



UNIVERSITÀ DEGLI STUDI DI NAPOLI
FEDERICO II



UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

PH.D. THESIS

IN

**INFORMATION TECHNOLOGY AND ELECTRICAL
ENGINEERING**

**ADVANCED FORECASTING METHODS FOR
RENEWABLE GENERATION AND LOADS IN
MODERN POWER SYSTEMS**

PASQUALE DE FALCO

TUTOR: PROF. GUIDO CARPINELLI

CO-TUTOR: DR. ANTONIO BRACALE

COORDINATOR: PROF. DANIELE RICCIO

XXX CICLO

**SCUOLA POLITECNICA E DELLE SCIENZE DI BASE
DIPARTIMENTO DI INGEGNERIA ELETTRICA E TECNOLOGIE
DELL'INFORMAZIONE**

To Marisa and Vincenzo

To Luisa

Index

List of figures	v
List of tables	vii
List of abbreviations.....	ix
List of symbols.....	xi
Acknowledgment	xix
INTRODUCTION.....	1
CHAPTER 1.	
FORECASTING AND POWER SYSTEMS.....	6
1.1. INTRODUCTION	6
1.2. CLASSIFICATIONS OF FORECASTING METHODS.....	7
1.3. FORECASTING METHODS APPLIED TO POWER SYSTEMS	11
1.3.1. Statistical approaches.....	12
1.3.1.1. <i>Naïve approaches</i>	12
1.3.1.2. <i>Regression analysis</i>	15
1.3.1.3. <i>Univariate stochastic time series</i>	16
1.3.1.4. <i>Exponential smoothing</i>	18
1.3.1.5. <i>Bayesian approaches</i>	20
1.3.1.6. <i>Markov Chains</i>	21
1.3.1.7. <i>Kalman filter</i>	22
1.3.1.8. <i>Artificial neural networks</i>	23
1.3.1.9. <i>Support vector regression</i>	24
1.3.1.10. <i>Fuzzy approaches</i>	26
1.3.1.11. <i>K-nearest neighbors</i>	27
1.3.1.12. <i>Kernel density estimation</i>	28
1.3.2. Physical approaches.....	29
1.3.3. Hybrid approaches	30
CHAPTER 2.	
ADVANCED PROBABILISTIC METHODS FOR SHORT-TERM PHOTOVOLTAIC	
POWER FORECASTING	33
2.1. INTRODUCTION	33
2.2. PROBABILISTIC METHODS FOR SHORT-TERM PHOTOVOLTAIC POWER	
FORECASTING: STATE OF THE ART.....	34
2.3. A NEW PROBABILISTIC BAYESIAN-BASED METHOD FOR SHORT-TERM	
PHOTOVOLTAIC POWER FORECASTING	38
2.3.1. Proposed method	39
2.3.1.1. <i>Relationships that link the PV active power to the hourly</i>	
<i>clearness index and the solar irradiance</i>	40
2.3.1.2. <i>Selection of the PDFs of the hourly solar irradiance and</i>	
<i>of the hourly clearness index irradiance</i>	41

2.3.1.3.	<i>Exogenous linear regression model</i>	43
2.3.1.4.	<i>Evaluation of the PDFs of the coefficients of the exogenous linear regression models and of the Beta distribution shape parameter</i>	44
2.3.1.5.	<i>Evaluation of the samples of the predictive PDF of PV power</i>	46
2.3.2.	Numerical applications.....	48
2.3.2.1.	<i>Data characteristics</i>	48
2.3.2.2.	<i>Assessment of the quality of forecasts</i>	50
2.4.	A NEW PROBABILISTIC ENSEMBLE METHOD FOR SHORT-TERM PHOTOVOLTAIC POWER FORECASTING.....	57
2.4.1.	Proposed method.....	57
2.4.1.1.	<i>Relationship that links the PV active power with the solar irradiance</i>	58
2.4.1.2.	<i>Selection of probabilistic base predictors</i>	58
2.4.1.3.	<i>Processing the outputs of single probabilistic base predictor</i>	62
2.4.2.	Numerical applications.....	67
2.4.2.1.	<i>Data characteristics</i>	67
2.4.2.2.	<i>Assessment of the quality of forecasts</i>	68
2.5.	CONCLUSIONS.....	72

CHAPTER 3.

AN ADVANCED METHOD FOR SHORT-TERM INDUSTRIAL LOAD FORECASTING.....	74
3.1. INTRODUCTION.....	74
3.2. SHORT-TERM INDUSTRIAL LOAD FORECASTING METHODS: STATE OF THE ART.....	76
3.3. A NEW DETERMINISTIC REGRESSION-BASED METHOD FOR SHORT-TERM INDUSTRIAL LOAD FORECASTING.....	79
3.3.1. Proposed method.....	80
3.3.1.1. <i>Multiple Linear Regression model</i>	80
3.3.1.2. <i>Support Vector Regression model</i>	81
3.3.1.3. <i>Data characteristics</i>	83
3.3.1.4. <i>Exploratory data analysis</i>	85
3.3.1.5. <i>Model selection techniques</i>	93
3.3.2. Numerical applications.....	95
3.3.2.1. <i>Assessment of the quality of active power forecasts of the aggregated industrial load</i>	96
3.3.2.2. <i>Assessment of the quality of active power forecasts of the industrial single loads</i>	100
3.3.2.3. <i>Assessment of the quality of reactive power forecasts of the industrial loads</i>	104
3.4. CONCLUSIONS.....	106

CHAPTER 4.	
ADVANCED PROBABILITY DISTRIBUTIONS FOR MODELING EXTREME VALUES OF WIND SPEED	107
4.1. INTRODUCTION	107
4.2. MODELS FOR EXTREME VALUES OF WIND SPEED: STATE OF THE ART	109
4.3. A NEW INVERSE BURR DISTRIBUTION FOR EXTREME VALUES OF WIND SPEED.....	111
4.3.1. Analytic formulation.....	112
4.3.2. Parameter estimation procedures	114
4.4. A NEW INVERSE BURR - INVERSE WEIBULL MIXTURE DISTRIBUTION FOR EXTREME VALUES OF WIND SPEED	116
4.4.1. Analytic formulation.....	116
4.4.2. Parameter estimation procedures	118
4.5. NUMERICAL APPLICATIONS.....	121
4.5.1. Benchmark distributions for modeling EWS	121
4.5.1.1. <i>The Generalized Extreme Value distribution</i>	121
4.5.1.2. <i>The Gumbel distribution</i>	123
4.5.1.3. <i>The Inverse Weibull distribution</i>	123
4.5.2. Inverse Burr distribution.....	124
4.5.2.1. <i>Data characteristics</i>	125
4.5.2.2. <i>Assessment of Gumbel, Inverse Weibull, and Inverse Burr distributions on measured EWS data</i>	127
4.5.2.3. <i>Assessment of Inverse Burr distribution on synthetic EWS data</i>	131
4.5.3. Mixture Inverse Burr – Inverse Weibull distribution.....	133
4.5.3.1. <i>Data characteristics</i>	134
4.5.3.2. <i>Assessment of Generalized Extreme Value, Gumbel, Inverse Weibull, Inverse Burr, and mixture Inverse Burr - Inverse Weibull distributions on measured EWS data</i>	135
4.5.3.3. <i>Assessment of mixture Inverse Burr - Inverse Weibull distribution on synthetic EWS data</i>	140
4.6. CONCLUSIONS.....	141
CONCLUSIONS.....	143
APPENDIX.....	146
A.1. DETERMINISTIC INDICES FOR THE ASSESSMENT OF THE QUALITY OF FORECASTS	146
A.2. PROBABILISTIC INDICES FOR THE ASSESSMENT OF THE QUALITY OF FORECASTS	148
A.2.1. Sharpness.....	148
A.2.2. Reliability.....	149
A.2.3. Reliability diagrams	149

A.2.4. Probability integral transform histograms	151
A.2.5. Proper scores	152
A.3. GOODNESS OF FITTING INDICES	154
BIBLIOGRAPHY	155
List of publications	170

List of figures

Figure 1.1 - Features of parametric and non-parametric methods.....	11
Figure 1.2 - Typical structure of an artificial neuron.....	24
Figure 1.3 - Features of competitive and cooperative approaches.....	30
Figure 2.1 - Forecast time scales.	39
Figure 2.2 - Flow chart of the probabilistic Bayesian method.	47
Figure 2.3 - Cross-correlation coefficients between solar irradiance (a), clearness index (b) and the available weather variables.....	49
Figure 2.4 - Next day PV power forecasts obtained through the Bayesian approaches, the PM method (solid lines), and actual PV power (dash line) on (a) 21 August 2012; (b) 23 September 2012.....	51
Figure 2.5 - Next day PV power forecasts obtained through the Bayesian approaches, the PM method (solid lines), and actual PV power (dash line) on (a) 29 November 2012; (b) 22 December 2012.....	52
Figure 2.6 - Reliability diagrams of the proposed Bayesian methods. Estimated coverages (solid lines) are compared to the ideal coverages (dash line).	57
Figure 2.7 - Next-day forecasting. Estimated coverages of single predictors: Bayesian method, Markov chain method, and quantile regression method in November 2014.	69
Figure 2.8 - Next-day forecasting. Estimated coverages of the linear ensemble predictor and the probabilistic persistence method in November 2014.	69
Figure 2.9 - Next-day forecasting. PIT histograms of single predictors: Bayesian method, Markov chain method, and quantile regression method in November 2014.	70
Figure 2.10 - Next-day forecasting. PIT histograms of the linear ensemble predictor with MO and SO procedures, compared to the probabilistic persistence method in November 2014.....	71
Figure 3.1 - Aggregate active and reactive powers from May 2, 2016 to May 8, 2016.....	84
Figure 3.2 - Autocorrelation of the aggregate industrial active power.....	86
Figure 3.3 - Scatter plot of the aggregate industrial active power versus ambient temperature.	87
Figure 3.4 - Scatter plot of the aggregate industrial active power for each type of day, versus the hour of the day. The red lines indicate the mean values of observations.	88
Figure 3.5 - Scatter plots of the aggregate industrial active power in January, February, March, and April, versus the hour of the day.	89
Figure 3.6 - Scatter plots of the aggregate industrial active power in May, June, July, and August, versus the hour of the day.	90

Figure 3.7 - Scatter plots of the aggregate industrial active power in September, October, November, and December, versus the hour of the day.	90
Figure 3.8 - Scatter plots of the aggregate industrial active power for each type of day versus the aggregate industrial active power measured one hour before.	92
Figure 3.9 - Scatter plots of the aggregate industrial active power for each type of day versus the aggregate industrial active power measured one day before.	92
Figure 3.10 - Scatter plots of the aggregate industrial active power for each type of day versus the aggregate industrial active power measured one week before.	93
Figure 3.11 - Normalized Mean Absolute Errors for the aggregate active power.	98
Figure 3.12 - Normalized Root Mean Squared Errors for the aggregate active power: (a) comparison with benchmarks; (b) zoom on the models' errors.	99
Figure 3.13 - Normalized Root Mean Squared Errors for the active power of the electrical pump: (a) comparison with benchmarks; (b) zoom on the models' errors.	101
Figure 3.14 - Normalized Root Mean Squared Errors for the active power of the carpentry feeder: (a) comparison with benchmarks; (b) zoom on the models' errors.	102
Figure 3.15 - Normalized Root Mean Squared Errors for the active power of the painting machine: (a) comparison with benchmarks; (b) zoom on the models' errors.	103
Figure 3.16 - Normalized Root Mean Squared Errors for the aggregate reactive power: (a) comparison with benchmarks; (b) zoom on the models' errors.	105
Figure 4.1 - Gumbel fitting of dataset D2 through (a) MLE procedure; (b) ME procedure; (c) QE procedure.	128
Figure 4.2 - Inverse Weibull fitting of dataset D2 through (a) MLE procedure; (b) ME procedure; (c) QE procedure.	128
Figure 4.3 - Inverse Burr fitting of dataset D2 through (a) MLE procedure; (b) ME procedure; (c) QE procedure.	129
Figure 4.4 - Generalized Extreme Value (a) and Gumbel (b) fitting of dataset D13 through MLE procedure.	136
Figure 4.5 - Inverse Burr (a) and Inverse Weibull (b) fitting of dataset D13 through MLE procedure.	136
Figure 4.6 - Mixture Inverse Burr – Inverse Weibull fitting of dataset D13 through (a) MLE procedure, and (b) EM procedure.	137
Figure A.1 - Examples of reliability diagrams.	150
Figure A.2 - Examples of Probability Integral Transform histograms.	151
Figure A.3 - Graphical interpretation of the Continuous Ranked Probability Score.	152

List of tables

Table 1.1 - Utility of forecasting methods in power system operation needs	8
Table 2.1 - Spot-value error indices obtained through Bayesian method (a), Bayesian method (b), and the persistence method for the considered days	53
Table 2.2 - Spot-value error indices obtained through Bayesian method (a), Bayesian method (b), and the persistence method for the considered test set.....	54
Table 2.3 - Probabilistic error indices obtained through Bayesian method (a), Bayesian method (b), and the persistence method for the considered days.	55
Table 2.4 - Probabilistic error indices obtained through Bayesian method (a), Bayesian method (b), and the persistence method for the considered test set.	56
Table 2.5 - Next-day forecasting. Values of weight coefficients for linear ensemble in November 2014	68
Table 2.6 - Next-day forecasting. Continuous ranked probability scores and maximum deviation from perfect reliability in November 2014	68
Table 2.7 - Next-day forecasting. Continuous ranked probability scores from February to December 2014.	72
Table 2.8 - Next-day forecasting. Maximum deviation from perfect reliability from February to December 2014.	72
Table 3.1 - Statistical parameters of the analysed load time series.....	85
Table 3.2 - Predictors candidate to build models for industrial load forecasts	95
Table 4.1 - Values of Gumbel, Inverse Weibull, and Inverse Burr parameters estimated through MLE, ME, and QE procedures.	127
Table 4.2 - Kolmogorov-Smirnov test statistics and Chi-square test statistics for Gumbel, Inverse Weibull, and Inverse Burr fitting distributions. Bold italic values denote failed tests, while underlined values correspond to the best-fitting distributions for each dataset.	130
Table 4.3 - Mean Absolute Errors in the Inverse Burr parameter estimation procedures.	132
Table 4.4 - Mean Absolute percentage Errors in the Inverse Burr parameter estimation procedures.....	132
Table 4.5 - Values of Generalized Extreme Value, Gumbel, Inverse Weibull, Inverse Burr, and mixture Inverse Burr – Inverse Weibull parameters.	135
Table 4.6 - Chi-square test statistics and values of the Determination Coefficients and Adjusted Determination Coefficients for datasets D3 and D7. Bold italic values denote failed tests, while underlined values correspond to the best-fitting distributions for each dataset.	137

Table 4.7 - Chi-square test statistics and values of the Determination Coefficients and Adjusted Determination Coefficients for datasets D8 and D9. Bold italic values denote failed tests, while underlined values correspond to the best-fitting distributions for each dataset.	138
Table 4.8 - Chi-square test statistics and values of the Determination Coefficients and Adjusted Determination Coefficients for datasets D10 and D11. Bold italic values denote failed tests, while underlined values correspond to the best-fitting distributions for each dataset.	138
Table 4.9 - Chi-square test statistics and values of the Determination Coefficients and Adjusted Determination Coefficients for datasets D12 and D13. Bold italic values denote failed tests, while underlined values correspond to the best-fitting distributions for each dataset.	139
Table 4.10 - Chi-square test statistics and values of the Determination Coefficients and Adjusted Determination Coefficients for datasets D8 when random initial points are chosen. Bold italic values denote failed tests, while underlined values correspond to the best-fitting distributions for each dataset.	140
Table 4.11 - Results of the error analysis in terms of MAPEs, averaged for each parameter of the mixture Inverse Burr – Inverse Weibull distribution estimated through the EM procedure, and for each value of the weight ω_M	141

List of abbreviations

ADC	Adjusted Determination Coefficient
AMA	Annual Maxima
ANN	Artificial Neural Network
AR	AutoRegressive
ARIMAX	AutoRegressive Integrated Moving Average eXogenous
ARX	AutoRegressive eXogenous
BI	Bayesian Inference
BM	Bayesian Method
CDF	Cumulative Distribution Function
CRPS	Continuous Ranked Probability Score
CS	Chi-square
DC	Determination Coefficient
D1,...,D13	Dataset 1,...,Dataset 13
EKF	Extended Kalman Filter
ELM	Extreme Learning Machine
EnKF	Ensemble Kalman Filter
EM	Expectation-Maximization
EWS	Extreme values of wind speed
GARCH	Generalized Autoregressive Conditional Heteroscedastic
GEFCom	Global Energy Forecasting Competition
GEV	Generalized Extreme Value
GOF	Goodness Of Fitting
GP	Generalized Pareto
GU	Gumbel
I	Integrated
IB	Inverse Burr
IW	Inverse Weibull
KDE	Kernel Density Estimation
KNN	K-Nearest Neighbor
KS	Kolmogorov-Smirnov
LFT	Long Term Forecast
LPE	Linear Pool Ensemble
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MC	Markov Chain
MCMC	Markov Chain Monte Carlo
ME	Moment Estimation
MLE	Maximum Likelihood Estimation
MLR	Multiple Linear Regression
MM	Markov chain Method

MO	Multi-Objective
MTF	Medium Term Forecast
M-IB-IW	Mixture Inverse Burr - Inverse Weibull
NB	Naïve Benchmark
NMAE	Normalized Mean Absolute Error
NRMSE	Normalized Root Mean Square Error
NWP	Numeric Weather Prediction
PDF	Probability Density Function
PICP	Prediction Interval Coverage Probability
PINAW	Prediction Interval Normalized Averaged Width
PIT	Probability Integral Transform
PLF	Pinball Loss Function
PM	Persistence Method
PMA	Period Maxima
POT	Peak Over Threshold
PV	Photovoltaic
QE	Quantile Estimation
QM	Quantile regression Method
QR	Quantile Regression
QRF	Quantile Random Forest
RG	Renewable Generator
RMSE	Root Mean Squared Error
RMSPE	Root Mean Squared Percentage Error
SG	Smart Grid
SN	Seasonal naïve
SO	Single Objective
STF	Short Term Forecast
SVR	Support Vector Regression
T1	model selection Technique 1
T2	model selection Technique 2
VSTF	Very Short Term Forecast
WG	Wind Generator
μ G	Micro Grid

List of symbols

at_{h-k}	ambient temperature at the forecast start time
$\{b_i\}$	i^{th} bin of the probability space
cc_{h-k}	cloud cover at the forecast start time
\mathbf{d}	unobservable data in the expectation-maximization algorithm
d_i	i^{th} unobservable value in the expectation-maximization algorithm
e_t	white noise term at specific time
$f(\cdot)$	probability density function
$f_{BETA}(\cdot)$	probability density function of the Beta distribution used to model the hourly irradiance
$f_{GAMMA}(\cdot)$	probability density function of the Gamma distribution used to model the hourly clearness index
$f_{GEV}(\cdot)$	probability density function of the Generalized Extreme Value distribution
$f_{GU}(\cdot)$	probability density function of the Gumbel distribution
$f_{IB}(\cdot)$	probability density function of Inverse Burr distribution
$f_{IW}(\cdot)$	probability density function of Inverse Weibull distribution
$f_{M-IB-IW}(\cdot)$	probability density function of the mixture Inverse Burr - Inverse Weibull distribution
$\hat{f}_{BETA}(\cdot)$	posterior predictive distribution of the hourly irradiance
$\hat{f}_{GAMMA}(\cdot)$	predictive posterior distribution of the hourly clearness index
h	forecast time horizon
$h^{(KDE)}$	smoothing parameter in kernel density estimation
$hod_h^{(i)}$	qualitative variable representative of the i^{th} hour of the day
\mathbf{hod}_h	vector of qualitative variables representative of the hour of the day
i	counter
j	counter
k	forecast lead time
k_m	order of the moment of the probability density function
$k^{(Tu)}$	parameter of the Tukey's test
l	counter
m_1	number of inputs of a multiple linear regression model
m_2	number of inputs of an artificial neural network
m_3	number of inputs of support vector regression model
m_4	number of inputs of k-nearest neighbors model
m_5	number of inputs of quantile regression model
m_6	number of inputs of lasso regression model
\bar{m}_{GU}	median of the Gumbel distribution
\bar{m}_{IB}	median of the Inverse Burr distribution
\bar{m}_{IW}	median of the Inverse Weibull distribution
n_{ijl}	number of observed occurrences of the transition between

	consecutive states i, j, l of the Markov chain
n_{par}	number of parameters of the considered probability distribution
$p_{\{b_i\}}$	theoretical frequency of observations that lie in the i^{th} bin of the probability space
p_{h-k}	pressure at the forecast start time
$p(\cdot)$	probability
$p_d(\cdot)$	un-normalized posterior distribution
$\hat{p}(\cdot)$	estimated probability
q_E	expected value of the log-likelihood of complete data in the expectation-maximization algorithm
r_{d_h}	ratio of the diffuse radiation in hours to the diffuse radiation in a day
r_h	residual white noise of the quantile regression model
rh_{h-k}	relative humidity at the forecast start time
s	counter
t	time
$tod_h^{(i)}$	qualitative variable representative of the i^{th} type of day
\mathbf{tod}_h	vector of qualitative variables representative of the type of day
\mathbf{w}	vector of weights of the linear pool ensemble
w_i	i^{th} weight of the linear pool ensemble
$w_i^{(ANN)}$	i^{th} synaptic weight of artificial neural networks
$\hat{\mathbf{w}}$	vector of estimated weights of the linear pool ensemble
\hat{w}_i	i^{th} estimated weight of the linear pool ensemble
\mathbf{x}	vector of predictor variables
x_{t_i}	i^{th} predictor variable at generic time
\mathbf{x}_t	vector of predictor variables at generic time
\mathbf{y}	vector of variables of interest
$y_{GU}^{(\alpha_q)}$	α_q -quantile of the Gumbel distribution
$y_{IB}^{(\alpha_q)}$	α_q -quantile of the Inverse Burr distribution
$y_{IW}^{(\alpha_q)}$	α_q -quantile of the Inverse Weibull distribution
\hat{y}_h	forecast of the variable of interest at time h
y_{ref}	reference value for normalized error indices
y_t	variable of interest at generic time
\hat{y}_t	forecast of the variable of interest at generic time
$\hat{y}_h^{(\alpha_q)}$	estimated α_q -quantile
$\bar{y}^{(\alpha_q)}$	sample α_q -quantile
y_h^*	actual value of the variable of interest at time h
\mathbf{z}	complete data in the expectation-maximization algorithm
\mathbf{A}	Markov chain transition matrix
\mathbf{A}^*	fuzzy set
$\hat{\mathbf{A}}$	estimation of the Markov chain transition matrix

B	backward shift operator
$B(\cdot; \cdot)$	Beta function
C	penalty coefficient of support vector regression
$CRPS_h$	hourly continuous ranked probability score
CS_{TS}	test statistics of the chi-square test
$E(\cdot)$	expected value
$F(\cdot)$	cumulative distribution function
$F_{GEV}(\cdot)$	cumulative distribution function of the Generalized Extreme Value distribution
$F_{GU}(\cdot)$	cumulative distribution function of the Gumbel distribution
$F_{IB}(\cdot)$	cumulative distribution function of Inverse Burr distribution
$F_{IW}(\cdot)$	cumulative distribution function of Inverse Weibull distribution
$F_{LPE}(\cdot)$	cumulative distribution function of the linear pool ensemble
$F_{M-IB-IW}(\cdot)$	cumulative distribution function of the mixture Inverse Burr - Inverse Weibull distribution
$F_{h_i}(\cdot)$	i^{th} cumulative distribution function combined in the linear pool ensemble
$\hat{F}(\cdot)$	estimated cumulative distribution function
G_{β_h}	hourly irradiance on a surfaced inclined by β degrees
$G_{\beta_h}^{(max)}$	upper bound of the observed hourly irradiances on a surface inclined by β degrees
$H(\cdot)$	Heaviside function
H_{0h}	extra-terrestrial total solar radiation
J	total number of considered quantiles
K_{dh}	diffuse fraction of total solar hourly radiation on a horizontal surface
K_h	hourly clearness index
$K_h^{(max)}$	upper bound of the observed hourly clearness indices
$K(\cdot)$	kernel function
KS_{TS}	test statistics of the Kolmogorov-Smirnov test
L_g	Lagrange function in support vector regression
L_{IB}	log-likelihood of the Inverse Burr distribution
$L_{M-IB-IW}$	log-likelihood of the mixture Inverse Burr - Inverse Weibull distribution
$\mathcal{L}_{M-IB-IW}$	likelihood of the mixture Inverse Burr - Inverse Weibull distribution
M	dimension of a time series
\mathcal{M}	set of vectors of predictor variables of a regression model
N_{bp}	number of base predictors of the ensemble method
N_B	number of bins of the probability space
$N_{\{b_i\}}$	number observations that lie in the i^{th} bin of the probability space

$\bar{N}_{\{b\}}$	mean value of observations that lie in each bin of the probability space
N_S	number of states in Markov chain frameworks
\mathcal{N}_t	subset of vectors of predictor variables of a regression model in k-nearest neighbors approaches
N_{tf}	number of performed forecasts
$N_{tr}^{(BM)}$	dimension of the training set of a Bayesian model
$N_{tr}^{(Exp)}$	dimension of the training set of an exponential smoothing model
$N_{tr}^{(G)}$	dimension of the training set of the Bayesian model for the hourly solar irradiance
$N_{tr}^{(GOF)}$	dimension of the set to evaluate the goodness of fitting of distribution
$N_{tr}^{(IB)}$	dimension of the training set to estimate parameters of the Inverse Burr distribution
$N_{tr}^{(K)}$	dimension of the training set of the Bayesian model for the hourly clearness index
$N_{tr}^{(KDE)}$	dimension of the training set of a kernel density estimation model
$N_{tr}^{(KNN)}$	dimension of the training set of a k-nearest neighbors model
$N_{tr}^{(LAS)}$	dimension of the training set of a lasso regression model
$N_{tr}^{(MLR)}$	dimension of the training set of a multiple linear regression model
$N_{tr}^{(M-IB-IW)}$	dimension of the training set to estimate parameters of the mixture Inverse Burr - Inverse Weibull distribution
$N_{tr}^{(NB)}$	dimension of the training set of a naïve benchmark model
$N_{tr}^{(Nie)}$	dimension of the training set of a Nielsen model
$N_{tr}^{(Ppm)}$	dimension of the training set of a probabilistic extension of the persistence approach
$N_{tr}^{(QM)}$	dimension of the training set of the quantile regression method
$N_{tr}^{(SVR)}$	dimension of the training set of a support vector regression model
$N(\cdot, \cdot)$	Gaussian distribution
P_h	active power at time h
$PLF_i^{(\alpha_q)}$	contribution to the pinball loss function of the α_q -quantile at time i
P_{PV_h}	photovoltaic power at the time horizon
P_{PV_r}	rated power of the photovoltaic generator
$P_{PV_h}^{(\alpha_q)}$	α_q -quantile of photovoltaic power
$P_{PV_h}^*$	actual value of photovoltaic power
\hat{P}_h	estimated active power at time h
$\hat{P}_{PV_{LPE_h}}^{(\alpha_q)}$	estimated α_q -quantile of the linear pool ensemble of photovoltaic power

$\hat{P}_h^{(PM)}$	persistence active power forecast
$\hat{P}_h^{(SN)}$	seasonal naïve active power forecast
R_{bh}	ratio of the beam radiation on a tilted surface to the beam radiation on a horizontal surface
R_{ADC}^2	adjusted determination coefficient
R_{DC}^2	determination coefficient
S_{PV}	surface area of the photovoltaic array
$S_t^{(i)}$	i^{th} state in which the variable of interest lies at specific time
S_{AT_h}	vector of ambient temperatures
S_{CC_h}	vector of cloud cover
$S_{G\beta_h}$	vector of hourly solar irradiances
S_{K_h}	vector of hourly clearness indices
S_{P_h}	vector of pressures
S_{RH_h}	vector of relative humidity
T	seasonal period
Z	universe of discourse
α_q	quantile level
$\alpha^{(Exp)}$	smoothing coefficient of exponential smoothing models
$\alpha_k^{(Nie)}$	coefficient of Nielsen
$\hat{\alpha}_q$	estimated quantile level
$\hat{\alpha}_{qLPE}$	estimated quantile level of the linear pool ensemble
β	degrees of inclination of photovoltaic modules
$\beta_i^{(G)}$	i^{th} parameter of the hourly solar irradiance regression model
$\beta_i^{(K)}$	i^{th} parameter of the hourly clearness index regression model
$\beta_i^{(LAS)}$	i^{th} parameter of a lasso regression model
$\beta_i^{(MLR)}$	i^{th} parameter of a multiple linear regression model
$\beta_i^{(SVR)}$	i^{th} parameter of a support vector regression model
$\hat{\beta}_i^{(G)}$	estimated sample of the i^{th} parameter of the hourly solar irradiance regression model
$\hat{\beta}_i^{(K)}$	estimated sample of the i^{th} parameter of the hourly clearness index regression model
$\hat{\beta}_i^{(LAS)}$	i^{th} estimated parameter of a lasso regression model
$\hat{\beta}_i^{(MLR)}$	i^{th} estimated parameter of a multiple linear regression model
$\hat{\beta}_i^{(SVR)}$	i^{th} estimated parameter of a support vector regression model
$\beta^{(LAS)}$	vector of parameters of the lasso regression model
$\beta^{(MLR)}$	vector of parameters of the multiple linear regression model
$\beta^{(SVR)}$	vector of parameters of the support vector regression model
$\beta^{(\alpha_q)}$	vector of parameters of the quantile regression model

$\widehat{\beta}^{(LAS)}$	vector of estimated parameters of the lasso regression model
$\widehat{\beta}^{(MLR)}$	vector of estimated parameters of the multiple linear regression model
$\widehat{\beta}^{(SVR)}$	vector of estimated parameters of the support vector regression model
$\widehat{\beta}^{(\alpha_q)}$	vector of estimated parameters of the quantile regression model
γ	shape parameter of the Inverse Burr distribution
$\bar{\gamma}_{Eu}$	Euler-Mascheroni constant
γ_{GAMMA_h}	parameter of the modified Gamma distribution
δ	shape parameter of the Inverse Weibull distribution
ε	error band of the support vector regression
ε_t	error term of a multiple linear regression model
ζ	shape parameter of the Inverse Burr distribution
η_i	i^{th} Lagrange variable in support vector regression
η_i^*	i^{th} Lagrange variable in support vector regression
η_{PV}	global efficiency of the photovoltaic array
$\theta(B)$	moving average operator
θ	vector of parameters in Bayesian frameworks
ι	scale parameter of the Generalized Extreme Value distribution
κ	shape parameter of the Generalized Extreme Value distribution
λ_{GAMMA_h}	parameter of the modified Gamma distribution
λ	vector of Lagrange variables in support vector regression
λ_i	i^{th} Lagrange variable in support vector regression
$\lambda^{(LAS)}$	degree of regularization of lasso regression
λ^*	vector of Lagrange variables in support vector regression
λ_i^*	i^{th} Lagrange variable in support vector regression
$\hat{\lambda}$	vector of estimated Lagrange variables
$\hat{\lambda}^*$	vector of estimated Lagrange variables
μ_{BETA_h}	mean value of the Beta distribution
μ_{GAMMA_h}	mean value of the modified Gamma distribution
μ_{GEV}	mean value of the Generalized Extreme Value distribution
μ_{GU}	mean value of the Gumbel distribution
μ_{IB}	mean value of the Inverse Burr distribution
μ_{IW}	mean value of the Inverse Weibull distribution
$\mu_{M-IB-IW}$	mean value of the mixture Inverse Burr - Inverse Weibull distribution
$\mu_{N_{tr}^{(G)}}$	mean value of the last measured $N_{tr}^{(G)}$ observed hourly solar irradiances
$\mu_{N_{tr}^{(K)}}$	mean value of the last measured $N_{tr}^{(K)}$ observed hourly clearness indices
$\mu^{A^*}(z^{(fuz)})$	grade of membership in fuzzy approaches

$\hat{\mu}_{BETA_h}$	estimated mean value of the Beta distribution
$\hat{\mu}_{GAMMA_h}$	estimated mean value of the modified Gamma distribution
ν	scale parameter of the Inverse Weibull distribution
ξ_t	augmentation coefficient of support vector regression at generic time
ξ_t^*	augmentation coefficient of support vector regression at generic time
ξ	vector of augmentation coefficients of support vector regression
ξ^*	vector of augmentation coefficients of support vector regression
$\xi_1^{(MO)}$	first weight of the multi-objective optimization problem
$\xi_2^{(MO)}$	second weight of the multi-objective optimization problem
$\hat{\xi}$	vector of estimated augmentation coefficients of support vector regression
$\hat{\xi}^*$	vector of estimated augmentation coefficients of support vector regression
o	shape parameter of the Gumbel distribution
$\pi_t^{(i)}$	probability of the variable of interest to lie in the i^{th} state at time h
$\boldsymbol{\pi}_h$	vector of state probabilities of the variable of interest at time h
ρ	scale parameter of the Inverse Burr distribution
$\zeta_t^{(Ppm)}$	error term of the probabilistic persistence approach
σ_{BETA_h}	shape parameter of the Beta distribution
$\hat{\sigma}_{BETA_h}$	estimated sample of the shape parameter of the Beta distribution
σ_{GEV}^2	variance of the Generalized Extreme Value distribution
σ_{GU}^2	variance of the Gumbel distribution
σ_{IB}^2	variance of the Inverse Burr distribution
σ_{IW}^2	variance of the Inverse Weibull distribution
$\sigma_{M-IB-IW}^2$	variance of the mixture Inverse Burr - Inverse Weibull distribution
$\sigma_t^{(Ppm)}$	standard deviation of the probabilistic persistence approach
$\bar{\tau}^{(ANN)}$	threshold of artificial neural networks
$\bar{\tau}^{(EM)}$	threshold of the expectation-maximization algorithm
$\bar{\tau}^{(SVR)}$	threshold of support vector regression
v	location parameter of the Generalized Extreme Value distribution
φ_{BETA_h}	shape parameter of the Beta distribution
χ	location parameter of the Gumbel distribution
ψ	maximum value of absolute deviations between estimated and actual coverages
ω_M	weight parameter of the mixture Inverse Burr - Inverse Weibull distribution
$\hat{\omega}_M^{(EM)}$	expectation-maximization estimation of the weight of the mixture Inverse Burr - Inverse Weibull distribution
$\Gamma(\cdot)$	Gamma function

θ	parameters estimated in the expectation-maximization algorithm
$\hat{\theta}_0$	initial set of parameters of the expectation-maximization algorithm
$\hat{\Lambda}_h^{(\alpha_q)}$	binary indicator of a generic estimated α_q -quantile
$\hat{\Lambda}_{LPE_h}^{(\alpha_q)}$	binary indicator of the estimated α_q -quantile of the linear pool ensemble
ϕ_1	first objective function to be minimized in the multi-objective ensemble method, i.e., the average value of hourly continuous ranked probability scores
ϕ_2	second objective function to be minimized in the multi-objective ensemble method, i.e., the maximum value of all absolute deviations between estimated and actual coverages
$\Phi(B)$	stationary AR operator
$\Psi(B)$	input operator
∇^d	backward difference operator

Acknowledgment

After these three years, I recognize that pursuing a Ph.D. was the best choice I ever made in my entire life. But when you reach such a great achievement, enjoying every moment of it, the least you can do is to express your gratitude to everyone who made it possible.

My tutor, prof. Guido Carpinelli, is more than an academic advisor. His dedication to research and teaching is exemplary, and his competence and passion towards every field of application in electrical engineering is an incitement for everyone. It is so blessed to work with him. His attitude of respect towards collaborators, professors, researchers, and students, should be a reference point for everybody who aims to pursue an academic career.

My co-tutor, dr. Antonio Bracale, is a role model in terms of proficiency and enthusiasm towards the academic research. His theoretical and technical skills are extraordinary and he dedicated himself to improve my skills too, through his daily support in my Ph.D. activities.

Although not officially, prof. Pierluigi Caramia was essentially my third tutor. I don't remember a single circumstance in which he was not able to support or help me. His positive and constructive attitude towards me was an added value during these years.

Guido, Antonio, and Pierluigi made me feel a member of the team, a member of the family.

I thank my coordinator, prof. Daniele Riccio; he played his institutional role with an uncommon willingness towards me and my colleagues, always understanding our needs. I am also grateful to prof. Mario Russo, prof. Filippo Spertino, and prof. Paola Verde, for critically reviewing my Ph.D. thesis. Their useful suggestions and advices allowed me to significantly improve my dissertation.

I have sincerely appreciated the opportunity to work with several excellent academic researchers during my Ph.D. course. I thank dr. Tao Hong, dr. Andrea Michiorri, and prof. Renato Rizzo for their priceless contributions in developing my skills and in my personal growth. I have really enjoyed working with prof. Mario Pagano and prof. Angela Russo; their expertise exactly matches their kindness.

Prof. Elio Chiodo deserves a special mention: he acts not only as a scientific partner, but also as a friend.

My daily experience would not have been so complete without prof. Amedeo Andreotti, prof. Davide Lauria, prof. Daniela Proto, and prof. Pietro Varilone; they are people of absolute value, and outstanding professors. I established a genuine friendship with dr. Luigi Pio Di Noia and dr. Fabio Mottola: two nice guys who share an extraordinary research versatility.

I could have not fully dedicate myself in pursuing my Ph.D. without the unconditional support of my parents, Marisa and Vincenzo. I hope to reward them with my love and with my accomplishments. I want to express my gratitude also to my “brothers” Ivan and Carmine, for the time spent together as ka-tet.

My greatest “thank you” definitely goes to Luisa, my font of happiness during these years and forevermore (and a bit more). She is the one who helps me, teaches me, questions me, loves me, but most importantly she gives new meanings to my life every day.

INTRODUCTION

In the new liberalized markets a multitude of operators is allowed to interact with transmission and distribution networks, both by purchasing or selling energy. In this context, renewable generators supply a significant share of the total electricity demand, saving thousands of tons of CO₂ and polluting agents from being released in the atmosphere each day. The old paradigm of static networks is going to be surpassed by intelligent structures, with a widespread diffusion of distributed generation, information technologies, and control devices that foster the optimal exploitation of energy resources.

These are great achievements in pursuing social wellness and technology advance, with care to the environment. However, the path to smarter electrical grids is very intricate, and many complications that were unimaginable two decades ago are now part of the daily routine of transmission and distribution grid managers.

One of the biggest problem is facing with forecasting.

Until '90s, the main topic of forecasting was, by far, the load consumption at aggregate national, regional or sub-station levels. Load forecasts were required for short-term and long-term scenarios, in order to respectively i) assure power balancing and ii) plan future expansions of the networks in decennial schedules.

Nowadays, many other variables (i.e., renewable generation powers, extreme values of weather variables, loads at disaggregate levels, and energy prices) are the subject of forecasting.

The spread of renewable generators has, in fact, extended the problem of forecasting also to the generation powers and extreme values of weather variables, such as wind speeds.

Renewable generation power forecasting, with particular focus on wind and photovoltaic (PV) systems, is yet to be fully explored in relevant literature. Indeed, forecasting systems found in literature for these variables are deterministic tools in the majority of cases: they

only provide a spot-value as forecast. The need of probabilistic tools, that are able to catch also information on the uncertainty linked to the wind and solar energy sources, has only recently been recognized by electrical practitioners, and researches in this framework are well encouraged since 2010s.

Extreme values of wind speed (EWS) play a key role in wind power forecasting, in the overhead line rating, and in the assessment of the mechanical reliability of system components; in such kinds of applications, risk analyses are encouraged and probabilistic tools are needed. An accurate statistical characterization of such events is mandatory to improve the quality of several EWS probabilistic forecasting tools, since they usually require the definition of an appropriate Probability Density Function (PDF) to perform with high accuracy.

Another new forecasting issue is the need of forecasts for loads at disaggregate levels, and of forecasts of both active and reactive powers. It is fostered by the growing ability of new management and regulation tools to push the optimal exploitation of energy sources in smart power systems. Having active and reactive power profiles available in advance is mandatory in order to perform the correct scheduling to manage grids at consumers level (e.g., micro-grids).

Eventually, also energy price forecasts are currently required for the convenient participation of operators to electric markets.

In summary, forecasting in power systems is a wide topic that today covers many and many needs, and that requires further research efforts.

In this wide and complex context, after a brief discussion on the classification of forecasting systems and on the methods that are currently available in literature for forecasting electrical variables, stressing pros and cons of each approach, the thesis provides four contributions to the state of the art on forecasting in power systems where literature is somehow weak.

The first contribution is a *Bayesian-based probabilistic method* to forecast PV power in short-term scenarios. The method transforms probabilistic forecasts of the hourly solar irradiance (or the hourly clearness index) into probabilistic forecasts of the PV power by means of well-known relationships, in an indirect approach. Solar irradiance

(or hourly clearness index) is modeled by means of an analytic PDF, whose parameters are estimated by means of the Bayesian inference of past available observations. An exogenous linear regression model is also defined in order to link one of the PDF parameters to the measurements of some influencing weather variables.

The second contribution is a *probabilistic competitive ensemble method* once again to forecast PV power in short-term scenarios. The idea is to improve the quality of forecasts obtained by means of some individual probabilistic predictors, by combining them in a probabilistic competitive approach. Since the probabilistic predictors may vary in terms of predictive outputs (e.g., they can provide predictive samples, predictive PDFs, or predictive quantiles), in the proposed ensemble method the forecasts obtained through base predictors are firstly properly combined through a linear pooling of predictive cumulative density functions. Then, in order to guarantee elevate sharpness and reliability characteristics, a multi-objective (MO) optimization method is proposed and applied during the training period in order to estimate coefficients of the linear pooling. The MO optimization is specifically devoted to overcome well-known problems resulting in the over-dispersion of forecasts coming from the probabilistic combination of probabilistic base predictors in the linear pool approach. The Bayesian method (i.e., the first proposed contribution), a Markov chain method, and a quantile regression method are selected as probabilistic base predictors to be merged.

The third contribution is aimed to the development of a *deterministic industrial load forecasting method* suitable in short-term scenarios, at both aggregated and single-load levels, and for both active and reactive powers. The *deterministic industrial load forecasting method* is based on Multiple Linear Regression (MLR) or Support Vector Regression (SVR) models. The selection of most adequate models is performed with two different techniques. The first technique is based on the 10-fold cross-validation of multiple MLR and SVR models that contain combinations of the informative inputs; the best MLR and the best SVR models (in terms of average errors) are selected for the test step. The second technique is instead based on the lasso analysis, in order to directly draw the most useful inputs among the informative ones; a 10-

fold cross-validation is performed also in this case, in order to provide coherent comparison with the first technique.

The fourth contribution provides advanced PDFs for the statistical characterization of EWS.

In particular, one of the PDFs proposed in this Thesis is an *Inverse Burr distribution for EWS modeling*. The derivation process of the Inverse Burr distribution is discussed, and a rigorous parameter estimation procedure based on the quantile estimation is provided and compared to classical maximum likelihood estimation and moment estimation procedures. In some conditions, the quantile estimation procedure consists in solving an analytic equation, thus avoiding the well-known convergence problems of classical estimation procedures.

The other PDF proposed in this Thesis is a *mixture Inverse Burr – Inverse Weibull distribution for EWS modeling*. The mixture of an Inverse Burr and an Inverse Weibull distribution allows to increase the versatility of the tool, although increasing the number of parameters to be estimated. This complicates the parameter estimation process, since traditional techniques such as the maximum likelihood estimation suffer from convergence problems. Therefore, an Expectation-Maximization (EM) procedure is specifically developed for the parameter estimation. The aim of the EM procedure is still to maximize the likelihood of an observed EWS dataset, although hypothesizing additional, hidden parameters to simplify the formulation of the likelihood function.

This thesis is organized in four Chapters and an Appendix.

The first Chapter provides an overview of the classification of forecasting methods in power systems based on the needs of electrical operators, and a brief explanation of the main methods available in relevant literature.

The second Chapter explores in details the state of the art on probabilistic PV power forecasting, and shows the two contributions of this thesis in that field.

The state of the art on industrial load forecasting is presented in details in the third Chapter; the related contribution (i.e., the *deterministic industrial load forecasting method*) is also presented in this Chapter.

The fourth Chapter deals with EWS modelling; after discussions on the state of the art, the two proposed models and their corresponding parameter estimation procedures are presented.

Numerical applications related to each of the proposed contributions are shown in the ending parts of second, third and fourth Chapters; actual data are used in all of the numerical applications, in order to effectively test the validity of the proposals in real-world scenarios.

Eventually, the main error indices and tools for the assessment of forecasts in both deterministic and probabilistic frameworks, and for the assessment of the Goodness Of Fitting (GOF) distributions, are shown in the Appendix.

Chapter 1.

FORECASTING AND POWER SYSTEMS

1.1. INTRODUCTION

The planning, operation and management of power systems are strongly affected by weather conditions, social factors, and economic factors [1-5].

Solar irradiance, wind speed and ambient temperature, in fact, are weather conditions that strongly affect the production of PV and wind (WGs) generators, and the ampacity of overhead lines [6,7]. EWS may seriously damage electrical installations such as wind towers, WG blades, overhead lines and trellis [3,8-10]. Ambient temperature also influences energy demand due to the spread of cooling and heating systems [11,12], and transformer loadability due the variations in thermal exchange [13].

Social and economic factors modify the human attitude toward energy consumption, as wealthier societies tend to consume more energy, and they also affect the supply from renewable sources, as more generators are typically installed in periods of high incentives [5,14]. Price variations on energy markets play a key role in the optimal management of transmission and distribution systems [15,16].

The increasing complexity of electrical networks makes the whole power system more vulnerable to the abovementioned factors and, then, power system operators would appreciate to perfectly prior know the future status of the grids, in order to plan and perform their actions with no miscalculations or approximations [17]. However, this is not feasible.

The above problems, in addition, will definitely grow in interest in next years due to the continuous development of Smart Grids (SGs) and Micro Grids (μ Gs): adequate criteria of management and planning of transmission and distribution networks should be developed in these

new frameworks, thus requiring more accurate and reliable forecasting methods to be applied in power systems. In particular, forecasting non-controllable generation, loads and market prices is therefore mandatory in order to help the future decision-makers to optimally exploit energy sources, assuring the balance and stability on networks, and favoring the risk assessment in reliability and maintenance tasks [18-20].

The forecasting methods applied in power systems are briefly summarized in this Chapter. First, their typical classifications are provided in Section 1.2. Then, forecasting approaches that have been widely applied in relevant literature to power systems are discussed in Section 1.3.

1.2. CLASSIFICATIONS OF FORECASTING METHODS

The diversity in forecasting needs has a direct, intuitive consequence: no forecasting method is universally able to fit any purpose, but it has to be selected case by case on the basis of particular needs [21]. The classifications of forecasting methods straightforwardly follows the diversity in terms of end user needs.

The first classification is made in terms of forecast lead time. Indeed, actions on power systems are performed on different time lines: e.g., improvement, replacement or realization of new infrastructures are planned several years before, while optimal management of distributed energy resources in SGs and μ Gs is scheduled some minutes to some hours before [1,22].

Few papers [23] classify forecasting methods in 2 categories (short-term and long-term); however, the most complete practice is to individuate *Very Short-Term Forecasting* (VSTF), *Short-Term Forecasting* (STF), *Medium-Term Forecasting* (MTF), and *Long-Term Forecasting* (LTF) methods [24-28].

VSTF lead times range up to 24 hours¹; they are usually involved in power balancing and system optimal management and control. The influence of external variables (e.g., ambient temperature for load

¹ There is a lack of standardized classification. Therefore, some papers refer to VSTF when lead times are up to 30 minutes; in this case, STF lead times range from 30 minutes to 24-48 hours ahead.

forecasting) is limited in this kind of applications, and therefore is often overlooked.

STF lead times range from 24 hours ahead to two weeks ahead; they are usually involved in power balancing for acquiring appropriate reserve, market participation, and system optimal management.

MTF lead times range from 2 weeks to 3 year ahead; this wide interval of time makes MTF methods useful for market participation, system optimal management, and planning. Social and economic factors should be carefully investigated in MTF, especially for monthly and yearly scenarios.

LTF lead times start from 3 years and reach 20 (or more) years. These forecasts are involved in power system planning, and weather, social and economic long-term evaluations are mandatory in order to cope with evolutionary trends.

Table 1.1 associates forecasting methods, classified in terms of lead times, to corresponding needs [28].

Table 1.1 - Utility of forecasting methods in power system operation needs

Classification of forecasting methods	Power system operation need			
	Power balancing	Participation to electrical markets	Optimal management and control	Planning
VSTF	yes	no	yes	no
STF	yes	yes	yes	no
MTF	no	yes	yes	yes
LTF	no	no	no	yes

A second classification involves the output of forecasting methods. This comes from the different risks linked to power system tasks that require forecasts to be completed.

Let's think of a wind plant owner, who wants to sell energy on electrical markets [29]. He has to submit a selling offer, stating the (exact) amount of energy he will be able to produce; in several Countries, he is penalized if the resulting production is too far from the declared one. If he disposes of a forecasting method that provides only a single value of wind power as output, the plant owner has no other choice than submitting a selling offer of as much energy as the forecasted one. Instead, if he disposes of a forecasting tool that

provides more values, or the probability distribution of wind powers, he can manage the forecasts and make the best choice for his needs.

In this context, *deterministic* forecasts provide as output only a single value of the variable of interest (point forecast). *Probabilistic* forecasts provide as output analytical distributions such as PDFs, Cumulative Density Functions (CDFs), sampled distributions (discrete probabilities), quantiles², or moments of the predictive distribution (e.g., mean, variance and skewness) [30]. Note that the variable of interest is still treated as a random entity in both frameworks: the main difference is that a single value is given as forecast of the variable of interest in deterministic framework, while more values, or a function, are given as forecast of the variable of interest in probabilistic framework.

Probabilistic forecasts are generally preferable, since they provide also information about the uncertainty linked to the forecast itself. Therefore, they allow the risk assessment and the optimal selection of a single value, on the basis of different frameworks [31,32]. Indeed, it is always possible to extract a single, spot-value (e.g., the mean value of the predictive distribution) from probabilistic forecasts, while the reciprocal is obviously not valid. The main drawbacks of probabilistic forecasts are the increase of method complexity, and their greater computational burden. Then, if the forecast end user gains no benefit in having a probabilistic forecast, deterministic methods are still the best choice.

It is worth noting that probabilistic methods sometimes rely on an underlying deterministic method [33,34]; e.g., some parameters of the predictive probabilistic distribution could be set from the output value of a deterministic method. In this case, improving the performance of the underlying deterministic method is compulsory in order to increase the overall quality of the probabilistic forecasts. Thus, research efforts in the deterministic framework are always encouraged.

A third classification of forecasting methods is based on the characteristics of models involved in the forecasting method, and

² Let's recall that the α_q -quantile of a probability distribution is the value of the variable of interest that is not exceeded with probability α_q , $0 \leq \alpha_q \leq 1$.

consequently on the solving procedure. The common classification is in terms of *parametric* and *non-parametric* methods.

Parametric methods are based on models that are univocally identified as several numerical parameters are known; e.g., a predictive analytical Gaussian distribution is univocally identified when its mean and variance are known. Therefore, solving a parametric forecasting method consists in finding estimations of unknown parameters, usually by minimizing or maximizing assigned objective functions (i.e., by minimizing an error index). In the particular case of parametric probabilistic methods, usually the problem of finding a prior probabilistic characterization of the variable of interest through a specific PDF has to be solved [33].

Non-parametric methods, instead, rely on the idea that forecasting future dynamics can be achieved by analogy with past dynamics. Indeed, the variable under study is not assessed through an analytic model, instead it is forecasted by means of a procedure that “learns” from the past. Note that the “non-parametric” definition could be misleading. It does not mean that no parameters are involved in non-parametric methods; indeed, some involved parameters could identify the order of the model, rather than the model itself.

