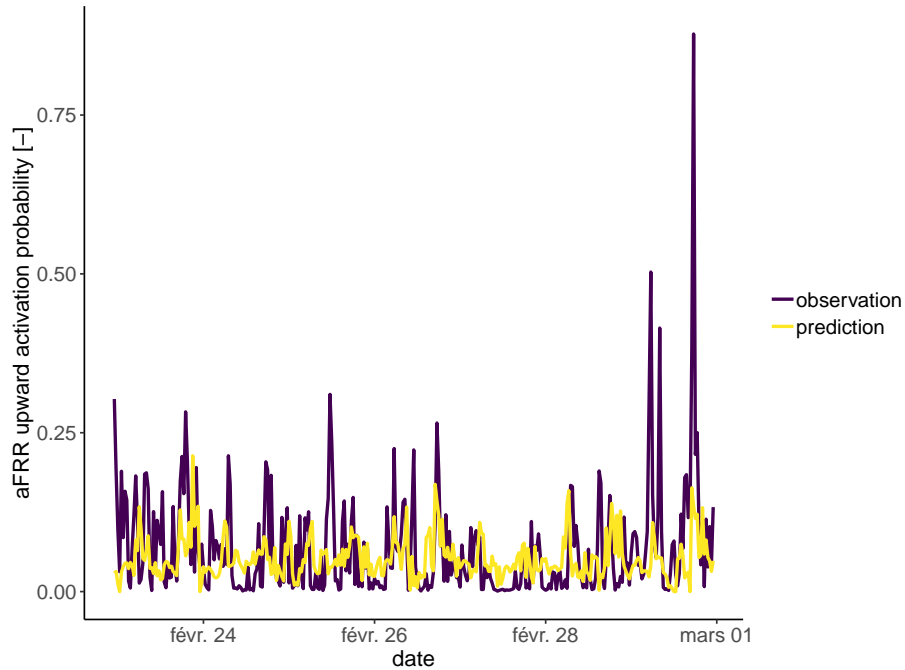
Figure 4.10: Forecasts of the aFRR activation probability, example for upward reserve

4.4.8 Optimal Quantile Dependent on Prices and VPP Production

In a context of very high penetration of renewables, it is possible that VPP production will show some correlation with energy and reserve prices. A significant degree of correlation has been

144

Price	MI Price-Production
aFRR upward	0.246
aFRR downward	0.209
day-ahead energy	0.207
FCR	0.156
real-time energy	0.053

Table 4.2: Mutual Information between VPP production and prices, German market 12/2018-03/2019

observed and modelled by [145] between renewable production and prices on energy markets penetrated largely by Wind. Now assuming that a given VRE plant has a production pattern quite close to the production at market level, then the correlation between may production on site and prices may be used to inform bidding strategies [145]. If this approach is valid at a plant level, then a VPP gathering plants in different regions of a market is even more likely to have a good match between its production and the VRE production at market level.

It is legitimate to question whether this idea is also valid for markets of balancing AS. Figure 4.11 shows the joint density of wind production of a VPP and prices observed on energy and markets of Germany during a 3-month winter period. Visually, zones of higher densities emerge at low production levels, which are also the most frequent. These are associated mostly to nonextremal prices. Reserve prices appear as correlated as energy prices. This is corroborated by the evaluation of the *Mutual Information* (MI) between production and prices. If the MI of two variables x and y is close to 0, then the dependence between x and y is low. The MI associated to Figure 4.11 are reported in Table 4.2: aFRR and FCR prices have significative dependence levels with the VPP production, even higher than the imbalance prices for real-time energy. One can then deduce that considering the dependence between prices and production is worth of interest.

With the assumption of dependence between prices and VRE production, the optimal quantile for reserve depends on the expected price spreads which are conditioned in (4.20) by the forecast VPP production, (ΔP_{RE}^{da} for the day-ahead and ΔP_{RE}^{rt}) for the imbalances in real-time).

$$\mathbb{E}(\Delta P_{RE}^{da}|Y = \hat{y}) = \int_{\pi=-\infty}^{\pi=+\infty} \pi \hat{f}_{\Delta P_{RE}^{da}|Y=\hat{y}} d\pi \tag{4.20}$$

In contrast with the previous approach, the uncertainty in prices is considered. If production and prices are independent, then the derivative of the expected loss gives an analytical solution similar to the previous section, and spread prices equal the expectation of the two random variables modelling spread prices:

$$\Delta \pi_{RE^{da}} = \mathbb{E}(\Delta \pi_{RE^{da}})$$
$$\Delta \pi_{RE^{rt}} = \mathbb{E}(\Delta \pi_{RE^{rt}})$$

When dependence between price spreads and VPP production is considered, the expected linear
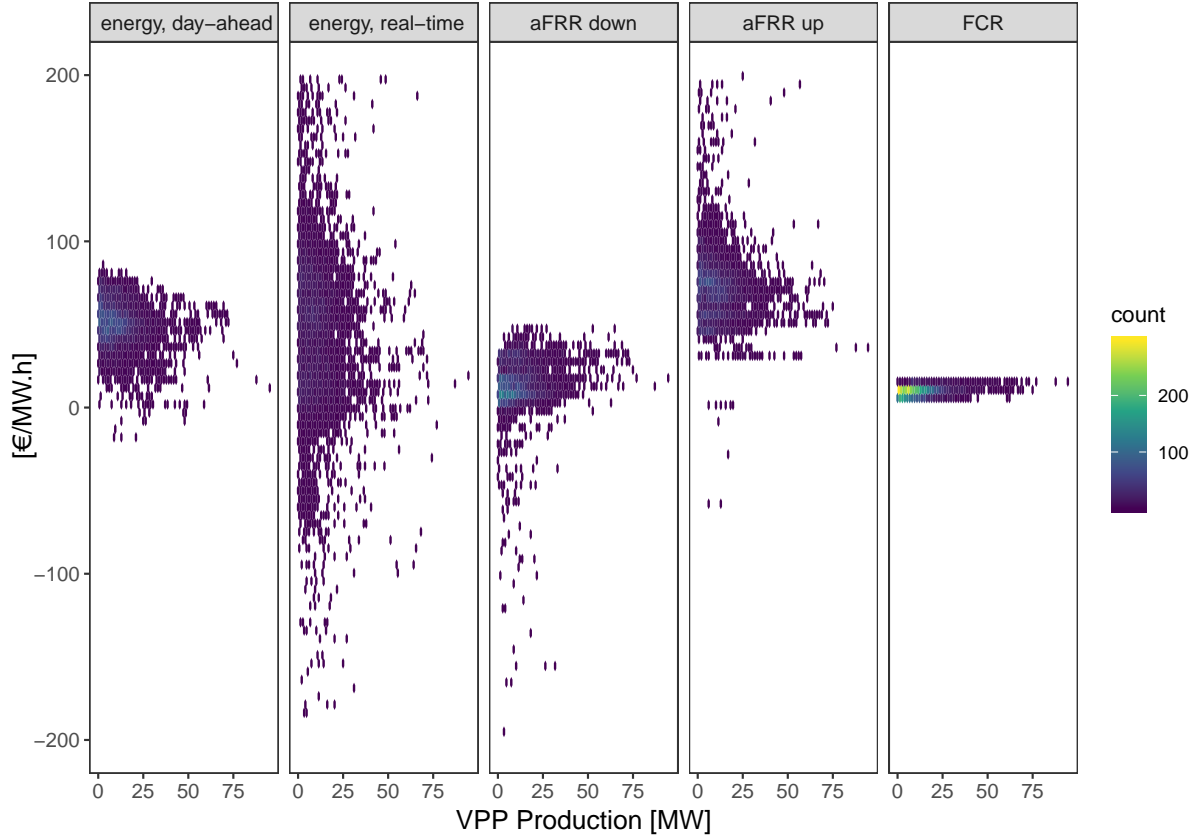
Figure 4.11: Joint densities of observations of VPP production with prices on the energy market (day-ahead and real-time imbalances) and on reserve markets (aFRR downward and upward, finally FCR), for the German part (100% wind capacity) of the VPP presented in Section 2.4.4 . Data from 06/2018 to 03/2019.

loss associated with the reserve offer becomes:

$$\tau_R^{opt} = \arg\min_R \mathbb{E}(l_R)$$

$$\mathbb{E}(l_R) = \int_{y=0}^{R} \mathbb{E}(\Delta P_{RE}^{rt}|Y=y)(R-y)f_Y(y)dy + \int_{y=R}^{y_{max}} \mathbb{E}(\Delta P_{RE}^{da}|Y=y)(y-R)f_Y(y)dy \qquad (4.21)$$

The minimum of the loss expressed in (4.21) can be found by sampling the density function of the forecast production f_Y using quantiles, and then constructing the discretized gradient of the loss on these quantiles [157]. The production level with the closest gradient to zero is selected as the optimal reserve offer.

The problem is analytically tractable if the joint densities of prices and production are modeled with simple approaches such as bivariate normal distribution, which are of limited use in practice because renewable production and prices are not normally distributed [157]. A regression approach could be employed to derive these conditional expected price spreads, for instance using a bivariate kernel density estimation. Instead, a density-based regression using a non-parametric copula is chosen here. Copulas have the advantage of decoupling the dependence model from the marginal

distributions, which can be forecast with various approaches without any impact on the dependence model. Copulas also deliver a probabilistic description of the dependence between production and prices, e.g. the most likely price spreads conditioned by renewable production forecast. If the VPP production is high, there will be probably a high share of renewable in the market, therefore-a high probability of low day-ahead energy price. Finally, this gives a high probability of favorable day-ahead spread for reserve.

In (4.22), the Kernel Density Estimation (KDE) copula evaluates the conditional density of the price spread ΔP_{RE} with respect to VPP production Y. The Epanechnikov Kernel is used for price spread (K_U) and power production (K_V). An alternative not tested here could be to use the Beta Kernel for the production, which is well suited to bounded variables [158]. The two smoothing bandwidths h_u, h_v are chosen following the Scott rule.

$$\hat{f}_{\Delta P_{RE}|Y=\hat{y}}(\pi) = \hat{f}_{\Delta P_{RE}}(\pi)\hat{c}(F_Y(\hat{y}), \hat{f}_{\Delta P_{RE}}(\pi)), \quad \pi \in \mathbb{R} \tag{4.22}$$

$$\hat{c}(u, v) = \frac{1}{Nh_uh_v}\sum_{i=1}^{N} K_U(\frac{u - F_Y(y_i)}{h_u})K_V(\frac{v - F_Y(\Delta P_{RE})}{\pi_i})$$

The distribution of price spread is obtained by a GBT model trained on the feature vector gathering in (4.23) the deterministic forecast of the prices composing the spread $\hat{\pi}_.$ and the associated errors $\epsilon_{\pi,.}$.

$$X_{\Delta P_{RE}} = \{\hat{\pi}_E^{da}, \hat{\pi}_R^{da}, \hat{\pi}_{\bar{R}}^{da}, \epsilon_{\pi,E}^{da}, \hat{\pi}_R^{da}, \hat{\pi}_{\bar{R}}^{da},\} \tag{4.23}$$

4.4.9 Derivation of net revenue

The revenue generated with this Method is computed using observed prices for energy and reserve capacity. The obtained revenue at a given time unit is computed for a single combination of prices and VPP operation. The VPP operation is defined by the forecasted and measured production. To increase the robustness of the evaluation of revenue, we sample realizations of VPP operation by B bootstraps and compute revenues at each time unit on all samples of VPP production. The penalties and revenues occurring in real-time are computed in (4.24) using observed reserve activation probabilities and prices for upward and downward reserve, and observed imbalance energy prices. If the VPP production in real-time is lower than the summed offer of reserve capacity and energy, the amount of energy delivered is reduced and energy penalties are paid to avoid failure on reserve deployment. The VPP net revenue sums up revenues and penalties from the day-ahead stage and the real-time stage.

$$rev_t^{(b)} = \pi_{E,t}^{da}E_{(b)}^{da} + (\pi_{R\downarrow,t}^{da} + a_{\uparrow,t}\pi_{R\uparrow,t}^{rt} + a_{\downarrow,t}\pi_{R\uparrow,t}^{rt})R_{(b)}^{da} - \pi_{E,t}^{rt}E_{(b)}^{rt}, \quad \forall b \in [1, B] \tag{4.24}$$

$$E_{(b)}^{rt} = y_{(b)} - E_{(b)}^{da} - (a_{\uparrow,t} - a_{\downarrow,t})R_{(b)}^{da}$$

4.5 Method 3: Stochastic optimization with chance-constraints

One limit of the optimal quantile method is that it is based on probabilistic forecasts which do not explicitly integrate the temporal variability of aggregated production. It was shown however in Section 2.3 that even if the aggregation smoothes the production profile, significant ramps are observed on periods up to 4 hours. The present method considers temporal variability by integrating the scenarios of aggregated VRE production developed in Chapter 2 into a stochastic optimization formalizing the joint bidding of energy and reserve. The better representation of temporal correlations is thought to improve the reliability of decisions over the validity period of the reserve product.

Stochastic optimization models for the joint bidding of energy and reserve have been already presented in the literature, see for instance [159] and [160]. However, the two following challenges have not been addressed in the reviewed works:

1. what is the best methodology for generating scenarios of aggregated VRE production, considering the stochastic optimization problem at hand?

2. how to integrate directly in the optimization the possibility to have limited under-fulfillments in reserve provision?

The first challenge finds elements of response in Chapter 2: scenarios generated from separate forecasts by source have lower errors than scenarios generated from direct aggregated forecast. The impact on revenue remains still to be quantified by resolving the stochastic optimization problem. A classical two-stage stochastic optimization can not solve the second challenge: reserve underfulfillments can only be observed a posteriori, depending on the realizations of uncertain aggregated production and the risk-averse character of the VPP. Lastly, in order to individuate clearly the impact of production scenarios on the result, market conditions are assumed to be known without uncertainty.

Originality of the method

Formulations of stochastic optimization for the bidding of energy and reserve by wind power plants have been proposed by [159]. A chance-constrained optimization of energy and reserve offer from an aggregator of distributed energy resources is proposed by [161] based on a distributionally robust approach. Compared to these publications, the present method innovates principally by its adaption to the case of reserve provision by a VRE-based VPP:

- The optimization integrates *VRE-based VPP production trajectories* that can be generated from the methods developed in Section 2.5.

- The *chance-constrained formulation is based on VPP trajectories*, instead of a distributionally robust approach, leading to a more realistic consideration of temporal variability in the problem.

- The chance constraint itself corresponds explicitly to a limitation of the frequency of under-fulfilments. By doing so, the optimization is designed to meet the high requirements of reserve

reliability expressed by TSOs for VRE-based reserve offers. This contrasts with the work of [161], in which chance constraints ensure feasible charging / discharging capacities.

4.5.1 Problem formulation

The 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 occurring 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. With the VPP being price-taker and risk-neutral, the objective function writes:

$$\arg\min_{\mathbf{B}} \mathbb{E}_T(\text{loss}(\mathbf{B}, \Omega)) = \sum_{\omega \in [1,\Omega]} p_\omega \text{loss}(\mathbf{B}, \omega) \tag{4.25}$$

where the bidding net penalty for scenario ω is

$$\text{loss}(\mathbf{B}, \omega) = \sum_{t \in [1,N]} [c_{da,t}^{\mathcal{T}} . \mathbf{B}_t^{da} + c_{rt,t}^{\mathcal{T}} . \mathbf{B}_{\omega t}^{rt}], \quad \forall \omega \in \Omega \tag{4.26}$$

with the following decision variables:

$$\mathbf{B}_t^{da} = (E_t^{da}, R_t^{\uparrow,da}, R_t^{\downarrow,da}) \tag{4.27}$$

$$\mathbf{B}_{\omega t}^{rt} = (E_{\omega t}^{rt,-}, E_{\omega t}^{rt,+}, R_{\omega t}^{\uparrow,rt}, R_{\omega t}^{\downarrow,rt}, \Delta R_{\omega t}^{\downarrow,-}, \Delta R_{\omega t}^{\uparrow,-}) \tag{4.28}$$

and their corresponding costs:

$$c_{da,t}^{\mathcal{T}} = (-\pi_{E,t}^{da}, -\pi_{R\uparrow,t}^{da}, -\pi_{R\downarrow,t}^{da}) \tag{4.29}$$

$$c_{rt,t}^{\mathcal{T}} = (\pi_{E,t}^{rt,-}, -\pi_{E,t}^{rt,+}, a_{R\uparrow,t}^{rt} . \pi_{R\uparrow,t}^{rt}, -a_{R\downarrow,t}^{rt} . \pi_{R\downarrow,t}^{rt}, -\pi_{R\uparrow,t}^{rt,-}, -\pi_{R\downarrow,t}^{rt,-}) \tag{4.30}$$

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.

$$E_t^{da} + E_{\omega t}^{rt,+} - E_{\omega t}^{rt,-} + a_{R\uparrow,t}^{rt} . R_{\omega t}^{\uparrow,rt} - a_{R\downarrow,t}^{rt} . R_{\omega t}^{\downarrow,rt} = Y_{\omega t}^{agg} \tag{4.31}$$

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

$$\Delta R_{\omega t}^{\uparrow,-} = R_t^{\uparrow,da} - R_{\omega t}^{\uparrow,rt} \quad \forall \omega, t \tag{4.32}$$

$$\Delta R_{\omega t}^{\downarrow,-} = R_t^{\downarrow,da} - R_{\omega t}^{\downarrow,rt} \quad \forall \omega, t \tag{4.33}$$

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, with risk-aversion parameter β. An economic Conditional Value-at-Risk (CVaR) is formulated as a function of a Value-at-Risk θ_{da}, which is defined as the upper bound of revenues rev having the probability $1 - \alpha_{CVaR}$ 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 (4.37) a non-anticipativity constraint to ensure that the day-ahead decisions are independent from the outcomes of the production scenarios.

$$\max o_\beta = (1 - \beta).\mathbb{E}(rev(\mathbf{B}, \omega)) + \beta.(\theta_{da} - \frac{1}{1 - \alpha_{CVaR}} \sum_{\omega \in [1,\Omega]} p_\omega \rho_\omega) \tag{4.34}$$

s.t.

$$\theta_{da} - rev(\mathbf{B}, \omega) \leq \rho_\omega, \quad \forall \omega \in \Omega \tag{4.35}$$

$$\rho_\omega \geq 0 \quad \forall \omega \in \Omega \tag{4.36}$$

$$\mathbf{B}_{\omega,t}^{da} = \mathbf{B}_{\omega',t}^{da} \quad \forall \omega, \omega' \in \Omega \tag{4.37}$$

$$(13)\text{-}(15) \tag{4.38}$$

4.5.2 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 under-fulfillment have a very low probability of occurrence over the scenario set.

The previous model is therefore augmented with a chance constraint on upward reserve (4.39) and downward reserve (4.40) 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 $1 - \epsilon$.

$$Pr(\Delta R_{\omega t}^{\uparrow;-} \leq 0) \geq 1 - \epsilon \quad \forall \omega, t \tag{4.39}$$

$$Pr(\Delta R_{\omega t}^{\downarrow;-} \leq 0) \geq 1 - \epsilon \quad \forall \omega, t \tag{4.40}$$

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 [91]. One option is to derive the divergence between the distribution of production and the normal distribution, then apply a distributionally robust chance-constrained programming model using this divergence [161]. As the distributions of renewable production show

significant divergences with the Gaussian distribution (right skews and fat tails), an alternative approach is taken which is distribution-free and scenario-oriented, i.e. the constraint is approximated by a non-decreasing convex function [162]. The constraint is conservatively approximated by a technical CVaR function on the distribution of reserve deficit $\text{CVaR}_{\Delta R_{\omega t}^{\uparrow,-}}$ such that:

$$\text{CVaR}_{\Delta R_{\omega t}^{\uparrow,-}}(1 - \epsilon) = \inf_{\alpha_{\Delta R}} \left(\frac{\mathbb{E}([\Delta R_{\omega t}^{\uparrow,-} + \alpha_{\Delta R}^+)}{\epsilon} \right) - \alpha_{\Delta R} \tag{4.41}$$

Then the chance constraint (4.39) can be conservatively linearized as in (4.42). The expected value of the ramp function is obtained in (4.43) by averaging its value over the number of scenarios.

$$\mathbb{E}([\Delta R_{\omega t}^{\uparrow,-} + \alpha_{\Delta R^-,t]}^{\uparrow}{}^+) \leq \alpha_{\Delta R^-,t}^{\uparrow} \cdot \epsilon \quad , \forall \omega, t \tag{4.42}$$

$$\mathbb{E}([\Delta R_{\omega t}^{\uparrow,-} + \alpha_{\Delta R^-,t}^{\uparrow}]^+) = \sum_{\omega \in [1,\Omega]} p_\omega [\Delta R_{\omega t}^{\uparrow,-} + \alpha_{\Delta R^-,t}^{\uparrow}]^+ \tag{4.43}$$

The positive ramp function appearing in (4.42) is approached using the big M constraint technique. Binary variables $b_{\omega t}$ and positive variables $\psi_{\omega t}$ approximate the positivity constraint for each scenario ω:

$$-M.(1 - b_{\omega t}^{\uparrow}) = \leq \Delta R_{\omega t}^{\uparrow,-} + \alpha_t^{\uparrow} \leq M.b_{\omega t}^{\uparrow} \tag{4.44}$$

$$\psi_{\omega t}^{\uparrow} = b_{\omega t}^{\uparrow}.(\Delta R_{\omega t}^{\uparrow,-} + \alpha_t^{\uparrow}) \tag{4.45}$$

$$-M.(1 - b_{\omega t}^{\uparrow}) \leq \psi_{\omega t}^{\uparrow} - (\Delta R_{\omega t}^{\uparrow,-} + \alpha_t^{\uparrow}) \leq M.(1 - b_{\omega t}^{\uparrow}) \tag{4.46}$$

$$-M.b_{\omega t}^{\uparrow} \leq \psi_{\omega t}^{\uparrow} \leq M.b_{\omega t}^{\uparrow} \tag{4.47}$$

4.6 Method 4: Bidding on multiple reserve markets

An aggregator is potentially interested in bidding on multiple reserve markets, which may impact differently the plants composing the VPP. Consider for instance that the aggregator wishes to bid on three markets: the wholesale energy market, the FCR reserve market and the aFRR reserve market. This situation is illustrated in Figure 4.12: two VPPs operate in their own region, in that case France and Germany, and optimize their bids of FCR (in the particular case FCR is procured in a common market to several European countries), aFRR and energy on their respective markets.

This problem could be solved by extending the dimensions of reserve products in the stochastic optimization formulated in Section 4.5. However, this complicates further the model chain. In contrast, in the present Method it is proposed to formulate the bidding of multiple reserve products as an extension of the newsvendor problem mentioned earlier for the derivation of optimal quantiles.

Originality of the method

Most of the publications found in the literature analyzed reserve products separately. An exception is the general formulation of [163] that corresponds to a multi-scale optimization on markets that are organized following multiple time scales from delivery to real-time. In that sense, multiple reserve products can be incorporated in the strategy. However, the challenge of finding the optimal allocation of an uncertain VRE aggregated production is left untouched. To the best of the author's
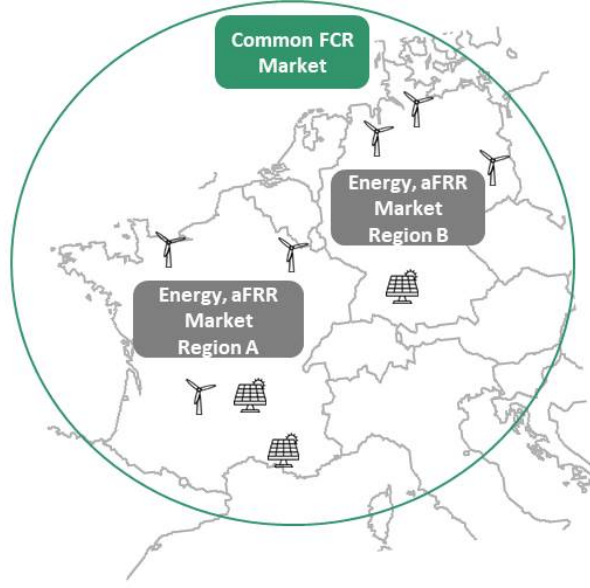
Figure 4.12: VPPs offering FCR on a common market, aFRR and energy on separate markets

knowledge, the present method innovates in its proposition of *adapting the framework of the multi-product newsvendor formulation*, as analyzed for instance in [164], in a simple optimization method of multiple reserve products based on a probabilistic prediction of aggregated production.

The following assumptions hold for the different markets:

- FCR, aFRR capacities must be guaranteed over defined validity periods that make differ. The problem is simplified by assuming a single validity period $T_{validity}$ to ease computation.

- FCR and aFRR may be offered in asymmetrical upward/downward products.

- Day-ahead clearing of energy, FCR and aFRR markets.

- The imbalances on the energy markets are paid according to a single-price mechanism.

- For the sake of simplicity a perfect knowledge of prices is assumed. Note that the proposed method remains valid if deterministic forecasts of spread prices are implemented.

The problem is now cast in (4.49) as a multi-product newsvendor problem under stock constraint [165]. A single stock constraint limits the total participation of plants to the multiple reserve markets.

$$\underset{R_i}{\arg\max} \, \mathbb{E}[\sum_{i \in FCR, aFRR} \Delta\pi_{RE,i}^{da} R_i - \Delta\pi_{RE,i}^{rt}(R_i - y)] \tag{4.48}$$

$$s.t. \sum_{i \in FCR, aFRR} s_i R_i - S_{multi} \leq 0 \tag{4.49}$$

In this equation, s_i represent the resource coefficient per reserve product i. The stock constraint S_{multi} is the maximum power capacity available for reserve. It is equal to the installed capacity if

only downward reserve is proposed, otherwise it is equal to the installed capacity reduced by the day-ahead energy bid.

$$S_{multi} = \begin{cases} y_{max}, & \text{if} \quad R^{\downarrow} \\ y_{max} - E_{da}, & \text{if} \quad (R^{\uparrow}, R^{\uparrow,\downarrow}) \end{cases} \tag{4.50}$$

The optimal reserve quantities R_i associated to each reserve market i can be found by formulating the Lagrangian L associated to this problem, with its multiplier λ:

$$L(R_i, \lambda) = \mathbb{E}[\sum_i \Delta\pi_{RE,i}^{da} R_i - \Delta\pi_{RE,i}^{rt}(R_i - y) - \lambda(\sum_i s_i R_i - S_{multi})] \tag{4.51}$$

Note that λ can be interpreted as a marginal price or shadow price [146] of adding an extra monetary unit to the available reserve capacity. Exploiting the stationarity Karun-Kush-Tucker (KKT) conditions, the equation system of (4.52) defines the resource coefficients and marginal reserve quantities R_i as a function of the predicted aggregated production CDF F_Y :

$$\frac{\partial L}{\partial \lambda} = -F_Y(R_i)\Delta\pi_{RE,i}^{da} R_i + (1 - F_Y(R_i)\Delta\pi_{RE,i}^{rt}(R_i - y) - \lambda s_i = 0, \qquad \forall i$$

$$\frac{\partial L}{\partial R_i} = \sum_i s_i R_i - S_{multi} = 0, \qquad \lambda > 0 \tag{4.52}$$

In the first equation of the system, an optimal quantile formulation can be identified easily as in the case of a single reserve product (see Section 4.4). The optimal quantile $\tau_{R,i}^{opt}$ obtained for a product i equals to the marginal optimal quantile $\tau_{R,i}^{marg} = \frac{\Delta\pi_{RE,i}^{da}}{\Delta\pi_{RE,i}^{da} + \Delta\pi_{RE,i}^{rt}}$ corrected by the contribution of the reserve product λs_i to the global reserve offer:

$$\tau_{R,i}^{opt} = \frac{\Delta\pi_{RE,i}^{da} - \lambda s_i}{\Delta\pi_{RE,i}^{da} + \Delta\pi_{RE,i}^{rt}} = \tau_{R,i}^{marg} - \frac{\lambda s_i}{\Delta\pi_{RE,i}^{da} + \Delta\pi_{RE,i}^{rt}} \tag{4.53}$$

$$R_i = F_Y^{-1}(\tau_{R,i}^{opt}) \tag{4.54}$$

The equation system defined in (4.52) can now be solved numerically with λ taking values that satisfy the stock constraint.

The present approach is limited by the formulation with a single stock constraint: if the aggregator wishes to impose several limits to its bidding volumes on the different reserve markets, the optimization problem must be reformulated with multiple stock constraints and solved with more complex methods involving quadratic approximations of the inverse CDFs [164].

4.7 Case studies

The four methods proposed in this Chapter are evaluated in 3 case studies. The aim of these case studies is to identify the best configuration for each method. Comparisons between Methods which

would bring little original contributions are avoided (e.g. comparing the benefits of the stochastic approach in Method 3 against an optimal quantile for reserve optimization, which was extensively studied in [89]). Instead, the decision-maker may be interested in knowing, after having chosen for a given bidding method, what is the impact of the different configurations of this method, for instance: how is the risk on revenue evolving when opting for an optimal quantile with dependence between price and production instead of only considering deterministic price forecasts, or which type of aggregated production trajectories (from direct or separate forecasting) delivers the highest revenue.

