



UNIVERSITY OF GENOVA

POLYTECHNIC SCHOOL

DEGREE IN ENVIRONMENTAL ENGINEERING

Exploiting the ECMWF Ensemble Prediction System for wind energy forecasting

December 2020

Supervisor

Prof. Andrea Mazzino

Co-Supervisor

Ing. Matteo Fattore

Candidate

Gabriele Casciaro

Contents

List of Figures	iv
List of Tables	xi
Abstract	xv
1 The Italian Electricity Market	1
1.1 The institutional subjects of the Electricity Market	1
1.2 The Electricity Market structure	2
1.3 The Electricity Market structure	2
1.3.1 The System marginal price (SMP)	3
1.3.2 Nodal and zonal markets	4
1.3.3 The The Day Ahead Market (MGP)	4
1.3.4 The Intraday Market (MI)	5
1.3.5 Dispatching of electricity	5
2 The European Centre for Medium-Range Weather Forecasts (ECMWF)	7
2.1 The European Centre for Medium-Range Weather Forecasts – an historical background	7
2.1.1 The creation of ECMWF	8
2.1.2 The ECMWF forecasting model since 1979- an overview	10
2.2 The ECMWF global atmospheric model	10
2.2.1 The model equations	10
2.2.2 The resolution in time and space	11
2.2.3 The numerical formulation	13
2.3 Parametrization of physical processes	14
2.3.1 The model orography	14
2.3.2 The Planetary Boundary Layer	14
2.3.3 Radiation	15
2.3.4 Clouds	16

2.3.5	The hydrological cycle	16
3	The Ensemble Prediction System (EPS)	19
3.1	Introduction	19
3.2	The ECMWF Ensemble Prediction System - an overview . . .	20
3.3	The performance of the ECMWF Ensemble Prediction System	22
3.4	Perturbations	22
3.4.1	The calculations of perturbations in the mid-latitudes .	22
3.4.2	The simulation of model errors	24
3.4.3	Definition of the initial perturbations	25
3.5	Spread-skill	26
3.6	EPS clustering	26
3.6.1	The operational clustering	26
3.6.2	The “tubing” clustering	27
3.6.3	No ideal clustering	28
4	The statistical indices	30
4.1	Introduction	30
4.2	The indices of the errors used	31
4.2.1	Bias and NBias	31
4.2.2	RMSE and NRMSE	32
4.2.3	MAE and NMAE	33
4.2.4	SI, Scatter Index	34
4.2.5	HH, Hanna and Heinold index	34
4.2.6	Pearson index	35
5	Analysis of forecasts and observations	37
5.1	Introduction	37
5.2	Observed data import for Marsical	37
5.2.1	Substitution of null records with R2 and R3 ones . . .	41
5.2.2	Substitution or elimination of wrong records with the average of the records around	41
5.3	ECMWF forecasts data import for EPS and HRES	44
5.3.1	EPS structure	45
5.3.2	HRES structure	48
5.4	Analysis and characteristics of the starting data	51
5.4.1	Seasonal trend of wind speed	51
5.4.2	Daily trend of wind speed	54
5.4.3	The wind direction	57
5.5	Quality and accuracy of the starting forecasts	60

6	Correction of the forecasts	68
6.1	Introduction	68
6.2	The best correction techniques	68
6.2.1	Linear regression	68
6.2.2	Quantile regression	69
6.2.3	Gamma distribution	71
6.3	The strategies for correcting the forecasts	76
6.3.1	The correction on the forecasts without conditioning	77
6.3.2	Correction dividing for the directions	81
6.3.3	Correction dividing for day hours	85
6.3.4	Correction dividing in intervals of wind speed	90
6.3.5	Correction using prediction error correlation	94
6.3.6	The moving average	98
6.3.7	Correction of negative values	103
6.3.8	The use of the 50 members of the EPS	103
6.4	Analysis of the best corrections	110
6.4.1	Comparison of the corrected forecasts starting from EPS or HRES	110
6.4.2	Correction using the gamma distribution	111
6.4.3	The variation of principal index as function of the fore- cast horizon before and after the best correction	115
6.4.4	Summary of how correction strategies affected the fore- cast.	118
6.5	Variation of indices as a function of the number of data for training	120
6.6	The prediction of power	122
6.7	Comparison with other forecast calibration techniques	125
6.7.1	Comparison with Machine Learning	125
6.7.2	Comparison with a scientific article of a forecast gen- erated in a site similar to the one analyzed	129
	Conclusions	133
	Bibliography	135

List of Figures

1.1	The Electricity Market structure. (Taken from [1])	3
1.2	Graphic representation of the difference between node and zone. (Taken from [1])	4
2.1	Artistic sketch of the Ensemble Prediction System.(Take from Ref. [3])	8
2.2	To the left the vertical resolution before 1999 (31 level to 10 hPa, and to the right after 1999 (60 levels to 0.1 hPa).(Take from Ref. [3])	12
2.3	The model height (in deksmeterd) for southwestern Europe.(Taken from Ref. [3])	15
2.4	Accumulated rainfall during the first 96 hours of the T213 operational forecast 5 July 1997 12 UTC. The floodings in eastern Europe summer 1997 were call forecast by the ECMWF model. However, the maximum rainfall of 400 mm in siuth-eastern Poland was slightly underastimated. (Taken from Ref. [3])	18
2.5	Accumulated rainfall during the first 96 hours of the T639 experimental model forecast 5 July 1997 12 UTC. With a higher resolution model the correct level of intensity is achieved, and the orographic effects more realistically treated. (Taken from Ref. [3]))	18
3.1	Representation of how the different perturbed initial conditions return a different forecast. (The ECMWF Ensemble Prediction System)	21

3.2	A comparison of the performance of all global ensemble prediction systems operational in the world demonstrates the leading position of the ECMWF EPS. The skill measure used here is the Ranked Probability Skill Score (RPSS), which is 1 for a perfect forecast and 0 for a forecast no better than climatology. (The ECMWF Ensemble Prediction System)	23
3.3	The performance of the EPS has improved steadily since it became operational in the mid-1990s. The skill measure used here is the Ranked Probability Skill Score (RPSS), which is 1 for a perfect forecast and 0 for a forecast no better than climatology. (The ECMWF Ensemble Prediction System) . . .	24
3.4	The five main clustering areas, the European and four sub-area. (Taken from Ref. [3])	28
4.1	Systematic difference between forecast and reality.	31
4.2	Residuals of the data from the diagonal line.	33
4.3	Variation of the correlation visualized as scatterplots.	36
5.1	Marsica wind farm photo. Note how the wind turbines are quite close to each other.	38
5.2	Wind speed time series measured from 08/01/2017 to 28/01/2017. There is a sudden zeroing of the wind speed for almost 4 days. This is clearly due to some kind of error in the data acquisition.	39
5.3	Empirical transfer function. Data that is not in the vicinity of the transfer curve are those that need to be corrected or modified.	40
5.4	Scatter plot of the speeds recorded with the anemometer R1 vs those recorded with the anemometer R2.	42
5.5	Scatter plot of the power recorded with the anemometer R1 vs those recorded with the anemometer R2.	43
5.6	Transfer function after operations to clean up the data.	44
5.7	The 50 forecasts of the EPS of 1 and 2 March 2017. Each different forecast has its own color, as it can be seen the width of the members tends to widen over time.	46
5.8	The 50 in red are the forecast referring to the run of the day before, while the black ones are the forecasts generated the same day.	46
5.9	Time series from 12/3/2018 to 16/4/2018 in which the forecasts 0-24 hours were superimposed on those 24-48 hours. . . .	47
5.10	Scatter plot of the average forecasts with a time horizon of 0-24 hours vs those of 24-48 hours.	47

5.11	Time series from 12/3/2018 to 16/4/2018 in which the HRES forecast and the average forecast of EPS with a time maturity of 0-24 hours were superimposed.	49
5.12	Scatter plot of the HRES forecasts vs average forecasts EPS.	50
5.13	Seasonal trend of the average wind speed observation in 2017 and 2018.	51
5.14	Seasonal trend of the average power product in 2017 and 2018.	52
5.15	Seasonal trend of the average wind speed forecast in 2017 and 2018.	53
5.16	Seasonal trend of the average variance of EPS forecast in 2017 and 2018.	53
5.17	Daily trend of the wind speed observed in 2017 and 2018.	54
5.18	Daily trend of the average power product in 2017 and 2018.	55
5.19	Daily trend of the average wind speed forecast in 2017 and 2018.	56
5.20	Daily trend of the variance of EPS forecasts in 2017 and 2018.	56
5.21	Mountain range near the Marsical wind farm.	57
5.22	Wind rose.	58
5.23	Distribution of observed wind directions.	58
5.24	Distribution of the forecast wind directions.	59
5.25	Comparison between observed and EPS and HRES forecasts wind speed.	61
5.26	Scatter plot: observed vs EPS and HRES forecasts wind velocity.	61
5.27	Probability density of wind velocity observed and forecast.	62
5.28	Scatter plot: observed vs EPS and HRES forecasts wind velocity.	64
5.29	Trend over time of the RMSE, MAE, NBias, Bias indices.	64
5.30	Trend over time of the RMSE, MAE, NBias, Bias indices.	66
5.31	Trend over time of the correlation index.	66
5.32	Transfer function with measured power and forecast wind speed.	67
6.1	Data fitted with quantile regression and linear regression (red line).	70
6.2	Density function probability of observed wind speed for different intervals of wind speed forecast.	73
6.3	Probability density in semilogarithmic scale (y axis) of the observations of the conditioned wind speed to: observed wind from hour 10 to 15, with a forecast speed between 1.5 and 2.5 m/s and with a variance of the EPS as between 0.3 and 0.7.	74
6.4	The probability density function on y axis of Gamma distribution for x variable for different parameters.	75

6.5	Scatter plot of forecasts before and after correction vs observation. The forecast are the EPS mean with a forecast horizon of 24-48 hours. The correction used is the quantile regression.	79
6.6	Time series of the forecasts before and after the correction from 12/3/2018 to 16/4/2018. The forecasts are the EPS mean with a forecast horizon of 24-48 hours. The correction used is the quantile regression.	80
6.7	Scatter plot: comparison between observations and forecasts 24-48 of the average of the EPS for the 4 main directions.	82
6.8	Scatter plot of forecasts (mean EPS, 24-48 hours) vs observations before and after the correction, influencing the directions.	84
6.9	Time series of the forecasts before and after the correction from 12/3/2018 to 16/4/2018. The forecasts are the EPS mean with a forecast horizon of 24-48 hours. The correction used is the quantile regression with two techniques: using all the dataset (tt), conditioning on the directions(dr).	84
6.10	Scatter plot of forecasts (mean EPS, 24-48 hours) vs observations for the different hours of the day (0, 3, 6, 9, 12, 15, 18, 21 and 23 hour)	86
6.11	Comparison between MAE and Bias before and after correction, depending on the time of day. They are shown how they change as the hour changes.	87
6.12	Time series of the forecasts before and after the correction from 12/3/2018 to 16/4/2018. The forecasts are the EPS mean with a forecast horizon of 24-48 hours. The correction used is the quantile regression with two techniques: conditioning on the directions(dr), conditioning on the directions and on hours(drdH).	89
6.13	Scatter plot of forecasts (average EPS 24-48 hours) vs observations with 2 different best fits (in red). On the left a best fit dividing the dataset into 3 speed ranges while on the right dividing by 12 speed ranges.	90
6.14	Time series of the forecasts before and after the correction from 12/3/2018 to 16/4/2018. The forecasts are the EPS mean with a forecast horizon of 24-48 hours. The correction used is the quantile regression with two techniques: conditioning on the directions and on hours(drh), conditioning on the directions, hours and velocity(drdHdv).	93

6.15	Trend of the correlation between the difference between forecasts and observations at 7 am with those in the 48 hours of forecasting. This with the following predictions: EPS mean 0-48 h raw and EPS mean 0-48 h corrected by conditioning on direction, time and speed (drdHdv).	96
6.16	Time series of the forecasts before and after the correction from 12/3/2018 to 16/4/2018. The forecasts are the EPS mean with a forecast horizon of 8-24 hours. The correction used is the quantile regression with two techniques: conditioning on the directions, hours and velocity(drdHdv) and using the error correlation after conditioning on the directions, hours and velocity(drdHdv-df). The last forecast is only for 8 to 24 hour.	98
6.17	How the NMAE, Pearson and HH indices vary as the floating window used varies. On the ordinate there is the value of the indices while on the abscissa there is the heat of the number of hours considered to make the moving average. The moving average was made on the forecast corrected with the best available correction (drdHdv-df) on the EPS mean 24-48 h forecast.	100
6.18	Time series of the forecasts before and after the correction from 12/3/2018 to 16/4/2018. The forecasts are the EPS mean with a forecast horizon of 24-48 hours. The correction used is the quantile regression with two techniques: conditioning for the directions, hours and velocity and using the correlation of the error(drdHdv-df) and conditioning for the directions, hours, velocity, using the correlation of the error and making a moving average(drdHdv-df-mv)	102
6.19	How the NMAE, Pearson and HH indices vary as the number of members used varies. On the ordinate there is the value of the indices while on the abscissa there is the number of members used. The moving average was made on the forecast corrected with the correction conditioning on direction, time and speed(drdHdv) on the EPS mean 24-48 h forecast.	105

6.20	Time series of the forecasts before and after the correction from 12/3/2018 to 16/4/2018. The green and red forecasts are the EPS mean with a forecast horizon of 24-48 hours. The correction used is the quantile regression with two techniques: conditioning for the directions, hours, velocity, using the correlation of the error and making a moving average(drdHdv-df-mv) and using the different members of EPS conditioning for the directions, hours, velocity, choosing the best 35 member of EPS and put them together, using the correlation of the error, making a moving average and finally making a quantile regression on all the forecasts(drdHdv-mx-df-mv-tt)	109
6.21	Time series of the forecasts before and after the correction from 12/3/2018 to 16/4/2018. The forecast generated starting from the average of the EPS for the 24-48 hour forecast, correcting by conditioning on direction, time and speed, exploiting the correlation of the error and making the moving average. This correction was done using the CRPS of the gamma distribution. The confidence bands were created considering a probability from the low of 5, 20 and 35% while from the top of 65, 80, 95%.	114
6.22	Zoom of figure 6.21 from 26/3/2018 to 9/4/2018.	114
6.23	Trend of the MAE, NMAE, RMSE and NRMSE indices as a function of the forecast horizon before and after the best correction.	116
6.24	Trend of the correlation index (Pearson) as a function of the forecast horizon before and after the best correction.	117

6.25	Graphical representation of the NMAE, NRMSE and Pearson indices of the forecast in the following order. Raw: Raw previews; tt: Predictions calibrated without any division of the same; dr: Forecasts calibrated by influencing the direction; drdH: Forecasts calibrated by conditioning on direction and time; drdHdv: Forecasts calibrated by conditioning on direction, time and speed; drdHdv-df: Forecasts calibrated by conditioning on direction, time and speed and then exploiting the correlation of the error; drdHdv-df-mv: Calibrated forecasts by conditioning on direction, time and speed, exploiting the correlation of the error and finally making a moving average; drdHdv-mx-df-mv-tt: Calibrated forecasts by conditioning on direction, time and speed, choosing the best spaghetti, exploiting the correlation of the error, making a moving average and finally redoing a calibration without conditioning. The forecasts in particular are those 24-48 hours that have been corrected using quantile regression.	119
6.26	Trend of the NMAE, NRMSE, HH and Pearson index as a function of the number of months used to make the training for the correction drdHdv(depending on direction, time and speed). The predictions were obtained starting from the average of the EPS 24-48 hours and using quantile regression. . . .	121
6.27	Time series of the raw predicted power (red) compared with the predicted power obtained by making the transfer function on the best forecast starting from the members of the EPS (green) and with that obtained by further adding a correction by conditioning on direction and power (black).	124
6.28	The figure shows the table taken from the cited article. There are indices of the error as the technique used varies. The indexes present only the Bias, the RMSE, the NMAE, and the NRMSE. The NMAE and NRMSE were normalized using the nominal power of the plant (Layout 1 6600 [kW], Layout 2 8700 [kW]). (Taken from Ref. [18]	131

List of Tables

5.1	Error index of the original transfer function.	40
5.2	Error index of original transfer function vs cleaned transfer function	44
5.3	Error index of EPS vs HRES forecast.	49
5.4	Mean value of wind speed	60
5.5	Error index of raw forecasts.	63
5.6	Table of the error index as the time limit changes, mean EPS 0-24 h forecast.	65
5.7	Index value of power prediction with raw wind speed forecasts.	67
6.1	Error indices of EPS mean before and after correction for the 0-24 and 24-48 h forecasts, QR represents the correction using quantile regression while GM using the gamma distribution.	78
6.2	Error indices of HRES before and after correction for the 0-24 and 24-48 h forecasts, QR represents the correction using quantile regression while GM using the gamma distribution.	78
6.3	Error indices for raw predictions (EPS average) 24-48 hours for: whole dataset merged and divided into 4 quadrants.	81
6.4	Comparison between the various indices of the error for the raw forecasts (mean EPS, 0-24 and 24-48 hours), corrected without divisions (tt), and corrected by conditioning for the directions.	83
6.5	Comparison between the various indices of the error for the raw forecasts (mean EPS, 0-24 and 24-48 hours), corrected without divisions (tt), and corrected by conditioning for the day hours.	85
6.6	Comparison between the various indices of the error for the raw forecasts (mean EPS, 0-24 and 24-48 hours), corrected by conditioning for the directions(dr), and corrected by conditioning for the directions and for the hours.	88

6.7	Error indices values, starting from mean EPS 24-48 hour forecast, of the raw predictions, corrected for the entire dataset (tt), corrected by conditioning for 3 speed intervals (dv3) and finally for 12 intervals (dv12).	91
6.8	Comparison between the various indices of the error for the raw forecasts (mean EPS, 0-24 and 24-48 hours), corrected by conditioning for the directions and hours(drdH), and corrected by conditioning for the directions, hours and velocity (drdHdv).	92
6.9	Comparison between the forecasts from 0 to 24 with the forecasts only from 8 to 24.	95
6.10	Comparison between the various indices of the error for the raw forecasts (mean EPS, 0-24 and 24-48 hours), corrected by conditioning for the directions, hours and velocity (drdHdv), and corrected by conditioning for the directions, hours, velocity and using the correlation of the error(drdHdv-df).	97
6.11	Comparison to see if it is better to do the correction of the error correlation before (df-drdHdv) or after the correction with the conditioning on direction, time and speed (drdHdv-df). All this starting from the forecast of the average EPS for 8-24 and 24-48 hours.	97
6.12	Comparison between the various indices of the error for the raw forecasts (mean EPS, 8-24 and 24-48 hours), corrected by conditioning for the directions, hours and velocity and using the correlation of the error(drdHdv-df), and corrected by conditioning for the directions, hours, velocity, using the correlation of the error and making a moving average(drdHdv-df-mv).	101
6.13	Comparison between the various indices of the error for the raw forecasts (mean EPS, 24-48 hours), corrected with the best correction(drdHdv-df-mv) and then later replacing negative values with their opposite (sub. opp.) or with a 0(sub. 0).	103
6.14	Comparison between the error indices calculated for the following cases: raw case with the average of EPS (8-24 and 24-48 hours) and the other two cases which both start from the best "static" correction (drdHdv, conditioning on directions, time and speed) in which a prediction is obtained using the correlation of the error (df) while the second one choosing the best 35 members and making the weighted average (mx).	106

6.15	Comparison between the best forecasts corrected starting from the average of the EPS (drdHdv-df-mv) with those correcting the single members and then after choosing the best ones by correcting them further thanks to the correlation of the error(drdHdv-mx-df) and then with the moving average(drdHdv-mx-df-mv).	107
6.16	Comparison between the best forecasts corrected starting from the average of the EPS (drdHdv-df-mv) with those correcting the single members and then after choosing the best ones by correcting them further thanks to the correlation of the error and with the moving average(drdHdv-mx-df-mv) and then making a quantile regression on all the forecasts(drdHdv-mx-df-mv-tt).	108
6.17	Comparison of the error indices between the best forecast starting from the average of EPS, the best forecast starting from the EPS and the best forecast starting from the HRES. All this done with quantile regressions.	111
6.18	Comparison of the error indices for the corrected predictions with the two best strategies (one starting from the average of the EPS and the other considering each single member) using the CRPS method for the gamma distribution.	113
6.19	Error indices varying the number of months used to train for the drdHdv strategy (depending on direction, time and speed). The predictions were obtained starting from the average of the EPS 24-48 hours and using quantile regression.	121
6.20	Comparison between the raw power prediction and the best prediction obtained starting from the average of the EPS and the individual members of the EPS using quantile regression.	123
6.21	Comparison between the raw power prediction and the best prediction obtained starting from the average of the EPS and the individual members of the EPS using quantile regression and making a further correction on the expected power by conditioning on direction and power.	123
6.22	Comparison between the best forecast obtained starting from the average of the EPS with the strategies shown previously with the Machine learning strategy.	126
6.23	Comparison between the raw power prediction of the EPS average with the best generated by the EPS average with the best one of machine learning.	127

6.24	Comparison between the best available wind forecast (drdHdv-mx-df-mv-tt) with the forecast obtained using machine learning in cascade at the best forecast	128
6.25	Comparison between the best prediction of the power obtained starting from the average of the EPS with that obtained by making a cascade Machine learning.	128
6.26	Value of the NMAE, NRMSE indices of the power forecast normalized using the nominal power of the plant (2000 kW). The forecast is the best obtained so far for a 24-48 hour forecast horizon.	132

Abstract

Over the past decades, wind energy has become an important power source. In 2018, wind energy accounted for 14% of the installed capacity in Europe. However, the volatility of wind and therefore the fluctuating energy supply complicates an integration of wind power into the power grid. To keep the energy demand and supply balanced, accurate wind power forecasts are important for energy traders, producers, and distributors.

Usually, wind power forecasts for lead times longer than 6 h are based on numerical weather prediction (NWP) models. NWP models differ widely in their model formulations, space and time resolutions, and parameterizations. Global models provide global forecasts with a limited spatial resolution. Limited-area models use the output of global models as boundary conditions and usually have higher resolutions. Furthermore, to estimate forecast errors arising from errors in the initial state or parametrization, many weather services also provide an ensemble of forecasts, with perturbed initial conditions and/or different model formulations.

During the past decade, the use of forecast ensembles for assessing the uncertainty of numerical weather predictions has become routine. Three operational methods for the generation of synoptic-scale ensembles have been developed: the breeding growing modes method used by the National Centers for Environmental Prediction (NCEP), the singular vector method used by the European Centre for Medium-Range Weather Forecasts (ECMWF), and the perturbed observations method used by the Canadian Meteorological Centre (CMC). In this thesis the forecasts of the European center will be used in particular.

Ensemble prediction systems usually has positive spread-error correlation, but are also subject to forecast bias and dispersion errors, and are therefore uncalibrated. With this study it will be shown the use of ensemble model output statistics (EMOS), an easy-to-implement postprocessing technique that addresses both forecast bias and underdispersion and takes into account the spread-skill relationship. The technique is based on multiple linear regression and is akin to the superensemble approach that has traditionally

been used for deterministic-style forecasts. The EMOS predictive mean is a bias-corrected weighted average of the ensemble member forecasts, with coefficients that can be interpreted in terms of the relative contributions of the member models to the ensemble, and provides a highly competitive deterministic-style forecast. The EMOS predictive variance is a linear function of the ensemble variance. For fitting the EMOS coefficients two different methods will be used. One is the method of minimum continuous ranked probability score (CRPS) estimation. This technique finds the coefficient values that optimize the CRPS for the training data thus providing an output as well as a forecast of the wind speed also its variance. The other method is to apply quantile regression. This regression consists in identifying the values of the coefficients that identify the median of the probability distribution of the training data. In this way, by definition, the absolute average value of the forecast error is minimized, which is also the most used index to be able to make an economic estimate of the damage of the forecast error. With this method, however, it is not possible to exploit the variance of the ensembles. Statistical relationships between NWP forecasts and corresponding observations are formulated from past data and then used to correct future forecasts. To get stable estimates of these relationships, the dataset of past forecasts and observations should be as long as possible. However, frequent model enhancements like increased resolution or improved parameterizations usually change the statistical properties of NWP model output. This limits the usable length of the dataset of past forecasts for commonly used statistical models. Reforecasts (or hindcasts) overcome this problem.

This work presents an application of the European Centre for Medium-Range Weather Forecasts (ECMWF) Ensemble Prediction System (EPS) to produce short-term probabilistic wind power forecasts. The ECMWF's EPS is based on running a meteorological model multiple times, starting from slightly perturbed initial conditions. The distribution of these different runs allows estimating the prediction uncertainty.

The study carried out derives from the collaboration between the University of Genoa and the company Ego S.r.l.. The analyzed case is a wind farm managed by the company located in Marsica, in Abruzzo, in an area where due to the complex orography it is particularly difficult to make wind forecasts. It were used about two years of data. Forecasts will be made with a maximum forecast horizon of 48 hours.

The thesis is structured as follows. In the first chapter there is a description of how the Italian electricity market works, to further understand the importance of having a good forecast of wind speed and to also be able to justify certain choices that will be made in the course of the treaty. Mainly the

need to have good forecasts for the company derives from the fact that the electricity market system is based on an offer made the day before for which it is necessary to have the best possible forecast. In the second chapter it will be describe the European Center for Medium-Range Weather Forecasts (ECMWF). It will therefore be described on which equations their models are based, on the resolution they have and on how they take into account the physical parameters involved. In the third chapter it will be explained how the ECMWF generates the Ensemble Prediction System (EPS) and also the forecast called HRES which is a unique forecast created at a higher definition. Will be described specifically the EPS forecast by showing how the perturbations of the initial condition and the clustering operations occur. The fourth chapter will show the error indices that will be used to check the quality of the results and of the raw data. The physical meanings of each of them will therefore be shown to better understand in the following chapters how predictions change. With the fifth chapter will be made an accurate description of the available wind observations taken from the anemometer of a turbine of the Marsica wind farm and of the wind forecasts of that site. The process that led to the cleaning of the input data from incorrect measurements will be analyzed, the main parameters will be seen that will then be used for the calibration of the forecasts and finally the quality of the input data will be shown as a function of the error indices defined in the previous chapter.

Finally in the sixth and last chapter the techniques used to calibrate the forecasts and the results deriving from them will be shown. A comparison and evaluation of the various techniques implemented (quantile regression and CRPS) will be made to identify the one deemed to be the best. All the available parameters introduced in the previous chapter will be exploited. In particular, it will use the wind direction, the hours of the day, the wind speed, the forecast error at a given time, the different EPS members and also the moving average. The power forecast will then be shown after calibration of the wind speed forecast. It will be discuss how much the number of data available for training influences. A comparison will also be made with a Machine Learning strategy and with a study regarding a site similar to that of this thesis in order to further evaluate the goodness of the results.

Chapter 1

The Italian Electricity Market

The “power exchange” is an organized system of offers, sales and purchases of electricity. The power exchange, provided for by legislative decree no. 79/1999 for the liberalization of the electricity market, was established in Italy starting from April 1, 2004 and is managed by the Electricity Market Operator, which later became the Energy Market Operator in November 2009. The sale of electricity is carried out every day for the following day by resorting to bargaining on an hourly basis where the meeting between supply and demand is carried out through the marginal price system.[1]

This mechanism remunerates producers by paying everyone the equilibrium price between supply and demand, which is equal to the price of the most expensive offer among those accepted to satisfy demand.[2]

1.1 The institutional subjects of the Electricity Market

In order to better understand the structure and nature of the Electricity Market, it is essential to identify which are the main subjects that make up this market:

- Ministry of Economic Development (MSE): which defines the strategic and operational guidelines to ensure the safety and economy of the national electricity system.
- Authority for Electricity, Gas and Water System (AEEGSI): this is an independent public body. The main function is the control and regulation of the national energy system. This takes place through the determination of tariff prices and the general rules of the electricity market, the quality control of the activities and services provided by

the operators and, finally, the supervision of whether or not their own rules are fulfilled.

- Energy Services Manager (GSE): is the public holding that supports the development of renewable sources by managing and providing the related incentive mechanisms. The company operates following the strategic lines defined by the MSE. GSE is the parent company of the three subsidiaries AU, GME and RSE.
- Electricity market operator (GME): which organizes and manages the energy market, according to criteria of neutrality, transparency, objectivity, as well as competition between producers. Furthermore, it must maintain a certain degree of transparency in the operations and choices made, as it has the duty not only to encourage competitiveness among market operators, but also to guarantee a determined and constant availability of power reserves.

1.2 The Electricity Market structure

The electricity market should be as crucial and instantaneous as possible. In fact, the value of energy it varies from node to node due to congestion and varies every moment. However, it is not possible to build a market that describes reality perfectly. Therefore, it is necessary introduce standardizations, both from a temporal and a spatial point of view. From a time point of view, one hour is generally assumed as the reference period.

Standardization of traded products brings benefits if liquidity increases markets and reduces transaction costs. However, it can generate new systemic costs if too much broad because it would no longer be representative of reality and would induce higher costs to guarantee the real-time balance between supply and demand of electricity.

The Electricity Market organized and managed by GME, aimed at programming the production and consumption units, is divided, figure 1.1, into the Spot Electricity Market (MPE), into the Forward Electricity Market with mandatory delivery and collection (MTE) and in the Platform for the physical delivery of financial contracts concluded on IDEX1.

1.3 The Electricity Market structure

The electricity market should be as crucial and instantaneous as possible. In fact, the value of energy it varies from node to node due to congestion

and varies every moment. However, it is not possible to build a market that describes reality perfectly. Therefore, it is necessary to introduce standardizations, both from a temporal and a spatial point of view. From a time point of view, one hour is generally assumed as the reference period.

Standardization of traded products brings benefits if liquidity increases markets and reduces transaction costs. However, it can generate new systemic costs if too much broad because it would no longer be representative of reality and would induce higher costs to guarantee the real-time balance between supply and demand of electricity.

The Electricity Market organized and managed by GME, aimed at programming the production and consumption units, is divided, figure 1.1, into the Spot Electricity Market (MPE), into the Forward Electricity Market with mandatory delivery and collection (MTE) and in the Platform for the physical delivery of financial contracts concluded on IDEX1.

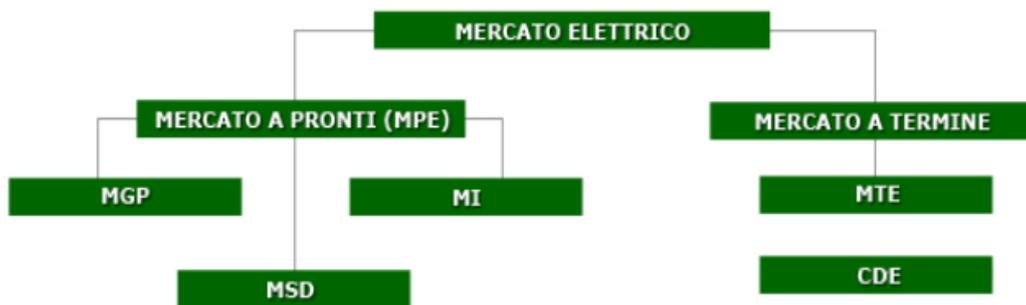


Figure 1.1: The Electricity Market structure. (Taken from [1])

1.3.1 The System marginal price (SMP)

The System marginal price (SMP) is a uniform price. For each relevant interval, the offers accepted are valued at the equilibrium price of the system, equal to the value of the last accepted offer (offer marginal). Each operator for sale obtains a revenue equal to the price of the marginal offer multiplied by the total volume of energy sold on the market. With this system an operator, in a competitive market, is induced to offer at his own cost variable since the valuation of the accepted volumes will in any case be carried out based on the price marginal of the system (the most expensive last accepted offer). The inframarginal production units will receive a remuneration higher than their own variable costs which will make it possible to cover fixed costs.

1.3.2 Nodal and zonal markets

Some markets refer to the network node, others to the market area (i.e. to an area plus extended internally), others to the entire nation:

- Nodal markets: presence of frequent and/or economically significant differences in value of electricity between the different nodes of the network
- Zonal markets: presence of large groups of nodes (zones) within which there are no differences in the value of electricity

The more the geographic extension is broad, the more it is possible that there are network constraints within it that cause electricity does not have the same value in all geographical points. In the event that markets are defined. No subsequent interventions are necessary aimed at making the programs operable in compliance with the network constraints (the so-called re-dispatching action which instead characterizes the markets is not necessary zonal).

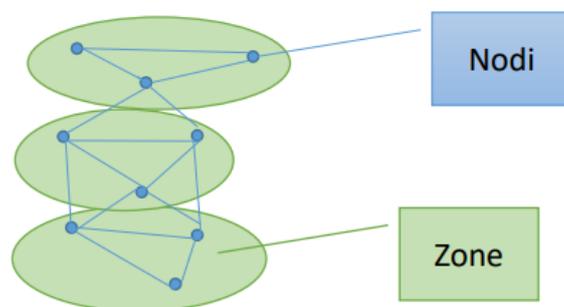


Figure 1.2: Graphic representation of the difference between node and zone. (Taken from [1])

1.3.3 The The Day Ahead Market (MGP)

The hourly schedule for the injection and/or withdrawal of electricity is negotiated on the MGP. The operators participate by submitting offers in which they indicate the quantity and the price maximum/minimum at which they are willing to buy/sell. The session of the MGP opens at 8.00 on the ninth day before the day of delivery and closes at 12.00 on the day before the day

of delivery.

Offers are accepted after the close of the market session, on the basis of economic merit and in compliance with the transit limits between the zones. MGP is therefore an auction market and not a market continuous bargaining. Accepted offers to sell are valued at the clearing price of the area to which they belong (hourly zone price). This price is determined, for each hour, by the intersection of the curve of supply and demand and differs from area to area in the presence of saturated transit limits (system marginal price). Accepted purchase offers are valued at the single national price (PUN), equal to the average hourly zonal prices weighted for zonal consumption. GME acts as a central counterparty.

1.3.4 The Intraday Market (MI)

The Intraday Market (MI) allows operators to make changes to the programs defined in the MGP through further offers to buy or sell. Here too the object of negotiation is the hourly schedule for the injection and / or withdrawal of electricity. It takes place in seven sessions: the first session takes place after the closing of the MGP, opens at 12.55 of the day before the day of delivery and closes at 15.00 of the same day. The last session opens at 5.30 pm the day before the delivery day e closes at 3.45 pm on the day of delivery.

The purchase and sale offers are selected on the basis of the same criteria described for MGP (system marginal price). Unlike MGP, accepted purchase offers are also valued at the zonal price schedule. GME acts as a central counterparty.

1.3.5 Dispatching of electricity

Electricity cannot be stored. It is therefore necessary to produce the quantity, instant by instant of energy required by all consumers and manage its transmission so that the offer and demand are always in balance, thus ensuring the continuity and security of supply of the service.

The coordinated management of the injections and withdrawals of electricity and the flows of electricity on the transmission grid for the purpose of maintaining the balance of the electricity system in conditions of security is what is called a dispatching service. This service is provided by Terna under conditions defined by the Authority.

The Market for the Dispatching Service (MSD) is the tool through which Terna is supplies the resources necessary for the management and control of the system (resolution of intrazonal congestion, creation of the energy reserve, balancing in real time). On the MSD Terna acts as a central coun-

terparty and accepted offers are remunerated at the price introduced. The MSD is divided into the programming phase (ex-ante MSD) and the Balancing Market (MB). The MSD ex-ante and MB takes place in several sessions.

Chapter 2

The European Centre for Medium-Range Weather Forecasts (ECMWF)

Weather observations are neither perfect nor complete. Also, because of limitations in computer power, the models inevitably approximate the exact equations for weather. Hence every single forecast is, to some extent, uncertain. But how much uncertain? Uncertainty will vary from day to day, depending on the atmospheric conditions at the start of the forecast. When the state of the atmosphere is such that forecasts are not very sensitive to uncertainties in the initial conditions, the forecasts can be made with confidence many days ahead. However, when the forecasts are particularly sensitive to the initial conditions, forecasts can be uncertain almost from the beginning. Is there a way to know beforehand whether a forecast is going to be accurate or not? The European Centre for Medium-Range Weather Forecasts (ECMWF) has pioneered a system to predict forecast confidence. This system, operational at ECMWF since 1992, is the Ensemble Prediction System (EPS).^{[3][4][5]}

2.1 The European Centre for Medium-Range Weather Forecasts – an historical back- ground

The European Centre for Medium Range Weather Forecasts (ECMWF) is the consequence of 100 years of development in dynamic and synoptic meteorology, and fifty years of development in numerical weather prediction

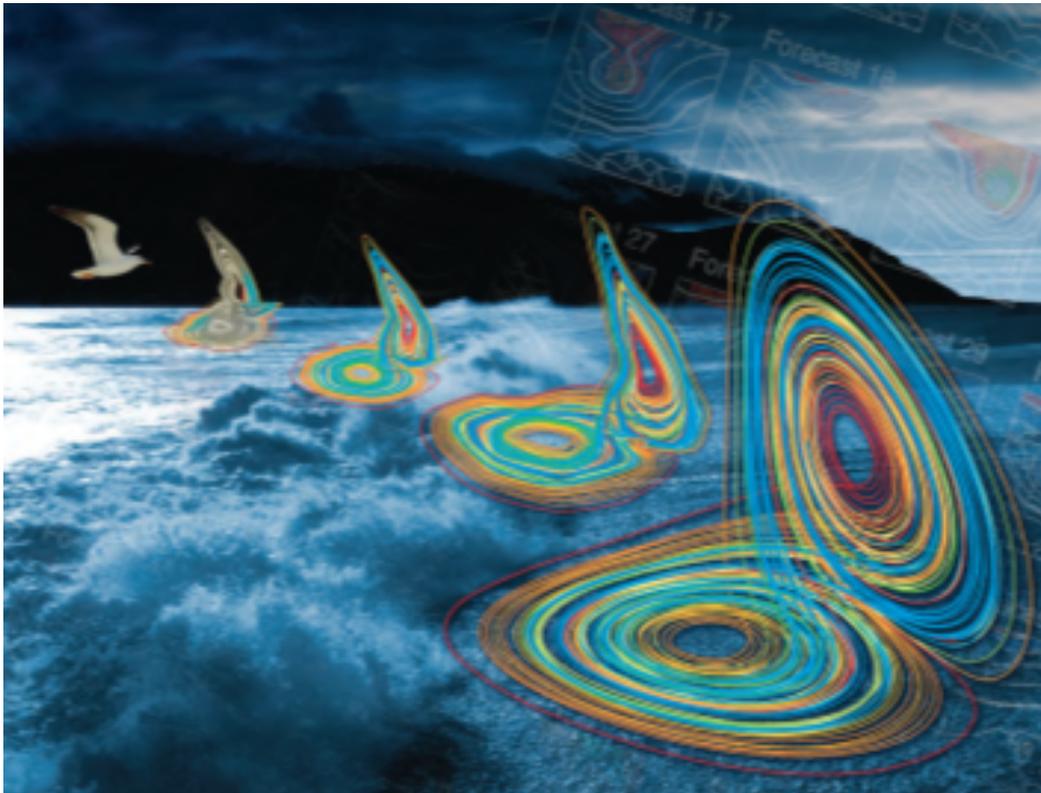


Figure 2.1: Artistic sketch of the Ensemble Prediction System.(Take from Ref. [3])

(NWP).

2.1.1 The creation of ECMWF

From the experience gathered with short-range and climatological simulations, there was, in the late 60's, enough know-how to motivate an attack on the medium-range forecast problem, defined as the interval from 3 to 10 days ahead. The scientific and technical problems were still formidable, and only few countries had enough expertise to tackle them. This made medium-range forecasting an ideal candidate for multi-national co-operation.

In October 1967, the Council of Ministers of the European Communities adopted a resolution to implement a programme to promote joint scientific and technical research. A proposal for a "Meteorological Computer Centre for Research and Operation" occupied the first place on a list of meteorological projects submitted by an expert group in April 1969. The proposal was accepted, and other European nations were invited to participate. In April

1970, an expanded expert group initiated two study groups to investigate the economic and scientific motivations for the project.

The reports from the two groups were completed in August 1971, and at the conference of ministers in the same year it was decided to create the European Centre for Medium–Range Weather Forecasts. The ambition, laid out in the plans, was to produce forecasts ten days ahead with the five–day forecasts having the same accuracy as subjective two–day forecasts in the 50’s.

