MINES PARIS

Collège doctoral

N°attribué par la bibliothèque

|__|__|__|__|__|__|__|__|__|__|

T H E S E

pour obtenir le grade de

Docteur de l'Ecole des Mines de Paris

Spécialité "Energétique"

présentée et soutenue publiquement par

Pierre Pinson

le 23 mars 2006

ESTIMATION OF THE UNCERTAINTY IN WIND POWER FORECASTING

(ESTIMATION DE L'INCERTITUDE DES PREDICTIONS DE PRODUCTION EOLIENNE)

Directeur de thèse : Georges KARINIOTAKIS

Jury :

Pr. Arthouros ZERVOS (NTUA, Greece) ....Président
Pr. Henrik MADSEN (DTU, Denmark)...Rapporteur
Pr. Vladimiro MIRANDA (INESC, Portugal)....................................Rapporteur
Dr. Frédéric ATGER (Météo-France, France) .... Examinateur
Dr. Etienne BRIERE (EDF, France) .... Examinateur
Dr. Georges KARINIOTAKIS (ENSMP, France) .... Examinateur

"Forecasting is very difficult, especially if it's about the future..."

– Niels Bohr –

"If a man will begin with certainties, then he shall end in doubts;
but if he will content to begin with doubts he shall end in certainties..."

– Francis Bacon –

Jury:

Pr. Arthouros Zervos, National Technical University of Athens, Greece (Président)
Pr. Henrik Madsen, Technical University of Denmark, Denmark (Rapporteur)
Pr. Vladimiro Miranda, Institute for Systems and Computer Engineering, Portugal (Rapporteur)
Dr. Frééric Atger, Météo France, France (Examinateur)
Dr. Etienne Brière, Electricité de France, France (Examinateur)
Dr. Georges Kariniotakis, Ecole des Mines de Paris, France (Examinateur)

Abstract

WIND POWER experiences a tremendous development of its installed capacities in Europe. Though, the intermittence of wind generation causes difficulties in the management of power systems. Also, in the context of the deregulation of electricity markets, wind energy is penalized by its intermittent nature. It is recognized today that the forecasting of wind power for horizons up to 2/3-day ahead eases the integration of wind generation. Wind power forecasts are traditionally provided in the form of point predictions, which correspond to the most-likely power production for a given horizon. That sole information is not sufficient for developing optimal management or trading strategies. Therefore, we investigate on possible ways for estimating the uncertainty of wind power forecasts. The characteristics of the prediction uncertainty are described by a thorough study of the performance of some of the state-of-the-art approaches, and by underlining the influence of some variables e.g. level of predicted power on distributions of prediction errors. Then, a generic method for the estimation of prediction intervals is introduced. This statistical method is non-parametric and utilizes fuzzy logic concepts for integrating expertise on the prediction uncertainty characteristics. By estimating several prediction intervals at once, one obtains predictive distributions of wind power output. The proposed method is evaluated in terms of its reliability, sharpness and resolution. In parallel, we explore the potential use of ensemble predictions for skill forecasting. Wind power ensemble forecasts are obtained either by converting meteorological ensembles (from ECMWF and NCEP) to power or by applying a poor man's temporal approach. A proposal for the definition of prediction risk indices is given, reflecting the disagreement between ensemble members over a set of successive look-ahead times. Such prediction risk indices may comprise a more comprehensive signal on the expected level of uncertainty in an operational environment. A probabilistic relation between classes of risk indices and the level of forecast error is shown. In a final part, the trading application is considered for demonstrating the value of uncertainty estimation when predicting wind generation. It is explained how to integrate that uncertainty information in a decision-making process accounting for the sensitivity of end-users to regulation costs. The benefits of having a probabilistic view of wind power forecasting are clearly shown.

Préface

I have started with the research works gathered in this thesis in October 2002, almost by coincidence. The path I was following was definitely not leading me towards pursuing Ph.D. studies. Though, series of fortunate events made that my resume ended up in the hands of Georges Kariniotakis, at the Centre for Energy and Processes of École des Mines de Paris, who proposed me to work with him in the field of wind power forecasting. I kindly thank Gilles Guérassimoff, François Neirac, Didier Mayer, and of course Georges for having given me that possibility. I must also thank all these fortunate events... The public defense of the thesis took place on March 23rd, 2006, in Sophia Antipolis.

'Professional' acknowledgements

All the participants in the European project ANEMOS are fully acknowledged for a fruitful collaboration. More particularly, I would like to further acknowledge ESB (Electricity Supply Board of Ireland), Elsam and IDAE (Instituto para la Diversificación y Ahorro de la Energía) for providing the datasets (both power measures and HIRLAM meteorological predictions) that are used in the present thesis.
In addition, I must thank the members of the jury, of which Pr. Zervos accepted to be the president, and of which Pr. Madsen and Pr. Miranda have accepted to be the censors. All the members of the jury have provided me with valuable comments for the present thesis, and comforted me in future directions to give to my research on wind power forecasting.

'Personal' acknowledgements

This research area is a perfect combination of several topics I have always dreamed to deal with, namely mathematical modeling, meteorological prediction, and wind energy. In addition, my aim was to carry out research in an international framework, and this has been

possible mainly thanks to my participation in the Anemos project. It has been a pleasure for me to work and collaborate with all the participants in that project: some of them allowed me to use data that were essential for my research work, and some others took time for discussing (or maybe better say arguing...) my opinions on some topics related to this work.

During these last years, I have had the chance to meet experts in wind energy, meteorology and mathematics. Most of these meetings have significantly influenced my view of forecasting and decision-making. For instance, I remember my meeting with Frédéric Atger from Météo France in June 2003. I am grateful that he spent time for reviewing my preliminary developments and for the discussion we had on his view of forecasting. Similarly, I must thank the WPPT team (i.e. Henrik Madsen, Henrik Aa. Nielsen and Torben S. Nielsen), at the Technical University of Denmark, for having allowed me to come and collaborate with them on some particular points of my work. This collaboration has considerably helped me in developing a broader view on wind power forecasting, from some basic to more challenging problems. Obviously, I am grateful to the various members of the jury, who spent time for reading the thesis, and giving me comments on my proposals, assumptions and developments. Finally, I have to thank Georges Kariniotakis for having given me part of his precious time in order to transmit his expertise in wind energy and distributed generation.

I am sorry to say that these experts have not been the only people to provide me with valuable contribution for developing this research work. For instance, I have had the pleasure to be assisted by two master students who participated in specific points. And, 'introducing' them to wind power forecasting obliged me to have a clear view on my objectives and ways to meet them. Thank you Nils and Christophe. I hope you will be successful with your Ph.D. studies. In addition, all my collegues at École des Mines de Paris are to be acknowledged for their professional and kind support. Special thanks are due to Alain who was always available for IT issues, and also because he saved my laptop few weeks before the defense.

Now is the more tricky part where I have to thank relatives, and people that I consider as relatives. First, there is that group of 'crazy guys' I met during my first year in Toulouse, when starting my studies at the Institut National des Sciences Appliquées. We are best friends since that time and I know they have had a great influence in my approach to life. Thanks for that. Then, I have to thank all my family for their support, whenever and whatever. More particularly, I think about my parents, who continuously encouraged my choices and were present in bad moments. I also need to thank my twin brother for so many things that I cannot mention them here... Last, and certainly not least, Pernille, 'mange tak' for your patience and your affection...

Contents

Abbreviations

a.g.l.	Above ground level
ADEME	French Environment and Energy Management Agency
APX	Amsterdam Power eXchange
AWPPS	Armines Wind Power Prediction System
CFD	Computational Fluid Dynamics
CO_2	Carbon Dioxide
EWEA	European Wind Energy Association
Fuzzy-NN	Fuzzy-Neural Networks
HIRLAM	High Resolution Limited Area Model
HIRPOM	HIRlam POwer prediction Model
i.i.d.	Independant and identically distributed
IPP	Independant Power Producer
ISET	Institut für Solare Energieversorgungstechnik
LOCALS	Local Circulation Assessment and Prediction System
LS-SVR	Least-Square Support Vector Regression
MAE	Mean Absolute Error
MOS	Model Ouput Statistics
MSE	Mean Square Error
NETA	New Electricity Trading Arrangement
NMAE	Normalized Mean Absolute Error
NN	Neural Network
NO_x	Oxides of Nitrogen
NPRI	Normalized Prediction Risk Index
NRMSE	Normalized Root Mean Square Error
NSDE	Normalized Standard Deviation of the Errors
NWP	Numerical Weather Prediction
PC	Probabilistic Choice
PCA	Principal Component Analysis
PTU	Program Time Unit
RA	Risk Analysis
RMSE	Root Mean Square Error
RUC	Rapid Update Cycle
SCADA	Supervisory Control And Data Acquisition
SDE	Standard Deviation of the Errors
SO_2	Sulfur Dioxide
SVM	Support Vector Machine
TenneT	TSO for the Netherlands
TSO	Transmission System Operator
WPPT	Wind Power Prediction Tool

Notations and Mathematical Symbols

$(1-\alpha)$	Nominal coverage rate
$a^{(\alpha)}$	Empirical coverage of the $(1-\alpha)$-confidence prediction intervals
A_r	Rotor swept area exposed to the wind
A_w	Weibull scale parameter
$\mathcal{A}$	Fuzzy set
bias	Bias of a prediction model
B	Number of bootstrap replications
c	Forecast condition
$\mathcal{C}$	Set of forecast conditions
$\delta^{(\alpha)}$	Width of the $(1-\alpha)$-confidence prediction intervals
d	Imbalance
ϵ	Normalized prediction error
e	Prediction error
E	Amount of wind energy
F	Distribution function
γ	Performance ratio associated to bidding strategies
Γ	Gamma function
G	Cumulative distribution function
$\hat{\mathrm{I}}^{(\alpha)}$	Prediction interval with nominal coverage rate $(1-\alpha)$
$\mathcal{I}^{(\alpha)}$	Indicator variable for the $(1-\alpha)$-confidence prediction intervals
κ	Kurtosis of a probability distribution
k	Prediction horizon (alternatively referred to as look-ahead time)
k_w	Weibull shape parameter
$\hat{L}^{(\alpha)}$	Lower bound of the $(1-\alpha)$-confidence prediction interval
$\mathcal{N}(\mu,\sigma^2)$	Normal distribution, with mean μ and variance σ^2
Ω	Set of prediction errors
π	Prices in the spot/regulation markets
p	Wind power
$\mathrm{P}(X)$	Probability for the occurence of the event X
P_n	Wind farm nominal power
μ	Mean of a probability distribution
m	Membership function related to a fuzzy set
ν	Skewness of a probability distribution
ρ	Coefficient of correlation
ρ_{air}	Air density
$r^{(\alpha)}$	Quantile with proportion α of a probability distribution
R	Revenue of a participant in an electricity market
R^2	Coefficient of determination

σ	Standard deviation of a probability distribution
S	Sample of prediction errors
Sc	Scoring rule for probabilistic forecasts
θ	Wind direction
t	Time
t_r	Temporal resolution of forecast series
T	Imbalance/regulation costs
u	Wind speed
$\hat{U}^{(\alpha)}$	Upper bound of the $(1-\alpha)$-confidence prediction interval
x_t	Measured value of the x-variable at time t
$\hat{x}_{t+k/t}$	Prediction of the value of the x-variable at time t for time $t+k$
v	Influential variable
w	Weight of a fuzzy rule
ζ	Wind power penetration
z	Height above ground level
z_0	Roughness length

1

Introduction

1.1 General context

TODAY, wind farm installations in Europe exceed 40 GW. Motivated by the Kyoto Protocol, the European Commission has set the target of doubling the share of renewables in gross energy consumption from 6% in 1997 to 12% in 2010 [66]. This directive targets 22,1% indicative share of electricity produced from renewable energy sources in total Community electricity consumption by 2010. To achieve this share, installed wind power capacity in the Member States should reach 45 GW. The European Wind Energy Association (EWEA) has set its own target to 60 GW, which was revised upwards in 2003 to 75 GW [242]. Wind energy is considered as the fastest growing technology in the landscape of the alternative power generating sources. Moreover, it appears to be a clean and cost-effective energy source [91].

Certain countries, such as Germany, Denmark and Spain, have managed to perform large-scale integration of wind generation on land[1]. Future major developments of wind power capacities are more likely to take place offshore. Higher and more regular wind speeds [189], availability of space that permits to install large wind farms, and less difficulties with local population acceptance, are the main advantages of going offshore to produce electricity. Important offshore projects are currently in progress with Horns Rev being a pioneer wind farm, in operation since end of December 2002. This wind farm supplies alone 2% of the whole electricity consumption of Denmark [212]. Several other ambitious offshore projects are under study in some European countries like United Kingdom and

[1] Installed wind power capacities in these three countries alone represent more than 80% of the total capacity installed in Europe.

Germany among others. Indeed, offshore wind energy could be sufficient to feed the local demand in countries like United Kingdom or Denmark [1].

France has a peculiar position in this context. Despite the fact that it has one of the best wind power potentials of Europe [14] (similar to the one of the United Kingdom), wind energy struggles to take its place in the French energetic landscape. To achieve its energetic independence, France has chosen three decades ago to invest in nuclear power, and is now a world leader in this field. Moreover, although it seems that the French population is in majority in favor of wind power, even in the areas with a lot of wind parks [236], anti-wind lobbies are very active in communicating their disagreement with the installation of wind farms[2]. However, perspectives drawn by the ADEME (French Environment and Energy Management Agency) are rather optimistic for the next years: forecasts of the installed capacities for wind generation (including offshore installations) reach 7 GW in 2010 and 28 GW in 2020 [37, 38]. To support this development, incentive feed-in tariffs were set in June 2001. In addition, studies about the advantages of geographically dispersed wind generation throughout the country [8] aim at showing how wind will behave in the current French power system and at providing guidelines concerning the future installations.

Either onshore or offshore, such a large-scale integration of wind generation is expected to cause several difficulties in the management of a power system, wind being highly variable by nature. Wind generation is traditionally seen as fatal by utilities: a high level of reserves is often allocated to account for the intermittent profile of wind production, thus reducing the benefits from the use of wind energy. Typically, a wind farm capacity factor — i.e. the ratio of actual energy output to the amount of energy a project would produce if it operated continuously at full rated power within a given time period — is between 25% and 40%. A first solution for smoothing this intermittent behavior is to combine the output of several wind farms. Indeed, by distributing the resource over a large area with different wind flows, wind generation variability can be significantly lowered [75]. Another solution would be to use energy storage devices [11, 12]. But at a large scale, mainly owing to investment costs, it does not appear to be feasible yet. Finally, the use of reserves is currently the most common way to palliate for the lack of wind generation. However, these reserves often come from conventional sources, implying some costs and additional emissions. It is thus crucial to optimally quantify the reserve needs to preserve the environmental benefits from the use of a clean energy source. First methodologies are developed today for reserve quantification taking into account the characteristics of wind [56, 151]. In countries with large-scale wind integration, it may also be required to define power exchanges through interconnections to compensate major changes in wind production.

Meanwhile, the development of wind energy in Europe takes place under particular circumstances characterized by a parallel process for the deregulation of electricity markets. This means that even if wind generation benefits of incentive measures today, in a few years it will have to be considered as equal to the other energies in the electricity markets. Actually, this implies that intermittent energy sources such as wind will be submitted to penalties for their imbalances [155]. This is already the case in the Scandinavian countries for

[2] cf. website `www.eoliennes.net`

instance, where wind power producers participate in the Nord Pool [164].

In this general context, forecasting of the wind power generation for the following hours ahead is actually considered as necessary, either for energy management or for trading [99]. Today, there is an increasing demand for operational forecasting tools, designed to meet the needs of end-users such as Independent Power Producers (IPPs), energy traders, Transmission System Operators (TSOs), and utilities.

1.2 Forecasting wind power

Predictions have been used by the human kind since the dawn of time. Today, they are essential in several areas of economy and industry, where they serve as a basis for making decisions and developing strategies. However, forecasts are of value only if they are specially tailored to the intended application. Forecasting methods must be developed in concert between users and analysts [42], in order to define the context and the objectives of their application.

The aim of short-term predictions of wind power output is to contribute to a secure and economic power system operation. Such predictions provide end-users with estimations of the future wind generation, usually for the next 24-72 hours, thus tackling the intermittent nature of wind that is feared by traditional energy actors. A crucial point is that wind power forecasting methods should be designed for operational use, for real-time application. Commonly, this real-time aspect is referred to as *online*, in opposition to *offline* when working on historic data for research purposes. Increasing the value of wind generation through the improvement of prediction systems' performance is one of the priorities in wind energy research needs for the coming years [220].

Forecasting wind generation is far from being a trivial problem. Roughly, it involves two stages: firstly, the 'meteorological' one, which consists in predicting wind at the level of the considered site for the next hours or days, and secondly the 'energy conversion' stage that involves the transformation of wind speed to power. Usually, the first operation is based on the refinement of Numerical Weather Predictions (abbreviated NWPs) that are provided on a grid around the wind farm and at various heights. This operation is referred to as statistical downscaling. The latter, in practice, corresponds to the modeling of the wind park power curve, which should take into account the individual turbine curves, the terrain characteristics, the shadowing effects inside the wind farm, and the meteorological parameters' influence (e.g. ambient temperature, air density, turbulence intensity) on the production.

Each one of the above steps in the forecasting procedure involves a modeling error that penalizes the accuracy of the final output, that is the wind power forecasts. At first sight, it seems that the most important of these aspect is the meteorological one, because it is easy to imagine that the better we predict wind, the better we will estimate the resulting electricity generatio from a wind farm. This has been confirmed by several studies that try to quantify the relative share of NWPs in the accuracy of power forecasts[3] [99, 129].

[3]Holttinen and Horvinen [99] mention that so far an accuracy of $\pm$2-3 m.s^{-1} in amplitude and of $\pm$3-4 hours

The meteorological forecasting errors are highly variable. These errors are then amplified or dampened by the energy conversion process modeling. Consequently, characterizing the uncertainty included in wind power forecasts appears to be a complex task. But, as forecasting errors are unavoidable, operational solutions have to be developed for helping end-users (i.e. utilities, transmission system operators or energy producers) to optimally take decisions related to wind power management (e.g. reserve estimation, power exchanges).

1.3 Estimating the uncertainty of wind power forecasts

Estimations of future wind generation are usually given as *point forecasts*: they provide a single value for a given lead time, which is 'the most likely outcome'. For instance, a wind power forecast would tell that a certain wind farm is expected to produce 250 kW the following day at noon. Actually, the probability that this 'event' (with such an accuracy) occurs is clearly close to zero: a point forecast is always subject to an error. The level of error may be acceptable for the forecast user, and in this case he will be content with that point forecast alone. But, if the risk of error is noteworthy and the expected consequences significant, it is then necessary to associate the forecast with an estimation of the expected uncertainty. Tennekes [219], interpreting the work of Popper (*The Open Universe*, 1982), states that "a unique deterministic prediction is not consistent with the 'scientific method', in the sense that the result from any scientific prediction is not complete without an estimate of the likely error associated with measurement and other experimental accuracy". It is expected that the additional information provided by uncertainty estimates can then be integrated in the decision-making process in order to optimize the benefits from the use of predictions.

Interval forecasts[4] are a kind of uncertainty estimate. They consist in a range of values within which the expected future value (of the variable of interest) is expected to lie, with a prescribed probability. Going back to the previous example, an interval forecast would be that the considered wind farm is expected to produce between 100 and 360 kW with a 90% probability the following day at noon.

Prediction intervals have only recently attracted attention in the statistical literature [46]. In practice, companies are quite reluctant to produce (or to exploit) interval forecasts in complement to (or instead of) point forecasts [41, 87]. A first reason for that is that intervals may be harder to interpret for a non-specialist. A typical concern of an end-user receiving an interval forecast would be the question of how to use it, how to make a decision from a range of values associated with a probability. Another concern, for both end-users and analysts, is about the performance of such type of forecasts. When it is rather easy to evaluate point predictions by comparing them with related measured values, the case of intervals is more delicate and requires the use of advanced skill scores and statistical tests [47, 81, 231]. Regarding in particular the wind power area, prediction intervals are a

in time was sufficient for wind speed forecasts applications though such a level of accuracy is not satisfactory for utilization in wind power prediction.

[4] Such forecasts are also referred to as *prediction intervals*. Both terms will be used in the document.

new topic that presents several challenges. The wind power production process, due to the particularities of the wind power curve and the weather predictability, makes that conventional approaches for estimating interval forecasts are not directly applicable. One of the objectives of the thesis is to provide solutions to that particular problem.

As explained above, meteorological forecast errors contribute to the final prediction error. Then, identifying meteorological situations that may lead to small (or alternatively large) NWP errors is a way to detect situations for which low (or high) wind power prediction errors may be expected. Ensemble forecasts of meteorological variables are a promising way to do so. They consist in a set of alternative predictions (scenarios) around the so-called control forecast, and aim at illustrating how the uncertainty in the initial conditions will grow through the meteorological forecasting model [182]. The use of the ensemble approach may be applied to the wind power forecasting problem by calculating wind power ensemble forecasts from ensemble NWPs. Then, it will be of particular interest to evaluate the uncertainty information that can be extracted from these alternative power production scenarios.

1.4 Purpose of the work

The purpose of this work is primarily to propose and describe original operational solutions for estimating the uncertainty of wind generation forecasts. This involves an exploratory analysis or characterization of the prediction errors, illustrating how the NWPs, the lead time, the power curve, and other factors influence these errors. Our choice is to propose uncertainty estimation methods with a general value, that is applicable to all the state-of-the-art point prediction methods. To illustrate this, we have selected five state-of-the-art methods based on different approaches. The errors of these methods are used for two purposes: firstly to investigate on the characteristics of wind power prediction errors, and secondly to evaluate the applicability of the developed uncertainty estimation techniques to these forecasting methods.

For estimating the uncertainty of wind power forecasts, we explore two complementary solutions:

prediction intervals **-** We concentrate on the design of interval forecasts of wind power production. When several prediction intervals are estimated at once, they permit to provide predictive distributions of wind power. Our contribution consists in developing an original statistical method that have an empirical nature and which is distribution-free (i.e. no restrictive assumption is made on the shape of prediction error distributions). The proposed method based on an advanced statistical treatment of the recent errors the prediction method made in similar conditions permits to evaluate the range of possible wind power production that can be expected in a near-future. The sole requirement for the application of this uncertainty estimation method to the wind power forecasting problem is the possibility to collect the errors online, which is possible only if a wind farm (or a group of wind farms) is equipped with a data acquisition system. This is the case for almost all new wind farms today (especially the large ones).

***prediction risk indices* -** Alternatively, ensemble forecasting is considered for assessing the predictability of wind generation, and thus for estimating the expected skill of the predictions provided by point prediction methods. A definition of prediction risk indices is proposed, and we study how they are related to the level of forecast uncertainty. Prediction risk indices may be a more comprehensive signal on expected uncertainty than interval forecasts in an operational context. They may be needed for making a choice among a set of alternative (and more or less conservative) decisions. The ensembles of wind power that are utilized here are derived from meteorological ensembles provided either by the European Center for Medium-range Weather Forecasts (ECMWF) or by the National Center for Environmental Prediction (NCEP) of the United States. In addition, a poor man's temporal approach consisting in making up ensemble predictions of wind power by associating forecasts for the same lead time, but issued at different time origins, is applied as a baseline approach.

As a second objective of the present thesis, we aim at demonstrating the value of the uncertainty information when using wind power forecasts in a decision-making process. We explained above that among the possible uses, forecasts may either be integrated in an energy management decision process or considered for bidding in an electricity market. In the present document, we focus on the latter possibility: we describe how the uncertainty estimates may be exploited for optimizing the benefits from the participation of wind power producers in European electricity markets.

1.5 Structure of the thesis

In order to describe the context in which the thesis has been initiated, Chapter 2 presents the state of the art concerning wind power forecasting. This literature survey describes the data and the main approaches that are currently in use for predicting wind power. The prediction methods that are considered in the following Chapters are described. Also, this Chapter justifies the motivations and objectives of the thesis by identifying the lack of uncertainty estimation methods in the state of the art.

The aim of Chapter 3 is to characterize and analyze the wind power forecasting errors, as a basis for the following developments. Initially, we introduce an evaluation framework consisting in a set of error measures and diagnostic tools, as well as the way they should be interpreted. In parallel, a distribution-oriented approach of forecast verification is described, the application of which will permit to better appraise the influence of some parameters on the characteristics of predictions error distributions. Then, we comment on the level of forecast uncertainty of the state-of-the-art wind power prediction methods. The contributions to the power prediction errors are investigated and we give a thorough description of their characteristics.

In Chapter 4 we introduce a new method for the estimation of prediction intervals, which is suitable for the case of nonlinear, nonstationary and bounded processes such as wind generation. The various hypotheses that led to an approach with an empirical nature and which is distribution-free are justified. Furthermore, we define what the required properties of prediction intervals are, and propose a method that integrates fuzzy logic concepts for integrating further expertise of the analyst in order to increase the resolution of produced interval forecasts. A non-parametric framework for the evaluation of interval (or alternatively quantile) predictions is proposed. It is then applied for assessing the statistical performance of the introduced method. A sensitivity analysis on its parameters follows. Finally, guidelines for an online application of the method are given.

Chapter 5 considers the possibility of estimating the uncertainty of wind power forecasts using prediction risk indices. Initially, it is explained how to produce wind power ensemble forecasts, either by post-processing meteorological ensemble predictions, or with a poor man's temporal approach. The added value provided by ensemble predictions is investigated, with focus to the contribution of their mean and spread. Then, we propose a definition of prediction risk indices, which reflect the spread of the ensembles over successive look-ahead times. On the case-study of a European wind farm, we show how these prediction risk indices may be related to the level of prediction error in a probabilistic manner, and how they may be used for forecasting the expected level of energy imbalance. 'Forecasting' uncertainty is thus a new concept introduced in this thesis for the area of wind power. A comparison of the results for the three types of ensemble predictions of wind generation is carried out.

In order to assess the applicability and benefits of the developed wind power forecasting methodology (i.e. integrating an uncertainty information), Chapter 6 concentrates on the value of forecasting in liberalized electricity markets. A generic formulation of the revenue of a wind power producer participating in a market is given. Then, we develop on decision-making approaches for trading, based on the use of probabilistic predictions of wind generation associated to a model of the participant's sensitivity to regulation costs. The related optimization problem for deriving optimal bids is given. For assessing the value of the uncertainty information, we evaluate the revenues of a wind power producer trading in the Dutch electricity pool over a one-year period. We compare what the benefits are when using point predictions to the alternative of using an advanced revenue-maximization strategy resulting from the introduced decision-making approach. The benefits obtained by applying the more advanced trading strategy are clearly shown.

Finally, Chapter 7 summarizes the overall conclusions from the present research works and gives perspectives for further research.

2

State of the Art in Wind Power Forecasting

Abstract

This Chapter presents the current status of methods for the short-term prediction of wind generation. We introduce here what the motivations for forecasting are, and explain how it appears as an essential feature for the economic and secure management of power systems integrating a significant share of wind power. Also, it is underlined why an uncertainty information is needed for optimizing the decision-making process resulting from the use of predictions. Though, this topic was not thoroughly addressed in the relevant literature at the beginning of the thesis.

2.1 Introduction

RESEARCH WORKS in the area of wind power forecasting have started in the eighties [152]. Since then, numerous research centers and companies have invested in the development of methods and operational tools, leading to a high number of prediction models. Few of them are commercially available today. These tools are characterized by the prediction horizon (few minutes, hours or days), the computational facilities (a small PC or a supercomputer), and the desired accuracy. Here, our aim is not to give a full survey on all existing models but more to describe the main approaches and their specificities. Also, one of the goals of the present Chapter is to explain why forecasting is considered today as necessary either for the management or for the trading of wind generation.

Prediction models can be sorted in different families. Simple methods based on climatology or averages of past production values may be considered as reference forecasting methods, since they are easy to implement. Such methods can be used to benchmark more advanced methods. Then, what we referred to as advanced methods can be divided into two groups, depending on the considerations that are made for going from the meteorological predictions to the expected wind power output. A first possibility is the so-called *physical approach* that focuses on the description of the wind flow around and inside the wind farm for proposing an estimation of the wind power output. Alternatively, the so-called *statistical approach* models the relation between a set of historical measurements and/or meteorological predictions, and the power output, without any assumption on the physical phenomena.

Whatever the considered approach, wind power prediction can be put back in its mathematical framework. We introduce hereafter the generic formulation of the forecasting problem and the basic concepts of forecasting from a mathematical point of view. The way this mathematical framework has been translated by the wind power forecasting community is consequently developed, by giving an overview of the main (and most employed in practice) prediction approaches.

Finally, this Chapter raises the question if the actual wind power forecasting methodologies are sufficient for designing optimal strategies for the management of wind generation. Indeed, the description of the main forecasting methods and of their characteristics allows us to identify the gaps in the research area and to introduce the goals of the present work. The conclusions will obviously point towards the need for associating an uncertainty information to power forecasts.

2.2 Describing the basis of the problem

Wind power cannot be managed like conventional power sources: its power production is not imposed by human intervention but by meteorological conditions. Because of its fatal nature, wind parks' generation is usually seen as a 'negative load' by utilities. They actually consider that the demand that should be met by conventional generation means is reduced by the proportion of wind power available on the electricity network [18].

In this Section, the specificities of the wind power production process are exposed. More particularly, our aim is in a first stage to outline the physical reasons for the variability of wind generation. This is done by discussing wind characteristics and the way wind is transformed to power. Then, by introducing the challenges wind power producers, grid operators, energy traders, etc. face due to wind variability, we explain why forecasting helps for the optimal integration of wind generation. Finally, the main features of wind power forecasting are described, i.e. the prediction of relevant meteorological variables at the level of the wind farm and the modeling of the energy conversion process.

2.2.1 The intermittent nature of wind generation

Wind speed constantly changes. It is hence necessary to use statistical tools for describing this nonstationary process. In order to model wind speed distributions, we generally use the Weibull distribution with two parameters [107], these two parameters being a scale factor A_w and a shape factor k_w. Such a kind of distribution is widely used for product lifetime analysis and reliability engineering. Its shape and properties have made it the most appropriate description of the wind speed behavior when studying potential sites for wind park installations. It allows engineers to optimize the turbines, to minimize generating costs as well as to evaluate economical returns. The probability density function of the Weibull distribution is given by:

$$f(u) = \frac{k_w}{A_w} \left(\frac{u}{A_w} \right)^{k_w - 1} \exp \left(- \left(\frac{u}{A_w} \right)^{k_w} \right), \tag{2.1}$$

where u is the wind speed.

The parameters A_w and k_w are estimated using a long period of wind speed data. Wind characteristics may significantly evolve from a year to another. Weibull distributions are right-skewed, reflecting the fact that strong winds are rare while moderate and fresh winds are more common. A description of the wind conditions in a large number of sites in Europe is gathered in the European Wind Atlas [225], in which these conditions are summarized with the two Weibull parameters given as a function of wind direction sectors at heights of around 10 meters above ground level (a.g.l.). In general, the scale parameter A_w takes values between 2 and 8, and the shape factor k_w between 1.5 and 2.2.

Wind generation is highly variable as it is directly related to wind speed: wind turbines convert the kinetic energy that is present in the wind into mechanical energy, consequently used for operating an electrical generator. The energy conversion process for a single wind turbine is described by its characteristic curve, also referred to as the *wind turbine power curve*. Characteristic curves have roughly the same shape whatever the manufacturer and the turbine type. Figure 2.1 gives a schematic example of such a curve. The power production is null below the cut-in wind speed (around 2-4 m.s^{-1}), then sharply augments between the cut-in and rated wind speeds (around 12-16 m.s^{-1}). At rated speed, it reaches a production level close to the nominal power P_n. The power production is almost constant between the rated and cut-off wind speeds (around 25-30 m.s^{-1}). At cut-off speed, the turbine stops for security reasons. There may be a difference between the maximum and nominal power values (up to 10-20%). The nominal power P_n serves as a reference value for denoting the capacity of a turbine and computing expected energy yields depending on site characteristics.

The sharp increase of the characteristic curve for low wind speed is explained by the cubic power law: the power p available in the free flowing stream of wind is a cubic function of the wind speed u:

$$p = \frac{1}{2} \rho_{air} A_r u^3, \tag{2.2}$$

where A_r is the rotor swept area exposed to the wind and ρ_{air} the air density.

Equation (2.2) only gives a theoretical expression of the available power. In practice, due

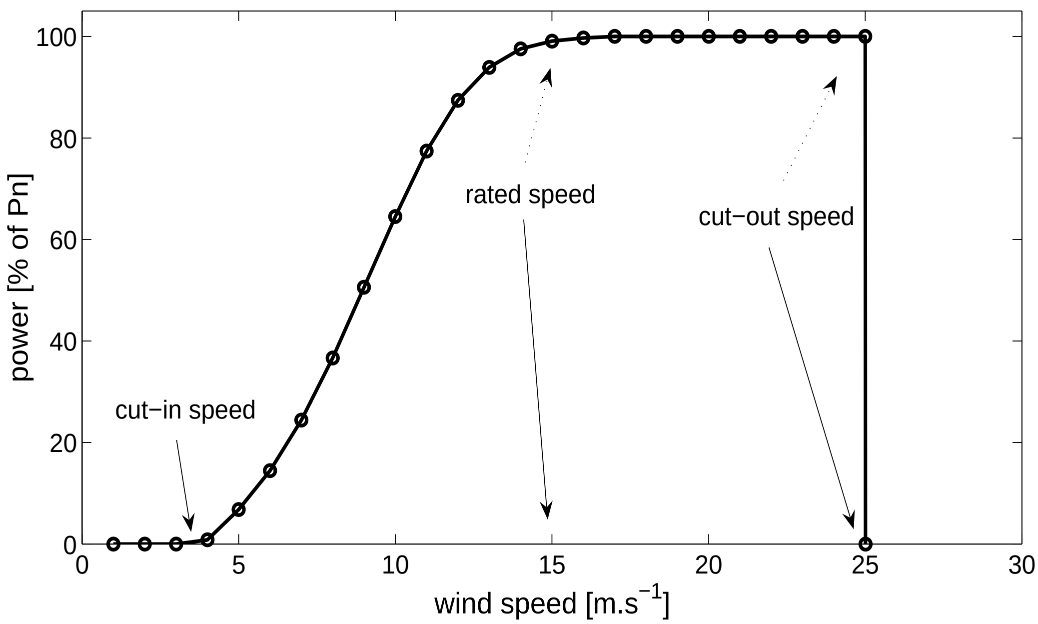


Figure 2.1: *Example of a typical wind turbine power curve. The power production is normalized with the turbine nominal power P_n, and is expressed as a function of wind speed.*

to a number of factors like the Betz limit[1], the generator and gearbox efficiencies as well as other losses, it is only possible to extract 20-30% of the original energy available in the wind:

$$p = \frac{1}{2}\rho_{air} C_p \eta_t A_r u^3, \tag{2.3}$$

with η_t the turbine efficiency ratio for the turbine's elements (up to 0.8) and C_p the performance coefficient for the wind turbine (related to aerodynamic considerations), which is bounded by the Betz limit. Such a coefficient is a function of both the blade pitch angle and the ratio of the rotor blade tip speed to wind speed. C_p is around 0.35 for a good turbine design.

Because low wind speeds are more common than strong breezes one understands that we are most of the times in the low and steep parts of the power curve. They correspond to a zone where a small wind speed variation induces a large variation of power production. The variability of wind speed leads to a highly dynamic behavior of wind generation. In fact, wind turbines do not often operate at their nominal power P_n, and the resulting energy yield is not given by P_n times the time of operation. Owing to that aspect, a wind farm *capacity factor* is defined as the ratio of actual energy output to the amount of energy the wind farm would produce if it operated continuously at full rated power, over a selected period. It can be seen as a performance ratio, which describes the profitability of a given project. For

[1] The Betz limit states that the maximum theoretical amount of energy that can be extracted from the wind is approximately 59%.

onshore projects, this number usually varies between 25 and 40%, while it goes from 45 to 60% for offshore projects. This is one of the reasons that justify the tendency for offshore developments. The capacity factor represents a mean production but does not describe the high variations around this mean. The day-to-day management of wind generation, due to these fluctuations, also greatly affects the likely income from a wind park project. Also, these variations induce a cost for maintaining the balance and the supply security of a power system. The widely used term for referring to the variable nature of the wind power production process is *intermittence*.

2.2.2 The various motivations for forecasting wind generation

Fluctuations of wind generation receive a great amount of attention. Intermittence can be regarded at various time scales. First, wind power production is subject to seasonal variations, i.e. it may be higher in winter in Northern Europe due to low-pressure meteorological systems or it may be higher in summer in the Mediterranean regions owing to strong summer breezes. There are also diurnal cycles, which may be substantial or not, mainly due to thermal effects. Finally, fluctuations are observed at the very short-term scale (at the minute or intra-minute scale). The variations are not of the same order for the three different time-scales. Managing the intermittence of wind generation is the key aspect associated to the integration of that renewable energy into electricity grids [103].

The challenges to face when wind generation is injected in a power system depends on the share of that renewable energy. It is a basic concept, the wind *penetration* (denoted by ζ), which allows one to describe the share of wind generation in the electricity mix of a given power system. Penetration can be defined in several ways. The first definition is the wind capacity penetration ζ_{cap} that corresponds to the percentage of installed wind capacities C_{wind} in the total system capacity C_{tot}:

$$\zeta_{cap} = \frac{C_{wind}}{C_{tot}}. \tag{2.4}$$

This type of penetration relates to the structure of the considered power system. Alternatively, the instantaneous penetration ζ_{inst}, which proves to be crucial when dealing with the stability of a power system, is the share of wind generation p_{wind} in the total available power p_{tot} (equal to the total electrical consumption plus losses) at a given moment t:

$$\zeta_{inst}(t) = \frac{p_{wind}(t)}{p_{tot}(t)}, \tag{2.5}$$

while a third manner is to define the average penetration ζ_{av} that is the ratio over a given time period $[t_i, t_f]$ between the energy E_{wind} generated from wind installations and the energy E_{tot} generated by the whole power system:

$$\zeta_{av}\left([t_i, t_f]\right) = \frac{E_{wind}([t_i, t_f])}{E_{tot}([t_i, t_f])}. \tag{2.6}$$

For illustrating these three definitions, let us consider the specific case of the Eltra system in Denmark. Eltra owns and operates the transmission system that covers mainland Jutland and the Island of Funen. At the end of 2004 this system had a wind capacity penetration ζ_{cap} of 33%, since there were 2315MW of installed wind farms for a total of 7018MW of installed generating capacities [27]. However, the average penetration ζ_{av} for 2002 was equal to 18%: from the 21TWh of annual consumption, 3.8GWh were met by the means of wind generation. Wind generation rapidly changes, thus the instantaneous penetration greatly varies as well. For the same example of the Eltra system, the instantaneous penetration even reached 100% during some hours of 2004.

The Transmission System Operator (TSO) is responsible for managing the electricity balance on the grid: at any time, electricity production has to match consumption. Therefore, the use of production means is scheduled in advance in order to respond to load profiles. The load corresponds to the total electricity consumption over the area of interest. Load profiles are usually given by load forecasts (produced from experience or by prediction methods), which are of high accuracy [31, 32]. The error is in the order of 1.5-2% of the level of forecast load for day-ahead predictions and does not significantly exceed 5% in the case of week-ahead predictions. Still, continuous efforts are made for improving the performance of load forecasting methods, since even a small reduction of the level of forecasting error will lead to substantial savings for utilities that manage large interconnected systems [94].

For making up the daily schedule, TSOs may consider their own power production means, if they have any, and/or they can purchase power generation from Independent Power Producers (IPPs) and utilities, via bilateral contracts or electricity pools. In the context of deregulation, more and more players appear on the market, thus breaking the traditional situation of vertically-integrated utilities with quasi local monopolies.

Two mechanisms compose electricity markets. The first one is the spot market where participants propose quantities of energy for the following day at a given production cost. An auction system permits to settle the electricity spot price for the various periods depending on the different bids. The second mechanism is the balancing of power generation, which is coordinated by the TSO. Depending on the energy lacks and surplus (e.g. due to power plant failures or to intermittence in the case of wind power installations), the TSO determines the penalties that will be paid by IPPs who missed in their obligations. In general, wind power producers are penalized by such market system since a great part of their production may be subject to penalties. To ease the development of wind power, certain countries have chosen to provide a guaranteed grid access for the electricity produced by wind farms, as well as to impose the purchase of all wind generation at a guaranteed price (also called feed-in tariff) [33], thus limiting the 'hazardous' nature of the expected revenues for wind farm owners. In addition, for encouraging wind power producers to participate in the market, some countries like Spain [72]) have chosen to add a premium to market clearing prices. This premium stands for the ecological advantages of that renewable energy. It appears obvious that an IPP cannot propose quantities of energy on the market without knowing what is going to be the output of his wind farms. While the IPP revenue is greatly enhanced by using advanced wind power forecasting approaches for bidding in the electricity market [227], it is preferable to also consider associated uncertainty estimates for defining optimal participation strategies [204]. Electricity markets and possible participation strategies for wind farm operators will be the central topic of Chapter 6.

Though wind capacity penetration in Europe is rapidly growing, the average penetration is still low. Wind power capacities produced only 2% of the generation necessary to meet electricity consumption for Europe in 2003. For some interconnected systems such as Denmark for instance or island systems such as the one of Crete, the average penetration ζ_{av} is already high, and this causes difficulties in their management. It is necessary to compensate for unexpected drops of wind production (e.g. in case of severe winds, for which turbines switch off for security reasons, or in case of unpredicted falls of wind speed) or sometimes to deal with 'surplus-production' situations.

In the first case, the challenge is to quantify the reserve needs for compensating the eventual lacks of wind production. Often, a high level of reserves is used, which diminishes the environmental benefits from the use of wind power [57] due to increased emissions. Studies have appeared in the literature on possible strategies for determining the way conventional power plants and wind production means may participate in the electricity mix, e.g. by Doherty and O'Malley [56] for the specific case of the Irish power system. The authors have studied the so-called 'fuel-saver' and 'forecasting' scenarios. The first one assumes that the conventional generation and required reserves are scheduled with no consideration given to the forecast wind production. When wind generation is present, the conventional generation is backed off in accordance with a merit order to lower operating costs. If wind generation attains a level such that no more conventional generation can be backed off, then the additional wind production is curtailed. In contrast to the first one, the 'forecasting' scenario uses wind power forecasts (and their associated uncertainty) for optimizing the quantities of both conventional generation and reserves. From analyzing the impact of these scenarios over a typical day, the authors show that integrating wind power predictions in the reserve management process significantly decreases CO_2, NO_x and SO_2 emissions from conventional plants. In addition, they show that the costs related to the management of power plants is lowered substantially. However, the way the uncertainty of wind power forecasts is modeled is basic: distributions of prediction errors are assumed to be Gaussian. This particular point is discussed in Chapter 3.

For compensating the impact owing to the intermittent behavior of wind generation, an alternative is to use energy storage devices. The association of storage with wind power plants is expected to augment the potential benefits of IPPs since it would allow them to decrease penalties on the balancing market. It permits to control the energy delivery, to store wind generation in low electricity-price periods (or during periods with high wind speeds) and then to sell the energy when market prices are high. Various methods have been proposed for the scheduling and operation of such wind-storage combined systems [12,121]. It is widely recognized that if such methods are based on advanced wind power forecasts, this would permit to optimize the resulting benefits. Though, the fact that large storage capacities may be needed makes that solution expensive and not conceivable in a near future [11].

However, hydro-storage facilities, in the form of pumped storage or hydro-reservoirs, already allow for large-scale electricity storage. In addition, they present fast response time and low operating costs. Thus, the wind-hydro combination appears as a promising option and is studied in the literature. For instance, Castronuovo and Peças Lopes [36] focus on the particular case of wind-hydro power plants for which they propose optimal management strategies for both improving the economic gains and reducing the power production intermittence. For developing these management strategies, wind power predictions and estimations of their associated uncertainty are a primary requirement [154, 206].

In the western part of Denmark, the production from wind occasionally exceeds consumption. This situation corresponds to the surplus-production problem we previously mentioned. Holttinen [97] states that “when wind power production exceeds the amount that can be safely absorbed while maintaining adequate reserve and dynamic control of the system, a part of wind energy produced may have to be curtailed”. Actually, studies have shown that such situations may occur when the average yearly penetration exceeds 10%. Also, when this average penetration is more than 20%, there may be up to 10% of the total wind power produced during the year that is lost [75]. This kind of situations may be even more problematic for the case of autonomous systems such the ones of islands. In general, if situations of very high instantaneous wind penetration can be predicted, this is expected to help for exporting the surplus production or for curtailing excess wind generation in order to ensure system security [143]. Interconnection with neighboring grid systems is an option for contributing to the management of wind power intermittence.

It is remarked that wind variability is also dependent on the spatial resolution one looks at. Distributing wind capacities over a large area (eventually with different wind regimes) helps in reducing the variability of the overall power production. Fluctuations of a single turbine generation are greater than those of a group of turbines, because these variations average out. This effect is also valid if we look at the power output from a single wind farm and the power output over a region. This effect is known as the smoothing effect of geographical dispersion. Giebel [75] investigated on the benefits of distributing the wind power capacities over Europe. He showed that having a global view of wind generation would allow more fuel savings, and would also give more capacity credit to wind power[2]. In addition, for the Nordic countries area, Holttinen [97] studied the variations of wind generation. It was shown that for this area, the hourly step variations are very smooth: about 98% of the times they lie between $\pm 5\%$ of installed capacity. Consequently, the consideration of wind generation over a group of wind farms or over a region is a way to reduce the variability problem. This contrasts with the idea of concentrating large-scale wind installations in isolated areas (or offshore). In such cases, fluctuations will remain of high magnitude and the status of the transmission system may become a critical aspect (e.g. bottleneck problems) [210].

[2] Despites the multiple definitions given in the literature, the capacity credit can be roughly defined as the capacity of conventional power plants that can be replaced by wind installations, without a loss of reliability.

2.2.3 The main aspects of the forecasting problem

In the literature, focus is given to the prediction of power output instead of energy. This is because early prediction methods were physical ones based on the direct use of a characteristic curve for the wind park to convert wind speed predictions to power. In the present document we also focus on power forecasts. If the energy information is needed, it can be easily obtained by integrating power values over the considered time interval.

Forecasting of the wind power generation may be considered at different time scales [78]:

- *from milliseconds up to a few minutes*, forecasts can be used for the turbine active control. Such a type of forecasts are usually referred to as *very short-term* forecasts.
- *for the following 48-72 hours*, forecasts are needed for the power system management or energy trading. They may serve for deciding on the use of conventional power plants (unit commitment) and for the optimization of the scheduling of these plants (economic dispatch). Regarding the trading application, bids are usually required during the morning of day d for day $d+1$ from midnight to midnight. These forecasts are called *short-term* forecasts.
- *for longer time scales* (up to 5-7 days ahead), they may be considered for planning the maintenance of wind farms, or conventional power plants or transmission lines. For the specific case of offshore wind farms maintenance costs may be prohibitive, and thus an optimal planning of maintenance operations is of particular importance.

In the next Paragraphs, we describe what the main aspects of the short-term wind power forecasting chain are: it typically gathers the meteorological aspect, i.e. the prediction of meteorological variables at the level of the wind farm, and the energy conversion process modeling, which tells what the wind farm power output is, given the wind. This model chain is strictly followed by the physical prediction methods, while the statistical ones are usually direct procedures: they directly go from NWPs and historical values of wind power production to forecasts of expected power output.

Prediction of weather variables at the level of the wind farm

Wind power generation is directly linked to weather conditions and thus the first aspect of wind power forecasting is the prediction of future values of the necessary weather variables at the level of the wind farm. This is done by using numerical weather prediction models. Such models are based on equations governing the motions and forces affecting motion of fluids. From the knowledge of the actual state of the atmosphere, the system of equations allows to estimate what will be the evolution of state variables such as temperature, velocity, humidity and pressure at a series of grid points. The first successful numerical forecasts were made by Charney et al. [40] in the early fifties, following the demonstration by Rossby et al. [195] that the linearized perturbation of the equation of motion could be used for weather prediction. For a thorough introduction to NWP models, we refer to [109].

The assessment of the system initial conditions of a NWP model is based on the retrieval of a large number of measurements from meteorological synoptic stations, weather balloons and radiosondes. Other types of observing systems may contribute to this data retrieval process, like satellites, ships and aircrafts for instance (which are not fixed in space in contrast to meteorological stations). A data assimilation procedure is then used for correcting the data and interpolating their values in areas where only few measurements are available (e.g. over Africa). The quality of the data that serve as initial conditions for a NWP model is paramount: without a good estimation of these initial conditions, the resulting forecasts from the model will be of poor accuracy.

Mathematically formulating the atmosphere evolution yields a system of nonlinear partial differential equations. Such large systems do not have analytical solutions though they can be resolved by means of numerical tools. The equations are solved both in time and in space: time derivatives are typically replaced by finite differences and spatial derivatives are numerically represented by finite-difference schemes or spectral methods [28]. The distance between the grid points is called the *spatial resolution* of the NWPs. The mesh typically has spacing that varies between few kilometers and up to 50 kilometers for mesoscale models. It is not possible to model the phenomena that are active at scales smaller than this resolution. This is the case of local thermal effects and turbulent mixing for instance. The unresolved processes are accounted for by parameterization at every grid point. Parameterization is certainly the most sensible and controversial area of weather modeling [28, 96].

Regarding the time axis, the forecast length of most of the operational models today is between 48 and 172 hours ahead, which is in adequacy with the applications we consider here. In general, models with higher forecast lengths have coarser grids. In the sixties, Lorenz [139, 140] used the chaos theory to show that the weather predictability had a finite limit of about two weeks. Indeed, even with perfect observations and perfect models, the chaotic nature of the atmosphere makes it impossible to predict the evolution of meteorological variables further than two weeks ahead. This is a theoretical limit that has no relevance for the horizons we consider here. But, both predictability and the chaotic nature of the atmosphere will be the central topic of a further chapter.

The temporal resolution is usually between 1 and 3 hours. NWP models impose their temporal resolution to power forecasting methods since they are used as a direct input. Thus, the short-term power prediction methods that are based on NWP input only have a temporal resolution of the same order. They all provide an estimate of the average power output over the considered period. Their aim is not to describe what are the power fluctuations inside the time interval. For that purpose, stochastic methods can be used to describe the power output behavior depending on wind speed fluctuations and turbulence intensity[3] [4, 190].

When considering the wind power forecasting application, the spatial resolution of the currently used NWP models is not enough for modeling all the local effects at the level of the wind farm. Therefore, a necessary step is to extrapolate the meteorological predictions to

[3]Periodic power variations due to the tower shadow are negligible in comparison with the effect of turbulence intensity [62].

the wind park level, taking into account the fine-scale processes such as thermally-induced breezes, terrain effects (e.g. winds accelerating over hills), and the influence of existing obstacles. This is the so-called *downscaling* procedure. For doing so, simple methodologies that consider only terrain effects and more advanced ones based on Computational Fluid Dynamics (CFD) have been proposed. They will be developed in Section 2.5.

The downscaling process also involves scaling the wind speed to hub height. Hence, this implies the modeling of the wind speed profile, which describes the variations of the mean wind speed $\overline{u}$ as a function of the height z above ground level. The most classical model is the logarithmic wind profile:

$$\overline{u}(z) = \frac{u_*}{\kappa} \ln\left(\frac{z}{z_0}\right), \qquad z \geq z_0, \tag{2.7}$$

where κ is the von Kármán constant, u_* the friction velocity and z_0 the roughness length. z_0 is related to surface roughness. Its magnitude varies depending on the terrain type (from 0.0002 for sea surface to 1 for areas covered with large obstacles). The von Kármán constant characterizes the dimensionless wind shear for statically neutral conditions. A value of 0.04 is commonly assumed. The friction velocity u_* is a reference wind velocity, i.e. the characteristic velocity scale of the flow in the surface boundary layer.

Actually, by writing Equation (2.7) for two different heights z_1 and z_2, and after a simple manipulation, a relation between the mean wind speeds at these two heights is obtained:

$$\overline{u}(z_2) = \frac{\ln(z_2/z_0)}{\ln(z_1/z_0)}\, \overline{u}(z_1). \tag{2.8}$$

The interest of this logarithmic modeling is hence that wind speed can be easily scaled to the hub height level, by multiplying the known wind speed by a factor that is a function of these two heights and of the local roughness length. Even if this modeling is widely applied and accepted, its use should be restricted to near neutral conditions, to flat terrains, and to wind modeling over land. In complex terrain situations, the logarithmic wind profile provides poor agreement with observations, and more advanced models based on computational fluid dynamics should be preferred [239]. In parallel, for offshore conditions the logarithmic wind profile based on the local surface roughness only is not sufficient for describing the wind shear. It appears that thermal effects should also be included [128].

Modeling the energy conversion process

The wind turbine characteristic curve has been introduced in Paragraph 2.2.1. However, this power curve must be seen as theoretical: it is given by manufacturers, and obtained from experiments in wind tunnels. When turbines are gathered in a wind farm, a single wind turbine characteristic curve cannot serve to model the energy conversion process. It is therefore necessary to propose solutions for estimating a *wind farm power curve* that encompass all the aspects that are not taken into account when estimating the theoretical power curve of a single turbine.

The first aspect is that a wind farm may be composed by several types of turbines, with different power curves. A simple sum of the various characteristic curves may be envisaged as a first solution. Also, both the choice of the site for a wind park and the arrangement of the turbines are the result of multiple studies for optimizing the power yields. It starts with the determination of the prevailing wind directions and the corresponding wind intensities from wind roses and Weibull distributions. Then, since a wind turbine uses the energy in the wind for generating electricity, it appears obvious that the wind leaving the turbine has a lower energy content. This effect is called *wake* or *shadowing effect*. It corresponds to the idea that behind the turbine, there is a zone with turbulent and lower wind, similar to the wake behind a boat in movement. Simple formulae are available for quantifying the decrease in wind speed as a function of the rotor size and some other parameters [119]. Wakes significantly influence wind farm power curves.

The shadowing effects are a primary concern when choosing the siting of turbines in a given wind farm. Turbines are usually separated by a distance of 5 times (or more) the rotor diameter in the prevailing wind direction and by a distance of 3 times (or more) the rotor diameter in the direction perpendicular to the prevailing winds. This is a first rule of thumb. Indeed, the cost of connecting the turbines to the grid is a restriction for spacing them too far ones from the others. Also, the terrain characteristics may impose the way turbines will be positioned (e.g. on a crest).

More advanced methods are available for investigating wake effects inside a wind farm [17, 51, 230]. They allow one to better arrange turbines than with the first rule of thumb introduced above. Also, if wind farms are gathered in clusters (i.e. groupings of turbines with a certain distance between them), a large distance may be needed for wind to recover its original energy. The use of turbine clusters is mainly envisaged for the case of very large offshore wind farms, for which the energy yield of the project have to be optimized. Thus, studies are in progress for assessing the so-called 'far-wake' effects that turbine clusters induce [10, 230].

For any wind farm layout, the function that gives power output depending on wind speed will not be the same whatever the wind direction, owing to these shadowing effects. Hence, a wind farm power curve cannot be envisaged as a sole function of wind speed, but also as a function of wind direction. There may also be some other parameters that should be considered as variables when estimating wind farm power curves, such as air density[4] or turbulence intensity.

The environment of a given wind farm greatly evolves during its lifetime. The maintenance or decommissioning of the turbines, or wind park extensions, are factors that significantly impact a wind farm power curve. Other factors, such as the ageing of the turbines, or changes in the surroundings of the park (e.g. vegetation depending on the season), are expected to affect the power curve, but with less effects. Therefore, the estimation of wind farm power curves must take into account these effects. A power curve model should be a function of time and be adapted with all the newly-available data.

In [34], Cabezon et al. describe several methods for estimating wind farm power curves.

[4]The power available is a linear function of air density ρ_{air}, cf. Equations (2.2) and (2.3).

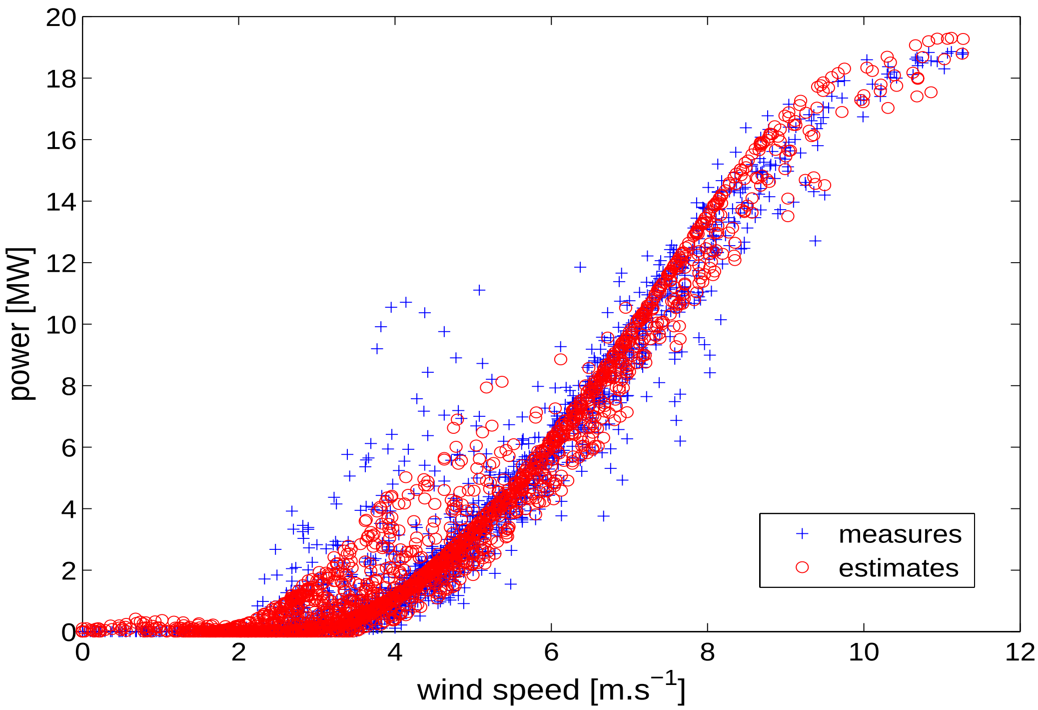


Figure 2.2: *Estimated wind farm power curve for a 21MW wind farm in Denmark. Results are shown over a 2000-hour evaluation period. The power curve model is set-up by using hourly wind data (speed and direction) from a meteorological mast and related power measures from a SCADA system over a 2000-hour training period. It is based on Least-Square Support Vector Regression.*

The wind farm of the study is located in a very complex terrain, for which the wind flow characteristics and the wind farm layout makes the characteristic curve estimation rather difficult. In that study, wind speed measurements are given by a meteorological mast at the level of the wind farm and by nacelle anemometers (i.e. anemometers on each turbine). Wind power measurements are retrieved thanks to Supervisory Control And Data Acquisition (SCADA) systems for every turbine. The various proposed methods range from a direct modeling of the wind farm power curve to the modeling of each turbine power curve from detailed measurements. In addition, more advanced methods are introduced, such as cluster analysis for modeling the power curves of groups of turbines with a similar behavior, and fuzzy logic as a nonlinear modeling alternative. The authors show that advanced statistical methods are of better accuracy (estimation error is reduced by almost 30% in this study), and also that detailed measurements permit to greatly improve the wind farm power curve estimation.

Figure 2.2 gives the example of an estimated power curve for a 21MW wind farm in Denmark, in flat terrain conditions. This estimation is based on a Least-Square Support Vector Regression (LS-SVR) method that allows nonlinear modeling by learning from examples. 2000 hourly-averaged values of wind speed and direction data (from a meteorological mast) and power measurements are utilized for tuning the model's parameters. The Figure compares estimated power production values against measures over an evaluation period of 2000 hours.

One sees from that Figure that the wind farm power output is significantly variable for a given wind speed. In this example the variation is up to 4MW (approximately 20% of the installed capacity) in the low and steep part of the power curve. Power estimates also exhibit substantial variations for a given wind speed thanks to the consideration of wind direction by the power curve model. Nowadays, most of the advanced wind power forecasting tools do not base their power curve modeling on wind speed only [76].

2.3 Generic formulation of the wind power forecasting problem

Forecasting is the art of telling what will happen in the future from the knowledge we have of the current and past situations. This means that a forecaster focuses on the evolution in time of a variable of interest p. Such a variable is sampled in time (usually with a constant time interval) and the evolution of p is then represented by a discrete *time-series* $\{p_t,\ t \in \mathbb{T}\}$ (where $\mathbb{T}$ is the ensemble of time indexes and is usually included in $\mathbb{Z}^+$). Values of $\{p_t\}_{t\in\mathbb{T}}$ may be instantaneous values or averages over the past time interval.

In the case of wind power forecasting, the variable of interest p is the available power output for a wind turbine, a wind farm, or for an area with several wind farms. When power forecasts are used for power system management or trading, power values are usually sampled with an hourly time resolution.

Wind power time-series $\{p_t\}_{t\in\mathbb{T}}$ can be characterized as *nonstationary* and *nonlinear* time-series. Nonstationarity of a time-series means that its behavior, which may be described by its moments (mean, standard deviation, etc.), evolves with time. Nonlinearity stands for the fact the time-series exhibits features that cannot be explained by linear models. Both of these properties can be deduced from expertise and from a visual inspection of wind power time-series. In our case, nonstationarity comes from the very nature of the wind. Also, nonstationarity stands for the temporal evolution of the wind farm environment, as was mentioned in Paragraph 2.2.3. The nonlinear and bounded nature of the time-series is mainly due to the energy conversion process: even if wind speed had a linear behavior, it would be turned into a nonlinear one when converted to power due to the shape of wind farm power curves.

The forecast of the p-value made at time t for lead time $t+k$, is denoted here by $\hat{p}_{t+k/t}$ — the notation $\hat{p}_t(k)$ can also be found in the forecasting literature. With respect to k, the terms *forecast horizon* or *look-ahead time* are used interchangeably. $\hat{p}_{t+k/t}$ is the most likely outcome one can estimate given the knowledge available at time t. This knowledge, constituted by the available data at that time is often called the *information set*.

Forecasting methods are procedures that permit to provide predictions from the knowledge one has of the past and the present. They may be based on simple algorithmic rules, on the judgment of experts on the phenomenon of interest, or on a given model that has been identified after an analysis of the available data. Here, we will follow the distinction expressed by Chatfield [41] between the terms 'model' and 'method': the model only describes the temporal evolution of the time-series — a model being literally a mathematical representation of reality — while the method encompasses the whole procedure for producing predictions. Indeed, time-series modeling and forecasting are two different branches of time-series analysis. Forecasting methods may not be based on a particular time-series model.

$\hat{p}_{t+k/t}$ is a *point forecast* in the sense that it is a single value. This is to be opposed to *probabilistic forecasts* for which events or ranges of possible outcomes are associated to a probability. They include density forecasts or prediction intervals, which will be the central topic of a following Chapter. Actually, even if prediction specialists usually reason in a probabilistic way, they traditionally provide deterministic forecasts in the form of point forecasts, because they have been found easier to interpret or to integrate in a decision-making process.

Whatever the type of forecast, it is essential to realize that time-series forecasting is a form of extrapolation: a prediction model is fitted on a set of data and is consequently used outside of that range of data. Forecasts are then conditional, and should be formulated in the following way: "given the information set and assuming that the identified behaviors continue in the future, we can predict that...". This means that a forecaster makes the crucial assumption that the future will be like the past. That is the first aspect of forecast uncertainty. Always feeding a forecasting method with the most recent available information is of particular importance.

In the statistical literature, the distinction is made between *univariate* models that consider only past values of p and *multivariate* models that use not only past values of that variable, but also past or present values of other variables. These additional variables are called *explanatory* variables since it is expected that their variation may explain the variation of p. In the present work, we will consider that multivariate models may also use forecast values of explanatory variables. For predicting wind generation, explanatory variables are mainly wind speed and direction, but they can also be air density, temperature, humidity, etc.

An univariate model states that p_t can be expressed as the sum of a function g_u of the past l values and of a random shock e_t:

$$p_t = g_u(p_{t-1}, p_{t-2}, \ldots, p_{t-l}) + e_t, \tag{2.9}$$

where $\{e_t\}$ is a purely random process, which can be defined as a sequence of uncorrelated and identically distributed random variables with zero mean and constant variance σ_e^2. This process is named *white noise* or innovation process in the forecasting literature. It should be noted that $\{e_t\}$ may be a sequence of independent variables, which is a stronger assumption than the sole uncorrelation. However, it is often assumed that the white noise is a Gaussian process, and in this case uncorrelation and independence are equivalent.

Alternatively to univariate models, multivariate ones express p_t as a function of past values of p, past values of a set of explanatory variables $\mathbf{x}$, as well as forecast values of these explanatory variables, plus a random shock e_t:

$$p_t = g_m(p_{t-1}, p_{t-2}, \dots, p_{t-l}, \mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \dots, \mathbf{x}_{t-m}, \hat{\mathbf{x}}_{t/t-1}) + e_t. \tag{2.10}$$

For both types of models, a lag can be introduced, i.e. models may be used to describe the relation between these same inputs and p_{t+k}. Also, in addition to a dependence of p_t with the other variables, it may be useful to include a trend (linear or not) and seasonal variations (seasonal standing here for periodic).

A forecasting method can be generally formulated as some function of the available data:

$$\hat{p}_{t+k/t} = f(p_t, p_{t-1}, \dots, p_{t-l}, \mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-m}, \hat{\mathbf{x}}_{t+k/t}) = f(\Phi_t). \tag{2.11}$$

In the following of the document we will denote by Φ_t the information set. Depending on the considered forecasting method (univariate or multivariate), Φ_t may or may not contain values of explanatory variables.

The work of the analyst is then to propose a function f that best describes the considered process. If a wealth of explanatory data is available, a first task is then to identify the ones that are really meaningful.

In a forecasting exercise, the dataset is usually defined as a time-series of N measurements of the variable of interest p, associated or not with a set of explanatory variables. In this document we refer to that situation as the *offline* case, as opposed to the *online* one for real-life forecasting. When working offline, the dataset is divided into a *learning* and a *test* set (with respectively N_L and N_T elements), which are independent. The learning set is used for fitting the model, while the test set is used for evaluating the performance of the forecasting method. Thus, in the test set, predictions must be produced in a way that mimics the real forecasting situation, for which information is available up to time t for predicting at time $t+k$. Predictions made in the learning phase are called *in-sample* forecasts and the ones produced over the test set are *out-of-sample* or *ex-ante* forecasts.

In the following Sections, we give an overview of the main methodologies for predicting wind power generation.

2.4 The reference forecasting methods

It is worthwhile to develop and implement an advanced wind power forecasting tool if it is able to beat reference methods, which are the result of simple considerations and not of modeling efforts. Probably the most common reference method used in the frame of wind power prediction, or in the meteorological field in general, is *persistence*[5]. This naive predictor — commonly referred to as 'what you see is what you get' — states that the future

[5] Such a forecasting method is known as the 'random walk' in the time-series forecasting literature.

wind generation will be the same as the last measured value, i.e.

$$\hat{p}^{\mathrm{P}}_{t+k/t} = p_t. \tag{2.12}$$

Hence, by definition the error for zero time step ahead is zero. Despite its apparent simplicity, this method might be hard to beat for the first look-ahead times (say up to 4-6 hours). This is due to the scale of changes in the atmosphere, which are actually slow, in the order of days (this is true for the case of Europe). It takes about one or three days for a low-pressure system to cross the continent. Since the pressure systems are the driving force for the wind, the rest of the atmosphere has time scales of that order. High-pressure systems can be even more stationary, but they are not associated with high winds and so not really interesting for wind power prediction.

A generalization of the persistence method is to replace the last measured value by the average of the last n measured values:

$$\hat{p}^{\mathrm{MA},n}_{t+k/t} = \frac{1}{n}\sum_{i=0}^{n-1} p_{t-i}. \tag{2.13}$$

Such types of methods are sometimes referred to as moving average predictors. Asymptotically (as n goes to infinity), they tend to the global average (also known as the *mean climatology*)

$$\hat{p}^{0}_{t+k/t} = \overline{p_t}. \tag{2.14}$$

where $\overline{p_t}$ is the average of all the available observations of wind power up to time t.

This last one can also be seen as a reference method, but since it is not dynamic, its performance may be very poor for the first prediction horizons. However, for further look-ahead times, its skill is far better than the one of persistence. The performance of these two reference methods has been analytically studied by Nielsen et al. [175], and it has been shown that for longer time horizons, the climatological method is twice as better as persistence. Consequently, the authors proposed to merge the two methods in order to get the best of their performance over the whole range of prediction horizons. The merging yields a new reference method:

$$\hat{p}^{\mathrm{NR}}_{t+k/t} \; = \; \rho_k p_t + (1-\rho_k)\overline{p_t}, \tag{2.15}$$

where ρ_k is the correlation coefficient between p_t and p_{t+k}.

The drawback of this new reference method is that the ρ_k coefficients have to be estimated or fixed by using some considerations or assumptions. This is therefore in disagreement with the definition we gave of the reference methods, and this is probably why this method is not really used as a reference in practice by the wind power forecasting community.

2.5 The physical approaches

2.5.1 The physical methodology

Meteorological forecasts are given at specific nodes of a grid covering an area. Since wind farms are not situated on these nodes, it is then needed to extrapolate these forecasts at the desired location. The main idea of the physical approach is to refine the NWPs to determine the wind field around the wind park and therefore at hub height, by using physical considerations about the terrain such as the roughness, orography and obstacles, and by making assumptions on the wind profile if needed. The two alternatives to do so are: *(i)* to combine the modeling of the wind profile (with a logarithmic assumption in most of the cases) and the geostrophic drag law for obtaining surface winds [124]; *(ii)* to use a CFD code that allows one to accurately compute the wind field that the farm will see, considering a full description of the terrain.

When the wind at the level of the wind farm and at hub height is known, the second step consists in converting wind speed to power. Usually, that task is carried out with theoretical power curves. However, since several studies have shown the interest of using empirically derived power curve instead of theoretical ones [34] (cf. Paragraph 2.2.3), theoretical power curves are less and less considered.

When applying a physical methodology, the modeling of the function which gives the wind generation from NWPs at given locations around the wind farm is done once for all. Then, the estimated transfer function is consequently applied to the available weather predictions at a given moment. In order to account for systematic forecasting errors that may be due to the NWP model or to their modeling approach, physical modelers often integrate Model Output Statistics (MOS) for post-processing power forecasts.

2.5.2 Overview of physical methods

CFD models are already fully tested state-of-the-art models. They can run for several hours (e.g. 7-8 hours for the LOCALS model [63]). LOCALS, standing for Local Circulation Assessment and Prediction System, serves to predict wind speed on a very fine spatial resolution (500m or 3km), taking into account localized conditions such as variations in the terrain, surface roughness, etc. Therefore, very detailed maps are needed to describe the terrain around wind farms. This approach aims at catching the turbulent behavior of wind due to the interruption of wind streams by the topographic features and frequent changes in surface roughness. The proposed approach has been tested on the case-study of a Japanese wind park in a complex terrain area. The authors found that better results where obtained when considering the spatial distribution of wind turbines in the park. Though, they witnessed a bias of their prediction method. This reveals the need of using MOS techniques for correcting physical-type forecasts. For that purpose, the authors proposed to implement Kalman filtering.

Some people directly use commercial codes like STAR-CD to solve local effects on wind fields [146]. The related model is a three-dimensional Navier-Stokes numerical solver with a $k - \epsilon$ closure. Advantages of using such a CFD model is that areas of interest, i.e. the surroundings of the turbines, can have a rather dense grid while areas of lower interest can be meshed with a coarser grid. In each grid cell three-dimensional calculations are performed for the three components of wind, as well as for pressure, turbulence, temperature, etc. The works carried out by Magnusson et Wern [146] comprised a first study to show the possibilities to use CFD codes for accurate wind speed forecasting and hence wind energy prediction. However, very few results are available for practical case studies. The aim of the authors has mainly been to demonstrate the importance of accounting for fine-scale effects when carrying out physical modeling for wind speed and power prediction.

The use of these CFD codes is mainly associated to complex terrain situations, either to carry out wind power forecasting or wind resource assessment. Indeed, they are of great interest mostly for the latter case, because they permit to assess the particularities of wind flows over cliffs for instance, when classical methods for wind resource assessment express their limits [73, 179].

A US-American society, TrueWind Inc., developed one of the first commercially available wind power forecasting software: *eWind* [6]. This tool runs a numerical weather model called *ForeWind*, which can produce accurate near-surface wind forecasts on a fine grid, using boundary conditions from a regional weather prediction model. More physical processes are captured and the prediction can be better tailored to the local site. *ForeWind* uses a subset of the equations and modules contained in the Mesoscale Atmospheric Simulation System (MASS), which is a state-of-the-art mesoscale weather model. Adaptive statistics are used to correct systematic errors in wind forecasts. This adaptive statistical corrections are updated in real-time through online wind and plant output data collected at the site of interest. In a follow-up paper [241], Zack presents an alternate approach that is based on a high resolution physics-based atmospheric model operating in a rapid-update cycle mode. This model aims at assimilating all the available data (forecasts and observations) in the vicinity of the wind farm for better describing the meteorological variables state in the area of interest. It is expected that this approach will also be able to assimilate satellite-sensored data in a near future.

The *Prediktor* method developed at Risø laboratory in Denmark is another kind of physical method that does not deal with CFD but with the WAsP methodology (which was designed first for wind resource evaluation, cf. European Wind Atlas [225]) to convert the forecast wind from NWPs into the wind seen by the wind park [125, 127]. This is an example of the simple approach based on the utilization of geostrophic winds we described above. It is pointed by Landberg [124] that even if simple linear function can be used for approximating the geostrophic drag law, one should be very careful concerning the turning of geostrophic winds. Consequently to wind speed predictions at the level of the wind farm, a park model permits to account for shadowing effects and wind power is calculated by using theoretical power curves. By post-processing *Prediktor* forecasts with MOS techniques, the related prediction error is significantly reduced. Evaluation results from the application of the *Prediktor* methodology are available for the case-study of set of Irish wind farms [233].

Beyer et al. [16] developed a similar forecasting procedure with NWPs resulting from the Deustchlandmodell as input (which as a temporal resolution of one hour and a spatial grid of approximately 14x14km^2). One of the conclusions of the authors is that measured power data should be further exploited for estimating empirical power curves. They actually noticed that theoretical power curves often drastically differ from empirical ones. Consequently, the proposed prediction methodology has been further exploited with the Lokalmodell of the German weather service. It has led to the *Previento* method [69], which includes a wind profile modeling that accounts for thermal stratification. Another particularity of this method is that it offers the possibility to predict at either the local scale (for a single wind farm) or at a more global level, i.e. for the whole Germany in this case. The prediction method is associated with an assessment of the risk of relying on produced wind power forecasts [70]. Moreover, Focken et al. have described the smoothing effect that is witnessed when forecasting the aggregated power of wind farms [68]. They actually looked at the cross-correlation of forecasting errors as a function of the distance between wind parks for the specific case of Germany. Recently, Lange et al. [132] have examined the possibility to adapt *Previento* for offshore. Their main concern was about the modeling of offshore wind profiles, which cannot be seen as logarithmic in neutral atmospheric conditions. The authors hence proposed an alternative approach that involves the inertial coupling of the Ekman layers of the atmosphere and sea via a wave-boundary layer with constant shear stress. This alternative modeling has been evaluated and validated for various locations at the North Sea [214].

Mörhlen et al. [157] checked a very simple approach that converts directly wind speed forecasts to power using theoretical power curves. But, it is done inside the HIRLAM NWP module so that one can use physical properties that are not available as output, such as direction-dependent roughness, actual density and stratification of the atmospheric boundary layer, etc. This new approach is called *HIRPOM*, standing for HIRlam POwer prediction Model [105]. In different runs with horizontal model resolutions of 30km, 15km, 5km and 1.4km for two months in the beginning of 2001, the most common statistical accuracy measures[6] (MAE, RMSE, correlation, etc.) did improve only slightly with higher resolution. However, peak wind speeds were closer to the measured values for the high-resolution forecasts. For these higher resolution forecasts, the best model layers were closer to the ground than in the coarser models. Regarding the errors, Mörhlen et al. pointed out that phase errors (the timing of the frontal system) has a much larger influence on the error scores (and eventual payments) than level errors. We will develop on that result when studying the NWP contribution to the power prediction error in Paragraph 3.5.1. After investigations on the case-studies of several wind farms in Ireland [160], Mörhlen et al. stated that all their results (about systematic model errors and phase front errors) point towards the use of ensemble forecasting for both better predicting wind generation and also for estimating the related uncertainty.

Note that complex terrain situations do not mandatory imply the use of CFD codes. For instance Marti et al. [149] applied a statistical approach (a conditional semi-parametric

[6]For a detailed description of the statistical error measures that may be considered for the verification of wind power forecasts, we refer to Paragraph 3.2.3.

model that we will describe in the next Section) with HIRLAM forecasts as input for a wind farm in La Muela (Spain), and obtained good results. In order to estimate contribution of both the input NWPs and of the statistical modeling approach in the quality of the results, two spatial resolutions (0.2° and 0.5°) and different sets of HIRLAM variables were used to predict wind speed and energy production. The authors concluded on the importance of the spatial resolution of NWPs for complex terrain in the accuracy of HIRLAM forecasts and therefore in the quality of the resulting power predictions.

CFD codes are obviously good tools for studying the motion of a fluid like air. Though, they are most of the times too computationally expensive. In addition, we have explained that physical prediction methods need to be post-processed with statistical methods for erasing the systematic part of prediction errors. For these two reasons, statistical forecasting methods may comprise an interesting alternative.

2.6 The statistical approaches

2.6.1 The statistical methodology

Statistical prediction methods are based on one or several models that try to establish the relation between historical values of power, as well as historical and forecast values of explanatory variables, and wind power measures. This mapping is carried out over the training set. The physical phenomena are not decomposed and accounted for, even if expertise of the problem is appreciated for choosing the right explanatory variables and designing suitable models. Consequently, the trained models are used over the testing set (for a given forecasting exercise), or for real-life forecasting situations. Often, auto-adaptation methods are envisaged for tuning the model parameters during real-life operation in order to reach an optimal performance.

All statistical methods are based on the fitting of a particular model to the data used for training, for defining the f-function of Equation (2.11). The fitting consists in an optimization problem, where the *loss function* to be minimized is a direct function of the forecast error. Let e denote a forecasting error[7] and $\mathbf{e} = \{e_t, t = 1, \ldots, N_L\}$ the sequence of prediction errors over the training set. The loss function $L(e)$ tells what is the 'regret' for a forecast user, associated to the error e. Such a function is typically continuous, has the property that $L(0) = 0$ and is usually symmetric (which means that the forecast user has the same sensitivity for positive or negative prediction errors). The most common loss functions that a statistical modeler aims at minimizing over the training set are the absolute error loss function

$$L_a(\mathbf{e}) = \sum_{t=0}^{N_L} c_1 |e_t|, \tag{2.16}$$

[7] The forecast error will be further defined and described in Chapter 3.

and the quadratic loss function

$$L_q(\mathbf{e}) = \sum_{t=0}^{N_L} c_2 e_t^2, \tag{2.17}$$

c_1 and c_2 being constants. Actually, it may be difficult to define a loss function that really reflects the context-specific sensitivity of forecast users to prediction errors. This is why generic loss functions such as the ones given by Equations (2.16) and (2.17) are mostly chosen in practice.

Considering a quadratic loss function translates to the forecast $\hat{p}_{t+k/t}$ being an estimate of the conditional expectation of P_{t+k} (the random variable of potential power outcomes), given the chosen model f and the information set Φ_t[8]. In contrast, considering an absolute error loss function leads to forecasts that correspond to estimates of the median of P_{t+k}-distributions, conditional to both the considered model and the information set.

Wind power forecasts are in most of the cases multi-step-ahead forecasts. When applying a statistical prediction methodology two alternatives may be envisaged. On the one hand, one may design models for each step-ahead. Then, a 48 step-ahead forecasting exercise would imply the design of 48 specific models. On the other hand, one may design a unique model for one-step ahead prediction, and use this model in an iterative manner. This means that the one-step-ahead forecast is fed back to replace the lag-one value as input in order to produce the two-step-ahead forecast, and so on until we reach the forecast length. The advantage of the latter solution is then that there is only one model to configure and to train. However, errors are cumulated from one-step-ahead to another, since the model is fed with estimations instead of measured values. If aiming at an optimal performance the former alternative should be preferred.

Besides the fact that there may be linear or nonlinear models, we would like to point here the difference between *structural* and *black-box* types of statistical models. Structural models try to use the analyst's expertise on the phenomenon of interest while black-box models require little subject-matter knowledge and are constructed from data in a fairly mechanical way. Concerning wind power forecasting, structural models would be those that include a modeling of the diurnal wind speed variations, or an explicit function of meteorological variable predictions. Black-box models include most of the artificial-intelligence-based models such as Neural-Networks (NNs) and Support Vector Machines (SVMs). However, some models are 'in-between' the two extremes of being completely black-box or structural. This is the case of expert systems, which learn from experience (from a dataset), and for which prior knowledge can be injected. We then talk about *grey-box* modeling.

Statistical models are usually composed by an autoregressive part, for seizing the persistent behavior of the wind, and by a 'meteorological' part, which consists in the nonlinear transformation of meteorological variable forecasts. The use of NWPs as input in plus of

[8]This follows from the fact that the conditional expectation is the 'best' forecast in the sense of minimizing the square error (see [41], ch. 4). For a mathematical proof regarding univariate models, we refer to Clements [47] (ch. 2) for one-step ahead predictions and to Clements and Hendry [48] (ch. 2) for k>1.

past production values appears obvious today, because it is impossible to consider power predictions up to two or three days ahead without such weather forecasts. Pure time-series models including the sole autoregressive part exhibit good performance for horizons up to 6-10 hours ahead.

Finally, it should be noted that as long as relevant data are available, statistical models may be envisaged for focusing on downscaling task only or for modeling the sole wind-to-power conversion process.

2.6.2 Overview of statistical methods

Statistical models such as ARMA, ARX and Box-Jenkins forecasting methods in general have been widely used for time-series prediction, either for finance, control, or other applications [148]. They have been applied for short-term wind generation forecasting up to few hours ahead [71]. Since these methods are based on past production data only they cannot have a good performance for further forecasting horizons. It was found that different models should be built depending on the season or even depending on the month because their level of performance was varying substantially between winter and summer for instance [71].

What may have been the first works on the use of statistical approaches for predicting wind power generation are those carried out by Brown et al. [24]. The authors focused in a primary stage on wind speed forecasting, accounting for several basic features of wind speed data, like autocorrelation, non-Gaussian distribution and even diurnal nonstationarity. Wind speed data were transformed following a clever remark by Dubey [58], which is that for shape parameters k_w close to 3.6, the Weibull distribution is similar in shape to a Normal distribution. By applying an autoregressive process to these transformed data, the authors produced hourly forecasts (up to 24-hours ahead) of wind speed that were then converted to power by passing them through a turbine power curve. Unfortunately, they only described their methodology but did not carry out an evaluation of their results due to a lack of relevant data. Another interesting point about that early paper is that Brown et al. already envisaged the uncertainty assessment aspect of wind power forecasting by proposing methods for estimating interval and probability forecasts.

It seems that the most ‘established’ statistical forecasting method is the one developed at the Informatics and Mathematical Modeling department of the Technical University of Denmark, *WPPT*, which is in operation in the control rooms at ELSAM and Eltra since 1994. *WPPT* (which stands for Wind Power Prediction Tool) is a conditional semi-parametric model that has been regularly improved during the last decade. It is now part of the *Zephyr* tool with the *Prediktor* model developed at Risø [79]. The aim is to merge these two models to obtain a synergy between the physical and the statistical approaches.

The version of *WPPT* described in [176] integrates information from NWPs and uses semi-parametric estimates of wind direction dependent power curves in the transformation of forecast (by HIRLAM) wind speed and wind direction to power. Indeed, power forecasts are the result of a two-stage procedure. In a first stage, a simple power curve model, which takes into account both predicted wind speed $\hat{u}_{t+k/t}$ and wind direction $\hat{\theta}_{t+k/t}$ gives a first estimation of expected generation:

$$\hat{p}^{pc}_{t+k/t} \;=\; g(\hat{u}_{t+k/t}, \hat{\theta}_{t+k/t}, k). \tag{2.18}$$

Then, in a second stage, the first estimation $\hat{p}^{pc}_{t+k/t}$ is fed into a second model that accounts for past production values and the hour of the day for which the forecast is made. Such a consideration allows one to model possible diurnal effects that are not always captured by NWPs. This second model can be written as

$$\begin{aligned} \hat{p}^{adv}_{t+k/t} \;=\;& a(\hat{\theta}_{t+k/t}, k)p_t \;+\; b(\hat{\theta}_{t+k/t}, k)\hat{p}^{pc}_{t+k/t} \\ &+ c^c(\hat{\theta}_{t+k/t}, k)\cos\left(\frac{2\pi h^{24}_{t+k/t}}{24}\right) \;+\; c^s(\hat{\theta}_{t+k/t}, k)\sin\left(\frac{2\pi h^{24}_{t+k/t}}{24}\right), \end{aligned} \tag{2.19}$$

where a, b, c^c, and c^s are smooth time-varying functions to be estimated either by non-recursive process or by a recursive one respectively for offline and online uses. An exponential forgetting factor permits a continuous adaptation of these model parameters.