In non-parametric methods, however, the complexity of the models (i.e., the number of parameters) grows with the dimension of the problem and, theoretically, is not constrained. The more the inputs (i.e., the elements of the training set) fed to the non-parametric method, the larger is the number of parameters to be estimated. Therefore, the structure of the model itself “grows” as the training set enlarges (Fig. 1.1). On the other hand, the structure of models in parametric methods is fixed with the dimension of the problem; the same number of parameters has to be estimated, regardless of the size of the training set³ (Fig. 1.1).

³ Training set stores all of the available input data used to estimate model parameters. Since the actual values of the variable of interest are known, forecasts are produced in-sample during the training period. A test period is therefore necessary to fairly evaluate the performance of the so-built forecasting model, in an out-of-sample context.

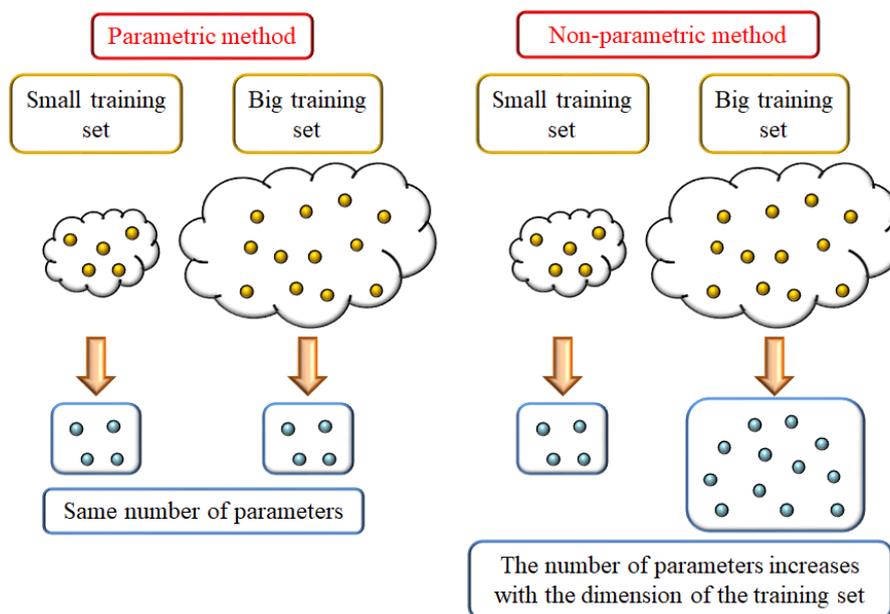


Figure 1.1 – Features of parametric and non-parametric methods.

The fourth and last classification is based on the approach used to build and solve the forecasting problem. *Statistical* approaches rely on measurement data acquired in the past to produce forecasts for the future, starting from the assumption that past conditions are informative for the future. *Physical* approaches rely instead on a specific formulation of the problem under study in a rigorous fashion, by exploiting mathematical formulas based on physical principles. *Hybrid* approaches can be a combination of statistical approaches, physical approaches, or statistical and physical approaches together.

1.3. FORECASTING METHODS APPLIED TO POWER SYSTEMS

A review of common forecasting methods applied to power systems is presented in this Section. The latter classification of Section 1.2, i.e., statistical approaches, physical approaches, and hybrid approaches, is considered in the following. The main features of each approach are briefly discussed in the corresponding sub-Sections.

1.3.1. Statistical approaches

Statistical approaches exploit information provided by the observed past history to describe and forecast the realization of a physical phenomenon.

In several kinds of applications, a mathematical model is formulated to characterize the variable of interest. If the past were able to exactly describe its behavior (e.g., if the mathematical model can be formulated from well-known physical laws), the model would be purely deterministic, and forecasting future values of the variables simply consists in solving an equation from assigned input data. However, in the majority of cases, there are many random events that could affect the realization of the phenomenon, and, therefore, models can deviate from classical physical laws, since they could fail to exactly predict the future conditions. Thus, models can be built on the basis of previous experiences in order to take into account the effect of external random events on the objective variable.

In other applications, no strict mathematical model is needed in order to forecast the future values of the variable of interest. This is the case, of machine learning approaches, as the artificial intelligence of computers “learns” from the past and tries to find the underlying relationships between inputs and outputs. Due their versatility, machine learning approaches are suitable for both deterministic or probabilistic forecasting.

1.3.1.1. Naïve approaches

Naïve approaches are techniques that allow to produce forecasts that can surprisingly be accurate in VSTF and STF frameworks, despite their superior simplicity. For this reason, they are usually adopted as benchmarks, to give a comparative reference for assessing the quality of forecasts coming from more complex models.

The first notable naïve approach is based on the persistence of the variable of interest. In other words, the variable of interest y is expected not to vary during the forecast lead time k , thus assuming the forecast \hat{y}_t at the time horizon t the same value experienced at the forecast start time $t - k$:

$$\hat{y}_t = y_{t-k} . \tag{1.1}$$

Note that, to take into account the eventual seasonal behavior of the variable of interest, a seasonal naïve (SN) benchmark, direct extension of the persistence approach, can be expressed as follows:

$$\hat{y}_t = y_{t-T}, \quad (1.2)$$

where T is the seasonal period (e.g., daily or weekly).

Both the abovementioned naïve benchmarks use only one past measured value of the variable. This could result in an over-sensitive approach, since it directly depends on a single value that could lose generality. The average naïve benchmark (NB) overcomes this problem by providing a forecast as an average of $N_{tr}^{(NB)}$ past measured values $y_{t-k-(N_{tr}^{(NB)}-1)}, y_{t-k-(N_{tr}^{(NB)}-2)}, \dots, y_{t-k}$ of the variable of interest, as follows:

$$\hat{y}_t = \frac{1}{N_{tr}^{(NB)}} \sum_{i=0}^{N_{tr}^{(NB)}-1} y_{t-k-i}. \quad (1.3)$$

Obviously, the average naïve benchmark could suffer from a dual problem: it could over-generalize the behavior of the variable of interest, losing valuable information linked to recency.

An interesting compromise between the persistence and the average naïve models was proposed by Nielsen et al. in [35]. In the Nielsen approach, the forecast \hat{y}_t is obtained as a weighted sum of the last $N_{tr}^{(Nie)}$ available observation and of the average of past observations:

$$\hat{y}_t = \alpha_k^{(Nie)} y_{t-k} + (1 - \alpha_k^{(Nie)}) \frac{1}{N_{tr}^{(Nie)}} \sum_{i=0}^{N_{tr}^{(Nie)}-1} y_{t-k-i}, \quad (1.4)$$

where the Nielsen weight coefficient $\alpha_k^{(Nie)}$ is a function of the lead time k , and it is analytically computed from the correlation between lagged observations as:

$$\alpha_k^{(Nie)} = \frac{\sum_{i=0}^{N_{tr}^{(Nie)}-1-k} \hat{y}_{t-k-i} \cdot \hat{y}_{t-2k-i}}{\sum_{i=0}^{N_{tr}^{(Nie)}-1-k} \hat{y}_{t-2k-i}^2}, \quad (1.5)$$

where $\tilde{y}_j = y_j - \frac{1}{N_{tr}^{(Nie)}} \sum_{i=0}^{N_{tr}^{(Nie)}-1} y_{t-k-i}$.

As said before, naïve benchmarks are often used as deterministic benchmarks due their intuitiveness and simplicity. Extensions of all of these naïve approaches to probabilistic frameworks exist, and have been extensively applied in relevant literature to build also probabilistic benchmarks.

In the probabilistic persistence approach, e.g., it is common to introduce a probabilistic error term $\zeta_t^{(Ppm)}$ from a Gaussian distribution with zero mean, and standard deviation $\sigma_t^{(Ppm)}$ equal to the variance of $N_{tr}^{(Ppm)}$ past errors:

$$\sigma_t^{(Ppm)} = \sqrt{\frac{1}{N_{tr}^{(Ppm)}-k} \sum_{i=0}^{N_{tr}^{(Ppm)}-1-k} (y_{t-k-i} - \hat{y}_{t-k-i})^2}. \quad (1.6)$$

Thus, from the probabilistic model:

$$\hat{y}_t = y_{t-k} + \zeta_t^{(Ppm)}, \quad (1.7)$$

it is possible to extract predictive samples or to evaluate predictive quantiles of the variable of interest. Note that also the predictive PDF could be built in an analogous way, e.g., assuming is a Gaussian distribution with mean y_{t-k} and standard deviation $\sigma_t^{(Ppm)}$.

The same approach could be adapted also to other distribution families, mirroring the prior assumption on the distribution family in the model that is to be compared with the naïve benchmarks. The assumption on the distribution family is, therefore, a key point.

Naïve approaches have been extensively applied in literature relevant to power systems, mostly as deterministic and probabilistic reference methods; relevant papers are recalled in surveys on load forecasting [6,24,28,36-44], PV forecasting [27,36,45-54], WG forecasting [30,47,48,55-64], and price forecasting [15,65-67].

1.3.1.2. Regression analysis

Regression analysis is a technique that allows to find and to model the intrinsic relationships between the variable of interest (i.e., the dependent variable) and one or more other variables (i.e., the independent variables⁴) [68,69].

The most common approach in regression analysis is to model the conditional expectance of probability distribution of the variable of interest, assuming that the variance does not changes with varying conditions. However, in many application the focus is transferred to modeling some quantiles of the probability distribution of the variable of interest (Quantile Regression, QR), or some other specific parameters of the underlying probability distribution. In all of these cases, however, the target is a mathematical function (regression function) of the independent variables.

Regression analysis can be both parametric and non-parametric. In the first case, unknown parameters have to be estimated in order to univocally determine the regression function. In the second case, specific techniques can be used to identify the model in an infinite-dimensional space of functions.

The simplest, parametric regression approach is the *linear regression*. The dependence between the dependent variable and independent variable(s) is linear in the parameters, meaning that parameters are only coefficients to independent variables. If two or more independent variables are considered, it is common to refer to MLR approaches. The general form of a MLR model is:

$$\hat{y}_t = \beta_0^{(MLR)} + \beta_1^{(MLR)} x_{t_1} + \dots + \beta_{m_1}^{(MLR)} x_{t_{m_1}} + \varepsilon_t, \quad (1.8)$$

where \hat{y}_t is the forecast at time t , $x_{t_1}, \dots, x_{t_{m_1}}$ are the m_1 predictors, $\beta_0^{(MLR)}, \beta_1^{(MLR)}, \dots, \beta_{m_1}^{(MLR)}$ are the $m_1 + 1$ unknown parameters of the model, and ε_t is a white noise term.

MLR models are fitted to data available in the training set trough least square approach (i.e., by minimizing the sum of squared errors in the

⁴ Dependent variables are also denoted in relevant literature as response variables or regressand. Independent variables are also denoted in relevant literature as explanatory variables, predictors, or regressors.

training set), although several approaches (such as lasso estimation [70]) provide some accurate results in many kind of situations by minimizing the lack of fit through different norms.

Further regression approaches are nonlinear in the parameters [69]. The model development can be performed on the basis of prior experience. It is important to note that nonlinear models are not so uncommon as one may think; e.g., if the response variable is assumed to be always positive, the constraint often leads to a nonlinear model. Obviously, estimating the parameters in this function is more challenging than estimating parameters in linear models. If nonlinear models can be brought to linear models through specific transformations, they are usually referred as intrinsically linear models. A well-known example is the following regression model:

$$\hat{y}_t = \beta_0^{(MLR)} x_t^{\beta_1^{(MLR)}}, \quad (1.9)$$

that can be easily linearized through the logarithmic transformation [69], resulting in:

$$\ln \hat{y}_t = \hat{y}_t^* = \ln \beta_0^{(MLR)} + \beta_1^{(MLR)} \ln x_t = \beta_0^{*(MLR)} + \beta_1^{(MLR)} x_t^*. \quad (1.10)$$

Regression approaches have been extensively applied in literature relevant to power systems, in both deterministic and probabilistic fashion; relevant papers are recalled in surveys on load forecasting [6,24,28,36-44], PV forecasting [27,36,45-54], WG forecasting [30,47,48,55-64], and price forecasting [15,65-67].

1.3.1.3. Univariate stochastic time series

A time series y_1, y_2, \dots, y_M is a set of M sequential observations of the variable of interest; the dependency on time is here implicit, as subscripts $1, 2, \dots, M$ stand for $t = 1, t = 2, \dots, t = M$. When the time series is the sample manifestation of a stochastic process, it is denoted as univariate stochastic time series [71,72].

The main characteristic in time series analysis is the dependency between “adjacent” observations; the exploitation (and coherent modeling) of such dependency is of practical use in building a model

that is representative of the stochastic process, starting from the available time series.

Models used to capture the features of a time series fall into the AutoRegressive (AR), Moving Average (MA), and Integrated (I) families; combining these models into ARMA, ARI, IMA, and ARIMA models can be of help to obtain better performance. For example, the integration effect is mandatory in order to capture non-stationary (e.g., seasonality) effects in the time series that AR models alone are not able to catch.

The analytic modeling of the ARIMA family is [71-73]:

$$\Phi(B)\nabla^d y_t = \theta(B)e_t, \quad (1.11)$$

where:

- B is the backward shift operator, defined as $B^i y_t = y_{t-i}$;
- $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ is the stationary AR operator of order p ;
- ∇^d is the backward difference operator of order d , defined as $\nabla^d y_t = y_t - y_{t-d}$;
- $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ is the MA operator of order q ;
- e_t the white noise term.

Note that expanding Eq. (1.11) in the so-called difference form [71], AR and ARI families (i.e., ARIMA family with $q = 0$) become particular cases of MLR, with predictors being lagged observations of the variable of interest, differences between observations of the variable of interest, and error terms. However, specific solving procedure and techniques (e.g., the Yule-Walker algorithm for AR models [71]) have been developed for such families due their wide utilization in forecasting and modeling problems, and therefore univariate stochastic time series are usually treated separately. Note also that the incorporation of the MA operator definitely separates time series approach from MLR approach, due the different error modeling. Time series models with exogenous variables (mostly falling into the ARX, ARMAX and ARIMAX family) can improve the quality of forecasts, by considering as input not only the time series of the

variable of interest, but also one (or more) time series of informative variables [74].

The analytic modeling of the ARIMAX family with a single exogenous variable x is:

$$\Phi(B)\nabla^d y_t = \Psi(B)x_t + \theta(B)e_t, \quad (1.12)$$

where $\Psi(B) = 1 - \Psi_1 B - \Psi_2 B^2 - \dots - \Psi_u B^u$ is the input operator of order u .

Both “pure” ARIMA models and ARIMAX models can be treated in nonlinear framework (some common models applied to power systems are NARIMA and NARX, respectively, where N stands for nonlinear). Time series approaches have been extensively applied in literature relevant to power systems, in both deterministic and probabilistic fashion; relevant papers are recalled in surveys on load forecasting [6,24,28,36-44], PV forecasting [27,36,45-54], WG forecasting [30,47,48,55-64], and price forecasting [15,65-67].

1.3.1.4. Exponential smoothing

Exponential smoothing approaches consist in providing forecasts through the weighted average of observations collected in the past, with weights that decrease as the observations get farther from the present [75]. In practice, the last observation available (i.e., the one collected at the start time of the forecast) is multiplied for the greatest weight, and the first observation ever collected is multiplied for the smallest weight. This approach is in the middle between the persistence and the average naïve benchmark, since the former is an “exponential smoothing” model with a unitary weight assigned to the most recent observation, and the latter is an “exponential smoothing” model with equal weight assigned to all of the available observed values.

The simplest exponential smoothing approach is also called single exponential smoothing; it is based on the following smoothing formula:

$$\hat{y}_t = \alpha^{(Exp)} y_{t-1} + (1 - \alpha^{(Exp)}) \hat{y}_{t-1}, \quad (1.13)$$

where \hat{y}_t, \hat{y}_{t-1} are the forecast values at time t and $t - 1$, respectively, and $0 \leq \alpha^{(Exp)} \leq 1$ is the smoothing parameter that weights more the recent observations if it is closer to 1, and that weights more the far observations if it is closer to 0. In the particular case of $\alpha^{(Exp)} = 1$, the exponential smoothing approach collapses in the persistence approach. Expanding Eq. (1.13), \hat{y}_t can be expressed as a function of all of the $N_{tr}^{(Exp)}$ previous available measurements $y_{t-N_{tr}^{(Exp)}}, y_{t-N_{tr}^{(Exp)}+1}, \dots, y_{t-1}$ of the variable of interest:

$$\begin{aligned} \hat{y}_t = & \alpha^{(Exp)} y_{t-1} + \alpha^{(Exp)} (1 - \alpha^{(Exp)}) y_{t-2} + \\ & + \alpha^{(Exp)} (1 - \alpha^{(Exp)})^2 y_{t-3} + \dots + \\ & + \alpha^{(Exp)} (1 - \alpha^{(Exp)})^{N_{tr}^{(Exp)}-1} y_{t-N_{tr}^{(Exp)}} + \\ & + (1 - \alpha^{(Exp)})^{N_{tr}^{(Exp)}} \hat{y}_{t-N_{tr}^{(Exp)}}, \end{aligned} \quad (1.14)$$

thus the term ‘‘exponential’’. Note that the forecast $\hat{y}_{t-N_{tr}^{(Exp)}}$ must be initialized in order to let the things work. In the single exponential smoothing approach, the unknown parameter $\alpha^{(Exp)}$ can be found as a solution of a non-linear programming, constrained problem, e.g., by minimizing the Root Mean Squared Error (RMSE) in the training period.

Several variants of the exponential smoothing approach exist and have been investigated in relevant literature; e.g., the double exponential smoothing approach iteratively applies two smoothing formula to take into account also an eventual trend of the variable of interest, while the triple exponential smoothing approach is able to deal with both trend and seasonality behaviors [75].

Exponential smoothing approaches have been extensively applied in literature relevant to power systems, in both deterministic and probabilistic fashion; relevant papers are recalled in surveys on load forecasting [6,24,28,36-44], PV forecasting [27,36,45-54], WG forecasting [30,47,48,55-64], and price forecasting [15,65-67].

1.3.1.5. Bayesian approaches

Bayesian approaches cover a wide range of forecasting procedures that make use of the Bayes' theorem to provide probabilistic predictions.

The formulation of the Bayes theorem is:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})}, \quad (1.15)$$

where:

- $p(\boldsymbol{\theta}|\mathbf{y})$ is the posterior probability of parameters $\boldsymbol{\theta}$ given a dataset $\mathbf{y} = \{y_{t-k-(N_{tr}^{(BM)}-1)}, y_{t-k-(N_{tr}^{(BM)}-2)}, \dots, y_{t-k}\}$ of $N_{tr}^{(BM)}$ past observed values of the variable of interest;
- $p(\boldsymbol{\theta})$ is the prior probability of parameters;
- $p(\mathbf{y}|\boldsymbol{\theta})$ is the likelihood of data \mathbf{y} given parameters $\boldsymbol{\theta}$;
- $p(\mathbf{y})$ is the probability of data \mathbf{y} .

In particular, the likelihood is computed as follows:

$$p(\mathbf{y}|\boldsymbol{\theta}) = \prod_{i=0}^{N_{tr}^{(BM)}-1} f(y_{t-k-i}|\boldsymbol{\theta}), \quad (1.16)$$

where $f(y_t|\boldsymbol{\theta})$ is the PDF of the variable of interest y_t given parameters $\boldsymbol{\theta}$, and the probability of data \mathbf{y} is given by:

$$p(\mathbf{y}) = \int p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}. \quad (1.17)$$

The posterior probability of parameters is afterward used in order to compute the predictive posterior distribution $p(y_t|\mathbf{y})$ of the variable of interest, as follows:

$$p(y_t|\mathbf{y}) = \int f(y_t|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}. \quad (1.18)$$

In practice, when using a Bayesian approach, parameters that appear in the forecasting model and that have to be estimated to forecast the variable of interest are probabilistically evaluated by modifying the forecaster's prior knowledge on parameters, according to the inference of the available training set [76-78].

The PDF $f(\cdot)$ plays a key role in evaluating the likelihood $p(\mathbf{y}|\boldsymbol{\theta})$, and in evaluating the predictive posterior distribution $p(y_t|\mathbf{y})$. A prior assumption on the type of the PDF $f(\cdot)$ is usually necessary to compute the likelihood in Bayesian approaches. Searching the most appropriate PDF family in such kind of approaches is very useful in order to improve the overall forecast performance, thus driving research interests toward probabilistic modeling and characterization. Note that the posterior predictive distribution $p(y_t|\mathbf{y})$ of the variable of interest does not necessarily belong to the family of the prior assumed PDF $f(\cdot)$.

The analytical solution of Eq. (1.17) and, therefore, Eq. (1.18) is viable only under particular assumptions on the prior distributions of the variable of interest and of the parameters. Gaussian processes are some examples of prior distributions in such Bayesian approaches; the underlying assumption of Gaussian Processes is that the joint probability distribution is a Gaussian distribution [79-81].

Bayesian approaches have been applied in literature relevant to power systems in probabilistic framework; relevant papers are recalled in surveys on load forecasting [6,24,28,36-44], PV forecasting [27,36,45-54], WG forecasting [30,47,48,55-64], and price forecasting [15,65-67].

1.3.1.6. Markov Chains

Contrarily to Bayesian approaches, Markov Chain (MC) approaches are non-parametric and they require no prior assumption on the probability distribution of the variable under study.

A basic formulation is the discrete MC, that relies on the hypothesis that the variable of interest y_t at time t can assume values in a finite number N_S of states, $S_t^{(1)}, \dots, S_t^{(N_S)}$ (discrete state space), and that the probability of the variable to be in the i^{th} state $S_t^{(i)}$ does not depend on the entire observed history, but only on s previous states. Therefore, the main assumption is the stochastic independence after the s^{th} time step [82,83]:

$$p(S_t|S_{t-1}, S_{t-2}, \dots, S_1) = p(S_t|S_{t-1}, S_{t-2}, \dots, S_{t-s+1}). \quad (1.19)$$

Forecasts for the variable y_t to be in state $S_t^{(i)}$ are obtained from the probability $p(S_t^{(i)} | S_{t-1}^{(j)}, S_{t-2}^{(r)}, \dots, S_{t-s+1}^{(l)})$, i.e., the probability that the variable of interest is in the i^{th} state, provided that it assumed values respectively in the $j^{\text{th}}, r^{\text{th}}, \dots, l^{\text{th}}$ states in the previous s time steps.

When applied to power system forecasting, MCs have been proposed and extended to a continuous, measurable state space.

MC approaches have extensively been applied in literature relevant to power systems, in both deterministic and probabilistic fashion; relevant papers are recalled in surveys on load forecasting [6,24,28,36-44], PV forecasting [27,36,45-54], WG forecasting [30,47,48,55-64], and price forecasting [15,65-67].

1.3.1.7. Kalman filter

If the forecasting problem can be put in a state-space form, Kalman filter becomes one of the most powerful tools to produce accurate forecasts [84]. Indeed, in such a case, the solution of the forecasting problem is analytical.

Kalman filter relies on two fundamental equations: the *measurement equation* and the *transition equation*.

The measurement equation is a mathematical model that links the variable of interest at the desired time horizon to explanatory variables, collected until the forecast start time, and to variables that represent the unavoidable disturbances in the model (e.g., forecast errors). It is worth noting that the measurement equation can be an arbitrary model (e.g., a MLR model or an univariate time series model), but also applications to machine learning approaches have been proposed in relevant literature [85,86].

The transition equation, instead, provides the mathematical formulation of the variation of the explanatory variables at the desired time horizon, with respect to past values of the explanatory variables (e.g., in a first-order Markov process).

Kalman filter, in its basic form, provides estimation of the future state variables through two *prediction equations* and an *updating equation*, for the evaluation of the covariance error matrix.

Several variants of the Kalman filter have been applied in order to appropriately describe forecasting problems also in nonlinear

frameworks; some examples are the Extended Kalman Filter (EKF), the Ensemble Kalman Filter (EnKF), and the unscented Kalman filter. In particular, the EKF linearizes the state-space system evaluating the Jacobian matrices of both measurement equation and transition equations. This usually results in a burdensome computational effort. The EnKF, instead, deals with the non-linearity by exploiting a set (ensemble) of state estimates, rather than a single state estimate; in this framework, there's no need to compute the Jacobian matrices to estimate the covariance error matrices, since the latter are evaluated from the estimates distribution. This allows to reduce the overall computational effort.

Kalman filter approaches have been applied in literature relevant to power systems; relevant papers are recalled in surveys on load forecasting [6,24,28,36-44], PV forecasting [27,36,45-54], WG forecasting [30,47,48,55-64], and price forecasting [15,65-67].

1.3.1.8. Artificial neural networks

Artificial neural networks (ANNs) are machine learning structures that exploit and resemble the functionalities of human brains. The elementary components of ANNs are artificial neurons, that act as a simplified version of human neurons [87]. External inputs $\mathbf{x} = \{x_1, x_2, \dots, x_{m_2}\}$ are fed to the artificial neuron, multiplied by the synaptic weights $w_1^{(ANN)}, w_2^{(ANN)}, \dots, w_{m_2}^{(ANN)}$ and then summed. The resulting value is diminished by a threshold value $\bar{\tau}^{(ANN)}$, and an activation function triggers the output y of the neuron if the value is greater than the activation limit. Activation functions may be step functions, linear or smoothed functions, or also truncated functions as the one portrayed in Figure 1.2. Therefore, in their general form, artificial neurons perform nonlinear operations on the inputs.

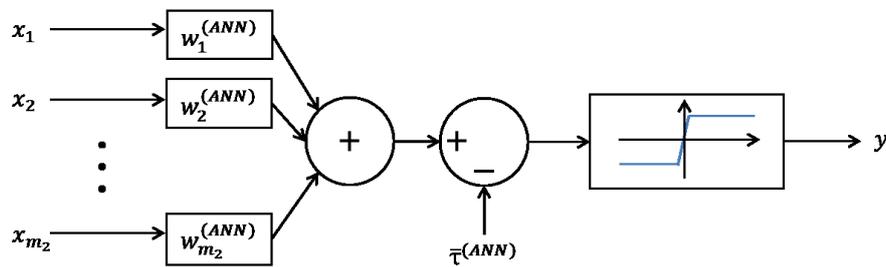


Figure 1.2 – Typical structure of an artificial neuron.

To resemble the activity of the human brain, usually more neurons are connected one each other in order to link the inputs of the ANN to the output of the ANN; neurons are therefore organized in layers. Input layers are the ones containing artificial neurons fed by the inputs of the ANN; output layers are the ones containing artificial neurons the output of which are the outputs of the ANN. One or more hidden layers are set between input and output layers; they are fed by the output of previous layers, and their outputs are fed to the following layers. This is the classical feed-forward ANN architecture, that is the most used structure in typical power system forecasting applications.

However, recurrent ANN architectures are also used in particular situations in order to improve the results. In recurrent ANN architectures, some of the output of the ANN are fed to the input layer after a feedback.

ANN approaches have been extensively applied in literature relevant to power systems, in both deterministic and probabilistic fashion; relevant papers are recalled in surveys on load forecasting [6,24,28,36-44], PV forecasting [27,36,45-54], WG forecasting [30,47,48,55-64], and price forecasting [15,65-67].

1.3.1.9. Support vector regression

SVR is a machine learning approach that is sometimes used as an alternative to ANNs. The high number of weights to be estimated in ANNs, indeed, is sometimes a difficult problem to cope with; classical techniques (e.g., back-propagation) provide estimations of ANN weights by solving a non-convex and unconstrained programming problem, with several local minima.

SVR structures allow to avoid these unwanted conditions; indeed, the algorithm solution is simple to be obtained, as the problem can be posed in a convex, constrained form, although being able to model non-linear relationships between inputs and outputs [88].

SVR links the output variable \hat{y}_t to a vector of m_3 explanatory variables $\mathbf{x}_t = \{x_{t_1}, \dots, x_{t_{m_3}}\}$ through a generic function selected among specific function classes (e.g., linear, Gaussian kernel, or polynomial kernels). The function is searched in order to let all of the outputs obtained through the SVR differ at most by an arbitrary threshold $\bar{\tau}^{(SVR)}$ from $N_{tr}^{(SVR)}$ available observations $y_{t-k-(N_{tr}^{(SVR)}-1)}, y_{t-k-(N_{tr}^{(SVR)}-2)}, \dots, y_{t-k}$; points that lie outside the $\pm\bar{\tau}^{(SVR)}$ band are penalized by a provided coefficient. In the linear case, the function is:

$$\hat{y}_t = \beta_0^{(SVR)} + \beta_1^{(SVR)} x_{t_1} + \dots + \beta_{m_3}^{(SVR)} x_{t_{m_3}}, \quad (1.20)$$

where \hat{y}_t is the forecast at time t , $x_{t_1}, \dots, x_{t_{m_3}}$ are the m_3 explanatory variables, and $\beta_0^{(SVR)}, \beta_1^{(SVR)}, \dots, \beta_{m_3}^{(SVR)}$ are the $m_3 + 1$ unknown parameters of the model. Eq. (1.20) is quite similar to classical MLR model (1.8), although differences arise in the modeling of the error term and, specifically, in the solving procedure. Indeed, estimations of the vector $\boldsymbol{\beta}^{(SVR)} = \{\beta_1^{(SVR)}, \dots, \beta_{m_3}^{(SVR)}\}$ and of parameter $\beta_0^{(SVR)}$ are found by solving the following quadratic, constrained programming problem:

$$\begin{aligned} \hat{\beta}_0^{(SVR)}, \hat{\boldsymbol{\beta}}^{(SVR)}, \hat{\xi}, \hat{\xi}^* = \\ = \operatorname{argmin}_{\beta_0^{(SVR)}, \boldsymbol{\beta}^{(SVR)}, \xi, \xi^*} \quad & \frac{1}{2} \boldsymbol{\beta}'^{(SVR)} \boldsymbol{\beta}^{(SVR)} + \\ & + C \sum_{i=0}^{N_{tr}^{(SVR)}-1} (\xi_{t-k-i} + \xi_{t-k-i}^*), \end{aligned} \quad (1.21)$$

$$\begin{aligned} \text{s.t. } y_{t-k-i} - \mathbf{x}'_{t-k-i} \boldsymbol{\beta}^{(SVR)} - \beta_0^{(SVR)} &\leq \bar{\tau}^{(SVR)} + \xi_{t-k-i} \quad \forall i, \\ \mathbf{x}'_{t-k-i} \boldsymbol{\beta}^{(SVR)} + \beta_0^{(SVR)} - y_{t-k-i} &\leq \bar{\tau}^{(SVR)} + \xi_{t-k-i}^* \quad \forall i, \\ \xi_{t-k-i}, \xi_{t-k-i}^* &\geq 0 \quad \forall i, \end{aligned}$$

where $\xi = \{\xi_{t-k-(N_{tr}^{(SVR)}-1)}, \dots, \xi_{t-k}\}$ and $\xi^* = \{\xi_{t-k-(N_{tr}^{(SVR)}-1)}^*, \dots, \xi_{t-k}^*\}$ are coefficients introduced to deal with points outside the $\pm \bar{\tau}^{(SVR)}$ band, and C is the given penalty coefficient. This approach, with adequate modifications, can be extended to other function classes (e.g., polynomial and Gaussian kernels). SVR approaches have been applied in literature relevant to power systems, in both deterministic and probabilistic fashion; relevant papers are recalled in surveys on load forecasting [6,24,28,36-44], PV forecasting [27,36,45-54], WG forecasting [30,47,48,55-64], and price forecasting [15,65-67].

1.3.1.10. Fuzzy approaches

Fuzzy logic has recently been applied also to forecasting methods. The idea of fuzzy logic is to assign a grade of membership to an element z , in order to individuate whether it belongs or not to a category [89]. For example, a person can be considered tall or non-tall; the concept is indeed vague (fuzzy). A nonnegative number, not greater than one, defined as the grade of membership, can be assigned to each person. The value 0 represents the full non-belongingness to the category “tall”, while the value 1 represents the full belongingness; intermediate values represent the “more or less” belongingness to the “tall” category. The grade of membership is clearly subjective; e.g., one could discriminate the belongingness degree on the basis of the height of the human being, assigning greater values to taller people and smaller values to shorter people.

To formalize fuzzy logic [89], a fuzzy set A^* in the finite universe of discourse $Z = \{z_1^{(fuz)}, z_2^{(fuz)}, \dots, z_{L_F}^{(fuz)}\}$ is the set of pairs $\{z^{(fuz)}, \mu^{A^*}(z^{(fuz)})\}$, with $z^{(fuz)} \in Z$. $\mu^{A^*}(z^{(fuz)}) \in [0,1]$ is the grade of membership of $z^{(fuz)}$ in A^* . The fuzzy set is usually indicated as follows:

$$A^* = \mu^{A^*}(z_1^{(fuz)})/z_1^{(fuz)} + \dots + \mu^{A^*}(z_{L_F}^{(fuz)})/z_{L_F}^{(fuz)}. \quad (1.22)$$

Fuzzy logic has been extended also to numbers, in the framework of fuzzy arithmetic. In particular, a fuzzy number is a fuzzy set A^* in \mathbb{R} .

As stated in [89], “fuzzy number may be exemplified by “about five”, “a little more than 7”, “more or less between 5 and 8”, and so on”. The extension of algebraic operations (sum, subtraction, multiplication and division) is possible in the framework of fuzzy operations. This allows to extend “standard” statistical approaches (such as time series, ANN, or MLR) in order to provide forecasts of a generic variable in the framework of fuzzy logic. For example, fuzzy logic applied to MLR is shown in the following [90].

In MLR, the deviations between observed values and estimated values are taken into account through the error term ε_t . In fuzzy logic, instead, the deviations are due the “fuzziness” of the model structure, i.e., due the fuzziness of model parameters. Therefore, the fuzzy linear function:

$$\mathbf{Y}_t^* = \mathbf{A}_1^* x_{t_1} + \dots + \mathbf{A}_m^* x_{t_{m_1}}, \quad (1.23)$$

is the extension of the MLR in a fuzzy framework; note that the output \mathbf{Y}_t^* is still a fuzzy set, from which the forecast can be drawn in probabilistic or deterministic frameworks. Obviously, the selection of the membership function strongly affects the resulting model, and it can be chosen by experts on the basis of the particular applications (e.g., triangular functions for parameters).

Fuzzy approaches have been extensively applied in literature relevant to power systems, in both deterministic and probabilistic fashion; relevant papers are recalled in surveys on load forecasting [6,24,28,36-44], PV forecasting [27,36,45-54], WG forecasting [30,47,48,55-64], and price forecasting [15,65-67].

1.3.1.11. K-nearest neighbors

Approaches based on K-nearest neighbors (KNNs) are among the most known techniques in forecasting [91]. Given a set $\mathcal{M} = \{\mathbf{x}_{t-k-(N_{tr}^{(KNN)}-1)}, \mathbf{x}_{t-k-(N_{tr}^{(KNN)}-2)}, \dots, \mathbf{x}_{t-k}\}$ of $N_{tr}^{(KNN)}$ m_4 -dimensional vectors of explanatory variables and a corresponding vector of $N_{tr}^{(KNN)}$ past outputs $\mathbf{y} = \{y_{t-k-(N_{tr}^{(KNN)}-1)}, y_{t-k-(N_{tr}^{(KNN)}-2)}, \dots, y_{t-k}\}$, the KNN approach provides a forecast \hat{y}_t of the variable of interest y_t , once the vector of

explanatory variables \mathbf{x}_t is known. The first step is to identify the set \mathcal{N}_t of the K vectors of explanatory variables, selected from the set \mathcal{M} , that are the “closest” to the vector \mathbf{x}_t . Obviously, a metric has to be introduced in order to give an interpretation to the “closeness” of vectors; e.g., the Euclidean distance in the vector space is one of the most common solution. Then, the forecast \hat{y}_t of the variable of interest is obtained from the K outputs corresponding to the K vectors in the set \mathcal{N}_t , e.g., through the arithmetic mean or through a mean weighted by the distances of explanatory variables vectors.

Obviously, the choice of the parameter K is critical, since low values of K tend to particularize the behavior (and therefore, the forecast) of the variable of interest, while greater values of K tend to generalize and smooth the behavior of the variable of interest. The computational burden linked to the selection of the nearest neighbors is obviously in favor of the former.

KNN smoothing approaches have been extensively applied in literature relevant to power systems, in both deterministic and probabilistic fashion; relevant papers are recalled in surveys on load forecasting [6,24,28,36-44], PV forecasting [27,36,45-54], WG forecasting [30,47,48,55-64], and price forecasting [15,65-67].

1.3.1.12. Kernel density estimation

Kernel Density Estimation (KDE) is applied to forecasting in order to build the predictive probability distribution of the variable of interest y_t , provided a set $\mathbf{y} = \{y_{t-k-(N_{tr}^{(KDE)}-1)}, y_{t-k-(N_{tr}^{(KDE)}-2)}, \dots, y_{t-k}\}$ of $N_{tr}^{(KDE)}$ past observations [92]. Contrarily to parametric density estimation, which consists in estimating the parameters of the (assumed) underlying PDF from the observed data (e.g., through the maximum likelihood or moment estimation procedures), KDE relies on a less rigid assumption on the family of the underlying probability distributions. In particular, a greater importance is given to observed data, since they are used not only to estimate parameters of a pre-assigned distribution, but also to *build* the predictive distribution $\hat{f}_{KDE}(y_t)$ in the following way:

$$\hat{f}_{KDE}(y_t) = \frac{1}{N_{tr}^{(KDE)} h^{(KDE)}} \sum_{i=1}^{N_{tr}^{(KDE)}} K\left(\frac{y_t - y_i}{h^{(KDE)}}\right), \quad (1.20)$$

where $K(\cdot)$ is the kernel function and $h^{(KDE)}$ is a smoothing parameter (usually referred as bandwidth). The peculiarity of kernel function classes is that they must be non-negative, and $\int_{-\infty}^{+\infty} K(t)dt = 1$. Some commonly used kernel functions are the uniform (in a compact support), Gaussian, Epanechnikov, Silverman, and logistic functions [92]. The selection of the bandwidth $h^{(KDE)}$ is a trade-off between the smoothing effect of the KDE and the corresponding estimation error; this effect is particularly enhanced in applications devoted to forecasting.

KDE approaches have been extensively applied in literature relevant to power systems, in both deterministic and probabilistic fashion; relevant papers are recalled in surveys on load forecasting [6,24,28,36-44], PV forecasting [27,36,45-54], WG forecasting [30,47,48,55-64], and price forecasting [15,65-67].

1.3.2. Physical approaches

Approaches based on physical models combine environmental and geographic aspects, such as terrain morphology, height and obstacles, to weather information such as air pressure, ambient temperature, and wind speed, in order to provide accurate forecasts for related variables [55]. These kind of approaches are widely used, e.g., in PV and WG power forecasting, since the main resource (irradiance and wind speed, respectively) can be accurately predicted from the knowledge of initial and boundary conditions, through the physical mathematical model of the surrounding area.

Numeric Weather Predictions (NWP) obtained through physical approaches are usually developed in a large-scale area framework, by splitting the globe in smaller areas through appropriate meshes. The thickness of the grid has an enormous impact on the computational burden of the physical approach. Indeed, the solution of the conservation equations at the given site is quite challenging, and dedicated computer architectures are mandatory in order to obtain forecasts in computational times adequate to forecast lead times [55].

NWP are usually provided at different lead time resolution (e.g., 1 hour, 3 hours, or 6 hours), for the next 3-20 days, according to the different model. The accuracy of NWPs in short-term scenarios usually is not competitive with forecasts coming from statistical approaches; however, the exploitation of NWPs as auxiliary inputs of statistical approaches (i.e., hybrid approaches) is a widely-used trick to increment the accuracy of the overall predictions.

Examples of physical forecasting models are developed by European and U.S. research centers, such as Met Office [93], European Centre for Medium-Range Weather Forecasts [94], National Oceanic and Atmospheric Administration [95], and National Centers for Environmental Prediction [96].

1.3.3. Hybrid approaches

Hybrid approaches consist in exploiting heterogeneous approaches in order to increase the robustness and quality of the resulting forecasts. Two types of hybrid methods find usual applications in power systems: cooperative approaches (series) and competitive approaches (parallel) (Fig. 1.3) [47].

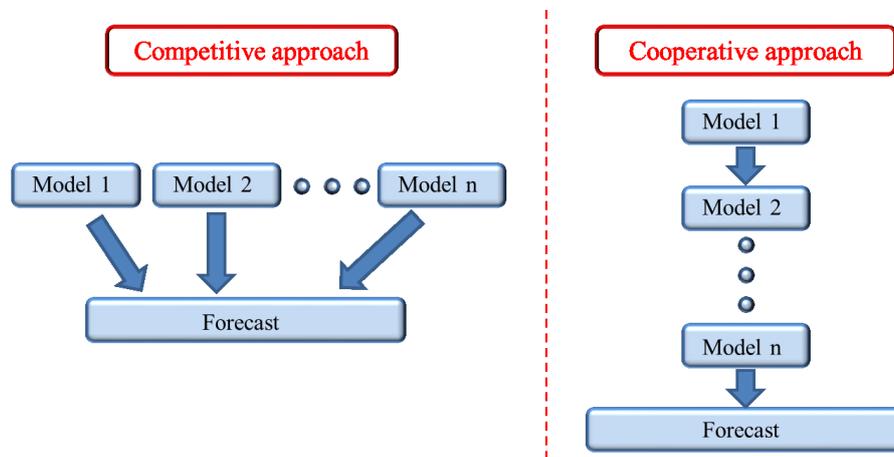


Figure 1.3 – Features of competitive and cooperative approaches.

Cooperative approaches enhance the performance of the forecasting system by sequentially performing transformations or forecasts, and using the outputs of previous steps as inputs of next steps. In

particular, the forecasting system is divided in several subtasks, and each of them is solved individually.

In particular, two kinds of cooperative approaches can be individuated: one based on the pre-processing of the data, and one based on post-processing of the forecasts.

Forecasting methods based on the initial filtering of input data (e.g., through the Wavelet decomposition or the empirical model decomposition [97]) fall in the pre-processing cooperative hybrid approach. Model output statistic is instead an example of cooperative post-processing hybrid approach; it performs corrections on short-term NWP forecasts through the application of a statistical model (e.g., an univariate time series model) on NWP forecasts.

One of the main drawbacks of cooperative approaches is the related computational burden. Indeed, the computational effort cannot be split in parallel activities, but each step of the approach must be run after the previous ones.

Competitive approaches combine the outputs of several models into a prediction that shows better performance than each base input. Usually they are also referred to as “ensemble” approaches. The concept of “diversity” introduced in [47] is vital in order to increase the performance of the overall forecasting system. Indeed, if few or no information is added by the introduction of a new base prediction, only few or no improvement is experienced in the final forecasts. Diversity can be introduced by adding forecasts coming from heterogeneous base models (e.g., different forecasting approaches) or homogeneous base models (e.g., ANNs with different architectures or NWP forecasts with different parameters).

The simplest combination of base predictions in a competitive approach is to select the average value, with equal or different weights for each base prediction, in a deterministic scenario. Further possible combinations are the median value, the minimum value, or the maximum value, on the basis of the forecasts’ end user needs. Different combinations (e.g., in a logarithmic scale) can be selected as the resulting predictions has to be searched to minimize a penalty or error function (e.g., the root mean squared error in the training period). Competitive methods, however, can also produce probabilistic forecasts by building the predictive PDF, starting from deterministic

base predictions provided at different scenarios, or from probabilistic base predictions. In the latter case, the combination of base predictive PDFs usually comes from a weighted or logarithmic average [98].

The main advantages of competitive approaches are related to the parallelization of several subtasks, resulting in reduced computational time.

Eventually, note that several approaches can be a combination of both competitive and cooperative subtasks (e.g., an initial filtering of the available data, followed by the combinations of multiple forecasting models); in this case, they will still be referred to as competitive approaches.

Chapter 2.

ADVANCED PROBABILISTIC METHODS FOR SHORT-TERM PHOTOVOLTAIC POWER FORECASTING

2.1.INTRODUCTION

The recent development of Renewable Generators (RGs) distributed across the whole electrical network traced the path towards a greener energy structure, with priceless benefits for the society. However, this strongly complicated the operation of electrical systems: contrarily to traditional, fossil-fueled plants, the majority of RGs cannot be finely regulated, since the primary energy source that is converted to electrical energy is not controllable and significantly linked to weather conditions. Among RGs, PV power plants are acknowledged to bring particular technical, environmental, and economic benefits to power systems, and their diffusion has straightforwardly grown during past years.

The power output of PV generators mainly depends on the solar irradiance in the specific site they are installed. Also, several more weather variables, such as air temperature and wind speed, have a decisive impact on the total energy production. These weather conditions that influence PV power cannot be exactly predicted, since they vary with random physical phenomena such as cloud motion. Therefore, the PV powers are treated as random variables and they are subject to forecasting.

This Chapter contributes to the existing literature on power forecasting by providing two new probabilistic PV power forecasting methods. The first method is based on a Bayesian approach, coupled with an underlying time series deterministic model in a cooperative fashion [99]. The second proposal, instead, is a competitive ensemble method

that optimally combines the probabilistic outputs of three heterogeneous base models that follow a Bayesian approach, a Markov chain approach, and a quantile regression approach [100].

This Chapter is organized as follows. The state of the art on probabilistic PV power forecasting methods is addressed in Section 2.2. The first provided contribution, i.e., the Bayesian-based probabilistic method, is discussed in Section 2.3. The second provided contribution, i.e., the competitive ensemble method, is discussed in Section 2.4. For both contributions, both the analytic model formulations and numerical applications based on real data are provided in the corresponding sub-Sections. The Chapter is concluded in Section 2.5.

2.2. PROBABILISTIC METHODS FOR SHORT-TERM

PHOTOVOLTAIC POWER FORECASTING: STATE OF THE ART

Hundreds of papers have dealt with PV forecasting methods in the last fifteen years. The most of the contributions have been developed in a deterministic framework, while the interest in probabilistic PV forecasting rapidly grew in the last 5 years. Several review papers [27,36,45,47-54,101] summarized the state of the art on PV power forecasting. Methods addressed to forecast solar radiation are also considered in this survey, since they can be used in an indirect approach to forecast also PV power through PV radiation-power models [102,103].

In the following, we will more specifically refer to methods that have been proposed for probabilistic STF applications, since the contributions of this thesis to the existing literature are in this field. Details about methods proposed for deterministic STF are reported in [27,36,45,47-54,101].

First attempts in PV power probabilistic forecasting are quite recent with respect to other power system variables (i.e., wind and load); only few papers on probabilistic PV forecasting are indeed dated before 2013.

The authors in [104] proposed a deterministic method based on AR and ARX time series model, with inputs provided by NWP, giving

also some hints on the probabilistic extension of such approach in a QR framework. The application was both for VSTF and STF, since lead times ranged from 1 to 36 hours ahead.

Interest in probabilistic PV forecasting definitely grew around 2013, as the number of publications in this field exponentially increased.

Similarities between past NWP were exploited in [105] through a genetic algorithm. The authors grouped PV powers in different bins, ranging from zero to the maximum producible power. Probabilistic forecasts consisted of the occurrence probabilities of PV power production to be each bin. No rigorous assessment on the quality of such probabilistic forecasts was however performed.

A probabilistic fuzzy approach was developed in [106], based on an Interval type-2 Takagi-Sugeno-Kang system. It exploited NWP forecasts of irradiance and ambient temperature; results were however compared only in terms of deterministic errors.

A probabilistic Bayesian approach with underlying ARX deterministic model was proposed in [33,107] to produce PV power forecasts. Measurements of weather variables, such as relative humidity, cloud cover and wind speed were used as exogenous inputs of the ARX model. ARX coefficients were probabilistically estimated in a Bayesian inference approach.

A benchmark model for probabilistic forecasts of the daily average of PV power was proposed in [108]. In particular, several QR and Quantile Regression Forest (QRF) models with NWP inputs were used and calibrated through a technique based on the rank histogram. Both rank histograms and Continuous Ranked Probability Score (CRPS) were used to assess the quality of forecasts. It is worth noting that the calibration step did not necessarily lead to an improvement in terms of reliability⁵.

NWPs were also used in [109] as inputs of a QRF model, together with PV power measurements. The assessment of the goodness of probabilistic forecasts up to 24 hours ahead was performed only in terms of deterministic errors, selecting the median as point forecast,

⁵ Details on such indices and tools to assess the quality of forecasts, and the definition of calibration and sharpness, are in the Appendix.

and through a performance error that took into account the total number of observations that lied outside the 0.1-0.9 quantiles.

NWPs were used in [110] in a probabilistic framework to build an analog ensemble for PV day-ahead forecasts; results were compared to a naïve benchmark based on the persistence and to results obtained through a QR model, in a rigorous probabilistic framework through a proper score (i.e., the CRPS) and rank histograms.

A vector AR model was developed in [111] to take into account spatial correlation among near PV installations in a Portuguese test SG. This model was used in a deterministic framework to build point forecasts through the recursive least squares solving method, and in a probabilistic framework by adopting the gradient boosting approach, with the Pinball Loss Function (PLF) as loss function to be minimized. Results were assessed through the CRPS.

An Extreme Learning Machine (ELM) method was proposed to forecast predictive intervals at 30-minutes resolution, exploiting only past values of PV power and influent meteorological variables such as temperature and cloud cover. Prediction Interval Coverage Probability (PICP) and Prediction Interval Normalized Averaged Width (PINAW) were used as probabilistic metrics.

A hybrid wavelet-ANN deterministic model was selected in [112] as underlying model to build probabilistic forecasts; in particular, bootstrap confidence intervals were evaluated from deterministic predictions. No rigorous assessment of the quality of probabilistic forecasts was however performed, apart from qualitative graphical evaluation of confidence intervals with respect to the observed values.

Prediction intervals of PV power generation were estimated in [113], based on the assumption about the distribution of the forecast error. In particular, since the true error distribution is not prior known, it is assumed to be a Gaussian distribution, the parameters of which are estimated through a maximum likelihood approach. The results are tested in terms of PICP, but a quantitative measure is missing.

Lower and upper bounds of predicted PV power were estimated in [114] through a SVR approach, considering only past observations of PV power and weather variables. The authors performed a probabilistic evaluation of the results through the PICP.

In [115] prediction intervals of solar irradiance were forecasted up to 6 hours through the combination of two linear models (ARMA and Generalized AutoRegressive Conditional Heteroscedastic (GARCH)). The only inputs of the method proposed by the authors were past observations of solar irradiance; results were assessed in terms of CRPS.

A method to construct probabilistic ensemble of NWP forecasts up to 3 days ahead was proposed in [116]; the authors also trained an ANN in order to reduce the bias error, and applied two further methods in order to improve the quality of forecasts: the variance deficit and the ensemble model output statistic. Results were assessed in terms of CRPS, reliability diagrams, and rank histograms.

A multi-model ensemble of several base predictors that used NWPs as input was developed in [117]. Base models were selected among a seasonal ARIMAX, a SVR and two different MLP ANNs. A rigorous assessment on the quality of probabilistic forecasts was not performed by the authors.

The authors of [118] developed an ensemble method based on NWP forecasts that was able to produce predictive intervals of solar irradiance. The empirical coverage proportions were compared to the nominal ones.

Feed-forward neural networks were exploited in an ELM framework in [119] to produce probabilistic forecasts; inputs were both past observations of PV power and weather variables. Results were assessed in terms of reliability diagrams and PLF.

The method proposed in [120] consisted in the state-space modeling the aggregate hourly PV power, taking into account the parameters of the PV system. An ensemble Kalman filter allowed to build a probabilistic forecasts of PV power for the considered time horizon. Results were assessed in terms of CRPS.

The raising interest in probabilistic PV power forecasting was driven also by the Global Energy Forecasting Competition 2014 (GEFCom2104) [12], that listed a track on PV power and made datasets publically available for reproducibility. Forecasts of several weather variables, together with observations of PV power, were given as inputs to contestants; results were compared in terms of PLF in order to rank the teams.

Gradient boosting and KNN were used in [121] to produce probabilistic forecasts of PV power in the framework of GEFCom2104. Clear sky models were also considered in order to improve the quality of forecasts. The authors performed a cross-validation strategy in order to avoid over-fitting.