The ECMWF convention was signed in October 1973. Seventeen European States are currently members: Belgium, Denmark, Germany, Spain, France, Greece, Ireland, Italy, the Netherlands, Norway, Austria, Portugal, Switzerland, Finland, Sweden, Turkey and the United Kingdom. The objectives of the Centre were laid down as follows:

- To develop dynamic models of the atmosphere with a view to preparing medium–range weather forecasts by means of numerical methods;
- To prepare, on a regular basis, the data necessary for the production of medium–range weather forecasts;
- To carry out scientific and technical research directed towards the improvement of these forecasts;
- To collect and store appropriate meteorological data;
- To make available to the meteorological offices of the Member States, in the most appropriate form, the results of the studies and research provided for in the first and third objectives above and the data referred to in the second and fourth objectives;
- To make available a sufficient proportion of its computing capacity to the meteorological offices of the Member States for their research, priority being given to the field of numerical forecasting. The allocation of the proportions would be determined by Council;
- To assist in implementing the programmes of the World Meteorological Organization;
- To assist in advanced training for the scientific staff of the meteorological offices of the Member States in the field of numerical weather forecasting.

2.1.2 The ECMWF forecasting model since 1979- an overview

The ECMWF forecasting system consists of three components: a general circulation model (coupled with an ocean wave model), a data assimilation system and, since 1992, an ensemble forecast system.

The first ECMWF numerical model was a grid-point model with 15 levels in the vertical up to 10 hPa. In April 1983, this grid-point model was replaced by a model with a spectral representation in the horizontal with a triangular truncation at wavenumber 63. The spectral technique was more accurate than the grid point model for the same computational cost.

In September 1991, a high-resolution spectral model was put into operations, where the spectral truncation was extended to wavenumber 213 and the number of levels increased to 31. The model used a computational grid with a resolution of about 60 km.

Until 1995 the ECMWF model did not contain any explicit clouds, only interpretations from other fields like relative humidity, precipitation, vertical motion, and vertical temperature gradients. A new cloud scheme was introduced in April 1995 with clouds as prognostic parameters, defined through the cloud fraction and the content of cloud liquid water and cloud ice.

Up to 1996 the analysis system was based on optimum interpolation. That year it was replaced by a three-dimensional variational system (3DVAR), which was upgraded to a four-dimensional variational system (4DVAR) in 1997.

In 1992 the ECMWF started its Ensemble Prediction System. In autumn 1996 the number of members was extended from 32 to 50 members and the model was upgraded from T63 to TL159, in autumn 2000 to TL255. The vertical resolution was increased from 31 to 40 levels in 1999. Crude allowance for the uncertainty of physical processes was made in autumn 1998 with the introduction of stochastic physics.

2.2 The ECMWF global atmospheric model

The ECMWF general circulation model, TL511L60, consists of a dynamical component, a physical component and a coupled ocean wave component.

2.2.1 The model equations

The model formulation can be summarized by six basic physical equations, the resolution in time and space and the way the numerical computations

are carried out.

Of the six equations governing the ECMWF primitive equation atmospheric model, two are diagnostic and tell us about the static relation between different parameters:

- The **gas law** gives the relation between pressure, density and temperature.
- The **hydrostatic equation** shows the relationship between the density of the air and the change of pressure with height. The other four equations are prognostic and describe the changes with time of the horizontal wind components, temperature and water vapour content of an air parcel, and of the surface pressure.
- The **equation of continuity** expresses the mass conservation and determines the vertical velocity and change in the surface pressure.
- The **equation of motion** describes how the momentum of an air parcel changes due to the pressure gradient and the Coriolis force. Included are also the effects of turbulent drag and gravity wave breaking
- The **thermodynamic equation** expresses how a change in an air parcel temperature is brought about by adiabatic cooling or warming due to vertical displacements. Other physical processes like condensation, evaporation, turbulent transport and radiative effects are also included
- The **conservation equation for moisture** assumes that the moisture content of an air parcel is constant, except for losses due to precipitation and condensation or gains by evaporation from clouds and rain or from the oceans and continents. Adding to this there are specific prognostic equations for the cloud fraction, water, ice content and ozone.

2.2.2 The resolution in time and space

The present system uses a temporal resolution of 15 minutes. The computational time step has to be chosen with care in order to avoid numerical instabilities and ensure enough accuracy. The vertical resolution (measured in geometric height) is highest in the planetary boundary layer and lowest in the stratosphere and lower mesosphere.

The atmosphere is divided into 60 layers up to 0.1 hPa (about 64 km). These so-called s-levels which follow the earth's surface in the lower and

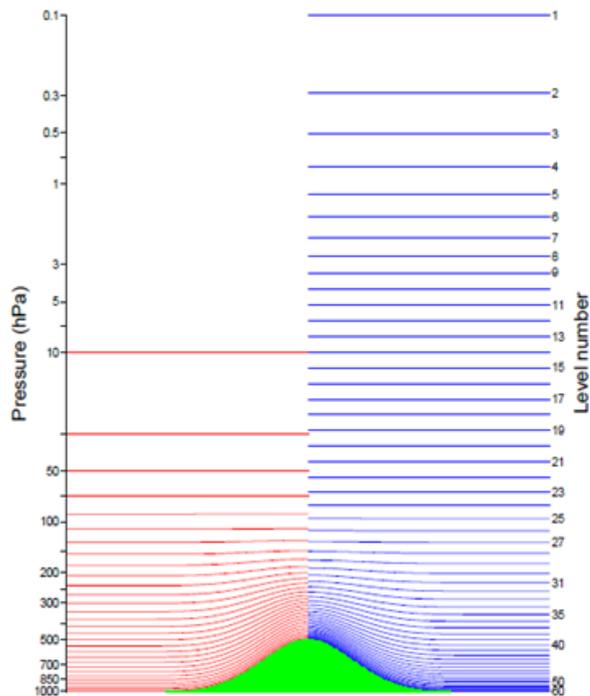


Figure 2.2: To the left the vertical resolution before 1999 (31 level to 10 hPa, and to the right after 1999 (60 levels to 0.1 hPa).(Take from Ref. [3])

mid-troposphere are used as vertical coordinates but are surfaces of constant pressure in the upper stratosphere and mesosphere. A smooth transition between these types of levels is ensured. For its horizontal resolution the ECMWF model uses two different numerical representations:

- A spectral method, based on a spherical harmonic expansion, truncated at total wave number 511, for the representation of upper air-fields and the computation of the horizontal derivatives. Apart from the operational TL511L60 model (511 spectral components and 60 levels), a TL255L40 is run for ensemble predictions (only up to 10 hPa), a TL159L40 for the 4DVAR assimilations and T63L31 for seasonal forecasts.
- In addition, there is a grid point representation used for computing dynamic tendencies and the diabatic physical parametrization. This so-called Gaussian grid, is regular in longitude and almost regular in latitude. Due to the convergence of the longitudes toward the poles, the east-west distance between the grid points decreases poleward.

The model surface is logically divided into sea and land points, by using a land–sea mask. A grid point is defined as a land point if more than 50% of the actual surface of the grid-box is land. With a TL511 resolution, islands like Corsica, Crete and Cyprus are represented by around five land grid points, Mallorca and Gotland by only two. The Faeroe Islands, the Shetland Island and Rhodos are not represented by any land point.

2.2.3 The numerical formulation

The choice of a semi-Lagrangian numerical scheme instead of an Eulerian is the result of partly the need to save computer time and speed up the forecast. The basic difference between an Eulerian and a Lagrangian formulation can be seen from the advection equation (in a one-dimensional space):

$$\frac{dQ}{dt} = \frac{\partial Q}{\partial t} + U \frac{\partial Q}{\partial x} = 0 \quad (2.1)$$

which in an Eulerian way expresses that the local changes in Q are due to the advection of Q by the wind U :

$$\frac{\partial Q}{\partial t} = -U \frac{\partial Q}{\partial x} \quad (2.2)$$

or in a Lagrangian way that Q is conserved for any fluid parcels:

$$\frac{dQ}{dt} = 0 \quad (2.3)$$

In a pure Lagrangian framework (following a set of marked fluid parcels) shear and stretching deformations tend to concentrate parcels inhomogeneously, so that it is difficult to maintain uniform resolution over the forecast region. A semi-Lagrangian scheme is used to overcome this difficulty. In this version, the grid points are stationary and at each time step the scheme computes a backward trajectory from every grid point. The point reached defines where the air parcel was at the beginning of the time step. The interpolated value of the variable in that point is then carried forward to the grid point, applying the various physical processes.

Whereas all Eulerian schemes require small time steps to avoid numerical instability, (the quantity Q must not be advected by more than one grid length per time step), the semi-Lagrangian scheme allows longer time steps. The limitation for stability is that the trajectories do not cross.

2.3 Parametrization of physical processes

The primary function of the forecast weather parameters in the ECMWF model, lies in their impact on the overall atmospheric flow. A ten-day integration makes it absolutely necessary to include effects with relatively long-time scale, even as subtle as the evaporation by vegetation, in order to handle the flow pattern more accurately. The different time scales and feed-back mechanisms between the various processes makes the computations extremely complex and expensive.

2.3.1 The model orography

The representation of the orography uses the mean orography and four additional fields describing the standard deviation, orientation, anisotropy and slope of the sub-grid orography. This takes into account some of the orographic variability but does not change the fact that for the usefulness of the weather parameters, the model orography is still significantly smoother than reality.

However, the parametrization allows a realistic representation of the mountain drag, which is important for the creation of large-scale atmospheric eddies. A novel and important part of the scheme is that, depending on dynamical criteria, it can block the low level flow rather than make the air go over the orography.

2.3.2 The Planetary Boundary Layer

The treatment of the Planetary Boundary Layer (PBL), plays a fundamental role for the whole atmosphere-earth system. It is through the surface exchanges of momentum, heat and moisture that the atmosphere “feel” that it moves over a rough land surface or a wet smooth sea.

The lowest 13 levels are at around 10, 30, 60, 100, 160, 240, 340, 460, 600, 760, 950, 1170 and 1400 m above the model surface. Even with this fairly high resolution the vertical gradients of temperature, wind, moisture etc. in the PBL cannot be described very accurately, and therefore it is even worse for the turbulent transports of momentum, heat and moisture. For the estimation of these parameters the model uses the larger scale variables such as wind, temperature and specific humidity, with the assumption that the transports are proportional to the vertical gradients.

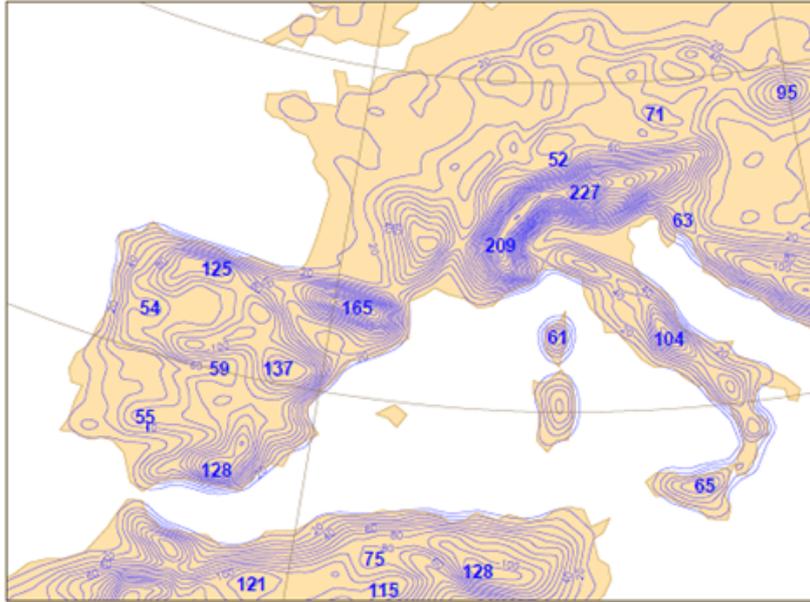


Figure 2.3: The model height (in deksmeterd) for southwestern Europe.(Taken from Ref. [3])

2.3.3 Radiation

In view of the importance of cloud–radiation interaction in both long- and short-term processes, ECMWF has placed high emphasis on the treatment of the absorption and scattering by clouds of solar and terrestrial radiation. About 15 percent of the overall computational time is devoted to the radiation scheme.

The radiation spectrum is divided into eight frequency bands: two in the short wave spectrum (direct from the sun and diffuse radiation), and 15 in the long wave spectrum (from the earth and within the atmosphere). The upward and downward diffused radiation is computed for each of the 16 spectral bands. The parameters influencing the emission and absorption are pressure, temperature, moisture, cloud cover and cloud water content, and carbon dioxide, ozone, methane, nitrous oxide, CFC–11 and CFC–12.

The radiation scheme is designed to take the cloud–radiation interactions into account in considerable detail. It allows partial cloud cover in any layer of the model.

2.3.4 Clouds

The main purpose of the cloud scheme is to provide input to the radiation computations and to calculate precipitation. The clouds are generated by large-scale ascent, cumulus convection, boundary layer turbulence and radiative cooling. They are dissipated through evaporation due to large-scale descent, cumulus induced subsidence, radiative heating and turbulence at both cloud tops and sides, as through precipitation processes.

The cloud scheme is unique in treating the main cloud-related processes in a consistent way by forecasting both cloud fraction and cloud water/ice content with their own prognostic equations. In the scheme the cloud processes are strongly coupled to other parametrized processes:

- **Convective clouds** are computed in parallel with the convective scheme
- **Deep convection** Deep convection predominantly occurs in disturbed situations with a deep layer of conditional instability and large-scale moisture convergence
- **Shallow convection** predominantly occurs in undisturbed flow, in the absence of large-scale convergent flow.
- **Mid-level convection** describes convective cells which originate at levels above the boundary layer
- **Stratocumulus clouds** are linked to the boundary layer moisture flux produced by the vertical diffusion scheme
- **Stratiform clouds** are determined by the rate at which the saturation specific humidity decreases due to upward vertical motion and radiative cooling.
- **Evaporation processes** in connection with clouds are accounted for in several ways: large-scale and cumulus-induced subsidence and radiative heating, evaporation at the cloud sides due to turbulent processes and turbulent motion at the cloud tops.

2.3.5 The hydrological cycle

Precipitation processes do not only take into account the local water/ice content, but also different precipitation enhancement processes. The effect of evaporation of falling precipitation is also included. Two mechanisms to generate precipitation are included in the ECMWF model, for convective and for stratiform (frontal or dynamical) precipitation:

- **Convective precipitation:** the condensate formed in the updrafts of the convection parametrization is water above 0°C , ice below -23°C and a mixture of the two in between. If the amount of condensate formed exceeds the value that can be sustained by the vertical velocity, precipitation is formed in the form of snow or water.
- **Stratiform precipitation:** cloud water and ice from the cloud scheme are converted into precipitation dependent on the water/ice content. Precipitation enhancement processes, such as collection of cloud water by precipitation and the Bergeron process are also taken into account.
- **Evaporation:** it is assumed that falling precipitation evaporates in non-saturated layers before reaching the ground. This may substantially reduce the surface precipitation. Evaporation of the precipitation is not assumed to take place within the cloud, but only in cloud free air besides or below the model clouds.
- **Melting:** melting of falling snow occurs in a thin layer of a few hundreds of metres below the freezing level. It is assumed that snow can melt in each layer whenever the temperature exceeds 0°C . The melting is limited not only by the snow amount, but also by keeping the induced cooling of the layer such that the temperature of the layer after melting is not below 0°C .

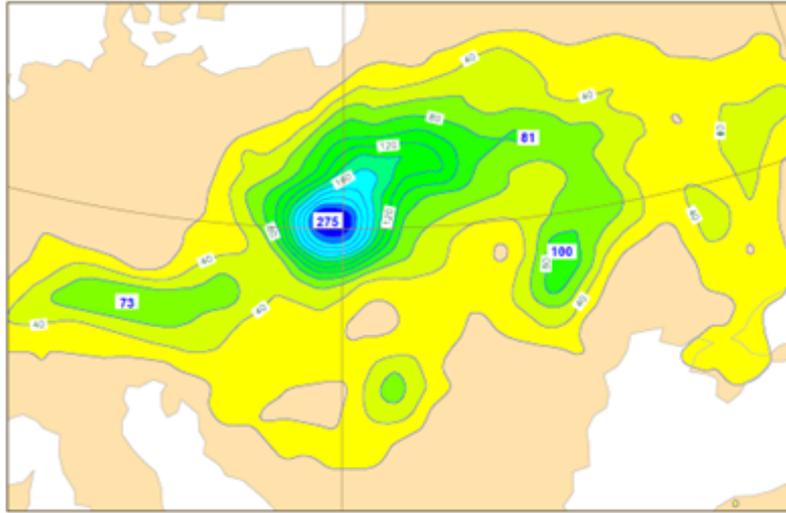


Figure 2.4: Accumulated rainfall during the first 96 hours of the T213 operational forecast 5 July 1997 12 UTC. The floodings in eastern Europe summer 1997 were call forecast by the ECMWF model. However, the maximum rainfall of 400 mm in siutheastern Poland was slightly underastimated. (Taken from Ref. [3])

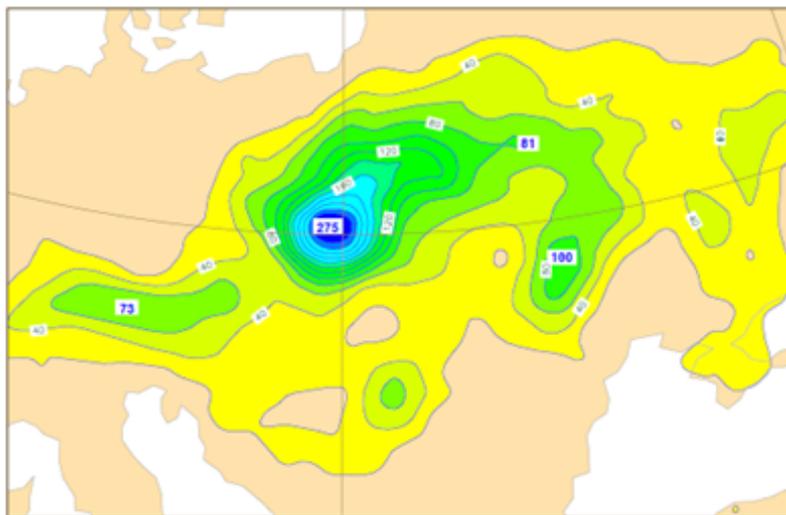


Figure 2.5: Accumulated rainfall during the first 96 hours of the T639 experimental model forecast 5 July 1997 12 UTC. With a higher resolution model the correct level of intensity is achieved, and the orographic effects more realistically treated. (Taken from Ref. [3])

Chapter 3

The Ensemble Prediction System (EPS)

3.1 Introduction

Twenty-thirty years ago the introduction of primitive equation models, the rapid advance in computer technology, remote sensing from satellites and an ever-increased sophistication of numerical methods fostered a sense of great optimism. But progress in predictive skill remained slow and gradually the question arose if there was an ultimate limit to atmospheric predictability. The interest came to focus on a strange result, first reported at a NWP meeting in Tokyo in 1960 by Edward Lorenz at the MIT. He had investigated if calculations based on non-linear differential equations could be replaced by statistical methods. The answer was “no”, but during one of his computational simulations he noticed how very small differences in the initial conditions could affect his extremely sensitive non-linear differential equations.

The consequences for NWP was that the limitations in the observational network and measurement accuracy would impose an upper limit in weather forecast quality. However, it was realized that in spite of this, the value of the NWP would be highly enhanced if the quality of the forecasts could be assessed a priori. The idea of including a stochastic element in NWP was born, but it had to wait until the late 1980' s until sufficient computer power made experiments possible.

For an analysis error to have a more wide spread impact it must occur in a dynamically sensitive region, in particular where young baroclinic systems develop. The errors from this weather system will in a few days' time spread to the next, in a process reminiscent of a “domino effect”.

The 2-4-day forecasts for the European area are therefore sensitive to the analysis over W Atlantic and eastern North America. The intensity and position of the cut-offs frequently forming in the eastern Atlantic, is highly influenced by the presence of a strong cyclogenesis over the Mexican Gulf and southern USA. Any error in the forecast of this upstream feature will be crucial for the success of the downstream cut-off.

The 5-7-day forecasts for the European area are sensitive to the initial conditions over the central and western part of North America, and the eastern part of the Pacific. Forecasts beyond a week are influenced by the initial conditions over central and western parts of the Pacific, and at day 10 from eastern Asia.

Since the initial state of the atmosphere is known with a limited accuracy, even small analysis errors in sensitive parts of the atmosphere may affect the very large scale flow during the course of the ten day forecast period. Another, equally accurate analysis with a slightly different geographical distribution of the initial errors, might yield a different forecast. The deterministic forecast is just one possible development of a number of alternatives, not necessarily the most likely.[6]

3.2 The ECMWF Ensemble Prediction System - an overview

The ECMWF Ensemble Prediction System (EPS) has been a part of the operational production since 1992. The EPS simulates possible initial uncertainties by adding, to the unperturbed analysis, small perturbations within the limits of uncertainty of the analysis. From these, a number of different forecasts are produced.

In 2008, the EPS was merged with the monthly prediction system and has been coupled to a dynamical ocean model. Since then, the EPS has been producing 15-day probabilistic forecasts daily at 00 and 12UTC. On Thursdays, forecasts are extended to 32 days, to provide users with monthly forecasts. Since 2010, the EPS probabilistic forecast has been based on 51 integrations with approximately 32-km resolution up to forecast day 10, and 65-km resolution thereafter, with 62 vertical levels.

The ECMWF EPS represents uncertainty in the initial conditions by creating a set of 50 forecasts starting from slightly different states that are close, but not identical, to our best estimate of the initial state of the atmosphere (the control). Each forecast is based on a model which is close, but not identical, to our best estimate of the model equations, thus representing also the

influence of model uncertainties on forecast error. The divergence, or spread, of the control plus 50 forecasts gives an estimate of the uncertainty of the prediction on that particular day. On some days, the spread might be small implying that the atmosphere is very predictable and users can trust that the reality will fall somewhere in the narrow range of forecasts.

On other days, the 51 forecasts might diverge considerably after just a few forecast days, indicating that the atmosphere is especially unpredictable. The variable ensemble spread gives users potentially very useful information on the range of uncertainty. Having a quantitative flow-dependent estimate of uncertainty allows users to make better informed weather-related decisions. The main sources of uncertainty in numerical weather prediction arise from our incomplete knowledge of the exact state of the atmosphere (the initial conditions) and unavoidable simplifications in the representation of the complexity of nature in the numerical weather models. Also the intricate vegetation and soil moisture processes can be described only by assuming a simplified description of vegetation and soil types and the associated processes.

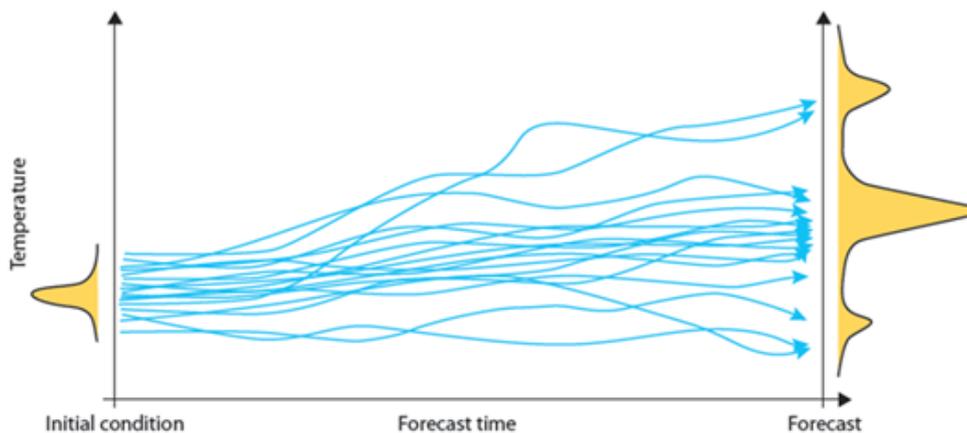


Figure 3.1: Representation of how the different perturbed initial conditions return a different forecast. (The ECMWF Ensemble Prediction System)

The basic principle of ensemble-based probabilistic forecasting is to make not only a single forecast from our best guess initial conditions, but also to perform a number of additional forecasts starting from slightly perturbed initial conditions, with each forecast created with a slightly perturbed model. This technique provides an estimate of the uncertainty associated with predictions from a given set of initial conditions compatible with observation

errors. If the atmosphere is in a predictable state, the spread will remain small; if the atmosphere is less predictable, the spread will be larger. In a reliable ensemble prediction system, reality will fall somewhere in the predicted range. This means that users get information on the actual predictability of the atmosphere, i.e. whether a particular forecast can be expected to be certain or less certain. In addition, they also get information on the range within which they can expect reality to fall.

3.3 The performance of the ECMWF Ensemble Prediction System

The ECMWF Ensemble Prediction System became fully operational in 1992. Since then, scientists at ECMWF have been constantly working to further improve the performance and utility of the EPS forecasts and products. Over the years, substantial improvements have been made in three key areas: in the model formulations and the data assimilation procedure used to estimate the initial conditions, in the use of more and better weather observations, and in the simulation of the effect of uncertainty in initial conditions and model equations. In 2010, two major changes were introduced: the simulation of initial uncertainties has been revised with the inclusion of perturbations defined by the ECMWF new Ensemble Data Assimilation system, and the schemes used to simulate model uncertainties have been revised substantially. As a result, the ECMWF EPS performance improved even further, and the EPS has kept its leadership position among the global, medium-range and monthly ensemble prediction systems operational in the world.

3.4 Perturbations

The success of any ensemble system depends on its ability to identify regions where small uncertainties in the analysis are likely to have significant impact on the forecast, and to create structures which will simulate these uncertainties.

3.4.1 The calculations of perturbations in the mid-latitudes

The EPS perturbation technique, based on a mathematical method called singular vector analysis, tries to identify the dynamically most unstable regions of the atmosphere by calculating where small initial uncertainties would

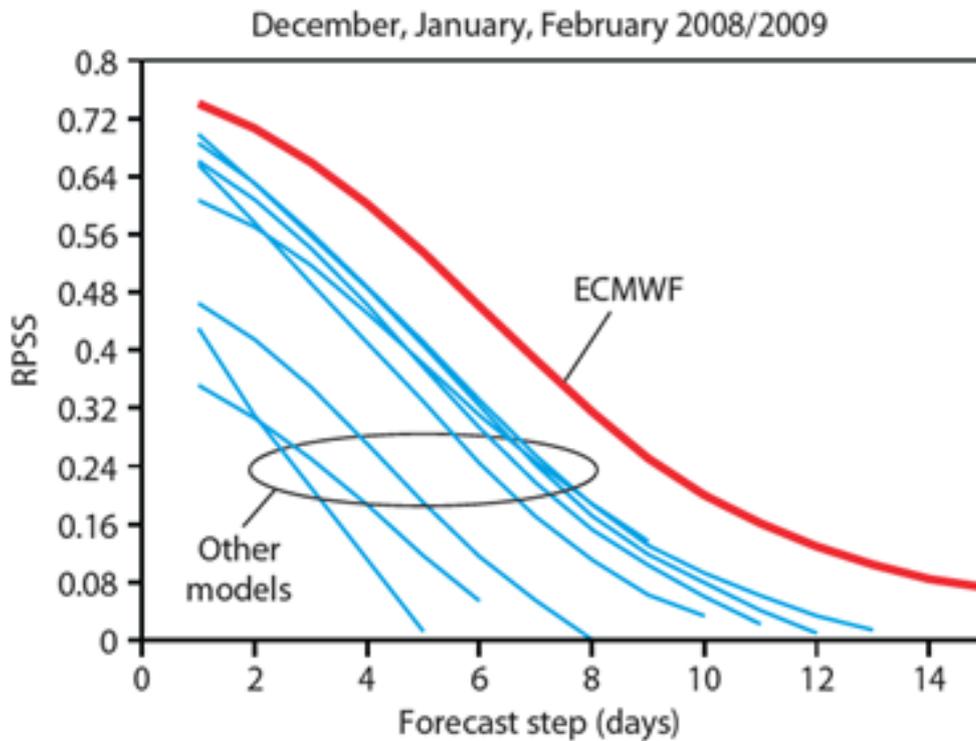


Figure 3.2: A comparison of the performance of all global ensemble prediction systems operational in the world demonstrates the leading position of the ECMWF EPS. The skill measure used here is the Ranked Probability Skill Score (RPSS), which is 1 for a perfect forecast and 0 for a forecast no better than climatology. (The ECMWF Ensemble Prediction System)

affect a 48 hour forecast most rapidly, i.e. both increasing or dampening the forecast amplification of a developing baroclinic system or unstable ridge. The first 25 of these singular vectors, chosen not to overlap too much, are combined in a linear way to calculate hemispheric structures (separately for each hemisphere) which are able to have a significant effect on the forecast after 48 hours.

By reversing the signs, 25 “mirrored” perturbations are produced, yielding a total of 50 global perturbation fields. These initial perturbations are scaled so that their local maxima are comparable to local analysis errors, and to have a realistic ensemble spread after 48 hours. The final perturbations are spatially uncorrelated. They are also considered a priori to be equally likely. The success of the EPS over Europe is determined to what degree it can correctly account for the uncertainties and alternative developments of an

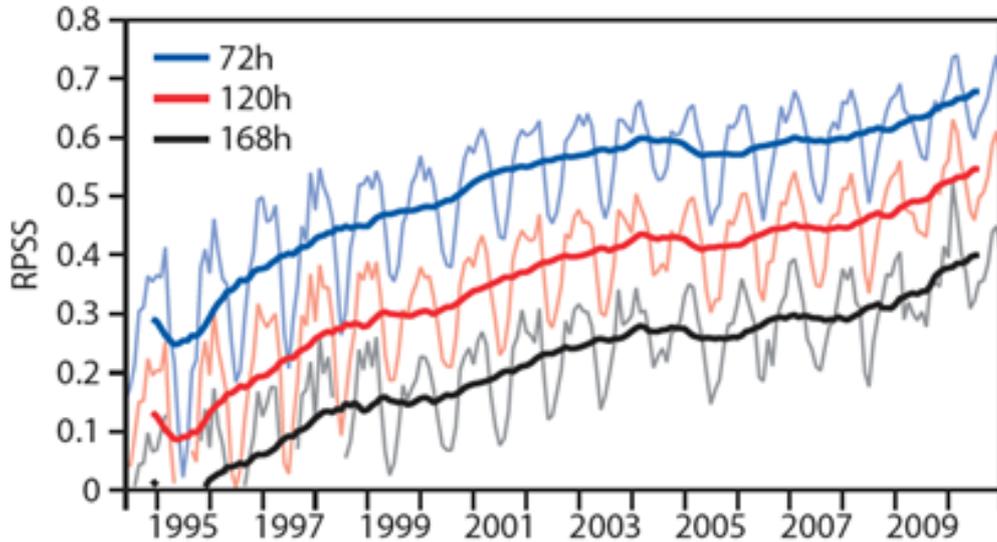


Figure 3.3: The performance of the EPS has improved steadily since it became operational in the mid-1990s. The skill measure used here is the Ranked Probability Skill Score (RPSS), which is 1 for a perfect forecast and 0 for a forecast no better than climatology. (The ECMWF Ensemble Prediction System)

upstream baroclinic development. Most of the EPS perturbations which are of importance for the medium range forecast over Europe on a weeks range are inserted in the analysis of baroclinic systems over the north Pacific.

3.4.2 The simulation of model errors

Although the main approach has been to simulate the effects of possible errors in the initial conditions, increasing research work is devoted to simulating the effect the finite resolution of the model grid or simplified representation of the physical processes. These will have importance in connection with strong physical forcing, for example when tropical cyclones enter the mid-latitudes and interact with the baroclinic development in the westerlies. The source of such errors has been addressed by the introduction of stochastic physics (Buizza et al, 1999). For each ensemble member, the stochastic physics perturbs grid point tendencies by up to 50%, with a spatial correlation radius of 10 latitude degrees and a time correlation interval of 6 hours. The whole globe is perturbed, including the Tropics. The non-perturbed Control forecast is run without stochastic perturbations.

3.4.3 Definition of the initial perturbations

Techniques to generate initial perturbation have been based on strategies to identify those direction in phase space where dynamical instabilities are strongest. One possibility is to assume that errors in the initial condition are dominated by those instabilities of the flow which have developed over a series of previous simulation cycles. However, assuming an isotropic PDF in phase space for the error at the initial time, the different amplification rates of perturbations along different axes would soon stretch the PDF along the directions of maximum instability during the early stages of the forecast.

In a meteorological context for any finite time interval in which the dynamics of perturbations is assumed to be linear the axis of maximum instability can be computed as the eigenvectors of the asymmetric operator defined as the product of a linear propagator by its adjoint. In linear algebra notation these eigenvectors are the singular vectors (SVs) of the linear propagator itself.

The methodology used in the Ensemble Prediction System to define these linear combinations is a modification of the procedure described in Palmer et al (1993). Its aim is to create perturbations which cover most of the Northern Hemisphere and have an amplitude comparable to the estimates of root-mean-square analysis error provided by the optimum-interpolation (OI) data assimilation. This proceeds as follows:

- The first SVs are always selected
- For each SV, a localisation function is defined in three-dimensional grid-point space, equal to 1 wherever the local energy(per unit mass) of the SV field is greater than 1% of its maximum value over the grid, 0 elsewhere.
- An overlap function is defined at each point as the sum of localisation functions of the first four SVs. In general, the overlap function gives the number of selected SVs which ‘cover’ any grid point.
- Each subsequent DV is examined in turn and selected only if more than half of its energy lies in regions where the current overlap function is less than 4. If this is the case, the localisation function for the new SV is used to update the overlap function.

The last step is repeated until all the SVs are selected. Once all the SVs have been selected, an orthogonal rotation in phase space and a final rescaling are performed to generate the ensemble perturbations. In practice the purpose of the phase-space rotation is to generate perturbations which have the same globally averaged energy as the ‘raw’ SVs, but a smaller maximum and a more uniform spatial distribution.

3.5 Spread–skill

Depending on the particular hemispheric flow pattern, forecasts originating from perturbed analyses develop more or less differently during the course of a ten day forecast.

If model errors played no role, and if initial uncertainties were fully included in the EPS initial perturbations, a small spread among the EPS members would be an indication of a very predictable situation. In other words, whatever small errors there might be in the initial conditions, they would not seriously affect the deterministic forecast. In these cases extended and/or detailed forecast interpretations are possible. By contrast, a large spread indicates a large uncertainty of the deterministic forecast, which prevents any extended or detailed forecast interpretation.

But the EPS does not limit the interpretation of the spread just as a measure of uncertainties. The information will also suggest possible alternative developments and their respective likelihood. Last but not least, it will also indicate what is not likely to happen, which at times might be as important as knowing what is likely to happen. When, on some rare occasions, the spread might cover most of the climatological range, then nothing can be deduced from the forecast about any significant deviations from climate.

The spread-skill interpretation of the EPS is complicated by the fact that in one and the same forecast the spread often varies considerably from one parameter to another. A small spread in the 500 hPa geopotential forecasts does not necessarily imply a small spread in for example the forecast precipitation, and vice versa.

3.6 EPS clustering

To compress the amount of information being produced by the EPS and highlight the predictable and thus relevant parts, individual EPS forecasts, which are “similar” according to some norm, are grouped together and averaged to constitute new forecast fields, so called clusters. The norm for judging this “similarity” can be the correlations between the fields or, as in the ECMWF system, their RMS differences.

3.6.1 The operational clustering

The ECMWF operational clustering algorithm is based on the RMS differences between the 500 hPa geopotential height ensemble forecasts, averaged from +120h to +168h taking the synoptic continuity into account. It is al-

ways the same members which make up the contents of each cluster. There are occasions when two members in the same cluster can be rather different at the beginning or end of the period, but sufficiently similar during the rest of the time interval to be placed in the same cluster. On the other hand, two members, being similar during a part of the period, may be placed in different clusters if they are sufficiently different during most of the period. The number of clusters depends on three factors:

- The spread of the day, i.e. the EPS standard deviation,. It is varying from day to day, but follows a seasonal trend as the forecast errors, with higher values in winter than in summer
- The clustering threshold used to limit the clusters standard deviation. It follows the same seasonal trend as the spread and errors
- The degree of “multi modality”, the tendency of the forecasts to form discrete alternatives. For the same spread and threshold a multi modal distribution might lead to a smaller number of clusters than a mono modal distribution.

A large spread in the ensemble does therefore not necessarily lead to more clusters, nor does a small spread necessarily lead to fewer. The clustering is performed separately for the whole of Europe plus four European sub-domain.

3.6.2 The “tubing” clustering

Another clustering method, called tubing, averages all ensemble members which are similar, on a RMS basis, to the ensemble mean and excludes members which are significantly different. The average of all these “similar” members provides a more refined ensemble mean, the central cluster mean. The excluded members are grouped together in a number of “tubes” (maximum 9) each represented by their most extreme member allowing to better visualize the different scenarios in the ensemble.

The central cluster mean and the tubes are computed for the whole forecast range. For each tubing reference step (+96h, +144h, +168h, +192h and +240h), tubing products are generated over a 48-hour sequence finishing on the reference step (for example +48/+72/+96h for the +96h tubing), allowing a sequential view of the different tendencies. Tubes are computed over each of the five geographical domains Europe, NW Europe, NE Europe, SW Europe and SE Europe. They do not intended to serve as probability alternatives, only to give an indication of what is not included in the central cluster.

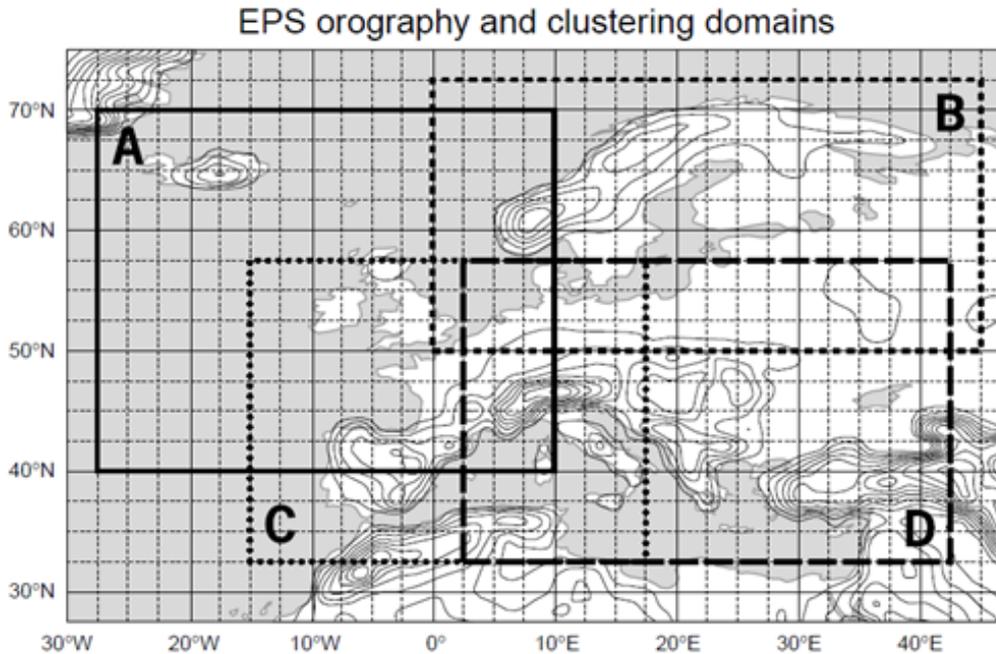


Figure 3.4: The five main clustering areas, the European and four sub-area. (Taken from Ref. [3])

3.6.3 No ideal clustering

Every possible clustering is a compromise; the advantage of condensing information has to be paid by the risk of losing information which on some occasions, in hindsight, might have been important. There is really no superior or objective measure of which type of clustering is “best”. Clustering can be performed over larger or smaller geographical areas, on different parameters, it can be done for each forecast time or for a longer period. Other possible ways:

- clustering using correlation measures will highlight similarities in the patterns but may group together forecasts which differ in the overall level of temperature and geopotential heights.
- clustering according to the 500 hPa flow might in a zonal situation give one cluster, whereas if the clustering had been performed on the MSLP pattern, the differences in the position and intensity of zonally moving baroclinic waves might have created 3–4 clusters.
- clustering on individual forecast days will have the advantage that each

day can be judged by its own merits; the disadvantage is that the temporal continuity and synoptical consistency will be lost.

Ideally the forecaster should have access to more than one clustering method, since what is the “best” clustering will vary according to the weather situation.