4.7.1 Case study 1: Optimal quantile approach (Methods 1 and 2)

This case study assesses the performance of the optimal quantile approach for bidding jointly VPP production on the energy market and on a specific reserve market. More specifically, the solutions proposed by Method 2 in Section 4.4 are implemented on a real-world case of a VPP, which jointly offers energy and symmetrical aFRR on a day-ahead auction with an aFRR validity period of 4 hours. The reserve bids are compared to those obtained with a constant low quantile (Method 1).

Real production data covering the period September to December 2015 are used to tune the forecasting model, while data from January to March 2016 are employed to evaluate them. The VPP offer has to be placed before gate closure time at 9h00 UTC, each day and is evaluated over the 3 month testing period. The VPP combines wind and PV plants operating in France, with a total capacity of 42.3 MW and a 24% share of PV. The distance between any two power plants varies between 30 km and 700 km. Forecasts of the VPP production are those presented in Section 2.4.3.

The VPP simultaneously offers energy and symmetrical aFRR before noon of the previous day, using the QRF aggregated production forecast, and price forecasts for the optimal quantile methods. Bids are based on forecasts issued at 00h00 UTC, so the forecast horizon is comprised between 24 and 48 hours. All price forecasts are trained on a sliding window of 150 days. Prices are taken from the German market, because at the time of writing Germany has a more market-oriented aFRR tender than France, where the price for aFRR is fixed. Note that the aFRR price data used here corresponds to week-ahead tenders [1]. Finally, the reserve penalty is assumed to be 5 times the reserve capacity price, as currently set by the French TSO [14].

4.7.2 Case study 2: Chance-constrained approach (Method 3)

The second case study presents the results of the chance-constrained optimization developed by Method 3 in Section 4.5. More specifically, the focus is placed on the impact on revenue of the methodology chosen for generating scenarios of aggregated production (cf Section 2.5).

The optimization is computed for 100 days using price data from the Portuguese Market (MIBEL for energy and REN for reserve), with 100 trajectories of aggregated production. Trajectories correspond to those generated in Section 2.5 for VPP1:

- from a Direct Aggregated production forecast and a Gaussian Copula (DG),

---

[1]day-ahead tenders have started in 2018 in Germany

- from an Indirect aggregated production forecast by energy source and a Gaussian Copula (IG)

- from an Indirect aggregated production forecast by energy source and a Vine Copula (IV)

The M coefficient is set to 60 MW, a number close to the aggregated installed capacity (53 MW) and ϵ is set to 1%. The MIP problem is solved with lpSolve, setting a MIP gap of 0.1%. The VPP is considered moderately risk-averse, $\beta = 0.5$. 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 Portuguese aFRR market shows high prices for aFRR compared to Germany or France, presumably because of a high concentration in reserve suppliers.

4.7.3   Case study 3: Multi-service bidding (Method 4)

The last case study quantifies the The multiple product bidding method presented in Section 4.6 is tested on a 3-month winter period (12/2018-02/2019), with the Wind-PV VPP presented in Section 2.4.4. The temporal resolution of production data is 10 min. In particular, two sub-aggregations are tested: the VPPs formed by all plants located in France and in Germany respectively. This allows to assess the sensitivity of the method to various price levels and forecasting conditions. Both VPPs offer upward FCR and upward aFRR, along with energy on their respective energy markets. Adding more services such as downward FCR and aFRR may also be pertinent, and leaves Method 4 unchanged. However, observed prices for downward reserve are so low that bidding downward reserve does not generate any additional revenue compared to bidding energy only.

Regarding prices, recall that FCR capacity prices are common between Germany and France, whereas aFRR prices are national. The aFRR capacity prices in France are fixed. The average aFRR prices in France, integrating balancing energy are lower than German prices on the same period (- 10 Eur/MW.h). In contrast, the average day-ahead energy price is higher in France on the period (+ 8 Eur/MWh).

The multiple product bidding method is compared with two reference methods:

1. **Constant extremely low quantile (Method 1)**

   The first reference method bids to minimize the total of risk of reserve under-fulfilment for FCR and aFRR. It is based on Method 1, where the reserve bid R_t is issued from a production forecast with a constant low quantile, taken here at 0.1%. The forecast is derived with the exponential distribution model presented in Section 3.3, with clusters of median production for the sake of simplicity. The total reserve capacity for all services R_0 is chosen as the minimum value over a validity period $T_{validity}$, here as 1 hour:

$$R_{0,t} = min_{t' \in T_{validity}(t)} F_{Y,t'}^{-1}(0.1\%), \quad \forall t \tag{4.55}$$

   Finally, the reserve capacity is equally split in FCR and aFRR bids:

$$Ri, t = \frac{1}{2} R_{0,t}, \quad \forall t, \forall i \in [\text{FCR}^{\uparrow}, \text{aFRR}^{\uparrow}] \tag{4.56}$$

2. **Marginal optimal quantiles (Method 2)**

A second reference method is built with reserve bids issued independently for each service. Marginal optimal quantiles $\tau_{R,i}^{marg}$ for each service i (FCR and aFRR) are computed according to prices observed on all markets (cf. Section 4.6):

$$\tau_{R,i}^{marg} = \frac{\Delta \pi_{RE,i}^{da}}{\Delta \pi_{RE,i}^{da} + \Delta \pi_{RE,i}^{rt}}$$
$$Ri, t = F_{Y,t}^{-1}(\tau_{R,i}^{marg}) \forall i \in [\text{FCR}^{\uparrow}, \text{aFRR}^{\uparrow}] \tag{4.57}$$

4.8 Evaluation metrics

4.8.1 Goodness of fit of the dependence model between price and production

The goodness of fit of the price-production copula is measured in by a Cramér-Von Mises statistic CvM [166] comparing the distribution of the copula $\hat{C}(u,v)$ with the distribution of the empirical copula $\hat{C}_{emp}(u,v)$.

$$CvM = N \int_{[0,1]^2} (\hat{C}(u,v) - \hat{C}_{emp}(u,v))^2 . d\hat{C}_{emp}(u,v) \tag{4.58}$$

$$\hat{C}_{emp}(u,v) = \frac{1}{N} \sum_{i=1}^{N} 1_{F_y(y_i) \leq u, F_{\Delta P_{RE}}(\pi_i) \leq v} \tag{4.59}$$

One may object that the non-parametric KDE copula proposed may overfit the particular relations observed in data. Is it therefore benchmarked against parametric copulas, fitted by Maximum Likelihood Estimation [167]. Tested copulas belong to the family of Archimedean copulas, because they enable to model asymmetrical dependencies between production and spread prices. The CDF of an Archimedean copula writes:

$$C(u,v) = \phi^{[-1]}(\phi(u) + \phi(v)) \tag{4.60}$$

with $\phi : [0,1] \longrightarrow \mathbb{R}^+$ is the generator function and $\phi^{[-1]}$ is its pseudo-inverse. The lowest CvM between the nonparametric KDE copula and the identified Archimedean parametric copula indicates the best fitted copula.

4.8.2 Scores of deterministic price forecasts

Deterministic price forecasts $\hat{\pi}$ are evaluated with respect to a simple benchmark forecast, in this case the daily persistence $\hat{\pi}_{persist}$:

$$\hat{\pi}_{persist,t} = \pi_{t-24h}, \quad \forall t \tag{4.61}$$

As prices are unbounded by nature, scaling forecasting scores by observed extrema is not pertinent. Instead, the deterministic scores (RMSE and MAE) are scaled by the score obtained with daily persistence [78], in order to obtain a versatile and simple representation of price forecasting performance. Given N observations in the evaluation set, the Root Mean Squared Scaled Error (RMSSE) and the Mean Average Scaled Error (MASE) write:

$$\text{RMSSE}(\hat{\pi}, \pi) = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{\pi}_i - \pi_i)^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{\pi}_{persist,i} - \pi_i)^2}} \tag{4.62}$$

$$\text{MASE}(\hat{\pi}, \pi) = \frac{\frac{1}{N} \sum_{i=1}^{N} |\hat{\pi}_i - \pi_i|}{\frac{1}{N} \sum_{i=1}^{N} |\hat{\pi}_{persist,i} - \pi_i|} \tag{4.63}$$

4.8.3 Evaluation of the technical reliability of reserve offers

The reliability of reserve offers is evaluated by a simple indicator denominated Rate of Under-Fulfilment (RUF) [168] which quantifies the frequency of events where the aggregated production is inferior to the offered reserve capacity:

$$RUF = \frac{1}{N} \sum_{i=1}^{N} 1_{y_i < R_i} \tag{4.64}$$

The principle is illustrated in Figure 4.13. In this Figure, the orange line represents the observed VPP production and the green curve is curtailed VPP production in case of a full activation of the downward reserve bid. The red dot highlights an underfulfilment: in that case the curtailed VPP production is negative, the reserve bid can not be provided fully. The evaluation is limited by the available temporal resolution of production time series. It also does not consider technical constraints that may impede a reliable regulation of power at low levels (wind turbines shutdowns were observed while down-regulating for a frequency response test in the Kombikraftwerk 2 project [95]).

4.8.4 Conditional Value-at-Risk of revenue

The benefits of the proposed methods in terms of revenue will be evaluated by means of two metrics:

1. *Average revenue*: Expectation of revenues over the timesteps of the evaluation period.

2. *Conditional Value-at-Risk* (CVaR): The CVaR quantifies in (4.65) the average revenue of revenues r below a quantile $F_{rev}^{-1}(\tau_{rev})$ at nominal value τ_{rev}:

$$CVaR_{\tau_{rev}} = \mathbb{E}(r), r \leq F_{rev}^{-1}(\tau_{rev}) \tag{4.65}$$
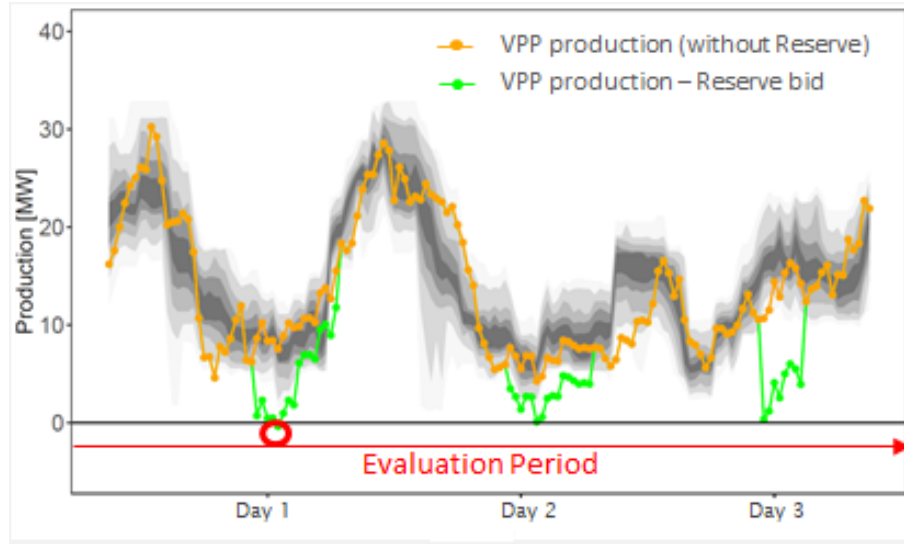
Figure 4.13: Illustration of the Rate of Under-Fulfilment

4.9 Results

4.9.1 Case study 1

Reserve offers and probabilistic production forecasts

The probabilistic forecast of the VPP production and the associated reserve (aFRR) offers obtained with Methods 1 and 2, as well as persistent prices, are illustrated in Figure 4.14. It is observed that methods generate diverse solutions, influenced either more by production forecasts (visible for the constant quantile in grey) or by market conditions (persistent price forecasts in green, price/production dependence model with copula in red).

Price Forecasts

The performance of the deterministic price forecasts are reported in Figure 4.15. The error levels for the energy price are coherent with similar studies based on auto-regressive or ARIMA models ([152], [151]). The forecast of aFRR capacity prices shows lower errors because daily persistence is strongly penalized on the first days of the week, which reproduce the valley-hour prices of the weekend. The error on imbalance price is lower than persistence mainly because the forecast has less bias than persistence. There is still significant room for improvement in capturing the variance correctly.

Next, the probabilistic forecast of spread price between energy and reserve presents sastifactory results if taken a first proposition. A training period of 60 days was found sufficient for the gradient boosting model, using 5000 trees and a shrinkage parameter of 0.01. The forecast of the spread for revenue prices is reliable on low quantiles (deviation of nominal rate below 3%), and underestimates higher quantiles (not surprising considering that the deterministic forecasts do not capture spikes). The quantile scores improve between 20% and 45% compared to climatology for quantiles lower
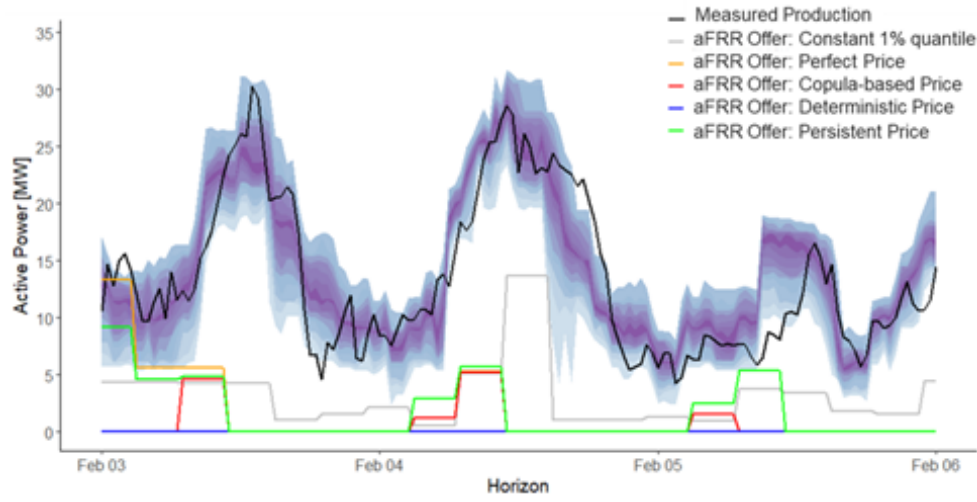
Figure 4.14: Visualization of reserve offers obtained with different quantile selections. The probabilistic forecast of aggregated production is displayed in violet/blue prediction intervals (10%-90%).

Price Forecasting Model	MASE [p.u.]	RMSSE [p.u.]
Energy Day-ahead Price ARIMAX(2,1,1)(1,0,1)$_{24}$	0.85	0.86
Energy Imbalance Price MRS(3,1)	0.72	0.73
aFRR upward capacity Random Forest	0.70	0.66
aFRR downward capacity Random Forest	0.74	0.62
aFRR upward activation probability PCA+SVR	0.77	0.82
aFRR downward activation probability PCA+SVR	0.74	0.77
aFRR upward average price SVR	0.84	0.87
aFRR downward average price SVR	0.84	0.82

Figure 4.15: Scores of deterministic price forecasting models for energy and aFRR
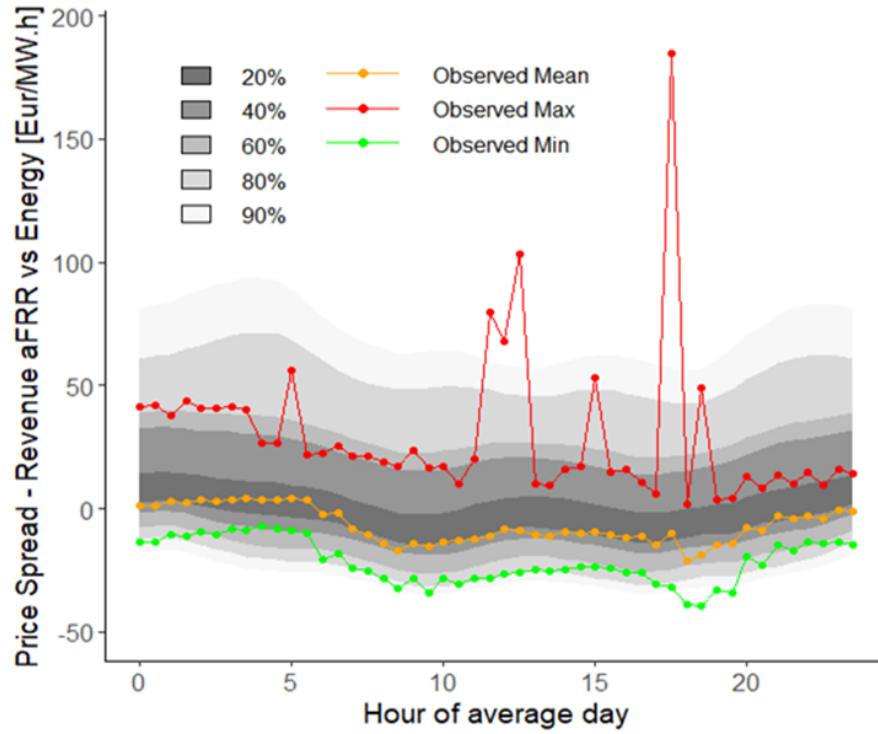
Figure 4.16: Day-ahead probabilistic forecast of revenue price spread between reserve and energy, obtained by a gradient boosting tree model. Prediction intervals between 20% and 90%. Forecasts averaged at same hour of day

or equal to 70%. Above this nominal value, the improvement is scarce which is in line with the findings for reliability.

Figure 4.16 shows that spreads observed over a 60-days period (February-March) fall on average within the central part of the forecast distribution, with a slightly positive spread. If we look at maximum spreads for a given hour of day (red curve), we observe spikes located outside the average forecast envelope during the two consumption peaks at mid-day and early evening. These are due to rare activation peaks where plants with a higher marginal price enter the merit-order. The minimum observed spreads for each hour of day (green curve) have lower levels during the daytime, and remain within the forecast envelope: the forecast model correctly captures low day-ahead prices.

Dependence between Price Spreads and Production

The KDE copulas are built on the VPP production and price spreads observed during the training period. The resulting copula density in the upper plot of Figure 4.17 indicates that high VPP production is frequently associated with high price spreads on revenue, whereas low production levels are mostly linked with average spreads. This nonparametric copula detects asymmetrical tail dependences (high density for high quantiles, low density for low quantiles) and dissymmetrical densities for low price spreads and low production. The high price spreads on revenue occur when
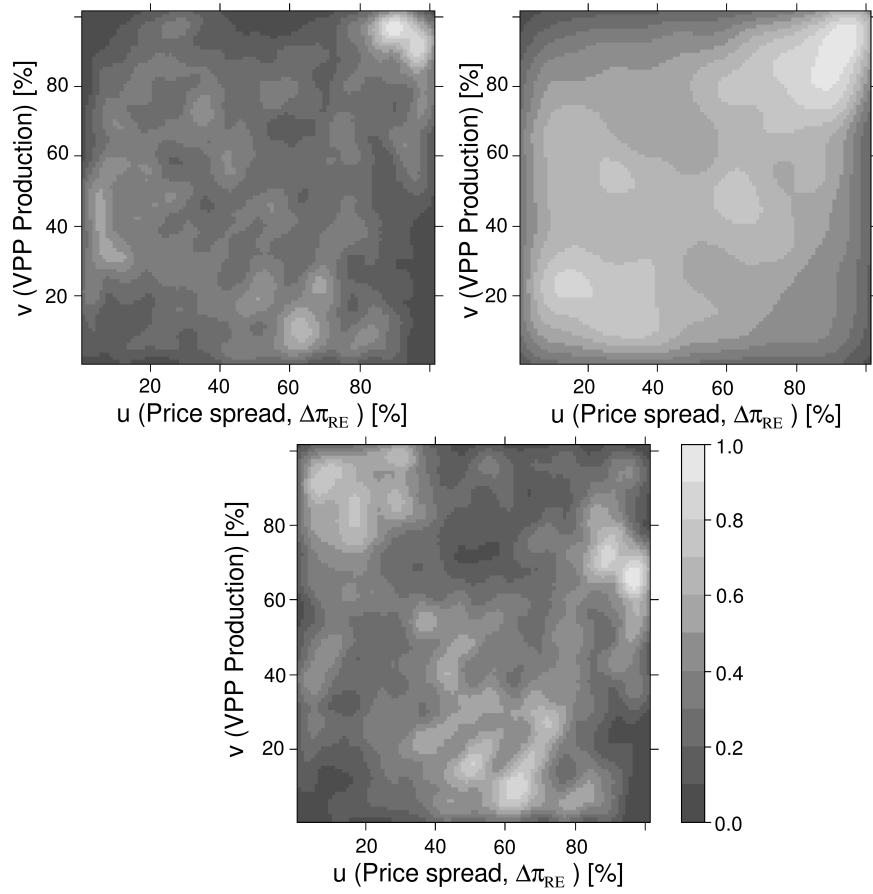
Figure 4.17: Copula modelling dependence between VPP production and price spread between aFRR and energy, (top left) 75% capacity from Wind, KDE Copula, (top right) 75% capacity from Wind, by Joe-Frank copula (1.39,0.95) (top right), (bottom) 75% capacity from PV, by KDE Copula

energy prices are low or the reserve activation price is high. High renewable production at very low marginal cost is known to produce low prices on energy markets, so the dependence structure detected by the copula in this zone is in line with practical experience. We also observe that medium levels of VPP production occur most frequently when revenue spreads are low, which usually occurs during the day.

The KDE non-parametric copula is compared with parametric copulas fitted by maximum likelihood estimation using the function "BiCopSelect" of the R package "VineCopula" [169]. A Joe-Frank copula (mostly symmetrical density and light upper-tail dependence) is selected for the dependence between day-ahead price spread ΔP_{RE}^{da} and production Y, while a Tawn 2 copula (asymmetrical density and asymmetrical tail dependences) is selected for the dependence between real-time price spread ΔP_{RE}^{rt} and production. The Cramér-von Mises statistic of the parametric copulas is higher than the KDE copulas (3.4 times higher for ΔP_{RE}^{da} and 2.3 times higher for ΔP_{RE}^{rt}), and thus the KDE copula is closer to the observed dependence structure in the learning data.
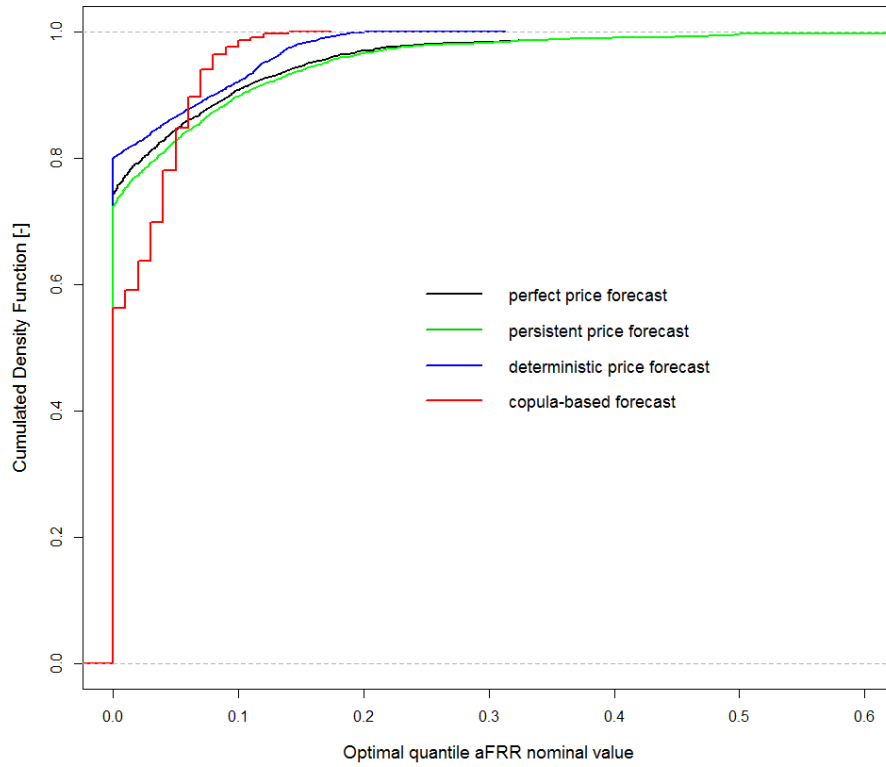
Figure 4.18: CDF of optimal quantiles for aFRR, according to price forecast

Does this tighter fit lead to overfitting on the estimated quantile for aFRR? Although the parametric copulas partly detect the tail behaviors in the dependences, they tend to put more weight on the central part of the price distribution than the KDE copula, hence delivering higher quantiles for aFRR. The mean absolute error on the forecast of the optimal quantile for aFRR is 4.3% for the KDE copula, lower than the 5.0% for the parametric copulas. The lower plot in Figure 4.17 represents the copula density with a modified VPP portfolio where PV accounts for 75% of the installed capacity. In this case the dependence between high productions and high spreads is less pronounced; it is now higher with low spreads (typically when the sun shines during the daytime).

Optimal quantile and net revenue

The cumulative density function (CDF) of optimal quantiles for aFRR obtained from the different price forecasts is presented in Figure 4.18. The nominal value is most frequently zero, when the revenue price spread is negative (occurs e.g. at high energy prices, or for large activations of downward reserve). We see that the optimal quantiles are distributed on low values (0%-40%). The distribution of deterministic price forecast is close to the distribution of perfect price in the 0%-10% range of nominal values, but does not capture the high tail of optimal quantiles. The copula-based forecast issues less dispersed quantiles, influenced by the forecast of VPP production. Finally, the quantile value of the constant quantile offer is set to 1 % (see Section 4.3), 1% being the maximum failure risk that a TSO is supposed to tolerate.

The failure risk for each offer strategy is evaluated by the Rate of Under-Fulfillments (RUF) criterion defined in Section 4.8.3. The offer strategies have a RUF of between 0.1% and 2%, mostly due to the reliability of the probabilistic production forecast and the low values of quantiles dedicated to reserve. The RUF of the constant quantile strategy is 1.3%. The highest RUF is 1.9% for the offer with persistence price forecast, and the lowest RUF is 0.1% for the strategy with the optimal quantile dependent on production, the latter having few under-fulfillments because its highest quantile values are lower than for the other strategies.

Realizations of the VPP operation are sampled during the test period to evaluate the revenues of each strategy for diverse VPP operation conditions, as seen in Section 4.4.9. The number of samples of VPP operations is 40 for each market time unit. Prices of the test period are characterized in Table 4.3.

	Day-ahead energy price	Imbalance energy price	aFRR upward average price	aFRR downward average price
Units	[Eur/MWh]	[Eur/MWh]	[Eur/MW.h]	[Eur/MW.h]
Average	23.7	20.7	41.5	0.9
Minimum	-20	-630.6	33.8	-152.7
Maximum	53.5	634.5	239.9	11.3

Table 4.3: Summary of prices in test period

A first analysis is conducted to evaluate if there exists a compromise between underfulfilments due to higher reserve bids than available production and increase of average revenue due to efficient capture of additional profit when the spread for reserve is positive. Results in Figure 4.19 indicate that the increase of revenue is not linearly correlated to the Rate of UnderFulfilment (RUF). Instead, reserve bids based on very low quantiles (Method 1, light blue and blue points) enable to limit RUF, but also fail to increase the average revenue over 5 %. Remarkably, the bidding strategies based on optimal quantile (Method 2, yellow and orange points) generate a lower RUF than a reserve bid based on a constant quantile of 1%. While strategies based on optimal quantiles perform comparably in terms of increase in average revenue, the distributions of the associated revenues shown in Figure 4.20 are distinct. In that case, the approach based on deterministic forecasts of prices (distibution in orange) shows more frequent losses and of higher amplitude than for the approach based on the copula between price and production (distribution in yellow).

More detailed results of the revenue calculation are shown in Table V. The theoretical maximum increase of mean revenue compared to selling energy only is +36%. The more realistic strategies based on Methods 1 and 2 show lower but significative increases in revenue, around +5%. The Conditional Value at Risk at 1% (1%-CVaR) is lower when offering both energy and reserve instead of energy only for all approaches. This indicates that adding aFRR increases the risk on net revenue, and advocates offering methods that are more risk conservative. The strategy with the production-dependent quantile obtained via the price-production copula is effective at mitigating both technical and financial risks (lower RUF, lower decrease of 1%-CVaR).