WPPT has also been adapted for producing regional forecasts [177]. The region is divided in sub-areas (following some heuristics). Then, the forecasting methodology is utilized for estimating wind generation at some reference wind farms. The power outputs for the reference wind farms are summed and upscaled for obtaining a first estimation of the sub-area wind generation. In parallel, a direct forecast of the sub-area power output is made from mesoscale meteorological predictions. In each branch, the sub-area estimates are summed, thus yielding to two different power forecasts for the region. The final power prediction is a weighted average of the forecasts from the two model branches.

Sanchez proposed a statistical approach [199] that consists in a dynamic combination of several prediction models ranging from reference models to conditional non-parametric models similar to the one described above. The author introduces a competition between the models, and depending on their recent performance, an combination procedure yield a weighted average of the best models' estimates. This prediction method is called *Sipreolico* and is operated by the Spanish utility REE (Red Electrica de España).

In contrast to classical statistical methods, some research teams have preferred the so-called NN-based models. They are trained over a long collection of production data using specific algorithms e.g. back propagation learning. Hence, they map the relation between the selected inputs and the wind power output. NNs are attractive owing to their ability to model nonlinear time-series, to their flexibility and finally due to the possibility one has to easily try very different architectures and learning schedules [110, 113]. Such methods have been widely used for the purpose of prediction. However, in their literature review on the use of NNs for forecasting, Zhang et al. [243] point that NNs should be carefully designed depending on the intended application. Another interesting point with NNs and artificial-intelligence based models in general is that they can learn the relation between any kind of input and a given output. This may be a danger for non-specialists who may imagine they can model everything from a given dataset [41]. Though, if used in combination with a subject-matter expertise, such a type of black-box models may allow one to model nonlinear behaviors for which classical linear models are known to fail.

As an example of the use of subject-matter expertise, Alexiadis et al. [2] (and later Damousis et al. [53]) noticed there was a strong correlation between winds at an upwind site and winds at the site of interest. Hence, the authors designed a network (and an alternative fuzzy-logic-based model) which is trained to give wind speed at the site of interest from wind speed and direction values from two upwind sites. Good results were obtained, and the advantage of this method is that it proved to have a good ability in forecasting oncoming fronts of steep increasing wind speed. Considering upwind data permits to produce accurate forecasts for short horizons (up to 2-3 hours ahead), but they are not sufficient for producing predictions for longer time-scales. This methodology has been considered more recently by Larson [135], with the aim of improving the ability of statistical models for the first prediction hours: upwind measurements are used in combination with wind farm measurements for better seizing fast-coming fronts. Another drawback of this methodology is that good results can only be obtained if specific wind regimes are present in the considered area. Also, for collecting upstream measurements, it is needed to install and maintain meteorological towers. This may be associated with a large cost [241].

In [137], Li et al. applied NNs for estimating power generation of a single turbine, from measured wind speeds and directions available for two meteorological masts located nearby. Their was not to predict wind generation but only to estimate that quantity in real-time in case that no data acquisition system is installed[9]. This approach outperformed traditional methods based on a conversion of the wind speed at the meteorological mast to power through theoretical (or empirical) power curves. The authors showed the general ability of neural networks for the downscaling problem as well as for power curve modeling. Alternatively, artificial-intelligence based method can be implemented for monitoring the wind generation for a group of wind farms or even for a region. This aspect is of major importance when wind integration reaches a large-scale status. In Germany for instance, country that has the largest installed wind capacity of the world (around 18GW today), it is not conceivable to have data acquisition systems for every wind farms. Thus, ISET (standing for Institut für Solare Energieversorgungstechnik) has designed an algorithm for giving online estimations of the total wind generation from data collected at few reference wind farms. The procedure described by Ernst et al. [65] uses data from 16 wind farms only (corresponding to a 425MW capacity), which are passed through an appropriate neural-network for providing current power output of about 3.2GW from all wind parks distributed over the utility supply area (around 5000 wind turbines). This extrapolation algorithm can also be used for predicting wind generation [64].

In parallel, a fuzzy-logic based approach has been developed. Expert systems are rather similar to artificial NNs, and were designed first for control purposes [232]. They appeared to be suitable for a large range of problems, and it turned out they were easily applicable for making predictions. The resulting forecasting methods have been tested on both the

[9]The real-time estimation of a given variable state when no measurements are available is usually referred to as *nowcasting*.

load forecasting problem and the wind power forecasting one [114]. These fuzzy-inference systems using 'if-then' rules are trained, like neural networks, over a long period of known production data. A simulated annealing algorithm controls the learning process and cross-validation is applied to terminate learning. An interesting advantage of using the expert system or NN-based approach with NWPs is the flexibility/adaptability of the models one can build. Outputs from NWPs are used as input parameters, as well as historical values of wind power (and eventually of some explanatory variables). An architecture optimization algorithm can then be applied to determine the contribution of all that variables in the model [110,112]. Therefore, this algorithm also carries out input selection. A final version of the forecasting method developed by Kariniotakis has led to *AWPPS* (standing for Armines Wind Power Prediction System), which is in operation in several countries Europe-wide. Evaluation results from this prediction platform were made available in 2003 for onshore sites [115]. This model has also been evaluated recently for offshore conditions [187]. Also, the flexibility of this approach made it possible to develop various strategies for regional forecasting that may be envisaged depending on the available data or on the required prediction accuracy at various time-scales [188].

Sfetsos [202,203] made a comparison between several models of the statistical and artificial-intelligence based approaches for the problem of wind speed forecasting, giving a brief description of the model design methodologies and of their performance compared to the one of persistence. From his research works, the author concluded on a superior ability of artificial-intelligence based methods against more traditional methods for online applications. ANFIS was one of the algorithm he implemented and evaluated, with encouraging results. ANFIS (Adaptive Network Based Fuzzy inference System) is a fuzzy inference system implemented in the framework of adaptive networks [194]. This dynamic algorithm permits an online evolution of the inputs when implemented for the prediction of chaotic time-series. DENFIS [118], standing for Dynamic Evolving Neural Fuzzy Inference System is another kind of dynamic algorithm: it is designed for adaptive online learning and time-series prediction. Fuzzy rules can evolve (insertion or removal) during the learning process and the online/offline utilization. Both of these inference systems are promising methods for application to wind power forecasting in an online environment.

Recently, another type of artificial-intelligence based forecasting methods has been described in the literature [134, 162], which involves the so-called Support Vector Machines (SVMs). Such models are less rigid in terms of architecture than NNs and may allow one to save even more modeling efforts when designing prediction models. While other statistical models are estimated following the empirical risk minimization principle, i.e. the minimization of loss function over the learning set and the checking of the generalization ability with some criteria, the SVM theory is based on the structural risk minimization principle, which consists on directly minimizing an upper bound on the generalization error, and thus on future points. For further reading, we refer to [52, 228]. There is today a large interest in applying SVMs for several purposes including forecasting. Though, even if they have many appealing features, many issues remained to be solved before they become indispensable tools [15] (one should remember here that these models are pure black-box models, which thus need to be handled with care, cf. discussion in previous Paragraph). Some literature is available for the case of load forecasting [240].

2.7 Conclusions

Wind power forecasting is today a wide research area that uses methods from weather forecasting, fluid mechanics, statistics, etc. Since the first attempt at it, a lot of research efforts have permitted to produce operational prediction tools that greatly enhance the position of wind energy in the landscape of alternative energy sources and ease its integration into the electrical grid.

The main concern about wind generation is its intermittent behavior. This yields new challenges for utilities that have the duty of grid operation. Indeed, situations for which wind generation suddenly drops or exceeds the load to satisfy have to be solved in the most technically and economically feasible way. Options that are envisaged for managing the wind intermittence are the use of conventional power plant providing operational and capacity reserves, electricity storage, a strong interconnection with neighboring grid systems and curtailment in extreme cases. Whatever the chosen alternative, the choice being made by each grid operator depending on their available means, forecasting of the wind power output appears necessary for defining management strategies. However, for optimizing these strategies, we have seen that an uncertainty information associated to each forecast is of particular importance. Some recent research works have developed on first methodologies for defining optimal management strategies by integrating an uncertainty information [57, 67, 154]. The same kind of comment stands for the case of wind farm owners aiming at participating in European electricity pools: point forecasts are necessary for proposing energy bids, but prediction uncertainty estimates are mandatory if one wants to maximize its expected revenues [13]. Such an information is not given by the actual wind power prediction methodologies, or assumptions on the uncertainty characteristics are too basic.

By introducing the mathematical framework related to forecasting, we have developed on the main characteristics of point forecasting, which consists in estimating the most likely outcome for a given horizon. This vision of forecasting is actually shared by 95% of the available wind power forecasting methods so far. Even if the providing of a single estimate for each look-ahead time may appear easier to handle for non-specialists, that sole information cannot permit to optimize a decision-making process. Consequently, research efforts have to be directed towards the development of uncertainty estimation methods specially dedicated to wind power forecasting. This will be the central topic of the following Chapters, with emphasis on the description of the uncertainty sources and their characteristics, on the development of specific uncertainty estimation methods, and finally on the demonstration of the resulting benefits from the integration of the uncertainty information in the decision-making process.

3

Characterizing the Uncertainty of Wind Power Predictions

Abstract

This Chapter aims at characterizing the uncertainty of wind power forecasts. For that purpose, the framework for the verification of wind power point predictions is introduced. A set of evaluation measures and diagnostic tools is given, which allows one to assess the inherent quality of prediction methods, to carry out comparisons among rival approaches, or alternatively to gain insight on uncertainty characteristics. The evaluation framework is utilized on a set of test cases consisting of five state-of-the-art methods, applied for forecasting the power generation of four European wind farms, characterized by various terrain conditions and wind climatologies. In a first stage, the general level of prediction uncertainty is described, by drawing conclusions on the performance of current point forecasting approaches. Then, focus is given to highlighting the characteristics of the prediction uncertainty, by studying the influence of some variables like the look-ahead time or the level of predicted power on the moments of forecast error distributions.

3.1 Introduction

FORECASTS always contain a part of error, they cannot be exact. Though, depending on the application, the sensitivity to their errors differs greatly. In the frame of wind power forecasting, predictions errors translate in general to economic losses for end-users, and to the management of energy imbalances for the specific case of grid operators. Evaluating forecasts is a crucial part of a forecasting exercise, not only for developing a critical view on the performance of the chosen approaches, but also for having a deeper insight on what characterizes the prediction uncertainty.

First, it is necessary to define what are a good and a bad forecast, and also what makes a forecasting method better than the other. What defines the skill of forecasts is still subject to debate between forecasters and forecast users, as well as among forecasters themselves. Actually, Murphy [165] identified three different types of goodness in weather forecasting applications. Firstly, predictions should correspond to forecasters' true beliefs, and should not be issued for maximizing a utility criterion defined by the forecast user — this type of goodness is referred to as *consistency*. Then, he defines as *quality* the correspondence between predictions and related observations. The quality assessment thus stands for the statistical evaluation of forecasts' performance out of the context in which they are produced or consequently integrated in a decision-making process. In this Chapter, the question of consistency is left aside since in our case forecasters' judgments only influence the way they develop their prediction approaches. This was indeed the particular focus of Chapter 2. Finally, the third aspect of forecast goodness (as defined by Murphy) lies in its *value*. The concept of value is tightly linked to the application for which a forecast is needed. It represents the resulting increased benefits (economical or not) from the use of a forecast in a decision-making process. Again, this type of goodness is not treated in this Chapter, but will instead be dealt with in Chapter 6 in which we will concentrate on the trading application.

Evaluating point forecasts may be seen as rather trivial since they can be directly compared with measurements. However, a wealth of evaluation criteria are available, which need to be applied and combined the right way if one wants to draw relevant conclusions. Here, the framework for the verification of wind power forecasts is initially described, by introducing both the measure-oriented and the distribution-oriented approaches. These sets of criteria are utilized for studying the offline performance of several state-of-the-art forecasting methods (following various modeling approaches). Then, our aim is to highlight the main contributions to and characteristics of the wind power prediction errors, with focus to the effect of the forecast horizon, as well as the role of the power curve. For that purpose, we will underline the influence of these variables on the moments of prediction error distributions. By doing so, we will better appraise what makes the uncertainty in wind power predictions, in order to propose appropriate methods for its online estimation.

3.2 Defining and measuring forecast accuracy

In this Section we develop the framework for the evaluation of the forecast accuracy of wind power prediction methods. This results in an evaluation protocol for measuring and commenting on that accuracy. Finally, the methodology based on a distribution-oriented approach for underlining prediction error characteristics is introduced.

3.2.1 The prediction error

In the field of time-series prediction in general, the *prediction error* is defined as the difference between the measured and the predicted value. Since we consider separately each forecast horizon, the prediction error related to the point prediction $\hat{p}_{t+k/t}$ is defined as

$$e_{t+k/t} := p_{t+k} - \hat{p}_{t+k/t}. \tag{3.1}$$

Very often it is convenient to introduce the *normalized prediction error*

$$\epsilon_{t+k/t} := \frac{1}{P_n}\big(p_{t+k} - \hat{p}_{t+k/t}\big), \tag{3.2}$$

where P_n is the installed capacity. Normalizing errors allows one to produce results that are comparable from one wind farm to another, independently of their rated capacity. Normalization by $P_n$contrasts with the possibility to provide the error as a percentage of measured (or predicted) power. In the case of wind power this is not obviously feasible since a measured power value may equal zero. Alternatively, if the prediction error is evaluated over a long period of time, it is then possible to normalize the prediction error by the average measured power production. This mode of normalization allows one to better assess the cost of the method errors as a function of the capacity factor of the wind farm. Both normalization modes are relevant to the intended applications: either on the market or for the management of energy imbalances, the cost of prediction errors is a function of the amount of power (and also of energy) in surplus or lack, and not a function of a relative error with respect to the level of predicted (or measured) power.

Any prediction error $e_{t+k/t}$ can be decomposed into a systematic error $\mu^e_{t+k/t}$ and a random error $\xi^e_{t+k/t}$,

$$e_{t+k/t} = \mu^e_{t+k/t} + \xi^e_{t+k/t}, \tag{3.3}$$

where $\mu^e_{t+k/t}$ is a constant and $\xi^e_{t+k/t}$ is a zero mean random variable. Ideally, when dealing with a perfect predictor, the systematic error should be equal to zero, and the random part should be a *white noise*. A white noise is usually seen as a sequence of independent random errors that can be modeled with a zero mean Gaussian distribution (cf. Central Limit Theorem). However, in practice this is seldom the case: consecutive forecasting errors are often correlated [170]. Also, they may not be normally distributed. The error correlation issue will not be treated in this Chapter, but we will thoroughly comment on the non-Gaussianity of error distributions.

3.2.2 Evaluation framework

Evaluating a set of forecasts implies that related observations are available. Today, most of the new wind farms are equipped with SCADA systems that regularly collect information on the power output (and sometimes meteorological variables e.g. wind speed). However, there may exist periods for which data have not been made available. Another concern is about the quality of these data because even if they are available, they may be of poor quality. If forecast verification is carried out against some figures that are not the truth, it is unlikely that the resulting conclusions (and maybe the tuning of the involved prediction models) would be of any relevance. Hence, before starting any evaluation process, it is of particular importance to check the dataset. Several degrees of quality may be defined: a given observation can be missing, it can be obviously non-reliable (e.g. outside of the range of possible values), it can be 'suspicious' (e.g. equal to the previous measure, but not equal to 0 or P_n), or it may appear to be correct. Then, the analyst decides on how to deal with the dataset depending on its overall quality. Sometimes, datasets are so large that some automatic checking procedures are preferred to the analyst verification. Whatever the procedure, it is likely that for these big datasets some erroneous observations remain when evaluating the related forecasts. Were the erroneous measures randomly and sparsely distributed that it would have only few consequences on the evaluation. Though, if non-reliable or suspicious data are present in a more systematic way (like regular or long periods over the evaluation set), it is expected that they will bias the performance assessment by significantly influencing the value of the various evaluation criteria. Finally, it is desirable for the evaluation set to be composed by power measures that span the whole range of possible values. A long evaluation period should also be preferred, since it would then include periods with various external conditions that may contribute to the level of performance of a given method.

It should be clear that the same dataset cannot be used both for developing (and optimizing) a forecasting approach and for making its evaluation. Such an approach would lead to over-optimistic conclusions on the performance of the prediction method. Training error consistently decreases with model complexity, typically dropping to zero if complexity is large enough (such a behavior is called *overfitting*). The forecasting community often refers to that 'abnormal' level of performance as *artificial skill* [41], since this same skill could never be reached in an operational context in which forecasts can only have a genuine nature. Therefore, the dataset must be divided into a *learning* and a *test* set, which are independent. The latter is considered for assessing the performance of the forecasting method only. When setting-up a forecasting competition between rival forecasting approaches, the bounds of these two periods are clearly specified. The ability of a method to perform well when it is applied for predicting new and independent data is defined as its *generalization* ability. Maximizing the generalization ability of statistical forecasting approaches has been a very active field of research in the past decades, with the introduction of generalization criteria, the use of cross-validation [209, 205] or resampling techniques, or alternatively with the definition of a new framework proposed by Vapnik [228], which is based on the structural risk minimization principle (in contrast with the empirical risk minimization one that is widely considered today). We will not go into details on the generalization aspect even though it is a captivating subject, which appears to be crucial for ensuring a good level of performance when applying forecasting methods in an operational context.

3.2.3 Definition of an appropriate evaluation protocol

The central question of *'how to determine what makes a forecast (or a forecasting method) better than the others'* still remains. The quality evaluation of prediction methods directly follows from the comparison of forecasts with related observations. Then, by going further than the binary "true"-or-"false" statement that is not suitable for the verification of continuous variable forecasts, one appraises the quality of a prediction by analyzing its deviation from the truth. A single forecast is not enough for deriving conclusions on a given forecasting method quality. It is indeed necessary to consider evaluation periods consisting of long series of predictions to assess the forecast quality in term of its *statistical performance*.

This Paragraph gathers several approaches and criteria for evaluating and comparing various methods' performance, from the statistical and the meteorological literature. Obviously, end-users would prefer that a single statistic or diagram could give the whole information. That does not appear to be possible, even if there are some attempts to gather several aspects of a method's performance in a single feature (see [218] for instance). Here, we focus on the criteria that are the most relevant (and hopefully the most employed by wind power forecasters) when evaluating wind power prediction methods. This set of criteria comprises an evaluation protocol that will serve as a basis for the remaining of the Chapter.

Measure-oriented approach for evaluation of wind power prediction methods

A method's *bias*, which corresponds to an estimate of the systematic error μ_k^e, is given by the mean error over the whole evaluation period and is computed for each horizon

$$\text{bias}(k) = \hat{\mu}_k^e = \frac{1}{N_T} \sum_{t=1}^{N_T} e_{t+k/t}, \tag{3.4}$$

where N_T is the number of predictions used for method evaluation.

The bias is a basic aspect of the forecasting method performance. It tells if the method tends to over- or under-estimate the forecast variable. When calculated over the whole test set, it only gives a primary information. Usually for wind power forecasts, the bias is kept at a very low level thanks to the use of statistical procedures (even physical methods use MOS techniques today for removing the bias). But, if that bias value is computed for various subsets with different weather conditions, or various levels of the predictand value, it allows one to carry out further analysis by spotting conditions for which the method produces predictions that are significantly overestimated for instance.

By only giving that primary information on a forecasting method's trend to under- or over-predict, the bias cannot tell what is the actual skill of the forecasting approach: it is very unlikely that a forecasting method with a null bias over an evaluation set provides perfect predictions. In this case, the bias cancels out only because positive and negative error values compensate over the test set.

Then, a common measure that reveals the contribution of both positive and negative errors to the lack of accuracy is the *Mean Square Error* (MSE), which is the average of the squared errors over the test set

$$\text{MSE}(k) = \frac{1}{N_T} \sum_{t=1}^{N_T} (e_{t+k/t})^2, \tag{3.5}$$

or its square root form (abbreviated RMSE for *Root Mean Square Error*):

$$\mathrm{RMSE}(k) \;=\; \mathrm{MSE}(k)^{\frac{1}{2}} = \left[\frac{1}{N_T}\sum_{t=1}^{N_T}(e_{t+k/t})^2\right]^{\frac{1}{2}}. \tag{3.6}$$

The RMSE is easier to interpret since it has the same unit as the predictand. It should be noted that the error measures we introduce here do not depend on the size of the test set.

The other common measure of forecast accuracy is the *Mean Absolute Error* (MAE), which is the average of the errors in their absolute values

$$\mathrm{MAE}(k) \;=\; \frac{1}{N_T}\sum_{t=1}^{N_T}|e_{t+k/t}|. \tag{3.7}$$

Actually the choice of the RMSE or the MAE as a main evaluation criterion depends on the sensitivity of the end-users to the errors. That sensitivity is represented by a loss function (also referred to as a cost function). For more details on loss functions' properties and interpretations, we refer to [41] (Section 4.1). The use of the RMSE implies the consideration of a quadratic loss function while the use of the MAE implies the consideration of a linear one. If methods are evaluated with the RMSE, this actually means that one expects the forecasting method to have been trained for producing minimum MSE forecasts. When the cost function related to the sensitivity of forecast users is not clearly defined, it appears preferable not to rely on the RMSE only, but to report both error measures.

It is worth mentioning that an alternative to the use of the RMSE is to consider the *Standard Deviation of the Errors* (SDE):

$$\mathrm{SDE}(k) \;=\; \hat{\sigma}^e_k \;=\; \left[\frac{1}{N_T-1}\sum_{t=1}^{N_T}(e_{t+k/t}-\hat{\mu}^e_k)^2\right]^{\frac{1}{2}}. \tag{3.8}$$

The SDE criterion is a measure that deals with the random part of the error $\xi^e_{t+k/t}$ (in complement with the bias that is an estimate of the systematic part of the error). In contrast, both systematic and random errors contribute to the MAE and RMSE values.

Statistically, the values of the bias and of the MAE are associated to the first order moment of prediction error distributions, and hence they are measures which are directly related to the produced power. The values of the RMSE and of the SDE are associated to the second order moment, and hence to the variance of the prediction error. They do not have a direct interpretation. Also, for these two measures, large prediction errors have the largest effect. For this reason, the RMSE is more sensitive to the presence of erroneous data in the evaluation set than the MAE. This latter measure, by being more robust, should be preferred as a main criterion if one is suspicious about the quality of the evaluation set. Otherwise, one may conclude on a weak accuracy of the evaluated prediction method when the high witnessed RMSE values are due to the poor quality of measured data.

All the error measures introduced above can be calculated using the prediction error

$e_{t+k/t}$ or the normalized prediction error $\epsilon_{t+k/t}$. The interest of using normalized error measures is to produce results independent of wind farm sizes. The resulting error measures are then referred to as Normalized bias (Nbias), Normalized MAE (abbreviated NMAE), etc.

Note that forecasting errors are *nonstationary*. Tong [221] stated that for the case of nonlinear models "how well we can predict depends on where we are" and that there are "windows of opportunity for substantial reduction in prediction errors". Owing to the inherent characteristics of nonlinear processes, there are periods or certain conditions for which it is easier to predict than for others. Also, some models may be better than others for representing given nonlinearities of the considered process. Carrying out a unique evaluation over the whole test set may lead to misleading conclusions and prevent from really seizing the advantages and drawbacks of a given method. Hence, it is useful to evaluate error measures for periods characterized by specific conditions e.g. high wind speed, summer, westerly wind.

Comparing the accuracy of several forecasting methods

When evaluating rival forecasting approaches, it might not be clear what is the best forecasting method, since a method can be the best under some criterion, but not under alternative one [211]. A given method may have a lower bias but an higher RMSE for instance. In addition, the performance of various methods may depend on the evaluation period: for the case of wind power forecasting, some method may better predict in the low-power range than the rival approaches and would thus be pointed as more accurate if evaluated over a low-power production period. This is why it is necessary to base a judgment about the quality of a given forecasting method on a complete analysis, consisting in a variety of error measures.

Then, a possibility to compare the level of performance of various methods and also to quantify the gain of preferring an advanced approach to the reference ones (cf. Section 2.4) is to use a criterion defined as the *improvement with respect to* the considered reference method. This improvement (sometimes called 'skill score') corresponds to the error reduction that is achieved by the advanced method, for a given error measure. It is defined as following:

$$\mathrm{Imp}_{\Upsilon}^{\mathrm{ref}}(k) = \frac{\Upsilon^{\mathrm{ref}}(k) - \Upsilon(k)}{\Upsilon^{\mathrm{ref}}(k)}, \tag{3.9}$$

where Υ is the considered evaluation criterion. $\Upsilon^{\mathrm{ref}}(k)$ denotes its value for the reference approach and $\Upsilon(k)$ for the advanced one, for horizon k. This criterion can be either MAE, RMSE, or even SDE (or the equivalent normalized versions). The improvement is often multiplied by 100 and then expressed as a percentage improvement against the performance of the reference approach. Values of $\mathrm{Imp}_{\Upsilon}^{\mathrm{ref}}$ are always less than 100%: such a value would mean that the advanced approach forecasts are actually perfect forecasts. But, they can go below zero if the advanced approach performs worse (in terms of the chosen criterion) than the reference one. We will see in a later part that it may be the case for some wind power forecasting methods in certain conditions. $\mathrm{Imp}^{\mathrm{ref}}_{\Upsilon}$ increases with forecast skill.

Another way to illustrate and compare the skill of various forecasting methods is to compute the coefficient of determination R^2 for each look-ahead time:

$$R^2(k) = \frac{\mathrm{MSE}^0(k) - \mathrm{MSE}(k)}{\mathrm{MSE}^0(k)}, \tag{3.10}$$

where MSE^0 is the Mean Squared Error for the global average model (cf. Equation (2.14)).

In statistics, the coefficient of determination represents the ability of a model to explain the variance of the data. The value of R^2 is between 0 for useless predictions and 1 for perfect ones. However, the R^2 criterion is primarily designed for model selection using the training set. For this reason, we suggest to avoid its use as a main tool for performance evaluations with a statistical-background point of view, since it might yield misleading conclusions. If, for instance, the naive predictor is used for large horizons, the resulting R^2-value will be negative. This is due to the fact that the asymptotic variance of the prediction errors for the naive predictor is twice the variance of the global mean prediction defined by Equation (2.14) [175]. The R^2 criterion can be considered for comparing the performance of various methods, and/or for various sites, but then it should be remembered that this is out of the scope of its primary use in statistics.

There exist several possible definitions of the R^2-value. One possibility that is frequently used is to define the R^2 from the correlation between the measured and predicted wind power. The problem of this definition is that although the predictions might be biased this definition will lead to $R^2 = 1$. The definition suggested in (3.10) does not pose this problem, since both the systematic and random parts of the error are embedded in the MSE values. Thus, if the R^2 criterion is reported, it is extremely important to describe exactly how it is calculated. Actually, one notes from Equation (3.10) and Equation (3.9) that

$$R^2(k) = \mathrm{Imp}^{0}_{\mathrm{MSE}}(k), \tag{3.11}$$

which actually means that this definition of R^2 corresponds to the improvement with respect to the climatology model (or global mean, cf. Equation (2.14)) in terms of MSE, referred to as *MSE skill score* in the meteorological literature. This score has been proposed by Murphy and Epstein [166] and is widely used for evaluating meteorological forecasts of continuous variables. Actually, the choice of the relevant reference approach depends on the temporal scale one looks at: for short-term forecasts persistence appears to be a suitable choice and for longer time ranges climatology may be more appropriate. In the area of weather forecasting, it is climatology that is usually seen as the benchmark prediction, whereas persistence is more often considered by wind power forecasters. Similarly to the other improvements criteria, R^2 can also be expressed as a percentage if multiplied by 100.

Distribution-oriented approach for performance evaluation and further understanding of prediction error characteristics

The previously introduced measures permit to summarize the performance of a forecasting method with a single number — for instance, "a given method has a NMAE of 3.5% of the considered wind park installed capacity for one-step ahead forecasts". However, as it was discussed above, a single error measure does not appear to be appropriate for judging of the quality of a method. Then, another way to characterize forecasting errors is to study their distributions. Analyses based on distributions are expected to give richer information on error characteristics [23].

Error histograms represent the empirical distributions of these errors. It is recommended to use the same size for all bins when plotting a histogram to avoid misleading interpretations of the error distributions. The optimal size for the bins can be derived from the number of observations (following for instance Sturges' formula or Doane's rule, see [200], p. 48), even though in practice the bin size is often chosen for convenience.

Figure 3.1 gives an example of a histogram of prediction errors. Errors are normalized by the considered wind farm nominal power P_n and sorted with bins representing 5% of that installed capacity. Errors are gathered after the application of an advanced forecasting method for the hourly prediction of the 12-hour ahead wind power production for a multi-MW wind farm. They are collected over a 4-month test set ($\sim$3000 forecasts). When looking at the (empirical) error distribution, one notes that it is slightly right-skewed and that the forecasting method almost never made errors larger than 50% of the wind farm nominal power over the test set. The normalized bias and SDE corresponding to that error distribution are equal to 1.97% and 16.37% of $P_n$respectively.

An important feature that can be derived from an error distribution is the frequency of occurrence of errors under or above a given threshold. This type of analysis gives an answer to a common question from the side of end-users: *"How often does your forecasting method make unacceptable errors?"*. This is an important aspect of a forecasting method robustness. Indeed, two forecasting methods with similar MAEs on the test set may present completely different error distribution shapes and thus different frequencies of large errors. Plots giving such an information will be utilized in the following of the Chapter and referred to as *error margin plots.*

The distribution-oriented approach is based on the notion that it is the joint distribution of forecasts $\hat{p}$ and observations p, $q(\hat{p}, p)$, which contains all the non-time-dependent information about a prediction method's quality [167]. Such a distribution-oriented approach is also known as the Murphy-Winkler verification framework, named after the two researchers who actually proposed it. While it is rather hard to directly study this joint distribution, one can instead focus on the various conditional and marginal distributions for deriving the necessary conclusions on the joint distribution properties. These various distributions include the conditional distributions of the observations given the forecasts $q(p|\hat{p})$, the conditional distributions of the forecasts given the observations $q(\hat{p}|p)$, the marginal distribution of the observations $q(p)$ and finally the marginal distribution of the forecasts $q(\hat{p})$. For all

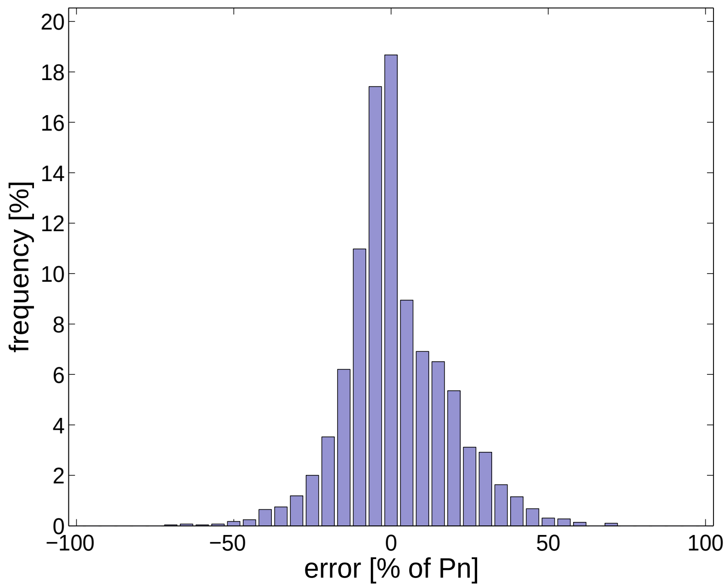


Figure 3.1: *Example of a histogram of errors. Prediction errors are normalized and sorted using bins representing 5% of the wind farm nominal power.*

the various aspects of forecast quality and the way they can be assessed from the analysis of these distributions, we refer to [165]. Some of these aspects will be mentioned throughout the present Chapter.

Following a distribution-oriented approach, we will apply a methodology consisting in studying the influence of a given variable (e.g. predicted power) on the moments of prediction error distributions (from the first to fourth order). This is because these moments correspond to different characteristics of prediction errors:

- the mean μ_k^e locates the 'center of gravity' of a distribution and provides an information on the systematic part of the error. It is given by the bias, as introduced by Equation (3.4),

- the standard deviation σ_k^e reflects the dispersion of a distribution, thus telling on the level of prediction uncertainty (cf. SDE criterion given by Equation (3.8)),

- the skewness ν_k^e describes the lack of symmetry of a distribution. It gives the most likely direction of expected prediction errors. A distribution with an asymmetric tail extending out to the right is referred to as positively skewed. The skewness is often estimated following Fisher's formula:

$$\hat{\nu}_k^e = \frac{N_T}{(N_T-1)(N_T-2)} \sum_{t=1}^{N_T} \left(\frac{e_{t+k/t} - \hat{\mu}_k^e}{\hat{\sigma}_k^e} \right)^3, \tag{3.12}$$

- the excess kurtosis κ_k^e informs on the shape of a given distribution, compared to the shape of normal distributions. The excess kurtosis for a normal distribution is equal to zero. Then, a positive excess kurtosis translates to a sharper peak and heavier tails. This moment is estimated by

$$\hat{\kappa}_k^e = \frac{N_T(N_T+1)}{(N_T-1)(N_T-2)(N_T-3)} \sum_{t=1}^{N_T} \left(\frac{e_{t+k/t} - \hat{\mu}_k^e}{\hat{\sigma}_k^e} \right)^4 - \frac{3(N_T-1)^2}{(N_T-2)(N_T-3)}. \quad (3.13)$$

3.3 Scope of the study

3.3.1 Prediction methods

Here, we consider five of the state-of-the-art approaches that were described in Chapter 2. Let denote by M1, M2,..., M5 the five forecasting methods. All these approaches have an operational nature and are actually part of wind power forecasting softwares installed in various European countries, for the day-to-day management of several gigawatts of wind generation. M1, M2 and M3 are statistical methods that are based either on artificial intelligence or structural modeling approaches. They all derive predictions from NWPs and online power production data. M4 and M5 are physical methods that use slightly different approaches for modeling wind characteristics and the energy-conversion process. Still, none of these two methods integrates CFD codes for modeling the wind in the vicinity of the wind farm. These two methods post-process their forecasts with model output statistics. Both persistence and the climatology model are seen as reference approaches for quantifying the benefits from the use of the more advanced ones.

3.3.2 Case-studies

We focus on 4 test cases consisting of real wind parks located in different zones of Europe, with various terrain conditions (offshore, flat and complex terrains). Tunø Knob and Klim are located in Denmark, Golagh in Ireland and Sotavento in Spain. Table 3.1 summarizes the main characteristics of these case-studies. Evaluation periods range from few months to several years depending on the available data from each wind farm.

Table 3.1: *Summary of the evaluation case-studies. The test cases span the different terrain types, with wind farms of various installed capacity. They all consist of several hundreds of predictions.*

Name (Type)	P_n [MW]	Beginning date	Ending date	# predictions
Tunø Knob (offshore)	5	16-12-2002	28-04-2003	536
Klim (flat terrain)	21	01-03-2001	28-04-2003	3156
Golagh (hilly terrain)	15	01-01-2003	30-03-2003	228
Sotavento (complex terrain)	17.56	01-09-2001	30-11-2001	2161

In the frame of this study, all the prediction methods have the same available data that can be used as input (online power production data and NWPs from HIRLAM). Meteorological forecasts are directly provided at the level of the wind farm. NWP data include at least wind speed and wind direction forecasts at 10 meters a.g.l., and in some cases predictions of other meteorological variables (e.g. temperature) and/or predictions at other heights (given as heights above ground level or at some atmospheric pressure levels). It is then the choice of the modelers to integrate or not these data in their modeling approach. The horizontal resolution of HIRLAM meteorological forecasts is of 0.15° for Golagh (approximately 16km) and the two Danish wind farms, whereas it equals 0.2° for the case of the Sotavento wind farm (approximately 22km).

Statistical and physical methods do not have the same frequency of update (hourly updates for statistical methods and updates only when NWPs are provided, i.e. every 6 hours here, for the physical ones). Therefore, in order to objectively compare prediction results from these approaches, we will only consider forecasts produced when NWPs are delivered for the test cases where results from both statistical and physical methods are available. This concerns the Tunø Knob, Klim and Golagh wind farms, even if for Klim results from the M4 method are missing. It should be noted though that statistical prediction methods were not optimized here for giving predictions in such an operational context. It is hence expected that their performance will be slightly lower than if they had been especially tuned for that particular operational set-up. Though, this slight decrease of performance will not affect our study of error characteristics. Regarding the Sotavento wind park, only results from the three statistical methods have been gathered. Thus, all the hourly predictions over the evaluation period are considered. Finally, the forecast length varies from 19 to 48-hour ahead depending on the wind farm and the prediction method.

3.4 Evaluating the quality of state-of the-art point prediction methods

As a first use of these various test cases, we have carried out an evaluation of the quality of the state-of-the-art point prediction methods. This evaluation follows from the evaluation framework developed in the above Section. We comment here on the main results of this evaluation study, in order to derive first conclusions on the level of uncertainty of state-of-the-art methods. A complete view of the evaluation results is given in Appendix D.

3.4.1 Analysis based on error measures

The NMAE and NRMSE are the most widely used error measures for having a first evaluation of prediction methods' performance. They are depicted here as a function of the look-ahead time for the Golagh test case (Figures 3.2 and 3.3). If no particular mention, the Golagh wind farm will be considered for illustration all along the present Section.

The NMAE measure can be directly interpreted. For instance, from Figure 3.2 one tells that M3 makes an average absolute error of 11.5% of the wind farm nominal power for 24-hour ahead forecasts, or that M4 makes an average absolute error of 12.5% of P_n for that same horizon. Since Golagh is a 15MW wind farm, this 1% difference between the two methods' NMAE represents a 150kW difference if errors were not normalized. Such an in-

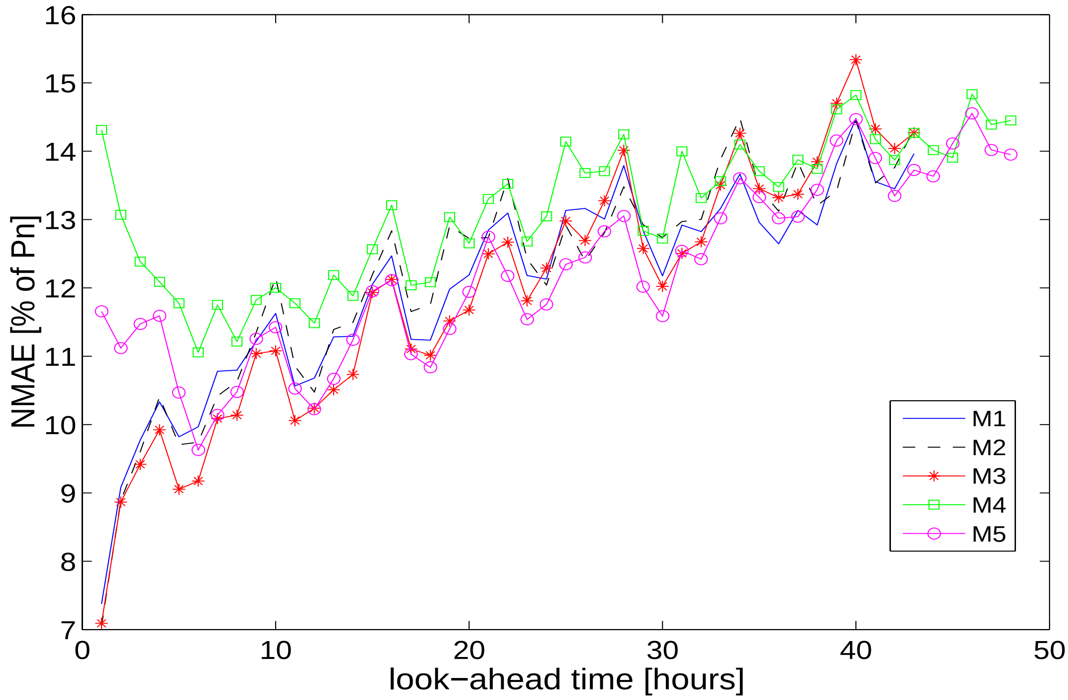


Figure 3.2: *Performance evaluation of the forecasting methods by the use of the NMAE measure as a function of the look-ahead time. Results are for the Golagh case-study.*

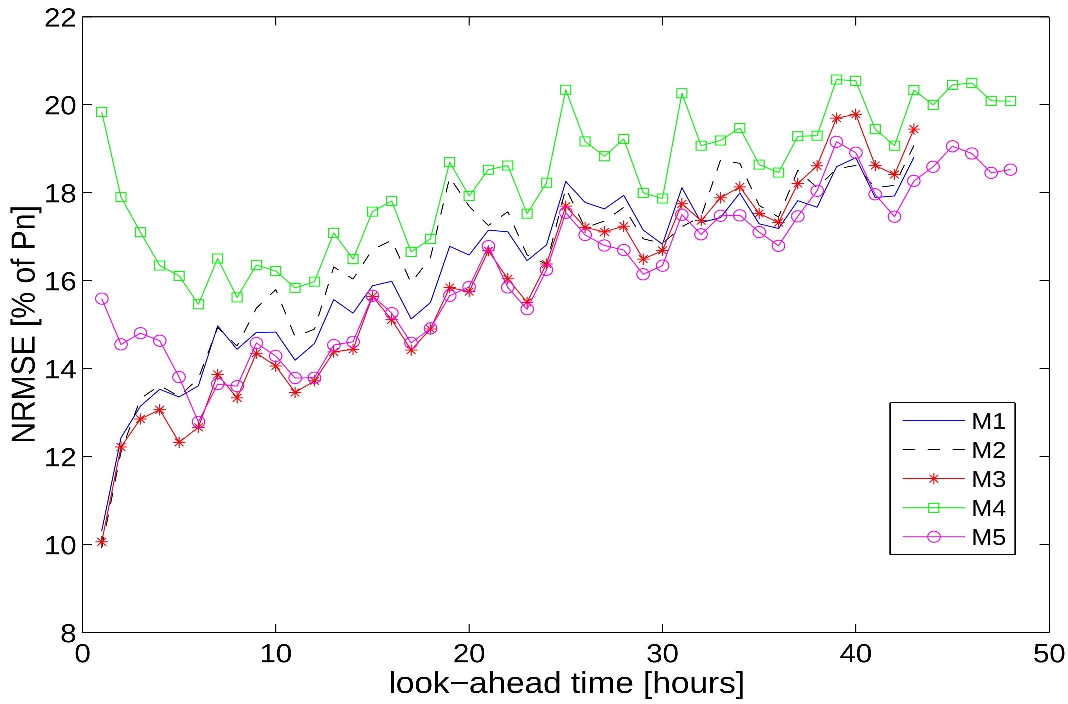


Figure 3.3: *Performance evaluation of the forecasting methods by the use of the NRMSE measure as a function of the look-ahead time. Results are for the Golagh case-study.*

formation is not given by the NRMSE measure since it has a quadratic form and thus gives more weight to large errors. Though, by comparing the NMAE and NRMSE plots, it may be possible to draw first conclusions on the methods' behaviors. For instance, one notices that M4's NMAE values over the whole range of horizons are just slightly higher than for the other approaches, while its NRMSE values are more significantly higher than for the other approaches. This may mean that M4 makes large errors more often, or that its large errors have a higher level. It will be revealed by studying the error distributions in a later part.

The first comment we can draw from the analysis of these two Figures is that both error measures augment in a constant manner with the forecast horizon, with a trend that is quasi linear. If we focus on Figure 3.2, The NMAE values for M1, M2 and M3 are approximately twice higher for 2-day ahead forecasts than for 1-hour ahead predictions (going from 7% to 14-15% of P_n). This cannot be stated for the two physical approaches M4 and M5. Indeed, there is a clear difference between physical and statistical methods for the first six look-ahead times: the level of prediction error of the statistical methods is lower for these horizons thanks to the use of power production data as input. This way, they manage to capture the persistent behavior of the wind. The physical approaches rely on NWPs only for estimating wind power generation. Such NWPs are of lower quality for the first look-ahead times. Meteorological forecasts focus on further horizons. Then, for look-ahead times between 6 and 48-hour ahead, the error measures for all the different approaches are contained in an envelope representing 2-3% of the installed capacity. The difference in wind-to-power modeling approaches is the sole responsible for the performance variability. One may consider that variability as reasonable, but it actually represents a gain of performance up to 20% in regard of the values of the error measures. It is expected that these differences will have a significant impact on the value of the various methods in an operational context.

Table 3.2: *Summary of the minimum NRMSE values over the whole range of horizons and of the envelopes of NRMSE values for 24-hour ahead and 48-hour ahead forecasts, for the various test-cases.*

Test case	Min. [% of P_n]	24-ahead [% of P_n]	48-ahead [% of P_n]
Tunø Knob	11	17.5-18.5	22-23
Klim	10	15.5-16	18.5-19
Golagh	10	16-18	18.5-20
Sotavento	8.5	16-19.5	–

Table 3.2 gives a summary of the lowest NRMSE values and of the NRMSE envelopes for one-day and two-day ahead forecasts among the various prediction approaches for the considered case-studies. The minimum NRMSE value always corresponds to the one achieved by statistical methods for 1-hour ahead forecasts. It ranges from 8.5 to 11% of the installed capacity. Note that the performance of statistical approaches for short horizons is not related to the terrain type. The lowest NRMSE is for the Sotavento case study which is located in a complex terrain and the highest one is for the offshore wind farm (Tunø Knob). For such look-ahead times (up to 3-6 hours ahead), since statistical predictions are mainly based on historical power measures, their level of prediction error is greatly influenced by

the persistent nature of local winds, and also by the quality of the online power production data. For further horizons, both of these aspects have less importance and this is more the quality of meteorological forecasts that is paramount. In parallel, NRMSE envelopes for one-day and two-day ahead predictions exhibit a substantial width variability depending on the case-study. NRMSE values attain a higher level for the two wind farms in semi-complex or complex terrain and for the offshore wind park. Also, these NRMSE envelopes are tighter for the two Danish wind farms. It seems that it is pretty easier to predict wind for Klim than for the other wind farms, because it is located in flat terrain and witnesses more stable wind conditions. Also, the modeling of the energy conversion process appears easier for the two Danish wind farms, since the performance variability is lower among the various approaches. This is in line with the conclusions derived by Kariniotakis et al. [111] for a larger number of case-studies and with more prediction approaches: as terrain complexity increases, both the average level of prediction error and the variability in the various approaches' performance augment.

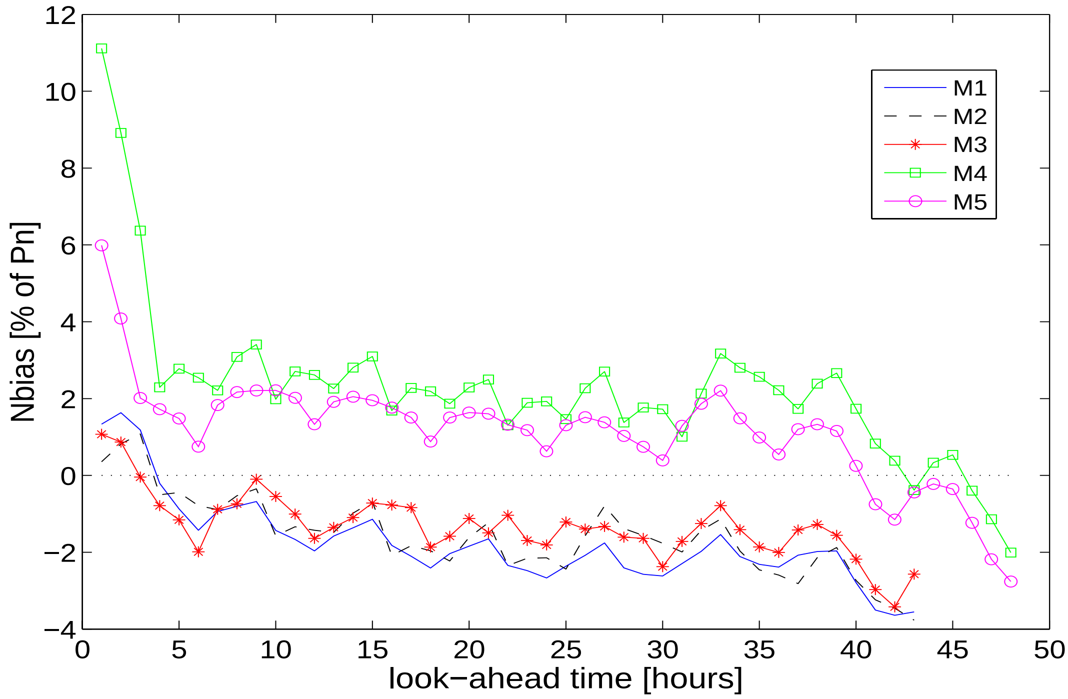


Figure 3.4: *Performance evaluation of the forecasting methods by the use of the Nbias measure as a function of the look-ahead time. Results are for the Golagh case-study.*

In Paragraph 3.2.1, we mentioned that the prediction error can be decomposed into a systematic and a random part (cf. Equation (3.3)). Both parts contribute to the MAE and RMSE measures[1]. Here, for highlighting what is the contribution of the systematic part to

[1] An analytical formulation of the contribution of both the systematic and random part of the error to the RMSE is given by Equation (3.14)

these error measures we focus on the normalized bias of wind power forecasts. Figure 3.4 depicts the evolution of this evaluation criterion as a function of the look-ahead time for the various methods.

A first glance at the Figure yields the surprising comment that the curves can be gathered in two groups, that of the physical and the statistical approaches. It seems that for this test case, the MOS used by M4 and M5 lead to similar bias corrections, whereas the various statistical modeling approaches also yield similar bias corrections. Such comment cannot be generalized, as we see that this kind of grouping does not hold for the other case-studies (Figures E.1, E.7 and E.19). The Nbias values for the two physical approaches reach a high level for the very first horizons (up to 12% of the wind farm nominal power for M4), while it is contained between -2% and 2% of P_n for further look-ahead times. This high contribution of the systematic part of the error is again due to meteorological forecasts that are not very accurate up to 3-6 hour ahead. In the worst cases, this phenomenon can be seen in the systematic part, but most of the time, this is not the case, and it is more the SDE measure that is slightly higher for the first horizons (see Figures E.2 and E.8). If we leave aside that particular point, one notices that Nbias values are quite low. Having a general view on the various case-studies, we can conclude that it is mostly the random part of the error that contributes to the NMAE and the NRMSE measures. Both the MOS techniques and the direct statistical modeling approaches have the ability of (almost) removing the systematic part of the error.

3.4.2 Performance against reference approaches

The skill of an advanced forecasting method is often defined as the improvement it proposes against reference approaches. We mentioned before that persistence and climatology forecasts consisted the two benchmarks, depending on the temporal scale one looks at. In this Paragraph, we will consider both of these reference methods for quantifying the benefits (if any) of using the various advanced prediction methods.

In a first stage, the improvement with respect to persistence is calculated for the NRMSE error measure, and depicted in Figure 3.5 as a function of the prediction horizon. The improvement proposed by all the methods sharply grows for the first look-ahead times. Persistence is a serious benchmark up to 6-hour ahead, but for further horizons all the methods propose improvements that are in the range of 40-60%. For the very first horizons, the skill of the two physical approaches M4 and M5 is even negative (down to -80%, though we have cut the y-axis for better seeing the positive part of the graph). In other words, it is preferable to use persistence than these two advanced methods if needing forecasts up to 3-hour ahead. Thanks to their autoregressive part, statistical methods (M1, M2 and M3) are always as good or better than persistence, even though their skill is rather low for the very first look-ahead times. It is actually not very hard to attain such improvements for horizons further that 12-hour ahead: Nielsen et al. [175] showed that even the climatology's MSE is half of the MSE of persistence for longer-range forecasts. This is why for evaluating the skill of advanced methods for longer-term forecasts it may be more suitable to use climatology as a reference approach.

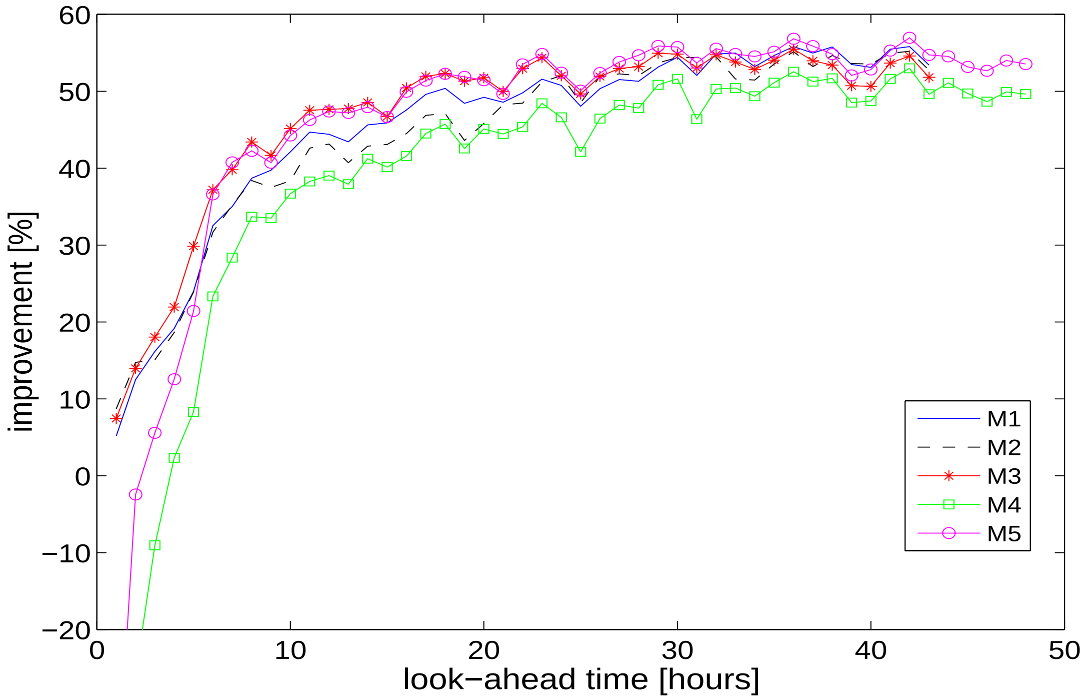


Figure 3.5: *Skill evaluation of the forecasting methods - Improvement with respect to persistence for the NRMSE criterion, as a function of the look-ahead time. Results are for the Golagh case-study.*

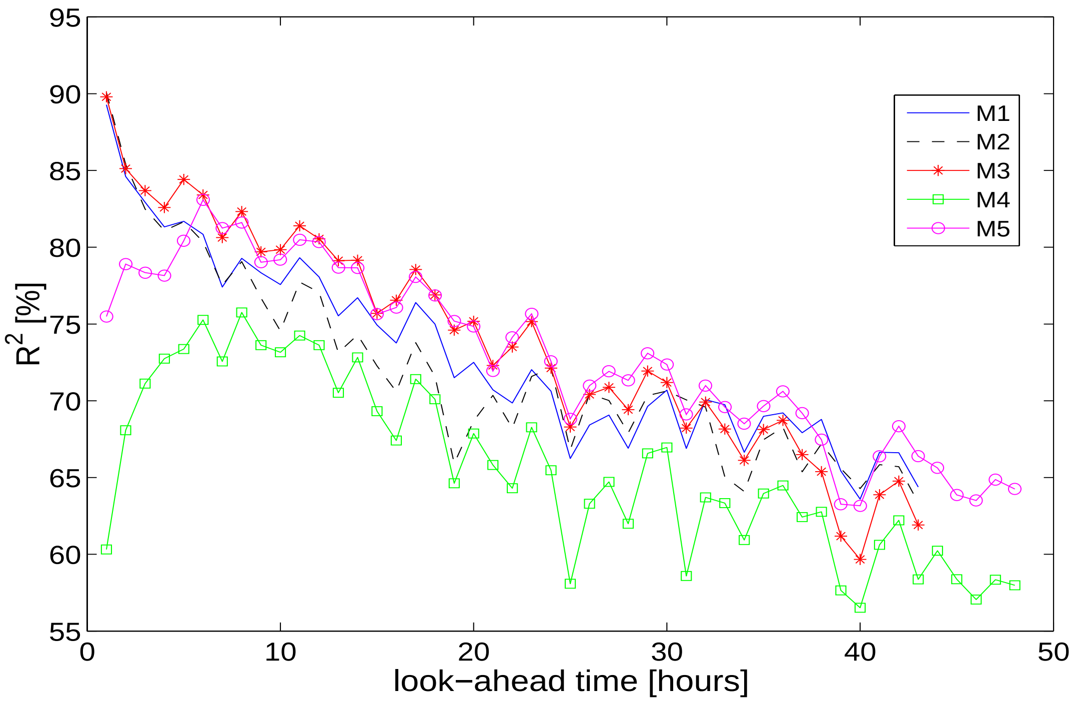


Figure 3.6: *Skill evaluation of the forecasting methods - R^2 as a function of the look-ahead time. Results are for the Golagh case-study.*

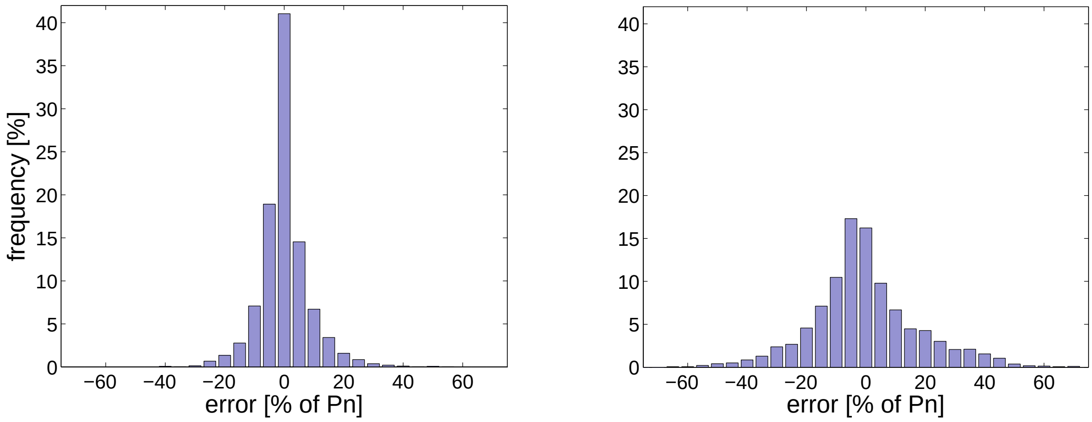
Alternatively, the improvement with respect to climatology is given by the R^2 criterion. Its evolution with the forecast horizon is displayed in Figure 3.6 for the various advanced methods. Inversely to the previous criterion, the general trend is that the methods' skill decreases with the look-ahead time. Remember that following the R^2 definition (cf. Equation (3.10)) a 100% score value means that the evaluated method produces perfect forecasts, and that a zero value translates to the advanced method having a level of performance similar with that of climatology. Up to 6-hour ahead, the physical methods M4 and M5 provide forecasts of lower quality than that from the statistical approaches. Though, their skill augments for these first look-ahead times. Then, for further horizons, the R^2 values steadily decrease for all the methods. Even though M4's skill is substantially lower over the whole range of horizons, all prediction methods propose very high improvements (between 55 and 90%) with respect to climatology.

Figure 3.6 is for the Golagh test case, which is located in hilly terrain. One sees from the other Figures depicting R^2 values in Appendix D (Figures D.4, D.10 and D.22) that their minima and maxima are significantly different from one test case to the other: the maxima are between 85 and 91% (the highest value being for the offshore wind farm) whereas the minima range from 45 to 65% for two-day ahead forecasts. The overall minimum even equal 32% for one-day ahead predictions for Sotavento. This firstly means that whatever the case-study the various methods produce forecasts that are by far more accurate than the ones from climatology. While persistence is a serious benchmark for the first look-ahead times and can thus be used for these horizons if one is reluctant with investing in an advanced prediction method, it is not the case for climatology (at least for horizons up to 48-hour ahead). Since this benchmark approach has no dynamics, its level of performance is directly related to the variability of the wind power production for the considered case-study.

In the frame of the verification of wind power forecasts, it is not clear what is the best reference approach to consider. Persistence could be seen as a more appropriate choice since we deal with short-range horizons. Though, as one has seen that the methods' improvements rapidly attain high levels with the lead time getting further, relying on a such a benchmark approach may mask the differences between the advanced prediction methods. This would be an argument in favor of the use of the R^2 for better resolving among the methods' skill, but similarly, very large improvements for the first horizons against that benchmark are inherent to its non-dynamic nature. It is indeed for that reason that Nielsen et al. [175] proposed a new reference model, which consists in a clever combination of both persistence and climatology. However, since this new reference model is not easy to implement, our advice is to carry out the skill assessment by considering both reference approaches. Here, we have seen that the performance of the state-of-the-art wind power forecasting models was always better than those of persistence and climatology, except for the case of physical-type methods that do not manage to outperform persistence for the very first look-ahead times. Though, the improvements with respect to these benchmark approaches may significantly vary among the advanced prediction methods, as well as among the test cases.

3.4.3 Analysis based on error distributions

After using the error measures, the analysis of the error distributions is expected to give more insight on the error characteristics. Marginal distributions of prediction errors are considered for each forecast horizon. Figure 3.7 gives the example of two error histograms from the application of the M2 forecasting method to the Tunø Knob case-study. They correspond to the errors for 1-hour ahead (left) and 24-hour ahead (right) predictions.



(a) 1-hour ahead **(b)** 24-hour ahead

Figure 3.7: *Empirical distributions of the prediction error for two look-ahead times. Prediction errors are normalized and sorted in bins representing 5% of P_n. These distributions result from the application of the forecasting approach M2 to the Tunø Knob test case.*

At first sight, it is clear that the error distribution for 1-hour ahead forecasts is much sharper than the other: the central part of the distribution is higher and its tails shorter. This was expected since the use of the error measures has already told us that the level of prediction error was increasing with the look-ahead time. Moreover, all the 1-step ahead prediction errors are less than 40% of P_n while the upper bound for 24-step ahead errors is close to 65% of P_n. That maximum prediction error one may expect when applying a forecasting method to a given case study is a first information about the prediction uncertainty.

Error margin plots ease the interpretation of error distributions: they give the proportion of errors within a given error margin (expressed as a percentage of the wind farm installed capacity P_n) as a function of the look-ahead time. Similarly, they can tell what is the proportion of the errors that are above a certain threshold: this proportion is simply given by 100% minus the proportion lower below that threshold. Evaluating these proportions will answer the end-user's concern about the frequency of occurrence of unacceptable errors we expressed before. Figure 3.8 is an example of such a plot for an error margin representing $\pm 5\%$ of P_n (thus corresponding to 750kW for Golagh), for all the forecasting methods.

In Figure 3.8, the proportion of errors within the error margin decreases with the look-ahead time, but not at the same rate for all the methods. When for M4 and M5 the proportion

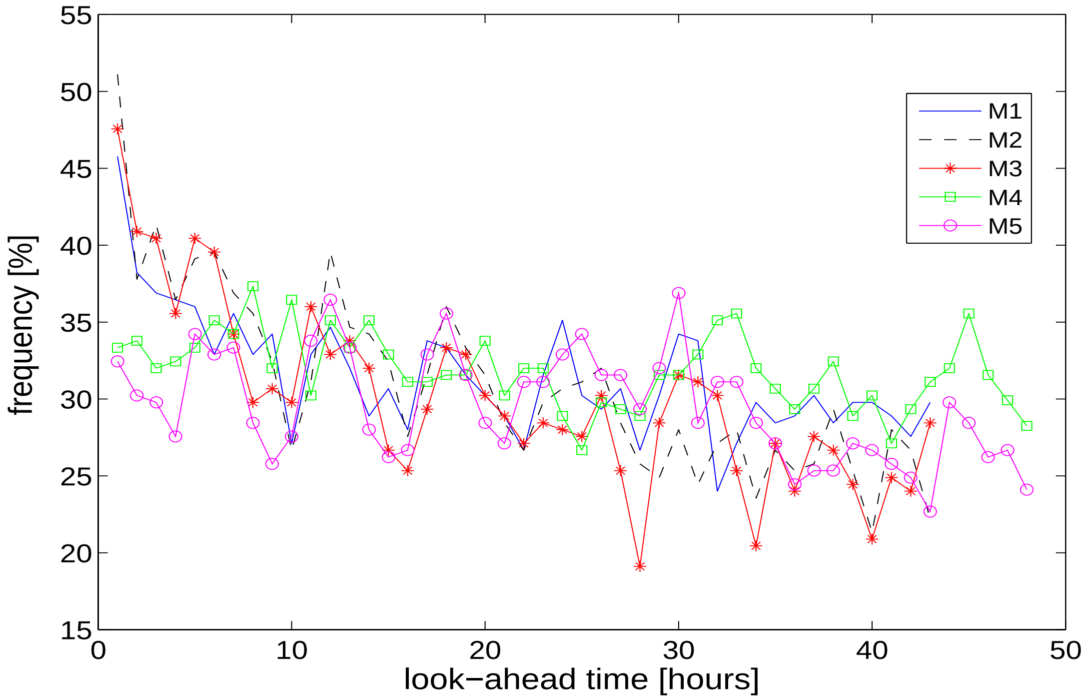


Figure 3.8: *Proportion of errors within a* $\pm 5\%$ *(of* P_n*) error margin as a function of the look-ahead time. Results are for the Golagh case-study.*

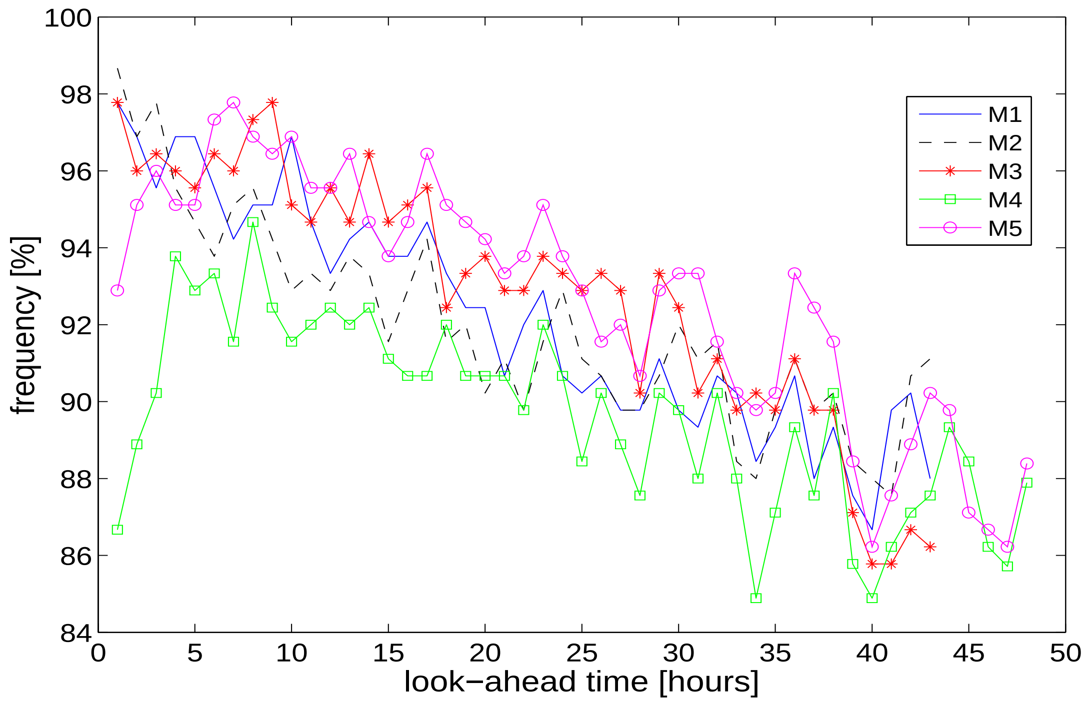


Figure 3.9: *Proportion of errors within a* $\pm 30\%$ *(of* P_n*) error margin as a function of the look-ahead time. Results are for the Golagh case-study.*

goes from 33-34% at the first look-ahead time to 24-28% for 2-day ahead predictions, this same proportion starts from 46-50% and drops to 23-30% for 2-day ahead forecasts for the statistical methods. Actually, all the statistical approaches have a larger proportion of small errors for the first 4-5 look-ahead times, but for further horizons that proportion diminishes steadily. The proportion for physical methods is slightly more constant over the forecast length. Still, approximately a half (or alternatively one third for the physical approaches) of the errors are at a very low level for the closer lead times, and between one fourth and one third of the predictions errors are within this margin for further horizons. This means that all the state-of-the-art approaches we consider here makes very low errors on a rather regular basis. These proportions are dependent on the case-study: frequencies of small errors are lower for the Sotavento case-study located in complex terrain for instance (Figure D.23).

If associating an error margin plot for larger errors, it is then possible to compare the central and tail parts of error distributions. Greater frequencies of errors larger than a high threshold would signify that these distributions have heavier tails. A 30%-error-margin plot (Figure 3.9) for the five methods serves as a comparison with the previous Figure. For the case of Golagh, 30% of the nominal power corresponds to 5MW. Here, we consider that errors larger than 30% of the installed capacity are extreme prediction errors. Hopefully, one notes from Figure 3.9 that there are only few occurrences of extreme prediction errors, less than 15% of the times whatever the approach or the look-ahead time. Though, there are significant differences among the various methods and over the forecast length: M4 and M5 makes two to six times more extreme errors than the statistical methods for the very first horizons for instance. Differences tend to vanish for further look-ahead times.

By comparing the two error margin plots, we notice that the methods' behaviors are quite different in the central and tail parts of the error distributions. If one takes the example of M4, this method makes as much or even more low errors than the others, but it is also the method that has the highest frequency of extreme prediction errors: the corresponding distributions seem to be sharper and higher in the central part, though they have heavier tails. Similar remarks can be drawn for the error distributions corresponding to the results of M4 for Tunø Knob (Figures D.5 and D.6) and of M1 for Klim (Figures D.11 and D.12). Here, the fact that M4 makes large prediction errors more often that the others actually explain our comment about the Figures depicting the NMAE and NRMSE error measures: knowing that the NRMSE measure penalizes more methods that make larger errors, one can already have a first clue on the error distributions' tails by thoroughly comparing several methods with these two measures.

3.5 Highlighting the characteristics of the prediction uncertainty

The previous assessment of the performance of state-of-the-art approaches informs on the level of prediction uncertainty one may expect when applying one of these approaches for a new wind farm. Such an evaluation may be sufficient for wind farm operators. But, in order to upgrade their forecasting methods, modelers are interested in a more thorough study of prediction errors. Alternatively, this better understanding of wind power forecasting uncertainty is necessary for designing appropriate methods for its online estimation. In the following, we develop first on the various contributions to prediction uncertainty. Then, we apply a generic methodology for studying how some variables influence the prediction uncertainty, by focusing on their impact on the moments of prediction error distributions. The particular effects of the look-ahead time and of the level of predicted power are underlined.

3.5.1 Contributions to the wind power prediction error

The role of meteorological forecasts

The main input for a wind power prediction method today is the information provided by the meteorological forecasts. This information is composed by estimates of wind speed and wind direction mainly (and eventually temperature, humidity, etc.) that are primarily given at the NWP model grid nodes around the site of interest. Statistical downscaling techniques (e.g. bilinear interpolation) are then applied, either by the meteorological office that supplies NWPs or by the wind power forecaster, to refine these meteorological forecasts to the level of the wind farm. For an overview of statistical downscaling methods and their performance, we refer to [229]. If online production data are available, they are used to enhance the model performance for the first look-ahead times (for the case of statistical-type methods), and also to remove the systematic error μ_k^e (and also a linear part of the error [129]) in the power forecasts when post-processing physical-type predictions with model output statistics. Therefore, the quality of wind power prediction models strongly relies on the quality of the NWPs: bad meteorological forecasts yield bad energy ones.

Let us consider for instance the case of a *phase error* in the wind speed forecasts. A phase error corresponds to a time-shift between a predicted meteorological front (or low) and the measurements. In such a case, the NWPs do not catch the temporal evolution of the situation properly: for instance the wind speed increase is well-predicted, but few hours before it actually materializes. The state-of-the-art wind power prediction models make an advanced conversion of wind speed to power, but they do not correct any temporal shift present in the meteorological forecasts (or maybe on a very-short time lag for statistical approaches that use past power values). It is thus possible to affirm that time-evolving patterns are imposed by the NWPs.

A typical example of a phase error is shown in Figure 3.10. All the five prediction methods base their estimate of future wind generation at Tunø Knob on the meteorological forecasts issued by HIRLAM on the 3$^{\text{rd}}$ February 2003 at 18:00. They are compared with hourly power measures for the same period. The various prediction series approximately follow the same time-evolving pattern. Despites the overall good agreement between predictions and measures one notes a phase shift when wind generation decreases between look-ahead times 35 and 40. Whatever the considered approach, the decrease in wind power output is predicted 6-7 hours too early. The resulting error is rather high: at horizons 35-36, while all the methods tell that wind generation is quasi null, it is actually between 40 and 45% of the installed capacity. The error corresponds here to only 2MW, but were we dealing with a 100MW wind park that this would translate to an error of around 40MW, for which the resulting penalties for imbalance on the market would be very high.

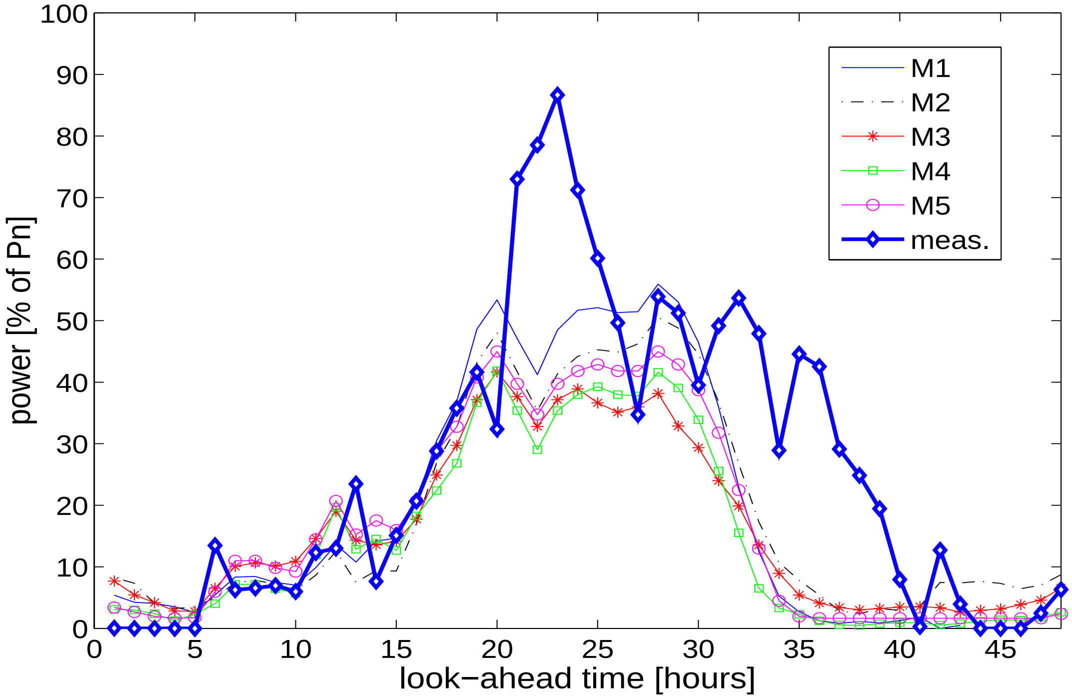


Figure 3.10: *Example of 48-hour ahead wind power forecasts for Tunø Knob, issued on the 3rd February 2003 at 18:00, with the five considered prediction methods. Forecasts are compared with power measures for the same period. This is a typical case of a "phase error".*

For better studying the power prediction errors and the contribution of phase errors, Lange [129] (following previous works by Takacs [213]) proposed to decompose the RMSE measure as following:

$$\mathrm{RMSE}^2 = \mathrm{bias}^2 + \mathrm{SDE}^2 = \mathrm{bias}^2 + \mathrm{sdbias}^2 + \mathrm{disp}^2, \tag{3.14}$$

where 'bias' is the prediction bias given by Equation (3.4), 'stdbias' the difference between the measure and prediction time-series variances, and 'disp', referred to as dispersion, involves the cross-correlation of the time-series. While a lack of variability of the forecasts increases the 'sdbias' part, phase errors mainly contribute to the dispersion value. By using this RMSE decomposition when studying the forecasting errors for several sites, Lange noticed that a very high part of the RMSE value was explained by the dispersion, and he hence concluded on the large contribution of phase errors in the final power prediction error. Such a contribution of phase errors has also been mentioned and discussed by Möhrlen [158].

In recent works, Louka et al. [141] implemented a Kalman filtering approach for improving numerical predictions of wind speed. Kalman filters [108] are a way to combine recent observations of a given process with a predicted value for a given lead-time. It can be seen as a MOS technique that aims at removing both the systematic and a linear part of the forecasting error. In that paper, Kalman filtering is applied for diminishing the wind speed prediction error for horizons up to 120-hour ahead (5 days). Then, non-corrected and filtered wind speed predictions are fed into a statistical wind power forecasting method. The authors showed that the wind speed forecasting error was greatly reduced and that the resulting decrease in the NMAE values for wind power forecasts was up to 22% for horizons between 48 and 120-hour ahead. This clearly shows that by improving the quality of the NWPs the performance of the resulting power output predictions is significantly enhanced. It should be noted though that such an approach is only applicable if local wind measurements are available at the level of the wind park, which is seldom the case nowadays.

The wind-speed-to-power model contribution

The second contribution to the power forecasting error is obviously the one introduced by the wind power prediction model itself. This error may be due to the way the NWP are refined at the level of the site, to the way the wind profile is modeled, to the way the park effects are considered, and finally to the way the power curve is estimated. As we explained in the above Paragraph, the time-evolving pattern is dictated by the NWPs. The following modeling steps act as nonlinear transformations of the predicted wind speed.

A wind farm power curve has two characteristics: it is bounded between 0 and the nominal power P_n, and it is nonlinear. The effect of the first characteristic on the forecasting error is that this error can vary between -100% and 100% of the wind park nominal power P_n. For a non-bounded prediction model it can take values even outside that range. The potential error of the prediction model, defined as *error margin*, depends on the level of predicted (or also measured) wind power. Figure 3.11 graphically represents the error margin as a function of a generic wind farm characteristic curve. For wind speeds below cut-in speed, the error margin is maximal since the model can predict a production up to the nominal wind turbine power. On the contrary, for higher wind speeds the model will show a negative error margin, i.e. the generated power is likely to be greater than the one proposed by the prediction model. Close to the cut-off wind speed the uncertainty is again maximal since the model can switch from a positive error margin to a negative one, or the inverse.

In [126, 133], it was shown that the power prediction errors depend on the errors involved in the prediction of wind speed by the NWP system. Lange [129] showed that the level of wind speed prediction error was not necessarily a function of the level of forecast wind speed. But, while there is an inherent modeling error associated to the chosen wind-speed-to-power conversion approach, the effect of the wind park power curve is to amplify (between cut-in and rated speed) or to reduce (below cut-in speed or between rated and cut-off speed) the uncertainty introduced by the NWPs. This effect can indeed be quantified by calculating the local power curve derivative [130]. Such an effect is particularly important in the steep part of the power curve where wind speed prediction errors can be amplified by a factor 2 or 3. Hence, it is expected that the NMAE, the NRMSE, or the standard deviation of forecast-power-dependent error distributions will vary in a similar way. In the following, we will address the question of whether the choice of a given approach has

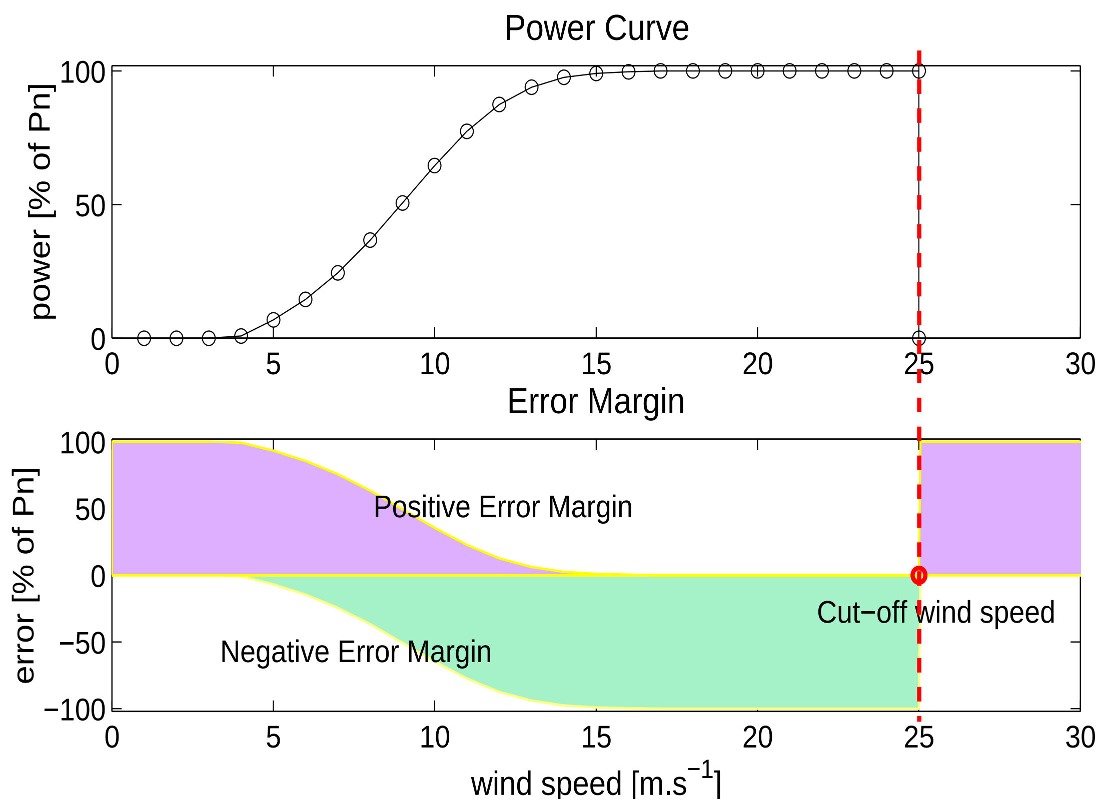


Figure 3.11: *The error margin as a function of the wind turbine power curve.*

or not an influence on the strength of this contribution.

The second contribution of the energy-conversion process modeling is that the nonlinear (and bounded) transformation which is applied to wind speed values actually acts on the very nature of error distributions. While it appears conceivable to test for the Normality of wind speed forecast errors (which would give them a nice property), we already know that wind power prediction errors cannot be Gaussian. Indeed, the nonlinear transformation completely changes the shape of error distributions. Lange [129] thoroughly studied the properties of unconditional wind speed and power prediction errors for 20 sites in Germany, over a 3-year period. He applied both a parametrical χ^2-test and a non-parametrical Lilliefors test for assessing the Gaussianity of the various error distributions. He concluded that a large part (between 80 and 90%) of the considered wind speed forecast error distributions could be considered as Gaussian, while none of the power prediction error distributions could pass the hypothesis tests. And, if considering conditional prediction errors (given the level of the predictand), it is commonly accepted for the case of nonlinear models that they are not distributed Gaussian anyway [41]. Consequently, it will be assumed in the remaining of this work that wind power prediction errors cannot be considered as Normally distributed.

Figure 3.12 gives the example of forecast series produced for Tunø Knob by the five state-of-the-art forecasting methods, based on the HIRLAM meteorological forecasts issued on the 29th January 2003 at 18:00. One notices that again the overall agreement is rather good. The increase of wind power generation for horizons between 10 and 15-hours ahead is well predicted, but the level of wind power output for the following 10 hours is underestimated by all the forecasting methods. Such an error is usually referred to as a *level error*. Also, even if all the forecasts series follow the same pattern, there is a difference in their amplitude (up to 20% of P_n in this example). Therefore, one notes from the Figure that the various modeling approaches do not lead to the same level error. It should be understood here that NWPs contribute to level errors too: when wind speed is significantly underestimated for instance, it is very unlikely that the power modeling approach manages to accurately predict the level of wind generation.