Chapter 4

The statistical indices

4.1 Introduction

Performance evaluation is an interdisciplinary research problem. Performance metrics (error measures) are vital components of the evaluation frameworks in various fields. In machine learning regression experiments, performance metrics are used to compare the trained model predictions with the actual (observed) data from the testing data set. Forecasting has a long history of employing performance metrics to measure how much forecasts deviate from observations in order to assess quality and choose forecasting methods, especially in support of supply chain or predicting workload for software development[7].

Classification is one of the main topics of scientific research. Each knowledge domain, as a subject of scientific research, requires classification systems (typology) to structure the contents in a systematic manner. Categories of the typology are defined based on resemblances (or differences) of items/objects in a specific context. Typologies are helpful in ordering and organizing knowledge, defining the scope and simplifying studies, facilitating information retrieval and detecting duplicative objects.

For the purposes of the study it is therefore necessary to identify the most suitable indices in order to verify the validity of the corrections that will be made.

4.2 The indices of the errors used

The i th discrepancy D_i between the i th forecast simulations S_i and the i th observation O_i of a vector variable is defined as:

$$D_i = S_i - O_i \quad (4.1)$$

A variety of verification procedures has been developed and a review of these can be found. Each measure has its own strengths and shortcomings, where the latter are not necessarily addressed by other diagnostics. The indices that have been selected for analysis are therefore shown.

4.2.1 Bias and NBias

The Bias indicates the overall systematic difference between forecast and reality so that useful guiding notions like “the model is too wet/dry or too warm/cold” can be derived, but what constitutes a large or small bias is hard to say from the value of the bias itself without a context [8]. The figure 4.1 shows the above concept of systematic error.

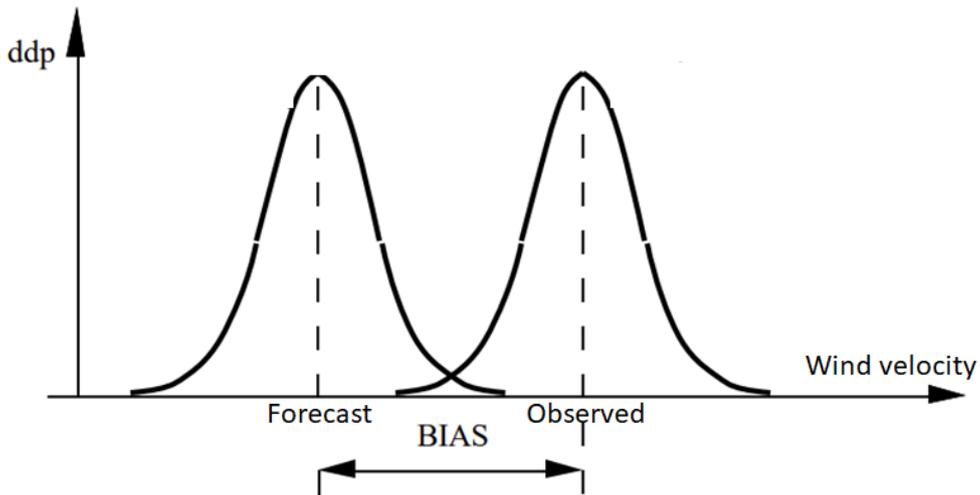


Figure 4.1: Systematic difference between forecast and reality.

From the figure it is clear that there is a constant error that generates an underestimation of the forecasts with respect to the real data. The Bias is therefore a very useful index because it can be easily corrected with a translation. Bias is determined as follows:

$$Bias = \frac{\sum_{i=1}^N D_i}{N} \quad (4.2)$$

Where:

- D_i is the discrepancy defined in the (4.1)
- N is the number of observation and forecast

The Bias can assume both positive and negative values. The positive value will highlight an overestimation of the forecast while the negative value an underestimate (having defined the discrepancy as in 4.1).

The NBias is obtained by carrying out the normalization of the Bias with the sum of the observations as follows:

$$NBias = \frac{\sum_{i=1}^N D_i}{\sum_{i=1}^N O_i} \quad (4.3)$$

By doing so, the percentage of the systematic error is obtained, which can still be negative.

4.2.2 RMSE and NRMSE

Root Mean Square Error (RMSE) is the standard deviation of the residuals (prediction errors). Residuals are a measure of how far from the regression line data points are; RMSE is a measure of how spread out these residuals are. In other words, it tells how concentrated the data is around the diagonal (Figure 4.2).

Root mean square gives a good estimate of the overall error between the model and the observations, but it tends to vary directly with the standard deviation of the observed quantities. This means the size of RMSE is not solely due to the model's performance per se, e.g. small errors for temperature and humidity in the tropics and large errors for wind in the upper troposphere are somewhat expected from the corresponding small or large variabilities in physical quantities themselves.

RMSE is determined as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N D_i^2}{N}} \quad (4.4)$$

As can be seen from the formula 4.4, the RMSE is a non-linear index. Therefore this index gives more weight to occasional but coarse errors than to small and frequent ones.

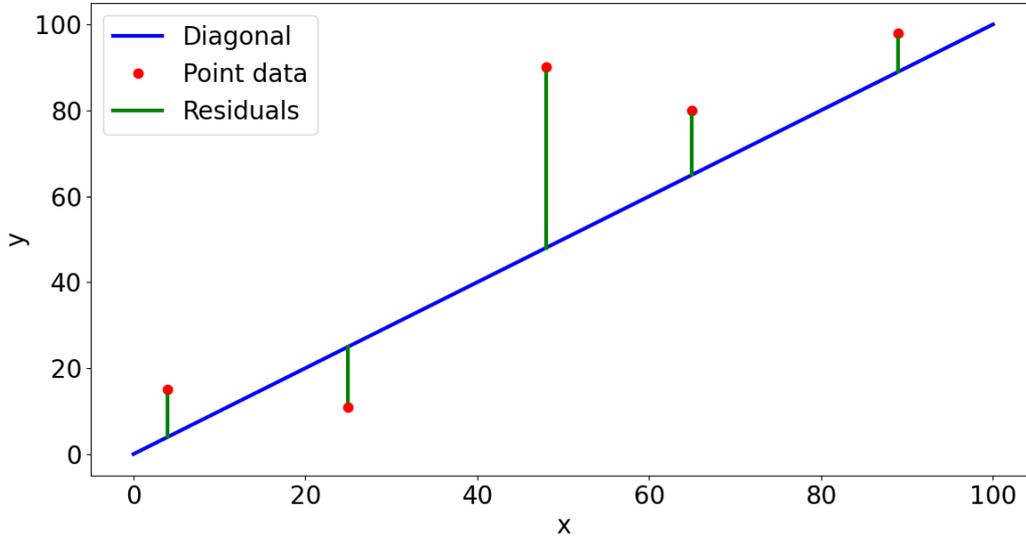


Figure 4.2: Residuals of the data from the diagonal line.

The NRMSE is obtained by carrying out the normalization of the RMSE with square root of the sum of the observations raised to the second as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N D_i^2}{\sum_{i=1}^N O_i^2}} \quad (4.5)$$

So the NRMSE gives a percentage of the root mean square error.

4.2.3 MAE and NMAE

The Mean Absolute Error (MAE) measures the average magnitude of the errors in a set of predictions, without considering their direction. It's the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight. MAE is determined as follows:

$$MAE = \frac{\sum_{i=1}^N |D_i|}{N} \quad (4.6)$$

[9]Is interesting to make a comparison between RMSE and MAE:

- **Similarities:** Both MAE and RMSE express average model prediction error in units of the variable of interest. Both metrics can range from 0 to ∞ and are indifferent to the direction of errors. They are negatively-oriented scores, which means lower values are better.

- **Differences:** Taking the square root of the average squared errors has some interesting implications for RMSE. Since the errors are squared before they are averaged, the RMSE gives a relatively high weight to large errors. This means the RMSE should be more useful when large errors are particularly undesirable. Differently MAE is a linear term and so, gives the same weight for small and large errors.
- **MAE \leq RMSE:** The RMSE result will always be larger or equal to the MAE. If all of the errors have the same magnitude, then RMSE=MAE.

The NMAE is obtained by carrying out the normalization of the MAE with the sum of the observations as follows:

$$NMAE = \frac{\sum_{i=1}^N |D_i|}{\sum_{i=1}^N O_i} \quad (4.7)$$

It returns a percentage of the Mean Absolute Error.

This is one of the indices that will have the greatest importance in assessing the skill of forecasts. This is due to the fact that it shows the precise value of how far the forecast deviates from the observation and therefore by multiplying for the cost of the electricity (€/kW), the economic value of the forecast error can easily be obtained.

4.2.4 SI, Scatter Index

The scatter index (SI) is defined as the standard deviation of the difference normalized by the mean of the observations. Is obtain subtracting the average component of the error. It is defined as:

$$SI = \sqrt{\frac{\sum_{i=1}^N [(S_i - \bar{S}) - (O_i - \bar{O})]^2}{\sum_{i=1}^N O_i^2}} \quad (4.8)$$

Where:

- \bar{S} is the mean value of the simulations (S_i)
- \bar{O} is the mean value of the observation (O_i)

4.2.5 HH, Hanna and Heinold index

In order to evaluate the reliability of numerical simulations in geophysical applications it is necessary to pay attention when using the root mean square error (RMSE) and two other indicators derived from it (the normalized root

mean square error (NRMSE), and the scatter index (SI)). In conditions of constant correlation coefficient, in fact, the RMSE index and its variants tend to be systematically smaller (hence identifying better performances of numerical models) for simulations affected by negative bias. Through a geometrical decomposition of RMSE in its components related to the average error and the scatter error it can be shown that the above mentioned behavior is triggered by a quasi-linear dependency between these components in the neighborhood of null bias. This result suggests that smaller values of RMSE, NRMSE and SI do not always identify the best performances of numerical simulations, and that these indicators are not always reliable to assess the accuracy of numerical models[10]. The corrected indicator HH proposed by Hanna and Heinold (1985) is demonstrated that provides a more reliable information about the accuracy of the results of numerical models. The corrected statistical indicator proposed by Hanna and Heinold is defined as:

$$HH = \sqrt{\frac{\sum_{i=1}^N (S_i - O_i)^2}{\sum_{i=1}^N S_i O_i}} \quad (4.9)$$

As it can be seen from the formula 4.9 the HH indicator overcome the problem of the RMSE, NRMSE, SI introducing a different normalization of the root mean square error.

4.2.6 Pearson index

A Pearson correlation is a number between -1 and +1 that indicates to which extent 2 variables are linearly related.

As shown in the figure 4.3 the correlation coefficient is useful to detect errors arising from phase lead or lag between forecast and observation but is independent of the difference in the variance of forecast and observation. So having a correlation of one is of dubious significance if forecast variance is much smaller than observed variance and is left uncorrected. It is defined as:

$$Pearson = \frac{\sum_{i=1}^N (S_i - \bar{S})(O_i - \bar{O})}{N\sigma_S\sigma_O} \quad (4.10)$$

With:

- $$\sigma_S = \sqrt{\frac{\sum_{i=1}^N (S_i - \bar{S})^2}{N}} \quad (4.11)$$

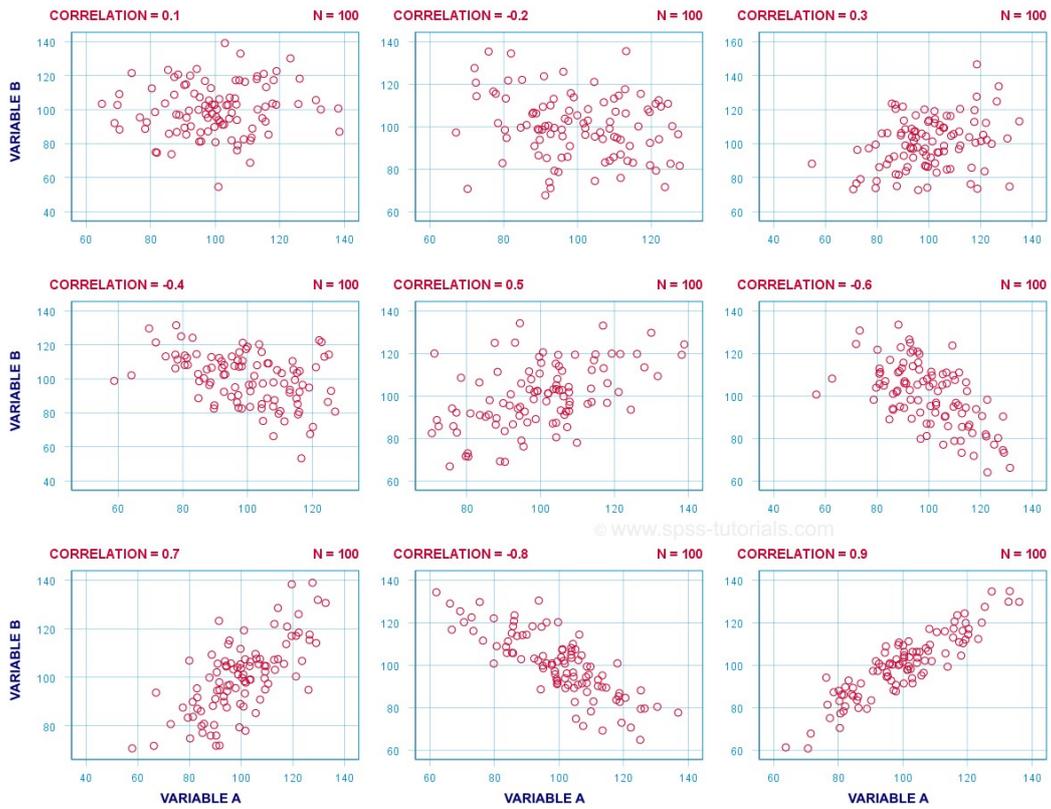


Figure 4.3: Variation of the correlation visualized as scatterplots.

•

$$\sigma_O = \sqrt{\frac{\sum_{i=1}^N (O_i - \bar{O})^2}{N}} \quad (4.12)$$

This index is very important because the more the predictions are related to the observations, the more it is possible to improve them.

Chapter 5

Analysis of forecasts and observations

5.1 Introduction

A wind farm, called Marsica1, was identified in Marsica, a geographic region of the Abruzzo hinterland, in the province of L'Aquila. In this wind farm was selected the anemometer of a turbine, called turbine R1, from which the observed data of the wind in that area were extrapolated. The measurements of the power produced in kW were also extrapolated from this turbine.

The wind forecasts were then downloaded from the European center referring to the node closest to the aforementioned turbine.

In this chapter it will be shown how the observed data were imported and cleaned up. The useful features in order to perform a correction of the forecasts will also be described.

The Python programming language was used for all analysis.

5.2 Observed data import for Marsica1

First, the files, with the .txt extension, containing the data of the Marsica1 wind farm were downloaded. Inside there were the wind speed data recorded by the anemometers of 3 wind turbines, including the R1, and also the data regarding the energy produced in terms of power [kW]. The data had been recorded with a time interval of 10 minutes. Data cover a total of two years: 2017 and 2018.

Since the forecasts provided by the European center have an interval of one hour, a moving average has been used on the observed data in order to make the two datasets uniform.



Figure 5.1: Marsica wind farm photo. Note how the wind turbines are quite close to each other.

Before the observations data could be used for analysis, it was necessary to clean them up. Regarding the wind speed measure there were indeed present several times records that were null. This appeared strange for several reasons:

1. The anemometer has a good sensitivity and can perceive even very light winds, up to $0,1m/s$.
2. Since the observations were averaged over the hour, it is statistically unlikely that there was no wind current in one hour.
3. The blades are located at a height of about 100 meters and are specially located in a very windy area. At those heights it is almost impossible not to have the minimum wind speed being very distant from the mainland and therefore having no nearby obstacles that can prevent it from moving.
4. Going to analyze one by one the records that marked a null value of the speed there was a clear discontinuity with the closest non-null values.

In figure 5.2 it can be seen how null values present an unnatural behaviour.

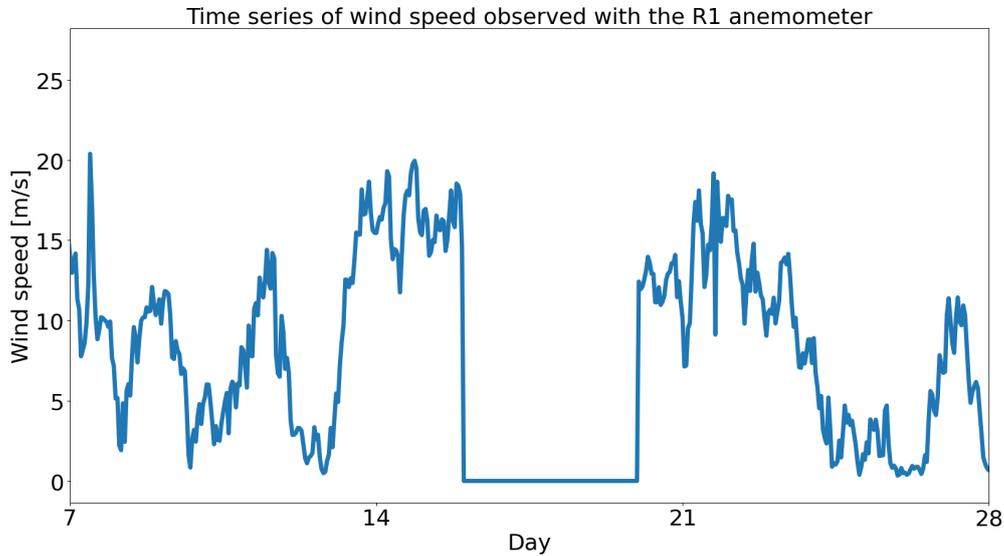


Figure 5.2: Wind speed time series measured from 08/01/2017 to 28/01/2017. There is a sudden zeroing of the wind speed for almost 4 days. This is clearly due to some kind of error in the data acquisition.

There are also errors in the measurements of the power produced. They are very evident when graphing the transfer function from the potential energy contained in the wind to the electricity produced. Figure 5.3 shows precisely how there are many values that are too far away from the transfer curve. The error is not necessarily due to the collection of power data but could also be due to incorrect measurements of the anemometer.

From figure 5.3 it is also noted that the transfer function is not perfect. It does not show a line but a point cloud. This means that even if the forecast of the wind speed was perfectly known, the perfect power forecast would never be reached. Table 5.1 shows the error indices of the powers calculated using the empirical transfer function with respect to the powers actually measured. Basically it is like simulating having reached a perfect wind forecast. As it can be seen, however, there is an error. There are mainly 2 indices: the NMAE which is approximately 6% and the NRMSE which is approximately 16%. This difference is due to the fact that NRMSE places much more weight on large errors than NMAE. Therefore, following the correction of the measurements, a greater improvement is expected from the NRMSE compared to the NMAE.

Much of the incorrect data mentioned above are not necessarily due to instrument or processing errors. Many are also due to the fact that during the

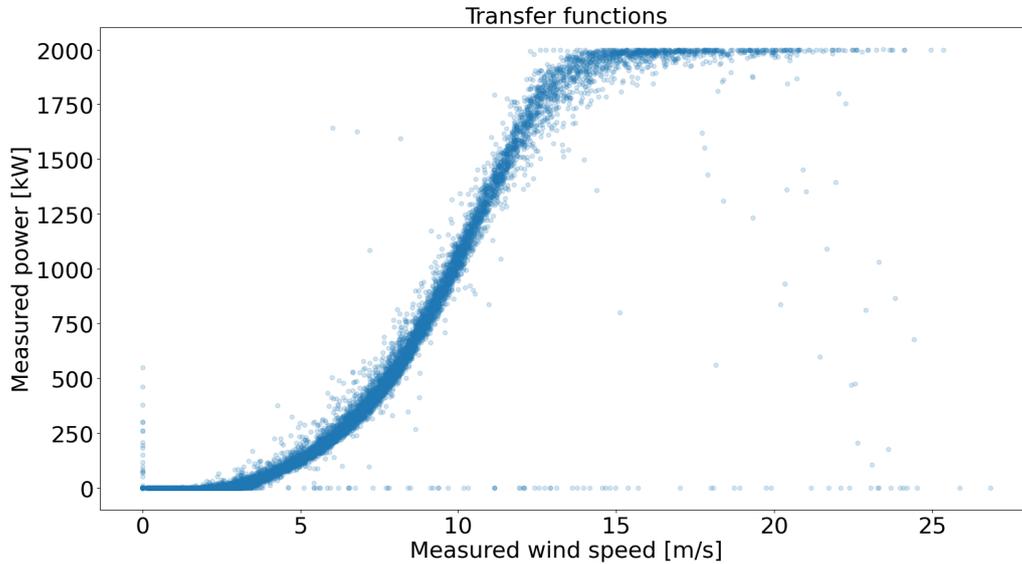


Figure 5.3: Empirical transfer function. Data that is not in the vicinity of the transfer curve are those that need to be corrected or modified.

Table 5.1: Error index of the original transfer function.

Bias	4.830
NBias	0.012
MAE	24.625
NMAE	0.062
RMSE	115.635
NRMSE	0.162
HH	0.163
SI	0.162
Pearson	0.981

period of activity of the turbine there is the need to carry out maintenance for which it is necessary to turn off the single turbine or even the entire plant. To clean this data, two procedures were used in order to throw away as little data as possible. In fact, to do some of the analyses it is necessary to have the records of the entire day without missing data. For each null data for which the replacement could not be made, the entire day of registrations and the corresponding forecasts were eliminated.

5.2.1 Substitution of null records with R2 and R3 ones

In figure 5.1 it has been pointed out that the wind turbines are quite close together. It can also be noted that there are no particular obstacles between them that can modify the flow of the wind in some way. It is therefore expected that the wind speed and therefore also the power produced by the turbines R2 and R3 is similar to that of the turbine R1. Figures 5.4 and 5.5 show how there is indeed a strong correlation and equality between the wind and the power produced by the three turbines for which the data is available.

This is due, as said before, to the fact that the turbines are located at a relatively short distance from the distance needed by the wind to be able to significantly change their motion at a height of 100 m where there are therefore no obstacles. Figures 5.4 and 5.5 show that there is not a perfect correspondence between the measurements of the turbine R1 and of the R2. In particular with regard to the power produced. This is due to the fact that the same conversion of wind energy into electrical energy is variable, as can also be seen from the transfer function (figure 5.3), and therefore this phenomenon is amplified when the two measurements are compared. After verifying the similarity of the records, the wrong ones of the R1 turbine were replaced with that of the R2 and R3 turbines.

Not in all the wrong values of R1, however, there could have been replacements because in some cases all the blades had anemometers that did not work.

5.2.2 Substitution or elimination of wrong records with the average of the records around

This substitution was only done where there were single wrong values next to valid records. In these cases the wrong records have been replaced with the average of the adjacent observations so as to avoid the elimination of the data of a whole day due to the lack of a single hour. Finally where the previous operations failed to create a meaningful data it was necessary to remove that data.

Figure 5.6 shows what the transfer function looks like after it has been cleaned up.

Table 5.2 shows the values of the error indices of the transfer function, by simulating having perfect wind forecasts, previously calculated in table 5.1 and those calculated using the cleaned data. All the indices have improved and in particular the NRMSE which, as expected, went from 16% to a 5. Also the NMAE had an improvement of about 2 percentage points and was around at 4%. This means that in spite of the cleaning works per-

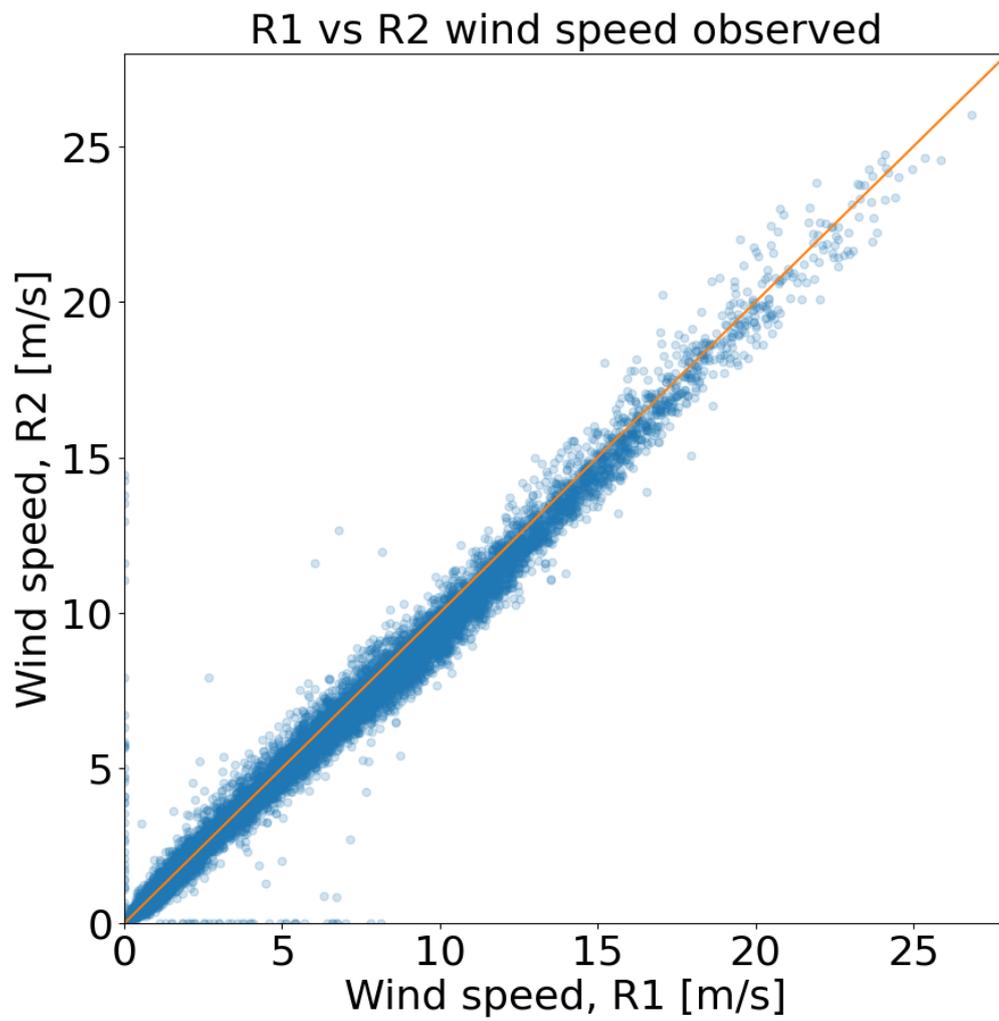


Figure 5.4: Scatter plot of the speeds recorded with the anemometer R1 vs those recorded with the anemometer R2.

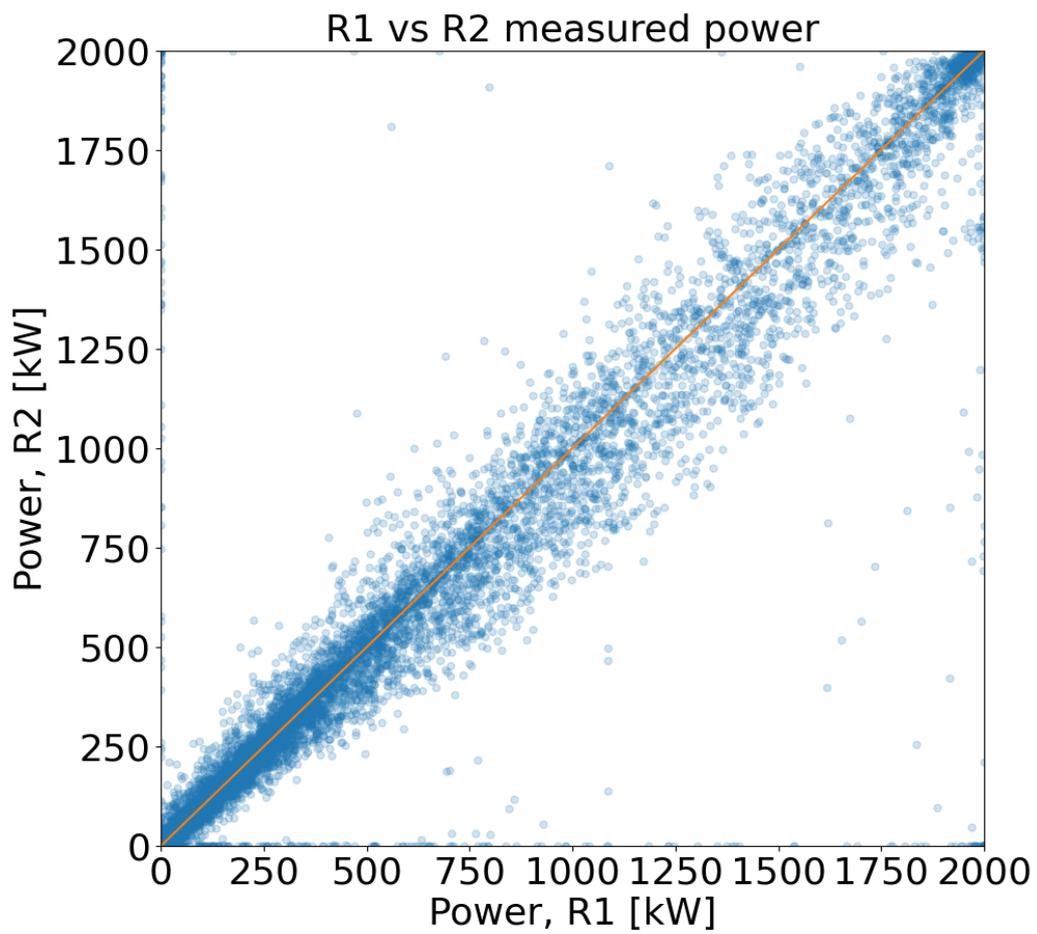


Figure 5.5: Scatter plot of the power recorded with the anemometer R1 vs those recorded with the anemometer R2.

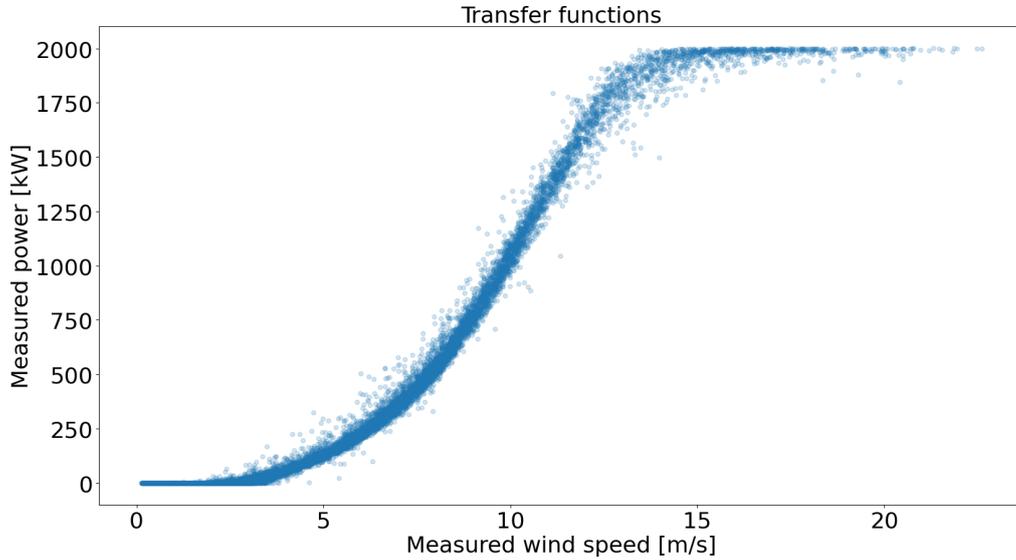


Figure 5.6: Transfer function after operations to clean up the data.

Table 5.2: Error index of original transfer function vs cleaned transfer function

Index	Original data	Cleaned data
Bias	4.830	-1.886
NBias	0.012	-0.005
MAE	24.625	16.426
NMAE	0.062	0.043
RMSE	115.635	33.219
NRMSE	0.162	0.048
HH	0.163	0.048
SI	0.162	0.048
Pearson	0.981	0.998

formed in the transfer function, an error persists due to the instrumentation or characteristics that are not controllable for the moment.

5.3 ECMWF forecasts data import for EPS and HRES

The wind forecast data was downloaded from ECMWF, European Center for Medium-Range Weather Forecasts. One are the so-called EPS, i.e. one

‘ensemble forecast’, that consists of 50 separate forecasts made by the same computer model equation, all activated from the same starting time. The other one is the HRES (High resolution) which provides a highly detailed description of future weather and it should be the most accurate forecast. Unlike EPS, the latter consists of a single forecast.

The data were taken by selecting the node closest to the R1 turbine of the Marsica 1 plant, that is, with a latitude of 42 and a longitude of 13.6. Since the resolution of the European center is relatively low (18 km for EPS and 9 km for HRES of resolution), the node is also suitable for being applied to the entire Marsica 1 plant. The data are sampled at the frequency of one hour.

The forecasts download start to run at midnight and have a maximum forecast horizon of 48 hours. For each day there are therefore 2 forecasts: one made the day before (24-48 hour forecasts) and the other made the same day (0-24 hour forecasts). It is important to underline that the forecasts are made available at least 7 hours after the calculation of the forecasts begins. Therefore, when the 0-24 hour forecasts are available in Italy it is around 7/8 in the morning. This means that the forecast from midnight to 7 is actually about the past.

The imported data had velocities along the two x and y directions, so it was necessary to find the velocity module for each prediction.

5.3.1 EPS structure

As mentioned in the previous chapters, EPS are 50 different predictions resulting from different boundary conditions obtained by perturbing the initial condition, considered to be the best, 50 times. It is therefore logical to expect that these small perturbations the more go forward in time the more they have an effect on the forecast.

Figure 5.7 shows the trend of the 50 forecasts in the 48-hour forecast. As it can be seen in the first hour the forecasts are all very close together although they have different values. This is due to the effect of the disturbance of the initial condition. Going forward in time, forecasts tend to move away, ending at the 48th hour with a greater variance than the initial one. It is obvious that the opening of the forecast is not due exclusively to the forecast horizon. In fact, between the 12th and the 28th hour there is the maximum opening probably due to the presence of a wind that is more difficult to predict than that of the following hours.

In Figure 5.8, forecasts 0-24 h were superimposed with those 24-48 h of the day before. It is clear that the forecasts generated on the day have a much greater variance than that of the forecasts generated on the same day.

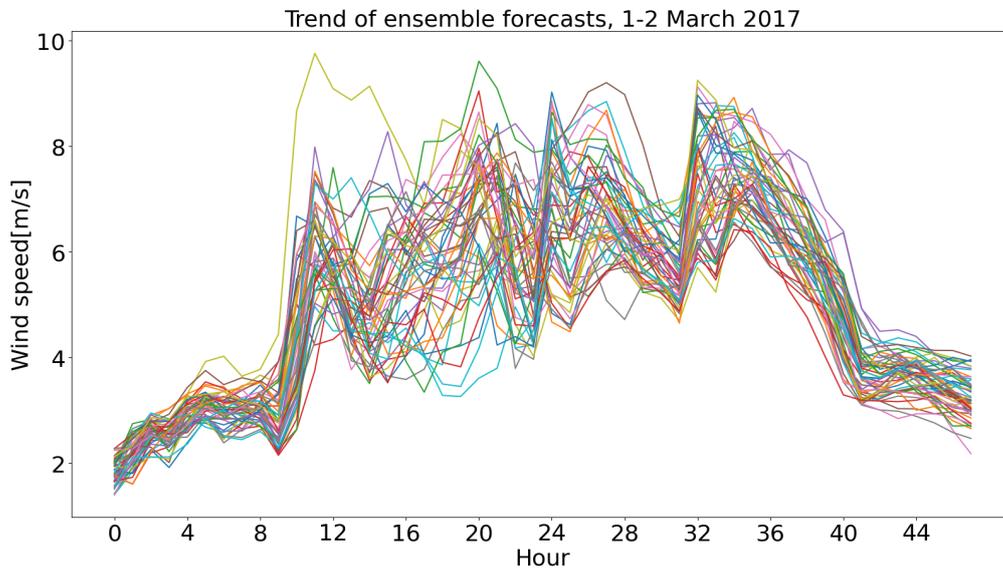


Figure 5.7: The 50 forecasts of the EPS of 1 and 2 March 2017. Each different forecast has its own color, as it can be seen the width of the members tends to widen over time.

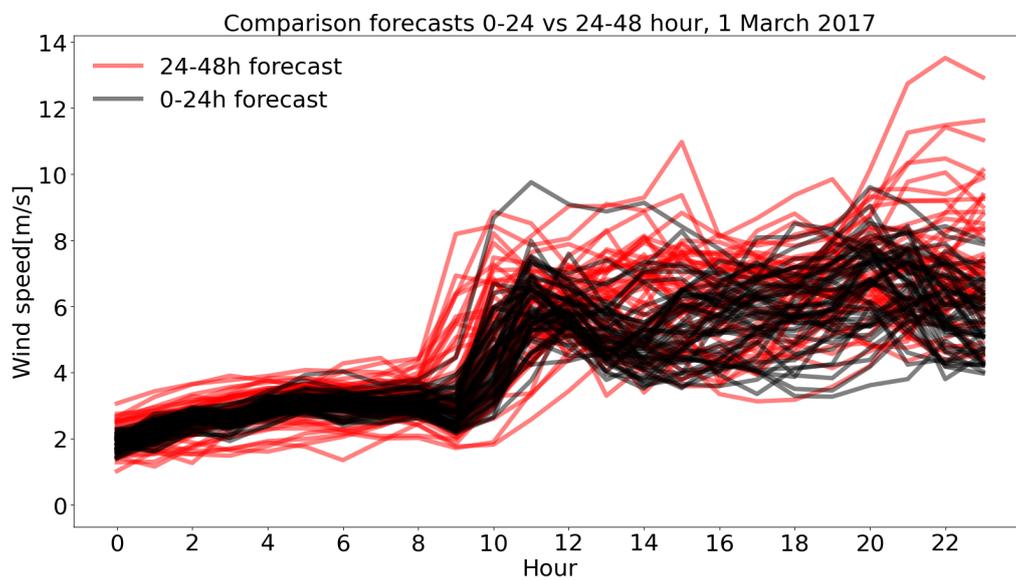


Figure 5.8: The 50 in red are the forecast referring to the run of the day before, while the black ones are the forecasts generated the same day.

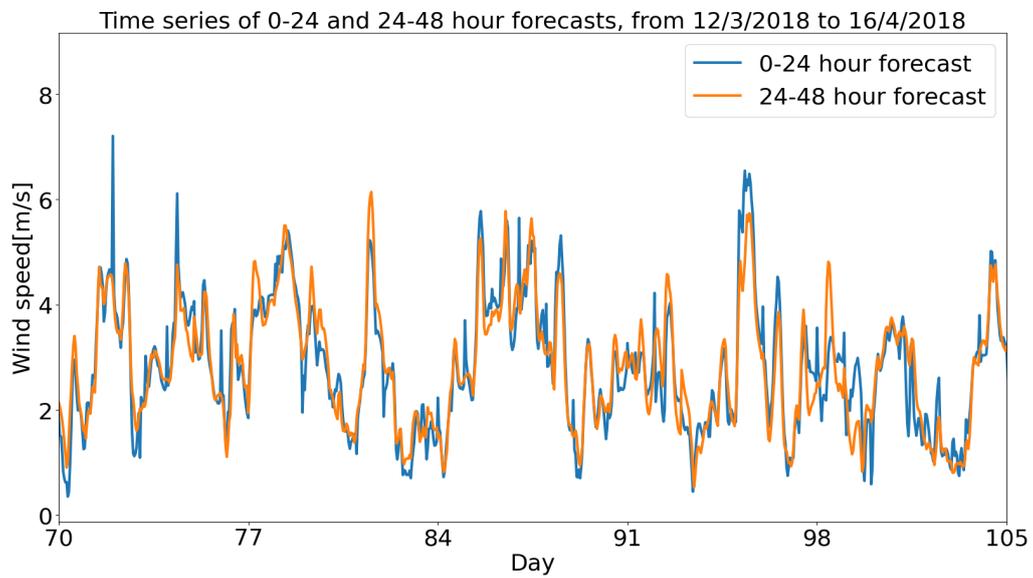


Figure 5.9: Time series from 12/3/2018 to 16/4/2018 in which the forecasts 0-24 hours were superimposed on those 24-48 hours.