Energy markets are oriented towards lower prices due to the penetration of renewable energy at near-zero marginal costs and other factors. To test the sensitivity of the present method in a
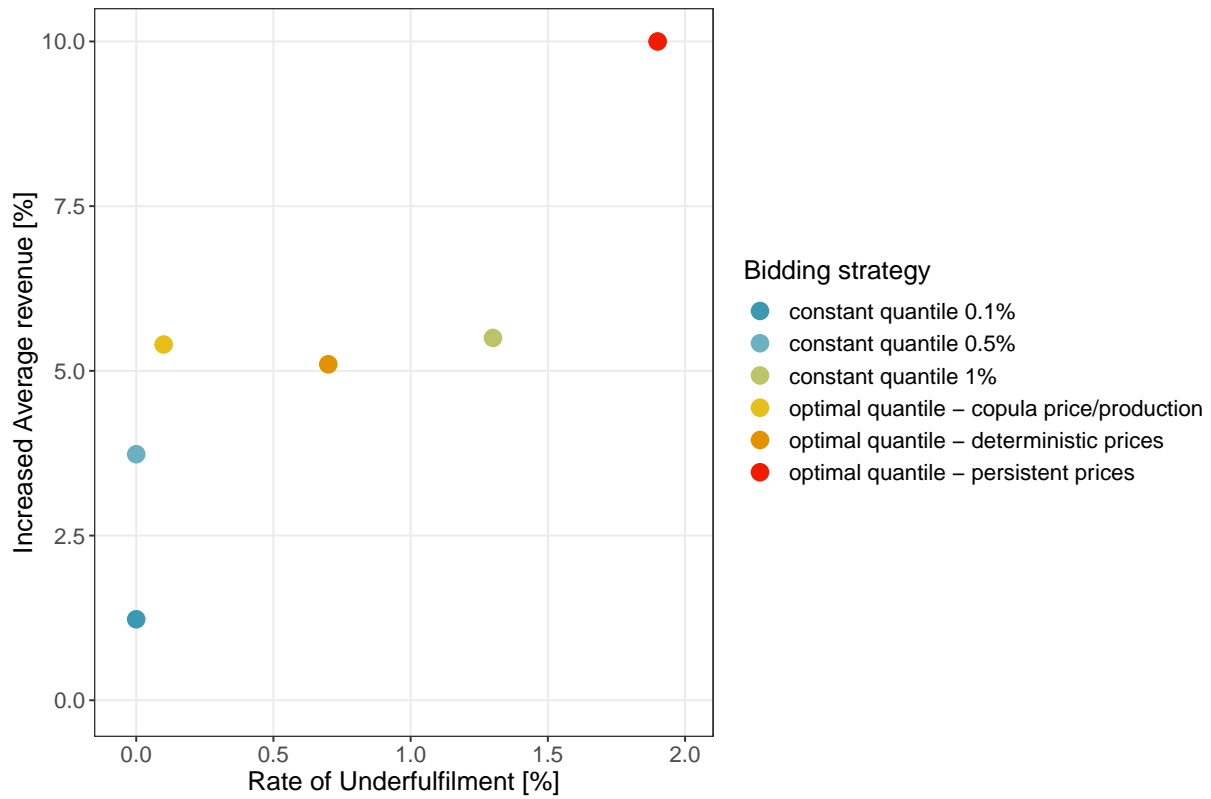
Figure 4.19: Results of proposed bidding strategies in terms of Rate of Underfulfilment (X axis) and Increased average revenue compared to bidding energy only (Y axis)

context of high penetration of renewables, we linearly reduce the energy prices (revenue price and penalty price), while keeping the reserve prices constant. We assume that reserve, being a product
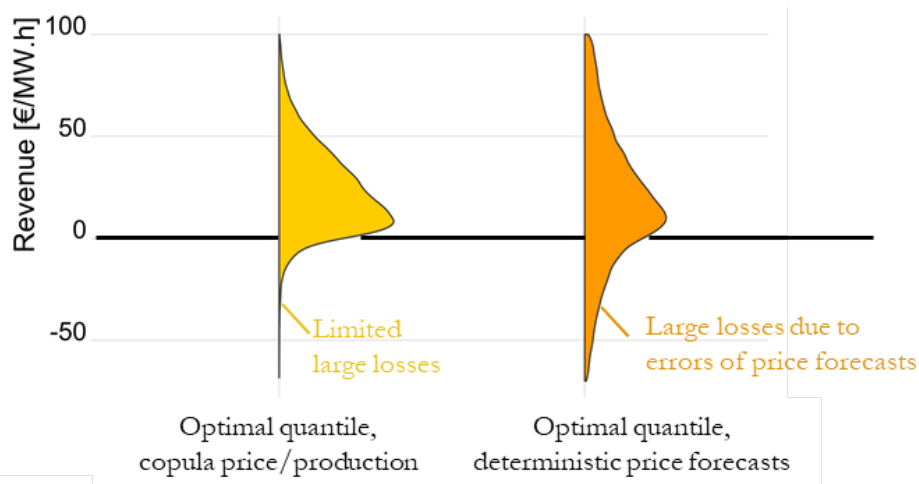
Figure 4.20: Distribution of observed revenues from energy + reserve depending on the optimal quantile variant strategies (left half-violins, in colors)

	Rate of under-fulfilment	Average reserve offered	Difference in mean revenue aFRR+Energy vs Energy only	Difference in 1%-CVaR of revenue aFRR+Energy vs Energy only
Units	%	% of Pn	%	[Eur/MW.h]
Method 1	1.3	**10.3**	+5.5	-25
Method 2, production-independent quantile	0.7	4	+5.1	-138
Method 2, production-dependent quantile	**0.1**	3.8	+5.4	**-8**
Perfect price forecast	Not tested	5.7	**+36**	-18
Persistent price forecast	1.9	5.5	+10	-12

Table 4.4: Results of offering strategies. Mean revenue for energy only is 25.05 Eur/MW.h

of high added value with limited availability, will see its price remain stable. We then compute offers based on deterministic price forecasts and the original (not sampled) VPP production and its forecast. The optimal quantile for aFRR increases as energy prices decrease. Table 4.5 shows that the additional mean revenue increases linearly with decreasing energy prices. An aFRR validity period of 1 day instead of 4 hours leads to a lower increase in net revenue. Short validity periods are therefore an important request for the profitability of reserve for a VRE-based VPP.

	Variation of mean revenue aFRR+Energy vs Energy
Units	%
Reference Energy Price - Validity period varying	
4 hours	+5.1
24 hours	+3.9
Varying Energy Price - Validity period constant	
Reference price - 10%	+8.2
Reference price - 20%	+9.2
Reference price - 30%	+10.2
Reference price - 40%	+11.4
Reference price - 50%	+12.5

Table 4.5: Sensitivity analysis on validity period and energy prices (deterministic forecast of prices)

4.9.2 Case study 2

The frequency of reserve deficit simulated by the model is null for all scenarios. Economic and technical results are reported in Table 4.6. The use of the direct aggregated forecast, which generates scenarios with higher amplitudes, reduces the amount of reserve offered to avoid penalties for reserve under-fulfillment. 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%-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 (see Results of Section 2.5). 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 4.6: Average profits and volumes of energy and reserve for the optimized bidding, depending on the scenario generation method

4.9.3 Case study 3

Figure 4.21 shows the profits obtained when bidding jointly energy (abbreviated as 'Nrj') and upward aFRR and FCR. The top panel shows results for the German VPP and the bottom panel for the French VPP. Columns represent the four bidding options:

- Energy only: Bidding only on the energy market (in dark blue)

- Reserve, low quantile: Bidding energy+aFRR upward+FCR upward, with reserve offers based on a constant low quantile

- Reserve, marginal: Bidding energy+aFRR upward+FCR upward with reserve offers based on marginal optimal quantiles for each reserve

Thanks to the good reliability of the aggregated production forecast at 0.1% generated with exponential distributions, the Rate of UnderFulfillment of the reserve capacity forecast is limited to 0.2 % for the German VPP and 0.09 % for the French VPP. Therefore the results of reserve bidding based on a low quantile (0.1 %), represented in the third column of 4.21, indicate that this bidding method exposes to limited penalties for not delivering FCR+aFRR (0 Eur/MW.h penalties in the German case, and -0.3 Eur/MW.h in the French case). However, one notes that the turnout of average net profit (see last rows) differs between the two countries. A modest increase against bidding energy only is seen in Germany (18.53 against 16.83 Eur/MW.h) while the average net revenue decreases in France (33.05 against 36.37 Eur/MW.h). This is associated to bidding reserve when it is not economical to do so (ie when prices for reserve are lower than those of energy, taking into account expected real-time prices). In contrast, it is observed that bidding methods taking into account prices (Lagrangian relaxation, second column and marginal optimal quantiles, fourth column) lead to an increase in net profit, when compared to bidding energy only, on both countries tested. The higher profit in France is a consequence of the larger dimension of the VPP: it can offer more reserve volume relatively to its capacity (production is less variable), and therefore increases more significantly its earnings when price conditions are favourable compared to a smaller VPP such as the German one. The 5%-CVaR of the net profit is improving with respect to bidding a low quantile for both approaches. Note that it does not necessarily increase with respect to bidding energy only like the average net profit: as the volume of energy sold in real-time is adapted to provide the day-ahead energy volume and requested reserves, it may be associated to diverse price levels for the energy imbalance. Also, most of the time the VPP does not bid reserve at all because prices are not favourable, so the 5-% CVaR of net profit is largely impacted by the conditions on the energy market.

The proposed Lagrangian method for bidding multiple balancing AS achieves average net profits which are close to the method based on marginal optimal quantiles (+10% increase w.r.t. energy only in Germany, +16 % in France). Its bidding behaviour is distinct in that it limits the total reserve volume of FCR+aFRR in order to limit penalties, a fact that the method with marginal optimal quantiles does not consider. As a result, sales on the day-ahead reserve markets and penalties tend to be lower (visible in the French case). The more conservative bidding in reserve results also in a lower average net profit. In conclusion, bidding multiple balancing AS improves

the average profit, and Lagrangian method and independent optimal quantiles deliver close results. However the Lagrangian method is able to limit sales of reserves in order to hedge against penalties associated with reserve under-fulfilments.
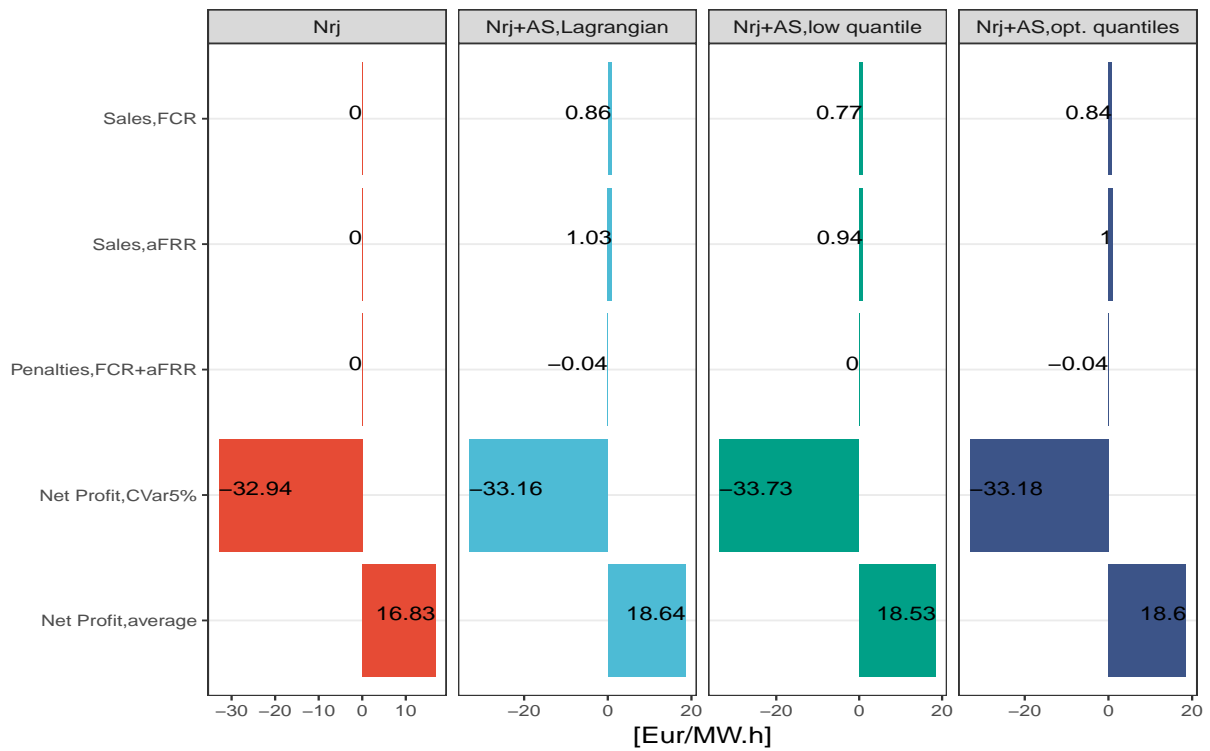
(a) German VPP
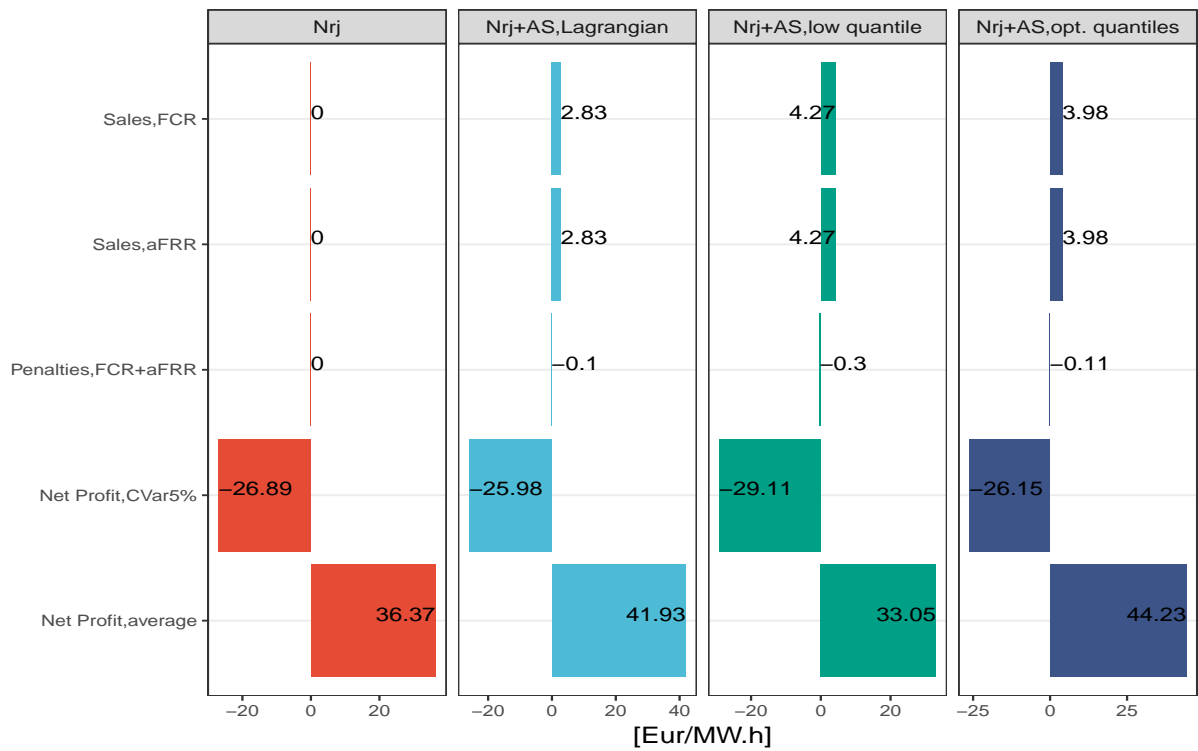
(b) French VPP

Figure 4.21: Revenues of bidding of energy and FCR + aFRR (upward) from a German VPP (top) and French VPP (bottom).

4.10   Conclusion

The VPP operator now disposes of three optimization methods for bidding on energy and balancing AS markets:

- **Optimal quantile for Reserve**

- **Stochastic optimization**

- **Multiple reserve product bidding**

Methods are assessed on several case studies, with different VPP configurations and markets. This diversity in test conditions may be a limit of the present work, but it is also a consequence of the lack of maturity in some balancing AS markets (e.g. in France where prices are still mostly constant, and in Germany where auctions rules have changed significantly): choosing markets with dynamic prices (Germany, Portugal) was necessary to evaluate the added value of the proposed methods. A positive aspect of having different test conditions is that it creates a broader perspective on bidding reserve.

The optimal quantile method presented in Section 4.4 improves the existing bidding strategies for the joint offer of reserve and energy by considering the dependence a VPP composed of wind and PV plants located in different climatic zones. The operator of the VPP can opt either for a strategy offering aFRR with a minimal risk of underfulfilment, or strategies aiming at a higher combined revenue from aFRR and energy. Offers of aFRR in the revenue-maximizing strategies are derived by an optimal quantile using price forecasts. Prices are forecasted with deterministic models, with moderate improvement relative to persistence (15 to 25 % improvement in forecasting absolute or squared errors). Energy offers are adjusted as a function of aFRR offers and expected production. No use of the intraday energy market (i.e. for hedging forecast errors) has been investigated. The mean daily revenue increases from 25.05 Eur/MW.h when offering energy only to 26.33 Eur/MW.h (+5.1%) when offering both energy and aFRR, if offers use deterministic price forecasts.

In the case study, revenue-maximizing offers based on a forecasted optimal quantile generate similar net revenue than reliability-maximizing offers based on a constant quantile (here 1%). Similar increases in average revenues can be obtained with different average volumes (4% to 10% of installed capacity), which shows the interest for producers to implement the most economically efficient strategies for aFRR deployment. The reliability-maximizing strategy has two advantages: simplicity and higher reserve volumes, because offers depend only on the uncertainty in production. As such it can be a compromise between the economic expectation of the renewable producers and the need to fulfill the reserve demand for grid operators. The strategy with optimal quantile independent from production leads to the highest risk for net revenue (highest decrease in conditional Value-at-risk at 1%). The high risk is due to forecast errors on several prices which give the wrong incentive for reserve (offering reserve when it pays less than energy). We observe the same behavior if daily persistent price is used instead of deterministic price forecasts, especially because it can lead to high reserve volumes one day after it would have been effectively needed. The aFRR capacity prices used in this study are highly persistent as they come from a weekly auction. This characteristic contributes to higher average increase in revenue than the other price-based strategies

which tend to offer less. However with persistence price forecast the rate of underfulfilment is the highest of all tested strategies (1.9%), which is an important drawback for practical implementation. All strategies of joint offer and energy increase the financial risk compared to offering energy only, mostly because we choose to avoid penalties in the reserve market by allowing large energy imbalances.

The operator of the VPP may wish to hedge against the price uncertainty while taking into account the uncertainty on production. This is the purpose of the strategy with optimal quantile dependent on production. Here a dependence model using a KDE copula has been proposed to capture the interaction between price spreads and VPP production. A Gradient Boosting Tree model generates probabilistic densities of the price spreads, which are combined with the conditional spread densities originating from the copula models. We obtain the optimal quantile conditional on VPP production by numerical minimization of the losses based on the expected conditional spreads. The offer based on the copula model focuses on events where forecast production is expected to coincide with favorable price spreads. This model seems more robust to large losses than deterministic approaches (lowest decrease in 1%-CVaR) because it combines the uncertainty of prices and production, neglecting separate spikes forecasted for either prices or production.

These results rely on the assumption that penalties paid for failed reserve are linear with the reserve failure. It may be more pertinent for the secure operation of grids to consider higher penalties for large deviations, for instance via a quadratic cost or an exponential utility function. The product length of aFRR has a major impact on the level of capacity that can be offered and the associated revenue. The case study results indicate that aFRR product lengths equal or superior to one day reduce the amount of capacity that can be offered and reduce the increase in revenue due to the combined offer of aFRR and energy.

The chance-constrained optimization presented in Section 4.5 demonstrates that scenarios from direct aggregated production forecast generate more revenue fro reserve bidding on average than those from separate forecasts, because bidding is 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.

Finally, the VPP operator disposes of a bidding strategy for multiple balancing AS services defined in Section 4.6. Formulated as an extension of the classical newsvendor problem, the optimization consists in statisfying the stationarity of KKT conditions associated with the Lagrangian of the problem. Results indicate that this method is effective at hedging against the total penalties possibly incurred when bidding reserve on a several markets. The average net profit increases compared to bidding energy only (up to +22 % assuming perfect knowledge of price), and is superior to the profit obtained when bidding constant very low quantiles based on forecasts of Chapter

3. Overall, the Lagrangian relaxation obtains results close to the alternative method relying on independent optimal quantiles for each balancing AS product. It is however still able to guarantee that the 5%-CVaR will not degrade compared to bidding energy only (improvement up to +4% in relative terms).

Optimal quantile	Stochastic optimization	Multi-product bidding
Maximize average revenue		
Production-independent optimal quantile (*+5.5% average revenue vs energy only*)	Scenarios from **direct forecast** of aggregated production (*+6% average revenue vs scenarios from forecasts by sources*)	Optimal quantiles or Lagrangian relaxation (*+10-+22% average revenue vs energy only*)
Minimize risk of underfulfilment on reserve capacities		
Production-dependent optimal quantile (*- 8 Eur/MW.h 1%-CVaR vs energy only*)	Scenarios of aggregated production from **separate forecasts by sources** (*+18% 5%-CVaR vs scenarios from direct forecast*)	Lagrangian Relaxation (*+4% 5%-CVaR vs energy only*)
Maxim. average revenue + Minimize under-fulfilments		
Production-dependent optimal quantile (*RUF = 0.1 %*)	Chance-constrained optimization (*RUF = 0 %*)	Lagrangian relaxation (*RUF = 0.2 %*)

Table 4.7: Summary of results on bidding methodologies for balancing AS from a renewable VPP

4.11 Résumé en français

Introduction

L'opérateur d'une centrale virtuelle renouvelable variable qui souhaite fournir des services système d'équilibrage doit décider de son offre de réserve en fonction des prévisions de production et des conditions de marché prévues. Il dispose de prévisions fiables de production agrégée générées au Chapitre 2. Il peut également, s'il donne la priorité à la minimisation du risque de défaut de réserve par rapport à la maximisation de son profit (choix qui peut être imposé par le gestionnaire de réseau), baser son offre de réserve sur les prévisions à quantiles bas obtenues en Chapitre 3. Dans le cas contraire, l'offre de réserve doit être optimisée selon les incertitudes de production et de marché. Il existe de nombreux travaux sur l'optimisation d'offre par les centrales renouvelable variables sur les marchés de l'énergie (voir [143]), et quelques travaux récents sur l'offre de réserve (production éolienne [89], production PV [170]). Toutefois, des verrous restent à lever pour l'optimisation d'offre de réserve par une centrale virtuelle renouvelable variable:

- Les marchés de réserve sont dimensionnés de façon statique et conservative. De plus, de nombreux pays disposent de surcapacités. La probabilité pour un fournisseur de réserve d'être activé est donc faible et incertaine, elle doit être prévue et intégrée à la stratégie d'offre de l'opérateur.

- Un facteur primordial dans le choix de privilégier la réserve par rapport à la vente sur le marché de l'énergie est l'écart de prix (ou *spread*) entre réserve et énergie. Cet écart de prix doit être prévu et l'incertitude de la prévision doit être quantifiée, afin de mesurer le risque économique associé.

- L'utilité des prévisions probabilistes de production agrégée pour l'optimisation d'offre de réserve doit être démontrée.

Méthodologie

Compte tenu de ces verrous, le présent Chapitre propose plusieurs méthodes pour optimiser l'offre en fonction des besoins de l'agrégateur gérant la centrale virtuelle. Les différentes solutions sont articulées sous forme d'un diagramme en Figure 4.22. Si l'agrégateur souhaite uniquement minimiser le risque technique de défaillance sur la réserve, il peut utiliser la Méthode 1 en Section 4.3 basée sur un quantile constant, avec des prévisions très fiables issues du Chapitre 3. S'il souhaite par contre optimiser son revenu global en fonction des prix attendus sur les différents marchés, il peut opter pour la méthode 2 qui propose d'optimiser l'offre de réserve à partir des prévision probabilistes de production du Chapitre 2 et de prévisions de prix. Cette méthode n'intègre pas explicitement les variations temporelles de production agrégée, ceci peut conduire à des défauts de réserve sur la durée de période de validité de la réserve. L'optimisation stochastique proposée en Section 4.5, permet d'adapter l'offre en fonction des variations modélisées par les scénarios de production. Enfin l'agrégateur peut utiliser la méthodologie définie en 4.6 pour offrir simultanément plusieurs services système.
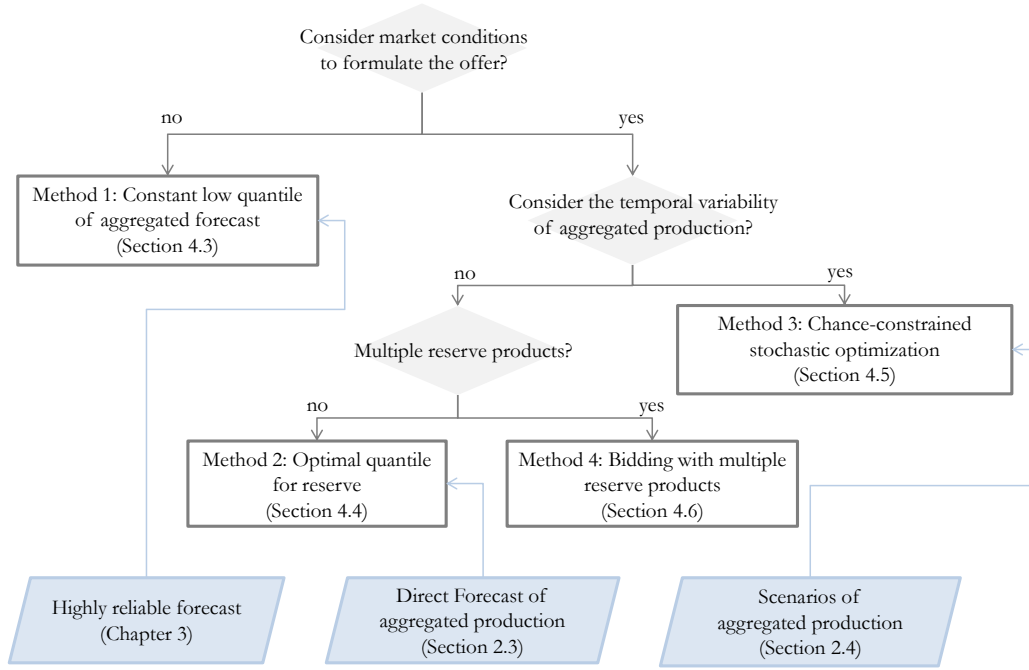
Figure 4.22: Diagramme de la méthodologie d'offre optimale de réserve

Quantile optimal pour la réserve, indépendance entre prix et production

La méthode du quantile optimal est décomposée en 2 approches distinctes. La première suppose que les prix sur les marchés de l'énergie sont indépendants de la production renouvelable de la centrale virtuelle. Il s'agit alors de déterminer le quantile optimal τ_R à partir de prévisions déterministes des prix composant les écarts entre marché de l'énergie et marché de la réserve, pour les ventes en day-ahead $\Delta\hat{\pi}_{RE}^{da}$ et les pénalités à la livraison ou *real-time* $\Delta\hat{\pi}_{RE}^{*}$.

$$\tau_R = \frac{\Delta\hat{\pi}_{RE}^{da}}{\Delta\hat{\pi}_{RE}^{da} + \Delta\hat{\pi}_{RE}^{*}} \tag{4.66}$$

Les prévisions de prix sont effectuées avec un modèle spécifique pour chaque prix, étant donné que chaque marché a ses dynamiques propres et qu'il n'est pas aisé de prévoir directement les écarts de prix avec un modèle unique. La Table 4.8 synthétise les modèles utilisés. A côté des modèles classiques pour les prévisions de prix déjà existantes, le modèle Support Vector Regression a été identifié comme adapté aux prévisions sur le marché de la réserve (ici secondaire, aFRR), pour lequel peu de références sont disponibles, en raison de sa capacité à bien généraliser des comportements à partir d'observations fortement variables [156].

Quantile optimal pour la réserve, dépendance entre prix et production

La pénétration croissante des renouvelables dans le mix électrique influe sur les prix des marchés de l'électricité car ils offrent leur production à des prix marginaux quasi nuls, ce qui déplace le prix marginal résultant des enchères. Sachant qu'il est plausible que la production de la centrale

Variable	Model	Explanatory variables
Energy price π_E^{da}	ARIMAX	lagged prices and volumes
Reserve capacity price π_R^{da}	Random Forests	lagged prices
Reserve average price $\pi_{\bar{R}}^{da}$	Support Vector Regression	reserve demand, national forecasts of load and renewable production
Reserve activation probability a_R	Support Vector Regression + PCA	past prices, reserve demand, national forecasts of load and renewable production
Energy imbalance price $\pi_E^{rt,-}$	Markov Regime Switching	past prices

Table 4.8: Résumé des modèles de prévision de prix disponibles

virtuelle se trouve en phase avec la production renoulevable totale à l'échelle du marché, et que la prévision des prix de chaque marché est délicate, il est ici proposé de pondérer le choix du quantile optimal par une dépendance des écarts de prix prévus avec la production prévue.