In [122], an ensemble of a QRF model and gradient boosting models was developed to predict the predictive probability distribution of PV power. Compared with other high-ranked GEFCom2014 methods, no spatial information was added (i.e., no variables coming from near PV installations) as input of the model.

A multiple QR method developed during GEFCom2014 was shown in [123]; it exploited an innovative solution of the regression problem formulation. Moreover, Radial Basis Function was used to select features that were able to catch non-linear dependencies on the inputs. The method proved its versatility by performing well also in wind, load and price tracks of GEFCom2014.

Available data from GEFCom2014 was used in [124] to build an ensemble of forecasts from statistical methods, such as KNN and gradient boosting. In particular, results were combined through normal distributions with different initial settings. The performance of the resulting probabilistic model was compared to base predictors in terms of PLF.

Further probabilistic approaches devoted to PV power VSTF were in [125-133]. Their suitability for longer lead times has not yet been tested in relevant literature.

2.3.A NEW PROBABILISTIC BAYESIAN-BASED METHOD FOR SHORT-TERM PHOTOVOLTAIC POWER FORECASTING

The probabilistic Bayesian method [99] is applied to predict the PDF of the hourly active power generated by PV systems, using two relationships of the PV active power that are most frequently used in the relevant literature: the first relationship links the PV power to the hourly solar irradiance, while the second relationship links the PV active power to the hourly clearness index. These relationships are used in the frame of a Bayesian-based procedure in which the forecast

of the PDF of the PV power at the time horizon h is performed at the origin hour $h - k$, with k the forecast lead time (Fig. 2.1).

Details about the proposed Bayesian method are reported in the following sub-Section 2.3.1; numerical applications based on real data are in sub-Section 2.3.2.

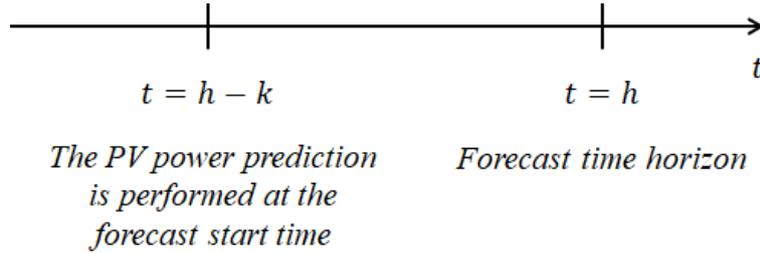


Figure 2.1 – Forecast time scales.

2.3.1. Proposed method

The Bayesian method [99] provides probabilistic forecasts of PV power by transforming probabilistic forecasts of the hourly solar irradiance (or the hourly clearness index) through well-known relationships, in an indirect approach.

Solar irradiance (or hourly clearness index) is modeled through an analytic PDF, whose parameters are estimated by means of the Bayesian Inference (BI) of past available observations. An exogenous linear regression model is also defined in order to link one of the PDF parameters to the measurements of the hourly solar irradiance (or the hourly clearness index), and to some influencing weather variables.

Eventually, the proposed method involves the following steps:

- i. selection of the relationships that link the PV active power to the hourly clearness index and the solar irradiance (*sub-Section 2.3.1.1*);
- ii. selection of the PDFs to model the hourly solar irradiance and the hourly clearness index irradiance (*sub-Section 2.3.1.2*);
- iii. selection of an exogenous linear regression model that links some of the PDFs parameters in (ii) to past observations of hourly solar irradiance (or clearness index), and to the

- measurements of influencing meteorological variables (*sub-Section 2.3.1.3*);
- iv. evaluation of the predictive samples of PDFs parameters and regression parameters through the BI of past observations of hourly solar irradiance (or clearness index) (*sub-Section 2.3.1.4*);
 - v. evaluation of the predictive samples of PV power (*sub-Section 2.3.1.5*).

2.3.1.1. Relationships that link the PV active power to the hourly clearness index and the solar irradiance

When the PV generation system is equipped with a maximum power point tracking system [134], the output active power P_{PV_h} of the PV system at the forecast time horizon h can be expressed as a linear function of the irradiance G_{β_h} at hour h on a surface with an inclination of β degrees with respect to the horizontal plane, as in [135]:

$$P_{PV_h} = S_{PV}\eta_{PV}G_{\beta_h}, \quad (2.1)$$

where S_{PV} is the surface area of the PV array, and η_{PV} is the efficiency of the PV system. The main advantage found in the application of Eq. (2.1) is connected to the linear relationship between P_{PV_h} and the related meteorological variable G_{β_h} ; the overall efficiency of the PV system is approximately considered constant with the solar irradiance, although with no loss of generalization.

Another relationship that is frequently used for the calculation of the PV power, and that was derived directly from Eq. (2.1), expresses the PV power as a function of the hourly clearness index K_h at the forecast time horizon h ; it is defined as the ratio between the hourly mean global solar irradiance on a horizontal plane at the surface of earth and the hourly mean extra-atmospheric solar irradiance [136,137]. The relationship is given by [137]:

$$P_{PV_h} = S_{PV}\eta_{PV}(T_h K_h - T'_h K_h^2), \quad (2.2)$$

where T_h and T'_h are respectively defined as:

$$T_h = \left[\left(R_{b_h} + \rho \frac{1 - \cos \beta}{2} \right) + \left(\frac{1 + \cos \beta}{2} - R_{b_h} \right) c_1 \right] r_{d_h} \frac{H_{0_h}}{3600}, \quad (2.3)$$

$$T'_h = \left(\frac{1 + \cos \beta}{2} - R_{b_h} \right) c_2 r_{d_h} \frac{H_{0_h}}{3600}. \quad (2.4)$$

In Eqs. (2.3) and (2.4), R_{b_h} is the ratio of beam radiation on a tilted surface to the beam radiation on a horizontal surface, ρ is the reflectance of the ground, r_{d_h} is the ratio of diffuse radiation in hours to diffuse radiation in a day, H_{0_h} is the extra-terrestrial total solar radiation, and c_1, c_2 are coefficients that link the diffuse fraction of the total hourly solar radiation on the horizontal plane K_{d_h} to the hourly clearness index K_h :

$$K_{d_h} = c_1 - c_2 K_h. \quad (2.5)$$

Note that sun hourly position plays a key role in the evaluation of coefficients T_h and T'_h in Eqs. (2.3) and (2.4), since the ratio R_{b_h} is strongly affected by this position. More details on this subject can be found in [138]. Also, models (2.1) and (2.2) could be improved by taking into account the saturation values of PV power introduced by DC/AC converter control system.

2.3.1.2. Selection of the PDFs of the hourly solar irradiance and of the hourly clearness index irradiance

The PDFs of the hourly solar irradiance (a) and the hourly clearness index (b) need to be selected in order to apply the Bayesian method.

a) The PDF $f_{BETA}(\cdot)$ of the hourly solar irradiance G_{β_h} at the hour h can be modeled using the well-known Beta distribution as proposed in [139]:

$$\begin{aligned} f_{BETA}(G_{\beta_h} | \sigma_{BETA_h}, \varphi_{BETA_h}) &= \\ &= \frac{\left(\frac{G_{\beta_h}}{G_{\beta_h}^{(max)}} \right)^{\sigma_{BETA_h} - 1} \left(1 - \frac{G_{\beta_h}}{G_{\beta_h}^{(max)}} \right)^{\varphi_{BETA_h} - 1}}{B(\sigma_{BETA_h}, \varphi_{BETA_h}) G_{\beta_h}^{(max)}}, \end{aligned} \quad (2.6)$$

where σ_{BETA_h} and φ_{BETA_h} are the shape parameters of the Beta distribution, $B(\cdot)$ is the Beta function, and $G_{\beta_h}^{(max)}$ is the upper bound of the observed values of G_{β_h} , which can be calculated from the measurements of the irradiance on a horizontal plane [140,141]. The mean value μ_{BETA_h} of Beta distribution can be calculated as follows:

$$\mu_{BETA_h} = \frac{\sigma_{BETA_h}}{\sigma_{BETA_h} + \varphi_{BETA_h}} G_{\beta_h}^{(max)}, \quad (2.7)$$

and therefore the shape parameter φ_{BETA_h} can be univocally determined if both the mean value μ_{BETA_h} and the shape parameter σ_{BETA_h} are known, as follows:

$$\varphi_{BETA_h} = \frac{\sigma_{BETA_h}(G_{\beta_h}^{(max)} - \mu_{BETA_h})}{\mu_{BETA_h}}, \quad (2.8)$$

As a result of the analysis of Eqs. (2.6), (2.7), and (2.8), the estimation $\hat{\mu}_{BETA_h}$ of the mean value μ_{BETA_h} and the estimation $\hat{\sigma}_{BETA_h}$ of the shape parameter σ_{BETA_h} are sufficient to estimate the predictive PDF $\hat{f}_{BETA}(G_{\beta_h} | \hat{\sigma}_{BETA_h}, \hat{\mu}_{BETA_h})$ at the desired time horizon h . In particular, $\hat{\mu}_{BETA_h}$ is estimated by applying the exogenous linear regression model described in sub-Section 2.3.1.3, while the evaluation of the shape parameter $\hat{\sigma}_{BETA_h}$ is committed to the BI algorithm described in sub-Section 2.3.1.4.

b) The PDF $f_{GAMMA}(\cdot)$ of the hourly clearness index K_h at the hour h is modeled using the following modified Gamma distribution [142]:

$$\begin{aligned} f_{GAMMA}(K_h | \gamma_{GAMMA_h}, \lambda_{GAMMA_h}) &= \\ &= \gamma_{GAMMA_h} \frac{K_h^{(max)} - K_h}{K_h^{(max)}} e^{\lambda_{GAMMA_h} K_h}, \end{aligned} \quad (2.9)$$

where $K_h^{(max)}$ is an assigned upper bound of the observed values of K_h , and $\gamma_{GAMMA_h}, \lambda_{GAMMA_h}$ are the parameters of the modified Gamma distribution; these parameters are univocally determined if the mean value μ_{GAMMA_h} of K_h at hour h is known. The relationships are:

$$\gamma_{GAMMA_h} = \frac{\lambda_{GAMMA_h}^2 K_h^{(max)}}{\left[e^{\lambda_{GAMMA_h}^2 K_h^{(max)}} - 1 - \lambda_{GAMMA_h}^2 K_h^{(max)} \right]}, \quad (2.10)$$

$$\lambda_{GAMMA_h} = \frac{(2\bar{K}_h - 17.519e^{-1.3118\bar{K}_h} - 1062e^{-5.0426\bar{K}_h})}{K_h^{(max)}}, \quad (2.11)$$

$$\text{with } \bar{K}_h = \frac{K_h^{(max)}}{K_h^{(max)} - \mu_{K_h}}.$$

As a result from the analysis of Eqs. (2.9), (2.10) and (2.11), the estimation $\hat{\mu}_{GAMMA_h}$ of the mean value μ_{GAMMA_h} is sufficient to estimate the predictive PDF $\hat{f}_{GAMMA}(K_h | \hat{\mu}_{GAMMA_h})$ at the desired time horizon h ; this is effected in the next sub-Section 2.3.1.3 by applying an exogenous linear regression model with weather inputs.

2.3.1.3. Exogenous linear regression model

Two linear regression models with exogenous inputs are considered to link the estimated mean value $\hat{\mu}_{BETA_h}$ of solar irradiance (a) and the estimated mean value $\hat{\mu}_{GAMMA_h}$ of clearness index (b), to explanatory exogenous variables observed at the forecast start time $h - k$.

a) The first proposed model links the estimated mean value $\hat{\mu}_{BETA_h}$ of solar irradiance, to the measurements $\mu_{N_{tr}^{(G)}}$, rh_{h-k} , cc_{h-k} , p_{h-k} and at_{h-k} of, respectively, the mean value of the $N_{tr}^{(G)}$ last observations of hourly solar irradiance collected until the forecast start time $h - k$, the relative humidity at $h - k$, the cloud cover at $h - k$, the pressure at $h - k$, and the ambient temperature at hour $h - k$. The model is expressed as follows:

$$\begin{aligned} \hat{\mu}_{BETA_h} = & \hat{\beta}_0^{(G)} + \hat{\beta}_1^{(G)} \mu_{N_{tr}^{(G)}} + \hat{\beta}_2^{(G)} rh_{h-k} + \hat{\beta}_3^{(G)} cc_{h-k} + \\ & + \hat{\beta}_4^{(G)} p_{h-k} + \hat{\beta}_5^{(G)} at_{h-k}, \end{aligned} \quad (2.12)$$

where $\hat{\beta}_0^{(G)}, \dots, \hat{\beta}_5^{(G)}$ are the estimated values of coefficients $\beta_0^{(G)}, \dots, \beta_5^{(G)}$ of the regression model. Since these coefficients are modeled in the BI approach (sub-Section 2.3.1.4) as random variables,

$\hat{\beta}_0^{(G)}, \dots, \hat{\beta}_5^{(G)}$ are actually samples drawn from the corresponding posterior distributions of parameters $\beta_0^{(G)}, \dots, \beta_5^{(G)}$. Note that the samples of posterior distributions of these coefficients are updated at each time horizon h , but the symbol h is omitted in the equations to simplify the analytic formulation of the problem.

b) The second proposed model links the mean value $\hat{\mu}_{GAMMA_h}$ of clearness index at the time horizon h to the mean value $\mu_{N_{tr}^{(K)}}$ of the last $N_{tr}^{(K)}$ observations of the hourly clearness index and to the same weather variables in Eq. (2.12) at the forecast start time $h - k$; it is expressed as follows:

$$\begin{aligned} \hat{\mu}_{GAMMA_h} = & \hat{\beta}_0^{(K)} + \hat{\beta}_1^{(K)} \mu_{N_{tr}^{(K)}} + \hat{\beta}_2^{(K)} r h_{h-k} + \hat{\beta}_3^{(K)} c c_{h-k} + \\ & + \hat{\beta}_4^{(K)} p_{h-k} + \hat{\beta}_5^{(K)} a t_{h-k}, \end{aligned} \quad (2.13)$$

where $\hat{\beta}_0^{(K)}, \dots, \hat{\beta}_5^{(K)}$ are the estimated values of coefficients $\beta_0^{(K)}, \dots, \beta_5^{(K)}$ of the regression model. Once again these coefficients are modeled in the BI approach as random variables (sub-Section 2.3.1.4), and therefore $\hat{\beta}_0^{(K)}, \dots, \hat{\beta}_5^{(K)}$ are actually samples drawn from the corresponding posterior distributions of parameters $\beta_0^{(K)}, \dots, \beta_5^{(K)}$. The samples of posterior distributions of these coefficients were updated at each time horizon h , being the symbol h omitted.

2.3.1.4. Evaluation of the PDFs of the coefficients of the exogenous linear regression models and of the Beta distribution shape parameter

The BI approach [76,77] allows to estimate the posterior distributions $\hat{p}(\beta_0^{(G)} | \mathbf{S}_{I_{\beta_h}}), \dots, \hat{p}(\beta_5^{(G)} | \mathbf{S}_{I_{\beta_h}}), \hat{p}(\sigma_{BETA_h} | \mathbf{S}_{I_{\beta_h}})$ and $\hat{p}(\beta_0^{(K)} | \mathbf{S}_{K_h}), \dots, \hat{p}(\beta_5^{(K)} | \mathbf{S}_{K_h})$ of each parameter of the linear regression models (2.12),(2.13) and of the Beta distribution shape parameter σ_{BETA_h} , given respectively the vector $\mathbf{S}_{G_{\beta_h}} = \{G_{\beta_{h-k-(N_{tr}^{(G)}-1)}}, \dots, G_{\beta_{h-k}}\}$ of $N_{tr}^{(G)}$ hourly observations of solar irradiance, and the vector $\mathbf{S}_{K_h} = \{K_{h-k-(N_{tr}^{(K)}-1)}, \dots, K_{h-k}\}$ of $N_{tr}^{(K)}$

hourly measurements of clearness index. All of the measurements are supposed to be available until the forecast start hour $h - k$. Actually, the BI allows the estimation of the joint posterior distributions $\hat{p}(\beta_0^{(G)}, \dots, \beta_5^{(G)}, \sigma_{BETA_h} | \mathbf{S}_{G\beta_h})$ and $\hat{p}(\beta_0^{(K)}, \dots, \beta_5^{(K)} | \mathbf{S}_{K_h})$. Once they are known, it's trivial to evaluate the posterior distributions of each parameter by applying the theory of the joint PDFs.

Unfortunately, only simplified expressions of the posterior distributions can be provided in closed form. They are known as the un-normalized posterior distributions $\hat{p}_q(\beta_0^{(G)}, \dots, \beta_5^{(G)}, \sigma_{BETA_h} | \mathbf{S}_{G\beta_h})$ and $\hat{p}_q(\beta_0^{(K)}, \dots, \beta_5^{(K)} | \mathbf{S}_{K_h})$ of the prior random parameters, and they are equal to:

$$\begin{aligned} \hat{p}_q(\beta_0^{(G)}, \dots, \beta_5^{(G)}, \sigma_{BETA_h} | \mathbf{S}_{G\beta_h}) = \\ \hat{p}(\mathbf{S}_{G\beta_h} | \hat{\beta}_0^{(G)}, \dots, \hat{\beta}_5^{(G)}, \hat{\sigma}_{BETA_h}) \prod_{i=0}^5 [p(\hat{\beta}_i^{(G)})] p(\sigma_{BETA_h}), \end{aligned} \quad (2.14)$$

$$\hat{p}_q(\beta_0^{(K)}, \dots, \beta_5^{(K)} | \mathbf{S}_{K_h}) = \hat{p}(\mathbf{S}_{K_h} | \hat{\beta}_0^{(K)}, \dots, \hat{\beta}_5^{(K)}) \prod_{i=0}^5 [p(\hat{\beta}_i^{(K)})], \quad (2.15)$$

where $\hat{p}(\mathbf{S}_{G\beta_h} | \hat{\beta}_0^{(G)}, \dots, \hat{\beta}_5^{(G)}, \hat{\sigma}_{BETA_h})$ and $\hat{p}(\mathbf{S}_{K_h} | \hat{\beta}_0^{(K)}, \dots, \hat{\beta}_5^{(K)})$ are likelihood functions, and $p(\beta_i^{(G)})$, $p(\beta_i^{(K)})$, $p(\sigma_{BETA_h})$ are the prior distributions of the random parameters, chosen with a large variance⁶. The likelihood functions are the PDFs in (2.6) and (2.9), evaluated for the vectors of measurements $\mathbf{S}_{G\beta_h}$ and \mathbf{S}_{K_h} :

$$\begin{aligned} \hat{p}(\mathbf{S}_{G\beta_h} | \hat{\beta}_0^{(G)}, \dots, \hat{\beta}_5^{(G)}, \hat{\sigma}_{BETA_h}) = \\ = \prod_{i=0}^{N_{tr}^{(G)} - 1 - k} \hat{f}_{BETA}(G_{\beta_{h-k-i}} | \hat{\sigma}_{BETA_h}, \hat{\mu}_{BETA_{h-k-i}}), \end{aligned} \quad (2.16)$$

⁶ This is the case of uninformative prior distributions, that are usually chosen so that the data, rather than the prior distributions, determine the values of the relevant parameters in the posterior distribution [143]; uniform distributions and Gaussian distributions with large variance are common uninformative prior distributions. If instead the forecaster disposes of strong prior knowledge on a parameter, he/she can rather select an informative prior distribution (e.g., a Gaussian with small variance) [77].

$$\hat{p}(\mathbf{S}_{K_h} | \hat{\beta}_0^{(K)}, \dots, \hat{\beta}_5^{(K)}) = \prod_{i=0}^{N_{tr}^{(K)}-1-k} \hat{f}_{GAMMA}(K_{h-k-i} | \hat{\mu}_{GAMMA_{h-k-i}}), \quad (2.17)$$

with $\hat{\mu}_{BETA_{h-k-i}}$ and $\hat{\mu}_{GAMMA_{h-k-i}}$ derived from Eqs. (2.12) and (2.13) using parameter samples $\hat{\beta}_0^{(G)}, \dots, \hat{\beta}_5^{(G)}$ and $\hat{\beta}_0^{(K)}, \dots, \hat{\beta}_5^{(K)}$, respectively. Measurements of the hourly solar irradiance, the hourly clearness index, the relative humidity, the cloud cover, the pressure, and the ambient temperature are contained in the vectors $\mathbf{S}_{G_{\beta_h}}, \mathbf{S}_{K_h}, \mathbf{S}_{RH_h}, \mathbf{S}_{CC_h}, \mathbf{S}_{P_h}, \mathbf{S}_{AT_h}$, respectively; measurements collected until the forecast start hour $h - k$. The dimensions of the historical data can be selected with adequate criteria, thus improving the accuracy of the forecasting method.

The estimation of the un-normalized posterior distributions in Eqs. (2.14) and (2.15) is sufficient for developing algorithms that provide information about the joint posterior distributions. In fact, the Monte Carlo Markov Chain (MCMC) simulation method based on the Metropolis-Hasting algorithm [76] can be directly applied to the un-normalized distribution to obtain samples of the posterior distribution of each parameter.

2.3.1.5. Evaluation of the samples of the predictive PDF of PV power

The samples of parameters drawn from posterior distributions $\hat{p}(\beta_0^{(G)} | \mathbf{S}_{G_{\beta_h}}), \dots, \hat{p}(\beta_5^{(G)} | \mathbf{S}_{G_{\beta_h}})$ can be used, together with the measurements $\mu_{N_{tr}}^{(G)}, rh_{h-k}, cc_{h-k}, p_{h-k}$ and at_{h-k} , to obtain samples $\hat{\mu}_{BETA_h}$ from Eq. (2.12) through the MCMC procedure. Once the samples $\hat{\mu}_{BETA_h}$ and the samples drawn from the posterior distribution $\hat{p}(\sigma_{BETA_h} | \mathbf{S}_{G_{\beta_h}})$ of the shape parameter are known, the predictive samples of solar irradiance from the predictive distribution $\hat{f}_{BETA}(G_{\beta_h} | \hat{\sigma}_{BETA_h}, \hat{\mu}_{BETA_h})$ are provided by applying Eqs. (2.6) and (2.8). Finally, the predictive samples of PV power are provided through a random rejection sampling of the predictive distribution $\hat{f}_{BETA}(G_{\beta_h} | \hat{\sigma}_{BETA_h}, \hat{\mu}_{BETA_h})$ and using Eq. (2.1).

Similarly, the samples of parameters drawn from posterior distributions $\hat{p}(\beta_0^{(K)} | \mathbf{S}_{K_h}), \dots, \hat{p}(\beta_5^{(K)} | \mathbf{S}_{K_h})$ can be used, together with

the measurements $\mu_{N_{tr}}^{(k)}, rh_{h-k}, cc_{h-k}, p_{h-k}, at_{h-k}$, to obtain samples $\hat{\mu}_{GAMMA_h}$ from Eq. (2.13). Once the samples $\hat{\mu}_{GAMMA_h}$ are known, indeed, the predictive samples of clearness index are drawn from the predictive distribution $\hat{f}_{GAMMA}(K_h|\hat{\mu}_{GAMMA_h})$ by applying Eq. (2.9). Finally, the predictive samples of PV power are provided through a random rejection sampling of the predictive distribution $\hat{f}_{GAMMA}(K_h|\hat{\mu}_{GAMMA_h})$ and using Eq. (2.2).

The random rejection sampling method is used in both cases in accordance with the Von Neumann method for the generation of random samples from a known probability function [144].

Eventually, the flow chart of the proposed probabilistic Bayesian method is shown in Figure 2.2.

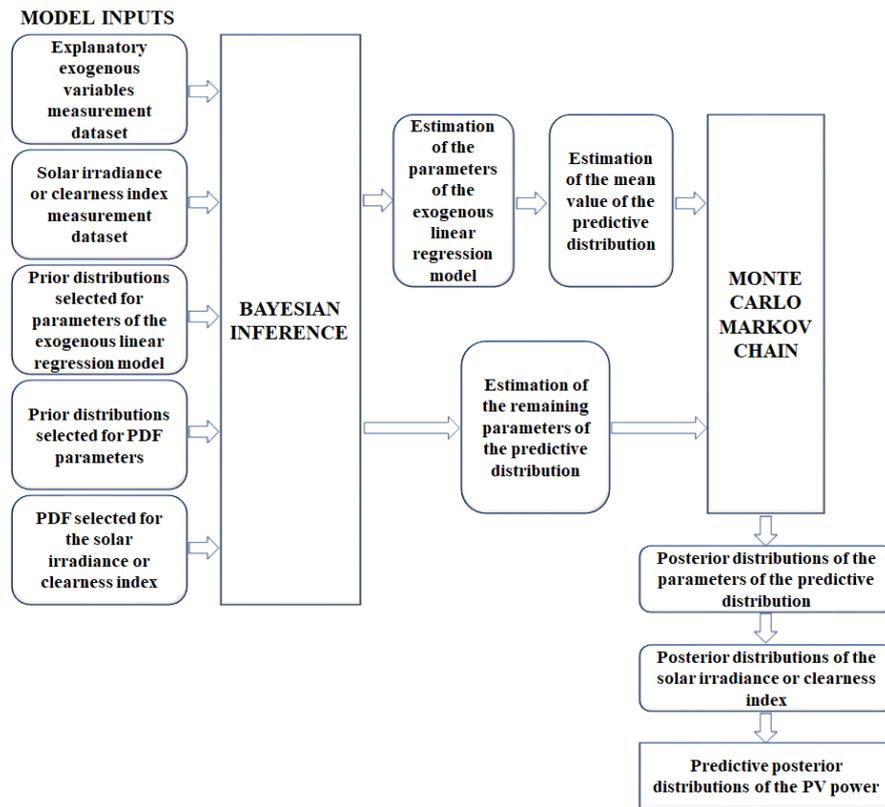


Figure 2.2 - Flow chart of the probabilistic Bayesian method.

2.3.2. Numerical applications

The probabilistic Bayesian method shown in Section 2.3.1 was tested with reference to the active power generated by an a-Si:H thin-film PV system of rated power $P_{PV_r} = 75$ kW. The total surface of PV modules was $S_{PV} = 700$ m², with an efficiency value of $\eta_{PV} = 0.09$ for the whole PV system. The forecasts were made by using both of the Bayesian-based approaches mentioned in Section 2.3.1, i.e.,:

- method (a): the Bayesian-based approach that uses the hourly solar irradiance;
- method (b): the Bayesian-based approach that uses the hourly clearness index.

The assessment of the quality of the forecasts was performed in both a spot-value framework, by selecting the mean value of the predictive distribution of PV power as spot-value forecast, and in a probabilistic framework. In particular, the indices used were the Mean Absolute Error (MAE), the RMSE, their normalized versions (NMAE and NRMSE), and the CRPS. Also, reliability diagrams were provided to assess the overall concordance between nominal coverages and estimated coverages of the probabilistic forecasts⁷.

The results of the proposed method were compared with benchmark values obtained through the well-known persistence method (PM), applied both in a spot-value and in a probabilistic frameworks [145,146]; in the last framework, the probability function was assumed to be a modified Gamma distribution.

The forecasts were performed for several lead times, from $k = 1$ to $k = 24$. For the sake of conciseness, only the results for $k = 24$ (*next day forecasts*) are shown.

Information concerning the data that was used to test the Bayesian method is presented in the following sub-Section 2.3.2.1. Then, the forecasting results are presented and discussed in sub-Section 2.3.2.2.

2.3.2.1. Data characteristics

The set of available measurements consists of observations of the meteorological variables (i.e., solar irradiance, clearness index, relative

⁷ Details on such indices and tools to assess the quality of forecasts are in the Appendix.

humidity, cloud cover, ambient pressure, and ambient temperature) collected in Pignataro Maggiore, Italy (latitude: 41.12° north and longitude: 14.10° east). Data was collected with a 15-minutes time resolution, from 7:00 to 20:00, from 1 August 2012 to 28 February 2013. In order to obtain hourly values, the original dataset was then averaged to obtain the values required for the models.

An “off-line” cross-correlation analysis between the solar irradiance/clearness index and the other meteorological variables was performed to individuate the most suitable underlying exogenous linear regression model. Fig. 2.3 shows the correlation coefficients between the solar irradiance (Fig. 2.3 a), clearness index (Fig. 2.3 b), and the other weather variables. Only the variables that provided the highest values of the cross-correlation coefficient were taken into account in the linear regression models. Total cloud cover and relative humidity are the most-correlated weather variables on both the solar irradiance and the clearness index; consequently, the linear regression models included only solar irradiance, total cloud cover, and relative humidity for method (a) and clearness index, total cloud cover, and relative humidity for method (b).

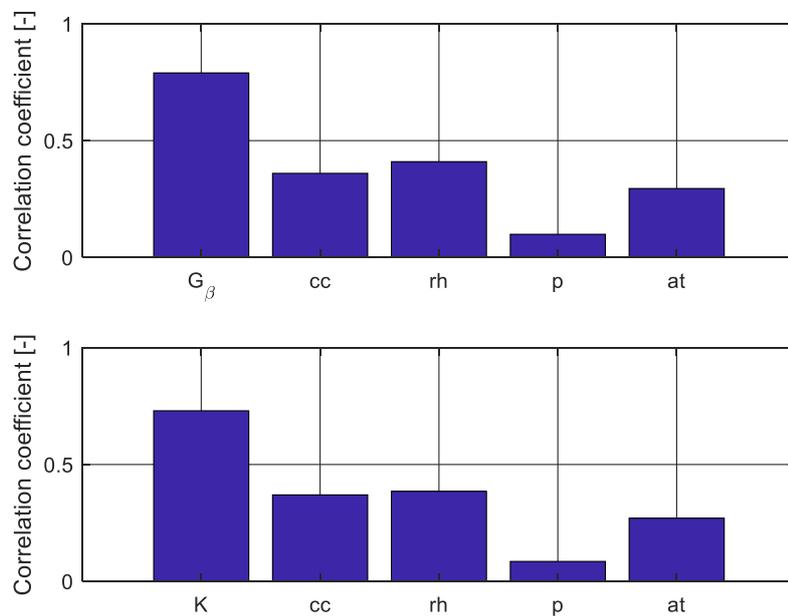


Figure 2.3 - Cross-correlation coefficients between solar irradiance (a), clearness index (b) and the available weather variables.

The measurements of the hourly solar irradiance, the hourly clearness index, the relative humidity and the cloud cover are contained in the sets $\mathbf{S}_{G\beta_h}, \mathbf{S}_{K_h}, \mathbf{S}_{RH_h}, \mathbf{S}_{CC_h}$, which included $N_{tr}^{(G)} = N_{tr}^{(K)} = 20$ observations collected at homologous hours of the day before the forecast start time $h - k$, to make inference about the prior random parameters, i.e., $\beta_0^{(G)}, \dots, \beta_3^{(G)}$, of the model in Eq. (2.12) and $\beta_0^{(K)}, \dots, \beta_3^{(K)}$ of the model in Eq. (2.13),

All of the prior distributions used in the BI are Gaussian distributions with large variance equal to 10^4 ; in particular, for method (a), they were assumed to be:

$$p(\beta_0^{(G)}) = N(0, 10^4), \quad p(\beta_1^{(G)}) = N(0.5, 10^4), \quad p(\beta_2^{(G)}) = N(0, 10^4), \\ p(\beta_3^{(G)}) = N(0, 10^4), \quad p(\sigma_{BETA_h}) = N(3, 10^4);$$

for method (b), they were assumed to be:

$$p(\beta_0^{(K)}) = N(0, 10^4), \quad p(\beta_1^{(K)}) = N(0.5, 10^4), \quad p(\beta_2^{(K)}) = N(0, 10^4), \\ p(\beta_3^{(K)}) = N(0, 10^4).$$

2.3.2.2. Assessment of the quality of forecasts

The results of the next day forecast were presented in details with reference to four specific days (e.g., 21 August 2012, 23 September 2012, 29 November 2012, and 22 December 2012), characterized by different meteorological conditions. In particular, the first day is a “clear-sky day” with no appreciable cloud cover, while the second day is slightly more unstable in terms of cloud cover; the latter two days are instead cloudy days characterized by very unstable weather conditions. Then, the results of forecasts performed in all of the test months (i.e., from September 2012 to February 2013) are shown in a more compact form.

With reference to the spot-value framework, Figs. 2.4 and 2.5 show the results of next day forecast performed through the Bayesian-based approaches, through the PM, and the actual hourly PV powers. In particular, Figs. 2.4a and 2.4b refer to 21 August and 23 September, respectively, and Figs. 2.5a and 2.5b refer to 29 November and 22 December, respectively.

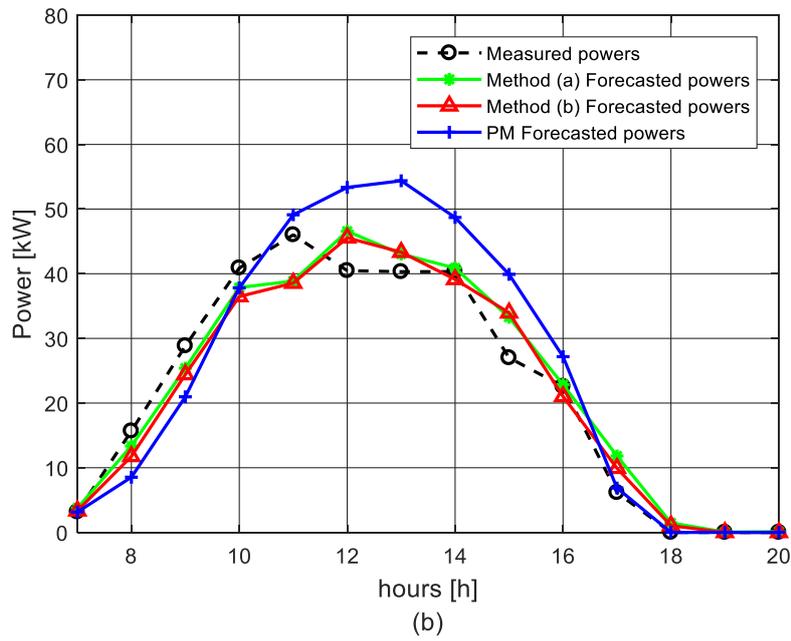
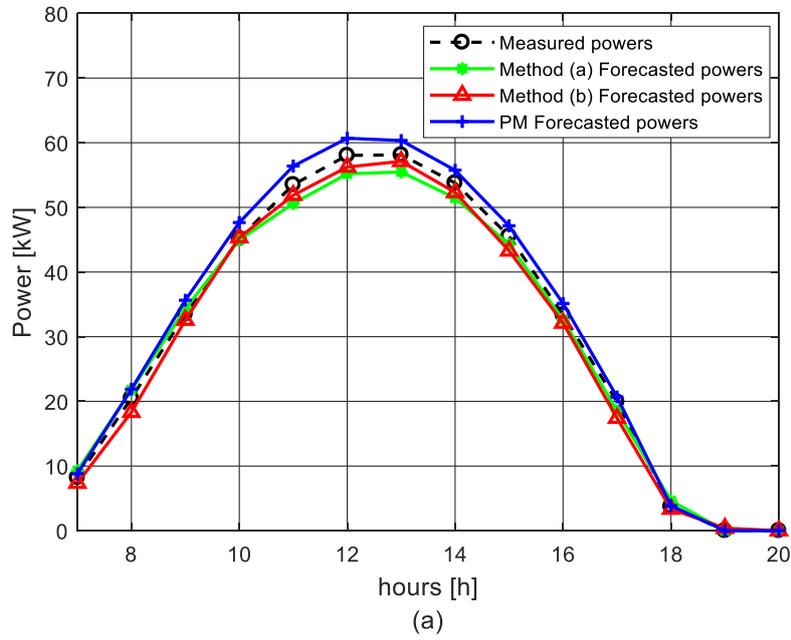


Figure 2.4 - Next day PV power forecasts obtained through the Bayesian approaches, the PM method (solid lines), and actual PV power (dash line) on (a) 21 August 2012; (b) 23 September 2012.

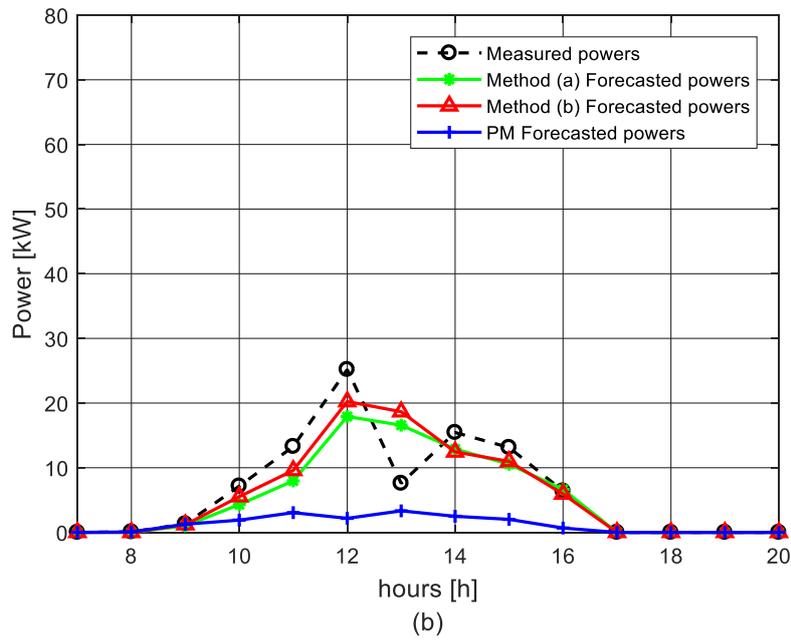
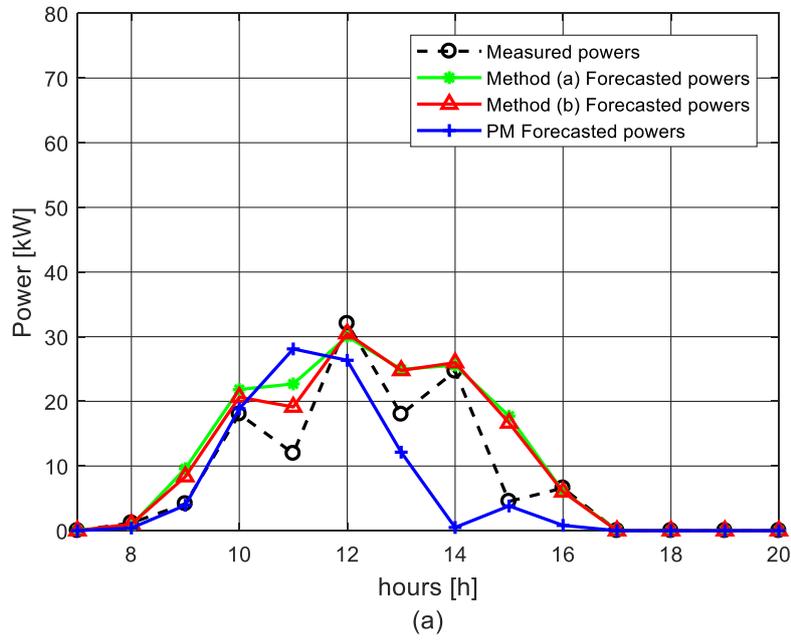


Figure 2.5 - Next day PV power forecasts obtained through the Bayesian approaches, the PM method (solid lines), and actual PV power (dash line) on (a) 29 November 2012; (b) 22 December 2012.

The analysis Figs. 2.4 and 2.5 indicated that the absolute error is always smaller than 5 kW during the clearest day (Fig. 2.4a), while it is quite greater (up to 7 kW) during the more variable day (Fig. 2.4b). The error strongly increases in the case of unstable weather conditions, in which the proposed method shows absolute errors up to about 12 kW (at 15:00, see Fig. 2.5a). The performance of the PM was particularly poor in unstable meteorological conditions (with a peak of absolute error greater than 22 kW at 14:00 on 29 November (Fig. 2.5a) and at 12:00 on 22 December (Fig. 2.5b).

For comparative purposes, Table 2.1 reports the mean values of the spot-value error indices, for all of the methods and for the four considered days; also the corresponding normalized versions of these indices are shown, to provide a scale-free reference. All of the indices were normalized to the rated power P_{PV_r} of the PV system (75 kW).

Table 2.1 – Spot-value error indices obtained through Bayesian method (a), Bayesian method (b), and the persistence method for the considered days

Day	Method	RMSE [kW]	MAE [kW]	NRMSE [%]	NMAE [%]
21 August	Method (a)	1.64	1.32	2.18	1.76
	Method (b)	1.43	1.20	1.91	1.60
	PM	1.76	1.47	2.35	1.96
23 September	Method (a)	3.74	2.84	4.99	3.78
	Method (b)	3.90	3.07	5.20	4.09
	PM	7.34	5.34	9.78	7.12
29 November	Method (a)	7.26	4.16	9.68	5.55
	Method (b)	6.42	3.61	8.56	4.81
	PM	10.23	5.30	13.64	7.07
22 December	Method (a)	3.62	2.16	4.82	2.88
	Method (b)	3.56	1.94	4.74	2.59
	PM	8.47	5.19	11.29	6.91

The analyses of the indices in Table 2.1 shows that:

- the values of spot-value indices obtained with method (a) were greater than the corresponding values obtained with method (b) in three upon four days, with the only exception on 23 September (a slightly unstable day).
- All error indices, as foreseeable, were greater during the unstable meteorological conditions, since the hourly-varying conditions had a negative impact on the forecasts. However, the proposed

methods showed their usefulness also in these conditions, since they significantly outperformed the PM. In particular, the quality gap between the proposed methods and the reference method is more enhanced than in the clear day conditions. For example, MAE obtained through method (b) was 17% smaller than MAE obtained with PM during the clear day (21 August), while MAE obtained through method (b) was 32% smaller than MAE obtained with PM during the most unstable meteorological conditions (on 29 November).

To verify the proposed methods on a larger number of forecasts, the values of the spot-value error indices, averaged in the considered test months, are shown in Table 2.2 for all of the methods.

Table 2.2 – Spot-value error indices obtained through Bayesian method (a), Bayesian method (b), and the persistence method for the considered test set

Period	Method	RMSE [kW]	MAE [kW]	NRMSE [%]	NMAE [%]
Entire test set	Method (a)	8.38	5.03	11.18	6.71
	Method (b)	8.14	4.79	10.86	6.39
	PM	11.70	6.49	15.60	8.65

Both Bayesian methods have, on average, better performances than PM. In particular, method (b) showed the best performance, since its MAE and RMSE decreased by 26% and 30%, respectively, while the MAE and RMSE of method (a) decreased by 23% and 29%, respectively.

With reference to the probabilistic framework, Table 2.3 reports the mean values of the CRPS, together with the corresponding normalized version, for all of the methods and for both forecasted days.

Table 2.3 - Probabilistic error indices obtained through Bayesian method (a), Bayesian method (b), and the persistence method for the considered days

Day	Method	CRPS [kW]	NCRPS [%]
21 August	Method (a)	1.74	2.32
	Method (b)	0.87	1.16
	PM	1.14	1.52
23 September	Method (a)	2.26	3.01
	Method (b)	2.31	3.08
	PM	3.96	5.28
29 November	Method (a)	2.73	3.64
	Method (b)	2.42	3.23
	PM	4.80	6.40
22 December	Method (a)	2.21	2.95
	Method (b)	2.09	2.79
	PM	4.23	5.64

The analyses of the indices in Table 2.3 shows that:

- the behavior in the probabilistic framework is similar to the behavior in the spot-value framework, since the probabilistic index CRPS is greater for method (a) in three upon four days; it means that the hourly CDFs of PV power in case (b) better fit the CDF of the measured power, delimiting smaller areas of error. The maximum improvement in terms of CRPS with respect to the benchmark occurs in the most unstable day (29 November), and it is about 60%.
- The CRPS values usually were small during the clearest day (less than 2.5% of the rated power); this means that the forecasted predictive distributions are sharp and concentrated around the actual PV power. The CRPS values instead were greater during the day with more unstable weather conditions (more than 3.5% of the rated power).
- Although in the spot-value framework method (a) always outperformed the benchmark, this does not happen in the probabilistic framework, since the CRPS calculated with PM was smaller than the CRPS obtained through the method (a) in the clearest sky day, while it was significantly greater than it during the unstable days. In particular, the reduction in terms of CRPS in the unstable days was up than 40%.

To verify the proposed methods in the probabilistic framework on a larger number of forecasts, the values of the CRPS and NCRPS, averaged in the considered test months, are shown in Table 2.4 for all of the methods. Results are similar to the spot-value framework, with method (b) outperforming method (a) and the benchmark by 7% and 39%, respectively.

Table 2.4 - Probabilistic error indices obtained through Bayesian method (a), Bayesian method (b), and the persistence method for the considered test set.

Period	Method	CRPS [kW]	NCRPS [%]
Entire test set	Method (a)	3.51	4.68
	Method (b)	3.27	4.36
	PM	5.34	7.12

For the sake of completeness, the reliability diagrams for the proposed methods are shown in Fig. 2.6. Given the observed values of the PV power, the empirical coverages of several quantiles were compared to the nominal coverages. The figure indicates that method (b) had better performance also in terms of reliability than method (a), since noticeable deviations from perfect reliability occur only under 50% of nominal coverages. Method (a) deviated from the ideal reliability with both positive and negative differences. The maximum deviation from perfect reliability was 5.12% for method (a), and 3.78% for method (b).

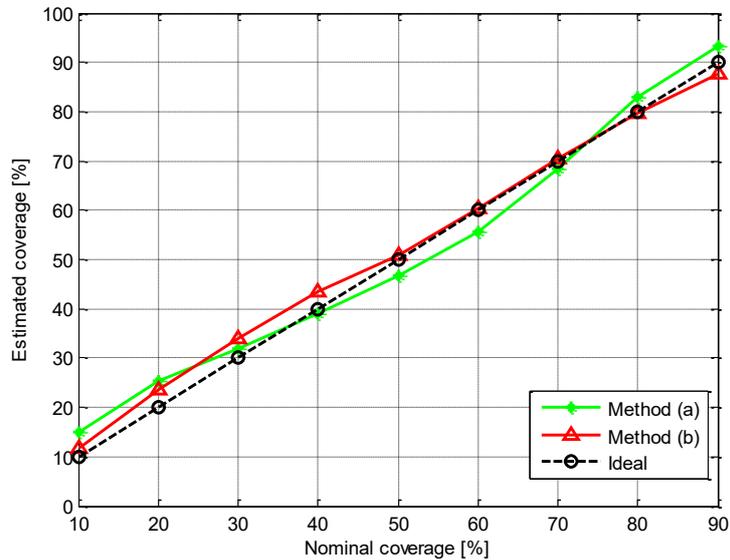


Figure 2.6 - Reliability diagrams of the proposed Bayesian methods. Estimated coverages (solid lines) are compared to the ideal coverages (dash line).

2.4. A NEW PROBABILISTIC ENSEMBLE METHOD FOR SHORT-TERM PHOTOVOLTAIC POWER FORECASTING

A competitive, multi-model ensemble method [100] is proposed aiming to improve the quality of forecasts obtained through some individual probabilistic predictors.

Details about the proposed ensemble method are reported in the following sub-Section 2.4.1; numerical applications based on real data are in sub-Section 2.4.2.

2.4.1. Proposed method

As stated in Chapter 1, probabilistic predictors may vary in terms of predictive outputs; e.g., they can provide predictive samples, predictive PDFs, or predictive quantiles. For this reason, in the proposed ensemble method [100] the forecasts obtained through base predictors are firstly properly combined through a linear pooling of predictive CDFs [98]. Then, in order to guarantee elevate sharpness and reliability characteristics, a MO optimization system is applied during the training period in order to estimate coefficients of the linear pooling. The MO optimization is specifically devoted to overcome

well-known problems resulting from the probabilistic combination of probabilistic forecasts in the linear pool approach [98]. A Bayesian Method (BM), a Markov chain Method (MM), and a Quantile regression Method (QM) are selected as probabilistic base predictors. Eventually, the proposed ensemble method involves the following steps:

- i. selection of the relationships that link the PV active power to the solar irradiance (*sub-Section 2.4.1.1*);
- ii. selection of the probabilistic base predictors (*sub-Section 2.4.1.2*);
- iii. proper combination of probabilistic forecasts obtained through base predictors in a MO optimization procedure (*sub-Section 2.4.1.3*).

2.4.1.1. Relationship that links the PV active power with the solar irradiance

The relationship that links the output active power P_{PV_h} of the PV system at the time horizon h to the solar irradiance I_{β_h} on a surface with an inclination of β degrees with respect to the horizontal plane is the same in Eq. (2.1). This relationship was selected due its linearity, in order to reduce the overall computational effort, although with no loss of generality.

2.4.1.2. Selection of probabilistic base predictors

A BM, MM and QM were selected as probabilistic base predictors; their diversity in both the types of probabilistic output and model formulations was of use in testing the functionalities of the ensemble method.

a) Bayesian probabilistic base predictor

A slightly-modified version of the BM [99], deeply described in sub-Section 2.3.1, is the first probabilistic base predictor selected for the multi-model ensemble method. In particular, the approach (a) was selected, considering the solar irradiance modeled through a Beta distribution (as in Eq. (2.6)); parameters of the linear regression model (2.12) are deterministically estimated through the least square method,

while the BI is applied to probabilistically estimate the posterior distribution of the shape parameter σ_{BETA_h} .

It is here reminded that the BM provides as output samples from the predictive posterior PDF of PV power, and the discrete predictive CDF is straightforwardly obtained.

(b) Markov chain probabilistic base predictor

The non-parametric, time-discrete MM, originally proposed in [25] for wind power forecasting, was used as second base predictor to forecast PV power. The main difference between BM and MM is that no prior assumption concerning the statistical characterization of the random variable is needed in MM. Instead, it is assumed that the PV power P_{PV_h} at the forecast time horizon h can assume a value in N_S states, $S_h^{(1)}, \dots, S_h^{(N_S)}$, with each state corresponding to an interval of possible values of PV power. In particular, assigned the rated power P_{PV_r} , the first and last states are assumed to be $S_h^{(1)} = 0$ and $S_h^{(N_S)} = P_{PV_r}$, whereas the remaining $N_S - 2$ states are $N_S - 2$ equally-spaced intervals from 0 to the rated power P_{PV_r} . Note that N_S is a calibration parameter for the MM.