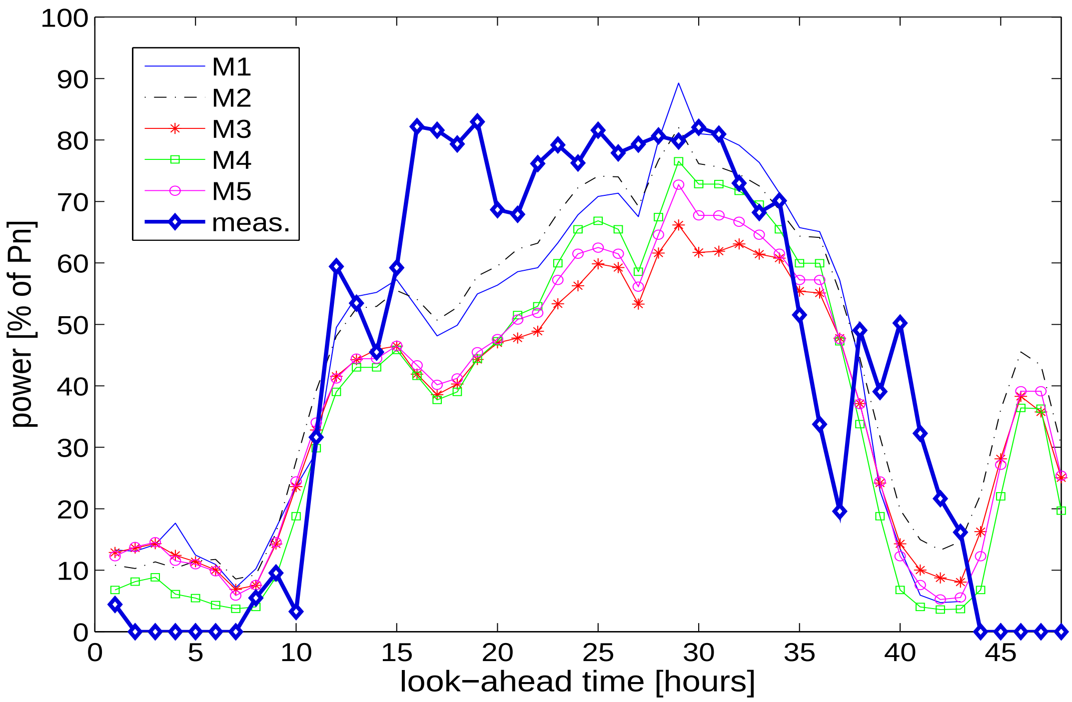


Figure 3.12: *Example of 48-hour ahead wind power forecasts for Tunø Knob, issued on the 29th January 2003 at 18:00, with the five considered prediction methods. Forecasts are compared with power measures for the same period. This is a typical example of a “level error”.*

Other factors and effects

Both the quality of the meteorological forecasts and the chosen power prediction method contribute to the value of the error measures. However, the site characteristics and period may also significantly influence the apparent performance of a given forecasting system. Results may greatly vary from summer to winter periods owing to the difference in the strength and characteristics of the winds. This has been shown in [117], in which the performance of an artificial-intelligence-based prediction method has been evaluated on a monthly basis for several wind farms in Ireland. Both the NMAE and NRMSE measures exhibited very high variations from a month to another (from simple to double). Significant variations in the level of performance of a given power prediction method may also occur if the considered site is located in flat or complex terrain, onshore or offshore. Recently, Kariniotakis et al. [111] underlined the influence of the terrain on prediction models' performance. The authors concluded on a strong link between terrain complexity and the level of performance end-users may expect when applying forecasting methods. This is why both analysts and end-users should be careful when comparing the performance of various methods for various sites, and clearly define the operational framework of the evaluation.

In addition, predicting the power output from a single site or for a group of wind farms (or even over a region) does not lead to the same level of forecast uncertainty. Due to the geographical smoothing effect, the variability of wind generation is lowered, and the average level of prediction error is diminished too. Focken et al. [70] studied that effect over the whole Germany area, and showed that the average level of prediction error (quantified with the NRMSE) was reduced by approximately 60% when predicting for the whole country in comparison with errors than can be seen for a single site.

3.5.2 Characteristics of prediction errors

The effect of the prediction horizon

A first aspect is that the characteristics of prediction error distributions depend on the look-ahead time. This statement is valid whatever the considered modeling approach. Wind power forecasting methods can be built on a per-horizon basis or in an iterative way, but all the different methods we consider here use meteorological forecasts as input. It is known that the magnitude of possible errors in meteorological forecasts grows as the horizon increases [182]. The time-evolution of the atmospheric system is usually described with finite-difference numerical schemes that diverge at a low rate. Both the uncertainty in the initial state estimation and in the meteorological model (often due to local parameterization) are responsible for this error growth [181].

The effect of the look-ahead time on the power production forecast uncertainty is depicted in Figure 3.13, in which this uncertainty is quantified by the standard deviation of prediction errors. Results are derived from the application of the 5 prediction methods to the Golagh case-study. Although there are differences in the level of NSDE among the various modeling approaches, the general trend is the same. The NSDE increase is almost linear. The minimum NSDE is for the first horizon (10% of P_n), and it reaches a maximum of 20% of P_n for 2-day ahead forecasts. The augmentation of the forecast uncertainty with the look-ahead time can also be seen for the other wind farms (Figures E.8, E.14 and E.20).

In parallel, the look-ahead time does not have a significant impact on the other moments of prediction error distributions. First, the systematic part of the prediction error should not be a function of the look-ahead time and should stay at a very low level, since statistical prediction methods can be made unbiased by applying appropriate estimation methods, or alternatively MOS techniques can be used for the case of physical methods. This was discussed in Paragraph 3.4.1 and is illustrated in Figures E.1, E.7, E.13 and E.19.

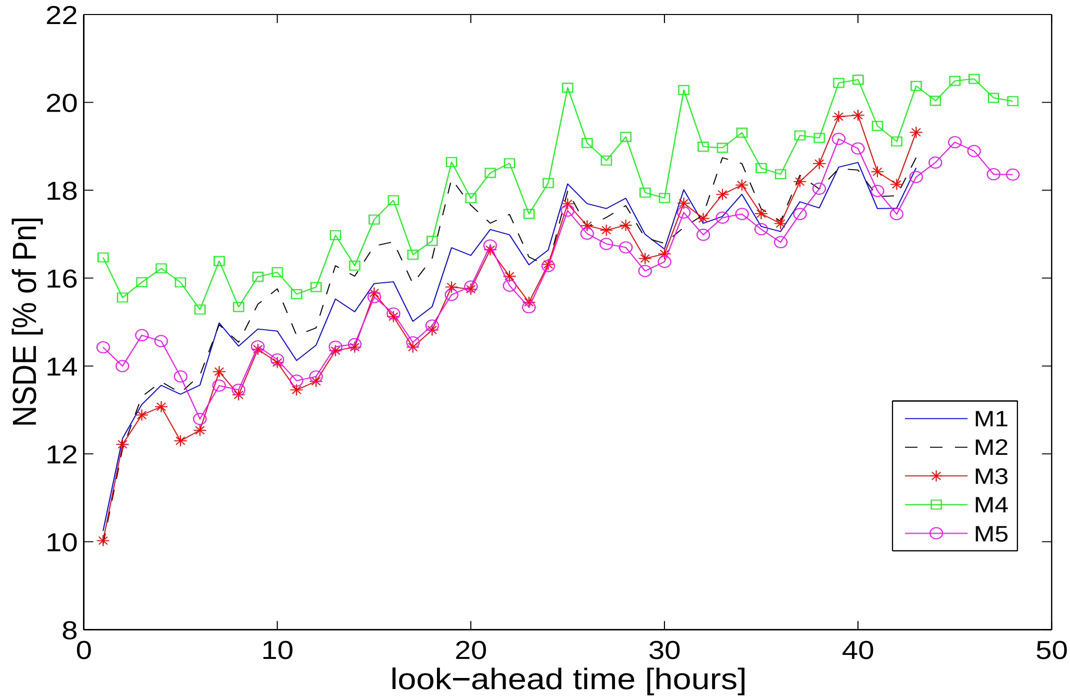


Figure 3.13: *Normalized standard deviation of the prediction error distributions as a function of the look-ahead time. Results are for Golagh.*

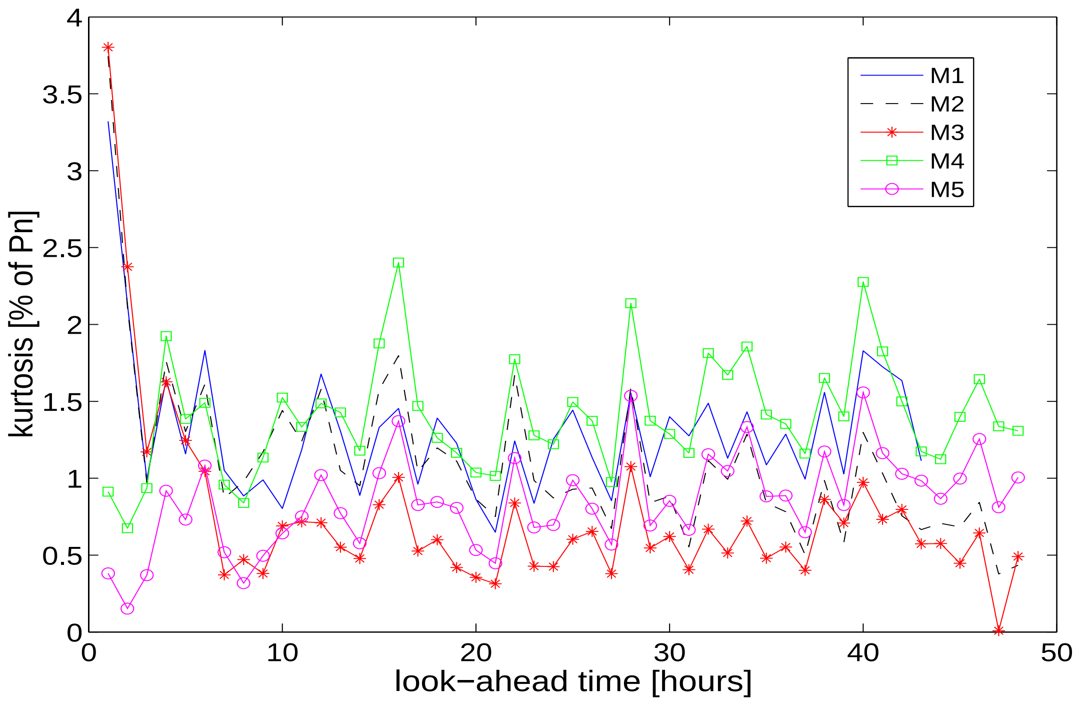


Figure 3.14: *Excess kurtosis of the prediction error distributions as a function of the look-ahead time. Results are for Tunø Knob.*

Second, the prediction horizon does not have any effect on the symmetry (or the lack of symmetry) of prediction error distributions: the skewness stays at a constant level over the whole range of prediction horizons. Note that this skewness is generally positive, due to the nonlinear and bounded shape of the power curve, allied to the fact that wind speed values are more often in a low than in a high range. This will be detailed in the following Paragraph. Note that we do not show the Figures related to the dependence of skewness with the forecast horizon since they do not present much interest.

Finally, the excess kurtosis, and hence the shape of error distributions, are only slightly influenced by the look-ahead time. For all the test cases and prediction methods, the excess kurtosis of horizon-dependent error distribution has been found to be positive, indicating that these distributions are more peaked than normal distributions and have longer tails. This is in line with the study carried out by Lange [129], showing that horizon-dependent distributions were not Gaussian. The evolution of the excess kurtosis as a function of the look-ahead time is displayed in Figure 3.14, for the Tunø Knob case-study. We do not show similar figures for the other test cases. For the first 2 horizons, statistical methods have much more peaked error distributions owing to the use of online data as input. Then, for further look-ahead times, the excess kurtosis exhibits some variations, but around a constant level, whatever the forecasting approach. This level is three times higher for M4 than for M3. A glance at the error margin plots shown in Figures D.5 (5% error margin) and D.6 (30% error margin) confirms the differences of excess kurtosis between these two approaches: even if the two methods have the same NSDE, M4 makes more very low prediction errors and at the same time makes very high prediction errors on a more regular basis.

The effect of the power curve

In Paragraph 3.5.1, the contribution of the power curve to the power forecasting errors has been described. Briefly, we have stated that it either amplifies or dampens wind speed prediction errors depending on level of predicted wind speed, and that it thus alters the shape of the wind speed error distributions. While previous studies (see [129] or [130] for instance) have only focused on the effect of the power curve on the general level of prediction error (expressed with measures), we want to go further here by basing our study on a distribution-oriented approach for better showing how the level of predicted power impacts error characteristics.

Wind speed distributions are transformed to wind power distributions through the wind-to-power conversion process. We refer to the power distribution for a given wind park as its *power climatology*[2]. Since low-wind-speed situations appear more often than periods with strong breezes, periods with low wind generation also occur on a more regular basis. Owing to the variety of wind climatologies and to the differences in wind-to-power conversion processes (due in turn to local effects, turbines siting, turbines' characteristic curves, etc.), the power climatology of each wind farm can be seen as unique. When forecasting wind power

[2]The 'capacity factor', which is a measure of the profitability of a given wind park, is actually the mean of that power climatology.

output with a given method, for a particular site and over a long period of time, one expects that the method follows the local power climatology. For instance, if power measurements tell that wind generation is 20% of the time between 0 and 10% of the installed capacity, it should also be seen that the method has forecasted power values in that particular range with approximately the same frequency. Comparing the site and prediction method climatologies relates to the comparison of the marginal distributions of the observations $q(p)$ and of the forecasts $q(\hat{p})$.

We concentrate on the Tunø Knob test case which comprises 536 forecasts over a period of four months and a half. That wind farm is used as an example all along the present Paragraph. Power measures are binned by ranges of 10% of the wind farm nominal power. A histogram depicting the frequency of occurrences of the various ranges of power production is shown in Figure 3.15. The bar for power outputs lower than 10% of P_n is much higher than the others: 37% of the measurements corresponds to low (or even null) wind generation whereas the frequencies for other bins are contained between 5 and 10%. Even though frequencies of low power outputs vary from one case-study to another (cf. Figures E.9, E.15 and E.21 in Appendix D), from 24% to 47% of the times, such a comment can be generalized.

Power climatologies for the various forecasting methods are superposed on the histogram. The frequencies of predicted power values within the different bins are plotted here for 18-hour ahead predictions (Figure 3.15). This could be done over the whole range of horizons, but one should rather differentiate the look-ahead times for verifying that the power climatology is unconditionally respected. The horizon is chosen randomly since we have not witnessed significant differences over the whole study. We will also concentrate on that particular prediction horizon in the remaining of the Paragraph, even if derived conclusions can be generalized over the whole range of look-ahead times.

Going back to the first bar which is higher than the others, one sees that all the forecasting methods do not predict very low power values ($<$10% o P_n) as often as requested. Also, there is a noteworthy variation between their frequencies of predicting this range of values. M4's frequency is higher than 30% while the one for M3 is just above 20%. Were we having a categorical vision of forecast verification that we would say M3 fails in predicting the 'very low power' event almost half of the times. Though, since we deal with a continuous variable, it is more the deviation from measurements that is important. The fact that frequencies of predicting the 'very low power' event are low tells us that all the forecasting methods tend to overestimate wind generation for low outputs (since the corresponding predictions obviously fit on some of the other bars). This explains why the methods' frequencies for the next bars are slightly above the local power climatology frequencies. A similar remark can be drawn for the bar corresponding to the power values between 90% and 100% of P_n. Since the frequencies for the methods are lower than the local power climatology frequency, we can state that all the prediction methods tend to underestimate power for high measured power values. One notices from the evaluation of the other power climatologies (Figures E.9, E.15 and E.21) that there are cases for which some prediction methods never predict more than 90% (or even 80%) of the installed capacity. This is mostly the case for wind farms that do not often generate high power outputs. Thus, the lack of data in the

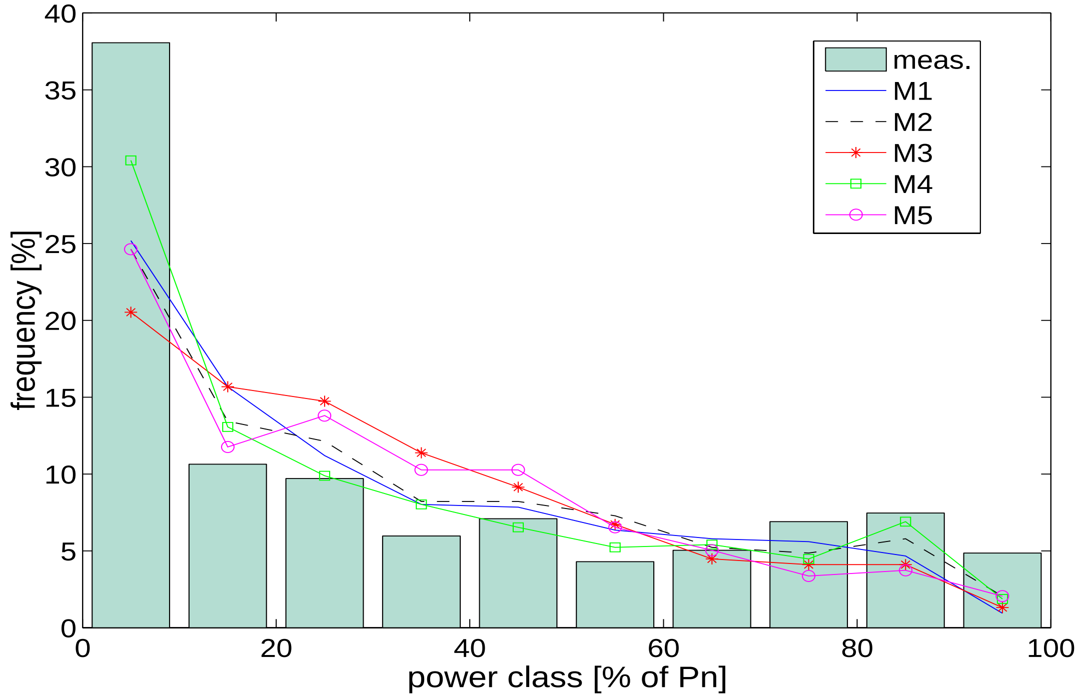


Figure 3.15: *Power climatology of Tunø Knob compared with the climatologies of the five forecasting methods. Power measurements and forecasts are sorted with bins representing 10% of P_n. Focus is given to 18-hour ahead predictions.*

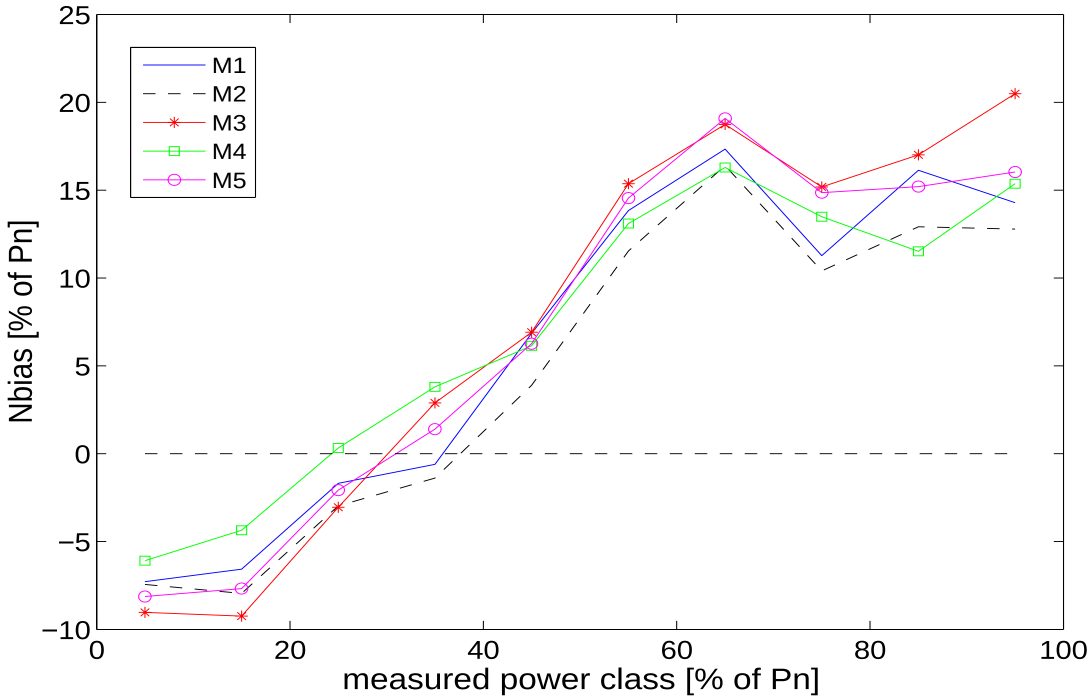


Figure 3.16: *Normalized bias of forecasting error distributions depending on power measures. Results are for Tunø Knob. Focus is given to 18-hour ahead predictions.*

high power range, combined to the fact that NWPs often underestimate wind speed, makes forecasting methods underestimating the actual power output.

This analysis is confirmed by Figure 3.16, which shows the normalized bias of prediction methods as a function of the measured power production for 18-hour ahead predictions. Such a plot relates to the idea of studying the conditional distributions of forecasts given power measures $q(\hat{p}|p)$. It is expected that these conditional distributions provide insight on the aspect of quality that is commonly referred to as *discrimination* [165]. This aspect corresponds to the ability of the forecasts to discriminate among the observations. An example of weak discrimination would be the case of a forecasting method that provides similar forecasts (say 80-85% of P_n) when actual power output varies between 80% and 100% of P_n. The bias as function of the true power output is a measure of discrimination.

In this Figure, the general behavior is similar for all the methods: the bias linearly increases as the measured power augments. It is substantially negative for low power values (down to -10%) and reaches very high levels for power values greater than 50% of P_n (up to 15-20% of the wind farm nominal power). All the forecasting approaches actually underestimate by approximately 15% on average the true power output (!) in that range of power observations. Such a behavior (maybe even worse) can also be seen for the other test cases (Figures E.16, E.10 and E.22), with a similar type of bias variations. After analyzing this Figure, one may be suspicious regarding the comment we made in Paragraph 3.4.1, in which we stated that the forecasting methods' biases were not significant for most of the test cases. The bias values depicted as a function of power measures are associated with frequencies of occurrence of these situations (given by Figure 3.15). When summing these bias values with their related weights, negative and positive values cancel out (with a large contribution of the bias for very-low measured power values) and the overall bias is obtained. To conclude, we can say that the state-of-the-art forecasting approaches do not span in a correct manner the whole range of power values: predictions are significantly concentrated in the medium-power part when the true effect lies in the more extreme parts of the power curve. They achieve a rather weak discrimination. It is a direct consequence of the bounded nature of the wind generation process.

However, one can see things the other way around by studying the conditional distributions of the measures given the forecasts $q(p|\hat{p})$. This permits to assess the *reliability* of wind power forecasts [165]. Reliability is defined as the correspondence between the mean of the observations associated to a particular forecast and that forecast. It therefore translates to studying the dependence of the systematic error to the level of the predictand. This is what is done in Figure 3.17, where the methods' normalized biases are given for each of the 10%-ranges of forecast values. Bias values exhibit significant variations over the range of possible predicted outcomes. These values are comprised between -7% and 9% of P_n, and seem to have a general trend to be positive in the lower part of the power curve and negative in its higher part. However, it does not appear possible to establish a relation like the linear one we previously observed when studying conditional distributions of the forecasts given the measures $q(\hat{p}|p)$. This is also the case for the other sites (Figure E.11, E.17 and E.23), with higher bias values for the wind farms located in semi-complex and complex terrain. In a

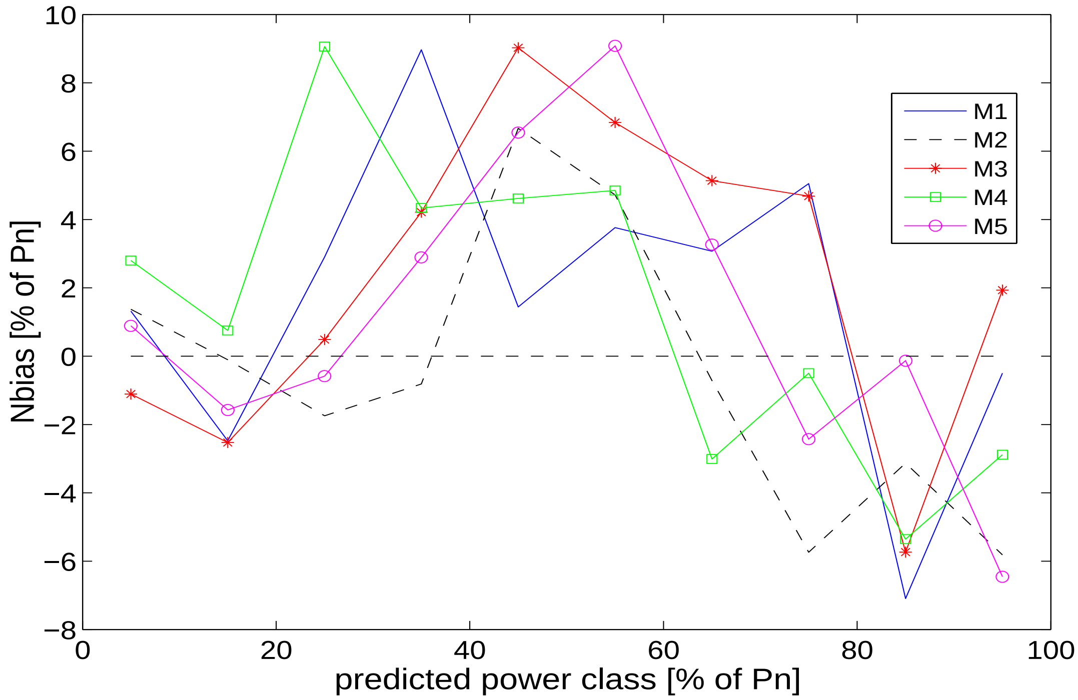


Figure 3.17: *Normalized bias of the forecasting error distributions depending on the predicted power range. Results are for Tunø Knob. Focus is given to 18-hour ahead predictions.*

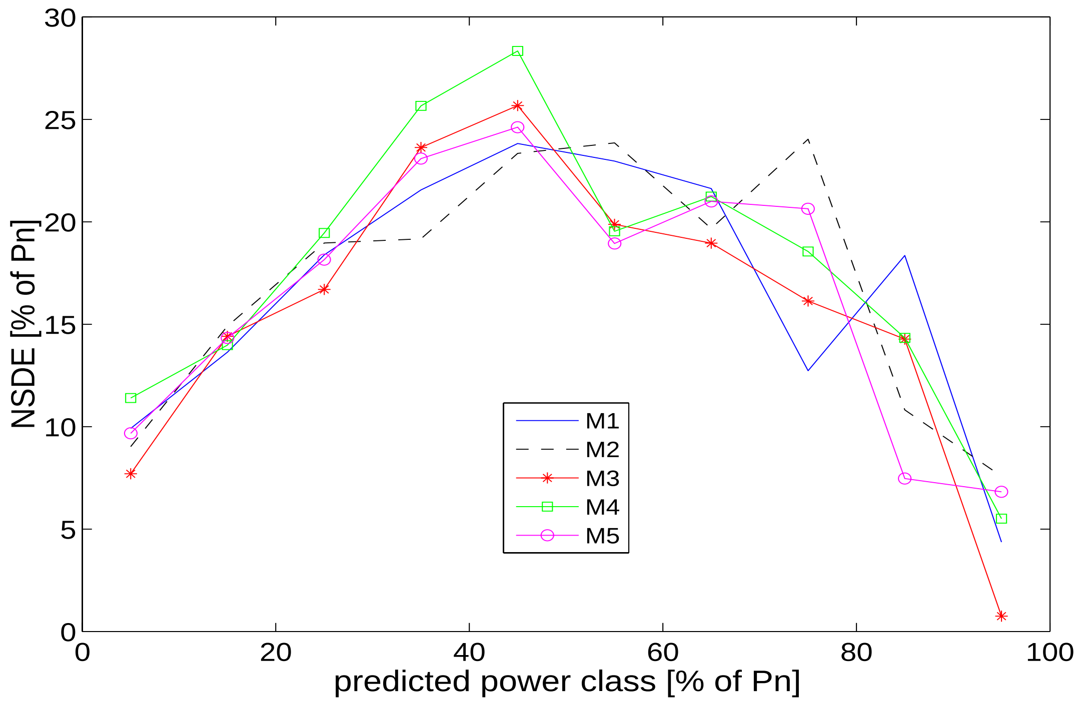


Figure 3.18: *Normalized standard deviation of forecasting error distributions depending on the predicted power range. Results are for Tunø Knob. Focus is given to 18-hour ahead predictions.*

general way, even if methods' reliability is not perfect, we cannot identify a systematic lack of reliability in certain zones of the power curve or for a given method, apart from the trend we have expressed above.

Consequently, we want to evaluate what are the variations in the shape of error distributions depending on the predictand value. First, the nonlinear and bounded nature of the energy conversion process makes that the skewness of error distributions evolves with the level of predicted power. Figure 3.19 depicts this evolution. Distributions are positively-skewed for low predicted values and then negatively skewed when these values are in the high part of the power curve. Moreover, the nonlinear process acts on both the spread and peakedness of error distributions. The spread dependence to the level of predicted power is shown in Figure 3.18, in which the spread is quantified by the standard deviation. Remember that the uncertainty of a given process is usually seen as the variability of its related observations. Then, studying the evolution of the spread of conditional distributions of the measures given the forecasts $q(p|\hat{p})$ relates to evaluating the predictand-dependent uncertainty. In parallel, excess kurtosis, as a function of predicted power, is depicted in Figure 3.20. High excess kurtosis values correspond to predicted power values close to minimum and maximum wind generation. Error distributions are highly peaked in these zones of the power curve. And, in the medium power range, slightly negative excess kurtosis values indicate that distributions are more flat than Gaussian distributions. At the same time, NSDE curves are almost symmetric with respect to the 50% power value. In the range of values related to the steep part of the power curve, the standard deviation is larger than for power values close to the power curve plateaus (say two or three times larger). Also, it can be seen that the standard deviations for these two plateaus are rather similar. The ratio between the uncertainty in the steep part of the power curve and the one in the low and high parts is approximately the same for all the prediction methods and the case studies (cf. Figure E.12, E.18 and E.24), even if the shape of the standard deviation curves slightly differs from one test case to another. This tells us that the variations of the wind power forecasting uncertainty are similar whatever the wind farm and are actually due to the wind-speed-to-power conversion process. Uncertainty levels may be higher when it is harder to predict wind speed (e.g. for complex terrain or offshore), but the way forecast uncertainty will vary as a function of the level of predicted power will be similar.

3.6 Conclusions

Evaluating the goodness of wind power point forecasts is not a trivial task indeed, even if one concentrates on their quality only, by a thorough analysis of their statistical performance. Since these predictions are issued from complex multivariate forecasting methods, it is of particular importance to focus on what makes their skill and what affects their performance. This is why we have highlighted the contributions of both the method inputs and the energy-conversion-process modeling to particular types of errors. Particularly, a part of the errors coming from meteorological forecasts e.g. phase errors cannot be removed by the state-of-the-art predictions methods, since they only consist in nonlinear transforma-

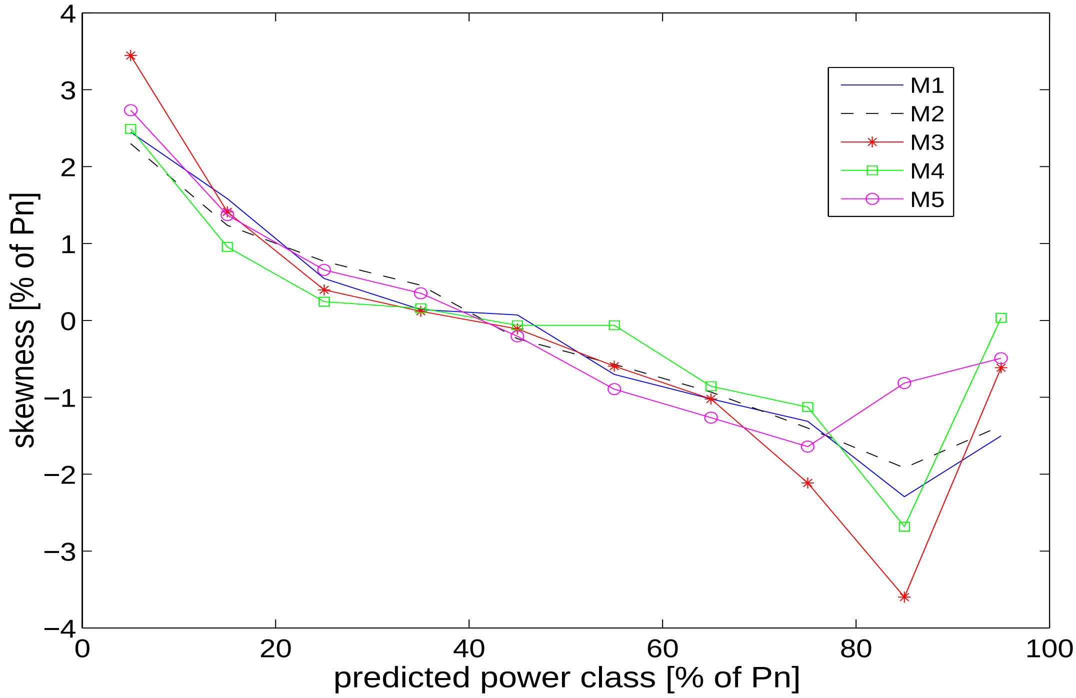


Figure 3.19: *Skewness of forecasting error distributions depending on the predicted power range. Results are for Tunø Knob. Focus is given to 18-hour ahead predictions.*

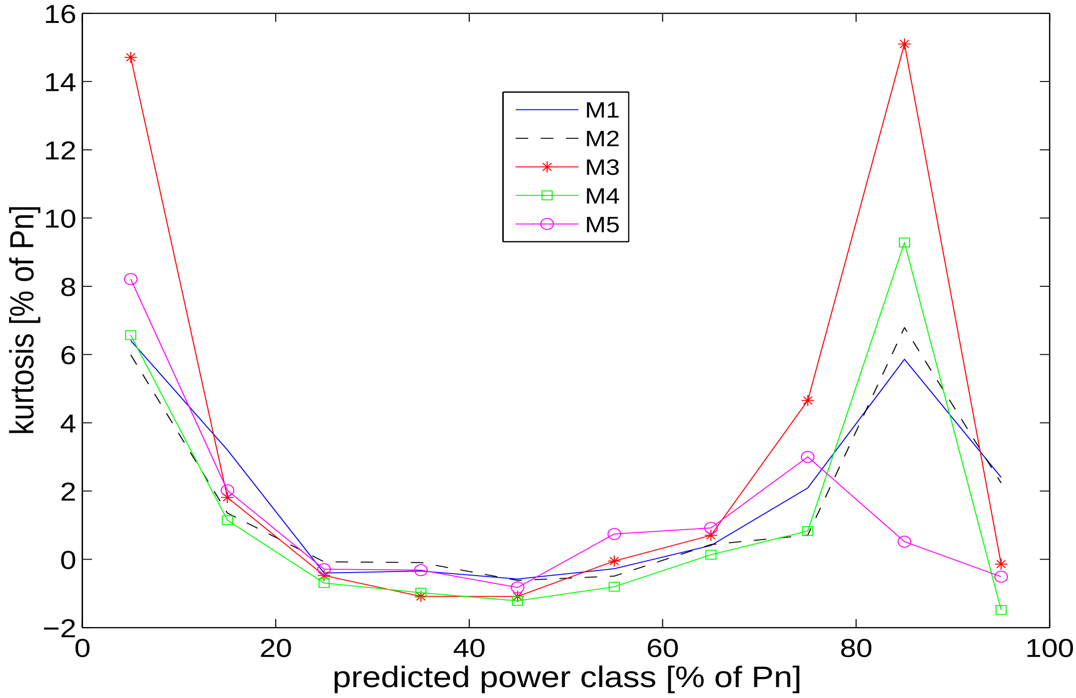


Figure 3.20: *Kurtosis of forecasting error distributions depending on the predicted power range. Results are for Tunø Knob. Focus is given to 18-hour ahead predictions.*

tions of wind speed to power, either by using physical or statistical approaches. NWPs also partly contribute to level errors, which relate to the misestimation of the magnitude of wind generation. Previous works [129, 158] already concluded on the strong contribution of meteorological predictions in the wind power forecasting errors. Improvements in the NWPs (especially of wind speed) will result in a noteworthy increase of wind generation forecasts' quality. However, we do not neglect the contribution of the energy-conversion-process modeling: we have shown there is a significant performance variability among the state-of-the-art approaches. Our opinion is that there is still room for improving these methods. In addition, we have mentioned some other factors that influence the level of performance of power predictions. Mainly, the type of terrain (offshore, flat or complex) where a wind park is located has an influence on the wind farm generation predictability.

We have had the chance to thoroughly compare the performance of the main state-of-the-art wind power forecasting approaches: this is the first forecasting exercise of that type. Moreover we have had the possibility to work with several sites in Europe, including an offshore wind farm in Denmark. Far from being skeptic about the performance of such or such method, our opinion is rather that they all have their qualities and drawbacks. Therefore, combining their forecasts for obtaining the best of each method appears in our sense as an option for ensuring the level of performance of wind power predictions. However, such a combination should also be envisaged with the use of various NWPs issued from different meteorological models, since forecast combination becomes more and more beneficial as one integrates various (and if possible independent) sources of information [93].

A general framework for the verification of point predictions of wind generation have been introduced in the present Chapter. Such a framework includes both a measure-oriented and a distribution-oriented approach. This distribution-oriented approach has been applied here for the first time and allowed us to derive new conclusions on some aspects of forecasting methods' quality following the Murphy-Winkler framework. A generic methodology was introduced, which consists in studying the influence of some variables on the moments of error distributions. This methodology has been consequently utilized for underlining the main characteristics of the forecasting errors. More precisely, our concern was about the effects of the look-ahead time and of the power curve. We have shown that the prediction uncertainty increases as the lead time gets further, even though it has no effect on both the systematic part of the error and the shape of error distributions. Also, the level of prediction uncertainty is 2 or 3 times higher for medium range power values than for low and high ones. And, error distributions are highly skewed and peaked in the more extreme parts of the power curve. A crucial conclusion is that all these characteristics are shared by the various state-of-the-art point prediction methods. One should notice though that even if we have not studied the potential effects of some other explanatory variables (e.g. wind direction), the same methodology can be used for that purpose in the future.

4

Estimation and Evaluation of Prediction Intervals of Wind Power

Abstract

In this central Chapter is developed and evaluated a generic method appropriate for the estimation of prediction intervals of wind generation. In order to avoid a restrictive assumption on the shape of prediction error distributions, we focus on an empirical and distribution-free approach. Also, a fuzzy inference model is introduced in order to integrate the expertise on the characteristics of prediction errors for providing conditional interval forecasts. The proposed method can be considered for providing full predictive distributions of wind generation. In parallel, the required properties of probabilistic predictors are given, followed by the description of a non-parametric framework for the verification of wind power probabilistic forecasts in the form of quantiles or intervals. This framework is consequently used for evaluating and analyzing the skill of the proposed approach. This one proves to be reliable and it is shown how its resolution may be enhanced by using the forecaster's expertise. Finally, some guidelines are given for the application of the method to online prediction exercises.

4.1 Introduction

PREDICTIONS of wind power output are traditionally provided in the form of point forecasts. They have the advantage of being easily understandable because this single number is expected to tell everything about future power generation. Today, a major part of the research efforts on wind power forecasting still focuses on point prediction only, with the aim of assimilating more and more observations in the models or refining the resolution of physical models for better representing wind fields at the very local scale for instance [76]. These efforts may lead to a significant decrease of the level of prediction error. However, even by better understanding and modeling both the meteorological and power conversion processes, there will always be an inherent and irreducible uncertainty in every prediction. This epistemic uncertainty corresponds to the incomplete knowledge one has of the processes that influence future events [207].

Therefore, in complement to point forecasts of wind generation in the next hours, of major importance is to provide means for assessing online the accuracy of these predictions. Error measures described in Chapter 3 only provide an assessment of a given point forecasting method performance over a large period of time. They tell what is the historic performance of the method, but they cannot give an estimation of the uncertainty related to a given prediction. In practice today, uncertainty is expressed in the form of *interval forecasts* that are associated to wind power point predictions. An interval forecast is a range of values within which the actual outcome is expected to lie with a pre-assigned probability. Such intervals are expected to be valuable for developing alternative strategies for the management or the trading of wind power generation. In a general manner, they are necessary for optimizing the decision-making process related to the use of wind power forecasts.

In the present Chapter, our aim is to develop an appropriate method for estimating prediction intervals of wind generation, which can be applied to any state-of-the-art point forecasting method. For that purpose we exploit the characterization of prediction errors carried out in Chapter 3. Since it was shown that these characteristics were shared by all point forecasting approaches (either of the physical or of the statistical type), we use that aspect for developing a generic method. Also, focus is given to the development of a distribution-free approach, suitable for nonstationary, nonlinear and bounded processes. It is explained why such an approach can be applied to any point prediction method. Also, the way it can be straightforwardly applied for the wind generation prediction problem is detailed.

The second part of the Chapter is devoted to the assessment of the quality[1] of the resulting prediction intervals. Interval forecasts have attracted attention only recently and the assessment of their quality is more complicated than for the case of point predictions. A non-parametric framework for carrying out this performance assessment is introduced. Then, an evaluation of the derived method quality is given by applying the proposed framework to various case-studies and on several state-of-the-art point prediction methods. In complement, we highlight the influence of the method parameters on some particular aspects of the quality of prediction intervals. This allows us to derive guidelines for the application of the developed approach to online forecasting exercises.

4.2 Different types of statistical intervals

Often, it is needed to draw conclusions on the characteristics of a process from a limited amount of available knowledge. Statistics are usually calculated from limited samples and

[1] Quality is still related here to a statistical performance, following Murphy's terminology [165].

may prove to be uncertain. Perhaps the most illustrative example is that of public opinion polls, for which panels composed by few hundreds or thousands people are used to tell what is the average opinion of millions of people in a country. Since this population sampling induces uncertainty, calculated statistics are therefore associated with estimates of their accuracy, in the form of intervals. Depending on the type of decision to make from a given statistic, several types of related intervals may be defined. For an introduction to these different types of statistical intervals, we refer to [87].

Our concern here is about the accuracy of point forecasts. Two types of intervals appear to be relevant for that purpose: *confidence intervals* and *prediction intervals*. There is a fundamental difference between these two. Given a sample population $\{p_t\}_{t=1,\dots,T}$, a confidence interval is meant for giving a measure of confidence on the estimate $\hat{\theta}(\{p_t\}_{t=1,\dots,T})$ of a parameter θ for the whole population, whereas a prediction interval is meant for giving the range of values within which the next randomly selected individual p_t $(t > T)$ from that population may lie, with a certain degree of confidence.

In order to describe how this can be translated to the forecasting problem, let us consider the case of a statistical model g designed for 1-step ahead prediction. The parameters $\mathbf{w}$ of that statistical model are estimated over a training set consisting of N_L pairs $\{\mathbf{y}_t, p_t\}_{t=1,\dots,N_L}$, where $\mathbf{y}_t$ is a vector containing past values of p (up to time $t-1$) plus eventually past values and forecasts of explanatory variables, and p_t is the observation at time t. Following the notations used in Equation (2.10):

$$\mathbf{y}_t = (p_{t-1}, p_{t-2}, \dots, p_{t-l}, \mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \dots, \mathbf{x}_{t-m}, \hat{\mathbf{x}}_{t/t-1}). \tag{4.1}$$

These data pairs are assumed to be generated according to the following process:

$$p_t = g(\mathbf{y}_t, \mathbf{w}) + e_t, \tag{4.2}$$

where $\mathbf{w}$ are the parameters of the chosen model g and e_t is a zero mean random variable. For $t > N_L$ the trained model $g(\mathbf{y}_t, \hat{\mathbf{w}})$ (with $\hat{\mathbf{w}}$ estimated by considering a quadratic loss function) will then produce at time t a forecast $\hat{p}_{t+1/t}$ that is an estimate of the mean $\bar{p}_{t+1}$ of the target distribution F^p_{t+1} at time $t+1$, given $\mathbf{y}_{t+1}$ (cf. Paragraph 2.6.1). The uncertainty in the estimate of the mean of the target distribution partly comes from the fact that one uses a finite sample for training the model, which consists in an incomplete knowledge of the true process. In addition, observations may integrate a noise component coming from data acquisition devices. This uncertainty is also due to the choice of the model that may not reflect the true behavior of the process and to the way the model parameters $\mathbf{w}$ are estimated. A confidence interval associated to $\hat{p}_{t+1/t}$ is hence a measure of the confidence in the estimation of the mean $\bar{p}_{t+1}$ of the target distribution. Since $\hat{p}_{t+1/t}$ is not an estimate of the true outcome p_{t+1}, this interval does not give the confidence in the estimation of the true effect p_{t+1}.

Alternatively, a prediction interval associated to a point forecast is a measure of the accuracy of that point forecast with respect to the true outcome p_{t+1}, by giving a range of potential values. A prediction interval necessarily encloses the corresponding confidence interval [92]. Figure 4.1 is an illustrative example of the difference between confidence and prediction intervals. The solid curve represents a probability distribution of expected wind generation at time $t+1$. The two dashed vertical lines correspond respectively to the mean $\bar{p}_{t+1}$ of that distribution (bold) and to a wind power point forecast $\hat{p}_{t+1/t}$, which is an estimate at time t of that mean. The dark shaded area stands for the confidence interval associated to $\hat{p}_{t+1/t}$, while the light shaded area is for the interval forecast. The solid vertical line gives the observed power value at time $t+1$.

Although we have taken the example of a statistical model designed for 1-step ahead predictions, this reasoning can be extended to the case of other types of models (i.e. multi-step ahead models and physical models): they all aim at estimating a particular point of the target distribution, which is its mean in most of the cases. Then, a confidence interval will always correspond to the confidence in the estimate of the expected outcome, whereas a prediction interval associated to a point forecast will give the accuracy of that estimate with respect to the true outcome. Because we are mostly interested in that second type of uncertainty assessment we will turn our attention to prediction intervals from now on.

Formally, a *prediction interval* $\hat{\mathrm{I}}^{(\alpha)}_{t+k/t}$, alternatively called interval forecast, estimated at time t for lead time $t+k$, is a range of values within which the true effect p_{t+k} is expected to lie with a certain probability $(1-\alpha)$, denoting its *nominal coverage rate*:

$$\mathrm{P}\left(p_{t+k} \in \hat{\mathrm{I}}^{(\alpha)}_{t+k/t}\right) = \mathrm{P}\left(p_{t+k} \in [\hat{L}^{(\alpha)}_{t+k/t}, \hat{U}^{(\alpha)}_{t+k/t}]\right) = 1-\alpha. \tag{4.3}$$

An interval forecast is then specified by its lower and upper bounds, denoted by $\hat{L}^{(\alpha)}_{t+k/t}$ and $\hat{U}^{(\alpha)}_{t+k/t}$ respectively. Note that we will prefer the term 'nominal coverage rate' (or alternatively 'degree of confidence') instead of the widely used 'confidence level' term when referring to the probability associated to interval forecasts, so that the reader does not confuse them with the more classical confidence intervals.

Most of the times prediction intervals are central prediction intervals: there is the same probability $(\alpha/2)$ for a non-covered outcome to be above or below the interval bounds. Then, these bounds correspond to the quantiles[2] with proportion $(\alpha/2)$ and $(1-\alpha/2)$ of the *predictive distribution* $\hat{F}^{p}_{t+k/t}$ of future events:

$$\hat{L}^{(\alpha)}_{t+k/t} = \hat{r}^{(\alpha/2)}_{t+k/t}, \quad \mathrm{P}\left(p_{t+k} < \hat{L}^{(\alpha)}_{t+k/t}\right) = \alpha/2, \tag{4.4}$$

$$\hat{U}^{(\alpha)}_{t+k/t} = \hat{r}^{(1-\alpha/2)}_{t+k/t}, \quad \mathrm{P}\left(p_{t+k} < \hat{U}^{(\alpha)}_{t+k/t}\right) = 1-\alpha/2. \tag{4.5}$$

Central prediction intervals are hence centered on the median of the predictive distribution $\hat{F}^{p}_{t+k/t}$.

Traditionally, emphasis is given in the literature to the computation of prediction intervals for a Normal distribution, or more generally for a symmetric target distribution [35, 54, 59, 87, 92]. Thus, estimated prediction intervals are centered on the point predic-

[2]The quantile $r^{(\alpha)}$ with proportion α of the distribution F^X of a random variable X is defined as the value x such that $\mathrm{P}(X \leq x) = \alpha$.

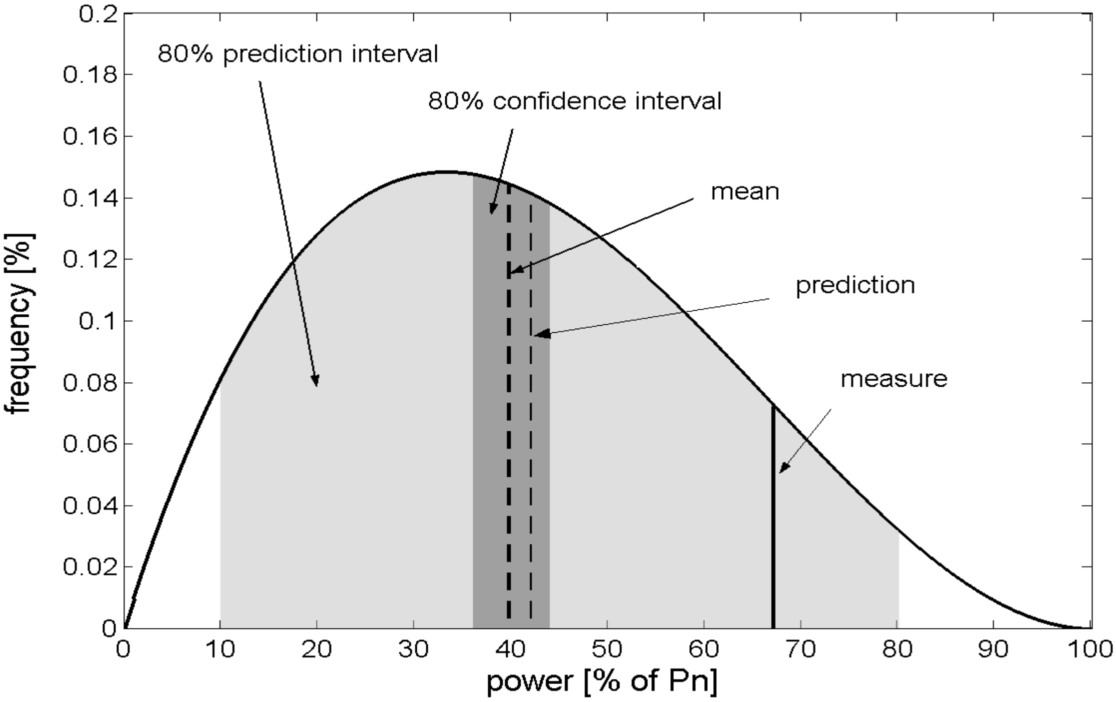


Figure 4.1: *Illustrative example of the difference between confidence (light shaded area) and prediction (dark shaded area) intervals. The confidence interval is a measure of the confidence in our estimate $\hat{p}_{t+1/t}$ of the expectation $\bar{p}_{t+1}$, whereas the prediction interval is related to the accuracy of the point forecast $\hat{p}_{t+1/t}$ with respect to the true effect p_{t+1}.*

tion itself and give the equally probable upward and downward margins in which the future outcome may lie. Due to symmetry, the mean and median of these target distributions are equal. Moreover, the upper and lower sides of the intervals have the same size. Therefore, whatever the nominal coverage rate, the point forecast is included in the interval forecast it is associated to. For a nonlinear and bounded process such as wind generation, probability distributions of future power output exhibit some skewness. For these asymmetric distributions, the median may significantly differ from the mean, and thus central prediction intervals (for rather low nominal coverage rate) may not even cover the point forecast value. This is why interval forecasts can be alternatively constructed in the form of intervals $\hat{\mathrm{I}}\mathrm{c}^{(\alpha)}$ centered on the point forecast itself

$$\hat{\mathrm{I}}\mathrm{c}^{(\alpha)}(\hat{p}_{t+k/t}) \;=\; [\hat{L}c^{(\alpha)}_{t+k/t}, \hat{U}c^{(\alpha)}_{t+k/t}], \tag{4.6}$$

as equally probable positive and negative margins in which the actual outcome may lie, for a given nominal coverage rate $(1-\alpha)$:

$$\mathrm{P}\left(p_{t+k} \in [\hat{L}c^{(\alpha)}_{t+k/t}, \hat{p}_{t+k/t}]\right) \;=\; \mathrm{P}\left(p_{t+k} \in [\hat{p}_{t+k/t}, \hat{U}c^{(\alpha)}_{t+k/t}]\right) \;=\; (1-\alpha)/2. \tag{4.7}$$

Such a type of intervals will be referred to as prediction-centered interval forecasts. They consist in separately modeling two different probability distributions, which are the ones for the positive and negative errors respectively. Then, one notes that even if $\hat{L}c^{(\alpha)}_{t+k/t}$ and $\hat{U}c^{(\alpha)}_{t+k/t}$ are quantiles of the whole predictive distribution, we do not know the proportions they correspond to. Since we aim at directing our work towards a probabilistic view of wind power forecasting, our preference goes to central prediction intervals, since they model the target distribution F^{p}_{t+k} as a whole. Consequently, by specifying a nominal coverage rate $(1-\alpha)$, we will then determine the quantiles with proportions $(\alpha/2)$ and $(1-\alpha/2)$ of $\hat{F}^{p}_{t+k/t}$.

Finally, as for point forecasts, prediction intervals issued at time t are produced from the information set Φ_t that gathers the available information up to that time. Therefore, even if we use the notation $\hat{\mathrm{I}}^{(\alpha)}_{t+k/t}$ in the following of the Chapter, it actually stands for $\hat{\mathrm{I}}^{(\alpha)}_{t+k/t}(\Phi_t)$.

4.3 Basic parametric approaches for prediction interval estimation

An approach is said to be *parametric* if there is an underlying assumption on the distribution one tries to model. Inversely, a *non-parametric* (or *distribution-free*) approach does not rely on such an assumption.

The simplest parametric approach for estimating prediction intervals is the method proposed by Box and Jenkins [21]. It follows the assumption that for a model (such as the multivariate one given by Equation (2.10)) the $\{e_t\}$ sequence is independent and identically distributed Gaussian with zero mean and variance $\sigma_e^2 < \infty$, $e_t \overset{\text{i.i.d.}}{\sim} N(0,\sigma_e^2)$. By using an estimate $\hat{\sigma}_e^2$ of the variance, the one-step ahead interval forecasts with nominal coverage rate $(1-\alpha)$ are such that:

$$\hat{\mathrm{I}}^{(\alpha)}_{t+1/t} \;=\; \left[\hat{p}_{t+1/t} - z_{\alpha/2}.\hat{\sigma}_e^2,\, \hat{p}_{t+1/t} + z_{1-\alpha/2}.\hat{\sigma}_e^2\right], \tag{4.8}$$

where $z_{\alpha/2}$ and $z_{1-\alpha/2}$ are the quantiles with proportion $(\alpha/2)$ an $(1-\alpha/2)$ of the standard Normal distribution $\mathcal{N}(0,1)$. Then, k-step ahead interval forecasts can be produced similarly, by considering estimates of the variance $\hat{\sigma}^2_{e,k}$ of the random shock for k-step ahead point predictions

$$\hat{\mathrm{I}}^{(\alpha)}_{t+k/t} \;=\; \left[\hat{p}_{t+k/t} - z_{\alpha/2}.\hat{\sigma}_{e,k}^2,\, \hat{p}_{t+k/t} + z_{1-\alpha/2}.\hat{\sigma}_{e,k}^2\right]. \tag{4.9}$$

By assuming that errors in consecutive step-ahead forecasts are mutually independent and distributed Gaussian with zero mean and constant variance, Makridakis et al. [147] proposed to compute the prediction intervals of Equation (4.9) with an 'approximate' formula for $\hat{\sigma}^2_{e,k}$, which states that

$$\hat{\sigma}^2_{e,k} \;=\; k.\hat{\sigma}^2_e. \tag{4.10}$$

It has been shown by Koehler [120] that there was no theoretical justification for Equation (4.10), and that the assumptions mentioned above about the error process could only be true for a random walk model. Therefore, using such a simple approximation of the k-step variance would yield very inadequate results.

Instead of relying on approximate formulae for estimating the k-step ahead variance, another possibility is to use the historical performance of the predictor:

$$\hat{\sigma}^2_{e,k} = \text{SDE}^2(k), \tag{4.11}$$

where SDE(k) is the standard deviation of the k-step ahead forecasting errors over a given evaluation period, as defined by Equation (3.8). Alternatively, one may consider the use of time-adaptive statistics for estimating recent SDEs of the prediction method.

Intervals estimated from Equation (4.9) are symmetric around the point prediction. Even if the Gaussian assumption does not hold, the Box-Jenkins method is often followed in practice. When considering nonlinear (and chaotic) time-series such a basic estimation of prediction interval bounds will lead to poor results [123]. This has recently been illustrated for the specific case of wind power forecasting [183].

We know that the nonlinearity aspect is due to the energy conversion process. When thoroughly studying conditional distributions of wind speed prediction errors (given predicted wind speed), Lange [130] noticed that they could be modeled with Gaussian distributions whose standard deviations equal the standard deviation of unconditional error distributions of wind speed forecasts. In parallel, he proposed a model based on the local derivative of the wind park power curve for describing the way wind speed intervals would be mapped to power intervals. He used that model for estimating the standard deviation of conditional distributions of power prediction errors given predicted power output. Then, a Gaussian assumption was considered for calculating the $(1 - \alpha)$ prediction interval of wind generation [129]. The first shortcoming of this approach is that power prediction errors are assumed to be distributed Gaussian. This could be easily overcome by estimating $(1 - \alpha)$ intervals on the error distributions of wind speed and by passing these intervals through the wind park power curve for obtaining non-Gaussian intervals of wind power. This idea has already been proposed by Brown et al. [24]. The second shortcoming is that the method does not account for the modeling error itself, owing to the spatial refinement of the NWPs or to the model used for the power curve. Also, such a method is limited for application to physical methods only since it requires an explicit power curve. Finally, standard deviations of wind speed error distributions are not easy to obtain since wind forecasts and related measurements are often not available at the level of a wind farm. They may be provided as a guess by meteorological offices based on their expertise, but it is unlikely that resulting prediction intervals would be accurate.

The nonlinear and bounded nature of the wind generation process is taken into account by the method proposed by Luig et al. [142], which is based on modeling predictive distributions of power output using β-distributions. Such distributions are bounded between 0 and 1 (like is normalized power production) and their shape is controlled by two parameters α and β. These two parameters are a function of the mean and the variance of the distribution. Luig et al. proposed to set the mean of predictive β-distributions equal to point forecast values given by a power prediction approach. In parallel, the variance of these distribution is determined from a study of the historical performance of the considered prediction method. Different estimates of the variance are considered depending on the range of predicted power (i.e. four ranges in this case). Central prediction intervals are provided consequently by quoting quantiles of estimated predictive distributions. This approach is expected to offer a significant improvement against intervals produced from a Gaussian assumption. However, considering only certain variance estimates for some ranges of predicted power values does not reflect the continuous variability of the power prediction uncertainty, as described in Paragraph 3.5.2. Moreover, the choice of β-distributions is also a restrictive assumption on the predictive distributions of wind generation.

4.4 Development of a distribution-free approach appropriate for non-linear and bounded processes

When it is not possible to use theoretical formulae, and in the case for which the hypothesis that prediction errors follow a known distribution appears to be a weak assumption, an alternative solution is to develop *non-parametric* approaches for the estimation of predictive distributions or interval forecasts [41]. More generally, distribution-free approaches are appealing since they are not related to any assumption concerning the error-generating process, i.e. to a particular model. Therefore, they are suitable for estimating the uncertainty of different types of forecasting methods either of the statistical or of the physical types. This is also valid if forecasts are the results of some combination procedure [217].

Quantile regression is a family of non-parametric methods that aim at estimating predictive quantiles for a given proportion. It has recently been considered for producing interval forecasts of wind generation [22, 171] in two different manners. Nielsen et al. [171] proposed a quantile regression method that uses as input point forecasts produced by a state-of-the-art forecasting method, plus some other explanatory variables e.g. the wind speed and direction forecasts that were previously utilized by the point prediction tool. That method can hence be considered for application to already installed point prediction methods in order to associate point forecasts with an estimation of their accuracy. Alternatively, Bremnes [22] developed a slightly different approach, which is based on linear quantile regression with only NWPs as input. This approach has the advantage that one then avoids the point prediction step for producing interval forecasts of wind generation. An important shortcoming of quantile regression approaches is that a specific model needs to be set-up and trained for every quantile of the predictive distribution to be estimated. Therefore, one already has to consider two different models for estimating a single prediction interval. And, for having an adequate estimation of complete predictive distributions, say by forecasting quantiles for every 0.05 proportion, this would lead to 19 models (!). Nielsen et al. [171] pointed out that since models are independently trained, they may result in inconsistent results in certain situations e.g. crossing quantiles. This is not desirable from both a theoretical and practical point of view. Moreover, these models are site-dependent: for each new wind farm they are applied to, a dataset must be collected in order to estimate their parameters through a training process.

In the present work, our aim is to propose not only a non-parametric approach, but also an approach that can be utilized for easily estimating multiple power prediction intervals (and thus several quantiles) at once. Consequently, the target method has to directly construct the predictive distribution of wind generation at once — this was one of the conclusions by Nielsen et al. [171] for avoiding the crossing-quantile problem. This is possible if one considers *empirical* approaches such as the one developed in the following. In a first stage, we introduce the main assumptions related to empirical approaches for prediction interval estimation, as well as the underlying methodology. Then, our contribution is to propose an upgrade of the empirical methods introduced in the literature, which is appropriate for non-linear and bounded processes such as wind generation. Paragraph 4.4.2 describes the classification of forecast conditions related to different characteristics of prediction error distributions. The fuzzy inference model developed in Paragraph 4.4.3 provides conditional distributions of prediction errors as a function of forecast conditions, in the form of combined probability distributions. By dressing point predictions with the estimated conditional distributions of prediction errors, one obtains predictive distributions of wind generations. Finally, two approaches for the combination of empirical distribution functions are given in Paragraph 4.4.4.

4.4.1 Hypothesis and development of empirical methods for prediction interval estimation

In general, as 'empirical' are characterized methods that origin from knowledge or experience. Here, the empirical nature of the prediction interval estimation method stands for the fact that interval forecasts are produced from the witnessed behavior of the point forecasting method it is applied to. The behavior of the point prediction approach is characterized by its recent performance.

The development and application of empirical-type approaches for prediction interval estimation can be traced back to the works by Williams and Goodman [235]. The authors fitted a regressive model for producing 18-step ahead forecasts of the number of phone lines in service over a dataset consisting of 169 data points, and envisaged to associate them with an estimation of their accuracy. Therefore, they estimated prediction intervals with the method described hereafter, by assuming that future forecast errors would be distributed in the same way than the recent ones. They noticed that prediction errors were not Normally distributed — they actually seemed to follow a Γ-distribution. And, despite the rather limited dataset, they showed that this basic empirical approach was much more efficient than usual Box-Jenkins methods for estimating prediction intervals, for various degrees of confidence. The Williams-Goodman method has been applied (with minor changes) on some other forecasting exercises for which prediction errors proved to be not-Normally distributed [106]. More particularly, Alves da Silva and Moulin [3] used a similar method for estimating prediction intervals associated to point forecasts produced with a neural-network-based method, for the short-term load forecasting problem. The authors compared the empirical interval forecasts with two other approaches, namely 'error output' (which is based on a second neural-network model trained for estimating the prediction error of the first neural network) and 'multilinear regression' (which is based on a regressive model with variables the output of the hidden layer neurons and coefficients the weights of the output neuron for estimating the prediction error variance — intervals are consequently computed following the Box-Jenkins method). Conclusions of the study were in favor of the use of the empirical approach.

The first step before computing prediction intervals is to collect the prediction errors the method made in the past. The intervals that are going to be computed will rely on the most recent information on the method's performance. For that purpose, a window in the past (a certain number of hours) is defined and used as a sliding window for storing the errors. The size n of this window determines the size of the samples of errors. At time t, a separate sample $S_{t,k}$ is defined for each prediction horizon k (i.e. for 1-hour ahead, 2-hour ahead, and so on) since we have shown that prediction uncertainty significantly varies with the look-ahead time. The collected errors are the most recent ones at a given time: when the actual measured wind power is known, that value is compared with all the past predictions made for that time. Using the most recent information on a given method performance for estimating future uncertainty is motivated by the non-stationary aspect of wind power prediction errors (cf. discussion in Chapter 3). Write $\Omega_{t,k}$ the set of prediction errors associated to k-step ahead point predictions up to time t:

$$\Omega_{t,k} = \{\epsilon_{t-i+k/t-i},\, i \in \mathbb{N},\, i \geq k\}, \tag{4.12}$$

where $\epsilon_{t-i+k/t-i}$ is the normalized prediction error related to the point forecast $\hat{p}_{t-i+k/t-i}$. Since the wind generation process is bounded, we will hereafter only deal with normalized errors and predicted values (both normalized by P_n). Straightforwardly, by renumbering the elements of $\Omega_{t,k}$, an error sample $S_{t,k}$ containing the last n k-step ahead point prediction errors at time t consists in

$$S_{t,k} = \{\epsilon_i \in \Omega_{t,k},\, i = 1, \ldots, n\}. \tag{4.13}$$

The *empirical distribution function* $\hat{F}^{\epsilon}_{t,k}$ of errors, at time t and for horizon k, is defined as the discrete distribution that puts probability $1/n$ on each element of $S_{t,k}$. It can be shown that $\hat{F}^{\epsilon}_{t,k}$ is the non-parametric maximum likelihood estimate of the true distribution function of errors $F^{\epsilon}_{t,k}$ (see [61], p. 310). Consequently, any parameter $\hat{\theta}(\hat{F}^{\epsilon}_{t,k})$ estimated from $\hat{F}^{\epsilon}_{t,k}$ is the non-parametric maximum likelihood estimate of the parameter $\theta(F^{\epsilon}_{t,k})$. For practical use, we introduce the cumulative distribution function $\hat{G}^{\epsilon}_{t,k}(\epsilon)$, which gives the fraction of errors less than or equal to ϵ

$$\hat{G}^{\epsilon}_{t,k}(\epsilon) = \frac{1}{n}\, \#\{\epsilon_i \in S_{t,k} \,|\, \epsilon_i \leq \epsilon\}. \tag{4.14}$$

The underlying assumption of the empirical approach is that future uncertainty can be expressed from the recently witnessed behavior of the point prediction method. This means that we consider here that the empirical distribution function of errors $\hat{F}^{\epsilon}_{t,k}$ can be seen as an estimate of the distribution of errors associated to the point forecast $\hat{p}_{t+k/t}$. Therefore, an empirical predictive distribution $\hat{F}^{p}_{t+k/t}$ of wind power output at lead time $t+k$ can be constructed as following:

$$\hat{F}^{p}_{t+k/t} \rightarrow \{\hat{p}_{t+k/t} + \epsilon_i,\, \epsilon_i \in S_{t,k}\}, \tag{4.15}$$

with an equal probability $1/n$ associated to each element of $\hat{F}^{p}_{t+k/t}$.

Since the bounds of the central prediction interval $\hat{\text{I}}^{(\alpha)}_{t+k/t}$ with nominal coverage rate $(1-\alpha)$ are defined as the quantiles with proportion $(\alpha/2)$ and $(1-\alpha/2)$ of the predictive distribution $\hat{F}^{p}_{t+k/t}$, they are given by:

$$\hat{L}^{(\alpha)}_{t+k/t} = \hat{p}_{t+k/t} + \hat{G}^{\epsilon}_{t,k}{}^{-1}(\alpha/2), \tag{4.16}$$

$$\hat{U}^{(\alpha)}_{t+k/t} = \hat{p}_{t+k/t} + \hat{G}^{\epsilon}_{t,k}{}^{-1}(1-\alpha/2). \tag{4.17}$$

Such a construction of the predictive distribution $\hat{F}^{p}_{t+k/t}$ of wind generation from recent performance implicitly assumes the *representativeness* of the sample data. Actually, this hypothesis cannot be completely exact and then the prediction intervals may only provide a lower bound on the real forecast uncertainty. Note that parametric interval estimation methods described in Section 4.3 also assume that near-future uncertainty will be like the historical one, since estimates of the error variances are based on past performance of the considered prediction method. Secondly, it is implicitly assumed that the sample is a *random sample*, that we do not apply any selection procedure that will then introduce a bias in the uncertainty estimation. This assumption is also not completely respected for the case of k-step wind power forecasting since consecutive prediction errors may be correlated[3] [170]. However, we will see that breaking this assumption will not have a significant influence on the performance of prediction intervals of wind generation.

4.4.2 Classification of forecast conditions

When predicting nonlinear processes, it is of common knowledge that the shape of the prediction error distributions evolves as a function of the value of the variable of interest [20, 101]. For the specific case of wind power forecasting, there may also be other variables that have an impact on the characteristics of forecast error distributions. We will refer to these variables as *influential variables*. They obviously include predicted power but they may also include forecast wind speed and direction, and eventually some other explanatory variables that are expected to have an influence on the characteristics of prediction error distributions. However, even by applying an empirical approach such as the one presented in the previous Paragraph, prediction intervals will be estimated in the same way whatever

[3] Actually, there exists a correlation between prediction errors for successive look-ahead times i.e. between $e_{t+k/t}$ and $e_{t+k+i/t}$, $i > 0$, as well as between predictions for the same look-ahead time but issued at consecutive time origins i.e. between $e_{t+k/t}$ and $e_{t+k+i/t+i}$, $i > 0$. Here, our concern is mainly about the first type of correlation, since interval forecasts are estimated independently for each prediction horizon.

the level of influential variables: they actually are *unconditional* interval forecasts. It is unlikely that samples of prediction errors would be representative of the current — and thus conditional — uncertainty. An illustrative example would be the case where collected errors correspond to situations for which the level of predicted power was low and where the current power prediction is in the medium power range. It is hence necessary to propose a more dynamic approach that would be appropriate for estimating *conditional* prediction intervals. Our proposal is then to enhance the empirical method initially described by Williams and Goodman [235] for giving an assessment of the prediction uncertainty related to current forecast conditions. The present Paragraph concentrates on the classification of these forecast conditions.

We define as a *forecast condition* $c_{t,k}$ at time t and horizon k the association of a set of values of the considered influential variables. Denote by $v^l_{t,k}$ the l^{th} influential variable (say that we consider L different variables, hence $l = 1, \ldots, L$) related to the point prediction $\hat{p}_{t+k/t}$. We make the assumption that all the influential variables are bounded[4] and can thus be normalized. Consequently, we have

$$v^l_{t,k} \in V_l = [0, 1] \quad \forall l, t, k. \tag{4.18}$$

Prediction errors are also normalized and bounded, though they lie in the range $[-1, 1]$.

What we referred to as a forecast condition at time t for lead time $t+k$ is uniquely defined by the association of the values of each of the L influential variables:

$$c_{t,k} = \{v^1_{t,k}, v^2_{t,k}, \ldots, v^L_{t,k}\}, \quad c_{t,k} \in \mathcal{C} = V_1 \times V_2 \times \ldots \times V_L, \tag{4.19}$$

where $\mathcal{C}$ is the set of possible forecast conditions at any time t and look-ahead time k.

Then, we map $\mathcal{C}$ with a finite number of subsets to which are associated different kinds of characteristics of prediction error distributions. For that purpose, consider J_l ranges of possible values for each of the influential variables v^l ($l = 1, \ldots, L$). Consequently, we define as $V_l^{j_l}$ the subset of V_l that contains the variable values in the $j_l{}^{th}$ range. By construction, V_l is the union of all of its subsets

$$V_l = V_l^1 \cup V_l^2 \cup \ldots \cup V_l^{J_l}, \quad \forall l, \tag{4.20}$$

such that none of these subsets are overlapping

$$V_l^i \cap V_l^j, = \emptyset, \quad \forall l, i, j, \, i \neq j. \tag{4.21}$$

Now that the sets of possible values for the various influential variables are split into subsets accounting for different characteristics of prediction error distributions, $\mathcal{C}$ can also be split into all the possible associations of the subsets for the various influential variables.

[4]This assumption about the bounded nature of influential variables appears reasonable: the range of physically possible values for both measured or forecast variables obviously have a lower and an upper bound. If outside of that range, these values can be deemed as suspicious or even as outliers.

Write

$$\mathcal{C}(\{(l, j_l)\}) \;=\; \mathcal{C}((1, j_1), \ldots, (L, j_L))) \;=\; V_1^{j_1} \times V_2^{j_2} \times \ldots \times V_L^{j_L}, \quad \forall j_l, \tag{4.22}$$

these subsets corresponding to the $j_l{}^{th}$ range of values for each of the L different influential variables. This hence yields N_s subsets, where

$$N_s \;=\; \prod_{l=1}^{L} J_l. \tag{4.23}$$

If for instance one considers two influential variables (say forecast wind power and forecast wind direction) for which sets of possible predicted values are split into two subsets, then $\mathcal{C}((1, 1), (2, 2))$ corresponds to the subset of forecast conditions for which predicted wind power lies in its first subset and predicted wind direction in its second subset. Again, by construction, $\mathcal{C}$ is the union of all of its subsets, such that none of them are overlapping. Note that this classification of the forecast conditions with different related characteristics of prediction error distributions can only be the result of a thorough analysis of the error-generating process. Analyses of forecasting errors are often very informative (cf. Chapter 3), and allow the analyst to gain expertise on the prediction problem.

Since our aim is to associate specific characteristics of prediction error distributions to each subset of $\mathcal{C}$, we extend here the empirical approach described in Paragraph 4.4.1, by associating a collection of recent prediction errors to each of these subsets. As introduced in Equation (4.12), $\Omega_{t,k}$ is the set of all the past k-step ahead prediction errors up to time t. Define now $\Omega_{t,k}(\{(l, j_l)\})$ the subset of past prediction errors corresponding to the subset of forecast conditions $\mathcal{C}(\{(l, j_l)\})$:

$$\Omega_{t,k}(\{(l, j_l)\}) \;=\; \{\epsilon_{t-i+k/t-i} \in \Omega_{t,k} \mid c_{t-i,k} \in \mathcal{C}(\{(l, j_l)\})\}, \quad \forall j_l. \tag{4.24}$$

And finally, as we did in Equation (4.13), we can extract from each subset $\Omega_{t,k}(\{(l, j_l)\})$ a sample $S_{t,k}(\{(l, j_l)\})$ of size n containing the last n forecasting errors, but in *similar forecast conditions*:

$$S_{t,k}(\{(l, j_l)\}) \;=\; \{\epsilon_i \in \Omega_{t,k}(\{(l, j_l)\}),\ i = 1, \ldots, n\}, \quad \forall j_l. \tag{4.25}$$

Therefore, each of the subsets $\mathcal{C}(\{(l, j_l)\})$ is characterized by its own empirical distribution function $\hat{F}^{\epsilon}_{t,k}(\{(l, j_l)\})$, drawn from a different sample of past errors. Note that $\hat{F}^{\epsilon}_{t,k}(\{(l, j_l)\})$ is a conditional distribution function since it is an estimate of the distribution function of prediction errors given that $c_{t,k}$ is an item of $\mathcal{C}(\{(l, j_l)\})$. This empirical distribution function puts probability $1/n$ on each element of $S_{t,k}(\{(l, j_l)\})$:

$$\hat{F}^{\epsilon}_{t+k/t}(\{(l, j_l)\}) \;\rightarrow\; \{\epsilon_i,\ \epsilon_i \in S_{t,k}(\{(l, j_l)\})\}, \tag{4.26}$$

4.4.3 The fuzzy inference model for producing conditional distribution functions

The previously described classification is the basis for deriving an empirical and distribution-free method that provides conditional prediction intervals, given particular forecast conditions. The choice of the influential variables, as well as the splitting of the sets of possible values into various subsets with different characteristics of related prediction error distributions, are the result of the expertise one has on the process of interest. It was explained in Paragraph 4.4.1 how to dress a point prediction $\hat{p}_{t+k/t}$ with an empirical distribution of prediction errors $\hat{F}^{\epsilon}_{t+k/t}$ for producing empirical distributions of wind generation $\hat{F}^{p}_{t+k/t}$. Hereafter, we develop a fuzzy inference model $h_f(c_{t,k})$ which gives conditional distributions of prediction errors $\hat{F}^{\epsilon,*}_{t+k/t}(c_{t,k})$ given the forecast condition $c_{t,k}$.

Fuzzy logic is an alternative paradigm to that of binary logic for which an event can only be associated to a true or false statement (and therefore 1 or 0). It considers instead that to each event can be associated a degree of truth, which is a continuous function between 0 and 1. For an introduction to the fuzzy logic theory, we refer to [232]. In the previous Paragraph, the set $\mathcal{C}$ of possible forecast conditions has been mapped with several subsets $\mathcal{C}(\{(l,j_l)\})$ related to different characteristics of the forecast uncertainty. Particularly, we have explained that a given subset $\mathcal{C}(\{(l,j_l)\})$ is defined as the association of the subsets $V_l^{j_l}$ $(l=1,\ldots,L)$ for the various considered input variables (Equation (4.22)). Here, we associate a fuzzy set $\mathcal{A}_l^{j_l}$ to each of these V-subsets. A fuzzy set is characterized by a membership function $m_l^{j_l}(v_{t,k}^l)$, which tells what the degree of truth of $v_{t,k}^l$ being an element of $V_l^{j_l}$ is:

$$m_l^{j_l} \,:\, v_{t,k}^l \,\rightarrow\, m_l^{j_l}(v_{t,k}^l) \in [0,1]. \tag{4.27}$$

The subset of forecast conditions $\mathcal{C}(\{(l,j_l)\})$ is defined as the association of the L subsets $V_l^{j_l}$. Therefore, the degree of truth of a given forecast condition $c_{t,k}=\{v_{t,k}^l\}_{l=1,\ldots,L}$ being an element of $\mathcal{C}(\{(l,j_l)\})$ is given by the product of the membership values for every influential variable:

$$m(c_{t,k},\{(l,j_l)\}) \;=\; m\left(c_{t,k}\in\mathcal{C}(\{(l,j_l)\})\right) \;=\; \prod_{l=1}^{L} m_l^{j_l}(v_{t,k}^l). \tag{4.28}$$

The basic element of the fuzzy inference model we develop here consists in fuzzy rules. Such a fuzzy rule can be expressed as

$$\text{“}\,\textbf{IF}\quad v_{t,k}^1\in\mathcal{D}(\mathcal{A}_1^{j_1})\ \text{ and }\ \ldots\ \text{ and }\ v_{t,k}^L\in\mathcal{D}(\mathcal{A}_L^{j_L})\quad\textbf{THEN}\quad \epsilon_{t+k/t}\sim F_{t,k}^{\epsilon}(\{(l,j_l)\})\,\text{”}, \tag{4.29}$$

where $\mathcal{D}(\mathcal{A}_l^{j_l})$ stands for the support of the fuzzy set $\mathcal{A}_l^{j_l}$. The ‘IF’ part is referred to as the premise of the rule, whereas the ‘THEN’ part is called the conclusion. Note that the above rule is equivalent to:

$$\text{“}\,\textbf{IF}\quad c_{t,k}\in\mathcal{D}(\mathcal{A}_{\mathcal{C}}(\{(l,j_l)\}))\quad\textbf{THEN}\quad \epsilon_{t+k/t}\sim\hat{F}_{t,k}^{\epsilon}(\{(l,j_l)\})\,\text{”}. \tag{4.30}$$

where

$$\mathcal{D}(\mathcal{A}_{\mathcal{C}}(\{(l,j_l)\})) \;=\; \mathcal{D}(\mathcal{A}_1^{j_l}) \times \ldots \times \mathcal{D}(\mathcal{A}_L^{j_L}). \tag{4.31}$$

Actually, the rule (4.30) states that if the forecast condition $c_{t,k}$ can be considered as being an item of a given subset $\mathcal{C}(\{(l,j_l)\})$ of $\mathcal{C}$, then the prediction error $\epsilon_{t+k/t}$ follows the distribution $F^{\epsilon}_{t,k}(\{(l,j_l)\})$.

Then, a rule base is composed by rules similar that given by (4.30), which span all the possible subsets of $\mathcal{C}$. The number of fuzzy rules is hence given by the number of subsets N_s used to map the set of possible forecast conditions. For convenience, we associate an index i to each of the N_s subsets, and we introduce the function $\eta(i)$ that returns the $\{(l,j_l(i))\}_{l=1,\ldots,L}$ pairs that serve to identify the corresponding subset:

$$\eta \;:\; i \in \{1,\ldots,N_s\} \;\rightarrow\; (\{(l,j_l(i))\}_{l=1,\ldots,L}), \tag{4.32}$$

such that each of the $\{(l,j_l(i))\}_{l=1,\ldots,L}$ pairs is given by a unique value of i. Consequently, the i^{th} rule of the fuzzy rule base is of the form:

$$\text{“ }\mathbf{IF} \quad c_{t,k} \in \mathcal{D}(\mathcal{A}_{\mathcal{C}}(\eta(i))) \quad \mathbf{THEN} \quad \epsilon_{t+k/t} \sim F^{\epsilon}_{t,k}(\eta(i)) \text{ ”}, \tag{4.33}$$

where

$$\mathcal{D}(\mathcal{A}_{\mathcal{C}}(\eta(i))) \;=\; \mathcal{D}(\mathcal{A}_1^{j_l(i)}) \times \ldots \times \mathcal{D}(\mathcal{A}_L^{j_L(i)}). \tag{4.34}$$

The inference procedure for the fuzzy logic model consists in applying the rule-base to the forecast condition $c_{t,k}$ in order to provide the overall conclusion as the weighted average of the conclusion of each rule. The weight w_i for each rule is given by the degree of truth of the related premise, normalized by the sum of the weights for each rule:

$$w_i(c_{t,k}) \;=\; \frac{m(c_{t,k},\eta(i))}{\sum_{i=1}^{N_s} m_{\eta(i)}(c_{t,k})}, \quad i=1,\ldots,N_s, \tag{4.35}$$

with $m(c_{t,k},\eta(i))$ defined by Equation (4.28).

By doing so, the fuzzy model tells what is the contribution of each of the $F^{\epsilon}_{t,k}(\eta(i))$ $(i = 1,\ldots,N_s)$ error distributions in the error distribution $F^{\epsilon}_{t,k}$ related to the current forecast condition $c_{t,k}$. Finally, the fuzzy logic model can be written as

$$h_f \;:\; c_{t,k} \;\rightarrow\; \epsilon_{t+k/t} \sim F^{\epsilon,*}_{t,k} = \sum_{i=1}^{N_s} w_i(c_{t,k}).F^{\epsilon}_{t,k}(\eta(i)). \tag{4.36}$$

Let us draw an illustration of the fuzzy inference process by going back to the example of the above Paragraph, in which predicted power and forecast wind direction were considered as influential variables. For both input variables, the sets of possible values were split into two subsets. This would yield four samples of prediction errors $S_{t,k}((1,1),(2,1))$,

$S_{t,k}((1,1),(2,2))$, $S_{t,k}((1,2),(2,1))$, $S_{t,k}((1,2),(2,2))$ related to different subsets of forecast conditions $\mathcal{C}((1,1),(2,1))$, $\mathcal{C}((1,1),(2,2))$, $\mathcal{C}((1,2),(2,1))$, $\mathcal{C}((1,2),(2,2))$, for a given time t and look-ahead time k. Moreover, the fuzzy logic model in that case would have a rule-base composed by four rules, one for each of the possible $\mathcal{C}$-subsets. Then, imagine that for given time t and horizon k the degrees of truth of the current forecast condition $c_{t,k}$ being part of the $\mathcal{C}$-subsets are evaluated to be equal to 0.3, 0.5, 0.15 and 0.05 respectively. The fuzzy rule-base (4.36) then defines the corresponding distribution of prediction errors as:

$$\begin{aligned}\epsilon_{t+k/t} \sim F^{\epsilon,*}_{t,k} \quad &= \quad 0.3\, F^{\epsilon}_{t,k}((1,1),(2,1)) \;+\; 0.5\, F^{\epsilon}_{t,k}((1,1),(2,2)) \\ &\quad\; + 0.15\, F^{\epsilon}_{t,k}((1,2),(2,1)) \;+\; 0.05\, F^{\epsilon}_{t,k}((1,2),(2,2)). \end{aligned} \tag{4.37}$$

4.4.4 Methods for combining error distributions

In the above Paragraph, we have developed a fuzzy inference model h_f that provides conditional distribution functions of prediction errors. Given a specific forecast condition $c_{t,k}$, it returns the distribution $F^{\epsilon,*}_{t,k}$ of prediction errors $\epsilon_{t+k/t}$ as a combination of several distributions, corresponding to different subsets of the forecast conditions.

Combining probability distributions is not a trivial task. Perhaps the area which is the most concerned with the probability-combination problem is the area of probabilistic risk analysis and decision science. It is often demanded to a panel of experts to provide their judgment on a particular event in the form of probability distributions. A decision maker has then to assimilate the various experts' judgments, which may be converging or conflicting. Hence, the corresponding probability distributions can have significantly different shapes, and, in a general manner, they cannot be seen as Gaussian or even symmetric. The assimilation procedure followed by the decision maker consists in summarizing the various experts' opinion in a single *combined probability distribution*. In the last decades, several methods have been developed for that purpose, either of the mathematical or of the behavioral types. These methods are reviewed by Clemen and Winkler in [45]. In the following, we describe two alternative approaches for combining probability distributions: the linear opinion pool and adapted resampling.