Cette pondération est effectué en (4.67) sur la modélisation d'une copule non-paramétrique $\hat{c}$. Cette copule quantifie la dépendance entre les écarts de prix entre réserve et énergie ΔP_{RE} d'une part, et de la production agrégée prévue F_Y d'autre part. Ceci nécessite une prévision probabiliste des écarts de prix $\hat{f}_{\Delta P_{RE}}$, qui se base sur les prévisions déterministes effectuées pour le cas précédent.

$$\hat{f}_{\Delta P_{RE}|Y=\hat{y}}(\pi) = \hat{f}_{\Delta P_{RE}}(\pi)\hat{c}(F_Y(\hat{y}), \hat{f}_{\Delta P_{RE}}(\pi)), \quad \pi \in \mathbb{R} \tag{4.67}$$

Optimisation stochastique

Si l'agrégateur souhaite intégrer la variabilité temporelle de production agrégée dans son offre, un modèle d'optimisation stochastique est proposé. Les scénarios de production agrégée produits en Chapitre 2 sont intégrés dans la formulation de l'équilibre entre ventes et production. De plus, une variante est proposée avec contrainte probabiliste sur la réserve fournie en temps-réel. Cette contrainte, définie en (4.68), permet de fixer une fréquence maximum admissible de défaut sur la réserve $\Delta R_{\omega t}^{\uparrow,-}$, avec une probabilité $1 - \epsilon$.

$$Pr(\Delta R_{\omega t}^{\uparrow,-} \leq 0) \geq 1 - \epsilon \quad \forall \omega, t \tag{4.68}$$
$$Pr(\Delta R_{\omega t}^{\downarrow,-} \leq 0) \geq 1 - \epsilon \quad \forall \omega, t$$

Offre de produits multiples

L'optimisation stochastique précédente peut être adaptée à une offre de produits multiples. Il est proposé ici une alternative qui ne nécessite pas de génération de scénarios, mais seulement l'utilisation de la prévision probabiliste de production agrégée. On peut formuler le problème d'optimisation une offre de produits de réserve multiples comme en (4.70), en y associant une

contrainte de ressource S représentant la bande de productible disponible pour la fourniture de réserve.

$$\underset{R_i}{\arg\max} \, \mathbb{E}[\sum_{i \in FCR,aFRR} \Delta\pi_{RE,i}^{da} R_i - \Delta\pi_{RE,i}^{rt}(R_i - y)] \tag{4.69}$$

$$s.t. \sum_{i \in FCR,aFRR} s_i R_i - S_{multi} \leq 0 \tag{4.70}$$

Dans cette équation, s_i représent le coefficient de ressource associée à la réserve i. La contrainte de ressource S_{multi} est égale à la puissance installée si la réserve est fournie uniquement à la baisse, et sinon à la différence entre puissance installée et offre d'énergie.

$$S_{multi} = \begin{cases} y_{max}, & \text{si} \quad R^{\downarrow} \\ y_{max} - E_{da}, & \text{si} \quad (R^{\uparrow}, R^{\uparrow,\downarrow}) \end{cases} \tag{4.71}$$

Ce problème d'optimisation est résolu par relaxation Lagrangienne, en considérant que la fonction objectif définie par la prévision de production est continument dérivable. et en recherchant les valeurs prises par les offres R_i et par le multiplicateur de Lagrange qui respectent les conditions de stationnarité de Karun-Kush-Tucker.

Cas d'étude

Les méthodologies ci-dessus sont évaluées à l'aide de différents cas d'étude, couvrant une variété de situations et de marchés (France, Allemagne, Portugal) afin d'obtenir une vision du potentiel de gain supplémentaire associé à l'offre de réserve.

Les métriques utilisée pour évaluer les stratégies d'offre sont les suivantes:

- Revenu moyen: espérance des revenus sur l'ensemble de la période d'évaluation.

- Valeur à risque conditionée ou Conditional Value at Risk (CVaR): espérance des revenus inférieurs à un quantile de référence de valeur nominale τ_R.

$$\tau_{rev}\%\text{-CVaR} = \mathbb{E}(r), r \leq F_{rev}^{-1}(\tau_{rev}) \tag{4.72}$$

- Fréquence de défaillance de fourniture de réserve ou *Rate of Underfulfilment*, (RUF): Fréquence des évènements pour lesquels la production agrégée y^{agg} est inférieure à l'offre de reserve R.

$$RUF = \frac{1}{N} \sum_{i=1}^{N} 1_{y_i^{agg} < R_i} \tag{4.73}$$
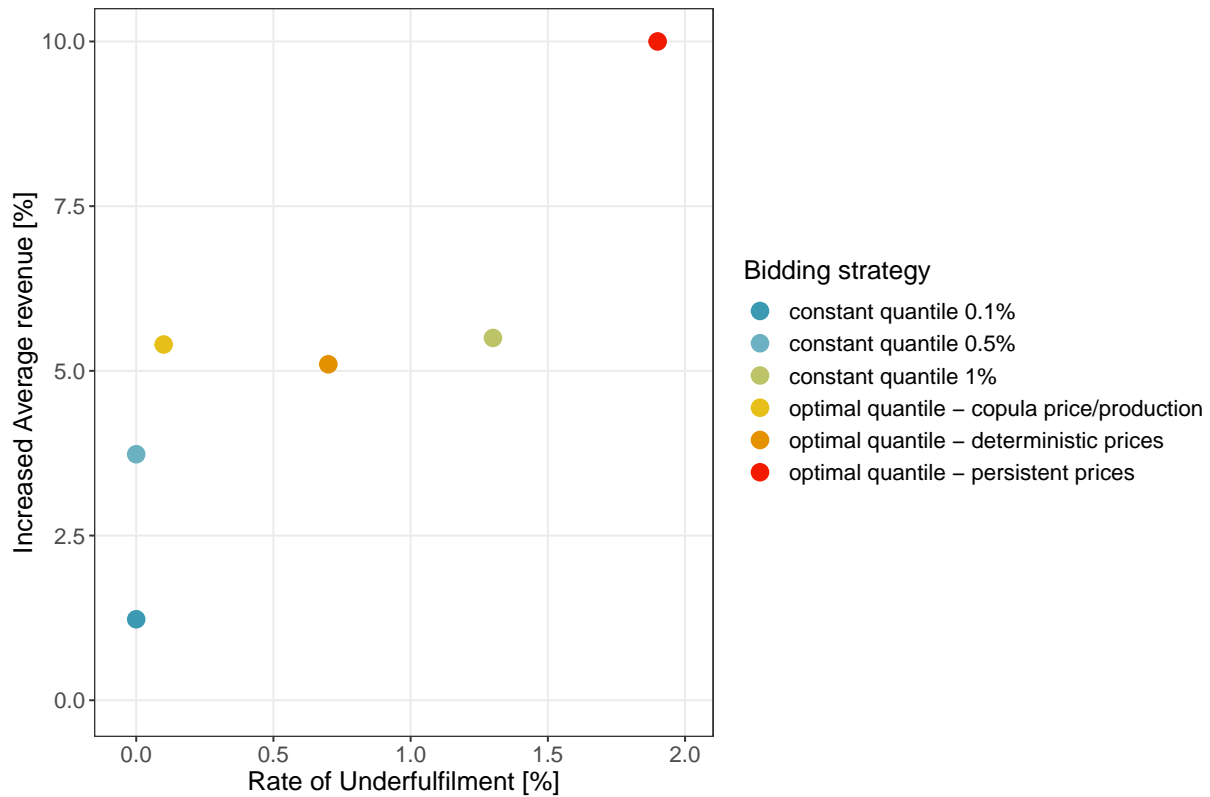
Figure 4.23: Résultats des stratégies d'offre d'énergie et de réserve basées sur une prévision probabiliste de production agrégée, en fonction du risque de défaillance sur la réserve (abscisse) et l'augmentation du revenu moyen comparé à la vente d'énergie seule (ordonnée)

Résultats

Les résultats principaux des différents cas d'étude sont reportés en Table 4.9. Les meilleures configurations pour chaque méthode sont sélectionnées en fonction de l'objectif principal de l'opérateur du VPP.

Si l'opérateur du VPP ne s'intéresse qu'à la maximisation du revenu moyen, il peut choisir les solutions les plus simples pour chaque méthode, à savoir le quantile optimal sans dépendance entre production et prix, l'optimisation stochastique avec scénarios issus de prévision agrégée directe, et enfin l'offre multi-produits basée uniquement sur des quantiles optimaux indépendants pour chaque réserve.

Ceci s'observe par exemple en Figure 4.23, qui représente l'augmentation de revenu moyen associée à l'offre conjointe d'énergie et de réserve comparée au revenu moyen obtenu par vente d'énergie seule. En effet, l'offre conjointe d'énergie et de réserve, avec une capacité de réserve égale au quantile optimal basé sur des prévisions de prix déterministes (sans dépendance entre production et prix) permet d'obtenir une augmentation du revenu de l'ordre de 5% et une fréquence de défaillance inférieure à 1% (cf. point orange), ce qui la classe parmi les solutions les plus profitables en termes de revenu moyen tout en étant acceptable techniquement.

Sinon, les solutions plus avancées ont des avantages quantifiés s'il souhaite par exemple min-

imiser le risque prix sur le revenu. Le quantile optimal dépendant de la production a par exemple la plus haute 1%-CVaR parmi l'ensemble des stratégies d'offre basées sur une prévision probabiliste de production agrégée, tout comme les scénarios de prévision séparée permettent d'augmenter la 5%-CVaR par rapport aux scénarios issus de prévision directe. Enfin l'optimisation multi-produits permet d'augmenter le revenu moyen de 10% à 22 % en comparaison d'une vente d'énergie seule, et la méthode par relaxation Lagrangienne permet de limiter l'offre totale de réserve, et ainsi de réduire le risque de non-fourniture de réserve tout comme le risque financier pris (jusqu' à +4% de 5%-CVaR comparé à la vente d'énergie seule).

Quantile optimal	Optimisation stochastique	Offre multi-produits
Maximisation du revenu moyen		
Quantile optimal **indépendant** de la production *(+5.5% de revenu moyen vs vente énergie seule)*	Scénarios par prévision **directe** de la production agrégée *(+6% de revenu moyen vs scénarios par sources)*	Quantiles optimaux ou Relaxation Lagrangienne *(+10-+22% de revenu moyen vs vente d'énergie seule)*
Minimisation du risque sur le revenu		
Quantile optimal dépendant de la production *(- 8 Eur/MW.h de 1%-CVaR vs vente d'énergie seule)*	Scénarios par prévision **séparée par sources** de la production agrégée *(+18% de 5%-CVaR vs scénarios par prévision directe)*	Relaxation Lagrangienne *(+4% de 5%-CVaR vs vente d'énergie seule)*
Maxim. du revenu + Minimisation du risque de défauts		
Quantile optimal dépendant de la production *(RUF = 0.1 %)*	Formulation avec contraintes probabilistes *(RUF = 0 %)*	Relaxation Lagrangienne *(RUF = 0.2 %)*

Table 4.9: Résumé des meilleures configurations pour chaque méthode d'optimisation d'offre, en fonction de l'objectif principal de l'opérateur de la centrale virtuelle.

Conclusion

L'opérateur d'une centrale virtuelle renouvelable variable dispose maintenant de 4 méthodologies pour optimiser son offre:

- Offre d'un quantile bas constant, basé sur les prévisions extrêmes du Chapitre 3.

- Offre selon un quantile optimal obtenu par prévision de la production et des prix

- Offre selon une optimisation stochastique avec scénarios de production agrégée

- Offre de services système multiples

Chaque méthode d'offre présente ses propres avantages et inconvénients qui sont résumés ici. L'offre par quantile bas constant minimise le risque technique de défaut mais fait perdre naturellement du revenu potentiellement captable lorsque les prix de la réserve sont inférieurs à ceux de l'énergie. L'offre par quantile optimal, avec sa déclinaison intégrant une dépendance entre production et écarts de prix, offre une solution basée intégralement sur des prévisions probabilistes. Ceci permet d'assurer une augmentation de revenu par rapport à la vente d'énergie seule, mais aussi de limiter le risque d'encourir des pénalités élevées.

L'optimisation stochastique permet de mieux intégrer les variations temporelles de production agrégée à la stratégie, et donc de réduire les volumes offerts en prévision des rampes ou des creux attendus durant la durée de livraison de réserve (de typiquement quelques heures). Sa variante avec contrainte probabiliste permet effectivement de contrôler la fréquence des défauts de réserve. L'utilisation de scénarios de prévision agrégée issus de prévisions séparées par source d'énergie est à privilégier par rapport aux prévisions directes de production agrégée. En effet, elles modélisent mieux la variabilité du processus et permettent ainsi de mieux valoriser le potentiel de réserve.

Enfin la méthode d'offre d'énergie et de services système multiples démontre l'accroissement de gain lorsque plusieurs services système sont offerts simultanément (+10 à +22 % selon le cas d'étude). Une option plus naïve basée sur des quantiles optimaux indépendants par réserve, si elle peut conduire à une augmentation du revenu moyen, fait courir un risque plus élevé d'encourir des pénalités importantes sur la réserve: la méthode par Relaxation Lagrangienne réduit les volumes totaux de réserve vendus et assure que la Conditional Value at Risk à 5% ne se dégrade pas comparé à la vente d'énergie seule.

Chapter 5

Conclusion

5.1 Summary

In Chapter 1, the interest for the provision of balancing Ancillary Services (AS) by renewables is presented. The necessity to aggregate variable producers is explained and illustrated with state-of-the art implementations in the industry, namely Virtual Power Plants (VPP). Such VPP have the technical capacity to provide reserve, but challenges still need to be overcome. In particular, the production uncertainty is still significant, and this impacts the possible bidding strategies of VPP on reserve and energy markets. The global objective of this thesis is therefore to optimize the bidding strategy of a VPP aggregating VRE plants of various energy sources and providing balancing AS.

Best candidates for the provision of reserve are aggregations with the lowest variability in production. Therefore the smoothing effect in a multi-source aggregation which gathers different energy sources must be evaluated, and can serve as a basis for the composition of a reserve-prone VPP. A lack of methods for the probabilistic forecasting of a multi-source aggregation has been identified in the existing literature. Such forecasts are needed to safely offer reserve on ancillary services markets, i.e. minimize the risk of having not enough available power at real-time to curtail or ramp up production in the real-time, following requests from grid operators. Given that ancillary services are critical for electrical networks, operators may require maximum levels of reliability, which is particularly challenging for common renewable forecasting models. Finally, reserve bidding strategies should specifically address the uncertainties on both aggregated production and market conditions.

After reviewing the existing literature, this work is articulated around seven research questions:

RQ1: How to develop a direct probabilistic forecast of multi-source aggregated VRE production?

RQ2: What is the optimal methodology to generate scenarios of aggregated multi-source VRE production?

RQ3: What is the influence of the mix of energy sources on the variability of VPP production and forecast performance ?

RQ 4: How to develop specific forecasting models for extremely low quantiles of aggregated VRE production, with reliability > 99% ?

RQ 5: How to integrate market uncertainty into reserve bidding strategies?

RQ 6: Which type of bidding strategy provides best results in terms of maximized revenue and minimized risk of reserve under-fulfiment?

RQ 7: How to optimize a bidding strategy integrating energy and multiple services?

The development of efficient probabilistic forecasting models of aggregated VRE production is the purpose of Chapter 2. Before building forecasting models, the variability of the aggregated production process is characterized. In response to RQ3, it is found that correlations between plants become low if plants are distant of hundreds of kilometers and depend largely on the energy sources. Moreover, the smoothing factor of an aggregation mixing energy sources is larger than aggregation of a single energy source. This reduction in variability is benefitial for the provision of balancing AS.

A direct approach for the aggregated forecasting is proposed, implementing models based on decision-trees and neural networks. A specific adaptation of convolutional neural networks (CNN) and Long-Short-Term-Memory networks (LSTM) for the specific problem is proposed. It is found that the CNN can outperform the state-of-the art Quantile Regression Forests model on deterministic errors (-1.5 % RMSE) and probabilistic score (- 0.4 % in CRPS), whereas the LSTM has a lower deterministic error but lower probabilistic score due to a lack of reliability. The developed direct forecasting models and the associated case studies answer Research Question 1.

The second part of the Chapter presents a methodology to generate short-term scenarios of aggregated production. It is tested whether scenarios generated from separate forecasts by energy sources, are reproducing better the temporal correlations in the aggregated production signal than scenarios generated directly from the aggregated production forecast. Case study results indicate that scenarios from separate forecasts by source, which can reproduce the specific variability of each source, but are unaware of the total resulting uncertainty at the VPP level, represent overall better the aggregated variability (lower Variogram Scores, generally lower Brier Scores on ramps). This is particularly true when the aggregated production is heavily dependent on the horizon, for instance when the PV share in the total capacity is large. This part answers RQ2, and brings additional material to Research Question 3. The methodology for the generation of scenarios has been published in *Applied Energy* [171].

As previously stated, network operators may require maximum reliability for the reserve offer. This advocates for advanced forecasting models for the extremely low quantiles of aggregated production. The afore-mentioned models, decision-tree based or neural networks, cannot be readily applied to forecast extremes. In Chapter 3, quantile regression models are adapted to the forecast of extremes: the CNN model incorporates a specific filtering step denominated *min-pooling*, and both LSTM and CNN are trained on the minimization of a *Skill score* instead of the classical quantile loss. As an alternative, parametric models fit a specific distribution to the lowest quantiles. The proposed distributions are the exponential distribution, which is simple to fit because governed by

a unique rate parameter, and the Generalized Pareto distribution, which can be inferred using the theoretical background of Extreme Value Theory (EVT). A clustering of conditions is proposed to increase the selectivity of parametric models, following the observation that the distribution of extremes may be conditioned by production levels or amplitudes of weather conditions across aggregated plants. Finally, mixture density networks are proposed to combine the learning capacity of neural networks with the flexibility offered by distribution mixtures. A Gaussian mixture, where each marginal distribution is parametrized by one plant of the aggregation, is compared to a specific Beta Mixture which enables to remain in the original bounded space of aggregated production. The evaluation of forecasts is based on a weighted version of the quantile score, to compare forecast quantiles on an even ground, regardless of their absolute value. It is found that Beta Mixture, the EVT and the CNN regression have the best score (18 % relative improvement compared to QRF). The score is evaluated by a weighted Quantile Score adapted to the left tail of the distribution. The best reliability on lower quantiles is achieved with the CNN incorporating a min-pooling layer (all quantiles between 0.1 % and 0.5 % lie within the uncertainty bars quantify the sampling effect by consistency resampling, unique model to so). The developed forecasting models and the associated case study answer to Research Question 4. The comparison of forecasting models for highly reliable forecasts of aggregated VRE production is integrated into an article in preparation for *IEEE Transactions on Smart Grid.*

The last Chapter 4 deals with optimal reserve bidding strategies, that make use of the forecasts developed earlier. When the VPP operator must decide which volumes to bid on the energy market and on a specific reserve market, such as aFRR for instance, it can base its decisions on an optimal quantile for reserve. In Section 4.4.1, the optimal quantile is derived analytically, following classical results from the inventory optimization theory. This approach is valid as long as prices and renewable production are considered independent. After determining the optimal quantile, the reserve offer is obtained as a function of probabilistic forecasts of production generated in Chapter 2 and deterministic forecasts of all prices involved in the markets for energy and reserve. Given that renewables are expected to bear a significant weight in the total generation capacity of most countries in the following years, it is likely that prices on energy and reserve markets will actually show some degree of dependence to the renewable production. A bivariate copula is proposed in Section 4.4 to capture the dependence between the production of the VPP and spread prices between reserve and energy markets. The distribution of spread prices are forecasted by means of a Gradient Boosting Tree model. The bidding revenue is found to increase on average (around +5 %) when a Wind/PV VPP bids jointly energy and reserve, whereas the Conditional Value at Risk at 1% level decreases (ie the revenue that is expected to be exceeded with 99% probability decreases). The optimal quantile with dependence model between VPP production and prices mitigates this risk, with the lowest decrease of 1-CVaR% among all alternatives. The Rate of Under-Fulfillment of reserve is also limited to 0.1%. This work has been published in *IEEE Transactions for Power Systems* [172]. By considering market uncertainties on energy and reserve, the bidding strategy replies to Research Question 5. The integration of forecasts of production and prices into a reserve bidding strategy, with the option of using probabilistic forecasts of prices, gives a strategy fully based on forecasts, without the need to employ a more complex optimization tool.

A limitation of the optimal quantile comes from the fact that it does not integrate the temporal variability of aggregated production, which may impact the availability of reserve over its validity period. In addition, the underfulfilment is observed a posteriori instead of being included in the optimization model. A stochastic optimization is therefore developed in Section 4.5, where scenarios of aggregated production enable to model the temporal variability. Furthermore, a chance-constrained variant of the optimization problem integrates the rate of underfulfillment as a probabilistic constraint. A case study on the joint bidding of energy and aFRR by a Wind/PV/Hydro VPP shows that the chance-constraint enables to minimize the rate of underfulfillment, with zero underfulfillment observed (the time resolution is 30 minutes, rapid production variations with a finer resolution could have created more underfulfillments). Lastly, scenarios generated by separate forecasts of production by energy source lead to a higher CVaR revenue, showing that their better description of the variability in VPP production leads to a more robust bidding outcome that scenarios from direct multi-source forecast. The stochastic optimization was included in the article mentioned earlier for scenarios [171]. This comparison of optimization outcomes in terms of revenues and technical risk of reserve underfulfilment, and the formulation of a chance-constrained optimization problem, is a contribution to Research Question 6.

Finally, the VPP operator may require a strategy for bidding multiple balancing AS products simultaneously. A strategy is proposed in Section 4.6, extending the optimal quantile with findings from the multi-product newsvendor theory. It permits to bid multiple products based on probabilistic forecasts of production (and prices, if the dependence between prices and production is considered). The problem is formulated as a stock-constrained maximization of profit, and solved by exploiting stationarity Karun-Kusch-Tucker conditions after Lagrangian relaxation. The case study of the multi-service bidding optimization shows that offering energy and multiple balancing ancillary services with the proposed Lagrangian resolution instead of bidding independently optimal quantiles for each balancing service limits the amount of penalties incurred, while ensuring an additional profit when compared to sales of energy only (up to 22% when compared to bidding energy only). The model proposed in this Section constitutes a simple answer to Research Question 7.

5.2 Analysis of contributions

Each chapter of the thesis contains elements that support the original contributions claimed in the Introduction.

The direct probabilistic forecasting of aggregated multi-source renewable production is an innovation that structures Chapter 2. The proposed forecasting models are adapted to the context of VRE-based VPPs by applying regression on specific data configurations adequate for VPPs (e.g. placement of plants of VPP shuffled or contiguously by energy source in the feature space of the CNN), and by evaluating them on different VPP situations in terms of mix of energy sources (PV-Wind-Hydro, PV-Wind...) and aggregation levels (aggregated forecast and forecasts by energy source). The set of forecasting models proposed corresponds to **Contribution 2**. Lastly, the approaches for generating trajectories of VPP production based on direct QRF forecast of total

production or separate QRF probabilistic forecasts by energy source are compared as announced in **Contribution 3**. The comparison shows that trajectories generated by energy source are more representative of the observed VPP variability, except for specific events such as rapid production ramps.

The forecasting models for very low quantiles developed in Chapter 3 correspond to **Contribution 4**: they form a set of models adapted to the context of aggregated VRE production, with various levels of complexity. A simple model such as the exponential distribution parametrized by clusterized production regimes and features summarizing weather conditions and their dispersion shows satisfactory results. It should suffice for an application to reserve bidding. More complex models such as the convolutional neural network adapted to very low quantiles have certainly more improvement potential than simple models. The regression on Skill Score is found to be a promising solution to issue reliable forecasts at very low quantiles. In the case of mixture density networks, the Beta mixture appears to be more adapted to the bounded process of VRE production compared to the Gaussian mixture parametrized by installed capacities within the VPP. Nonetheless, it would be useful to analyze in more detail how learning is done and shared between the mixture density and the underlying neural network.

The bidding strategies including ancillary services provision for renewables presented in Chapter 4 have participated to the following contributions:

Contribution 1: Each bidding strategy constitutes a solution for the last step in the value chain of AS bidding by a renewable VPP. The last step consists in an optimization of revenue and associated technical risks, which considers realistically technical constraints and is adapted to the context of vRES aggregated forecasting:

- The *rate of under-fulfilment* is measured ex-post and is explicitly minimized in (1) the bidding strategy based on a low constant quantile obtained by a dedicated extreme production forecasting model and in (2) a strategy based on chance-constrained optimization.

- Bidding strategies are *adapted to vRES aggregated forecasting* in the sense that they incorporate either a direct probabilistic forecast of aggregated production or trajectories of aggregated production, that may be obtained from forecasts by energy source to depict the total variability more accurately.

Contribution 5: Bidding strategies integrate uncertainties on variable production and electricity markets:

- *Uncertainties on energy and reserve markets* are modelled by means of prediction models dedicated to each market quantity influencing the offer strategy, including the activation probability of the reserve offered by the VPP.

- *Uncertainties on the variable production* of the VPP serve are modelled by the probabilistic prediction models of production and as inputs to the bidding strategies.

Contribution 6: Trajectories of aggregated production generated and evaluated in Chapter 2 in terms of forecasting error are integrated here in the bidding strategy based on stochastic

optimization. By doing so, *the impact of trajectories of aggregated production on revenues* at the end of the value chain is assessed.

Contribution 7: The bidding strategy for multiple AS products developed in Section 4.6 constitutes a *simple solution for the joint offer of energy and multiple AS*, because it does not need to rely on stochastic optimization, which requires more modelling steps (scenario generation, solving the linearized problem with possible integer variables, etc.).

5.3 Perspectives

This section describes briefly possible perspectives for further work, building on the propositions formulated in this thesis.

Optimal real-time dispatch of reserve.

An untouched issue in the present work is the dispatch in real-time of the offered reserve to the individual power plants composing the aggregation. Several solutions exist for the disaggregation of a control setpoint based on simple merit-order activation [26] or model predictive control [94]. However, a robust optimal algorithm specifically designed for a 100% VRE-based VPP should be developed. The algorithm could integrate the probability of plant failure, the real-time dynamics of generators and the latest information on available active power. Models inspired by tree search of faults or reinforcement learning could beat state-of-the-art solutions.

Multivariate forecast of extremes.

It has been found in this work that multivariate scenarios from separate forecasts by energy source reproduce the aggregated production process more accurately than scenarios from direct forecast. It may be interesting to see if a multivariate forecast of extremes, based on separate forecasts by energy source and merged at a later step, would similarly improve the modelling of extremes compared to the direct models of aggregated production proposed in this work. Models presented here such as EVT and Mixtures could be extended to the multivariate context (see for instance[138] for bivariate EVT model on stock prices, and [139] for Dirichlet mixtures, adapted to multiple bounded processes).

Reconciliation of hierarchical forecasts of VPP production

A possible extension of the model chain presented in this thesis is to produce forecasts at different hierarchical levels of the VPP, from the plant to the total aggregation, while ensuring coherence of the forecasted densities or trajectories. Reconciliation applied to day-ahead deterministic forecasts of wind power production has shown to improve forecasting performance on the aggregated production level [56]. A cross-validation method for reconciling probabilistic forecasts has been proposed by [173] as a score-oriented alternative to the mean-coherent approach of [63]. Additional research is however needed to develop a tailored approach in the context of multi-source aggregated VRE production, where unreconciled base forecasts at the level of plants or of energy sources can have very different patterns of error.

Improve price forecasts in energy and balancing AS markets, or get rid of price forecasts altogether.

The offering strategies presented here relying on production and price forecasts. Forecast-free

approaches, such as reinforcement learning, optimizing Alternative Direction Method of Multipliers or neural networks directly on prices have the capacity to formulate interesting bidding offers, bypassing the forecasting stages. However, decision-makers will probably request anyhow forecasts of production and prices. These forecasts constitute a source of information which has an intrinsic value, and can be integrated in various models (optimal quantiles or stochastic optimisations). A simplification of the proposed price forecasting models could consist in developing a direct forecast of spread prices between energy and reserve, for instance relying on neural networks [174] or statistical methods tailored to the specific dynamics of all markets.

Bidding strategy integrating the intraday horizon and market power.

Markets for the provision of balancing energy are evolving towards gate closures close to the operative window of the energy intraday market, from one hour to 15 minutes before delivery. The statistical forecasting of intraday energy prices is challenging, because exchanges are realized continuously in some countries (e.g. in France). Furthermore, prices for balancing energy are expected to change levels when their gate closure is displaced closer to delivery. A fundamental modelling of the formation of prices in balancing markets, integrating possible intraday exchanges, would be valuable for reserve bidding strategies. Energy and balancing markets are clearly interdependent, formalizing relationships between those markets should improve the potential bidding strategies. Finally, a large VRE-based VPP (e.g. disposing of 100 MW production capacity) may have some degree of market power in balancing markets. Programming the bidding problem with equilibrium constraints or game theory approaches may lead to specific strategical decisions for VRE-based reserve. The competition or collaboration with storage systems should be properly investigated.

Résumé en français

Les renouvelables intermittents sont destinés à fournir des services système de réserve en fréquence, car ils remplacent progressivement les générateurs conventionnels programmables dans le mix électrique. Toutefois l'incertitude de production étant élevée au niveau des centrales intermittentes prises individuellement, la fourniture de réserve intermittente doit se baser sur l'agrégation de centrales au sein d'une centrale virtuelle ou *Virtual Power Plant, VPP*. De premières expérimentations industrielles de VPP fournissant de la réserve ont démontré la faisabilité technique de réguler en quelques secondes voire quelques minutes un pool de centrales dispersées au sein d'un pays ou d'une zone plus large. L'incertitude de production agrégée est réduite du fait du lissage et des complémentarités entre sources d'énergie (éolien, PV, voire hydraulique fil de l'eau), mais elle reste significative. Ceci impacte fortement les stratégies d'offre de ces centrales virtuelles sur les marchés d'énergie et de réserve: les écarts entre production observée et volumes offerts sont pénalisés. Les pénalités sont fortes sur les marchés de réserve car une insuffisance de productible met en péril la sécurité du réseau.