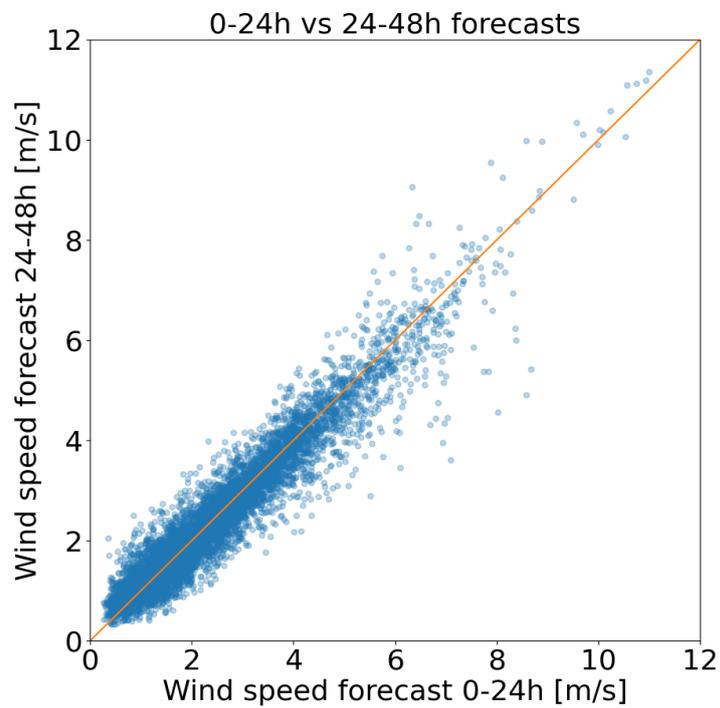


Figure 5.10: Scatter plot of the average forecasts with a time horizon of 0-24 hours vs those of 24-48 hours.

Finally it is useful to make some comparison between the forecasts of the same day (0-24 hours) with those of the following day (24-48 hours). Figure 5.9 shows the 5-week time series of the average of the two forecasts. It can be seen quite clearly that the values of the forecasts are very much in agreement even from day to day. This is also confirmed in the scatter plot of figure 5.10 in which the values tending to position themselves around the diagonal show that there is a strong similarity of the forecasts. Calculating the Pearson index confirms the strong correlation between the two forecasts. In fact, a Pearson equal to 0.96 is obtained.

This may mean that the models built by the European center turn out to be very robust as the forecast horizon advances.

5.3.2 HRES structure

HRES is another forecast that is provided by ECMWF. Unlike EPS, it has only one prediction obtained by running a higher resolution model using the initial conditions considered to be the best.

To compare this forecast with that deriving from the EPS, the average of the EPS was used. This because from the literature the EPS mean is considered as a better forecast than that obtained by selecting the forecast of 50 of the EPS with the best initial condition.

Figure 5.11 shows the comparison between the time series of the forecasts generated by HRES with the average of the EPS. The two forecasts appear very similar albeit different.

Figure 5.12 instead shows the scatter plot in which the two forecasts seen above are compared. There is a good, albeit not perfect, correlation. In fact, the cloud is quite enlarged. Furthermore, for low speeds the values seem to be around the diagonal while for speeds greater than 2 m/s there is an underestimation of the EPS forecast compared to that of HRES.

Table 5.3 shows the error indices calculated by comparing the two forecasts. The Bias confirms the fact that there is an underestimation of the EPS forecasts compared to the HRES one. It is also noted that the indices do not vary particularly between 0-24 and 24-48 hours. Finally, it is emphasized that the Pearson is worth about 0.88 which means that there is a fairly high correlation between the two forecasts.

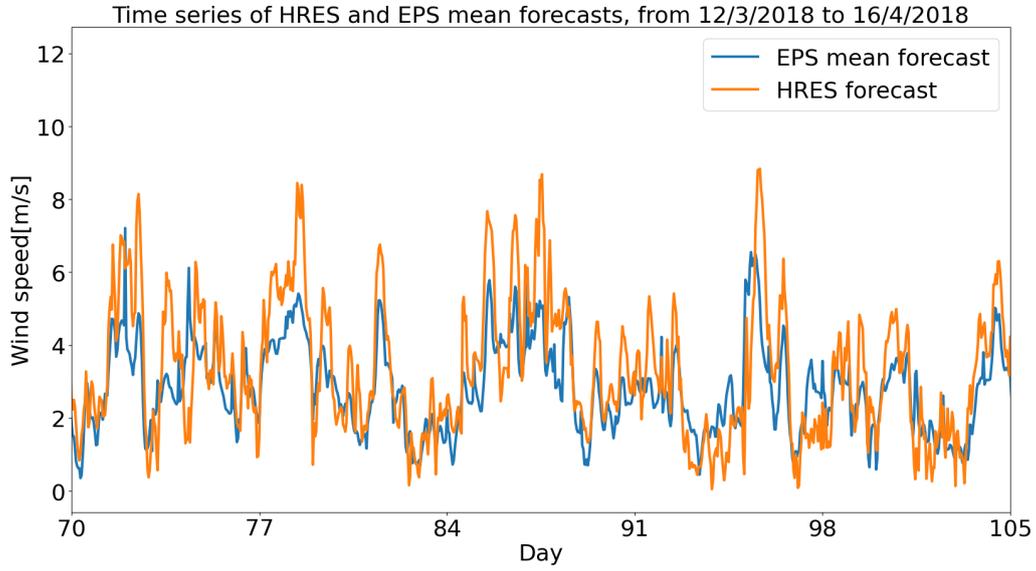


Figure 5.11: Time series from 12/3/2018 to 16/4/2018 in which the HRES forecast and the average forecast of EPS with a time maturity of 0-24 hours were superimposed.

Table 5.3: Error index of EPS vs HRES forecast.

Index	0-24 h	24-48 h
Bias	-0.412	-0.418
NBias	-0.145	-0.147
MAE	0.823	0.833
NMAE	0.288	0.293
RMSE	1.106	1.121
NRMSE	0.315	0.322
HH	0.358	0.367
SI	0.293	0.299
Pearson	0.882	0.877

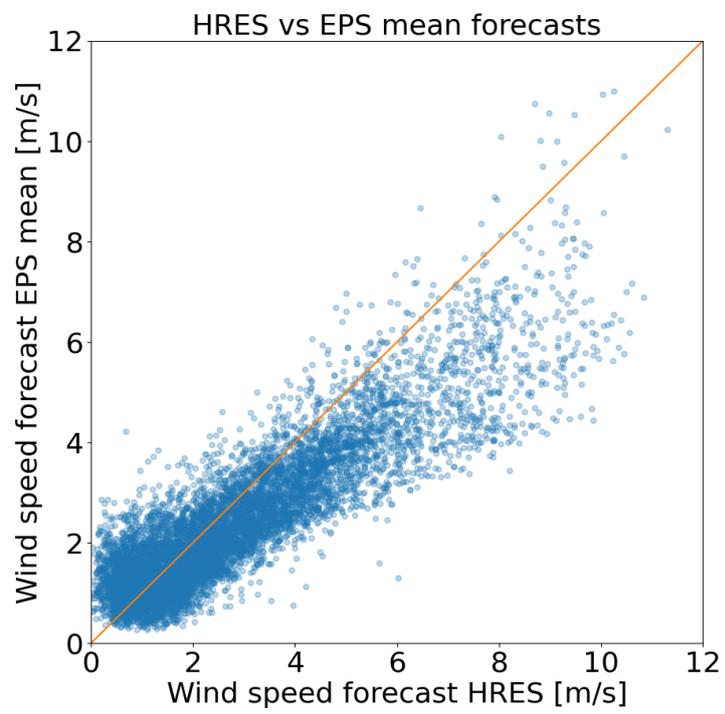


Figure 5.12: Scatter plot of the HRES forecasts vs average forecasts EPS.

5.4 Analysis and characteristics of the starting data

To make an effective improvement of the wind and power forecasts it is important first of all to know the characteristics of the wind in the site under analysis.

The characteristics will be analysed most of the time for both observations and predictions. This is to see if the two things coincide or not.

5.4.1 Seasonal trend of wind speed

The first step in these analyses is to know how the wind behaves over a whole year. Therefore it is necessary if it is subject to seasonal cycles like many other natural forces, in particular in the site under analysis.

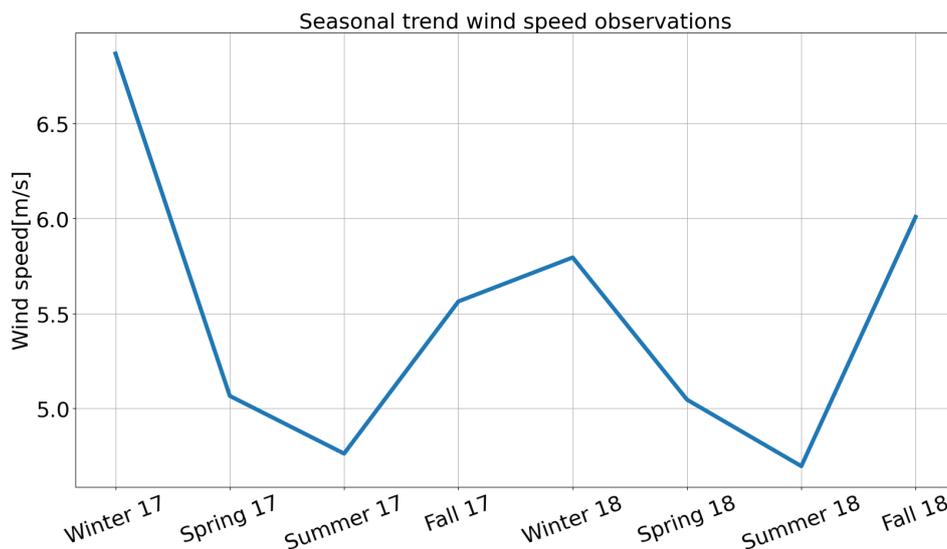


Figure 5.13: Seasonal trend of the average wind speed observation in 2017 and 2018.

Figures 5.13 and 5.14 show the average trend of the measurements of wind speed and power produced for each season. It is noted how both elements agree in denoting a seasonality of wind speed and power. In fact, the graphs show a greater intensity of the wind during autumn and winter as opposed to spring and summer in which there are lower values. This conclusion is obviously only and exclusively referred to the territory that is being analysed and does not represent an absolute truth being based on only two years of

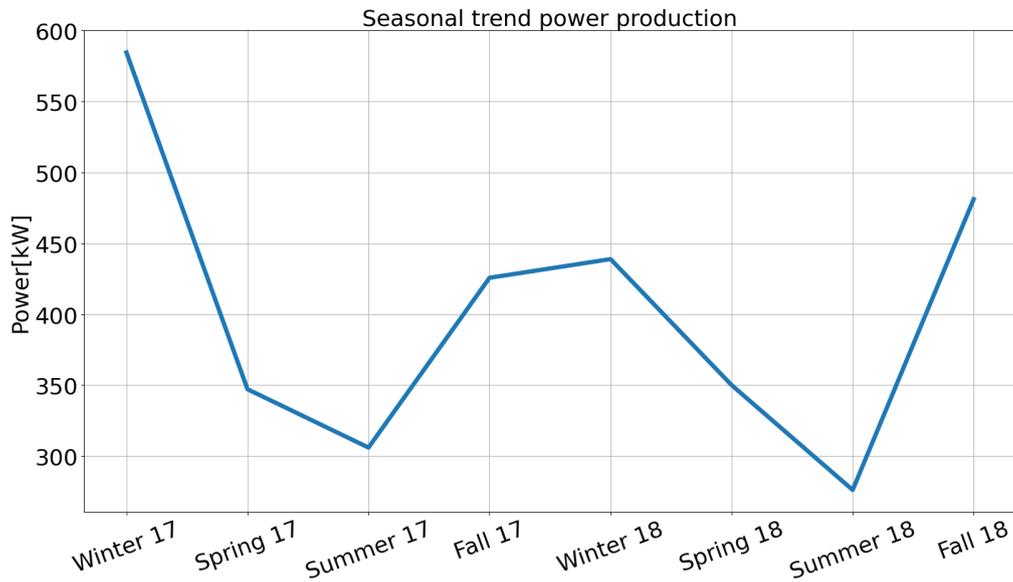


Figure 5.14: Seasonal trend of the average power product in 2017 and 2018.

observations. If this were the case, however, it is a point to the advantage of the wind farm in question since precisely in autumn and winter it is the period in which for example there is less sun and therefore there is less renewable energy deriving from the sun. Furthermore, winter is usually the period in which, as there are lower temperatures, there is more demand for energy.

It is also interesting to see if the forecasts confirm what seen from the observations. Another interesting aspect regarding forecasts is to see if the variance of EPS is also influenced by the seasons. Figure 5.15 shows the trend of the average forecasts as the season changes and confirms what was said previously. Figure 5.16 shows the change in variance as a function of the season. As it can be seen, there seems to be no particular correlation between the season and the EPS variation.

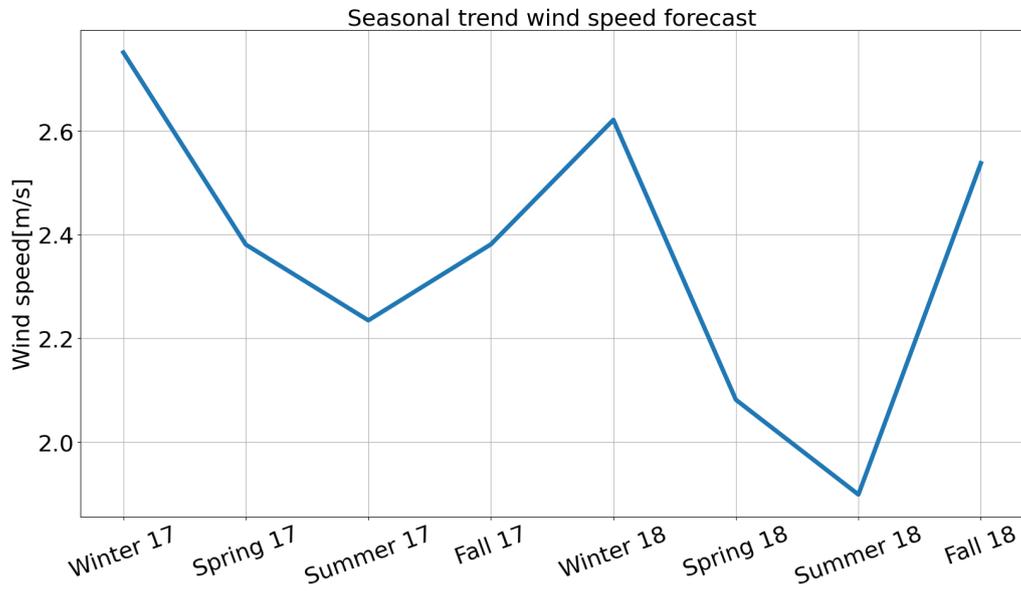


Figure 5.15: Seasonal trend of the average wind speed forecast in 2017 and 2018.

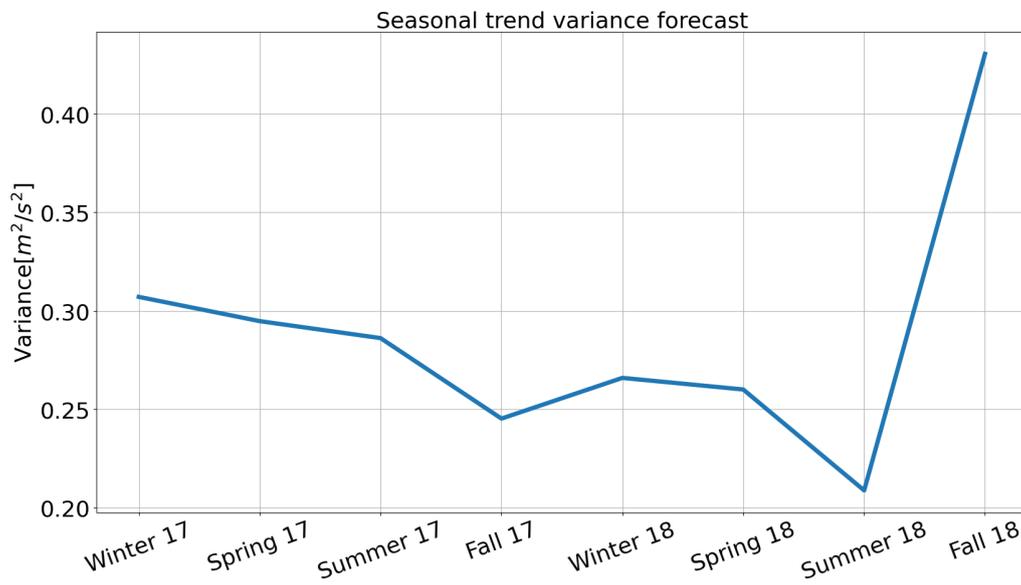


Figure 5.16: Seasonal trend of the average variance of EPS forecast in 2017 and 2018.

5.4.2 Daily trend of wind speed

Another important wind cycle is the so-called diurnal cycle. Typically, in fact, the wind tends to be more intense during the day while at night it is weaker.

Figures 5.17 and 5.18 show which are the average speeds and the average power observed during the different hours of the day. As expected there is a higher average speed during the day but still during the night there is a considerable energy production.

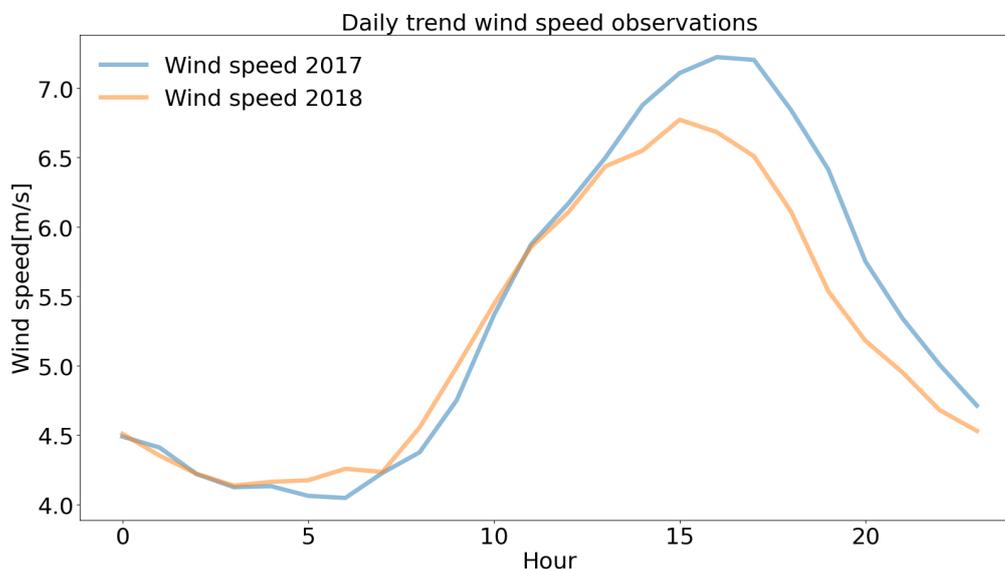


Figure 5.17: Daily trend of the wind speed observed in 2017 and 2018.

Figure 5.19 shows how the daily trend of the forecasts for the years 2017 and 2018 is. Visually the trend is in agreement with that seen as regards the observations. It is also interesting to reiterate how the 0-24 hour and 24-48 hour forecasts are practically totally overlapping with the exception of the first two hours. In fact, the first two hours of the 0-24 hour forecast show a sort of discontinuity with respect to the rest of the graph. This is due to the fact that in those two, having just started the run of the prediction, the model needs a couple of hours to settle and run correctly.

Figure 5.20 shows what the mean trend in the variance of the EPS forecast looks like during the day. Both holding years are shown for both 0-24 hour and 24-48 hour predictions. First of all, it is important to underline that there is a consistently greater variance of the forecasts 24-48 compared to

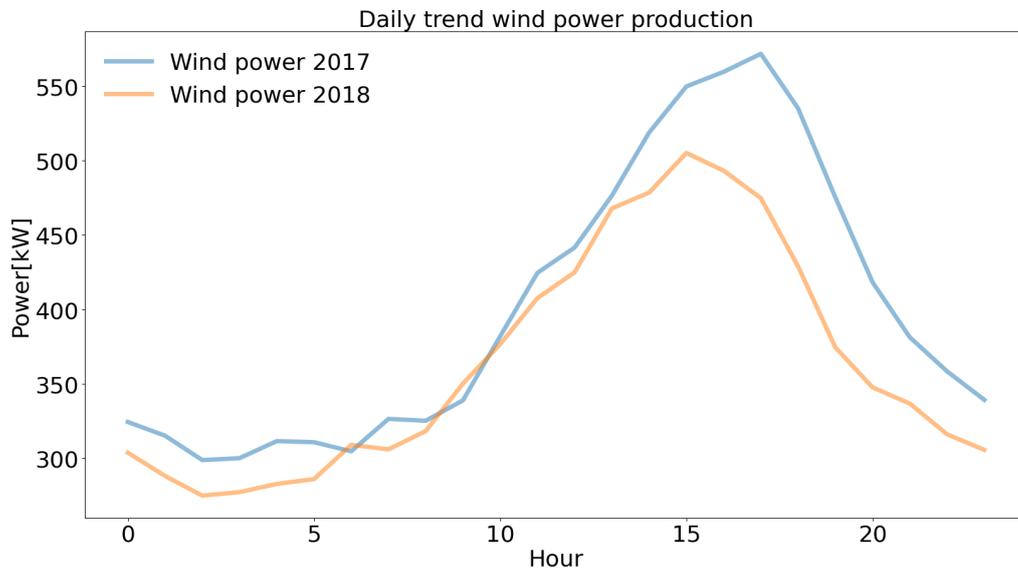


Figure 5.18: Daily trend of the average power product in 2017 and 2018.

0-24. This confirms what was said in the previous sections. It is also noted that the variance, without considering its variation due to the alternation between day and night, has a constant increase. In fact, the hour 23 of the forecasts is greater than the hour 0; this difference cannot be justified only by the fact that there is a physical difference between the two hours but above all by the simple increase in the forecast horizon. It should also be noted that the variance is also affected by the first two hours of settling the run of the opening 0-24 hours.

The other important point that is denoted in this graph is the presence of a strong variation of the variance due to the time in which it is considered. In fact, as for the wind speed, it also seems to be subject to a sort of diurnal cycle. This can be divided into two factors: the first is that the variance during the daytime is greater because the wind speed has higher values and therefore the forecasts tend to move away from each other; the second may be simply due to the fact that during the day the model of the European center is more difficult to describe the physical phenomena present. This is normal because during the day there are and more complex more physical phenomena due to the presence of the sun, which does not happen at night.

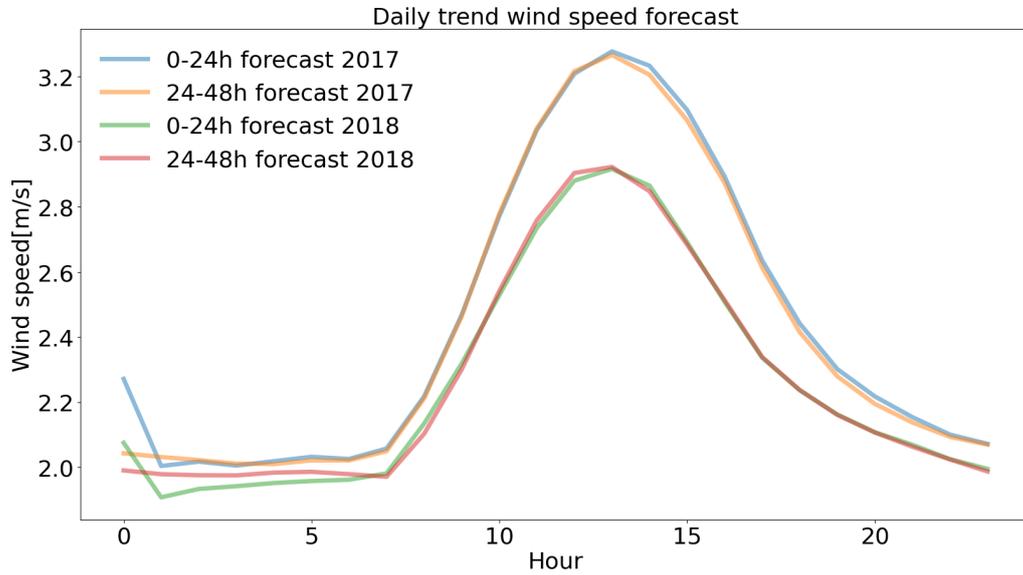


Figure 5.19: Daily trend of the average wind speed forecast in 2017 and 2018.

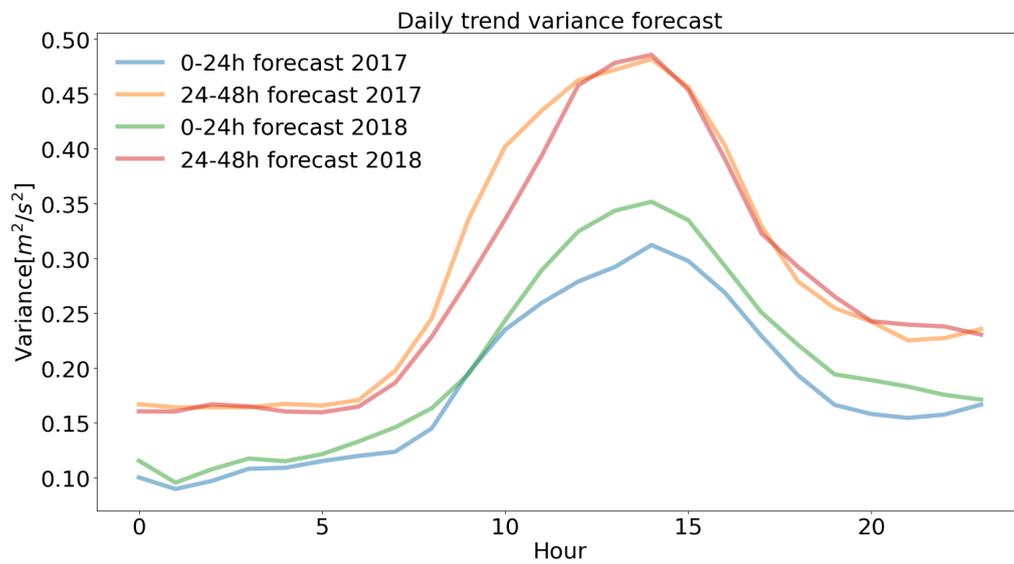


Figure 5.20: Daily trend of the variance of EPS forecasts in 2017 and 2018.

5.4.3 The wind direction

Wind direction is a very important component. In fact, depending on the morphology and characteristics of the analysed territory, the wind can change its intensity depending on the direction from which it arrives or in any case its characteristics. In fact, there are often winds considered more or less dangerous depending on the territory in which you are located. To give an example for Liguria the strongest wind comes from “Libeccio”. This is because the “Libeccio” before arriving in Liguria does not encounter obstacles for more than a thousand kilometres. However, winds from other directions are usually more docile.

As can be seen in figure 5.21, the Marsica 1 wind farm is located in a territory with a complex morphology. In fact, differently from what often happens around it there is no plain but rather mountain ranges.



Figure 5.21: Mountain range near the Marsica1 wind farm.

Figure 5.23 shows the observed wind speed and direction. It denotes very clearly how most of the wind is distributed in the second and fourth quadrant. In particular, looking at the wind rose (figure 5.22) they correspond to the winds called “Mistral” and “Scirocco”. Looking at figure 5.24 in which the predicted directions are shown, it confirm what was seen for the observed wind. Unlike observations, in which a very precise and well-defined direction appeared, with forecasts the direction is less defined but in any case it is clear that most of the wind belongs to quadrants 2 and 4.

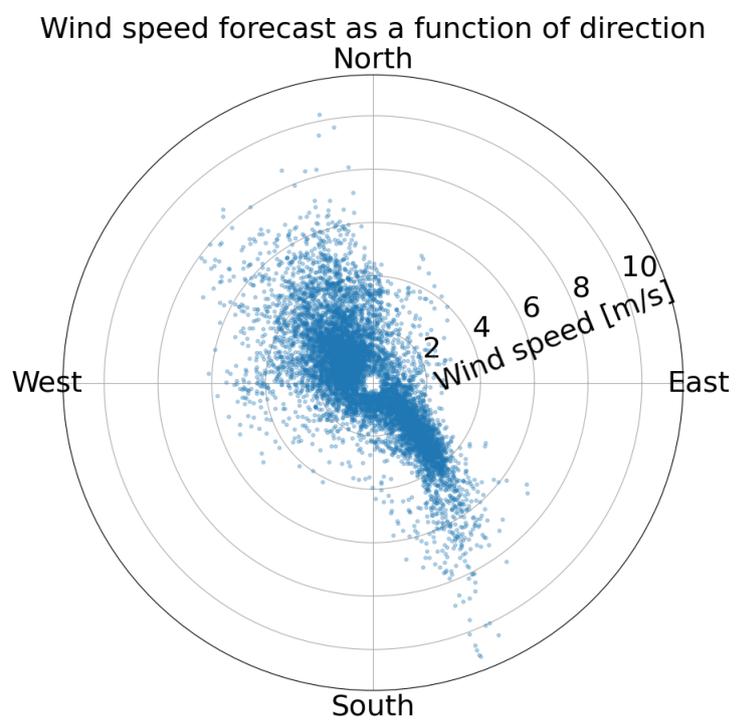


Figure 5.24: Distribution of the forecast wind directions.

5.5 Quality and accuracy of the starting forecasts

To correct the forecasts, it was necessary to divide the dataset into two parts: one part, corresponding to 2017 dataset, to calibrate the parameters for the correction, and the other, corresponding to 2018 dataset, to verify the correctness of the correction.

As seen in the previous sections, between the two years there are no particular variations in the wind and they are absolutely comparable as shown in table 5.4 which shows the average of the wind measured and forecast in the two years. There is a difference but it is absolutely natural due to the normal variation from year to year.

Table 5.4: Mean value of wind speed

	2017	2018
Observed [m/s]	5.38	5.2
HRES [m/s]	2.85	2.57
EPS [m/s]	2.43	2.25

In this section the dataset corresponding to the year 2018 will be analysed. This is because it will be the dataset that will then be used for the correction and it will be possible to make a comparison.

First of all it is useful to make a visual comparison between forecasts and observations.

Figure 5.25 shows the trend of wind speeds for forecasts and observations. From a simple visual analysis it is clear that the forecasts tend to be underestimated in most cases. Despite the underestimation, however, a certain correlation seems to remain which, as previously mentioned, will allow for a better correction. This assumption is confirmed by table 5.4 which shows an average underestimation of the forecast of more than 2 m/s. It is also noted that the underestimation is more marked as regards the forecast of EPS than that of HRES.

Figure 5.26 represents the scatter plot of EPS mean and HRES forecasts and observations. If the predictions were perfect, all points on the graph would lie on the black diagonal line. The fact that the points tend to be more to the left of the diagonal shows once again that there is an underestimation of the forecasts; in this case it can be said that there is a systematic error and therefore a negative Bias will be expected. Furthermore, the scatter plot provides a first view on the correlation. In fact, the more the point cloud is

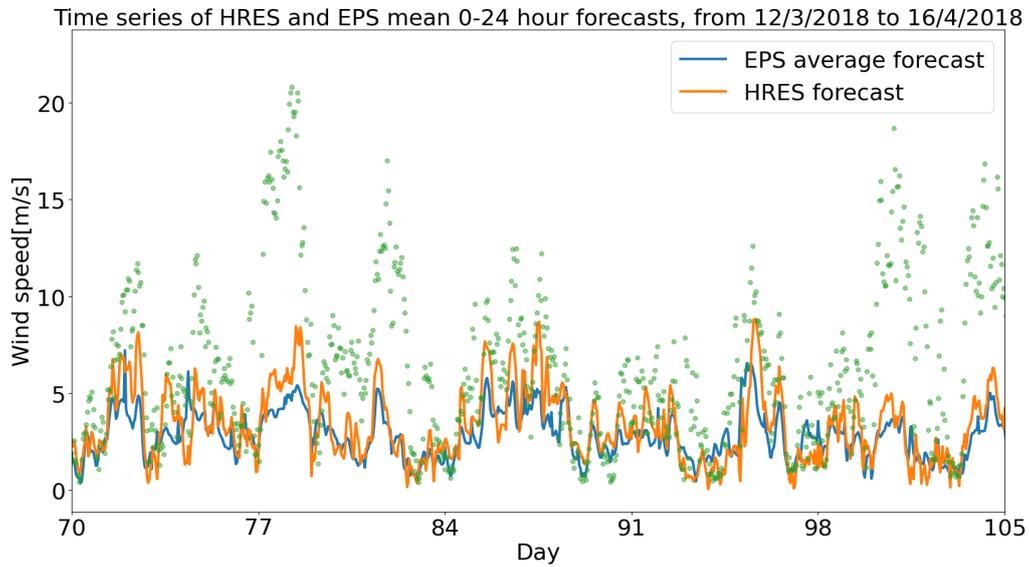


Figure 5.25: Comparison between observed and EPS and HRES forecasts wind speed.

restricted, the better the two datasets are correlated. On the contrary, the larger the cloud, the less the data are correlated with each other. From this figure, the forecast obtained with the average of the EPS seems to have a narrower point cloud than that of the HRES forecasts.

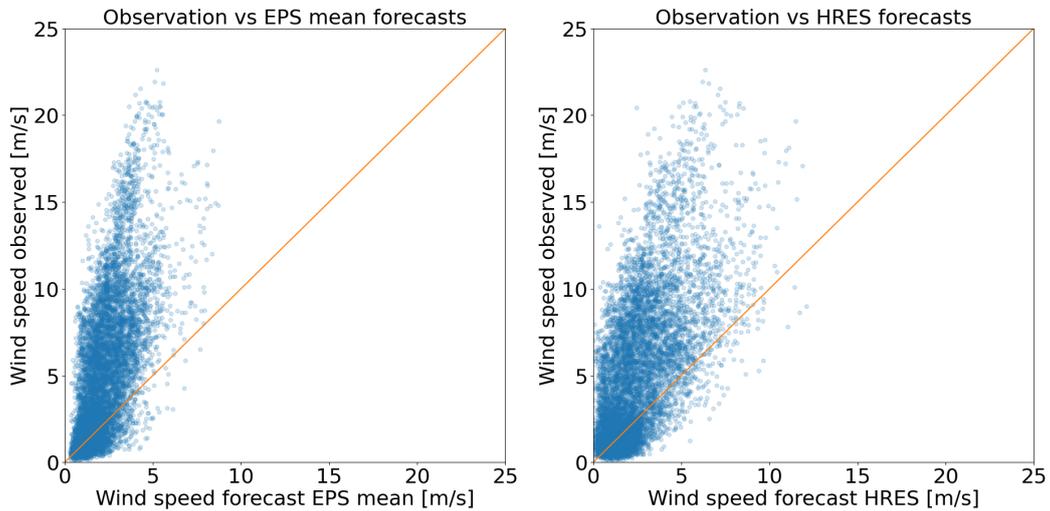


Figure 5.26: Scatter plot: observed vs EPS and HRES forecasts wind velocity.

Figure 5.27 shows the probability distribution of the EPS and HRES

predictions and of the wind speed observations. It is noted that there are velocities that are more likely to be both predicted and observed. This graph once again confirms the strong underestimation between forecasts and observations.

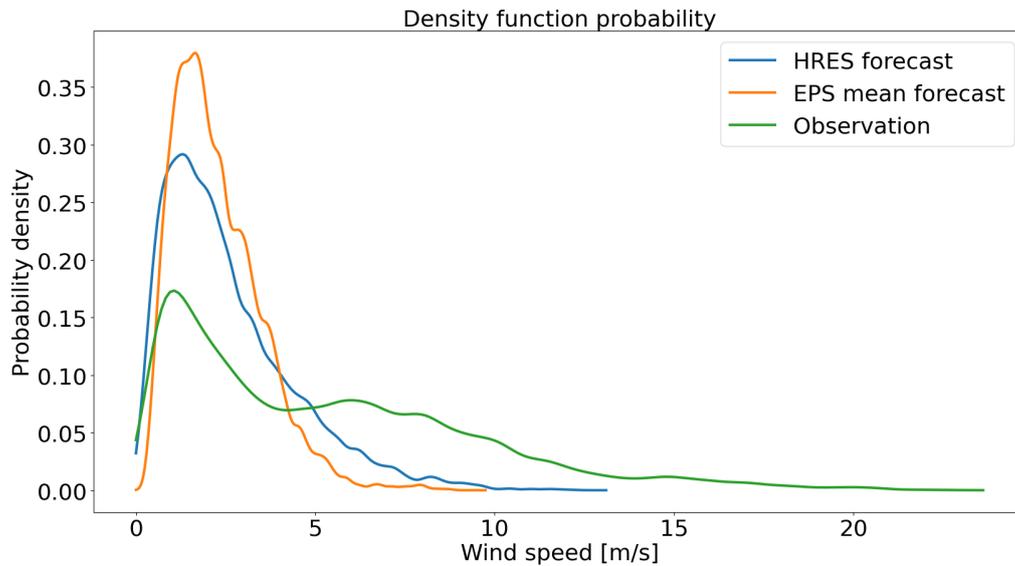


Figure 5.27: Probability density of wind velocity observed and forecast.

Table 5.5 shows the values of the error indices for each forecast and also. The Bias always has a negative value, which confirms what was deduced from the scatter plot.

It is also noted that for both types of forecast, the more you go forward in the forecast horizon, the more the forecast gets worse, albeit slightly. The table confirms what was previously assumed: the HRES forecast shows how, by exploiting the higher resolution of the model, it generally obtains better indices. In particular, all the indices apart from the Pearson have values closer to 0 than the EPS forecasts. For example, although the Bias is still negative, it shows an underestimation of the wind speed is better than that generated by the EPS; therefore HRES underestimates less than the EPS forecast, as previously assumed. Therefore, if by hypothesis one were to choose the best forecast without having the possibility to carry out any operation on it, the HRES forecast would certainly be chosen. However, the only index that is shown to be in favor of the EPS average is the Pearson. As mentioned in the previous chapter, Pearson shows what the correlation between forecasting and observation is. Therefore, in addition to confirming what is supposed based on figure 5.26, it leads to the supposition that following corrections,

the EPS forecasts may have a better result than that of HRES thanks to the greater correlation that links them with the observations.

Table 5.5: Error index of raw forecasts.

Index	Mean EPS 0-24h	Mean EPS 24-48h	HRES 0-24h	HRES 24-48h
Bias	-2.945	-2.941	-2.637	-2.656
NBias	-0.566	-0.565	-0.507	-0.510
MAE	3.215	3.223	3.013	3.055
NMAE	0.618	0.619	0.579	0.587
RMSE	4.572	4.599	4.247	4.322
NRMSE	0.687	0.692	0.639	0.650
HH	1.182	1.196	0.999	1.029
SI	0.526	0.532	0.501	0.513
Pearson	0.632	0.615	0.621	0.586

At this point it is interesting to see how the error indices vary according to the period in which the forecast is calculated. This is in order to identify some correlation between the period and the quality of the forecast. As regards the seasonal cycle, Figure 5.28 shows the correlation trend for both years in possession. There seems to be no clear correlation between the forecast and the season in which it is made. This analysis, however, is very inconsistent since two years are too short to be able to determine this aspect. It would be necessary to have at least ten years to compensate for the random variability that the variables under analysis.

One aspect for which sufficient data is instead available to be able to carry out a robust analysis is the correlation between errors and the daily cycle. Table 5.6 shows the value of the indices for all hours of the day for the forecast obtained from the average of the EPS with a forecast horizon of 24 hours.

To better understand how the various indices vary, it may be useful to see them in a graph. Figure 5.29 shows how the RMSE, MAE, NBias, Bias values vary according to the time expiration, overlapping with the value calculated on the whole series, without distinguishing the different time expirations. It is important to underline that the further the values are from zero, the more the forecast tends to be wrong. As can be seen, the indices worsen during the day while they tend to improve at night.

In figure 5.30 the HH, NRMSE, NMAE, SI values vary according to the time expiration, overlapping with the value calculated on the whole series, without distinguishing the different time expirations. As before the further the values are from zero, the more the forecast tends to be wrong. As can be seen, the indices worsen during the day while they tend to improve at night.