The linear opinion pool

An appealing approach to the aggregation of probability distributions is the *linear opinion pool*, which consists in saying that a combined distribution is the weighted average of the individual probability distributions, the weights being non-negative and summed to one [208]. One notices then that this is exactly what is given by the fuzzy logic inference model described by Equation (4.36): the probability distribution of errors is given as the weighted average of probability distributions for various forecast condition subsets. Genest and McConway [74] discussed the interpretation of the weights to be assigned to individual probability distributions. While it appears obvious that they relate to the confidence one may have in such or such distribution to give more information on the true effect, it is not evident how they should be calculated. In our case, the weights are derived from the fuzzy-inference model, which, for a given forecast condition $c_{t,k}$, tells to what extent we expect the error distribution for each subsets of $\mathcal{C}$ to represent the actual error distribution. Moreover, we assume that by considering non-overlapping subsets of forecast conditions the related error distributions $F^{\epsilon}_{t,k}(\eta(i))$ $(i = 1\ldots, N_s)$ can provide independent and relevant information on the true distribution.

In Paragraph 4.4.1 it was explained that an error distribution function could be approximated by its related empirical distribution function that puts an equal probability to every item of a sample of past errors. Straightforwardly, we approximate here each $F^{\epsilon}_{t,k}(\eta(i))$ by the related $\hat{F}^{\epsilon}_{t,k}(\eta(i))$ $(i = 1\ldots, N_s)$. Consequently, an estimate of the distribution $F^{\epsilon,*}_{t,k}$ follows from Equation (4.36):

$$\hat{F}^{\epsilon,*}_{t,k}(c_{t,k}) = \sum_{i=1}^{N_s} w_i(c_{t,k}).\hat{F}^{\epsilon}_{t,k}(\eta(i)). \tag{4.38}$$

By gathering all the error sample $S_{t,k}(\eta(i))$ $(i = 1\ldots, N_s)$ available at time t and for horizon k, we define $S^*_{t,k}$ such that

$$S^*_{t,k} = S_{t,k}(\eta(1)) \cup \ldots \cup S_{t,k}(\eta(N_s)). \tag{4.39}$$

Therefore, given that the size of the error sample $S_{t,k}(\eta(i))$ $(i = 1\ldots, N_s)$ is set to n, $S^*_{t,k}$ is composed by $n.N_s$ elements. An estimate of the distribution $F^{\epsilon,*}_{t,k}$ is given by the discrete distribution that puts a probability $w_j = w_i(c_{t,k})/n$ to every element ϵ_j of $S^*_{t,k}$ that is originally an element of $S_{t,k}(\eta(i))$:

$$\hat{F}^{\epsilon,*}_{t,k}(c_{t,k}) \;\rightarrow\; \{\epsilon_j \in S^*_{t,k},\; \mathrm{P}(\epsilon_j \mid \epsilon_j \in S^*_{t,k} \cap S_{t,k}(\eta(i))) \;=\; w_j \;=\; w_i(c_{t,k})/n\}. \tag{4.40}$$

As in the previous developments, the predictive distribution of wind generation is constructed by associating the estimate of the distribution of prediction errors to the point forecast itself:

$$\hat{F}^{p,*}_{t,k}(c_{t,k}) \;\rightarrow\; \{\hat{p}_{t+k/t} + \epsilon_j,\; \mathrm{P}(\epsilon_j \mid \epsilon_j \in S^*_{t,k} \cap S_{t,k}(\eta(i))) \;=\; w_j \;=\; w_i(c_{t,k})/n\}. \tag{4.41}$$

Note that $\hat{F}^{p,*}_{t,k}$ is now a continuous function of forecast conditions.

The cumulative distribution function $\hat{G}^{p,*}_{t,k}$ related to $\hat{F}^{p,*}_{t,k}$ has a slightly different form than that of Equation (4.14), since the items of $S^*_{t,k}$ do not have the same probabilities:

$$\hat{G}^{p,*}_{t,k}(\epsilon) \;=\; \sum_{j=0}^{n.N_s} w_j.\mathbf{1}_{\epsilon_j<\epsilon}, \tag{4.42}$$

where $\mathbf{1}_{\epsilon_j<\epsilon}$ takes the value 1 if $\epsilon_j < \epsilon$ and 0 otherwise.

However, $\hat{G}^{p,*}_{t,k}$ can be used similarly for estimating the lower and upper bounds of the central prediction interval $\hat{\mathrm{I}}^{(\alpha)}_{t+k/t}$ with nominal coverage rate $(1-\alpha)$ by picking the quantiles of the predictive distribution $\hat{F}^{p,*}_{t+k/t}$, with proportion $(\alpha/2)$ and $(1-\alpha/2)$ respectively:

$$\hat{L}^{(\alpha)}_{t+k/t} = \hat{p}_{t+k/t} + \left(\hat{G}^{p,*}_{t,k}\right)^{-1}(\alpha/2), \tag{4.43}$$

$$\hat{U}^{(\alpha)}_{t+k/t} = \hat{p}_{t+k/t} + \left(\hat{G}^{p,*}_{t,k}\right)^{-1}(1-\alpha/2). \tag{4.44}$$

The adapted resampling method

The aim of methods like resampling (or bootstrapping, following the terminology of its inventor Efron [60]) is to have a better idea of a population distribution parameter (e.g. its mean or standard deviation) by going through a representative sample a high number of times. This manipulation of the representative sample can serve to associate a measure of accuracy to the estimate of this population parameter. Actually, bootstrapping has also been considered in the forecasting literature for estimating prediction intervals associated to point forecasts (see Clements and Taylor [49], Grigoletto [83], or Reeves [192] among others). Such a method has the advantage of being non-parametric, but it needs to have access to the analytic model. This is not conceivable here, since the approach we aim at developing assumes that the point prediction method is a kind of black-box, and thus that we do not have access to the underlying model. Resampling is used here as an alternative to the linear opinion pool approach for estimating quantiles of combined probability distributions.

Write $S = \{\epsilon_j\}_{j=1,\dots,n}$ a random sample from a probability distribution F. The observations ϵ_j $(j = 1,\dots,n)$ are assumed to be i.i.d. (independent and identically distributed) F. Following Efron's terminology, the *plug-in estimate* of a parameter $\theta = h(F)$ is defined to be $\hat{\theta} = h(\hat{F})$. This means that we estimate the true parameter of F by applying the same function to the empirical distribution function $\hat{F}$. This is what we have done in Equations (4.16) and (4.17) for estimating the lower and upper bounds of the prediction intervals. The elements of S are used for setting up an estimate $\hat{G}$ of the cumulative distribution function associated to F.

Denote by $X = \{x_j\}_{j=1,\dots,n}$ a random sample that is i.i.d. $U[0,1]$. The theory of probabilities tells us that the sample $G^{-1}(X) = \{G^{-1}(x_j)\}_{j=1,\dots,n}$ is i.i.d. F. Then, the idea of resampling states that since $\hat{G}$ is an estimate of the true cumulative distribution associated to F, one can use it for drawing alternative samples that would lead to other empirical distribution functions of the true distribution F. In practice, this alternative sample $S^{(b)}$ $(b = 1,\dots,B)$ is called a *bootstrap sample* and is obtained by picking *randomly* and *with replacement* n values out of the original sample S. $\hat{\theta}^{(b)}$ is a *bootstrap replication* of the θ statistic. Since all the bootstrap replications are potential estimates of the true parameter θ, one can consider them for calculating the bias or standard deviation associated to the original estimate $\hat{\theta}$, or even confidence intervals.

Here, we propose to apply the idea of resampling for estimating a given parameter θ of a combined probability distribution, by having a slightly different interpretation of the combination given by the fuzzy inference model (4.36) than that of the linear opinion pool. Remember that the fuzzy inference model gives a weight to each of the N_s distributions $F^{\epsilon}_{t,k}(\eta(i))$. The distributions $F^{\epsilon}_{t,k}(\eta(i))$ can be approximated by the empirical distributions

$\hat{F}^{\epsilon}_{t,k}(\eta(i))$. The linear opinion pool approach states that these weights can be seen as probabilities and that one can construct a combined distribution by associating these probabilities to each sample. The difference we introduce here is that these weights w_i $(i = 1, \dots, N_s)$ are to be used for defining the share of each of the representative samples of errors $S_{t,k}(\eta(i))$ for defining a representative sample drawn from the combined distribution. We will use that interpretation for creating B bootstrap sample $S^{(b)}_{t,k}$ and compute a bootstrap replication $\hat{\theta}^{(b)}$ for each of them. Given n the size of the error samples, a bootstrap sample $S^{(b)}_{t,k}$ (also of size n) is constructed as following:

$$S^{(b)}_{t,k} = \{S^{(b)}_{t,k}(\eta(i))\}_{i=1,\dots,N_s}, \tag{4.45}$$

such that

$$S^{(b)}_{t,k}(\eta(i)) = \{\epsilon_j \mid \epsilon_j \in S_{t,k}(\eta(i))\}_{j=1,\dots,w_i.n}, \quad i = 1, \dots, N_s, \tag{4.46}$$

where the items of $S^{(b)}_{t,k}(\eta(i))$ are picked randomly and with replacement from $S_{t,k}(\eta(i))$.

The parameters of interest are the quantiles of the combined probability distribution $\hat{F}^{\epsilon}_{t,k}$. Therefore, write $\hat{G}^{\epsilon,(b)}_{t,k}$ the cumulative distribution function associated to the empirical distribution function $\hat{F}^{\epsilon,(b)}_{t,k}$ (following the definition of Equation (4.14)). The bootstrap replications of the lower and upper bounds of the interval forecast $\hat{\mathrm{I}}^{(\alpha)}_{t+k/t}$ with nominal coverage rate $(1-\alpha)$ are given by:

$$\hat{L}^{(\alpha)(b)}_{t+k/t} = \hat{p}_{t+k/t} + \left(\hat{G}^{\epsilon,(b)}_{t,k}\right)^{-1}(\alpha/2), \tag{4.47}$$

$$\hat{U}^{(\alpha)(b)}_{t+k/t} = \hat{p}_{t+k/t} + \left(\hat{G}^{\epsilon,(b)}_{t,k}\right)^{-1}(1-\alpha/2). \tag{4.48}$$

Finally, we approximate the bootstrap expectation by taking the mean of all the bootstrap replications, in order to obtain an estimate of the interval limits:

$$\hat{L}^{(\alpha)}_{t+k/t} = \frac{1}{B}\sum_{b=1}^{B} \hat{L}^{(\alpha)(b)}_{t+k/t}, \tag{4.49}$$

$$\hat{U}^{(\alpha)}_{t+k/t} = \frac{1}{B}\sum_{b=1}^{B} \hat{U}^{(\alpha)(b)}_{t+k/t}. \tag{4.50}$$

Note that by constituting these B bootstrap samples, we actually use all the information included in the individual samples by drawing alternatives scenarios. Also, while Efron and Tibshirani (see [61], pp. 124-126) explain that the bootstrap expectation serves for calculating the bias associated to the original estimate of a distribution parameter from a single sample, it has a completely different meaning here, since we apply that form of resampling for a multi-sample problem. In the remaining of the document, this approach is referred to as *adapted resampling* owing to the similarities with the original resampling approach.

4.5 Application to the wind power forecasting problem

In this Section, we detail how the previously introduced methods can be straightforwardly applied to the specific case of the wind power forecasting problem, by describing a particular configuration that accounts for both the nonlinearity related to the level of forecast power and the one related to the level of forecast wind speed (owing to the cut-off risk). Therefore, following the notations used in the previous Section, let us consider two influential variables ($L = 2$):

$$v^1_{t,k} = \hat{p}_{t+k/t}, \quad v^1_{t,k} \in V_1, \tag{4.51}$$

$$v^2_{t,k} = \hat{u}_{t+k/t}, \quad v^2_{t,k} \in V_2. \tag{4.52}$$

The forecast condition $c_{t,k}$ at time t, for lead time $t+k$, is then given by the pair consisting of the forecast wind speed and predicted power values

$$c_{t,k} = \{v^1_{t,k}, v^2_{t,k}\} = \{\hat{p}_{t+k/t}, \hat{u}_{t+k/t}\}, \quad c_{t,k} \in \mathcal{C} = V_1 \times V_2. \tag{4.53}$$

To account first for the power curve effects detailed in Paragraph 3.5.2, the set V_1 of possible power values is divided into three subsets ($J_1 = 3$), corresponding to the power ranges 'low' (V_1^1), 'medium' (V_1^2) and 'high' (V_1^3):

$$V_1 = V_1^1 \cup V_1^2 \cup V_1^3. \tag{4.54}$$

In parallel, V_2 is divided into two subsets ($J_2 = 2$), corresponding to the range of forecast wind speed values for which a cut-off event is not expected (V_2^1), and to the range of values for which a cut-off is probable (V_2^1):

$$V_2 = V_2^2 \cup V_2^2. \tag{4.55}$$

This constitutes six different subsets of the forecast conditions ($N_s = 6$):

$$\mathcal{C}_1 = V_1^1 \times V_2^1, \tag{4.56}$$

$$\mathcal{C}_2 = V_1^2 \times V_2^1, \tag{4.57}$$

$$\mathcal{C}_3 = V_1^3 \times V_2^1, \tag{4.58}$$

$$\mathcal{C}_4 = V_1^1 \times V_2^2, \tag{4.59}$$

$$\mathcal{C}_5 = V_2^2 \times V_2^2, \tag{4.60}$$

$$\mathcal{C}_6 = V_3^3 \times V_2^2. \tag{4.61}$$

However, if considering a theoretical power curve such as the one depicted in Figure 4.2 it appears unlikely that a cut-off event occurs when predicted power values are in the 'low' or 'medium' ranges, and that the possibility of a cut-off event is only dictated by the forecast wind speed, the three subsets formed with V_2^2 are grouped to form only one:

$$\mathcal{C}_{4+} = \mathcal{C}_4 \cup \mathcal{C}_5 \cup \mathcal{C}_6. \tag{4.62}$$

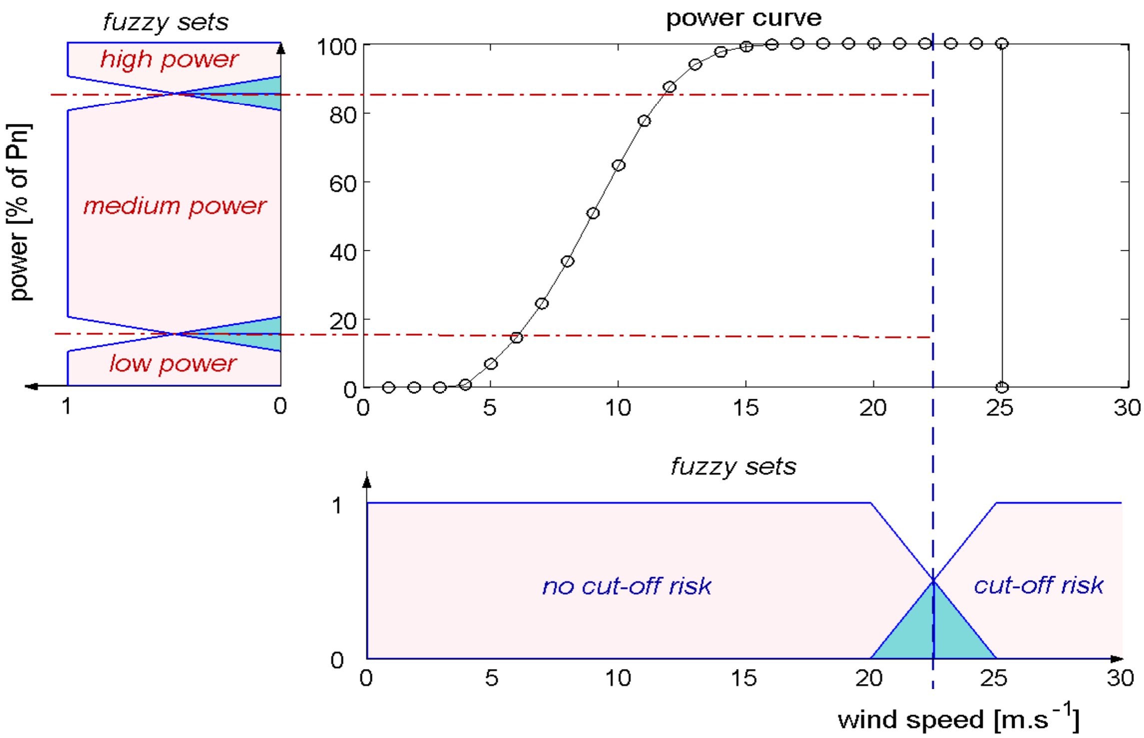


Figure 4.2: *Mapping of the forecast uncertainty introduced by the power curve. The range of possible predicted power values is divided into three ranges ('low', 'medium' and 'high'), to which are associated three trapezoidal fuzzy sets, in order to account for the nonlinearity introduced by the power variable. Similarly, the range of possible forecast wind speed values is divided into two ranges ('no cut-off risk' and 'cut-off risk'), owing to the nonlinearity introduced by the cut-off, to which are associated two trapezoidal fuzzy sets. This yields four zones of the power curve related to different characteristics of power prediction error distributions.*

Denote by $S^1_{t,k}$, $S^2_{t,k}$, $S^3_{t,k}$ and $S^{4+}_{t,k}$ the samples (of size n) of prediction errors corresponding to the four subsets of forecast conditions introduced above. At a given time t, each of these samples contain the last n k-step ahead prediction errors made by the point prediction approach in the forecast conditions defined by $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$ and $\mathcal{C}_{4+}$ respectively.

To every subsets of V_1 and V_2 are associated trapezoidal fuzzy sets. Figure 4.2 illustrates this mapping of a theoretical power curve into various zones corresponding to different characteristics of the prediction error distributions. Then, denote by $\mathcal{A}^i_{\mathcal{C}}$ the two-dimensional fuzzy sets related to the subset of forecast conditions $\mathcal{C}_i$, $i = 1, \ldots, 4+$. Each two-dimensional fuzzy set $\mathcal{A}^i_{\mathcal{C}}$ is characterized by its membership function $m(., i)$, $i = 1, \ldots, 4+$. The analytical form of these membership functions is not given here.

The fuzzy rule base inference model is composed by four fuzzy rules, which can be expressed as:

$$\text{“ \textbf{IF} } \quad c_{t,k} \in \mathcal{D}(\mathcal{A}^1_{\mathcal{C}}) \quad \textbf{THEN} \quad \epsilon_{t+k/t} \sim F^{\epsilon,1}_{t,k} \text{ ”}, \tag{4.63}$$

$$\text{“ \textbf{IF} } \quad c_{t,k} \in \mathcal{D}(\mathcal{A}^2_{\mathcal{C}}) \quad \textbf{THEN} \quad \epsilon_{t+k/t} \sim F^{\epsilon,2}_{t,k} \text{ ”}, \tag{4.64}$$

$$\text{“ \textbf{IF} } \quad c_{t,k} \in \mathcal{D}(\mathcal{A}^3_{\mathcal{C}}) \quad \textbf{THEN} \quad \epsilon_{t+k/t} \sim F^{\epsilon,3}_{t,k} \text{ ”}, \tag{4.65}$$

$$\text{“ \textbf{IF} } \quad c_{t,k} \in \mathcal{D}(\mathcal{A}^{4+}_{\mathcal{C}}) \quad \textbf{THEN} \quad \epsilon_{t+k/t} \sim F^{\epsilon,4+}_{t,k} \text{ ”}, \tag{4.66}$$

where $F^{\epsilon,i}_{t,k}$ is the empirical probability distribution that puts probability $1/n$ on each element of $S^i_{t,k}$, $i = 1, \ldots, 4+$. For instance, the first rule (given by (4.63)) states that if predicted power $\hat{p}_{t+k/t}$ is in the ‘low’ range and forecast wind speed $\hat{u}_{t+k/t}$ is in the ‘no cut-off’ range, prediction errors $e_{t+k/t}$ for that look-ahead time are distributed $F^{\epsilon,i}_{t,k}$.

The fuzzy inference model, which gives the conditional distributions of prediction errors as a function of the forecasts conditions, can thus be written as:

$$h_f \; : \; c_{t,k} \;\rightarrow\; \epsilon_{t+k/t} \sim F^{\epsilon,*}_{t,k} = \sum_{i=1}^{4+} w_i(c_{t,k}).F^{\epsilon,i}_{t,k}, \tag{4.67}$$

where the weights w_i of each of the fuzzy rules are calculated as following:

$$w_i(c_{t,k}) \;=\; \frac{m(c_{t,k}, i)}{\sum_{i=1}^{4+} m(c_{t,k}, i)}, \quad i = 1, \ldots, 4+ \,. \tag{4.68}$$

Let us now imagine an operational wind power forecasting application in which point predictions are produced from a state-of-the-art method (say one of the M1, M2,..., M5 methods introduced in Section 3.3). Denote by k_{max} the forecast length. The size n of the error samples is defined by the end-user, as well as the nominal coverage rate $(1-\alpha)$ of the interval forecasts. The Algorithm 4.1 describes the steps for the estimation of prediction intervals of wind generation at prediction time t. In a first stage, one retrieves the power measure p_t at time t and the series of predictions $\hat{p}_{t+k/t}$ $k = 1, \ldots, k_{max}$ produced from the point forecasting method. The power measure is used for calculating the errors $e_{t/t-k}$ related to the predictions $\hat{p}_{t/t-k}$ issued in the past for time t. It is thus necessary to store the series of predictions for a time period equal to the forecast length of the considered point prediction method. This is also valid for the case of the considered influential variables. Forecast conditions $c_{t-k,k}$ related to $e_{t/t-k}$ are determined in order to decide to which error samples the prediction errors $\hat{e}_{t/t-k}$ belong to. These samples are then updated by discarding the oldest error value and by adding the new one as it becomes available[5]. This makes the method *adaptive*, since it always considers the most recent information on the process. It can accommodate temporal modifications of the characteristics of prediction error distributions, owing to the season of the year, changes in the wind farm environment, etc. The

[5] At the beginning of the application, error samples are empty. But, as new predictions are provided and related power measures made available, these samples are filled and updated. Even if the number of items in each sample has not reached n, it is possible to apply the previously described methods by modifying the necessary steps, i.e. by considering the number of available errors instead of the required n elements. After a certain time of operation (a minimum of $n.N_s$ forecasting steps), all the samples attain their defined size.

fuzzy inference model (Equation (4.67)) is used independently for each prediction horizon, for determining how to estimate conditional prediction error distributions $F_{t,k}^{\epsilon,*}$ given the forecast conditions $c_{t,k}$. The prediction intervals $\hat{\mathrm{I}}_{t+k/k}^{(\alpha)}$ with nominal coverage rate $(1-\alpha)$ are finally estimated by applying either the linear opinion pool or the resampling approach for the combination of probability distributions given by the fuzzy inference model.

Algorithm 4.1: *The necessary steps at time t for producing the empirical and distribution-free prediction intervals of wind generation*

step 1.	Retrieve the power measure p_t for time t
step 2.	Retrieve and store the power predictions $\hat{p}_{t+k/t}$, $k>0$, provided by a point prediction method, as well as related influential variables values
step 3.	Calculate the prediction errors $e_{t/t-k}$ related the power predictions $\hat{p}_{t/t-k}$, $k>0$, issued at time $t-k$ for time t
step 4.	Update the relevant error samples given the forecast condition $c_{t-k,k}$ related to the prediction error $e_{t/t-k}$
step 5.	For each look-ahead time k, use the fuzzy inference model given by Equation (4.67) to determine the distribution of prediction errors $F_{t,k}^{\epsilon}$ given the forecast conditions $c_{t,k}$
step 6.	For each look-ahead time k, apply either the linear opinion pool or the adapted resampling approach for estimating the bounds of the prediction intervals of wind generation $\hat{\mathrm{I}}_{t+k/t}^{(\alpha)}$

The proposed methods for the estimation of prediction intervals of wind generation has originally been developed for online application. In Appendix C, we detail the characteristics of the module we have implemented and which is integrated in the ANEMOS prediction platform.

4.6 Discussion on operational aspects

What type of prediction intervals?

The above methods permit to estimate predictive distributions of wind generation. We proposed to summarize the uncertainty information by quoting prediction intervals, which consist of two particular quantiles of these distributions. Actually, even by assigning a certain nominal coverage rate, the resulting intervals with that pre-assigned probability can be intervals around the mean, the median, or intervals with the shortest length for instance. It was already explained in Section 4.2 that it was more appropriate to provide central prediction intervals than intervals around the distribution mean, since they correspond to quantiles with known proportions. An alternative described by Hyndman [102] is to provide highest-density interval forecasts, which are defined as intervals with the shortest length given that any point inside the intervals has a probability density at least as large as every point outside the intervals. They can be of practical interest when working with non-normal and multimodal distributions. But, again, the endpoints of these intervals do not correspond to quantiles with known proportions. For instance, if focusing on the lower bound of intervals with a 80% degree of confidence, it may correspond to a quantile with proportion 0.05 for a given forecast and then to a quantile with proportion 0.15 for the following one. Thus, the risk of the true effect lying below that lower bound will change for every prediction. This does not appear appropriate from an operational point of view. Therefore, we will focus on central prediction intervals only.

Choice of an optimal coverage rate

An important question concerning the intervals arises: how to choose an optimal nominal coverage rate? When this pre-assigned probability is higher than 90%, intervals can be 'embarrassingly' wide, because they will contain extreme prediction errors (or even outliers). Working with high-coverage intervals means that we are aiming at modeling the very tails of the error distributions. Thus, the robustness of the uncertainty estimation methods becomes a critical aspect. However, if one defines lower pre-assigned probabilities (50% for instance), intervals will be much more narrow and more robust with respect to extreme prediction errors. But, this would translate to future observations being equally likely to lie inside or outside these bounds. In both cases, prediction intervals appear hard to handle and that is why an intermediate degree of confidence (75-85%) seems to be a good compromise [41].

Marginal or simultaneous prediction intervals

Moreover, the fact that prediction intervals are designed for multi-step ahead forecasts imposes to define what is the real required degree of confidence. As a matter of fact, there is a difference between a nominal coverage rate defined for each predicted value and a nominal coverage rate that would be defined over the whole forecast length. For instance, if a 85% degree of confidence is required for one-day ahead hourly predictions, the former corresponds to "each of the 24 intervals will contain the true value 85% of the times" (referred to as *marginal intervals* in the forecasting literature, though the term pointwise may be more appropriate), while the latter translates to "the 24 intervals will contain all the 24 true values 85% of the times" (referred to as *simultaneous intervals* in the literature). The second way of reasoning is obviously much more restrictive and seems less applicable in our case. As we explained in previous Sections, the method for interval estimation is applied separately for every look-ahead time. Therefore, the observed confidence will be verified accordingly. For a more thorough discussion about multi-step ahead prediction intervals, we refer to Chan et al. [39] and Ravishankar et al. [191].

Multiple intervals for providing predictive distributions of wind generation

Instead of focusing on a particular nominal coverage rate, it seems that producing a number of prediction intervals for a range of nominal coverage rates would be a better solution. This would allow one to build the whole probability distribution of expected wind generation for each look-ahead time. As explained in Section 4.4, this may involve the development of several models e.g. if considering quantile regression methods, but with the methodology described above, it is not more computationally expensive to estimate one or thirty quantiles. Wind power forecast users may request not only a single interval forecast but also predictive distributions of future wind generation. Indeed, the decision-making methods appearing in the literature need a complete approximation of the density function for providing an optimal management [57] or trading strategy [13, 67]. This will be explained in greater details in Chapter 6. Therefore, we will consider that when possible, several interval forecasts with various degrees of confidence should be provided.

4.7 A non-parametric framework for the evaluation of prediction intervals

Evaluating probabilistic forecasts (either density or interval forecasts) is more complicated than evaluating deterministic ones. When it is easy to say that a point forecast is false because the deviation between the predicted and the real values is of practical magnitude, an individual probabilistic forecast cannot be deemed as incorrect [150]. Indeed, when an interval forecast states there is a 90% confidence that expected power generation (for a given horizon) will be between 100 and 250kW and that the actual outcome equals 90kW, how to tell if this case should be part or not of the 10% of cases for which intervals miss?

In this Section, our aim is to describe what the required properties of interval forecasts are, and how they can be evaluated in terms of their statistical performance. For that purpose, we present relevant skill scores and diagrams that were introduced in the statistical and meteorological literature. We consider here a non-parametric framework that is suitable for evaluating either intervals or series of quantiles. Moreover, in the following, all criteria are evaluated as a function of the look-ahead time, or as an average over the forecast length. If the evaluation set is large enough, it would also be appropriate to assess the skill of probabilistic forecasting methods as a function of some other parameters (e.g. level of power).

4.7.1 Required properties for interval forecasts

Prediction intervals are associated to a probability, which is their nominal coverage rate. The first requirement for interval forecasts is that their empirical coverage should be close to the nominal one. Actually, if considering infinite series of interval forecasts, that empirical coverage should exactly equal the pre-assigned probability. That first property is referred to as *reliability* or *calibration* in the literature [5, 47, 150].

Besides this first requirement, it is necessary that prediction intervals provide a situation-dependent assessment of the forecast uncertainty. Their size should then vary depending on various external conditions. For the example of wind prediction, it is intuitively expected that prediction intervals (for a given nominal coverage rate) should not have the same size when predicted wind speed equals zero and when it is near cut-off speed. The most simple type of intervals is constant-size intervals (e.g. produced from climatology). Advanced methods for their estimation are expected to produce variable-size intervals. This property is commonly named *sharpness* or *resolution* of the intervals [5, 150]. Note that here, we will introduce a nuance between sharpness and resolution: the former will relate to the average size of intervals while the latter will be associated to their size variability.

Actually, the traditional view of interval forecast evaluation, which mainly comes from the econometric forecasting community, is based on the testing of correct conditional coverage. This means intervals have to be unconditionally reliable, and independent (see for instance [44], or [47] ch. 3)). However, in the case of wind power forecasting, we know there exists a correlation among forecasting errors (at least for short time-lags) [170]. Thus, we do not expect prediction intervals to be independent. Then, it appears preferable to develop an evaluation framework that is based on an alternative paradigm. We propose to consider reliability as a primary requirement and then sharpness and resolution as an added value. It should be noted here that reliability can be increased by using some re-calibration methods (e.g. conditional parametric models [172] or smoothed bootstrap [88]), while sharpness/resolution cannot be enhanced with post-processing procedures. This second aspect is the inherent (and invariant) ability of a probabilistic forecasting method to distinctly resolve future events [223].

4.7.2 Methods for the evaluation of prediction intervals

The following methods focus on the evaluation of predictive quantiles or prediction intervals of wind generation in a hierarchical manner: reliability has to be assessed first, followed by a study of sharpness and then resolution. A skill score is introduced in a second stage, which allows one to directly assess the overall quality of these predictions.

The indicator variable

Before going further with the evaluation of interval forecasts, it is necessary to introduce the *indicator variable* $\mathcal{I}_{t,k}^{(\alpha)}$ (following the definition by Christoffersen [44]), which is defined for a prediction made at time t and for the horizon k as follows

$$\mathcal{I}_{t,k}^{(\alpha)} = \mathbf{1}_{p_{t+k} \in \hat{\mathrm{I}}_{t+k/t}^{(\alpha)}} = \begin{cases} 1, & \text{if} \quad p_{t+k} \in [\hat{L}_{t+k/t}^{(\alpha)}, \hat{U}_{t+k/t}^{(\alpha)}] \\ 0, & \text{otherwise} \end{cases} . \tag{4.69}$$

This indicator variable tells if the actual outcome p_{t+k} at time $t+k$ lies ("*hit*") or not ("*miss*") in the prediction interval estimated for that lead time. We would like to mention that this definition of the indicator variable can be easily adapted when working with quantiles of a probabilistic distribution. Indeed, the value of p_{t+k} lying or not inside the interval is replaced by the test of p_{t+k} being below or above the estimated quantile $\hat{r}_{t+k/t}^{(\alpha)}$. Then, $\mathcal{I}_{t,k}^{(\alpha)}$ can alternatively be defined with

$$\mathcal{I}_{t,k}^{(\alpha)} = \mathbf{1}_{p_{t+k} \leq \hat{r}_{t+k/t}^{(\alpha)}} = \begin{cases} 1, & \text{if} \quad p_{t+k} \leq \hat{r}_{t+k/t}^{(\alpha)} \\ 0, & \text{otherwise} \end{cases} . \tag{4.70}$$

Let then define as $n_{k,1}^{(\alpha)}$ the sum of hits and $n_{k,0}^{(\alpha)}$ the sum of misses (for a given horizon k)

over the N_T realizations:

$$n_{k,1}^{(\alpha)} = \#\{\mathcal{I}_{t,k}^{(\alpha)} = 1\} = \sum_{t=1}^{N_T} \mathcal{I}_{t,k}^{(\alpha)}, \tag{4.71}$$

$$n_{k,0}^{(\alpha)} = \#\{\mathcal{I}_{t,k}^{(\alpha)} = 0\} = N_T - n_{k,1}^{(\alpha)}. \tag{4.72}$$

It is by studying the series of indicator variable $\{\mathcal{I}_{t,k}^{(\alpha)}, t = 1, \ldots, N_T\}$ over the test set that we will assess the reliability and overall skill of interval forecasts.

Reliability

The easiest way to check the reliability of interval forecasts is to compare their empirical coverage to the nominal one (i.e. the required probability (1-α)). An estimation $\hat{a}_k^{(\alpha)}$ of the actual coverage $a_k^{(\alpha)}$, for a given horizon k, is obtained by calculating the mean of the $\{\mathcal{I}_{t,k}^{(\alpha)}\}_{t=1,\ldots,N_T}$ time-series over the test set:

$$\hat{a}_k^{(\alpha)} = \frac{1}{N_T} \sum_{t=1}^{N_T} \mathcal{I}_{t,k}^{(\alpha)} = \frac{n_{k,1}^{(\alpha)}}{n_{k,0}^{(\alpha)} + n_{k,1}^{(\alpha)}}. \tag{4.73}$$

This standard measure for evaluating prediction intervals' reliability was first proposed by Ballie et al. [7] and by McNees [153]. This is the idea used in *reliability diagrams* which give the empirical coverage versus the nominal coverage for various nominal coverage rates. The closer to the diagonal the better. They can alternatively be depicted as the deviation from the 'perfect reliability' case for which empirical coverage would equal the nominal one (calculated as the difference between these two quantities). This idea is similar to the use of Probability Integral Transform histograms as proposed by Gneiting et al. [80], except that reliability diagrams directly provide that additional information about the magnitude of the deviations from the 'perfect reliability' case.

Reliability diagrams allow one to summarize the calibration assessment of several quantiles or intervals and thus to see at one glance if a given method tends to systematically underestimate (or overestimate) the uncertainty. Figure 4.3 shows an example of a reliability diagram for the evaluation of a given estimation method of wind power predictive distributions. Deviations from the 'perfect reliability' case are given as a function of the quantile nominal proportions, as an average over the forecast length. Here, one notices a rather good calibration of the method since deviations are lower than 2%. However, the fact that quantiles are slightly overestimated for proportions lower than 0.5 and slightly underestimated for proportions above that value indicates that corresponding predictive distributions are a bit too narrow.

Using that kind of comparison between the nominal and empirical coverage introduces subjectivity in the evaluation: the decision of whether the intervals have correct coverage or not is left to the analyst. This is why a more objective framework based on hypothesis testing has been introduced in the forecasting literature (mainly in econometric forecasting). For

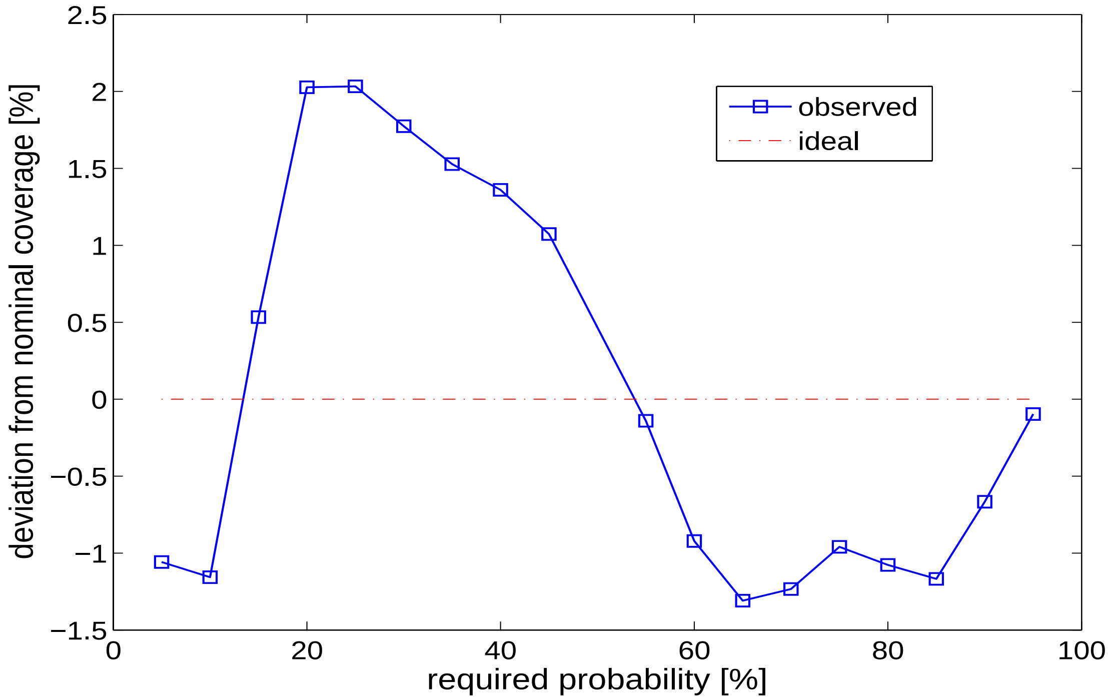


Figure 4.3: *Example of a reliability diagram depicting deviations as a function of the nominal coverage rate, for the reliability evaluation of a method providing probabilistic forecasts of wind generation.*

instance, Christoffersen [44] proposed a likelihood ratio χ^2-test for evaluating the unconditional coverage of interval forecasts of economic variables, accompanied by another test of independence. In the area of wind generation forecasting, Bremnes [22] recently used a Pearson χ^2-test for evaluating the reliability of the quantiles produced from a local quantile regression approach. However, χ^2-tests rely on an independence assumption regarding the sample data. Owing to the correlation of wind power forecasting errors, it is expected that series of interval hits and misses can come clustered together in a time-dependent fashion. This actually means that independence of the indicator variable sequence cannot be assumed in our case (except if independence is proven in a prior analysis). In such cases, serial correlation invalidates the significance level of hypothesis tests. In general, it is known that statistical hypothesis tests cannot be directly applied for assessing the reliability of probabilistic forecasts due to the either serial or spatial correlation structures [89].

Sharpness and Resolution

When dealing with sharpness or resolution, focus is given to the size of prediction intervals, or in a more general manner to the shape of predictive distributions. In the meteorological literature, the sharpness of probabilistic forecasts correspond to the ability of these forecasts to deviate from the climatological mean probabilities, whereas resolution stands for the ability of providing different conditional probability distributions $q(p|\hat{p})$ depending on the level of the predictand. For probabilistic forecasts with perfect reliability, these two notions are equivalent [223]. Here, we introduce a slightly different view of these two aspects. Given that the reliability of probabilistic forecasts is assessed in a prior analysis, we then propose to study the evolution of the shape of probabilistic distributions. Distributions that are narrower should be rewarded, since it will increase their value in a decision-making context. This is what we will regard as the sharpness of probabilistic forecasts. And, if rival probabilistic prediction methods produce distributions with a similar sharpness, then distributions whose shape exhibits larger variations over the evaluation period, hence showing a better ability for discriminating among future events, should be preferred. This is in line with the resolution aspect defined in the meteorological literature.

Define

$$\delta_{t,k}^{(\alpha)} = \hat{U}_{t+k/t}^{(\alpha)} - \hat{L}_{t+k/t}^{(\alpha)} = \hat{r}_{t+k/t}^{(1-\alpha/2)} - \hat{r}_{t+k/t}^{(\alpha/2)} \tag{4.74}$$

the size of the central interval forecast (with pre-assigned probability $(1-\alpha)$) estimated at time t for lead time $t+k$.

If two uncertainty estimation methods provide intervals at an acceptable level of reliability, we explained that it is the method that yields the narrowest intervals that is to be preferred. Here, the sharpness aspect is evaluated by calculating the average size $\bar{\delta}_k^{(\alpha)}$ of the prediction intervals for a given horizon k:

$$\bar{\delta}_k^{(\alpha)} = \frac{1}{N_T}\sum_{t=1}^{N_T} \delta_{t,k}^{(\alpha)} = \frac{1}{N_T}\sum_{t=1}^{N_T} \left(\hat{r}_{t+k/t}^{(1-\alpha/2)} - \hat{r}_{t+k/t}^{(\alpha/2)}\right). \tag{4.75}$$

Both Bremnes [22] and Nielsen et al. [171] used such a measure for evaluating the sharpness of the their probabilistic forecasts as a function of the horizon. When focusing on the distance between the quantiles for proportions 0.25 and 0.75 (i.e. the quartiles), this measure is commonly known as the inter-quartile range. However, since in a non-parametric framework probabilistic forecasts may consist in a set of prediction intervals, it would be interesting not to focus only on these two particular quantiles but also to look at the size of intervals corresponding to the very central and to the tail parts of the predictive distributions — say $\bar{\delta}_k^{(0.8)}$ and $\bar{\delta}_k^{(0.2)}$ for instance, which are the average size of the 20%- and 80%-confidence central intervals respectively.

In parallel, the resolution concept stands for the ability of providing a situation-dependent assessment of the uncertainty. If two approaches have similar sharpness, then a higher resolution translates to a higher quality of related interval forecasts. It is not possible to directly verify that property, though we can study the variation in size of the intervals by using an appropriate summary statistic such as the standard deviation $\sigma_k^{(\alpha)}$ of the interval size (for a given horizon k and nominal coverage rate $(1-\alpha)$), where

$$\sigma_k^{(\alpha)} = \left[\frac{1}{N_T-1}\sum_{t=1}^{N_T}\left(\delta_{t,k}^{(\alpha)}-\bar{\delta}_k^{(\alpha)}\right)^2\right]^{\frac{1}{2}}. \tag{4.76}$$

Because of the nonlinear and conditionally heteroskedastic nature of the wind generation process, the forecast uncertainty is highly variable and it is thus expected that the interval size also greatly varies.

Finally, δ-diagrams and σ-diagrams, which give respectively $\bar{\delta}_k^{(\alpha)}$ and $\sigma_k^{(\alpha)}$ as a function of the nominal coverage rate for a given look-ahead time k (or over the forecast length), permit to better visualize the shape (and shape variations) of predictive distributions. We will underline the interest of such diagnostic tools in the following Section.

Defining a unique skill score

As for point-forecast verification, it is often demanded that a unique skill score would give the whole information on a given method performance. Such a measure would be given by scoring rules that associate a single numerical value $\mathrm{Sc}(\hat{q},p)$ to a predictive distribution $\hat{q}$ if the event p materializes. Then, we can define as

$$\mathrm{Sc}(\hat{q}',\hat{q}) = \int \mathrm{Sc}(\hat{q}',p)\hat{q}(p)dp \tag{4.77}$$

the score under $\hat{q}$ when the predictive distribution is $\hat{q}'$.

A scoring rule should reward a forecaster that expresses his true beliefs. It is said to be *proper* if it does so. One remembers here that Murphy [165] referred to that aspect as the forecast *consistency* and stated that a forecast (probabilistic or not) should correspond to the forecaster's judgment. If we assume that a forecaster wishes to maximize his skill score over an evaluation set, then a scoring rule is said to be proper if for any two predictive distributions $\hat{q}$ and $\hat{q}'$ we have

$$\mathrm{Sc}(\hat{q}',\hat{q}) \leq \mathrm{Sc}(\hat{q},\hat{q}), \quad \forall \hat{q},\hat{q}'. \tag{4.78}$$

The scoring rule Sc is said to be strictly proper if Equation (4.78) holds with equality if and only if $\hat{q}' = \hat{q}$. Hence, if $\hat{q}$ corresponds to the forecaster's judgment, it is by quoting this particular predictive distribution that he will maximize his skill score.

If we consider that a predictive distribution $\hat{q}$ is characterized by its quantiles $\hat{\mathbf{r}} = \{\hat{r}_1,\hat{r}_2,\ldots,\hat{r}_l\}$ at levels $\alpha_1,\alpha_2,\ldots,\alpha_l$, Gneiting et Raftery [81] recently showed that any scoring rule of the form

$$\mathrm{Sc}(\hat{\mathbf{r}},p) = \sum_{i=1}^{l}\left(\alpha_i s_i(r_i) + (s_i(p)-s_i(r_i))\mathcal{I}^{(\alpha_i)} + f(p)\right), \tag{4.79}$$

with $\mathcal{I}^{(\alpha_i)}$ the indicator variable (for the quantile with proportion α_i) introduced above, s_i

non-decreasing functions and f arbitrary, was proper for evaluating this set of quantiles. Here $\mathrm{Sc}(\hat{\mathbf{r}}, p)$ is a positively rewarding score: a higher score value stands for an higher skill. The specific case of central prediction intervals corresponds to the case where only two quantiles are quoted (cf. Equations (4.4) and (4.5)). Note that for a unique quantile, the scoring rule given by Equation (4.79) generalizes the loss functions considered in quantile regression [171] and local quantile regression [22].

Actually, Gneiting and Raftery [81] also noticed that for the specific case of central prediction intervals with nominal coverage rate $(1-\alpha)$, by putting $\alpha_1 = \alpha/2$ and $\alpha_2 = 1-\alpha/2$, $s_i(p) = 4p$, $(i = 1, 2)$, and $f(p) = -2p$, one retrieves an interval score that has already been proposed by Winkler [237]. Such an interval score $\mathrm{Sc}_{t,k}^{(\alpha)}$ used for evaluating the interval $\hat{\mathrm{I}}_{t+k/t}^{(\alpha)}$ has the following form:

$$\mathrm{Sc}_{t,k}^{(\alpha)} = \begin{cases} -2\alpha\delta_{t,k}^{(\alpha)} - 4(\hat{L}_{t+k/t}^{(\alpha)} - p_{t+k}), & \text{if} \quad p_{t+k} < \hat{L}_{t+k/t}^{(\alpha)} \\ -2\alpha\delta_{t,k}^{(\alpha)}, & \text{if} \quad p_{t+k} \in \hat{\mathrm{I}}_{t+k/t}^{(\alpha)} \\ -2\alpha\delta_{t,k}^{(\alpha)} - 4(p_{t+k} - \hat{U}_{t+k/t}^{(\alpha)}), & \text{if} \quad p_{t+k} > \hat{U}_{t+k/t}^{(\alpha)} \end{cases}, \tag{4.80}$$

where $\delta_{t,k}^{(\alpha)}$ is the size of the interval forecast $\hat{\mathrm{I}}_{t+k/t}^{(\alpha)}$ as defined in Equation (4.74).

This score is appealing since it considers the size of the intervals (by rewarding tight intervals) and gives a penalty if the observation does not lie inside the estimated interval. The score is calculated at each prediction time and then averaged over the test set in order to obtain the final score value $\mathrm{Sc}_k^{(\alpha)}$ for every horizon k

$$\mathrm{Sc}_k^{(\alpha)} = \frac{1}{N_T} \sum_{t=1}^{N_T} \mathrm{Sc}_{t,k}^{(\alpha)}. \tag{4.81}$$

Using a unique proper skill score allows one to compare the overall skill of rival approaches, since scoring rules such as the one given by Equation (4.79) encompass all the aspects of probabilistic forecast evaluation. It can also be utilized as a criterion for optimizing the parameters of a given quantile estimation method. However, a unique score does not tell what are the contributions of reliability or sharpness/resolution to the skill (or to the lack of skill)[6]. Though, if reliability is verified in a prior analysis, relying on a skill score permits to carry out an assessment of all the remaining aspects, namely sharpness and resolution.

4.8 Results

The evaluation framework introduced in Section 4.7.2 is applied here for assessing the quality of the prediction intervals produced from both the linear opinion pool and the adapted resampling method. For that purpose, we have selected the statistical prediction meth-

[6] This has already been stated by Roulston et al. [197] when introducing the 'ignorance score', which despites its many justifications and properties has no ability to tell why a given method is better than another.

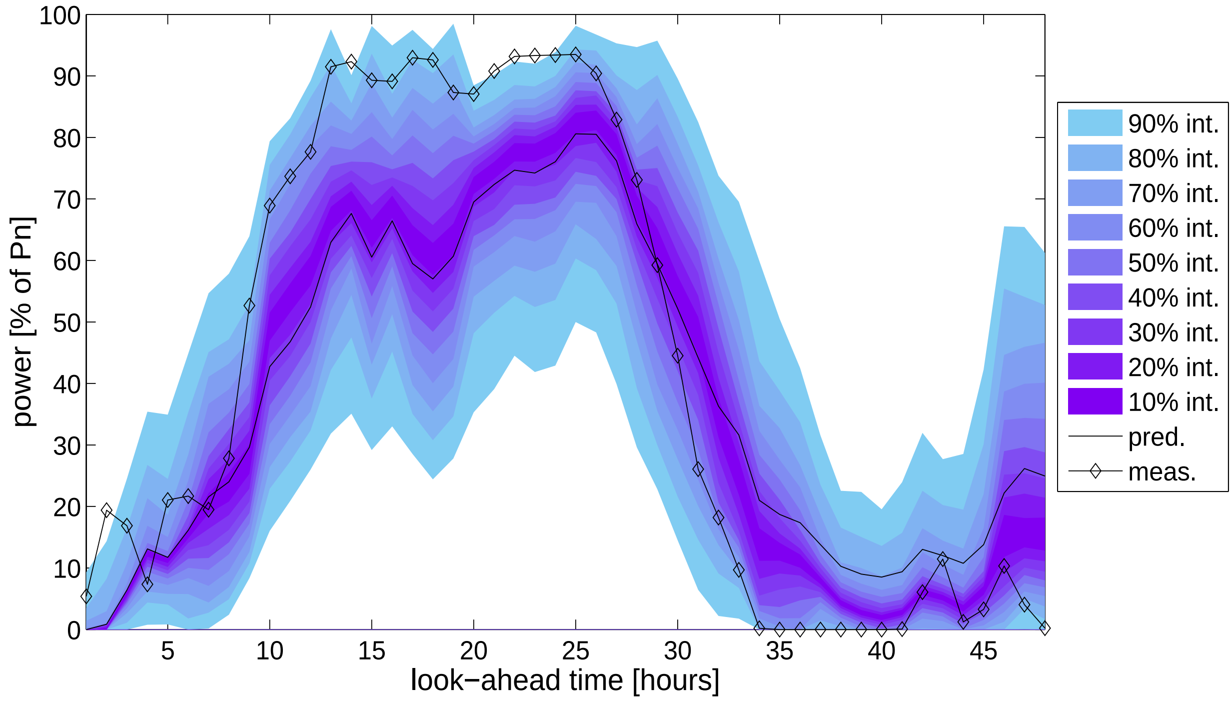


Figure 4.4: *Example of wind power point predictions associated with a set of interval forecasts. The point predictions are given by M2 and the central interval forecasts are estimated consequently with the adapted resampling method. Nominal coverage rates range from 10 to 90%. These sets of predictions and intervals were issued on the 28th March 2003 at 10:00, for the Tunø Knob wind farm.*

ods M1, M2 and M3 and their application to the Tunø Knob and Klim case-studies (see Section 3.3 for a description of the point forecasting methods and test-cases). By selecting these methods that provide predictions every hour (in contrast to every six hours for M4 and M5), there are more available data for evaluating interval forecasts. The Klim and Tunø Knob test cases consists of respectively 18943 and 3220 series of wind power point predictions and associated interval forecasts. In this Section, we assess the skill of the proposed interval estimation methods by showing and commenting on some selected results from the full verification procedure. Note that we do not consider any benchmark intervals based on an assumption about the shape of error distributions. We have already demonstrated the superiority of the proposed distribution-free approach against Box-Jenkins intervals [183] and also against intervals derived from the assumption that predictive distributions of wind generation can be modeled with β-distribution [184].

Regarding the mapping of the forecast uncertainty, since the available dataset only prove very few occurrences of cut-off events, it has not appeared appropriate to consider the non-linearity introduced by the cut-off risk. At the end of Chapter 3 we concluded on the effects of the level of predicted power on the uncertainty. We did not study the influence of meteorological variables e.g. wind direction on prediction error distributions. We assume here that these effects can be neglected since there is no evidence in the literature of their impact on forecast uncertainty. Therefore, the mapping of the forecast conditions only concerns the range of possible predicted power values.

Figure 4.4 depicts an episode consisting in a set of wind power predictions provided by M2, issued on the 28$^{\text{th}}$ March 2003 at 10:00, for the Tunø Knob wind farm. The related power measures are also shown. Moreover, a set of interval forecasts produced from the adapted resampling method is associated to the point predictions, in the form of a fan chart. The nominal coverage rates for these intervals were set to 10, 20, . . ., 90%. This illustrative example does not have any statistical value for assessing the quality of the intervals, but serves instead to show some of the nice properties of the designed approach.

When describing the uncertainty estimation methods, we explained that these methods were non-parametric (i.e. the intervals are estimated without assuming a specified distribution), and that this would permit to produce asymmetric prediction intervals. From the example of Figure 4.4, one clearly sees that interval forecasts are not symmetric around the point predictions. Also, one verifies a comment we made in Section 4.2: since intervals are central prediction intervals, they are centered around the median of the predictive distributions of wind generation and hence do not necessarily cover the point predictions themselves (which in turn are estimates of the mean of these distributions). Therefore, when the asymmetry of error distribution is more pronounced, for low and high predicted power values for instance, the difference between the center of the intervals and the point prediction is higher. This is clear here for horizons between 35- and 45-hour ahead. Note that the developed methods for estimating interval forecasts can be straightforwardly adapted if one wants to build prediction-centered intervals — this is done by considering separately distributions of positive and negative prediction errors.

Moreover, the effects of both the lead time and the level of predicted power can be seen from the Figure. Prediction intervals are very tight for the very first horizons, owing to the low level of predicted power and also because it is easier to predict for short-range horizons with statistical methods. Then, they get rather large when predicted power is in the medium-range: the forecast uncertainty is higher in such a case. Finally, they become narrower for horizons between 37- and 45-hour ahead, since predicted power is again at a low level. However, for the very last look-ahead times, one notices that intervals for nominal coverage rates greater than 80% have high upper bounds. This reflects the possibility of large negative prediction errors, even if such errors are unlikely.

4.8.1 Linear opinion pool vs. Adapted resampling

Two approaches for the combination of error distributions have been introduced, based on the linear opinion pool and adapted resampling methods. While the first one is based on the weighting of the probability distributions in a probabilistic sense, the second uses the weights from the fuzzy inference model (4.36) for defining the share of each sample in the multi-sample resampling scheme. Our first aim is to compare the intervals resulting from these two approaches. For that purpose, we evaluate the quality of the interval forecasts produced from the point predictions given by M1, M2 and M3 for the two wind farms. In all the following evaluation exercises, the predictive distributions of wind generation will consist in a set of interval forecasts with nominal coverage rates ranging from 10 to 90%, with an increment of 10%. Regarding the set-up of both methods, the mapping of the forecast uncertainty is done by dividing the range of possible predicted power values into five zones, to which we associate triangular fuzzy sets. The size of the error samples is set to 300. Finally, we consider the case of 50 bootstrap replications for the adapted resampling approach.

Focus is given first to the reliability aspect, since we expect that the choice between the two approaches for combining probability distributions will mainly have an effect on the reliability of the resulting predictive distribution quantiles. Therefore, we estimate the actual coverage of the predictive distributions, and summarize this information in reliability diagrams that give the difference between the empirical and the nominal coverage rates, for the various estimated quantiles. Figures 4.5 and 4.6 depict the results for Tunø Knob and Klim respectively. These two diagrams are for the whole forecast length: displayed values correspond the average deviations over all the prediction horizons.

Consider in a first stage Figure 4.5 for explaining how to read such reliability diagrams and what kind of conclusions can be derived from their study. The x-axis gives the required probability, i.e. the nominal coverage rate of the predictive quantiles, and the various curves display the deviation (in %) from the 'perfect reliability' situation for which the empirical coverage of the quantiles would equal the nominal one. This ideal situation is represented by the dash-dot straight line. Then, a +1%-deviation for the quantile with nominal coverage rate 30% (for instance) actually tells that the empirical coverage estimated with Equation (4.73) is equal to 31%. Figures in the legend correspond to the average absolute deviation from the ideal case, over the range of nominal coverage rates (and also over the forecast length).

For the Tunø Knob case-study, the deviations from 'perfect reliability' are contained in a $\pm 3\%$ envelope whatever the considered point prediction method or the interval estimation approach. The reliability of the intervals could be expected to be lower for low and high nominal coverage rates since it is harder to model the very tails of error distributions. This is not the case here. However, one notices a general trend, which is that quantiles for proportions below 0.5 are overestimated while quantiles above the median are underestimated. Prediction intervals are slightly too narrow on average. It should be understood here that having too narrow intervals is more likely than having too large intervals: methods for estimating future uncertainty usually rely on past experience of a given model performance and therefore do not integrate the additional uncertainty of predicting new data [41]. Average absolute deviations are between 0.86 and 1.23%, with slightly better results obtained from the application of the linear opinion pool approach. In a general manner, we conclude on an acceptable reliability of the probabilistic forecasts produced by both methods.

The Klim test case consists in a longer evaluation period (almost 19.000 series of two-days ahead forecasts) and can thus give more insight on the reliability of predictive distributions. In Figure 4.6, deviations from nominal coverage are in general lower than the ones witnessed when studying Tunø Knob. Average absolute deviations range from 0.32 to 0.89% only. These deviations are significantly lower for intervals estimated with the adapted re-

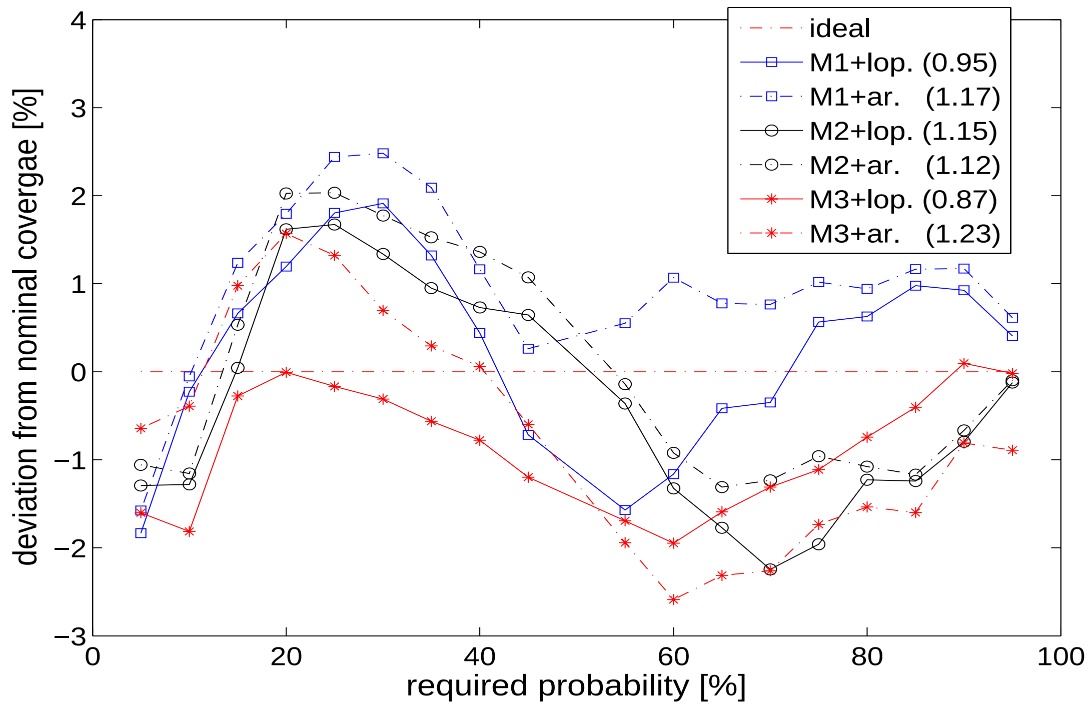


Figure 4.5: *Reliability diagrams for Tunø Knob. Results are given for the three point prediction methods and for the two implemented approaches (lop: linear opinion pool; ar: adapted resampling).*

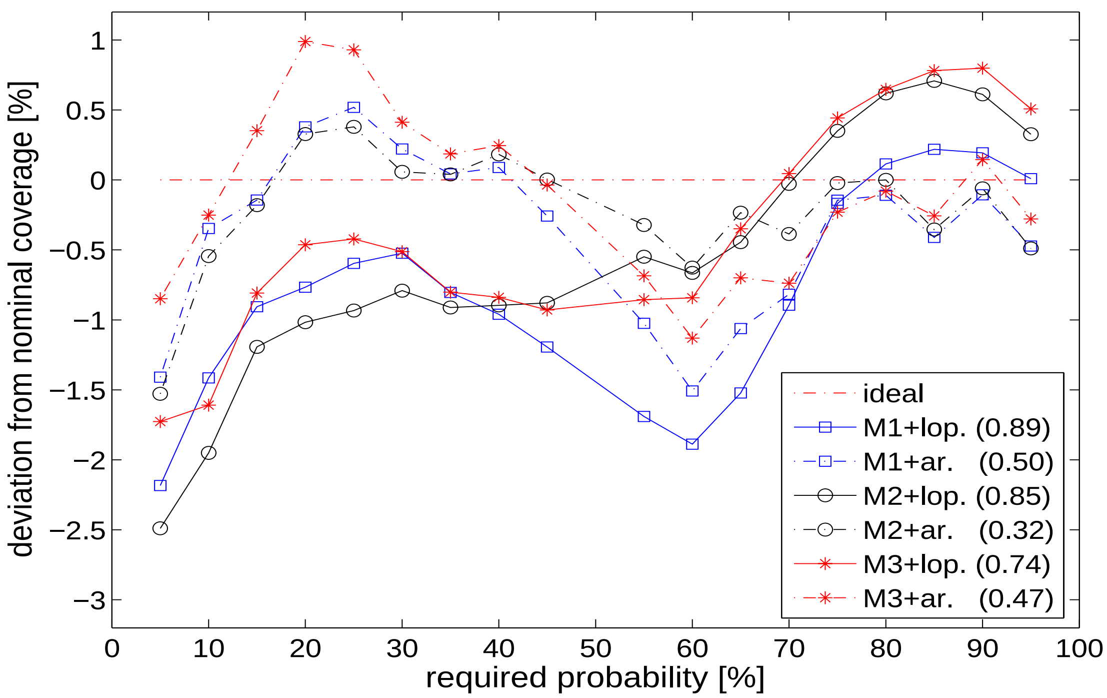


Figure 4.6: *Reliability diagrams for Klim. Results are given for the three point prediction methods and for the two implemented approaches (lop: linear opinion pool; ar: adapted resampling).*

sampling approach. There is a trend that the linear opinion pool quantiles underestimate the true quantiles (cf. left part of the reliability diagrams). The difference between the two approaches is less pronounced for quantiles with proportions higher than 0.5. Even if the calibration of adapted-resampling quantiles appears better than the one of linear-opinion-pool quantiles, we consider here also that both approaches yield reliable interval forecasts.

The second stage of the evaluation of the predictive distributions is carried out by using a scoring rule of the form of Equation (4.79) where the s_i functions are such that $s_i(p) = 4p$, $(i = 1, \dots, 18)$, and $f(p) = -2p$ following Gneiting and Raftery [81], calculated as a function of the look-ahead time. The resulting score summarizes the skill of the predictive distributions described by the 18 quantiles estimated from both interval estimation methods. Given that we have accepted quantiles to be reliable (even if it is only a subjective result), the positively-oriented score can tell which method (and also which point prediction method used as input) leads to the 'best' predictive distributions.

Figure 4.7 gives the evolution of the skill score as a function of the horizon for Tunø Knob. Figures in the legend correspond to the average skill score values over the forecast length. The skill score steadily decreases as the look-ahead time augments. This meets the general statement that it is harder to predict for lead times further in the future, which was already discussed and illustrated in Section 3.5.2 for the case of wind power point forecasts. Also, we see that the skill scores of predictive distributions generated from the linear opinion pool and adapted resampling approaches are rather close: they actually coincide when point predictions are provided by the M3 method, though there are significant differences for the cases of M1 and M2. The average values shown in the legend tell that adapted-resampling predictive distributions are better than the ones resulting from the linear-opinion-pool combination approach.

The way the skill score evolves as a function of the look-ahead time for the Klim case-study is shown in Figure 4.8. Similarly, the skill score values of predictive distributions are rather close, with a slight advantage for the ones estimated with adapted resampling. But, an interesting point is that the choice of a given point prediction method as input has an influence on the quality of the resulting predictive distributions. Indeed, it appears that considering M1 leads to better probabilistic forecasts for Tunø Knob (but not over all the forecast length), whereas considering M2 is better for Klim. Note that since the predictive distributions are actually estimations of the error distributions related to point forecasts, a point prediction method with sharper error distributions will yield sharper probabilistic forecasts. Though, this comment is of course only valid if the prediction interval estimation approach has a real ability to reflect the error distribution associated to a given point forecast.

To conclude on that comparison of the two approaches for the probabilistic distribution combination problem, we can say that predictive distributions estimated from the adapted resampling approach prove to have a higher skill than the ones resulting from the more classic linear opinion pool approach. On the case-study with the longer evaluation period, the reliability of adapted resampling quantiles is significantly higher. Also, for both case-studies and for the three point prediction methods considered as input, this method leads

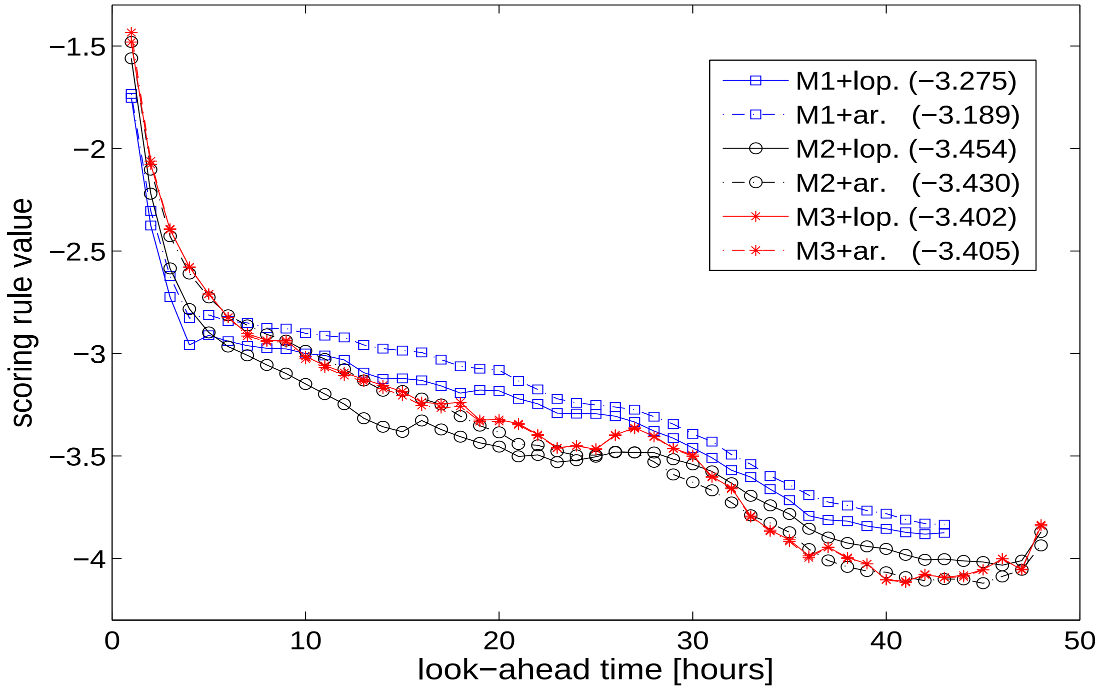


Figure 4.7: *Skill score as a function of the horizon for Tunø Knob. Results are given for the three point prediction methods and for the two implemented approaches (lop: linear opinion pool; ar: adapted resampling).*

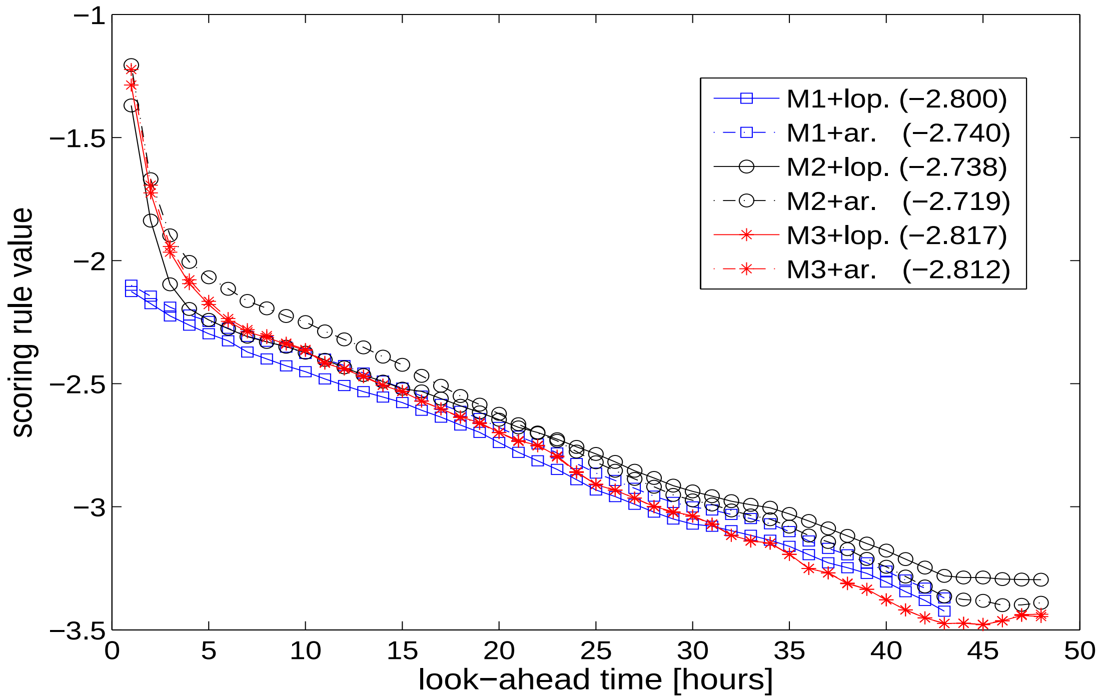


Figure 4.8: *Skill score as a function of the horizon for Klim. Results are given for the three point prediction methods and for the two implemented approaches (lop: linear opinion pool; ar: adapted resampling).*

to higher values of the skill score, which encompasses all the aspects of the evaluation of probabilistic forecast quality. This is why we will focus on this approach in the following Paragraphs, and illustrate the influence of its degrees of freedom, on reliability, sharpness and resolution. This study comprises a sensitivity analysis of its performance and will result in general guidelines for its application to further case-studies or alternatively for online forecasting exercises.

4.8.2 Influence of the fuzzy mapping of the forecast conditions

The idea of introduced method is to propose a situation-dependent assessment of the forecast uncertainty: fuzzy logic is used for mapping several zones with different characteristics of the prediction error distributions. As explained previously, we concentrate here on the variation of the forecast uncertainty as a function of the level of predicted power. The range of possible predicted power values is divided into several ranges, to which we associate triangular fuzzy sets. It is expected that increasing the number of fuzzy sets will mainly have a positive effect on increasing the resolution of predictive distributions. In a general manner, increasing the resolution of probabilistic forecasting methods will augment their value for the management or the trading of wind generation (as long as they are still reliable). Three possibilities are envisaged: using only one fuzzy set on the power range (which is equivalent to using the classical Williams-Goodman empirical approach, cf. Paragraph 4.4.1), and a mapping with alternatively 3 or 5 fuzzy sets. We set the sample size to 300 elements and the number of bootstrap replications to 50. The considered case-study for that sensitivity analysis is Tunø Knob. The point predictions used as input are provided by the M2 method.

For assessing in a first stage the reliability of the probabilistic forecasts produced with these three settings, we use the reliability diagrams depicted in Figure 4.9. The deviations from nominal coverage are given as the average deviations over the whole range of look-ahead times. The figures given in the legend are the average absolute deviations from 'perfect reliability' (over the various nominal coverage rates and forecast horizons).

One sees from Figure 4.9 that the deviations from 'perfect reliability' are of the same order for the various settings: they are within $\pm 2.5\%$. Even if it is not the primary aim of the mapping, it seems that using several fuzzy sets permits to increase the overall reliability of estimated quantiles. In this example, the average absolute deviation is 1.40 % for Williams-Goodman intervals, whereas it is equal to 1.05 and 1.12% for the two other settings with 3 and 5 fuzzy sets respectively.

Then, we turn our attention to sharpness and resolution. Since sharpness proved to be similar for the three settings and that the power curve mapping is mainly expect to impact the resolution of predictive distributions, we focus here only on the latter quality aspect. We base our evaluation of the resolution property of the intervals on the σ-diagrams depicted in Figure 4.10, which give the standard deviation of the interval size as a function of the interval nominal coverage rate. As an example, we focus on the σ-diagrams of 24-hour ahead probabilistic forecasts, but similar conclusions could be derived if looking at σ-diagrams for 24-hour ahead probabilistic forecasts, but similar conclusions could be derived if looking at σ-diagrams for other look-ahead times. When going from Williams-Goodman to adapted

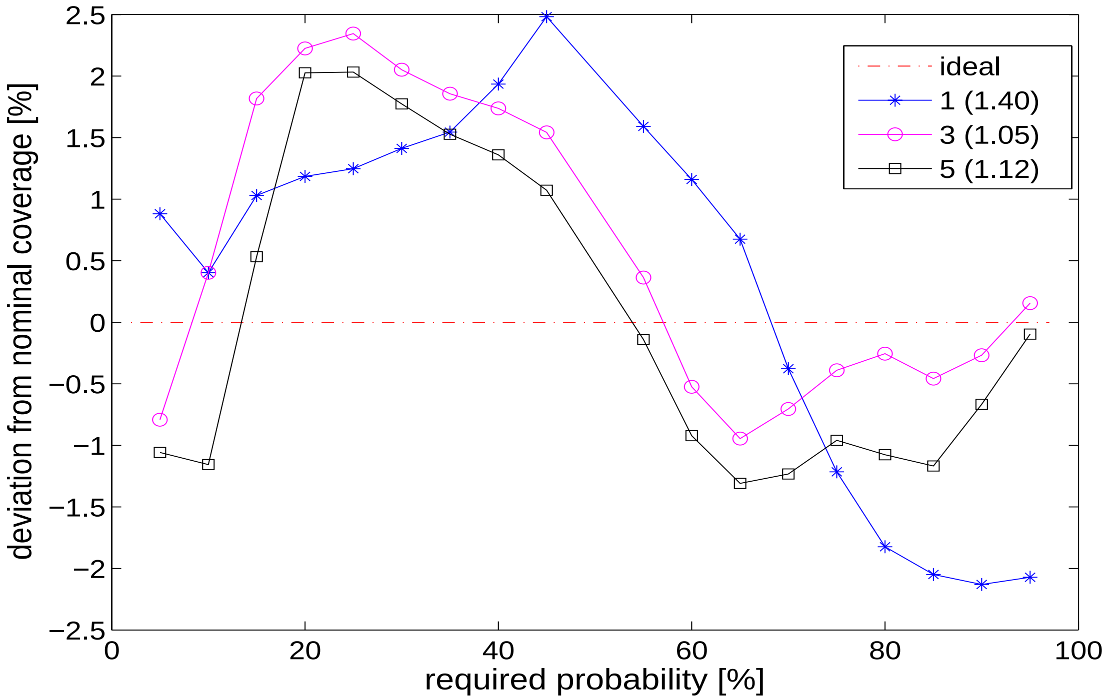


Figure 4.9: *Reliability diagrams for evaluating the influence of the power curve mapping on the resulting probabilistic forecasts' reliability.*

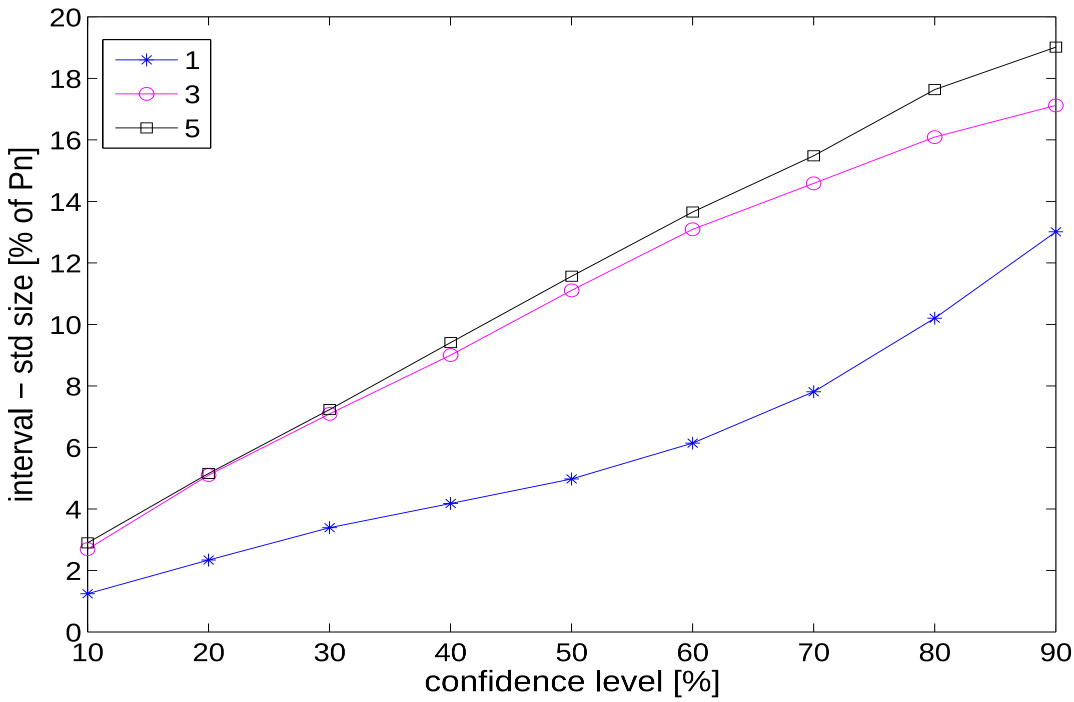


Figure 4.10: *σ-diagrams for 24-hour ahead forecasts for evaluating the influence of the power curve mapping on the resulting probabilistic forecasts resolution.*

resampling intervals, the resolution is significantly augmented, whatever the degree of confidence. For the example of the 50%-confidence prediction interval, the standard deviation of the interval size is actually multiplied by a factor 3. Also, one sees that by using more fuzzy sets for mapping the power curve, the resolution can be increased even more, mostly for high degrees of confidence. This means that the method has a better ability to differentiate the tail ends of predictive distributions. However, considering 3 or 5 subsets of forecast conditions leads to the constitution of the same number of error samples (3 and 5 samples of prediction errors respectively). Therefore increasing the method resolution has a cost, which is the time needed for filling the error samples. Here a minimum of 900 and 1500 series of point predictions respectively are necessary for filling the samples. Note that this is does not appear to be a restriction for the application of the method, since intervals can be estimated even if all the samples are not full. The only consequence is that predictive distributions may not be as reliable as it would be expected in a first period of the application. This is certainly a reason why predictive distributions produced with the 5-fuzzy-set configuration are slightly less reliable than the ones from the 3-fuzzy-set configuration here. Finally, even if we have focused on the resampling method, we have noticed that the fuzzy mapping of the forecast conditions has a similar influence on the skill of the linear opinion pool approach.

4.8.3 Influence of the sample size

The second part of the study concerns the influence of the sample size, i.e. the number of past prediction errors, on the skill of the estimated intervals provided by the adapted resampling method. Intuitively, considering more past errors should permit to better understand the uncertainty of the process and thus to augment the reliability of estimated predictive distributions. However, relying on very large error samples would make the method less dynamic. Here, the number of fuzzy sets is set to five and the number of bootstrap replications to 50. We produce probabilistic forecasts with error samples containing 50, 100, 200 and 300 elements.

The reliability diagrams displayed in Figure 4.11 show how the sample size affects the predictive distributions' reliability. The absolute average deviation from 'perfect reliability' greatly diminishes as we use more elements in the adapted resampling procedure. This absolute average deviation is divided by 2 if considering the last 300 errors instead of dealing with the last 50 only. Also, one notes from the reliability diagrams that the trend to have too narrow intervals (i.e. overestimated quantiles if below the median and underestimated if above the median) diminishes when the sample size is increased.

For evaluating the sharpness of the interval forecasts, we superpose δ-diagrams for the various method settings (Figure 4.12). This Figure is for 24-hour ahead probabilistic forecasts. In a general manner, the average interval size ranges from $\sim$5% of nominal power for intervals at a 10% degree of confidence to $\sim$50% for those associated with a 90% degree of confidence. Whatever the nominal coverage rate, the average size decreases when considering more past prediction errors for estimating predictive distributions. This diminution in the mean size is up to 10% when going from 50 to 300 sample elements. Therefore, by in-

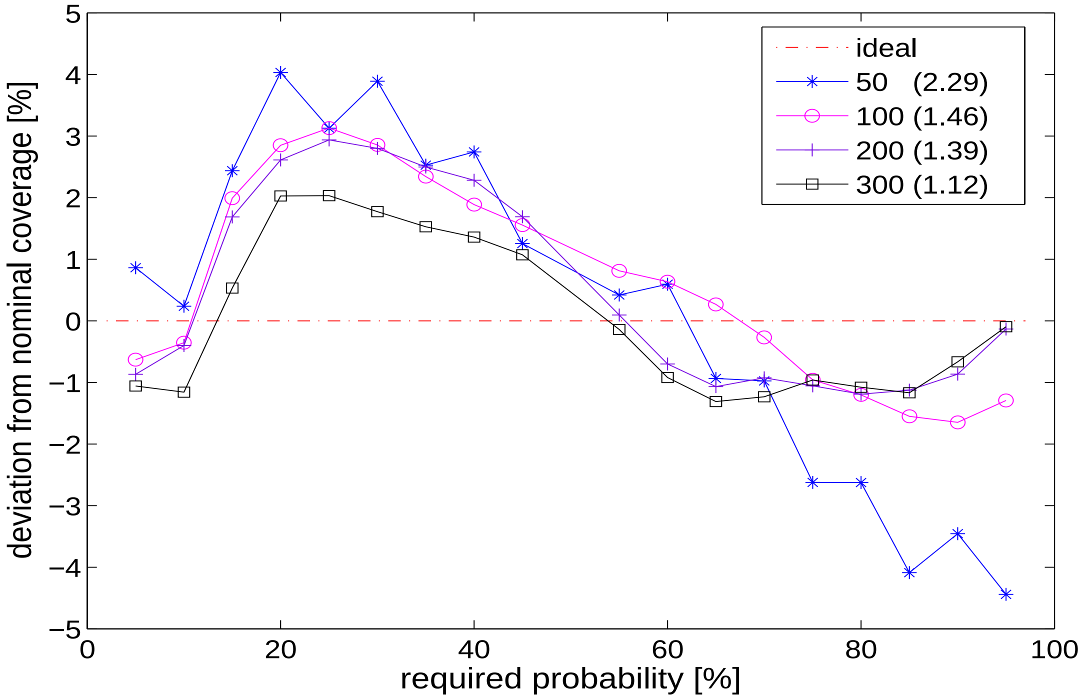


Figure 4.11: *Reliability diagrams for evaluating the influence of error sample size on the resulting probabilistic forecasts' reliability.*

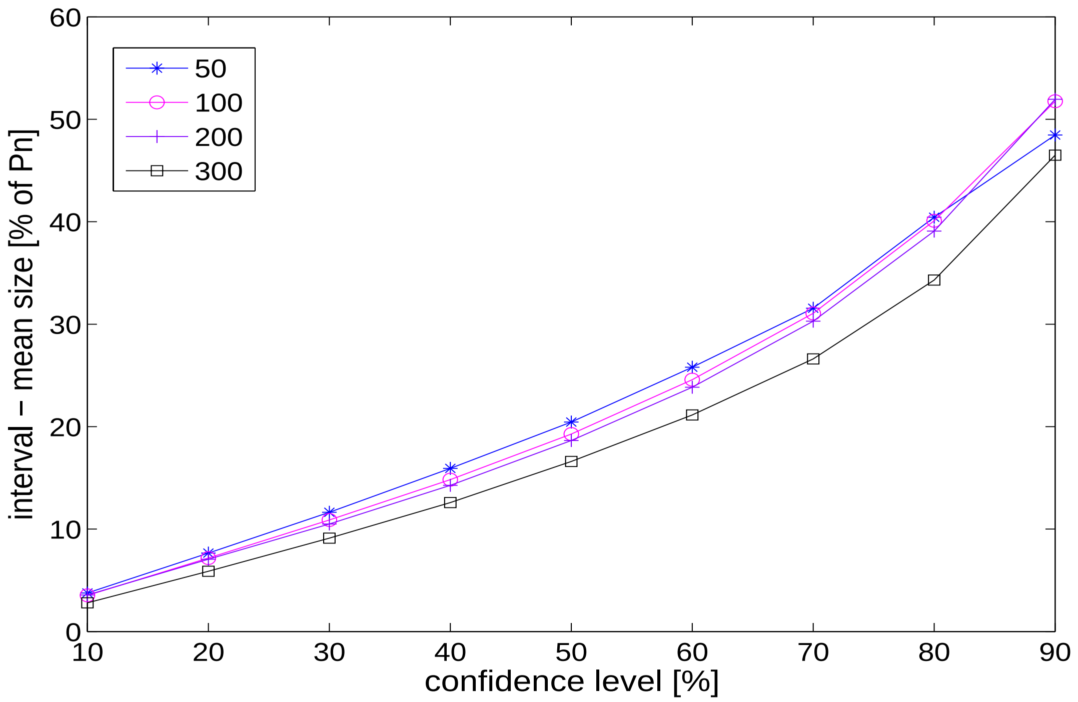


Figure 4.12: *δ-diagrams for 24-hour ahead forecasts for evaluating the influence of the error sample size on the resulting probabilistic forecasts' sharpness.*

creasing the sample size, we improve both the forecasts reliability and their sharpness. That parameter does not have a significant effect on the resolution of predictive distributions.