Markov chains are based on the hypothesis that each unknown state S is statistically dependent only on its s previously-observed states (s^{th} order Markov chain)⁸. For PV power forecasting, accurate results were obtained experimentally for $s \geq 2$. For the sake of conciseness, the case $s = 2$ (second order Markov chain) is discussed here, being trivial the extension to higher orders. The formulation of stochastic independence after the second time step is:

$$p(S_h | S_{h-k}, S_{h-2k}, \dots, S_1) = p(S_h | S_{h-k}, S_{h-2k}), \quad (2.18)$$

⁸ The previously-observed states are conveniently selected on the basis of the forecast lead time k , e.g., for next hour forecasting and for $s = 2$, the previous observed states refer to one and two hours before, while for next day forecasting and for $s = 2$ the previous observed states refer to one and two days before. Please note straightforwardly that in the notation used in this thesis the step of MM time discretization is the same value as the forecast lead time, but this is not the unique solution; indeed, things change if the forecast lead time varies (e.g., for 2-days ahead forecasting and for $s = 2$, the previous observed states refer to two and three days before).

and the forecast for the state S_h is obtained by calculating an estimation $\hat{\mathbf{A}}$ of the second order transition matrix \mathbf{A} of size $[N_S \times N_S \times N_S]$. The element a_{ijl} of the matrix \mathbf{A} is the probability $p(S_h^{(i)} | S_{h-k}^{(j)}, S_{h-2k}^{(l)})$, i.e., the probability that the PV power is in the i^{th} state at time $t = h$, given that the j^{th} state was observed at time $t = h - k$ and the l^{th} state was observed at time $t = h - 2k$. Obviously, only estimations \hat{a}_{ijl} of a_{ijl} can be provided; given the observations of PV power in a training period, the maximum likelihood estimates are:

$$\hat{a}_{ijl} = \frac{n_{ijl}}{\sum_{i=1}^{N_S} n_{ijl}}, \quad (2.19)$$

where n_{ijl} is the number of observed occurrences of the transition between consecutive states i, j, l during the training period of the MM. Also, if $\sum_{i=1}^{N_S} n_{ijl} = 0$ (i.e., no consecutive states j, l have ever been observed in the training period), it is common to set:

$$\hat{a}_{ijl} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}, \quad (2.20)$$

giving credit to the persistence assumption. With these positions, it results that $\sum_{i=1}^{N_S} \hat{a}_{ijl} = 1 \forall j, l$, and, therefore, $\hat{\mathbf{A}}$ is a valid estimate of the Markov chain transition matrix. It should be noted that an accurate estimation of $\hat{\mathbf{A}}$ can be obtained only using a significant number of previously observed samples. Once $\hat{\mathbf{A}}$ is known, it is trivial to obtain forecasts for the PV power state at the desired time horizon; in particular, the output of MM is the vector $\boldsymbol{\pi}_h = \{\pi_h^{(1)}, \dots, \pi_h^{(N_S)}\}$ of state probabilities at time $t = h$, given the vectors $\boldsymbol{\pi}_{h-k}$ and $\boldsymbol{\pi}_{h-2k}$ of state probabilities at time $t = h - k$ and $t = h - 2k$, respectively. The i^{th} element $\pi_h^{(i)}$ of the vector $\boldsymbol{\pi}_h$ is simply obtained as follows:

$$\pi_h^{(i)} = \sum_{j=1}^{N_S} \sum_{l=1}^{N_S} \hat{a}_{ijl} \cdot \pi_{h-k}^{(j)} \cdot \pi_{h-2k}^{(l)}. \quad (2.21)$$

Obviously, in the first step of MM, both vectors $\boldsymbol{\pi}_{h-k}$ and $\boldsymbol{\pi}_{h-2k}$ are known, and, in each of them, all elements are zero except the unitary element that corresponds to the observed state at times $t = h - k$ and $t = h - 2k$. Eventually, if necessary, the procedure is repeated by updating vectors $\boldsymbol{\pi}_{h-k}$ and $\boldsymbol{\pi}_{h-2k}$ at each iteration. Samples of PV power can be drawn from the predicted state probabilities for the desired time horizon [147]. Then, as for BM, the discrete predictive CDF is directly obtained from the predictive samples of PV power.

(c) Quantile regression probabilistic base predictor

Also the QM, selected as third probabilistic base predictor, does not rely on prior hypothesis about the distribution of the PV power data [148-150]. Indeed, instead of forecasting the parameters of a selected PDF for the desired time horizon h , the outputs of QM are J estimated quantiles of PV power. The input of the model is the column vector of m_5 explanatory variables, $\boldsymbol{x}_h = \{x_{h_1}, \dots, x_{h_{m_5}}\}$; in the most general form, the α_q -quantile $P_{PV_h}^{(\alpha_q)}$ of PV power can be expressed as:

$$P_{PV_h}^{(\alpha_q)} = \boldsymbol{\beta}^{(\alpha_q)} \cdot \boldsymbol{x}_h + r_h, \quad (2.22)$$

where $\boldsymbol{\beta}^{(\alpha_q)}$ is a row vector of m_5 coefficients to be estimated, and r_h is a residual white noise at time $t = h$. Obviously, the main problem is to find an estimation $\widehat{\boldsymbol{\beta}}^{(\alpha_q)}$ of vector $\boldsymbol{\beta}^{(\alpha_q)}$, since the expected value $\widehat{P}_{PV_h}^{(\alpha_q)}$ of $P_{PV_h}^{(\alpha_q)}$ is given by:

$$\widehat{P}_{PV_h}^{(\alpha_q)} = \widehat{\boldsymbol{\beta}}^{(\alpha_q)} \cdot \boldsymbol{x}_h + r_h. \quad (2.23)$$

If the dataset $\{P_{PV_{h-k-(N_{tr}^{(QM)}-1)}}, \dots, P_{PV_{h-k}}\}$ of $N_{tr}^{(QM)}$ past measurements of PV power and the $N_{tr}^{(QM)}$ vectors of explanatory variables are available for a given training period, $\widehat{\boldsymbol{\beta}}^{(\alpha_q)}$ is obtained by minimizing the sum of values $PLF_i^{(\alpha_q)}$:

$$\widehat{\boldsymbol{\beta}}^{(\alpha_q)} = \arg \min_{\boldsymbol{\beta}^{(\alpha_q)}} \sum_{i=1}^{N_{tr}^{(QM)}} PLF_i^{(\alpha_q)}, \quad (2.24)$$

where each value $PLF_i^{(\alpha_q)}$ is calculated as⁹:

$$PLF_i^{(\alpha_q)} = \begin{cases} (\alpha_q - 1) \cdot (P_i - \boldsymbol{\beta}^{(\alpha_q)} \cdot \mathbf{x}_i) & \text{if } P_i < \boldsymbol{\beta}^{(\alpha_q)} \cdot \mathbf{x}_i \\ \alpha_q \cdot (P_i - \boldsymbol{\beta}^{(\alpha_q)} \cdot \mathbf{y}_i) & \text{if } P_i \geq \boldsymbol{\beta}^{(\alpha_q)} \cdot \mathbf{x}_i \end{cases}, \quad (2.25)$$

An elegant solution to problem (2.25) can be found in [149]. As in [150], 0-quantile and 1-quantile are set to 0 and P_{PV_r} , respectively. Once the selected J quantiles of PV power are estimated, the predictive CDF can be obtained through linear interpolation.

2.4.1.3. Processing the outputs of single probabilistic base predictor

Base predictors are combined through the combination of predictive CDFs, since they are easier to manage when different kinds of predictors are to be aggregated. One of the acknowledged methods for the combination of predictive CDFs is the Linear Pool Ensemble (LPE) method [98,151,152]; the ensemble predictive CDF $F_{LPE_h}(P_{PV_h})$ of the PV power for the horizon time h can be obtained as a weighted sum of each CDF $F_{h_i}(P_{PV_h})$:

$$F_{LPE_h}(P_{PV_h}) = \sum_{i=1}^{N_{bp}} w_n \cdot F_{h_i}(P_{PV_h}), \quad (2.26)$$

where N_{bp} is the number of base predictors (in this case, $N_{bp} = 3$), $0 \leq w_i \leq 1$, and the sum of weights is $\sum_{i=1}^{N_{bp}} w_i = 1$ in order to guarantee that the output function is a CDF defined in the interval $[0, P_{PV_r}]$. In [98] it was pointed out that LPE predictions may be over-dispersed¹⁰ if single predictors are neutrally dispersed. Since neutral dispersion is a necessary condition for calibration and since the aim of the forecaster is to produce sharp forecasts, subject to calibration, some techniques shown in [98] may be applied to overcome this problem.,

⁹ Note that Eq. (2.25) is the same of Eq. (A12) in the Appendix.

¹⁰ Exhaustive definitions of over-dispersion, neutral dispersion, and under-dispersion of forecasts are provided in the Appendix of this thesis.

Instead, a new approach based on a MO optimization problem is proposed in the ensemble method shown in this thesis.

Obviously, the main problem in applying relationship (2.26) lies in the evaluation of weights $\mathbf{w} = \{w_1, \dots, w_{N_{bp}}\}$. An elegant solution to the problem was proposed in the relevant literature [153], and it has been extensively used in approaches that involve probabilistic forecasting of weather variables, such as temperature and precipitation. In these applications, the weight coefficients are estimated by solving a single-objective (SO) optimization problem that consists in minimizing a proper score (objective function) during a training period; in particular, the CRPS has been considered in many applications as an adequate objective function to be minimized¹¹.

The analytic formulation of the hourly CRPS for the forecasted CDF $F_h(P_{PV_h})$ at time $t = h$ is recalled here, for sake of clearness:

$$CRPS_h(F_h) = \int_0^{P_{PV_r}} [F_h(P_{PV_h}) - H(P_{PV_h} - P_{PV_h}^*)]^2 \cdot dP_{PV_h}, \quad (2.27)$$

where $P_{PV_h}^*$ is the actual value of PV power at time horizon h , and $H(P_{PV_h} - P_{PV_h}^*)$ is the Heaviside function centered in $P_{PV_h}^*$. If $F_h(P_{PV_h})$ is the ensemble CDF $F_{LPE_h}(P_{PV_h})$ in Eq. (2.26), the SO minimization problem to be solved is:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \sum_{h=1}^{N_{tf}} CRPS_h \left(F_{LPE_h} | w_1, \dots, w_{N_{bp}} \right) \\ \text{s. t.} \quad & w_n \geq 0 \quad \forall n, \\ & \sum_{n=1}^{N_{bp}} w_n = 1 \end{aligned}, \quad (2.28)$$

where N_{tf} is the total number of forecasts performed during the training period.

However, even if the proposed approach based on the minimization of CRPS seems to be appropriate, reliability should be explicitly taken

¹¹ The choice of CRPS is not univocal. Indeed, other proper scores could be used in the optimization procedure, such as the PLF. However, the results of numerical applications performed with the PLF are very similar to those obtained with CRPS, according to naïve intuition.

into account in dealing with an ensemble probabilistic forecast. It is known that the CRPS value can be decomposed into three terms, one of which is related to the reliability of the forecast [154]. However, the sharpness and reliability components of the CRPS cannot be separately identified, as one influences each other. Thus, minimizing an entire proper score in the SO problem without splitting it into its single contributions can lead to unreliable forecasts (as shown in the numerical application in sub-Section 2.4.2). Indeed, in [18], it was shown that a reduction in sharpness may increase the reliability of the forecast and vice versa. The same problem arose also in [119]; in that case, the PLF was used in the evaluation procedure, and it was stated that “...the score is to be seen as a proper skill score, hence allowing to objectively sort rival forecasting approaches. However, both reliability and sharpness should also be examined separately and visually via reliability diagrams and quantiles' plots to make sure there is a reasonable balance between these two attributes.”

Hence, the proposal of this Thesis is to estimate optimal values $\hat{\mathbf{w}}$ of coefficients \mathbf{w} by solving the following MO optimization problem, instead of the SO minimization problem (2.28):

$$\begin{aligned} \min_{\mathbf{w}} & \left[\phi_1 \left(F_{LPE_h} | w_1, \dots, w_{N_{bp}} \right), \phi_2 \left(F_{LPE_h} | w_1, \dots, w_{N_{bp}} \right) \right] \\ \text{s. t.} & \quad w_i \geq 0 \quad \forall i = 1, \dots, N_{bp} \\ & \quad \sum_{i=1}^{N_{bp}} w_i = 1 \end{aligned}, \quad (2.29)$$

where ϕ_1 and ϕ_2 are the objective functions to be minimized. In particular, ϕ_1 is the average value of the hourly CRPS scores (2.27) in the total number of forecasts performed during the training period, and ϕ_2 is an index that takes into account the deviations from perfect reliability in the same conditions.

In order to find an index that explicitly takes into account the reliability (function ϕ_2 in Eq. (2.29)), it is important to observe that the reliability of a probabilistic forecast may be estimated in terms of consistency between estimated coverages of prediction and nominal coverages [151]; ideal reliability is obtained when all estimated coverages are equal to the corresponding nominal coverages.

The definition of coverage is the following. The α_q -quantile $\hat{P}_{PV_{LPE_h}}^{(\alpha_q)}$ extracted from the ensemble prediction is compared to the corresponding actual value of PV power $P_{PV_h}^*$. The binary indicator $\hat{\Lambda}_{LPE_h}^{(\alpha_q)}$ is conveniently introduced as:

$$\hat{\Lambda}_{LPE_h}^{(\alpha_q)} = \begin{cases} 1 & \text{if } P_{PV_h}^* \leq \hat{P}_{PV_{LPE_h}}^{(\alpha_q)} \\ 0 & \text{if } P_{PV_h}^* > \hat{P}_{PV_{LPE_h}}^{(\alpha_q)} \end{cases}, \quad (2.30)$$

and an estimation $\hat{\alpha}_{q_{LPE}}$ of the nominal coverage α_q is obtained as:

$$\hat{\alpha}_{q_{LPE}} = \frac{1}{N_{tf}} \sum_{h=1}^{N_{tf}} \hat{\Lambda}_{LPE_h}^{(\alpha_q)}, \quad (2.31)$$

Then, to take into account the reliability of the forecasts, the maximum value ψ of all absolute deviations between the estimated and actual coverages can be introduced in the MO optimization problem as the objective function ϕ_2 to be minimized. An alternative objective function ϕ_2 could be the sum of all absolute deviations between the estimated and actual coverages¹². Both indices have been used in relevant literature to provide estimations of the forecast reliability [119].

Once the objective functions and constraints are known, an algorithm to solve the MO optimization problem (2.29) must be used to determine a suitable solution (values $\hat{\mathbf{w}}$ of coefficients \mathbf{w}).

Indeed, the MO problem (2.29) does not have a unique solution, but a set of points are available, all of which fit a pre-determined definition of an optimum. Pareto optimality is the predominant concept in defining an optimal point.

Theoretically, there is an infinite number of Pareto optimal solutions (Pareto front), so it is often necessary to incorporate the forecaster's preferences to obtain a single, suitable solution. The methods used to obtain a solution can be classified on the basis of the criteria used to

¹² Numerical applications suggest that the results obtained with maximum deviation and sum of deviations are quite similar.

articulate preferences and methods with prior articulation, with posterior articulation, and with no articulation of preferences. In the methods with prior articulation of preferences, the forecaster specifies its preferences, typically in terms of the relative importance of different objectives whose values reflect forecaster preferences [155]. This is the approach followed in the solution of the MO problem (2.29).

The most popular method with prior articulation of preferences is the weighted sum method, which allows the MO optimization problem (2.29) to be solved as an equivalent SO minimization problem formulated as:

$$\begin{aligned} \min_{\mathbf{w}} & \left[\xi_1^{(MO)} \phi_1(F_{LPE_h} | \mathbf{w}) + \xi_2^{(MO)} \phi_2(F_{LPE_h} | \mathbf{w}) \right] \\ \text{s. t.} & \quad w_i \geq 0 \quad \forall i = 1, \dots, N_{bp} \\ & \quad \sum_{i=1}^{N_{bp}} w_i = 1, \end{aligned} \quad (2.32)$$

where $\xi_1^{(MO)}$ and $\xi_2^{(MO)}$ are appropriate positive weights, the values of which reflect the relative importance of the objective functions. Typically, the sum of weights $\xi_1^{(MO)} + \xi_2^{(MO)}$ is equal to 1. Several methods can be used to choose the weights $\xi_1^{(MO)}$ and $\xi_2^{(MO)}$; the simplest procedure is the same weight estimation, considering all weights equal when no prior articulation of preference is made by the forecaster. Other suitable procedures can be applied when the forecaster is able to articulate preferences (i.e., the rank order of the true weights and the rank sum method). More details on the methods of choosing the weights can be found in [155].

Note that the objective functions in problem (2.32) usually must be normalized to provide a uniformly distributed solution set; in the appliance to the considered case, the normalization was achieved by assuming as references the mean values of CRPS and the maximum values of all absolute deviations from perfect reliability calculated on the three basic predictors.

2.4.2. Numerical applications

The multi-model ensemble procedure shown in the previous sub-Section was applied for the probabilistic forecast of a $P_{PV_r} = 110$ kW PV generator at a U.S. site.

Different values of lead time k were considered in the numerical applications; for the sake of conciseness, only the results for $k = 24$ hours (*next day forecast*) are shown in this sub-Section. Moreover, the performances of single predictors and of the ensemble predictor were compared with a benchmark method, i.e., the probabilistic extension of the PM.

Comparisons among methods were conducted in terms of probabilistic indices and diagrams; in particular, CRPS, reliability diagrams, and Probability histograms were used to check the goodness of the forecasts. The same-weight estimation was used in the MO procedure; also, the SO procedure that minimizes only the CRPS (i.e., the limit case of the MO procedure with weights $\xi_1^{(MO)} = 1$ and $\xi_2^{(MO)} = 0$) was considered for comparative purposes.

In the next subsections, firstly the input data characteristics are shown. Then, results of forecasts performed for one month for next-day forecasting are presented with extensive details, showing also reliability diagrams and Probability Integral Transform (PIT) histograms. Finally, the results of several further months are shown in a more compact form, for sake of conciseness.

2.4.2.1. Data characteristics

Input data were collected from January 1, 2013 to December 31, 2014 at the latitude of 39.74° north and the longitude of 105.18° west by the U.S. National Renewable Energy Laboratory [156]. Available measurements consisted of 17520 hourly observations of solar irradiance, air temperature, cloud cover, relative humidity, and wind speed. The latter measurements were used as explanatory variables in the underlying regression model of the BM, and in the regression in the QM. Only 14 daytime hours (from 6:00 A.M. to 8:00 P.M.) were considered in the PV power forecasting application.

2.4.2.2. Assessment of the quality of forecasts

The results of forecasts made for a lead time $k = 24$ hour are shown in this sub-Section. For the sake of conciseness, results are exhaustively shown only for one month, i.e., for November 2014. Eleven months (from 01/11/2013 to 30/09/2014) were used as the training interval for each predictor and the following month (from 01/10/2014 to 31/10/2014) was used to calibrate base predictors and to find the weights of the ensemble method. Only air temperature and cloud cover were selected as explanatory variables.

Table 2.5 shows the estimated values $\hat{\mathbf{w}}$ of coefficients \mathbf{w} for the LPE method. Table 2.6 shows the CRPS values for single predictors, LPE, PM and for the SO procedure.

As shown in Table 2.5, weights are quite uniformly distributed for three base predictors (with a rough proportion of 0.3, 0.3 and 0.4 for BM, MM and QM, respectively). With these values of weight coefficients, LPE provides forecasts with a reduced CRPS than base predictors (from about 4% to 8% with respect to QM and BM, respectively). The SO procedure provided a smaller value of CRPS than the MO procedure (about 3% smaller), but performance in terms of reliability was significantly poorer (maximum deviation increased up to 10.95%). Also, the ensemble method outperformed the PM by 50%.

Table 2.5 - Next-day forecasting. Values of weight coefficients for linear ensemble in November 2014

Ensemble method	BM weight \hat{w}_1 [-]	MM weight \hat{w}_2 [-]	QM weight \hat{w}_3 [-]
LPE - MO	0.29	0.29	0.42

Table 2.6 - Next-day forecasting. Continuous ranked probability scores and maximum deviation from perfect reliability in November 2014

Score	BM	MM	QM	PM	LPE - MO	LPE - SO
CRPS [kW]	8.68	8.54	8.32	10.54	8.04	7.77
ψ [%]	4.05	3.81	3.57	7.14	3.57	10.95

Fig. 2.7 shows the reliability diagrams for single predictors compared to ideal reliability; Fig. 2.8 shows the reliability diagrams for the LPE predictor compared to ideal reliability and to PM.

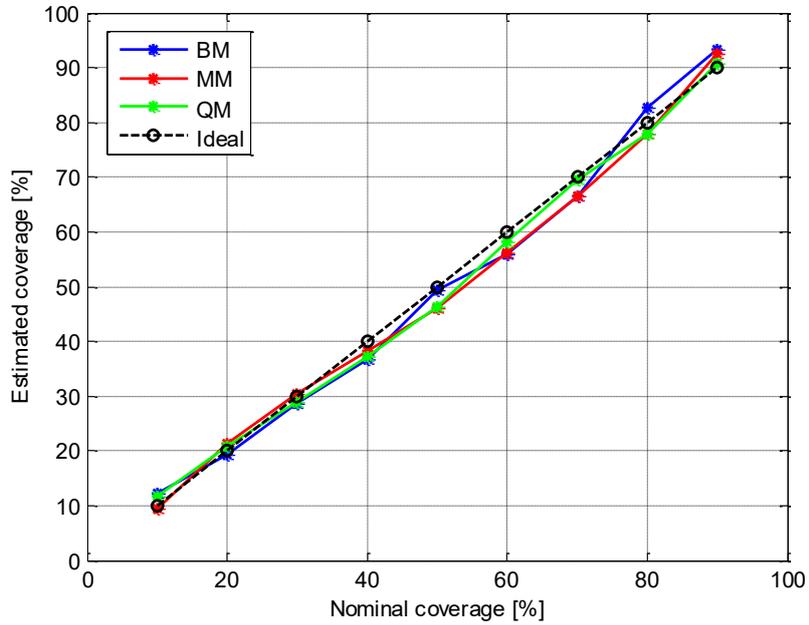


Figure 2.7 - Next-day forecasting. Estimated coverages of single predictors: Bayesian method, Markov chain method, and quantile regression method in November 2014.

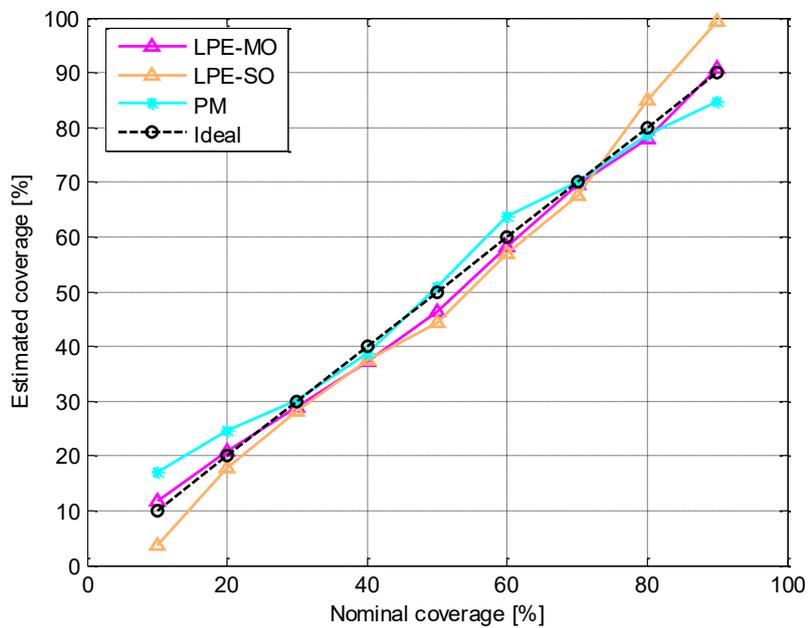


Figure 2.8 - Next-day forecasting. Estimated coverages of the linear ensemble predictor and the probabilistic persistence method in November 2014.

Single predictors appear to be calibrated, as only negligible deviations can be found in Fig. 2.7. Also, the MO reliability appears to be more consistent than the SO reliability, as the latter appear to be heavily over-dispersed.

To confirm this behavior, Fig. 2.9 shows the relative frequencies for BM, MM, and QM, and Fig. 2.10 shows the relative frequencies for PM and LPE with MO and SO procedures. PIT histograms¹³ of single predictors suggest an overall normal dispersion, in particular for MM and QM. The proposed MO approach proved to reduce the over-dispersion that was expected for the LPE in presence of normally-dispersed base predictors with respect to the SO procedure, as shown in Fig. 2.10.

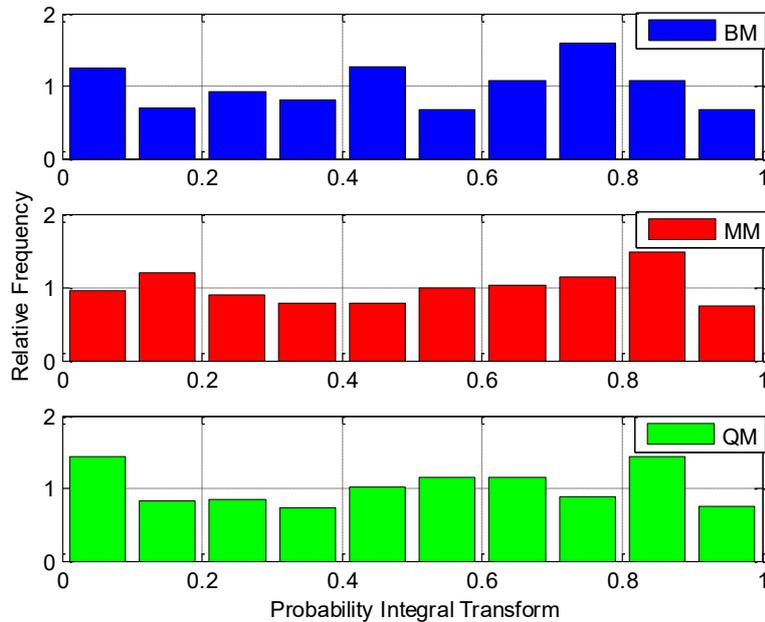


Figure 2.9 - Next-day forecasting, PIT histograms of single predictors: Bayesian method, Markov chain method, and quantile regression method in November 2014.

¹³ Details on PIT histograms are in the Appendix.

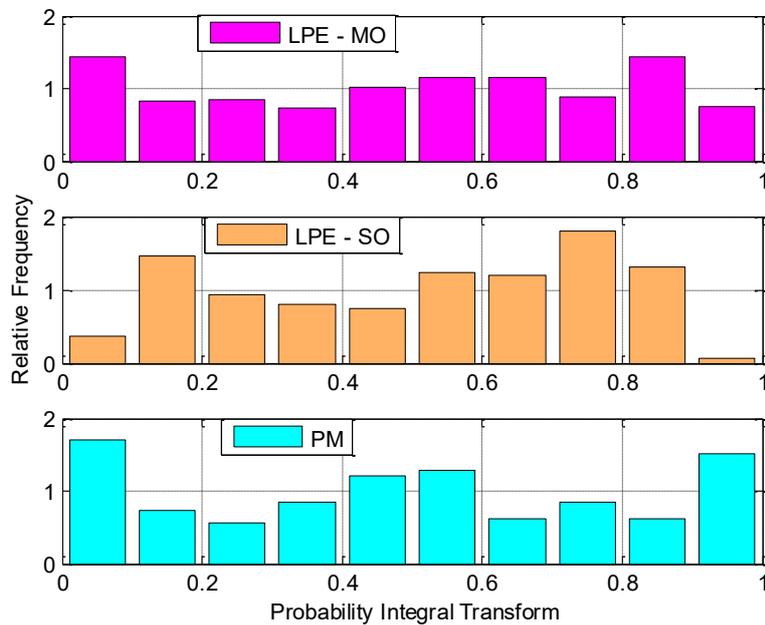


Figure 2.10 - Next-day forecasting. PIT histograms of the linear ensemble predictor with MO and SO procedures, compared to the probabilistic persistence method in November 2014.

Eventually, further results from February 2014 to December 2014 are shown in this in a compact form. The training procedure was the same used for forecasting the single months in the previous Section. Tables 2.7 and 2.8 show the results for next day forecasting in terms of CRPS and maximum deviation from perfect reliability, respectively.

The analysis of the results shows that, on average, BM and MM appear to perform worse than QM in terms of CRPS, but slightly better in terms of reliability. As shown in Table 2.7 and 2.8, in 8 months the LPE-MO produced forecasts with a smaller CRPS than single predictors for next-day forecasting; also, the LPE-MO performed better than the SO procedure in 8 months, allowing to reduce the index ψ up to 8.52% in December.

Table 2.7 - Next-day forecasting. Continuous ranked probability scores from February to December 2014.

Month	CRPS [kW]					
	BM	MM	QM	PM	LPE - MO	LPE - SO
Feb	9.71	9.74	9.63	11.99	9.29	8.99
Mar	10.80	11.67	11.29	15.59	11.02	10.76
Apr	11.80	11.63	11.70	15.5	11.56	11.28
May	11.36	11.71	11.13	13.54	11.09	10.99
Jun	11.02	10.23	10.52	12.88	10.15	9.78
Jul	9.93	9.83	9.84	12.39	9.48	9.20
Aug	11.05	10.84	10.65	13.1	10.41	10.02
Sep	9.88	10.19	9.38	11.2	9.39	9.14
Oct	8.63	8.69	8.50	10.54	8.33	8.05
Nov	8.68	8.54	8.32	10.54	8.04	7.77
Dec	6.85	8.14	7.43	9.12	7.01	6.77

Table 2.8 - Next-day forecasting. Maximum deviation from perfect reliability from February to December 2014.

Month	ψ [%]					
	BM	MM	QM	PM	LPE - MO	LPE - SO
Feb	4.08	5.36	3.57	10.20	3.83	8.42
Mar	3.92	4.84	4.84	14.52	3.69	2.30
Apr	4.05	4.05	4.05	14.29	4.76	3.81
May	3.92	4.38	4.61	8.29	3.00	3.00
Jun	3.57	4.29	4.05	7.86	5.71	5.48
Jul	3.23	3.69	4.61	10.83	4.15	5.53
Aug	3.23	4.15	4.61	10.14	3.92	8.06
Sep	4.29	3.81	3.81	10.95	4.29	6.67
Oct	4.38	4.61	3.69	6.45	6.91	7.83
Nov	4.05	3.81	3.57	7.14	3.57	10.95
Dec	4.15	4.61	3.46	7.14	3.46	11.98

2.5. CONCLUSIONS

Two contributions to the state of the art on probabilistic PV power forecasting were presented in this Chapter.

The first contribution was a new Bayesian method based on an underlying exogenous linear regression model. Two different approaches were compared in the framework of the Bayesian method: the first (a) links the probabilistic PV power forecasts to the solar irradiance, modeled through a three-parameters Beta distribution, and the latter (b) links the probabilistic PV power forecasts to the clearness index, modeled through a modified Gamma distribution. Both the proposed Bayesian-based approaches showed great flexibility in their

use for developing models. They allowed the inclusion of exogenous input (e.g., cloud cover, relative humidity, ambient temperature, and pressure) that may influence both the hourly solar irradiance and the clearness index.

Numerical applications based on real data on a 75-kW PV power plant indicated that both proposed methods, i.e., methods (a) and (b), provided good-quality forecasts in both a spot-value and a probabilistic framework, with method (b) providing slightly better results. Both methods however provided better results than the probabilistic benchmark method, thus proving to be useful in short-term probabilistic forecasting of the production of power by PV systems.

Note that the proposed Bayesian method proved its versatility also in other forecasting field. Indeed, the conceptual forecasting system was also adapted to produce forecasts of wind power [157] and of loads [158,159]; results confirmed the suitability of the Bayesian approach in forecasting variables for power systems .

The second contribution was a new probabilistic method based on a competitive ensemble of different base predictors, aimed to obtain sharp and reliable short-term PV power forecasts. Optimal characteristics of reliability and sharpness were obtained by aggregating the base predictors on the basis of a new MO optimization procedure. Comparisons with a SO procedure, already proposed in relevant literature, and with a probabilistic benchmark method, were also performed in order to assess the usefulness of the MO optimization.

The MO procedure performed significantly better than the SO procedure in terms of the reliability of the output forecasts, with only little losses in terms of CRPS. The proposed method had, on average, better performance in terms of CRPS also compared to the base predictors, but it had on average slightly poorer performance in terms of reliability. The over-dispersion of linear pool ensemble forecasts in presence of normally-dispersed base predictors was effectively reduced through the proposed MO procedure.

Chapter 3.

AN ADVANCED METHOD FOR SHORT-TERM INDUSTRIAL LOAD FORECASTING

3.1. INTRODUCTION

Load¹⁴ forecasts are always needed to perform all of the tasks in which the future status of the systems has to be estimated, i.e., grid planning, optimal management and control, participation to electrical markets, and power balancing [28]. Indeed, the majority of electrical loads cannot be exactly predicted, since two main factors mainly influence the electrical consumption.

The first is a human factor. Intuitively, when some devices are turned on, the total load increases; when some devices are turned off, the total load decreases. Unless monitoring and scheduling the activities of people that utilize these devices, it is almost impossible to exactly predict the utilization of each electrical device, thus it is impossible to exactly predict the total load.

The second factor is linked to weather conditions. Due the spread of air cooling and heating electrical systems, the total load of a building strongly varies as the air temperature varies, with peaks in hot and cold days. As well known, weather conditions are complex phenomena that influence the randomness of loads.

All the above factors have a different impact when dealing with aggregated loads (i.e., at substation level) or industrial loads, and things change significantly also when dealing with single/small groups

¹⁴ Load is an ambiguous term, with different meanings on the basis of the context. For example, load may indicate the device that absorbs energy, the active power, the apparent power, or the energy. In usual load forecasting notation, the term “load” usually means demand (in kW) or energy (kWh). In this Chapter, however, the term “load” is intended as the total consumption of the device, made of both active and reactive powers. Active and reactive powers are clearly distinct, to avoid confusion.

of loads. Moreover, the impact on active powers can be significantly different from the impact on reactive powers, due to the different utilizations of electrical motors, devices, and electronic equipment.

While most papers in the literature have been devoted to forecast active powers at aggregated levels, very few are on industrial load forecasting, only with applications to their active powers. This should not appear strange, since developing a forecasting system for industrial loads is a challenging task; these loads, in fact, usually follow different daily and weekly patterns and could benefit from analyzing the patterns that are uncommon or not as important in classical weather-sensitive load profiles.

To avoid verbose presentation, the Bayesian probabilistic forecasting system developed in [158,159] for aggregated and single loads is overlooked in this Chapter. Note that also further probabilistic methods based on the QR were specifically developed and used in [160] as tools to forecast the allowable current of distribution transformers.

This Chapter therefore contributes to the existing literature on load forecasting by providing deterministic active and reactive power forecasting methods based on MLR and SVR models, that are specifically addressed to industrial applications [161]. In order to properly select the predictors of the models, an exploratory data analysis is first performed to discard uninformative data; then, two model selection techniques are applied to build the models, and their results are compared on actual data. The first technique is based on “classical” k-fold cross-validation of several models, each of them containing combinations of informative candidate predictors selected in the exploratory data analysis. The second technique is instead based on lasso regression analysis.

This Chapter is organized as follows. The state of the art on short-term load forecasting methods is addressed in Section 2.2, with particular focus on applications to industrial load forecasting. The development of the proposed deterministic methods for industrial load forecasting is discussed in Section 2.3; the analytic model formulations and numerical applications based on real data are provided in the corresponding sub-Sections. The Chapter is concluded in Section 2.4.

3.2. SHORT-TERM INDUSTRIAL LOAD FORECASTING

METHODS: STATE OF THE ART

Interest in load forecasting grew since the second half of the past centuries; since then, hundreds of papers have dealt with active power forecasting methods, both in deterministic and probabilistic frameworks. Comprehensive surveys on short-term load forecasting are in [28,36-44,162,163].

The very first contributions found in relevant literature were only addressed to aggregated active power forecasting at national, regional or substation levels, since power systems still were conceived in their traditional, one-directional structure. However, a very large number of works [28,36-44,162,163] dealt with this topic also in the last decade, since it is still of great value in power system balancing and market participation. In several of these papers, hierarchical load forecasting was developed in order to take into account spatial and regional information; however, the contributions that dealt with reactive powers were very limited.

With the spread of SG and μ G concepts, indeed, the interest in developing forecasting methods able to catch the variability of different types of loads at different levels of aggregation, and for both active and reactive powers, definitely grew.

The modeling of the unpredictable behavior of a single load or only few loads usually is a challenging task, complicated by the disaggregation in smaller loads. Several papers [40,42-44] presented forecasting methods addressed for households, commercial and residential loads at appliance or building aggregation levels, or for SG and μ G loads.

Fewer papers instead focused their aim on industrial load forecasting. The problem is particularly complicated and worthy of attention, since, as stated in [164], *“typical forecasting methods tend to be not well adapted when applied at an industrial site level. Seasonality, calendar events, and weather dependency are parameters usually taken into account when modeling a national electricity consumption curve. However, due to the radically different nature of industrial sites, these parameters are inconsistent from site to site and may not be reflected in the consumption curve.”* Indeed, as previously evidenced, different

challenges arise in developing a forecasting system for industrial loads, because they usually follow different daily and weekly patterns.

The relevant literature on industrial load forecasting is quite limited.

A deterministic ANN-based method for the STLF of three substations in Taiwan was developed in [165], taking into account the different shares and typologies of served loads (residential, industrial and commercial). This helped to increase the overall performance of the forecasting system, due to the different needs and patterns of energy consumers.

A tool to analyze and model the load patterns in industrial parks was presented in [166], although without any application of the identified model to load forecasting. The tool exploited self-organizing maps and the clustering k-means algorithm in a cooperative hybrid approach. The tool performance was assessed on actual data taken at a Spanish industrial park.

A similar approach to industrial load pattern analysis was proposed in [167], exploiting self-organizing maps and fuzzy k-means; also in that case, the authors did not apply the results to load forecasting.

Another approach developed for pattern recognition of industrial customers was proposed in [168]; in particular, the first step of the approach allows to estimate industrial customers' by means of fuzzy k-means and hierarchical clustering techniques. The authors only stated that the results of this step could be used in load forecasting, although without any further application.

The STF method based on periodic AR models in [169] actually exploited customer clustering and segmentation, in order to individuate the different types of customers (industrial among them) in 245 aggregate time series taken at a Belgian substation.

Fuzzy-ANN hybrid deterministic approaches were applied in [170] to forecast loads in industrial framework, and they were compared to ARIMA models in terms of RMSEs. The approaches took into account dummy variables such as day of the week and hour of the day.

The authors of [171] exploited smart meter data in order to improve STF of loads for residential and industrial customers. The proposed method exploits a polynomial fitting of the load curve; numerical applications were based on a porcelain industry fed by a medium voltage grid.

In [172] an improved, fast-converging ANN industrial load forecasting model was developed by selecting candidate inputs through mutual information technique, and then by estimating parameters through the enhanced differential evolution algorithm. It is not clear, however, if the datasets considered in the numerical applications actually referred to industrial sites.

The forecasting performances of ANN and ARIMA models were compared in [173] with appliance to a gas processing plant. Results were assessed in terms of MAE and RMSE, and denoted a superiority of ANN.

A day-ahead probabilistic load forecasting model for energy-intensive enterprises, such as big factories, based on Gaussian process was developed in [174]. The model proposed by the authors was a probabilistic tool, and load series were treated as heteroscedastic time series, due to the start-up and shutdown of units that absorbed a notable share of power. The probabilistic model was compared to classical Gaussian process and quantile regression in terms of reliability and sharpness of forecasts.

An industrial load zone was included in the load forecasting track of the Global Energy Forecasting Competition 2012 (GEFCom2012) [11]. The forecast error for that zone was experienced to be quite large when applying classical load forecasting models that work well on other weather-sensitive zones. Several papers presented the forecasting methods that resulted as top entries in the GEFCom2012 and that, therefore, were also applied to industrial load forecasting.

A gradient boosting method with different additive models for each hour was presented in [175]; the authors used univariate regression splines as base learners. Past values and forecasts of ambient temperatures provided by NWPs, together with past values of load, were used as inputs.

A MLR model was developed in [176] to produce load forecasts in the GEFCom2012 framework; the MLR model was furthermore refined by combining models from different weather stations, treating outliers and holidays.

The author of [177] developed a tool based on Gaussian processes and gradient boosting in order to solve the hierarchical load forecasting

problem of GEFCom2012. Only few domain-specific adjustments were added to improve the quality of forecasts.

A multi-scale model that performed a combination of three components was proposed in [178]. The three components took into account a long term trend, the medium term sensitivity to ambient temperature, and short term local behaviors.

Eventually, the number of papers specifically dedicated to industrial load forecasting is much smaller than the ones dedicated to aggregate, residential, or building load forecasting. Also, no application to industrial reactive power forecasting was found in relevant literature.

3.3.A NEW DETERMINISTIC REGRESSION-BASED METHOD FOR SHORT-TERM INDUSTRIAL LOAD FORECASTING

Different challenges arise in developing a short-term forecasting system of the electricity needs of factories. Industrial loads, in fact, may depend on many factors, such as scheduled processes and work shifts, that are uncommon or not so important as in classical load forecasting models. Moreover, the choice of inputs is critical in order to avoid over-fitting and bad-modeling, resulting in inaccurate forecasts.

In this Section there are some insights on developing models suited for forecasting industrial active and reactive powers, at both aggregated and single-load levels. The forecasting is based on MLR or SVR models [161]. The selection of most adequate models is performed with two different techniques, i.e., a “classical” technique based on the k-fold cross-validation of multiple models and a technique based on the lasso analysis.

Numerical applications based on real data collected at an Italian industrial site confirmed the improvements in terms of accuracy of the proposed models for forecasts from 1 to 48 hours ahead, outperforming two benchmark models. Details about the proposed methodology for model development are reported in the following sub-Section 2.3.1; numerical applications based on real data are in sub-Section 2.3.2.

3.3.1. Proposed method

The proposed methodology [161] involves the following steps:

- i. analytic formulation of MLR (*sub-Section 3.3.1.1*) and SVR models (*sub-Section 3.3.1.2*). Analytic expressions are provided only for active powers, to avoid redundancy being trivial the extension to reactive powers;
- ii. individuation of the available time series of data collected until the forecast start time (*sub-Section 3.3.1.3*);
- iii. exploratory data analysis to discard uninformative inputs and to select informative candidate inputs (*sub-Section 3.3.1.4*);
- iv. selection of the most adequate MLR and SVR models for industrial active and reactive power forecasting through two techniques (*sub-Section 3.3.1.5*).

The first technique (T1) of step iv is based on the 10-fold cross-validation of multiple MLR and SVR models that contain combinations of the informative inputs; the best MLR and the best SVR models (in terms of RMSE) are selected for the test step. The second technique (T2) of step iv. is instead based on the lasso analysis [70,179], in order to directly draw the most useful inputs among the informative ones; a 10-fold cross-validation was performed also in this case, in order to provide coherent comparison with the first technique.

3.3.1.1. Multiple Linear Regression model

The basic form of MLR model is here recalled for sake of completeness:

$$P_h = \beta_0^{(MLR)} + \beta_1^{(MLR)} x_{h_1} + \dots + \beta_{m_1}^{(MLR)} x_{h_{m_1}} + \varepsilon_h, \quad (3.1)$$

where P_h is the active power at the time horizon h , $x_{h_1}, \dots, x_{h_{m_1}}$ are the m_1 predictor variables of the model, $\beta_0^{(MLR)}, \beta_1^{(MLR)}, \dots, \beta_{m_1}^{(MLR)}$ are the $m_1 + 1$ parameters to be estimated in the training step of the procedure, and ε_h is a white noise term.

Given $N_{tr}^{(MLR)}$ active power observations $P_{h-k-(N_{tr}^{(MLR)}-1)}, P_{h-k-(N_{tr}^{(MLR)}-2)}, \dots, P_{h-k}$ collected until the forecast start time $h - k$, estimations $\hat{\beta}_0^{(MLR)}, \hat{\beta}_1^{(MLR)}, \dots, \hat{\beta}_{m_1}^{(MLR)}$ of parameters

$\beta_0^{(MLR)}, \beta_1^{(MLR)}, \dots, \beta_{m_1}^{(MLR)}$ are obtained in this approach via the ordinary least square method in the training step:

$$\begin{aligned} & \hat{\beta}_0^{(MLR)}, \hat{\beta}_1^{(MLR)}, \dots, \hat{\beta}_{m_1}^{(MLR)} = \\ & = \underset{\beta_0^{(MLR)}, \beta_1^{(MLR)}, \dots, \beta_{m_1}^{(MLR)}}{\operatorname{argmin}} \sum_{i=1}^{N_{tr}^{(MLR)}} \left[P_{h-k-(i-1)} - \beta_0^{(MLR)} + \right. \\ & \quad \left. - \beta_1^{(MLR)} x_{h-k-(i-1)_1} - \dots - \beta_{m_1}^{(MLR)} x_{h-k-(i-1)_{m_1}} \right]^2, \end{aligned} \quad (3.3)$$

and the maximum-likelihood estimation \hat{P}_h of the active power P_h is then:

$$\hat{P}_h = \hat{\beta}_0^{(MLR)} + \hat{\beta}_1^{(MLR)} x_{h_1} + \dots + \hat{\beta}_{m_1}^{(MLR)} x_{h_{m_1}}; \quad (3.3)$$

it is the forecasted active power at the time horizon h .

3.3.1.2. Support Vector Regression model

As for MLR models, SVR models link the forecasted active power \hat{P}_h at hour h to m_3 predictor variables $\mathbf{x}_h = \{x_{h_1}, x_{h_2}, \dots, x_{h_{m_3}}\}$, although being able to represent features also through a non-linear relationship. Given a set of $N_{tr}^{(SVR)}$ observations $P_{h-k-(N_{tr}^{(SVR)}-1)}, P_{h-k-(N_{tr}^{(SVR)}-2)}, \dots, P_{h-k}$ of active power, the estimation procedure aims at finding parameters of the model in order to assure that all of the available observations are so-well reconstructed through the model that they lie in a $\pm \varepsilon$ band around the estimated values. Linear, Gaussian and polynomial SVR were tested in numerical applications; linear SVR appeared to produce the best results in terms of forecasting error, and it also appeared to us a more “fair” comparative approach with MLR models. Therefore, only results of the linear SVR are shown in this Chapter.

The generic linear SVR model is:

$$\hat{P}_h = \langle \boldsymbol{\beta}^{(SVR)}, \mathbf{x}_h \rangle + \beta_0^{(SVR)}; \quad (3.4)$$

where $\boldsymbol{\beta}^{(SVR)}, \beta_0^{(SVR)}$ are the parameters of the model, and the symbol $\langle \cdot, \cdot \rangle$ stands for the dot product between inputs. Parameters $\boldsymbol{\beta}^{(SVR)}, \beta_0^{(SVR)}$ are estimated by minimizing the norm $\frac{1}{2} \|\boldsymbol{\beta}^{(SVR)}\|^2$ in the training period, under the aforesaid constraint on the distance between observations and reconstructed values:

$$\begin{aligned} \widehat{\boldsymbol{\beta}}^{(SVR)}, \widehat{\beta}_0^{(SVR)} &= \operatorname{argmin}_{\boldsymbol{\beta}^{(SVR)}, \beta_0^{(SVR)}} \frac{1}{2} \|\boldsymbol{\beta}^{(SVR)}\|^2 \\ \text{s.t. } P_{h-k-(i-1)} - \langle \boldsymbol{\beta}^{(SVR)}, \mathbf{x}_{h-k-(i-1)} \rangle - \beta_0^{(SVR)} &\leq \varepsilon \quad \forall i, \\ \langle \boldsymbol{\beta}^{(SVR)}, \mathbf{x}_{h-k-(i-1)} \rangle + \beta_0^{(SVR)} - P_{h-k-(i-1)} &\leq \varepsilon \quad \forall i. \end{aligned} \quad (3.5)$$

Since constraints are quite strict in this form, two auxiliary parameters ξ_i, ξ_i^* are added for each of the $N_{tr}^{(SVR)}$ observations in the training period; this allows to permit reconstructed values to lie outside the $\pm\varepsilon$ band, although penalizing them through a constant C ; the augmented optimization problem is:

$$\begin{aligned} \widehat{\boldsymbol{\beta}}^{(SVR)}, \widehat{\beta}_0^{(SVR)}, \widehat{\xi}, \widehat{\xi}^* &= \\ &= \operatorname{argmin}_{\boldsymbol{\beta}^{(SVR)}, \beta_0^{(SVR)}} \frac{1}{2} \|\boldsymbol{\beta}^{(SVR)}\|^2 + C \sum_{i=1}^{N_{tr}^{(SVR)}} (\xi_i + \xi_i^*), \\ \text{s.t. } P_{h-k-(i-1)} - \langle \boldsymbol{\beta}^{(SVR)}, \mathbf{x}_{h-k-(i-1)} \rangle - \beta_0^{(SVR)} &\leq \varepsilon + \xi_i \quad \forall i, \\ \langle \boldsymbol{\beta}^{(SVR)}, \mathbf{x}_{h-k-(i-1)} \rangle + \beta_0^{(SVR)} - P_{h-k-(i-1)} &\leq \varepsilon + \xi_i^* \quad \forall i, \\ \xi_i &\geq 0 \quad \forall i, \\ \xi_i^* &\geq 0 \quad \forall i. \end{aligned} \quad (3.6)$$

The Lagrange function L_g linked to Eq. (3.6) is:

$$\begin{aligned} L_g &= \frac{1}{2} \|\boldsymbol{\beta}^{(SVR)}\|^2 + C \sum_{i=1}^{N_{tr}^{(SVR)}} (\xi_i + \xi_i^*) + \\ &- \sum_{i=1}^{N_{tr}^{(SVR)}} \lambda_i [\varepsilon + \xi_i - P_{h-k-(i-1)} + \langle \boldsymbol{\beta}^{(SVR)}, \mathbf{x}_{h-k-(i-1)} \rangle + \beta_0^{(SVR)}] + \\ &- \sum_{i=1}^{N_{tr}^{(SVR)}} \lambda_i^* [\varepsilon + \xi_i^* + P_{h-k-(i-1)} - \langle \boldsymbol{\beta}^{(SVR)}, \mathbf{x}_{h-k-(i-1)} \rangle - \beta_0^{(SVR)}] + \\ &- \sum_{i=1}^{N_{tr}^{(SVR)}} (\eta_i \xi_i + \eta_i^* \xi_i^*), \end{aligned} \quad (3.7)$$

where Lagrange variables $\lambda_i, \lambda_i^*, \eta_i, \eta_i^*$ are non-negative values. Setting partial derivatives of L_g to zero and substituting the solutions in Eq. (3.7), the dual problem (3.8) is obtained:

$$\begin{aligned}
\hat{\lambda}, \hat{\lambda}^* = \operatorname{argmin}_{\lambda, \lambda^*} & -\frac{1}{2} \sum_{i=1}^{N_{tr}^{(SVR)}} \sum_{j=1}^{N_{tr}^{(SVR)}} (\lambda_i - \lambda_i^*)(\lambda_j - \lambda_j^*) \\
& \langle \mathbf{x}_{h-k-(i-1)}, \mathbf{x}_{(h-k-(j-1))} \rangle - \varepsilon \sum_{i=1}^{N_{tr}^{(SVR)}} (\lambda_i^* + \lambda_i) + \\
& \sum_{i=1}^{N_{tr}^{(SVR)}} P_{h-k-(i-1)} (\lambda_i - \lambda_i^*), \\
\text{s.t. } & \sum_{i=1}^{N_{tr}^{(SVR)}} (\lambda_i^* - \lambda_i) = 0, \\
& 0 \leq \lambda_i \leq C, i = 1, 2, \dots, N_{tr}^{(SVR)}, \\
& 0 \leq \lambda_i^* \leq C, i = 1, 2, \dots, N_{tr}^{(SVR)}.
\end{aligned} \tag{3.8}$$

The j^{th} solution coefficient can be rewritten as:

$$\hat{\beta}_j^{(SVR)} = \sum_{i=1}^{N_{tr}^{(SVR)}} (\hat{\lambda}_i - \hat{\lambda}_i^*) \mathbf{x}_{h-k-(i-1)_j}, \tag{3.9}$$

and therefore:

$$\hat{P}_h = \sum_{i=1}^{N_{tr}^{(SVR)}} (\hat{\lambda}_i - \hat{\lambda}_i^*) \langle \mathbf{x}_{h-k-(i-1)}, \mathbf{x}_h \rangle + \hat{\beta}_0^{(SVR)}, \tag{3.10}$$

where the remaining parameter $\hat{\beta}_0^{(SVR)}$ can be computed from Karush-Kuhn-Tucker conditions:

$$\begin{aligned}
\hat{\beta}_0^{(SVR)} &= P_{h-k-(i-1)} - \langle \hat{\boldsymbol{\beta}}^{(SVR)}, \mathbf{x}_{h-k-(i-1)} \rangle - \varepsilon \text{ if } 0 \leq \hat{\lambda}_i \leq C, \\
\hat{\beta}_0^{(SVR)} &= P_{h-k-(i-1)} - \langle \hat{\boldsymbol{\beta}}^{(SVR)}, \mathbf{x}_{h-k-(i-1)} \rangle + \varepsilon \text{ if } 0 \leq \hat{\lambda}_i^* \leq C.
\end{aligned} \tag{3.11}$$

3.3.1.3. Data characteristics

The factory of interest is located in southern Italy, manufacturing MV/LV transformers. It is fed by a 20-kV distribution network through a 630-kVA, 20/0.4-kV transformer. Four main feeders supply power to four sections of the factory, i.e., metal carpentry, assembly, wiring and offices, and the testing laboratory. The factory operates from Monday

through Friday with two shifts (from 06:00 A.M. to 02:30 P.M. and from 02:30 P.M. to 11:00 P.M.). On Saturday, there is only one work shift (from 06:00 A.M. to 02:30 P.M.).