L'étude de la littérature fait émerger 7 questions de recherche orientant le travail de thèse:

RQ1 Comment développer une prévision probabiliste directe de la production intermittente agrégée multi-source ?

RQ2 Quelle est la méthode optimale de génération des scenarios de production intermittente agrégée multi-source ?

RQ3 Quelle est l'influence du mix de sources d'énergie sur la variabilité de la production d'une centrale virtuelle et sur la performance de prévision ?

RQ4 Comment développer des modèles de prévision spécifiques aux quantiles extrêmement bas de production agrégée, avec une fiabilité > 99 %?

RQ5 : Comment intégrer les incertitudes de marchés dans les stratégies d'offre de réserve ?

RQ6 :Quel type de stratégie d'offre permet les meilleurs résultats en termes de maximisation de revenu et de minimisation du risque de non-fourniture de réserve ?

RQ7 : Comment optimiser une stratégie d'offre d'énergie et de services système multiples ?

L'objectif global de cette thèse est donc d'optimiser la valeur de la production agrégée intermittente dans le contexte de la fourniture de services système, en proposant des prévisions probabilistes fiables de la production et des stratégies d'offre de réserve par des centrales virtuelles intermittentes, qui peuvent contenir des proportions variables des différentes sources d'énergie intermittente, notamment photovoltaïque, éolien et hydraulique au fil de l'eau.

Propositions et résultats

Pour répondre à l'objectif global de la thèse, le Chapitre 2 caractérise l'incertitude de production agrégée, puis propose des modèles probabilistes de prévision directe de la production agrégée. La

188

prévision est effectuée à l'horizon journalier ou *day-ahead*. C'est principalement à cet horizon (au moment de la rédaction de cette thèse) que les offres d'énergie et de capacité de réglage à offrir sur les marchés sont réalisées. Les cas d'étude montrent que les modèles standards basés sur des arbres de décision comme le Quantile Regression Forests (QRF) et le Gradient Boosting Trees (GBT) sont adaptés à la prévision probabiliste d'une agrégation comportant plusieurs dizaines de centrales intermittentes, avec une fiabilité comprise dans les intervalles de confiance canoniques sur l'ensemble de la distribution (quantiles de 1% à 99%). Des modèles de régression quantile basés sur des réseaux de neurones sont ensuite proposés pour creuser l'apprentissage des interactions complexes entre les facteurs explicatifs. Des architectures spécifiques de réseaux convolutifs (ou *Convolutional Neural Networks, CNN*) et récurrents (en particulier *Long Short Term Memory, LSTM*) sont proposées pour effectuer une régression quantile de production agrégée. Les cas d'études montrent que le modèle CNN est capable de prédire avec un meilleur score probabiliste que le QRF la production agrégée. Cette amélioration est due en particulier à une plus grande finesse des prévisions. Le modèle LSTM lui améliore le score déterministe (RMSE) mais est moins bon que le QRF du à une grande finesse combinée à un biais significatif. Les réseaux convolutifs, permettant de filtrer progressivement les corrélations entre centrales de même source puis entre sources différentes, paraissent être adaptés au problème de prévision agrégée multisource. L'architecture de CNN proposée, où les variables explicatives sont rangées par sources d'énergie contigues, améliorent notablement le RMSE et le CRPS, score global de prévision probabiliste.

Considérant que les gestionnaires de réseau peuvent requérir une fiabilité supérieure aux 99% atteints au maximum dans le Chapitre 2, des modèles de prévision spécifiques aux quantiles bas sont proposés en Chapitre 3. Le modèle QRF servant de référence est assez performant dans le contexte de production agrégée bornée, mais tend à produire des prévisions trop élevées aux quantiles les plus bas (0.1 à 0.5 %). Les modèles paramétriques par distribution exponentielle donnent des résultats satisfaisants, en particulier dans le cas où les données sont clustérisées par k-means, ce qui permet d'atteindre une meilleure fiabilité qu'en clusterisant par niveaux de prévision médiane. Un modèle paramétrique plus flexible basé sur la théorie des valeurs extrêmes améliore le score quantile pondéré, mais dégrade la fiabilité. Le CNN est adapté aux prévisions extrêmes par l'introduction d'une couche de *min-pooling*, soit de sélection des minimums sur la fenêtre de filtrage convolutif. Ceci donne les meilleurs résultats du panel en termes d'équilibre entre fiabilité sur les quantiles les plus bas et score global de prévision. Enfin des modèles par mélanges de densité sont proposés, avec inférence des paramètres par réseau de neurones. Le cas d'étude montre que le modèle paramétrique utilisant des distributions exponentielles permet d'effectuer une prévision fiable si cette exponentielle est paramétrée sur la base d'un clustering des conditions météorologiques et de production agrégée prévue. Enfin, les réseaux de neurones à mélange de densité présentent également un score global de prévision intéressant, tout comme les modèles des valeurs extrêmes. En conclusion, la prévision des quantiles de 0.1% à 0.9 % peut être effectuée à l'aide d'un modèle paramétrique relativement simple (distributions exponentielles + clusterisation par k-means) ou à l'aide d'un modèle plus complexe, permettant d'atteindre de meilleures performances en fiabilité et finesse (régression quantile par réseau convolutif).

Enfin le Chapitre 4 propose une série de méthodes pour l'optimisation de l'offre de réserve:

- Offre d'un quantile bas constant, basé sur les prévisions extrêmes du Chapitre 3.

- Offre selon un quantile optimal obtenu par prévision de la production et des prix

- Offre selon une optimisation stochastique avec scénarios de production agrégée

- Offre de produits de réserve multiple

La méthodologie d'offre par quantile optimal a été publiée dans IEEE Transaction of Power Systems [172]. Elle propose de baser la décision d'offrir l'énergie et la réserve sur la prévision directe de la production agrégée proposée dans cette thèse, ainsi que sur des prévisions de prix. Des prévisions probabilistes de l'écart de prix entre réserve et énergie permettent de pondérer l'offre de réserve à l'aide d'une copule, qui modélise la dépendance entre écarts de prix sur les marchés et production de la centrale virtuelle. Le cas d'étude présenté démontre une augmentation moyenne du revenu de 5% pour une offre jointe énergie + aFRR, et la dégradation de la CVaR à 1% la plus faible par rapport aux solutions concurrentes, comme l'offre de réserve basée sur un quantile faible et constant.

L'optimisation stochastique avec contraintes probabilistes a été publiée dans Applied Energy [171]. Le résultat principal de l'étude est que les scénarios de production agrégée issus de prévision séparée par source d'énergie reflètent mieux la variabilité de la production agrégée qu'une approche par prévision directe, et que leur intégration dans l'optimisation stochastique donne une vente d'énergie plus modérée et augmente la vente de réserve. L'intérêt d'ajouter des contraintes probabilistes est de pouvoir définir explicitement un seuil maximum de défaut de fourniture de réserve, tout en recherchant la maximisation du revenu.

La dernière méthode proposée concerne l'offre de réserves multiples par relaxation Lagrangienne. Elle constitue une alternative simple à l'optimisation stochastique lorsqu'un opérateur de centrale virtuelle intermittente cherche à valoriser son productible sur plusieurs marchés de réserve. L'augmentation de gain par rapport à la vente d'énergie seule (+10 à +22 %), ainsi que la non-dégradation de la 5%-CVaR du revenu, est vérifiée sur un cas d'étude.

Analyse des contributions

Pour terminer, on définit ci-dessous les éléments apportés par chaque chapitre aux contributions annoncées dans l'Introduction en Section 1.4.

La prévision probabiliste directe de production renouvelable agrégée multi-source est une innovation qui structure le Chapitre 2. Les différents modèles sont adaptés au contexte de la centrale virtuelle renouvelable en appliquant la régression sur des configurations de données spécifiques à la centrale virtuelle (par ex. centrales placées aéalatoirement ou par source d'énergie dans le volume d'entrée des réseaux convolutifs), et en les évaluant sur différentes compositions de source d'énergie dans la centrale virtuelle (PV-Wind-Hydro, PV-Wind...) et sur différents niveaux d'aggrégation (prévision agrégée et prévision par source d'énergie). L'ensemble des modèles développés et adaptés au contexte de la production agrégée multi-source répond ainsi à la **Contribution 2**. Enfin les méthodes de génération de trajectoires basées sur des prévisions probabilistes QRF par sources d'énergie ou sur une prévision directe QRF de la production agrégée sont comparées conformément

à la **Contribution 3**. La comparaison montre que les trajectoires générées par source d'énergie sont plus représentatives de la variabilité observée pour une centrale virtuelle.

Les modèles de prévision des quantiles faibles développés dans le Chapitre 3 correspondent à la **Contribution 4**, à savoir qu'ils constituent un ensemble de modèles adaptés à la prévision des quantiles faibles de production renouvelable agrégée, ayant des niveaux de complexité différents. Les modèles simples comme la distribution exponentielle avec paramétrisation clustérisée sur les régimes de production et un échantillon réduit caractérisant les conditions météorologiques et leur dispersion obtiennent de bons résultats et peuvent être suffisants pour être appliqués dans une stratégie d'offre de réserve. Les modèles nécessitant des développements plus complexes comme le réseau convolutif adapté aux quantiles faibles ont pour leur part probablement plus de potentiels d'amélioration. La régression de réseaux de neurones entraînés sur Skill Score paraît prometteuse. Dans le cas des réseaux à mélange de densité, le mélange de densités Beta semble mieux capter les niveaux de production associés aux quantiles faibles que le mélange Gaussien paramétré sur les puissances installées de la centrale virtuelle. Toutefois il serait utile d'analyser plus en détail la répartitoin de l'apprentissage entre le mélange des densités et la couche sous-jacente du réseau de neurones.

Le Chapitre 4 abonde plusieurs contributions se rapportant à la prise de décision sous incertitude et sa place dans la chaîne de valeur de l'offre de services système:

Contribution 1: Chacune des stratégies d'offre fournit une solution pour la dernière brique de la chaîne de valeur de l'offre de services systèmes. Cette dernière brique consiste à optimiser le revenu ainsi que les risques techniques associés, en prenant en compte de façon réaliste l'ensemble des contraintes techniques et en adaptant les stratégies au contexte de la prévision agrégée:

- la *fréquence de défaut de réserve* est mesurée a posteriori, et sa minimisation est intégrée aux stratégies (1) d'offre de réserve basée sur un quantile constant bas issue d'une prévision adaptée aux extrêmes et (2) d'optimisation sous contraintes probabilistes.

- *les* stratégies sont *adaptées au contexte de la production agrégée* car elles incorporent soit une prévision probabiliste directe de la production agrégée, soit des trajectoires de production agrégée qui peuvent être décomposées par sources d'énergies composant la centrale virtuelle afin de reproduire de façon plus réaliste sa variabilité.

Contribution 5: Les stratégies d'offre de services système intègrent les incertitudes liées à la production variable et aux marchés:

- les *incertitudes liés aux marchés d'énergie et de réserve* sont modélisées pas des modèles de prévision dédiés à chaque quantité de marché influant sur la stratégie d'offre, y compris la probabilité que la réserve offerte par la centrale virtuelle soit activée.

- les *incertitudes de production variable* de la centrale virtuelle sont intégrées en entrée des stratégies d'offre par les modèles probabilisites de prévision de production agrégée.

Contribution 6: Les trajectoires de production agrégée développées et évaluées en Chapitre 2 en termes d'erreur de prévision sont intégrées à la stratégie d'offre basée sur l'optimisation stochas-

tique. Ainsi, *l'impact des trajectoires de production agrégée sur les revenus* obtenus à la fin de la chaîne de valeur est évalué.

Contribution 7: La méthodologie d'offre de services système multiples développée en Section 4.6 constitue une *solution simple pour l'offre conjointe de services système multiples et d'énergie*, car elle ne nécessite pas de recourir à des méthodes d'optimisation stochastique.

Appendix A

Supplementary material for the characterisation of a multi-source VRE VPP production

A.1 Cross-source correlations within a multi-source VRE VPP

The correlations observed between single power plants and higher aggregation levels are represented as a function of the concerned energy sources in A.1. Cross-source correlations are low and centered around zero (see points associated to "PV-Wind" correlations in Figure A.1), but can not be discarded. The correlation of wind plants with aggregated productions of the VPP is significant (see "Wind-Wind+PV" correlations in Figure A.1), because aggregations at national and European levels are dominated by wind capacities. Interestingly, some negative correlations occur, for instance between Wind plants and higher aggregation levels (see points associated to "Wind-Wind" and "Wind-Wind+PV" correlations in Figure A.1). Negative correlations illustrate the interest to join different climates into the VPP: surges in production may compensate drops of production in distant regions.
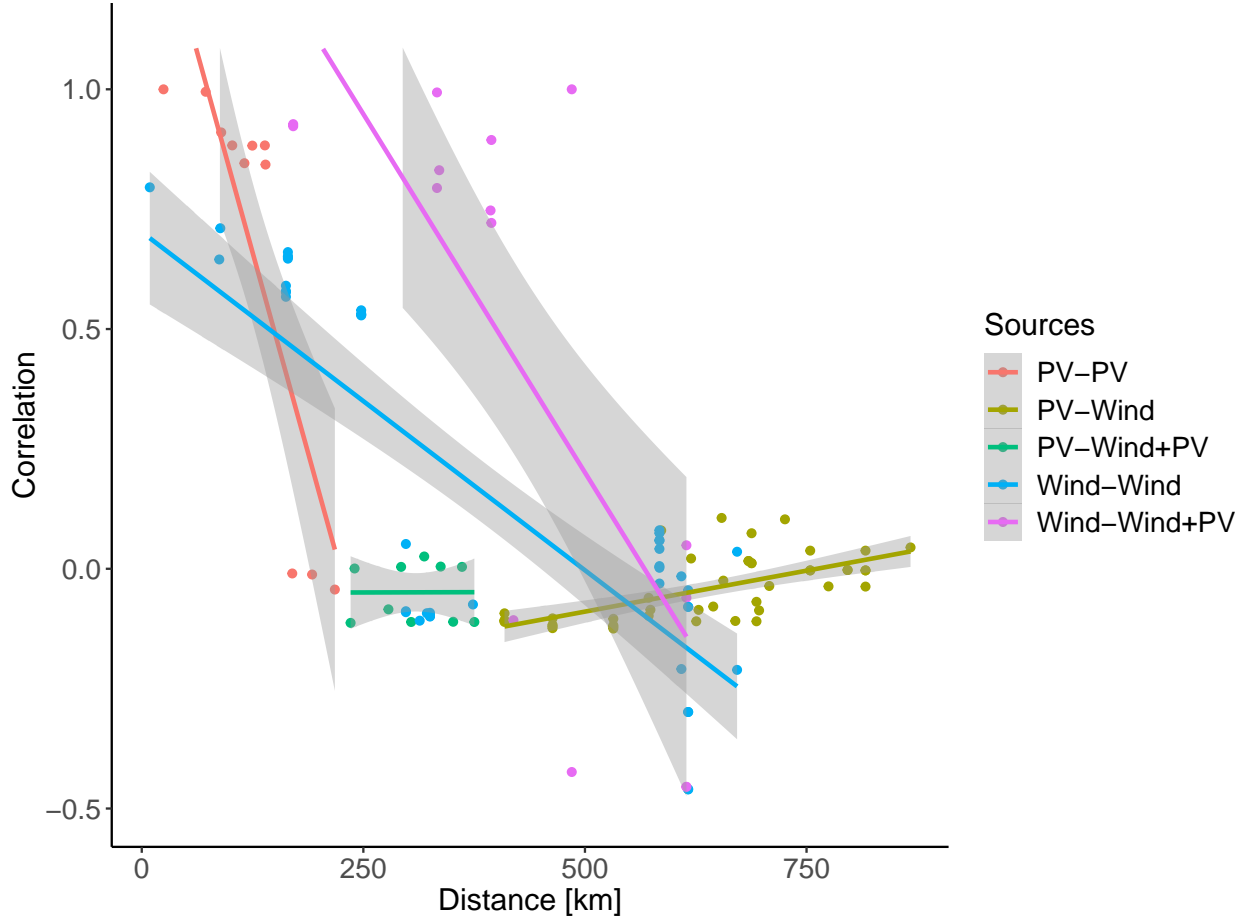
Figure A.1: Correlations between plants and higher aggregation levels in the VPP, by energy sources

A.2 Mutual information between production at various aggregation levels

Linear correlations do not capture the full range of dependence between uncertain productions. The relation between distributions of production levels is better quantified by the Mutual Information (MI). The MI, denoted $I(x; y)$, measures the distance between the distribution of two random variables of interest x and y, for instance two production levels. It is expressed in (A.1) as a Kullback-Leibler divergence, giving the amount of information lost when the two variables are assumed independent with distribution $f_x.f_y$ instead of considering their joint distribution $f_{x,y}$. The MI is positive: if it is equal to zero, the two random variables can be considered independent. For practical use, the distributions can be discretized over the probability spaces of both variables and estimated via kernel density estimators.

$$I(x; y) = D_{KL}(f_{x,y}||f_x f y) = \int_{x,y} f_{x,y} log \frac{f_{x,y}}{f_x f y} dx dy \tag{A.1}$$

The MI between production of individual plants and aggregated production on the VPP is presented in Figure A.2. After evaluating MI on all possible pairs of the aggregation, attention is
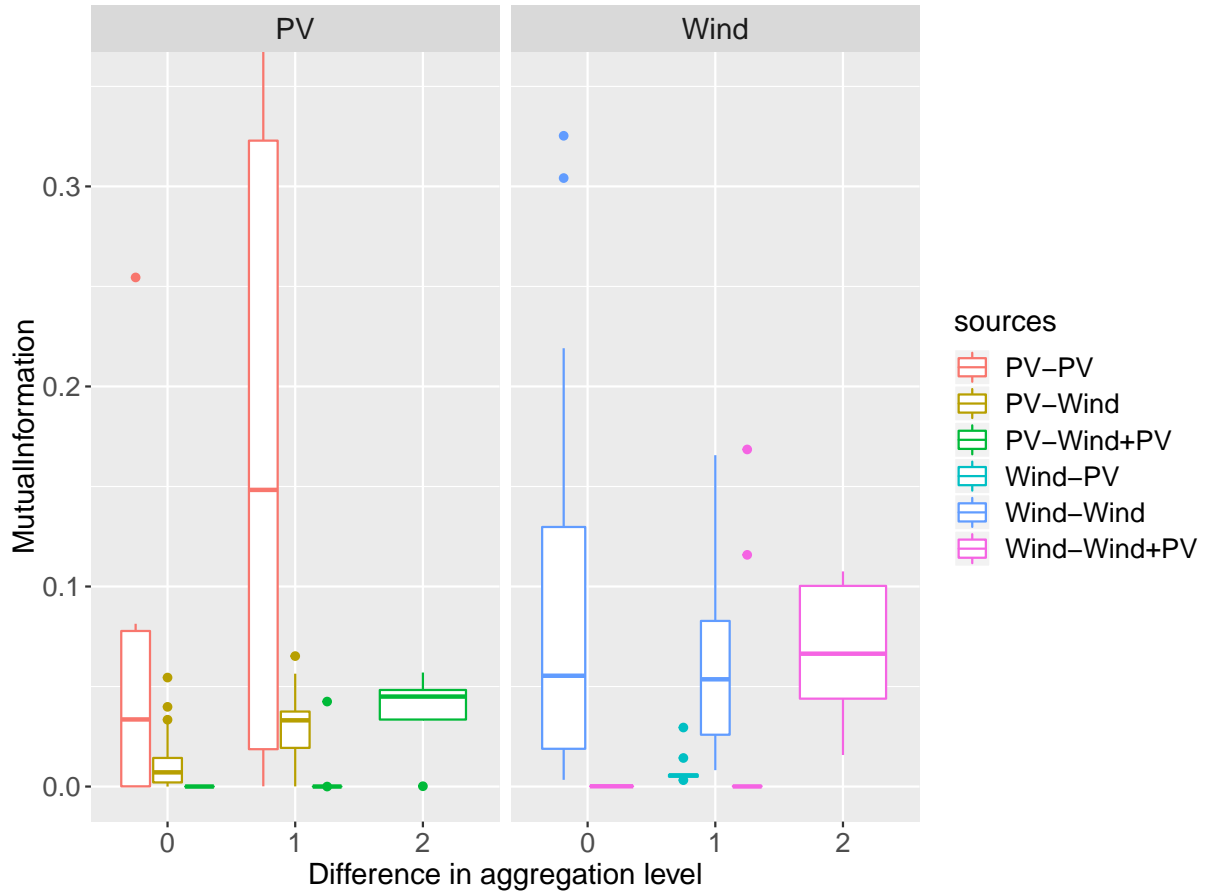
Figure A.2: Mutual information of plants and sub-aggregations with higher levels of aggregated production. The x-axis describes the difference in aggregation level: 1 is for a single plant informing on a BRP aggregation or a BRP aggregation on the total VPP production; 2 is for a single plant informing on the total VPP production.

placed on the influence of production at sub-levels (plant or BRP aggregation) on higher aggregation levels, up to the VPP production. The MI of PV and Wind single power plants with the total VPP production is low, between 0.06 and 0.10 (see boxplots for difference in aggregation level = 2). It is however of a comparable range with the MI between single power plants, PV or Wind. Lastly the MI between different energy sources (see boxplots of "PV-Wind", "Wind-PV") is small but superior to 0. This is expected as both production processes are distinct. In conclusion, all plants of an aggregation have a significant contribution to the total VPP production.

A.3 Analysis of temporal variability

Sparse representations are useful in characterizing the temporal variability. In the next subsection a Fourier spectrum analysis produces a sparse representation of the aggregated production. The Fourier spectrum assumes that the production signal is stationary. This assumption can be valid when one investigates the constitution of power plant pooling, but is of limited interest when a

decision must be taken over a limited time period (e.g. reserve bid on one day). A transient analysis will then be performed in the last subsection by applying a wavelet transform, which characterizes the temporal locations of power variations.

The Power Spectrum Density PSD_y of a production signal y is the limit at period T of the average signal density. This average can be estimated by a periodogram, an auto-regressive model or a taper as in (A.2), applying a tapering sequence h_t to each each observation $y(t)$, in order to reduce noise.

$$\hat{y}^{FT}(\omega) = \int_{-\infty}^{+\infty} y(t)e^{-i\omega t}dt$$

$$\hat{y}^{FT,trunc}(\omega) = \frac{1}{\sqrt{T}} \int_0^T h_t(y(t))e^{-i\omega t}dt \tag{A.2}$$

$$PSD_y(\omega) = lim_{T\infty}\mathbb{E}(|\hat{y}^{FT,trunc}(\omega)^2|) \tag{A.3}$$

The PSD of the Wind-PV VPP (dominated by Wind in terms of capacity), is presented in Figure A.3. The harmonics associated to the solar curve are dampened in the spectrum of the aggregated production in dark blue), compared to the spectrum of an individual PV plant (in yellow), due to the reduced PV capacity in the VPP, but it is still noticeable. The aggregated production exhibits a smoother spectrum than individual wind and PV plants at frequencies superior to $\frac{1}{24h}$, even if the temporal variability remains significant.
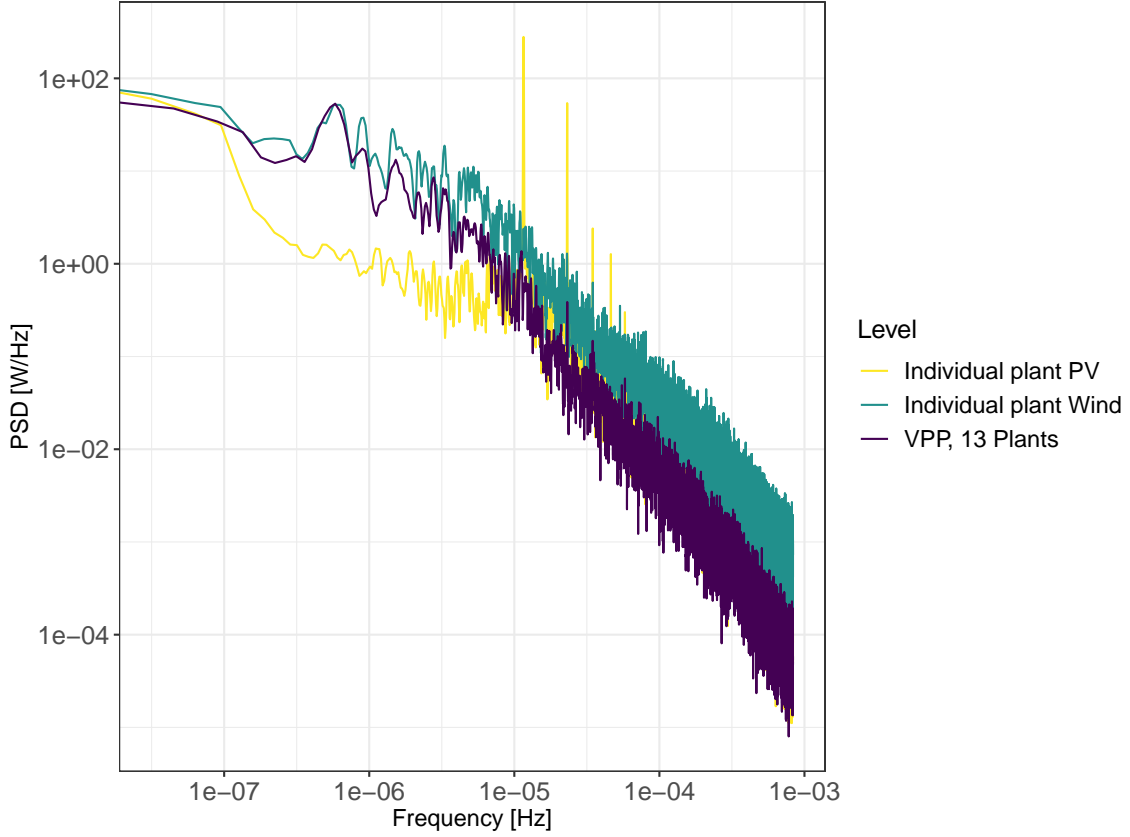
Figure A.3: Spectrum analysis of production of single PV plant, Wind plant and a VPP comprising
13 plants (10% PV capacity)

A.4   Localizing variability events through Wavelet Transforms

The spectrum developed above informs on how frequently a variation amplitude occurs, but is
unable to localize it in time. To precisely locate when production variations occur, the aggregated
production signal can be processed with a Wavelet Transform. The Wavelet Transform decomposes
the variation of production following high and low frequency components, enabling to identify trends
in the low frequencies and discard noise in the high frequencies [175]. The core operation of the
transform is a convolution of the signal $y(t)$ at scale s and shift k with the wavelet function for the
given scale, writing:

$$w_t(s, k) = 2^{-s/2} . \sum_{t=0}^{T-1} y_t . \psi((t - k.2^s)/2^s) \tag{A.4}$$

with ψ being the mother wavelet. We implement the Morlet wavelet described below, because it
preserves shape in case of frequency shifts in the original signal [176], so that separation in frequency
bands can be done efficiently.

$$\psi(t) = \pi^{-1/4} . e^{i\omega t} . e^{-t^2/2} \tag{A.5}$$

The transform applied on the previous Wind-PV VPP in Figure A.4 detects clearly periods of idle
production, and informs on their duration: the transform (middle panel in Figure A.4 has low

power levels in the first two weeks of September, over the whole range of transform scales (y-axis of the image). The wavelet transform of a single wind power plant in the aggregation is marked by a similar idle period, more pronounced, but also by more frequent ramps lasting several hours (see bottom panel, periods inferior to 0.5 days), with higher intensities and different locations in time. Hence a forecasting model the temporal variability in aggregated production should be modelled via specific methods, which distinguish from the case of single power plants.