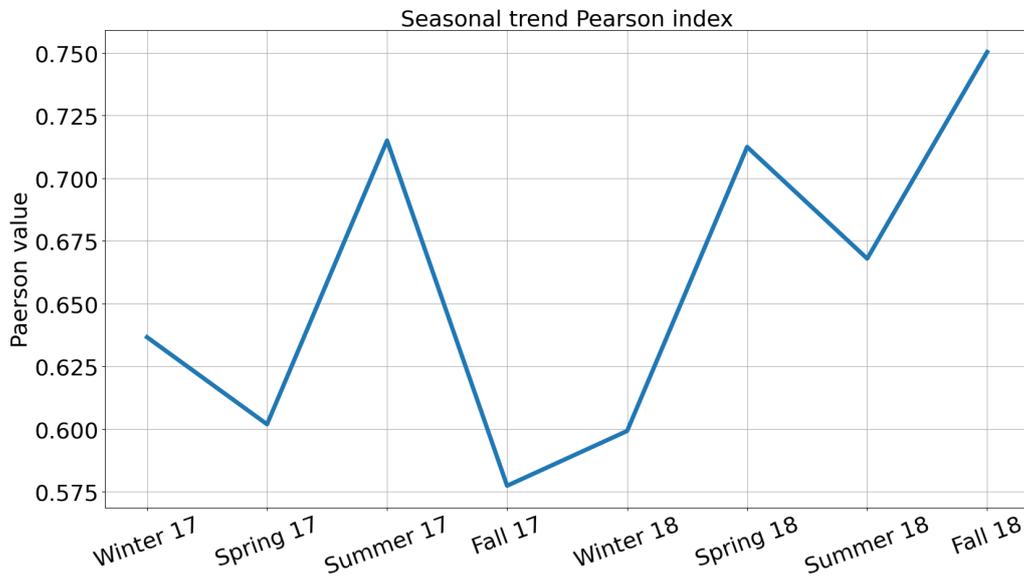


Figure 5.28: Scatter plot: observed vs EPS and HRES forecasts wind velocity.

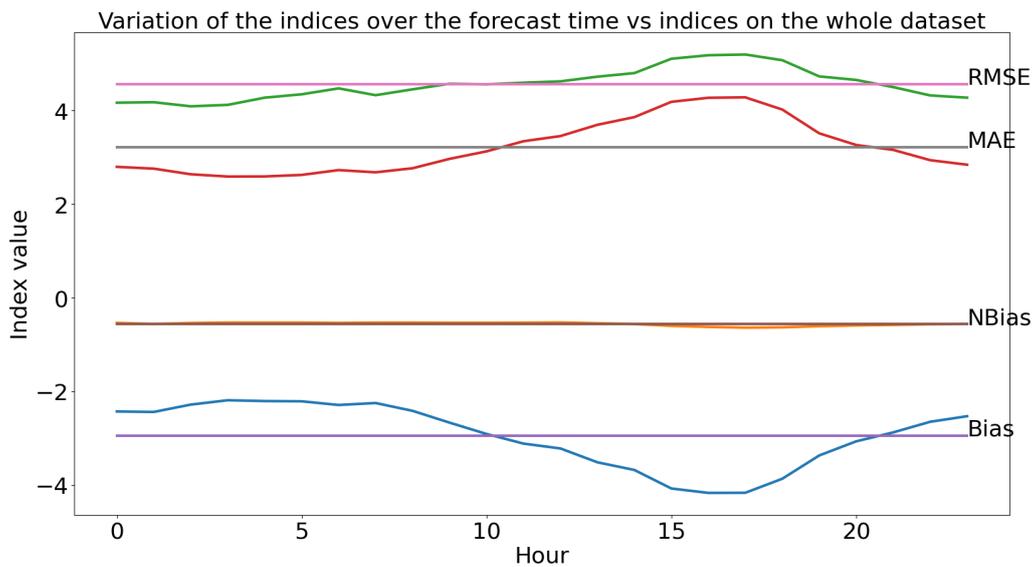


Figure 5.29: Trend over time of the RMSE, MAE, NBias, Bias indices.

Differently from before, the normalized indices have a worsening during the night and an improvement during the day. This is due to the fact that during the day, as seen in figure 5.17, wind speeds are greater than at night. It is therefore likely that non-normalized indices show greater accuracy when

Table 5.6: Table of the error index as the time limit changes, mean EPS 0-24 h forecast.

Indici	Bias	NBias	MAE	NMAE	RMSE	NRMSE	HH	SI	Pearson
0H	-2.433	-0.540	2.799	0.621	4.170	0.692	1.176	0.562	0.575
1H	-2.443	-0.561	2.760	0.634	4.181	0.705	1.250	0.572	0.703
2H	-2.286	-0.542	2.640	0.626	4.094	0.697	1.217	0.578	0.733
3H	-2.193	-0.530	2.592	0.627	4.124	0.699	1.223	0.592	0.737
4H	-2.210	-0.531	2.593	0.623	4.278	0.703	1.240	0.602	0.760
5H	-2.215	-0.531	2.626	0.629	4.351	0.708	1.257	0.609	0.752
6H	-2.294	-0.539	2.730	0.641	4.477	0.715	1.286	0.614	0.730
7H	-2.252	-0.532	2.682	0.633	4.331	0.708	1.249	0.604	0.694
8H	-2.419	-0.531	2.769	0.608	4.457	0.703	1.226	0.590	0.640
9H	-2.671	-0.535	2.969	0.595	4.576	0.693	1.185	0.563	0.582
10H	-2.914	-0.535	3.130	0.575	4.563	0.671	1.112	0.516	0.560
11H	-3.120	-0.533	3.346	0.572	4.599	0.654	1.061	0.480	0.532
12H	-3.224	-0.528	3.457	0.566	4.627	0.646	1.035	0.463	0.481
13H	-3.520	-0.547	3.698	0.575	4.727	0.641	1.028	0.428	0.495
14H	-3.684	-0.563	3.863	0.590	4.805	0.648	1.051	0.416	0.467
15H	-4.080	-0.603	4.189	0.619	5.111	0.669	1.129	0.403	0.515
16H	-4.174	-0.625	4.276	0.640	5.186	0.683	1.184	0.405	0.581
17H	-4.172	-0.641	4.285	0.658	5.202	0.696	1.236	0.416	0.631
18H	-3.872	-0.634	4.024	0.659	5.080	0.703	1.261	0.455	0.638
19H	-3.373	-0.609	3.516	0.635	4.733	0.700	1.241	0.491	0.647
20H	-3.071	-0.593	3.266	0.631	4.657	0.708	1.266	0.532	0.634
21H	-2.880	-0.582	3.161	0.639	4.504	0.707	1.257	0.543	0.622
22H	-2.653	-0.567	2.942	0.629	4.327	0.702	1.237	0.554	0.661
23H	-2.534	-0.559	2.844	0.628	4.278	0.702	1.238	0.566	0.681

speeds are lower and less accuracy when greater than normalized indices.

Figure 5.31 shows how the correlation varies. The first value (hour 0) is a bit low but it is due to the fact that the model is still setting itself. It is quite clear how the correlation decreases during the day and increases during the night; this is normal, in fact the maximum errors are during the day, due to convection that the models struggle to describe accurately with a low resolution. It is however important to note that, for when it resumes, the correlation at after 6 pm is lower than that of the hours before 8 am. This result is correct because as the hours advance the forecasts become more difficult and therefore there is a loss in correlation.

Finally, table 5.7 is shown containing the indices of the error of the power forecast with the raw 0-24 hour forecasts. As can also be seen from figure

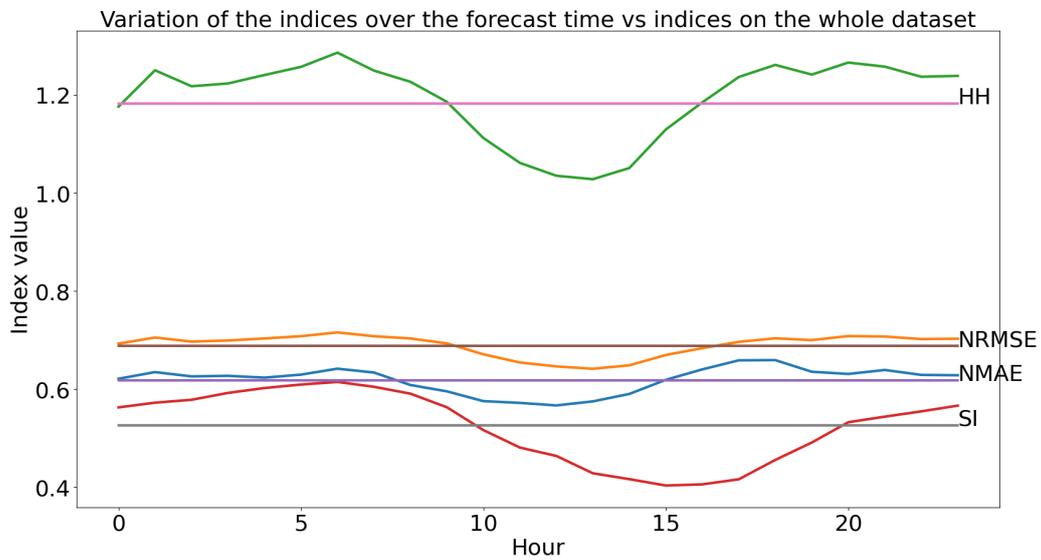


Figure 5.30: Trend over time of the RMSE, MAE, NBias, Bias indices.

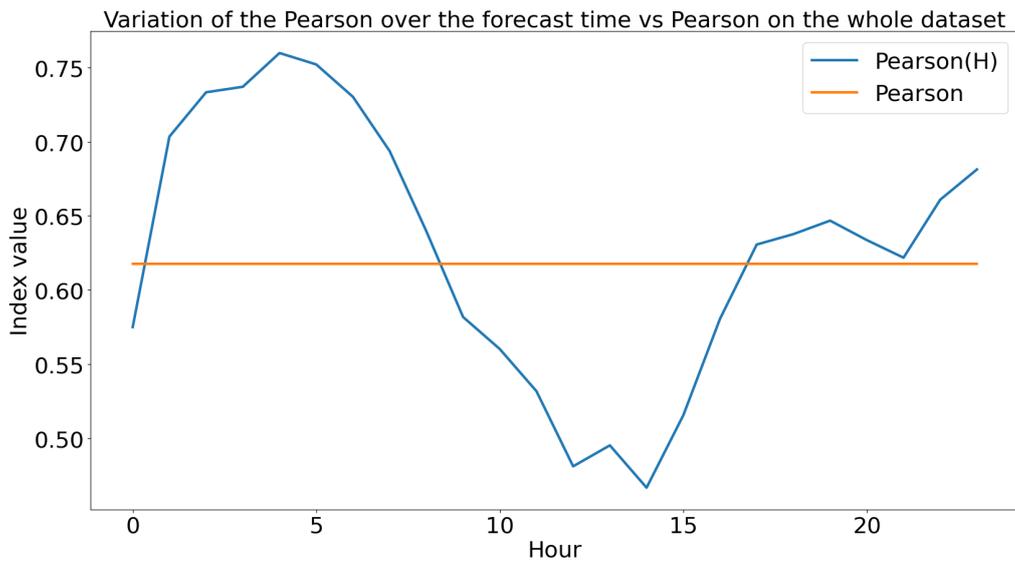


Figure 5.31: Trend over time of the correlation index.

5.32, with the transfer function based on observed power and wind forecasts, there are very large errors.

Once the forecasts provided by the European center have been analyzed, it is possible to correct them.

Table 5.7: Index value of power prediction with raw wind speed forecasts.

Bias	-347.382
NBias	-0.954
MAE	349.149
NMAE	0.959
RMSE	638.459
NRMSE	0.962
HH	4.841
SI	0.8071
Pearson	0.410

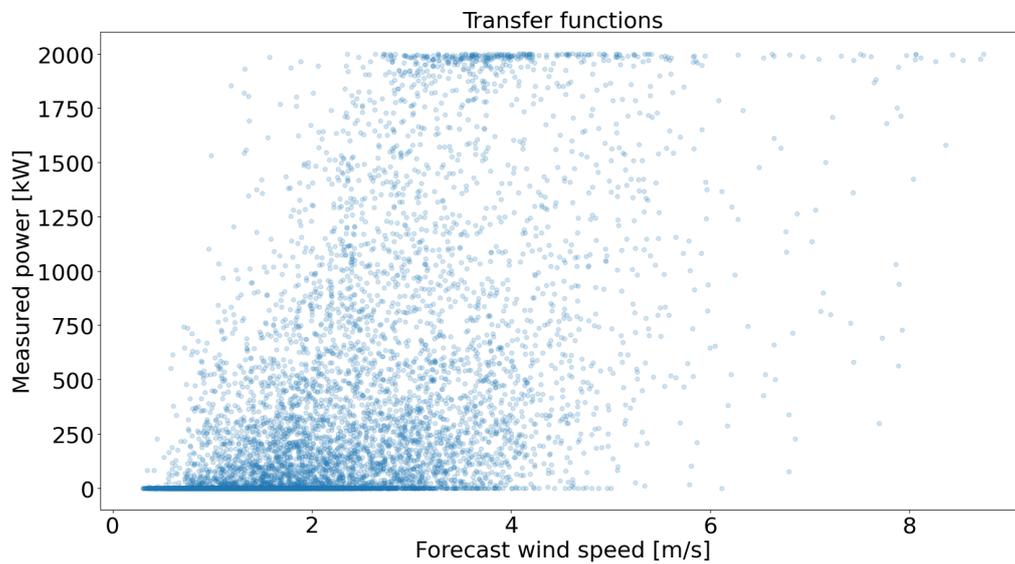


Figure 5.32: Transfer function with measured power and forecast wind speed.

Chapter 6

Correction of the forecasts

6.1 Introduction

In this chapter it will be shown step by step how the forecasts were corrected. First, the algorithms considered to be the best for correction will be evaluated and identified. The corrections will then be made by exploiting all the available parameters that have been presented in the previous chapter. In fact, the objective is to condition the forecasts as a function of other variables such as the time or speed in order to find non-linear correlations that are hidden considering the whole forecasts. Finally, a comparison will be made with corrections made through the Machine Learning algorithm in order to evaluate which and when is the best technique to use.

6.2 The best correction techniques

6.2.1 Linear regression

Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable.

Before attempting to fit a linear model to observed data, a modeller should first determine whether or not there is a relationship between the variables of interest. This does not necessarily imply that one variable causes the other, but that there is some significant association between the two variables. A scatter plot can be a helpful tool in determining the strength of the relationship between two variables.

Closely linked to regression is the concept of correlation, in fact:

- In the theory of (simple) regression it is assumed that a variable X assumes certain values and the relationship that links the second variable Y to the first is sought: in other words, an attempt is made to establish a functional link between the two variables.
- In the theory of correlation the degree of interdependence between the two variables is determined, that is, it is determined whether a variation of the character X corresponds to a more or less sensitive variation of the character Y

The type of regression that will be used is called least squares regression. By denoting the estimated independent variable with X and the estimated dependent variable with Y , the aim is to determining real coefficients a and b for which the following linear relationship exists between the two variables:

$$Y = a + bX \quad (6.1)$$

The coefficient a is called intercept and represents the value of the variable Y when $X = 0$; while b is called the angular coefficient or regression coefficient or, again, slope of the line and represents the variation undergone on average by the character Y due to a unit increase of the character X . [?] Basically, linear regression identifies the mean of the variables using the least squares technique. In this way, by definition, the best values of the intercept and of the angular coefficient are obtained to best minimize the values of the Bias and RMSE indices.

6.2.2 Quantile regression

Quantile regression is a type of regression analysis used in statistics. Whereas the method of least squares estimates the conditional mean of the response variable across values of the predictor variables, quantile regression estimates the conditional median (or other quantiles) of the response variable. Quantile regression is an extension of linear regression used when the conditions of linear regression are not met.

One advantage of quantile regression relative to ordinary least squares regression is that the quantile regression estimates are more robust against outliers in the response measurements. However, the main attraction of quantile regression goes beyond this and is advantageous when conditional quantile functions are of interest. Different measures of central tendency and statistical dispersion can be useful to obtain a more comprehensive analysis of the relationship between variables.[11]

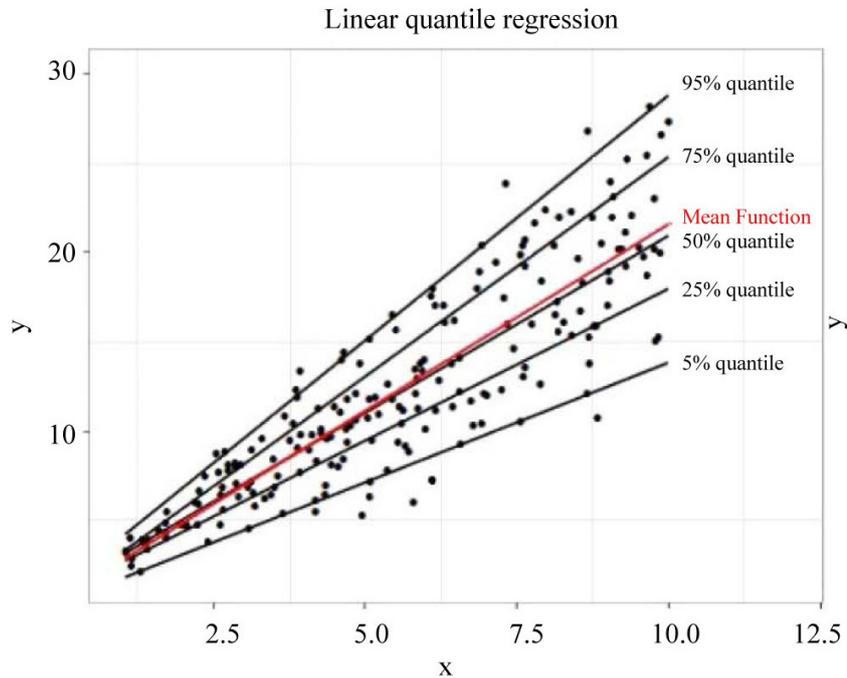


Figure 6.1: Data fitted with quantile regression and linear regression (red line).

Figure 6.1 shows how if the data have a non-Gaussian probability distribution, in which mean and median correspond, different values are obtained using linear regression and quantile regression. Quantile regression is also not limited to averaging. In fact, the calculation of the median corresponds very simply to calculating the quantile of 50%. It is therefore also possible to calculate quantiles corresponding to other percentages. In this way, in addition to obtaining the best forecast, it is also possible to provide confidence intervals with the respective probabilities of exceeding.

In conclusion, quantile regression will be used to correct the wind forecast at the expense of linear regression. This is mainly due to 2 reasons.

- The first is that quantile regression also providing the confidence intervals gives the possibility to understand how reliable the forecast is or not.
- The second point is that the median is identified with this regression. Identifying the median means finding the optimal value to minimize the NMAE and in view of the fact that the forecasts will then be used to predict the power produced by a wind power plant, the index to be

taken into consideration (as seen in the previous chapters) is just the NMAE.

6.2.3 Gamma distribution

During the past decade, the use of forecast ensembles for assessing the uncertainty of numerical weather predictions has become routine. Three operational methods for the generation of synoptic-scale ensembles have been developed: one is the singular vector method used by the European Centre for Medium-Range Weather Forecasts. The ability of ensemble systems to improve deterministic-style forecasts and to predict forecast skill has been convincingly established. Statistically significant spread-error correlations suggest that ensemble variance and related measures of ensemble spread are skilful indicators of the accuracy of the ensemble mean forecast.

Case studies in probabilistic weather forecasting have typically focused on the prediction of categorical events. Ensembles also allow for probabilistic forecasts of continuous weather variables, such as air pressure and temperature, which are ideally expressed in terms of predictive cumulative distribution functions (CDFs) or predictive probability density functions (PDFs). However, ensemble systems are finite and typically include of 5 to 50 member models. Hence, raw ensemble output does not provide predictive PDFs, and some form of postprocessing is required.

In this section, is shown the use of ensemble model output statistics (EMOS), an easy to implement statistical postprocessing technique that addresses the aforementioned issues. The method is a variant of multiple linear regression or model output statistics (MOS) techniques that have traditionally been used for deterministic-style and probability of precipitation forecasts.[14] Specifically, suppose that X_1, \dots, X_m denotes an ensemble of individually distinguishable forecasts for a univariate weather quantity Y . A multiple linear regression equation for Y in terms of the ensemble member forecasts can be written as

$$Y = a + b_1X_1 + \dots + g_mX_m + \varepsilon \quad (6.2)$$

where a and b_1, \dots, b_m are regression coefficients, and where ε is an error term that averages to zero. Regression approaches of this type have been shown to improve the deterministic-style forecast accuracy of synoptic weather and seasonal climate ensembles, and the associated forecast systems have been referred to as superensembles. The use of regression techniques for probabilistic forecasting has not received much attention in the literature, except for forecasts of binary events. With this approach, is obtain full

predictive PDFs and CDFs from ensemble forecasts of a continuous weather variable. Standard regression theory suggests a straightforward way of constructing predictive PDFs and CDFs from a regression equation, by taking them to be Gaussian with predictive mean equal to the regression estimate, and predictive variance equal to the mean squared prediction error for the training data. This approach corrects for model biases and takes account of dispersion errors. However, the resulting assessment of uncertainty is static, in that the predictive variance is independent of the ensemble spread, thereby negating the spread-skill relationship. Hence, the variance was modelled of the error term in the following equation as a linear function of the ensemble spread, that is,

$$Var(\varepsilon) = c + dS^2 \quad (6.3)$$

where S^2 is the ensemble variance, and where c and d are nonnegative coefficients. Combining 6.2 and 6.8 yields the Gaussian predictive distribution

$$\mathfrak{N}(a + b_1X_1 + \dots b_mX_m, c + dS^2) \quad (6.4)$$

whose mean derives from the regression equation and forms a bias-corrected weighted average of the ensemble member forecasts, and whose variance depends linearly on the ensemble variance. It refers to the resulting predictive PDFs and CDFs as ensemble model output statistics or EMOS forecasts.

For estimating the EMOS coefficients it was used the novel approach of minimum CRPS estimation, which forms a special case of minimum contrast estimation (MCE). This method is best explained in terms of verification measures.[12][13] The CRPS is the integral of the Brier scores at all possible threshold values t for the continuous predictand. Specifically, if F is the predictive CDF and y is the verifying observation, the continuous ranked probability score is defined as

$$crps(F, y) = \int_{-\infty}^{\infty} [F(t) - H(t - y)]^2, dt \quad (6.5)$$

where $H(t-y)$ denotes the Heaviside function and takes the value 0 when $t < y$ and the value 1 otherwise. Applications of the continuous ranked probability score have been hampered by a lack of closed-form expressions for the associated integral.

It is therefore necessary to identify a probability distribution that best represents the data under analysis.

Figure 6.2 shows how the probability distribution of conditional observations at different speed intervals of the predictions is shown. The distributions show a similar trend between them. In fact, they do not have a symmetrical

trend but instead seem to be subject to an important skewness. The left tail is in fact for all cases shorter than the right curve which decays very slowly.

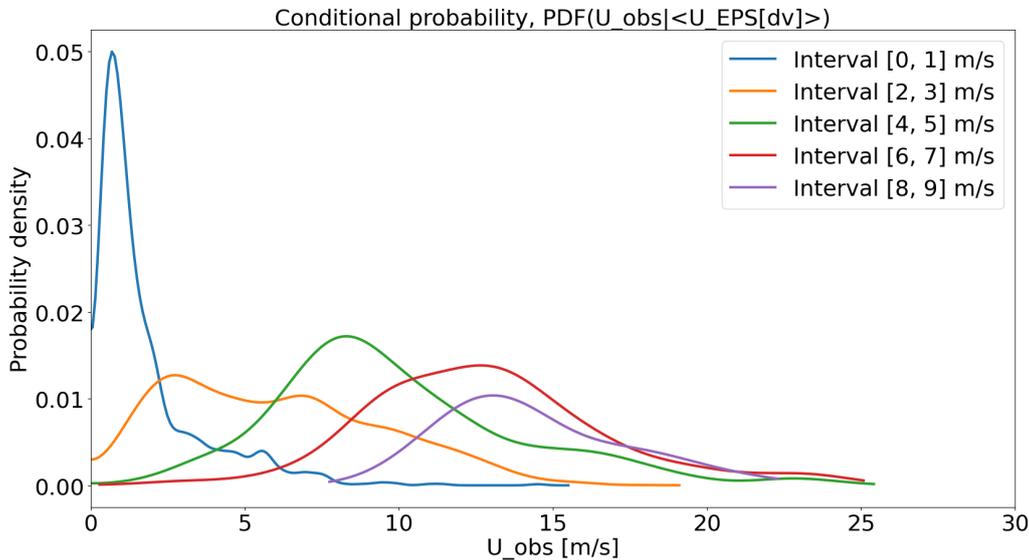


Figure 6.2: Density function probability of observed wind speed for different intervals of wind speed forecast.

There are several probability distributions with such characteristics. To identify the probability distribution that best describes the data in possession, it is necessary to look at figure 6.3. This graph was made in a semi-logarithmic scale to better evaluate the behaviour of the queues. A comparison is therefore shown between a gamma distribution and the probability density of the observations conditioning to the period of the day between 10 and 15, to a forecast of wind speed between 1.5 and 2.5 m/s and also conditioning to a variance of the EPS between 0.3 and 0.7 m^2/s^2 . In this comparison it is evident that the two curves are very similar and therefore it can be assumed that the gamma distribution is the best that can describe the data under analysis. This is good because of the gamma distribution it is possible to calculate its parameters through techniques such as that of maximum likelihood or even CRPS which, as previously mentioned, is the technique that is considered the best.

In probability theory and statistics, the gamma distribution is a two-parameter family of continuous probability distributions. There are three different parametrizations in common use:

- With a shape parameter k and a scale parameter θ .

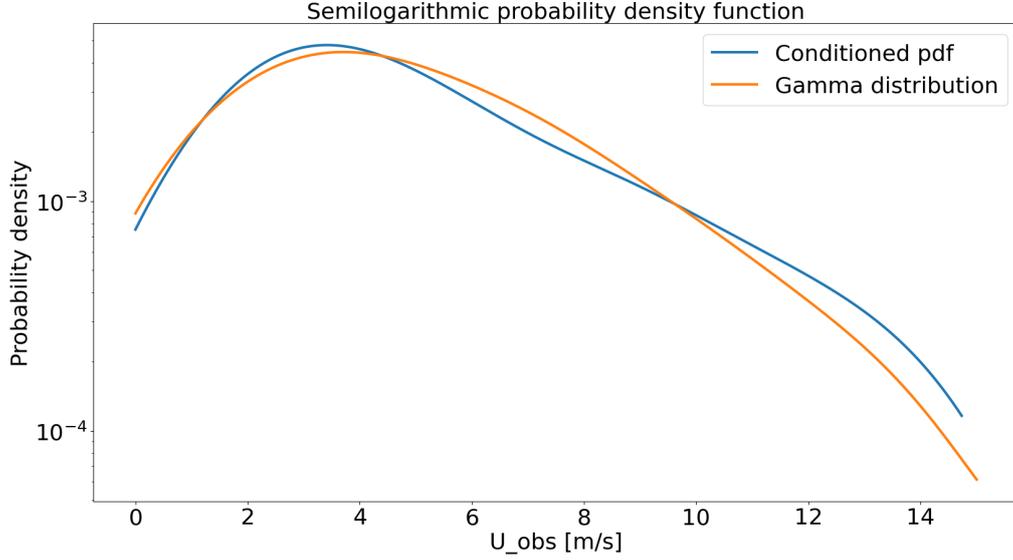


Figure 6.3: Probability density in semilogarithmic scale (y axis) of the observations of the conditioned wind speed to: observed wind from hour 10 to 15, with a forecast speed between 1.5 and 2.5 m/s and with a variance of the EPS as between 0.3 and 0.7.

- With a shape parameter $\alpha = k$ and an inverse scale parameter $\beta = 1/\theta$, called a rate parameter.
- With a shape parameter k and a mean parameter $\mu = k\theta = \alpha/\beta$.

The parametrization with k and θ appears to be more common in econometrics and certain other applied fields, where for example the gamma distribution is frequently used to model waiting times. For instance, in life testing, the waiting time until death is a random variable that is frequently modelled with a gamma distribution.[15] The figure 6.4 shows how the gamma distribution vary at the variation of the parameters k and θ .

For future analyses, the k and θ parameters will be used. Where:

- $k = M^2/\sigma^2$
- $\theta = \sigma^2/M$

With M representing the mean of the predictions and θ the variance. So, going to minimize the CRPS function:

$$CRPS = \frac{1}{N} \sum_{i=1}^N \left\{ O_i \left[2P \left(K_i, \frac{O_i}{\theta_i} \right) - 1 \right] - K_i \theta_i \left[2P \left(K_i + 1, \frac{O_i}{\theta_i} \right) - 1 \right] - \frac{\theta_i}{\beta \left(\frac{1}{2}, k \right)} \right\} \quad (6.6)$$

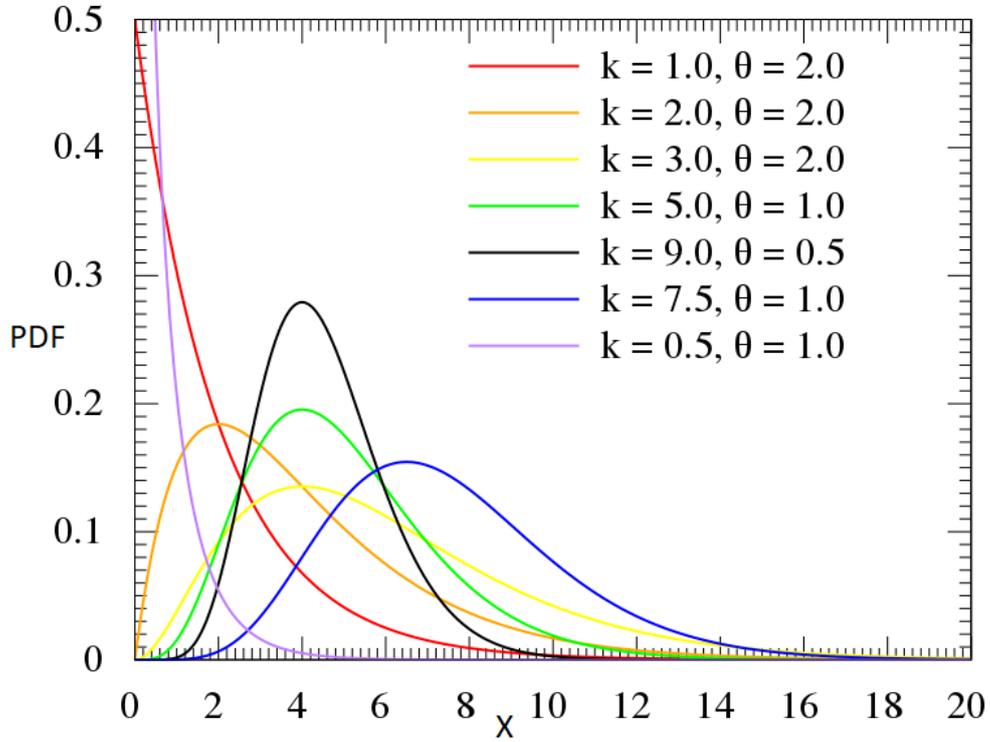


Figure 6.4: The probability density function on y axis of Gamma distribution for x variable for different parameters.

Where:

- $K_i = (a + bF_i) / (c + dS_i^2)$
- $\theta_i = (c + dS_i^2) / (a + bF_i)$
- P = the incomplete function of gamma
- β = is the beta function
- O_i = the i observation

So knowing F_i (the forecasts) and S_i^2 (the variance of the forecasts) it is possible to determine the parameters a, b, c, d which allow to calibrate the forecast (F_i^{cal}) and its variance (S_i^{2cal}) as follow:

- $F_i^{cal} = a + bF_i$
- $S_i^{2cal} = c + dS_i^2$

Therefore, to conclude, two techniques will be used for the correction of the forecasts: the first is that which, using quantile regression, provides the best result in terms of NMAE, however, giving a non-dynamic variance as the reliability of the forecast; the second is that using the gamma distribution with which the best punctual forecast is not obtained but rather the best probability distribution of the forecast by exploiting the variance of the EPS dynamically.

6.3 The strategies for correcting the forecasts

This section goes into actual forecast correction. In the various subsections it will be shown how the error indices vary according to the correction strategy adopted. The goal is to find elements that lead to a correlation between forecasts and observation to better calibrate the forecast.

All corrections apart for the first and the one that will be considered to be the best, will be shown using the one obtained from the average of the EPS as a starting forecast and using the quantile regression to calibrate the parameters. Only with the best correction strategy will it also be shown how the HRES forecast varies to compare it with that of EPS. The correction will also be shown using the gamma distribution both to make a comparison but above all to see the resulting distribution.

As regards the correction of the forecasts with a forecast horizon of 0-24 hours, they will be recorded as follows. The first strategies will be of the “static” type. In other words, strategies will be used that provide for the division of the datasets according to different variables but without ever considering the observed data of the same day. For these strategies, the full correction of the forecasts from 0 to 24 hours will be shown. Further on, other strategies will be used that exploit the fact that 0-24 hour forecasts are made available after about 7 hours and therefore there are already field observations ranging from 0 to 7 hours; this strategies are of the “dynamic” type. For this reason, it makes no sense to correct the forecasts previous to 7 because it would be enough to replace the value of the observation as it would be the past. Therefore, only the hours ranging from 8 to 24 will be corrected and shown.

This assumption is absolutely legitimate not only because the forecasts arrive 7 hours late due to the time needed to make the models of ECMWF appear but also because the electricity market closes after this time and therefore there is time to enter in possession of the wind observations in those hours.

6.3.1 The correction on the forecasts without conditioning

This first subsection shows how the error indices improve by calibrating the parameters without using any kind of division or particular technique.

Table 6.1 shows the error indices calculated before and after the correction for cases 0-24 and 24-48 hours. The Bias and the NBias decrease considerably but without tending to 0, they also continue to have a non-negligible value. In fact, the NBias is between 17 and 18%. This is due to the fact that quantile regression and gamma distribution regression do not aim to minimize the Bias, as does linear regression. Furthermore, it must be considered that even in the case of linear regression, an absolute 0 would not be reached because, since the training dataset is different from the verification one, they have slightly different characteristics due to the normal variation between year and year. If hypothetically infinite years were available in the past and in the future, then one could search for the 0 fulfilled (assuming that there are no variables such as global warming etc ...). The indices of MAE and NMAE, as expected, reach their lowest value thanks to quantile regression. On average there is an improvement of about 15% and between quantile regression (QR) and regression due to gamma distribution (GM) there is a difference of only 1.5 percentage points. This confirms the fact that the gamma distribution is suitable for the data in possession. Regarding the RMSE and the NRMSE there is a better result using the gamma distribution. As confirms the HH index the gamma distribution, between the two, is the most suitable technique to correct the forecast. However, as the most important index is NMAE, quantile regression remains necessary. SI improves without showing particular differences between the two techniques. Finally, it is interesting to note that the Pearson has not changed its value in the slightest. This is absolutely normal; in fact the Pearson does not undergo any variation if a certain dataset is multiplied or added to some value. This is because doing so changes the dataset linearly. The Pearson changes when there are variations that are not linear.

Table 6.2 shows the error indices starting from the HRES forecast. Also in this case there is the same results as in table 6.1 at the qualitative level. Going however to compare the two tables, it is noticed something that was previously supposed. That is, the indices obtained following the correction are worse despite starting from better values. This is precisely due to Pearson and therefore confirms the fact that this index is the most important to understand which forecast after a calibration contains the most correct and reliable information.

Figures 6.5 and 6.6 show how the forecasts have varied due to the correc-

Table 6.1: Error indices of EPS mean before and after correction for the 0-24 and 24-48 h forecasts, QR represents the correction using quantile regression while GM using the gamma distribution.

Index	0-24 h forecast			24-48 h forecast		
	Raw	QR	GM	Raw	QR	GM
Bias	-2.945	-0.923	-0.067	-2.941	-0.896	-0.038
NBias	-0.566	-0.177	-0.013	-0.565	-0.172	-0.007
MAE	3.215	2.443	2.532	3.223	2.477	2.572
NMAE	0.618	0.469	0.486	0.619	0.476	0.494
RMSE	4.572	3.337	3.217	4.599	3.383	3.266
NRMSE	0.687	0.502	0.484	0.692	0.509	0.491
HH	1.182	0.619	0.561	1.196	0.628	0.569
SI	0.526	0.482	0.484	0.532	0.491	0.491
Pearson	0.632	0.632	0.632	0.615	0.615	0.615

Table 6.2: Error indices of HRES before and after correction for the 0-24 and 24-48 h forecasts, QR represents the correction using quantile regression while GM using the gamma distribution.

Index	0-24 h forecast			24-48 h forecast		
	Raw	QR	GM	Raw	QR	GM
Bias	-2.637	-0.996	-0.058	-2.656	-1.061	-0.057
NBias	-0.507	-0.191	-0.011	-0.510	-0.204	-0.011
MAE	3.013	2.531	2.612	3.055	2.612	2.698
NMAE	0.579	0.486	0.502	0.587	0.502	0.518
RMSE	4.247	3.393	3.259	4.322	3.516	3.361
NRMSE	0.639	0.510	0.490	0.650	0.529	0.505
HH	0.999	0.636	0.571	1.029	0.670	0.593
SI	0.501	0.488	0.490	0.513	0.504	0.505
Pearson	0.621	0.621	0.621	0.586	0.586	0.586

tion used. In the first there is the scatter plot with respect to the observations before and after the correction of the forecasts. As it can be seen the points tend to be more on the diagonal following the correction while before they were conspicuously more to the left to indicate the underestimation of the forecasts. It is not appreciable to the eye but as there is still a negative NBias (almost 20%) the points should tend to stay a little to the left of the diagonal. The second figure provides a view over a time series of 5 weeks

in which the large difference between the forecasts before and after the correction is shown. Here, too, the forecasts become much more reliable and less underestimated, although as mentioned earlier an underestimate of the observation persists.

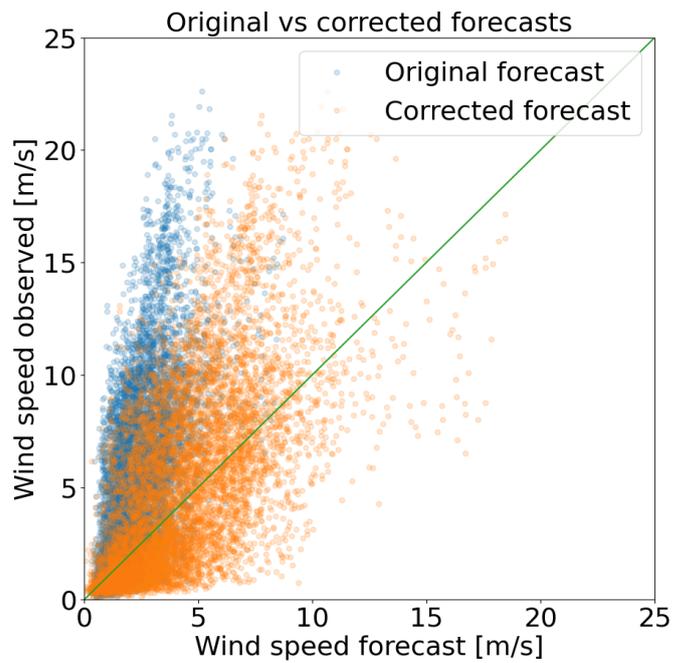


Figure 6.5: Scatter plot of forecasts before and after correction vs observation. The forecast are the EPS mean with a forecast horizon of 24-48 hours. The correction used is the quantile regression.

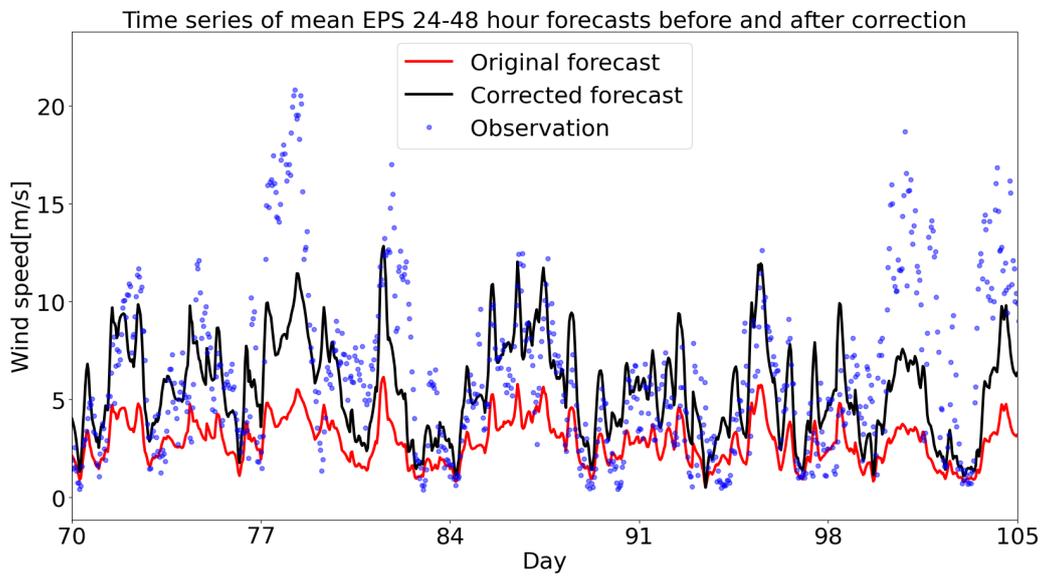


Figure 6.6: Time series of the forecasts before and after the correction from 12/3/2018 to 16/4/2018. The forecasts are the EPS mean with a forecast horizon of 24-48 hours. The correction used is the quantile regression.