Measurements were taken for the total load, loads of each feeder, and single loads (such as individual machines, the office building, electrical pumps, and robots). The measurement system included 19 measurement points recording the data of the average active and reactive powers, currents, voltages, frequency, and power factor in 15-minute intervals. These values were then averaged in groups of four to provide hourly measurements from April 1, 2016 to July 31, 2017. Fig. 3.1 shows typical profiles of the total active and reactive powers during a 7-day period (from May 2, 2016 to May 8, 2016). Statistical parameters of the four load time series (i.e., of the aggregate load, an electrical pump, the carpentry feeder, and a painting machine) considered in the following applications in Section 3.3.2 are shown in Table 3.1.

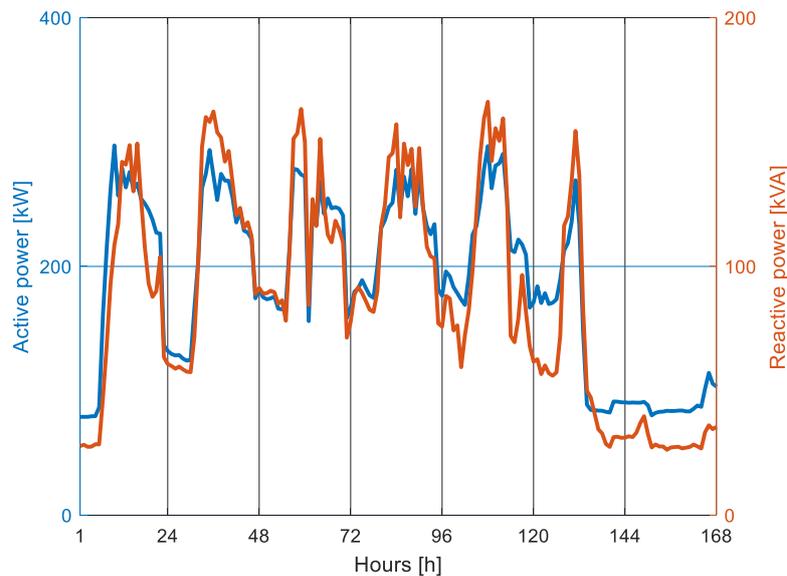


Figure 3.1 – Aggregate active and reactive powers from May 2, 2016 to May 8, 2016.

Table 3.1 – Statistical parameters of the analysed load time series

Load		Mean value	Median value	Standard deviation	Minimum value	Maximum value
Aggregate load	P [kW]	214.42	230.15	77.15	56.41	404.73
	Q [kVA]	109.41	104.74	50.48	14.73	335.75
Electrical pump	P [kW]	13.49	9.17	11.74	0	51.68
	Q [kVA]	8.92	6.75	7.71	0	29.95
Carpentry feeder	P [kW]	38.87	31.02	36.69	0.37	169.72
	Q [kVA]	28.16	17.54	29.19	0	101.50
Painting machine	P [kW]	1.78	0	2.84	0	16.02
	Q [kVA]	2.10	0	3.32	0	10.18

3.3.1.4. Exploratory data analysis

The exploratory data analysis aims at evaluating the information levels of available inputs, in order to discard uninformative inputs and to select instead only candidate inputs among the informative ones. Obviously, the analysis depends on the number of available inputs (weather variables, manufactured units, load measurements, work schedules, and so on), thus varying case-by-case.

The exploratory data analysis results based on the active power of the aggregate load under study are presented in the following; the same analysis was performed for the reactive power of the aggregate load, and for active and reactive powers of single loads to build forecasting systems for each individual load; those results are excluded to avoid verbose presentation.

Firstly, the past values of active power (or functions of the values) that were measured until the forecast origin were considered as candidate quantitative variables to be included in the MLR and SVR models. To provide a “prior” estimation of the effectiveness of this choice, Fig. 3.2 shows the autocorrelation function plot of the aggregate industrial active power. As expected, the pattern is seasonal with daily and weekly periods, and the strongest periodicity is weekly. High values of autocorrelation are set for low values of lags and around the weekly lag; thus, it is beneficial to consider past measurements of power (as in

autoregressive models) and past power differences (as in integrated models) as candidate quantitative predictor variables.

Due to the significant share of energy used to supply the heating and cooling systems, the ambient temperature is usually an additional candidate quantitative predictor variable for modeling loads. Therefore, it is also worthwhile to perform a preliminary investigation to determine whether such a dependence exists also for the industrial active power under study. Fig. 3.3 is a scatter plot of the aggregate industrial active power versus the ambient temperature, which does not show a strong relationship between the two. In fact, the portion of energy used to supply the heating and cooling systems in this factory is quite small compared to the total consumption, as foreseeable due to the working process. Therefore, ambient temperature is treated as an uninformative variable, and therefore it is not included in the forecasting models.

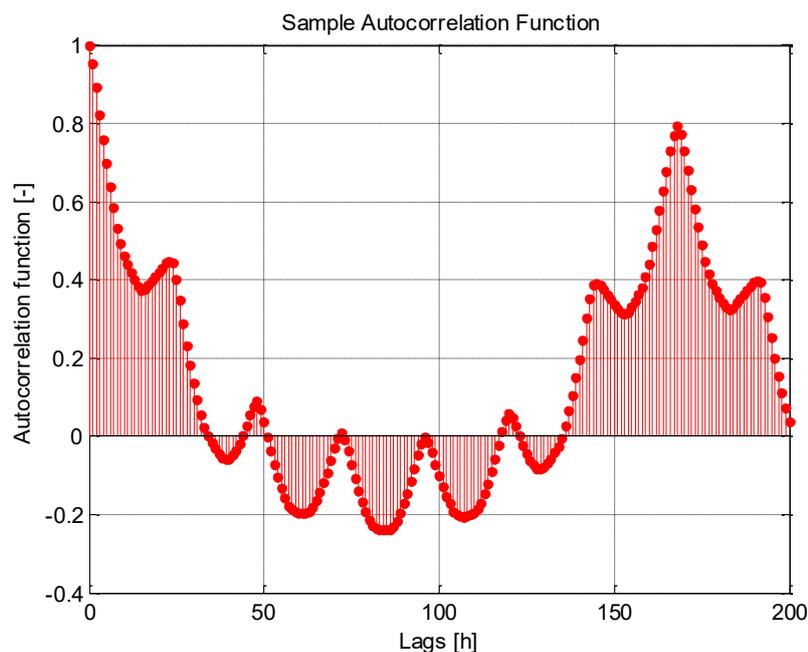


Figure 3.2 – Autocorrelation of the aggregate industrial active power.

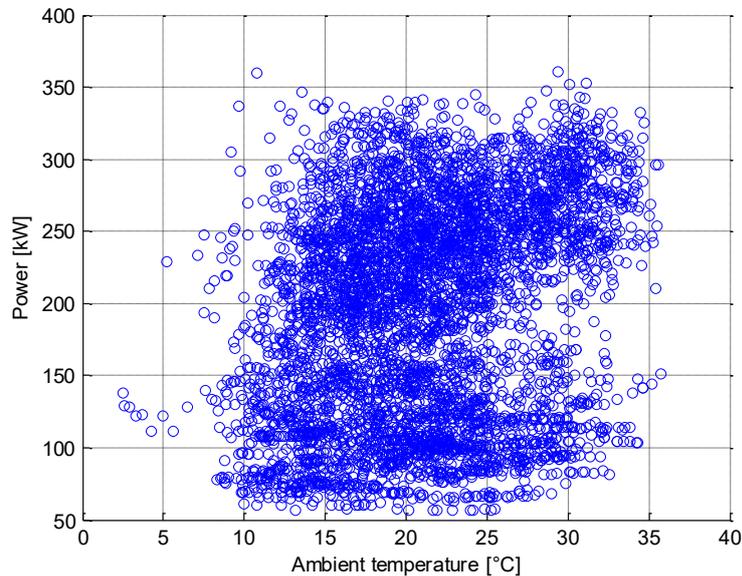


Figure 3.3 – Scatter plot of the aggregate industrial active power versus ambient temperature.

To possibly improve the forecasts in the industrial facility, work shifts and variability during the hours of the day and days of the week were analyzed. The work shifts in the factory are scheduled as 6:00 A.M.-11:00 P.M. from Monday through Friday, and 6:00 A.M.-3:00 P.M. on Saturday. Outside work hours, the demand of machines that must be controlled manually (e.g., welding machines or wood cutters) is automatically set to zero, while the power required by some of the other loads (e.g., offices and furnaces) is significantly reduced. However, there are some exceptions to this rule. For instance, a production phase might have to be scheduled out of the usual work shift due to urgent orders. The schedule of industrial production might not require the use of a particular machine from day x to day y.

Fig. 3.4 shows the active power profiles for different day types: 1) Mondays and days after a holiday; 2) work days; 3) Saturdays and days before a holiday; and 4) Sundays and holidays. The plots in Fig. 3.4 indicate different patterns of the aggregate load versus the hour of the day for each type of day.

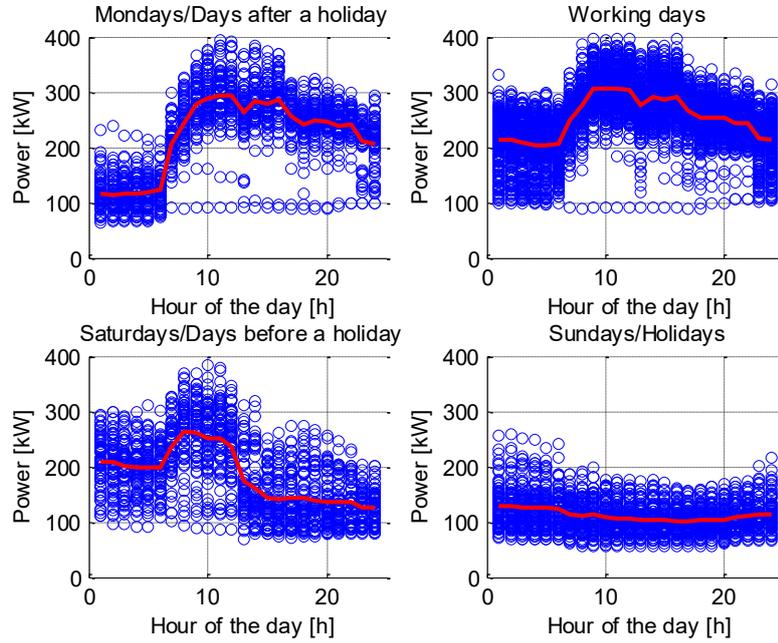


Figure 3.4 – Scatter plot of the aggregate industrial active power for each type of day, versus the hour of the day. The red lines indicate the mean values of observations.

In particular, Mondays and days after a holiday are introduced since the demand in the early hours may be smaller than the demand in the early hours of work days. Saturdays and days before a holiday are introduced since the demand in the late hours may be smaller than the demand in the late hours of work days. Both considerations also suggest an interaction effect¹⁵ between the hour of the day and the type of day.

Therefore, the qualitative variable “hour of the day” $hod_h = [hod_h^{(1)} \quad hod_h^{(2)} \quad \dots \quad hod_h^{(24)}]$ was considered as a first candidate qualitative predictor variable to capture the variability of active power during the day. Variable hod_h is a vector of binary values: the i^{th} element $hod_h^{(i)} = 1$ if h is the i^{th} hour of the day; $hod_h^{(i)} = 0$

¹⁵ One way to add interactions to models is by multiplying the interacting predictor variables by each other. Qualitative variables are sometimes added as stand-alone input variables, but they are more often selected as inputs of interactions. Indeed, if a quantitative variable interacts with a qualitative variable, the quantitative variable can be withdrawn from the model; if two qualitative variables interact with each other, both can be withdrawn from the model [21].

otherwise. Additionally, other candidate qualitative predictor variables selected for the models were the “type of day”

$$tod_h = [tod_h^{(1)} \quad tod_h^{(2)} \quad tod_h^{(3)} \quad tod_h^{(4)}]:$$

- $tod_h = [1 \ 0 \ 0 \ 0]$ if hour h occurs on a Monday or on a day after a holiday;
- $tod_h = [0 \ 1 \ 0 \ 0]$ if hour h occurs on a work day;
- $tod_h = [0 \ 0 \ 1 \ 0]$ if hour h occurs on a Saturday or on a day before a holiday;
- $tod_h = [0 \ 0 \ 0 \ 1]$ if hour h occurs on a Sunday or on a holiday.

The opportunity to introduce a qualitative predictor variable representative of the month of the year was also explored. Nevertheless, the scatter plots of active powers versus the hour of the day for each month (Figs. 3.5, 3.6 and 3.7) do not suggest a monthly dependence given that the patterns are very similar to each other. This is an additional confirmation of the low dependence of aggregate active power on ambient temperature, since the various seasons did not seem to have impacts on total demand.

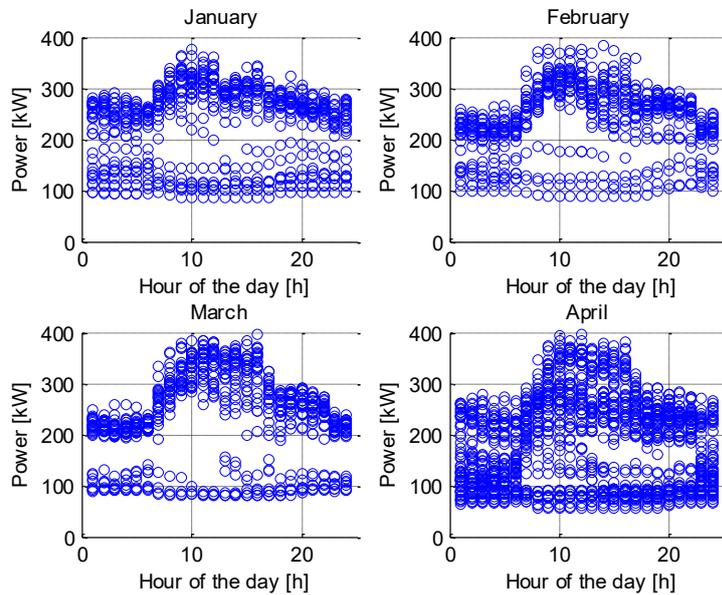


Figure 3.5 – Scatter plots of the aggregate industrial active power in January, February, March, and April, versus the hour of the day.

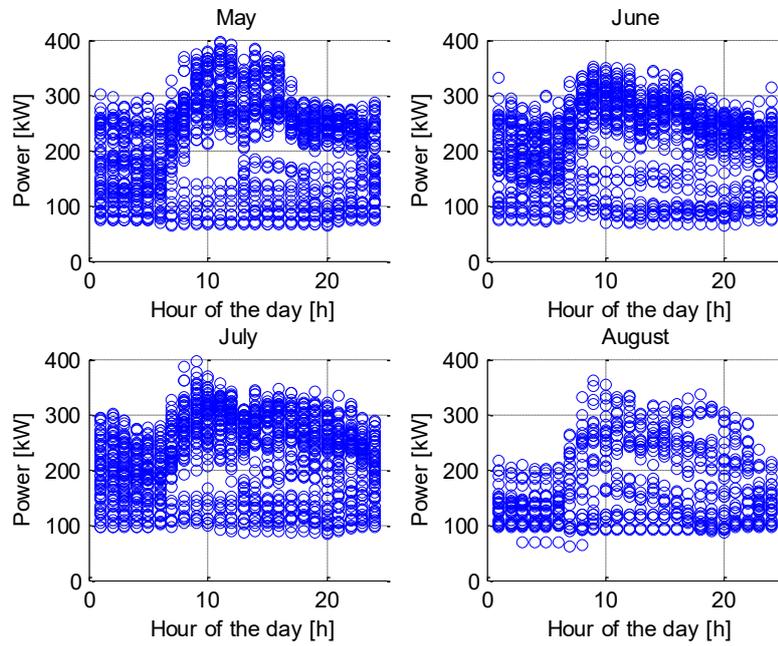


Figure 3.6 – Scatter plots of the aggregate industrial active power in May, June, July, and August, versus the hour of the day.

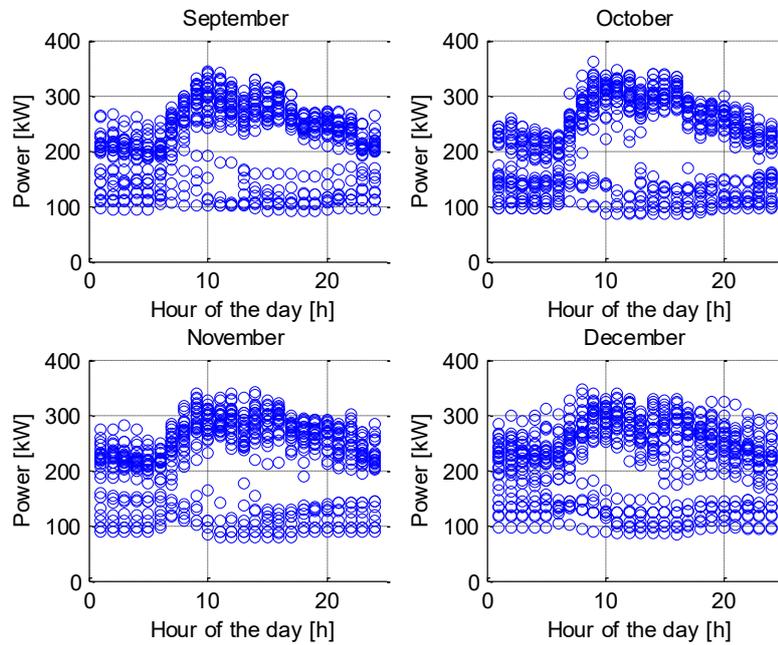


Figure 3.7 - Scatter plots of the aggregate industrial active power in September, October, November, and December, versus the hour of the day.

Moreover, it suggests that the industrial factory had a “standard” production during the considered months (except for August, which had an entire week of holiday). Thus, the month of the year was not considered as a candidate qualitative variable.

Indications of the interaction effects between hour of day, type of day, and past values of measured active power were extracted from the inspection of Figs. 3.8, 3.9, and 3.10, showing aggregate industrial active power versus past measurements of aggregate active power (collected one hour, one day, and one week before, respectively) for each type of day. As expected, when assigning the lag of past values of power, the scatter plots varied significantly with the type of day.

In particular, the scatters of active powers versus previous-hour active powers lie along the diagonal. Moreover, the scatter plots appear to be quite different by day type. For instance, the plots of Mondays and days after a holiday are more concentrated around greater values, while the plots of Sundays and holidays are more concentrated around low values. Additionally, the plots of work days are more spread around the diagonal, while the plots of Sundays and holidays are more concentrated around the diagonal.

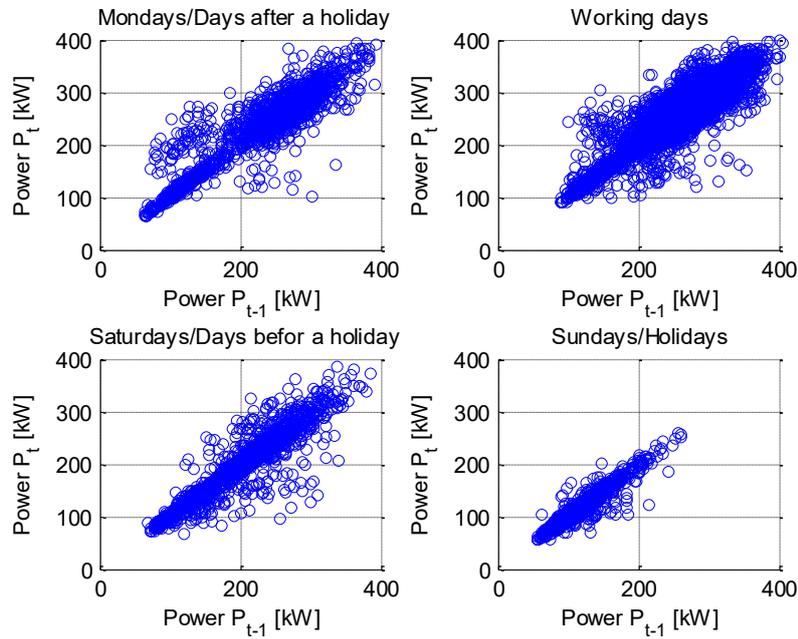


Figure 3.8 – Scatter plots of the aggregate industrial active power for each type of day versus the aggregate industrial active power measured one hour before.

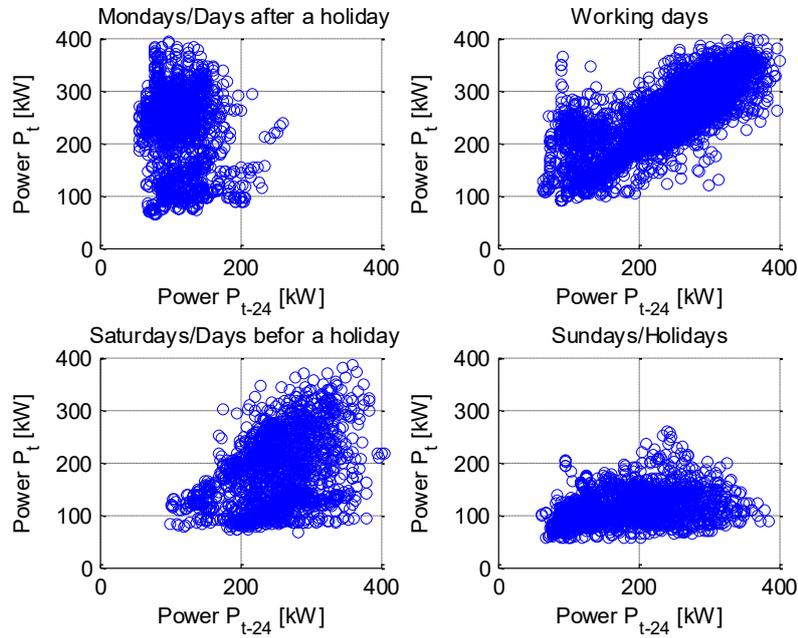


Figure 3.9 – Scatter plots of the aggregate industrial active power for each type of day versus the aggregate industrial active power measured one day before.

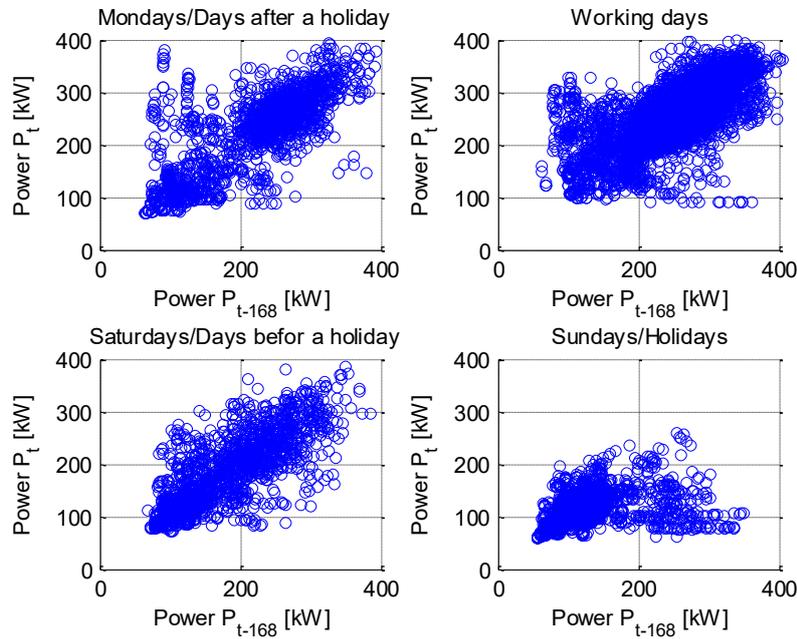


Figure 3.10 – Scatter plots of the aggregate industrial active power for each type of day versus the aggregate industrial active power measured one week before.

The spread of scatters around the diagonal increases as the lag increases. The scatter plots for power versus day-before and week-before power are no longer identifiable as lying around the diagonal. Additionally, the scatter plots for the same type of day differed when considering day-before or week-before measurements. For example, the scatter plot of Mondays and days after a holiday when considering weekly lag differed from the analogous scatter plot considering daily lag. These observations suggested that the interaction effects should be carefully investigated.

Eventually, we also tried to build MLR and SVR forecasts in a “hierarchical” fashion, exploiting past active power measurements taken by sensors in proximity to the load under study. Also in this case, we observed no correlations between such variables, and therefore we did not include them as candidate predictor variables.

3.3.1.5. Model selection techniques

In the model selection technique T1, several MLR models were trained with several possible combinations of informative variables drawn

from the exploratory data analysis; a 10-fold cross validation procedure allowed to select the best among these MLR models in terms of RMSE, avoiding over-fitting issues.

The 10-fold cross validation consists in dividing the available training dataset in 10 equally-dimensioned subsets. At turn, one subset is left out and all of the models under evaluation are trained through ordinary least squares, using the data contained in the remaining 9 subsets; the performance of the model is validated using the data contained in the subset that was left out. The procedure is repeated 9 times, obviously varying the subset left out, in order to pick all the 10 available subsets. Average performances of the models under evaluation are obtained, e.g., by averaging error values in the 10 validation steps; the model with the highest average performance is selected for the test step, using different data since the test dataset is blind to each of the 10 training subsets.

In the model selection technique T2, instead, parameters of a MLR model with all of the informative variables drawn from the exploratory data analysis are estimated via the lasso analysis [70,179].

Lasso analysis is based on the lasso regression; given the set of $N_{tr}^{(LAS)}$ observations $P_{h-k-(N_{tr}^{(LAS)}-1)}, P_{h-k-(N_{tr}^{(LAS)}-2)}, \dots, P_{h-k}$ of active power, the aim of lasso analysis is to find estimations $\hat{\beta}_0^{(LAS)}, \hat{\beta}_1^{(LAS)}, \dots, \hat{\beta}_{m_6}^{(LAS)}$ by solving the following minimization problem:

$$\begin{aligned} & \hat{\beta}_0^{(LAS)}, \hat{\beta}_1^{(LAS)}, \dots, \hat{\beta}_{m_6}^{(LAS)} = \\ & = \operatorname{argmin}_{\beta_0^{(LAS)}, \beta_1^{(LAS)}, \dots, \beta_{m_6}^{(LAS)}} \left\{ \frac{1}{2N_{tr}^{(LAS)}} \sum_{i=1}^{N_{tr}^{(LAS)}} \left[P_{h-k-(i-1)} + \right. \right. \\ & \left. \left. - \beta_0^{(LAS)} - \beta_1^{(LAS)} x_{h-k-(i-1)_1} - \dots - \beta_{m_6}^{(LAS)} x_{h-k-(i-1)_{m_6}} \right]^2 + \right. \\ & \left. + \lambda^{(LAS)} \sum_{i=1}^{m_6} |\beta_i^{(LAS)}| \right\}, \end{aligned} \quad (3.12)$$

where $\lambda^{(LAS)}$ is an assigned non-negative parameter that is linked to the degree of regularization. The peculiarity of lasso analysis is that the addition of the term $\lambda^{(LAS)} \sum_{i=1}^{m_6} |\beta_i^{(LAS)}|$ is analogous to adding an upper-bound constraint on the sum of absolute values of the estimated

parameters; therefore, several parameters are estimated as null, or quasi null values. A threshold can be set to discern between zero and non-zero values; the number of estimated parameters $\hat{\beta}_0^{(LAS)}, \hat{\beta}_1^{(LAS)}, \dots, \hat{\beta}_{m_6}^{(LAS)}$ that are non-zero increases as $\lambda^{(LAS)}$ increases. $\lambda^{(LAS)}$ can be empirically determined, on the basis of the dimension of the problem (e.g., it can be varied with the number of parameters m). Inputs associated to parameters whose values are zero were discarded, and the resulting inputs were selected to build the MLR model (with parameters estimated via the ordinary least squares) for the test step. For SVR, the model selection was performed using the two same conceptual techniques T1 and T2 described for the MLR model selection, although parameters were estimated in the SVR framework.

3.3.2. Numerical applications

The techniques for MLR and SVR model selection were applied to the eight time series described in sub-Section 3.3.1.1, for both active and reactive powers. The cross-validation period was from May 1, 2016 to June 30, 2017; the RMSEs of each of the 10 cross-validation folds were averaged, and the model with the lowest average RMSE was selected for the test period (from July 1, 2017 to July 31, 2017). For all of the test cases, the pool of candidate predictors from which the variables were selected to build models is shown in Table 3.2.

Table 3.2 – Predictors candidate to build models for industrial load forecasts

Variables
$P_{h-1}, P_{h-1}^2, P_{h-1}^3, P_{h-2}, P_{h-2}^2, P_{h-2}^3, P_{h-3}, P_{h-3}^2, P_{h-3}^3,$ $P_{h-24}, P_{h-168}, P_{h-1} - P_{h-2}, P_{h-2} - P_{h-3},$ hod_h, tod_h

Forecasts were performed for lead times from 1 hour to 48 hours; when dealing with multi-step time horizons, the last forecasted values were used as inputs of sub-sequential models. A comparison with two benchmarks was performed to validate the usefulness of the procedure at different lead times in terms of both NRMSE and NMAE; normalization is to the maximum observed values in Table 3.1.

A SN model is selected as first benchmark. The SN model forecasts the unknown load taking the last observed value at the same time of

the last same season. Industrial loads under study showed two seasonal behaviors, with daily and weekly periodicity. However, two motivations led to select the load measured one week before. First, forecasts were performed with the proposed methods for lead times that exceed one day (i.e., up to 48 hours), and therefore the SN model could not be applied for comparison for lead times ranging from 25 to 48 hours. Also, selecting the load from one week before leads to results that are better on average than selecting the pervious-day measurements. Therefore, given the hourly resolution in the load data, the active power forecast $\hat{P}_h^{(SN)}$ at hour h is:

$$\hat{P}_h^{(SN)} = P_{h-168}, \quad (3.12)$$

being trivial the extension to SN reactive power forecasts.

A PM is selected as second benchmark. The PM assumes the unknown load to be equal to the last observed value, whatever the forecast lead time k is. This is a common benchmark in load forecasting, that works particularly well for small lead times (e.g., 1 hour ahead). The active power forecast $\hat{P}_h^{(PM)}$ at hour h is:

$$\hat{P}_h^{(PM)} = P_{h-k}, \quad (3.13)$$

being trivial the extension to PM reactive power forecasts.

The forecasting results of the aggregate active power are analyzed in details in the following sub-Section, while forecasting results of single loads' active powers are presented in a more compact form in sub-Section 3.3.2.2, only for sake of conciseness. Also, several hints on results of the reactive power forecasting are shown in sub-Section 3.3.2.3.

3.3.2.1. *Assessment of the quality of active power forecasts of the aggregated industrial load*

Figs 3.11 and 3.12 respectively show the NMAEs and the NRMSEs versus lead times of active power forecasts obtained with MLR and SVR models, applying model selection techniques T1 and T2. Errors coming from PM and SN model are also shown for benchmarking in

Figs 3.11a and 3.12a, while the details on the MLR and SVR forecast errors are captured in the zoomed Figs. 3.11b and 3.12b.

As expected, all of the MLR and SVR methods provided better results for short lead times, and tended to degrade as the lead time increase, although the errors did not rapidly diverge as the lead time increased. T2 MLR was the overall best model for all of the lead times. T2 SVR errors were, on average, slightly smaller than the T1 SVR.

Benchmarks were always outperformed by all of the proposed methods; PM was a better comparing tool from 1 to 3 hours ahead, while for greater lead times the quality of SN forecasts was considerably higher than PM's forecasts.

The NMAE maximum and minimum improvements of the active power forecasts obtained through the T2 MLR were around 55% and 16% when compared to SN; the minimum improvement was instead 18% when compared to PM. These numbers were very similar when obtained with reference to the NRMSE: indeed, the maximum and minimum improvements of the T2 MLR were around 50% and 20% when compared to SN, while the minimum improvement was 18% when compared to PM. Obviously, NRMSEs are greater than NMAEs since they "weight" more the greatest errors; however, the average difference between a NRMSE and its corresponding NMAE is only around 2.5%, thus suggesting that, on average, the magnitude of absolute errors do not significantly differ from the mean of absolute errors.

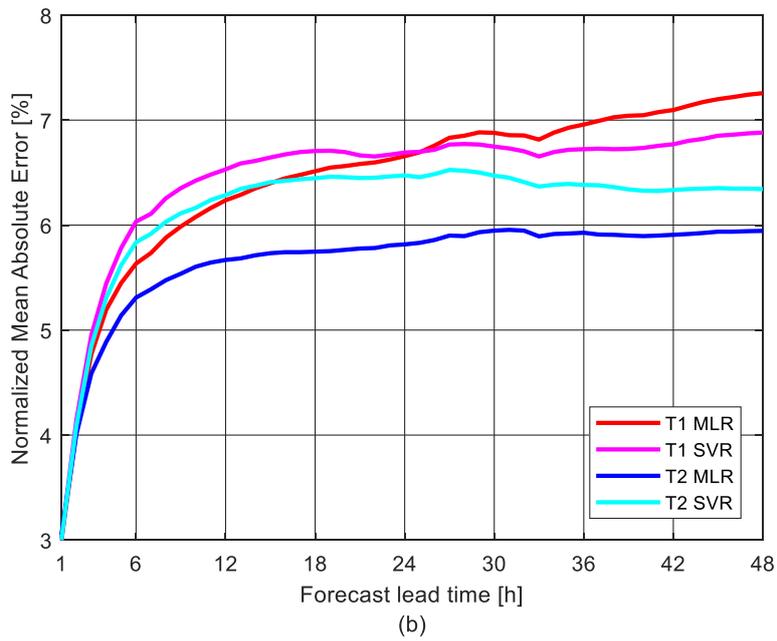
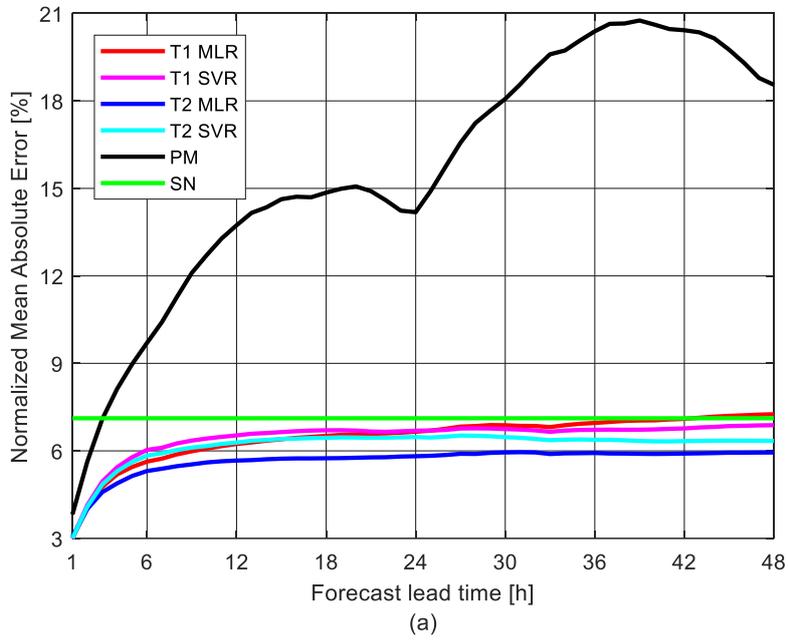


Figure 3.11 - Normalized Mean Absolute Errors for the aggregate active power: (a) comparison with benchmarks; (b) zoom on the models' errors.

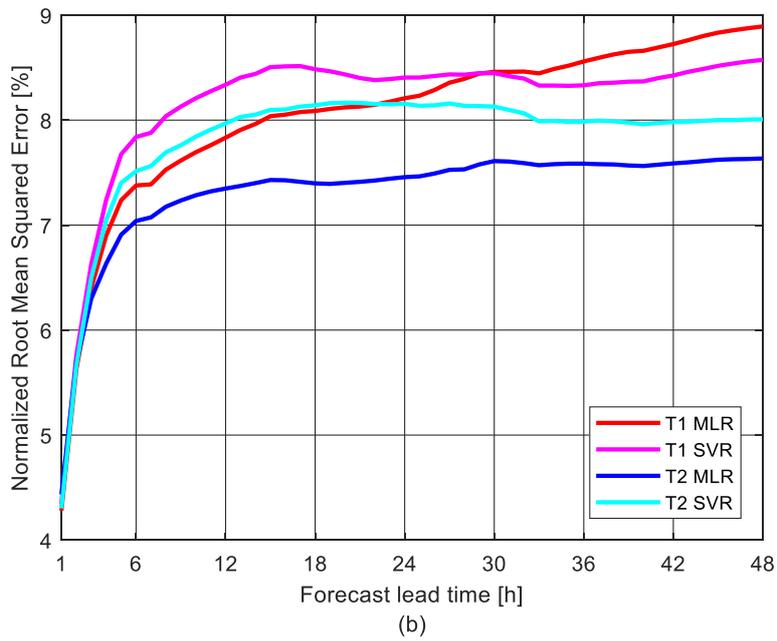
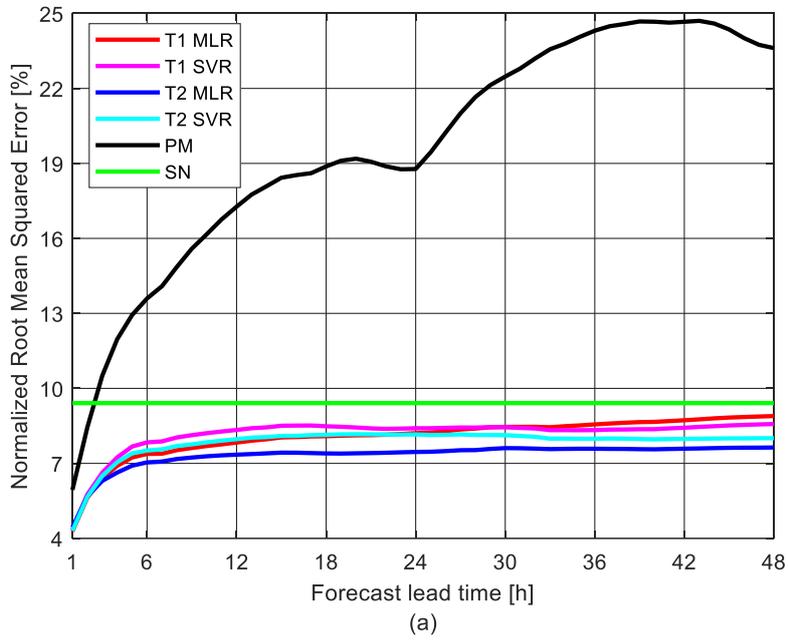


Figure 3.12 - Normalized Root Mean Squared Errors for the aggregate active power: (a) comparison with benchmarks; (b) zoom on the models' errors.

3.3.2.2. *Assessment of the quality of active power forecasts of single industrial loads*

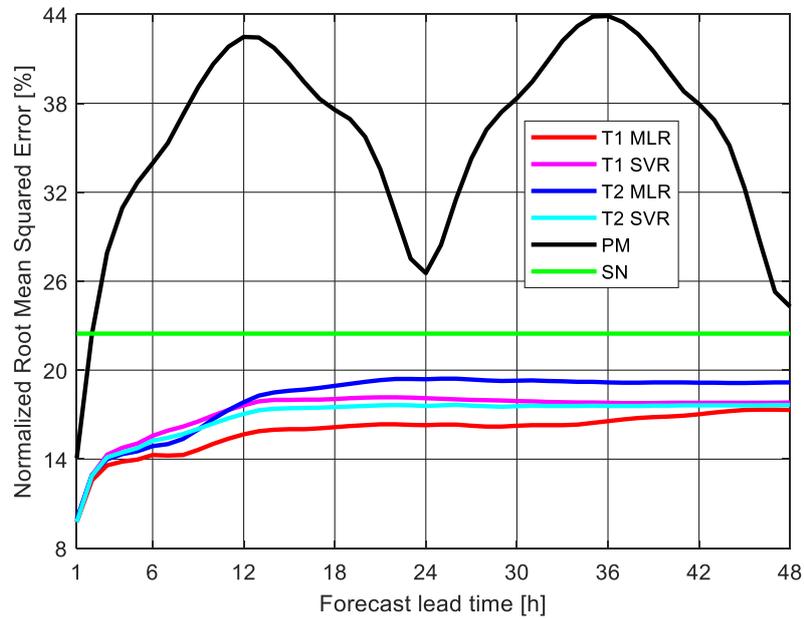
The quality of single loads' active power forecasts is assessed in this sub-Section only by inspecting the NRMSEs, to avoid verbose presentation.

Figs. 3.13 shows the NRMSEs versus lead times of the active power forecasts of the electrical pump. T1 MLR provided the best active power forecasts for all of the lead times. Using this method the improvements with respect to benchmark models were more intense than in the case of aggregate active load; they always exceeded 31% when compared to PM, and they exceeded 23% and 21% when compared to SN.

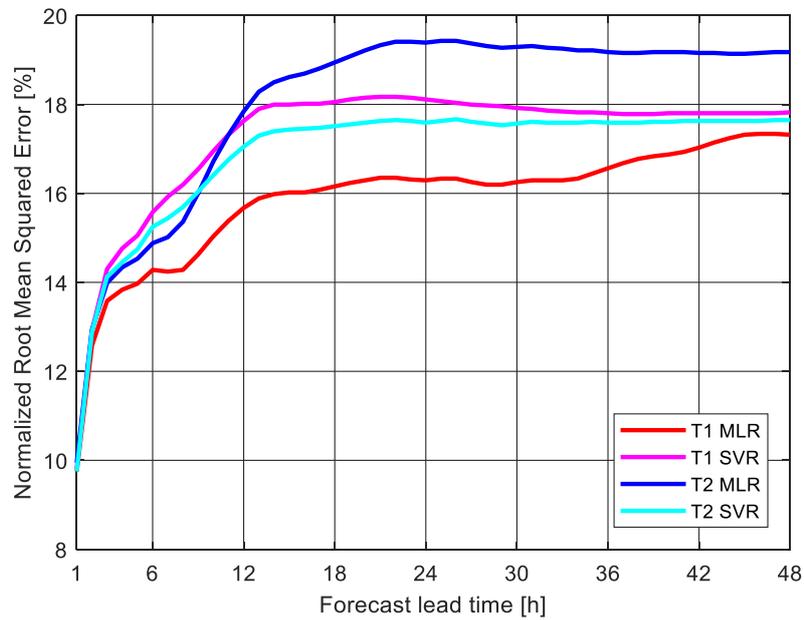
Fig. 3.14 show the NRMSEs versus lead times of the active power forecasts of the carpentry feeder. Although models outperformed PM benchmark, the convenience of MLR and SVR models was limited to lead times up to 27 and 5 hours, respectively; the quality of longer-term forecasts was indeed inferior to the SN one.

Fig. 3.15 shows the NRMSEs versus lead times of the active power forecasts of the painting machine. In this case, the error trends significantly differed from a model to another, while in the previous examples the errors were quite similar for the proposed models. The worst performance was achieved through the T2 SVR, while the T1 MLR provided the best active power forecasts for all of the lead times. The MLR models performed, on average, better than SVR models, and model selection technique T1 led to better improvements with respect to T2.

Although the overall trends of the proposed models were on average quite similar, some straightforward general considerations could be drawn. SVR appeared to perform slightly worse than MLR for active power forecasting both at aggregate and single-load levels; model selection technique T1 was preferable for single loads, while technique T2 led to better aggregate active power forecasts.

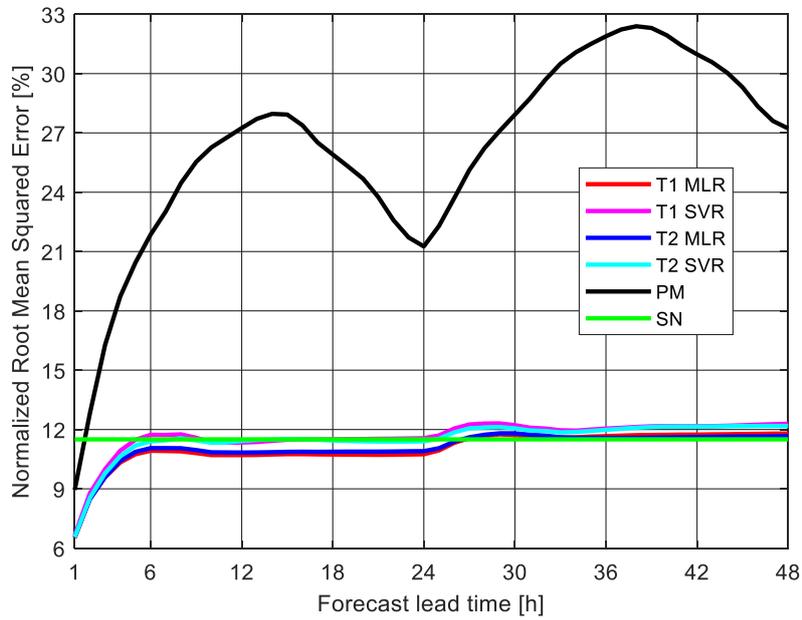


(a)

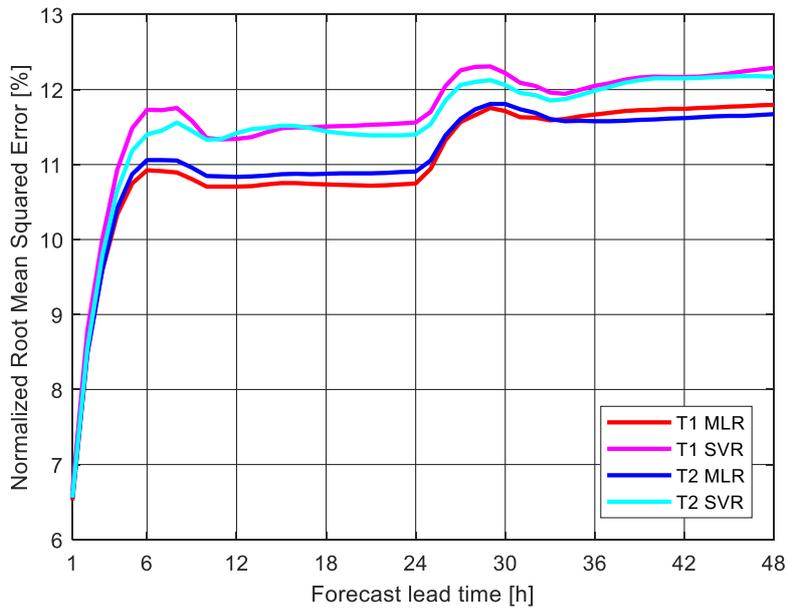


(b)

Figure 3.13 - Normalized Root Mean Squared Errors for the active power of the electrical pump: (a) comparison with benchmarks; (b) zoom on the models' errors.



(a)



(b)

Figure 3.14 - Normalized Root Mean Squared Errors for the active power of the carpentry feeder: (a) comparison with benchmarks; (b) zoom on the models' errors.

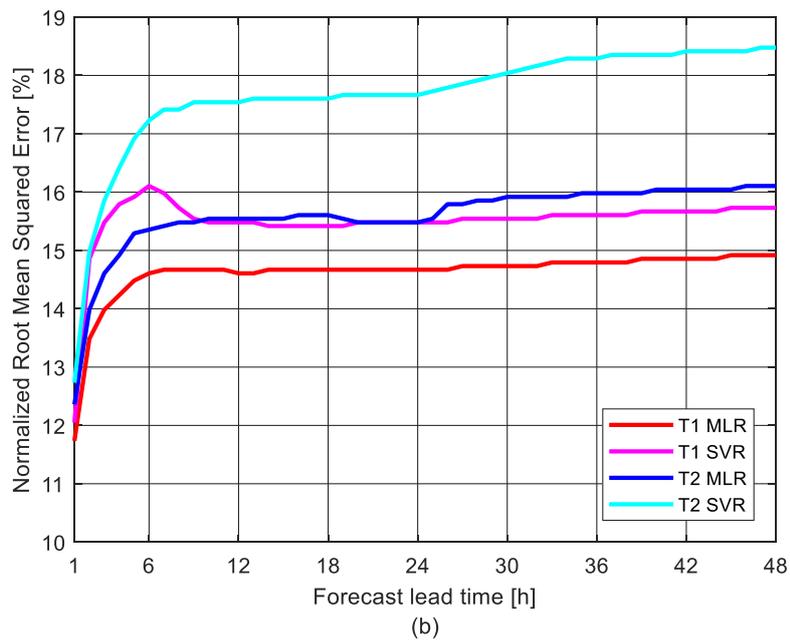
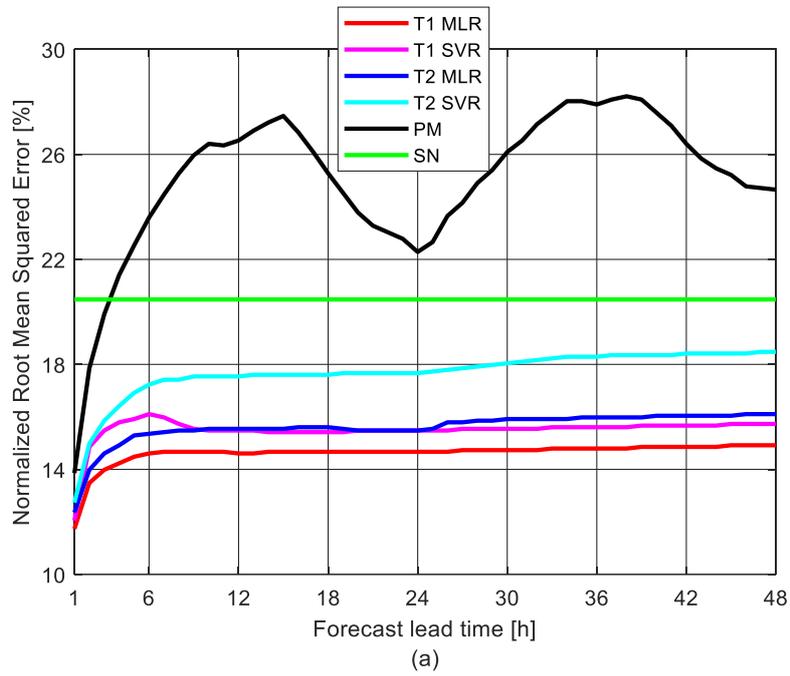


Figure 3.15 - Normalized Root Mean Squared Errors for the active power of the painting machine: (a) comparison with benchmarks; (b) zoom on the models' errors.

3.3.2.3. Assessment of the quality of reactive power forecasts of industrial loads

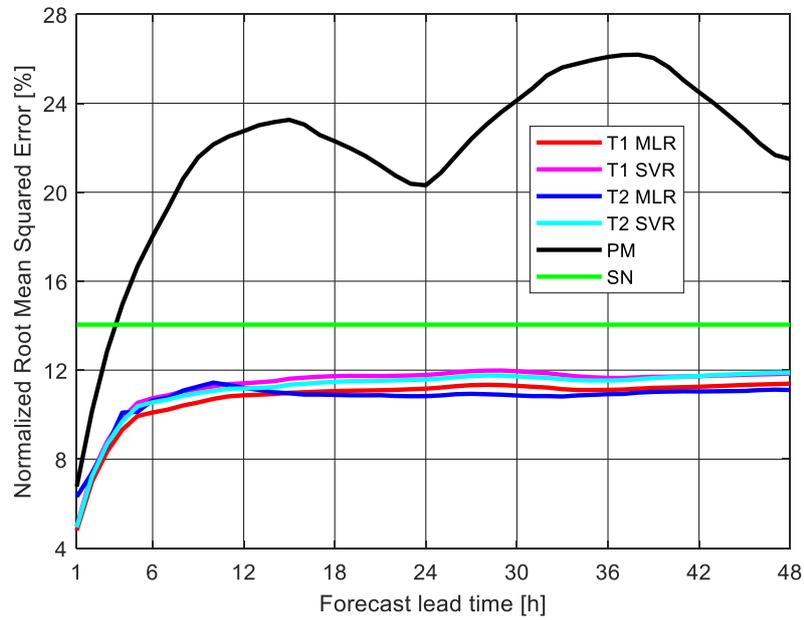
The same procedures applied to build models for forecasting active powers were applied to reactive powers of the four considered loads.