This study of the influence of the sample size also tells what can be the expected quality of probabilistic forecasts in a first period of an online forecasting exercise, as error samples are filled when new point predictions are provided. At the very beginning, error samples are empty and the approach we have developed cannot be used for estimating interval forecasts. However, one understands by looking at Figures 4.11 and 4.12 that an acceptable performance is already attained with a minimum number of 50 elements. Therefore, defining a larger sample size does not affect the quality of estimated interval in that sample-filling period. Note that one may also envisage to extract error samples from an offline forecasting exercise with the wind farm of interest, and use these samples for initializing the prediction interval estimation method for the online application.

4.8.4 Influence of the number of bootstrap replications

In the last part of the present sensitivity analysis, we turn our attention to the influence of the number of bootstrap replications on the quality of resulting predictive distributions of wind generation. For the adapted resampling method, the number of bootstrap replications correspond to the number of combined error samples created by following the fuzzy inference model (4.36). This degree of freedom is not present in the linear opinion pool approach. Augmenting the number of bootstrap replications then translates to considering more alternative scenarios for estimating the quantiles of predictive distributions. For better illustrating the influence of that parameter, we focus on smaller samples or errors. Here, that sample size is set to 50. The range of possible predicted power values is still mapped with 5 triangular fuzzy sets.

Primarily, we concentrate on the way the reliability of predictive distributions evolves with the number of bootstrap replications. Figure 4.13 gathers the reliability diagrams of these distributions when estimated after 1, 10, 100 and 1000 resampling steps. Again, average absolute deviations from 'perfect reliability' are given in the legend. Results are again for the whole forecast length. One notices that reliability is significantly increased when augmenting the number of bootstrap replications, up to 100 replications. However, it seems that increasing that parameter value is not necessary, since reliability remains at a similar level. Figure 4.14 then depicts the evolution of the skill score with the forecast horizon, where the skill score is defined in the same manner than in Paragraph 4.8.1. The curves for the various number of bootstrap replications are quite close, but one sees from the average values in the legend that the score values augment when considering more resampling steps. Better reliability contributes to augmenting the overall quality of predictive distributions. And, by considering 1000 resampling steps this overall quality is even slightly higher, certainly because probabilistic forecasts get sharper. But, as resampling methods are CPU-demanding, it is not desirable to use more and more bootstrap replications if not necessary. Therefore, in the frame of an online application, one has to find a trade-off between the desired quality and the time that may be needed for computing probabilistic forecasts.

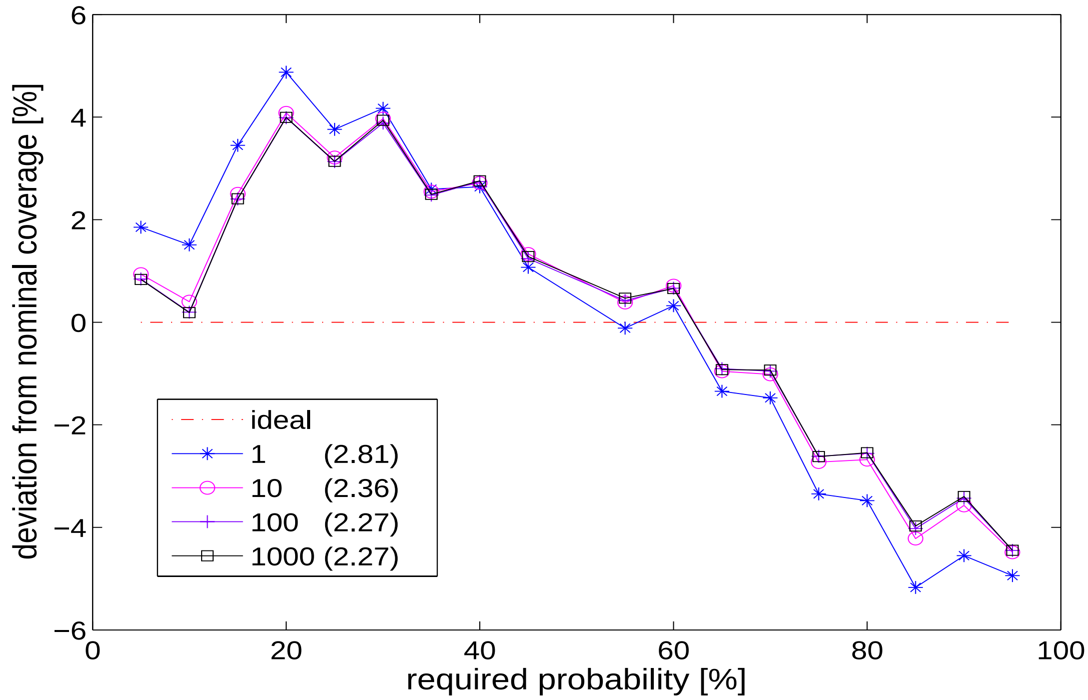


Figure 4.13: *Reliability diagrams for evaluating the influence of the number of bootstrap replications on the resulting probabilistic forecasts' reliability.*

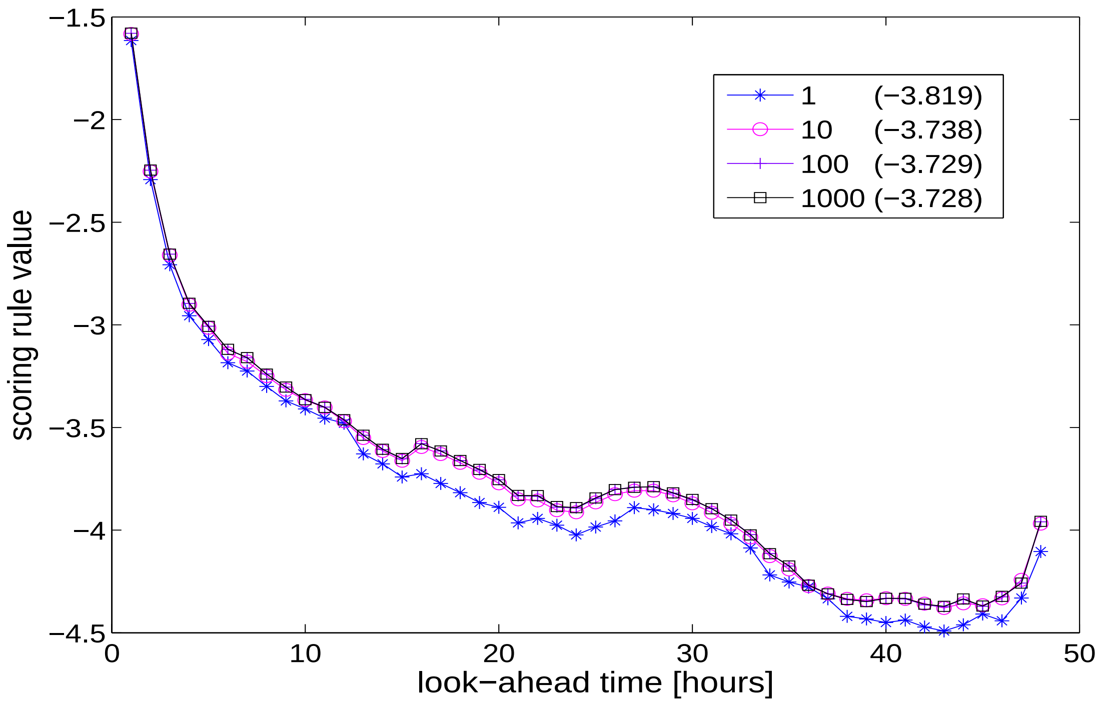


Figure 4.14: *Influence of the number of bootstrap replications on the overall quality of predictive distributions. Overall quality is assessed with the skill score, given as a function of the forecast horizon.*

4.9 Conclusions

A generic method for assessing online the uncertainty of short-term wind power forecasts in the form of prediction intervals has been introduced. The developed method is designed for the case of nonstationary, nonlinear and bounded time-series of prediction errors. It has a non-parametric and empirical nature, since it considers recent prediction errors for estimating predictive distributions of wind generation. A great advantage of that method is that it permits to construct the whole distribution of errors at once, and can thus be used for estimating several prediction intervals without needing several models. Also, by having this empirical view, we encompass all the sources of uncertainty that may come from the input data, the chosen prediction model and its parameters, etc. However, the proposed method requires a subject-matter expertise of the process of interest, since the mapping of the forecast conditions related to various characteristics of the prediction error distributions has to be carried out by the analyst. Here, that expertise follows from the study of the characteristics of the state-of-the-art point prediction methods described in Chapter 3. A fuzzy inference model is proposed, which permits to produce conditional error distributions given the forecast conditions, in the form of combined probability distributions. Predictive distributions of wind generation are then obtained by dressing point predictions with related conditional distributions of forecast errors. Two approaches for the combination of probability distributions resulting from the fuzzy inference model have been described. On the one hand, we applied the so-called linear opinion pool, which is a classical method in the probability combination literature. On the other hand, we have introduced an original approach referred to as adapted resampling. Such a method follows from the basic idea of resampling methods, which consists in thinking that more information can be extracted from a sample of data by cleverly going through that sample a certain number of times. In our case, the multi-sample resampling scheme is used for estimating the quantiles of the combined error distributions from samples representing the individual ones.

We have thoroughly demonstrated the quality of this method for the estimation of prediction intervals of wind power by evaluating its statistical performance. For that purpose, we have gathered a set of relevant skill scores, measures and diagrams, in a non-parametric framework suitable for assessing the developed method's properties. These properties include the reliability, sharpness and resolution of interval forecasts. The verification framework allowed us to conclude on the superiority of the adapted resampling method on the linear opinion pool approach. Also, we have considered some of the criteria for illustrating the influence of the method parameters on the various required properties for prediction intervals: mapping the forecast conditions increase the resolution of the resulting probabilistic predictions, while augmenting the sample size or the number of bootstrap replications mainly has an effect on the reliability property. Finally, we have given guidelines regarding the method configuration if applied for online forecasting exercises, since it definitely has an operational nature.

It was clearly shown that the approach is suitable for application to current state-of-the-art point prediction methods of wind generation. Our opinion is that the quality (more precisely the sharpness) of predictive distributions produced from such methods is bounded by the quality of the point predictions used as input. This follows from the fact that here probabilistic forecasts are based on the modeling of the point predictor's error distributions. The sharper these distributions, the sharper the resulting probabilistic forecasts. Therefore, further research works should go towards direct probabilistic forecasting of wind generation, in order to verify if by releasing the constraint of using power point predictions as input one can further increase the quality of predictive distributions.

Normally, forecasts should be evaluated in a specific decision context, i.e. their quality should be assessed in terms of the gains and losses that result from their use in a particular application. Here, we have carried out an evaluation of interval forecasts based on their statistical performance. In Chapter 6, we will consider their use for a specific end-user application, that of bidding in an electricity market. We will then assess the value of predictive distributions of wind generation for a wind power trader.

5

Ensemble Predictions of Wind Power for Skill Forecasting

Abstract

In this Chapter, we investigate on alternative approaches for providing uncertainty estimates associated to point predictions of wind generation. More precisely, focus is given to prediction risk indices aiming at giving a comprehensive signal on the expected level of forecast uncertainty. For that purpose, the use of ensemble predictions of wind generation is investigated. Such ensemble predictions consist in several alternative forecasts for the same lead time. After demonstrating the additional information provided by ensemble predictions with respect to the sole point predictions that have been considered so far, a proposal for the definition of prediction risk indices is given. Such 'skill forecasts' are based on the dispersion of the ensemble members over a set of successive horizons. We show on a real-world case-study how these prediction risk indices may be related to several levels of forecast uncertainty (and energy imbalances). Wind power ensemble predictions are derived from the transformation to power of ECMWF and NCEP meteorological ensembles, as well as by a lagging poor man's alternative.

5.1 Introduction

IT APPEARS that low quality forecasts of wind generation are partly due to the power prediction model, and partly to the numerical weather prediction system. Indeed, during some periods, the weather dynamics can be relatively more predictable, and during some other periods, they may prove to be unpredictable. Since power predictions are derived from nonlinear transformations of wind speed forecasts, a high level of uncertainty in NWPs directly translates to a high level of uncertainty in the resulting power predictions. Relying on a single deterministic forecast in periods that are deemed as unpredictable would lead to misleading decisions. Therefore, providing forecast users with an *a priori* warning on the expected level of prediction uncertainty may allow them to develop alternative strategies. Providing such signals relates to the assessment of the *weather predictability*.

In complement to point predictions of wind generation, we concentrate here on the use of ensemble forecasts, which comprise a set of alternative predictions over the period of interest. Such ensemble forecasts are produced by integrating the uncertainty in the computation of NWPs [182]. Different types of meteorological ensembles predictions are considered, provided either by the European Center for Medium-range Weather Forecasts (ECMWF) or by the National Center for Environmental Prediction (NCEP), as well as the so-called poor man's ensemble. The way ensemble predictions of wind power are obtained and how they can be interpreted is described in the first part of the Chapter. Then, focus is given to the additional information captured by such a type of predictions about expected wind generation. More particularly, an investigation is carried out first for evaluating the reduction of the level of prediction error of a single-valued prediction derived from ensemble forecasts. On the other hand, it is shown that the dispersion of the ensemble members can give valuable information on the forecast uncertainty.

In Chapter 4, the prediction uncertainty has been considered as a whole. The introduced statistical method permits to associate point forecasts of wind generation with prediction intervals. These intervals are derived from the past performance of the point prediction method. This approach does not account for the weather predictability, and hence does not make any difference between more or less predictable weather situations. Alternatively, our aim in the present Chapter is to focus on the weather predictability only, and to demonstrate how to benefit from ensemble predictions for providing a comprehensive signal on the level of uncertainty associated to point predictions of wind generation. For this purpose, we introduce *prediction risk indices* that can be used as skill forecasts, i.e. forecasts of the distributions of expected prediction errors. In an operational context, a skill forecast associated to a given point prediction may be more easily understood [244]. Also, skill forecasts are not directly related to a given prediction method: since they consist in an assessment of the inherent predictability of the weather dynamics, they are expected to provide an a priori warning whatever the considered single-valued forecast. In a second part of the present Chapter, a definition of prediction risk indices is proposed, based on the dispersion of wind power ensemble members over a set of consecutive look-ahead times. They thus provide an uncertainty estimate for a certain period of time, in contrast with the prediction intervals introduced in Chapter 4 which are pointwise[1] uncertainty estimates. The relation between these prediction risk indices and the related level of prediction error is expressed in a probabilistic manner, through the use of conditional probability diagrams.

[1]The term 'pointwise' has been introduced in Section 4.6, when explaining the difference between marginal and simultaneous interval forecasts. As pointwise is characterized an uncertainty estimate for a given look-ahead time.

5.2 Ensemble predictions of wind power

For the present study, we concentrate on the prediction of wind generation at a single location, the Tunø Knob wind farm in Denmark. That offshore wind farm has already been considered as a test case in Chapters 3 and 4. It is located in the shallow waters between Jutland and the island of Samsø. Power production data consists of 15-minutes power averages over the period for which meteorological ensemble predictions are available, i.e. over the first 10 months of 2003. Power values range from 0 to 5MW. Though, as it is done in a large part of the document, they are normalized by the wind farm rated capacity P_n.

The assessment of weather predictability follows from the analysis of *ensemble forecasts*. Such ensemble predictions correspond to multiple runs of NWP models, which differ in the initial conditions and/or in the numerical representation of the atmosphere (cf. Figure 5.1). Therefore they address the two major sources of forecast uncertainty [82]. Actually, the strength of the contribution of both deficiencies in the models and the inaccuracy of initial conditions to the forecast error is still subject to debate [181]. In parallel, the traditional single-valued forecasts provided by meteorological offices from a 'best' estimate of the initial conditions and by applying their operational prediction model are referred to as *control* or *unperturbed* forecasts. Meteorological ensembles from both ECMWF and NCEP are used in the present Chapter.

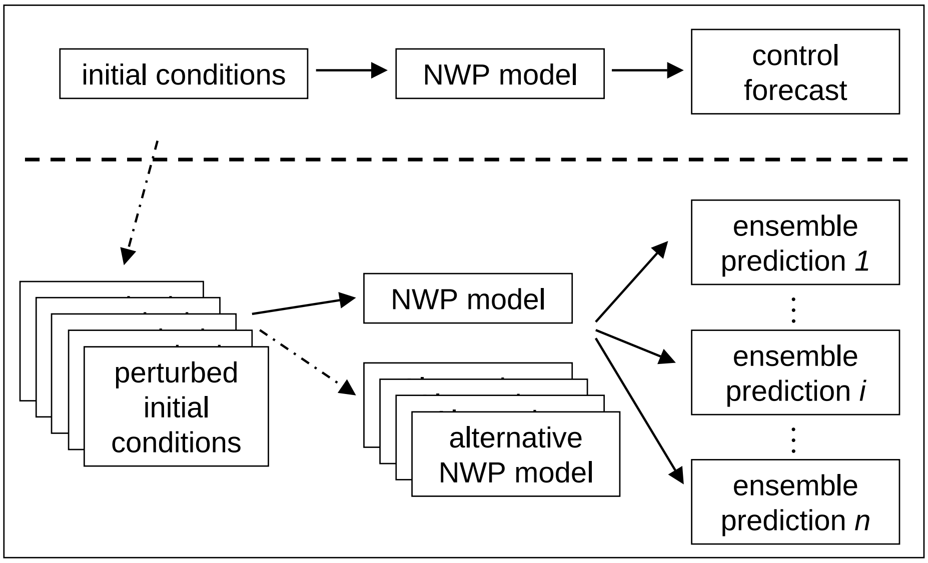


Figure 5.1: *Schematic representation of the procedure for producing meteorological ensemble predictions.*

5.2.1 The meteorological ensemble predictions from ECMWF and NCEP

The ECMWF ensembles are produced from a T255 model, which is spectral model with a truncation at wave number 255. The horizontal resolution of this model is of approximately 80km. 50 ensemble members are associated to the control forecast. They actually go by pairs: an analysis based on singular vectors is used for perturbating the initial conditions in both the positive and negative directions, leading to 25 pairs of perturbed members. Singular vectors are the perturbations with the largest energy growth during the first two days [161]. This growth is assumed to be linear in time. The perturbations of the initial conditions are given by linear combination of the singular vectors: because singular vectors tend to be localized in space, they are combined such that they have a more uniform spatial distribution over the considered area. Note that in addition to that representation of the uncertainties in the initial conditions, the ECMWF ensemble prediction system also integrates a stochastic representation of the model uncertainties [30]. The input meteorological ensemble for the present study have a maximum look-ahead time of 168-hour ahead (7 days), with a temporal resolution of 6 hours. Calculations are initiated everyday at 12:00. In an operational context, forecasts are available 17 hours after initial time.

NCEP meteorological ensembles are obtained by applying the breeding method, which is an alternative to singular vectors for estimating the subspace of fastest growing perturbations [222]. This method simulates the development of growing errors in the analysis cycle. A set of ensemble members is created by adding or subtracting bred modes to the unperturbed analysis. Note that the NCEP ensembles only attempts to sample the analysis error and do not account for model deficiencies. In our case, this set is composed of the unperturbed forecasts plus 10 perturbed members. The horizontal resolution is of approximately 100km (T126 truncation). NCEP ensembles with such a resolution are issued once a day at 00:00. The forecasts have a 6-hour temporal resolution up to 84-hour ahead. An operational advantage of NCEP ensembles with respect to the ECMWF ones is their ease of implementation, and consequently a shorter delay between initial time of computation and actual forecasts delivery. More details on both NCEP and ECMWF approaches are given in [29].

5.2.2 Conversion to ensembles of wind power

Either from ECMWF or NCEP, the meteorological variables of interest are some of the near-surface weather variables, i.e. wind speed and direction at 10m a.g.l. Both dataset consist of 300 ensemble predictions, for a period covering the first 10 months of 2003. Power predictions are produced with the statistical method described by Nielsen et al. [172]. Because the NWP temporal resolution is of 6 hours, and power measures are averages over 15 minutes, the meteorological forecasts are linearly interpolated in order to have them corresponding to the available observations. However, it is obvious that such an interpolation cannot increase the variability of power predictions to the level one could expect from the wind generation process on a 15-minute basis. It would have been a better compromise to convert both to an hourly resolution. The 15-minute resolution is motivated by the end-user requirements in the Danish case-study we consider here. The statistical method is characterized first by a logarithmic transformation of estimated power values to force these estimates to span the whole range of possible values (i.e. between 0 and nominal power). Moreover, in addition to being a function of wind speed and direction, the power curve model accounts for the prediction horizon. The non-parametric approach for power curve modeling and coefficient estimation is described in [178]. From now on, the ECMWF-based and NCEP-based ensembles predictions of wind generation will be referred to as *ECMWF-EPW* and *NCEP-EPW*.

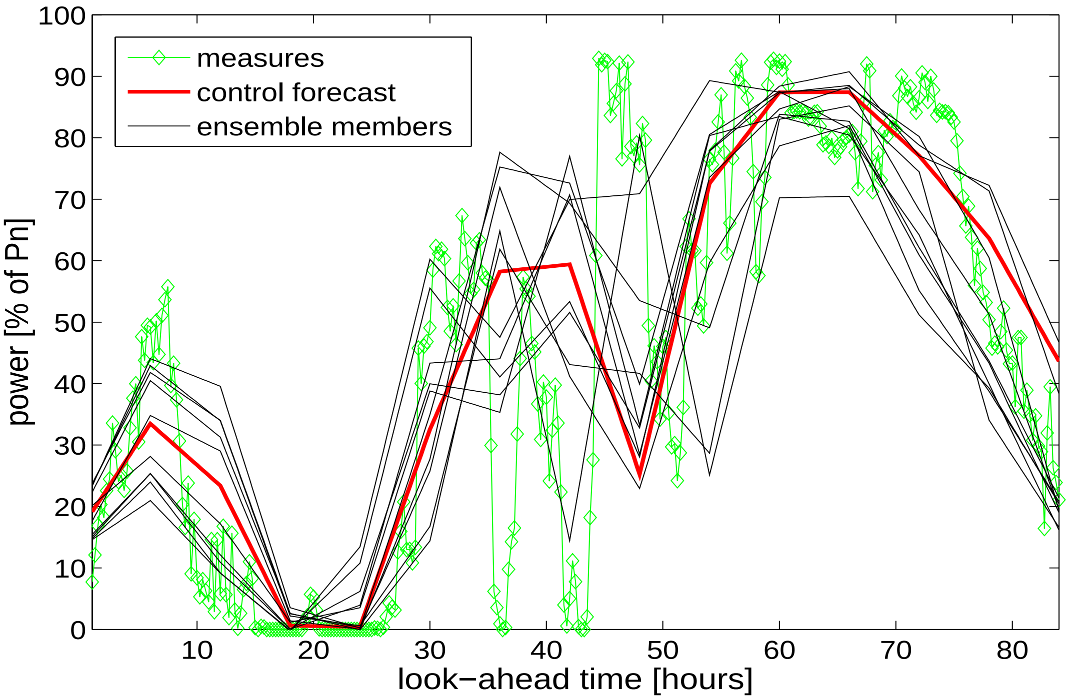


Figure 5.2: *Example of ensemble predictions of wind generation. Meteorological ensembles are provided by NCEP. They are then used for wind power forecasting following the method described in [172]. The ensemble predictions consist of the control forecast and 10 perturbed members.*

Figure 5.2 displays an episode with point predictions of wind generation produced from a NCEP control forecast (with a maximum look-ahead time of 84-hour ahead), compared with power measurements. Are also depicted the 10 NCEP-EPW ensemble members. The general agreement between the control predictions and measures is rather good over the whole forecast length, except for prediction horizons from 35 to 50-hour ahead. In parallel, one notices that the agreement between the ensemble members varies throughout this episode. Both the meteorological ensemble prediction system and the nonlinear model that is applied for converting the meteorological variables to power contribute to the variability of the discrepancy among ensemble members. The former reflects the growth of the uncertainty in the estimation of the initial state for the NWP model while the latter amplifies or dampens that growth depending on the level of predicted power, thus accounting for the nonlinear nature of the energy conversion process. For instance, when predicted wind generation is low the envelope containing the alternative predictions is much tighter (horizons between 16 and 26-hour ahead in the Figure). And, for look-ahead times from 35 to 55-hour ahead, the disagreement between members is more pronounced, with a criss-crossing of some of these members, while some others stay at a similar level. Hereafter, we will explain how wind power forecasting can benefit from these sets of alternative predictions.

5.2.3 Poor man's ensembles of wind power

Finally, as a gratis alternative to the NCEP-based and ECMWF-based ensemble predictions of wind power, we also consider the so-called *temporal ensembles* (sometimes referred to as poor man's lagging ensembles). Such ensembles consist in a set of forecasts with common lead time, but issued at different time origins. Therefore, these forecasts are obtained from different initial conditions, but with the same prediction model. Here, temporal ensembles are made up by lagging the power prediction series estimated from the ECMWF unperturbed forecasts. The maximum look-ahead time for these ensembles is set to 3-day ahead, which is a relevant with the current needs for the management or trading of wind generation. Figure 5.3 gives an illustrative example of such temporal ensembles.

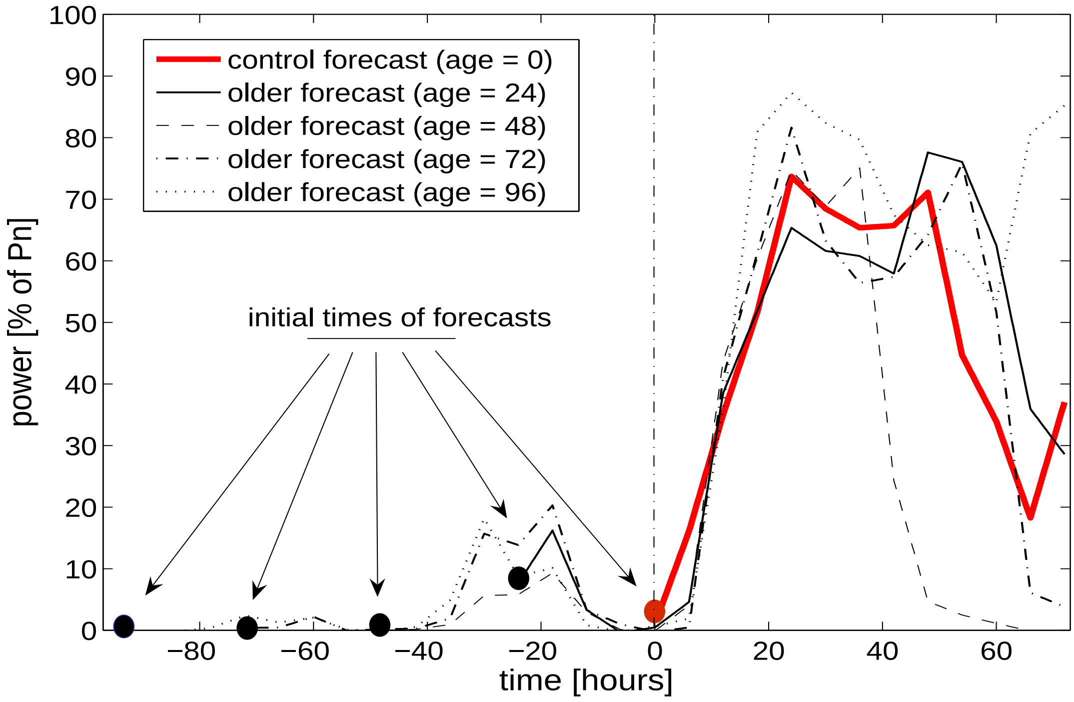


Figure 5.3: *Example of temporal ensemble predictions of wind generation. Since ECMWF-based power predictions are issued every 24 hours with a forecast length of 7-day, there is always a 72-hour period with an overlap of 5 forecasts of various ages, i.e. issued at different time origins.*

Since power predictions estimated from the ECMWF control forecasts have a maximum horizon of 7-day ahead and are issued every 24 hours, there are always 5 forecast series overlapping over the following 72 hours. Therefore, the 5 members composing the temporal ensembles include the last available forecast, plus the forecasts issued 24, 48, 72 and 96 hours before. For the specific case of temporal ensembles, we will use the term of control forecast for the most recent power prediction.

5.3 Ensembles vs. spot forecasts

In this Section, we investigate on the additional information that can be provided by wind power ensembles, information that is not given by the sole point predictions we have considered so far. Indeed, we expect ensemble predictions to contribute first to the accuracy of the point predictions that can be consequently derived (thus decreasing the level of uncertainty). Second, as it is their primary purpose, these wind power ensemble predictions should give an information on weather predictability, and therefore on the expected level of forecast uncertainty. By focusing on the Tunø Knob test case, we aim at demonstrating hereafter that ensembles indeed contain these additional information.

5.3.1 The possibility to derive more accurate point predictions

Focus on ECMWF-EPW and NCEP-EPW

The easiest and most popular way of utilizing ensemble forecasts is to derive a single point prediction by calculating the mean of all the ensemble members for each look-ahead time. It is often said that the mean of an ensemble of forecasts has a smaller error than that of any individual forecast composing this ensemble [93, 136, 168]. Therefore, it is expected that the statistical mean of ensemble predictions will prove to be the most accurate single forecast one may extract from ensembles, if evaluated over a certain period of time. For instance, Möhrlen in [158] developed a probabilistic multi-trend filter that permits to derive a 'statistical best guess' forecast from a set of ensemble members. She noticed that, even if in extreme events the skill of the best guess was higher than the one of the mean, the ensemble mean always had the lowest level of prediction error in the long run. Alternatively, Grimit and Mass [85] focused on 10m a.g.l. wind direction forecasting over the Pacific Northwest, with a mesoscale short-range ensemble forecasting system, composed by five members. They showed (over a six months evaluation period) that the ensemble mean had the highest prediction accuracy on average. The authors also noticed that the mean verified as the best forecast with the same frequency than the ensemble members, though it was never deemed as the worst forecast. Note that here, the mean is calculated from the ensemble members of wind generation, hence after the transformation of all meteorological ensemble members to power. An alternative would be to compute the mean of the meteorological ensembles for every look-ahead time and then pass it through the wind-to-power conversion model. The former possibility is to be preferred, first because of a higher level of prediction accuracy [159], and also since we are interested in maintaining all the power members for skill forecasting.

The improvements in terms of statistical accuracy of the ensemble mean with respect to the unperturbed forecast are verified here. We have already discussed the fact that point predictions of wind generation are estimates of the expectation of future generation. Besides that, calculating the mean of the ensemble members relates to estimating the expec-

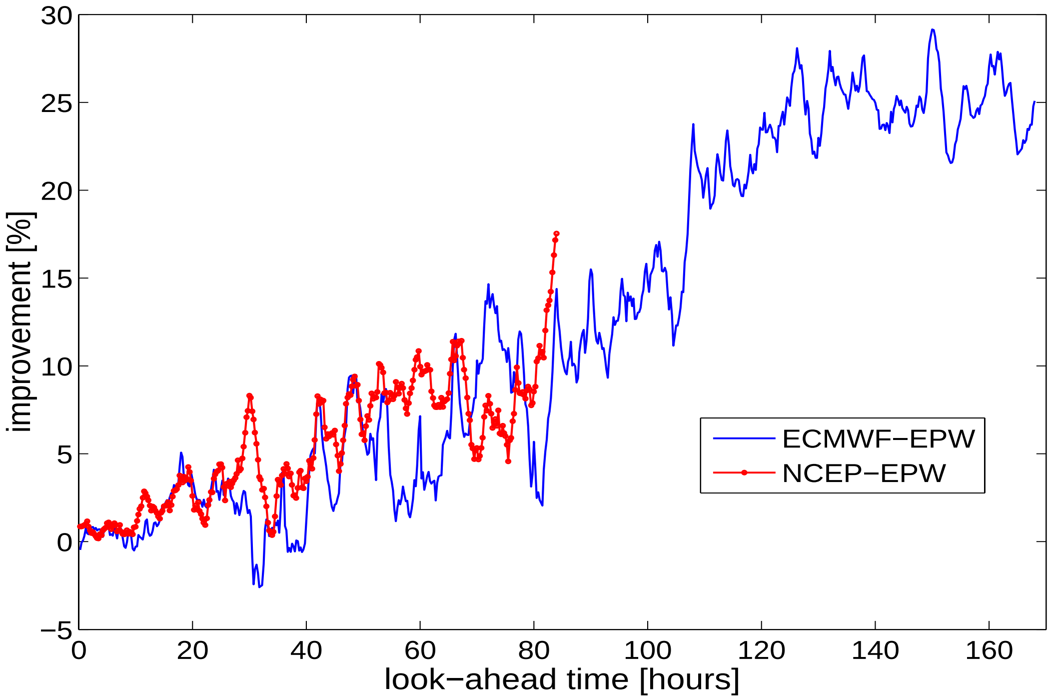


Figure 5.4: *Improvement of the ensemble mean with respect to the control forecast. Results are for wind power ensembles obtained from both ECMWF and NCEP meteorological ensembles. The error measure is the RMSE. The improvement is displayed as a function of the look-ahead time.*

tation of the probability density function sampled by these members. Therefore, we concentrate on using the RMSE criterion (Equation (3.6)) for evaluating these rival single predictions, since they both aim at minimizing a quadratic loss function. The improvement is calculated following Equation (3.9), in which the control forecast is seen as the reference method and the ensemble mean as the advanced approach. Figure 5.4 depicts the evolution of this improvement as a function of the prediction horizon, for both ECMWF-EPW and NCEP-EPW. In complement, Table 5.1 summarizes the average performance of these rival power point forecasting approaches for each day ahead.

In Figure 5.4, positive improvement values for almost all the look-ahead times translate to ensemble mean predictions exhibiting lower RMSE values than control forecasts, whatever the considered meteorological ensemble prediction system. Also, the improvement steadily increases as the lead time gets further. It reaches 17% for the case of NCEP-EPW (for 3.5-day ahead predictions) and is up to 29% for ECMWF-EPW (for 7-day ahead predictions). Table 5.1 tells that the level of prediction error is rather high for predictions up to 48-hour ahead (in comparison to the performance evaluation results for Tunø Knob, given in Appendix D). It is partly due to the low resolution of the considered meteorological predictions. For further look-ahead times, the average NRMSE of the power predictions derived from the ECMWF control forecasts even goes up to 40% of P_n. It stays below 30% of the nominal capacity for the ensemble mean, which corresponds to a 25% improvement. The improvements proposed by the NCEP-EPW mean are slightly higher than those for the ECMWF-EPW case.

Table 5.1: *Performance of unperturbed forecasts and ensemble mean forecasts for both ECMWF-EPW and NCEP-EPW. The performance is evaluated with the NRMSE criterion, expressed in percentage of P_n. NRMSE values are averaged for each day ahead.*

Day	1	2	3	4	5	6	7
ECMWF control forecast	18.25	20.52	23.53	27.31	31.90	36.33	39.94
ECMWF-EPW mean forecast	17.96	20.00	22.07	24.47	26.08	27.27	29.98
Improvement [%]	1.59	2.53	6.20	10.41	18.24	24.92	24.94
NCEP control forecast	16.73	19.78	24.27	26.00	-	-	-
NCEP-EPW mean forecast	16.44	18.85	22.23	23.55	-	-	-
Improvement [%]	1.72	4.67	8.41	9.42	-	-	-

Single-valued predictions derived from temporal ensembles

It has been mentioned that a baseline approach would consist in lagging the available control forecasts for making up ensembles. From these temporal ensembles, single point predictions could be derived as the statistical mean of the various members [95]. When considering ensembles obtained by perturbating initial conditions (and/or by stochastic perturbations of the model), it is meaningful to give the same weight to the various members, and eventually more weight to the control forecast since it is expected to be the 'best' prediction. But, if using temporal forecasts, it makes more sense to derive the single prediction as a weighted mean, the weights being function of the predictions' age [25]. This method is applied for extracting a single-valued prediction from the 5 alternative members. The weights were determined by trial and error on the first month of data, with the aim of minimizing a quadratic loss function over the whole range of look-ahead times. Temporal ensemble members are combined in the same manner whatever the prediction horizon. The resulting combination strategy is then evaluated on the full dataset consisting of 10 months of data. This method is not as rigorous as the use of cross-validation with consideration of an independent testing set. Though, we do not aim here at optimizing the performance of the obtained combination. We just want to show that this combination has the potential for proposing a significant improvement (in terms of statistical accuracy) over the control forecast. The combination weights are given in Table 5.2.

Then, we evaluate the performance of both the control forecasts (i.e. the most recent ones) and of the weighted mean predictions with the RMSE criterion, over the 10-month evaluation period. The improvement of the latter with respect to the control forecasts is displayed in Figure 5.5, as a function of the prediction horizon (up to 72-hour ahead). For comparison, we plot the previously given improvements for both the ECMWF-EPW and the NCEP-EPW mean predictions over the same range of look-ahead times.

Table 5.2: *Weights in the combination for obtaining a single-valued prediction from the 5-member temporal ensembles of wind generation. These weights are a function of the aging of the ensemble members.*

Forecast age [hours]	0	24	48	72	96
Weight	0.71	0.1	0.07	0.06	0.05

The improvement proposed by the weighted mean is not positive for all the prediction horizons, though one notices that it is positive for almost all look-ahead times further than 6-hour ahead. Actually, it could be made positive over the whole range of horizons, and at a higher level, if considering more advanced combination strategies e.g. on a per-horizon basis. The forecast combination obtained by trial and error yields a lower RMSE on average, but sacrifices the combined forecast performance for the first horizons. The improvement with respect to control forecasts is at a similar level with the one for the ECMWF-EPW and NCEP-EPW mean predictions, reaching 5-7% for 3-day ahead predictions.

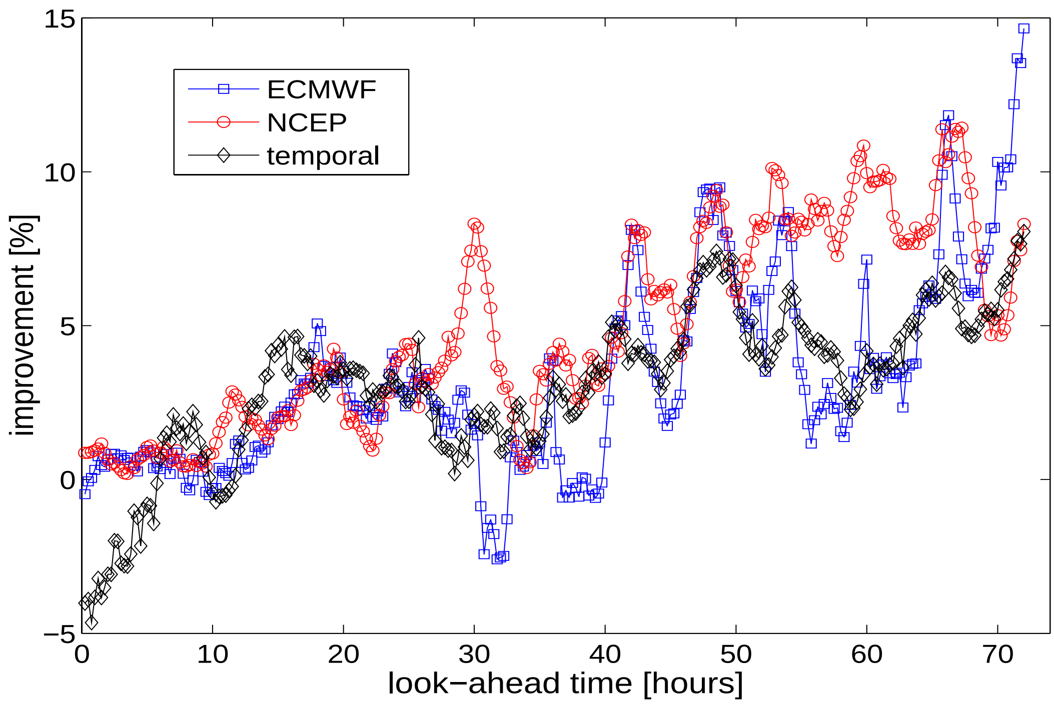


Figure 5.5: *Improvement of the weighted mean of temporal ensembles with respect to control forecasts. Results are compared with the improvements for the ECMWF-based and NCEP-based power ensembles. The error measure is the RMSE. The improvement is displayed as a function of the look-ahead time.*

So far, ECMWF and NCEP meteorological ensembles can only be produced on a rather coarse grid. Thus, the main interest of this poor man's approach stands for the possibility one has to easily apply it to higher resolution forecasts (e.g. derived from HIRLAM meteorological forecasts). For instance, such an approach consisting in deriving a single-valued prediction as the weighted mean of temporal ensemble members has been applied very recently (end of 2005) by Brundage et al. [25] for near-surface wind speed forecasting. The considered NWP model is the Rapid Update Cycle (RUC) model, with a 40-km horizontal resolution. The authors studied several combination approaches and showed the interest of this weighted combination.

From a general point of view, it appears that the three different types of ensemble predictions of wind power contain an additional information that allows one to decrease the level of uncertainty associated to control forecasts.

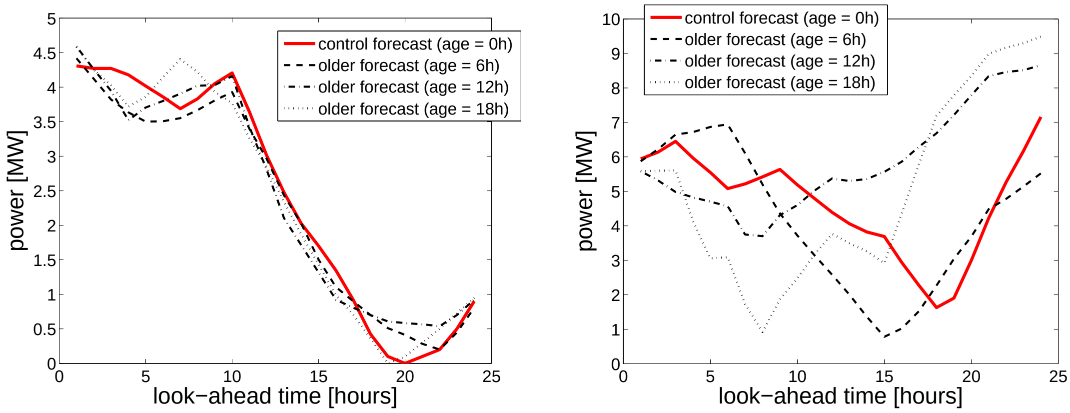
5.3.2 The ensembles' ability to reflect the forecast uncertainty

Deriving a single prediction from a set of alternative ensemble predictions is perhaps the simplest and most common method of utilizing ensemble forecasts in an operational context. Though, the major additional value expected from ensemble predictions is related to their ability to resolve between more and less easily predictable situations, thus giving valuable information on the expected level of forecast uncertainty. In the following, the way the spread of ensemble predictions may provide that information on forecast uncertainty is described. Then, the relation between spread and prediction uncertainty is demonstrated for the three types of wind power ensemble forecasts considered in the present Chapter.

Using the ensemble spread as an indicator of forecast uncertainty

The ensemble members should be sampling the distribution of expected events for every look-ahead time. The disagreement between ensemble members should then give a qualitative estimate of the uncertainty in wind power predictions [234]. This is obviously valid only if the uncertainty in both the model and the initial conditions are well represented, by the perturbations of these initial conditions, and by the use of stochastic physics in the NWP model [238]. Note that since for all types of wind power ensembles the same model is used for the conversion of meteorological ensembles to power, the uncertainty in the modeling of the wind-to-power process is not represented.

Figure 5.6 gives the illustrative example of situations that could be deemed as 'more easily predictable' and 'rather unpredictable'. Both cases consist of 24-hour ahead temporal ensemble predictions of wind power for a multi-MW wind farm in Ireland, with HIRLAM meteorological forecasts as input, issued every 6 hours. Are represented the control forecast and the temporal ensemble members, corresponding to power forecasts issued 6, 12 and 18 hours before the control forecast. In Figure 5.6 (a) all the 4 sets of predictions exhibit a rather good agreement. In such a situation it is expected that the predictability is quite high and that the level of uncertainty is low. At the contrary, the various forecasts of Figure 5.6 (b) significantly disagree over the considered period. This second type of situation would be classified as less predictable, and then the related level of uncertainty would be higher than for the first one. Note that we use the term 'uncertainty' and not 'error'. Therefore, when we



(a) more easily predictable situation **(b)** rather unpredictable situation

Figure 5.6: *Illustrative example of more and less easily predictable situations depending on the agreement of the temporal ensemble prediction members.*

mention that the level of uncertainty is higher (and respectively lower), this does not directly translate to a high (and respectively low) prediction error. Indeed, It tells that probabilities of witnessing large prediction errors are higher in this case.

Focus on ECMWF-EPW and NCEP-EPW

In this Paragraph, we aim at evaluating how the disagreement among ensemble members relates to the level of prediction uncertainty. Point predictions are given by the mean of ECMWF-EPW and NCEP-EPW. As explained in Paragraph 3.5.2, the level of forecast uncertainty is given by the standard deviation of prediction errors (calculated with the SDE criterion introduced in Equation (3.8)), since that uncertainty is related to the sharpness of prediction error distributions. The disagreement between ensemble members is quantified by the ensemble spread, in turn given by the standard deviation of the ensemble members. Were the ensembles perfectly reliable in a probabilistic sense[2] that the standard deviation of ensemble members would actually correspond to the one of the prediction errors made by the ensemble mean. Evaluating the correspondence between ensemble spread and the sharpness of error distributions is a manner to evaluate the ensemble's ability to resolve between situations with low and high uncertainty levels in the long run. Note that this could also be seen as a kind of spread-skill relationship since the standard deviation of the errors can be considered as a performance evaluation criterion (see Paragraph 3.2.3).

Figure 5.7 depicts the results of that investigation for both NCEP-EPW and ECMWF-EPW. The relationship between spread and sharpness of related distributions of prediction errors is evaluated on a daily basis. This means that datasets for each curve contain all the predictions and measures for the related day, and used to draw the trend between the en-

[2] cf. the definition of reliability given in Paragraph 4.7.1. For the specific case of ensemble predictions, reliability is the property of statistical consistency between predicted probabilities and observed frequencies of the event under consideration.

semble spread and the standard deviation of the errors. In order to obtain the curves, the datasets are split in ten equally populated classes of ensemble spread. They all contain 2880 spread-error pairs. Average ensemble spread and standard deviation of the prediction errors are calculated for each class. In addition to showing the trend of the spread-sharpness relationship (if any), these curves inform on the distribution of ensemble spread values.

Focusing first on NCEP-EPW, one sees from Figure 5.7 (a) that there is clear trend according to which the standard deviation of prediction errors increases as the ensemble spread is higher. But, this standard deviation is larger than that of the ensemble members. In a probabilistic sense, this means that wind power ensembles are not well-calibrated: they underestimate the prediction uncertainty. Like for the case of reliability diagrams, the diagonal dashed line corresponds to the ideal case for which the ensemble spread perfectly reflects the standard deviation of prediction errors. Actually, wind power ensembles could be recalibrated for having an acceptable reliability [172, 173]. This has not been done here since we aim at studying the inherent value of ensemble predictions only. The underestimation of the actual uncertainty is also present for the case of ECMWF-EPW, but not over the whole range of spread values. For predictions further than 2/3-day ahead (and for spread values higher than approximately 20% of P_n), it is actually the inverse: they tend to underestimate the forecast uncertainty. It has already been noticed that the spread of ECMWF meteorological ensembles grows faster than that of NCEP meteorological ensembles [29]. The method developed by Nielsen et al. [172] for going from meteorological to wind power ensembles conserves that property, since the wind-to-power conversion model is estimated with the unperturbed forecast only, and applied consequently to the ensemble members. The ensemble spread is amplified (in the medium power range) or dampened (in the low or high power range) depending on predicted wind conditions, owing to the nonlinear and bounded shape of the estimated power curve. Though, again, the method does not carry out any statistical correction on power ensemble members.

In Figure 5.7 (b), which is related to ECMWF-EPW, the trend is almost linear for all the 7 curves, but not along the diagonal. The curves for the case of NCEP-EPW do not exhibit such a behavior: their trend is composed by a sharp increase for low spread values, and then by a much weaker augmentation for larger spread values. Moreover, it appears that spread values are shared out more homogeneously (and also reach larger levels) for the former than for the latter. For instance for day 2, the maximum average spread is equal to 25% of P_n for NCEP-EPW while this value is of 35% of P_n for ECMWF-EPW. And, while 50% of the spread values for this day are below 6% of P_n for NCEP-EPW, this same proportion of spread values are up to 14% of P_n for ECMWF-EPW. Hence, spread values (and their variability) for the two ensemble prediction systems cannot be interpreted in a similar way. The growth of the ECMWF-EPW ensemble spread is more pronounced certainly because it integrates both perturbations of initial conditions and stochastic model perturbations. Perturbations in the initial conditions mainly have an impact on short-term look-ahead times, while the influence of the stochastic model perturbations is more on the medium range (further than 3-day ahead) [234, 29]. From these two plots, one concludes on a higher statistical reliability of ECMWF-EPW. Note that the size of the ensembles is known to have an impact on their

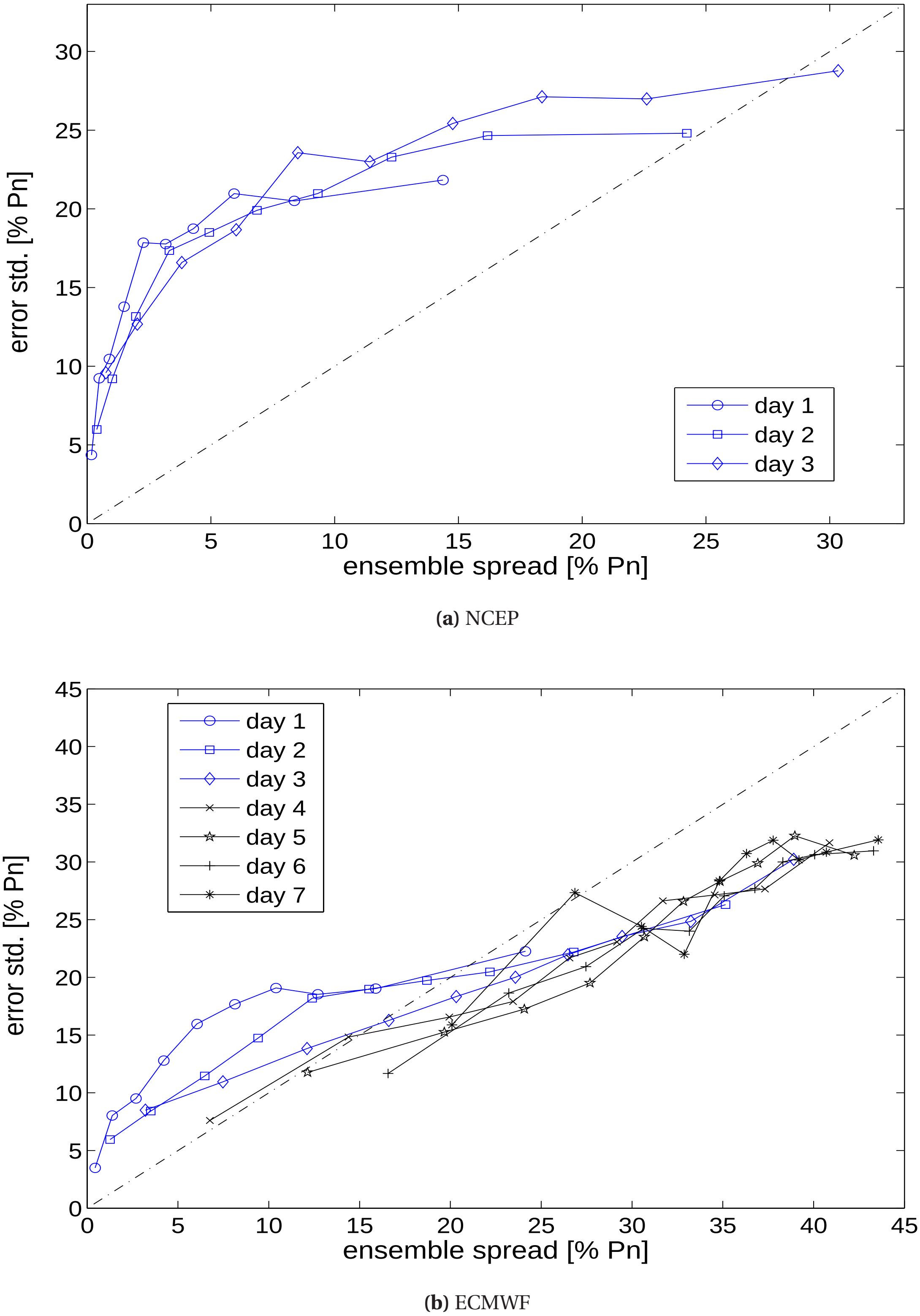


Figure 5.7: *Relation between the ensemble spread and the error standard deviation of the ensemble mean, for NCEP-EPW and ECMWF-EPW. That relation is evaluated on a daily basis. The dataset for each day is split in 10 equally-populated classes.*

skill [5, 168]. ECMWF-based ensembles of wind power includes 5 times more members that the NCEP-based ensembles.

And what if considering temporal ensembles?

In Paragraph 5.3.1, it was shown that the poor man's alternative was valuable for deriving single-valued predictions that propose a significant improvement with respect to the control forecasts of wind power. Then, it is of particular interest to check if the trend between the ensemble spread and the standard deviation of prediction errors depicted for ECMWF-EPW and NCEP-EPW can also be seen for the case of temporal ensembles. There are actually few studies focusing on the ability of temporal ensembles to dissociate between various levels of forecast uncertainty. And, the results of these investigations yield heterogeneous conclusions on that particular point. For instance, the temporal ensemble spread of 500-hPa geopotential height forecast has proven to be related to forecast skill [95], but this relation was characterized as being weak. Also, Hamill [90] concluded that there was almost no information on expected uncertainty from the amount of discrepancy between short-term lagged forecast, when studying daily 850-hPa temperature forecasts over a period of 23 years. In contrast, Roebber [193] noticed that the variability of successive forecasts of marine cyclogenesis in the Western Atlantic, using a 4-member ensemble of lagged sea-level pressure forecasts, could provide a valuable estimation of forecast skill. And more recently, for the specific case of wind forecasting at a single site (with the RUC forecasting model), Brundage et al. [26] presented results that support the idea of using the spread of temporal ensembles for skill forecasting. Despite these contradictory remarks, we investigate here on the possible value of temporal ensembles for reflecting the uncertainty in wind power predictions.

For that purpose, we carry out a study similar to that presented above for ECMWF-EPW and NCEP-EPW, with a dataset composed by the spread of the 5-member temporal ensembles and the prediction errors achieved by the ensemble mean. This dataset is split into three parts for horizons belonging to the first day-ahead up to the third day-ahead forecasts. As we did in Paragraph 5.3.1 for deriving single-valued predictions from the 5-member temporal ensembles, we expect that all the alternative predictions cannot have the same weight when evaluating the discrepancy between the members. Fast-changing weather situations may lead to a large spread of temporal ensembles, but some of these situations are not compulsorily associated to a lower predictability. Therefore, the ensemble spread is quantified by the weighted standard deviation of the ensemble members. As a first approach, the weights we consider are the same than the ones used in Paragraph 5.3.1 for computing the weighted mean (cf. Table 5.2). The standard deviation of prediction errors is calculated for 10 equally-populated classes. It is then plotted against the average spread for every class in Figure 5.8 (solid curves).

The relation between the ensemble spread and the standard deviation of prediction errors is almost linear for the three days. The curves do not lie along the diagonal, but they are closer to this ideal case than the related curves for the cases of ECMWF-EPW and NCEP-EPW. The standard deviation of the errors goes from 7% to 30% of P_n for day 1, from 8% to

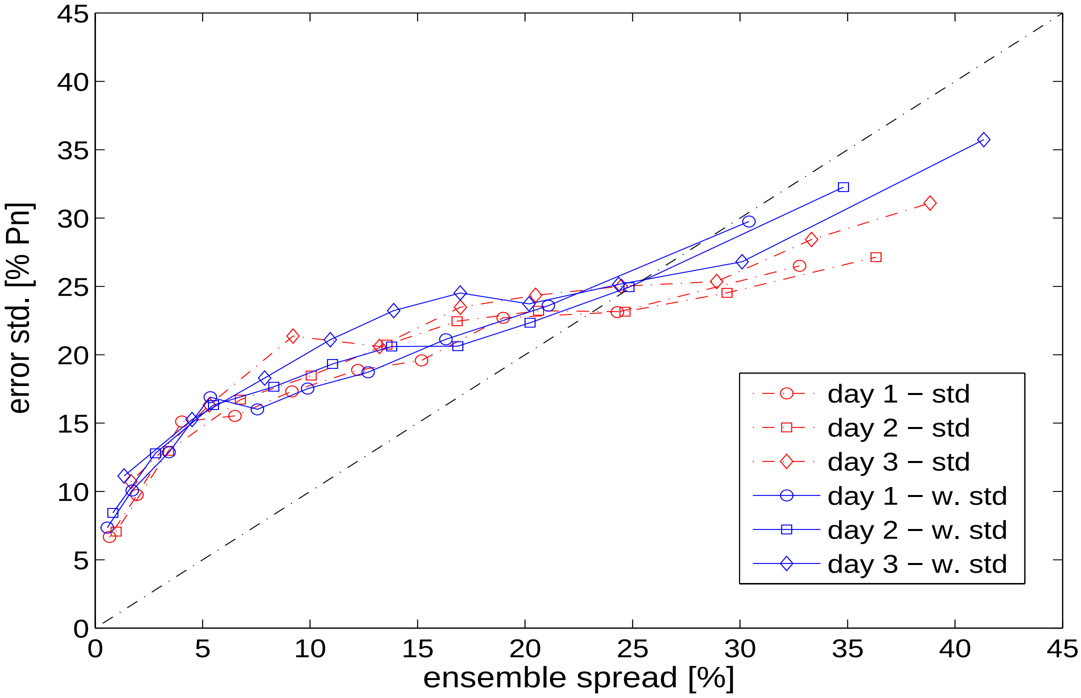


Figure 5.8: *Relation between the ensemble spread and the error standard deviation of the ensemble mean for the temporal ensemble predictions of wind generation. That relation is evaluated on a daily basis. The dataset for each day is split in 10 equally-populated classes. Comparison is made between the normal (*std*) and weighted (*w. std*) standard deviation of ensemble members for quantifying the ensemble spread.*

32% of P_n for day 2, and from 11% to 36% of P_n for day 3, when going from the first to the last class of spread values. Therefore, this clear trend between the spread of the temporal ensembles and the sharpness of prediction error distributions comforts us in using poor man's ensembles as an alternative to ECMWF-EPW and NCEP-EPW for informing on the expected level of prediction uncertainty. Note that spread values are homogeneously scattered. In addition, they reach high levels (up to 42% for day 3), significantly larger than for both ECMWF-EPW and NCEP-EPW (especially for day 1). Though, these large spread values do not seem to substantially overestimate the actual uncertainty, as it was the case for ECMWF-EPW. For comparison, Figure 5.8 also depicts the relationship between ensemble spread and sharpness of prediction error distributions when the spread is quantified by calculating the normal (i.e. equally weighted) standard deviation of ensemble members (dashed curves). Using the weighted standard deviation for quantifying the disagreement between members is beneficial since the resulting curves are closer to the diagonal, mostly for large spread values. The weighting alternative better reflects the prediction uncertainty by diminishing the influence of older forecasts. Though, like for the forecast combination procedure we proposed in Paragraph 5.3.1, the weights could be optimized so that the resulting quantification of the temporal ensemble spread better reflects the prediction uncertainty.

For the three types of wind power ensemble predictions we have shown that the ensemble spread gives valuable information on the level of forecast uncertainty (quantified by the standard deviation of prediction error distributions, where predictions are given by the ensemble mean). However, the spread values distributions significantly differ for the various types of ensembles, as well as the characteristics of the relation between ensemble spread and prediction uncertainty. We will explain in the following Section how to account for these effects when deriving a skill forecasting methodology.

5.4 Skill forecasts based on wind power ensembles

In this Section is developed a methodology for skill forecasting, appropriate for the wind power forecasting application. First, we propose a definition for prediction risk indices, based on the dispersion of wind power ensembles over a single prediction horizon, or over a set of successive look-ahead times. Then, it is explained how such prediction risk indices may be used as skill forecasts, i.e. forecasts of the distributions of expected prediction errors. The relation between prediction risk indices and the level of prediction error is described with conditional probability diagrams. We study the ability of prediction risk indices to differentiate between situations with low and high uncertainty depending on the use of ECMWF-EPW, NCEP-EPW, or temporal ensembles as input. Also, comparison is made between prediction risk indices defined on a per look-ahead time basis or over periods of 24 hours.

5.4.1 Skill forecasting in the wind power prediction literature

For assessing weather predictability and consequently forecast the expected level of wind power prediction uncertainty, the investigations described in the literature follow two different paths, which may prove to be complementary. The first idea consists in finding indicators that characterize the actual weather dynamics and to link these typical weather patterns to several levels of expected error. The second and more popular alternative is based on the consideration of ensemble predictions of wind generation, as it is done in the present Chapter.

The developments towards the definition of weather dynamics indicators are the research works by Lange [129, 131]. He utilizes methods from synoptic climatology to classify the local weather conditions based on measurements of wind speed and direction, as well as pressure. This classification, through principal component analysis and cluster analysis methods, allows one to reveal characteristic meteorological patterns that can be associated to various levels of forecast uncertainty. Indeed, when low pressure systems are dominant the average level of wind speed prediction error is higher and inversely when high pressure systems govern, that average error proves to be much lower. However, no link with wind power prediction errors is shown, and this study is based on meteorological measurements, which are often not available online. Therefore, the derived method does not appear directly applicable for detecting in a preventive way, and in an online environment, situations for which high level of forecasting uncertainty is expected. That is the reason why the ideas developed in the present Chapter exploit the information included in the NWPs (and not in the measurements) in order to develop tools for online estimation of the expected level of uncertainty in power predictions.

In parallel, ensemble predictions have gained increased interest for energy-related applications in the recent years. Several projects focusing on ensemble forecasting for wind power are ongoing [77, 104]. Their aim is to exploit the information contained in multi-scenario forecasts to derive an uncertainty estimate and/or a single 'best' forecast. Related developments have mainly been directed towards the estimation of predictive distributions of wind generation [173]. The underlying idea is that the resulting predictive distributions of wind generation would have a better resolution than the ones obtained from statistical methods, such as the method developed in Chapter 4 or methods based on quantile [171] and local quantile [22] regression. Such an approach for producing probabilistic predictions based on meteorological ensembles has already given promising results for the load forecasting problem [215, 216].

In contrast to these developments directed towards probabilistic prediction of wind generation, our aim is to propose to consider wind power ensemble predictions for skill forecasting. Indeed, this consists in deriving a signal (in the form of a numerical value) from these ensemble predictions, reflecting the current predictability of weather dynamics, which will inform forecast users on the confidence they may have in the point prediction they will use for making a decision.

5.4.2 Definition of prediction risk indices

Owing to the relation between the spread of ensemble members and the standard deviation of the errors described in the previous Section, we propose hereafter to define prediction risk indices as a measure of that ensemble spread. This measure is a continuous one, in contrast to some categorical measures[3] introduced in the meteorological literature, such as the mode population [224] or the ensemble statistical entropy [244]. Our choice is motivated by the conclusions from Grimit [84], stating that continuous measures of ensemble spread are more appropriate if forecast's users have a continuous utility function[4]. We assume that this is case for users of wind power predictions, either for the management or trading of wind generation.

If the ensemble predictions of wind power are issued at time t, there are then J alternative predictions $\hat{p}^{(j)}_{t+k/t}$ $(j = 1, \ldots, J)$ for any lead time $t + k$. Computing the weighted standard deviation $\tilde{\sigma}_{t,k}$ of the set of alternative predictions is a way to estimate the ensem-

[3]The basic idea of categorical measures of ensemble spread consists in dividing the range of possible forecast values in several bins, and to count the numbers of ensemble members falling in each bin.

[4]The utility function for a forecast's user is introduced and further discussed in Chapter 6.

ble spread for that look-ahead time. $\tilde{\sigma}_{t,k}$ is calculated as following:

$$\tilde{\sigma}_{t,k} = \left[\frac{J}{J-1} \sum_{j=1}^{J} w_j \left(\hat{p}^{(j)}_{t+k/t} - \bar{p}^{J}_{t+k/t} \right)^2 \right]^{\frac{1}{2}}, \tag{5.1}$$

such that the sum of the weights w_j totals 1, and with $\bar{p}^{J}_{t+k/t}$ the mean of the J alternative predictions for that lead time, that is

$$\bar{p}^{J}_{t+k/t} = \frac{1}{J} \sum_{j=1}^{J} \hat{p}^{(j)}_{t+k/t}. \tag{5.2}$$

In Equation (5.1), the weights serve to reflect the ability of the various alternative predictions to give an assessment of the predictability. If considering for instance an algorithm that derives a best-guess forecast as a weighted average of the ensemble members, these weights can be directly used in the calculation of $\tilde{\sigma}_{t,k}$. A similar remark is valid if considering time-lagged ensembles, for which the weights in the optimal combination of the alternative predictions are a function of their age. This is what was done in Paragraph 5.3.2.

The temporal resolution of prediction ensembles and measures are not the same. Even if we have interpolated predictions for having a correspondence between power predictions and measures, the actual temporal resolution of ECMWF and NCEP meteorological ensembles is of six hours. In addition, the weather predictability does not have an instantaneous nature: it is very unlikely that wind generation would be easily predictable for a given look-ahead time, and then highly unpredictable for the following one. This is the reason why it is envisaged here to estimate predictability over a certain range of time. Therefore, we generalize the use of the weighted standard deviation by computing the average of $\tilde{\sigma}_{t,k}$ over a set of consecutive horizons, from look-ahead time k_1 to k_2. This average weighted standard deviation defines a *Normalized Prediction Risk Index*, abbreviated NPRI, which is calculated as

$$\mathrm{NPRI}(k_1, k_2) := \frac{1}{k_2 - k_1 + 1} \sum_{i=k_1}^{k_2} \tilde{\sigma}_{t,i}, \tag{5.3}$$

with $\tilde{\sigma}_{t,i}$ given by Equation (5.1). In the following, we denote by NPRI_h the prediction risk index if calculated on a per-horizon basis (i.e. such that $k_1 = k_2$) and by NPRI_d if defined over a period of 24 hours.

5.4.3 On the relation between NPRI and energy imbalance

Considering prediction errors as energy imbalances

Prediction errors are expressed in the form of energy imbalances. This is because we aim at showing that the NPRI introduced in Equation (5.3) can be used for informing on the level of expected prediction error over a certain period of time. Energy imbalances are defined here as the difference, in absolute value, between the predicted and measured amounts of energy over a period of interest. The amount of energy $\hat{E}_{t+k1}^{t+k_2}$ predicted at time t for the period between lead times $t+k_1$ and $t+k_2$ can be calculated as

$$\hat{E}_{t+k_1}^{t+k_2} = t_r \sum_{i=k_1}^{k_2} \hat{p}_{t+i/t}, \tag{5.4}$$

where t_r is the temporal resolution of the wind power predictions. In parallel, the measured quantity of energy $E_{t+k_1}^{t+k_2}$ over that same period is

$$E_{t+k_1}^{t+k_2} = t_r \sum_{i=k_1}^{k_2} p_{t+i}. \tag{5.5}$$

Note that both measured and predicted amounts of energy are normalized quantities, since power forecasts and measures are normalized values (normalized by the wind farm nominal power P_n). Then, the energy imbalance $d_{t+k_1}^{t+k_2}$ over the period going from lead time $t+k_1$ to lead time $t+k_2$ is given by

$$d_{t+k_1}^{t+k_2} = |E_{t+k_1}^{t+k_2} - \hat{E}_{t+k_1}^{t+k_2}| = t_r \sum_{i=k_1}^{k_2} |p_{t+i} - \hat{p}_{t+i/t}|. \tag{5.6}$$

In the above Equation, by calculating energy imbalances in absolute value, we similarly integrates production surplus and shortage. Prediction risk indices are meant for estimating the expected level of uncertainty, but cannot give the sign of forecast errors. If considering a single look-ahead time, the normalized imbalance equals the prediction error in absolute value. And, for successive horizons, it is equivalent to the average absolute error over that time interval.

Prediction risk indices should give information on the expected level of forecast uncertainty whatever the considered point prediction. Therefore, we do not concentrate hereafter on the use of the best available point forecast of wind generation that can be derived from ensembles, i.e. given by the ensemble mean, or alternatively by the weighted mean for temporal ensembles (cf. Paragraph 5.3.1). Instead, the considered point predictions are produced as it is commonly done today, that is by applying a wind-to-power conversion model to the control forecasts provided by meteorological offices. The point predictions thus correspond to those used as benchmark in Paragraph 5.3.1, i.e. derived from the ECMWF or NCEP control forecasts.

Conditional probability diagrams for relating NPRI with the level of expected prediction error

Following the works by Grimit [84], the relationship between prediction risk indices and the level of prediction error of the considered point prediction method is drawn from a probabilistic perspective. This proposal goes against the traditional approach consisting in fitting a linear regressor between the measures quantifying the ensemble spread and the point predictor's skill, associated with a correlation coefficient that assesses the strength of this relation (see [9, 86, 234] among others). The inconsistency of using the correlation coefficient for that purpose has been discussed by Grimit and Mass [86]: considering it for measuring the strength of the relationship between the ensemble spread and the predictor's skill implicitly assumes a linear relation between the spread estimator and the evaluation criterion for the prediction error[5]. Note that the linear relation we were evaluating in Paragraph 5.3.2 must not be seen as an attempt to draw a deterministic relation between these two variables: it was showed how the sharpness of error distributions is related to the ensemble spread.

A possibility for expressing the relation between NPRI and the related prediction error in a probabilistic manner would be to use contingency tables, which give the probabilities of events defined by the occurrence of NPRI-range/error-range pairs. Such an idea has been proposed first by Houtekamer [100] and consequently applied by Whitaker and Loughe [234]. Though, our choice goes for conditional probability diagrams similar to those used by Moore and Kleeman [163], which easily give a visual information on the relation between prediction risk indices and the level of forecast uncertainty.

Conditional probability diagrams summarize the distribution of energy imbalances given the NPRI value. The range of NPRI values are divided in various categories, defined as equally populated classes. This follows from the idea that it is not the value of NPRI by itself that tells if the situation is more or less uncertain, but more where this value is located in the climatological distribution of the NPRI values [234, 244]. Also, considering equally populated classes of NPRI values will permit to compare skill forecasts made from ECMWF-EPW, NCEP-EPW or the temporal ensembles as input, independently of the range of their ensemble spread values. We apply a similar reasoning to energy imbalances, which are normalized by their climatological value depending on the look-ahead period. This climatological value corresponds to the average imbalance over the 10-month evaluation period for a given look-ahead period. When mentioning imbalance levels, they will indeed be relative and expressed in percentage of their climatological value. Thus, we will study how NPRI has the ability to tell if these imbalances are lower or higher than usual, independently of the global performance of the considered point prediction method.

5.4.4 Pointwise estimation of expected uncertainty

In a first stage, the ability of NPRI to inform on the level of prediction uncertainty when calculated for each look-ahead time is evaluated, for the three sets of wind power ensemble predictions. As explained in Paragraph 5.4.2, NPRI_h corresponds to the weighted standard deviation of the ensemble members for a given look-ahead time. The weights for its calculation are set to $1/J$ for ECMWF-EPW and NCEP-EPW (where J the number of ensemble

[5]Actually, it has been shown for an ideal ensemble of infinite size that the spread-error correlation can be written analytically, as a function of the temporal variability of the ensemble spread [100]. In this model, the prediction error is in absolute value. For an infinite spread variability, this spread-error correlation asymptots to 0.8 [234].

members equals 11 and 51 respectively). Alternatively, the weights given in Table 5.2 are used for the case of the temporal ensembles. Owing to the limited amount of data available (only 300 series of wind power predictions, over a 10-month evaluation period), and also for comparison with results we will present in the following Paragraph, we gather NPRI and energy imbalance values for each day ahead.

The NPRI$_h$ ability to inform on the expected imbalance level

The way conditional probability diagrams are used for relating NPRI and the imbalance level has been described in the above Paragraph. Figure 5.9 is an example of such a diagram for ECMWF-EPW for day 3 (i.e. for look-ahead times between 48 and 72-hour ahead). It depicts particular points of conditional probability distributions of energy imbalances depending on the class of NPRI$_h$ values. For each NPRI$_h$ class are given the median $r^{(0.5)}$ (crosses), the mean μ (circles), the lower $r^{(0.25)}$ and upper $r^{(0.75)}$ quartiles (squares), as well as the 10% $r^{(0.1)}$ and 90% $r^{(0.9)}$ quantiles (downward and upward triangles), of related imbalance distributions. All these symbols are centered on the average NPRI$_h$ value for each class. Five different NPRI classes are considered, for which the related empirical distributions of imbalances are composed by 1440 items each.

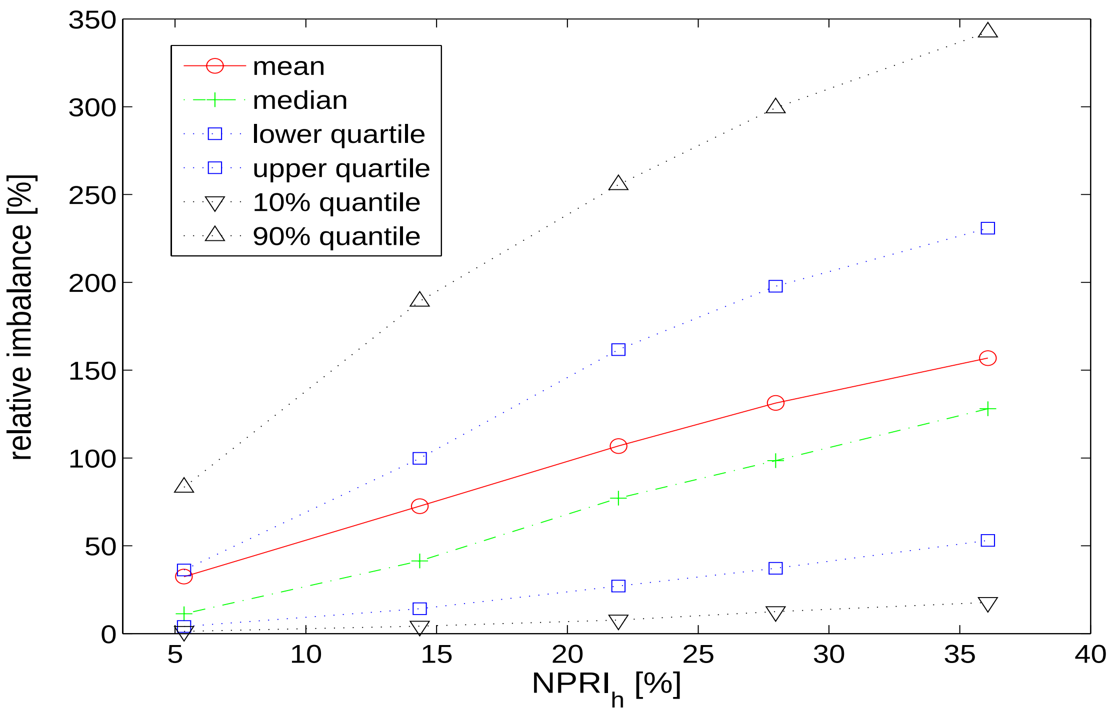


Figure 5.9: *Conditional probability diagram giving the relation between NPRI$_h$ and the level of energy imbalance. NPRI$_h$ are calculated from ECMWF-EPW. Results are for day 3 (prediction horizons from 48 to 72-hour ahead). Empirical distributions are made up with 1440 elements.*

The mean values give the general trend between NPRI_h and the level of prediction error. There is a steady (and quasi linear) increase in the mean imbalance level when going from the lowest to the highest NPRI_h class. When NPRI_h values belong to the NPRI_h class 1, the average imbalance level equals 30% of the climatological imbalance level. Though, for class 5, this average imbalance is more than 5 times larger, reaching 155% of the climatological value. Therefore, using NPRI_h with ECMWF-EPW appears to be a possibility for resolving between situations with various expected imbalance levels.

The most interesting information comes from the quoted quantiles of the conditional probability distributions given the NPRI_h class, since it informs on the range of expected imbalances. More precisely, they give both a lower and an upper bound for these expected imbalances. In Figure 5.9, one sees for instance that if NPRI_h lies in the first class, then 90% of imbalances are below 90% of the climatological imbalance level (for the considered look-ahead time), while there is still a 10% probability that the level of imbalance exceeds 340% of the climatological imbalance value if the NPRI_h value is in the fifth class. The imbalance distributions become much wider when NPRI_h values are larger: the 10% quantiles are still close to zero, but the 90% ones get much higher. It is indeed that upper bound on the expected imbalance level that tells what may be the risk of relying on the provided wind power point prediction. From a risk aversion point of view, it would be preferable to make conservative decisions if NPRI_h values lie in class 5. Note that here, we compare imbalances to their climatological level for better illustrating the fact that prediction uncertainty is lower or higher than usual. Though, in an operational context, expected levels of imbalance could be expressed with physical units (e.g. in MWh or in percentage of the maximum possible generation over the time range).

We have chosen the example of day 3 for describing the relationship between the introduced NPRI applied to ECMWF-EPW and the level of imbalance in a probabilistic manner. The same kind of relation can be witnessed for the other days (from day 1 to day 7 since we concentrate on ECMWF-EPW), or for the other types of ensembles. Though, that relationship may exhibit slightly different characteristics, which are studied in the following Paragraph.

Comparing the cases for which NPRI_h is calculated with ECMWF-EPW, NCEP-EPW and the temporal ensembles

In the above Paragraph, the relation between NPRI_h (for ECMWF-EPW) and imbalance on a per-look-ahead time basis has been described. Here, a comparison is made between the information given by the three types of ensemble predictions of wind generation we consider in the present Chapter. This comparison would be possible for the first 3 days ahead: they are the days for which all the various ensembles are available. The specific case of day 1 is left aside since ECMWF-EPW are only available at the end of this first day in an operational context. Hereafter, we concentrate on day 2 (look-ahead times between 24 and 48-hour ahead) for highlighting the differences between the various wind power ensembles. Similar analyses were carried out for the other days, and the following comments are representative for the whole study.

Even if the NPRI_h values are not scattered in the same manner if considering NCEP-EPW, ECMWF-EPW, or the temporal ensembles, using classes of NPRI values enables to study the inherent ability of the various ensemble approaches to resolve between situations with low and high uncertainty. These categories of NPRI_h values allow us to leave aside the problem of their distributions and to see how their variations may have an indicative value for estimating the relative level of imbalance. Therefore, when comparing the various approaches, we do not mention the ranges of NPRI_h values, but only the NPRI_h class, numbered from 1 to 5. Table 5.3 gathers some of the quantiles and the mean of the conditional probability distributions of imbalances, given the NPRI_h class, for the three types of ensemble predictions.

Table 5.3: *Characteristics of the conditional imbalance distributions given the NPRI_h class. Both NPRI_h and imbalance values are gathered over day 2 (look-ahead times between 24 and 48-hour ahead). Results are for ECMWF-EPW, NCEP-EPW, and the temporal ensembles respectively. Empirical distributions are made up with 1440 elements. Imbalance values correspond to relative imbalances (in % of their climatological level).*

(a) - ECMWF-EPW

NPRI_h class	$r^{(0.1)}$	$r^{(0.25)}$	$r^{(0.5)}$	$r^{(0.75)}$	$r^{(0.9)}$	μ
1	1.3	2.9	10.2	28.7	68.9	27.7
2	4.2	15.2	42.7	89.6	176.7	70.5
3	10.8	30.1	78.4	157.7	277.7	113.7
4	15.2	44.3	113.5	202.7	298.3	137.4
5	16.0	47.5	116.4	223.5	333.6	150.7

(b) - NCEP-EPW

NPRI_h class	$r^{(0.1)}$	$r^{(0.25)}$	$r^{(0.5)}$	$r^{(0.75)}$	$r^{(0.9)}$	μ
1	1.4	4.7	10.8	29.3	84.8	30.4
2	10.5	22.4	40.3	91.3	192.6	75.9
3	17.2	39.8	76.2	146.1	243.4	108.7
4	24.9	56.5	103.3	176.5	275.7	130.93
5	26.8	69.4	135.2	218.7	305.1	154.0

(c) - Temporal ensembles

NPRI_h class	$r^{(0.1)}$	$r^{(0.25)}$	$r^{(0.5)}$	$r^{(0.75)}$	$r^{(0.9)}$	μ
1	1.4	3.7	13.0	40.2	99.8	40.0
2	5.1	20.0	57.3	117.0	223.5	89.1
3	9.2	28.9	88.4	171.6	264.4	114.9
4	11.2	35.3	101.8	192.6	290.2	127.6
5	8.7	27.1	76.7	187.8	323.5	128.3

The variability of the mean imbalance can be seen as a criterion for evaluating the ability of the different approaches for dissociating between several levels of forecast uncertainty. We quantify that variability by the ratio between the mean imbalances for NPRI$_h$ class 5 and 1. This ratio is equal to 5.4, 5.1 and 3.2 for ECMWF-EPW, NCEP-EPW and the temporal ensembles respectively. It appears thus that ECMWF-EPW and NCEP-EPW have a higher differentiation ability (with a slight advantage for ECMWF-EPW), far better than the one of the temporal ensembles. Even if we showed in Paragraph 5.3.2 that the spread of temporal ensemble was a good indicator of the skill of the temporal ensemble mean, it seems that it is not as good for estimating the uncertainty associated to the control forecasts, especially for large NPRI$_h$ values.

Then, we focus on the quantiles of conditional imbalance distributions. The increase in the spread[6] of these distributions when going from the first to the fifth class is more significant for ECMWF-EPW, followed by NCEP-EPW and the temporal ensembles. This can also be seen as another criterion for stating that ECMWF-based ensembles better resolve among situations, since the evolution of the uncertainty in the expected level of imbalance is more pronounced. If looking separately at lower ($r^{(0.1)}$ and $r^{(0.25)}$) and upper ($r^{(0.75)}$ and $r^{(0.9)}$) quantiles, one sees that lower quantiles are more variable for NCEP-EPW while upper quantiles are more variable for ECMWF-EPW. The first one better resolves the low uncertainty situations when the second better differentiates the high uncertainty situations. Therefore, if having a risk aversion point of view, NPRI$_h$ used with ECMWF-EPW gives a more valuable information on the risk one may face when relying on the provided point prediction.

The increase in the mean imbalance depending on the NPRI$_h$ class is not as steady for the temporal ensembles than for the others. Also, the four lowest quantiles decrease between class 4 and 5. If the mean imbalance is higher for NPRI class 5, it is only because this NPRI class contains very large prediction errors. But, it also contains more low prediction errors than the fourth class. Note that this temporal approach, even if less skillful for skill forecasting on a per-step ahead basis, has the great advantage of being a gratis and easily applicable alternative to the use of ECMWF-based and NCEP-based ensemble predictions.

5.4.5 Estimation of the uncertainty for a look-ahead period

In a second part of the study, the possibility of providing an estimation of expected uncertainty for a look-ahead period is considered, by calculating NPRI over a set of successive look-ahead times. And, we evaluate the benefits of this temporal integration of uncertainty estimation, by looking at the relation between NPRI$_d$ classes and energy imbalances. Both quantities are calculated over 24 hours (thus for 96 consecutive look-ahead times). While Möhrlen [158] discussed the benefits of considering a larger area when assessing the spread-skill relationship of wind power ensembles, our aim here is to show how skill forecasting can benefit from temporal averaging of spread and skill. Also, adding this temporal component would be relevant for real-world applications, since NPRI would be related to

[6]Here, the spread of the imbalance distributions can be quantified by the inter quartile range, or alternatively by the distance between the quantiles with proportion 0.1 and 0.9. In general, the inter quartile range is preferred, since it consists a more robust measure of the spread of an empirical distribution.

levels of energy imbalance over the considered time range. The current methods for uncertainty estimation of wind power predictions focus on providing pointwise uncertainty estimates (cf. discussion about marginal and simultaneous prediction intervals in Section 4.6). Showing a relation between NPRI and energy imbalance over a set of successive horizons will be an innovative contribution in that sense.

The 300 series of ensemble predictions over the 10-month evaluation period are considered. Both NPRI_d values and energy imbalances are calculated for look-ahead times between 0 and 24-hour ahead (day 1), 24 and 48-hour ahead (day 2), etc. NPRI_d values are sorted in 5 equally populated classes. To each of these classes are associated the empirical distributions of related energy imbalances, which contain 60 items each. We quote the same quantities than in Paragraph 5.4.4 for summarizing the characteristics of conditional probability distributions (i.e. mean, median, quartiles and 10 and 90% quantiles).