6.3.2 Correction dividing for the directions

In this section will be analyse the correction that exploits the different wind directions. As shown in figure 5.21, the wind farm is located in a territory with a very complex orography. This complexity is due to the presence of the mountains around the site. Therefore, analysing the wind directions both observed and predicted it was found that there were two main directions from which most of the winds came and which statistically had greater intensity. In the forecasts the distinction was slightly less sophisticated, but the four quadrants were clearly distinguished.

As already announced, the tactic is to condition the forecasts according to direction in order to find correlations that would otherwise be hidden.

Figure 6.7 shows the scatter plots for the 4 wind directions. Compared to the scatter plot that was shown previously, it can be seen how the point cloud, for quadrants 1 and 3, is much narrower. It is also noted that scatter plots have very different characteristics from each other. This leads to think that the technique of conditioning the directions can bring great benefits because it would seem to have been able to identify a parameter that gives the possibility to find those correlations that were previously hidden.

Table 6.3 shows the error indices of the 24-48 hour forecasts all together and conditioned by the 4 directions. The most important index is precisely the Pearson which varies considerably from direction to direction. As mentioned above, in fact, the correlation of the 1st and 3rd quadrant is much greater than the total one as well as the 2nd and 3rd quadrant have a much lower correlation.

Table 6.3: Error indices for raw predictions (EPS average) 24-48 hours for: whole dataset merged and divided into 4 quadrants.

Index	24-48 h forecast				
	Total	1 quad	2 quad	3 quad	4 quad
Bias	-2.945	-1.757	-1.134	-5.191	-1.246
NBias	-0.566	-0.419	-0.356	-0.692	-0.438
MAE	3.215	2.028	1.788	5.326	1.651
NMAE	0.618	0.484	0.562	0.709	0.580
RMSE	4.572	2.846	2.576	6.562	2.606
NRMSE	0.687	0.547	0.638	0.740	0.674
HH	1.182	0.781	0.946	1.431	1.082
SI	0.526	0.431	0.573	0.453	0.592
Pearson	0.632	0.766	0.362	0.708	0.512

Scatter plot: forecasts vs observations for the 4 quadrants

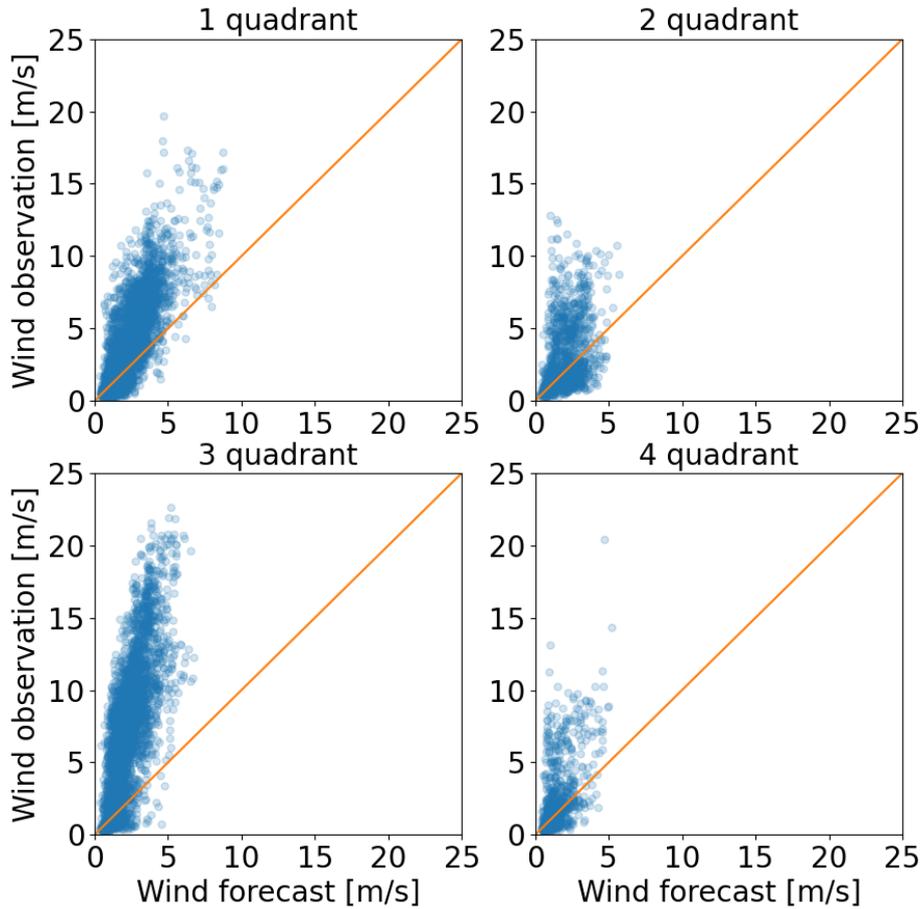


Figure 6.7: Scatter plot: comparison between observations and forecasts 24-48 of the average of the EPS for the 4 main directions.

Going to perform a quantile regression (and even if not shown using the gamma distribution) is obtained the results reported in table 6.4. All the indices have undergone a clear improvement compared to the simple calibration carried out in the previous sub-chapter and consequently also with respect to the values of departure. The NBias begins to tend towards zero, settling around 6/7% (10% less than the tt correction). The NMAE undergoes an improvement of about 8/9% on the tt correction for an overall -23% about the raw forecast. It is very interesting to underline the strong variation that correlations have undergone. In fact, the Pearson, which had previously re-

mained constant, has undergone an increase of almost 15 percentage points. It is a notable improvement that shows once again how conditioning on the directions has been a winning tactic. The increase in Pearson is since compared to before, having been performed 4 different regressions, the linearity condition necessary to keep the Pearson constant has been lost.

Table 6.4: Comparison between the various indices of the error for the raw forecasts (mean EPS, 0-24 and 24-48 hours), corrected without divisions (tt), and corrected by conditioning for the directions.

Index	0-24 h forecast			24-48 h forecast		
	Raw	tt	dr	Raw	tt	dr
Bias	-2.945	-0.923	-0.349	-2.941	-0.896	-0.329
NBias	-0.566	-0.177	-0.067	-0.565	-0.172	-0.063
MAE	3.215	2.443	2.009	3.223	2.477	2.072
NMAE	0.618	0.469	0.386	0.619	0.476	0.398
RMSE	4.572	3.337	2.659	4.599	3.383	2.760
NRMSE	0.687	0.502	0.400	0.692	0.509	0.415
HH	1.182	0.619	0.451	1.196	0.628	0.469
SI	0.526	0.482	0.396	0.532	0.491	0.412
Pearson	0.632	0.632	0.773	0.615	0.615	0.750

Finally in Figure 6.8 the scatter plots of the raw predictions vs the observations were superimposed with that of the predictions corrected with the direction vs the observations. Compared to the scatter plot 6.5 can appreciate how the point cloud is much better around the diagonals. In fact, it is less wide (due to the raising of Person) better distributed between the right and left side of the diagonal (due to the lowering of the Bias) and finally better lying along the diagonal (due to the lowering of MAE, RMSE and HH).

Finally, figure 6.9 shows the difference between the forecast obtained without conditioning (tt) with respect to that conditioning on the directions (dr). Briefly the difference is not as evident as it is for the error indices. The improvement in forecasts is very evident also from a graphic point of view.

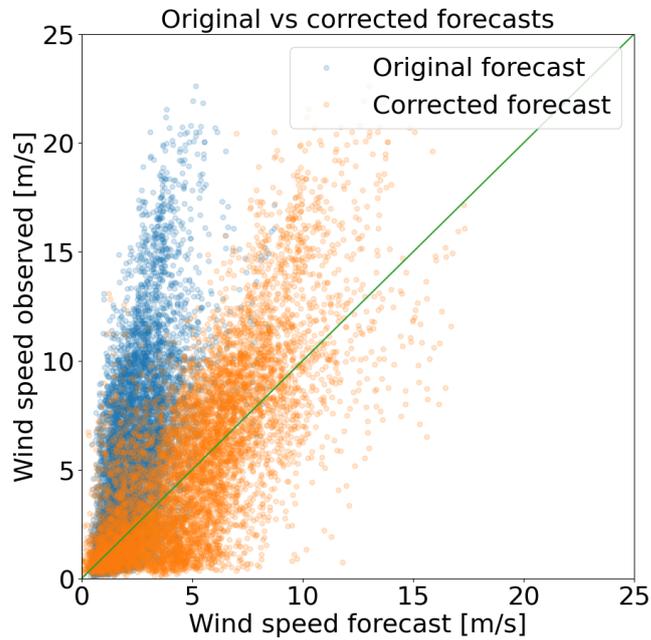


Figure 6.8: Scatter plot of forecasts (mean EPS, 24-48 hours) vs observations before and after the correction, influencing the directions.

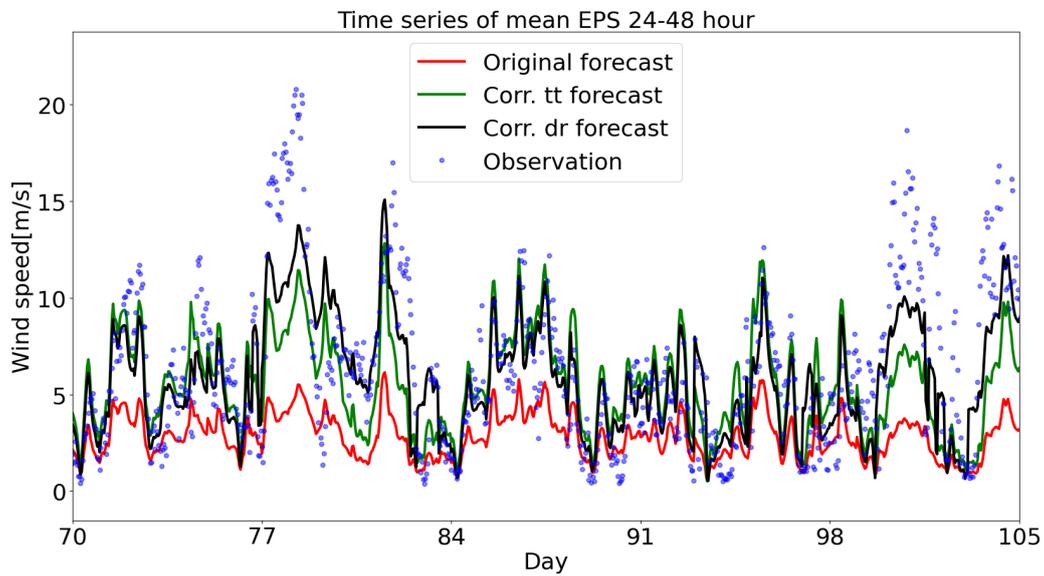


Figure 6.9: Time series of the forecasts before and after the correction from 12/3/2018 to 16/4/2018. The forecasts are the EPS mean with a forecast horizon of 24-48 hours. The correction used is the quantile regression with two techniques: using all the dataset (tt), conditioning on the directions(dr).

6.3.3 Correction dividing for day hours

As shown in table 5.6 and in figure 5.31, the Pearson index does not remain constant throughout the day. It has been highlighted that there is less correlation in the daytime hours and that moving away from the time in which the forecast was made, the correlation decreases.

It may therefore make sense to try to divide the dataset according to the time limit, in order to make the most of the correlation present. It is probable that, if there is a systematic error that depends on the time, by dividing the dataset, it would be possible to correct it. By dividing by the hours of the day, a correlation is sought mainly linked to two aspects: the first is due to the physical phenomena that have different behaviours as the hours vary; the second is due to the time that has passed since the start of the forecast made by the model for which there could be significant variations.

Figure 6.10 shows how actually going to divide for hours the scatter plots take on a more defined shape. Furthermore, as already seen thanks to the Pearson index, it is clear that the correlation in daylight hours is lower; in fact the point cloud is much larger than that of the daytime hours.

Table 6.5 shows the error indices with the correction of the forecasts without conditions and of that made by conditioning the hours of the day. All the indices, including the Pearson, undergo an improvement albeit limited. The NMAE, NRMSE and Pearson all improved by about 2%. Compared to conditioning with the management it is undoubtedly worse but there is still a not negligible improvement.

Table 6.5: Comparison between the various indices of the error for the raw forecasts (mean EPS, 0-24 and 24-48 hours), corrected without divisions (tt), and corrected by conditioning for the day hours.

Index	0-24 h forecast			24-48 h forecast		
	Raw	tt	dH	Raw	tt	dH
Bias	-2.945	-0.923	-0.768	-2.941	-0.896	-0.710
NBias	-0.566	-0.177	-0.148	-0.565	-0.172	-0.136
MAE	3.215	2.443	2.341	3.223	2.477	2.382
NMAE	0.618	0.469	0.450	0.619	0.476	0.458
RMSE	4.572	3.337	3.225	4.599	3.383	3.273
NRMSE	0.687	0.502	0.485	0.692	0.509	0.492
HH	1.182	0.619	0.588	1.196	0.628	0.595
SI	0.526	0.482	0.471	0.532	0.491	0.481
Pearson	0.632	0.632	0.654	0.615	0.615	0.635

Scatter plot: forecasts vs observations for forecasts horizon

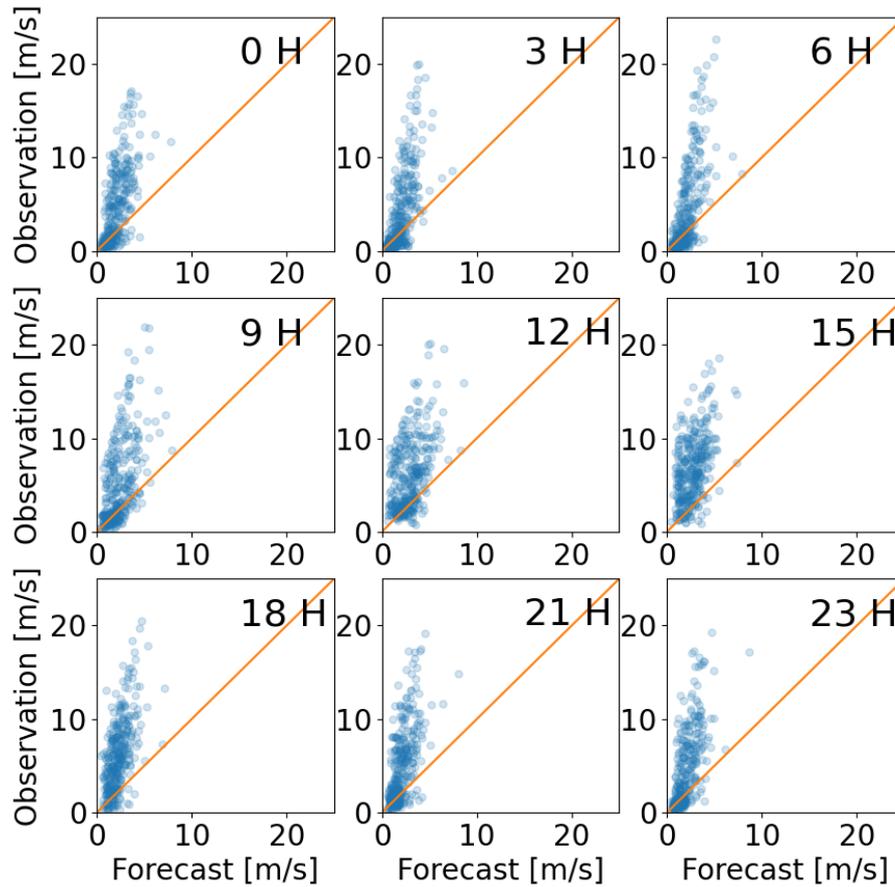


Figure 6.10: Scatter plot of forecasts (mean EPS, 24-48 hours) vs observations for the different hours of the day (0, 3, 6, 9, 12, 15, 18, 21 and 23 hour)

It is interesting to analyse the MAE from a graphic point of view. Figure 6.11 plots the Bias and MAE indices before and after making the correction as the hours change.

As already seen in the previous chapter and as the graph reconfirms, daytime hours are those with the worst forecasts. Looking at the graph, the MAE undergoes a greater improvement during the hours of the day. The fact is curious because from graph 5.31 it was seen that during the hours of the day it was where there was less correlation. This improvement must

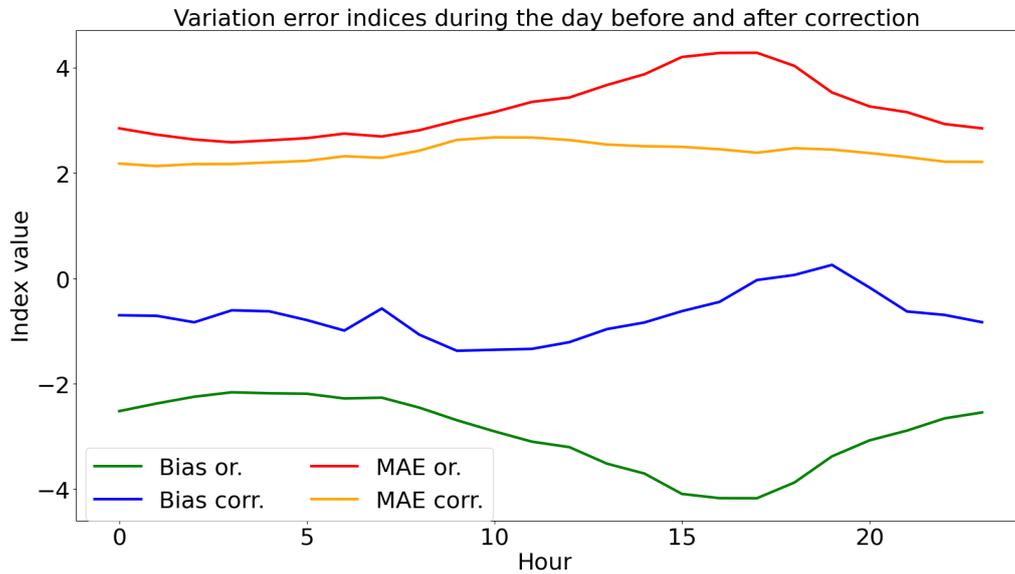


Figure 6.11: Comparison between MAE and Bias before and after correction, depending on the time of day. They are shown how they change as the hour changes.

therefore be sought in another index, the Bias. In fact, the Bias shows the systematic error of the forecast. The systematic error most of the time is the most easily correctable in the predictions. Therefore, the improvement of the MAE in the day's hours is due to the fact that in those hours the greatest systematic error is recorded and therefore the other indices such as the MAE benefit more from the correction in these hours than the others. From an economic point of view this is good news as the wind during the day is more difficult to predict. Furthermore, during the day it is the time when there is more demand for energy, and it is therefore necessary to have a good forecast.

At this point, having verified that conditioning on the hours leads to a benefit in terms of forecasts, it is possible to combine the two conditions made up to now, thus simultaneously conditioning for the direction and for the time.

Table 6.6 shows the error indices for forecasts 0-24 and 24-48 hours calculated on the raw forecast, on the forecast calculated by conditioning on the directions (dr) and on the forecast calculated by conditioning on the direction and time (drdH). As expected, the indices all improve even more than the difference between the "tt" and "dH" correction (table 6.5). This means that there is also a correlation between time and direction of the day. So if

you take a wind with a certain direction, as the time of day changes, it will show a different phenomenology.

Table 6.6: Comparison between the various indices of the error for the raw forecasts (mean EPS, 0-24 and 24-48 hours), corrected by conditioning for the directions(dr), and corrected by conditioning for the directions and for the hours.

Index	0-24 h forecast			24-48 h forecast		
	Raw	dr	drdH	Raw	dr	drdH
Bias	-2.945	-0.349	-0.241	-2.941	-0.329	-0.268
NBias	-0.566	-0.067	-0.046	-0.565	-0.063	-0.051
MAE	3.215	2.009	1.837	3.223	2.072	1.919
NMAE	0.618	0.386	0.353	0.619	0.398	0.369
RMSE	4.572	2.659	2.501	4.599	2.760	2.637
NRMSE	0.687	0.400	0.376	0.692	0.415	0.397
HH	1.182	0.451	0.416	1.196	0.469	0.442
SI	0.526	0.396	0.374	0.532	0.412	0.395
Pearson	0.632	0.773	0.800	0.615	0.750	0.774

Finally, figure 6.12 shows the difference between the forecast obtained by conditioning only on the directions (dr) with respect to that also conditioning on the hours (drh). At a glance the difference is not as evident as it is for the error indices. It is good to underline that even if this difference is not particularly appreciable on a visual level, since it is instead appreciable from the point of view of the indices, it will be useful when the wind speed is transformed into energy power since often this operation involves an amplification of the mistakes of forecasts.

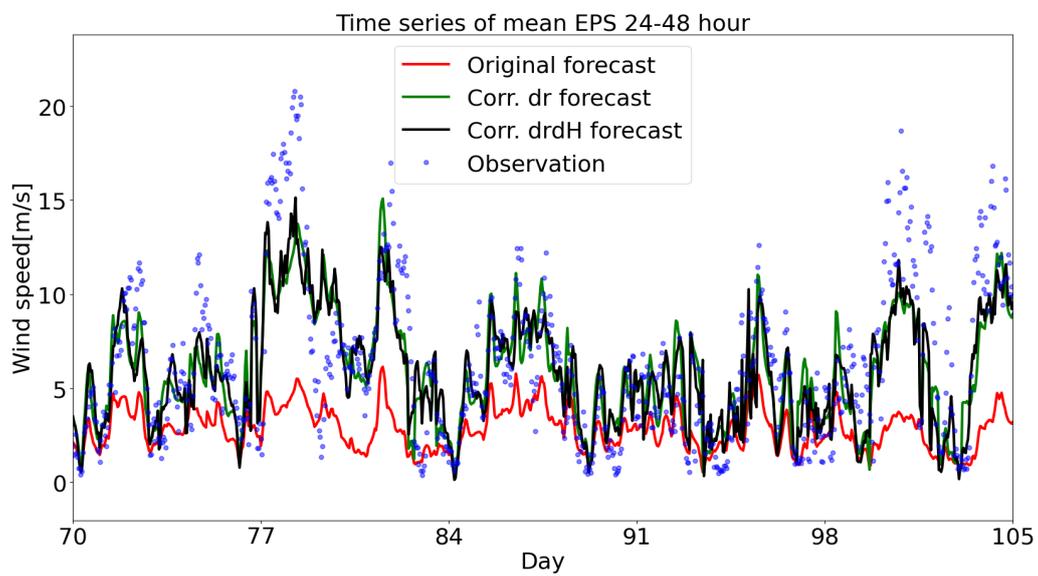


Figure 6.12: Time series of the forecasts before and after the correction from 12/3/2018 to 16/4/2018. The forecasts are the EPS mean with a forecast horizon of 24-48 hours. The correction used is the quantile regression with two techniques: conditioning on the directions(dr), conditioning on the directions and on hours(drdH).

6.3.4 Correction dividing in intervals of wind speed

The last conditioning that will be performed is the one on the speed of predictions. In fact it may be that by dividing the forecasts according to speed intervals, the fit lines have different slopes. To understand how many speed ranges to use and how effective they are two tests were done:

- Division into 3 intervals: $(0 - 3)$, $(3 - 6)$, $(6 - \infty)[m/s]$
- Division into 12 intervals: $(0 - 0.5)$, $(0.5 - 1)$, $(1 - 1.5)$, $(1.5 - 2)$, $(2 - 2.5)$, $(2.5 - 3)$, $(3 - 3.5)$, $(3.5 - 4)$, $(4 - 4.5)$, $(4.5 - 5)$, $(5 - 5.5)$, $(5.5 - \infty)[m/s]$

Figure 6.13 shows the scatter plots of the 24-48 hour forecasts. The left figure in red shows the broken line of best fit created by dividing the dataset into the 3 intervals, while on the right the best fit broken into the 12 intervals. Both graphs show that for speed values below 5/5.5 m/s the best fit tends to be substantially linear. The difference between lower and higher speeds of 5 m/s is more evident.

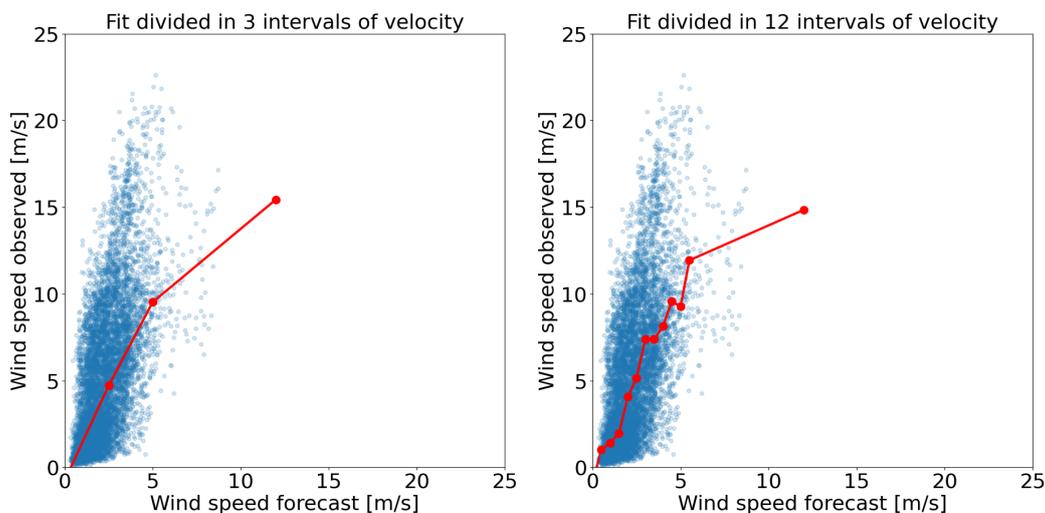


Figure 6.13: Scatter plot of forecasts (average EPS 24-48 hours) vs observations with 2 different best fits (in red). On the left a best fit dividing the dataset into 3 speed ranges while on the right dividing by 12 speed ranges.

Table 6.7 shows the index values of the raw predictions, corrected for the entire dataset, corrected by conditioning for 3 speed intervals and finally for 12 intervals. Comparing the tt correction with the dv3 correction shows that there is an improvement in the indices. In particular, the NBias improves

by about 2.5%, the NMAE by about 0.5% and the NRMSE by almost one percentage point. Confirming the overall improvement, HH improves by as much as 2%. Finally, it is emphasized that the Pearson also improves, attesting in fact that inserting a conditioning on the speed leads to better correlations.

A comparison is now made between which of the two strategies implemented is the best and therefore whether it is better to condition on 12 intervals or 3 intervals. From the table it appears that some of the indices would seem to be better if the forecasts are divided into 12 speed intervals while others in the case where it is divided by 3 intervals, paying particular attention to the HH index which shows a perfectly equal value. These differences are absolutely insignificant. This was expected given that, as mentioned before, in graph 6.13 a best fit line was shown which, apart from the two macrozones above and below 5 m/s, was particularly linear. Therefore, for the future this conditioning will be used by dividing only and exclusively into 2 intervals in order to detect the significant non-linearity shown now without however fragmenting the dataset too much, thus losing its robustness.

Table 6.7: Error indices values, starting from mean EPS 24-48 hour forecast, of the raw predictions, corrected for the entire dataset (tt), corrected by conditioning for 3 speed intervals (dv3) and finally for 12 intervals (dv12).

Index	24-48 h forecast			
	Raw	tt	dv3	dv12
Bias	-2.941	-0.896	-0.761	-0.746
NBias	-0.565	-0.172	-0.146	-0.143
MAE	3.223	2.477	2.458	2.455
NMAE	0.619	0.476	0.472	0.472
RMSE	4.599	3.383	3.332	3.340
NRMSE	0.692	0.509	0.501	0.502
HH	1.196	0.628	0.607	0.607
SI	0.532	0.491	0.488	0.490
Pearson	0.615	0.615	0.622	0.619

Now it is possible to make a correction by joining all the conditionings that have been presented so far. Therefore, the forecasts will be conditioned simultaneously on directions, time and speed intervals.

Table 6.8 shows how they improve by also inserting the speed condition compared to the best predictions obtained so far. It is clear how the improvement is clear for all the indices and how it is amplified compared to that seen in

the previous table. In particular, if highlighted an NBias that begins to be close to 0, settling at 3%, an NMAE improved by almost a percentage point and overall, by almost 30% compared to the raw forecast and finally a Pearson also improved by almost 1% compared to the correction drdH and 20% of the raw forecast.

Table 6.8: Comparison between the various indices of the error for the raw forecasts (mean EPS, 0-24 and 24-48 hours), corrected by conditioning for the directions and hours(drdH), and corrected by conditioning for the directions, hours and velocity (drdHdv).

Index	0-24 h forecast			24-48 h forecast		
	Raw	drdH	drdHdv	Raw	drdH	drdHdv
Bias	-2.945	-0.241	-0.158	-2.941	-0.268	-0.124
NBias	-0.566	-0.046	-0.030	-0.565	-0.051	-0.024
MAE	3.215	1.837	1.802	3.223	1.919	1.878
NMAE	0.618	0.353	0.346	0.619	0.369	0.361
RMSE	4.572	2.501	2.446	4.599	2.637	2.559
NRMSE	0.687	0.376	0.368	0.692	0.397	0.385
HH	1.182	0.416	0.400	1.196	0.442	0.419
SI	0.526	0.374	0.367	0.532	0.395	0.384
Pearson	0.632	0.800	0.808	0.615	0.774	0.787

Finally, figure 6.14 shows a comparison between the raw forecasts, those with the correction conditioning on direction and hours, the correction made now conditioning on direction, time and speed and the observations in a time series of 5 weeks. The two correct predictions are very similar since even the error indices have not diverged very much. However, it is very interesting to note what the effect of conditioning was on speeds lower and higher than 5 m/s. In fact, while the lower speeds seem to have had few variations, the higher speeds are those that seem to have benefited most from this conditioning. It can be said that this result was expected because, as seen from figure 5.26, most of the forecast speeds were below 5 m/s and that means that when the best fit was carried out on the whole dataset, these speeds presented a different trend from the others but were not very represented and therefore the best fit that was created was more similar to that for speeds below 5 m/s.

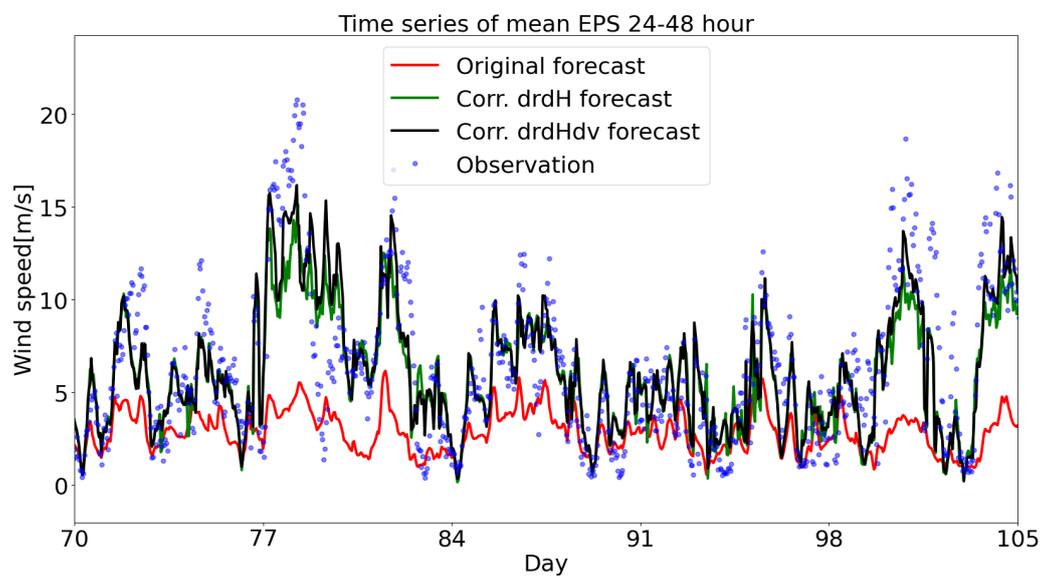


Figure 6.14: Time series of the forecasts before and after the correction from 12/3/2018 to 16/4/2018. The forecasts are the EPS mean with a forecast horizon of 24-48 hours. The correction used is the quantile regression with two techniques: conditioning on the directions and on hours(drh), conditioning on the directions, hours and velocity(drdHdv).

6.3.5 Correction using prediction error correlation

As previously mentioned, the wind forecast ranges from 0 to 48 hours. These forecasts begin to be calculated at midnight by the ECMWF. Logically, for all the forecasts to be calculated and distributed, several hours of calculation are required as the models are computationally very heavy. Thus, the first available hour of this forecast turns out to be after 7 in the morning. In addition to this fact is added that the electricity market closes at certain established times. For example, the one for the next day closes at 11.45 am. This means that there are forecasts available which refer to the past and are therefore apparently useless. From this point on, it is simulated to have to calculate the forecast at 7:30 in the morning. This means that for the 0-24 hour forecast only the 8-24 hour forecast will be considered. In doing so, however, you come into possession of a data that was not known before and so it can be known how and how much the forecasts from 00 to 07 hours were wrong. In this section it will analyse if it exists and how to exploit the correlation between the errors of the forecast at 7 am with that of the following hours.

First of all, it is necessary to take a look at how the 0-24 hour forecasts change by considering only those from 8 to 24. They are expected to be worse as the hours have been eliminated as the least forecast horizon and the hours that, as seen in the previous chapters, being mainly nocturnal they are easier to predict.

Table 6.9 shows this comparison. As expected, many of the indices have deteriorated but not all. However, the indices can be divided into two categories. In fact, going to see which are the improved indices, they are only those that have been normalized. In fact, since the diurnal hours are typically windier than the nocturnal ones, the moment in which it is normalized there is a higher denominator and therefore the index is lower. On the other hand, looking at the non-normalized indices as expected, they have worsened. In conclusion, the Pearson also gets worse.

At this point the first thing to do is to see if and how much correlation there is between the error at 7 and those at other hours. Figure 6.15 shows precisely how this correlation varies throughout the forecast period. Two correlations are shown. The first (in green) shows the correlation of the error (forecasts - observations) of the raw predictions without any calibration. The second in red shows the correlation of the error with the forecasts calibrated with the best technique currently available (drdHdv, conditioning on direction, time, and speed). The error correlation of the raw predictions is much greater than that of the corrected predictions. This is normal and is since the correct prediction, by virtue of the fact that conditioning has been made,

Table 6.9: Comparison between the forecasts from 0 to 24 with the forecasts only from 8 to 24.

Index	Raw		drdHdv	
	0-24 h	8-24 h	0-24 h	8-24 h
Bias	-2.945	-3.273	-0.158	-0.097
NBias	-0.566	-0.576	-0.030	-0.017
MAE	3.215	3.484	1.802	1.897
NMAE	0.618	0.613	0.346	0.334
RMSE	4.572	4.723	2.446	2.533
NRMSE	0.687	0.681	0.368	0.365
HH	1.182	1.162	0.400	0.394
SI	0.526	0.491	0.367	0.365
Pearson	0.632	0.584	0.808	0.771

has changed its raw nature. So, for example, when the correction is made at 7 it will have a direction “d7” and a speed “v7”, while on the same day the hour 10 will have its direction “d10” and its speed “v10”. Therefore, the two forecasts will be subjected to different corrections due to the fact that they have different hours, directions and speeds. Having two different corrections therefore changes the relationship between the two errors. For this reason, the raw predictions have a greater correlation of the error, because they have the same origin. Logically, since the predictions are correct, however derived from the raw ones, they maintain a certain correlation of the error. Analysing the graph further, it is noted first that at 7 o’clock there is a perfect correlation and how the more one moves away from 7 the more it decreases. There would seem to be a slowdown in the decay of the correlation in the hours around 7 of the following day, even in the correct case there is an increase in the correlation. This phenomenon could be traced back to the fact that repeating the same time of day there are more similar weather conditions. By virtue of such an evident loss of correlation, it is legitimate to ask whether it is not worthwhile to first make a correction using this correlation and then do that by viewing the various terms. It will later be shown that this assumption is incorrect.

At this point it was possible to make a correction by exploiting this correlation. In other words, taking the error of the forecast at hour 7, a quantile regression was made to calibrate it on the error of all the following hours one by one. Once the new calibrated error for each hour was obtained, it was simply added to the forecast thus obtaining the correct forecast. Logically,

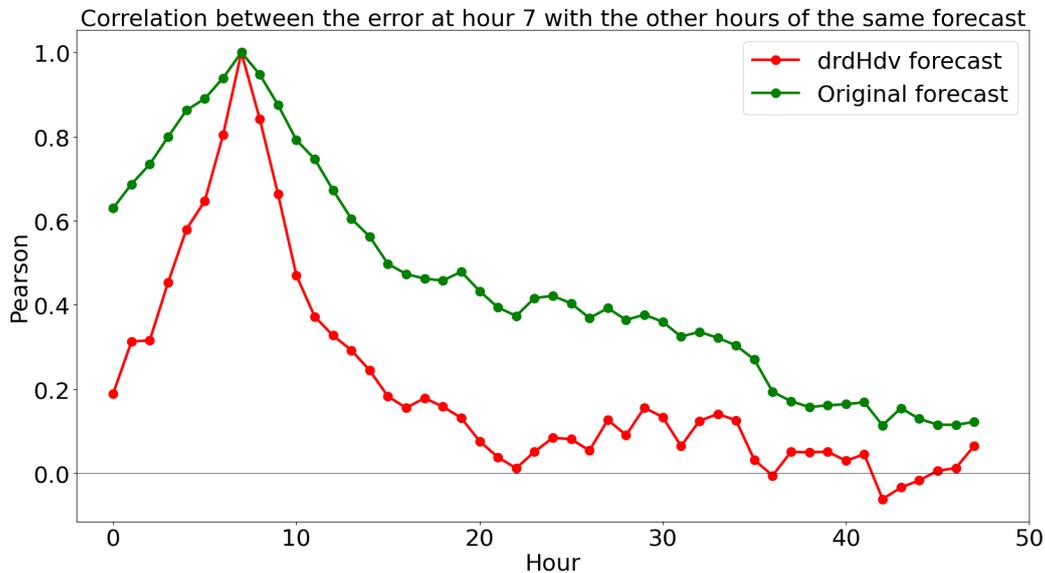


Figure 6.15: Trend of the correlation between the difference between forecasts and observations at 7 am with those in the 48 hours of forecasting. This with the following predictions: EPS mean 0-48 h raw and EPS mean 0-48 h corrected by conditioning on direction, time and speed (drdHdv).

given the poor correlation that exists the next day (forecasts 24-48 hours) a great improvement is not expected, while for the hours closest to 7 (8-24 hours) yes.

Table 6.10 shows the results of this correction after the correction already made previously. As expected, the hours 8-24 have undergone a marked improvement. The NBias is practically 0, the NMAE is improved by more than 2% as well as the NRMSE. HH also improved by nearly 3 percentage points, which confirms that forecasts have improved overall. Finally, the Pearson also made a jump of more than 3 percentage points. As regards the forecasts at 24-48 hours, there was no benefit from this correction. In fact, the indices have remained almost identical.

Table 6.11 answers the question posed previously, that is, whether it was more convenient to put this last correction downstream since it showed high correlations. Well, the indices show indisputably how best the previously used approach is. In fact, all the indices have a marked deterioration.

Finally, in figure 6.16, the time series of the raw predictions is shown, corrected with the conditionings on direction, time and speed (drdHdv) and further corrected with the correlation of the error (drdHdv-df). This forecast is that of 0-24 hours as regards the drdHdv while for the drdHdv it goes from

Table 6.10: Comparison between the various indices of the error for the raw forecasts (mean EPS, 0-24 and 24-48 hours), corrected by conditioning for the directions, hours and velocity (drdHdv), and corrected by conditioning for the directions, hours, velocity and using the correlation of the error(drdHdv-df).