The main outcome is that the methods that provided the best performances for active powers not necessarily provided the best performances also for reactive powers.

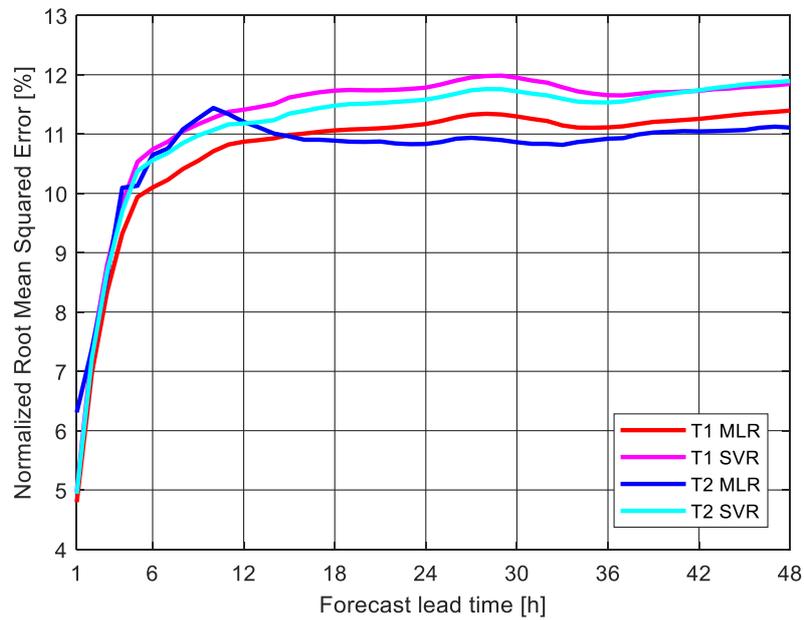
As a significant example, Fig. 3.16 shows the NRMSEs versus lead times of the aggregate reactive power forecasts. The comparison of these errors to those shown in Fig. 3.12 for the active power is quite interesting. T1 MLR was the best method for the aggregate active power; however, it outperformed the other methods for reactive power forecasting only up to 14-hour lead times, but T2 MLR performed slightly better from 15-hour to 48-hour lead times.

As a general consideration, the performances of both benchmarks and of the proposed methods worsened when applied to reactive power forecasting with respect to active power forecasting. For example, the best model provided a NRMSE of about 4.6% for active power forecasts, while the best model provided a NRMSE of about 5% for reactive power forecasts at the same lead time (1 hour). This trend was confirmed also for different lead times, and for all of the single loads.

The different scenarios confirm the utility of research efforts in active and reactive power forecasting, due the increasing needs of industrial systems to be contextualized in smart power systems.



(a)



(b)

Figure 3.16 - Normalized Root Mean Squared Errors for the aggregate reactive power: (a) comparison with benchmarks; (b) zoom on the models' errors.

3.4.CONCLUSIONS

The contributions presented in this Chapter focused on the development of forecasting systems for industrial loads. Forecasting systems need to be accurately built case-by-case, even for loads within the same factory, due the different nature and schedule of machines and devices. Selecting only informative inputs is mandatory in order to build high-performance models, with low computational effort. Therefore, two model selection techniques were compared in terms of produced forecast errors. The first technique was based on the 10-fold cross-validation of several MLR and SVR models that contain combinations of the informative inputs; the best models in terms of RMSEs were selected for the test step. The second technique was instead based on the lasso analysis through a 10-fold cross-validation, in order to directly draw the most useful inputs among the informative ones.

Both techniques were compared using actual data from an Italian factory; results were presented for both the aggregate load and for several single loads, in order to assess the performance in different frameworks having two benchmark models as references.

Although the overall trend of all of the proposed models was similar as the lead times varies, results differed as the typologies of load varied, thus justifying the need for ad-hoc forecasting systems. Results of MLR models were more accurate on average than results of SVR models; technique T1 was the best pick for aggregate load, while T2 was the best pick for single loads.

As a final note, the Bayesian probabilistic forecasting system developed in [158,159] was overlooked in this Chapter, only to avoid verbose presentation. Further probabilistic methods based on the QR were specifically developed and used in [160] as tools to forecast the allowable current of distribution transformers.

Chapter 4.

ADVANCED PROBABILITY DISTRIBUTIONS FOR MODELING EXTREME VALUES OF WIND SPEED

4.1. INTRODUCTION

EWS are potentially destructive events that affect also power systems in three ways.

The first aspect is the power production. The energy output of WGs mainly depends on the magnitude of wind speeds. Wind uncertainty must be taken into account by wind producers when preparing selling offers to minimize penalties, and also by the authority companies that presents the corresponding aggregate selling offers (e.g., this role is played by the “Gestore dei Servizi Energetici” in Italy [180]). In this context, EWS are usually unfavorable, as wind speeds that exceed the cut-off value of generators correspond to no power output. Also, values of wind speed fluctuating in proximity of the cut-off values are translated into rapid oscillations of power from zero to values close to the rated power, negatively influencing the real-time balancing and power quality [181,182].

The second aspect is the overhead line rating. As well known, the total electrical demand is constantly increasing, but infrastructures planning does not follow this trend. Moreover, the economical and environment impact of transmission lines is not negligible; the economical investment for a km of line is expected to be hundreds of thousands of euros, and it is often opposed by public opinion [7]. Thus, a better exploitation of existing transmission lines is mandatory in order to satisfy the increasing demand, without severe congestion in normal and emergency conditions. Dynamic line rating allows power flows to be constrained in a less-conservative way. In this context, wind speed

plays a key role, since it can modify the maximum allowable loading thanks to its strong influence on the convective thermal exchange [7]. In particular, EWS can significantly increase the maximum allowable loading of the overhead lines; moreover, wind farms in proximity of overhead lines can be, in turn, subject to an increased amount of power production. Therefore, a great attention to the combined effects must be paid.

The third aspect is linked to the mechanical reliability of system components. EWS jeopardize sensible structures, such as wind towers and generator blades, and also overhead transmission lines [9,10]. The effects on towers and blades can be destructive, leading also to large intervals of time of inactivity and expensive maintenance or repair; obviously, offshore wind plants are particularly sensitive to this problem [8]. Transmission lines can rarely suffer mechanical damages, but can hazardously oscillate, reducing the minimum distance between active parts and inert parts or between two different active parts. The failure of a single component can affect the functioning of the whole system, and could lead to cascade trips.

The problems and requirements indicated above have recently increased the attention of researchers and system operators towards EWS forecasting [183], in order to timely operate to restore acceptable levels of reliability and to assure real time balancing.

As stated in Chapter 1, some parametric probabilistic tools for forecasting EWS require prior assumptions on the statistical distribution of the random variable; therefore, an accurate characterization of EWS datasets is mandatory in these kind of applications [184-188]. Statistical distributions for EWS should be versatile but also easy to be handled by forecasters; therefore, when classical parameter estimation procedures (such as Maximum Likelihood Estimation (MLE) and Moment Estimation (ME)) risk to produce unreliable results, each new distribution should come with an appropriate parameter estimation procedure.

Bearing this in mind, this Chapter adds two contributions to the existing literature on probability distributions for EWS.

The first contribution is the proposal of an Inverse Burr (IB) distribution for the statistical modeling of EWS [189]. Several classical procedures for the IB parameters estimation are compared to a new

contribution, the Quantile Estimation (QE) that is sometimes easier to be implemented, implying in particular cases a simple algebraic computation.

The second contribution is the proposal of a finite mixture of Inverse Burr - Inverse Weibull distributions (M-IB-IW) that could suit different scenarios [190]. Since mixture distributions usually introduce severe difficulties in their parameter estimation, an ad-hoc EM procedure is proposed as it appears to be a particularly suitable solution for estimating parameters of the M-IB-IW distribution.

This Chapter is organized as follows. The state of the art on models for the statistical characterization of EWS is addressed in Section 4.2. The first proposal, i.e., the IB distribution, is presented in Section 4.3, while the second proposal, i.e., the M-IB-IW distribution, is presented in Section 4.4. EWS models that are already available in literature are presented in Section 4.5 as benchmarks. Numerical applications based on real data are provided in Section 4.6. The Chapter is then concluded in Section 4.7.

4.2. MODELS FOR EXTREME VALUES OF WIND SPEED: STATE OF THE ART

Researches on EWS statistical modeling started when practitioners understood that “traditional” wind speed statistical modeling was unable to exactly capture the behavior of the phenomenon in its extreme manifestations. Many scientific papers have dealt with the statistical characterization of wind speed in the last decades; comprehensive surveys and comparative studies on the topic are in [191-194]. The fundamental outcome of all of these researches is that, even if some PDFs such as Weibull distribution, Rayleigh and Lognormal distributions are widely used to characterize wind speeds, no model is universally recognized as the best fit for all of the wind speed datasets, since the GOF varies case-by-case. Moreover, when applied to EWS, “traditional” wind speed distributions proved to fail GOF checks, as they can severely underestimate or overestimate upper quantiles [195-198]. Therefore, the need for models that are specifically aimed to EWS characterization increased, in order to avoid

such a kind of problems in a world that is more and more conditioned by extreme events.

Reviews and comparison of papers dedicated to EWS analysis and modeling are in [192,199-201]. Basically, research papers explored two approaches for determining the EWS data under test: Period Maxima (PMA) and Peak Over Threshold (POT).

PMA divides the entire available dataset of wind speed in several equally-spaced subsets (e.g., annually, monthly, or weekly spaced); it then extracts one maximum value of wind speed for each subset.

POT instead picks all of the values of wind speed that exceed an assigned threshold (e.g., 40 m/s), although avoiding to pick observations too close one another in time, to assure the statistical independency between observations in the EWS dataset.

The size of a PMA EWS dataset is usually much smaller than the size of a POT EWS dataset, if they are drawn from the same wind speed dataset. In particular, Annual Maxima (AMA) are usually made of very few samples, since only one wind speed observation per year is selected.

Two well-established distribution families, such as the Generalized Extreme Value¹⁶ (GEV), the Inverse Weibull (IW), and the Generalized Pareto (GP), are usually applied for PMA and POT applications in relevant literature. In the majority of the papers, parameter estimation procedures are discussed and some novelties are proposed to overcome typical issues of bad estimation.

Annual extreme wind gusts measured at 143 weather stations in USA were analyzed in [202] and modeled through GU and IW distributions; in 139 out of 143 cases, the IW distribution was the most suitable model for EWS characterization.

The authors of [203] tested several procedures for the estimation of GU and IW parameters on EWS data; also in this case, IW distribution proved to perform better than GU distribution in the majority of cases.

¹⁶ GEV distributions have three different structures. GEV Type-I distribution is the so called Gumbel (GU) distribution; GEV Type-II distribution is the so called Frechet distribution; GEV Type-III distribution is linked to the Weibull distribution. The Type is determined by the sign of the shape parameter, as illustrated in sub-Section 4.5.1. Since GU distribution is widely used in EWS applications, it will be treated as an individual distribution, although belonging to the GEV family.

A Bayesian estimation procedure of IW parameters was proposed in [204]; the GU distribution was selected as a comparative benchmark in Monte Carlo simulations on synthetic datasets.

IW and GU distributions were also compared in [205] using 100 maximum yearly speeds datasets; in 88 out of 100 cases, IW outperformed GU in terms of GOF. These results were coherent to those shown in [206].

Measurements collected at 5 Danish weather stations were used in [207] to compare GEV and GP distributions. Results showed that GEV distribution had a better GOF than GP distribution.

Several modifications and generalizations of the GU distribution were tested on synthetic and actual monthly peak wind speed data in [208]; among them, the “parent” GEV distribution was also analyzed. The distribution that appeared to be the most flexible in the presented applications was indeed the GEV; a two-component extreme value distribution was also a suitable pick for EWS modeling.

GEV and GP distributions were compared using EWS data taken at offshore locations in [209]. GP appeared to fit available data better than GEV in all of the locations.

The authors of [210] proposed a distribution that mirrored the traditional Weibull distribution until an assigned threshold; after that threshold, the proposed distribution followed the GP distribution. The proposed model was compared to traditional Weibull distribution.

4.3.A NEW INVERSE BURR DISTRIBUTION FOR EXTREME VALUES OF WIND SPEED

The IB distribution has recently been used in some extreme values studies [211,212], proving its versatility in economics and as a stress-strength model for reliability applications [213], but it has never been used for EWS.

The proposal is to test the validity of the distribution for modelling EWS [189].

Parameters of the IB distribution were estimated in relevant literature through the well-known MLE and ME procedures. A new parameter estimation procedure is proposed in this sub-Section: the QE

procedure, that simply consists of the algebraic solution of an equation in specific conditions, thus allowing a faster and easier evaluation.

To provide fair comparisons, results of the three parameter estimation procedures will be compared in terms of both GOF tests on real EWS data, and also in terms of errors coming from the estimation based upon synthetic EWS samples from known distributions.

Details about the proposed IB distribution for EWS characterization and on the corresponding parameter estimation procedures are reported in the following sub-Section 4.3.1; numerical applications based on real and synthetic data are in sub-Section 4.6.

4.3.1. Analytic formulation

The IB distribution is also known as ‘‘Dagum’’ distribution in literature [211,212]; the analytic expressions of IB PDF and CDF for the generic random variable y are, respectively:

$$f_{IB}(y|\rho, \zeta, \gamma) = \frac{\zeta \gamma \rho^\zeta}{\left[1 + \left(\frac{\rho}{y}\right)^\zeta\right]^{\gamma+1} y^{\zeta+1}}, \quad (4.1)$$

$$F_{IB}(y|\rho, \zeta, \gamma) = \frac{1}{\left[1 + \left(\frac{\rho}{y}\right)^\zeta\right]^\gamma}, \quad (4.2)$$

where ρ is the scale parameter, and ζ, γ are the shape parameters, defined as positive numbers. The median value of the IB distribution is:

$$\bar{m}_{IB} = \frac{\rho}{\left(2^{\frac{1}{\gamma}} - 1\right)^\zeta}; \quad (4.3)$$

if the shape parameter $\gamma = 1$, then $\rho = \bar{m}_{IB}$, whatever the value of ζ . Indeed, also from Eq. (4.2), it is trivial to note that $F_{IB}(\rho|\rho, \zeta, 1) = 0.5$.

The generic α_q -quantile $y_{IB}^{(\alpha_q)}$ of the IB distribution is:

$$y_{IB}^{(\alpha_q)} = \frac{\rho}{\left[\left(\frac{1}{\alpha_q} \right)^{\frac{1}{\gamma}} - 1 \right]^{\frac{1}{\zeta}}}. \quad (4.4)$$

Quantile formulation in Eq. (4.4) is very simple to handle; hence, random samples can be easily generated from IB distribution through the well-known inversion method [214]. Obviously, Eq. (4.4) falls into Eq. (4.3) when $\alpha_q = 0.5$. Moreover, if $\gamma = 1$ (and therefore $\rho = \bar{m}_{IB}$), the following, simple relation between the generic α_q -quantile $y_{IB}^{(\alpha_q)}$ and the median $\rho = \bar{m}_{IB}$ holds:

$$y_{IB}^{(\alpha_q)} = \frac{\rho}{k_\alpha^{\zeta'}}, \quad (4.5)$$

where:

$$\zeta' = \frac{1}{\zeta}; \quad k_\alpha = \frac{1}{\alpha_q} - 1. \quad (4.6)$$

Note that the position $\gamma = 1$ is not a strict assumption, since, from the experience made on real datasets, γ is usually not far from this value in many EWS applications.

It is useful to express also the mean value μ_{IB} and the variance σ_{IB}^2 of the Inverse Burr distribution:

$$\mu_{IB} = \gamma \rho B \left(\gamma + \frac{1}{\zeta}, 1 - \frac{1}{\zeta} \right), \quad (4.7)$$

$$\sigma_{IB}^2 = \frac{\rho^2 [\Gamma(\gamma) \Gamma(\gamma + \frac{2}{\zeta}) \Gamma(1 - \frac{2}{\zeta}) - \Gamma^2(\gamma + \frac{1}{\zeta}) \Gamma^2(1 - \frac{1}{\zeta})]}{\Gamma^2(\gamma)}, \quad (4.8)$$

where $B(\cdot, \cdot)$ is the Beta function.

The expression of the generic k_m^{th} order moment is:

$$E[X^{k_m}]_{IB} = \gamma \rho^{k_m} B \left(\gamma + \frac{k_m}{\zeta}, 1 - \frac{k_m}{\zeta} \right). \quad (4.9)$$

4.3.2. Parameter estimation procedures

Since the topic was already discussed in literature [211], only brief hints to the classical MLE and ME of IB distribution are shown here, for sake of conciseness.

The MLE procedure consists in maximizing the likelihood of a given dataset $\mathbf{y} = \{y_1, y_2, \dots, y_{N_{tr}^{(IB)}}\}$ of $N_{tr}^{(IB)}$ independent samples of the variable y . Since the Log-likelihood is usually an easier function to be handled, maximizing the likelihood is often achieved by maximizing the log-likelihood function. For an IB distribution, the Log-likelihood function $L_{IB}(\mathbf{y}|\rho, \zeta, \gamma)$ is:

$$L_{IB}(\mathbf{y}|\rho, \zeta, \gamma) = N_{tr}^{(IB)} \log \gamma + N_{tr}^{(IB)} \log \zeta + N_{tr}^{(IB)} \zeta \log \rho + \\ -(\gamma + 1) \sum_{i=1}^{N_{tr}^{(IB)}} \log \left[1 + \left(\frac{\rho}{y_i} \right)^\zeta \right] - (\zeta + 1) \sum_{i=1}^{N_{tr}^{(IB)}} \log y_i ; \quad (4.10)$$

the MLE of parameters ρ, ζ, γ of IB distribution is therefore achieved by solving the following equation system:

$$\begin{cases} \frac{N_{tr}^{(IB)} \zeta}{\rho} - (\gamma + 1) \sum_{i=1}^{N_{tr}^{(IB)}} \frac{\zeta \rho^{\zeta-1}}{\left[1 + \left(\frac{\rho}{y_i} \right)^\zeta \right] y_i^\zeta} = 0 \\ \frac{N_{tr}^{(IB)}}{\zeta} + N_{tr}^{(IB)} \log \rho + \\ -(\gamma + 1) \sum_{i=1}^{N_{tr}^{(IB)}} \frac{\rho^\zeta (\log \rho - \log y_i)}{\left[1 + \left(\frac{\rho}{y_i} \right)^\zeta \right] y_i^\zeta} - \sum_{i=1}^{N_{tr}^{(IB)}} \log y_i = 0 \\ \frac{N_{tr}^{(IB)}}{\gamma} - \sum_{i=1}^{N_{tr}^{(IB)}} \log \left[1 + \left(\frac{\rho}{y_i} \right)^\zeta \right] = 0. \end{cases} \quad (4.11)$$

The MLE procedure can sometimes yield convergence problems [211]. The ME procedure consists instead in solving a system of three Eqs. (4.9) for three different values of k_m ; it is common to select the lowest orders (first, second and third) to avoid numerical singularities. The equation system to be solved is:

$$\begin{cases} \frac{1}{N_{tr}^{(IB)}} \sum_{i=1}^{N_{tr}^{(IB)}} y_i - \gamma \rho B\left(\gamma + \frac{1}{\zeta}, 1 - \frac{1}{\zeta}\right) = 0 \\ \frac{1}{N_{tr}^{(IB)}} \sum_{i=1}^{N_{tr}^{(IB)}} y_i^2 - \gamma \rho^2 B\left(\gamma + \frac{2}{\zeta}, 1 - \frac{2}{\zeta}\right) = 0 \\ \frac{1}{N_{tr}^{(IB)}} \sum_{i=1}^{N_{tr}^{(IB)}} y_i^3 - \gamma \rho^3 B\left(\gamma + \frac{3}{\zeta}, 1 - \frac{3}{\zeta}\right) = 0. \end{cases} \quad (4.12)$$

However, the solution of the ME system (4.12) is rather cumbersome due to the special function Beta involved, and can be found only numerically.

To avoid such problems, the proposed QE procedure consists in solving a system of three Eqs. (4.4) for three different values of α_q , e.g., $\alpha_{q_1}, \alpha_{q_2}, \alpha_{q_3}$. The sample quantiles $\bar{y}^{(\alpha_{q_1})}, \bar{y}^{(\alpha_{q_2})}, \bar{y}^{(\alpha_{q_3})}$ are calculated from dataset \mathbf{y} for each of the quantile levels, and the equation system to be solved therefore is:

$$\begin{cases} \bar{y}^{(\alpha_{q_1})} - \rho \left[\left(\frac{1}{\alpha_{q_1}} \right)^{\frac{1}{\gamma}} - 1 \right]^{-\frac{1}{\zeta}} = 0 \\ \bar{y}^{(\alpha_{q_2})} - \rho \left[\left(\frac{1}{\alpha_{q_2}} \right)^{\frac{1}{\gamma}} - 1 \right]^{-\frac{1}{\zeta}} = 0 \\ \bar{y}^{(\alpha_{q_3})} - \rho \left[\left(\frac{1}{\alpha_{q_3}} \right)^{\frac{1}{\gamma}} - 1 \right]^{-\frac{1}{\zeta}} = 0. \end{cases} \quad (4.13)$$

A comparison between equation systems (4.11), (4.23), and (4.13) suggests that QE can be a simpler and more intuitive alternative procedure to the MLE and ME procedures, since no special functions are involved in the solution of the system. Also, if $\gamma = 1$, the solution of the QE procedure can be expressed in closed analytic form, which yields to a significant simplification.

4.4. A NEW INVERSE BURR - INVERSE WEIBULL MIXTURE

DISTRIBUTION FOR EXTREME VALUES OF WIND SPEED

The proposal shown in this sub-Section is a finite M-IB-IW that combines the IB and the IW distributions in the mixture framework, in order to suit different scenarios [190].

Mixture distributions are sometimes questionable, since they usually introduce severe difficulties in their parameter estimation [194,215]. To cope with this point of concern, an EM procedure that is particularly suitable for estimating parameters of mixture distributions is also proposed.

The analytic formulation of the proposed M-IB-IW distribution for EWS characterization is reported in the following sub-Section 4.4.1, with hints on its classical MLE procedure; details on the proposed EM parameter estimation procedure are in sub-Section 4.4.2.

4.4.1. Analytic formulation

The M-IB-IW distribution proposed for EWS modeling in this Chapter is obtained through a weighted sum of IB and IW distributions. Obviously, in order to satisfy that the resulting function is still a probability distribution, the sum of both weights must be unitary.

Assuming the same parameter formulation for IW and IB distributions depicted in sub-Sections 4.5.3 and 4.3.1, respectively, the PDF and CDF of the M-IB-IW distribution are expressed as follows, respectively:

$$f_{M-IB-IW}(y|\rho, \zeta, \gamma, \nu, \delta, \omega_M) = \omega_M \frac{\zeta \gamma \rho^\zeta}{\left[1 + \left(\frac{\rho}{y}\right)^\zeta\right]^{\gamma+1} y^{\zeta+1}} + (1 - \omega_M) \frac{\delta}{\nu^\delta y^{\delta+1}} e^{-\frac{1}{(\nu y)^\delta}}; \quad (4.14)$$

$$F_{M-IB-IW}(y|\rho, \zeta, \gamma, \nu, \delta, \omega_M) = \omega_M \frac{1}{\left[1 + \left(\frac{\rho}{y}\right)^\zeta\right]^\gamma} + (1 - \omega_M) e^{-\frac{1}{(\nu y)^\delta}}; \quad (4.15)$$

where ν, δ are the scale and shape parameters of the IW distribution, respectively, and $0 \leq \omega_M \leq 1$ is the weight parameter. The random variable y is defined only for positive values; the mean value

$\mu_{M-IB-IW}$ of the M-IB-IW distribution is defined only if $\delta > 1$ and $\zeta > 1$, and it is:

$$\mu_{M-IB-IW} = \omega_M \gamma \rho B\left(\gamma + \frac{1}{\zeta}, 1 - \frac{1}{\zeta}\right) + (1 - \omega_M) \frac{\Gamma(1 - \frac{1}{\delta})}{\nu}. \quad (4.16)$$

The MLE of mixture distributions is not easily formulated and can suffer from convergence problems [194,215]. Indeed, the main peculiarity is that the logarithmic transformation of the mixture likelihood function does not simplify the problem of finding the solutions of the maximization problem, as the weighted sum of two probability functions cannot be separated in distinct terms. The approach followed to obtain numerical solutions of the M-IB-IW problem is the maximization of the likelihood function $\mathcal{L}_{M-IB-IW}(\mathbf{y}|\rho, \zeta, \gamma, \nu, \delta, \omega_M)$, given the dataset $\mathbf{y} = \{y_1, y_2, \dots, y_{N_{tr}^{(M-IB-IW)}}\}$ of $N_{tr}^{(M-IB-IW)}$ independent samples of the variable y . Estimations of parameters are found in their corresponding domains by maximizing the likelihood function, that obviously is:

$$\begin{aligned} \mathcal{L}_{M-IB-IW}(\mathbf{y}|\rho, \zeta, \gamma, \nu, \delta, \omega_M) = \\ = \prod_{i=1}^{N_{tr}^{(M-IB-IW)}} \omega_M \frac{\zeta \gamma \rho^\zeta}{\left[1 + \left(\frac{\rho}{y_i}\right)^\zeta\right]^{\gamma+1} y_i^{\zeta+1}} + (1 - \omega_M) \frac{\delta}{\nu^\delta y_i^{\delta+1}} e^{-\frac{1}{(\nu y_i)^\delta}}. \end{aligned} \quad (4.17)$$

The solution of the M-IB-IW MLE can be found numerically, e.g., in MATLAB environment; however, it sometimes yields convergence problems. Indeed, the accuracy of the results is strongly affected by the choice of the initial values of variables provided to the maximization algorithm; in fact, if the initials points were arbitrarily assigned, the M-IB-IW MLE procedure could lead to misleading results, or could not converge. In order to avoid the above problems, the MLEs of the two constituent distributions were necessary in order to provide suitable initial values of the IB and IW parameters; the weight ω_M is initially set to 0.5. However, it is worth noting that performing initial MLEs for constituent distributions is not a trivial task for mixtures of complex

distributions, or for mixtures of more than two constituent distributions. Therefore, further solutions need to be developed to avoid these problems.

4.4.2. Parameter estimation procedures

A more elaborate technique that does not fail to find solutions of the likelihood maximization is desirable when a mixture distribution is considered, due to the aforementioned problems of the classical MLE procedure. The EM procedure [215] was considered for this task, as its formulation avoids convergence problems in case of mixture distributions, and it also provides results that were, on average, slightly more accurate in numerical experiments.

The aim of the EM procedure is still to maximize the likelihood function $\mathcal{L}_{M-IB-IW}(\mathbf{y}|\rho, \zeta, \gamma, \nu, \delta, \omega_M) = \mathcal{L}_{M-IB-IW}(\mathbf{y}|\boldsymbol{\theta})$, where $\boldsymbol{\theta} = \{\rho, \zeta, \gamma, \nu, \delta, \omega_M\}$ is the vector of M-IB-IW parameters that is conveniently introduced to lighten the notation. The EM is based on the hypothesis of additional, hidden parameters used to simplify the likelihood function.

Let's assume that a set of unobservable (and, therefore, unobserved) data $\mathbf{d} = \{d_1, d_2, \dots, d_{N_{tr}^{(M-IB-IW)}}\}$ exists, and let's assume that the unobserved data is able to provide some hints about which component distribution (the IB or the IW) generated each data point of the dataset \mathbf{y} (i.e., $d_i = 1$ or $d_i = 2$ if the value y_i was drawn from the IB or the IW distribution, respectively, $\forall i$). Obviously, given the set of observed (incomplete) data \mathbf{y} , the complete data is $\mathbf{z} = \{\mathbf{y}, \mathbf{d}\}$.

The log-likelihood function $L_{M-IB-IW}(\mathbf{y}|\boldsymbol{\theta})$ of the incomplete data \mathbf{y} modeled through a M-IB-IW distribution is the natural logarithm of Eq. (4.17). If one could know the data \mathbf{d} , the joint probability theorem would assure that the log-likelihood $L_{M-IB-IW}(\mathbf{z}|\boldsymbol{\theta})$ of the complete data is:

$$\begin{aligned} L_{M-IB-IW}(\mathbf{z}|\boldsymbol{\theta}) &= L_{M-IB-IW}(\mathbf{y}, \mathbf{d}|\boldsymbol{\theta}) = \\ &= \sum_{i=1}^{N_{tr}^{(M-IB-IW)}} \log[f_{M-IB-IW}(y_i|d_i, \boldsymbol{\theta})p(d_i|\boldsymbol{\theta})], \end{aligned} \quad (4.18)$$

where:

$$p(d_i|\boldsymbol{\theta}) = \begin{cases} \omega_M & \text{if } d_i = 1 \\ 1 - \omega_M & \text{if } d_i = 2 \end{cases}; \quad (4.19)$$

$$f_{M-IB-IW}(y_i|d_i, \boldsymbol{\theta}) = \begin{cases} \frac{\zeta \gamma \rho^\zeta}{\left[1 + \left(\frac{\rho}{y_i}\right)^\zeta\right]^{\gamma+1}} y_i^{\zeta+1} & \text{if } d_i = 1 \\ \frac{\delta}{v^\delta y_i^{\delta+1}} e^{-\frac{1}{(vy_i)^\delta}} & \text{if } d_i = 2. \end{cases} \quad (4.20)$$

Obviously, the unobservable data \mathbf{d} is not available, and therefore Eq. (4.18) cannot be directly computed. However, given an initial point estimate $\widehat{\boldsymbol{\theta}}_0$ of $\boldsymbol{\theta}$, both corresponding terms in Eq. (4.20) still can be found:

$$\begin{aligned} f_{IB}(y_i|\widehat{\boldsymbol{\theta}}_0) &= f_{IB}(y_i|\widehat{\rho}_0, \widehat{\zeta}_0, \widehat{\gamma}_0) = \\ &= f_{M-IB-IW}(y_i|d_i = 1, \widehat{\rho}_0, \widehat{\zeta}_0, \widehat{\gamma}_0) = \frac{\widehat{\zeta}_0 \widehat{\gamma}_0 \widehat{\rho}_0^{\widehat{\zeta}_0}}{\left[1 + \left(\frac{\widehat{\rho}_0}{y_i}\right)^{\widehat{\zeta}_0}\right]^{\widehat{\gamma}_0+1}} y_i^{\widehat{\zeta}_0+1}; \end{aligned} \quad (4.21)$$

$$\begin{aligned} f_{IW}(y_i|\widehat{\boldsymbol{\theta}}_0) &= f_{IW}(y_i|\widehat{v}_0, \widehat{\delta}_0) = \\ &= f_{M-IB-IW}(y_i|d_i = 2, \widehat{v}_0, \widehat{\delta}_0) = \frac{\widehat{\delta}_0}{\widehat{v}_0^{\widehat{\delta}_0} y_i^{\widehat{\delta}_0+1}} e^{-\frac{1}{(\widehat{v}_0 y_i)^{\widehat{\delta}_0}}}. \end{aligned} \quad (4.22)$$

If the weights in Eq. (4.19) are considered as prior probabilities for IB and IW distributions, the Bayes' formula can be applied to find the posterior probability $p(\mathbf{d}|\mathbf{y}, \widehat{\boldsymbol{\theta}}_0)$ of the unobservable data \mathbf{d} , given the observed data \mathbf{y} and the initial point estimate $\widehat{\boldsymbol{\theta}}_0$:

$$\begin{aligned} p(\mathbf{d}|\mathbf{y}, \widehat{\boldsymbol{\theta}}_0) &= \prod_{i=1}^{N_{tr}^{(M-IB-IW)}} p(d_i|y_i, \widehat{\boldsymbol{\theta}}_0) = \\ &= \prod_{i=1}^{N_{tr}^{(M-IB-IW)}} \frac{f_{M-IB-IW}(y_i|d_i, \widehat{\boldsymbol{\theta}}_0) p(d_i|\widehat{\boldsymbol{\theta}}_0)}{f_{M-IB-IW}(y_i|\widehat{\boldsymbol{\theta}}_0)}. \end{aligned} \quad (4.23)$$

It is trivial to note that:

$$p(d_i = 1|y_i, \widehat{\boldsymbol{\theta}}_0) = \frac{f_{IB}(y_i|\widehat{\boldsymbol{\theta}}_0) \widehat{\omega}_{M_0}}{f_{M-IB-IW}(y_i|\widehat{\boldsymbol{\theta}}_0)}; \quad (4.24)$$

$$p(d_i = 2|y_i, \hat{\boldsymbol{\theta}}_0) = \frac{f_{IW}(y_i|\hat{\boldsymbol{\theta}}_0)(1-\hat{\omega}_{M_0})}{f_{M-IB-IW}(y_i|\hat{\boldsymbol{\theta}}_0)}. \quad (4.25)$$

The EM algorithm aims to maximize the expected value $q_E(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}_0)$ of the log-likelihood $L_{M-IB-IW}(\mathbf{z}|\boldsymbol{\theta})$ of the complete data in Eq. (4.18), given the observed data \mathbf{y} and the initial point estimate $\hat{\boldsymbol{\theta}}_0$:

$$\begin{aligned} q_E(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}_0) &= \mathbb{E}[L_{M-IB-IW}(\mathbf{z}|\boldsymbol{\theta})|\mathbf{y}, \hat{\boldsymbol{\theta}}_0] = \\ &= \int L_{M-IB-IW}(\mathbf{y}, \mathbf{d}|\boldsymbol{\theta})p(\mathbf{d}|\mathbf{y}, \hat{\boldsymbol{\theta}}_0)d\mathbf{d}. \end{aligned} \quad (4.26)$$

The second factor of the function to be integrated in Eq. (4.26) corresponds to Eq. (4.23). With application to the M-IB-IW distribution, Eq. (4.26) is transformed in the following form through some manipulations shown in the Appendix of [190]:

$$\begin{aligned} q_E(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}_0) &= \sum_{i=1}^{N_{tr}^{(M-IB-IW)}} \log \omega_M \cdot p(d_i = 1|y_i, \hat{\boldsymbol{\theta}}_0) + \\ &+ \sum_{i=1}^{N_{tr}^{(M-IB-IW)}} \log(1 - \omega_M) \cdot p(d_i = 2|y_i, \hat{\boldsymbol{\theta}}_0) + \\ &+ \sum_{i=1}^{N_{tr}^{(M-IB-IW)}} \log[f_{IB}(y_i|\rho, \zeta, \gamma)] \cdot p(d_i = 1|y_i, \hat{\boldsymbol{\theta}}_0) + \\ &+ \sum_{i=1}^{N_{tr}^{(M-IB-IW)}} \log[f_{IW}(y_i|\nu, \delta)] \cdot p(d_i = 2|y_i, \hat{\boldsymbol{\theta}}_0). \end{aligned} \quad (4.27)$$

The only unknown parameter in the first two sums of Eq. (4.27) is the weight ω_M ; the unknown parameters in the remaining two sums of Eq. (4.27) are $\rho, \zeta, \gamma, \nu, \delta$. Thus, it is possible to decouple the maximization of $q_E(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}_0)$ in two steps, i.e., the maximization of the first two sums (expectation step) and the maximization of the remaining two sums (maximization step).

An analytic solution of the expectation step can be provided by setting the partial derivative in ω_M of the first two sums in Eq. (4.27) to zero; it is:

$$\hat{\omega}_M^{(EM)} = \frac{1}{N_{tr}^{(M-IB-IW)}} \sum_{i=1}^{N_{tr}^{(M-IB-IW)}} p(d_i = 1|y_i, \hat{\boldsymbol{\theta}}_0), \quad (4.28)$$

while the solution of the maximization step is found numerically.

The EM algorithm therefore is the following. Starting from an initial point estimate $\hat{\boldsymbol{\theta}}_0$, solve the expectation step applying Eq. (4.28) and the maximization step numerically in order to find $\hat{\boldsymbol{\theta}}$. Then, a test of convergence is run. Given an arbitrary threshold $\bar{\tau}^{(EM)}$, the EM algorithm stops if $\|\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_0\| < \bar{\tau}^{(EM)}$; otherwise, the initial point estimate $\hat{\boldsymbol{\theta}}_0$ is refreshed by setting it equal to the parameters estimated in the expectation and maximization steps ($\hat{\boldsymbol{\theta}}_0 = \hat{\boldsymbol{\theta}}$). If the test of convergence is not satisfied, the whole process is repeated until the condition on the threshold is satisfied. Obviously, threshold selection derives from a compromise between accuracy and computational effort of the EM procedure. Considering that both vectors $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_0$ are made of six components, after few runs of the algorithm on test cases, a reasonable threshold $\bar{\tau}^{(EM)} = 10^{-4}$ is selected for all of the analyses reported in Chapter.

4.5. NUMERICAL APPLICATIONS

Numerical applications based on real and synthetic EWS data are shown in this sub-Section to test the usefulness of the proposed models. In particular, benchmark distributions used to compare the proposed models to acknowledged models are shown in sub-Section 4.5.1. The validity of the IB distribution is assessed in sub-Section 4.5.2, while the validity of the M-IB-IW distribution is assessed in sub-Section 4.5.3.

4.5.1. Benchmark distributions for modeling EWS

An accurate comparison with the acknowledged distributions that are already available in relevant literature EWS modelling is obviously mandatory in order to catch the usefulness of the new proposals. The GEV, the GU and the IW distributions are briefly recalled in this sub-Section, since they were used as benchmark models to compare the performances of the IB and the M-IB-IW distributions in terms of GOF.

4.5.1.1. *The Generalized Extreme Value distribution*

The analytic expressions of GEV PDF and CDF for the generic random variable y are, respectively:

$$f_{GEV}(y|l, v, \kappa) = \frac{1}{l} \left(1 + \kappa \frac{y-v}{l}\right)^{-1-\frac{1}{\kappa}} e^{-\left(1+\kappa\frac{y-v}{l}\right)^{-\frac{1}{\kappa}}} \quad \text{for } \kappa \neq 0; \quad (4.29)$$

$$f_{GEV}(y|l, v, \kappa) = \frac{1}{l} e^{-\frac{y-v}{l}} e^{-\frac{y-v}{l}} \quad \text{for } \kappa = 0; \quad (4.30)$$

$$F_{GEV}(y|l, v, \kappa) = e^{-\left(1+\kappa\frac{y-v}{l}\right)^{-\frac{1}{\kappa}}} \quad \text{for } \kappa \neq 0; \quad (4.31)$$

$$F_{GEV}(y|l, v, \kappa) = e^{-e^{-\frac{y-v}{l}}} \quad \text{for } \kappa = 0; \quad (4.32)$$

where l is a positive scale parameter, v is a real location parameter and κ is a real shape parameter that determines the distribution Type (i.e., Type I for $\kappa = 0$, Type II for $\kappa > 0$ and Type III for $\kappa < 0$). Note that Type I distribution is GU distribution; it is treated in details in sub-Section 4.5.1.2 since it is a widely-used distribution for EWS. Therefore, only Type II and III GEV distributions are treated in the following. In such conditions, the random variable y is defined in the intervals:

$$y \geq v - \frac{l}{\kappa} \quad \text{for } \kappa > 0; \quad (4.33)$$

$$y \leq v - \frac{l}{\kappa} \quad \text{for } \kappa < 0. \quad (4.34)$$

The mean value μ_{GEV} of the Type II or Type III GEV distribution is defined only for $\kappa < 1$, and it is:

$$\mu_{GEV} = v + l \frac{\Gamma(1-\kappa)-1}{\kappa} \quad (4.35)$$

where $\Gamma(\cdot)$ is the Gamma function. The variance σ_{GEV}^2 of the Type II or Type III GEV distribution is defined only for $\kappa < 0.5$, and it is:

$$\sigma_{GEV}^2 = l^2 \frac{\Gamma(1-2\kappa)-\Gamma^2(1-\kappa)}{\kappa^2}. \quad (4.36)$$

4.5.1.2. The Gumbel distribution

The analytic expressions of GU PDF and CDF for the generic random variable y are, respectively:

$$f_{GU}(y|\chi, o) = \frac{1}{o} e^{-\left[\frac{y-\chi}{o} + e^{-\frac{y-\chi}{o}}\right]}, \quad (4.37)$$

$$F_{GU}(y|\chi, o) = e^{-e^{-\frac{y-\chi}{o}}}, \quad (4.38)$$

where χ, o are the location parameter, defined as a real number, and the scale parameter, defined as a positive number, respectively. The median value of GU distribution is:

$$\bar{m}_{GU} = \chi - o \log(\log 2). \quad (4.39)$$

The generic α_q -quantile $y_{GU}^{(\alpha_q)}$ of the GU distribution is:

$$y_{GU}^{(\alpha_q)} = \chi - o \log\left(\log \frac{1}{\alpha_q}\right). \quad (4.40)$$

It is useful to express also the mean value μ_{GU} and the variance σ_{GU}^2 of the GU distribution, respectively, as:

$$\mu_{GU} = \chi + o\bar{\gamma}_{Eu}, \quad (4.41)$$

$$\sigma_{GU}^2 = \frac{\pi^2 o^2}{6}. \quad (4.42)$$

where $\bar{\gamma}_{Eu} \approx 0.5772$ is Euler-Mascheroni constant.

4.5.1.3. The Inverse Weibull distribution

The analytic expressions of IW PDF and CDF for the generic random variable y are, respectively:

$$f_{IW}(y|v, \delta) = \frac{\delta}{v^\delta y^{\delta+1}} e^{-\frac{1}{(vy)^\delta}}, \quad (4.43)$$

$$F_{IW}(y|v, \delta) = e^{-\frac{1}{(vy)^\delta}}, \quad (4.44)$$

where v, δ are the scale parameter and the shape parameter, respectively, both defined as positive numbers. The median value of IW distribution is:

$$\bar{m}_{IW} = \frac{1}{v(\log 2)^{\frac{1}{\delta}}}. \quad (4.45)$$

The generic α_q -quantile $y_{IW}^{(\alpha_q)}$ of the IW distribution is:

$$y_{IW}^{(\alpha_q)} = \frac{1}{v(-\log \alpha_q)^{\frac{1}{\delta}}}. \quad (4.46)$$

It is useful to express also the mean value μ_{IW} and the variance σ_{IW}^2 of the IW distribution:

$$\mu_{IW} = \frac{\Gamma(1-\frac{1}{\delta})}{v}, \quad (4.47)$$

$$\sigma_{IW}^2 = \frac{\Gamma(1-\frac{2}{\delta}) - \Gamma^2(1-\frac{1}{\delta})}{v^2}, \quad (4.48)$$

where $\Gamma(\cdot)$ denotes the Gamma function. Note that the mean value is defined only if $\delta > 1$, and the variance is defined only if $\delta > 2$. The expression of the generic k_m th order moment is:

$$E[X^{k_m}]_{IW} = \frac{\Gamma(1-\frac{k_m}{\delta})}{v^{k_m}}. \quad (4.49)$$

4.5.2. Inverse Burr distribution

The IB distribution was tested in order to prove its usefulness for EWS characterization. Datasets of actual measurements of wind speed collected in different sites were selected and modeled through the GU and IW distributions, and through the proposed IB distribution, in order to justify the proposal of the latter for EWS characterization. For

each dataset, Kolmogorov-Smirnov (KS) test [216] and Chi-square (CS) test [217] were performed in order to compare the models; the parameters of each distribution were estimated through the MLE, ME and QE procedures shown in sub-Section 4.3.2. In particular, the IB QE was performed considering 0.1-, 0.5- and 0.9-quantiles of each dataset.

A further comparison among the results obtained through the different estimation procedures was performed also on synthetic samples of EWS, drawn from known IB distributions through the inversion method [214]. This helps to test the accuracy of the generic estimated parameter, comparing it to a known value; it is possible to directly compare the results obtained through the different estimation procedures for IB distributions.

Information on the characteristics of measured EWS data are provided in sub-Section 4.5.2.1; results of the fitting GU, IW, and IB on measured EWS data are shown in sub-Section 4.5.2.2. Eventually, IB parameter estimation procedures applied to synthetic EWS samples are compared in sub-Section 4.5.3.3.

4.5.2.1. Data characteristics

Different datasets were collected from several locations in different years, in order to validate the proposed IB distribution in many different conditions.

The first and second datasets (D1 and D2) were drawn from a set of 1578240 observations of peak values of wind speed, collected each minute from 1 January 2012 to 31 December 2014 by NREL National Wind Technology Center M2 Tower; Boulder, Colorado, USA at latitude $39^{\circ}54'$ north and longitude $105^{\circ}14'$ west [218]. Two smaller datasets, D1 and D2 respectively, were then extracted in a PMA framework from this set of measurements, in order to be used for the EWS statistical characterization. In particular, the 36 maximum values registered in each of the 36 months of observation were selected to build dataset D1, while the 52 maximum values registered in each of the 52 weeks of year 2014 were selected to build dataset D2.

The third and fourth datasets (D3 and D4) were drawn from a set of 1578240 observations of peak values of wind speed, collected each minute from 1 August 2010 to 31 July 2013 in the framework of

NREL Solar Resource & Meteorological Assessment Project in Milford, Utah, USA (latitude 38°41' north and longitude 113°03' west) [219]. Then, the 36 maximum values registered in each of the 36 months of observation were selected to build dataset D3, while the 52 maximum values registered in each of the 52 weeks of year 2010 were selected in a PMA framework to build dataset D4.

The fifth and sixth datasets (D5 and D6) were drawn from a set of 105120 observations of wind speed, averaged each 10 minutes from 1 January 2001 to 31 December 2002. Measurements were publically provided by Gestore dei Servizi Energetici (the Italian state-owned company that promotes and supports renewable energies) and were collected in Forenza, Italy (latitude 40°52' north and longitude 15°51' east) [180]. Then, the 24 maximum values registered in each of the 24 months of observation were selected to build dataset D5, while the 52 maximum values registered in each of the 52 weeks of year 2001 were selected to build dataset D6.

The seventh dataset (D7) was drawn from a set of 8760 observations of wind speed, averaged each hour from 1 January 2006 to 31 December 2006. Measurements were publically provided by the Sustainable Energy Authority of Ireland and were collected in Sligo, Ireland (latitude 54°16' north and longitude 8°28' west [220]). The 52 maximum values registered in each of the 52 weeks of year 2006 were selected to build dataset D7.

All of the seven datasets were initially filtered by first eliminating missing and bad data. Then, the Tukey's test [221] was performed to individuate possible outliers and suspicious values: values external to the following interval were considered as probable outliers, and then required for further investigation:

$$\left[\begin{array}{l} \bar{y}^{(0.25)} - k^{(Tu)}(\bar{y}^{(0.75)} - \bar{y}^{(0.25)}) \\ \bar{y}^{(0.75)} + k^{(Tu)}(\bar{x}^{(0.75)} - \bar{y}^{(0.25)}) \end{array} \right]. \quad (4.50)$$

In Eq. (4.50), $\bar{y}^{(0.25)}$ and $\bar{y}^{(0.75)}$ are the 0.25-quantile and the 0.75-quantile of each dataset, respectively; the coefficient $k^{(Tu)}$ was set as $k^{(Tu)} = 1.5$ to determine the outliers, and then as $k^{(Tu)} = 3$ to determine suspicious values. However, for the seven datasets D1-D7

considered in our applications, no values were found outside the interval (4.50) with coefficient $k^{(Tu)} = 3$; therefore the initial filtering of data appeared to eliminate bad values that would have led to incorrect estimations.

4.5.2.2. *Assessment of Gumbel, Inverse Weibull, and Inverse Burr distributions on measured EWS data*

GU, IW, and IB distributions were used to fit the datasets D1-D7. Parameters of the corresponding distributions estimated through the MLE, ME and QE procedures are shown in Table 4.1. Examples of the corresponding CDFs are shown in Figs. 4.1-4.3 only for one datasets, to avoid unnecessary redundancy: the estimated CDFs for dataset D2 are therefore shown in Figs. 4.1, 4.2, and 4.3.

Table 4.1 - Values of Gumbel, Inverse Weibull, and Inverse Burr parameters estimated through MLE, ME, and QE procedures.

Dataset and estimation procedure	Estimated distribution parameters						
	Gumbel		Inverse Weibull		Inverse Burr		
	χ	σ	ν	δ	ρ	ζ	γ
D1 - MLE	27.901	5.787	0.037	4.983	24.208	5.937	2.745
D1 - ME	28.076	5.412	0.036	6.709	34.223	10.712	0.526
D1 - QE	27.153	6.970	0.037	4.307	8.380	0.642	33.851
D2 - MLE	22.620	5.933	0.046	3.651	23.824	6.106	1.209
D2 - ME	22.692	5.652	0.044	5.532	28.519	8.212	0.564
D2 - QE	22.835	5.462	0.044	4.596	4.577	5.485	15.811
D3 - MLE	18.896	2.597	0.053	7.167	20.936	15.032	0.701
D3 - ME	18.967	2.261	0.053	9.909	21.156	15.030	0.646
D3 - QE	18.979	2.200	0.053	9.062	22.192	22.222	0.323
D4 - MLE	15.574	4.000	0.067	3.403	19.903	11.879	0.422
D4 - ME	15.760	3.122	0.064	6.546	20.028	11.753	0.415
D4 - QE	15.953	3.132	0.063	5.516	20.647	17.215	0.257
D5 - MLE	23.517	4.305	0.043	5.386	24.586	8.702	1.204
D5 - ME	23.561	4.048	0.042	7.383	26.513	10.209	0.780
D5 - QE	24.096	3.815	0.042	6.744	26.231	8.702	0.782
D6 - MLE	17.507	5.621	0.061	2.886	19.812	5.382	0.954
D6 - ME	17.414	5.311	0.057	4.837	24.244	7.733	0.438
D6 - QE	18.115	5.787	0.056	3.532	16.468	4.489	1.663
D7 - MLE	15.289	3.695	0.068	4.044	17.987	8.023	0.738
D7 - ME	15.397	3.320	0.065	6.142	19.163	9.665	0.521
D7 - QE	15.439	3.180	0.065	5.276	19.502	8.121	0.550

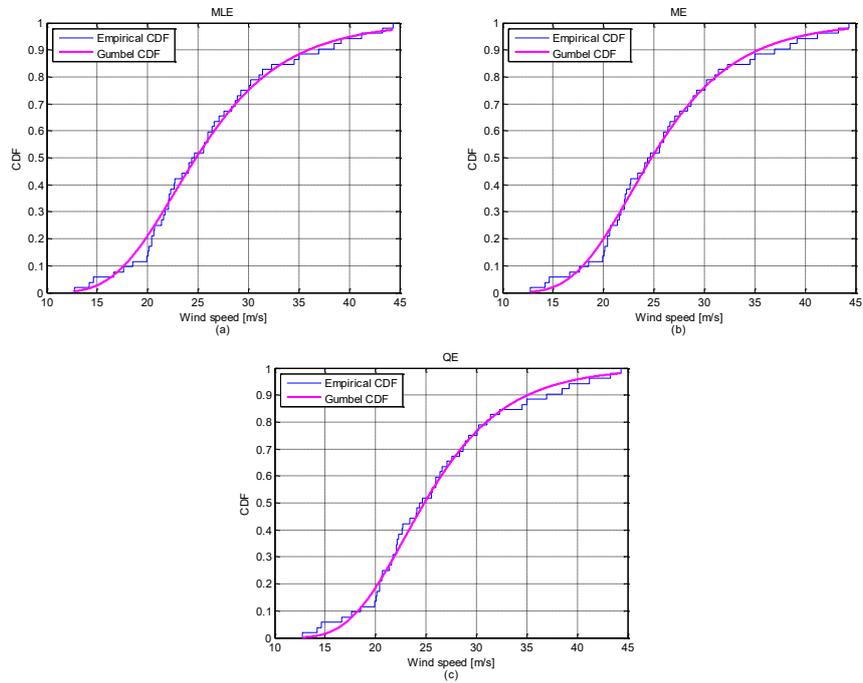


Figure 4.1 – Gumbel fitting of dataset D2 through (a) MLE procedure; (b) ME procedure; (c) QE procedure

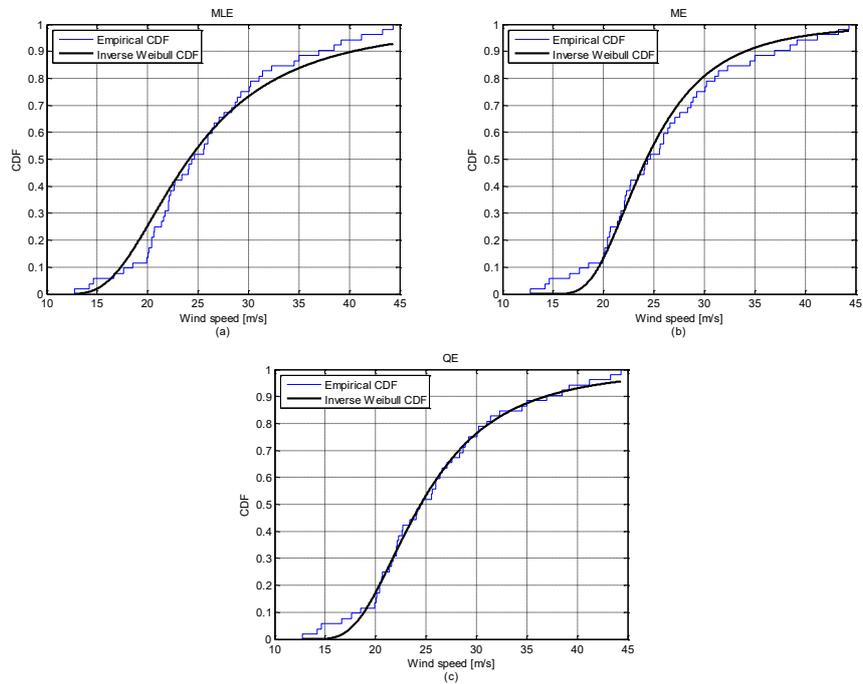


Figure 4.2 - Inverse Weibull fitting of dataset D2 through (a) MLE procedure; (b) ME procedure; (c) QE procedure.