In conclusion, the aggregated multi-source VRE production exhibits a smoother variance than individual plants, due to its size and the diversity in production profiles from various sources. The higher auto-correlation is due to the diverse locations of plants, which experience weather conditions with delays, but can also stem from the deterministic trend of PV (course of sun) when the PV capacity share is high. In VPPs offering balancing AS, which are usually aggregations of modest size (tens of plants), the information carried by a single plant is valuable to characterize the aggregated production and should not be discarded. This will be implemented in the next section, where models will base their forecasts on features from all plants in the aggregation.
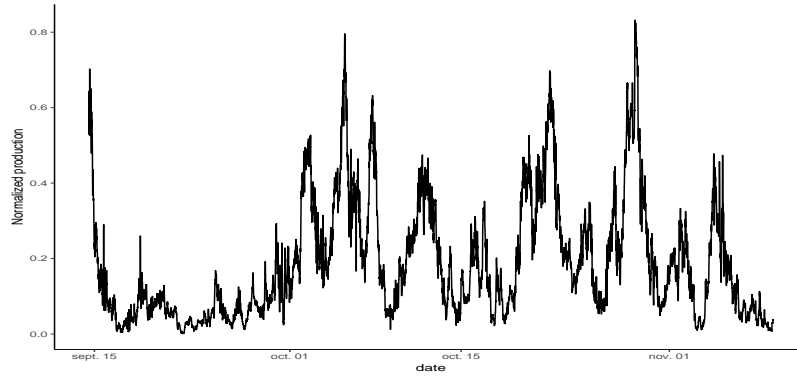
(a) Aggregated Production
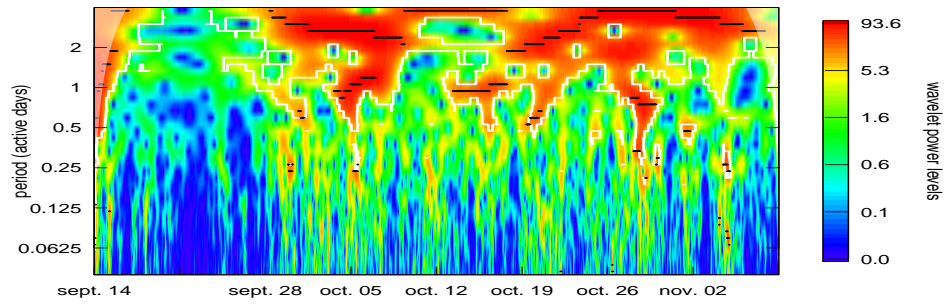
(b) Wavelet transform of aggregated production
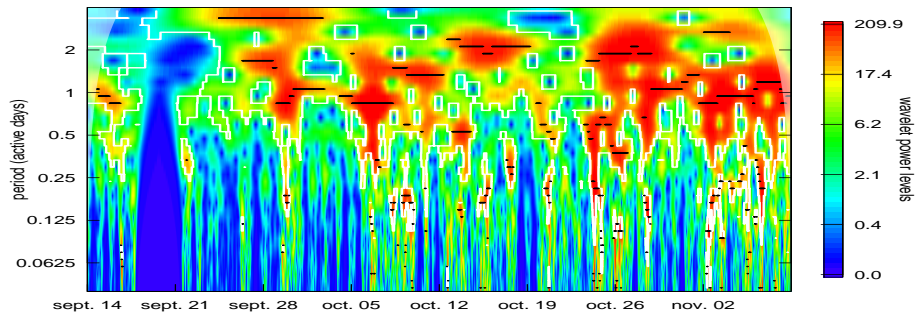
(c) Wavelet transform of a single wind plant in aggregation

Figure A.4: Wavelet transform on a Wind-PV VPP production, 2 months data (09-2015/10-2015)
at 10 minutes resolution

Appendix B

Forecast of reserve capacity from a renewable VPP using a bivariate Kernel Density Estimator

A simple forecasting model is proposed here to forecast the reserve capacity of a renewable VPP. It relies on a bivariate Kernel Density Estimator (KDE), able to issue a probabilistic forecast of the production of a VPP consisting of two energy sources, such as Wind and PV. The production forecast is done at two levels, at the plant level and at the aggregated production level. Both forecasts are translated into reserve capacity forecasts, by retaining the minimum value of the production forecast over a reserve validity period.

The forecasting performance of the KDE is evaluated on a case study comprising a VPP aggregating Wind and PV plants in France. The quality of reserve forecasts is evaluated by counting the frequency of reserve under-fulfilments.

The following pages present the details of this work, presented at the *IEEE ISGT Europe 2017* Conference in Torino, Italy [168].

Short-term Forecast of Automatic Frequency Restoration Reserve from a Renewable Energy Based Virtual Power Plant

Simon Camal, Andrea Michiorri
Georges Kariniotakis
MINES ParisTech, PSL – Research University
Center PERSEE
Sophia-Antipolis, France
simon.camal@mines-paristech.fr

Andreas Liebelt

FRAUNHOFER Institute for Wind Energy
and Energy System Technology IWES
Kassel, Germany

Abstract— **This paper presents the initial findings on a new forecast approach for ancillary services delivered by aggregated renewable power plants. The increasing penetration of distributed variable generators challenges grid reliability. Wind and photovoltaic power plants are technically able to provide ancillary services, but their stochastic behavior currently impedes their integration into reserve mechanisms. A methodology is developed to forecast the flexibility that a wind-photovoltaic aggregate can provide. A bivariate Kernel Density Estimator forecasts the probability to provide reserve. The methodology is tested on a case study where volumes of automatic Frequency Restoration Reserve (aFRR) are forecasted on a day-ahead horizon. It is found that the wind-photovoltaic aggregate can dedicate a limited share of its forecast production to aFRR. The frequency of insufficient reserve capacity is assessed, by comparing the capacities offered with the measured production.**

Index Terms-- **Aggregation, Ancillary Services, Forecasting, Photovoltaics, Wind power.**

I. INTRODUCTION

Variable renewable power plants substitute conventional synchronous generators at a fast growing rate. The increased intermittency among available generation impacts significantly the stability of power systems. Due to the spatiotemporal uncertainties associated with their production, variable renewable generators are currently restrained by operators in their provision of Ancillary Services (AS), for which maximum reliability is a firm pre-requisite. However studies have identified that wind and photovoltaics (PV) power plants show technical capabilities to provide AS [1],[2]. The Irish TSO has issued a specific regulation for frequency control from wind power plants [3]. This paper investigates the capacity of renewables to offer frequency control services.

Frequency control services follow a multi-level sequence. At a first control level (activation time 0-5 s), the inertia of generators contains instantaneous frequency perturbations.

Wind and solar plants can emulate synthetic inertia [1]. The next level is Frequency Control Reserve (FCR), where generators connected to a synchronous area regulate their power output in function of the frequency deviations they capture. Frequency Restoration Reserve (FRR) follows FCR. It is activated both automatically (aFRR, full activation time 5-15 min) and manually (mFRR, full activation time 13-15 min). TSOs activate aFRR by first evaluating centrally the Area Control Error, then calling for modification of the active-power setpoint of generators. In France, the TSO updates the aFRR sizing at a 30-min timestep [4]. The amount of aFRR is expected to vary in the coming years: the International Grid Control Cooperation, triggered by European TSOs, increases its exchanges in order to lower the overall need for aFRR [5]. In contrast, the stochastic behavior of renewables is expected to lead to higher aFRR levels. Improvement in wind and solar power forecasting could mitigate their impact on the sizing of this reserve [6]. The last level of response to frequency deviations is Replacement Reserve (RR) which is in place in several European countries. RR is manually activated in case the restoration reserve is not sufficient to ensure stability, within a time frame of 15 to 30 min.

Procurement schemes and markets for AS are diverse among European countries. Some services may be mandatory in some countries (e.g. FCR, [7]), tendered following economic merit-order (RR in France, [4]), or traded on a market with different upward and downward prices (RR in Portugal, [8]). For an energy trader, participation in reserve markets is economically interesting if the reserve price is superior to the average price for energy. While current prices in Europe tend to incentivize more energy than reserve [6], the AS markets are profoundly evolving and promising for renewables as their marginal cost is close to zero. Bidding strategies for participation of wind farms in an AS market have been studied recently. The reserve strategies proposed in [9] keep a share of the active power forecast to the FCR market. The optimal bid of wind power is analytically derived as a quantile of the production forecast for the day ahead,

considering high penalties in the balancing market if failing to provide the FCR service. Moderate increase in revenue is found (<12%) compared to participation in energy market only.

The European project REstable, which motivates this paper, aims at demonstrating renewable-based ancillary services through better interaction of European control zones [10]. The central idea of the project is that the aggregation of distributed power plants with distinct features (geography, time, resource, market regulations) can offer reliable AS via an adaptive European Virtual Power Plant (VPP).

Reliable offers of AS suppose adequate power forecasting methods. For variable generation, assuming that the distribution of forecast errors depends only on historic performance does not capture the uncertainty inherent to the forecast model [11]. The nonlinearity of wind and solar generation induces that conditional distribution of forecast errors is easier to model with nonparametric approaches [12]. According to a review of probabilistic methods for reserve requirements [11], density forecasts can be applied to both wind and solar power, and give more reliability on reserve allocation problems than approaches based on historical forecast only. Kernel Density Estimation (KDE) is a density forecast method which figures among the top-ranked methods for wind and PV forecasting [13]. Ensemble forecasts represent an alternative to density forecasts. They can incorporate temporal and spatial interdependence of prediction errors, and perform well on short-term horizons [14]. A density approach has been chosen here to model the problem of aFRR capacity forecasting, formulated in Section II. The forecast method is described in Section III.

II. PROBLEM FORMULATION

The objective of this work is to forecast a reliable day-ahead offer of aFRR, provided by a VPP aggregating photovoltaic and wind power plants located in France. This geographical limitation is due to the availability of data for tests. It is deemed acceptable to study a VPP concerned by the French control area, because most of the aFRR need in France is covered currently by plants located within this area [15].

Producers who supply aFRR must comply with strict regulations defined by TSOs. For instance, German TSOs ask that deployed reserve capacities are never lower than the contracted volume over the whole product length [16]. In France penalties apply if the measured deployed capacity is more than 10% lower than the contracted capacity, over an evaluation period > 100 h [4].

This problem poses two main challenges:

1. Propose a reliable production forecast over the product length, so that the risk of failing to provide reserve due to overestimation of production is minimal.

2. From this forecast, derive volumes of reserve that are significant (superior or equal to 1% of aggregated capacity) during intervals of sufficient production.

The temporal resolution of the forecast must be at least equal to the temporal resolution of the AS product, in order to qualify the aggregate as a potential AS provider. In the case of aFRR it is 15 minutes in Germany and 30 minutes in France. The production forecast has generally a coarser temporal resolution than the grid signals that the service must react to (e.g. Automatic Generation Control (AGC) signal). It also does not capture the very-short term variations in the power output of the aggregate. The forecasting error is therefore dependent on the temporal resolution.

It is assumed in the present study that plants can effectively communicate to a distant VPP control center and regulate their power output following setpoints sent by the center without significant discrepancies. Experiments conducted within the project Kombikraftwerk 2 have shown that renewable plants can be controlled within a 3-second time lapse. Power regulation of aggregated plants showed also some limits: regulation was unsuccessful on wind turbines operating close to cut-in wind speed, and deviated from the emulated AGC signal [17], similarly to another experiment [16]. In the next section, a methodology is presented to solve the afore-mentioned problem. Three approaches are proposed with increasing level of complexity: a basic approach using individual probabilistic production forecasts as inputs, a deterministic aggregated forecast issued by a bivariate KDE, and a probabilistic forecast computed by the same KDE model.

III. METHODOLOGY

A. Basic Approach for Aggregated Flexibility Forecast

1) Probabilistic production forecasts at plant level

Probabilistic production forecasts are issued for each plant of the aggregate. The forecasts are based on a KDE k-Nearest Neighbors (k-NN) model for PV plants, and bivariate conditional KDE on wind speed and wind direction for Wind plants. A description of the k-NN algorithm is reported in Subsection B.3. Both forecasting models have been validated using deterministic and probabilistic criteria [18], [19]. The forecasts are issued at a runtime t for a horizon interval Δh.

2) Individual flexibility forecasts at plant level

For each plant of the aggregate, a share of the active power forecast is dedicated to reserve. This target share is considered here as a quantile of the probabilistic forecast, at nominal value α. The choice of the nominal value α can be realized through an optimization on expected gains and losses associated with energy and reserve, similarly to optimal bidding strategies used by renewable producers [20]. The risk of failing to provide reserve decreases with α. The total reserve volume is then chosen as the minimum value of the quantile on the horizon interval Δh of the day to predict. The minimum is chosen in (1) to minimize the risk of failure. The offer of symmetrical reserve $\hat{r}_{i,t}$ by the plant i at runtime t, is equal to half of the total reserve volume:

$$\hat{r}_{i,t} = \frac{1}{2} \min_{\forall h \in \Delta h} F^{-1}_{i,t+h|t}(\alpha) \qquad (1)$$

where $F^{-1}_{i,t+h|t}$ is the inverse Cumulative Distribution Function (CDF) of the power forecast of plant i at horizon h.

3) Aggregation of Individual Flexibility Forecasts

The aggregated day-ahead offer $\hat{r}_t$ issued at runtime t equals in (2) to the sum of the individual forecasts from the plants of the aggregate.

$$\hat{r}_t = \sum_{i=1}^{N_{plants}} \hat{r}_{i,t} \qquad (2)$$

B. Aggregated KDE Probabilistic Forecast

In this section, a bivariate KDE model is proposed to forecast the power production of an aggregate of PV plants and Wind plants. Weather conditions are reduced to a joint distribution of solar radiation and wind speed, therefore this model applies to aggregates in which plants of same technology experience similar weather conditions. Future work could include more diversity in weather forecasts via reduction techniques such as Principal Component Analysis (PCA), or ensembles using weather forecast at multiple sites.

1) Retrieve weather forecasts and select reference data

In this work, a central reference site is derived for all plants of similar technology. The coordinates of the site minimize Euclidean distance among plants. Numerical Weather Prediction forecasts (NWP) are retrieved for the reference sites. For photovoltaic plants, a unique NWP, Solar Surface Radiation Downwards, is selected. For Wind plants, meridional and zonal wind components are selected as they appear in the literature to be the principal influential variables [19], [21], [22].

2) Select bi-variate conditional explanatory variables

The bivariate explanatory variable X, based on NWP forecasts, is constructed in (3) following a simple regime-switching approach. The contribution of solar radiation to the bi-variate condition at the current timestamp is estimated by $w_{I,0}$, the proportion of PV power installed in the aggregate. This static contribution is associated to the contribution $\hat{w}_I$ of the solar radiation forecast $\hat{I}$, which is compared to its expected value over the available learning data set $E(\hat{I})$:

$$w_I = w_{I,0}.\hat{w}_I, \quad w_{I,0} = \frac{P_{PV,peak}}{P_{aggreg,peak}}, \quad \hat{w}_I = \frac{\hat{I}}{E(\hat{I})}$$
$$w_I > w_{I,0} \Rightarrow X = (\hat{I}, \hat{W}), \quad w_I \le w_{I,0} \Rightarrow X = (\hat{U}, \hat{V}) \qquad (3)$$

where $\hat{W}$ is the wind speed, $\hat{U}$ and $\hat{V}$ the meridional and zonal wind components respectively.

3) Training data selection by k-Nearest Neighbours

The estimator is trained on a sample of similar bivariate points of NWP forecast, following a k-NN approach. The k-NN algorithm has been applied to probabilistic wind power forecasting [21], [22], and has the advantage of maintaining constant the size of the training data as new forecasts are added [12]. Nearest neighbors minimize in (4) the Manhattan Distance between the available historical learning set X and the forecast condition x. Distances are weighted by the contributions of PV and Wind in the case where solar contribution is significant. The weights for meridional and

zonal wind components w_U and w_V, are obtained via an optimization on the sum of square errors from the deterministic output of the KDE, presented in the next section. The optimization is realized with a Particle Swarm Optimization algorithm [23], following a cross-validation on weekdays as presented by [21].

$$D(x,X) = \begin{cases} |w_I.(x_I - X_I) + w_W.(x_W - X_W)|, w_I > w_{I,0} \\ |w_U.(x_U - X_U) + w_V.(x_V - X_V)|, w_I \le w_{I,0} \end{cases} \qquad (4)$$

The training set is populated in (5) with nearest neighbors until the set size k_T has reached a ratio r_{kNN} of the size of the learning set L. An upper bound of 125 is applied to contain computational burden. Such a value appears sufficient in the context of k-NN wind power forecasting [21].

$$k_T = r_{kNN}.Card\{(X_k,Y_k),k \in L\}, \quad k_T \le 125 \qquad (5)$$

4) Conditional Kernel Density Estimator

The obtained training sample of conditional explanatory variable is compared to the forecast condition x within a bi-variate KDE, with the Kernel proposed in [19]. A bandwidth matrix H is computed following the Smooth Cross Validation (SCV) method to model directions of correlation between covariates [24]. In conditions when wind is dominant, a diagonal SCV matrix is selected. The explanatory weight $\hat{w}_X$ in (6) associates the weather forecast condition x with its position relative to the learning set of size N, and filters older points by an exponential forgetting factor fgt of constant parameter λ.

$$\hat{w}_X(x,X) = fgt(\lambda).\frac{K(H^{-1}(x-X))}{\sum_{j=1}^{N} K(H^{-1}(x-X_j))} \qquad (6)$$

The conditional density forecast of aggregated production $\hat{f}_{Y_{t+h}|X_t}$ is obtained in (7) via a KDE with a bandwidth h_Y equal to the k-nearest neighbour, using the same ratio r_{kNN}. An Epanechnikov Kernel and a reflection method model the bounded behavior of the production.

$$\hat{f}_{Y_{t+h}|X_t}(y,x) = \frac{1}{h_Y}.\sum_{i=1}^{N} \hat{w}_X(x,X_i) K_Y\left(\frac{y-y_i}{h_Y}\right) \qquad (7)$$

5) Selection of minimal quantile of aggregated forecast

The forecast flexibility offer $\hat{r}_t$ is chosen in (8) as the minimum of the aggregate production for the α-quantile, on the horizon interval Δh of the day to predict.

$$\hat{r}_t = \frac{1}{2} \min_{\forall h \in \Delta h} F_{t+h|t}^{-1}(\alpha) \qquad (8)$$

C. Aggregated Deterministic Forecast

The third approach uses the KDE model presented in Section B to derive an aggregated deterministic forecast. Predictive densities produced by the model are optimized with respect to the expectancy E, so the mean of the predictive density for each horizon constitutes the

deterministic forecast. The forecast flexibility for the day-ahead horizon is chosen in (9) as the minimum of the deterministic aggregated forecast on the day to predict.

$$\hat{r}_t = \frac{1}{2} \min_{\forall h \in \Delta h} E\left(\hat{f}_{Y_{t+h}|X_t}\right) \qquad (9)$$

IV. CASE STUDY

A. Workflow

The methodology is evaluated on the following case study. A VPP is constituted by 2 wind farms located in North-East France and 10 photovoltaic plants located in a region of West France. The wind farms account for 95% of the aggregated power capacity. The VPP offers symmetrical aFRR on a day-ahead horizon. Offers must be communicated to the TSO prior to 13:00 UTC, similarly to the existing rules in France [4]. Offered volumes of reserve are computed given the three approaches described in the methodology, for each day of the test period (December-January). The nominal value α of the production quantile forecast is set to 10%, for the individual plants and the aggregate. Note that only the amount of power proposed in the offer is calculated in this paper, not the price.

B. Input Data

Measured production from all plants is available at a 5-min resolution. Forecasts of solar radiation, meridional and zonal wind speeds at 100-m hub height are retrieved at runtime 12:00 UTC from ECMWF HRES, for horizons +12 h to +36 h.

C. Performance Indicators

The KDE aggregated forecast is evaluated using the following criteria: Root Mean Square Error (RMSE) for its deterministic version, reliability and Continuous Ranking Probability Score (CRPS) for the probabilistic version. Reliability informs on the probabilistic bias of the forecast while the CRPS produces an overall probabilistic forecast score [20]. We normalize here errors by the sum of the installed capacities in the VPP.

The risk of failing to provide offered reserve levels is evaluated with a simple criterion. If the measured production is lower than the offered reserve levels at a given timestamp, then this event is counted as an *under-fulfillment*. The *Rate of Under-Fulfillments* (RUF) is the frequency of *under-fulfillments* on the evaluation interval. The RUF is chosen as a criterion to assess the reliability of the ancillary service forecast. It is inspired by current practice of TSOs, eg. in France where aFRR deployment is judged insufficient if the reserve deployed is below the day-ahead offered capacity – 10%, during more than 10% of the evaluation interval [4]. The accuracy of this criterion is limited by the temporal resolution of the measured production. The amount of flexibility offered is another performance indicator. It is expressed by the Cumulated Distribution Function (CDF) of the volumes on the evaluation interval.

V. RESULTS

A. KDE Aggregated Production Forecast

The training period of the KDE forecast is July–November and the test period is December – February. The forecast model issues predictions at a 30-min resolution over a 12-h to 36-h horizon. The deterministic output of the KDE aggregated forecast shows a RMSE comprised between 10% and 20% over the forecast horizon, with significant improvement compared to persistence (Fig. 1). These levels are coherent with state-of-the art forecasting for wind and PV at these horizons [18], [21]. The CRPS is comprised between 7% and 11% over the forecast horizon, with similar improvement compared to climatology.

The reliability of the probabilistic KDE forecast is evaluated by a calibration diagram (Fig. 2). The model shows fair reliability, in particular for the 10%-quantile, which is of particular interest for the present application. The calibration analysis shows room for improvement of the model, and could be extended to larger intervals.
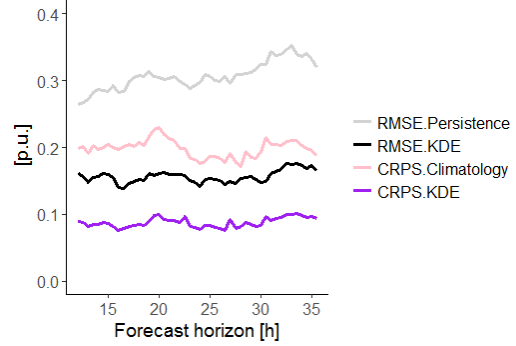
Fig. 1: RMSE of deterministic KDE forecast and Persistence model. CRPS of probabilistic KDE forecast and Climatology model.
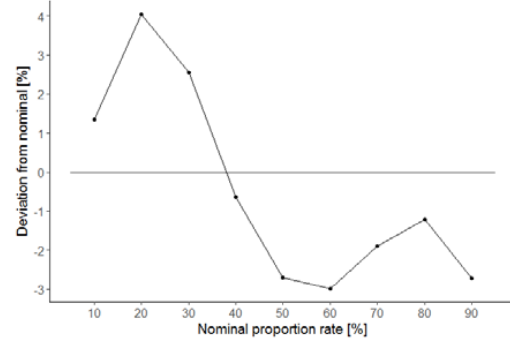
Fig. 2: Calibration diagram (over all horizons)

B. aFRR Offer Forecast

The offers of aFRR flexibility obtained with the three approaches are presented on an interval of high wind production in Fig. 3. The basic approach based on individual flexibility forecasts, in red on Fig. 3, offers volumes mostly inferior to 1% p.u., except when the forecasted production is constantly high during the whole day. The approach based on aggregated deterministic forecasts, in brown, leads to higher

flexibility levels, greater than zero for most of the days tested. Finally, the approach based on the aggregated probabilistic forecast, in orange, differs slightly from the individual forecast approach because it models the correlation between plants: its average flexibility is higher, whereas the dispersion of levels around the average is lower than for the individual forecast approach. Note that in the present day-ahead framework, the individual flexibility offered by PV plants is zero (no production at night).

The CDFs of aFRR offers for the three approaches show that the potential offer levels span from 0 to 0.30 p.u. on the evaluation interval (Fig. 4). The median offers of the probabilistic approaches are 0.015 p.u. and 0.010 p.u., for the aggregated approach and the individual approach respectively. The cumulated Rate of Under-Fulfillment (RUF) compared to 5-min monitored production reaches 26% for the deterministic approach (Fig. 5). The cumulated RUF for the aggregated forecast approach is 7%, close to the 5% for the individual forecast approach. As a first conclusion, the probabilistic approaches appear to be better candidates than the deterministic approach for a reliable aFRR capacity forecast.

VI.   Discussion and conclusion

A bivariate KDE has been developed to predict an aggregated production of wind and PV plants, with encouraging results and insights for future improvements, especially on calibration. Optimization of the model for conditions with significant solar radiation could improve reliability. Other models such as Quantile Regression Forests, gradient boosting and copulas will be tested to forecast the aggregated production. Improvement of the aggregated production forecast, as well as shorter product length, would lower the risk of failing to provide reserve. The aggregated flexibility forecast approach helps formulate offers for product lengths superior or equal to 24h: the individual forecast approach will lead to zero levels for PV plants (zero production at night). With high-resolution measurements of production data and weather conditions on site, the response of a VPP to frequency deviations or AGC signals can be modeled, and the probability of failure to deploy reserve can be assessed.

The rates of under-fulfillment observed in the case study could be assessed on a larger framework: higher plant diversity in the aggregate (higher PV penetration, several climates per technology), combined offers of aFRR with FCR. This would help calibrate the qualification tests a renewable aggregate must pass to provide AS. Future work may include conditional analysis on the technical capacity of an aggregate to deliver AS. Finally the Rate of Under-Fulfillment can be optimized given prices for energy and reserve.

A methodology to forecast the provision of aFRR by aggregates of wind and photovoltaic plants has been presented. Reserve levels are obtained through a simple minimal quantile selection on the production forecast. It is found that renewable power plants can offer volumes of automatic Frequency Restoration Reserve on a day-ahead mechanism, if their aggregate production is forecasted accurately. A trade-off is noticed between the volume of flexibility offered and the expected RUF. Potential aFRR capacities are identified on a case study, yielding volumes from 0 to 0.30 p.u. with medians of 0.01 to 0.13 p.u (Fig. 4).
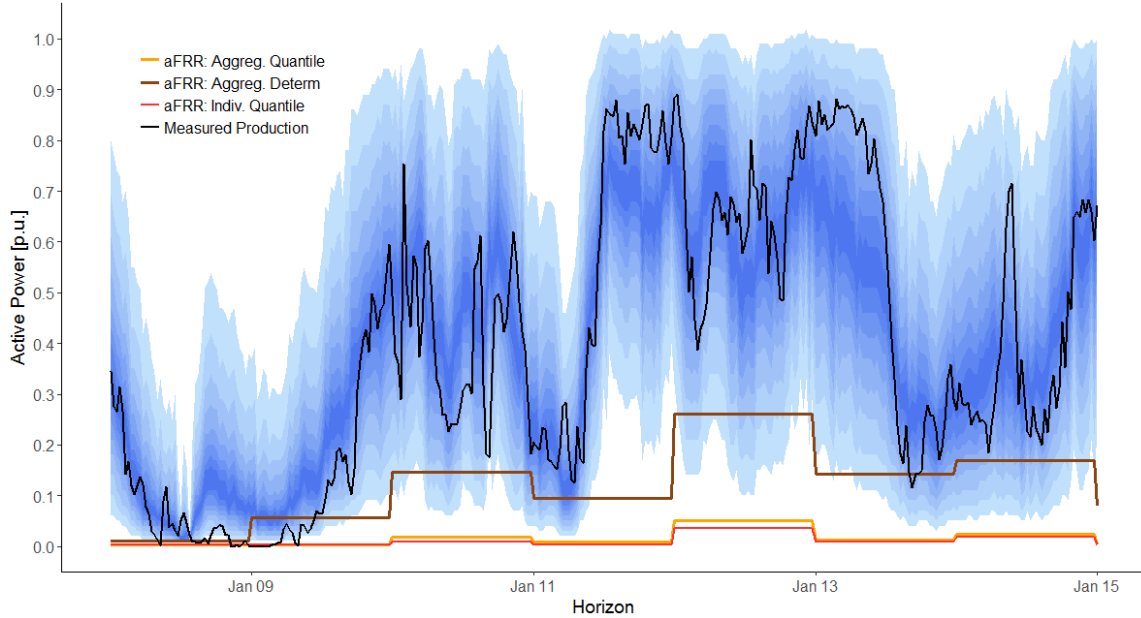

Fig. 3: aFRR capacity forecasts compared to production measurement and forecast (predictive intervals in blue scale, from 10% to 90%)
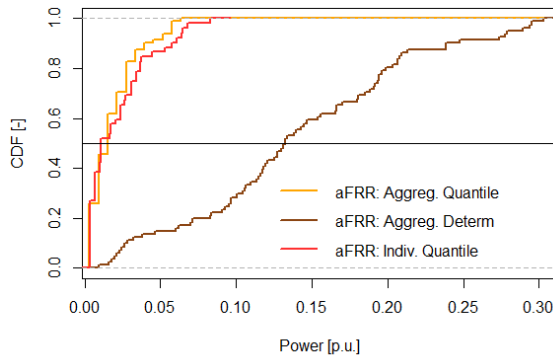
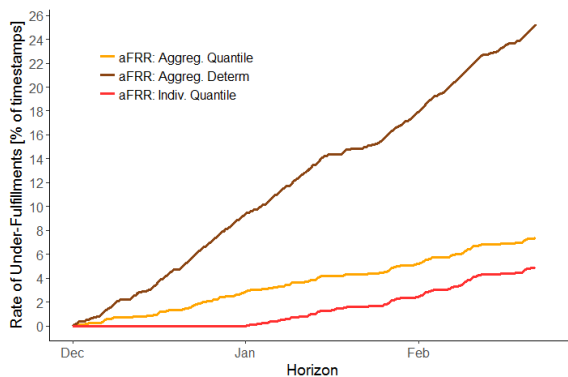Fig. 4: CDFs of aFRR offers for the three approaches

Fig. 5: Rate of Under-Fulfilment of aFRR for the three approaches

ACKNOWLEDGMENT

REFERENCES

[1] M. Faiella, T. Hennig, N. A. Cutululis, and F. Van Hulle, "Capabilities and costs for ancillary services provision by wind power plants," ReServices Project Deliverable, 2013. [Online]. Available: http://orbit.dtu.dk/fedora/objects/orbit:127131/datastreams/file_ac1a56 c2-b722-439b-bcfb-fb6790c1f4fc/content

[2] B. I. Craciun, T. Kerekes, D. Sera, and R. Teodorescu, "Frequency Support Functions in Large PV Power Plants With Active Power Reserves," IEEE J. Emerg. Sel. Top. Power Electron., vol. 2, no. 4, pp. 849–858, décembre 2014.

[3] Eirgrid, Eirgrid Grid Code. 2015.