Index	8-24 h forecast			24-48 h forecast		
	Raw	drdHdv	drdHdv-df	Raw	drdHdv	drdHdv-df
Bias	-3.273	-0.097	-0.041	-2.941	-0.124	-0.123
NBias	-0.576	-0.017	-0.007	-0.565	-0.024	-0.024
MAE	3.484	1.897	1.768	3.223	1.878	1.878
NMAE	0.613	0.334	0.311	0.619	0.361	0.361
RMSE	4.723	2.533	2.382	4.599	2.559	2.561
NRMSE	0.681	0.365	0.344	0.692	0.385	0.385
HH	1.162	0.394	0.365	1.196	0.419	0.419
SI	0.491	0.365	0.344	0.532	0.384	0.385
Pearson	0.584	0.771	0.802	0.615	0.787	0.787

Table 6.11: Comparison to see if it is better to do the correction of the error correlation before (df-drdHdv) or after the correction with the conditioning on direction, time and speed (drdHdv-df). All this starting from the forecast of the average EPS for 8-24 and 24-48 hours.

Index	8-24 h forecast			24-48 h forecast		
	Raw	drdHdv-df	df-drdHdv	Raw	drdHdv-df	df-drdHdv
Bias	-3.273	-0.041	-0.090	-2.941	-0.123	-0.241
NBias	-0.576	-0.007	-0.016	-0.565	-0.024	-0.046
MAE	3.484	1.768	1.874	3.223	1.878	2.131
NMAE	0.613	0.311	0.330	0.619	0.361	0.409
RMSE	4.723	2.382	2.546	4.599	2.561	2.961
NRMSE	0.681	0.344	0.367	0.692	0.385	0.445
HH	1.162	0.365	0.395	1.196	0.419	0.501
SI	0.491	0.344	0.367	0.532	0.385	0.444
Pearson	0.584	0.802	0.769	0.615	0.787	0.703

8 to 24, in fact there are holes. Well, it appears that the correlation of the error has led to an improvement in the data, especially as regards the hours closest to 7. Such improvements are seen particularly where there are peaks or particularly low values.

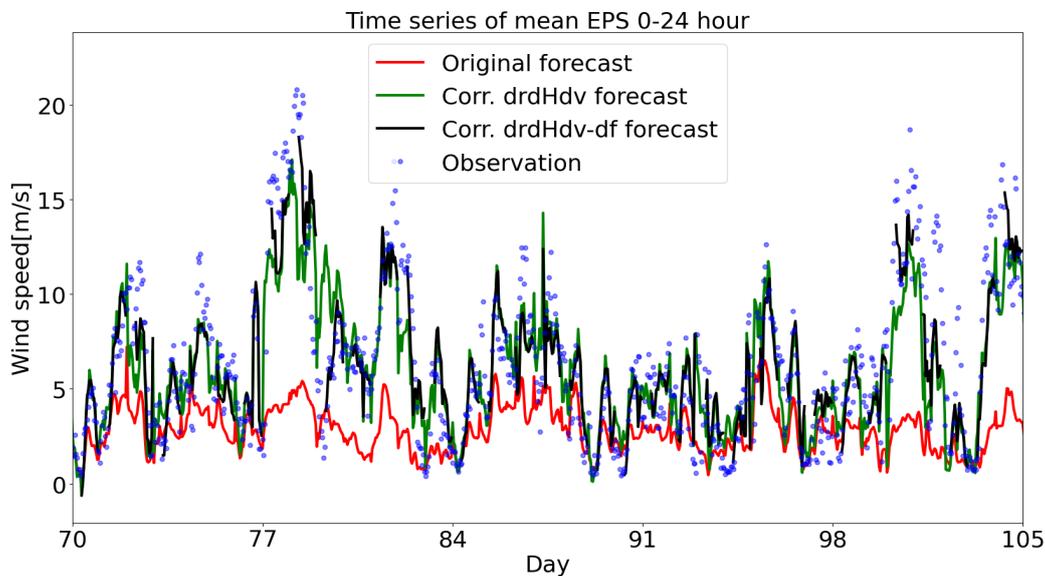


Figure 6.16: Time series of the forecasts before and after the correction from 12/3/2018 to 16/4/2018. The forecasts are the EPS mean with a forecast horizon of 8-24 hours. The correction used is the quantile regression with two techniques: conditioning on the directions, hours and velocity(drdHdv) and using the error correlation after conditioning on the directions, hours and velocity(drdHdv-df). The last forecast is only for 8 to 24 hour.

6.3.6 The moving average

As seen by the graphs depicting the time series of the forecasts, they have a rather fluctuating aspect. This is certainly due to the nature of the wind which has large variations even in small periods of time. Furthermore, the fact of having conditioned and calibrated the forecasts could have led to an amplification of the noise now exposed. It may therefore make sense to try to make moving averages to see if the forecast can gain an advantage.

It is therefore necessary to first understand how many hours it is convenient to use for the moving average. Several cases have been made with the following number of hours:

- 3 hour
- 5 hour
- 7 hour
- 9 hour

- 11 hour
- 13 hour

The number of hours is always odd because excluding the current time you want to have the same distance back and forth in time. Logically close to the forecast limits (when approaching the 48 hour forecast) the moving average must necessarily consist of fewer hours because for example for the hour 47 with a moving average composed of 7 hours it would also be necessary to know the hours 49 and 50 which, however, do not know each other. So, for 47 hours the best can be done is use a 3 hour moving average.

The formula for calculating the moving average($\bar{V}(t)$) is as follows:

$$\bar{V}(t) = \frac{1}{k} \sum_{i=m_1}^{m_2} V(t+i) \quad (6.7)$$

With:

- m_1 the number of periods before t .
- m_2 the number of periods following t .
- $k = m_1 + m_2 + 1$ is the period or order of the moving average, and is equal to the number of addends.
- V is the forecast of the wind speed.

Figure 6.17 shows how the NMAE, HH and Pearson indices change by the number of hours used to make the moving average. First of all, the graph shows that even if only using 3 hours of moving average there is a considerable improvement of all indices. It also appears that the optimal number of hours to run the moving average is 7. So 3 hours forward and 3 hours back.

At this point it is possible to make the moving average for all forecasts. Table 6.12 shows the error indices for the raw forecast, the best forecast so far and the same after making a 7-hour moving average. Well all the indices have improved apart from the NBias which has a very slight deterioration. The NMAE instead improves by more than 1% as well as NRMSE and HH. The Pearson also shows a good improvement of up to 2 percentage points.

Finally, the usual time series is shown in figure 6.18. It is quite clear what the effect of the moving average is. In fact, the trend of the forecast is much sweeter than before and all the noise that previously existed has been eliminated. Of course, in doing so, some small speed peaks are lost that were previously taken better but overall, as confirmed by the error indices, the result is better.

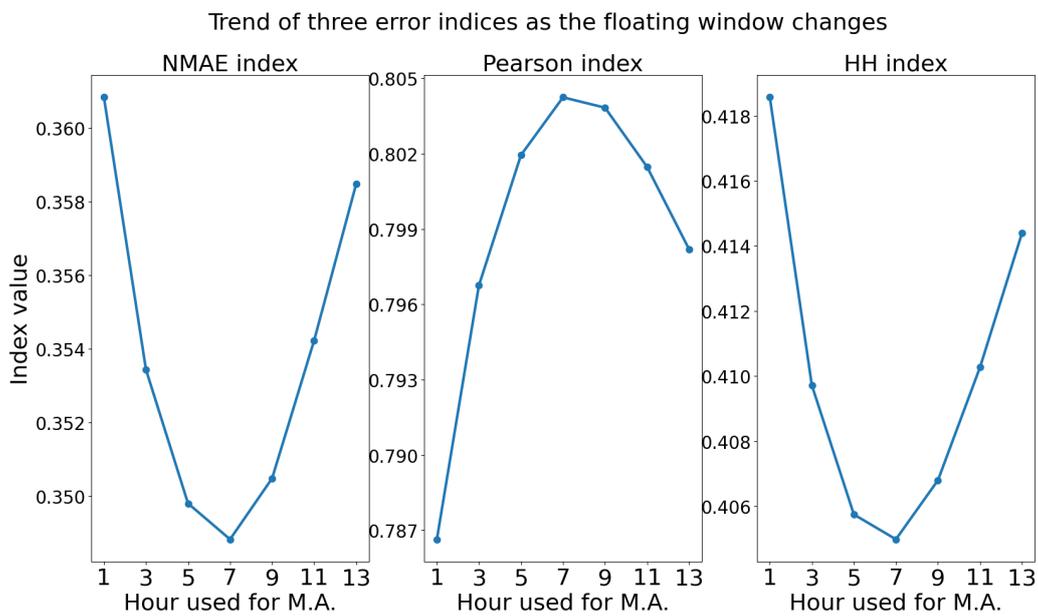


Figure 6.17: How the NMAE, Pearson and HH indices vary as the floating window used varies. On the ordinate there is the value of the indices while on the abscissa there is the heat of the number of hours considered to make the moving average. The moving average was made on the forecast corrected with the best available correction (drdHdv-df) on the EPS mean 24-48 h forecast.

Table 6.12: Comparison between the various indices of the error for the raw forecasts (mean EPS, 8-24 and 24-48 hours), corrected by conditioning for the directions, hours and velocity and using the correlation of the error(drdHdv-df), and corrected by conditioning for the directions, hours, velocity, using the correlation of the error and making a moving average(drdHdv-df-mv).

Index	8-24 h forecast			24-48 h forecast		
	Raw	drdHdv- -df	drdHdv- -df-mv	Raw	drdHdv- -df	drdHdv- -df-mv
Bias	-3.273	-0.041	-0.107	-2.941	-0.123	-0.142
NBias	-0.576	-0.007	-0.019	-0.565	-0.024	-0.027
MAE	3.484	1.768	1.686	3.223	1.878	1.816
NMAE	0.613	0.311	0.297	0.619	0.361	0.349
RMSE	4.723	2.382	2.262	4.599	2.561	2.466
NRMSE	0.681	0.344	0.326	0.692	0.385	0.371
HH	1.162	0.365	0.349	1.196	0.419	0.405
SI	0.491	0.344	0.326	0.532	0.385	0.370
Pearson	0.584	0.802	0.823	0.615	0.787	0.804

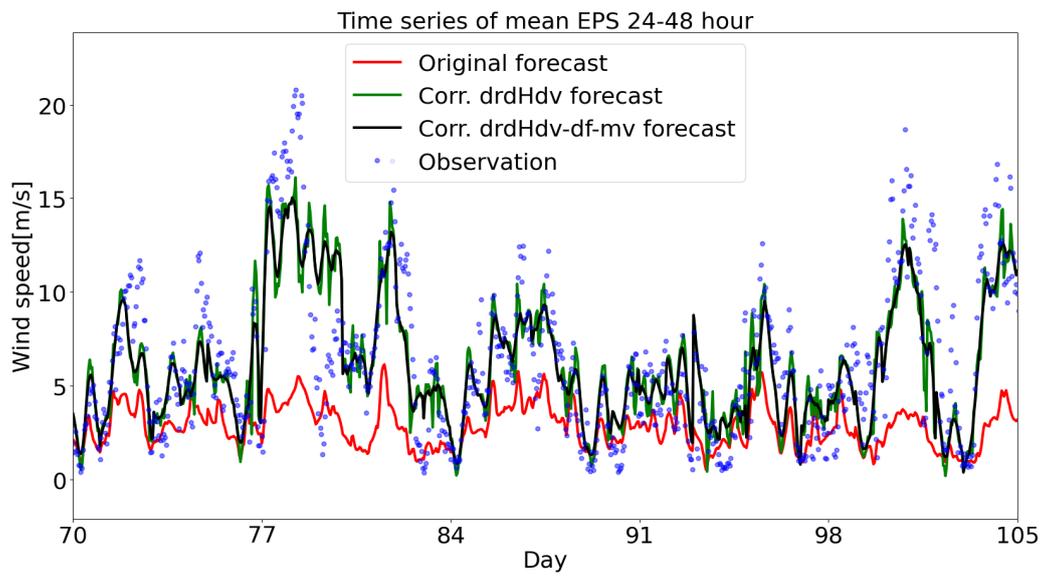


Figure 6.18: Time series of the forecasts before and after the correction from 12/3/2018 to 16/4/2018. The forecasts are the EPS mean with a forecast horizon of 24-48 hours. The correction used is the quantile regression with two techniques: conditioning for the directions, hours and velocity and using the correlation of the error(drdHdv-df) and conditioning for the directions, hours, velocity, using the correlation of the error and making a moving average(drdHdv-df-mv)

6.3.7 Correction of negative values

Due to the corrections previously made it may have happened that some wind speed predictions have assumed a negative value. For example, when correlating the error, it can happen that at 7 o'clock there is an error opposite to the ratio it usually has with that of any hour later and therefore if a forecast is close to 0 it can easily become negative. This perhaps does not happen in the time series shown so far but it is something that for example for the 24-48 hour forecast has occurred 29 times. Since these predictions are mathematically wrong, it makes sense to change them manually. From the literature there are 2 different methods, one that replaces the negative values with a zero and the other that replaces them with opposite values. In this case, being so few points, one method or the other is absolutely indifferent. From the table 6.13 there is practically no type of improvement in general and that the difference between replacing with 0 and the opposite value does not change anything. The technique of replacing with 0, which seems to be slightly better, will therefore be adopted.

Table 6.13: Comparison between the various indices of the error for the raw forecasts (mean EPS, 24-48 hours), corrected with the best correction(drdHdv-df-mv) and then later replacing negative values with their opposite (sub. opp.) or with a 0(sub. 0).

Index	24-48 h forecast			
	Raw	drdHdv-df-mv	sub. opp.	sub. 0
Bias	-2.941	-0.142	-0.140	-0.139
NBias	-0.565	-0.027	-0.027	-0.027
MAE	3.223	1.816	1.814	1.814
NMAE	0.619	0.349	0.349	0.348
RMSE	4.599	2.466	2.466	2.465
NRMSE	0.692	0.371	0.371	0.371
HH	1.196	0.405	0.405	0.405
SI	0.532	0.370	0.370	0.370
Pearson	0.615	0.804	0.804	0.804

6.3.8 The use of the 50 members of the EPS

EPS are 50 different wind speed predictions. As previously explained, they are different because they start from 50 different initial conditions generated going to perturb the initial condition considered to be the best. The average

of these 50 forecasts has been used so far. The goal of this sub-chapter is to try to make the most of all 50 members.

As seen in the sub-chapter where forecast was corrected using error correlation, it was exploited the fact that forecasts, when made available, have hours of forecasts that are now in the past. The idea is therefore to identify one or more members who in the first 7 hours seem to have predicted the wind better. Once identified, these n members will be used to predict the next 40 hours.

First of all, it is therefore necessary to understand how many members should be taken. In fact, since the members are very variable (much more than their average), taking too few would risk having a forecast with too many fluctuations. On the contrary, by taking too few of them it risks inserting too many who have a too wrong forecast. The members will be chosen according to the NRMSE index. In fact, those who will have the best NRMSE in the first 7 hours of forecasting will be the members who will be used. To put them together, will not be use a simple average but a weighted average ($\langle V \rangle (t)$) according to the NRMSE:

$$\langle V \rangle (t) = \sum_{i=1}^N NRMSE_i(t) \sum_{i=1}^N V_i(t) \frac{1}{NRMSE_i(t)} \quad (6.8)$$

- t is the instant considered
- $\langle V \rangle (t)$ is the result of the prediction at time t
- N is the number of members used
- $NRMSE_i(t)$ is the NRMSE at time t of member i

Logically, to give more weight to the best members it is necessary to point them a lot by the inverse of the NRMSE since the more NRMSE tends to zero the better.

To understand how many members, it is best to consider starting from the correct predictions with the drdHdv (conditioning on direction, time and speed).

Figure 6.19 shows how the NMAE, HH and Pearson indices vary as a function of how many members of the EPS are considered. It is immediately evident how there is a huge difference between taking only one member compared to taking even only 2. In fact, as said before, taking too few members leads to a too disturbed and “turbulent” forecast, while increasing the number the forecast becomes more “stable”. The best condition seems to be taking a number of members between 30 and 40. In fact, after 40 there appears to be

a counter-trend in the indices. It is quite indifferent which number to take between 30 and 40 as the differences are very tiny. From now on, therefore, the best 35 members will be considered.

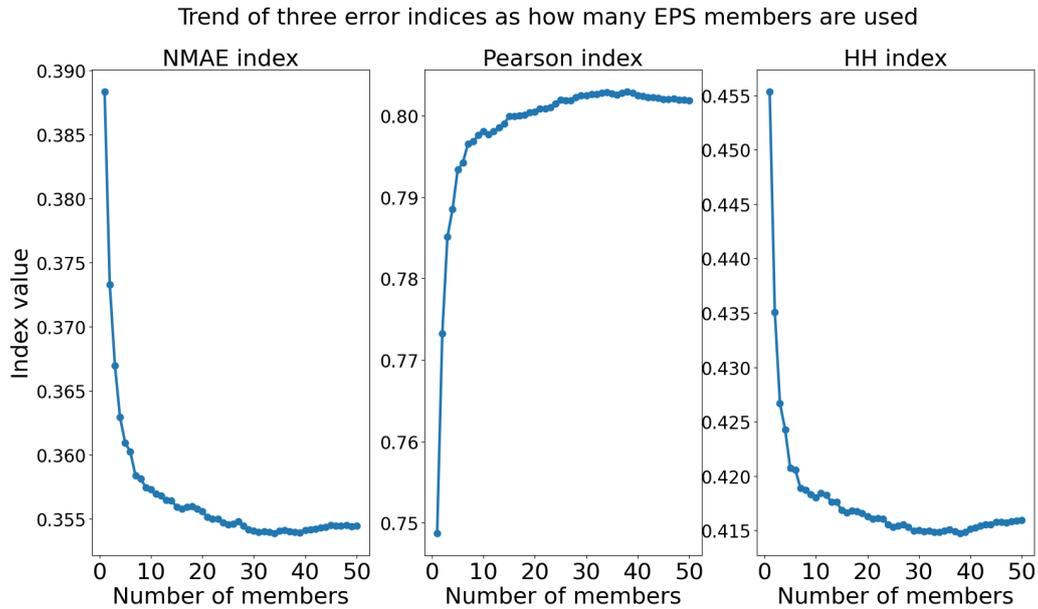


Figure 6.19: How the NMAE, Pearson and HH indices vary as the number of members used varies. On the ordinate there is the value of the indices while on the abscissa there is the number of members used. The moving average was made on the forecast corrected with the correction conditioning on direction, time and speed(drdHdv) on the EPS mean 24-48 h forecast.

Table 6.14 shows a comparison between the forecasts starting from the best static forecast (drdHdv) to which one exploits the correlation of the error while the second the choice of the best members. The result is quite interesting. In fact, for the 8-24 hours case the best correction is the one that starts from the average of the EPS and uses the error correlation while for the 24-48 hours the best is the one that chooses the best 35 members and puts them together. This is probably since for the hours 8-24 the correlation of the error, as seen previously, has a big effect that cannot be matched by the choice of the best member. For the 24-48 hour case, on the other hand, the correlation of the error had not led to any benefit. On the contrary, the choice of the best members seems to maintain a good solidity even 24 hours later. In fact, all the indices in this case improve between 0.5 and 1%.

Table 6.15 shows how the indices vary using the techniques seen in the previous sections on the forecast obtained by choosing the best members

Table 6.14: Comparison between the error indices calculated for the following cases: raw case with the average of EPS (8-24 and 24-48 hours) and the other two cases which both start from the best “static” correction (drdHdv, conditioning on directions, time and speed) in which a prediction is obtained using the correlation of the error (df) while the second one choosing the best 35 members and making the weighted average (mx).

Index	8-24 h forecast			24-48 h forecast		
	Raw	drdHdv- -df	drdHdv- -mx	Raw	drdHdv- -df	drdHdv- -mx
Bias	-3.273	-0.041	-0.073	-2.941	-0.123	-0.167
NBias	-0.576	-0.007	-0.013	-0.565	-0.024	-0.032
MAE	3.484	1.768	1.838	3.223	1.878	1.842
NMAE	0.613	0.311	0.324	0.619	0.361	0.354
RMSE	4.723	2.382	2.446	4.599	2.561	2.495
NRMSE	0.681	0.344	0.353	0.692	0.385	0.375
HH	1.162	0.365	0.382	1.196	0.419	0.415
SI	0.491	0.344	0.353	0.532	0.385	0.374
Pearson	0.584	0.802	0.791	0.615	0.787	0.803

after correcting them. These forecasts are compared with the best dynamic correction that was obtained starting from the average of the EPS. Using only the correlation of the error, the forecasts are worse under all indices (drdHdv-mx-df) while when the moving average is also used (drdHdv-mx-df-mv) the situation changes. In fact, some indices are better while others are not. In particular, the NMAE fails to fall below the value found previously while the NRMSE does. HH is best for 8-24 hours but worse for 24-48. However, all these variations remain in the order of one thousandth. The index that undergoes a greater variation is the Pearson. This bodes well in the fact that since the correlation increases, doing a simple quantile regression on the whole forecast without any conditions can be an overall improvement of the indices.

Table 6.16 shows the error indices of the best forecast starting from the average of the EPS, of the best correction using the different members that has just been described and finally with the forecast obtained by making a quantile regression on the whole forecast just mentioned. This was done considering that a good increase in Pearson was noted, but not followed by an equally increase in the other indices. In doing so it was possible to make the most of the correlation that had been added. In fact, all the indices (apart from Pearson of course) are improving by a percentage ranging from 0.5 to

Table 6.15: Comparison between the best forecasts corrected starting from the average of the EPS (drdHdv-df-mv) with those correcting the single members and then after choosing the best ones by correcting them further thanks to the correlation of the error(drdHdv-mx-df) and then with the moving average(drdHdv-mx-df-mv).

Index	8-24 h forecast			24-48 h forecast		
	drdHdv-df-mb	drdHdv-mx-df	drdHdv-mx-df-mv	drdHdv-df-mb	drdHdv-mx-df	drdHdv-mx-df-mv
Bias	-0.107	-0.004	-0.076	-0.142	-0.175	-0.194
NBias	-0.019	-0.001	-0.013	-0.027	-0.034	-0.037
MAE	1.686	1.718	1.690	1.816	1.842	1.827
NMAE	0.297	0.303	0.298	0.349	0.354	0.351
RMSE	2.262	2.297	2.240	2.466	2.494	2.458
NRMSE	0.326	0.331	0.323	0.371	0.375	0.370
HH	0.349	0.353	0.347	0.405	0.414	0.411
SI	0.326	0.331	0.323	0.370	0.374	0.369
Pearson	0.823	0.816	0.830	0.804	0.803	0.814

1. This is interesting because in this way it was possible to fall, especially as regards the NMAE, under the best forecast that had been obtained so far.

Figure 6.20 shows the time series of the best forecast obtained previously with the one obtained now. The differences are very low as they were in the indices but there is still some difference. The three major peaks present seem to have been better predicted with this latter prediction, albeit slightly.

To conclude the search for the best spaghetti it can be said that it has led to an improvement. However, this improvement proved to be quite limited and in the face of the much greater computational effort it requires compared to the correction starting from the average of the EPS (about 50 times more) it is not necessarily worth it.

Table 6.16: Comparison between the best forecasts corrected starting from the average of the EPS (drdHdv-df-mv) with those correcting the single members and then after choosing the best ones by correcting them further thanks to the correlation of the error and with the moving average(drdHdv-mx-df-mv) and then making a quantile regression on all the forecasts(drdHdv-mx-df-mv-tt).

Index	8-24 h forecast			24-48 h forecast		
	drdHdv-df-mb	drdHdv-mx-df-mv	drdHdv-mx-df-mb-tt	drdHdv-df-mb	drdHdv-mx-df-mv	drdHdv-mx-df-mb-tt
Bias	-0.107	-0.076	0.033	-0.142	-0.194	-0.037
NBias	-0.019	-0.013	0.006	-0.027	-0.037	-0.007
MAE	1.686	1.690	1.662	1.816	1.827	1.784
NMAE	0.297	0.298	0.293	0.349	0.351	0.343
RMSE	2.262	2.240	2.221	2.466	2.458	2.407
NRMSE	0.326	0.323	0.320	0.371	0.370	0.362
HH	0.349	0.347	0.338	0.405	0.411	0.390
SI	0.326	0.323	0.320	0.370	0.369	0.362
Pearson	0.823	0.830	0.830	0.804	0.814	0.814

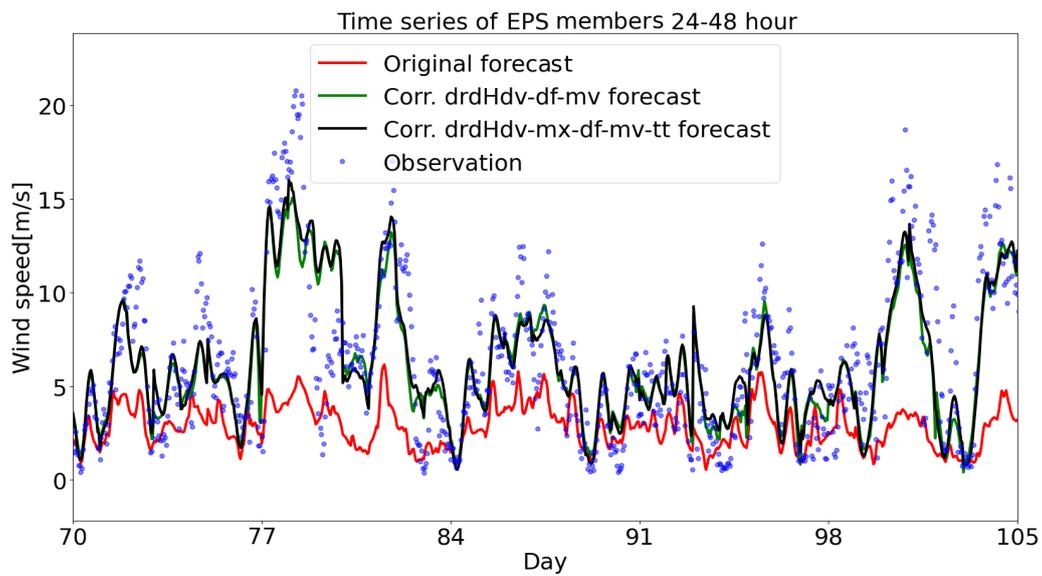


Figure 6.20: Time series of the forecasts before and after the correction from 12/3/2018 to 16/4/2018. The green and red forecasts are the EPS mean with a forecast horizon of 24-48 hours. The correction used is the quantile regression with two techniques: conditioning for the directions, hours, velocity, using the correlation of the error and making a moving average(drdHdv-df-mv) and using the different members of EPS conditioning for the directions, hours, velocity, choosing the best 35 member of EPS and put them together, using the correlation of the error, making a moving average and finally making a quantile regression on all the forecasts(drdHdv-mx-df-mv-tt)

6.4 Analysis of the best corrections

In this section the corrected forecasts will be seen starting from the two downloaded ECMWF forecasts, the EPS ones (already shown in the previous ones) and the HRES ones, in order to make a comparison between them. It will then be shown how the predictions change depending on which methodology is used, that is, if the quantile regression or the CRPS of the gamma distribution is used. Finally, it will be shown how the confidence bands appear to understand how much a prediction can be variable and reliable.

6.4.1 Comparison of the corrected forecasts starting from EPS or HRES

First, a comparison will be shown between the results of the forecasts obtained starting in one case from the EPS forecast and in the other from the HRES forecast. As for the EPS, the last described technique will be used as the best technique which exploits the choice of the best members downstream of the correction that conditions on direction, hours, speed, uses the correlation of the error and makes the moving average, and then downstream of all this it makes a further quantile regression. As for the HRES forecast, obviously it cannot be made a choice of the best members since it is a single forecast, and therefore the same technique will be used, stopping before the choice of the best members (therefore conditioning on direction, hours, speed, use error correlation, do the moving average). Quantile regression will be used to make this comparison.

Table 6.17 shows the comparison between the best forecasts obtained starting from the average of the EPS and the individual members of the EPS, seen in the previous chapters, with the best forecast starting from the HRES forecast. In the previous chapter it was seen how the correlation of the mean of the EPS was better than the correlation of the forecast of the HRES. Given the difference in correlation, it was assumed that following the correction strategies, the forecast that would have turned out better would have been the one starting from the average of the EPS. From the table this hypothesis seems to be partially confirmed. In fact, for the forecast from 0 to 48 hours, the best forecast seems to be the one deriving from HRES, even if the differences are minimal. Instead, this theory is confirmed for the forecasts from 8 to 24. This is probably since when conditioning in HRES, hidden correlations were found that were not highlighted by looking at the correlation in full.

As for the comparison of the forecasts obtained starting from the individual members of the EPS with those obtained starting from the HRES, it is quite clear that the forecast deriving from the EPS is the best. In fact, all

the indices assume a higher value. NMAE is between 0.5 and 1% better as well as HH between 0.5 and 1.5%. Furthermore, Pearson also shows a not insignificant difference.

Table 6.17: Comparison of the error indices between the best forecast starting from the average of EPS, the best forecast starting from the EPS and the best forecast starting from the HRES. All this done with quantile regressions.

Index	8-24 h forecast			24-48 h forecast		
	EPS		HRES	EPS		HRES
	drdHdv- -df-mv	drdHdv- -mx-df-mv-tt	drdHdv- -df-mv	drdHdv- -df-mv	drdHdv- -mx-df-mv-tt	drdHdv- -df-mv
Bias	-0.107	0.033	0.037	-0.142	-0.037	-0.080
NBias	-0.019	0.006	0.007	-0.027	-0.007	-0.015
MAE	1.686	1.662	1.715	1.816	1.784	1.805
NMAE	0.297	0.293	0.302	0.349	0.343	0.347
RMSE	2.262	2.221	2.308	2.466	2.407	2.429
NRMSE	0.326	0.320	0.333	0.371	0.362	0.365
HH	0.349	0.338	0.354	0.405	0.390	0.395
SI	0.326	0.320	0.333	0.370	0.362	0.365
Pearson	0.823	0.830	0.815	0.804	0.814	0.810

6.4.2 Correction using the gamma distribution

The correction strategy exploiting the gamma distribution can be used with all the forecasts available. In fact, it is not necessary to have the variance of the forecast to use it. This is because when the CRPS is minimized, if the variance is not available, it is sufficient to determine one less parameter corresponding to the parameter that should be very sensitive precisely for the missing variance. In doing so, however, a variance is obtained that depends only on the conditioning carried out and on the expected wind speed. It therefore turns out to be a static variance. By doing so, however, a confidence band is obtained for the forecast also for the HRES forecast. If the EPS forecast is used, the variance is available and therefore by using the CRPS of the gamma distribution a new variance is obtained and therefore a confidence band of the so-called dynamic forecast. Which means that it is a function of how large the forecast members are.

Table 6.18 shows the value of the indices obtained with the two best strategies obtained so far using the CRPS for the gamma distribution instead of

quantile regression. The table shows how, although having more correlation, using single spaghetti leads to less precise forecasts. This is probably due to the fact that the gamma distribution was selected because it was considered that it best represented the distribution of observations as a function of the velocity and variance of the predictions obtained from the mean of the EPS. It is therefore probable that this distribution is not the most optimal for all 50 members of the EPS and therefore the error indices are affected, albeit to a limited extent.

Therefore, looking at the forecast column only, starting from the average of the EPS (drdHdv-df-mv) it can be seen how the Bias and the NBias strongly try to 0. This leads us to think once again that the choice of the gamma distribution has been correct. At this point it is interesting to compare this prediction with those obtained in table 6.17 (using quantile regression). Well, looking at the drdHdv-df-mv columns, corresponding to the best forecast obtained from the average of the EPS, is clear how the NMAE values are very close. Once again this confirms how well the gamma distribution describes the data held. As previously said using quantile regression it is logical to arrive at the absolute lowest value of NMAE by definition. It is therefore correct that the forecast obtained using quantile regression has better NMAEs. Looking at the other indices, however, it is noted that some are better for the prediction obtained using the gamma function. Considering the best prediction obtained starting from the individual members using quantile regression and the one starting from the average of the EPS using the gamma distribution (two of the best 2 predictions for the two techniques used), the precision of the forecast is a little more in favour of prediction obtained with quantile regression. However, these are non-abysmal differences, indeed they can be considered negligible. Finally, it can be said that the prediction obtained from the gamma distribution is a prediction that can very well be substituted for that obtained from quantile regression by virtue of the fact, although less precise, it possesses the accuracy of the dynamic rather than static prediction.

Finally, figure 6.21 and figure 6.22 show the time series of the forecast obtained with the gamma distribution with the relative confidence bands of 30, 60, 90%. As can also be seen from the graphical point of view, the wind observations are almost always within the confidence band.

Table 6.18: Comparison of the error indices for the corrected predictions with the two best strategies (one starting from the average of the EPS and the other considering each single member) using the CRPS method for the gamma distribution.

Index	8-24 h forecast			24-48 h forecast		
	Raw	drdHdv- -df-mv	drdHdv- -mx-df-mv-tt	Raw	drdHdv- -df-mv	drdHdv- -mx-df-mv-tt
Bias	-3.273	0.001	0.458	-2.941	-0.016	0.370
NBias	-0.576	0.000	0.081	-0.565	-0.003	0.071
MAE	3.484	1.692	1.806	3.223	1.848	1.901
NMAE	0.613	0.298	0.318	0.619	0.355	0.365
RMSE	4.723	2.248	2.317	4.599	2.473	2.474
NRMSE	0.681	0.324	0.334	0.692	0.372	0.372
HH	1.162	0.345	0.348	1.196	0.407	0.397
SI	0.491	0.324	0.328	0.532	0.372	0.368
Pearson	0.584	0.826	0.828	0.615	0.807	0.814

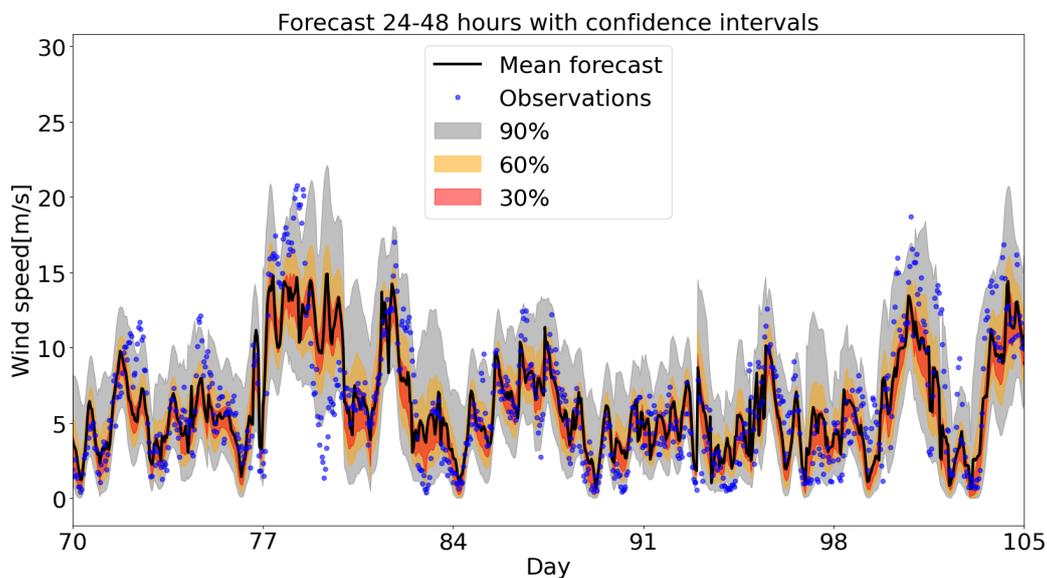


Figure 6.21: Time series of the forecasts before and after the correction from 12/3/2018 to 16/4/2018. The forecast generated starting from the average of the EPS for the 24-48 hour forecast, correcting by conditioning on direction, time and speed, exploiting the correlation of the error and making the moving average. This correction was done using the CRPS of the gamma distribution. The confidence bands were created considering a probability from the low of 5, 20 and 35% while from the top of 65, 80, 95%.

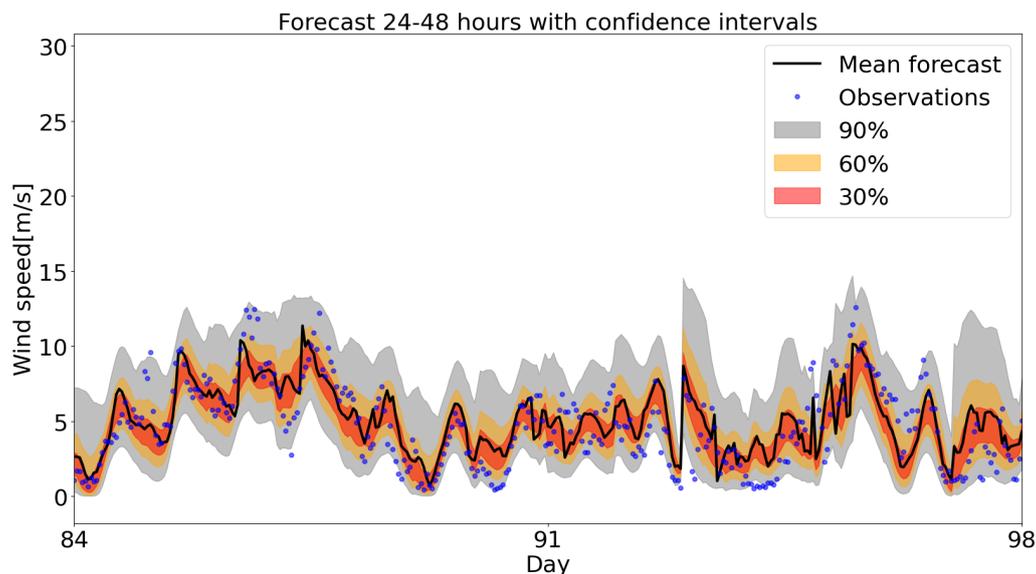


Figure 6.22: Zoom of figure 6.21 from 26/3/2018 to 9/4/2018.

6.4.3 The variation of principal index as function of the forecast horizon before and after the best correction

In the previous chapter it was seen how throughout the day the error indices did not have constant values but variable values. The aim of this section is to see if and how following the corrections the indices have changed their trend according to the forecast horizon.

Figure 6.23 shows the trend of the two main indices, the EAW and the RMSE, normalized and not. The errors were calculated using the best correction starting from the EPS members. The forecast horizons considered are from 8 to 48, this because it is hypothesized to be positioned at 7 in the morning.

First, it is noted that, although with different values, both indices show similar trends. With the raw forecast it is noted that for the non-normalized case during the day there was a worsening of both indices. Following the corrections, however, the absolute error of the indices does not seem to show any correlation with the diurnal cycle. This had already been noted previously when, influencing the time, it was seen that the daytime hours were those that had improved the most since there was a greater systematic error (Bias) which was therefore more easily correctable. It is also noted that in the first hours of the forecast there is a very rapid increase in the value of the indices which then tends to decrease. This is due to the fact that, as seen before, the first hours after 7 were those that had a high correlation with the error at 7 o'clock. This correlation was seen to be lost very quickly within a few hours and therefore it could not benefit from it for too long.

Finally, it can be seen how apart from the first few hours the trend of the two indices tends to increase their value but in a very slow way. This agrees with the trend of the raw forecasts. In fact, as seen, the 24-48 hour forecasts showed a very slight decline in the precision of the forecast. This is since the Central European model is a very stable and robust model over time. Moving on to the two normalized indices, there is an average trend very similar to that seen for non-normalized ones. In fact, it can be seen how the indices tend to worsen very slowly as the time horizon advances. Even here in the hours closest to 7 there is a very low value of the indices. Unlike the non-normalized indices, however, this time a daily trend is shown. This daily trend is the opposite of that seen for the MAE and NRMSE indices of the raw forecasts. This is simply since while the correct MAE and NRMSE are constant the intensity of the wind is not. Therefore, following the normalization, the hours in which the wind blows the most will be the hours that will

have a greater denominator and therefore in which the normalized indices will have their lowest value.