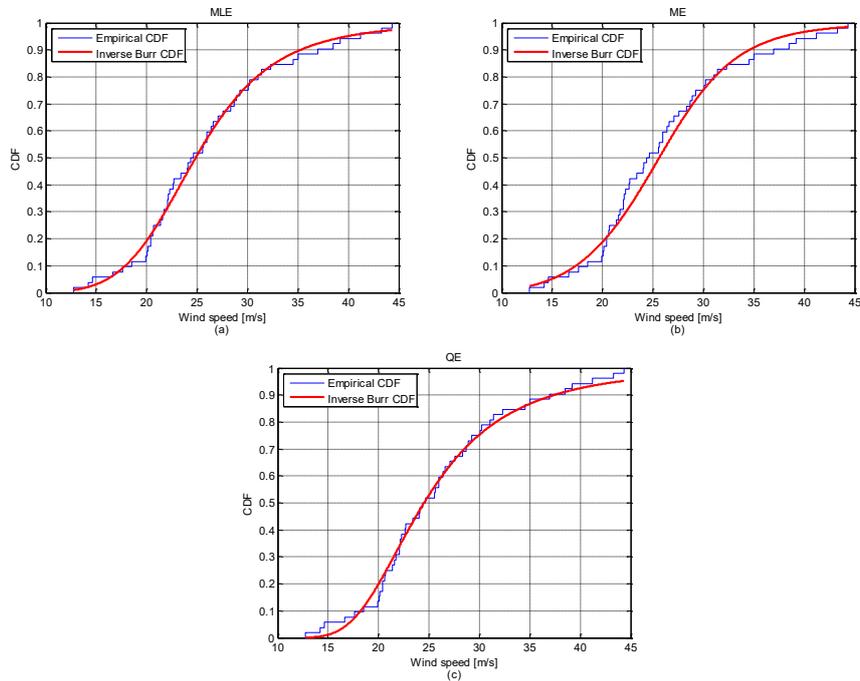


Figure 4.3 - Inverse Burr fitting of dataset D2 through (a) MLE procedure; (b) ME procedure; (c) QE procedure.

As shown from the results reported in Table 4.1 and in Figs. 4.1-4.3, even if the estimated parameters of IB distributions are different in value using different estimation procedures, the corresponding CDFs appear to properly fit the EWS samples. Parameter χ of GU distribution shows no appreciable differences on the basis of the estimation procedures, while parameter ρ appears to be slightly different by using MLE, ME or QE procedures. This behaviour is the same also in IW fitting, since parameter ν shows no appreciable differences, while parameter δ appears to be slightly different by using different procedures. From a graphical comparison between Figs. 4.1-4.3, IB distributions appear in some cases to better fit actual samples of EWS than the corresponding GU and IW distributions. However, further analyses are needed to reach a conclusion in terms of both quality of estimation and GOF.

The GOF of estimated GU, IW, and IB distributions was also assessed through KS test and CS test at 5% significance level. Once the significance level is fixed, the critical values of KS test depend only on

the size of the sample dataset to be fitted. Thus, the critical value for datasets D1 and D3 is 0.2212; the critical value for datasets D2, D4, D6 and D7 is 0.1848, and the critical value for dataset D5 is 0.2693. Results of the KS test statistics (KS_{TS}) and results of CS test statistics (CS_{TS}) are shown in Table 4.2.

Table 4.2 - Kolmogorov-Smirnov test statistics and Chi-square test statistics for Gumbel, Inverse Weibull, and Inverse Burr fitting distributions. Bold italic values denote failed tests, while underlined values correspond to the best-fitting distributions for each dataset.

Dataset and estimation procedure	Test statistics for estimated distributions					
	Gumbel		Inverse Weibull		Inverse Burr	
	KS_{TS}	CS_{TS}	KS_{TS}	CS_{TS}	KS_{TS}	CS_{TS}
D1 - MLE	0.153	2.745	0.174	3.543	0.152	3.016
D1 - ME	0.144	3.034	0.185	7.810	<u>0.118</u>	6.713
D1 - QE	0.148	1.852	0.170	2.731	0.125	0.121
D2 - MLE	0.090	2.783	0.131	1.136	0.071	2.699
D2 - ME	0.078	2.240	0.080	3.398	0.103	5.219
D2 - QE	<u>0.065</u>	1.977	0.070	1.920	0.067	2.311
D3 - MLE	0.149	4.495	0.171	2.719	0.086	1.130
D3 - ME	0.155	3.365	0.175	2.472	<u>0.083</u>	1.216
D3 - QE	0.156	3.624	0.168	2.976	0.086	0.786
D4 - MLE	0.151	5.595	0.181	14.908	0.079	1.183
D4 - ME	0.161	2.862	0.201	6.367	0.076	1.255
D4 - QE	0.140	3.752	0.159	3.988	<u>0.071</u>	2.243
D5 - MLE	0.112	0.202	0.147	0.520	<u>0.080</u>	0.339
D5 - ME	0.108	0.150	0.132	1.642	0.103	0.710
D5 - QE	0.096	0.562	0.113	0.357	0.082	0.358
D6 - MLE	0.074	4.002	0.137	6.233	0.068	4.043
D6 - ME	<u>0.066</u>	4.404	0.134	2.309	0.117	4.957
D6 - QE	0.117	4.782	0.091	1.649	0.086	4.111
D7 - MLE	0.088	6.463	0.127	8.292	0.059	2.862
D7 - ME	0.098	5.920	0.144	4.547	<u>0.055</u>	1.846
D7 - QE	0.102	6.133	0.119	5.380	0.090	5.755

Bold values in Table 4.2 correspond to unsuccessful tests; only IW distribution failed to pass one of the GOF tests in three cases. In particular, the null hypothesis of the CS test was rejected twice (once for the MLE of D4 and once for the MLE of D7), while the null hypothesis of the KS test was rejected only for the ME of D4. In all other cases, the null hypothesis was accepted, since the small amount

of samples contained in each dataset. However, further indications about the goodness of fitting can be drawn from the values of test statistics. KS_{TS} is a negatively-oriented test statistic; values closer to zero are likely to indicate a better level of fitting. Underlined values in Table 4.2 correspond to the lowest values of KS_{TS} for each dataset. In five of seven considered dataset, the proposed IB model provided the lowest value of KS_{TS} ; GU distribution provided instead the lowest values of KS_{TS} for D2 and D6 only. These results suggest that IB model and GU models appear to be more indicated than IW model for all the considered datasets, with a clear better behavior of IB model. CS test is used here to confirm results of KS test in terms of acceptance or refusal of models. Since CS test is conveniently used for categorical variables [217], the number of degrees of freedom was obtained numerically in MATLAB environment.

4.5.2.3. *Assessment of Inverse Burr distribution on synthetic EWS data*

Samples of EWS were drawn from IB distributions with known parameters through the inversion method, and the corresponding parameters were then estimated through the MLE, ME and QE procedures to test their evaluation performances. The goodness of the generic parameter estimated through the different procedures was quantified in terms of MAE and MAPE indices, averaging the errors obtained with respect to the actual value of the parameter at each iteration.

Median value of EWS was set to $\bar{m}_{IB} = 38$ for each cycle of variation of parameters ζ, γ . In particular, γ was assumed to be equal to five values ($\gamma = 1, 1.5, 2, 2.5, 3$), and for each cycle of variation of γ the parameter ζ was changed 50 times, keeping the ratio between theoretical standard deviation and theoretical mean under 20%. Tables 4.3 and 4.4 show the results of accuracy of estimation of the generic parameter in terms of MAE and MAPE respectively.

Table 4.3 - Mean Absolute Errors in the Inverse Burr parameter estimation procedures.

γ	Mean Absolute Errors in the parameter estimation procedures								
	MLE			ME			QE		
	ζ	ρ	γ	ζ	ρ	γ	ζ	ρ	γ
1	1.568	0.095	0.069	1.624	0.107	0.069	2.463	0.104	0.137
1.5	1.387	0.137	0.125	1.394	0.146	0.129	2.046	0.201	0.176
2	1.284	0.112	0.162	1.328	0.112	0.166	1.756	0.281	0.211
2.5	1.338	0.145	0.261	1.269	0.163	0.269	2.099	0.483	0.245
3	1.500	0.146	0.355	1.575	0.139	0.349	2.107	0.789	0.354

Table 4.4 - Mean Absolute percentage Errors in the Inverse Burr parameter estimation procedures.

γ	Mean Absolute Percentage Errors in the parameter estimation procedures								
	MLE			ME			QE		
	ζ	ρ	γ	ζ	ρ	γ	ζ	ρ	γ
1	2.746	0.244	6.835	3.110	0.282	6.885	4.638	0.456	13.732
1.5	2.770	0.362	8.332	2.795	0.378	8.636	4.085	0.466	11.738
2	2.186	0.282	8.096	2.212	0.275	8.297	4.060	0.719	10.546
2.5	2.296	0.414	10.436	2.217	0.429	10.759	3.586	0.900	9.813
3	2.526	0.420	11.840	2.634	0.391	11.656	3.871	0.969	11.796

From the analysis of Tables 4.3 and 4.4, classical estimation procedures (i.e., MLE and ME) appear to perform in similar way, since the corresponding values of indices are very similar for each of the three IB parameters. In particular, percentage indices allow a direct comparison between estimation procedures for different values of parameters. As a significant example, values of the MAPE for γ estimated through MLE and ME appear to increase as the value of the parameter γ increases, and therefore the estimation appear to be less reliable for greater values of γ . However, values of MAPE for γ estimated through QE do not show the same behavior; QE indices are slightly greater on average than the corresponding MLE and ME indices. Nevertheless, QE procedure is way much simpler to be solved, and the differences in terms of parameter estimation errors are not very remarkable, especially when parameter γ is considered known; this justifies its application in some non-unusual situations. On average,

parameter ρ is estimated with the lowest values of percentage error with all the considered procedures.

4.5.3. Mixture Inverse Burr – Inverse Weibull distribution

The M-IB-IW distribution was tested in order to prove its usefulness for EWS characterization. A comparison with the GU, IW, IB, and GEV distributions is performed in terms of “classical” CS GOF test, and also in terms of the Determination Coefficient (DC) R_{DC}^2 and Adjusted Determination Coefficient (ADC) R_{ADC}^2 [222], using real EWS data¹⁷. The ADC is specifically selected in order to assess if the addition of an increased number of parameters in the mixture distribution is justified by an according overall increased GOF, assigning a sort of penalization to distributions that have a high number of parameters.

The parameters of GU, IW, IB, and GEV distributions are estimated through the MLE procedure on several actual public datasets described in sub-Section 4.5.3.1; the parameters of the proposed M-IB-IW distribution are estimated through both MLE and EM procedures.

The results of the comparison with benchmark models are shown in sub-Section 4.5.3.2.

Eventually, an error analysis is performed in sub-Section 4.5.3.3 on synthetic EWS samples drawn from credible M-IB-IW distributions with known parameters. The EM procedure was applied to these samples, and the estimated parameters are compared to the actual, known values. The MAPEs are displayed to provide a quantitative evaluation of the quality the proposed procedure.

A sensitivity analysis was also run to individuate the impact of the variation of each parameter of the M-IB-IW distribution on the resulting GOF. In particular, keeping the values of five parameters constant, the other parameter is varied in the range of the corresponding estimated confidence interval, and the CS test statistics CS_{TS} , DC, and ADC are provided to individuate the most-influencing parameters in terms of GOF. Results are not shown in this sub-Section, only for sake of conciseness.

¹⁷ Definitions of DC and ADC are in the Appendix.

4.5.3.1. *Data characteristics*

Different EWS are considered in several regions and for different intervals of time, in order to test the proposed M-IB-IW distribution and its EM estimation procedure in different conditions. Values of wind speed are initially collected and post-processed, in order to avoid missing and bad data. The extreme events in the PMA framework are stored in order to form eight EWS datasets that were to be fitted through the GEV, GU, IW, IB, and M-IB-IW distributions. Extreme values should accurately be selected in order to exclude outliers; as in the previous analyses of sub-Section 4.5.2.1, the Tukey's test was run to identify values that could be outliers, in order to singularly investigate them. However, in all of the eight datasets, no suspect outliers were identified, and therefore the EWS datasets were directly build.

Two EWS datasets (D3 and D7) are the same as the ones described in sub-Section 4.5.2.1, for comparative purposes.

The third and fourth EWS datasets (D8 and D9) consisted respectively of the 36 monthly maximum values and of the 156 weekly maximum values of wind speed registered in Boulder, Colorado, USA (latitude 39°54' north, longitude 105°14' west) by the NREL National Wind Technology Center M2 Tower [218], from 1 January 2012 to 31 December 2014.

The fifth EWS dataset (D10) consisted of the 156 weekly maximum values of wind speed collected in Milford, Utah, USA (latitude 38°41' north, longitude 113°03' west) by the NREL Solar Resource & Meteorological Assessment from 1 August 2010 to 31 July 2013 [219].

The sixth and seventh EWS datasets (D11 and D12, respectively) consisted of the 52 weekly maximum values of wind speed provided in the context of the Gefcom2014 [12] for the site no. 1 and for the site no. 10, respectively, from 1 January 2012 to 31 December 2012.

The eighth EWS dataset (D13) consisted of the 52 weekly maximum values of wind speed collected in Forenza, Italy (latitude 40°52' north, longitude 15°51' east) by the Gestore dei Servizi Energetici (the Italian state-owned company that promotes and supports renewable energies) from 1 January 2002 to 31 December 2002 [180].

4.5.3.2. *Assessment of Generalized Extreme Value, Gumbel, Inverse Weibull, Inverse Burr, and mixture Inverse Burr - Inverse Weibull distributions on measured EWS data*

Parameters of GEV, GU, IB, IW and M-IB-IW distributions estimated through MLE are presented in Table 4.5 for each dataset described in the previous sub-Section. M-IB-IW distributions are also fitted through the proposed EM procedure; the initial points provided to both the M-IB-IW MLE and M-IB-IW EM procedures were calculated from initial MLEs of IB and IW constituent distributions, to provide a fair comparison of the results.

Table 4.5 – Values of Generalized Extreme Value, Gumbel, Inverse Weibull, Inverse Burr, and mixture Inverse Burr – Inverse Weibull parameters.

Distribution parameter	Dataset								
	D3	D7	D8	D9	D10	D11	D12	D13	
GEV	ι	2.65	3.83	3.97	4.57	3.45	2.01	1.99	4.94
	ν	19.11	15.59	23.03	18.53	15.27	11.07	11.15	16.06
	κ	-0.15	-0.15	0.05	-0.04	-0.20	-0.16	-0.17	-0.23
GU	χ	18.90	15.29	23.14	18.44	14.89	10.90	10.97	15.47
	o	2.60	3.70	4.06	4.52	3.44	1.97	1.91	4.56
	ρ	20.94	17.99	16.21	18.34	18.68	11.85	12.00	23.29
IB	ζ	15.03	8.02	6.29	5.92	12.04	10.04	9.90	11.58
	γ	0.70	0.74	9.27	1.51	0.45	0.95	0.93	0.26
IW	ν	0.05	0.07	0.04	0.06	0.07	0.09	0.09	0.07
	δ	7.17	4.04	5.93	4.09	3.91	5.26	5.73	3.39
	ρ	19.77	17.11	18.17	16.28	18.27	13.02	11.64	22.62
M-IB-IW MLE	ζ	23.27	8.35	8.20	5.64	11.61	11.04	10.51	14.16
	γ	2.92	1.36	22.63	2.72	0.57	0.62	1.66	0.79
	ν	0.06	0.09	0.05	0.08	0.10	0.09	0.11	0.08
EM	δ	7.25	8.33	9.10	27.9	5.28	14.26	11.38	4.59
	ω_M	0.50	0.82	0.51	0.96	0.94	0.76	0.81	0.45
	ρ	14.70	14.18	19.77	18.52	19.45	12.38	11.58	22.72
M-IB-IW EM	ζ	20.46	7.59	8.16	6.37	13.54	10.07	10.52	14.28
	γ	0.98	4.96	11.19	2.07	0.36	0.70	1.81	0.75
	ν	0.06	0.09	0.05	0.07	0.07	0.09	0.11	0.08
EM	δ	7.34	6.55	8.66	5.32	10.55	8.91	10.75	4.61
	ω_M	0.50	0.73	0.50	0.75	0.88	0.61	0.78	0.45

Examples of empirical CDFs and fitted CDFs are shown in Figs. 4.4-4.6 only for dataset D13, for sake of conciseness.

The visual comparison for dataset D13 suggests that the proposed M-IB-IW distribution fits particularly well the data D13, both through the MLE and the EM procedures. Indeed, the estimated parameters with both procedures are very similar, as shown in the last column of Table 4.5. In general, M-IB-IW parameters estimated through the MLE and EM procedures are similar, but differences are not always unintelligible (e.g., the shape parameter γ for dataset D8 estimated through the MLE procedure is twice the one estimated through the EM procedure); this however is not always easy to be identified graphically. Thus, a quantitative tool is necessary to perform an objective comparison in order to select the best-fitting CDF; also the MLE needs for a further investigation on the corresponding confidence intervals, that can provide hints on the trust put in each of the fitting parameters.

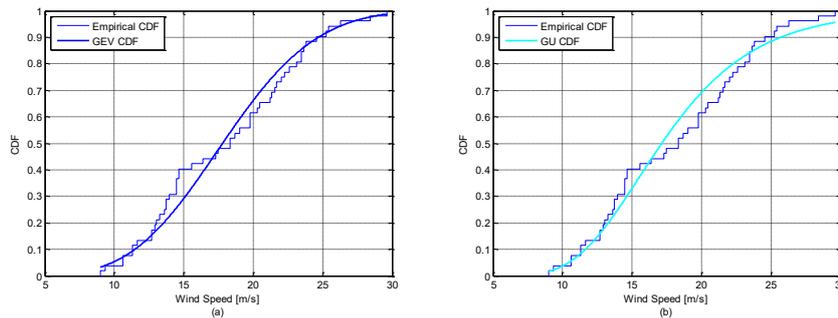


Figure 4.4 – Generalized Extreme Value (a) and Gumbel (b) fitting of dataset D13 through MLE procedure.

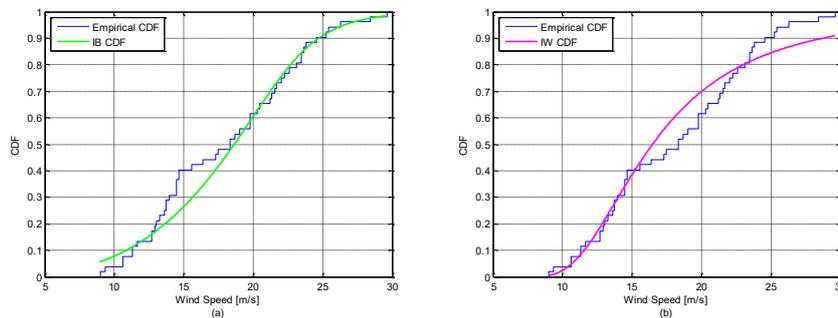


Figure 4.5 - Inverse Burr (a) and Inverse Weibull (b) fitting of dataset D13 through MLE procedure.

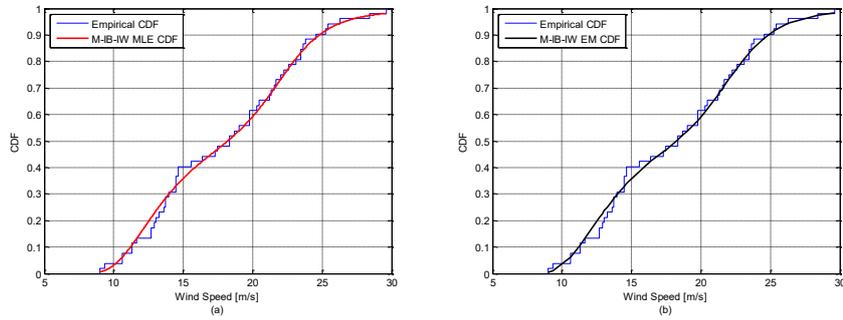


Figure 4.6 – Mixture Inverse Burr – Inverse Weibull fitting of dataset D13 through (a) MLE procedure, and (b) EM procedure.

The results of the CS test at 5% of significance level, the corresponding values of CS test statistics CS_{TS} , the DC and ADC are shown in Tables 4.6-4.9 for datasets D3, D7-D13, allowing for an immediate comparison between fitting distributions.

Table 4.6 - Chi-square test statistics and values of the Determination Coefficients and Adjusted Determination Coefficients for datasets D3 and D7. Bold italic values denote failed tests, while underlined values correspond to the best-fitting distributions for each dataset.

Distribution	Dataset D3			Dataset D7		
	CS_{TS}	Determination coefficients		CS_{TS}	Determination coefficients	
		R_{DC}^2	R_{DAC}^2		R_{DC}^2	R_{DAC}^2
GEV	9.1152	0.9728	0.9712	10.0804	0.9903	0.9899
GU	11.0339	0.9579	0.9567	8.6918	0.9838	0.9835
IB	6.1941	0.9861	0.9853	8.7313	0.9926	0.9923
IW	<i>14.6073</i>	0.9376	0.9358	<i>13.5971</i>	0.9558	0.9549
M-IB-IW MLE	4.7830	<u>0.9953</u>	<u>0.9945</u>	7.1561	0.9943	0.9937
M-IB-IW EM	4.2512	0.9945	0.9936	7.0241	<u>0.9952</u>	<u>0.9947</u>

Table 4.7 – Chi-square test statistics and values of the Determination Coefficients and Adjusted Determination Coefficients for datasets D8 and D9. Bold italic values denote failed tests, while underlined values correspond to the best-fitting distributions for each dataset.

Distribution	Dataset D8			Dataset D9		
	CS_{TS}	Determination coefficients		CS_{TS}	Determination coefficients	
		R_{DC}^2	R_{ADC}^2		R_{DC}^2	R_{ADC}^2
GEV	2.4495	0.9869	0.9861	6.7129	0.9974	0.9974
GU	2.4310	0.9858	0.9854	7.0262	<u>0.9977</u>	<u>0.9977</u>
IB	2.5962	0.9864	0.9856	9.1318	0.9966	0.9966
IW	2.3177	0.9862	0.9858	17.6829	0.9871	0.9870
M-IB-IW	2.0862	0.9886	0.9867	10.1191	0.9962	0.9961
MLE						
M-IB-IW	2.1130	<u>0.9890</u>	<u>0.9872</u>	9.4319	0.9968	0.9967
EM						

Table 4.8 - Chi-square test statistics and values of the Determination Coefficients and Adjusted Determination Coefficients for datasets D10 and D11. Bold italic values denote failed tests, while underlined values correspond to the best-fitting distributions for each dataset.

Distribution	Dataset D10			Dataset D11		
	CS_{TS}	Determination coefficients		CS_{TS}	Determination coefficients	
		R_{DC}^2	R_{ADC}^2		R_{DC}^2	R_{ADC}^2
GEV	11.6086	0.9922	0.9921	6.9502	0.9894	0.9890
GU	28.4556	0.9710	0.9708	6.3639	0.9850	0.9847
IB	8.8935	0.9957	0.9956	5.1538	0.9936	0.9933
IW	59.3152	0.9213	0.9208	10.3665	0.9594	0.9586
M-IB-IW						
MLE	7.6053	0.9956	0.9955	4.8772	0.9942	0.9936
M-IB-IW	8.7223	<u>0.9968</u>	<u>0.9967</u>	4.9595	<u>0.9945</u>	<u>0.9939</u>
EM						

Table 4.9 - Chi-square test statistics and values of the Determination Coefficients and Adjusted Determination Coefficients for datasets D12 and D13. Bold italic values denote failed tests, while underlined values correspond to the best-fitting distributions for each dataset.

Distribution	Dataset D12			Dataset D13		
	CS_{TS}	Determination coefficients		CS_{TS}	Determination coefficients	
		R_{DC}^2	R_{ADC}^2		R_{DC}^2	R_{ADC}^2
GEV	3.0901	0.9936	0.9933	13.4349	0.9705	0.9693
GU	3.3251	0.9869	0.9866	11.8078	0.9691	0.9685
IB	2.1568	0.9937	0.9934	<i>17.7939</i>	0.9639	0.9624
IW	5.8916	0.9698	0.9692	10.1175	0.9610	0.9602
M-IB-IW	1.1509	<u>0.9958</u>	<u>0.9953</u>	5.6044	<u>0.9933</u>	<u>0.9926</u>
MLE						
M-IB-IW	1.2032	<u>0.9958</u>	<u>0.9953</u>	5.6021	<u>0.9933</u>	<u>0.9926</u>
EM						

The CS test is performed in MATLAB environment. The IW distribution failed to pass the CS test in 4 on 8 cases; the IB and the GU failed to pass the CS test in only one case; the GEV and the M-IB-IW always passed the test. We recall here that the rejection or acceptance of the null hypothesis depends on the number of the parameters of the fitting distributions; distributions with a large number of parameters are “penalized” with respect to distributions with fewer parameters.

GOF can be evaluated also in terms of DC and ADC. The highest value of both DC and ADC is reached through the proposed M-IB-IW distribution in 7 on 8 cases. However, the same applies also for the ADC, thus confirming that the augmented number of parameters of the M-IB-IW distribution is not a disadvantage of the model. In 5 on 8 cases, the M-IB-IW EM performed better than the M-IB-IW MLE; in 2 on 8 cases, they showed the same value of DC (and ADC); in only one case, the M-IB-IW MLE performed better than the M-IB-IW EM distribution in terms of DC (and ADC).

As a further example, in Table 4.10 one can find the results of the M-IB-IW MLE and M-IB-IW EM fitting for dataset D8 when the same, arbitrary initial point (i.e., not provided through initial MLEs of constituent IB and IW distributions) was chosen in both procedures. From the comparison between the results shown in Tables 4.7 and

4.10, it is clear that the choice of initial point really matters in the MLE; indeed, the performance drastically drops when an arbitrary point is selected. Note that the MLE procedure in these conditions exceeded the maximum number of iterations (i.e., 10^6), thus not providing any assurance on the convergence of the maximum likelihood maximization. This instead does not apply for the EM procedure, since convergence was reached in only 30 expectation and maximization steps; also, the differences with the results obtained assigning the IB and IW MLEs as initial points (ref. Table 4.7) are almost unintelligible.

Table 4.10 – Chi-square test statistics and values of the Determination Coefficients and Adjusted Determination Coefficients for datasets D8 when random initial points are chosen. Bold italic values denote failed tests, while underlined values correspond to the best-fitting distributions for each dataset.

Distribution	Dataset D8		
	CS_{TS}	Determination coefficients	
		R_{DC}^2	R_{ADC}^2
M-IB-IW MLE	4.1508	0.5207	0.4408
M-IB-IW EM	2.0441	<u>0.9900</u>	<u>0.9883</u>

4.5.3.3. Assessment of mixture Inverse Burr - Inverse Weibull distribution on synthetic EWS data

The error analysis was performed on synthetic EWS datasets, each constituted by 2000 samples drawn from several M-IB-IW distributions with known parameters through the inversion method [214]. Different scenarios were developed by setting the theoretical medians of the constituting IB and IW distributions to $38 \text{ m}^2/\text{s}$ and $20 \text{ m}^2/\text{s}$, respectively, and by setting the weight coefficient ω_M to 0.1, 0.2, ..., 0.9; each scenario was analysed for five values of parameter γ (in particular, $\gamma = 1, 1.5, 2, 2.5, 3$) and twenty independent different values of both parameters ζ and ν , selected in credible ranges. A total number of 900 scenarios was considered in the whole error analysis. The ratios between theoretical standard deviations and theoretical means of the resulting distributions was kept under 20%. The proposed EM procedure was performed on these different-scenarios synthetic EWS datasets. The estimated value of each parameter in each scenario

was compared to the actual value of the parameter, and the corresponding MAPEs are shown in Table 4.11 in order to quantify the GOF for each weight-coefficient scenario. As suggested by these numerical values, the EM procedure appears to be reliable for the parameter estimation of M-IB-IW distributions in all of the considered weight-coefficient scenarios, as MAPE is always smaller than 5% for each parameter. On average, the lowest MAPEs occurred in the scenario $\omega_M = 0.4$.

Table 4.11 – Results of the error analysis in terms of MAPEs, averaged for each parameter of the mixture Inverse Burr – Inverse Weibull distribution estimated through the EM procedure, and for each value of the weight ω_M .

Weight coefficient scenario	ρ MAPE [%]	ζ MAPE [%]	γ MAPE [%]	ν MAPE [%]	δ MAPE [%]	ω_M MAPE [%]
$\omega_M = 0.1$	4.42	0.96	3.34	1.98	1.06	2.72
$\omega_M = 0.2$	4.10	0.89	2.24	0.54	4.51	0.51
$\omega_M = 0.3$	4.97	0.90	0.97	2.66	2.93	3.37
$\omega_M = 0.4$	1.06	0.57	1.28	0.51	0.74	0.65
$\omega_M = 0.5$	0.97	0.69	0.72	1.69	3.14	3.97
$\omega_M = 0.6$	1.59	3.57	1.21	1.03	1.06	2.09
$\omega_M = 0.7$	0.88	1.25	1.36	1.68	0.72	3.27
$\omega_M = 0.8$	0.70	4.62	3.01	4.10	1.21	0.72
$\omega_M = 0.9$	0.92	1.73	1.03	0.83	0.70	1.02

4.6. CONCLUSIONS

EWS play a key role in power system planning and operation, due the increased penetration of wind power plants and due the EWS effect on sensible structures, such as wind towers and overhead lines. A correct characterization and estimation of the EWS potential is then mandatory in order to make decisions that allow power systems to be operated in reliable, efficient way.

Two contributions to the state of the art on EWS probabilistic modeling were presented in this Chapter.

The first contribution was the proposal of the IB distribution as a rational and efficient alternative to more popular models for EWS, such as the GU and the IW distributions. The problem of the estimation of IB parameters was discussed by applying different

procedures, such as classical MLE and ME, to several real EWS data; moreover, a new proposal, the QE which is sometimes easier to be used, was compared to traditional parameter estimation procedures. The estimators were tested on both datasets of real EWS measured during different years and at different locations, and on synthetic samples of EWS extracted from known distributions. The results of a large set of numerical simulations confirmed the usefulness of the proposed model in terms of both accuracy and efficiency.

The second contribution was the proposal a new finite M-IB-IW distribution for EWS characterization, and the development of its EM procedure to cope with convergence problems of the classical MLE, that often occur when dealing with mixture distributions.

Several real datasets of EWS have been considered in order to compare the proposed model to the existing models in different conditions; results showed that the proposed M-IB-IW distribution is a versatile tool that allows for a suitable characterization in the majority of cases. Also, the EM procedure performed on average better than the classical MLE for M-IB-IW parameters, both on real and synthetic EWS datasets, without suffering from typical convergence issues.

In all of the performed analyses, the M-IB-IW proved to outperform the IB distribution in terms of GOF tests.

CONCLUSIONS

The transformation of power systems into smart, multi-device structures inevitably implies the need for advanced forecasting systems, to take into account the random nature of influencing variables such as wind speed, solar irradiance, and electric loads.

This Thesis provided four contributions to the state of the art of forecasting in power systems.

The first and second contributions were a *Bayesian-based probabilistic method* and a *probabilistic competitive ensemble method* to forecast photovoltaic generation in short-term scenarios.

With reference to the proposed *Bayesian-based probabilistic method*, the main innovative contributions were:

- (i) the use of two models to respectively link the hourly active power generated by photovoltaic systems to the hourly clearness index and to hourly solar irradiance. Note that two different distributions were selected to model the probabilistic behaviour of such variables, namely a modified Gamma distribution and a Beta distribution; some parameters of these distributions were estimated in a Bayesian framework;
- (ii) the development of new time series models with exogenous inputs (cloud cover, ambient temperature, pressure, and relative humidity) to predict the future mean value of the input random variable, i.e., solar irradiance or hourly clearness index;
- (iii) the critical comparison of two different hybrid Bayesian-based approaches for the photovoltaic power forecasting, one considering the clearness index as input and the other considering solar irradiance as input.

With reference to the proposed *probabilistic competitive ensemble method*, the main original contributions were:

- (i) the proposal of a linear pooling tool to combine probabilistic forecasts coming from different base predictors, in order to improve the overall quality of the probabilistic forecasts;
- (ii) the development of a multi-objective optimization system to overcome the well-known problems resulting from the linear pooling of forecasts. Indeed, linear pooling proved to produce forecasts that were over-dispersed, even if base predictors were normally dispersed;
- (iii) the comparison of the results obtained through the multi-objective optimization system with those obtained through the classical minimization of a proper score in the training step.

The third contribution was a *deterministic industrial load forecasting method* based on regression models.

The main original contributions on this topic were:

- (i) the development of methods that were able to forecast industrial loads at both aggregate and single-load levels, considering inputs that could be very different from those usually selected in aggregated national, regional or sub-station load forecasting;
- (ii) the comparison of multiple linear regression and support vector regression models built through two different model selection techniques, one based on classical 10-fold cross-validation and the other based on lasso analysis;
- (iii) the applications of such models not only to active power forecasting, but also to reactive power forecasting.

The fourth contribution dealt with the proposals of an *Inverse Burr distribution* and a *mixture Inverse Burr – Inverse Weibull distribution* for modeling extreme values of wind speed; both distributions were presented with appropriate parameter estimation procedures.

With reference to the proposed *Inverse Burr distribution for modeling extreme values of wind speed*, the main original contributions were:

- (i) the application of the Inverse Burr distribution to model extreme values of wind speed. Its comparison to the most used distributions (i.e., Inverse Weibull and Gumbel distributions) in this field appeared to validate its usefulness in the majority of the analyzed cases;

- (ii) the development of a parameter estimation procedure based on the quantile estimation, that in particular cases resulted in the algebraic solution of an equation, allowing to avoid convergence problems.

With reference to the proposed *mixture Inverse Burr – Inverse Weibull distribution for modeling extreme values of wind speed*, the main original contributions were:

- (i) the application of the mixture Inverse Burr – Inverse Weibull distribution to model extreme values of wind speed. The additional complexity added by the increased number of the parameters seemed to be fully justified by the increased versatility and better fitting;
- (ii) the development of an expectation-maximization parameter estimation procedure, that is specifically aimed to overcome problems resulting in the maximum likelihood and moment estimations of mixture distributions.

The results of comprehensive numerical applications, always based on real data, proved the validity of all of the proposals with respect to benchmarks that are commonly used in relevant literature.

The researches that led to the proposals presented in this Thesis will be further pursued and enhanced.

Future works on photovoltaic power forecasting will focus on the improvement of the *probabilistic competitive ensemble method*, e.g., by merging more probabilistic base predictors and considering new techniques to combine probabilistic forecasts.

The results of the *deterministic industrial load forecasting method* suggested to explore different techniques for the model selection in industrial load forecasting scenarios. Also, load forecasting methods will be improved in their probabilistic framework, in order to be used as parts of sophisticated procedures aimed at forecasting the dynamic rating of electric components, such as lines and transformers, in the context of an optimal operation of a smart grid.

APPENDIX

The error indices and tools used in this Thesis for the assessment of forecasts in both deterministic and probabilistic frameworks are shown in this Appendix. In particular, deterministic error indices for the assessment of the quality of forecasts in a spot-value framework are shown in sub-Section A.1. Probabilistic error indices and graphical tools for the assessment of the quality of forecasts in a probabilistic framework are shown in sub-Section A.2. Indices for the assessment of the GOF of PDFs on available independent samples are shown in sub-Section A.3.

A.1. DETERMINISTIC INDICES FOR THE ASSESSMENT OF THE QUALITY OF FORECASTS

The quality of deterministic forecasts is immediately assessed by comparing the single (spot) value provided by the deterministic method to the realization y_h^* of the generic random variable y at the desired forecast time horizon h . The spot-value framework can be easily extended also to probabilistic forecasts; indeed, the quality of probabilistic forecasts can be assessed by extracting a single value (e.g., the mean, the median, or a specific quantile) from the predictive distribution given by the probabilistic method, and comparing it to the realization y_h^* of the generic random variable y at the desired forecast time horizon h . Therefore, the symbol \hat{y}_h is here treated as both a deterministic forecast, or a spot value extracted from a probabilistic forecast for the desired forecast time horizon h .

Indices for this type of forecast verification are well known and have been extensively used in relevant literature. The MAE, the RMSE, the NMAE, the NRMSE, the Mean Absolute Percentage Error (MAPE), and the Root Mean Squared Percentage Error (RMSPE) are considered in this thesis. These indices are respectively defined as:

$$MAE = \frac{1}{N_{tf}} \sum_{h=1}^{N_{tf}} |y_h^* - \hat{y}_h|, \quad (A1)$$

$$RMSE = \sqrt{\frac{1}{N_{tf}} \sum_{h=1}^{N_{tf}} (y_h^* - \hat{y}_h)^2}, \quad (A2)$$

$$NMAE = \frac{100}{N_{tf}} \sum_{h=1}^{N_{tf}} \frac{|y_h^* - \hat{y}_h|}{y_{ref}}, \quad (A3)$$

$$NRMSE = 100 \sqrt{\frac{1}{N_{tf}} \sum_{h=1}^{N_{tf}} \left(\frac{y_h^* - \hat{y}_h}{y_{ref}} \right)^2}, \quad (A4)$$

$$MAPE = \frac{100}{N_{tf}} \sum_{h=1}^{N_{tf}} \frac{|y_h^* - \hat{y}_h|}{y_h^*}, \quad (A5)$$

$$RMSPE = 100 \sqrt{\frac{1}{N_{tf}} \sum_{h=1}^{N_{tf}} \left(\frac{y_h^* - \hat{y}_h}{y_h^*} \right)^2}, \quad (A6)$$

where y_{ref} is a reference value for the normalization (e.g., the rated power of a generator or of a load, or the average value of past observations of the variable y), and N_{tf} is the total number of forecasts.

MAE and RMSE provide indications on the absolute deviation from the actual value of the variable; they are unadvisable when comparing results of forecasts performed for variables of different order of magnitude (e.g., two generators with very different rated powers). Percentage errors (MAPE and RMSPE) overcome in part this problem; however, they suffer the presence of singularities in correspondence of quasi-zero actual values. In some applications, a threshold is fixed in order to discard contributions to the percentage errors in correspondence of these quasi-zero actual values.

Normalized errors (NMAE and NRMSE) are advisable when comparing results of forecasts performed for variables of different order of magnitude, and do not suffer from singularities. However, the

reference value y_{ref} must be accurately chosen, in order to avoid uninformative too-low or too-high error values.

The abovementioned indices are not able to catch the economic “value” of forecasting errors. For example, electrical energy is sold and, thus, has an economic value; moreover, the price of energy varies with time, i.e., with the hours of the day (for example in the day-ahead market). Therefore, when producing power forecasts, the same error can cause different economic consequences depending on the hour of the day. Novel cost-based indices that take into account not only the magnitude of the errors, but also the economic value of these errors, are proposed in [99] to deal with this aspect, although further researches are still to be pursued.

A.2. PROBABILISTIC INDICES FOR THE ASSESSMENT OF THE QUALITY OF FORECASTS

Two major requirements must be met simultaneously by all probabilistic forecasts, i.e., the forecasts must be sharp and also calibrated (or equivalently reliable) [223,224]. Note that sharpness and reliability are not distinct one each other, as one property significantly influences the other, and vice versa. After a brief introduction on these features of probabilistic forecasts (sub-Sections A.2.1 and A.2.2, respectively), indices and tools aimed to assess sharpness and reliability are presented in sub-Sections A.2.3, A.2.4, and A.2.5.

A.2.1. Sharpness

Sharpness is a property of the forecast alone, as the realization of the random variable is not involved in its definition. Sharpness, in the case of forecasts for a real-value variable, can be easily assessed in terms of the associated prediction intervals. The narrower the intervals, the better is the forecast (if the corresponding coverage is however coherent, as shown in [224]). Usually 50%, 90%, 95% and 99% prediction intervals are considered for probabilistic forecasting. Prediction intervals can be easily extracted from a forecasted predictive distribution.

A.2.2. Reliability

Reliability is a property of the probabilistic forecast and of the realization. It involves the correspondence between estimated coverages and actual coverages.

Indeed, let's suppose that a 50% prediction interval is provided for a random variable; the forecast is therefore considered reliable if the observation of the random variable lies in that interval with probability 0.5 for the given time horizon.

The same property can be defined also for predictive quantiles; e.g., if the 0.5-quantile (median) is predicted for a given horizon time, the realizations should be equal or smaller than the 0.5-quantile in 50% of cases [149,225,226].

A.2.3. Reliability diagrams

Reliability diagrams are very effective tools to evaluate the reliability of a probabilistic method [225,227,228]; they show the estimated coverage versus the nominal one, for various nominal coverage values (usually from 0.05 to 0.95, with a 0.05 step, or from 0.1 to 0.9 with a 0.1 step).

The estimated coverages can be found from a predictive distribution in a very intuitive manner. Let $\hat{y}_h^{(\alpha_q)}$ be the forecasted α_q -quantile extracted from the forecasted distribution of the generic random variable y at the desired time horizon h . The indicator $\hat{\Lambda}_h^{(\alpha_q)}$ is defined from the comparison between the actual value y_h^* and the forecasted quantile $\hat{y}_h^{(\alpha_q)}$, as follows:

$$\hat{\Lambda}_h^{(\alpha_q)} = \begin{cases} 1, & \text{if } y_h^* \leq \hat{y}_h^{(\alpha_q)} \\ 0, & \text{if } y_h^* > \hat{y}_h^{(\alpha_q)} \end{cases}, \quad (\text{A7})$$

and, consequently, the estimation $\hat{\alpha}_q$ of the actual coverage α_q based on a set of N_{tf} forecasts is:

$$\hat{\alpha}_q = \frac{1}{N_{tf}} \sum_{h=1}^{N_{tf}} \hat{\Lambda}_h^{(\alpha_q)}. \quad (\text{A8})$$

Obviously, the probabilistic forecasting method is considered reliable if the estimated coverages do not significantly differ from the nominal ones. A necessary condition for the probabilistic calibration is the normal dispersion of forecasts, and this results in a reliability curve that is close to the 45-degree diagonal line (representing the ideal reliability). Instead, over-dispersed forecasts (usually due to lack of sharpness) result in an inverse-S-shaped reliability curve, while under-dispersed forecasts (usually due to too much sharpness) result in a S-shaped reliability curve. Biased forecasts are easily recognized, as the corresponding reliability diagrams strongly differ from perfect curve. Fig. A.1 shows examples of reliability diagrams for reliable, over-dispersed, under-dispersed and biased forecasts.

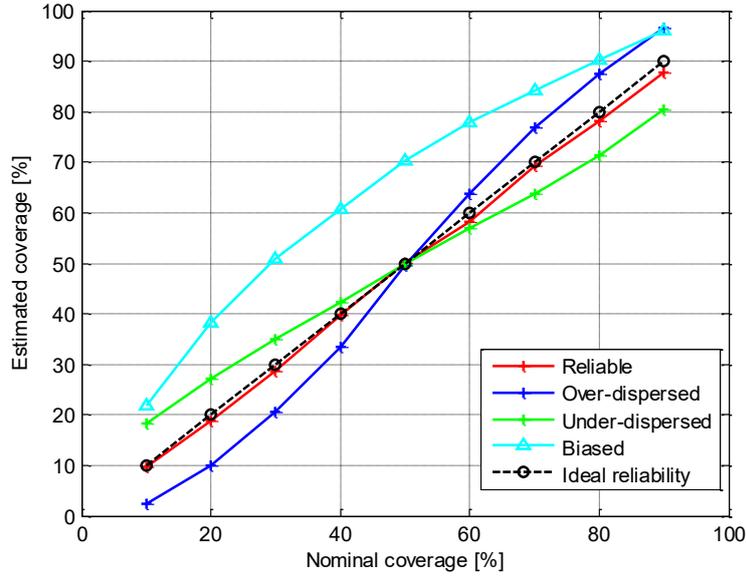


Figure A.1 – Examples of reliability diagrams.

The maximum deviation from perfect reliability ψ is straightforwardly defined as the maximum error between estimated coverages and nominal coverages; i.e.:

$$\psi = \max \left\{ \left| \alpha_{q_1} - \hat{\alpha}_{q_1} \right|, \dots, \left| \alpha_{q_J} - \hat{\alpha}_{q_J} \right| \right\}, \quad (\text{A9})$$

where J is the total number of considered coverages (e.g., $J = 19$ if coverages are from 0.05 to 0.95, with a 0.05 step, or $J = 9$ if coverages are from 0.1 to 0.9 with a 0.1 step).

A.2.4. Probability integral transform histograms

PIT histograms [223,229] can also be used to empirically check the calibration of forecasts. In these histograms the PIT values¹⁸ are plotted: for a probabilistically calibrated forecast, the PIT histogram is statistically uniform. Even if the uniformity of PIT histograms is a necessary, but not sufficient condition for the forecast to be perfect [229], from the behavior of PIT histograms, can be derived useful information; in particular, U-shaped histograms indicate under-dispersed predictive distributions as well as inverse-U shaped histograms suggest that the predictive distributions are over-dispersed. Biased predictive distributions have a very irregular PIT histograms. Fig. A.2 shows examples of PIT histograms for reliable, over-dispersed, under-dispersed and biased forecasts.

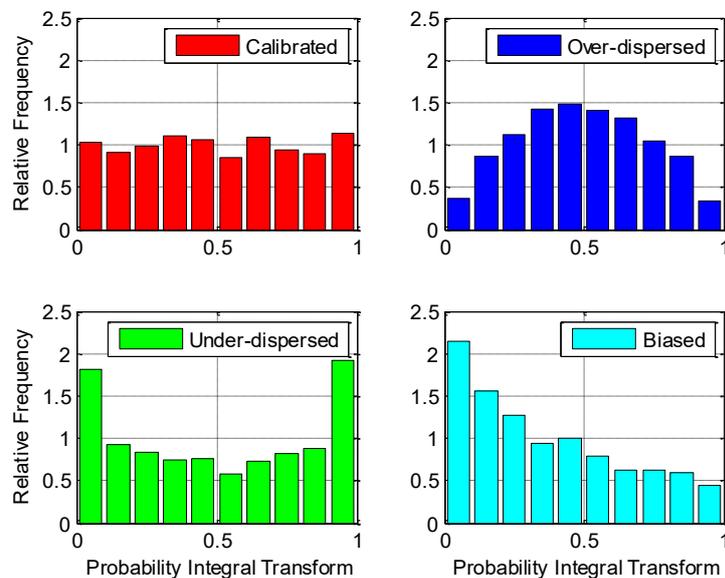


Figure A.2 – Examples of Probability Integral Transform histograms.

¹⁸ In a nutshell, the PIT is the value that the predictive CDF attains at the observation, with suitable adaptations at any points of discontinuity [223].

Anyway, formal tests of the hypothesis that a given forecasting method is probabilistically calibrated are also available, provided that these tests account for complex dependence structures. The reader can refer to specialized literature to deepen this subject [223,225].

A.2.5. Proper scores

Probabilistic forecasts can be assessed numerically through the evaluation of proper scores [230]. Two of the most common and versatile proper scores are the CRPS and the PLF; they simultaneously address both calibration and sharpness [154,230].

In practice, the CRPS compares the predictive distribution with the observation, both in terms of cumulative distribution functions. In particular, the CDF of the observation is a Heaviside function $H(\cdot)$ centered in the observation y_h^* , and the CRPS probabilistically accounts for the error area between predictive and actual CDFs (Fig. A.3).

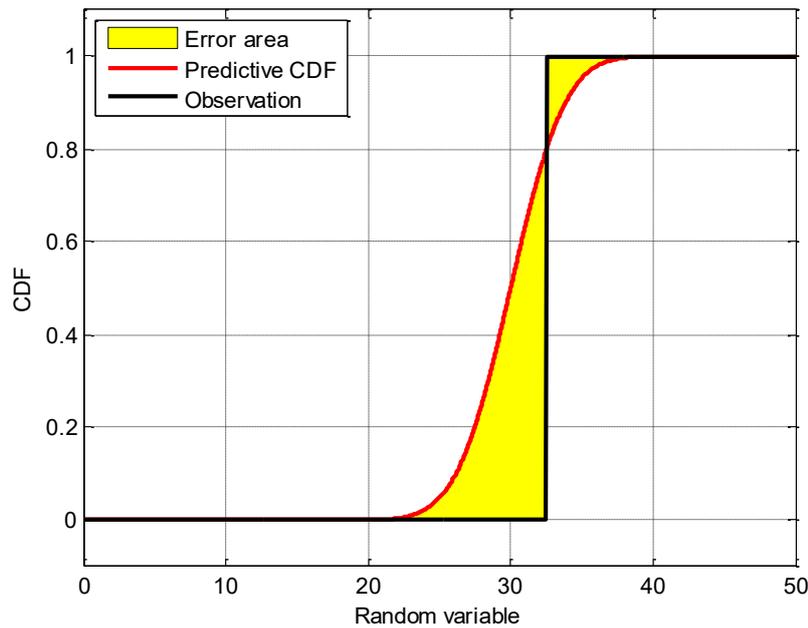


Figure A.3 – Graphical interpretation of the Continuous Ranked Probability Score.