[4] RTE, Règles Services Système Fréquence. 2017. [Online]. Available: https://clients.rte-france.com/htm/fr/offre/telecharge/20170101_Regles_services_system e_frequence.pdf

[5] IGCC Expert Group, "Stakeholder document for the principles of IGCC," Sep. 2016. [Online]. Available: https://www.entsoe.eu/Documents/Network%20codes%20documents/I mplementation/IGCC/20161020_IGCC_Stakeholder_document.pdf

[6] L. Hirth and I. Ziegenhagen, "Balancing power and variable renewables: Three links," Renew. Sustain. Energy Rev., vol. 50, pp. 1035–1051, Oct. 2015.

[7] ENTSO-E, "Survey on Ancillary services procurement, Balancing market design 2015," May 2016. [Online]. Available: https://www.entsoe.eu/Documents/Publications/Market%20Committee %20publications/WGAS%20Survey_04.05.2016_final_publication_v2 .pdf?Web=1

[8] ERSE, Manual de Procedimentos da Gestão Global do Sistema do Setor Elétrico. 2014. [Online]. Available: http://www.mercado.ren.pt/PT/Electr/InfoMercado/DocReg/BibSubreg ula/MP_GGS_SE.pdf

[9] T. Soares, P. Pinson, T. V. Jensen, and H. Morais, "Optimal Offering Strategies for Wind Power in Energy and Primary Reserve Markets," IEEE Trans. Sustain. Energy, vol. 7, no. 3, pp. 1036–1045, Jul. 2016.

[10] "REstable Project Overview," REstable Project Website, 2017. [Online]. Available: https://www.restable-project.eu/overview/. [Accessed: 24-Feb-2017].

[11] J. Dowell, G. Hawker, K. Bell, and S. Gill, "A Review of Probabilistic Methods for Defining Reserve Requirements," 2016 IEEE Power and Energy Society General Meeting (PESGM), Boston, MA, 2016, pp. 1-5

[12] P. Pinson and G. Kariniotakis, "Conditional Prediction Intervals of Wind Power Generation," IEEE Trans. Power Syst., vol. 25, no. 4, pp. 1845–1856, Nov. 2010.

[13] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond," Int. J. Forecast., vol. 32, no. 3, pp. 896–913, Jul. 2016.

[14] Y. Ren, P. N. Suganthan, and N. Srikanth, "Ensemble methods for wind and solar power forecasting—A state-of-the-art review," Renew. Sustain. Energy Rev., vol. 50, pp. 82–91, Oct. 2015.

[15] RTE, "RTE - Portail clients - Volumes et prix." [Online]. Available: http://clients.rte-france.com/lang/fr/visiteurs/vie/mecanisme/volumes_prix/equilibrage.j sp. [Accessed: 17-Mar-2017].

[16] Fraunhofer IWES, "Regelenergie durch Windkraftanlagen - Abschlussbericht," Kassel, Mar. 2014. [Online]. Available: https://www.energiesystemtechnik.iwes.fraunhofer.de/content/dam/iwe s-neu/energiesystemtechnik/de/Dokumente/Studien-Reports/20140822_Abschlussbericht_rev1.pdf

[17] Fraunhofer IWES, Siemens AG, Universität Hannover, and CUBE Engineering GMBH, "Kombikraftwerk 2 - Final Report," Aug-2014. [Online]. Available: http://www.kombikraftwerk.de/fileadmin/Kombikraftwerk_2/English/ Kombikraftwerk2_FinalReport.pdf.

[18] A. Bocquet, A. Michiorri, A. Bossavy, R. Girard, and G. Kariniotakis, "Assessment of probabilistic PV production forecasts performance in an operational context," in 6th Solar Integration Workshop - International Workshop on Integration of Solar Power into Power Systems, Nov 2016, Vienna, Austria. Energynautics GmbH, pp.6.

[19] J. Juban, L. Fugon, and G. Kariniotakis, "Probabilistic short-term wind power forecasting based on kernel density estimators," in European Wind Energy Conference and exhibition, EWEC 2007, May 2007, MILAN, Italy.

[20] Y. Zhang, J. Wang, and X. Wang, "Review on probabilistic forecasting of wind power generation," Renew. Sustain. Energy Rev., vol. 32, pp. 255–270, Apr. 2014.

[21] Y. Zhang and J. Wang, "K-nearest neighbors and a kernel density estimator for GEFCom2014 probabilistic wind power forecasting," Int. J. Forecast., vol. 32, no. 3, pp. 1074–1080, Jul. 2016.

[22] E. Mangalova and O. Shesterneva, "K-nearest neighbors for GEFCom2014 probabilistic wind power forecasting," Int. J. Forecast., vol. 32, no. 3, pp. 1067–1073, Jul. 2016.

[23] M. Zambrano-Bigiarini, R. Rojas, and M. M. Zambrano-Bigiarini, "Package 'hydroPSO,'" 2013.

[24] Chiu, "Bandwidth selection for Kernel Density Estimation," Ann. Stat., vol. 19, 1991.

Bibliography

[1] ENTSO-E (the European Network of Transmission System Operators for Electricity). POWERFACTS Europe 2019. Technical report, 2019. URL `https://docstore.entsoe.eu/Documents/Publications/ENTSO-Egeneralpublications/ENTSO-E{_}PowerFacts{_}2019.pdf`.

[2] Réseau de transport d'électricité (RTE). RTE - Bilan électrique 2018. Technical report, 2018. URL `https://bilan-electrique-2018.rte-france.com/reseau-de-transport-qualite-delectricite/{#}`.

[3] Guillaume Denis, Thibault Prevost, Marie-Sophie Debry, Florent Xavier, Xavier Guillaud, and Andreas Menze. The Migrate project: the challenges of operating a transmission grid with only inverter-based generation. A grid-forming control improvement with transient current-limiting control. *IET Renewable Power Generation*, 12(5):523–529, apr 2017. ISSN 1752-1416. doi: 10.1049/iet-rpg.2017.0369. URL `http://digital-library.theiet.org/content/journals/10.1049/iet-rpg.2017.0369`.

[4] Lion Hirth and Inka Ziegenhagen. Balancing power and variable renewables: Three links. *Renewable and Sustainable Energy Reviews*, 50:1035–1051, oct 2015. ISSN 13640321. doi: 10.1016/j.rser.2015.04.180. URL `https://linkinghub.elsevier.com/retrieve/pii/S1364032115004530`.

[5] Benjamin Kroposki, Brian Johnson, Yingchen Zhang, Vahan Gevorgian, Paul Denholm, Bri Mathias Hodge, and Bryan Hannegan. Achieving a 100% Renewable Grid: Operating Electric Power Systems with Extremely High Levels of Variable Renewable Energy. *IEEE Power and Energy Magazine*, 2017. ISSN 15407977. doi: 10.1109/MPE.2016.2637122.

[6] Yann Rebours. A Comprehensive Assessment of Markets for Frequency and Voltage Control Ancillary Services A thesis submitted to The University of Manchester for the degree of. 2009.

[7] Alexandre Teninge. *Participation aux services système de parcs éoliens mixtes: application en milieu insulaire.* PhD thesis, 2009.

[8] Entsoe. an Overview of the European Balancing Market and Electricity Balancing Guideline. (November), 2018.

[9] RTE. Documentation Technique de Référence Chapitre 4– Contribution des utilisateurs aux performances du RPT; Article 4.1 - Réglage Fréquence/puissance. pages 1–17, 2011.

[10] Karin Kreuzer. NOTE DE SYNTHÈSE Participation des éoliennes aux services système en Allemagne et en France Rôle , cadres réglementaires et défis futurs Novembre 2016. pages 1–29, 2016.

[11] Javier Saez-Gallego, Juan M. Morales, Henrik Madsen, and Tryggvi Jónsson. Determining reserve requirements in DK1 area of Nord Pool using a probabilistic approach. *Energy*, 74 (C):682–693, 2014. ISSN 03605442. doi: 10.1016/j.energy.2014.07.034.

[12] Fei Teng and Goran Strbac. Full Stochastic Scheduling for Low-Carbon Electricity Systems. *IEEE Transactions on Automation Science and Engineering*, 14(2):461–470, apr 2017. ISSN 1545-5955. doi: 10.1109/TASE.2016.2629479. URL http://ieeexplore.ieee.org/document/7833096/.

[13] RTE. Règles Services Système. pages 1–210, 2016.

[14] RTE. Règles Services Système Fréquence. 2018.

[15] The European Commission. System Operations Guideline 2017. *Official Journal of the European Union*, L 220(25 August 2017), 2017. URL http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R1485{&}from=EN.

[16] Christian Redl, Dimitri Pescia, Vincent Rious, Nicolas Hary, and Marcelo Saguan. Refining Short-Term Electricity Markets to Enhance Flexibility. page 76, 2016. doi: 099/02-S-2016/EN. URL www.agora-energiewende.de.

[17] European Commission. Commission Regulation - Establishing a guideline on Electricity Balancing, mar 2017.

[18] 50hertz, Amprion, APG, Elia, Energinet, RTE, Swissgrid, Tennet, and Transnet BW. TSOs' proposal for the establishment of common and harmonised rules and processes for the exchange and procurement of Balancing Capacity for Frequency Containment Reserves (FCR) in accordance with Artticle 33 of Commission Regulation (EU) 2017/2195 establ. (April):1–9, 2018. URL https://consultations.entsoe.eu/markets/draft-proposal-for-fcr-cooperation-market-design/supporting{_}documents/FCRDraftProposal.pdf.

[19] Reinhard Mackensen, Yves-Marie Saint-Drenan, Dominik Jost, Rafael Fritz, Nazgul Asanalieva, Martin Widdel, and Markus Hahler. *Regelenergie durch Wind- und Photovoltaikparks Abschlussbericht.* 2017.

[20] Julia Strahlhoff, Andreas Liebelt, Stefan Siegl, and Simon Camal. Development and Application of KPIs for the Evaluation of the Control Reserve Supply by a Cross-border Renewable Virtual Power Plant. 2019. ISSN 1617-5468. URL https://dl.gi.de/handle/20.500.12116/25020.

[21] Minghui Yin, Yan Xu, Chun Shen, Jiankun Liu, Zhao Yang Dong, and Yun Zou. Turbine Stability-Constrained Available Wind Power of Variable Speed Wind Turbines for Active Power Control. *IEEE Transactions on Power Systems*, 32(3):2487–2488, 2017. ISSN 08858950. doi: 10.1109/TPWRS.2016.2605012.

[22] Ye Wang, Ye Wang, Evaluation De, and Ye Wang. Evaluation de la Performance des Réglages de Fréquence des Eoliennes à l ' Echelle du Système Electrique : Application à un Cas Insulaire. 2012.

[23] Jacob Aho, Lucy Pao, and Andrew Buckspan. An Active Power Control System for Wind Turbines Capable of AGC and Primary Response. In *Proc. of 51st AIAA Aerospace Sciences Meeting*, page 456, 2013.

[24] Tuhfe Göçmen, Gregor ; Giebel, Pierre-Elouan Réthoré, Murcia Leon, Juan Pablo, and Gregor Giebel. Uncertainty Quantification of the Real-Time Reserves for Offshore Wind Power Plants. 2016. URL http://orbit.dtu.dk/files/132541609/TuhfeGocmen{_}WIW2016.pdf.

[25] Anderson F. Hoke, Mariko Shirazi, Sudipta Chakraborty, Eduard Muljadi, and Dragan Maksimovic. Rapid Active Power Control of Photovoltaic Systems for Grid Frequency Support. *IEEE Journal of Emerging and Selected Topics in Power Electronics*, 5(3):1154–1163, 2017. ISSN 21686785. doi: 10.1109/JESTPE.2017.2669299.

[26] Kaspar Knorr, Britta Zimmermann, Markus Speckmann, Manuela Wunderlich, Dirk Kirchner, Florian Steinke, Philipp Wolfrum, Thomas Leveringhaus, Thomas Lager, Lutz Hofmann, Dirk Filzek, Tina Göbel, Bettina Kusserow, Lars Nicklaus, and Peter Ritter. Kombikraftwerk 2 Abschlussbericht. (August):218, 2014. URL http://www.kombikraftwerk.de/fileadmin/Kombikraftwerk{_}2/Abschlussbericht/Abschlussbericht{_}Kombikraftwerk2{_}aug14.pdf.

[27] Andreas Linder, Ana Kosareva, Konstantin Kunz, David Mcmullin, and Bettina Lenz. Balancing by Wind and Energy Storage. *17th International Wind Integration Workshop*, (October), 2018.

[28] Gregor Giebel, G Giebel, Göçmen Bozkurt, and N K Kristoffersen. Experimental verification of a real-time power curve for down-regulated offshore wind power plants. URL http://orbit.dtu.dk/files/107150137/Experimental{_}verification{_}of{_}a{_}real{_}time{_}power{_}curve.pdf.

[29] Zhao Wang, Weisheng Wang, Chun Liu, Zheng Wang, and Yunhe Hou. Probabilistic Forecast for Multiple Wind Farms Based on Regular Vine Copulas. *IEEE Transactions on Power Systems*, 33(1):578–589, 2018. ISSN 0885-8950. doi: 10.1109/TPWRS.2017.2690297. URL http://ieeexplore.ieee.org/document/7909035/.

[30] Pierre Pinson, Henrik Aa Nielsen, Jan K. Møller, Henrik Madsen, and George N. Kariniotakis. Non-parametric probabilistic forecasts of wind power: Required properties and evaluation. *Wind Energy*, 10(6):497–516, 2007. ISSN 10954244. doi: 10.1002/we.230.

[31] Howard Raiffa and Robert Schlaifer. *Applied statistical decision theory*. Studies in managerial economics. Div. of Research, Graduate School of Business Administration, Harvard Univ, Boston, 6. print edition, 1974. ISBN 978-0-87584-017-8.

[32] Georges Kariniotakis. *Renewable Energy Forecasting - From Models to Applications*. Woodhead Publishing Series in Energy. Woodhead Publishing Series in Energy - Elsevier, 2017. ISBN 978-0-08-100504-0.

[33] D. W. van der Meer, J. Munkhammar, and J. Widén. Probabilistic forecasting of solar power, electricity consumption and net load: Investigating the effect of seasons, aggregation and penetration on prediction intervals. *Solar Energy*, 171(July):397–413, 2018. ISSN 0038092X. doi: 10.1016/j.solener.2018.06.103. URL https://doi.org/10.1016/j.solener.2018.06.103.

[34] Athanasios Papakonstantinou, Georgia Champeri, Stefanos Delikaraoglou, and Pierre Pinson. Optimal trading strategies for wind power producers in futures and short-term electricity markets. URL http://pierrepinson.com/docs/papakonstantinouetal2017.pdf.

[35] Jethro Browell. Risk Constrained Trading Strategies for Stochastic Generation with a Single-Price Balancing Market. *arXiv:1708.02625 [q-fin]*, jun 2017.

[36] Pierre Pinson, Christophe Chevallier, and George N Kariniotakis. Trading Wind Generation From Short-Term Probabilistic Forecasts of Wind Power. 22(3):1148–1156, 2007.

[37] Franck Bourry. Management of uncertainties related to renewable generation participating in electricity markets. pages 1—-254, 2010. URL https://tel.archives-ouvertes.fr/tel-00508345/.

[38] Frans Faiella, Mariano;Hennig, Tobias;Cutululis, Nicolaos Antonio;Van Hulle. Capabilities and costs for ancillary services provision by wind power plants. Technical report, 2013. URL http://orbit.dtu.dk/files/71169605/Capabilities{_}and{_}costs{_}for{_}ancillary.pdf.

[39] 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 pro-c. (October):1–9, 2018. URL https://docstore.entsoe.eu/Documents/Networkcodesdocuments/NCEB/FCR{_}Proposal-Article{_}33{_}1EBGL{_}20181018{_}FV.PDF.

[40] Yao Zhang, Jianxue Wang, and Xifan Wang. Review on probabilistic forecasting of wind power generation. *Renewable and Sustainable Energy Reviews*, 32:255–270, 2014. ISSN

13640321. doi: 10.1016/j.rser.2014.01.033. URL http://dx.doi.org/10.1016/j.rser.2014.01.033.

[41] Dennis Van Der Meer, Joakim Widén, Joakim Munkhammar, D W Van Der Meer, J. Widén, and J. Munkhammar. Review on probabilistic forecasting of photovoltaic power production and electricity consumption Solar resource for high penetration and large scale applications View project Review on probabilistic forecasting of photovoltaic power production and electri. *Renewable and Sustainable Energy Reviews*, 2017. doi: 10.1016/j.rser.2017.05.212. URL https://www.researchgate.net/publication/317507439.

[42] Ricardo J. Bessa, Corinna Möhrlen, Vanessa Fundel, Malte Siefert, Jethro Browell, Sebastian Haglund El Gaidi, Bri Mathias Hodge, Umit Cali, and George Kariniotakis. Towards improved understanding of the applicability of uncertainty forecasts in the electric power industry. *Energies*, 10(9):1–48, 2017. ISSN 19961073. doi: 10.3390/en10091402.

[43] Tao Hong, Pierre Pinson, Shu Fan, Hamidreza Zareipour, Alberto Troccoli, and Rob J. Hyndman. Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. *International Journal of Forecasting*, 32(3):896–913, jul 2016. ISSN 01692070. doi: 10.1016/j.ijforecast.2016.02.001.

[44] Marcelo Pinho Almeida, Oscar Perpiñán, and Luis Narvarte. PV power forecast using a nonparametric PV model. *Solar Energy*, 115:354–368, 2015.

[45] Gábor I. Nagy, Gergő Barta, Sándor Kazi, Gyula Borbély, and Gábor Simon. GEFCom2014: Probabilistic solar and wind power forecasting using a generalized additive tree ensemble approach. *International Journal of Forecasting*, 32(3):1087–1093, 2016. ISSN 01692070. doi: 10.1016/j.ijforecast.2015.11.013.

[46] Ciaran Gilbert, Jethro Browell, and David McMillan. A hierarchical approach to probabilistic wind power forecasting. *2018 International Conference on Probabilistic Methods Applied to Power Systems, PMAPS 2018 - Proceedings*, pages 1–6, 2018. doi: 10.1109/PMAPS.2018.8440571.

[47] Jing Huang and Matthew Perry. A semi-empirical approach using gradient boosting and k-nearest neighbors regression for GEFCom2014 probabilistic solar power forecasting. *International Journal of Forecasting*, 2016. ISSN 01692070. doi: 10.1016/j.ijforecast.2015.11.002.

[48] G. N. Kariniotakis, G. S. Stavrakakis, , and E. F. Nogaret. Wind power forecasting using advanced neural networks models. *IEEE Transactions on Energy Conversion*, 11(4):762–767, 1996. ISSN 08858969. doi: 10.1109/60.556376. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=556376.

[49] Huai-zhi Wang, Gang-qiang Li, Gui-bin Wang, Jian-chun Peng, Hui Jiang, and Yi-tao Liu. Deep learning based ensemble approach for probabilistic wind power forecasting. *Applied Energy*, 188:56–70, feb 2017. ISSN 0306-2619. doi: 10.1016/J.APENERGY.2016.11.111. URL https://www.sciencedirect.com/science/article/pii/S0306261916317421.

[50] Margarida Solas. Convolutional Neural Network for Short-term Wind Power Forecasting. *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, pages 1–5. doi: 10.1109/ISGTEurope.2019.8905432.

[51] Aurélien Géron. *Hands-on Machine Learning with Scikit-Learn& TensorFlow*. 2017. ISBN 9781491962299. doi: 10.3389/fninf.2014.00014.

[52] Guido Cervone, Laura Clemente-Harding, Stefano Alesandrini, and Luca Delle Monache. PhotoVoltaic Power Forecasts Using Artificial Neural Networks and an Analog Ensemble. *Applied Energy (Under Review)*, 2016. ISSN 09601481. doi: 10.1016/j.renene.2017.02.052.

[53] Jinhua Zhang, Jie Yan, David Infield, Yongqian Liu, and Fue 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*, 241(January):229–244, 2019. ISSN 03062619. doi: 10.1016/j.apenergy.2019.03.044.

[54] Mahdi Khodayar and Jianhui Wang. Spatio-temporal Graph Deep Neural Network for Short-term Wind Speed Forecasting. *IEEE Transactions on Sustainable Energy*, 10(2):670–681, 2018. ISSN 19493029. doi: 10.1109/TSTE.2018.2844102.

[55] Yue Wang. An Adaptive Importance Sampling Method for Spinning Reserve Risk Evaluation of Generating Systems Incorporating Virtual Power Plants, 2018. ISSN 08858950.

[56] Li Bai and Pierre Pinson. Distributed Reconciliation in Day-Ahead Wind Power Forecasting. *Energies*, 12(6):1112, 2019. ISSN 1996-1073. doi: 10.3390/en12061112. URL https://www.mdpi.com/1996-1073/12/6/1112.

[57] Jakob W. Messner and Pierre Pinson. Online adaptive lasso estimation in vector autoregressive models for high dimensional wind power forecasting. *International Journal of Forecasting*, 35(4):1485–1498, 2018. ISSN 01692070. doi: 10.1016/j.ijforecast.2018.02.001. URL https://doi.org/10.1016/j.ijforecast.2018.02.001.

[58] Xwégnon Ghislain Agoua, Robin Girard, and Georges Kariniotakis. Probabilistic Model for Spatio-Temporal Photovoltaic Power Forecasting. *IEEE Transactions on Sustainable Energy*, 2018. doi: 10.1109/TSTE.2018.2847558. URL https://hal.archives-ouvertes.fr/hal-01817629.

[59] Amanda Lenzi, Ingelin Steinsland, Pierre Pinson, and Et Al Lenzi. Benefits of spatio-temporal modelling for short term wind power forecasting at both individual and aggregated levels. doi: 10.1002/env.XXXX. URL http://pierrepinson.com/docs/Lenzietal2017.pdf.

[60] Thordis Thorarinsdottir, Anders Løland, and Alex Lenkoski. Probabilistic Forecasting of Temporal Trajectories of Regional Power Production - Part 2: Photovoltaic Solar. pages 1–12, 2019. URL http://arxiv.org/abs/1903.01186http://arxiv.org/abs/1903.01188.

[61] Kumar Shridhar, Felix Laumann, Adrian Llopart Maurin, Martin Olsen, and Marcus Liwicki. Bayesian Convolutional Neural Networks with Variational Inference. 2018. URL http://arxiv.org/abs/1806.05978.

[62] Tim Van Erven, Jairo Cugliari, and Game-theoretically Optimal Reconciliation. Game-theoretically Optimal Reconciliation of Contemporaneous Hierarchical Time Series Forecasts To cite this version : Game-theoretically Optimal Reconciliation of Contemporaneous Hierarchical Time Series. 2013.

[63] Souhaib Ben Taieb, James W Taylor, and Rob J Hyndman. Coherent probabilistic forecasts for hierarchical time series. *Proceedings of the 34th International Conference on Machine Learning*, 70(April):3348–3357, 2017. ISSN 1938-7228. URL http://proceedings.mlr.press/v70/taieb17a.html.

[64] Kenneth Bruninx and Erik Delarue. Endogenous probabilistic reserve sizing and allocation in unit commitment models: cost-effective, reliable and fast. *IEEE TRANSACTIONS ON POWER SYSTEMS*, 2016. URL https://lirias.kuleuven.be/bitstream/123456789/554972/3/PUCMV10{_}final.pdf.

[65] P. Pinson and R. Girard. Evaluating the quality of scenarios of short-term wind power generation. *Applied Energy*, 96:12–20, 2012. ISSN 03062619. doi: 10.1016/j.apenergy.2011.11.004.

[66] Faranak Golestaneh, Hoay Beng Gooi, and Pierre Pinson. Generation and evaluation of space–time trajectories of photovoltaic power. *Applied Energy*, 176:80–91, aug 2016. ISSN 0306-2619. doi: 10.1016/J.APENERGY.2016.05.025. URL https://www.sciencedirect.com/science/article/pii/S0306261916306079.

[67] W. Wu, K. Wang, B. Han, G. Li, X. Jiang, and M. L. Crow. A Versatile Probability Model of Photovoltaic Generation Using Pair Copula Construction. *IEEE Transactions on Sustainable Energy*, 6(4):1337–1345, oct 2015. ISSN 1949-3029. doi: 10.1109/TSTE.2015.2434934.

[68] Marie Pier Côté and Christian Genest. A copula-based risk aggregation model. *Canadian Journal of Statistics*, 43(1):60–81, mar 2015. ISSN 1708945X. doi: 10.1002/cjs.11238. URL http://doi.wiley.com/10.1002/cjs.11238.

[69] Jin Dong, Teja Kuruganti, and Seddik M. Djouadi. Very Short-term Photovoltaic Power Forecasting using Uncertain Basis Function. *Information Sciences and Systems (CISS), 2017 51st Annual Conference on*, pages 1–6, 2017.

[70] Stylianos I. Vagropoulos, G. I. Chouliaras, E. G. Kardakos, C. K. Simoglou, and A. G. Bakirtzis. Comparison of SARIMAX, SARIMA, modified SARIMA and ANN-based models for short-term PV generation forecasting. *2016 IEEE International Energy Conference, ENERGYCON 2016*, 2016. doi: 10.1109/ENERGYCON.2016.7514029.

[71] Kenneth Bruninx. *Improved modeling of unit commitment decisions under uncertainty.* PhD thesis, 2016. URL https://www.mech.kuleuven.be/en/tme/research/energy{_}environment/Pdf/doctoraat-kenneth-bruninx.pdf.

[72] Guo Chen and Jueyou Li. A fully distributed ADMM-based dispatch approach for virtual power plant problems. *Applied Mathematical Modelling*, 58:300–312, 2018. ISSN 0307904X. doi: 10.1016/j.apm.2017.06.010.

[73] Marita Blank. *Reliability Assessment of Coalitions for.* PhD thesis, 2015.

[74] Roger Koenker, Samantha Leorato, and Franco Peracchi. Distributional vs. Quantile Regression. *Ssrn*, 11(15), 2013. doi: 10.2139/ssrn.2368737.

[75] Dong Jun Seo, Miah Mohammad Saifuddin, and Haksu Lee. Conditional bias-penalized Kalman filter for improved estimation and prediction of extremes. *Stochastic Environmental Research and Risk Assessment*, 32(1):183–201, jan 2018. ISSN 14363259. doi: 10.1007/s00477-017-1442-8. URL http://link.springer.com/10.1007/s00477-017-1442-8.

[76] M. A. Matos, R. J. Bessa, C. Goncalves, 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 - Proceedings*, pages 1–7. IEEE, oct 2016. ISBN 9781509019700. doi: 10.1109/PMAPS.2016.7764145. URL http://ieeexplore.ieee.org/document/7764145/.

[77] Romain Dupin. *DLR Forecasting PhD Thesis.* PhD thesis, PSL Research University, 2018.

[78] Tryggvi Jónsson. *Forecasting and Decision-Making in Electricity Markets with Focus on Wind Energy.* PhD thesis, 2012. URL file:///Users/markus/Dropbox/Papers/2012/J{ó}nsson{_}2012{_}PhDThesis.pdf{%}5Cnpapers://6470de79-5287-45a9-8e4f-b629919aff7a/Paper/p5455.

[79] Myriam Charras-garrido and Pascal Lezaud. Extreme Value Analysis : an Introduction To cite this version : Extreme Value Analysis : an Introduction. *Journal de la Societe Francaise de Statistique*, 154 No.2 (, 2013.

[80] Manuel A. Matos and R. J. Bessa. Setting the operating reserve using probabilistic wind power forecasts. *IEEE Transactions on Power Systems*, 26(2):594–603, 2011. ISSN 08858950. doi: 10.1109/TPWRS.2010.2065818.

[81] Christian P. Robert. *The Bayesian Choice*, volume 102. 2006. ISBN 9780387781884. doi: 10.1016/j.peva.2007.06.006. URL http://books.google.com/books?id=9tv0taI8l6YC.

[82] Jean Michel Marin, Kerrie Mengersen, and Christian P. Robert. Bayesian Modelling and Inference on Mixtures of Distributions. *Handbook of Statistics*, 25:459–507, 2005. ISSN 01697161. doi: 10.1016/S0169-7161(05)25016-2.

[83] Ching Wai (Jeremy) Chiu, Haroon Mumtaz, and Gábor Pintér. Forecasting with VAR models: Fat tails and stochastic volatility. *International Journal of Forecasting*, 33(4): 1124–1143, oct 2017. ISSN 01692070. doi: 10.1016/j.ijforecast.2017.03.001. URL https://www.sciencedirect.com/science/article/pii/S016920701730033X?via{%}3Dihub.

[84] Christian Contino and Richard H. Gerlach. Bayesian tail-risk forecasting using realized GARCH. *Applied Stochastic Models in Business and Industry*, 33(2):213–236, 2017. ISSN 15241904. doi: 10.1002/asmb.2237. URL http://doi.wiley.com/10.1002/asmb.2237.

[85] Feng Li and Yanfei Kang. Improving forecasting performance using covariate-dependent copula models. *International Journal of Forecasting*, 34(3):456–476, jul 2018. ISSN 01692070. doi: 10.1016/j.ijforecast.2018.01.007. URL https://www.sciencedirect.com/science/article/pii/S0169207018300323?via{%}3Dihub.