The NPRI_d ability to inform on the expected imbalance level

First, we concentrate on the same example we considered in Paragraph 5.4.4, which is the use of NPRI_d with ensemble predictions of wind power derived from ECMWF meteorological ensembles, for horizons between 48 and 72-hour ahead. Related conditional probability diagrams are depicted in Figure 5.10, given NPRI_d class. As an interpretation of these conditional distributions, one sees for instance that the relative imbalance over day 3 is between 55 and 295% of its climatological level when NPRI_d lies in its fifth class. This same relative imbalance ranges from 5 to 85% of the climatological value only when NPRI_d belongs to the first class.

Similarly to the analysis carried out in Paragraph 5.4.4, the increase of the mean energy imbalance with the NPRI_d class is steady and quasi linear, with mean imbalance levels ranging from 35 to 150% of their climatological value. Therefore, a first remark is that the ability of NPRI_d to be an indicator of the level of expected uncertainty is still valid when considering temporal averaging. In addition, imbalance distributions for every NPRI_d class appear to be sharper than the ones when using NPRI_h as an uncertainty indicator (cf. Figure 5.9). For instance, the inter quartile range for the NPRI class 4 equals 165% in the latter case, while it is only of 105% for the former one. These distributions are sharper first because the upper quartiles are at lower level and also because the lower quartiles are at a higher level (the same remark is valid for the 10% and 90% quantiles). Temporal averaging of skill smoothes the differences between low and large prediction errors. From a skill forecasting point of view, sharper distributions of expected imbalance are beneficial, since they give more confidence in the estimation of forecast uncertainty.

Comparing the cases for which NPRI_d is calculated with ECMWF-EPW, NCEP-EPW and the temporal ensembles

Comparison is made for day 2 like in Paragraph 5.4.4 between the three types of ensemble predictions of wind generation. Table 5.4 gather the results some quantiles and the mean

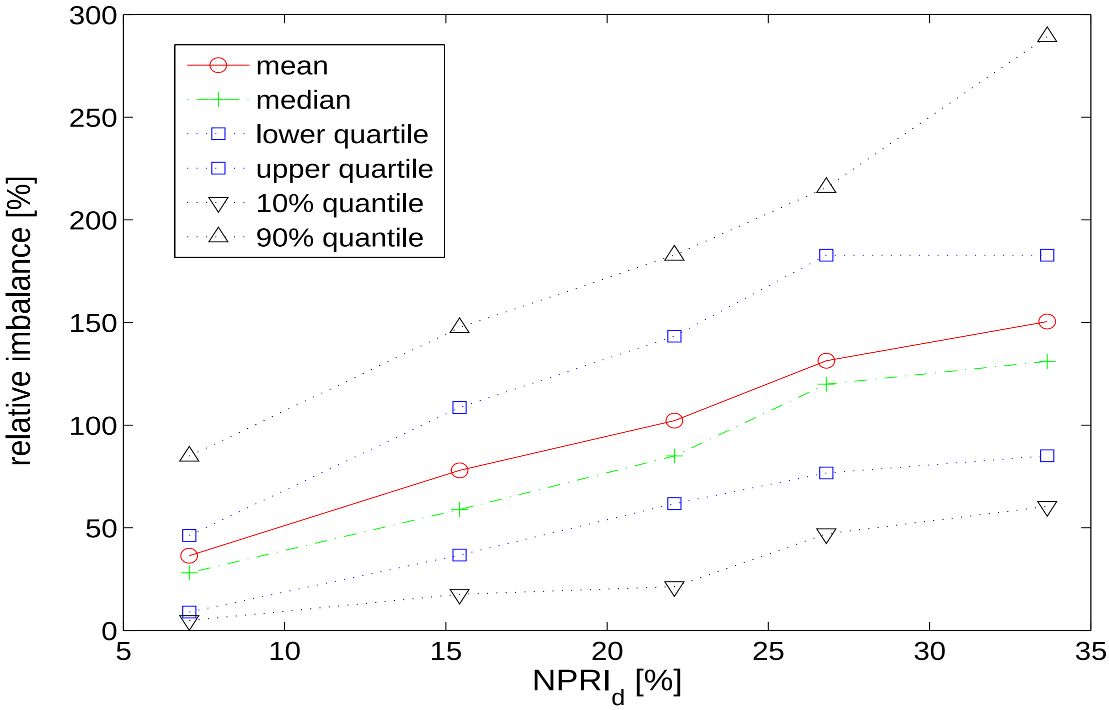


Figure 5.10: *Conditional distributions of energy imbalance given NPRI$_d$ classes. Both imbalances and NPRI values are calculated over day 3 (prediction horizons from 48 to 72-hour ahead). Empirical distributions are made up with 60 elements.*

of condition imbalance distributions, given the NPRI$_d$ class. If no particular mention, the following remarks are also valid for day 1 and day 3.

The ratios between the mean imbalance for NPRI$_d$ class 5 and NPRI$_d$ class 1 equal 4.2, 4 and 3.2 for ECMWF-EPW, NCEP-EPW and temporal ensembles respectively. This ratio is similar to the one we gave when focusing on NPRI$_h$ for temporal ensembles, when it is significantly lower for the two others. This decrease is mainly due to the smoothing of skill, not to a diminution in the ensembles' ability to resolve between situations. In a general manner, ECMWF-EPW and NCEP-EPW are still more skillful for indicating the expected level of prediction uncertainty. Though, one notices that temporal ensembles gain from the consideration of a temporal component for uncertainty estimation.

Imbalance distributions when considering NPRI$_d$ are much sharper than when considering NPRI$_h$ for the three types of ensembles. The inter quartile range is here between 26 (for NCEP-EPW) and 40% (for ECMWF-EPW) for the first class of NPRI$_d$ values. These values are slightly higher than the ones in Table 5.4. But for the other NPRI$_d$ classes, it is actually the inverse: the inter quantile range is much lower when imbalance values are sorted depending on NPRI$_d$ values. The reduction of the inter quartile range is up to 50%. Therefore, in term of skill forecasting, NPRI$_d$ appears to be a better indicator, owing to these sharper distributions.

Table 5.4: *Characteristics of the conditional imbalance distributions given the NPRI$_d$ class. Both NPRI$_d$ and imbalance values are gathered over day 2 (look-ahead times between 24 and 48-hour ahead). Results are for ECMWF-EPW, NCEP-EPW, and the temporal ensembles respectively. Empirical distributions are made up with 60 elements. Imbalance values correspond to relative imbalances (in % of their climatological level).*

(a) - ECMWF-EPW

NPRI$_d$ class	$r^{(0.1)}$	$r^{(0.25)}$	$r^{(0.5)}$	$r^{(0.75)}$	$r^{(0.9)}$	μ
1	3.6	8.6	24.4	48.8	88.3	34.9
2	23.5	41.2	61.4	100.8	171.1	76.9
3	29.7	60.9	94.7	140.3	192.7	112.3
4	67.3	91.1	115.0	164.1	203.0	130.1
5	60.1	102.5	125.6	176.4	222.6	144.6

(b) - NCEP-EPW

NPRI$_d$ class	$r^{(0.1)}$	$r^{(0.25)}$	$r^{(0.5)}$	$r^{(0.75)}$	$r^{(0.9)}$	μ
1	3.5	10.9	24.0	36.9	71.9	35.8
2	23.8	41.0	80.6	110.4	140.7	81.6
3	53.6	70.9	92.4	140.3	189.7	108.9
4	70.5	87.2	118.8	158.0	198.1	129.1
5	61.8	95.6	134.3	181.0	234.5	143.5

(c) - Temporal ensembles

NPRI$_d$ class	$r^{(0.1)}$	$r^{(0.25)}$	$r^{(0.5)}$	$r^{(0.75)}$	$r^{(0.9)}$	μ
1	7.1	12.9	30.4	49.1	97.9	41.9
2	8.1	44.3	75.8	118.4	170.1	88.9
3	47.4	77.9	96.8	153.2	229.1	116.3
4	29.9	77.4	115.6	154.8	194.9	117.8
5	55.1	88.9	120.9	176.3	195.0	135.5

We have already published results on the use of NPRI$_d$ with temporal ensembles of wind power for two other case-studies, in [116] and [185]. These two test cases consist in the prediction of the wind power production for the Golagh and the Klim wind farms, which we considered in Chapters 3 and 4, over periods of one and two years respectively. In these studies, the temporal ensembles are made up with 4 members, for a maximum look-ahead time of 24-hour ahead. NWPs are provided by HIRLAM (with a spatial resolution of approximately 20km) and issued every 6 hours. Point predictions of wind generation are produced with the forecasting method `M1`. The ability of NPRI$_d$ to inform of the level of expected imbalance is assessed with both contingency tables and conditional probability diagrams. For both test cases, it was clearly shown that using NPRI$_d$ with temporal ensembles have a value for the skill forecasting of wind power point predictions.

5.5 Conclusions

An investigation on the use of ensemble predictions of wind generation revealed its potential for associating skill forecasts to point predictions of wind generation. Focus was given to the possibility of providing a different information than the one provided by prediction intervals. It consists in prediction risk indices, which may be seen as estimates of the predictability of wind generation, and therefore as a comprehensive signal on the confidence forecast users may have on the forecasts provided by a point prediction method. The prediction risk index NPRI we introduced reflects the spread of ensemble members for a single or several consecutive look-ahead times. Various types of wind power ensemble forecasts have been considered: temporal ensembles (obtained by lagging the ECMWF control forecasts, 5 members), as well as ensembles derived from ECMWF (51 members) and NCEP (11 members) ensemble predictions of wind speed and direction (following the method proposed by Nielsen et al. [172]). The investigation has been carried out on the case-study of the Tunø Knob wind farm, over a period of 10 months. We focused on prediction horizons up to 72-hour ahead.

The methodology developed in this Chapter has consisted in considering various equally populated classes of NPRI values (more precisely 5 classes) and to establish their relation with related distributions of energy imbalance. It has been shown that for all the three different types of wind power ensembles the introduced NPRI could provide a useful information on the expected level of forecast uncertainty. In parallel, we have explained how to derive single-valued forecasts from wind power ensembles, by computing their mean or weighted mean. The single-valued predictions derived from ensembles of wind generation are significantly more accurate than those produced from meteorological control forecasts. For the test case we have considered, the error reduction (for the RMSE criterion) is up to 9% for 2-day ahead predictions and up to 29% for look-ahead times of 7-day ahead.

An important feature we have introduced is the possibility and interest of defining prediction risk indices for a look-ahead period. Such prediction risk indices permit to inform on the level of expected energy imbalance over the considered period. This consists an innovative contribution that contrasts with the works on the providing of pointwise uncertainty estimates. Moreover, an important conclusion is also that the gratis alternative of making up wind power ensembles by lagging available point predictions proved to be valuable, both for reducing the level of prediction error and for estimating the level of expected prediction uncertainty. Considering NCEP-based or ECMWF-based ensembles of wind generation is justified by their better ability of resolving between low and high predictability situations.

Perspectives regarding follow-up studies include: *(i)* a validation of the results on various types of test-cases located in zones with different meteorological characteristics (for which predictability may be more or less easily estimated); *(ii)* a further investigation on other possibilities for estimating the disagreement between ensemble members, e.g. the categorical measures (mode population and ensemble entropy); *(iii)* a study of other ensemble prediction systems, which may be more appropriate for short-range applications than the ones considered in the present Chapter; *(iv)* the use of such prediction risk indices as influential variables for the prediction interval estimation method described in Chapter 4.

Finally, a last perspective concerns the real-world utilization of prediction risk indices by end-users of wind power forecasts. As a first step, they can be communicated as a complement to point predictions. This way, forecast users will get used to that information, as a signal on the confidence they may have on the provided forecasts. Then, a second step will be to define how to make alternative decisions (more or less conservative depending on the risk aversion of end-users) depending on the value of the prediction risk index, and to demonstrate the resulting operational benefits.

6

The Value of Forecasting and the Benefits from Uncertainty Estimation

Abstract

This Chapter concentrates on the value of the prediction of wind generation and more particularly on the added value resulting from the use of an uncertainty information. For that purpose, the case of the trading of wind generation in electricity markets is considered. Several bidding strategies are described, either based on point forecasts only, or on more advanced strategies integrating an uncertainty information in the form of probabilistic forecasts as well as the modeling of the market participant's sensitivity to regulation costs. All strategies are applied on a real-world case-study consisting in the participation of a multi-MW wind farm operator in the Dutch electricity market over a year. It is shown how a higher accuracy of advanced prediction approaches translates to a reduction of regulation costs — of 38% here — by diminishing the quantities of energy subject to the regulation mechanism . Applying revenue-maximization strategies derived from probabilistic predictions of wind generation yields a further decrease of regulation costs — up to another 39%.

6.1 Introduction

Due to the deregulation of electricity markets, wind farm operators have the possibility to dispatch their production through electricity pools, instead of having recourse to bilateral contracts. The main characteristic of these markets is that one has to propose bids in advance — typically at noon for the following day starting at midnight — and is then charged for any *imbalance*, defined hereafter as the deviation between the bid and the actual production. Therefore, owing to the problem of predictability, the market value of wind power is reduced by the cost of these imbalances [174].

The participation of wind power producers in electricity pools is a decision-making problem: it is necessary to define what can be their optimal participation strategies, in order to maximize their revenues. For that purpose, wind farm owners need to use forecasts of future wind generation, typically for the next 24-48 hours, that may be produced from either reference or advanced methods. Chatfield [41] argued on the fact that a forecast (or a forecasting method) being considered as better than the others greatly depends on the operational context. A forecasting method that is slightly less accurate but which is much faster than the others may be preferred for some applications for instance. While forecast accuracy is the principal focus of forecasters, forecasts' consumers prefer prediction methods that are tailored to their needs, and which allow them to maximize their benefits (economical or not). This is what Murphy [165] referred to as the forecast *value*. The aim of the following developments is to assess the value of forecasts of wind generation, and to show how integrating associated uncertainty estimates in the decision-making process increases this value.

So far, several studies concerning the participation of wind energy in electricity markets have been carried out, considering different market mechanisms and various prediction methodologies. For instance, Usaola et al. [227] focused on the Spanish electricity market and tried to draw a relation between the accuracy of wind power prediction tools and the resulting income. Holttinen et al. [98] described the participation of Eltra and Elkraft (the independent system operators for western and eastern Denmark) in the Nord Pool and evaluated the cost of forecasting errors for these market players. Roulston et al. [196] envisaged the use of ensemble weather forecasts for better seizing the forecasting uncertainty and enhancing the position of wind generation in electricity markets. Bremnes [22] considered a simple model of the Nord Pool for evaluating the value of the probabilistic forecasting method he developed against basic point prediction approaches (i.e. based on the direct conversion of available NWPs into wind generation with a theoretical power curve). However, both Bremnes and Roulston et al. worked with virtual markets and have not applied their methodologies on real-world data. Finally, Bathurst et al. [13] concentrated on the uncertainty of wind generation and the resulting imbalance costs under the New Electricity Trading Arrangement (NETA) in the United Kingdom. From the use of probabilistic expected wind generation tables (giving the probabilities of the power production being in various energy bands for each lead time), the authors defined several bidding strategies accounting for the asymmetric structure of imbalance prices and the relative difference between imbalance and contract prices. However, the introduced developments are based on categorical forecasts of wind power, and should now be generalized to continuous predictions.

In this Chapter, different bidding strategies on a real-world day-ahead electricity market are evaluated. These strategies are based either on the use of advanced point forecasting methods or on the use of decision-making methods that consider the wind power forecasting uncertainty and a model of the wind farm operator's sensitivity to regulation costs. It is explained why point forecasts provided by state-of-the-art methods are not obviously the best bids one may propose on electricity pools.

Initially, the various European market mechanisms are concisely described and it is explained how wind power may be penalized in comparison with easily dispatchable generation. Then, we consider the participation of wind energy producers in an electricity pool with various bidding strategies. The assumptions for the present study are given and the problem is formulated in a general manner, so that the proposed bidding strategies can be applied to the case of the various European markets. The chosen short-term exchange market is the Dutch APX electricity market (APX standing for Amsterdam Power eXchange), which is associated to the regulation market run by the Transmission System Operator (TSO) TenneT for the Netherlands. For evaluating the bidding strategies, the participation of a real multi-MW wind farm in the market is simulated over a year. Our aim is to clearly show and quantify the interest of associating an uncertainty estimate to point forecasts of wind generation.

6.2 Trading wind generation in electricity markets

6.2.1 Describing the European electricity markets

At any time, the total amount of produced electricity must meet consumption. This balance is usually guaranteed by two different mechanisms: *(i)* production and consumption programs that are established in advance guarantee an a priori overall balance; *(ii)* a real-time balancing mechanism that allows the TSO to compensate any deviations in the electricity delivery programs. Electricity markets may be considered as an alternative solution to power units scheduling, since they permit a cost-effective match between supply and demand bids.

Wind is a highly variable energy production source, which can be seen as non-dispatchable (cf. Chapter 2). And, as discussed in Chapter 3, there will always be a deviation between predictions and actual power output. This is why wind power producers have to consider their revenue on the electricity markets as the combination of the income from the spot market and of the cost of imbalances.

Each electricity pool has its own rules, which determine the way electricity is to be sold or purchased, how the prices are settled, and the obligations that the participants (producers or consumers) are committed to. In order to stimulate the development of renewables, some pools have special rules supporting wind generation, such as guaranteed prices or no program-responsibility. For instance in Spain, wind generation is included in a special regime, with different ways of payment for the energy injected in the grid [227]. In parallel in Denmark, wind generation may be covered or not by prioritized dispatch (power balance handled by the TSO) depending on the turbines' age [164]. At the inverse, in UK all energy producers participating in the market are considered as equal [13]. An overview of European electricity markets is given in [144] and the regulatory frameworks for the integration of wind energy in some of the European countries are described in [226].

In this study, we consider that all energy producers (i.e. wind or conventional energy producers) participate in the electricity market under the same rules: they have to propose their bids on the *spot market* (no fixed price), and they are then financially responsible of their deviations from schedule. The costs of keeping the balance are charged to the participants, proportionally to their imbalances. On these spot markets, bids are to be given before gate closure, which occurs between 10:00 and 14:00 depending on the market. Quantities of energy are to be proposed for the following day from midnight to midnight, for every unit of time usually referred to as *Program Time Unit* (PTU). The PTU length ranges from 15 minutes to 1 hour. Therefore for a wind energy producer aiming at participating in these electricity pools, relevant prediction horizons are between 10 and 36 hours ahead, with a temporal resolution ideally corresponding to the PTU length. Note that most of the forecasting tools provide power predictions, and that market bids consist in quantities of energy. Then, an assumption and calculation is necessary to go from power to energy predictions (e.g. constant power production over the time period).

Because of the significant delay between gate closure and the beginning of the energy delivery period, certain electricity pools also integrate *intradaily markets* (also referred to as balancing markets), where it is possible to take corrective actions. The gate closure for these balancing markets often occurs between 30 minutes and 2 hours before the time of delivery. For certain countries e.g. the Netherlands participants may have to pay for proposing corrective bids in the intradaily markets [198].

Finally, a last aspect is the *regulation market* (or better say mechanism), which is managed by the TSO. This mechanism deals with the real-time energy delivery and ensures the match of generation and load at any time. Even if the price settlement and the determination of fees for not respecting the generation program vary from one country to the other, the general principle remains the same. There are actually three regulation scenarios:

- production may perfectly match consumption, and then no regulation is needed,
- if the production is not sufficient to meet the load, this leads to an up-regulation situation, for which it is necessary to increase generation. Alternatively, it may be envisaged to lower consumption, if there are special agreements between energy retailers and customers. Here, it is assumed that regulation only consists in acting on the production level,
- at the contrary, if the production level is higher than the consumption level, then one faces a down-regulation situation, for which it is necessary to decrease generation. If the grid has good interconnections with neighboring systems, it may be possible to export the energy surplus, but more often some curtailment is envisaged for lowering the surplus generation.

The connections between various areas of the overall power system may also step in the regulation problem: due to a limitation of these connections, it may be necessary to dispatch both positive power in a certain area and negative power in the next area. Also, in certain countries (such as the ones participating in the Nord Pool for instance), if a given deviation from contract actually helps the TSO for regulation, then the power producer is not charged for that imbalance. All these considerations make that imbalance prices are highly variable and hardly predictable. However, these imbalance prices should normally reflect the production (or curtailment) costs, include a premium for readiness, and discourage power producers to plan imbalances.

6.2.2 Assumptions for the present study

Either for simulating the participation of a wind power producer in European electricity pools or for developing advanced bidding strategies, it is necessary to formulate some assumptions about the impact of wind generation on the markets' behavior, i.e. its influence on both the spot and imbalance prices. In addition, the framework of the present study has to be thoroughly described, in order to define to what extent it will be possible to generalize the obtained results.

It is expected that if large amounts of wind power are introduced in a power system, this would tend to lower the spot price on the related electricity pool [164] in the long-term . However, our concern here is about the influence of the amount of generated wind power on the market clearing price in the short-run: does it have a positive or a negative impact on that price, and is that impact significant? For the case of the Nord Pool, which is a market highly penetrated by wind generation, several studies have been carried out in order to answer these two questions. Kristoffersen et al. [122] developed a market simulation model at Eltra (the system operator for western Denmark), and simulated the market consequences of large-scale integration of wind power. One of the main conclusions is that wind power affects the Nord Pool spot market by producing a downward pressure on market price. In parallel, Morthorst [164] observed in an analysis of the market data over the years 2001-2002 there was a tendency that more wind power in the system leads to relatively lower spot prices (and vice versa), though no strong relation is established. Also, this trend was not found statistically significant. In the case of the present study, we assume that the electricity pool is not highly penetrated by wind power and consequently that potential effects related to wind penetration in the market can be neglected. Wind farm operators are considered as *price-takers*, i.e. as economic entities that are too small relative to the market to affect its clearing price. They have no market power: the market clearing price can hence be seen as fatal for such participants. This is in line with the assumptions formulated by Skytte [204] and Usaola et al. [226, 227].

Similar hypotheses are made regarding the influence of wind generation on the imbalance prices. Since wind energy is non-dispatchable and not easily predictable with a high level of accuracy, the participation of wind generation in electricity pools compulsorily yields a certain volume of imbalances. Such imbalances would not be witnessed if only conventional units were proposing bids on the market. The imbalances resulting from wind generation are not necessarily in the same direction than the imbalances one would witness in a market not penetrated by wind. Though, as wind prediction errors are spatially correlated, it appears likely that deviations from contracted energy will have the same sign for neighboring wind production sites (assuming that their operators follow similar bidding policies). Morthorst [164] noticed that for the case of the regulation in the western part of Denmark over 2002, the amount of generated wind power had indeed an influence on the quantity (and also the direction) of regulation needs. Though, it turned out that the regulation unit costs in this area were almost independent of the level of wind generation. As we have made the assumption that a single wind power producer has no market power, we also assume that the bidding policy of this power producer alone cannot impact imbalance prices. Note that this assumption is implicit in the developments by Bremnes [22] and Linnet [138] when deriving their optimal bidding quantile strategies, and was explicitly formulated by Barthurst et al. [13].

Moreover, it is assumed that wind power producers act in electricity markets as conventional producers and do not benefit of derogatory rules such as guaranteed price, no program-responsibility, or premiums for nature-friendly electricity generation. So far, some markets provide subsidies to wind generation for easing its integration, but in the future wind is expected to compete as equal with conventional means. Therefore, this assumption is consistent with future developments of electricity markets. In addition, due to the stochastic and intermittent nature of wind, we have considered that they do not make any bids for regulation and reserve power supply. Eventual corrective actions in intradaily markets are also not considered. Holttinen [98] already showed that participating in such balancing markets would significantly increase the revenue of wind power producers. Though, Barthurst et al. [13] showed that even by diminishing the window between gate closure and actual delivery, the amount of energy in imbalance for wind power producers was still substantial, and thus that a financial risk was still remaining. The methodology introduced hereafter concentrates on the providing of optimal decisions on the day-ahead market when imbalances are expected. It could be further extended for considering the possibility of taking corrective actions in balancing markets, in a stochastic programming framework.

We make the assumption that the only control wind power producers have on their production is binary: supplying or not supplying the energy to the grid. They do not have possibilities to down-regulate the wind generation, or to couple that power output with conventional generation means, or even to use energy storage devices. Combining wind generation with conventional means or storage would permit to lower the amount of imbalances on the market. Hence, this is a restrictive assumption on the value of wind generation in electricity markets in comparison with the case for which wind is incorporated into a broader resource portfolio [55]. The possible coupling of wind generation is not considered since the present study focuses on the value of the sole forecasting. However, the way the methodology could be further extended for integrating these aspects is described.

The price limit in the bids sent to the spot market will be set to the minimum so that all of the generated energy is sold. By doing so, the wind power producer determines his own program. This producer is then paid the hourly system marginal price, referred to as the *market clearing price* or *spot price*, for the corresponding amount of energy stipulated in that program.

The case of bids for the production from a single wind farm is considered. Note that this is on the pessimist side since prediction errors are expected to be higher than for the case of multiple wind farms, for which aggregation smoothes errors [68, 188]. Also, the choice of a given point forecasting method would obviously have an impact on the resulting income. In the present Chapter, we only consider predictions provided by the point forecasting method `M1` and related predictive distributions estimated with the adapted resampling approach. Since the superiority of a given forecasting method or of associated prediction intervals has not been demonstrated, `M1` is chosen at random. Again, our main goal is to carry out a qualitative analysis and illustrate how a wind power producer may increase his revenues by utilizing advanced forecasting approaches and also by applying more advanced bidding strategies integrating an uncertainty information. This is why we do not attach much importance to the choice of the advanced point forecasting approach.

6.2.3 Formulation of the problem

For any PTU $t+k$, a market participant has to propose an amount of energy E^c_{t+k}, referred to as the *contracted energy*. The $(t+k)$-index is used for designating a given PTU since it indeed corresponds to a lead time $t+k$ when bids are proposed at time t. The revenue of a market participant proposing an amount of energy E^c_{t+k} but actually generating E^*_{t+k} can be formulated as

$$\mathrm{R}_{t+k}(E^c_{t+k}, E^*_{t+k}) \;=\; \pi^c_{t+k}E^c_{t+k} \;+\; T^c_{t+k}(E^c_{t+k}, E^*_{t+k}), \tag{6.1}$$

where π^c_{t+k} is the market spot price for the PTU $t+k$, and $T^c_{t+k}(E^c_{t+k}, E^*_{t+k})$ is the imbalance cost on the regulation market, resulting from an imbalance d^*_{t+k} defined as

$$d^*_{t+k} \;=\; E^*_{t+k} - E^c_{t+k}, \tag{6.2}$$

such that

$$T^c_{t+k}(E^c_{t+k}, E^*_{t+k}) \;=\; \begin{cases} \pi^{c,+}_{t+k}d^*_{t+k}, & d^*_{t+k} \geq 0 \\ \pi^{c,-}_{t+k}d^*_{t+k}, & d^*_{t+k} < 0 \end{cases}, \tag{6.3}$$

with $\pi^{c,+}_{t+k}$ and $\pi^{c,-}_{t+k}$ the imbalance prices for positive and negative deviations from contract respectively. Often, it is said that excess supply is sold at the *spill price*, and that missing energy is bought at the *top-up price*. Note that these two imbalance prices depend on the considered regulation mechanism. In certain cases, they can be equal to a certain proportion τ of the market clearing price:

$$\pi^{c,+}_{t+k} \;=\; -\pi^{c,-}_{t+k} \;=\; -\tau\pi^c_{t+k}, \quad \forall t, k. \tag{6.4}$$

For the specific example of the Spanish trading mechanism, that proportion τ equals 10% [227] (if the market participant fulfills certain requirements). This mechanism makes that whatever the overall system imbalance, a participant is charged for his own imbalance. However, for some other regulation mechanisms, imbalance prices depend on the sign of the imbalance as a whole, so that participants who offset the system imbalance are not penalized. In practice, on the Nord Pool for instance, if the overall system needs up-regulation and a wind power producer actually generates more energy than contracted, he is not penalized for that deviation from contract, but instead that amount of energy in exceedance is sold at the spot price [138]. This would translate to $\pi_k^{c,+}$ being equal to π_k^c in the imbalance cost formula given by Equation (6.3).

Equation (6.1) expresses the participant's revenue as the income resulting from the level of contracted energy on the spot market, to which is added the revenue or costs from imbalances traded on the regulation market. However, since by definition we have

$$E_{t+k}^c \;=\; E_{t+k}^* \;-\; (E_{t+k}^* - E_{t+k}^c) \;=\; E_{t+k}^* \;-\; d_{t+k}^*, \tag{6.5}$$

Equation (6.1) can be reformulated such that a participant revenue for the PTU $t+k$ becomes

$$\mathrm{R}_{t+k}(E_{t+k}^c, E_{t+k}^*) \;=\; \pi_{t+k}^c E_{t+k}^* \;-\; T_{t+k}^*(E_{t+k}^c, E_{t+k}^*), \tag{6.6}$$

where

$$T_{t+k}^*(E_{t+k}^c, E_{t+k}^*) \;=\; \begin{cases} \pi_{t+k}^{*,+} d_{t+k}^*, & d_{t+k}^* \geq 0 \\ \pi_{t+k}^{*,-} d_{t+k}^*, & d_{t+k}^* < 0 \end{cases}. \tag{6.7}$$

Therefore, this revenue can be seen as the income coming from the selling of actual wind generation at the spot price, minus the costs for regulation, which is a function of the market clearing price π_{t+k}^c, the imbalance prices $\pi_{t+k}^{c,+}$ and $\pi_{t+k}^{c,-}$, and the amount of energy produced in imbalance d_{t+k}^*. From now on, $\pi_{t+k}^{*,+}$ and $\pi_{t+k}^{*,-}$ will be referred to as the *regulation unit costs* for downward and upward regulation respectively. Such regulation unit costs are simply given by

$$\pi_{t+k}^{*,+} \;=\; \pi_{t+k}^c - \pi_{t+k}^{c,+}, \tag{6.8}$$
$$\pi_{t+k}^{*,-} \;=\; \pi_{t+k}^c - \pi_{t+k}^{c,-}, \tag{6.9}$$

The formulation given by Equation (6.6) has the advantage that the first component of the revenue indeed corresponds to the income one would receive if using perfect predictions. Also, since the contracted energy only appears in the second component of the participant's revenue, then maximizing $\mathrm{R}_{t+k}(E_{t+k}^c, E_{t+k}^*)$ translates to minimizing the costs for regulation $T_{t+k}^*(E_{t+k}^c, E_{t+k}^*)$. Normally, for the income on the day-ahead market and the resulting regulation costs the following related properties should hold:

$$T_{t+k}^*(E_{t+k}^c, E_{t+k}^*) \geq 0, \qquad \forall E_{t+k}^c, E_{t+k}^*, \tag{6.10}$$
$$\mathrm{R}_{t+k}(E_{t+k}^c, E_{t+k}^*) \leq \mathrm{R}_{t+k}(E_{t+k}^*, E_{t+k}^*), \quad \forall E_{t+k}^c, E_{t+k}^* \tag{6.11}$$

which means that a participant is always penalized for the imbalances he is responsible for, and hence that the maximum income he can receive is the one he would get by bidding the actual amount of generated energy. If these properties do not hold, it may encourage participants to plan imbalances. In the specific case for which a market participant is not charged for his imbalance since this imbalance helps the regulation process, then we simply have $T_k^*(E_{t+k}^c, E_{t+k}^*) = 0$. It will be seen in a following Section that properties (6.10) and (6.11) do not hold for all the markets, and thus that a market participant might be rewarded for his imbalance.

Owing to the nature of regulation costs, which are mainly penalizing, the revenue can also be written in the form of a performance ratio γ. It will be used for the evaluation of the rival bidding strategies. The performance ratio γ is calculated over a certain period of time by normalizing the actual revenue by the revenue that would be obtained if one had the possibility to use perfect forecasts. Considering en evaluation period of N_T participation days in the electricity pool, each participation day covering N_{PTU} daily, γ writes

$$\gamma = 1 - \frac{1}{N_T.N_{\mathrm{PTU}}} \frac{\sum_{t=1}^{N_T} \sum_{k=1}^{N_{\mathrm{PTU}}} T_{t+k}^*}{\sum_{t=1}^{N_T} \sum_{k=1}^{N_{\mathrm{PTU}}} \pi_{t+k}^c E_{t+k}^*} \tag{6.12}$$

The proposed performance ratio is such that $\gamma \in (-\infty, 1]$. It is obvious that for perfect prediction $\gamma = 1$ since deviations are null.

The optimization problem then consists in finding the bid $\tilde{E}_{t+k}^c$ for each PTU $t+k$ that would permit to maximize the participant's *utility function*, i.e. a function that links decisions to their related benefits for the market participant. These benefits are not necessarily economical. They correspond more to a measure of happiness or satisfaction for the market participant. Hence, maximizing this utility function may signify maximizing the participant's revenue on the market, but it can also integrate a risk aversion component, or other views of the participant on preferred bidding strategies. In the following Section, we describe how this may be done by considering point or probabilistic forecasts of wind generation.

6.3 Definition of advanced bidding strategies

6.3.1 Point predictions as the best bids

If point predictions consist the only information one has about future wind generation, they then comprise the best bids one can propose on the electricity pool. Point prediction methods usually provide estimates of the expected power production with a forecast length of 2-3 days ahead, with a forecast resolution ranging from 15 minutes to 1 hour. These estimates correspond to the average power production over the previous time period. For instance, a 12-hour ahead forecast of 6MW provided by a forecasting method with an hourly resolution means that the average power production between 11 and 12-hour ahead is expected to be 6MW. Therefore, it appears reasonable to give a forecast of the wind energy generated during that period as the product of the average power production by the length of the time period t_r:

$$\hat{E}_{t+k/t} = \hat{p}_{t+k/t}.t_r. \tag{6.13}$$

These energy forecasts are then directly used for defining an optimal contract level $\tilde{E}^c_{t+k}$ for each PTU $t+k$:

$$\tilde{E}^c_{t+k} = \hat{E}_{t+k/t}. \tag{6.14}$$

By applying such bidding strategies, the prediction errors from the considered forecasting method directly translate to regulation costs. Usaola et al. [227] applied these point-prediction-based bidding strategies for evaluating the value of the Sipreólico method if used for participating in the Spanish electricity market. In a follow-up study, Usaola and de Arriba Segurado [226] extended that study to the case of various state-of-the-art point forecasting methods, and by simulating the participation of wind energy in European electricity markets in general. In both cases, the authors witnessed a reduction from 30% up to 50% of regulation costs when using advanced prediction approaches instead of the persistence method. Also, they showed that participants' revenues could be significantly increased if bids were updated through the intradaily market. This is in line with the conclusions by Holttinen [98], who explained that a more flexible electricity market, with a shorter period between gate closure and energy delivery would be beneficial for wind power producers. This is due to the effect of the lead time on the accuracy of point forecasting methods, as shown in Paragraph 3.5.2: whatever the considered advanced approach, its level of prediction error significantly increases with the look-ahead time. More precisely, the standard deviation of prediction error distributions augments as the lead time gets further, which signifies an increase of the frequencies of occurrence of larger prediction errors, and therefore a higher financial risk for the market participant.

6.3.2 Advanced bidding strategies based on probabilistic forecasts

Instead of seeing E^*_{t+k} as a true effect one wants to predict as accurately as possible with $\hat{E}_{t+k/t}$, one may have a probabilistic view of the problem, by considering that E_{t+k} is a random variable and that E^*_{t+k} is a realization of that random variable. It was discussed in previous Chapters that state-of-the-art point prediction methods, which aims at minimizing a quadratic loss function, provide point forecasts $\hat{E}_{t+k/t}$ that are indeed estimates of the expectation $\bar{E}_{t+k}$ of the distribution of E_{t+k}. Denote by F^E_{t+k} the density function of E_{t+k}. Here, we consider that we work with normalized variables, and hence E_{t+k} can only take values between 0 and 1. Its expectation is then given by

$$\bar{E}_{t+k} = \int_0^1 x.F^E_{t+k}(x)dx. \tag{6.15}$$

The expectation is only a summary statistics of what can be the realization E^*_{t+k}, it cannot give an information on what could happen. Note that from now on the deviation from contract d_k is also seen a random variable, defined as

$$d_{t+k} = E_{t+k} - E^c_{t+k}, \tag{6.16}$$

such that d^*_{t+k} denotes a realization of that random variable.

After thoroughly studying the imbalance prices on the regulation market of the Nord Pool, Skytte [204] concluded that the regulation unit costs characteristics may encourage traders with fluctuating production to be more strategic in their way of bidding on the day-ahead market. Such strategic bidding would permit to limit the extra costs that would be charged for large imbalances. But, for optimizing these bids, it is necessary to know not only the expectation of future wind generation but also what *could* be this wind generation at a given lead time. Hereafter, we will suppose that a forecast $\hat{F}^E_{t+k/t}$ of the density function of E_{t+k} is available at time t. Such a forecast can be produced from the methodology developed in Chapter 4 for instance. Then, in the following is described a generic approach which allows one to derive optimal bidding strategies from the modeling of the sensitivity of the market participant to regulation costs, consequently integrated in the decision-making process.

Modeling the market participant's utility function

The utility assigns a degree of happiness in the form of a numerical value to every possible outcome a decision-maker may be faced with. Because we do not have any control on the first component of the participant's revenue formulated by Equation (6.6), the problem of maximizing a participant's utility function directly translates to minimizing the *loss function* associated to the second component of this revenue. Note that like for the definition of 'utility', the loss does not compulsorily have an economical meaning: it reflects the sensitivity of the market participant to regulation costs. Ideally, this model of the loss function results from a discussion between an analyst and the forecast user.

In a first stage, let us define the basic properties of a function g, which gives the loss associated to a given deviation from contract d^*:

$$g \;:\; d^* \in [-1,1] \;\rightarrow\; g(d^*) \in \mathbb{R}^+. \tag{6.17}$$

The deviations d^* are contained in the range $[-1,1]$ since we work with normalized variables. Values of $g(d^*)$ are always greater than or equal to 0, following the assumption that a market participant does not plan imbalances, and thus does not expect to be rewarded for an imbalance. It would be possible to build a loss function depending on k, in order to account for diurnal trends in regulation unit costs. Also, g could be made a function of t for modeling the influence of the time of the year on regulation prices. The loss functions introduced hereafter are independent of k. And, we will study the possibility of introducing a different model for losses depending on the season of the year.

A fair assumption consists in saying that the loss cannot decrease when deviations from level of contracted energy increase. This translates to g being a non-increasing function for negative deviations, and similarly a non-decreasing function for positive deviations from contract. Denoting by g' the derivative of g, this writes

$$g'(d^*) \leq 0, \qquad \forall d^*,\, d^* < 0, \tag{6.18}$$
$$g'(d^*) \geq 0, \qquad \forall d^*,\, d^* > 0. \tag{6.19}$$

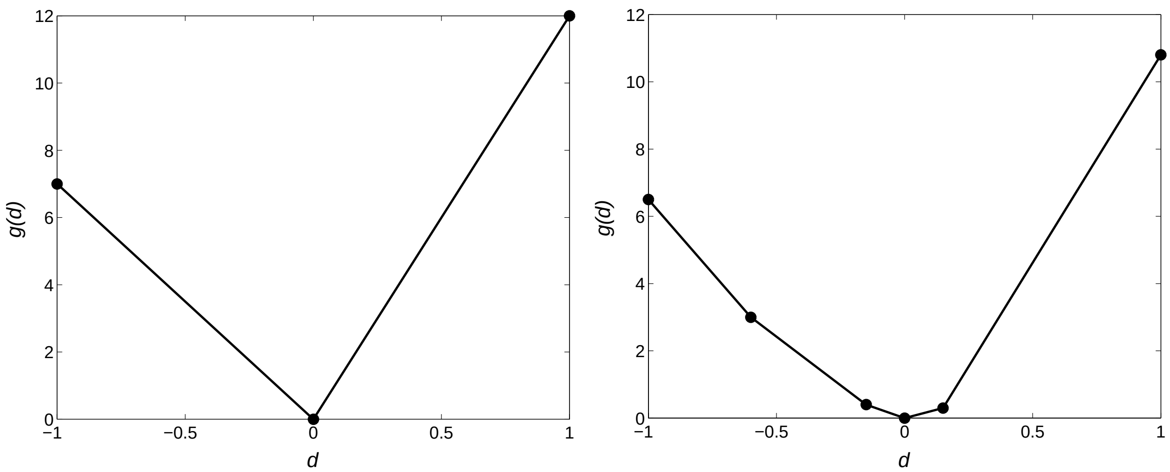
Though, if g is defined as a piecewise function, in order to reflect different level of losses for different ranges of deviations, then g' is not defined for certain values of d^*. In addition, g must be such that

$$g(0) \;=\; 0, \tag{6.20}$$

since the market participant does not have to buy or sell energy on the regulation market if there is no deviation from contract. Also, this is in line with the idea that the use of perfect predictions would lead to the maximum income. Finally, note that one may prefer to have g continuous, though this does not appear as a constraint in the definition of the loss function. Continuity, in complement to the positivity of the second-order derivative, would translate to g being a convex function.

The most simple way of defining g is to consider that the loss is directly given by the regulation unit costs on the market and thus by $\pi^{*,+}_{t+k}$ and $\pi^{*,-}_{t+k}$ as defined in Equations (6.8) and (6.9). These two regulation unit costs for positive and negative deviations then define the slope of linear functions for positive and negative values of d:

$$g\,:\, d^* \;\rightarrow \left\{ \begin{array}{ll} \pi^{*,+}_{t+k}|d^*|, & d^* \geq 0 \\ \pi^{*,-}_{t+k}|d^*|, & d^* < 0 \end{array} \right. . \tag{6.21}$$



(a) Market-based loss function

(b) More advanced loss function

Figure 6.1: *Example of two different loss functions based on either on the regulation unit costs only, or refined after a discussion between the analyst and the market participant. The latter loss function reflects the possibility for that wind power producer to face imbalances in a different manner depending on their magnitude.*

Both the market clearing price π^{c}_{t+k}, and the prices for negative $\pi^{c,-}_{t+k}$ and positive deviations from contract $\pi^{c,+}_{t+k}$ for PTU $t+k$, are not known when bidding at time t. Therefore, the regulation unit costs $\pi^{*,+}_{t+k}$ and $\pi^{*,-}_{t+k}$ are also unknown[1]. Consequently, regulation unit costs in Equation (6.21) are replaced by forecasts or estimates:

$$g:\ d^* \ \rightarrow \begin{cases} \hat{\pi}^{*,+}_{t+k}|d^*|, & d^* \geq 0 \\ \hat{\pi}^{*,-}_{t+k}|d^*|, & d^* < 0 \end{cases} . \tag{6.22}$$

When focusing on the regulation market for the Nord Pool, Skytte [204] showed that it was possible to model the regulation unit costs as a function of the market clearing price (which is settled few hours after gate closure). Similarly, Olsson and Söder [180] presented a model of the regulating power market prices, based on an ARIMA process, which considers the correlation of successive market clearing prices, the structure of the regulation mechanism, as well as the submission time. This ARIMA model can be applied for predicting regulation unit costs. In both cases, regulation prices cannot be forecast before the market clearing price is known. However, one may use in a first stage a model for forecasting spot prices (see Conejo et al. [50] for instance), which can then be used as input to a model able to predict the regulation unit costs. Alternatively, after an analysis of the Dutch regulation mechanism, Saint-Drenan [198] concluded that although imbalance prices were highly volatile, an estimation strategy consisting of climatology-like forecasts was already valuable. In the present study, we will only assume that it is possible to predict a trend (on an annual or a quarterly basis) for the regulation unit costs.

Figure 6.1 gives an example of two loss functions that could be used for modeling the sensitivity of a market participant to deviation costs. Here, we follow the analysis of the Nord Pool carried out by Morthorst [164], in which the average unit costs for up-regulation and down-regulation over 2002 were estimated as 7€/MWh and 12€/MWh respectively. These values have been used afterwards by Bremnes [22] for illustrating what could be the optimal quantile from a probabilistic forecast for bidding on the Nord Pool for that particular year. If these average unit costs are used to build the loss function, they define the slope of the two linear parts of the function shown in Figure 6.1 (a).

But then, imagine that an exchange between the analyst and the wind power producer lead to the definition of the cheapest way to deal with positive and negative deviations depending on their magnitude. Note that in this case we leave aside the assumption we formulated regarding the possibility of coupling wind production with conventional generation (or storage). From that exchange can be defined more advanced loss functions in the form of piecewise linear functions (or maybe even quadratic for large deviations, reflecting the fact that these large deviations may be really costly. An example of such an advanced loss function is given in Figure 6.1 (b), for which the slope of the linear portions increases when deviations get larger. For the case of this loss function, the wind power producer always envisages the cheapest solutions in a first stage for dealing with imbalances.

[1]Even for the specific case of the Spanish market, for which $\pi^{c,+}_{t+k}$ and $\pi^{c,-}_{t+k}$ are fixed to a proportion of the market clearing price (cf. Paragraph 6.2.3), regulation unit costs cannot be known in advance.

Bidding strategies tailored to end-user needs

Given the probabilistic distribution $F^E_{t+k}$of wind generation for PTU $t+k$, and given the loss function g, an optimal contract level can be determined in various ways. Indeed, the definition of an optimal bidding strategy, like the modeling of the loss function, depends on the sensitivity of the market participant to penalties. The market participant may want to optimize his utility on the electricity pool over a certain period of time, or alternatively to minimize the losses resulting from large deviations on a day-to-day basis since he would not be able to pay for the large related costs. Some other bidding strategies can also be defined, specially tailored to the wind power producer's needs, but the ones presented above are the most popular. The first strategy which consists in maximizing utility over a certain period is often referred to as the *Probabilistic Choice* (PC) [156] or *Minimum Expected Imbalance Cost* policy [13]. Since the latter type of strategies focuses on minimizing the risk of large losses, it is called *Risk Analysis* (RA) [156] or *Risk Adverse* policy [13] in the literature. This type of strategies is at the expense of a maximum utility on the long-term though.

Maximizing the utility of the market participant over a certain period of time is equivalent to maximizing the expectation of that utility for every PTU over that period. Following the previous developments, it is also equivalent to minimizing the expectation of the introduced loss function g for each of these PTUs. The loss expectation for PTU $t+k$ can be written as:

$$\mathbb{E}\left[g(d_{t+k})\right] = \int_0^1 g(x - E^c_{t+k}) F^E_{t+k}(x) dx, \tag{6.23}$$

And, the optimization problem to be solved for every PTU writes

$$\tilde{E}^c_{t+k} = \underset{E^c_{t+k}}{\arg\min}\ \mathbb{E}\left[g(d_{t+k})\right]. \tag{6.24}$$

In the case for which the loss function is a direct model of the regulation unit costs (cf. Equation (6.21), minimizing the expectation of g for PTU $t+k$ directly translates to minimizing the expectation of the regulation costs for that PTU. The optimal bid $\tilde{E}^c_{t+k}$ given by (6.24) is actually equivalent to:

$$\tilde{E}^c_{t+k} = \underset{E^c_{t+k}}{\arg\min}\ \mathbb{E}\left[T^*_{t+k}(E^c_{t+k}, E_{t+k})\right]. \tag{6.25}$$

Both Bremnes [22] and Linnet [138] showed that in such a case the theoretical solution of the optimization problem 6.24 for PTU $t+k$ is calculated as

$$\tilde{E}^c_k = {G^E_{t+k}}^{-1}\left(\frac{\pi^{*,+}_k}{\pi^{*,+}_k + \pi^{*,-}_k}\right), \tag{6.26}$$

where G^E_{t+k} is the cumulative distribution function for the random variable E_{t+k}.

In practice, G^E_{t+k} is replaced by an approximation of the cumulative distribution function, which can be derived from a predictive distribution of wind generation $\hat{F}^E_{t+k/t}$, given for instance by the methodology developed in Chapter 4. Also, regulation unit costs in Equation (6.26) have to be replaced by estimates. From the assumption that wind generation and regulation costs are independent random variables, $\pi^{*,+}_k$ and $\pi^{*,-}_k$ can be replaced by estimates of their expectation [13, 138], for instance given by average values over a past period. However, this theoretical solution only holds when the loss function is modeled as a function of the regulation prices only, as it is done in Equation (6.21). If considering more advanced loss functions (e.g. piecewise linear or quadratic functions), numerical optimization methods have to be envisaged for determining the optimal bid $\tilde{E}^c_k$. Owing to properties (6.18) and (6.19), and since G^E_{t+k} is a monotically increasing function, the optimization problem 6.24 admits a unique minimum over the range of possible contract levels. Local optimization methods are thus sufficient for determining optimal bids.

Actually, the main interest of the theoretical result given by Equation (6.26) lies in the fact it tells that the bid $\tilde{E}^c_k$ one can make on an electricity pool for maximizing the participant's income is not given by a point forecast (which is an estimate of the wind generation expectation). It is instead given by a particular quantile of predictive distributions of wind generation. The proportion of this quantile is a direct function of the regulation prices. With the average up- and down-regulation unit costs calculated by Morthorst [164] (mentioned in Paragraph 6.3.2), it is indeed by bidding the quantile with proportion 0.63 of predictive distributions of wind generation that one would have maximized his income on the Nord Pool over 2002. Therefore, when defining market bids from point forecasts only, it is not necessarily the most accurate point forecasting method that would lead to the higher revenue on the market. This is because the criterion used for estimating the accuracy (or quality) of point predictions is not the same than the one considered for assessing their value in an operational context. This has recently been illustrated by Nielsen et al. [169].

Alternatively, a wind power producer might not focus on maximizing his income only, but would actually prefer to minimize the risk of large losses. If this producer is not a big player on the market, he will have a higher loss aversion than preference for related gains. This would translate to applying a risk adverse policy as introduced above. In such a case, the optimization problem does not consist in minimizing the expectation of the loss function for each PTU. Indeed, this problem consists in finding the bid $\tilde{E}^c_{t+k}$ which minimizes the worst possible scenario. Such a minimax problem can be formulated as

$$\tilde{E}^c_{t+k} = \arg\min_{E^c_{t+k}} \max_x g(x - E^c_{t+k}) F^E_{t+k}(x), \tag{6.27}$$

where probabilistic distributions F^E_{t+k} are again to be replaced by predictive distributions $\hat{F}^E_{t+k/t}$. The optimization problem (6.27) can be solved with appropriate numerical methods.

Note that when applying a risk adverse policy the loss function must reflect the participant aversion for losses, which is obviously not optimally represented by linear functions with slopes given by averages of regulation unit costs. In fact, it would be preferable to consider upper bounds on the value of expected imbalance costs, or alternatively to use quadratic loss functions.

6.4 Evaluation of bidding strategies on a European electricity pool

In order to simulate the participation of wind energy in a European electricity market, we use series of wind power forecasts over a one-year period (from January 1st to December 31st 2001), as obtained from the application of the forecasting method M1 to a 15MW wind farm. This farm is the Golagh wind farm in Ireland, located in hilly terrain, which we already considered in Chapter 3, when studying the error characteristics of some state-of-the-art prediction approaches. The performance of the method M1 over this one-year period is actually at a level similar that given in the description of the models' performance in Appendix D.

Regarding the choice of the electricity market, we consider here the Dutch electricity market over the years 2002 and 2003. The specificities of that particular market are briefly described in a first part of this Section. Then, following the assumptions formulated in Paragraph 6.2.2, we simulate the participation of the operator of the Golagh wind farm in that market. Our first aim is to show how the use of an advanced wind power prediction tool can substantially increase the wind power producer's income, by reducing the amount of imbalances. Secondly, we explain how to implement the advanced bidding strategies developed in the previous Section, and quantify the related benefits for the wind power producer. The rules and regulations of the considered market evidently affect the results, and this is taken into account in the analysis of the simulation results. However, a roughly similar architecture between these rules and regulations can be found, so that the conclusions from our work may be generalized (with cautiousness) to other electricity pools.

6.4.1 Specificities of the Dutch electricity market

The Dutch electricity market we consider here consists in a day-ahead market APX (standing for Amsterdam Power eXchange), a regulation market run by the transmission system operator TenneT, and more recently in an adjustment market that we will not deal with in this Chapter.

The APX spot market enables the participants to buy and sell electricity for any of the 24 hours of a day one day in advance. Everyday, APX participants electronically send before 10:30 their buy/sell bids for each hour between 00:00 and 24:00 of the next day. This means that wind power producers must base their bids on 14-38 hours ahead wind generation forecasts (if the most recent forecast is provided at 10:00). APX runs the algorithm that matches demand and supply for determining the hourly system marginal price and the program of each participant. This program defines the amount of energy a participant is committed to produce or consume each hour for the following day. Producers are paid by APX the spot price for the quantity of energy specified in the program, independently of their actual production.

Producers supplying power to the Dutch power grid is responsible for the balance of their program (balance between program and actual production). APX spot and TenneT regulation markets are independent. There is no constraint on the sign or the magnitude of imbalance prices. Though, they are mainly penalizing: spill price is on average lower and top-up price on average higher than the spot price. For more details and analyses of TenneT imbalance prices, we refer to the works by Saint-Drenan [198], Chevallier [43] and Boogert and Dupont [19]. In this particular regulation market, there are situations for which imbalance prices are negative: electricity then becomes a waste good. For the example of the TenneT imbalance prices over 2001, the analysis of Saint-Drenan [198] reveals that spill and top-up prices were negative 25% and 10% of the times respectively. This appears surprising, but is indeed easily explainable by the so-called must-run character of some non-flexible generators. For instance combined-cycle installations are basically installed for the generation of heat whereby electricity is a co-product. Reducing the must-run output would be hardly possible from a technical perspective or it would involve high shutdown costs. Therefore, negative prices are acceptable to these power suppliers since the costs of a shutdown period are sometimes much higher [201]. Compared to other countries, cogeneration is a significant part of the Dutch electricity production.

For intermittent generation, by a wind farm for instance, APX spot revenues are globally reduced by the regulation costs due to forecasting errors. Moreover, since imbalance prices are determined for every 15 minutes and dependent on the actual state of balance of the grid, they are very volatile and they can reach very high levels. They appear as hardly predictable. If combined with a large prediction error, high regulation prices expose wind power producers to excessive regulation costs on a short-term basis, even if in the long run low and high-level regulation penalties may balance. Therefore, the participation for wind power producers in this market appears risky and will permit to clearly illustrate the gains of preferring more advanced bidding strategies.

Normally, it is possible to make changes in the programs in the Dutch electricity market. As formulated in the list of assumptions for the present study, we have considered that energy producers make only one bid a day, at 10:30 for the following day. Thus, the most recent prediction of wind generation at that hour is used to define an optimal bidding strategy. Finally, any participant in the APX market has to pay an entrance and annual fees, which may be prohibitive for small capacity generators. These fees are not taken into account in our study.

6.4.2 Results and discussion

In this Paragraph, we first analyze some characteristics of the Dutch electricity market, i.e. of both market clearing prices and regulation unit costs. Then, the participation of the Golagh wind farm in that market is simulated over 2002.

Characteristics of market prices

Table 6.1 summarizes some of the Dutch market characteristics for 2002 and 2003. It gathers the average prices on the day-ahead exchange market and the average regulation unit costs for these two years. The unit costs for positive and negative deviations from contract are not the same: there is a trend that penalties for downward regulation are higher than those for upward regulation. Here, a positive deviation costs in average almost three times more than a negative one over 2002. This ratio is slightly higher than the one observed for 2003 or for instance by Morthorst [164] for the Nord Pool over 2002.

Table 6.1: *The Dutch electricity market characteristics for 2002 and 2003: average spot prices and penalties for upward and downward regulation.*

year	$\bar{\pi}_k^c$ (€/MWh)	$\bar{\pi}_k^{*,-}$ (€/MWh)	$\bar{\pi}_k^{*,+}$ (€/MWh)
2002	29.99	4.03	10.93
2003	46.47	8.93	11.39

Moreover, the average spot price is substantially higher in 2003 than in 2002, while penalties for deviations stay at a similar level. The market is more severe in 2002: the unit costs of deviations from contract are relatively higher. This is the reason why we have chosen to simulate the participation of the Golagh wind farm over 2002. The fact that the ration between the unit regulation costs and the spot price are higher will reveal more easily the benefits of using more advanced bidding strategies.

Then, we can further look at the various prices on that market for 2002, by following their quarterly and monthly averages. Table 6.2 gathers these values. Market clearing prices tend to be higher in the summer period that in winter: monthly average prices vary from 10€/MWh in February to 58.02€/MWh in June. Such a clear trend is not present for penalties for upward and downward regulation.

Table 6.2: *Focus on the Dutch electricity market characteristics for 2002: monthly (m.) and quarterly (q.) average spot prices and penalties for upward and downward regulation. All prices are in €/MWh.*

Month	$\bar{\pi}_k^c$ (m.)	$\bar{\pi}_k^{*,-}$ (m.)	$\bar{\pi}_k^{*,+}$ (m.)	$\bar{\pi}_k^c$ (q.)	$\bar{\pi}_k^{*,-}$ (q.)	$\bar{\pi}_k^{*,+}$ (q.)
1	14.50	-2.10	18.29			
2	10.00	-0.67	17.96	11.65	0.33	16.22
3	10.43	3.77	12.40			
4	17.92	-6.66	18.49			
5	39.21	0.93	9.06	38.38	1.34	11.13
6	58.02	9.74	5.83			
7	48.56	12.97	2.90			
8	41.00	23.06	-4.30	41.17	8.22	8.51
9	33.94	-11.38	26.93			
10	38.25	9.61	6.38			
11	29.09	-4.40	18.92	29.38	6.97	7.61
12	20.81	15.71	-2.47			

Also, the ratio between the two regulation penalties is highly variable from a quarter to the other, and even more from a month to the other. During the first four months of 2002, average downward regulation unit costs are even higher than the average spot prices on the day-ahead pool. In parallel, the penalties for upward dispatch are very low, even negative on average for three out of these four months. The inverse situation, i.e. high upward and negative downward regulation unit costs on average, occurred two months during that year, in August and December. The fact that regulation costs averaged over a month period can be negative shows the recurrent character of negative regulation prices. Table 6.3 gives the percentage of times these two regulation unit costs are negative over 2002 and 2003.

One sees that these percentages are very high and at similar levels for these 2 years: a market participant is rewarded for a negative imbalance almost half of the times, and for a positive imbalance between 20 and 25% of the times. This is contradictory with the principle we formulated such that a participant should never be rewarded for an imbalance, to prevent participants from intending imbalances. However, Boogert and Dupont [19] argued that gaming strategies on the Dutch electricity market cannot be profitable, since it would be necessary to predict at least the sign of the overall system imbalance to apply these strategies. This is hardly feasible. Here, our aim is not to describe gaming strategies for wind power producers, but is instead to show how they can take advantage of their knowledge on the trends for imbalance prices for better bidding in electricity pools.

Table 6.3: *The Dutch electricity market characteristics for 2002 and 2003: percentage of situations with negative regulation unit costs for either upward or downward dispatch.*

year	up-regulation unit cost $\pi_k^{*,-}$ (%)	down-regulation unit cost $\pi_k^{*,+}$ (%)
2002	47.42	19.81
2003	49.11	24.85

Results with point predictions

To carry out this simulation, point predictions provided by the forecasting method M1 are considered. Persistence is the reference model we have chosen as benchmark. The series of M1-forecasts consist of 48 hour-ahead predictions with a time resolution of one hour. Also, predictions are updated every hour. Thus, only the power predictions produced at 10:00 every day of the year are needed for deriving the various bidding policies. They are translated to energy forecasts with the simple formula given by Equation (6.13). Relevant look-ahead times range from 14 to 38-hour ahead.

The PTU length on the day-ahead market APX is of 1 hour, while it shortens to 15 minutes on the regulation market run by TenneT. Because we do not want to dilute the variations of imbalance prices, we use the TenneT PTU as the time unit for this simulation. Therefore, we make the assumption that the quantities of energy proposed on the spot market can be divided in four equal amounts of energy, for the four TenneT PTUs included in an APX trading hour, sold at the same hourly spot price. A similar assumption is made for measured quantities, for comparison between contracted and actual levels of energy.

Note that even if we have described the prices' characteristics of both day-ahead and regulation markets, participation strategies based solely on point predictions do not integrate any expertise on the behavior of the markets. Wind power producers are passive with regard to these markets and their prices are considered as fatal. The only possible enhancement of such strategies would consist in applying correction coefficients to energy forecasts. These correction coefficients would be the result of a rule of thumb defined by some experts[2]. However, such corrections are not the result of rigorous decision-making processes, and consequently we do not envisage this possibility here.

Results from the simulation of the market participation of the operator of the Golagh wind farm are gathered in Table 6.4.

Table 6.4: *Simulation results over 2002 with predictions given by both persistence and the M1 forecasting approach. Results are compared to the case for which perfect predictions would be used for bidding.*

	Pers.	Adv. model	Perfect pred.
Contracted energy (GWh)	44.37	45.49	46.41
Surplus (GWh)	18.12	9.87	0
Shortage (GWh)	16.08	8.95	0
Down-regulation costs (10^3 €)	195.72	119.99	0
Up-regulation costs (10^3 €)	79.59	52.01	0
Total revenue (10^3 €)	1041.38	1145.69	1317.69
Av. down-regulation unit cost (€/MWh)	10.80	12.15	0
Av. up-regulation unit cost (€/MWh)	4.95	5.81	0
Av. regulation unit cost (€/MWh)	8.05	9.13	0
Av. energy price (€/MWh)	22.44	24.68	28.37
Part of imbalance (% of produced energy)	73.69	40.55	0
Performance ratio (γ, %)	79.1	86.99	100

Prediction biases of both the reference and the advanced prediction methods over this one-year evaluation period can be given by comparing the amounts of contracted energy. The bias is low for both of these methods, being equal to 4.4% and 1.98% for persistence and M1 respectively. Note that the mode of normalization is not equivalent to the case for which it is done with the wind farm nominal power (cf. Chapter 3). Here, it is relative to the measured energy. These biases tell that both methods tend to slightly under-predict wind generation. This is confirmed by looking at the amounts of energy in surplus and in shortage over 2002: the amount of energy in surplus is higher than the one in shortage. Therefore, the wind farm operator is more exposed to penalties for downward dispatch.

In the previous Paragraph, where we described the characteristics of both the spot and regulation markets, it was explained that penalties for downward regulation were much higher than the one for upward dispatch over 2002 (the ratio is slightly lower than 3). This

[2]For instance, Barthurst et al. [13] stated that applying a correction coefficient equal to 0.9 to persistence forecasts leads to 'more valuable' predictions ('more valuable' in the sense that they yield higher revenues).

remark, combined to the fact that by using the M1 forecasts the wind farm operator has a higher amount of energy in surplus over this year, makes that the regulation costs associated to the energy in surplus are 2.3 times higher than the ones related to shortage. This remark is also valid for the reference prediction method. Though, the overall regulation costs are substantially reduced thanks to the use of an advanced approach instead of persistence. The performance ratio γ, which gives the wind farm operator income, in comparison with the case for which he had used perfect predictions, equals 79.1% if using persistence forecasts and 86.99% if using M1 forecasts. In other words, regulation costs are diminished by 37.75% in this particular case-study when preferring the advanced prediction method.

In a general manner, it is owing to their higher level of accuracy that forecasting approaches have a greater value than simple reference methods. In Chapter 3, it was shown that whatever the considered prediction approach, the improvement with respect to persistence (which corresponds to an error reduction, cf. Equation (3.9)) reaches high levels for horizons further than 6-10 hours ahead. This improvement usually reaches levels from 40 up to 60%. Therefore, this reduction of the level of prediction error directly translates to a diminution of the amount of energy subject to regulation. Even if the revenue on the day-ahead market is rather similar, the amounts of energy to buy and sell on the regulation market are much lower. Here, while the wind farm operator has a quantity of energy representing 73.6% of the produced energy subject to the regulation mechanism if using persistence as prediction method, this proportion drops to 40.55% if defining bids from M1 forecasts. Hence, preferring an advanced prediction method significantly decreases the financial risks that the market participant may have to face when bidding in an electricity pool.

In Table 6.1 were given the average spot prices and regulation unit costs for 2002 and 2003. The average spot price on APX is 29.99€/MWh for the year 2002. Therefore, if always bidding the same amount of energy on APX every day for every PTU, this would be the average price one would receive over that year. But, due to production fluctuations, a wind farm operator cannot always propose the same quantities of energy, and the average price he would receive from the day-ahead market may be different from the one given by simple statistics. From Table 6.4, one sees that if using perfect predictions, which corresponds to the case for which all generation is sold on the spot market and is not subject to regulation, then the average price per produced MWh is equal to 28.37€/MWh. This is significantly lower than the statistical average. This means there are higher quantities of wind generation that are sold when the market clearing price is low and less when this price is high. When considering persistence or M1 forecasts for bidding on APX, that average price per produced MWh lowers to 22.44 and 24.68€/MWh repectively, due to regulation costs. In addition, one notices that the average cost per MWh that is subject to regulation is not the same in both cases: when using persistence, even if the amount of energy in imbalance is much higher over the year, the average regulation unit cost is significantly lower (8.05 instead of 9.13€/MWh).

Results from the application of different PC strategies

The case of the participation of a wind farm operator in the Dutch electricity market is further investigated here by applying PC strategies. Therefore, this means we consider that the operator's aim is only to maximize his revenue over the considered period. In order to compare the results with the ones given and commented in the above Paragraph, we focus again here on the participation of the Golagh wind farm in this market over the year 2002.

Firstly, for deriving what we referred to as advanced bidding strategies, it is necessary to have probabilistic forecasts of wind generation for the relevant horizons, over the period of interest. Here, we apply the adapted resampling method developed in Chapter 4 for producing predictive distributions of wind generation from M1 point forecasts. In order to define the settings of the uncertainty estimation method, we follow the guidelines given in Chapter 4: the size of error samples is set to 300 items, the number of bootstrap replications to 50. Also, because it was shown that by using more fuzzy sets for mapping the range of possible power values the sharpness and resolution of the resulting predictive distributions are increased, we use here 5 triangular fuzzy sets on the power axis. An offline simulation is carried out over the end of 2001 (2 months of data) for initializing the error samples. Interval forecasts are produced for several nominal coverage rates, from 10 to 90%, with 10% increments. This translates to estimating 18 quantiles of the predictive distributions of wind power for each look-ahead time. An example of such probabilistic forecasts is depicted in Figure 4.4 in Chapter 4.

In a second stage, one needs to model the operator's sensitivity to regulation costs. Since we aim at applying PC strategies for minimizing the regulation costs over 2002, and since we consider that the operator cannot couple the wind generation with conventional generation, only the penalties for upward and downward dispatch are utilized for defining the loss function g. We make the assumption that it is possible to estimate (or forecast) the annual or quarterly trends for regulation unit costs. Therefore, we define two probabilistic choice strategies PC1 and PC2, either based on a single loss function for the whole year (PC1), or on 4 different costs functions for the 4 quarters of 2002 (PC2). In Equation (6.22), estimates of regulation unit costs for upward and downward dispatch are replaced by the annual and quarterly averages presented in Table 6.2. Hence, following results rely on the assumption that we can perfectly predict trends for regulation unit costs. This assumption follows from [198], where Saint Drenan has shown that it is indeed possible to estimate trends for imbalance prices in the Netherlands with climatology-like forecasts. Note that when the loss function g is only a piecewise linear function with different slopes for positive and negative deviations, it is only the ratio between these two slopes that is important. For every day of the year, and for each PTU, bids are defined by solving the optimization problem formulated by Equation (6.24). The results from the application of these strategies are summarized in Table 6.5.

These results can be directly compared to those presented in Table 6.4. One notices that even with bidding strategies defined with annual averages of regulation unit costs, the revenue of the market participant is higher than when using simple strategies based on point forecasts only. While the performance ratio γ equals 86.99% when using M1 forecasts, this

Table 6.5: *Simulation results over 2002 with predictions given by the M1 forecasting approach and two different PC strategies (PC1: annual estimate of average regulation penalties - PC2: quarterly estimates). Results are compared to the case for which perfect predictions would be used for bidding.*

	PC1	PC2	Perfect pred.
Contracted energy (GWh)	57.23	62.37	46.41
Surplus (GWh)	5.19	4.89	0
Shortage (GWh)	16.03	20.85	0
Down-regulation costs (10^3 €)	55.92	42.61	0
Up-regulation costs (10^3 €)	87.15	61.46	0
Total revenue (10^3 €)	1173.62	1212.61	1317.69
Av. down-regulation unit cost (€/MWh)	10.77	8.71	0
Av. up-regulation unit cost (€/MWh)	5.44	2.95	0
Av. regulation unit cost (€/MWh)	6.74	4.04	0
Av. energy price (€/MWh)	25.29	26.13	28.37
Part of imbalance (% of produced energy)	45.72	55.46	0
Performance ratio (γ, %)	89.14	92.1	100

ratio reaches 89.14% with the strategy PC1. From Table 6.2, it was seen that the ratio between penalties for positive and negative deviations was not constant throughout the year, with significant variations from a quarter to the other or even from a month to the other. Therefore, it was expected that by using several loss functions depending on the period of the year, it would be possible to propose more strategic bids on the electricity pool and to further increase the resulting income. This is shown here with the bidding policy PC2, which is based on quarterly estimates of regulation unit costs. The performance ratio reaches 92.1% in this case. This means indeed that the regulation costs over 2002 are reduced by 39% when considering the PC2 strategy instead of the one based on M1 predictions only. Consequently, the more one can integrate information on the regulation unit costs behavior, by increasing the resolution of forecasts of these costs, the more would increase revenue from participation in the electricity market. Though, one must remember that it is particularly difficult to estimate regulation penalties in advance (cf. discussion by Boogert and Dupont [19] or by Skytte [204] for instance). In addition, the proposed strategies follow from the assumption that wind generation has no influence on regulation unit costs in the short-run, and thus that the corresponding random variables can be considered as independent. This may not be true for markets significantly penetrated by wind. Then, it would be necessary to model and integrate this influence in the definition of the loss function g.

The participant's revenue is not increased by reducing the amount of energy in imbalance. It is actually the inverse: while this amount was equal to 40.55% of the contracted energy for the bidding strategy based on M1 forecasts, it rises to 45.72 and 55.46% for the cases of strategies PC1 and PC2. Comparing Tables 6.4 and 6.5 allows one to see that further integrating information on the regulation market's behavior leads to an orientation of the imbalances of the wind farm operator. Since on average the unit costs for downward regulation are higher than the ones for upward regulation, it appears preferable to propose quantities of energy that are more possibly subject to shortage than to surplus regulations. This is preferable for both the wind farm operator and the TSO, since the definition of imbalance penalties also directly reflects the TSO's sensitivity to the system's balance. Therefore, for other electricity markets with different behaviors, or if a given regulation market behavior evolves, it can be accounted for by modifying the loss function g.

Going from bidding strategies based on M1 forecasts only to PC strategies PC1 and PC2, the costs supported by the wind farm operator over 2002 for upward and downward dispatch are completely different. Although we stated that the ratio between these costs was equal to 2.3 for strategies using M1 predictions, here the costs for up-regulation are higher than those for down-regulation, because the quantities of energy in shortage are much higher. Though, the cost of a regulated MWh lowers as we implement more advanced bidding strategies. This is true for cost per regulated MWh related to both up- and down-regulation. On average this cost per MWh subject to regulation is divided by 2 when applying the bidding strategy PC2 (equal to 4.04€/MWh) in comparison with the one based on point forecasts. In this simulation, the average cost of regulation per produced MWh[3] ranges from 2.24€/MWh for PC2 to 5.93€/MWh for the persistence-forecast-based policy. These numbers cannot be generalized for other years on the Dutch market or for other electricity pools, but since all European markets have a similar structure, applying more advanced bidding strategies is also expected to result in a diminution of regulation costs per produced MWh.

6.5 Conclusions

Wind power forecasting has an interest for wind power producers aiming at participating in electricity markets. The aim of this Chapter has been to show that the more we know about what could be the wind generation in the following hours, the more can be increased the income of this wind power producer. The study has been a two-fold study, by considering first the gains from preferring advanced point forecasting approaches and then bidding strategies that integrates uncertainty estimates associated to the point predictions. An important aspect is that point predictions provided by state-of-the-art wind power forecasting methods are not the optimal bids one can propose on electricity pools.

Because wind power forecasts contain a part of uncertainty, there will always be an imbalance cost related to forecasting errors. In this Chapter, we have described a generic method that takes into account the forecast uncertainty and a model of the market participant's sensitivity to regulation costs, for deriving optimal bidding strategies. This method is flexible in the sense that it can be defined following an exchange between the analyst and the market participant, in order to tailor it to his specific needs. Indeed, the loss function, which models the wind farm operator's sensitivity to the regulation costs, can integrate considerations on the day-ahead and regulation markets, but also the possibility to couple wind generation with conventional generators or storage means, for facing imbalances in

[3]The average cost of regulation per produced MWh is given by the difference in the average energy price obtained by applying a given bidding strategy and by using perfect predictions

the most cost-efficient manner. In addition, the bidding strategy definition method can be focused on maximizing the participant's income only, or integrate a risk aversion component, i.e. a preference for avoiding large regulation costs. Such a possibility to tailor bidding policies with respect to end-user needs is not possible if considering point forecasts only.

Results have been presented for the real-world case study of the participation of a wind farm operator in the Dutch electricity pool. Focus has been given to the amount of energy produced in imbalance and to the participant's revenue compared to that he would have obtained if using perfect predictions. It has been shown that using advanced forecasting approaches permits to increase the income of a wind power producer by diminishing the amount of energy subject to regulation, and hence the financial risk he may have to face. Though, if still considering that a wind farm operator aims at maximizing his revenue only, we have demonstrated that applying more advanced bidding strategies derived from the introduced methodology was actually more effective. This is because we integrate the knowledge one has about possible wind generation and the market's behavior in the decision-making process. In the previous Chapters, we have explained how to produce probabilistic forecasts of wind generation, but regarding the electricity market, we have assumed here that it would be possible to estimate trends for regulation unit costs, on a annual or quarterly basis. It would be of particular interest to further investigate on that topic in order to see to what extent it is possible to model and forecast the regulation penalties.

We have only implemented PC strategies that aim at minimizing the expected regulation costs in this Chapter, so that the resulting benefits are better highlighted. Though, we have not discussed the possibility to apply different bidding strategies depending on the expected level of uncertainty. Considering prediction risk indices such as the ones derived in Chapter 5 (and also risk indices related to expected regulation costs) may be a possibility for making a choice between the application of either PC or RA strategies, in order to avoid excessive regulation costs in situations exhibiting low predictability.

For assessing the value of uncertainty estimates associated to forecasts of wind power, we have concentrated here on the trading application. However, such a kind of optimization methodologies in a stochastic programming framework can be extended to other decision-making problems related to the management or trading of wind power, for the participation in several markets with various gate closures (see [138] for instance), for the coupling of wind with hydro or conventional generation (see [36] among others), etc. All management and trading problems involving wind generation prove to be optimization under uncertainty problems.

7

General Conclusions

7.1 Overall conclusions and contribution

The present thesis has proposed a methodology for the uncertainty estimation of wind power forecasts. Going further, we clearly demonstrated the value of that uncertainty information for an optimal integration of wind generation. When we initiated our research works few years ago, the research efforts were mainly focused on the development or improvement of point prediction methods, either of the physical or the statistical type. Very few results were then available concerning the forecasting error characteristics or ways to estimate the uncertainty in an appropriate manner. We explained the drawbacks of these methods being deterministic. Indeed, the sole information given by point predictions is not sufficient for both the optimal management and trading of wind generation. This has been confirmed by the requirements of forecast users, who expressed their need for meaningful information on the uncertainty of wind power predictions. Therefore, the lack of developments for uncertainty estimation motivated our research works.

In order to develop methods appropriate for the wind power application, it has been necessary in a first stage to gain insight on the uncertainty of wind power prediction errors. A first general conclusion from our research works is that characterizing this forecast uncertainty is not a trivial task. For that purpose, we have introduced a framework for the verification of point predictions of wind power output, encompassing a set of statistical measures and diagnostic tools. This framework can be utilized for evaluating point forecasting approaches or alternatively for highlighting the error characteristics. It was clearly explained that the quality assessment of forecasting methods actually follows from underlying assumptions on the forecast user's cost function, i.e. his sensitivity to prediction errors. Therefore, the most relevant criteria may not be obvious. And, considering that a single measure could provide the whole information on a given method skill is simply not possible. In contrast, a complete framework with insight on the meaning of the various criteria allows one to have a critical view on the performance of considered predictors. We recently contributed to the definition of an evaluation protocol consisting in a restricted set of evaluation criteria, published in [145]. For better characterizing the general features of prediction errors, we have considered several state-of-the-art forecasting methods (belonging to the physical and statistical families) that are already in use for predicting wind power production in several European countries. Moreover, a set of case-studies has been selected for this purpose, consisting of wind farms located in various terrain and climatological conditions, the power production of which have been predicted for periods ranging from several months to several years.

With the application of the verification framework we went further in the understanding of forecasting errors, by underlining their sources and highlighting their characteristics. Especially, we commented on the contributions of the meteorological predictions and of the wind-to-power conversion models. A part of prediction errors directly comes from the NWPs. And, the conversion model, because of its non-linear and bounded nature, amplifies or dampens the errors coming from the NWPS, thus acting on the shape of prediction error distributions. We have confirmed that the characteristics of the prediction uncertainty are common for the various point prediction approaches. Consequently, we proposed to develop generic methods that could be used in association to any point prediction approach. In addition, we explained how the evaluation of prediction methods following a Murphy-Winkler framework is indeed beneficial for learning on the influence of some variables on the level of prediction uncertainty. We focused on the effect of the lead time and of the level of predicted power only, but the initiated study could be further extended to the influence of wind direction, air density, etc.

For estimating the uncertainty of wind power point forecasts, we have envisaged in a first stage the possibility to accompany them with prediction intervals. At time t, such intervals give a lower and an upper bound for the power production at lead time $t+k$, with a pre-assigned probability. If estimating several prediction intervals at once, one can then dress point forecasts with predictive distributions of wind power. The original method we developed in the present document focuses on the time-series of prediction errors, which were characterized as nonstationary, nonlinear and bounded time-series. The only requirement for its application is that power measures can be made available online (for calculating prediction errors). This is possible if the considered wind farm (or group of wind farm) is equipped with real-time data acquisition systems. This is indeed the case for almost all new wind farms today. Also, the introduced method is distribution-free: no restrictive assumption is made regarding the shape of error distributions. It is a dynamic method since the estimation of prediction intervals is always based on the most recent information. In order to account for the non-linear and bounded aspects, the core of this method consists in a fuzzy logic inference model that permits to integrate the previously gained expertise on the characteristics of prediction error distributions.

In parallel to these developments, the required properties of interval forecasts have been described, as well as a non-parametric framework for their evaluation. This framework is of general value since it can be utilized for the quality assessment of quantile, interval and probabilistic predictions of wind power output. From the application of the evaluation framework to several case-studies, we showed that the proposed method for interval forecasting is reliable, and we discussed the influence of its degrees of freedom on the reliability, sharpness and resolution aspects. The developed method is generic, and can thus be used for post-processing any state-of-the-art point prediction approach. It is now implemented in a baseline module as part of the ANEMOS prediction platform, as described in Appendix C. An online evaluation phase is under realization in the frame of this project, in order to further confirm its operational value. But, more than the introduced method, our proposal consists in having a probabilistic view of wind power forecasting. Wind generation is then seen as a random variable whose characteristics (e.g. moments, quantiles or complete probability density function) are to be predicted for every look-ahead time. Probabilistic forecasting encompasses the aspects of both point forecasting and uncertainty estimation.

Given the importance of the role of the NWPs in the power prediction uncertainty, we have decided to investigate on methods developed in the meteorological literature, and to consider ensemble predictions for predictability assessment and skill forecasting. Wind power ensemble forecasts were obtained by converting either ECMWF or NCEP meteorological ensembles. Also, as a benchmark (and gratis) alternative for making up ensembles of wind power, we applied a poor man's temporal approach, by superposing forecasts for the same lead time, but issued at different time origins. We proposed a definition of prediction risk indices based on the dispersion of the ensemble members (for a given look-ahead time or over a set of consecutive prediction horizons). Prediction risk indices are a simple and comprehensive signal that may be more easily apprehended by forecast users than probabilistic predictions. We showed how these prediction risk indices can be related to the level of prediction error in a probabilistic manner. Then, we explained how they may be utilized for forecasting the expected level of imbalance over a look-ahead period. This consists an innovative contribution in the area of the uncertainty estimation in wind power forecasting. Such information can be utilized by end-users for making a choice among more or less conservative decisions depending on the risk aversion of the forecast users. For the example of the considered test case, the absolute error was on average 5 times higher when situations were classified (by the mean of the prediction risk index) as highly uncertain in comparison with the most easily predictable cases. In addition, we showed that even if advanced ensemble prediction systems have a higher ability for resolving between situations with low and high uncertainty, the poor man's alternative is already valuable. Existing wind power forecasting tools can be easily upgraded for providing such ensembles.

Finally, the benefits of having a probabilistic view of wind power forecasting have been demonstrated for the specific case of the trading application. Even if we focused on a particular electricity pool (the Dutch one), a generic formulation of the revenue of a wind farm operator participating in European electricity markets has been introduced. Different bidding strategies have been described, based either on the use of the sole point predictions, or derived from probabilistic predictions of wind power. It has been explained that the latter bidding strategies are to be preferred since they can be tailored to the sensitivity of the market participant to regulation costs. This cannot be the case if relying on point forecasts only. On the case study of the participation of a wind farm in the Dutch electricity pool with a revenue-maximization strategy, we have clearly demonstrated that applying more advanced bidding strategies based on wind power probabilistic predictions allows one to substantially lower the regulation costs induced by the intermittent nature of wind generation. One of the main conclusions from our research works is that since the wind power forecasting accuracy cannot reach an acceptable level in a near future (acceptable in the sense that the forecast users would not be sensitive to the cost of prediction errors), then an uncertainty estimation appears as important as the forecast itself. Further developments in that direction are a necessary step for going towards an optimal integration of wind generation.

7.2 Perspectives

A first perspective obviously concerns the application of the proposed methods for uncertainty estimation in an operational context. The implementation of the interval forecasting approach in the ANEMOS prediction platform and its evaluation on various test-cases is of particular importance for verifying if a level of performance similar to that we witnessed from our offline evaluation can be reached in an online operation. Moreover, robustness issues may step in, as the quality of estimated intervals directly relates to the quality of collected power measures. This will be accounted for when evaluating the operational value of the developed approach. In parallel, it would be beneficial to start communicating prediction risk indices in combination to point predictions of wind power, so that forecast users have a simple and comprehensive signal on the confidence they may have in the information given by point forecasts.

The contribution of the meteorological prediction errors to the wind power forecasting uncertainty has been underlined. Some previous investigations on this contribution already led to the conclusion that work should be done for better understanding wind characteristics and consequently produce more accurate predictions of both wind speed and direction (see [129] for instance). Research efforts in that area will certainly be motivated by potential commercial applications.

However, it would be rather easy to transfer responsibility of improving the quality of wind power predictions to meteorological research centers only. Further developments for better refining the wind field at the local scale and for better modeling the energy conversion process will obviously be beneficial. In addition, considering the possibility of combining forecasts provided by different methods (and eventually with different NWPs as input) is expected to enhance the quality of wind power point forecasts. However, as explained in the previous Section, improving the accuracy of point predictions is not the only path for easing the integration of wind generation in power systems. Providing a more complete description of expected wind generation by producing probabilistic forecasts goes in that direction. In the present work, predictive distributions of wind power are produced by dressing point forecasts produced from a state-of-the-art method with conditional prediction error distributions, given the forecast conditions. As explained, the quality of these predictive distributions is highly influenced by the quality of the point prediction method considered as input. More precisely, the sharpness of predictive distributions is directly related to the level of accuracy of the point prediction method. Therefore, it would be of particular interest to evaluate the alternative that consists in directly estimating predictive distributions of wind power.

A very positive aspect is that there exist various possibilities for producing probabilistic predictions of wind power, and that the expected competition between research teams developing on these rival approaches will be highly stimulating. First, a wealth of statistical methods may be directly applied to our problem, for providing probabilistic forecasts in the form of quantile, interval or density predictions. These methods may be parametric or not, based on structural models or black-box approaches. Alternatively, methods following from ensemble forecasting of meteorological variables and their conversion to wind power will benefit from the experience of meteorological centers. Though, the wind-to-power transformation, as well as the ensembles' recalibration through appropriate statistical methods, will comprise subtle aspects. We have carried out a first comparison on the case-study of an offshore wind farm (over a 10-month period) between a statistical and an ensemble-based method for probabilistic forecasting of wind power [186], in terms of their reliability, sharpness and resolution. The nice properties of both methods have been revealed, but it has not been possible so far to conclude on the superiority of a given approach.

Then, whatever the chosen alternative, a complete framework for the evaluation of probabilistic predictions of wind power has to be proposed in a near future. Existing frameworks such as the one described by the econometrical forecasting community are not directly applicable in our case. This is indeed why we proposed an alternative framework in the present thesis. Such a framework is appropriate for evaluating quantile and interval forecasts. Though, we have seen for the case of the verification of wind power point predictions that alternative frameworks may allow one to derive new conclusions on the properties and bottlenecks of forecasting methods. Similarly, alternative frameworks for the verification of probabilistic predictions of wind power are expected to provide new insight on the characteristics of the rival proposed approaches. Therefore, the wind power prediction community must realize the importance of contributing to the definition of verification frameworks, since they are the basis for justifying new developments and quantifying their related benefits.