MAE, NMAE, RMSE, NRMSE in function of forecast horizon before and after correction

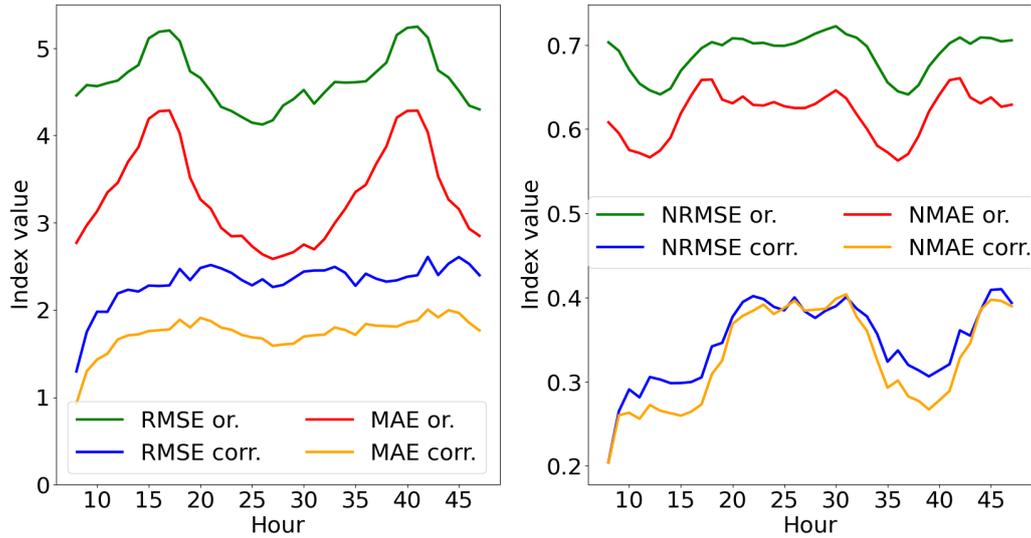


Figure 6.23: Trend of the MAE, NMAE, RMSE and NRMSE indices as a function of the forecast horizon before and after the best correction.

Figure 6.24 shows the Pearson trend as a function of the forecast horizon. The trend of the raw forecast was highly correlated with the diurnal cycle. Well, following the correction it would seem that this oscillation has flattened out considerably. Except for the first few hours in which correlation is at its maximum for the reasons set out above, also in this case the decline of Pearson is very slow as the forecast horizon advances.

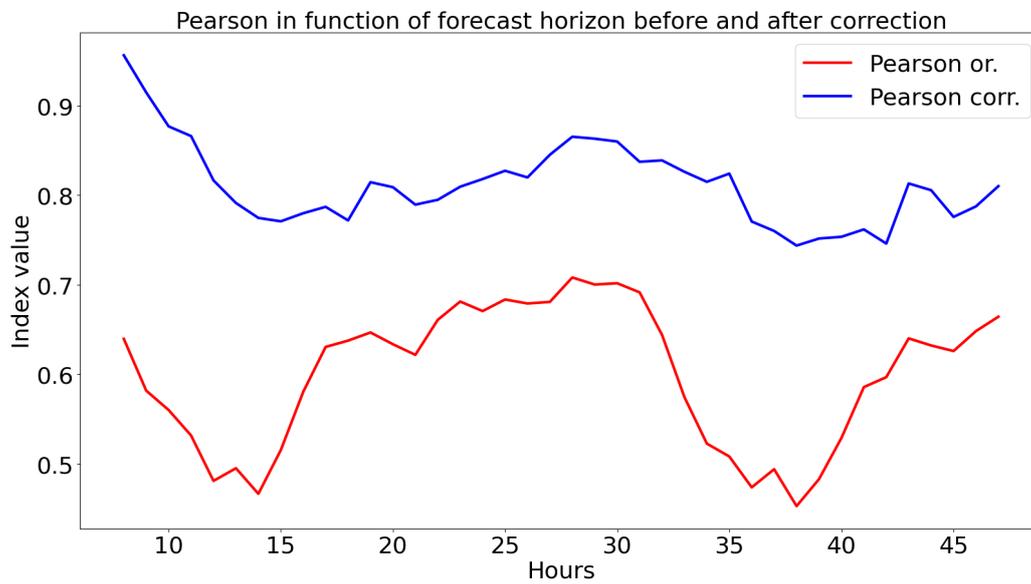


Figure 6.24: Trend of the correlation index (Pearson) as a function of the forecast horizon before and after the best correction.

6.4.4 Summary of how correction strategies affected the forecast.

In this section will be shown how the different strategies affected the final correction. To do this, the three error indices considered to be the main ones will be considered as well as those that provide a more intuitive physical meaning.

Figure 6.25 shows the NMAE, NRMSE and Pearson indices in relation to the type of strategy that was used to correct the forecast. Remember that the NMAE and the NRMSE tend to zero the more they are synonymous with good prediction while the Pearson more tends to 1 plus it is a symptom of good forecast (more specifically of forecast correlated to observation).

As seen, the simple calibration without any particular conditioning or strategy leads to significant improvements as regards the NMAE and NRMSE indices but not for the Pearson which remains constant. This is since these forecasts have a very strong systematic error that can be easily corrected by translating and greatly referring to these forecasts. The conditioning on the direction instead is that which has given the greatest contribution to the increase of the correlation; it is no coincidence that there has also been a clear improvement in the other two indices. For the rest, the other strategies were less effective and more or less all the same. In particular, the one that exploits the correlation of the error from this graph does not seem to have brought about any improvement, but this is due to the fact that unlike the other strategies it depended more on the proximity to the forecast time. The minor improvement of the indices that was seen downstream of the “dr” is however since a lot had already been considered with the simple regression “tt”.

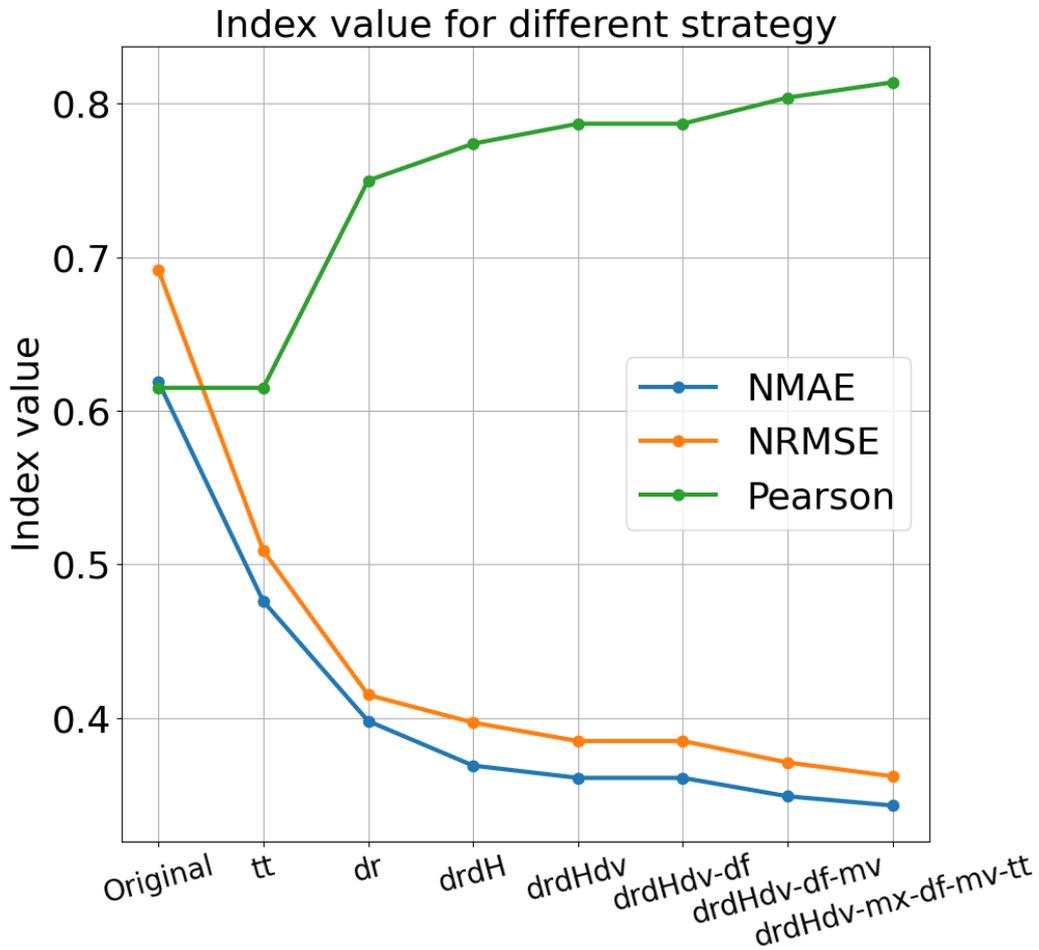


Figure 6.25: Graphical representation of the NMAE, NRMSE and Pearson indices of the forecast in the following order. Raw: Raw previews; tt: Predictions calibrated without any division of the same; dr: Forecasts calibrated by influencing the direction; drdH: Forecasts calibrated by conditioning on direction and time; drdHdv: Forecasts calibrated by conditioning on direction, time and speed; drdHdv-df: Forecasts calibrated by conditioning on direction, time and speed and then exploiting the correlation of the error; drdHdv-df-mv: Calibrated forecasts by conditioning on direction, time and speed, exploiting the correlation of the error and finally making a moving average; drdHdv-mx-df-mv-tt: Calibrated forecasts by conditioning on direction, time and speed, choosing the best spaghetti, exploiting the correlation of the error, making a moving average and finally redoing a calibration without conditioning. The forecasts in particular are those 24-48 hours that have been corrected using quantile regression.

6.5 Variation of indices as a function of the number of data for training

The aim of this sub-chapter is to analyse how the result obtained so far would vary according to how much data are available for training. For the analyses carried out so far, almost 1 year had been used to train. In truth, following the cleaning of the data, done in the previous chapter, there were 346 days left which is 19 days in less than a year. To see how the results obtained vary, it was decided to take the best static forecast as a reference point, i.e. the one that affected direction, time and speed. This is because this strategy is the one that, in addition to having generated the greatest improvement in forecasts, is the one that can be most affected by the variation of the training set. This is because undergoing many conditions, and therefore dividing the dataset into many parts, when there are few data, this aspect is amplified. For this analysis only the forecast deriving from the average of the EPS with a forecast horizon of 24-48 hours will be used. Furthermore, only quantile regression will be used.

Table 6.19 shows the values of the indices calculated as the months available for training vary. Several months ranging from 11 to 1 were used. In particular, fewer months were used when conditioning on hours instead of distinguishing hour by hour they were grouped into groups of 2, 3, 4, 6, 8. It is possible to do this because, as seen above, from now to now there is no substantial change in the phenomenological characteristics of the wind. It is therefore convenient to give up this conditionality a little to favour a more robust training. The table shows how the effectiveness of the correction strategy is logically progressively decreasing the months at worst dispositions. For greater clarity on the trend of the main error indices it is useful to see them in a graph.

Figure 6.26 shows how the NMAE, NRMSE, HH, Pearson indices vary according to the number of months available for training. It is quite clear that up to 7/6 months the strategy remains very valid and is very little affected by the decrease in the months of training. From 6 months to 3 months, the trend of the indexes to worsen increases a little but without then leading to excessively gross errors. For example, using 3 months of training, the NMAE still remains below 40%, as does the NRMSE below 45 and the Pearson above 72/73. From 3 months to 1 there is a more marked worsening which makes the strategy much less effective.

Table 6.19: Error indices varying the number of months used to train for the drdHdv strategy (depending on direction, time and speed). The predictions were obtained starting from the average of the EPS 24-48 hours and using quantile regression.

T.Months	Bias	NBias	MAE	NMAE	RMSE	NRMSE	HH	SI	Pearson
11	-0.271	-0.052	1.909	0.367	2.659	0.400	0.440	0.398	0.771
10	-0.285	-0.055	1.929	0.371	2.692	0.405	0.446	0.403	0.765
9	-0.274	-0.053	1.937	0.372	2.705	0.407	0.448	0.405	0.763
8	-0.283	-0.054	1.939	0.372	2.702	0.406	0.447	0.404	0.763
7	-0.298	-0.057	1.954	0.375	2.726	0.410	0.453	0.407	0.759
6	-0.303	-0.058	1.972	0.379	2.769	0.416	0.460	0.414	0.751
5	-0.312	-0.060	2.001	0.385	2.805	0.422	0.467	0.419	0.744
4	-0.304	-0.058	2.021	0.388	2.839	0.427	0.472	0.424	0.739
3	-0.281	-0.054	2.052	0.394	2.900	0.436	0.480	0.434	0.730
2	-0.283	-0.054	2.109	0.405	3.018	0.454	0.498	0.452	0.715
1	-0.374	-0.072	2.222	0.427	3.196	0.481	0.531	0.477	0.690

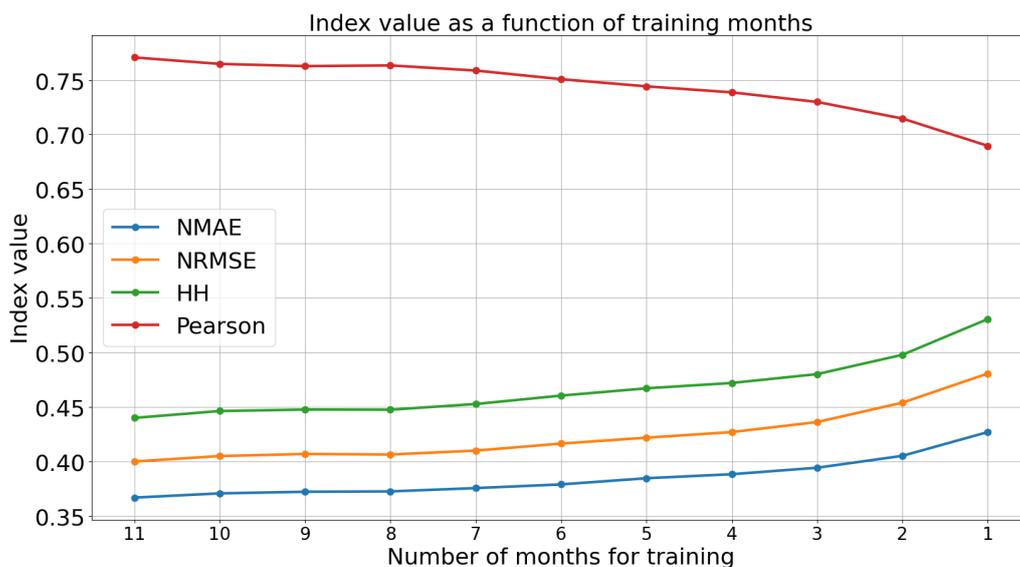


Figure 6.26: Trend of the NMAE, NRMSE, HH and Pearson index as a function of the number of months used to make the training for the correction drdHdv (depending on direction, time and speed). The predictions were obtained starting from the average of the EPS 24-48 hours and using quantile regression.

6.6 The prediction of power

Once the best strategies to correct the wind forecast have been defined, it is possible to transform this forecast into a power forecast. To do this, it is sufficient to use the empirical transfer function calculated on the observed data.

Table 6.21 shows the value of the indices calculated on the power forecast before and after correcting the wind forecast. As already seen, the raw forecast has very poor indices, even with NMAE and NRMSE that reach 100% (which can also be exceeded by being normalized for the average of the observations and not for the nominal power of the wind turbine). As for the correct forecasts, it is noted that the indices have improved to the point of being practically halved in the cases of the NMAE and NRMSE. In particular, the NBias still shows an underestimation of the power between 10 and 20%. The NMAE fluctuates around 50% as well as the NRMSE and the HH which is slightly above. The Pearson also undergoes a noticeable improvement from about 40 to 80%. By comparing these indices with those used for wind forecasting, it can be seen that, except for the Pearson and the non-normalized indices that it makes no sense to compare, they have worsened considerably. This is since, as seen in the previous chapter, the transfer function has a basic error due to the uncertainty of the instrument and also due to the fact that the transfer function is not linear and leads to the amplification of the error made on the predictions of the wind. It is curious, however, that Pearson has kept its value quite well without undergoing great variations. This is an excellent point in favour that it may be that following correct corrections there may still be an improvement in the forecast.

The only test that was carried out to see if there were any corrections that could be easily performed like those already carried out on the weather forecast was to condition on the direction, since it may be that the power produced may somehow be correlated to the direction with which the wind meets the wind turbines, and also conditioning on the power itself as it was done on the speed for the wind.

Table 6.21 shows the result of this correction. By comparing tables 6.20 and 6.21 this strategy did not lead to great improvements apart from a little bit for the Bias. However, in some cases the NMAE and NRMSE have undergone an improvement of half a percentage point.

Finally, figure 6.27 shows the time series of the power forecasts obtained starting from the members of the EPS with a forecast horizon of 24-48 hours. First, comparing the correction made, no particular differences appear with the forecast without the drdp correction. On the contrary, the difference between the raw forecast and the best one obtained is very marked.

Table 6.20: Comparison between the raw power prediction and the best prediction obtained starting from the average of the EPS and the individual members of the EPS using quantile regression.

Index	8-24 h power forecast			24-48 h power forecast		
	Raw	Best from EPS: mean	members	Raw	Best from EPS: mean	members
Bias	-381.0	-61.1	-41.3	-348.3	-70.4	-52.6
NBias	-0.951	-0.152	-0.103	-0.957	-0.193	-0.145
MAE	383.2	197.1	193.4	350.1	205.9	200.0
NMAE	0.957	0.492	0.483	0.962	0.566	0.550
RMSE	659.6	336.4	327.8	640.7	366.1	355.8
NRMSE	0.960	0.490	0.477	0.966	0.552	0.537
HH	4.647	0.566	0.536	5.089	0.667	0.630
SI	0.784	0.482	0.473	0.811	0.542	0.531
Pearson	0.411	0.807	0.817	0.384	0.764	0.777

Table 6.21: Comparison between the raw power prediction and the best prediction obtained starting from the average of the EPS and the individual members of the EPS using quantile regression and making a further correction on the expected power by conditioning on direction and power.

Index	8-24 h power forecast			24-48 h power forecast		
	Raw	Best from EPS: mean +drdp	members +drdp	Raw	Best from EPS: mean +drdp	members +drdp
Bias	-381.0	-47.4	-42.9	-348.3	-46.8	-48.7
NBias	-0.951	-0.118	-0.107	-0.957	-0.128	-0.134
MAE	383.2	195.4	191.9	350.1	205.3	199.9
NMAE	0.957	0.488	0.479	0.962	0.564	0.549
RMSE	659.6	335.2	327.6	640.7	363.5	352.9
NRMSE	0.960	0.488	0.477	0.966	0.548	0.532
HH	4.647	0.554	0.535	5.089	0.644	0.624
SI	0.784	0.483	0.473	0.811	0.544	0.527
Pearson	0.411	0.809	0.818	0.384	0.766	0.780

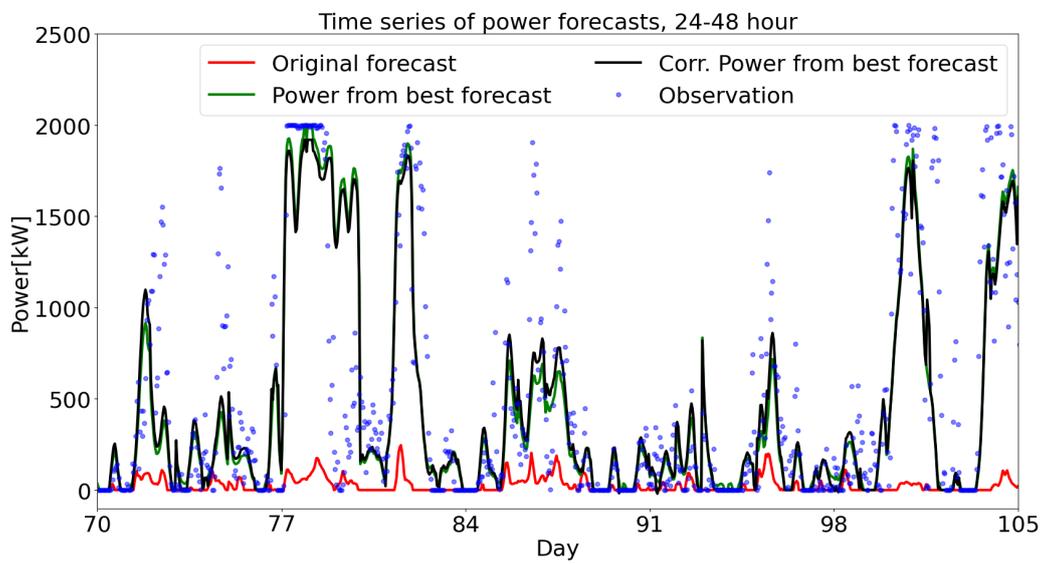


Figure 6.27: Time series of the raw predicted power (red) compared with the predicted power obtained by making the transfer function on the best forecast starting from the members of the EPS (green) and with that obtained by further adding a correction by conditioning on direction and power (black).

6.7 Comparison with other forecast calibration techniques

In this section comparisons will be made with other techniques with which predictions can be corrected. Two cases will be seen. The first will be applied to the same data that has been used up to now using a machine learning technique. The second will be a case taken from a scientific article that concerns another wind farm.

6.7.1 Comparison with Machine Learning

Machine learning (ML) is the study of computer algorithms that improve automatically through experience. It is seen as a subset of artificial intelligence. Machine learning algorithms build a model based on sample data, known as “training data”, in order to make predictions or decisions without being explicitly programmed to do so.[16]

The intent is to make a comparison between the results obtained so far with those that can be obtained with a particular machine learning technique. To do this it is therefore necessary to start from the same data. The data taken as reference only the average forecasts of the EPS. The best strategy will therefore initially be compared, starting from the average of the EPS with the result of Machine learning.

Machine learning was performed by the company that dispatches the energy of the wind farm analysed. It is therefore not possible to provide many details of how this machine learning was performed. Machine learning was implemented using the python XGBoost algorithm. XGBoost is an optimized distributed gradient boosting library designed to be highly efficient, flexible, and portable. It implements machine learning algorithms under the Gradient Boosting framework. Gradient boosting is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function.[17] XGBoost provides a parallel tree boosting (also known as GBDT, GBM) that solve many data science problems in a fast and accurate way.

Table 6.22 shows the values of the error indices of the raw forecast (average of EPS), of the best forecast obtained starting from the average of the EPS and of the forecast obtained with machine learning always starting from the average of the EPS. As it can be seen, with the use of Machine learning index

values are got compared to the starting ones. In fact, the differences are not particularly big. NBias for the 0-24 hour case is better with the drdHdv-df-mv technique while with the 24-48 hour case it is better with Machine learning. The other indices are all in favour of the drdHdv-df-mv strategy; in particular, the NMAE is better than 2/3%, the NRMSE between 0.5 and 3% as well as for HH and Pearson. It is interesting to underline that in the 8-24 hours there is a greater difference than in the 24-48 hours. This could be due to the strategy that exploits error correlation as it was very effective in the first 24 hours and very little from 24 to 48 hours.

Table 6.22: Comparison between the best forecast obtained starting from the average of the EPS with the strategies shown previously with the Machine learning strategy.

Index	8-24 h forecast			24-48 h forecast		
	Raw	drdHdv-df-mv	Machine learning	Raw	drdHdv-df-mv	Machine learning
Bias	-3.273	-0.107	-0.189	-2.941	-0.142	-0.013
NBias	-0.576	-0.019	-0.033	-0.565	-0.027	-0.002
MAE	3.484	1.686	1.877	3.223	1.816	1.925
NMAE	0.613	0.297	0.331	0.619	0.349	0.370
RMSE	4.723	2.262	2.456	4.599	2.466	2.497
NRMSE	0.681	0.326	0.354	0.692	0.371	0.376
HH	1.162	0.349	0.380	1.196	0.405	0.406
SI	0.491	0.326	0.353	0.532	0.370	0.376
Pearson	0.584	0.823	0.789	0.615	0.804	0.797

Table 6.23 shows the comparison between the best expected power obtained so far starting from the average of the EPS with the forecast of the power generated by Machine learning always starting from the average of the EPS. NBias is clearly in favour of Machine learning with a difference of about 10%. Even the NRMSE is better with the prediction produced by Machine learning as well as the HH albeit in low percentages. The only index that remains in favour of the forecast generated in this treaty is the NMAE. This is a point in favour of this strategy because as mentioned previously when it comes to forecasting power, NMAE is the most important index. Therefore, this result could be due to the fact that using quantile regression this index has been minimized to the detriment of the others. It is very interesting that Pearson is also better for the Machine learning strategy. This means that the ML is able to find correlations that were not found with the simple regres-

sion conditioning on direction and power that was performed. Therefore, it is possible that by investigating where Machine learning finds this correlation, a strategy can be identified to further improve the forecast.

Table 6.23: Comparison between the raw power prediction of the EPS average with the best generated by the EPS average with the best one of machine learning.

Index	8-24 h power forecast			24-48 h power forecast		
	Raw	Best from EPS mean	Machine learning	Raw	Best from EPS mean	Machine learning
Bias	-381.0	-47.4	3.9	-348.3	-46.8	4.0
NBias	-0.951	-0.118	0.010	-0.957	-0.128	0.011
MAE	383.2	195.4	223.6	350.1	205.3	216.8
NMAE	0.957	0.488	0.558	0.962	0.564	0.596
RMSE	659.6	335.2	333.3	640.7	363.5	332.2
NRMSE	0.960	0.488	0.485	0.966	0.548	0.501
HH	4.647	0.554	0.548	5.089	0.644	0.578
SI	0.784	0.483	0.485	0.811	0.544	0.501
Pearson	0.411	0.809	0.803	0.384	0.766	0.800

Table 6.24 shows the error indices of the best correction starting from the individual members of the EPS, and therefore the absolute best forecast so far, with the forecast generated by feeding Machine learning this last forecast. The improvement due to machine learning leads to almost zero benefit; in fact, the indices have almost equal values. This confirms how the strategies implemented to arrive at this forecast have been able to collect almost all the correlation available.

Finally, table 6.25 shows the result of having applied Machine learning to the best prediction of the power starting from the average of the EPS. It is very interesting to make a comparison between this table and 6.23. In fact, against all predictions, instead of improving, the forecast seems to have worsened compared to that obtained by machine learning starting from the raw data. This can therefore be an excellent clue as to how a strategy to improve forecasting could be adopted in the future. In fact, this result leads us to think that the transition from wind forecast to power forecast through the transfer function leads to the loss of important information that can be used for a possible correction. For this reason, therefore, Machine learning has a greater result starting from the wind; it may therefore be that he has found a correlation that exists between the forecast of the wind and the

Table 6.24: Comparison between the best available wind forecast (drdHdv-mx-df-mv-tt) with the forecast obtained using machine learning in cascade at the best forecast

Index	8-24 h forecast			24-48 h forecast		
	Raw	drdHdv-mx-df-mv-tt	+Machine learning	Raw	drdHdv-mx-df-mv-tt	+Machine learning
Bias	-3.273	0.033	0.048	-2.941	-0.037	0.023
NBias	-0.576	0.006	0.008	-0.565	-0.007	0.004
MAE	3.484	1.662	1.661	3.223	1.784	1.764
NMAE	0.613	0.293	0.292	0.619	0.343	0.339
RMSE	4.723	2.221	2.236	4.599	2.407	2.409
NRMSE	0.681	0.320	0.323	0.692	0.362	0.362
HH	1.162	0.338	0.340	1.196	0.390	0.389
SI	0.491	0.320	0.323	0.532	0.362	0.362
Pearson	0.584	0.830	0.827	0.615	0.814	0.813

observed power which, however, does not exist between the forecast of the power and the observed power.

Table 6.25: Comparison between the best prediction of the power obtained starting from the average of the EPS with that obtained by making a cascade Machine learning.

Index	8-24 h power forecast			24-48 h power forecast		
	Raw	Best from EPS mean	+Machine learning	Raw	Best from EPS mean	+Machine learning
Bias	-381.0	-47.4	15.2	-348.3	-46.8	-14.0
NBias	-0.951	-0.118	0.038	-0.957	-0.128	-0.038
MAE	383.2	195.4	211.8	350.1	205.3	218.0
NMAE	0.957	0.488	0.529	0.962	0.564	0.599
RMSE	659.6	335.2	327.5	640.7	363.5	355.3
NRMSE	0.960	0.488	0.477	0.966	0.548	0.536
HH	4.647	0.554	0.534	5.089	0.644	0.635
SI	0.784	0.483	0.476	0.811	0.544	0.535
Pearson	0.411	0.809	0.811	0.384	0.766	0.768

6.7.2 Comparison with a scientific article of a forecast generated in a site similar to the one analyzed

In this section a comparison will be made with a scientific article of an analysis on the predictions made for a site similar to the one analysed so far. The chosen article is: Wind Power Forecasting techniques in complex terrain: ANN vs. ANN-CFD hybrid approach.[18]

The following explanation shows the strategy used in this article. Wind power forecast is calculated applying a post-processing on the NWP (numerical weather prediction) output. In this study, the weather research forecast Model - WRF is used. Each of the two proposed methods is composed of different steps, leading to the estimate of the power production of a wind farm. The two approaches employed in this study can be summarized as follows:

- A single ANN (Artificial Neural Networks, a statistical method) processes the output of the NWP model and directly calculates the power production of each single turbine or the whole wind farm. This is the pure ANN approach: the ANN has wind speed and direction of the wind as input variables and power production as output variable.
- An ANN processes the wind conditions, as predicted by the NWP model, targeting the wind conditions on site. These are used as input to the CFD (computational fluid dynamics), in order to transfer the forecast from the reference wind measurement position to the positions of the turbines. The nominal power curve is employed for estimating the power output. This is the hybrid approach: the ANN has NWP wind speed and wind direction as input variables and observed wind speed and direction as output variable. The CFD flow simulations enable to transfer the wind conditions in the layout area up to the position of each turbine, and to calculate the power production.

The first approach is purely statistical: the ANN stores the correlation information between wind speed and direction from NWP and the measured power production. Such an approach can be seen as an artificial neural network power curve (ANN wind-power). The second approach is more complex, a hybrid of statistical and deterministic methods. The ANN acts as an MCP (measure correlate predict), detecting and using the correlation of the wind data between two time series. In this case, the correlation is used between NWP data and observed wind speed and direction at target site. The output of the ANN is used, within the CFD framework, to transfer the forecast from the wind measurement position to the positions of the turbines,

considering the simulated behaviour of the wind depending on the direction of the flow. Also, the wake effects between turbines are taken into account using the Jensen model and the nominal turbine thrust coefficient C_t . The calculation is performed using WindSim software. In short, this approach can be defined as ANN windwind + CFD method. In both methods, the ANN is fed with the same data in the input layer: time series of wind speed and direction obtained by the NWP model at a defined position in the wind farm layout. Depending on the method, the ANN gives different outputs: the power production of each single turbine (or of the complete wind farm) for the pure ANN approach, or the wind speed and direction at the target reference point for the hybrid method. Therefore, the two ANNs differ in the number and type of the variables set at the output layer. In both cases, the ANNs are single layer perceptrons, trained by feed-forward back-propagation method, unsupervised training. The ANNs can be set with different number of neurons in the inner layer and the performance is sensitive to such a setting. Therefore, many configurations are tested, and the best is chosen. To simulate the run of a real day ahead, as the forecast has to be done in the morning for the day after, 18 hours of each forecast run are cut out and the following 24 hours are used

The wind farm used for validation is sited in southern Italy. On site, 24 turbines are installed. The terrain is extremely complex: the presence of mountains is important in all the directions and there are severe slopes all around.

Therefore, the site considered in this article is very similar to the one considered in the thesis. In fact, both wind farms are located in a complex orography environment due to the constant presence of nearby mountain ranges.

Figure 6.28 shows the table of results present in the article cited. This table shows the Bias, RMSE, NMAE and NRMSE indices. Logically it is not possible to make a comparison with non-normalized data as the dimensions of the plants considered are clearly different. It is possible instead to make a comparison with the two normalized indices. To do this, however, it is necessary to normalize the MAE and RMSE indices seen so far with the same technique used in the article. In fact, while up to now these indices have been normalized as a function of the average of the observations, for the MAE, and of the average of the observations squared, for the RMSE, in the article they are normalized as a function of the nominal power.

The nominal power of the plant considered is 2000 kW. It will therefore be sufficient to divide the values of MAE and RMSE by this value. Since the article only considers the forecast for the next day (24-48 hours), only the 24-48 hour forecast will be considered to make the comparison more meaningful

NWP height [m] AGL	Layout	Technique	Bias [kW]	RMSE [kW]	NMAE	NRMSE	Points
10	1	ANN	-336.682	1723.981	0.209	0.261	2196
100	1	ANN	-318.798	1694.480	0.206	0.257	2196
200	1	ANN	-307.642	1682.663	0.205	0.255	2196
300	1	ANN	-346.803	1707.495	0.207	0.256	2196
400	1	ANN	-342.965	1769.535	0.216	0.268	2196
10	1	ANN+CFD	-416.747	1865.727	0.206	0.283	2196
100	1	ANN+CFD	462.121	1803.351	0.199	0.273	2196
200	1	ANN+CFD	-417.957	1789.776	0.200	0.271	2196
300	1	ANN+CFD	-504.929	1821.040	0.205	0.276	2196
400	1	ANN+CFD	-562.539	1902.911	0.213	0.288	2196
10	2	ANN	-193.235	1973.729	0.168	0.227	1847
100	2	ANN	-166.046	1920.192	0.162	0.221	1847
200	2	ANN	-198.062	1909.207	0.160	0.219	1847
300	2	ANN	-206.821	1900.868	0.160	0.219	1847
400	2	ANN	-237.339	1912.797	0.162	0.220	1847
10	2	ANN+CFD	-36.982	2136.938	0.169	0.246	1847
100	2	ANN+CFD	7.079	2068.746	0.161	0.238	1847
200	2	ANN+CFD	-121.554	2068.857	0.161	0.238	1847
300	2	ANN+CFD	-78.994	2085.767	0.162	0.240	1847
400	2	ANN+CFD	-115.361	2114.197	0.164	0.243	1847

Figure 6.28: The figure shows the table taken from the cited article. There are indices of the error as the technique used varies. The indexes present only the Bias, the RMSE, the NMAE, and the NRMSE. The NMAE and NRMSE were normalized using the nominal power of the plant (Layout 1 6600 [kW], Layout 2 8700 [kW]). (Taken from Ref. [18])

and correct.

Table 6.26 therefore shows the NMAE and NRMSE normalized with the nominal power for the prediction of the power at 24-48 hours. These values are therefore comparable with those seen in figure 6.28. Well, the lowest value of the article cited of the NMAE is greater by about 6 percentage points than that found with the analyses of this study. The same goes for the NRMSE which is below the 20% threshold as opposed to the NRMSE of the article which always exceeds this threshold.

Table 6.26: Value of the NMAE, NRMSE indices of the power forecast normalized using the nominal power of the plant (2000 kW). The forecast is the best obtained so far for a 24-48 hour forecast horizon.

Index	Value
NMAE	0.100
NRMSE	0.176

Conclusions

In this thesis, the strategies implemented to improve the forecast of wind speed were analysed step by step in order to have a forecast of the best possible power output. These analyses led to different results and different perspectives for future studies.

The first derive from the comparison between the EPS forecasts and the HRES forecast. Although the HRES forecast is the result of a model with a resolution double than of EPS, it has been seen that the second forecast is better, albeit slightly, than the first. This highlights the fact that a strategy involving multiple predictions with less resolution appears to be better than a single prediction but with better resolution.

Another result is that the use of the CRPS technique with the gamma distribution has led to results very similar to those obtained using quantile regression which by definition is the best to obtain the absolute lowest NMAE. The difference that leads to the use of one of the two techniques is small enough to be ignored since with the CRPS not only a precise forecast is obtained but also the variance of this forecast dynamically.

Carrying on the third result is about the strategies used. In fact, it has been seen that in a site with a complex topography such as the one considered, distinguishing the forecast according to the direction of the event is essential to improve it. That also the time of day leads to different correlations between forecast and observation and finally that distinguishing in speed intervals does not lead instead to particular characteristics. It has also been seen how exploiting the correlation of the prediction error is very effective in the hours immediately after but that after a while there is a rapid decay of the correlation in reverse instead of choosing the right member that seems to maintain a better correlation with advancing the forecast horizon.

Furthermore, it has been shown that using 50 different forecasts and averaging them results in a slower decay in time of the quality of the forecast compared to a single forecast but with higher resolution.

It was also analysed how the strategy of conditioning as a function of direction, time and speed varies according to how many historical data on which

it is possible to train exist, resulting in that up to 3 months of historical series it is possible to use the strategy effectively.

Finally, it was shown by a comparison with a Machine Learning technique and with the forecast of another study, in a site similar to the one analysed, the goodness of the calibration of the prediction made. Moreover, from the comparison with Machine Learning it appeared that it could be convenient, to predict the power, to carry out calibrations starting directly from the wind forecast and not from that already transformed into power.

Some prospects for the future also appeared from this study.

The forecasts of the ECMWF in fact do not only concern the wind speed but also all the other physical phenomena. One perspective is to extend this method to the forecast of the waves in the Mediterranean. In fact, even the waves have a strong dependence on the direction. More generally, the calibration technique can also be extended to other physical variables such as the ground temperature or even for irradiation. The latter is in fact fundamental for the forecast of the energy production from a solar system which, as for the wind, foresees the same market rules.

Another idea is to combine several forecasts to exploit the correlation of both; for example, can be successfully combine the HRES and EPS forecasts with coefficients.

Another aspect is to use the forecasts at different altitudes or even in the nodes close to the one considered to see if by inserting this additional variable there is no greater correlation.

Yet another field to explore is that of the correlation of error. In fact, in this analysis the error was correlated without any conditioning; on the other hand, it may be that by inserting conditions on it, it is possible to arrive at better predictions. For example, a condition that could be useful is that of direction because doing so would only correlate the wind of the same “family” and not two winds with two different directions which obviously cannot have the same characteristics.

Finally, it is necessary to deepen the prediction related to power. In fact, from the comparison with Machine Learning, it appeared that using the simple transfer function leads to the loss of information useful for finding the correlation. In fact, it could be possible to identify a non-linear function that describes the transfer curve in order to be able to make fits from the wind to the power directly without having to go through the transfer function.

Bibliography

- [1] Rivarolo M., 2018
I mercati elettrici. University of Genoa, Technical report

- [2] RSE, Politecnico di Milano, 2018
Partecipazione della domanda flessibile al mercato del servizio di dispacciamento. Book.

- [3] Persson A., 2001
User Guide to ECMWF forecast products. Meteorological Bulletin M3.2

- [4] Buizza R., Miller M., and Palmer T. N., 1999
Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, 125, 2887–2908.

- [5] Molteni F., Buizza R., Palmer T. N., and Petroliagis T., 1996
The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy Meteor. Soc.*, 122 , 73–120.

- [6] Leutbecher M., and Palmer T. N., 2008
Ensemble forecasting. *J. Comput. Phys.*, 227 , 3515–3539

- [7] Botchkarev A., 2018
Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology, arXiv preprint.

- [8] Koh T.Y., Wang S., Bhatt B.C., 2012.
A diagnostic suite to assess NWP performance. *J. Geophys. Res.* 117:

D13109

- [9] Chai T., Draxler R.R. ,2014
Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. *Geosci Model Dev* 7:1247–1250
- [10] L. Mentaschi L., Besio G., Cassola, Mazzino A., 2013
Problems in RMSE-based wave model validations, *Ocean Modelling* 72, 53-58
- [11] Wei Y., Pere A., Koenker R., et al., 2006
Quantile regression methods for reference growth charts, *Stat Med*, 2006, vol. 25, 1369-82
- [12] Gneiting, T., Raftery A. E., A. H. Westveld A. H. III, and Goldman T., 2005
Calibrated probabilistic forecasting using ensemble Model Output Statistics and minimum CRPS estimation. *Mon. Wea. Rev.*, 133 , 1098–1118.
- [13] Sloughter J. McLean, Gneiting Tilmann, Raftery Adrian E., 2010
Probabilistic Wind Speed Forecasting Using Ensembles and Bayesian Model Averaging. *Journal of the American Statistical Association* 105:489, 25-35.
- [14] Baran S., and Nemoda D., 2016
Censored and shifted gamma distribution based EMOS model for probabilistic quantitative precipitation forecasting. *Environmetrics*, 27, 280–292.
- [15] Hogg R., Craig A.T., 1995
Introduction into mathematical statistics Prentice Hall, Englewood Cliffs, NJ
- [16] Mitchell T. M., 1999
Machine learning and data mining, *Commun. ACM*, vol. 42, no. 11.

- [17] Friedman J., 2001
Greedy function approximation: A gradient boosting machine. *Ann. Statist.* 29, 1189-1232.
- [18] Castellani F., Astolfi D., Mana M., Burlando M., Meißner C., and Piccioni E., 2016.
Wind power forecasting techniques in complex terrain: ANN vs. ANN-CFD hybrid approach. In *Journal of Physics: Conference Series* (Vol. 753, No. 8, p. 082002)