[86] Can Wan, Jin Lin, Jianhui Wang, Yonghua Song, and Zhao Yang Dong. Direct Quantile Regression for Nonparametric Probabilistic Forecasting of Wind Power Generation. *IEEE Transactions on Power Systems*, 32(4):2767–2778, 2017. ISSN 08858950. doi: 10.1109/TPWRS.2016.2625101.

[87] Emilio Soria-Olivas, Juan Gómez-Sanchis, José D. Martín, Joan Vila-Francés, Marcelino Martínez, José R. Magdalena, and Antonio J. Serrano. BELM: Bayesian extreme learning machine. *IEEE Transactions on Neural Networks*, 22(3):505–509, 2011. ISSN 10459227. doi: 10.1109/TNN.2010.2103956.

[88] Sonia Migliorati, Agnese Maria Di Brisco, and Andrea Ongaro. A new regression model for bounded responses. *Bayesian Analysis*, 13(3):845–872, 2018. ISSN 19316690. doi: 10.1214/17-BA1079.

[89] Joao Soares, Zita Vale, Nuno Borges, Fernando Lezama, and Nelson Kagan. Multi-objective robust optimization to solve energy scheduling in buildings under uncertainty. *2017 19th International Conference on Intelligent System Application to Power Systems, ISAP 2017*, 2017. doi: 10.1109/ISAP.2017.8071417.

[90] Ye Wang, Herman Bayem, Maria Giralt-Devant, Vera Silva, Xavier Guillaud, and Bruno Francois. Methods for Assessing Available Wind Primary Power Reserve. *IEEE TRANSACTIONS ON SUSTAINABLE ENERGY*, 6(1), 2015. doi: 10.1109/TSTE.2014.2369235. URL http://l2ep.univ-lille1.fr/pagesperso/francois/IEEE{_}trans{_}SustDev{_}2015.pdf.

[91] Hao Fu, Zhi Wu, Xiao-Ping Zhang, and Joachim Brandt. Contributing to DSO's Energy-Reserve Pool: A Chance-Constrained Two-Stage μVPP Bidding Strategy. *IEEE Power and Energy Technology Systems Journal*, 4(4):1–1, dec 2017. ISSN 2332-7707. doi: 10.1109/JPETS.2017.2749256. URL http://ieeexplore.ieee.org/document/8058436/.

[92] Carlos Adrian Correa-florez, Andrea Michiorri, George Kariniotakis, and Senior Member. Optimal Participation of Residential Aggregators in Energy and Local Flexibility Markets. *IEEE Transactions on Smart Grid*, PP(c):1, 2019. doi: 10.1109/TSG.2019.2941687.

[93] H Ding, P Pinson, Z Hu, and Y Song. Optimal Offering and Operating Strategies for Wind-Storage Systems with Linear Decision Rules. *IEEE Transactions on Power Systems*, 31(6): 4755–4764, 2017. ISSN 0885-8950. doi: 10.1109/TPWRS.2016.2521177. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-84957709390{&}doi=10.1109{%}2FTPWRS.2016.2521177{&}partnerID=40{&}md5=f292cc7e5f7c7b881dcdd9a76fc04fbe.

[94] Evangelos Vrettos, Frauke Oldewurtel, and Goran Andersson. Robust Energy-Constrained Frequency Reserves from Aggregations of Commercial Buildings. *IEEE Transactions on Power Systems*, 31(6):4272–4285, 2016. ISSN 08858950. doi: 10.1109/TPWRS.2015.2511541.

[95] Kaspar Knorr, Britta Zimmermann, Markus Speckmann, Manuela Wunderlich, Dirk Kirchner, Florian Steinke, Philipp Wolfrum, Thomas Leveringhaus, Thomas Lager, Lutz Hofmann, Dirk Filzek, Tina Göbel, Bettina Kusserow, Lars Nicklaus, and Peter Ritter. Kombikraftwerk 2 - Final Report. Technical report, aug 2014. URL http://www.kombikraftwerk.de/fileadmin/Kombikraftwerk{_}2/Abschlussbericht/Abschlussbericht{_}Kombikraftwerk2{_}aug14.pdf.

[96] EPEX SPOT and OFA EnR. Vente directe des énergies renouvelables sur la Bourse Européenne de l ' Électricité NOTE DE SYNTHÈSE Un retour d ' expérience sur la transition énergétique française et allemande Février 2015. 2015.

[97] Thomas Ackermann, Enrico Maria Carlini, Bernhard Ernst, Frank Groome, Antje Orths, Jon O'Sullivan, Miguel De La Torre Rodriguez, and Vera Silva. Integrating Variable Renewables in Europe : Current Status and Recent Extreme Events. *IEEE Power and Energy Magazine*, 13(6):67–77, 2015. ISSN 15407977. doi: 10.1109/MPE.2015.2461333.

[98] The European Commission. Commission Regulation (EU) 2017/2195 of 23 November 2017 establishing a guideline on electricity balancing. *Official Journal of the European Union*, 2017(November):312/6 – 312/53, 2017. URL https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L{_}.2017.312.01.0006.01.ENG{&}toc=OJ:L:2017:312:TOC{#}d1e2376-6-1.

[99] Amprion, 50Hertz, Tennet, and BW Transnet. Leitfaden zur Präqualifikation von Windenergieanlagen zur Erbringung von Minutenreserveleistung im Rahmen einer Pilotphase. 1.1: 1–8, 2016.

[100] Pierre Pinson, Henrik Madsen, Henrik Aa Nielsen, George Papaefthymiou, and Bernd Klöckl. From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy*, 12(1):51–62, 2009. ISSN 10954244. doi: 10.1002/we.284. URL www.imm.dtu.dk/\protect\begingroup\immediate\write\@unused\def\MessageBreak˙\let\protect\edefYoumayprovideadefinitionwith\MessageBreak\DeclareUnicodeCharacter\errhelp\let\def\MessageBreak˙(inputenc)\def\errmessagePackageinputencError:UnicodecharacterâĹ(U+223C)\MessageBreaknotsetupforusewithLaTeX.

``Seetheinputencpackagedocumentationforexplanation.´TypeH<return> forimmediatehelp\endgrouppp.

[101] Nikolay Laptev, Jason Yosinski, Li Erran Li, and Slawek Smyl. Time-series Extreme Event Forecasting with Neural Networks at Uber. *International Conference on Machine Learning - Time Series Workshop*, pages 1–5, 2017. URL http://roseyu.com/time-series-workshop/ submissions/TSW2017{_}paper{_}3.pdf.

[102] Juan M. Morales, P. Pinson, and M. Zugno. *Integrating Renewables in Electricity Markets*. 2011. ISBN 9781441962805. doi: doi:10.1007/978-1-4419-6281-2_13. URL http://www.springer.com/series/6161http://books.google.com/books?hl= en{&}lr={&}id=Q6OdkTmpvE8C{&}oi=fnd{&}pg=PR2{&}dq=Profiles+in+Operations+ Research{&}ots=vufk95gzp2{&}sig=VGSPABTkW7JMlIvvtuuvD5Kxlso.

[103] T. Soares, P. Pinson, T. V. Jensen, and H. Morais. Optimal Offering Strategies for Wind Power in Energy and Primary Reserve Markets. *IEEE Transactions on Sustainable Energy*, 7(3):1036–1045, jul 2016. ISSN 1949-3029. doi: 10.1109/TSTE.2016.2516767.

[104] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision?, 2017.

[105] Nils Siebert. *Development of methods for regional wind power forecasting*. PhD thesis, 2008. URL https://pastel.archives-ouvertes.fr/tel-00287551.

[106] Jussi Ekstrom, Matti Koivisto, Ilkka Mellin, Robert John Millar, and Matti Lehtonen. A Statistical Model for Hourly Large-Scale Wind and Photovoltaic Generation in New Locations. *IEEE Transactions on Sustainable Energy*, 8(4):1383–1393, 2017. ISSN 19493029. doi: 10.1109/TSTE.2017.2682338.

[107] Matti Koivisto, Petr Maule, Nicolaos Cutululis, and Poul Sorensen. Effects of Wind Power Technology Development on Large-scale VRE Generation Variability. pages 1–6, 2019. doi: 10.1109/ptc.2019.8810687.

[108] Leo Breiman. No Title. pages 1–33, 2001.

[109] Greg Ridgeway. Generalized Boosted Models: A guide to the gbm package. *Compute*, 1(4):1–12, 2007. ISSN 14679752. doi: 10.1111/j.1467-9752.1996.tb00390. x. URL http://www.saedsayad.com/docs/gbm2.pdfhttp://cran.r-project.org/web/ packages/gbm/vignettes/gbm.pdf.

[110] G. Juban, J., Fugon, L., Kariniotakis. Probabilistic Short-Term Wind Power Forecasting Based on Kernel Density Estimators. *European Wind Conference and Exhibition*, pages 1–11, 2007. URL https://hal-mines-paristech.archives-ouvertes.fr/hal-00526011/ document.

[111] Nicolai Meinshausen and Maintainer Nicolai Meinshausen. Package 'quantregForest'. 2016.

[112] Erwan Scornet, Gérard Biau, and Jean-Philippe Vert. Consistency of random forests. may 2014. doi: 10.1214/15-AOS1321. URL http://arxiv.org/abs/1405.2881http://dx.doi.org/10.1214/15-AOS1321.

[113] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 1997. ISSN 08997667. doi: 10.1162/neco.1997.9.8.1735.

[114] Huaizhi Wang, Haiyan Yi, Jianchun Peng, Guibin Wang, Yitao Liu, Hui Jiang, and Wenxin Liu. Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network. *Energy Conversion and Management*, 153:409–422, dec 2017. ISSN 01968904. doi: 10.1016/j.enconman.2017.10.008. URL https://www.sciencedirect.com/science/article/pii/S019689041730910X?{_}rdoc=1{&}{_}fmt=high{&}{_}origin=gateway{&}{_}docanchor={&}md5=b8429449ccfc9c30159a5f9aeaa92ffb{&}dgcid=raven{_}sd{_}recommender{_}email.

[115] Anastasia Borovykh, Sander Bohte, and Cornelis W. Oosterlee. Conditional Time Series Forecasting with Convolutional Neural Networks. mar 2017. URL http://arxiv.org/abs/1703.04691.

[116] Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. Weight Uncertainty in Neural Networks. may 2015. URL http://arxiv.org/abs/1505.05424.

[117] Meire Fortunato, Charles Blundell, and Oriol Vinyals. Bayesian Recurrent Neural Networks. pages 1–14, 2017. ISSN 1944-8252. doi: 10.1021/am301964y. URL http://arxiv.org/abs/1704.02798.

[118] Duong D. Le, George Gross, and Alberto Berizzi. Probabilistic Modeling of Multisite Wind Farm Production for Scenario-Based Applications. *IEEE Transactions on Sustainable Energy*, 6(3):748–758, jul 2015. ISSN 1949-3029. doi: 10.1109/TSTE.2015.2411252. URL http://ieeexplore.ieee.org/document/7079532/.

[119] Jie Yan, Hao Zhang, Yongqian Liu, Shuang Han, Li Li, and Zongxiang Lu. Forecasting the High Penetration of Wind Power on Multiple Scales Using Multi-to-Multi Mapping. *IEEE Transactions on Power Systems*, 33(3):3276–3284, may 2018. ISSN 0885-8950. doi: 10.1109/TPWRS.2017.2787667. URL http://ieeexplore.ieee.org/document/8240639/.

[120] Jakob W Messner, Pierre Pinson, Jethro Browell, and Mathias B Bjerreg. Evaluation of Wind Power Forecasts – An up-to-date view. (2011).

[121] Philippe Lauret, Mathieu David, and Pierre Pinson. Verification of solar irradiance probabilistic forecasts. *Solar Energy*, 194:254–271, dec 2019. ISSN 0038092X. doi: 10.1016/j.solener.2019.10.041. URL https://linkinghub.elsevier.com/retrieve/pii/S0038092X19310382.

[122] Tilmann Gneiting, Fadoua Balabdaoui, Adrian Raftery, Tilmann Gneiting, Fadoua Balabdaoui, and Adrian Raftery. Probabilistic forecasts , calibration and sharpness . To cite this version :. 2009.

[123] Ludger Rüschendorf. *Mathematical Risk Analysis*, volume 2013. 2013. ISBN 978-3-642-33589-1. doi: 10.1007/978-3-642-33590-7. URL http://link.springer.com/10.1007/978-3-642-33590-7.

[124] Alexis Gérossier. Prévision à court terme de la demande électrique des maisons intelligentes et des réseaux de distribution Short-Term Forecasting of Electricity Demand of Smart Homes and Distribution Grids Soutenue par. 2019.

[125] Jochen Bröcker and Leonard A Smith. Increasing the Reliability of Reliability Diagrams. 2006. URL http://www.personal.reading.ac.uk/{~}pt904209/publications/reliable{_}reldiags.pdf.

[126] Michael Scheuerer and Thomas M Hamill. Variogram-Based Proper Scoring Rules for Probabilistic Forecasts of Multivariate Quantities*. *Monthly Weather Review*, 143(4):1321–1334, 2015. ISSN 0027-0644. 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.

[127] Kenneth Bruninx, Hrvoje Pandzic, Hélène Le Cadre, Erik Delarue, K Bruninx, H Pandžipandžic, H Le Cadre, and E Delarue. On the Interaction between Aggregators, Electricity Markets and Residential Demand Response Providers. (August), 2018. URL http://www.mech.kuleuven.be/tme/research/.

[128] Tryggvi Jónsson, Pierre Pinson, Henrik Madsen, and Henrik Aalborg Nielsen. Predictive Densities for Day-Ahead Electricity Prices Using Time-Adaptive Quantile Regression. *Energies*, 7(9):5523–5547, aug 2014. doi: 10.3390/en7095523.

[129] Pierre Pinson and George Kariniotakis. Conditional Prediction Intervals of Wind Power Generation. *IEEE Transactions on Power Systems*, 25(4):1845–1856, nov 2010. ISSN 0885-8950, 1558-0679. doi: 10.1109/TPWRS.2010.2045774.

[130] D. W. van der Meer, M. Shepero, A. Svensson, J. Widén, and J. Munkhammar. Probabilistic forecasting of electricity consumption, photovoltaic power generation and net demand of an individual building using Gaussian Processes. *Applied Energy*, 213:195–207, mar 2018. ISSN 03062619. doi: 10.1016/j.apenergy.2017.12.104. URL https://www.sciencedirect.com/science/article/pii/S0306261917318275.

[131] G. Sideratos and N. D. Hatziargyriou. Wind power forecasting focused on extreme power system events. *IEEE Transactions on Sustainable Energy*, 3(3):445–454, 2012. ISSN 19493029. doi: 10.1109/TSTE.2012.2189442.

[132] Yunjie Liu, Evan Racah, Prabhat, Joaquin Correa, Amir Khosrowshahi, David Lavers, Kenneth Kunkel, Michael Wehner, and William Collins. Application of Deep Convolutional Neural Networks for Detecting Extreme Weather in Climate Datasets. may 2016. URL http://arxiv.org/abs/1605.01156.

[133] Jethro Dowell, Graeme Hawker, Keith Bell, and Simon Gill. A Review of probabilistic methods for defining reserve requirements. *IEEE Power and Energy Society General Meeting*, 2016-Novem, 2016. ISSN 19449933. doi: 10.1109/PESGM.2016.7741361.

[134] Zhanyu Ma and Arne Leijon. Bayesian estimation of beta mixture models with variational inference. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(11):2160–2173, 2011. ISSN 01628828. doi: 10.1109/TPAMI.2011.63.

[135] Matt Hoffman, David M. Blei, Chong Wang, and John Paisley. Stochastic Variational Inference. 14:1303–1347, 2012. URL http://arxiv.org/abs/1206.7051.

[136] Michael Betancourt. A Conceptual Introduction to Hamiltonian Monte Carlo. 2017. URL http://arxiv.org/abs/1701.02434.

[137] Yuling Yao, Aki Vehtari, Daniel Simpson, and Andrew Gelman. Yes, but Did It Work?: Evaluating Variational Inference. 2018. ISSN 1520-0442. doi: 10.1175/2009JCLI3025.1. URL http://arxiv.org/abs/1802.02538.

[138] Abhay K. Singh, David E. Allen, and Robert J. Powell. Tail dependence analysis of stock markets using extreme value theory. *Applied Economics*, 49(45):4588–4599, sep 2017. ISSN 14664283. doi: 10.1080/00036846.2017.1287858. URL https://www.tandfonline.com/doi/full/10.1080/00036846.2017.1287858.

[139] Anne Sabourin and Philippe Naveau. Bayesian Dirichlet mixture model for multivariate extremes: A re-parametrization. *Computational Statistics & Data Analysis*, 71:542–567, mar 2014. ISSN 01679473. doi: 10.1016/j.csda.2013.04.021. URL https://linkinghub.elsevier.com/retrieve/pii/S0167947313001680.

[140] Faranak Golestaneh, Pierre Pinson, and Hoay Beng Gooi. Polyhedral Predictive Regions for Power System Applications, 2018. ISSN 08858950. URL https://arxiv.org/pdf/1803.04669.pdf.

[141] Title Extreme, Value Analysis, and Author Eric Gilleland. Package ' extRemes '. 2019.

[142] Tilmann Gneiting and Roopesh Ranjan. Comparing density forecasts using thresholdand quantile-weighted scoring rules. *Journal of Business and Economic Statistics*, 29(3):411–422, 2011. ISSN 07350015. doi: 10.1198/jbes.2010.08110.

[143] Pierre Pinson. Renewables in Electricity Markets (in less than 3 hours!... Part II). (September), 2014.

[144] Malte Jansen. Economics of control reserve provision by fluctuating renewable energy sources. *International Conference on the European Energy Market, EEM*, 2016-July(December), 2016. ISSN 21654093. doi: 10.1109/EEM.2016.7521342.

[145] Hunyoung Shin, Duehee Lee, and Ross Baldick. An Offer Strategy for Wind Power Producers That Considers the Correlation between Wind Power and Real-Time Electricity Prices. *IEEE*

Transactions on Sustainable Energy, 9(2):695–706, 2018. ISSN 19493029. doi: 10.1109/TSTE. 2017.2757501.

[146] Juan M. Morales, Antonio J. Conejo, Henrik Madsen, Pierre Pinson, and Marco Zugno. *Integrating Renewables in Electricity Markets*, volume 205 of *International Series in Operations Research & Management Science*. Springer US, Boston, 2014. ISBN 978-1-4614-9410-2.

[147] Juan M. Morales, Antonio J. Conejo, Henrik Madsen, Pierre Pinson, and Marco Zugno. *Integrating Renewables in Electricity Markets*, volume 205 of *International Series in Operations Research & Management Science*. Springer US, Boston, 2014. ISBN 978-1-4614-9410-2.

[148] Jethro Browell, Ciaran Gilbert, and David McMillan. Use of turbine-level data for improved wind power forecasting. In *2017 IEEE Manchester PowerTech*, pages 1–6. IEEE, jun 2017. ISBN 978-1-5090-4237-1. doi: 10.1109/PTC.2017.7981134. URL http://ieeexplore.ieee. org/document/7981134/.

[149] EPEX Spot. EPEX SPOT Exchange Rules. (January):1–28, 2018.

[150] Elec-Marché De. How the European day-ahead electricity market works Starting question. (October), 2014.

[151] Rafał Weron. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4):1030–1081, oct 2014. ISSN 01692070. doi: 10.1016/j.ijforecast.2014.08.008.

[152] Ricardo J. Bessa and Manuel A. Matos. Forecasting issues for managing a portfolio of electric vehicles under a smart grid paradigm. In *IEEE PES Innovative Smart Grid Technologies Conference Europe*, pages 1–8, Berlin, 2012. IEEE. ISBN 9781467325974. doi: 10.1109/ ISGTEurope.2012.6465724.

[153] Joanna Janczura and R. Weron. Efficient estimation of Markov regime-switching models: An application to electricity spot prices. *AStA Advances in Statistical Analysis*, 2012. ISSN 18638171. doi: 10.1007/s10182-011-0181-2.

[154] Najeh Chaâbane. A novel auto-regressive fractionally integrated moving average–least-squares support vector machine model for electricity spot prices prediction. *Journal of Applied Statistics*, 41(3):635–651, mar 2014. ISSN 0266-4763. doi: 10.1080/02664763.2013.847068.

[155] Mark Stitson, Alex Gammerman, Vladimir Vapnik, Volodya Vovk, Chris Watkins, and Jason Weston. Support vector regression with ANOVA decomposition kernels. *Advances in kernel methods—Support vector learning*, pages 285–292, 1999.

[156] Alex J Smola and Bernhard Sc Olkopf. A tutorial on support vector regression *. *Statistics and Computing*, 14:199–222, 2004. URL https://alex.smola.org/papers/2004/SmoSch04. pdf.

[157] H. Shin, D. Lee, and R. Baldick. An Offer Strategy for Wind Power Producers that Considers the Correlation between Wind Power and Real-Time Electricity Prices. *IEEE Transactions on Sustainable Energy*, PP(99):1–1, 2017. ISSN 1949-3029. doi: 10.1109/TSTE.2017.2757501.

[158] Ricardo J. Bessa, V. Miranda, A. Botterud, Z. Zhou, and J. Wang. Time-adaptive quantile-copula for wind power probabilistic forecasting. *Renewable Energy*, 40(1):29–39, 2012. ISSN 09601481. doi: 10.1016/j.renene.2011.08.015. URL http://dx.doi.org/10.1016/j.renene.2011.08.015.

[159] Tiago Soares and P. Pinson. *Renewable energy sources offering flexibility through electricity markets*. PhD thesis, Technical University of Denmark, 2017.

[160] Jianxiao Wang, Haiwang Zhong, Wenyuan Tang, Ram Rajagopal, Qing Xia, Chongqing Kang, and Yi Wang. Optimal bidding strategy for microgrids in joint energy and ancillary service markets considering flexible ramping products. *Applied Energy*, 205:294–303, 2017. ISSN 03062619. doi: 10.1016/j.apenergy.2017.07.047. URL https://doi.org/10.1016/j.apenergy.2017.07.047.

[161] Hongcai Zhang, Zechun Hu, Eric Munsing, Scott J Moura, and Yonghua Song. Data-driven Chance-constrained Regulation Capacity Offering for Distributed Energy Resources. *IEEE Transactions on Smart Grid*, 2018. ISSN 19493053. doi: 10.1109/TSG.2018.2809046. URL https://arxiv.org/pdf/1708.05114.pdf.

[162] Yang Liu, Prajit Ramachandran, Qiang Liu, and Jian Peng. Stein Variational Policy Gradient. apr 2017. URL http://arxiv.org/abs/1704.02399.

[163] Alexander W. Dowling, Ranjeet Kumar, and Victor M. Zavala. A multi-scale optimization framework for electricity market participation. *Applied Energy*, 190:147–164, 2017. ISSN 03062619. doi: 10.1016/j.apenergy.2016.12.081. URL http://dx.doi.org/10.1016/j.apenergy.2016.12.081.

[164] Bin Zhang and Zhongsheng Hua. *Handbook of Newsvendor Problems*, volume 176. 2012. ISBN 978-1-4614-3599-0. doi: 10.1007/978-1-4614-3600-3. URL http://www.scopus.com/inward/record.url?eid=2-s2.0-84955147896{&}partnerID=tZOtx3y1.

[165] Zhang, Hodge, and Florita. Joint Probability Distribution and Correlation Analysis of Wind and Solar Power Forecast Errors in the Western Interconnection. *Journal of Energy Engineering*, 139(2):120929002655007, 2012. ISSN 0733-9402. doi: 10.1061/(ASCE)EY.1943-7897. URL http://ascelibrary.org/doi/abs/10.1061/(ASCE)EY.1943-7897.0000103.

[166] Christian Genest, Bruno Rémillard, and David Beaudoin. Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44:199–213, 2009. doi: 10.1016/j.insmatheco.2007.10.005. URL www.elsevier.com/locate/ime.

[167] Eike Christian Brechmann. Statistical inference of vine copulas using the R-package VineCopula. 2013. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.648.9696{&}rep=rep1{&}type=pdf.

[168] 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 - Proceedings*, volume 2018-Janua, 2018. ISBN 9781538619537. doi: 10.1109/ISGTEurope. 2017.8260311.

[169] Eike Christian Brechmann and Ulf Schepsmeier. Modeling Dependence with C- and D-Vine Copulas: The <i>R</i> Package CDVine. *Journal of Statistical Software*, 2013. ISSN 1548-7660. doi: 10.18637/jss.v052.i03.

[170] Amaia Gonzalez-Garrido, Andoni Saez-de Ibarra, Haizea Gaztanaga, Aitor Milo, and Pablo 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):1–10, 2018. ISSN 08858950. doi: 10.1109/TPWRS.2018.2869626.

[171] S. Camal, F. Teng, A. Michiorri, G. Kariniotakis, and L. Badesa. Scenario generation of aggregated Wind, Photovoltaics and small Hydro production for power systems applications. *Applied Energy*, 242:1396–1406, may 2019. ISSN 03062619. doi: 10.1016/j.apenergy.2019.03. 112. URL https://www.sciencedirect.com/science/article/pii/S0306261919305203? via{%}3Dihub.

[172] Simon Camal, Andrea Michiorri, and Georges Kariniotakis. Optimal Offer of Automatic Frequency Restoration Reserve from a Combined PV/Wind Virtual Power Plant. *IEEE Transactions on Power Systems*, 99, 2018. ISSN 08858950. doi: 10.1109/TPWRS.2018. 2847239.

[173] Jooyoung Jeon, Anastasios Panagiotelis, and Fotios Petropoulos. Probabilistic forecast reconciliation with applications to wind power and electric load. *European Journal of Operational Research*, 2019. ISSN 03772217. doi: 10.1016/j.ejor.2019.05.020. URL https: //linkinghub.elsevier.com/retrieve/pii/S0377221719304242.

[174] Jesus Lago, Fjo De Ridder, and Bart De Schutter. Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. *Applied Energy*, 221:386–405, jul 2018. ISSN 03062619. doi: 10.1016/j.apenergy.2018.02.069. URL https: //linkinghub.elsevier.com/retrieve/pii/S030626191830196X.

[175] Stéphane Mallat. Wavelets - Sparse Edition. 2009. ISSN 1098-6596. doi: 10.1017/CBO9781107415324.004. URL http://ebooks.cambridge.org/ref/id/ CBO9781107415324A009.

[176] Angi Rösch and Harald Schmidbauer. WaveletComp: Computational Wavelet Analysis. R package version 1.1. 2018. URL http://www.hs-stat.com/projects/ WaveletComp/WaveletComp{_}guided{_}tour.pdfhttps://cran.r-project.org/ package=WaveletComp.

RÉSUMÉ

Les énergies renouvelables variables prennent une part croissante de la production raccordée aux réseaux électriques. Par conséquent, elles doivent s'intégrer aux mécanismes de services système qui assurent l'équilibre entre production et consommation de puissance sur les réseaux. Toutefois la forte incertitude de la production variable est un obstacle à la fourniture de ces services qui requièrent une fiabilité élevée. L'agrégation de centrales renouvelables dispersées et contrôlées par une centrale virtuelle permet de diminuer cette incertitude en profitant du foisonnement entre les centrales. Cette thèse propose plusieurs modèles de prévision probabiliste afin d'évaluer la capacité d'une centrale virtuelle renouvelable variable à offrir des services système avec une fiabilité maximale: ces modèles sont des adaptations d'arbres de décisions, de réseaux de neurones récurrents et convolutifs, ainsi que de distributions dédiées aux quantiles extrêmement faibles. Une attention particulière est portée à la combinaison de sources d'énergie (Photovoltaïque, éolien, hydraulique au fil de l'eau). Ensuite, des stratégies d'offre optimale d'énergie et de réserve par une agrégation renouvelable sont établies en utilisant les prévisions de production et en considérant les incertitudes associées aux différents marchés. Ces stratégies explorent plusieurs options de modélisation: dépendance entre production renouvelable et prix par une copule, taux de défaillance contrôlé par optimisation sous contraintes probabilistes, et enfin offre de services système multiples à l'aide d'une formulation Lagrangienne.

MOTS CLÉS

Prévision, Optimisation, Machine Learning, Energies renouvelables, Services système, Centrale virtuelle, Agrégation, Smart Grids

ABSTRACT

As variable renewable energy plants penetrate significantly the electricity generation mix, they are expected to contribute to the supply of reserve power, albeit the high uncertainty levels on their production. A solution to reduce the uncertainty consists in aggregating renewable plants dispersed over several climates to obtain a smoother production profile and operate them within a Virtual Power Plant control system. In this thesis, a series of probabilistic forecasting models are proposed to assess the capacity of a variable renewable Virtual Power Plant to provide ancillary services with maximum reliability: these models are adapted decision-tree regression models, recurrent and convolutional neural networks, as well as distributions dedicated to extremely low quantiles. The combination of energy sources (Photovoltaics, Wind, Run-of-river Hydro) is considered in detail. Optimal strategies for the joint offer of energy and ancillary services by a variable renewable Virtual Power Plant are later defined, based on production forecasts and market uncertainties. Offer strategies explore several modelling options: dependence between renewable production and prices via a copula, controlled rate of reserve underfullfilment with a chance-constraint optimization, and finally offer of multiple ancillary services thanks to a Lagrangian formulation.

KEYWORDS

Forecasting, Optimisation, Machine Learning, Renewable energy, Ancillary services, Virtual Power Plant, Aggregation, Smart Grids