Finally, a close collaboration between wind power forecasters and forecast consumers is necessary in the future. Benefits from the use of predictions can only be optimal if they are specially tailored to the end-user requirements. Focus should then be given to decision-making processes. There are multiple types of decisions that may be envisaged from the use of wind power predictions — we only considered the participation in electricity pools in the present work. And, each of these kinds of decision asks for different and specific decision-making methodologies. Having a more global view of the wind power management problem will be beneficial for both forecasters and forecast users.

A

List of Publications

Refereed Publications

P. Pinson, G. Kariniotakis. "On-line assessment of prediction risk for wind power production forecasts". *Wind Energy*, vol. 7, no. 2, pp. 119-132, April-June 2004.

H. Madsen, P. Pinson, T.S. Nielsen, H.Aa. Nielsen, G. Kariniotakis. "Standardizing the performance evaluation of short-term wind power prediction models". *Wind Engineering*, vol. 29, no. 6, pp. 475-489, December 2005.

Papers in progress or already submitted to Journals

P. Pinson, C. Chevallier, G. Kariniotakis. "Trading wind generation with short-term probabilistic forecasts of wind power". submitted to *IEEE Trans. on Power Systems*.

P. Pinson, H.Aa. Nielsen, J.K. Møller, H. Madsen, G. Kariniotakis. "Probabilistic forecasts of wind power: required properties and evaluation". submitted to *Wind Energy*.

P. Pinson, G. Kariniotakis. "Conditional prediction intervals of wind power from an expert model". to be submitted to *Int. Journal of Forecasting*.

P. Pinson, G. Kariniotakis. "Skill forecasting from ensemble predictions of wind generation". to be submitted to *IEEE Trans. on Power Systems*.

Refereed papers in Conference Proceedings

P. Pinson, J. Juban, G. Kariniotakis. "On the quality and value of probabilistic forecasts of wind generation". In *Proc. of the 2006 PMAPS Conference, IEEE Conference, 'Probabilistic Methods Applied to Power Systems'*, Stockholm, Sweden, June 2006. Invited paper.

P. Pinson, H.Aa. Nielsen, T.S. Nielsen, H. Madsen, G. Kariniotakis. "Properties of interval and quantile forecasts of wind generation and their evaluation". In *Proc. of the European Wind Energy Conference 2006, Scientific Track*, Athens, Greece, February 2006.

G. Kariniotakis, P. Pinson. "Uncertainty of short-term wind power forecasts - A methodology for on-line assessment". In *Proc. of the 2004 PMAPS Conference, IEEE Conference, 'Probabilistic Methods Applied to Power Systems'*, pp. 729-736, Ames, Iowa (USA), September 2004. Invited paper.

P. Pinson, G. Kariniotakis, D. Mayer. "Uncertainty and prediction risk assessment of short-term wind power forecasts". In *Proc. of the 2004 EAWE Conference, 'The science of making torque from wind'*, Delft, The Netherlands, April 2004.

P. Pinson, G. Kariniotakis. "Wind power forecasting using fuzzy-neural networks enhanced with on-line prediction risk assessment". In *Proc. of 2003 Power Tech Conference, IEEE Conference*, vol. 2, Bologna, Italy, June 2003.

Selection of Conference Proceedings

G. Kariniotakis, I. Marti, P. Pinson, et al. "What performance can be expected by wind power prediction models depending on site characteristics?". In *Proc. of the European Wind Energy Conference 2004*, London, United Kingdom, November 2004.

P. Pinson, C. Chevallier, G. Kariniotakis. "Optimizing benefits from wind power participation in electricity markets using advanced tools for wind power forecasting and uncertainty assessment". In *Proc. of the European Wind Energy Conference 2004*, London, UK, November 2004.

G. Kariniotakis, P. Pinson, N. Siebert, G. Giebel, R. Barthelmie. "State of the art in short-term prediction of wind power - From an offshore perspective". In *Proc. of the 2004 Ocean Energy Conference: Offshore Wind Energy, Marine Currents and Waves*, Brest, France, October 2004.

H. Madsen, H.Aa. Nielsen, T.S. Nielsen, G. Kariniotakis, P. Pinson. "A protocol for standardizing the performance evaluation of short-term wind power prediction models". In *Proc. of the 2004 Global Windpower Conference*, Chicago, Illinois (USA), March 2004.

P. Pinson, G. Kariniotakis. "On-line adaptation of confidence intervals based on weather stability for wind power forecasting". In *Proc. of the 2004 Global Windpower Conference*, Chicago, Illinois (USA), March 2004.

P. Pinson, T. Ranchin, G. Kariniotakis. "Short-term wind power prediction for offshore wind farms - Evaluation of fuzzy-neural network based models". In *Proc. of the 2004 Global Windpower Conference*. Chicago, Illinois (USA), March 2004.

P. Pinson, G. Kariniotakis. "On-line assessment of prediction risk for wind power production forecasts". In *Proc. of the European Wind Energy Conference 2003*, Madrid, Spain, June 2003.

P. Pinson, N. Siebert, G. Kariniotakis. "Forecasting of regional wind generation by a dynamic fuzzy-neural networks based upscaling approach". In *Proc. of the European Wind Energy Conference 2003*. Madrid, Spain, June 2003.

G. Kariniotakis, P. Pinson. "Evaluation of the MORE-CARE wind power prediction platform. Performance of the fuzzy logic based models". In *Proc. of the European Wind Energy Conference 2003*. Madrid, Spain, June 2003.

B

Résumé en français

B.1 Introduction

Contexte et objectifs

L'énergie éolienne connait un développement considérable en Europe, avec une capacité installée de plus de 40 GW au début de l'année 2006. A l'horizon 2010, les prévisions de l'EWEA (association Européenne pour l'énergie éolienne) attaignent 70-75 GW. La contribution de la France (qui a un des meilleurs potentiels éoliens d'Europe) à ces objectifs, devraient être de 10 GW.

Etant donnée la nature intermittente de la production de puissance électrique d'origine éolienne, les prédictions éoliennes[1] pour des horizons allant jusqu'à 48-72 heures apparaissent comme une information essentielle pour faciliter son intégration sur le réseau électrique et dans les marchés de l'électricité. Communément, ces prédictions sont fournies sous la forme de prédictions 'point', qui consistent en une valeur unique par horizon qui correspond à la production la plus probable. Les modèles de prédiction éolienne pour ces horizons incluent deux étapes: dans un premier temps la prédiction de différentes variables météorologiques au niveau du site considéré, et dans un deuxième temps la modélisation de la conversion (pour ce site) du vent en puissance électrique. Et l'erreur de prévision météorologique, et l'erreur de modélisation du processus de conversion contribuent au

[1]Dans le reste de ce document, nous utiliserons le terme de 'prédiction éolienne', qui fera référence à la prédiction de la puissance disponible au niveau du site considéré. Aussi, si ces prédictions sont des prédictions d'énergie produite pour une certaine période, ceci sera clairement mentionné.

manque de précision des prédictions éoliennes. Ce niveau de précision est très variable, et est fonction de différents facteurs tels que l'horizon de prédiction, les conditions météorologiques, etc.

Les erreurs de prédiction ont un coût significatif pour leurs utilisateurs, qu'ils soient opérateurs du réseau électrique, producteurs d'électricité, ou participants aux marchés de l'électricité. En marge d'autres travaux qui consistent à chercher à améliorer la précision des prédictions éoliennes, notre proposition ici est de développer des méthodes permettant d'estimer l'incertitude spécifique à chacune de ces prédictions. Aussi, nous introduisons deux alternatives pour communiquer cette information sur l'incertitude des prédictions à leurs utilisateurs. Ces deux alternatives sont:

- des intervalles de prédiction, qui donnent pour chaque horizon une plage de production possible pour un niveau de probabilité donné,

- des indices de risque, qui consistent en une simple valeur numérique informant l'utilisateur sur la confiance qu'il peut avoir dans la prédiction point qui lui a été fournie.

Finalement, un des objectifs de ce travail est aussi de montrer comment intégrer cette information sur l'incertitude des prédictions éoliennes dans des processus de prise de décision, ainsi que de mettre en avant les bénéfices résultant de l'utilisation de cette information. Dans ce but, nous nous concentrons sur le cas de la participation d'un producteur éolien à un marché de l'électricité.

Organisation

Afin de mieux décrire les motivations et l'origine de nos travaux de recherche, le Chapitre 2 présente un état de l'art de la prédiction éolienne. Les méthodes actuelles sont détaillées. Nous insistons sur l'intérêt des prédictions éoliennes pour les acteurs du secteur énergétique, mais nous expliquons aussi que ces acteurs ont eux-même exprimé le besoin d'avoir une information sur l'incertitude spécifique à chaque prédiction. Il n'y avait que très peu de développements dans ce sens au moment où nous avons initié ces travaux de recherche.

Les différentes propositions pour l'estimation de l'incertitude de prédiction découle de l'idée que l'erreur de prédiction est inévitable. Nous nous attachons dans le Chapitre 3 à caractériser cette erreur. Quelques définitions sont données. Puis nous décrivons un protocole d'évaluation permettant d'apprécier la qualité des méthodes de prédiction point existantes, et permettant aussi de montrer l'influence de certains facteurs (e.g. niveau de puissance prédit) sur les distributions d'erreurs. Ce protocole est alors utilisé pour un certain nombre de cas d'étude. Plusieurs fermes éoliennes sont considérés, situées dans différents pays Européens. Pour chacune de ces fermes éoliennes, plusieurs méthodes de prédiction sont employées, afin de montrer quelles sont leurs caractéristiques communes. Aussi, nous commentons la contribution des prévisions météorologiques et de la modélisation de la conversion vent-puissance au niveau d'incertitude. Dans une dernière partie, le rôle de l'horizon et du niveau de puissance prédit sont mis en valeur.

Notre première proposition pour communiquer l'incertitude spécifique à chaque prédiction éolienne est de leur associer des intervalles de prédiction. Pour cela, nous introduisons dans le Chapitre 4 une méthode non-paramétrique, et expliquons comment utiliser des concepts de logique floue afin d'intégrer l'expertise d'un analyste pour améliorer la qualité de ces intervalles. En parallèle, les différentes propriétés souhaitées pour ce type de prédictions sont décrites, et nous expliquons comment les évaluer. Nous utilisons alors ces différents critères pour déterminer la qualité de la méthode introduite pour plusieurs cas d'étude, ainsi que pour montrer l'influence des paramètres de la méthode sur cette qualité.

Dans le Chapitre 5, nous proposons une autre possibilité pour estimer et communiquer l'incertitude des prédictions: il s'agit d'indices de risque qui informent sur la confiance à apporter aux prédictions point fournies. Pour produire ces indices de risque, nous considérons l'utilisation de prévisions ensemblistes. Tout d'abord, plusieurs méthodes pour obtenir des prédictions ensemblistes éoliennes sont décrites. Puis, nous évaluons l'apport de ce type de prédiction, que ce soit pour améliorer la qualité des prédictions point, ou pour estimer le niveau d'incertitude. Une définition d'indices de risque basée sur la dispersion des ensembles pour un ou plusieurs horizons consécutifs est introduite. Dans une dernière partie, une relation probabiliste entre l'indice de risque et le niveau d'erreur est établie, et nous comparons la capacité des différentes prédictions ensemblistes éoliennes à informer sur le niveau d'incertitude attendue.

Afin d'évaluer la valeur de la prédiction éolienne pour ses utilisateurs, nous nous concentrons dans le Chapitre 6 sur la participation d'un opérateur de ferme éolienne au marché de l'électricité néerlandais. En plus d'évaluer la valeur de la prédiction éolienne, notre but est surtout de montrer quelle est la valeur ajoutée apportée par une estimation de l'incertitude. Dans une première partie de ce Chapitre, nous décrivons les mécanismes de marché et formulons les hypothèses de notre étude. Ensuite, nous décrivons plusieurs types de stratégies de participation au marché, basées sur les seules prédictions point, ou utilisant des prédictions probabilistes et un modèle de la sensibilité de l'opérateur aux pénalités résultant des déviations par rapport au programme. Ces stratégies sont alors comparées en évaluant les revenus de l'opérateur d'une ferme éolienne participant au marché néerlandais sur l'année 2002. Les bénéfices résultant de l'utilisation de prédictions probabilistes sont clairement démontrés.

Pour finir, le Chapitre 7 couvre les conclusions générales de la thèse, et propose des perspectives concernant de futurs travaux de recherche.

B.2 Etat de l'art de la prédiction éolienne

Ce Chapitre propose une vision globale de l'état de l'art concernant la prédiction éolienne. Une premiere partie se concentre sur la description des motivations pour la prédiction de la production éolienne à des horizons de quelques minutes à quelques jours, ainsi que les différents aspects mis en jeu, que ce soient les aspects météorologiques ou statistiques. Ensuite, la deuxième partie de ce Chapitre permet de formuler de façon mathématique le problème de prédiction en lui-même. On s'attache finalement à décrire les différentes méthodes existantes aujourd'hui. L'accent est mis sur le fait qu'une large majorité de ces méthodes ont été développées dans le but de fournir des prédictions point, et que la question de leur incertitude n'était pas traitée au moment où ont été initiés les travaux de thèse.

Le problème de la prédiction éolienne

La puissance électrique disponible au niveau d'une ferme éolienne est fortement variable, étant donné que cette puissance est une fonction directe de la vitesse du vent. Cette fonction, communément appelée courbe de puissance, est non-linéaire et bornée. Aussi, pour des vitesses de vent supérieures à la vitesse de décrochage (environ 25-30 $m.s^{-1}$), les turbines éoliennes sont arrêtées, et il n'y donc plus de puissance électrique disponible. D'autres variables météorologiques peuvent influencer la courbe de puissance d'une ferme éolienne, telle que la direction du vent, la température ou le taux d'humidité par exemple. D'une façon générale, les variations de la vitesse du vent sont amplifiées, ou atténuées, par la courbe de puissance.

Cette variabilité de la puissance électrique pose des difficultés pour l'intégration de l'énergie éolienne sur le réseau électrique. En effet, cette puissance n'est pas 'controlable' comme peut l'être celle de générateurs électriques conventionnels. Ces difficultés sont de nature différente en fonction des échelles de temps considérées. Ici, on s'intéresse aux problèmes de la production et du commerce de l'électricité, et aussi à celui de la gestion quotidienne du réseau électrique. Les échelles de temps correspondantes vont de quelques minutes à plusieurs jours. Pour les producteurs d'électricité, il est nécessaire de planifier de façon économique l'utilisation de leurs différents moyens de production, voire de définir leurs stratégies de participation au marché de l'électricité. Et, étant donnée que la puissance électrique disponible doit en permanence répondre à la demande, le gestionnaire du réseau doit prévoir des réserves suffisantes qui pemettent un ajustement en temps réel de l'équilibre du réseau. Dans tous les cas, des prédictions de la puissance éolienne disponible sont nécessaires. Des travaux récents, concernant le dimensionnement des réserves ou la gestion de systèmes électriques ayant une part significative d'énergies éolienne et hydraulique, ont montré qu'en plus de ces prédictions, une information sur leur incertitude est nécessaire pour optimiser les méthodes de gestion employées. Des conclusions similaires sont disponibles concernant les stratégies de participation au marché de l'électricité.

Le problème de la prédiction éolienne pour des horizons allant jusqu'à plusieurs jours regroupe des aspects métérologiques et mathématiques. Il est en effet nécessaire de prédire l'évolution des différentes variables météorologiques, qui vont avoir une influence sur la production de puissance électrique, et en parallèle de modéliser la courbe de puissance, qui permet de donner la puissance électrique disponible en fonction des valeurs des différentes variables météorologiques. Concernant l'aspect météorologique, les prévisions des différentes variables sont fournies par des modèles météorologiques. Il peut être nécessaire d'extrapoler les valeurs fournies par ces modèles au niveau du site considéré. Ensuite, le modèle mathématique de la courbe de puissance doit prendre en compte une multitude de facteurs. Les effets d'ombrage à l'intérieur d'une ferme éolienne peuvent avoir une influence significative sur sa production électrique, et ceux-ci doivent être intégrés dans le modèle. Aussi, cette courbe de puissance peut évoluer dans le temps, dû au vieillissement des turbines ou aux modifications saisonnières de l'environnement de la ferme éolienne par exemple.

Les méthodes de prédiction

Les méthodes de prédiction éolienne peuvent être classées en deux catégories qui correspondent à une approche physique d'une part, et à une approche statistique d'autre part. A celles-ci s'ajoutent les méthodes dites de référence, qui sont des méthodes simples permettant de fournir des prédictions éoliennes à un faible coût, et qui servent alors de référence pour évaluer les bénéfices provenant de l'utilisation de méthodes plus avancées. Les deux méthodes de référence les plus communes sont la persistence, qui consiste à dire que la production éolienne dans le futur sera égale à la dernière valeur de puissance mesurée, et la 'climatologie', pour laquelle toutes les prévisions sont égales à la moyenne de toutes les mesures de puissance disponible au niveau du site considéré.

L'approche dite physique utilise des lois physiques pour raffiner les prévisions météorologiques au niveau du site, ainsi qu'un modèle explicite de la courbe de puissance. Certaines méthodes se basent sur des lois physiques simplifiées, provenant d'une hypothèse logarithmique sur le profil de vent par exemple, alors que d'autres nécessitent l'utilisation de codes de calcul en mécanique des fluides, qui peuvent permettre d'avoir des prédictions plus précises (notamment pour des terrains à topologie complexe), mais qui s'avèrent être très couteux. Afin de corriger la part d'erreur systématique qui peut être présente en appliquant ces lois physiques, les prédictions sont communément corrigées avec des modèles régressifs linéaires.

L'approche dite statistique se base sur des modèles purement mathematiques qui décrivent l'évolution des séries temporelles de production éolienne à partir de valeurs historiques de ces séries temporelles (ce qui permet de capturer le caractère persistent de la production éolienne), ainsi que des valeurs historiques ou prédites de variables explicatives. Les paramétres de ces modèles sont estimés en définissant un problème d'optimisation sur un jeu de données historiques. Ces paramètres peuvent évoluer lors d'une utilisation en conditions opérationnelles, ce qui permet de refléter l'aspect non-stationnaire de la production éolienne. Une multitude de modèles peuvent être considérés, prenant plus ou moins en compte l'expertise physique du modélisateur. Certains modèles de type 'boîte noire', tels que les réseaux de neurones, ne requièrent aucune connaissance physique du problème, sauf pour le choix des variables explicatives à utiliser. A l'inverse, la famille des modèles structurels nécessite une description complète de la relation entre les différentes variables explicatives et la production éolienne.

B.3 Caractérisation de l'incertitude de prédiction

Toute prédiction contient obligatoirement une part d'erreur. En fonction de l'application considérée, la sensibilité de l'utilisateur de prédictions pour ces erreurs ne sera pas la même. Tout d'abord, il est nécessaire de définir ce qui fait qu'une prédiction peut être jugée comme étant 'bonne' ou 'mauvaise'. Il existe deux façons de qualifier les prédictions, soit en considérant leur performance d'un point de vue statistique (on parle alors de leur *qualité*), soit en déterminant leur *valeur*, i.e. les bénéfices (économiques ou autres) pour leurs utilisateurs. Ce Chapitre traite uniquement de la qualité des prédictions éoliennes, alors que la question de leur valeur est étudiée dans le Chapitre 6. Une méthodologie pour évaluer cette qualité est proposée dans une première partie du Chapitre, comprenant un ensemble de mesures et une approche basée sur l'étude des distributions conditionnelles d'erreurs de prédiction. Cette méthodologie est alors appliquée aux cas d'études de plusieurs fermes éoliennes, dont la production est prédite à partir de différentes méthodes. Les deux buts de cette étude sont de définir le niveau de performance des méthodes de prédiction éolienne de l'état de l'art, ainsi que de caractériser l'incertitude de ces prédictions, afin de développer par la suite des méthodes appropriées pour son estimation.

Méthodologie pour l'évaluation de la qualité des prédictions éoliennes

L'erreur de prédiction est définie comme étant la différence entre la valeur mesurée p_{t+k} au temps $t+k$ et la valeur $\hat{p}_{t+k/t}$ prédite au temps t pour le temps $t+k$. Cette erreur est communément normalisée par la puissance nominale P_n du site considéré.

Un point important de l'évaluation des prédictions éoliennes concerne la définition du cas d'étude et du cadre de travail. La qualité des mesures de production d'une ferme éolienne doit être vérifiée. Ensuite, les différentes méthodes de prédiction que l'on veut comparer doivent être appliquées de la même façon, de manière à ce que la comparaison soit juste. Par exemple, les jeux de données sont séparés en deux parties, la premiére pour estimer les paramètes des différents modèles considérés, et la seconde pour l'évaluation en elle-même, et donc pour laquelle on simule une utilisation en temps réel et en conditions opérationnelles de ces méthodes.

L'évaluation de la qualité des prédictions éoliennes ne peut se résumer à l'utilisation d'une seule mesure statistique : l'utilisation de plusieurs critères permet de tirer des conclusions de nature complémentaire. La première approche pour l'évaluation des prédictions éoliennes que nous décrivons dans ce Chapitre est basée sur une série de mesures statistiques. Chacune de ces mesures donne une information différente sur les caractéristiques des erreurs: le biais correspond à la part d'erreur systématique, le critère NMAE indique l'erreur moyenne en valeur absolue sur la période considérée, alors que le critère NRMSE correspond à l'erreur quadratique moyenne sur cette même période, etc. Ces différentes mesures peuvent alors être utilisées pour comparer les performances d'une méthode considérée par rapport à celles de méthodes de référence, la persistence par exemple. Aussi, elles peuvent être calculées en fonction de différents facteurs (période de l'année, direction du vent, etc.) afin de déterminer l'influence de ces facteurs sur la performance des méthodes de prédiction.

La deuxième approche que nous décrivons se base sur les distributions d'erreurs de prédictions, et sur l'idée que ce sont les distributions conjointes des mesures et des prédictions qui contiennent toute l'information sur l'erreur de prédiction. Une première (et très utile) information qui peut être dérivée à partir de ces distributions est la fréquence d'occurence d'erreurs supérieures ou inférieures à certains seuils, pour quantifier le risque d'erreurs extrêmes par exemple. Mais, au-delà de cette simple information, on peut aussi étudier l'influence de certains paramètres, tel que le niveau de puissance prédit, sur ces distributions d'erreurs. Pour cela, les distributions sont résumées par leur 4 premiers moments, qui donnent des indices complémentaires sur la forme des distributions: la moyenne correspond au 'centre de gravité', l'écart-type à la dispersion des erreurs, le moment d'ordre 3 à l'assymétrie et le moment d'ordre 4 au ratio entre la finesse de la partie centrale des distributions par rapport à l'épaisseur de leurs pattes.

Méthodes de prédiction et cas d'étude

Afin de donner une portée générale à cette étude, nous avons choisi de considérer 5 différentes méthodes de prédiction éolienne, que l'on appellera M1, M2,. . ., M5. Ces méthodes sont parmi celles de l'état de l'art les plus utilisées en pratique en Europe. M1, M2 et M3 appartiennent à la famille des méthodes statistiques, tandis que M4 et M5 sont de type physique. Des prédictions produites à partir de la persistence et de la climatologie servent de référence.

Notre étude se concentre sur quatre fermes éoliennes situées dans différents pays européens, avec des climatologies de vent différentes, ainsi que des conditions topographiques différentes (mer, plaine, collines, zone montagneuse). Les jeux de données couvrent des périodes de plusieurs mois à plusieurs années. Pour chacune de ces fermes sont disponibles des prédictions météorologiques et des mesures de puissance avec une résolution horaire.

Qualité des méthodes de l'état de l'art

Le biais des méthodes de prédiction doit être faible, sinon cela signifie qu'elles commettent une erreur systématique. Les méthodes de type statistiques sont globalement non-biaisée, de par la nature des approches utilisées pour l'estimation de leur paramètres. Quand aux méthodes physiques, les prédictions qu'elles produisent sont post-traitées avec des méthodes statistiques afin de corriger une part linéaire de leur erreur de prédiction. Notre étude révèle que toutes les méthodes considérées ont un général un biais faible, avec un léger avantage pour les méthodes de type statistique.

Les critères NMAE et NRMSE englobent et la part systématique et la part aléatoire de l'erreur de prédiction. Leur valeur augmente avec l'horizon de prédiction. Même si la différence entre les niveaux de NMAE (et NRMSE) pour les méthodes de prédiction considérées parait faible, cette différence peut avoir un coût significatif pour l'utilisateur de ces prédictions. Les méthodes de type statistiques ont un niveau d'erreur plus faible pour les horizons compris entre 1 et 6-8 heures, car elles utilisent les mesures de production en entrée, et permettent donc de modéliser le caractère persistent de la production éolienne. Pour des horizons plus lointains, rien ne permet de donner un avantage clair à tel ou tel type de méthode. Aussi, la topographie a une influence sur le niveau de performance des méthodes de prédiction, et surtout sur la variabilité de ces niveaux de performance. Pour des terrains complexes, le niveau d'erreur est globalement plus élevé et la différence entre les niveaux d'erreur plus importante. Si on compare les méthodes de l'état de l'art avec des méthodes de référence telles que la persistence ou la climatologie, la réduction de l'erreur de prédiction est très importante, ce qui justifie le choix de méthodes plus avancées.

Caractéristiques de l'incertitude de prédiction

Et les prévisions météorologiques et la modélisation de la conversion du vent en puissance contribuent à l'incertitude des prédictions éoliennes. Une part de l'erreur (erreur de phase) provient directement des prévisions météorologiques, et ne peut être corrigée par la suite par les modéles de conversion vent-puissance. Par contre, ces deux étapes contribuent à l'erreur d'amplitude. Notamment, de par la nature non-linéaire du procédé de conversion du vent en puissance électrique, sa contribution consiste plus particulièrement à amplifier, ou à atténuer, les erreurs provenant des prévisions météorologiques.

L'application de la méthodologie basée sur l'étude des distributions conjointes des mesures et des prédictions permet de mieux caractériser l'incertitude de prédiction. Ici, nous ne nous intéressons qu'à l'effet de l'horizon et du niveau de puissance prédit, mais cette étude pourrait être étendue pour étudier l'influence d'autres facteurs e.g. la direction du vent, sur l'incertitude des prédictions. Un point important est que les distributions conditionelles d'erreurs, en fonction du niveau de puissance prédit ou en fonction de l'horizon, ne sont pas Gaussiennes. Même si les erreurs de prédiction de la vitesse du vent pouvaient être caractérisées comme Gaussiennes, la nature non-linéaire et bornée du processus de conversion vent-puissance change complètement les caractéristiques de ces distributions. Cet effet se ressent sur la dispersion des erreurs en fonction du niveau de puissance prédit, mais aussi et surtout a un impact direct sur l'assymétrie des distributions d'erreur. L'horizon de prédiction a aussi une influence sur la dispersion des erreurs, celle-ci augmentant quand l'horizon est plus lointain. Tous ces aspects seront pris en compte par la suite pour le développement de méthodes d'estimation de l'incertitude spécifique à chaque prédiction éolienne.

B.4 Estimation et évaluation d'intervalles de prédiction

Dans ce Chapitre, notre proposition pour communiquer l'incertitude spécifique à chaque prédiction point éolienne est de leur associer des intervalles de prédiction. Un intervalle de prédiction correspond à une plage de valeurs possibles pour une probabilité donnée. Une méthode pour leur estimation est développée dans une premiere partie, avec pour but d'être applicable quelle que soit la méthode utilisée en amont pour produire les prédictions point. C'est pour cela que les seules hypothèses qui sont formulées découlent des conclusions générales données dans le Chapitre précédent, qui concernent à la fois les approches physiques et statistiques. La seule condition pour pouvoir appliquer cette méthode est que des mesures de puissance soient disponibles au niveau du site considéré, ce qui est le cas pour quasiment toutes les nouvelles fermes éoliennes mises en opération aujourd'hui.

Afin de commenter la qualité de la méthode proposée, on introduit dans une deuxième partie le cadre de travail pour l'évaluation d'intervalles de prédiction. Ce cadre de travail peut être utilisé plus généralement pour évaluer des prédictions probabilistes qui seraient données par plusieurs quantiles de la densité de probabilité de la variable considérée. Ce cadre de travail nous permet aussi de commenter l'influence des paramètres de la méthode sur ses performances.

Développement de la méthode d'estimation d'intervalles de prédiction

Un intervalle de prédiction est caractérisé par ses deux bornes, qui correspondent à des quantiles de la densité de probabilité de la variable considérée. Nous nous intéressons ici à l'estimation d'intervalles de prédiction centrés: il y a la même probabilité qu'une mesure non couverte par un intervalle de prédiction donné soit supérieure ou inférieure aux bornes de cet intervalle.

La méthode développée est non-paramétrique: on ne fait pas d'hypothèse sur la forme des distributions de probabilité des erreurs de prédiction. Aussi, cette méthode se base sur une approche empirique: ce sont les valeurs d'erreurs récemment commises par la méthode de prédiction point considérée qui sont utilisées pour estimer les intervalles de prédiction. De ce fait, on engloble les souces d'incertitude pouvant venir du choix du modèle, de l'estimation de ses paramètres, etc. De plus, en utilisant les erreurs récentes, on prend en compte l'aspect non-stationnaire de l'incertitude. Une hypothèse est donc que l'incertitude associée aux prédictions actuelles peut être déterminée à partir de la performance récente de la méthode de prédiction point. En parallèle, à l'inverse des méthodes de régression permettant de prédire un unique quantile, la méthode introduite peut être utilisée pour estimer plusieurs quantiles à la fois. Elle peut donc servir à construire d'une façon peu coûteuse des prédictions probabilistes sous forme de densités de probabilité.

La base de la méthode consiste à collecter pour chaque horizon les erreurs récentes de la méthode de prédiction point considérée, d'utiliser ces échantillons d'erreurs pour construire la distribution de probabilité cumulée de production éolienne, et de se servir de cette fonction pour déterminer les quantiles d'intérêt. Ensuite, afin de produire des intervalles de prédiction qui soient plus spécifiques aux conditions de prédiction (au niveau de puissance prédit par la méthode de prédiction point par exemple), nous proposons un modèle utilisant des concepts de logique floue. Ce modèle permet de formuler des densités de probabilité conditionnelles aux conditions de prédiction. Ces densités de probabilité conditionnelles prennent la forme d'une combinaison de distributions de probabilité, les poids étant déterminés par le degré d'appartenance à des ensembles flous de conditions de prédiction. Deux approches sont alors décrites pour la combinaison de distributions de probabilité: une approche classique dite 'linear opinion pool', et une approche originale dite 'adapted resampling', qui se base sur un réchantillonage croisé des échantillons d'erreurs disponibles. Les performances de la méthode d'estimation d'intervalles de prédiction utilisant ces deux approches sont discutées dans la suite du Chapitre.

Pour son application au probléme de la prédiction éolienne, on ne considère par la suite que l'effet que peut avoir le niveau de puissance prédit (par la méthode de prédiction point) sur l'incertitude des prédictions. Toutefois, la méthode peut être étendue dans le futur pour prendre en compte les effets d'autres variables, comme la vitesse ou la direction du vent prédits par les modèles de prévisions météorologiques par exemple.

Cadre de travail pour l'évaluation de prédictions probabilistes

Evaluer la qualité de prédictions probabilistes s'avère être encore plus subtile que celle des prédictions point. En effet, si une prédiction point peut être jugée peu fiable au regard de la différence importante en cette prédiction et la mesure associée, il n'est possible de juger de la qualité d'une unique prédiction probabiliste. La qualité de ce type de prédictions ne peut être appréciée que si étudiée pour un de jeu de données d'une taille significative.

Une nécessité pour des prédictions probabilistes est que les probabilitées nominales soit respectées. Cette propriété des prédictions probabilistes est leur 'fiabilité'. Par exemple, un intervalle de prédiction avec un niveau de confiance nominal de 80% doit couvrir les mesures pour 80% des cas. De plus, comme les intervalles de prédiction considérés sont des intervalles de prédiction centrés, il est important de vérifier que les proportions nominales des deux quantiles définissant les bornes d'un intervalle sont respectées. D'une façon générale, si les prédictions probabilistes sont non-paramétriques et que les prédictions de densités de probabilité correspondent à des séries de quantiles avec des proportions nominales différentes, l'évaluation de la fiabilité de ces prédictions s'effectue en vérifiant individuellement la fiabilité de chaque quantile. En pratique, une mesure de fiabilité est donnée par un calcul comptable de la couverture des quantiles. Pour des prédictions comprenant des séries de quantiles, l'évaluation globale de tous les quantiles peut être résumée par des diagrammes de fiabilité qui donne les probabilités observées en fonction des probabilités nominales.

En parallèle, une propriété désirée est que les prédictions probabilistes donnent une estimation de l'incertitude spécifique aux conditions de prédiction. Les intervalles de prédiction doivent être plus ou moins larges en fonction d'incertitude plus ou moins forte. Cette abilité à différencier les situations ayant différents niveaux d'incertitude est appelée 'finesse' ou 'résolution'. Cette propriété correspond à la qualité inhérente d'une méthode de prédiction probabiliste. Si la fiabilité d'une méthode peut être améliorée par des traitements statistiques, sa finesse et sa résolution sont invariantes. Ici, nous proposons d'apprécier ces propriétés en étudiant la taille des intervalles de prédiction. Plus particulièrement, la taille moyenne des intervalles donne une information sur leur finesse, et l'écart-type de la taille des intervalles sur leur résolution. Intuitivement, on souhaite que la taille moyenne des intervalles soit minimale, et que cette taille soit variable.

Etant donnée que la fiabilité des prédictions est une nécéssité, alors que leur finesse et résolution représentent la qualité inhérente de la méthode utilisée, notre proposition, est de vérifier la fiabilité des prédictions probabilistes dans un premier temps, et si cette fiabilité est jugée acceptable, de passer ensuite à l'étude des autres propriétés. Nous introduisons aussi un score approprié pour l'évaluation de la qualité globale d'une méthode de prédiction probabiliste, donc incluant les trois propriétés énoncées. Il alors possible de résumer l'évaluation de la finesse et de la résolution d'une méthode en utilisant ce score, à partir du moment où sa fiabilité est avérée.

Résultats et discussion

Afin d'apprécier la qualité de la méthode proposée et de discuter l'influence des paramètres de la méthode sur cette qualité, on considère deux cas d'études qui consistent en un suffisament grand nombre de séries de prédictions. Aussi, on utilise en entrée de la méthode des prédictions point produites avec trois différentes méthodes déjà considérées dans le Chapitre précédent. Il est choisi d'estimer des séries d'intervalles de prédiction avec des niveaux de confiance nominaux de 10%, 20%, ..., 90%, afin de pouvoir reconstruire des densités de probabilité.

Tou d'abord, on compare l'utilisation des approches 'linear opinion pool' et 'adapted resampling' pour l'estimation de distributions de probabilité conditionelles dans la méthode proposée. Pour cette étude, la taille des échantillons d'erreurs servant de base pour l'estimation des intervalles est fixée à 300 éléments, le nombre d'ensembles flous couvrant la plage de valeurs possibles pour les prédictions éoliennes à 5. Finalement, on considère 50 rééchantillonage pour le cas spécifique de l'approche 'adapted resampling'. On observe que les prédictions probabilistes obtenues avec les deux approches sont fiables, avec un faible avantage pour l'approche 'adapted resampling'. Si on utilise alors le score introduit pour l'appréciation de la qualité globale de la méthode, on note là aussi que quelque soient les prédictions point utilisées en entrée, les prédiction probabilistes obtenues avec l'approche 'adapted resamling' ont une qualité plus élevée. Et, sachant que l'estimation des intervalles de prédiction est basée sur les performance récentes de la méthode de prédiction point considérée, un résultat est que la qualité de cette méthode de prédiction point influe fortement la qualité des prédictions probabilistes associées.

Ensuite, on se concentre plus particulièrement sur le cas de la méthode proposée utilisant l'approche 'adapted resampling' et on étudie l'influence de ses paramétres sur la qualité des prédictions probabilistes. On constate que en faisant varier le nombre d'ensembles flous sur la plage des valeurs possibles pour les prédictions point, on augmente par la même occasion la résolution de la méthode sans en affecter la fiabilité. En parallèle, augmenter la taille des échantillons d'erreurs permet d'élever le niveau de fiabilité des prédictions probabilistes, tout en les rendant plus fines. Finalement, on note que l'augmentation du nombre de rééchantillonage utilisé dans l'approche 'adapted resampling' améoliore et la fiabilité et la qualité globale des prédictions probabilistes. Toutefois, ces différents paramètres doivent tout des même être choisis avec attention. Premièrement, le partitionnage de l'ensemble des conditions de prédiction ne peut se faire qu'à partir de l'expertise du prévisionniste concernant les caractéristiques de l'incertitude. Ensuite, il n'est peut être pas toujours possible d'augmenter la taille des échantillons d'erreurs (en fonction de l'application considérée), et avoir des échantillons de taille infinis effacerait de toutes façons l'aspect nonstationnaire de l'estimation de l'incertitude. Enfin, avoir un nombre de rééchantillonage élevé peut entraîner des coûts de calculs importants, ce qui est à prendre en compte pour des applications en temps réel.

La méthode développée pour l'estimation d'intervalles de prédiction a été implémenté sous la forme d'un module opérationnel intégré dans la plateforme de prédiction ANEMOS. Des détails concernant cette implémentation sont donnés en annexe.

B.5 Prédiction ensemblistes et indices de risque

Dans ce Chapitre est introduite la possibilité d'estimer et de communiquer l'incertitude des prédictions éoliennes sous la forme d'indices de risque. Alors que les intervalles de prédiction discutés dans le Chapitre précédent donnent pour chaque horizon une plage de valeurs possibles avec une probabilité donnée, un indice de risque a pour vocation de donner une information sur la confiance à apporter à une prédiction point. Cet indice de risque peut alors permettre de faire un choix entre un panel de décisions plus ou moins conservatrices, e.g. pour la définition du niveau de réserves nécessaires pour faire face à un possible aléa provoqué par une chute de la production éolienne. Ces indices de risque peuvent paraître plus facilement compréhensibles pour les utilisateurs de prédictions que les prédictions probabilistes.

Une part significative de l'incertitude de prédiction provient directement des prévisions météorologiques utilisées en entrée des méthodes de prédiction éolienne. Les variables météorologiques peuvent être plus ou moins facilement prévisibles, et cette incertitude se répercute directement dans les prédictions éoliennes. Les prévisions ensemblistes sont de plus en plus utilisées par les météorologues afin d'estimer le niveau d'incertitude associée à leurs prévisions à court et à moyen terme (jusqu'à 10 jours). Nous transposons ici ce concept pour le cas de la prédiction éolienne en utilisant des prédictions ensemblistes éoliennes. Les méthodes employées pour obtenir ces prédictions ensemblistes sont décrites dans une première partie. Ensuite, nous nous attachons à montrer les apports de ce type de prédictions par rapport aux plus 'classiques' prédictions point. Ils incluent la possibilité de fournir une prédiction point plus précise, et aussi la possibilité de quantifier le niveau d'incertitude associé aux prédictions point. Enfin, nous définissons un indice de risque, et montrons sa capacité à différencier des prédictions ayant des niveaux d'incertitude plus ou moins élevés.

Les prédictions ensemblistes éoliennes

Les prédictions ensemblistes éoliennes consistent en un ensemble de scénarios alternatifs pour une période considérée. Elles sont obtenues ici en convertissant des prévisions ensemblistes de vent (à 10 mètres) fournies par ECMWF ou NCEP à l'aide d'une méthode statistique, ou en appliquant une méthode de décalage temporel aux prédictions point disponibles.

Les prévisions ensemblistes d'ECMWF sont constituées de 51 membres qui incluent les prévisions de contrôle, i.e. celles habituellement fournies par leur modèle opérationnel, ainsi que 50 prévisions alternatives obtenues par la méthode des vecteurs singuliers. Celles provenant de NCEP, qui sont composées de 11 membres, prévision de contrôle incluse, sont le résultat de l'application de la méthode dite de 'breeding'. La méthode statistique utilisée pour leur conversion en prédictions éoliennes inclue une transformation des données afin que les prédictions éoliennes obtenues puissent prendre des valeurs sur toute la plage des valeurs possibles, i.e. entre zero et la puissance nominale de la ferme éolienne considérée. Ces deux types de prédictions ensemblistes ont des horizons maximaux de respectivement

144 et 84 heures. Mais, nous nous limitons ici aux premières 72 heures, qui sont les horizons qui nous intéressent pour la problématique éolienne.

Etant donné que l'horizon maximum des prédictions point est plus long que le temps séparant la fourniture de deux séries de prédictions successives, il est toujours possible de superposer plusieurs prédictions pour un même moment. Ces prédictions sont produites à partir du même modèle, mais à partir de différentes conditions initiales. Ce principe de superposition pour obtenir des prédictions ensemblistes est appelé méthode de décalage temporel. Ici, les prévisions de contrôle de ECMWF ayant un horizon maximum de 144 heures et étant fournies toutes les 24 heures, il y a toujours 5 prédictions éoliennes alternatives produites à partir de ces prévisions météorologiques pour les 72 heures suivantes.

Le cas d'étude que nous considérons est celui d'une ferme éolienne d'une puissance nominale de 5MW, située en mer proche des côtes du Danemark, pour laquelle des prédictions ensemblistes éoliennes sont produites sur une période couvrant les 10 premiers mois de 2003.

Apport des prédictions ensemblistes

Deux possibilitées peuvent être envisagées quant à l'utilisation de prédictions ensemblistes éoliennes. La première est de combiner l'information fournie par cet ensemble de scénarios alternatifs, afin de dériver une prédiction point qui aurait une meilleure performance que la prédiction point traditionnellement produite à partir de la prévision de contrôle.

Si on considère le cas des prédictions ensemblistes obtenues à partir des prévisions d'ECMWF ou de NCEP, il n'y a pas de raison apparente de donner plus de poids à certains membres de l'ensemble. Donc, nous appliquons un schéma de combinaison qui consiste à prendre la moyenne de tous les membres de l'ensemble pour chaque horizon. Par contre, pour le cas des ensembles obtenus par décalage temporel, il paraît raisonnable d'envisager que la qualité des prédictions puissent être fonction de leur âge. Dans ce cas, le schéma de combinaison consiste à calculer une moyenne pondérée des différents membres, les poids étant choisis en fonction de l'âge des prédictions.

Les résultats de cet exercice montrent que pour les trois types d'ensembles, il est en effet possible de déterminer une prédiction de meilleure qualité en combinant les membres de l'ensemble. Si on se concentre sur la diminution de l'erreur quadratique (donnée par le critère NRMSE), cette réduction atteint, pour le cas d'étude considéré, 5-7% à un horizon de 48 heures, et jusqu'à 25-29% à un de horizon de 7 jours.

Le deuxième apport des prédictions ensemblistes est qu'elles permettent de quantifier le niveau d'incertitude à associer aux prédictions point fournies. Ici, les prédictions point considérées sont les prédictions points qui peuvent être obtenues pour chaque type d'ensemble avec le schéma de combinaison décrit précédemment. Ensuite, nous étudions la relation entre la dispersion des membres d'un ensemble donné et la variation de l'erreur de prédiction associée. En effet, c'est une plus forte variation de l'erreur de prédiction qui définit une situation incertaine, sachant que l'erreur moyenne devrait toujours être nulle. La dispersion d'un ensemble est quantifiée par l'écart-type de ses membres — écart-type pondéré pour le cas des ensembles obtenus par décalage temporel, et la variation de l'erreur par son écart-type. La relation entre dispersion des ensembles et variation de l'erreur de prédiction est étudiée en regroupant les horizons sur une base journalière, par exemple l'horizon 'jour 1' est défini en regroupant tous les horizons entre 0 et 24 heures.

Cette étude révèle que les trois types d'ensembles permettent de dissocier des situations plus ou moins incertaines. Cette propriété des ensembles est communément définie comme étant leur résolution. On observe que la relation entre dispersion des ensembles et variabilité de l'erreur présente des caractéristiques différentes en fonction du type d'ensemble considéré. Les prédictions ensemblistes obtenues à partir des prévisions d'ECMWF paraissent avoir une meilleure résolution.

Indices de risque et estimation du niveau d'incertitude

Etant donnée la relation observée entre la dispersion des prédictions ensemblistes éoliennes et le niveau d'incertitude, cette dispersion sert de base à la définition d'un indice de risque, que l'on dénomme NPRI. Cette indice peut être défini pour chaque horizon de prédiction (NPRI_h). Mais, puisqu'il est peu probable que l'incertitude de prédiction change radicalement d'un horizon au suivant, il est aussi envisagé que l'indice de risque soit défini pour un ensemble d'horizons consecutifs, en utilisant la dispersion moyenne des membres de l'ensemble sur la période considérée. Ici, cet indice est calculé pour 96 horizons consécutifs, c'est à dire pour des périodes de 24 heures (NPRI_d, pour les périodes 'jour 1', 'jour 2' et 'jour 3').

Afin d'évaluer la capacité du NPRI à informer sur le niveau d'erreur attendu, nous utilisons une approche probabiliste. En effet, chercher à déterminer une relation linéaire entre NPRI et niveau d'erreur, avec un coefficient de corrélation donnant la force de cette relation, ne paraît pas approprié. Quantifier le risque revient à donner des probabilités d'erreurs plus ou moins grandes. Une faible valeur d'un indice de risque n'assure pas que l'erreur soit faible, et inversement, une forte valeur de l'indice n'implique pas obligatoirement que le niveau d'erreur soit élevé. L'erreur est calculée comme étant la différence en valeur absolue entre mesure et prédiction si on considére un seul horizon, et comme étant la différence moyenne entre ces deux quantités pour le cas des périodes de 24 heures. Ces erreurs sont alors normalisées par l'erreur moyenne sur les 9 mois constituant la période d'évaluation. Nous utilisons ces erreurs normalisées par l'erreur 'climatologique' car un indice de risque ne donne pas une information sur la valeur de l'erreur, mais plutôt sur le fait que l'on peut s'attendre à une erreur plus faible ou plus élevée que d'habitude. L'approche adoptée consiste alors à trier les valeurs de NPRI en classes de même taille (cinq classes, plus précisément), en fonction de leur amplitude, et à étudier les distributions d'erreurs associées. La première classe contient un cinquième des valeurs, qui sont les valeurs les plus faibles du jeu de données, la deuxième classe un cinquième des valeurs suivantes, etc. La raison pour cela est que ce n'est pas la valeur de l'indice en elle-même qui informe sur l'incertitude, mais plutôt le placement de cette valeur dans la distribution des valeurs possibles pour l'indice de risque.

Dans un premier temps, on évalue la capacité de l'indice NPRI_h à informer sur le niveau d'incertitude, et ce pour les trois types de prédictions ensemblistes éoliennes. On observe que l'erreur moyenne augmente significativement au fur et à mesure que l'indice est dans une classe plus élevée. Pour le cas des prédictions ensemblistes produites à partir des prévisions d'ECMWF, le ratio entre erreurs moyennes pour la plus faible et plus forte classe de NPRI_h est de 5.4. De plus, en regardant l'évolution des quantiles des distributions d'erreurs pour chaque classe de NPRI_h, on note que ces distributions s'élargissent au fur et à mesure que le NPRI_h augmente, montrant que le risque d'erreurs élevées augmente lui-aussi. Si on compare les trois types de prédictions ensemblistes, il apparaît que les indices de risque calculés à partir des prévisions d'ECMWF et de NCEP permettent de mieux différencier les situations plus ou moins incertaines que l'indice NPRI_h calculé à partir des ensembles obtenus par une approche de décalage temporel.

Dans un deuxième temps, on effectue la même étude avec l'indice NPRI_d, et on observe aussi sa capacité à différencier des situations plus ou moins incertaines. L'intérêt de cet indice est qu'il informe sur l'incertitude attendue pour une période (de 24 heures) dans le futur, qui correspond alors à une quantité d'énergie en surplus ou en déficit. Cette information n'est pas donnée par les prédictions probabilistes telles qu'elles sont définies dans le Chapitre précédent ou plus généralement dans la littérature. De plus, de par l'effet de moyennage temporelle, les distributions d'erreurs pour chaque classe de NPRI_d sont plus fines que celles liées aux classes de NPRI_h. Cet élément signifie que le NPRI_d a une meilleure résolution que le NPRI_h. Enfin, comme pour le NPRI_h, on voit que se baser sur les prévisions ensemblistes fournies par ECMWF ou NCEP, plutôt que sur les prédictions ensemblistes produites par décalage temporel, permet une meilleure différenciation des situations plus ou moins incertaines.

B.6 Valeur des prédictions éoliennes et de l'information sur leur incertitude

Avec la libéralisation des marchés de l'électricité, les producteurs d'énergie éolienne ont la possibilité de vendre leur production sur des bourses de l'électricité. Cependant, étant donné que le niveau de prédictabilité de cette ressource n'est pas parfait, ces producteurs sont nécessairement sujet à des pénalités pour non-respect de leur programme de production.

Ce Chapitre aborde deux problémes. Tout d'abord, nous expliquons que la participation d'un producteur éolien à une bourse de l'électricité est un problème de prise de décision. Plus précisément, ce producteur va vouloir déterminer, pour des conditions données, quel est le programme optimal qu'il va pouvoir proposer, afin de maximiser ses bénéfices. Le processus de prise de décision diffère en fonction de l'information disponible concernant la production probable pour les heures ou jours suivants. Ici, nous considérons l'utilisation de prédictions point ou probabilistes. Cela nous amène au deuxième problème abordé dans ce Chapitre, qui est celui de la valeur des prédictions éoliennes. Notre but est de montrer que des prédictions éoliennes probabilistes, si utilisées dans un processus de prise de décision adapté, ont plus de valeur pour leurs utilisateurs que des prédictions point.

Description du problème

Une bourse de l'électricité est composée par plusieurs mécanismes de marché, avec des horizons et des résolutions temporelles différentes. Les deux mécanismes qui nous intéressent ici sont le marché journalier, sur lequel les producteurs éoliens proposent des quantités d'électricité généralement la veille pour le lendemain, et le marché de régulation, qui est lié à la correction en temps réel des déviations des participants par rapport à leur programme. Le revenu d'un producteur éolien est donc composé par les gains résultant de la vente d'électricité sur le marché journalier auquels sont déduits les coûts engendrés par la correction de ses déviations par rapport au programme. Le revenu maximum que pourrait obtenir un producteur éolien est celui qu'il engendrerait en utilisant des prédictions parfaites. Ce revenu maximum correspond à un revenue qui ne serait donc pas entâché par des pénalités.

Nous avons formulé un jeu d'hypothèses pour cette étude, qui sont pour certaines nécessaires pour le développement des méthodes de prise de décisions, ou juste simplificatrices pour les autres. Tout d'abord, on considère que les volumes d'électricité proposés par un producteur éolien seul ne peuvent avoir d'influence sur les prix du marché journalier ou du marché de régulation : ce sont donc des variables indépendantes. Aussi, on fait l'hypothèse que le producteur éolien n'a pas de contrôle sur sa production. Il n'a pas la possibilité de stocker de l'électricité, ou de coupler sa production éolienne avec des moyens de productions conventionnels. Nous expliquons dans ce Chapitre comment cette hypothèse simplificatrice pourra être retirée dans de futurs développements de la méthode.

Stratégies de participation dans une bourse de l'électricité

La façon la plus simple d'utiliser des prédictions pour participer au marché de l'électricité est de considérer que le niveau de contrat optimal pour un horizon donné correspond à la prédiction point pour cet horizon. En utilisant cette stratégie, les pénalités sur le marché de régulation sont directement liées aux erreurs de prédiction de la méthode utilisée.

Toutefois, si le producteur éolien utilise plus d'information concernant le comportement du marché et l'incertitude des prédictions éoliennes, il peut alors développer des stratégies de participation plus avancées. Idéalement, l'information concernant l'incertitude de prédiction est représentée par une prédiction probabiliste de production éolienne, i.e. donnant la densité de probabilité complète de cette variable aléatoire. Ensuite, il est nécessaire de modéliser la sensibilité du producteur éolien aux coûts de régulation. Ce modèle prend la forme d'une fonction, qui renvoit pour toute déviation par rapport au contrat une seule valeur numérique, valeur qui reflète l'insatisfaction du producteur éolien liée au coût engendré par cette déviation. Ces deux informations peuvent alors être intégrées dans un problème d'optimisation afin de déterminer la niveau de contrat optimum qui permet soit de maximiser les revenus sur le long terme, soit de minimiser le risque de fortes pénalités sur le court-terme. Pour le premier cas, cela revient à minimiser l'espérance des pénalités pour chaque horizon considéré, alors que pour le deuxième, on cherche à minimiser la pénalité maximale potentielle pour chacun de ces horizons. Un point important est que les prédictions point, qui consistent en des estimations de l'espérance de la production éolienne pour chaque horizon, ne donnent pas forcément le meilleur niveau de contrat à proposer sur le marché.

Résultat et discussion

Afin d'évaluer la valeur des prédictions éoliennes, et afin de démontrer les bénéfices provenant de la méthodologie introduite, nous nous intéressons à la participation du propriétaire d'une ferme éolienne de 15MW au marché de l'électricité néerlandais, composé du marché journalier APX, et du marché de régulation de TenneT, le gestionnaire du réseau néerlandais, sur l'année 2002. Le but de ce producteur éolien est de maximiser ses bénéfices. Nous considérons la possibilité de participer au marché avec des prédictions point fournies par la persistence ou une méthode avancée, ou enfin en appliquant deux stratégies de maximisation des revenues basées sur des prédictions probabilistes. Ces deux stratégies utilisent des modèles de fonction coût définis soit pour toute l'année, ou pour chaque trimestre. On fait ici l'hypothèse que l'on peut parfaitement estimer les coûts de régulation moyens pour ces deux résolutions temporelles. Les prédictions probabilistes sont produites avec la méthode développée au Chapitre 4, en estimant 9 intervalles de prédiction avec des niveaux de confiance de 10, 20, ..., et 90%.

En comparant les deux méthodes de prédiction point, on voit que la méthode avancée a plus de valeur. En effet, étant donné que le niveau de précision est bien plus élevé, les volumes d'électricité qui sont soumis à des pénalités sont beaucoup moins importants (passant de 73.7% à 40.6% du volume total d'électricité produite), et cela se traduit finalement par des coûts de régulation sur toute l'année qui diminue de 38%. Ces coûts de régulation sont encore fortement réduits en appliquant des stratégies avancées basées sur les prédictions probabilistes (jusqu'à 37% de réduction supplémentaire pour ce cas). Pourtant, ils ne sont pas réduits en diminuant les volumes d'électricité soumis au marché de régulation, mais en orientant ces volumes. En effet, si les pénalités pour surproduction sont plus fortes que celles pour sous-production, il apparait préférable de proposer un niveau de contrat pour lequel on sera plus probablement sujet à des pénalités pour sous-production. Aussi, on se rend compte qu'utiliser plusieurs fonctions coût au cours de l'année permet d'augmenter les revenus car on modélise plus finement le comportement du marché de régulation. D'un point de vue général, les prédictions probabilistes éoliennes ont plus de valeur que les prédictions point pour le problème de la participation au marché de l'électricité. Pour ce cas d'étude, le prix moyen obtenu par MWh produit sur l'année 2002 et vendu sur le marché néerlandais passe de 22.44€ en utilisant les prédictions fournies par la persistence à 26.13€ en utilisant les stratégies plus avancées utilisant les prédictions probabilistes. Pour comparaison, s'il était possible d'avoir des prédictions parfaites, ce prix moyen par MWh serait de 28.37€.

B.7 Conclusions

Conclusions générales et contribution à la prédiction éolienne

Dans cette thèse, nous avons introduit des méthodes pour estimer l'incertitude spécifique à chaque prédiction éolienne. Aussi, nous avons clairement démontré quels pouvaient être les bénéfices résultant de l'utilisation de cette information pour un utilisateur de prédictions. Quand nous avons initié ces travaux de thèse, la plupart des équipes de recherches se concentraient seulement sur l'amélioration de la précision des prédictions point, et très peu de travaux se penchaient sur le problème de l'estimation de l'incertitude de ces prédictions. Nous avons clairement expliqué dans une première partie de la thèse pourquoi cette information sur l'incertitude est primordiale, que ce soit pour un opérateur de ferme éolienne, le gestionnaire du réseau, etc. Au cours de ces dernières années, n'étant pas satisfait du niveau de précision des prédictions éoliennes, ces acteurs ont d'ailleurs eux-même exprimé la nécessité d'avoir une information sur l'incertitude spécifique à chaque prédiction. Les travaux décrits dans la thèse permettent d'avoir une meilleure compréhension de l'erreur de prédiction, et proposent des méthodes originales (et qui peuvent être utilisées dans un cadre opérationel) pour l'estimation de l'incertitude.

Dans l'objectif de mieux comprendre l'incertitude des prédictions éoliennes, nous avons introduit un protocole qui permet d'évaluer la précision des méthodes de prédiction point, mais aussi de mieux caractériser l'influence de certains facteurs sur l'erreur de prédiction. Les cas d'étude de plusieurs fermes éoliennes, dont la production a été prédite à l'aide de cinq méthodes différentes, ont été considérés afin de pouvoir tirer des conclusions générales et non pas liées à une méthode en particulier. Nous avons étudié dans la thèse l'influence de l'horizon et du niveau de puissance prédit uniquement, mais la méthodologie décrite peut être utilisée dans le futur pour révéler l'influence d'autres facteurs comme la vitesse ou la direction du vent prédits par exemple. Les séries temporelles d'erreurs de prédiction ont été caractérisées comme étant non-stationnaires, non-linéaires et bornées. Un point important est qu'il n'apparait pas raisonnable de formuler une hypothèse Gaussienne à propos des distributions d'erreurs. Ceci est une conséquence directe du caractère non-linéaire et borné du procédé de conversion du vent en puissance électrique.

Pour estimer l'incertitude des prédictions éoliennes, nous avons considéré dans un premier temps la possibilité de leur associer des intervalles de prédiction, qui donnent une plage de production possible, pour une probabilité donnée. En estimant plusieurs intervalles avec des niveaux de probabilité répartis sur l'intervalle $[0, 1]$, on peut alors estimer pour chaque horizon la densité de probabilité de production éolienne. La méthode que nous avons proposée dans la thèse pour estimer les intervalles se base sur les erreurs récentes d'un modèle considéré, et est non-paramétrique: nous ne faisons pas d'hypothèses sur la forme des distributions d'erreurs. La seule condition pour l'application de cette méthode est que ces erreurs soient disponibles, et cela est possible si le site considéré est équipé d'un système de mesure de puissance (ce qui est le cas pour la plupart des nouvelles fermes éoliennes aujourd'hui). Afin de prendre en compte le fait que l'incertitude est conditionnelle, nous avons utilisé des concepts de logique floue, et cela a permis d'intégrer l'expertise développée concernant les caractéristiques des erreurs de prédiction. Plus que la méthode décrite, notre proposition dans cette thèse consiste à avoir une vision probabiliste de la prédiction éolienne, qui couvre donc les aspects de prédiction point et d'estimation de l'incertitude.

En parallèle, nous avons exposé les propriétés requises et voulues pour des intervalles de prédiction (ou des prédictions probabilistes en général) et expliqué comment les évaluer. Afin de valider la méthode proposée, nous avons considéré son application à plusieurs cas d'études, pour lesquelles nous avons construit des prédictions éoliennes probabilististes. Nous avons montré l'influence de la qualité du modèle de prédiction point considéré sur celle des prédictions probabilististes associées. Aussi, nous avons discuté l'influence des différents degrés de liberté de la méthode proposée sur les résultats obtenus.

Etant donné le role primordial des prévisions météorologiques dans l'incertitude des prédictions éoliennes, nous avons choisi de considérer un outil de plus en plus utilisé par les météorologues: les prédictions ensemblistes. A partir de ce prédictions ensemblstes éoliennes, notre but a été de définir des indices de risque, qui puissent donner de façon simple une information sur la confiance à apporter à des prédictions point fournies. Nous avons expliqué comment obtenir des prédictions éoliennes ensemblistes à partir d'ensembles provenant de ECMWF et NCEP, ou plus simplement en appliquant une approche de décalage temporel. Une définition d'indices de risque a été proposée, reflétant la dispersion des ensembles, et donnant une information sur la prédicatibilité de la production pour les prochaines heures ou prochains jours. Ces indices de risque donnent une information qui peut être plus facilement appréhendée par des utilisateurs de prédictions que des prédictions probabilistes. En nous focalisant sur le cas d'étude d'une ferme éolienne située en mer au Danemark, nous avons montré en nous basant sur une approche probabiliste comment ces indices de risque peuvent informer sur l'incertitude des prédictions éoliennes. L'utilisation dans ce but des différents types d'ensembles a été comparée. Une conclusion est que même si l'utilisation de 'vrais' ensembles provenant de centres météorologiques permettent de mieux différencier les situations ayant des niveaux d'incertitude plus ou moins élevés, l'approche de décalage temporel permet déjà de fournir une information valable. De plus, cette approche peut directement être appliquée aux modèles de prédiction point existants, ce qui lui donne un avantage pratique.

Dans une dernière partie de la thèse, nous nous sommes concentrés sur la question de la valeur des prédictions éoliennes pour leurs utilisateurs, et plus particulièrement sur la valeur ajoutée qu'apporte une information sur l'incertitude. Pour cela, nous avons considéré le cas de la participation d'un opérateur de ferme éolienne au marché de l'électricité néerlandais, et comparé ses bénéfices sur une année complète (l'année 2002) en utilisant différentes stratégies de participation, utilisant ou non l'information sur l'incertitude (sous forme de prédictions probabilistes). La façon dont ces stratégies peuvent être déterminées a été détaillée. Un point important est que ces stratégies peuvent être adaptées aux besoins spécifiques de leur utilisateur en utilisant des prédictions probabilistes, alors que ce ne peut être le cas si seules des prédictions point sont disponibles. En prenant l'exemple d'une stratégie de maximisation des revenues sur l'année considérée, les bénéfices supplémentaires obtenues avec les stratégies utilisant des prédictions probabilistes sont significatifs.

La conclusion principale de ces travaux de recherche est que, étant donné que les prédictions éoliennes ne peuvent atteindre un niveau de précision suffisant pour leurs utilisateurs dans un futur proche (et voire un futur plus lointain), une information sur l'incertitude des prédictions est primordiale pour faciliter l'intégration de l'énergie éolienne sur le réseau et dans les mécanismes de marché.

Perspectives

Une première perspective concerne évidemment l'application des méthodes proposées dans un cadre opérationnel. La méthode pour l'estimation d'intervalles de prédiction a été intégrée dans la plateforme de prédiction ANEMOS, et sera évaluée sur plusieurs cas d'études. Aussi, il serait intéressant de commencer à communiquer des indices de risque associés aux prédictions, pour que les utilisateurs aient un signal facilement compréhensible sur l'incertitude attendue dans les heures ou jours suivants.

L'amélioration de la qualité des prévisions météorologiques est bien sûr une perspective importante, et il sera aussi important de mieux raffiner ces prévisions au niveau des sites considérés, notamment pour les terrains complexes. Pourtant, comme nous l'avons expliqué, améliorer la précision des prédictions point n'est pas la seule piste pour faciliter l'intégration de l'énergie éolienne sur le réseau ou dans les marchés. Fournir une meilleure description de la production attendue avec des prédictions probabilistes par exemple va dans ce sens.

La qualité des prédictions probabilistes obtenues par la méthode proposée est fortement influencée par celle des prédictions point utilisées en entrée. Il serait donc intéressant de proposer des méthodes qui fournissent directement des prédictions probabilistes et de voir si cela améoliore les résultats obtenus. Ces méthodes pourront être basées sur la conversion d'ensembles météorologiques en puissance éolienne ou sur de nouvelles approches purement statistiques.

Dans tous les cas, il sera aussi nécessaire d'enrichir les méthodes d'évaluation des prédictions probabilistes éoliennes. Les critères et outils proposées dans la thèse permettent déjà de tirer des conclusions intéressantes, mais par expérience, nous avons vu que l'utilisation d'un grand nombre de critères alternatifs permet de tirer des conclusions nouvelles quant à la valeur des prédictions. Ces protocoles d'évaluation ont une importance primordiale, puisqu'ils permettent de justifier de nouveaux développements et d'en démontrer les bénéfices.

Finalement, une collaboration plus étroite entre les prévisionnistes et les utilisateurs de prédictions éoliennes est nécessaire dans le futur. Les bénéfices liés à l'utilisation de prédictions ne peuvent être optimaux que si ces prédictions sont spécifiques à l'application voulue et au processus de prise de décision associé. Ces applications sont très nombreuses, et chacune est liée à différentes méthodologies de prise de décision. Avoir une vision plus globale du problème de la gestion de la production éolienne sera bénéfique et pour les prévisionnistes et pour les utilisateurs de prédictions.

C

Implementation of a Module for Online Estimation of Prediction Intervals of Wind Generation

The methods for the estimation of prediction intervals of wind generation have been developed for online operation. In the frame of the thesis, we have also developed a module (in C++ programming language) which is integrated in the ANEMOS prediction platform. In the present Appendix, we present the main characteristics of this module, from its configuration by the analyst to the visualization of prediction intervals, via some details about its operation in an online environment.

Module setup

In the ANEMOS prediction platform, the necessary information for the operation of the different prediction modules are stored in a Static Data Repository (SDR), which is the database with all the static data. These information include the characteristics of the considered wind farm (e.g. geographical coordinates), of the NPWs (e.g. temporal resolution), etc. Then, each of the prediction modules may utilize a local configuration file in which are stored its specific parameters. The configuration file for the interval forecasting module contains the following parameters:

***Method*:** defines the chosen approach for the computation of prediction intervals, i.e. *linear opinion pool* or *adapted resampling*,

***File path*:** defines the path to the directory where are stored the local memory files necessary for the operation of the module (i.e. history of predictions and influential variables, and error samples),

***Number of intervals*:** defines the number of prediction intervals to be computed. Whatever the chosen approach, several intervals can be computed at once since they model the whole predictive distribution of wind generation for every horizon,

***Nominal coverage rate (i)*:** defines the nominal coverage rate of the i^{th} prediction interval to be computed. Such value is comprised between 0 and 100%,

***Sample size*:** defines the size of the samples of past prediction errors,

***Resampling times*:** defines the number of bootstrap replications if one chooses to utilize the adapted resampling approach,

***Number of influential variables*:** defines the number of influential variables,

***Influential variable (i)*:** gives the type of the i^{th} influential variable, such as predicted wind power or forecast wind speed for instance,

***Number of ranges (i)*:** defines the number of ranges of values that have to be considered for the i^{th} influential variable. For instance, if this parameter is set to 5 for the predicted power variable, then the range of possible predicted power values is divided into 5 ranges. To each of these ranges is associated a triangular fuzzy set,

***[low,up] (i,j)*:** defines the lower and upper bound of the j^{th} range of values for the i^{th} influential variable. The fuzzy set related to this range of values is defined accordingly.

One sees from this list of parameters that the configuration of the module can be tailored to the considered application, depending on the analyst's expertise on the specificities of that application.

Operation

The module provides interval forecasts with the same forecast length and resolution than the point prediction module it is associated to. In most of the cases, predictions are produced for look-ahead times up to 48-hour ahead, with an hourly resolution. Also, prediction intervals are provided with the same frequency of update than the point prediction module it is associated to. In general, forecasts are updated on an hourly basis for statistical methods and only when NWPs are provided for the case of physical methods (e.g. every 6 hours when considering HIRLAM meteorological predictions as input).

A scheduler is at the heart of the ANEMOS prediction platform. It manages the retrieval of onsite measures and meteorological forecasts, as well as the operation of the various modules. Since interval forecasts are associated to series of point predictions, the module for uncertainty estimation is run just after the point forecasting one. The chain of tasks to be carried out by the uncertainty estimation module is similar to the chain described by

Algorithm 4.1, with some additional steps dedicated to the communication with the Time-Series Data Repository (TSDR), which is the database containing all the dynamic data, as well as file management issues. This chain is given by Algorithm B.1.

Algorithm B.1: *The chain of tasks to be carried out by the uncertainty estimation module at each prediction time.*

step 1.	Load the module configuration file and the relevant parameters from the SDR
step 2.	Retrieve the new power measures, power predictions and influential variable values from the TSDR
step 3.	Load the memory files containing stored values of power predictions and influential variables
step 4.	Load the memory files containing the error samples
step 5.	Calculate the prediction errors from collected power measures and stored predictions
step 6.	Determine the forecast conditions related to calculated prediction errors
step 7.	Update and save the files containing power predictions and influential variables
step 8.	Update error samples given the forecast conditions, and save them in the memory files
step 9.	Use the fuzzy inference model for determining the distributions of prediction errors associated to every power predictions
step 10.	Apply either the linear opinion pool or the adapted resampling method for estimating the bounds of the prediction intervals with the required nominal coverage rates
step 11.	Save the prediction intervals in the TSDR

During online operation, power measurements, power predictions and influential variables values (e.g. wind speed forecasts) may be erroneous or missing. Hence, the second step of the above Algorithm, which consists in the retrieval of these data, also integrates a data checking procedure. If power measurements are missing, the prediction errors cannot be calculated and thus error samples are not updated. And, if power predictions or influential variable values are missing or erroneous, interval forecasts are not computed. Instead, series of ”-99” values are returned (in step 11), as well as a message indicating that interval computation was not possible.

Results and Visualization

The ANEMOS platform is also composed by a man-machine interface. Such an interface allows the end-user to visualize historic power production, power predictions, as well as associated prediction intervals. Also, the end-user may use that interface for consulting reports on the performance of the various prediction methods over a given period.

Figure C.1 shows a general view of the man-machine interface. In the upper window is represented the island of Crete, with the 12 wind farms for which wind generation is predicted. For wind farms that are equipped with SCADA systems and thus for which power measurements are regularly stored in the TSDR, the interval forecasts produced from the previously described uncertainty estimation module can be visualized at the same time than the point predictions. The two lower windows of Figure C.1 display 48-ahead point predictions to which are associated prediction intervals with a nominal coverage rate of 80%.

__Figure C.1:__ The man-machine interface of the ANEMOS prediction platform for a Windows XP operating system.

D

Point Forecasting Methods - Evaluation Results

D.1 Content description

In the following are gathered all the results from the verification of the point forecasting methods on the Tunø Knob, Klim, Golagh and Sotavento case-studies. Are shown:

- the NMAE as a function of the look-ahead time (Figures D.1, D.7, D.13, and D.19),
- the NRMSE as a function of the look-ahead time (Figures D.2, D.8, D.14, and D.20),
- the improvement with respect to persistence as a function of the look-ahead time for the NRMSE criterion (Figures D.3, D.9, D.15, and D.21),
- the R^2 as a function of the look-ahead time (Figures D.4, D.10, D.16, and D.22),
- error margin plots that give the proportion of errors less than 5% of P_n(Figures D.5, D.11, D.17, and D.23),
- error margin plots that give the proportion of errors less than 30% of P_n(Figures D.6, D.12, D.18, and D.24).

D.2 Tunø Knob

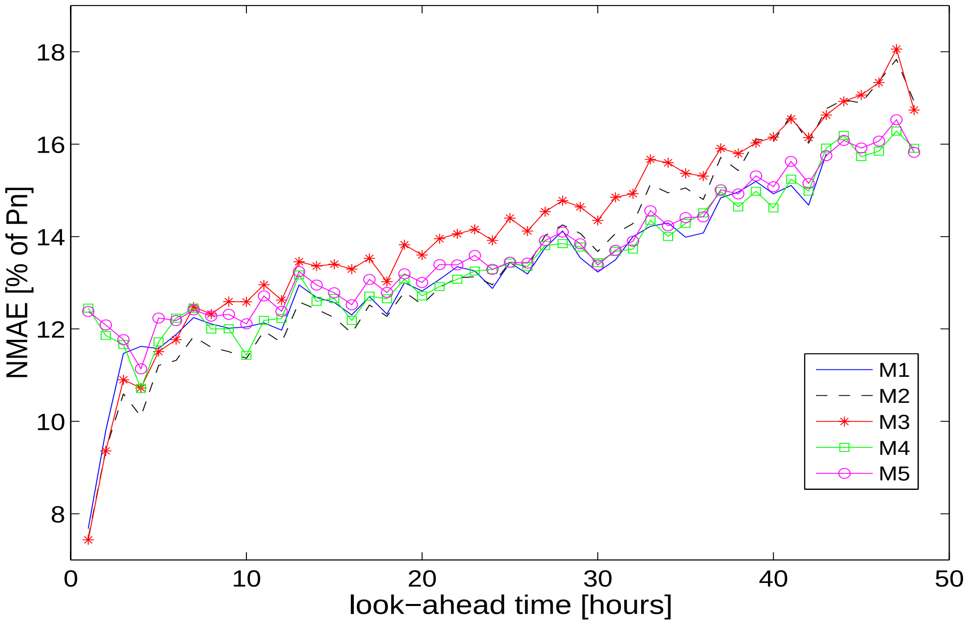


Figure D.1: *Performance evaluation of the forecasting methods with the NMAE measure as a function of the look-ahead time.*

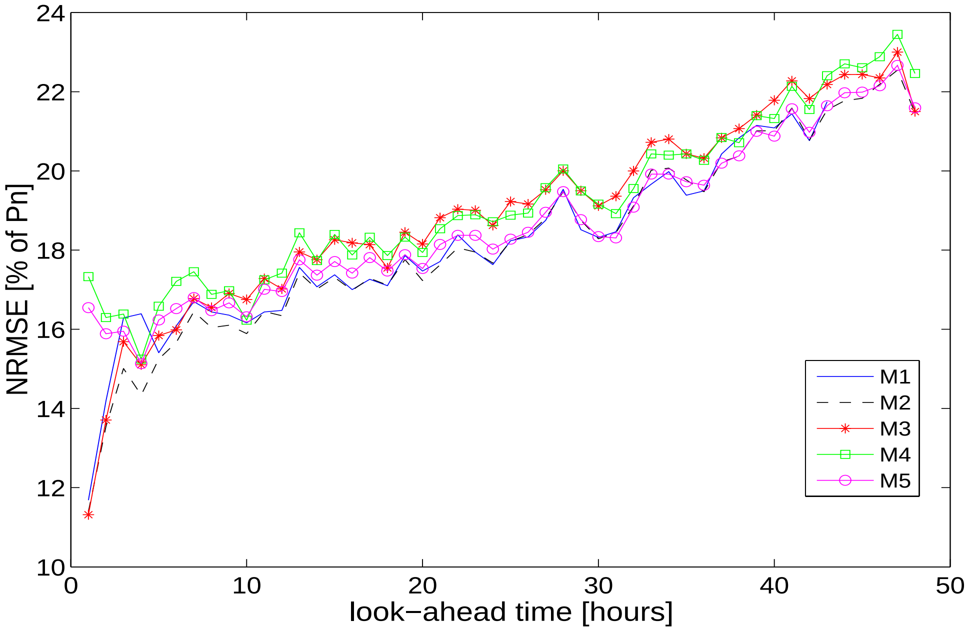


Figure D.2: *Performance evaluation of the forecasting methods with the NRMSE measure as a function of the look-ahead time.*

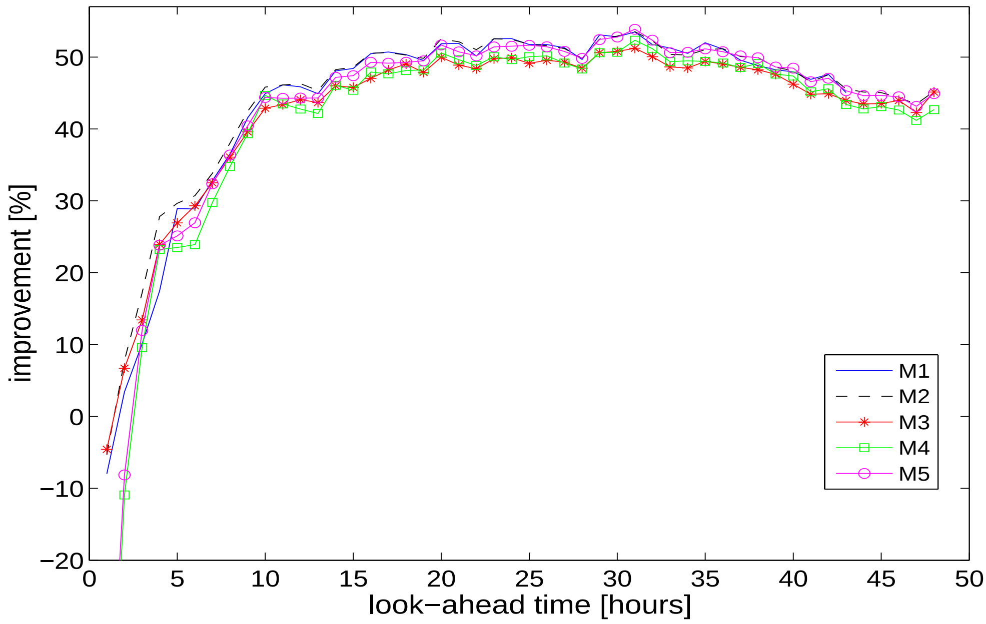


Figure D.3: *Skill evaluation of the forecasting methods - Improvement with respect to persistence for the NRMSE criterion, as a function of the look-ahead time.*

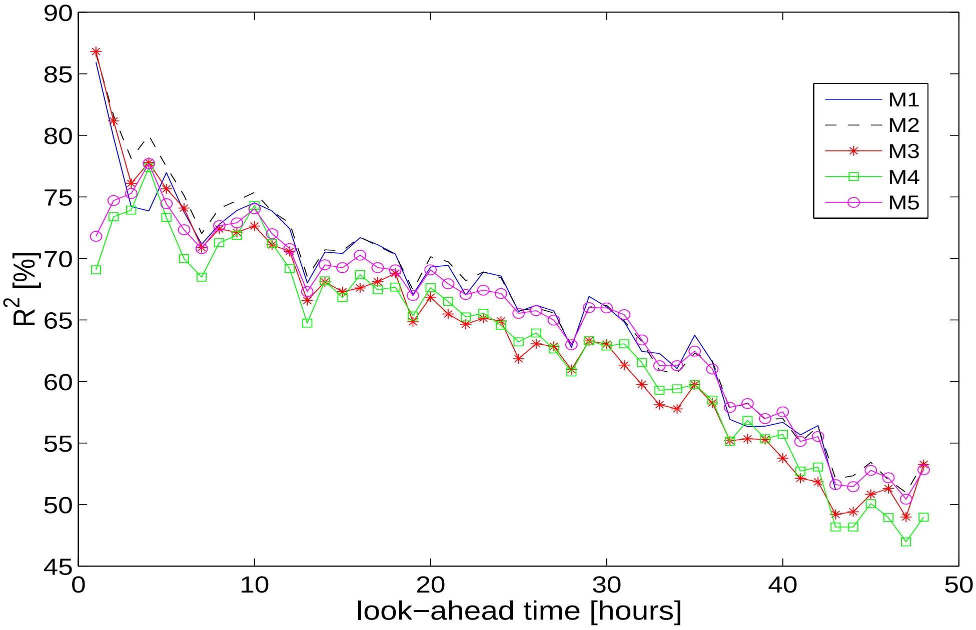


Figure D.4: *Skill evaluation of the forecasting methods - R^2 as a function of the look-ahead time.*

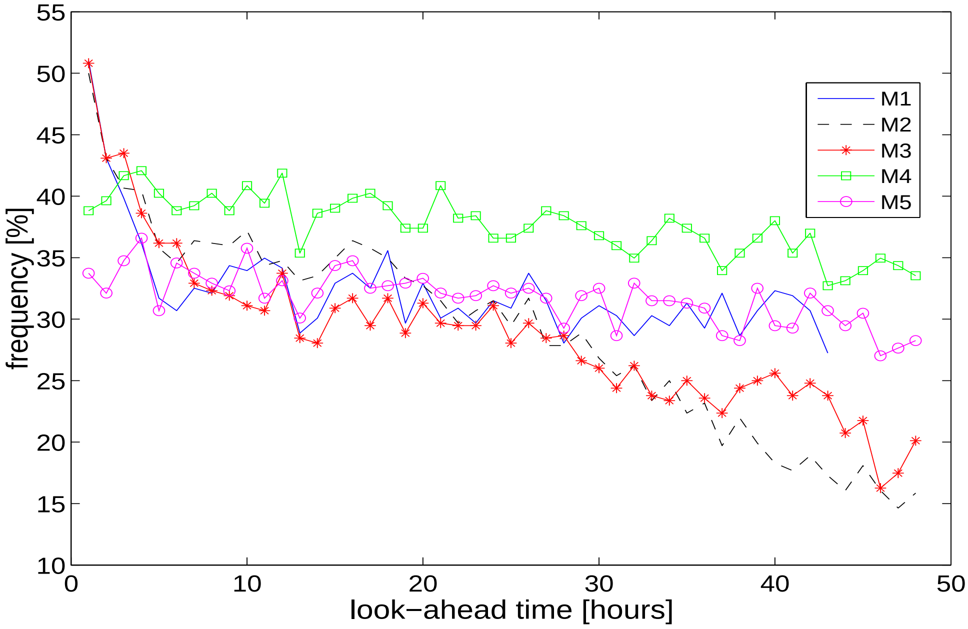


Figure D.5: *Proportion of errors within a $\pm 5\%$ (of P_n) error margin as a function of the look-ahead time.*

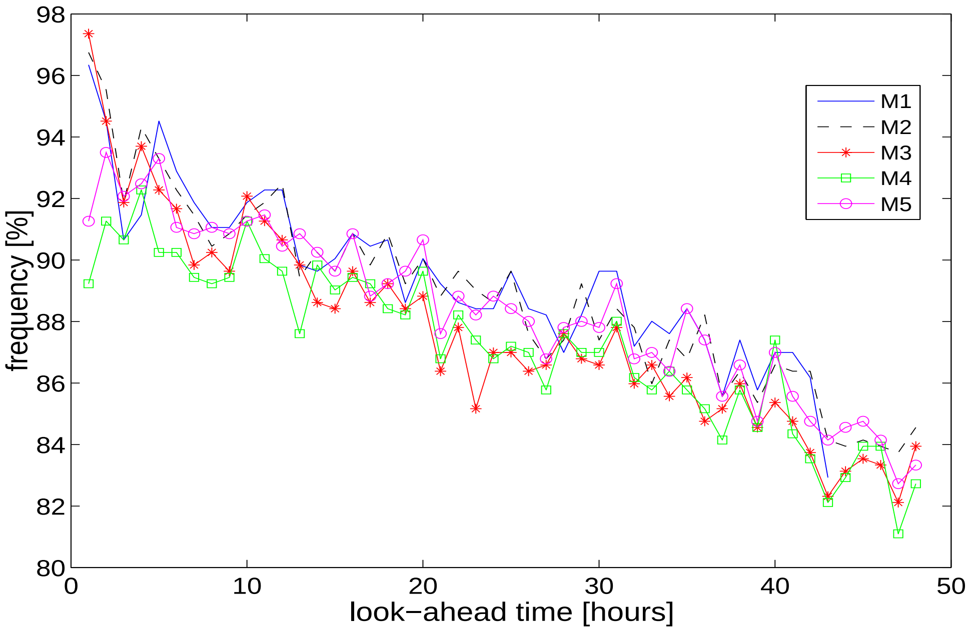


Figure D.6: *Proportion of errors within a $\pm 30\%$ (of P_n) error margin as a function of the look-ahead time.*

D.3 Klim

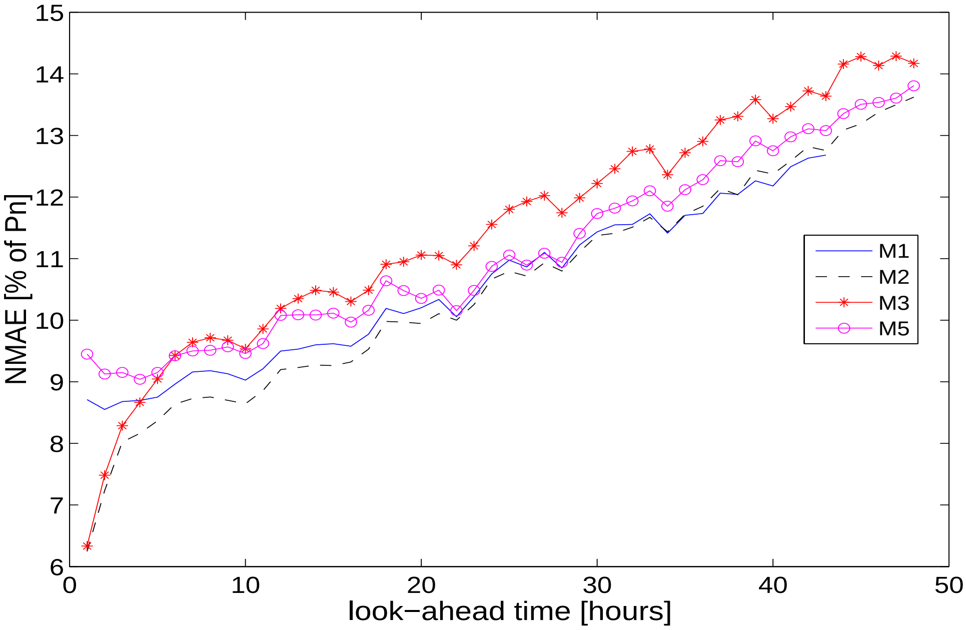


Figure D.7: *Performance evaluation of the forecasting methods with the NMAE measure as a function of the look-ahead time.*