Indeed, let the Heaviside function $H(y_h - y_h^*)$ centered in y_h^* be the CDF of the actual value of the random variable, and $\hat{F}(y_h)$ be the

predictive CDF at the time horizon h ; hourly CRPS can be evaluated as follows:

$$CRPS_h = \int_{-\infty}^{+\infty} [\hat{F}(y_h) - H(y_h - y_h^*)]^2 dy_h . \quad (A10)$$

From the analysis of Eq. (A4), it clearly appears that the CRPS is linked to the total area between the predictive CDF and the Heaviside function. It can be seen that the area (and, consequently, the $CRPS_h$) decreases as the predictive distribution approximates the step function. The calculation of the $CRPS_h$ will result in a value that has the units of the forecast variable. For a total number N_{tf} of forecasts, the average CRPS is:

$$CRPS = \frac{1}{N_{tf}} \sum_{h=1}^{N_{tf}} \int_{-\infty}^{+\infty} [\hat{F}(y_h) - H(y_h - y_h^*)]^2 dy_h , \quad (A11)$$

and it can be interpreted as a probabilistic version of the mean absolute error [230].

Similarly to cost-based deterministic indices, the cost-based extension of the CRPS is developed in [99] to provide also information on the value of probabilistic forecast errors.

The PLF is another widely-used proper score [12,230]. The contribution to the PLF of the α_q -quantile at time h is defined as follows:

$$PLF_h^{(\alpha_q)} = \begin{cases} (1 - \alpha_q) (\hat{y}_h^{(\alpha_q)} - y_h^*), & \text{if } y_h^* < \hat{y}_h^{(\alpha_q)} \\ \alpha_q (y_h^* - \hat{y}_h^{(\alpha_q)}), & \text{if } y_h^* \geq \hat{y}_h^{(\alpha_q)} \end{cases} . \quad (A12)$$

Summing up the PLFs across all of the J considered quantiles and averaging them throughout the total number N_{tf} of forecasts, the PLF of the corresponding probabilistic forecasts is obtained.

A.3. GOODNESS OF FITTING INDICES

The DC and the ADC, used to assess the GOF of PDFs and CDFs on a given dataset $\mathbf{y} = \{y_1, y_2, \dots, y_{N_{tr}^{(GOF)}}\}$ of $N_{tr}^{(GOF)}$ independent samples, are introduced in this sub-Section.

The DC R_{DC}^2 and the ADC R_{ADC}^2 [222] can directly be used for comparison; they are both positively oriented, i.e., the greater they are, the better the fitting.

Initially, the probability space in which the variable of interest y can range must be clustered in N_B bins $\{b_1\}, \{b_2\}, \dots, \{b_B\}$. The indices compare the empirical number of observations $N_{\{b_i\}}$ that lie in the i^{th} bin, with the theoretical frequency $p_{\{b_i\}}$ for the same bin of the hypothesized distribution.

The expressions of R_{DC}^2 and R_{ADC}^2 are, respectively:

$$R_{DC}^2 = 1 - \frac{\sum_{i=1}^{N_B} (N_{\{b_i\}} - p_{\{b_i\}})^2}{\sum_{i=1}^{N_B} (N_{\{b_i\}} - \bar{N}_{\{b\}})^2}, \quad (\text{A13})$$

$$R_{ADC}^2 = 1 - (1 - R_{DC}^2) \frac{N_{tr}^{(GOF)} - 1}{N_{tr}^{(GOF)} - n_{par}}, \quad (\text{A14})$$

where $\bar{N}_{\{b\}}$ is the mean value of $N_{\{b_i\}}$ for $i = 1, 2, \dots, B$, and n_{par} is the number of parameters of the hypothesized distribution. The ADC is indeed introduced to compare the performances of fitting distributions that differ in terms of number of parameters, penalizing the ones that have an increased number of parameters.

BIBLIOGRAPHY

- [1] A. Muñoz, E.F. Sánchez-Úbeda, A. Cruz, and J. Marín, “Short-term forecasting in power systems: a guided tour,” in *Handbook of power systems II*, Springer, 2010.
- [2] T. Hong, “Energy forecasting: Past, present, and future,” *Foresight: The International Journal of Applied Forecasting*, vol. 32, pp. 43-48, 2014.
- [3] A. Troccoli, L. Dubus, and S.E. Haupt, “Weather matters for energy,” Springer, 2014.
- [4] F. Ueckerdt, R. Brecha, and G. Luderer, “Analyzing major challenges of wind and solar variability in power systems,” *Renewable Energy*, vol. 81, pp. 1-10, 2015.
- [5] F. Ribeiro, P. Ferreira, and M. Araújo, “The inclusion of social aspects in power planning,” *Renewable and Sustainable Energy Reviews*, vol. 15, no. 9, pp. 4361-4369, 2011.
- [6] J. Khan, and M.H. Arsalan, “Solar power technologies for sustainable electricity generation—A review,” *Renewable and Sustainable Energy Reviews*, vol. 55, pp. 414-425, 2016.
- [7] A. Michiorri, et al., “Forecasting for dynamic line rating,” *Renewable and Sustainable Energy Reviews*, vol. 52, pp. 1713-1730, 2015.
- [8] D.H. Kim, and S.G. Lee, “Reliability analysis of offshore wind turbine support structures under extreme ocean environmental loads,” *Renewable Energy*, vol. 79, pp. 161-166, 2015.
- [9] A.T. Mathaios Panteli, and P. Mancarella, “Influence of extreme weather and climate change on the resilience of power systems: impacts and possible mitigation strategies,” *Electric Power System Research*, vol. 127, pp. 259–270, 2015.
- [10] I. Abdallah, A. Natarajan, and J.D. Sorensen, “Influence of the control system on wind turbine loads during power production in extreme turbulence: Structural reliability,” *Renewable Energy*, vol. 87, pp. 464-477, 2016.
- [11] T. Hong, P. Pinson, and S. Fan, “Global Energy Forecasting Competition 2012,” *International Journal of Forecasting*, vol. 30, no. 2, pp. 357-363, 2014.
- [12] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R.J. Hyndman, “Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond,” *International Journal of Forecasting*, vol. 32, no. 3, pp. 896-913, 2016.
- [13] IEEE Guide for Loading Mineral-Oil-Immersed Transformers and Step-Voltage Regulators, IEEE Std. C57.91-2011, 2012.
- [14] E.K. Stigka, J.A. Paravantis, and G.K. Mihalakakou, “Social acceptance of renewable energy sources: A review of contingent valuation applications,” *Renewable and Sustainable Energy Reviews*, vol. 32, pp. 100-106, 2014.

- [15] R. Weron, "Electricity price forecasting: A review of the state-of-the-art with a look into the future," *International journal of forecasting*, vol. 30, no. 4, pp. 1030-1081, 2014.
- [16] R. Weron, "Modeling and forecasting electricity loads and prices: A statistical approach," John Wiley & Sons, 2007.
- [17] M.E. Khodayar, and H. Wu, "Demand forecasting in the smart grid paradigm: features and challenges," *The Electricity Journal*, vol. 28, no. 6, pp. 51-62, 2015.
- [18] J. Catalão, "Forecasting Models and Tools for Load and Renewables Generation," in *Smart and sustainable power systems: operations, planning, and economics of insular electricity grids*, CRC Press, 2015.
- [19] M. Negnevitsky, P. Mandal, and A.K. Srivastava, "An overview of forecasting problems and techniques in power systems," *IEEE Power & Energy Society General Meeting*, 2009.
- [20] A. Gomez-Expósito, A.J. Conejo, and C. Canizares, "Electric energy systems: analysis and operation," CRC Press, 2016.
- [21] T. Hong, "Short term electric load forecasting," Ph.D. dissertation, North Carolina State University, USA, 2010.
- [22] H. Jiayi, J. Chuanwen, and X. Rong, "A review on distributed energy resources and MicroGrid," *Renewable and Sustainable Energy Reviews*, vol. 12, no. 9, pp. 2472-2483, 2008.
- [23] J. Xie, T. Hong, and J. Stroud, "Long-Term Retail Energy Forecasting With Consideration of Residential Customer Attrition," *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2245-2252, 2015.
- [24] L. Hernández, C. Baladrón, J.M. Aguiar, B. Carro, A. Sánchez-Esguevillas, and J. Lloret, "Artificial neural networks for short-term load forecasting in microgrids environment," *Energy*, vol. 75, 2014.
- [25] A. Carpinone, M. Giorgio, R. Langella, and A. Testa, "Markov chain modeling for very-short-term wind power forecasting," *Electric Power Systems Research*, vol. 122, pp. 152-158, 2015.
- [26] T. Hong, and P. Wang, "Fuzzy interaction regression for short term load forecasting," *Fuzzy optimization and decision making*, vol. 13, no. 1, pp. 91-103, 2014.
- [27] F. Barbieri, S. Rajakaruna, and A. Ghosh, "Very short-term photovoltaic power forecasting with cloud modeling: A review," *Renewable and Sustainable Energy Reviews*, vol. 75, pp. 242-263, 2017.
- [28] T. Hong, and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *International Journal of Forecasting*, vol. 32, no. 3, 2016.
- [29] T. Jónsson, P. Pinson, and H. Madsen, "On the market impact of wind energy forecasts," *Energy Economics*, vol. 32, no. 2, pp. 313-320, 2010.
- [30] Y. Zhang, J. Wang, and X. Wang, "Review on probabilistic forecasting of wind power generation," *Renewable and Sustainable Energy Reviews*, vol. 32, 2014.
- [31] M. Zugno, T. Jónsson, and P. Pinson, "Trading wind energy on the basis of probabilistic forecasts both of wind generation and of market quantities," *Wind Energy*, vol. 16, no. 6, pp. 909-926, 2013.

- [32] P. Pinson, C. Chevallier, and G.N. Kariniotakis, "Trading wind generation from short-term probabilistic forecasts of wind power," *IEEE Transactions on Power Systems*, vol. 22, no. 3, pp. 1148-1156, 2007.
- [33] A. Bracale, P. Caramia, G. Carpinelli, A.R. Di Fazio, and G. Ferruzzi, "A Bayesian method for short-term probabilistic forecasting of photovoltaic generation in smart grid operation and control," *Energies*, vol. 6, no. 2, pp. 733-747, 2013.
- [34] B. Liu, J. Nowotarski, T. Hong, and R. Weron, "Probabilistic Load Forecasting via Quantile Regression Averaging on Sister Forecasts," *IEEE Transactions on Smart Grid*, vol. 8, no. 2, pp. 730-737, 2017.
- [35] T.S. Nielsen, A. Joensen, and H. Madsen, "A new reference for wind power forecasting," *Wind Energy*, vol. 1, pp. 29-34, 1998.
- [36] D.W. van der Meer, J. Widén, and J. Munkhammar, "Review on probabilistic forecasting of photovoltaic power production and electricity consumption," *Renewable and Sustainable Energy Reviews*, vol. 81, part 1, pp. 1484-1512, 2018.
- [37] L. Suganthi, and A.A. Samuel, "Energy models for demand forecasting-A review," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 2, pp. 1223-1240, 2012.
- [38] H. Hahn, S. Meyer-Nieberg, and S. Pickl, "Electric load forecasting methods: Tools for decision making," *European Journal of Operational Research*, vol. 199, no. 3, pp. 902-907, 2009.
- [39] A. K. Singh, Ibraheem, S. Khatoon, M. Muazzam, and D.K. Chaturvedi, "Load forecasting techniques and methodologies: A review," in *proc. of 2nd International Conference on Power, Control and Embedded Systems*, Allahabad, pp. 1-10, 2012.
- [40] A.S. Ahmad, M.Y. Hassan, M.P. Abdullah, H.A. Rahman, F. Hussin, H. Abdullah, and R. Saidur, "A review on applications of ANN and SVM for building electrical energy consumption forecasting," *Renewable and Sustainable Energy Reviews*, vol. 33, pp. 102-109, 2014.
- [41] M.Q. Raza, and A. Khosravi, "A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings," *Renewable and Sustainable Energy Reviews*, vol. 50, pp. 1352-1372, 2015.
- [42] H.X. Zhao, and F. Magoulès, "A review on the prediction of building energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 6, pp. 3586-3592, 2012.
- [43] A. Fouquier, S. Robert, F. Suard, L. Stéphan, and A. Jay, "State of the art in building modelling and energy performances prediction: A review," *Renewable and Sustainable Energy Reviews*, vol. 23, pp. 272-288, 2013.
- [44] B. Yildiz, J.I. Bilbao, and A.B. Sproul, "A review and analysis of regression and machine learning models on commercial building electricity load forecasting," *Renewable and Sustainable Energy Reviews*, vol. 73, pp. 1104-1122, 2017.
- [45] A. Mellit, and S.A. Kalogirou, "Artificial intelligence techniques for photovoltaic applications: a review," *Progress in Energy and Combustion Science*, vol. 34, pp. 574-632, 2008.

- [46] M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids," *Renewable and Sustainable Energy Reviews*, vol. 27, pp. 65-76, 2013.
- [47] Y. Ren, P.N. Sugathan, and N. Srikanth, "Ensemble methods for wind and solar power forecasting—A state-of-the-art review," *Renewable and Sustainable Energy Reviews*, vol. 50, pp. 82-91, 2015.
- [48] J. Widén, et al., "Variability assessment and forecasting of renewables: a review for solar, wind, wave and tidal resources," *Renewable and Sustainable Energy Reviews*, vol. 44, pp. 356-375, 2015.
- [49] Y. Kashyap, A. Bansal, and A.K. Sao, "Solar radiation forecasting with multiple parameters neural networks," *Renewable and Sustainable Energy Reviews*, vol. 49, pp. 825-835, 2015.
- [50] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F.J. Martinez-de-Pison, and F. Antonanzas-Torres, "Review of photovoltaic power forecasting," *Solar Energy*, vol. 136, pp. 78-111, 2016.
- [51] R.H. Inman, H.T.C. Pedro, and C.F.M. Coimbra, "Solar forecasting methods for renewable energy integration," *Progress in Energy and Combustion Science*, vol. 39, no. 6, pp. 535-576, 2013.
- [52] C. Wan, J. Zhao, Y. Song, Z. Xu, J. Lin, and Z. Hu, "Photovoltaic and solar power forecasting for smart grid energy management," *CSEE Journal of Power and Energy Systems*, vol. 1, no. 4, pp. 38-46, 2015.
- [53] M.Q. Raza, M. Nadarajah, and C. Ekanayake "On recent advances in PV output power forecast," *Solar Energy*, vol. 136, pp. 125-144, 2016.
- [54] C. Voyant, G. Notton, S. Kalogirou, M.L. Nivet, C. Paoli, F. Motte, and A. Foulloy, "Machine learning methods for solar radiation forecasting: A review," *Renewable Energy*, vol. 105, pp. 569-582, 2017.
- [55] M. Lei, L. Shiyang, J. Chuanwen, L. Hongling, and Z. Yan, "A review on the forecasting of wind speed and generated power," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 4, pp. 915-920, 2009.
- [56] S.S. Soman, H. Zareipour, O. Malik, and P. Mandal, "A review of wind power and wind speed forecasting methods with different time horizons," in *proc. of North American Power Symposium (NAPS) 2010*, pp. 1-8, 2010.
- [57] X. Wang, P. Guo, and X. Huang, "A review of wind power forecasting models," *Energy procedia*, vol. 12, pp. 770-778, 2011.
- [58] W.Y. Chang, "A literature review of wind forecasting methods," *Journal of Power and Energy Engineering*, vol. 2, no. 4, 2014.
- [59] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa, "A review on the young history of the wind power short-term prediction," *Renewable and Sustainable Energy Reviews*, vol. 12, no. 6, pp. 1725-1744, 2008.
- [60] A.M. Foley, P.G. Leahy, A. Marvuglia, and E.J. McKeogh, "Current methods and advances in forecasting of wind power generation," *Renewable Energy*, vol. 37, no. 1, pp. 1-8, 2012.
- [61] X. Zhao, S. Wang, and T. Li, "Review of evaluation criteria and main methods of wind power forecasting," *Energy Procedia*, vol. 12, pp. 761-769, 2011.

- [62] A. Tascikaraoglu, and M. Uzunoglu, "A review of combined approaches for prediction of short-term wind speed and power," *Renewable and Sustainable Energy Reviews*, vol. 34, pp. 243-254, 2014.
- [63] I. Colak, S. Sagiroglu, and M. Yesilbudak, "Data mining and wind power prediction: A literature review," *Renewable Energy*, vol. 46, pp. 241-247, 2012.
- [64] D.R. Chandra, M.S. Kumari, and M. Sydulu, "A detailed literature review on wind forecasting," in *proc. of IEEE Power, Energy and Control (ICPEC)*, 2013.
- [65] S.K. Aggarwal, L.M. Saini, and A. Kumar, "Electricity price forecasting in deregulated markets: A review and evaluation," *International Journal of Electrical Power & Energy Systems*, vol. 31, no. 1, pp. 13-22, 2009.
- [66] L. Hu, G. Taylor, H.B. Wan, and M. Irving, "A review of short-term electricity price forecasting techniques in deregulated electricity markets," in *proc. of IEEE Universities Power Engineering Conference (UPEC)*, 2009.
- [67] M. Cerjan, I. Krzelj, M. Vidak, and M. Delimar, "A literature review with statistical analysis of electricity price forecasting methods," in *proc. of IEEE EUROCON 2013*, pp. 756-763, 2013.
- [68] G.A.F. Seber, and A.J. Lee, "Linear regression analysis," John Wiley and Sons, 2012.
- [69] J.O. Rawlings, S.G. Pantula, and D.A. Dickey, "Applied regression analysis: a research tool," Springer, 2001.
- [70] A. Miller, "Subset selection in regression," Chapman & Hall/CRC Press, Boca Raton, USA, 2002.
- [71] G. Box, G.M. Jenkins, G.C. Reinsel, and G.M. Ljung, "Time series analysis: forecasting and control," John Wiley and Sons, 2015.
- [72] S. Makridakis, S.C. Wheelwright, and R.J. Hyndman, "Forecasting: Methods and Applications," John Wiley and Sons, 1998.
- [73] P.J. Brockwell, and R.A. Davis, "Time series: theory and methods," Springer Science & Business Media, 2013.
- [74] S.A. Billings, "Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains," John Wiley & Sons, 2013.
- [75] R.J. Hyndman, and G. Athanasopoulos, "Forecasting: principles and practice," OTexts, 2014.
- [76] D. Gamerman, "Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference," Chapman & Hall, 1997.
- [77] A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin, "Bayesian Data Analysis," Chapman & Hall, 1995.
- [78] J.O. Berger, "Statistical decision theory and Bayesian analysis," Springer, 2013.
- [79] A. Girard, C.E. Rasmussen, J.Q. Candela, and R. Murray-Smith, "Gaussian process priors with uncertain inputs application to multiple-step ahead time series forecasting," *Advances in neural information processing systems*, pp. 545-552, 2003.
- [80] C.E. Rasmussen, and C.K. Williams, "Gaussian processes for machine learning," Cambridge: MIT press, 2006.
- [81] S. Roberts, Osborne, Ebden, "Gaussian processes for time series modeling," *Philosophical Transactions of the Royal Society A*, vol. 371, 2013.

- [82] I. Florescu, "Probability and Stochastic Processes," John Wiley and Sons, 2014.
- [83] G. Modica, and L. Poggiolini, "A First Course in Probability and Markov Chains," John Wiley and Sons, 2012.
- [84] A.C. Harvey, "Forecasting, structural time series models and the Kalman filter," Cambridge university press, 1989.
- [85] L. Zhang, and P.B. Luh, "Neural network-based market clearing price prediction and confidence interval estimation with an improved extended Kalman filter method," IEEE Transactions on Power Systems, vol. 20, no. 1, pp. 59-66, 2005.
- [86] C.N. Ko, and C.M. Lee, "Short-term load forecasting using SVR (support vector regression)-based radial basis function neural network with dual extended Kalman filter," Energy, vol. 49, pp. 413-422, 2013.
- [87] I.N. Da Silva, D. Spatti, R.A. Flauzino, L.H.B. Liboni, and S.F. dos Reis Alves, "Artificial Neural Networks: A Practical Course," Springer, 2016.
- [88] A.J. Smola, and B. Schölkopf, "A tutorial on support vector regression," Statistics and computing, vol. 14, no.3, pp. 199-222, 2004.
- [89] A.P. Rotshtein, and H.B. Rakytyanska, "Fuzzy evidence in identification, forecasting and diagnosis," Springer, 2012.
- [90] H. tanaka, S. Uejima, and K. Asai, "Linear regression analysis with fuzzy model," IEEE Transactions on systems, man, and cybernetics, vol. 12, no. 6, 1982.
- [91] G. Biau, and L. Devroye, "Lectures on the nearest neighbor method," Springer, 2015.
- [92] B.W. Silverman, "Density estimation for statistics and data analysis," Chapman & Hall, 186.
- [93] Met Office website: <http://www.metoffice.gov.uk/>.
- [94] Website of European Centre for Medium-Range Weather Forecasts (ECMWF): <https://www.ecmwf.int/>.
- [95] National and Oceanic Atmospheric Administration (NOAA) website: <http://www.noaa.gov/>.
- [96] National Centers for Environmental Protection (NCEP) website: <http://www.ncep.noaa.gov/>.
- [97] P. Ribeiro, "Time-varying waveform distortions in power systems," John Wiley and Sons, 2009.
- [98] R. Ranjan, and T. Gneiting, "Combining probability forecasts," Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 72, no. 1, pp. 71-91, 2010.
- [99] A. Bracale, G. Carpinelli, P. De Falco, R. Rizzo, and A. Russo, "New advanced method and cost-based indices applied to probabilistic forecasting of photovoltaic generation," Journal of Renewable and Sustainable Energy, vol. 8, no. 2, 023505, 2016.
- [100] A. Bracale, G. Carpinelli, and P. De Falco, "A probabilistic competitive ensemble method for short-term photovoltaic power forecasting," IEEE Transactions on Sustainable Energy, vol. 8, no. 2, pp. 551-560, 2017.
- [101] M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids," Renewable and Sustainable Energy Reviews, vol. 27, pp. 65-76, 2013.

- [102] A. Tascikaraoglu, et al., “Compressive Spatio-Temporal Forecasting of Meteorological Quantities and Photovoltaic Power,” *IEEE Transactions on Sustainable Energy*, vol. 7, no. 3, pp. 1295-1305, 2016.
- [103] G. Chicco, V. Cocina, P. Di Leo, F. Spertino, and A.M. Pavan, “Error assessment of solar irradiance forecasts and AC power from energy conversion model in grid-connected photovoltaic systems,” *Energies*, vol. 9, no. 1, 8, 2015.
- [104] P. Bacher, H. Madsen, and H.A. Nielsen, “Online short-term solar power forecasting,” *Solar Energy*, vol. 83, no. 10, 2009.
- [105] C. Monteiro, T. Santos, L. Fernandez-Jimenez, I. Ramirez-Rosado, and M. Terreros-Olarte, “Short-term power forecasting model for photovoltaic plants based on historical similarity,” *Energies*, vol. 6, pp. 2624–2643, 2013.
- [106] S. Jafarzadeh, M.S. Fadali, and C.Y. Evrenosoglu, “Solar Power Prediction Using Interval Type-2 TSK Modeling,” *IEEE Transactions on Sustainable Energy*, vol. 4, no. 2, pp. 333-339, 2013.
- [107] A. Bracale, P. Caramia, G. Carpinelli, A. R. Di Fazio, and P. Varilone, “A Bayesian-Based Approach for a Short-Term Steady-State Forecast of a Smart Grid,” *IEEE Transactions on Smart Grid*, vol. 4, no. 4, pp. 1760-1771, 2013.
- [108] M. Zamo, O. Mestre, P. Arbogast, and O. Pannekoucke, “A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production. Part II: Probabilistic forecast of daily production,” *Solar Energy*, vol. 105, pp. 804–816, 2014.
- [109] M.P. Almeida, O. Perpiñán, and L. Narvarte, “PV power forecast using a nonparametric PV model,” *Solar Energy*, vol. 115, pp. 354-368, 2015.
- [110] S. Alessandrini, L. Delle Monache, S. Sperati, and G. Cervone, “An analog ensemble for short-term probabilistic solar power forecast,” *Applied Energy*, vol. 157, pp. 95-110, 2015.
- [111] R.J. Bessa, A. Trindade, Cátia S.P. Silva, and V. Miranda, “Probabilistic solar power forecasting in smart grids using distributed information,” *International Journal of Electrical Power & Energy Systems*, vol. 72, pp. 16-23, 2015.
- [112] D. AlHakeem, P. Mandal, A. Haque, A. Yona, T. Senjyu, and T. Tseng, “A new strategy to quantify uncertainties of wavelet-GRNN-PSO based solar PV power forecasts using bootstrap confidence intervals,” in *proc. of IEEE Power & Energy Society General Meeting*, Denver, USA, 2015.
- [113] J.G. Fonseca, T. Oozeki, H. Ohtake, T. Takashima, and K. Ogimoto, “On the use of maximum likelihood and input data similarity to obtain prediction intervals for forecasts of photovoltaic power generation,” *Journal of Electrical Engineering and Technology*, vol. 10, no. 3, pp. 1342–1348, 2015.
- [114] M. Rana, I. Koprinska, and V.G. Agelidis, “2D-interval forecasts for solar power production,” *Solar Energy*, vol. 122, pp. 191-203, 2015.
- [115] M. David, F. Ramahatana, P.J. Trombe, and P. Lauret, “Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models,” *Solar Energy*, vol. 133, pp. 55-72, 2016.

- [116] S. Sperati, S. Alessandrini, and L. Delle Monache, "An application of the ECMWF Ensemble Prediction System for short-term solar power forecasting," *Solar Energy*, vol. 133, pp. 437-450, 2016.
- [117] M. Pierro, F. Bucci, M. De Felice, E. Maggioni, D. Moser, A. Perotto, F. Spada, and C. Cornaro, "Multi-Model Ensemble for day ahead prediction of photovoltaic power generation," *Solar Energy*, vol. 134, pp. 132-146, 2016.
- [118] Y. Liu, S. Shimada, J. Yoshino, T. Kobayashi, Y. Miwa, and K. Furuta, "Ensemble forecasting of solar irradiance by applying a mesoscale meteorological model," *Solar Energy*, vol. 136, pp. 597-605, 2016.
- [119] F. Golestaneh, P. Pinson, and H. B. Gooi, "Very Short-Term Nonparametric Probabilistic Forecasting of Renewable Energy Generation—With Application to Solar Energy," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3850-3863, 2016.
- [120] H. Takeda, "Short-term ensemble forecast for purchased photovoltaic generation," *Solar Energy*, vol. 149, pp. 176-187, 2017.
- [121] J. Huang, and M. Perry, "A semi-empirical approach using gradient boosting an -nearest neighbors regression for GEFCom2014 probabilistic solar power forecasting," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1081-1086, 2016.
- [122] G.I. Nagy, G. Barta, S. Kazi, G. Borbély, and G. Simon, "GEFCom2014: Probabilistic solar and wind power forecasting using a generalized additive tree ensemble approach," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1087-1093, 2016.
- [123] R. Juban, He. Ohlsson, M. Maasoumy, L. Poirier, and J. Zico Kolter, "A multiple quantile regression approach to the wind, solar, and price tracks of GEFCom2014," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1094-1102, 2016.
- [124] A.A. Mohammed, W. Yaqub, and Z. Aung, "Probabilistic forecasting of solar power: an ensemble learning approach," in *Intelligent Decision Technologies*, pp. 449-458, 2015.
- [125] S. Chai, Z. Xu, and W. K. Wong, "Optimal Granule-Based PIs Construction for Solar Irradiance Forecast," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 3332-3333, 2016.
- [126] C. Wan, J. Lin, Y. Song, Z. Xu, and G. Yang, "Probabilistic Forecasting of Photovoltaic Generation: An Efficient Statistical Approach," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 2471-2472, 2017.
- [127] E. Scolari, F. Sossan, and M. Paolone, "Irradiance prediction intervals for PV stochastic generation in microgrid applications," *Solar Energy*, vol. 139, pp. 116-129, 2016.
- [128] E. Scolari, D. Torregrossa, J. Y. Le Boudec, and M. Paolone, "Ultra-short-term prediction intervals of photovoltaic AC active power," 2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), China, 2016.

- [129] Y. Chu, M. Li, H.T.C. Pedro, and C.F.M. Coimbra, "Real-time prediction intervals for intra-hour DNI forecasts," *Renewable Energy*, vol. 83, pp. 234-244, 2015.
- [130] Y. Chu, and C.F.M. Coimbra, "Short-term probabilistic forecasts for Direct Normal Irradiance," *Renewable Energy*, vol. 101, pp. 526-536, 2017.
- [131] J. Boland, "Spatial-temporal forecasting of solar radiation," *Renewable Energy*, vol. 75, pp. 607-616, 2015.
- [132] A. Grantham, Y.R. Gel, and J. Boland, "Nonparametric short-term probabilistic forecasting for solar radiation," *Solar Energy*, vol. 133, pp. 465-475, 2016.
- [133] S. Wang, and C. Jia, "Prediction Intervals for Short-Term Photovoltaic Generation Forecasts," 2015 Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control (IMCCC), China, 2015.
- [134] T. Esum, and P.L. Chapman, "Comparison of photovoltaic array maximum power point tracking techniques," *IEEE Transactions on energy conversion*, vol. 22, no. 2, pp. 439-449, 2007.
- [135] A.A. El Desouky, "Security Constrained Generation Scheduling for Grids Incorporating Wind, Photovoltaic and Thermal Power", *Electric Power System Research*, vol. 116, pp. 284-292, 2014.
- [136] B. Kroposki, L. Mrig, C. Whitaker, and J. Newmiller, "Development of a Photovoltaic Module Energy Ratings Methodology", *AIP conference proceedings*, vol. 535, pp. 587-594, 1996.
- [137] G. Tina, S. Gagliano, and S. Raiti, "Hybrid Solar/Wind Power System Probabilistic Modelling for Long-term Performance Assessment", *Solar Energy*, vol. 80, pp. 578-588, 2006.
- [138] M. Kacira, M. Simsek, Y. Babur, and S. Demirkol, "Determining optimum tilt angles and orientations of photovoltaic panels in Sanliurfa, Turkey," *Renewable Energy*, vol. 29, no. 8, pp. 1265-1275, 2004.
- [139] S.H. Karaki, R.B. Chedid, and R. Ramadan, "Probabilistic performance assessment of autonomous solar-wind energy conversion systems", *IEEE Transactions on Energy Conversion*, vol. 14, pp. 766-772, 1999.
- [140] V.A. Graham, and K.G.T. Hollands, "A Method to Generate Synthetic Hourly Solar Radiation Globally", *Solar Energy*, vol. 44, pp. 333-341, 1990.
- [141] K.G.T. Hollands, and J.F. Orgill, "Correlation Equation for Hourly Diffuse Radiation on Horizontal Surface", *Solar Energy*, vol. 19, pp. 357-359, 1977.
- [142] K.G.T. Hollands, and R.G. Huget, "A Probability Density Function for the Clearness Index, with Applications", *Solar Energy*, vol. 30, pp. 195-209, 1983.
- [143] J. Zhang, J. Pu, J.D. McCalley, H. Stern, and W.A. Gallus, "A Bayesian approach to short term transmission line thermal overload risk assessment", *IEEE Transactions on Power Delivery*, vol. 17, pp. 770-778, 2002.
- [144] J. Von Neumann, "Various Techniques Used in Connection With Random Digits", *Natural Bureau of Standards*, vol. 12, pp. 36-38, 1951.
- [145] P. Pinson, G. Reikard, and J.R. Bidlot, "Probabilistic forecasting of the wave energy flux", *Applied Energy*, vol. 93, pp. 364-370, 2012.

- [146] P. Pinson, "Very-short-term probabilistic forecasting of wind power with generalized logit-normal distributions", *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 61, pp. 555-576, 2012.
- [147] H. Nfaoui, H. Essiarab, and A.A. M. Sayigh, "A stochastic Markov chain model for simulating wind speed time series at Tangiers, Morocco," *Renewable Energy*, vol. 29, no. 8, pp. 1407-1418, 2004.
- [148] P. Pinson, H.A. Nielsen, J.K. Møller, H. Madsen, and G.N. Kariniotakis, "Non-parametric probabilistic forecasts of wind power: required properties and evaluation," *Wind Energy*, vol. 10, no. 6, pp. 497-516, 2007.
- [149] J.K. Møller, H.A. Nielsen, and H. Madsen, "Time-adaptive quantile regression," *Computational Statistics & Data Analysis*, vol. 52, no. 3, pp. 1292-1303, 2008.
- [150] J. Tastu, P. Pinson, P.J. Trombe, and H. Madsen, "Probabilistic forecasts of wind power generation accounting for geographically dispersed information," *IEEE Trans. on Smart Grid*, vol. 5, no. 1, pp. 480-489, 2014.
- [151] S.G. Hall, and J. Mitchell, "Combining density forecasts," *International Journal of Forecasting*, vol. 23, no. 1, pp. 1-13, 2007.
- [152] A.S. Jore, J. Mitchell, and S.P. Vahey, "Combining forecast densities from VARs with uncertain instabilities," *Journal of Applied Econometrics*, vol. 25, no. 4, pp. 621-634, 2010.
- [153] T. Gneiting, A.E. Raftery, A.H. Westveld III, and T. Goldman, "Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation," *Monthly Weather Review*, vol. 133, no. 5, pp. 1098-1118, 2005.
- [154] H. Hersbach, "Decomposition of the continuous ranked probability score for ensemble prediction systems," *Weather and Forecasting*, vol. 15, no. 5, pp. 559-570, 2000.
- [155] R.T. Marler, and J.S. Arora, "Survey of multi-objective optimization methods for engineering," *Structural and multidisciplinary optimization*, vol. 26, no. 6, pp. 369-395, 2004.
- [156] NREL Solar Radiation Research Laboratory (SRRL): Baseline Measurement System (BMS), Golden, Colorado, USA. Available: <http://dx.doi.org/10.5439/1052221>.
- [157] A. Bracale, and P. De Falco, "An Advanced Bayesian Method for Short-Term Probabilistic Forecasting of the Generation of Wind Power," *Energies*, vol. 8, pp. 10293-10314, 2015.
- [158] A. Bracale, G. Carpinelli, and P. De Falco, "A Bayesian-based approach for the short-term forecasting of electrical loads in smart grids. Part I: theoretical aspects," in *proc. of 23rd IEEE International Symposium on Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM 2016)*, Anacapri, Italy, 2016.
- [159] A. Bracale, G. Carpinelli, and P. De Falco, "A Bayesian-based approach for the short-term forecasting of electrical loads in smart grids. Part II: numerical applications," in *proc. of 23rd IEEE International Symposium on Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM 2016)*, Anacapri, Italy, 2016.

- [160] A. Bracale, G. Carpinelli, P. De Falco, and M. Pagano, "A probabilistic approach for forecasting the allowable current of oil-immersed transformers," of interest of IEEE Transactions on Power Delivery (second round review).
- [161] A. Bracale, G. Carpinelli, P. De Falco, and T. Hong, "Short-term Industrial Load Forecasting: A Case Study in an Italian Factor," in *proc. of IEEE PES International Conference on Innovative Smart Grid Technologies (ISGT) Europe 2017*, Turin, Italy, 2017.
- [162] A.R. Khan, A. Mahmood, A. Safdar, Z.A. Khan, and N.A. Khan, "Load forecasting, dynamic pricing and DSM in smart grid: A review," *Renewable and Sustainable Energy Reviews*, vol. 54, pp. 1311-1322, 2016.
- [163] L. Hernandez, et al., "A Survey on Electric Power Demand Forecasting: Future Trends in Smart Grids, Microgrids and Smart Buildings," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1460-1495.
- [164] J. Blancarte, M. Batton-Hubert, X. Bay, M.A. Girard, and A. Grau, "Short Term Load Forecasting in the Industry for Establishing Consumption Baselines: A French Case," in "Modeling and Stochastic Learning for Forecasting in High Dimensions," Springer, Switzerland, 2015.
- [165] C.S. Chen, Y.M. Tzeng, and J.C. Hwang, "Application of Artificial Neural Networks to Substation Load Forecasting," *Electric Power Systems Research*, vol. 38, pp. 153-160, 1996.
- [166] L. Hernandez, C. Baladron, J. M. Aguiar, B. Carro, and A. Sanchez-Esguevillas, "Classification and Clustering of Electricity Demand Patterns in Industrial Parks," *Energies*, vol. 5, no. 12, pp. 5215-5228, 2012.
- [167] T. Zhang, G. Zhang, J. Lu, X. Feng, and W. Yang, "A New Index and Classification Approach for Load Pattern Analysis of Large Electricity Customers," *IEEE Transactions on Power Systems*, vol. 27, no. 1, pp. 153-160, 2012.
- [168] G.J. Tsekouras, N.D. Hatziargyriou, and E.N. Dialynas, "Two-Stage Pattern Recognition of Load Curves for Classification of Electricity Customers," *IEEE Transactions on Power Systems*, vol. 22, no. 3, pp. 1120-1128, 2007.
- [169] M. Espinoza, C. Joye, R. Belmans, and B. De Moor, "Short-Term Load Forecasting, Profile Identification, and Customer Segmentation: A Methodology Based on Periodic Time Series," *IEEE Transactions on Power Systems*, vol. 20, no. 3, pp. 1622-1630, 2005.
- [170] D.A. Gundin, C. Garcia, E. Gomez, and G. Vega, "Short-term load forecasting for industrial customers using FasArt and FasBack neuro-fuzzy systems," in *Proc. of 14th Power Systems Computation Conference (PSCC)*, Sevilla, Spain, 2002.
- [171] K.S.K. Weranga, D.P. Chandima, S.R. Munasinghe, S.P. Kumarawadu, and A.M. Harsha S. Abeykoon, "Short-term electricity demand forecasting method for smart meters," *6th IEEE International Conference on Information and Automation for Sustainability*, Beijing, China, pp. 266-272, 2012.
- [172] A. Ahmad, N. Javaid, M. Guizani, N. Alrajeh, and Z.A. Khan, "An Accurate and Fast Converging Short-Term Load Forecasting Model for Industrial Applications in a Smart Grid," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 5, pp. 2587-2596, 2017.

- [173] S.V. Rodygina and A.V. Rodygin, "Comparative analysis of electrical load forecasting models of industrial plants," in proc. of Dynamics of Systems, Mechanisms and Machines (Dynamics), Omsk, Russia, pp. 1-4, 2016.
- [174] P. Kou, and F. Gao, "A sparse heteroscedastic model for the probabilistic load forecasting in energy-intensive enterprises," International Journal of Electrical Power & Energy Systems, vol. 55, pp. 144-154, 2014.
- [175] S. Ben Taieb, and R.J. Hyndman, "A gradient boosting approach to the Kaggle load forecasting competition," International Journal of Forecasting, vol. 30, no. 2, pp. 382-394, 2014.
- [176] N. Charlton, and C. Singleton, "A refined parametric model for short term load forecasting," International Journal of Forecasting, vol. 30, no. 2, pp. 364-368, 2014.
- [177] J.R. Lloyd, "GEFCom2012 hierarchical load forecasting: Gradient boosting machines and Gaussian processes," International Journal of Forecasting, vol. 30, no. 2, pp. 369-374, 2014.
- [178] R. Nedellec, J. Cugliari, and Y. Goude, "GEFCom2012: Electric load forecasting and backcasting with semi-parametric models," International Journal of Forecasting, vol. 30, no. 2, pp. 375-381, 2014.
- [179] G.A.F. Seber, and A.J. Lee, "*Linear regression analysis*," vol. 936, John Wiley & Sons, 2012.
- [180] Gestore dei Servizi Energetici website: <http://www.gse.it/en/Pages/default.aspx>.
- [181] P. Caramia, G. Carpinelli, and P. Verde, "Power quality indices in liberalized markets," John Wiley and Sons, UK, 2009.
- [182] M. Bollen, and I.Y.H. Gu, "Signal Processing of Power Quality Disturbances," Wiley-IEEE Press, New Jersey, USA, 2006.
- [183] D.J. Cannon, et al., "Using reanalysis data to quantify extreme wind power generation statistics: a 33 year case study in Great Britain," Renewable Energy, vol. 75, pp. 767-778, 2015.
- [184] P. Friederichs, and L.T. Thorndis, "Forecast Verification for Extreme Value Distributions with an Application to Probabilistic Peak Wind Prediction," Environmetrics, vol. 23, n. 7, pp. 579-594, 2012.
- [185] T. Ishihara, and A. Yamaguchi, "Prediction of the extreme wind speed in the mixed climate region by using Monte Carlo simulation and measure-correlate-predict method," Wind Energy, vol. 18, no. 1, pp. 171-186, 2015.
- [186] P.C. Chang, R.Y. Yang, and C.M. Lai, "Potential of offshore wind energy and extreme wind speed forecasting on the west coast of Taiwan," Energies, vol. 8, no. 3, pp. 1685-1700, 2015.
- [187] M. Oesting, M. Schlather, and P. Friederichs, "Statistical post-processing of forecasts for extremes using bivariate brown-resnick processes with an application to wind gusts," Extremes, vol. 20, no. 2, pp. 309-332, 2017.
- [188] B. Mahmoudian, and M. Mohammadzadeh, "A spatio-temporal dynamic regression model for extreme wind speeds," Extremes, vol. 17, no. 2, pp. 221-245, 2014.

- [189] E. Chiodo, and P. De Falco, "Inverse Burr Distribution for Extreme Wind Speed Prediction: Genesis, Identification and Estimation," *Electric Power System Research*, vol. 141, pp. 549-561, 2016.
- [190] A. Bracale, G. Carpinelli, and P. De Falco, "A new finite mixture distribution and its expectation-maximization procedure for extreme wind speed characterization," *Renewable Energy*, vol. 113, pp. 1366-1377, 2017.
- [191] J.A. Carta, P. Ramirez, S. Velazquez, "A review of wind speed probability distributions used in wind energy analysis. Case studies in the Canary islands," *Renewable and Sustainable Energy Reviews*, vol. 13, pp. 933-955, 2009.
- [192] J. Wang, S. Qin, S. Jin, and J. Wu, "Estimation methods review and analysis of offshore extreme wind speeds and wind energy resources," *Renewable and Sustainable Energy Reviews*, vol. 42 pp. 26-42, 2015.
- [193] D. Villanueva, and A. Feijoo, "Wind Power Distributions: A review of their Applications," *Renewable and Sustainable Energy Reviews*, vol. 14, pp. 1490-1495, 2010.
- [194] J.Y. Shin, T. Ouarda, and T. Lee, "Heterogeneous mixture distributions for modeling wind speed, application to the UAE," *Renewable Energy*, vol. 91, pp. 40-52, 2016.
- [195] E. Castillo, "Extreme value theory in engineering", Academic Press, New York, 1988.
- [196] E. Castillo, A.S. Hadi, N. Balakrishnan, and J.M. Sarabia, "Extreme value and related models with applications in engineering and science," John Wiley and Sons, New Jersey, USA, 2005.
- [197] S. Coles, "An Introduction to Statistical Modeling of Extreme Values", Springer Series in Statistics. Springer-Verlag, London, 2001.
- [198] O. Perrin, H Rootzen, and R. Taesler, "A discussion of statistical methods used to estimate extreme wind speeds," *Theoretical and applied climatology*, vol. 85, no. 3, pp. 203-215, 2006.
- [199] Y. An, and M.D. Pandey, "A comparison of methods of extreme wind speed estimation," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 93, pp. 535-545, 2005.
- [200] J.P. Palutikof, B.B. Brabson, D.H. Lister, and S.T. Adcock, "A review of methods to calculate extreme wind speeds," *Meteorological applications*, vol. 6, no. 2, pp. 119-132, 1999.
- [201] I.R. Young, J. Vinoth, S. Zieger, and A.V. Babanin, "Investigation of trends in extreme value wave height and wind speed," *Journal of Geophysical Research*, vol. 117, 2012.
- [202] E. Cheng, and C. Yeung, "Generalized Extreme Gust Wind Speeds Distributions," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 90, pp. 1657-1669, 2002.
- [203] J. Gross, A. Heckert, J. Lechner, and E. Simiu, "Novel Extreme Value Estimation Procedures: Application to Extreme Wind Data," in *Extreme Value Theory and Applications*, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 139-158, 1994.

- [204] E. Chiodo, G. Mazzanti, and M. Karimian, "Bayes Estimation of Inverse Weibull Distribution for Extreme Wind Speed Prediction," in *proc. of IEEE International Conference on Clean Electrical Power Renewable Energy Resources Impact (ICCEP 2015)*, Taormina, Italy, 2015.
- [205] E. Simiu, N.A. Heckert, J.J. Filliben, and S.K. Johnson, "Extreme wind load estimates based on the Gumbel distribution of dynamic pressures: an assessment," *Structural Safety*, vol. 23, no. 3, 2001.
- [206] J.D. Holmes, and W.W. Moriarty, "Application of the generalized pareto distribution to extreme value analysis in wind engineering," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 83, pp. 1–10, 1999.
- [207] X.G. Larsen, J. Mann, O. Rathmann, and H.E. Jorgensen, "Uncertainties of the 50-year wind from short time series using generalized extreme value distribution and generalized Pareto distribution," *Wind Energy*, vol. 18, pp. 59-74, 2015.
- [208] E.C. Pinheiro, and S. Ferrari, "A comparative review of generalizations of the Gumbel extreme value distribution with an application to wind speed data," *Journal of Statistical Computation and Simulation*, vol. 86, no. 11, pp. 2241-2261, 2016.
- [209] M.M.F. de Oliveira, N.F.F. Ebecken,, J.L.F. de Oliveira, and E. Gilleland, "Generalized extreme wind speed distributions in South America over the Atlantic Ocean region," *Theoretical and applied climatology*, vol. 104, no. 3-4, pp. 377-385, 2011.
- [210] G. D'Amico, F. Petroni, and F. Pratico, "Wind speed prediction for wind farm applications by Extreme Value Theory and Copulas," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 145, pp. 229-236, 2015.
- [211] C. Kleiber, "A guide to the Dagum distributions," Springer, New York, USA, 2008.
- [212] C. Schluter, and M. Trede, "Tails of Lorenz curves," *Journal of Econometrics*, vol. 109, n. 1, pp. 151-166, 2002.
- [213] E. Chiodo, and P. De Falco, "Bayesian Estimation of Inverse Burr Stress-Strength Model for Power System Components Reliability Assessment," in *proc. of 23rd IEEE International Symposium on Power Electronics, Electrical Drives, Automation and Motion, (SPEEDAM 2016)*, Anacapri, Italy, June 2016.
- [214] J.E. Gentle, "Random number generation and Monte Carlo methods," Springer Science & Business Media, 2013.
- [215] J.A. Bilmes, "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models," *International Computer Science Institute*, vol. 4, no. 510, 1998.
- [216] F.J. Massey Jr., "The Kolmogorov-Smirnov test for goodness of fit," *Journal of the American statistical Association*, vol. 46, no. 253, pp. 68-78, 1951.
- [217] H.O. Lancaster and E. Seneta, "Chi-Square Distribution," John Wiley & Sons, New York, USA, 2005.
- [218] NREL National Wind Technology Center website: <http://dx.doi.org/10.5439/1052222>.
- [219] NREL Solar Resource & Meteorological Assessment Project website: <http://dx.doi.org/10.5439/1052449>.

- [220] Sustainable Energy Authority of Ireland Wind Mapping System website: <http://maps.seai.ie/wind>.
- [221] J.W. Tukey, "Exploratory data analysis," Addison-Wesley Publishing Company, USA, 1977.
- [222] T.N. Krishnamurti, V. Kumar, A. Simon, A. Bhardwaj, T. Ghosh, and R. Ross, "A review of multimodel superensemble forecasting for weather, seasonal climate, and hurricanes," *Reviews of Geophysics*, vol. 54, pp. 336–377, 2016.
- [223] T. Gneiting, F. Balabdaoui, and A.E. Raftery, "Probabilistic forecasts, calibration and sharpness," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 69, no. 2, pp. 243-268, 2007.
- [224] A.E. Raftery, T. Gneiting, F. Balabdaoui, and M. Polakowski, "Using Bayesian model averaging to calibrate forecast ensembles," *Monthly Weather Review*, vol. 133, no. 5, pp. 1155-1174, 2005.
- [225] P. Pinson, H.A. Nielsen, J.K. Møller, H. Madsen, and G.N. Kariniotakis, "Non-parametric probabilistic forecasts of wind power: required properties and evaluation," *Wind Energy*, vol. 10, no. 6, pp. 497-516, 2007.
- [226] J.B. Bremnes, "Probabilistic wind power forecasts using local quantile regression," *Wind Energy*, vol. 7, no. 1, pp. 47-54, 2004.
- [227] T.M. Hamill, "Reliability diagrams for multicategory probabilistic forecasts," *Weather and forecasting*, vol. 12, no. 4, pp. 736-741, 1997.
- [228] P. Pinson, J. Juban, and G.N. Kariniotakis, "On the quality and value of probabilistic forecasts of wind generation," in *proc. of International Conference on Probabilistic Methods Applied to Power Systems (PMAPS 2006)*, Stockholm, Sweden, 2006.
- [229] T.M. Hamill, "Interpretation of rank histograms for verifying ensemble forecasts," *Monthly Weather Review*, vol. 129, pp. 550-560, 2001.
- [230] T. Gneiting, and A.E. Raftery, "Strictly proper scoring rules, prediction, and estimation," *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 359-378, 2007.

List of publications

The research efforts during the Ph.D. course produced the following relevant publications:

Journal papers

- 1) A. Bracale, and P. De Falco, "An Advanced Bayesian Method for Short-Term Probabilistic Forecasting of the Generation of Wind Power," *Energies*, vol. 8, pp. 10293-10314, 2015.
- 2) A. Bracale, G. Carpinelli, P. De Falco, R. Rizzo, and A. Russo, "New advanced method and cost-based indices applied to probabilistic forecasting of photovoltaic generation," *Journal of Renewable and Sustainable Energy*, vol. 8, no. 2, 023505, 2016.
- 3) E. Chiodo, and P. De Falco, "Inverse Burr Distribution for Extreme Wind Speed Prediction: Genesis, Identification and Estimation," *Electric Power System Research*, vol. 141, pp. 549-561, 2016.
- 4) A. Bracale, G. Carpinelli, and P. De Falco, "A probabilistic competitive ensemble method for short-term photovoltaic power forecasting," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 2, pp. 551-560, 2017.
- 5) A. Bracale, G. Carpinelli, and P. De Falco, "A new finite mixture distribution and its expectation-maximization procedure for extreme wind speed characterization," *Renewable Energy*, vol. 113, pp. 1366-1377, 2017.
- 6) A. Bracale, G. Carpinelli, P. De Falco, and M. Pagano, "A probabilistic approach for forecasting the allowable current of oil-immersed transformers," of interest of *IEEE Transactions on Power Delivery* (second round review).

Book chapters

- 7) A. Bracale, G. Carpinelli, and P. De Falco, "A new ensemble probabilistic method for short-term photovoltaic power forecasting," book chapter in *Sustainable Energy - Technological Issues, Applications and Case Studies*, InTechOpen, 2016.

Conference papers

- 8) A. Bracale, G. Carpinelli, and P. De Falco, "A Bayesian-based approach for the short-term forecasting of electrical loads in smart grids. Part I: theoretical aspects," in proc. of 23rd IEEE International Symposium on Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM 2016), Anacapri, Italy, 2016.
- 9) A. Bracale, G. Carpinelli, and P. De Falco, "A Bayesian-based approach for the short-term forecasting of electrical loads in smart grids. Part II: numerical applications," in proc. of 23rd IEEE International Symposium on Power Electronics, Electrical Drives, Automation and Motion, (SPEEDAM 2016), Anacapri, Italy, 2016.

- 10) E. Chiodo, and P. De Falco, "Bayesian Estimation of Inverse Burr Stress-Strength Model for Power System Components Reliability Assessment," in proc. of 23rd IEEE International Symposium on Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM 2016), Anacapri, Italy, 2016.
- 11) A. Bracale, G. Carpinelli, P. De Falco, and T. Hong, "Short-term industrial load forecasting: a case study in an Italian factory," in proc. of IEEE PES Innovative Smart grid Technologies (ISGT) Europe 2017, Turin, Italy, 2017.