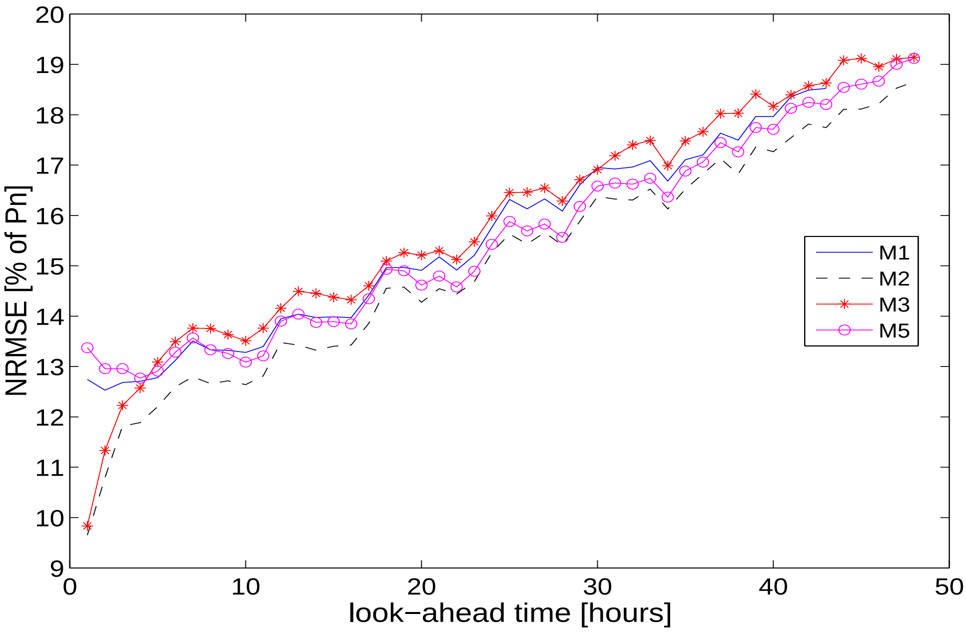


Figure D.8: *Performance evaluation of the forecasting methods with the NRMSE measure as a function of the look-ahead time.*

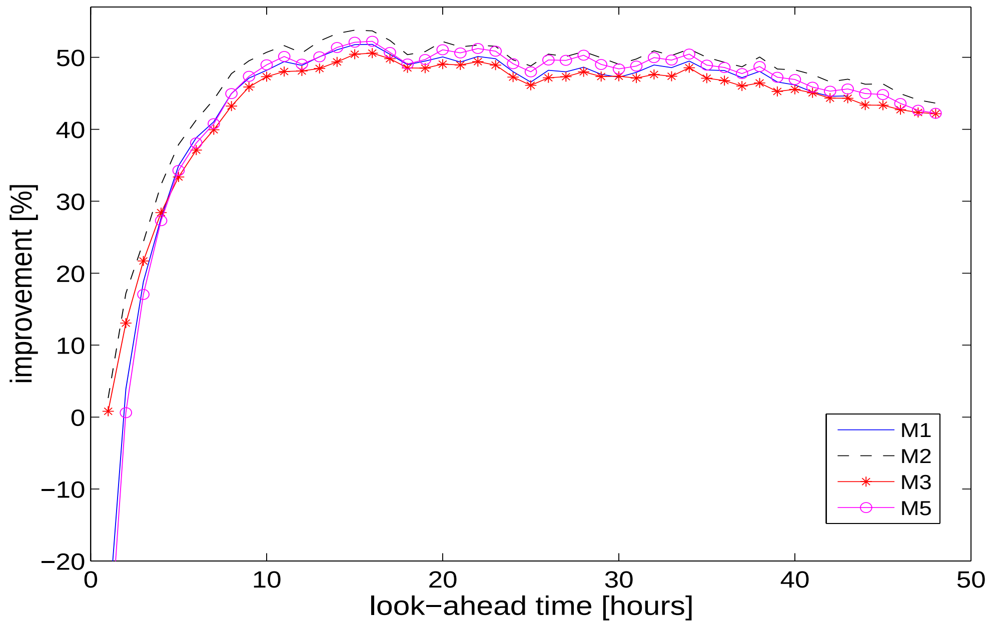


Figure D.9: *Skill evaluation of the forecasting methods - Improvement with respect to persistence for the NRMSE criterion, as a function of the look-ahead time.*

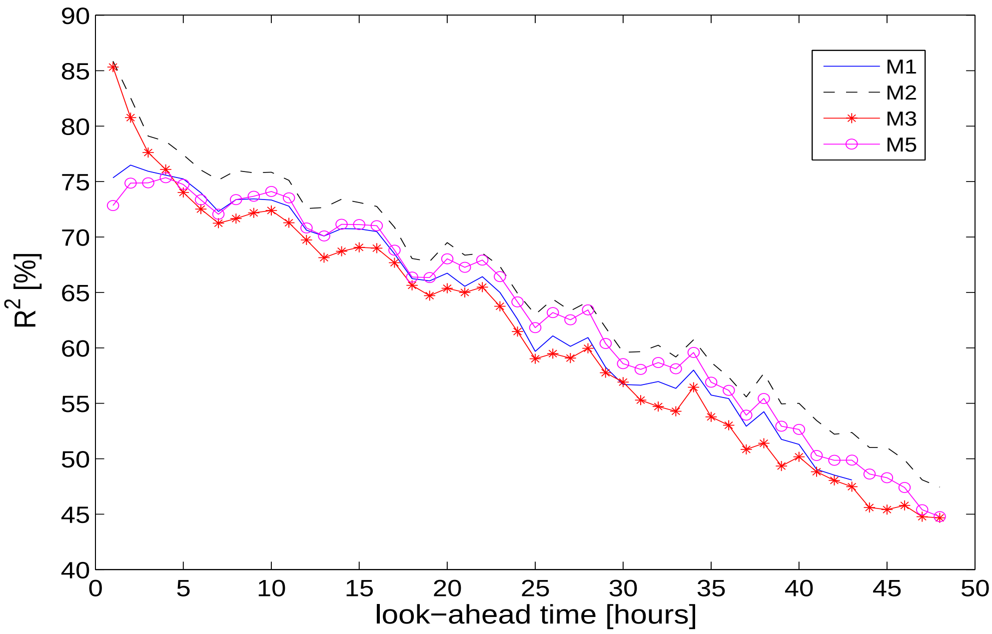


Figure D.10: *Skill evaluation of the forecasting methods - R^2 as a function of the look-ahead time.*

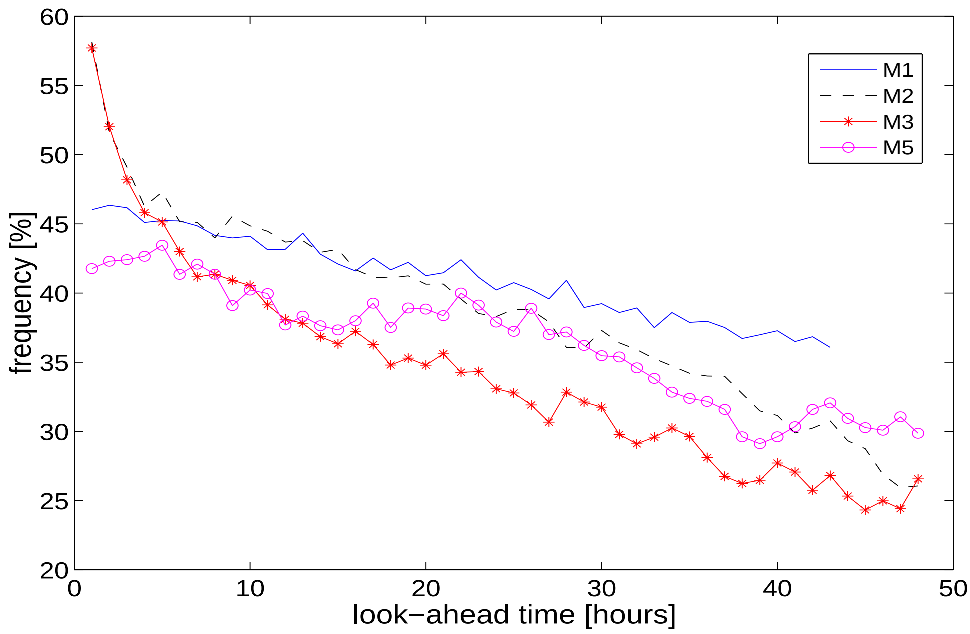


Figure D.11: *Proportion of errors within a ±5% (of P_n) error margin as a function of the look-ahead time.*

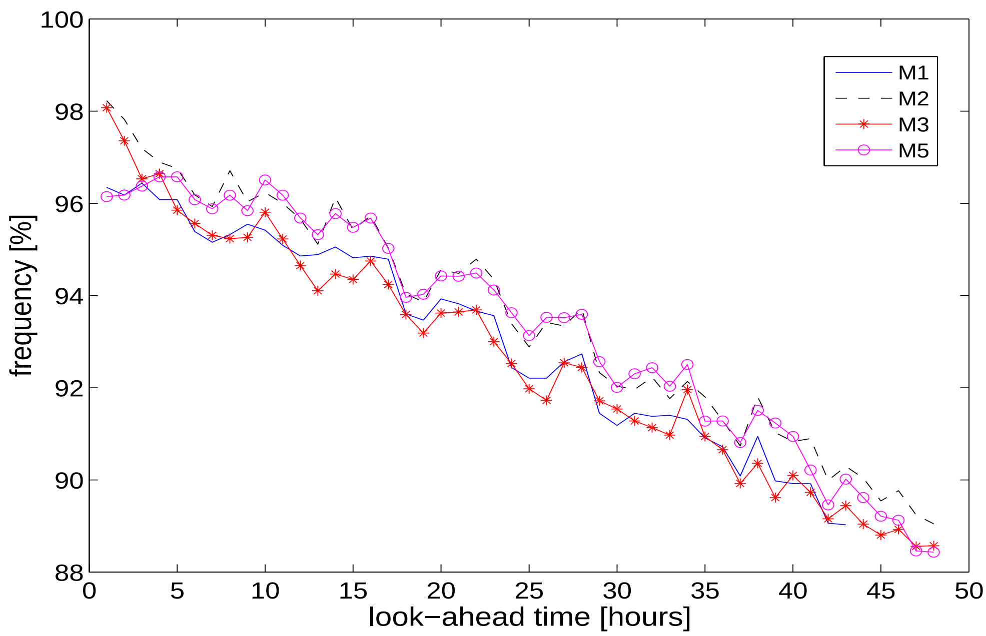


Figure D.12: *Proportion of errors within a ±30% (of P_n) error margin as a function of the look-ahead time.*

D.4 Golagh

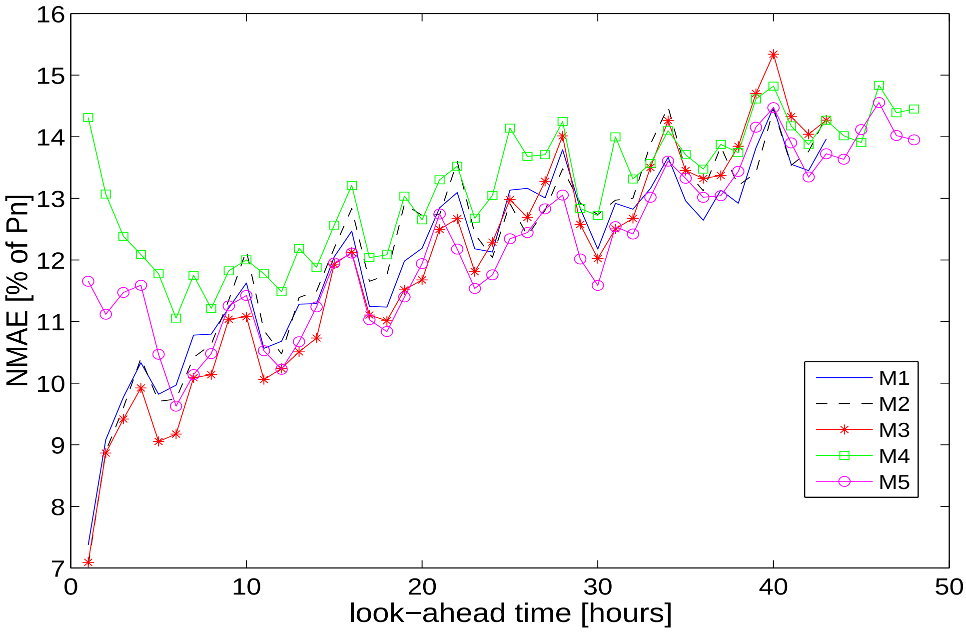


Figure D.13: *Performance evaluation of the forecasting methods with the NMAE measure as a function of the look-ahead time.*

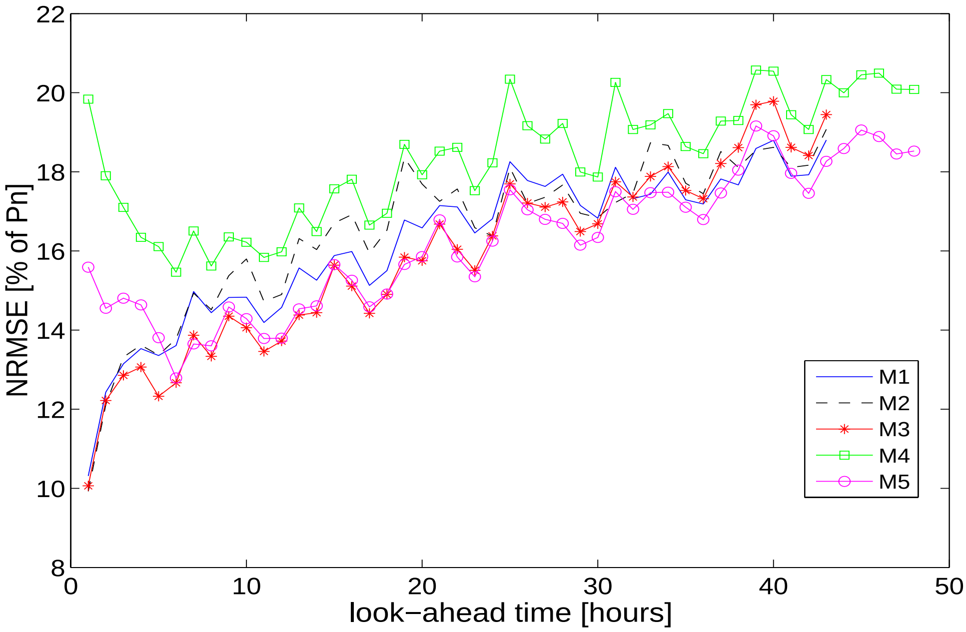


Figure D.14: *Performance evaluation of the forecasting methods by the use of the NRMSE measure as a function of the look-ahead time.*

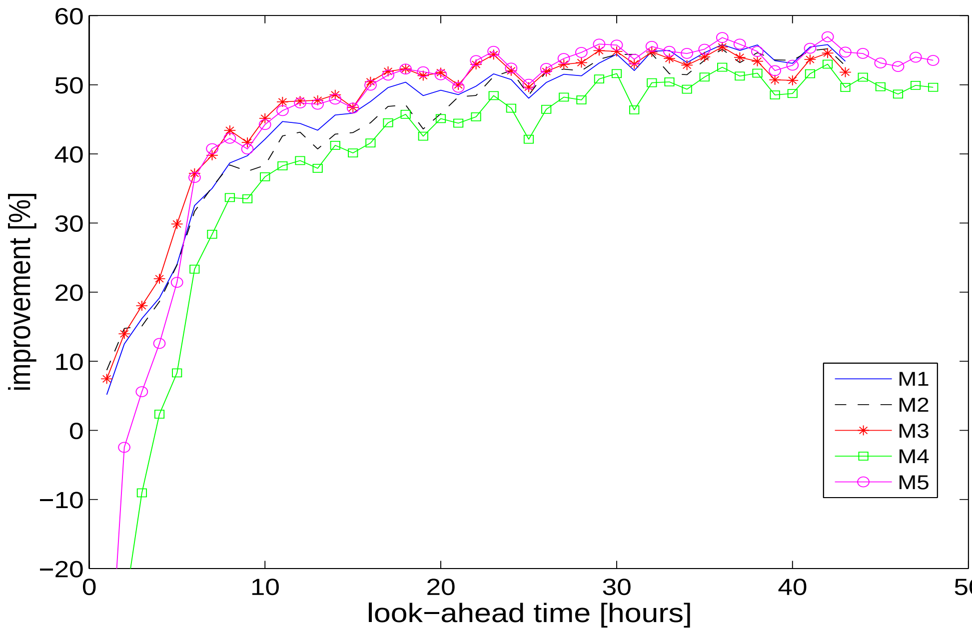


Figure D.15: *Skill evaluation of the forecasting methods - Improvement with respect to persistence for the NRMSE criterion, as a function of the look-ahead time.*

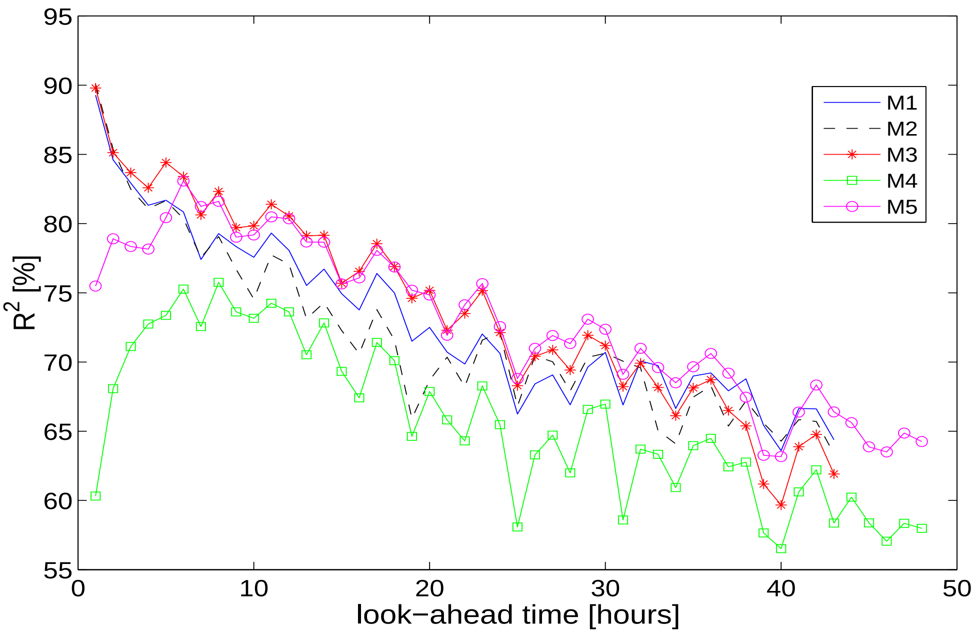


Figure D.16: *Skill evaluation of the forecasting methods - R^2 as a function of the look-ahead time.*

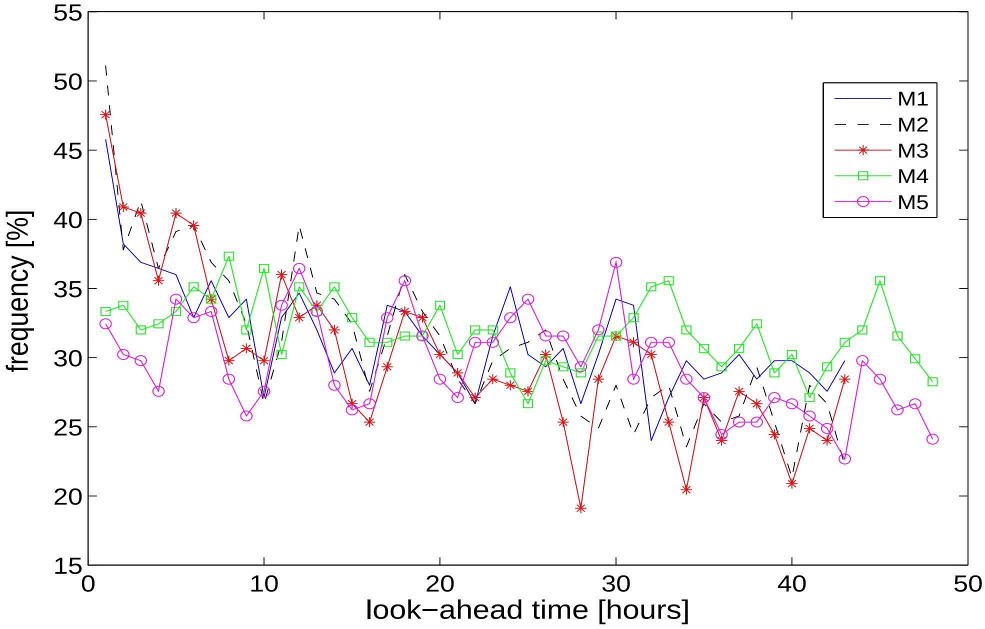


Figure D.17: *Proportion of errors within a ±5% (of P_n) error margin as a function of the look-ahead time.*

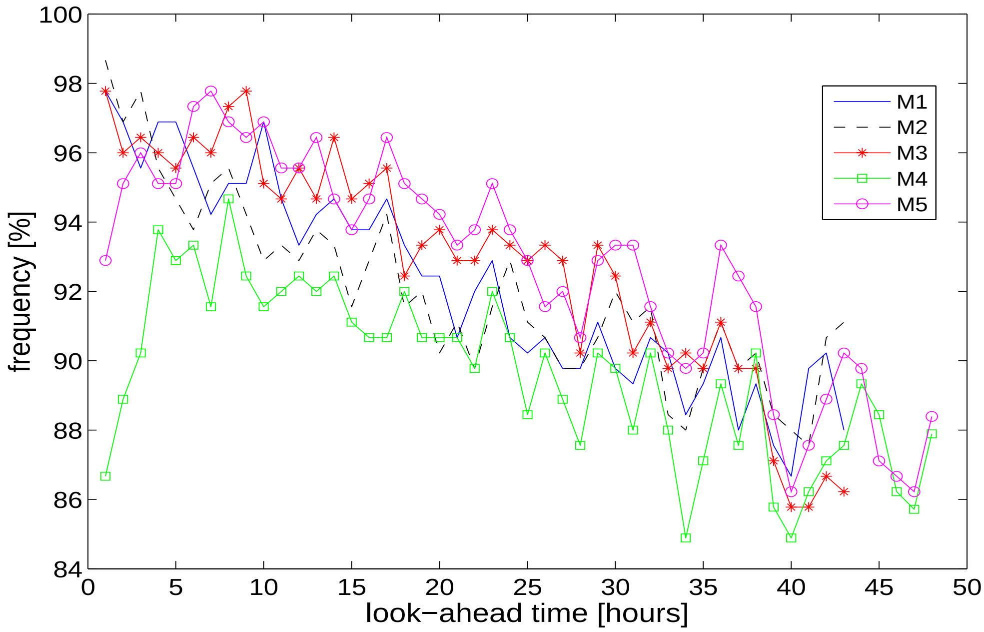


Figure D.18: *Proportion of errors within a ±30% (of P_n) error margin as a function of the look-ahead time.*

D.5 Sotavento

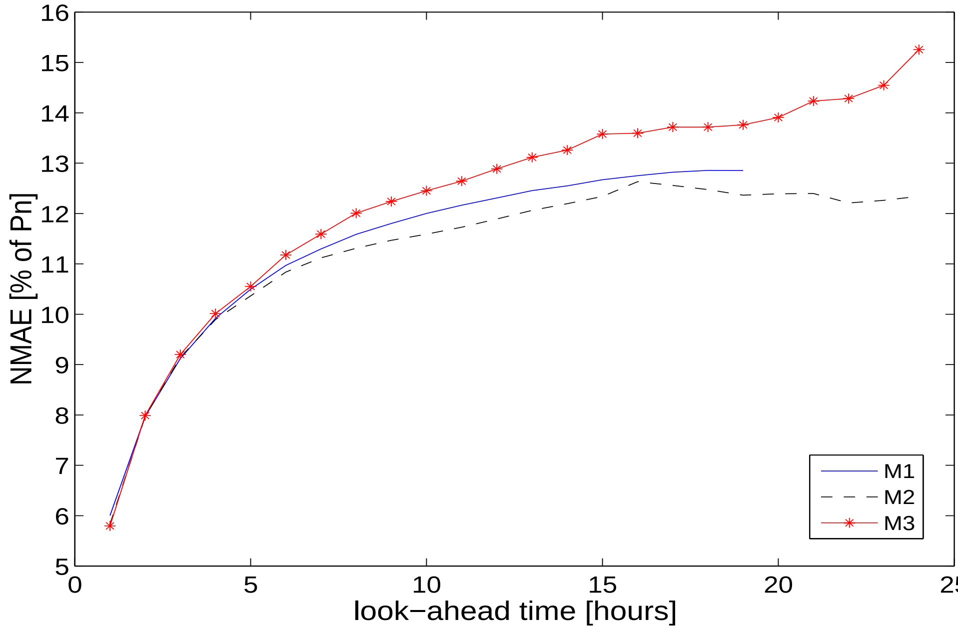


Figure D.19: *Performance evaluation of the forecasting methods with the NMAE measure as a function of the look-ahead time.*

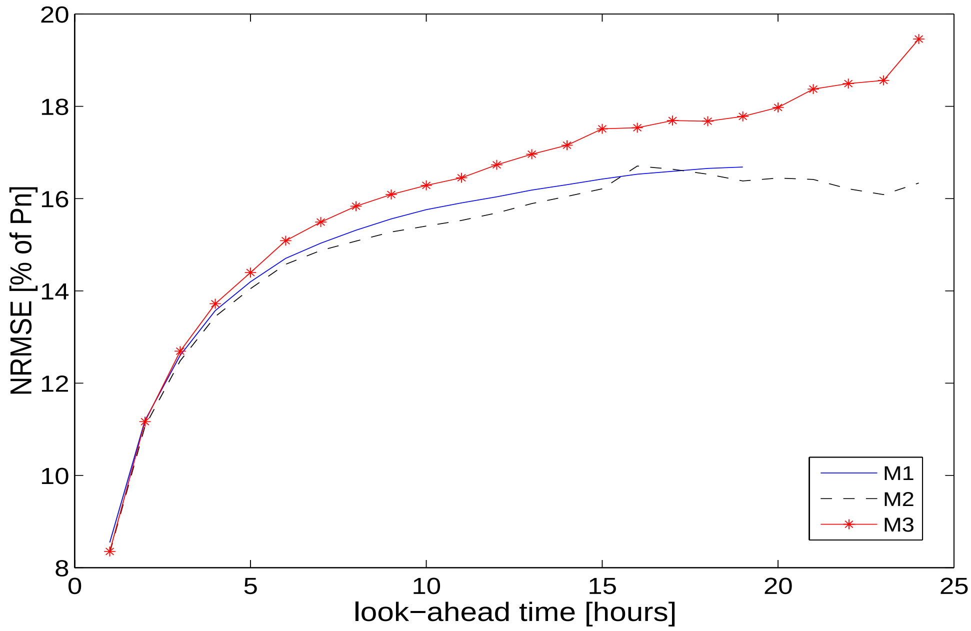


Figure D.20: *Performance evaluation of the forecasting methods by the use of the NRMSE measure as a function of the look-ahead time.*

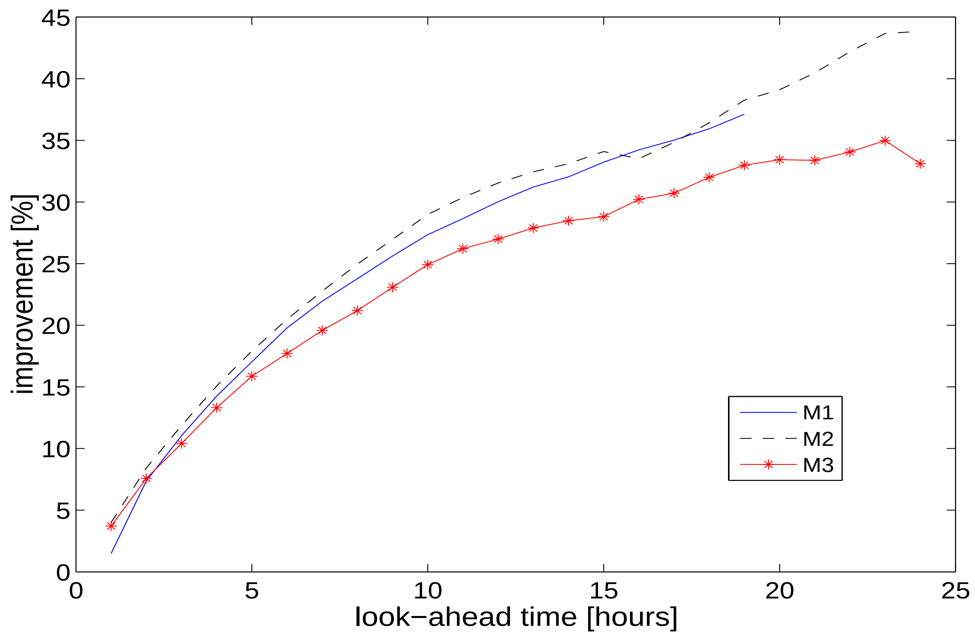


Figure D.21: *Skill evaluation of the forecasting methods - Improvement with respect to persistence for the NRMSE criterion, as a function of the look-ahead time.*

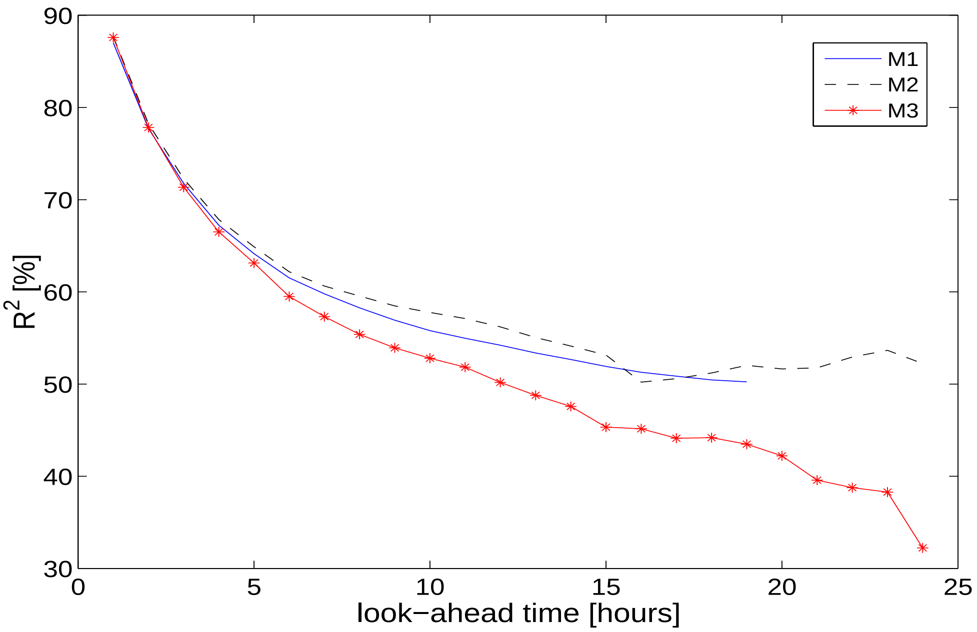


Figure D.22: *Skill evaluation of the forecasting methods - R^2 as a function of the look-ahead time.*

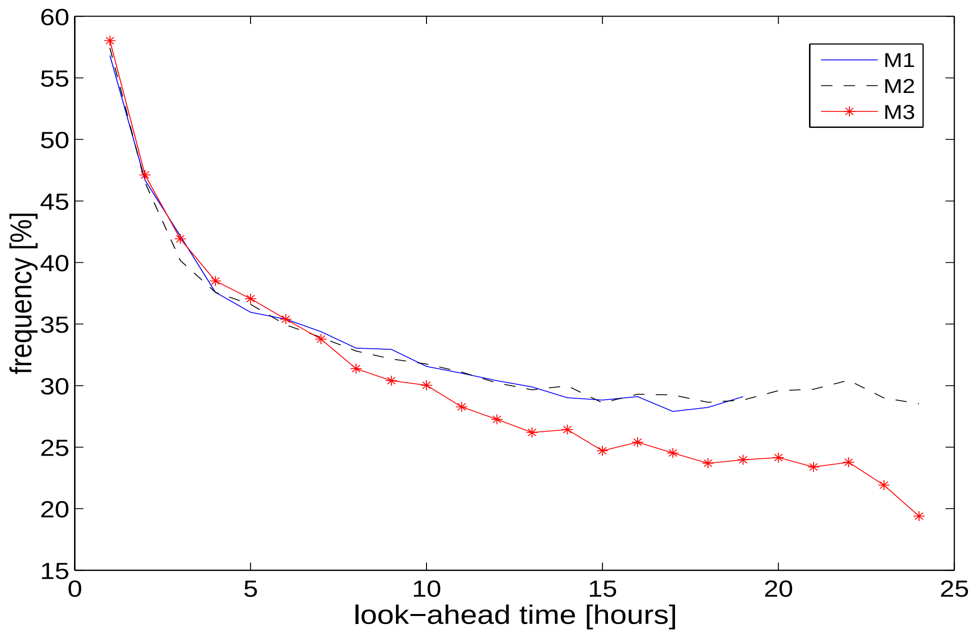


Figure D.23: *Proportion of errors within a $\pm 5\%$ (of P_n) error margin as a function of the look-ahead time.*

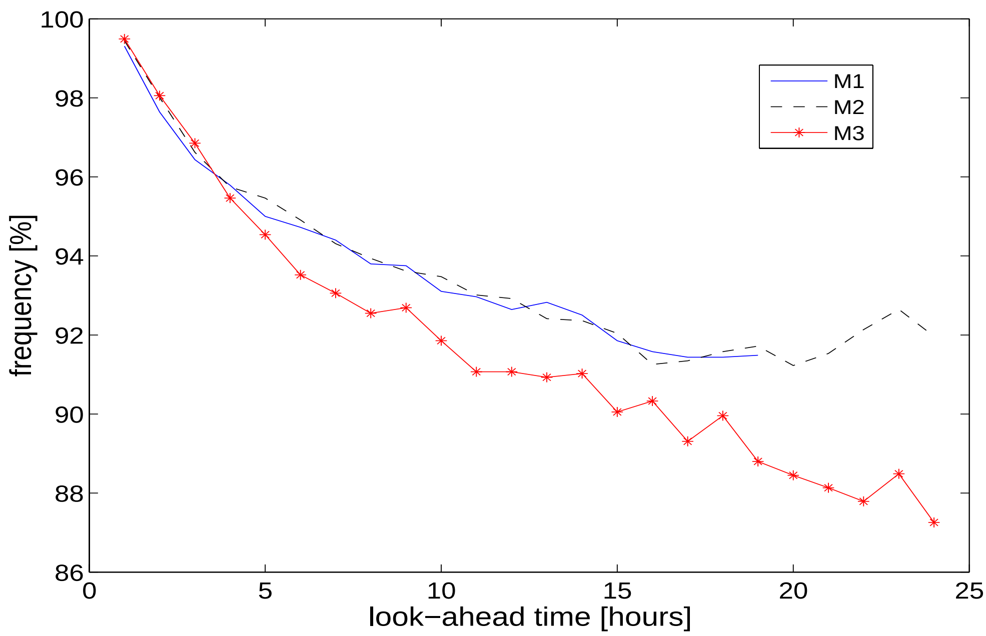


Figure D.24: *Proportion of errors within a $\pm 30\%$ (of P_n) error margin as a function of the look-ahead time.*

E

Uncertainty Characteristics - Full Survey

E.1 Content description

In the following are gathered all the results from the study of the characteristics of wind power prediction errors. Are shown:

- the Nbias as a function of the look-ahead time (Figures E.1, E.7, E.13, and E.19),
- the NSDE as a function of the look-ahead time (Figures E.2, E.8, E.14, and E.20),
- the comparison of power climatologies for the considered sites and the considered forecasting methods (Figures E.3, E.9, E.15, and E.21),
- the Nbias as a function of the level of measured power, for the assessment of the discrimination aspect (Figures E.4, E.10, E.16, and E.22),
- the Nbias as a function of the level of predicted power, for the assessment of the reliability aspect (Figures E.5, E.11, E.17, and E.23),
- the NSDE as a function of the level of predicted power in order to describe the dependence of the forecast uncertainty on the predictand level (Figures E.6, E.12, E.18, and E.24).

E.2 Tunø Knob

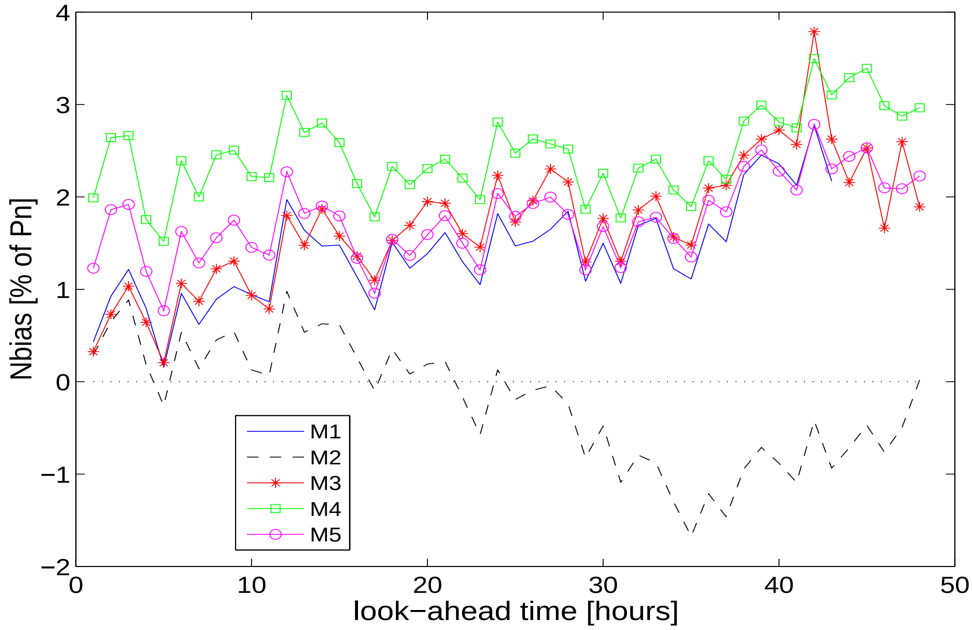


Figure E.1: *Performance evaluation of the forecasting methods with the Nbias measure as a function of the look-ahead time.*

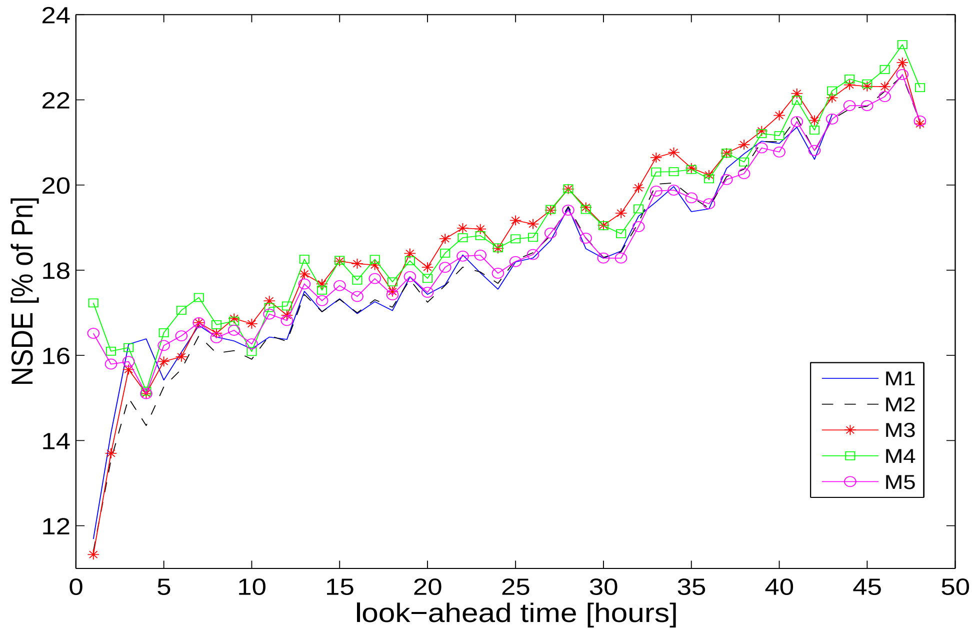


Figure E.2: *Performance evaluation of the forecasting methods with the NSDE measure as a function of the look-ahead time.*

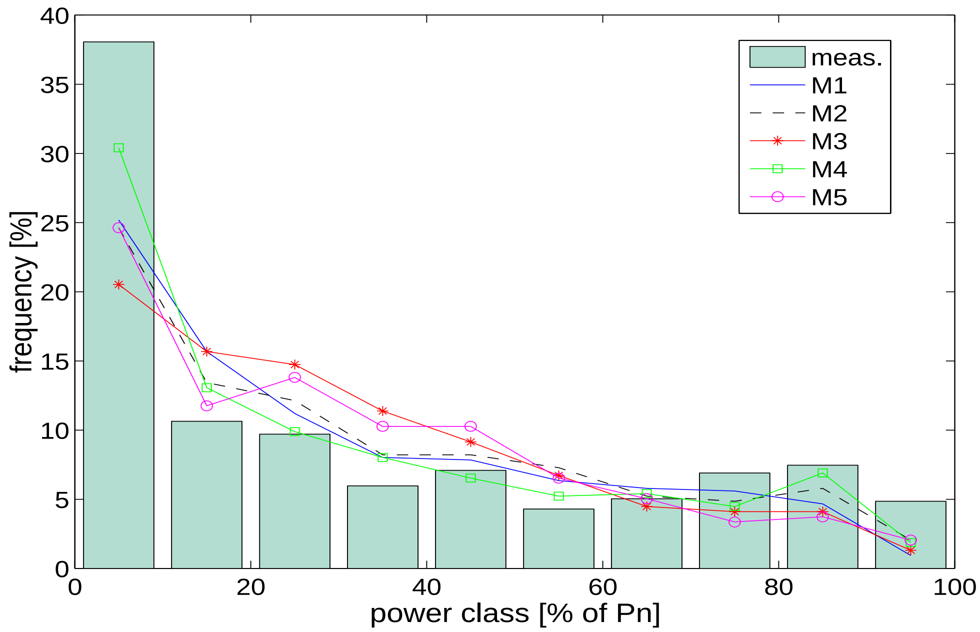


Figure E.3: *Power climatology of Tunø Knob compared with the climatologies of the five forecasting methods. Power measurements and forecasts are sorted with bins representing 10% of P_n. Focus is given to 18-hour ahead predictions.*

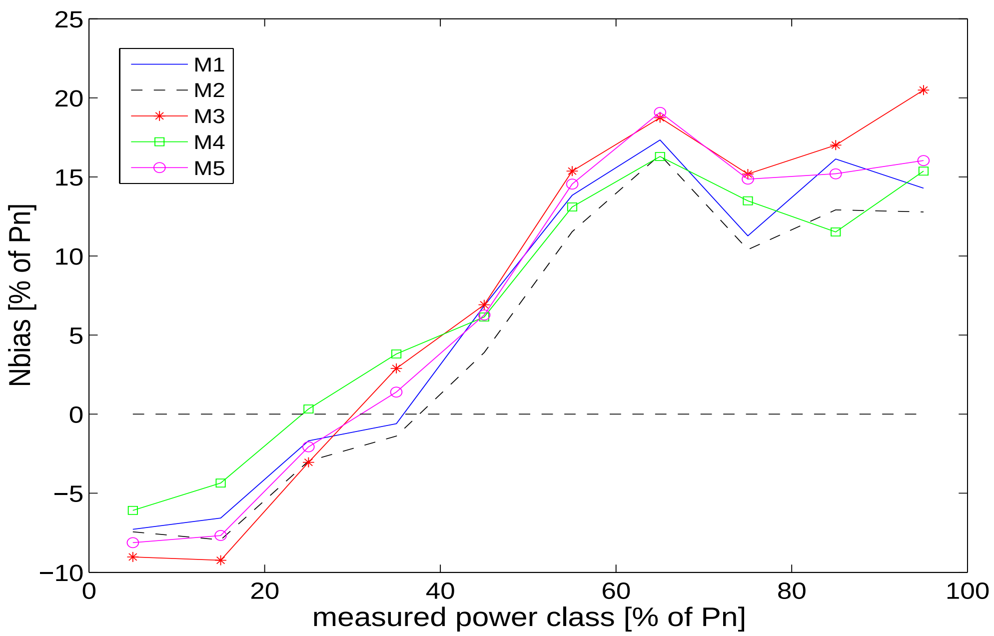


Figure E.4: *Normalized bias of the forecasting error distributions depending on power measures. Focus is given to 18-hour ahead predictions.*

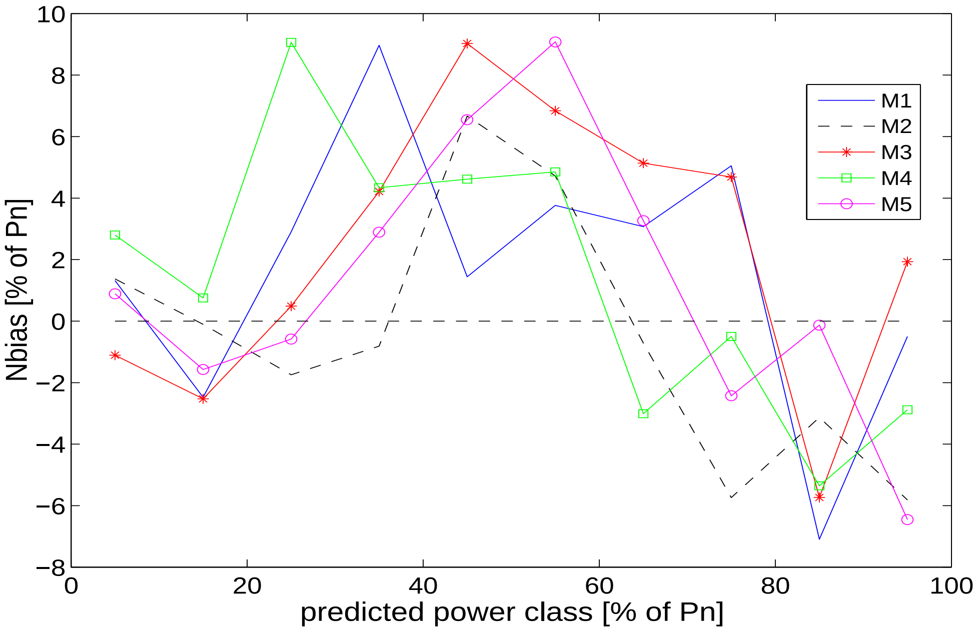


Figure E.5: *Normalized bias of the forecasting error distributions depending on the predicted power range. Focus is given to 18-hour ahead predictions.*

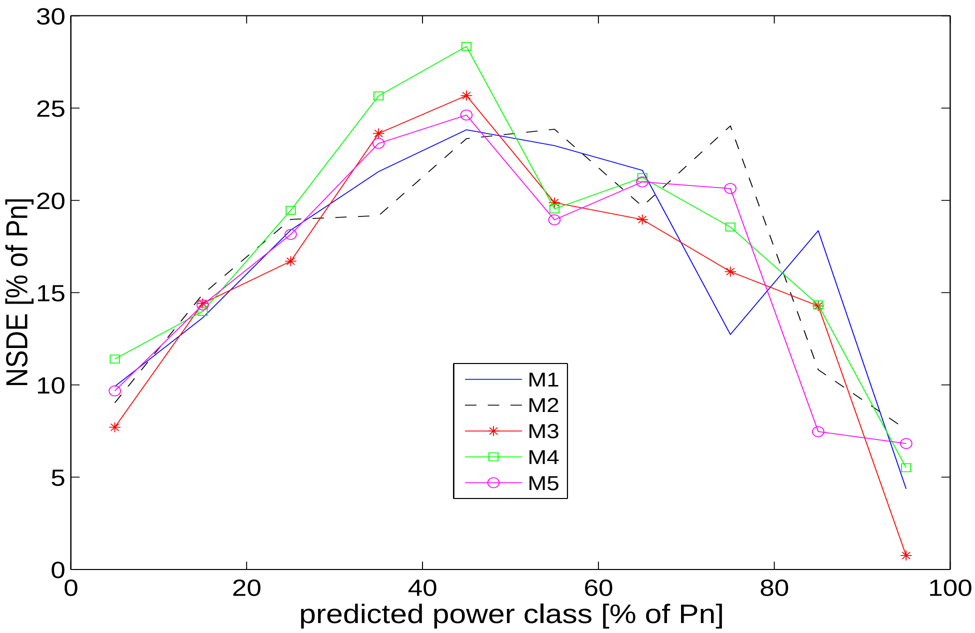


Figure E.6: *Normalized standard deviation of the forecasting error distributions depending on the predicted power range. Focus is given to 18-hour ahead predictions.*

E.3 Klim

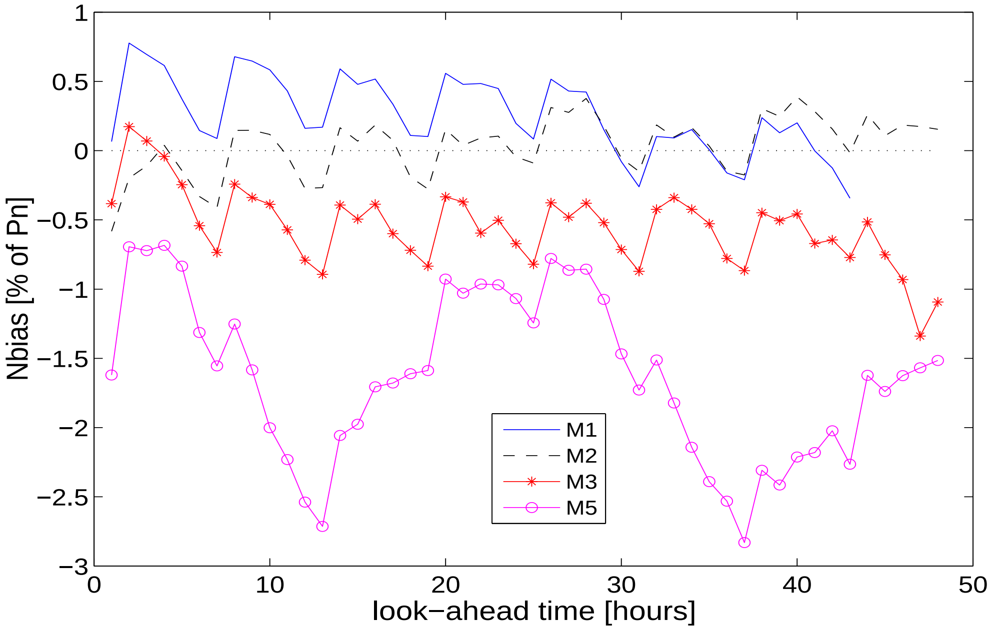


Figure E.7: *Performance evaluation of the forecasting methods with the Nbias measure as a function of the look-ahead time.*

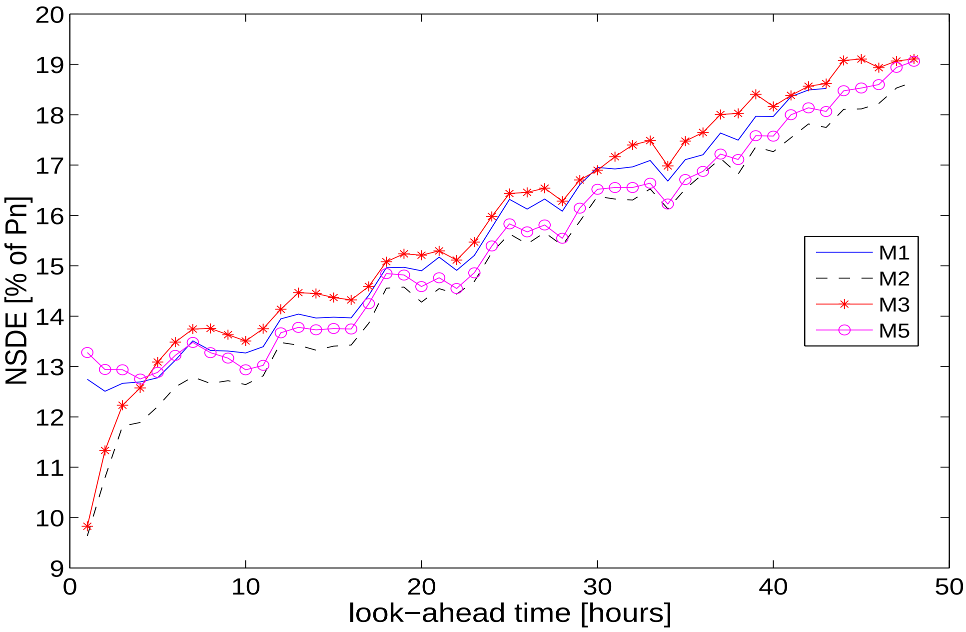


Figure E.8: *Performance evaluation of the forecasting methods with the NSDE measure as a function of the look-ahead time.*

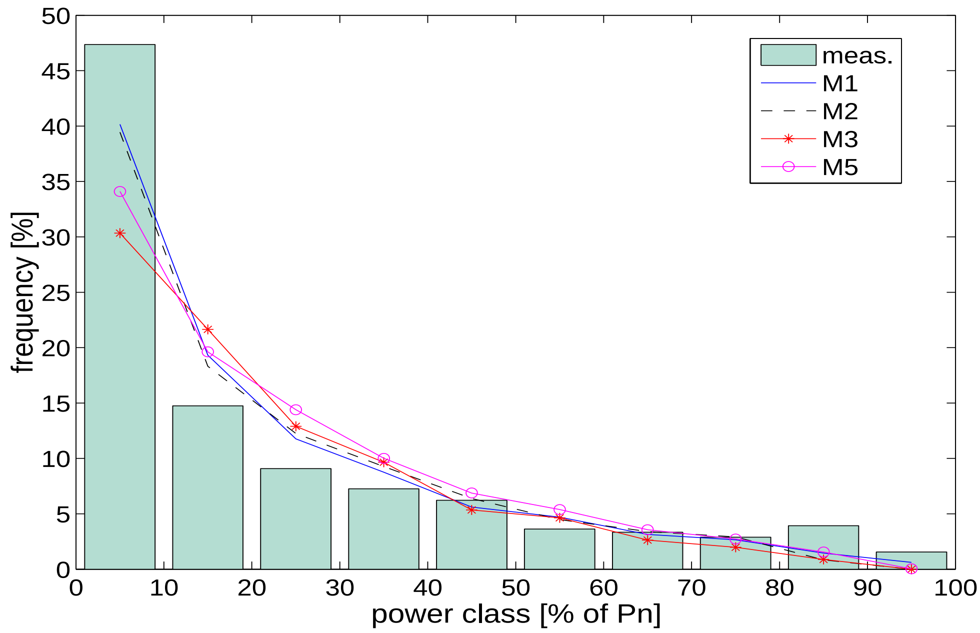


Figure E.9: *Power climatology of Klim compared with the climatologies of the five forecasting methods. Power measurements and forecasts are sorted with bins representing 10% of P_n. Focus is given to 18-hour ahead predictions.*

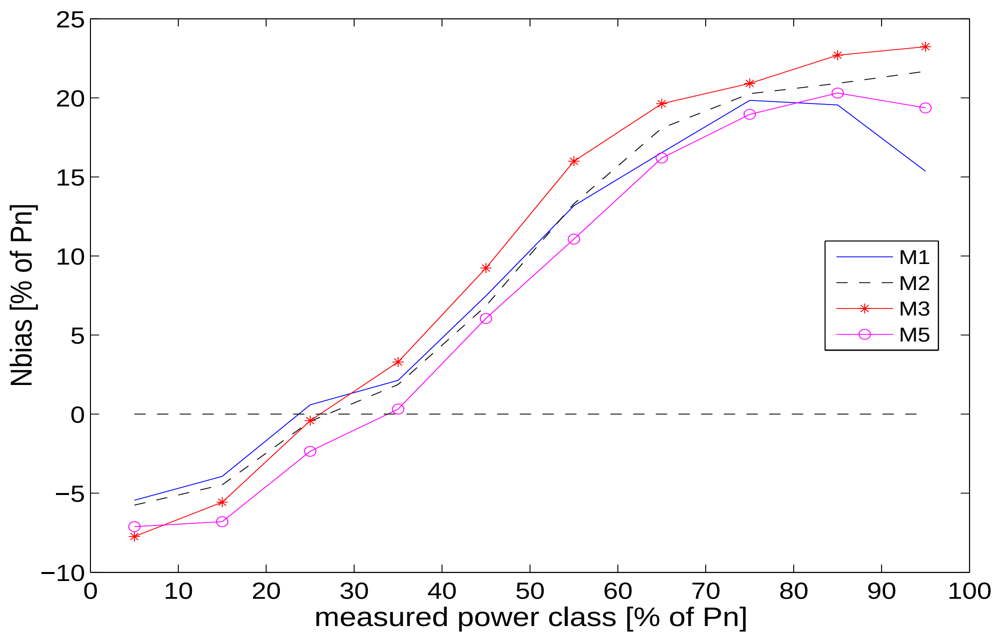


Figure E.10: *Normalized bias of the forecasting error distributions depending on power measures. Focus is given to 18-hour ahead predictions.*

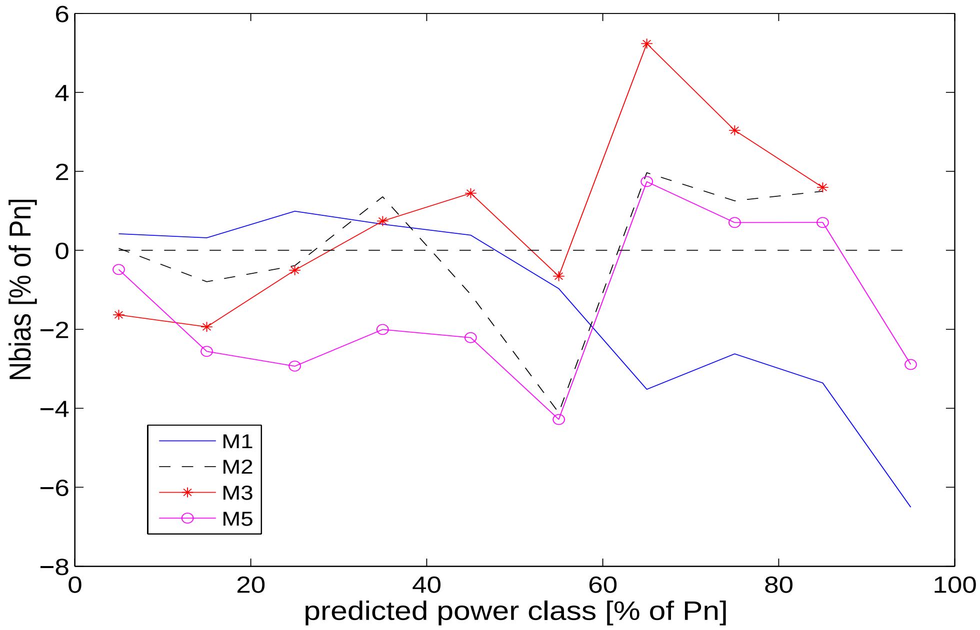


Figure E.11: *Normalized bias of the forecasting error distributions depending on the predicted power range. Focus is given to 18-hour ahead predictions.*

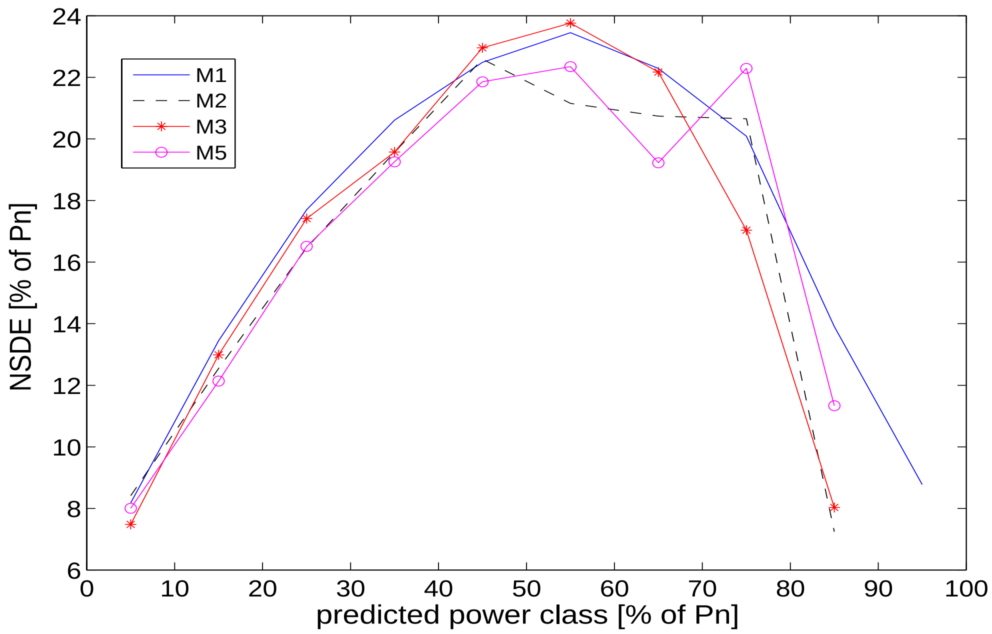


Figure E.12: *Normalized standard deviation deviations of the forecasting error distributions depending on the predicted power range. Focus is given to 18-hour ahead predictions.*

E.4 Golagh

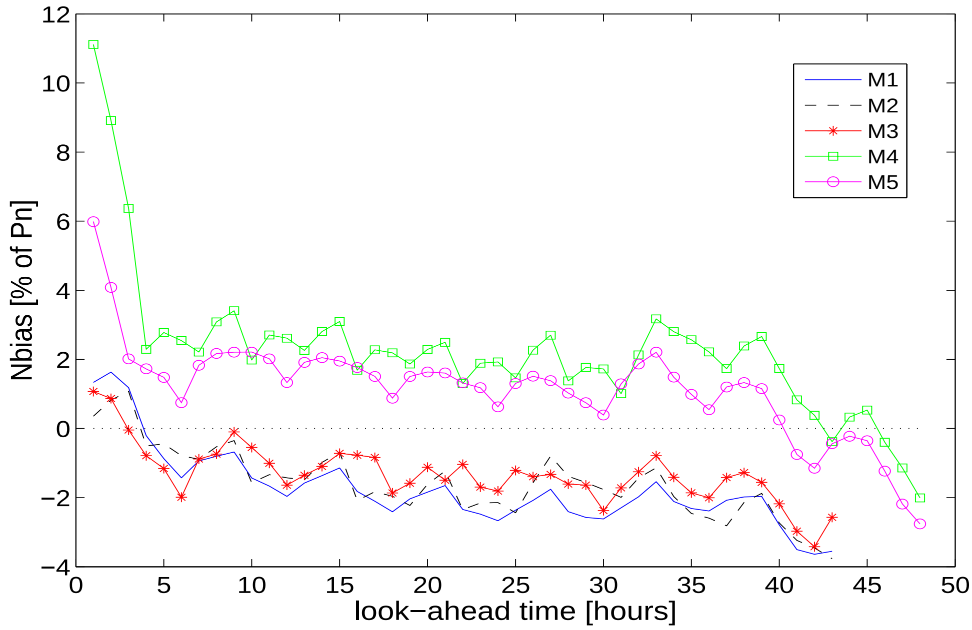


Figure E.13: *Performance evaluation of the forecasting methods with use of the Nbias measure as a function of the look-ahead time.*

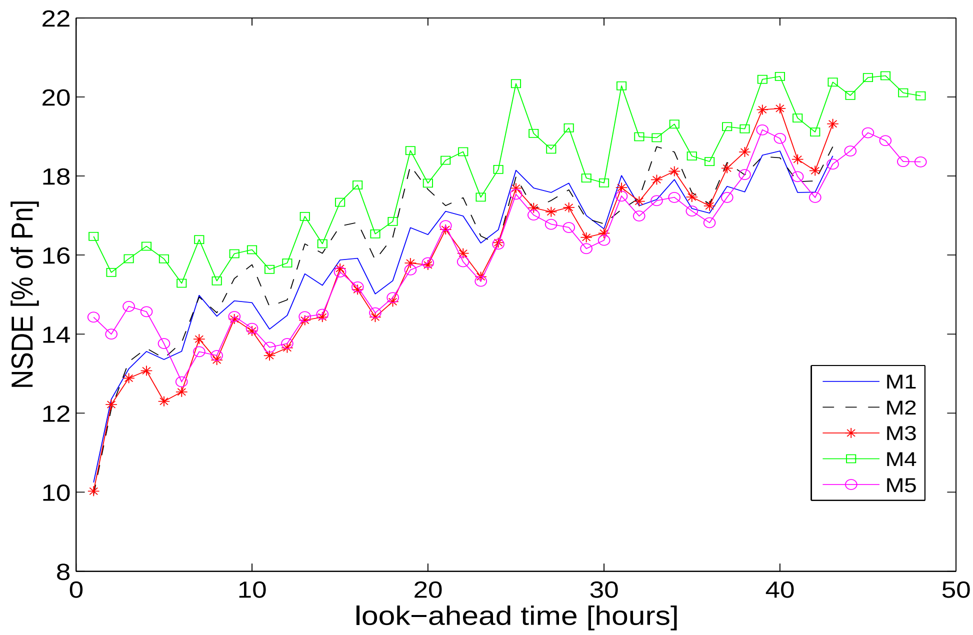


Figure E.14: *Performance evaluation of the forecasting methods with the NSDE measure as a function of the look-ahead time.*

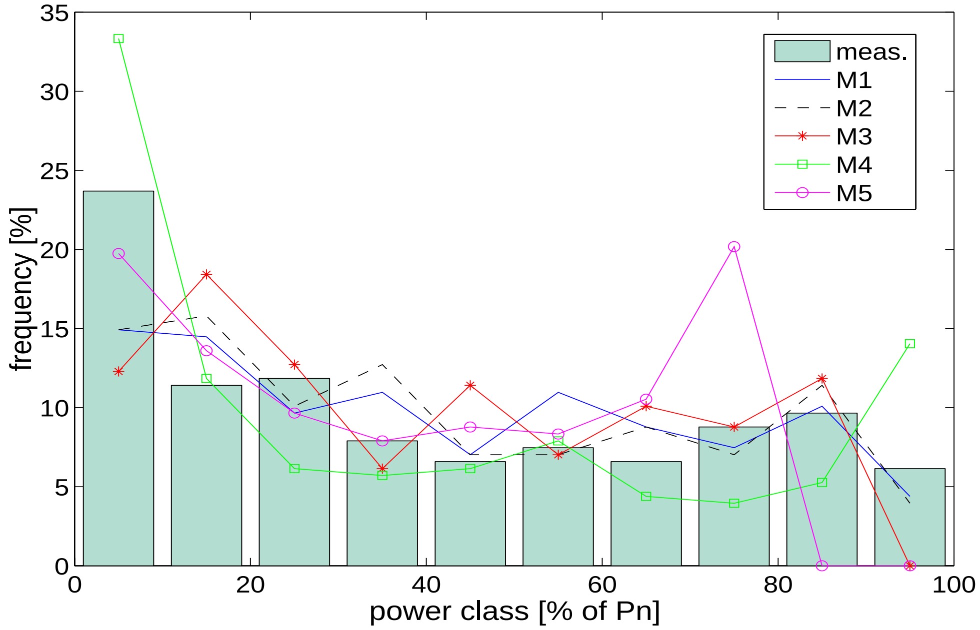


Figure E.15: *Power climatology of Golagh compared with the climatologies of the five forecasting methods. Power measurements and forecasts are sorted with bins representing 10% of P_n. Focus is given to 18-hour ahead predictions.*

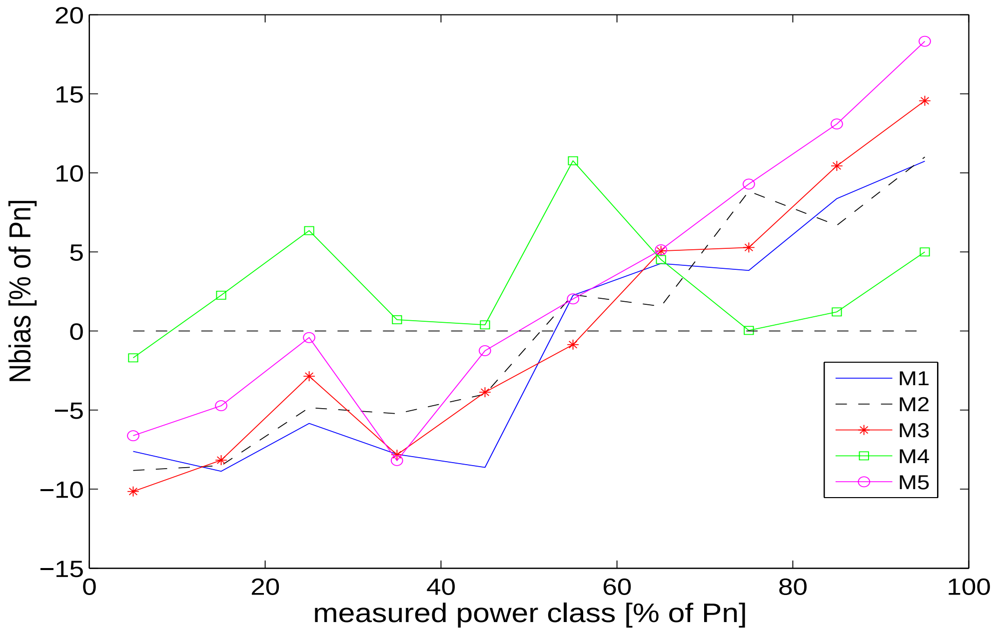


Figure E.16: *Normalized bias of the forecasting error distributions depending on power measures. Focus is given to 18-hour ahead predictions.*

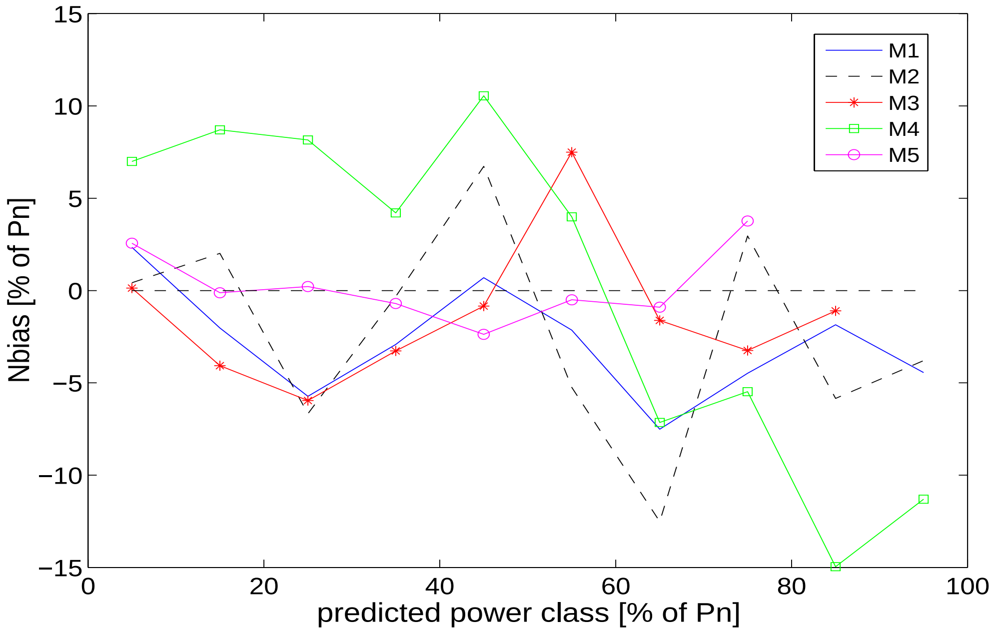


Figure E.17: *Normalized bias of the forecasting error distributions depending on the predicted power range. Focus is given to 18-hour ahead predictions.*

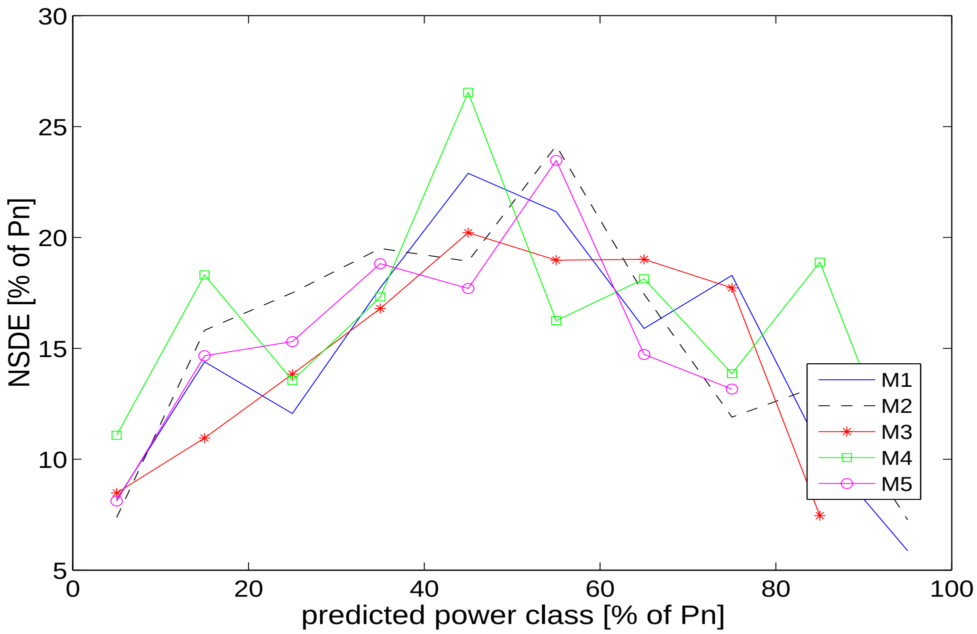


Figure E.18: *Normalized standard deviation of the forecasting error distributions depending on the predicted power range. Focus is given to 18-hour ahead predictions.*

E.5 Sotavento

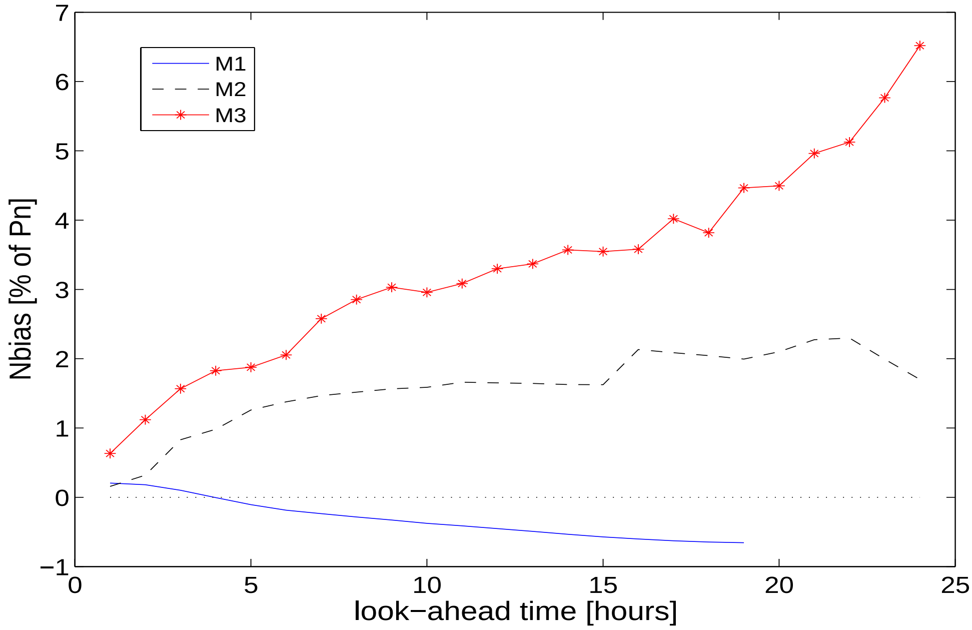


Figure E.19: *Performance evaluation of the forecasting methods with use of the Nbias measure as a function of the look-ahead time.*

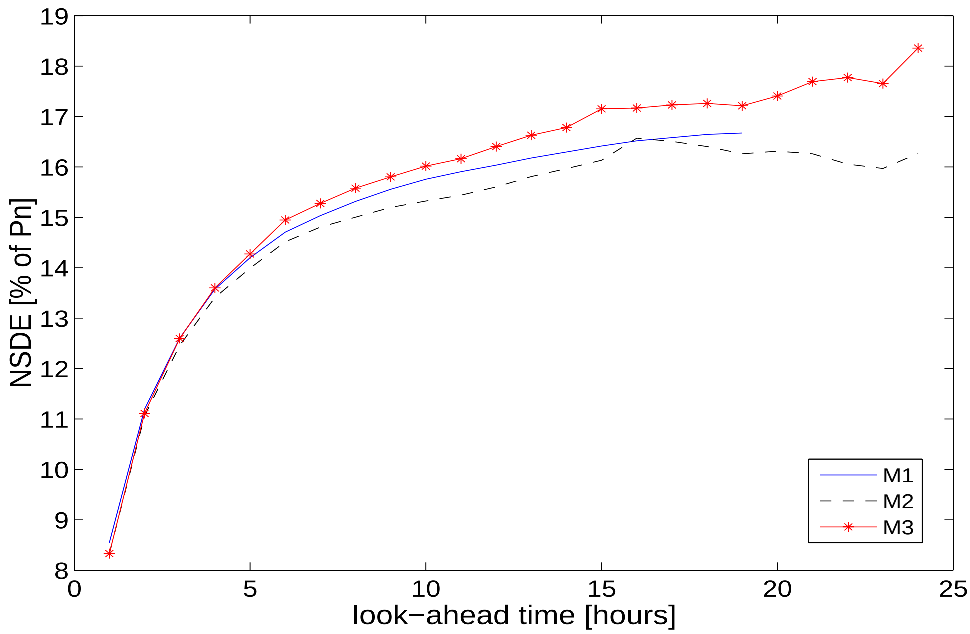


Figure E.20: *Performance evaluation of the forecasting methods with the NSDE measure as a function of the look-ahead time.*

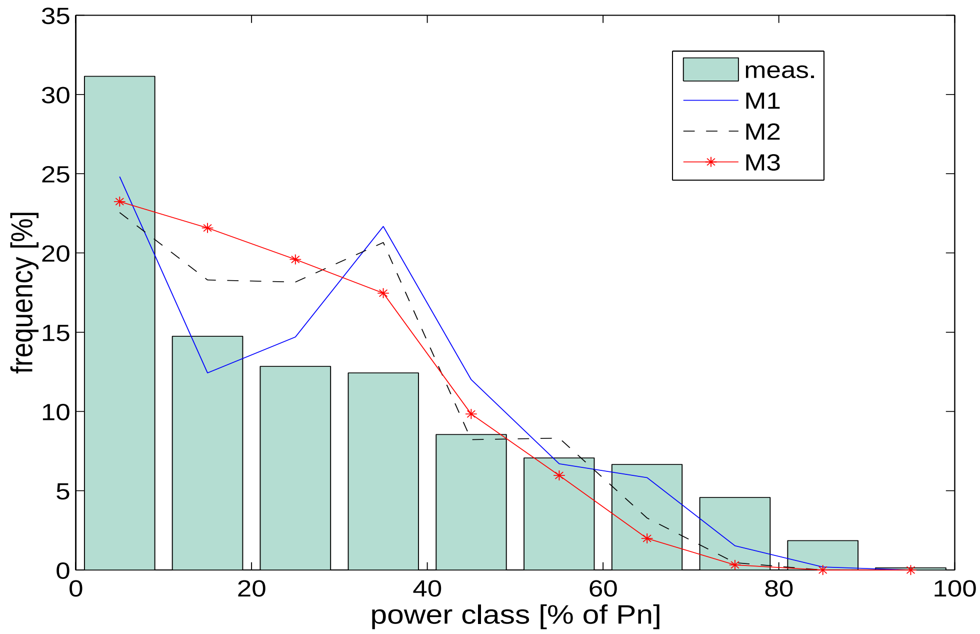


Figure E.21: *Power climatology of Sotavento compared with the climatologies of the five forecasting methods. Power measurements and forecasts are sorted with bins representing 10% of P_n. Focus is given to 12-hour ahead predictions.*

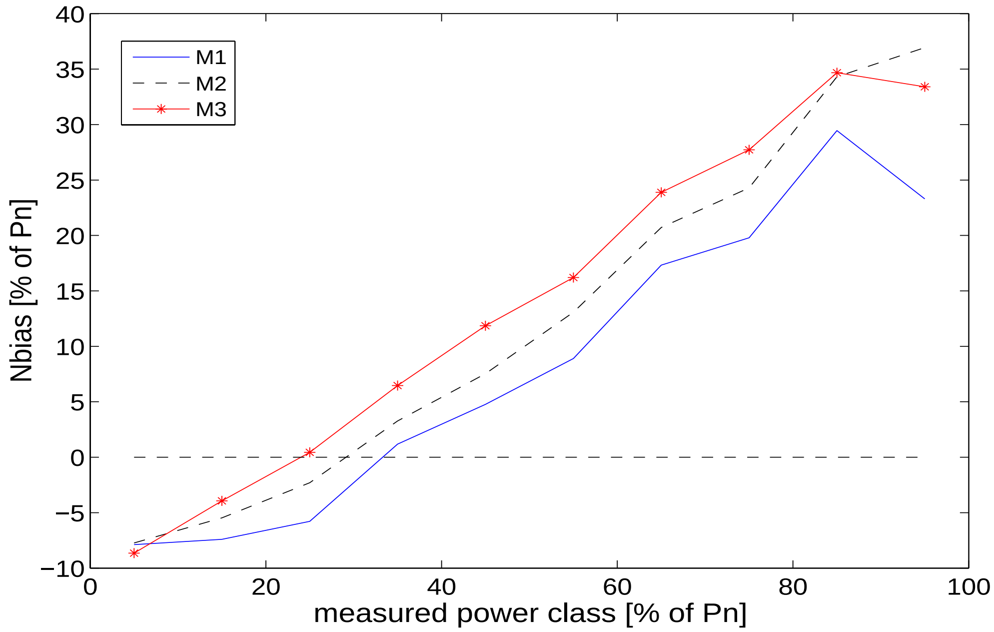


Figure E.22: *Normalized bias of the forecasting error distributions depending on power measures. Focus is given to 12-hour ahead predictions.*

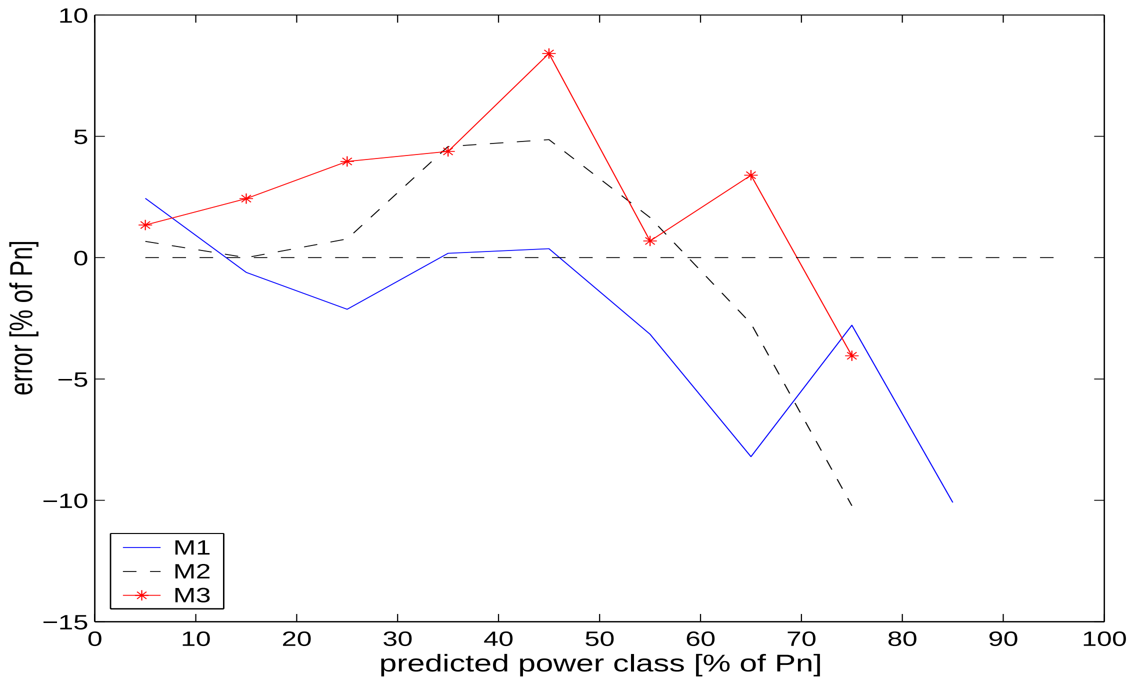


Figure E.23: *Normalized bias of the forecasting error distributions depending on the predicted power range. Focus is given to 12-hour ahead predictions.*

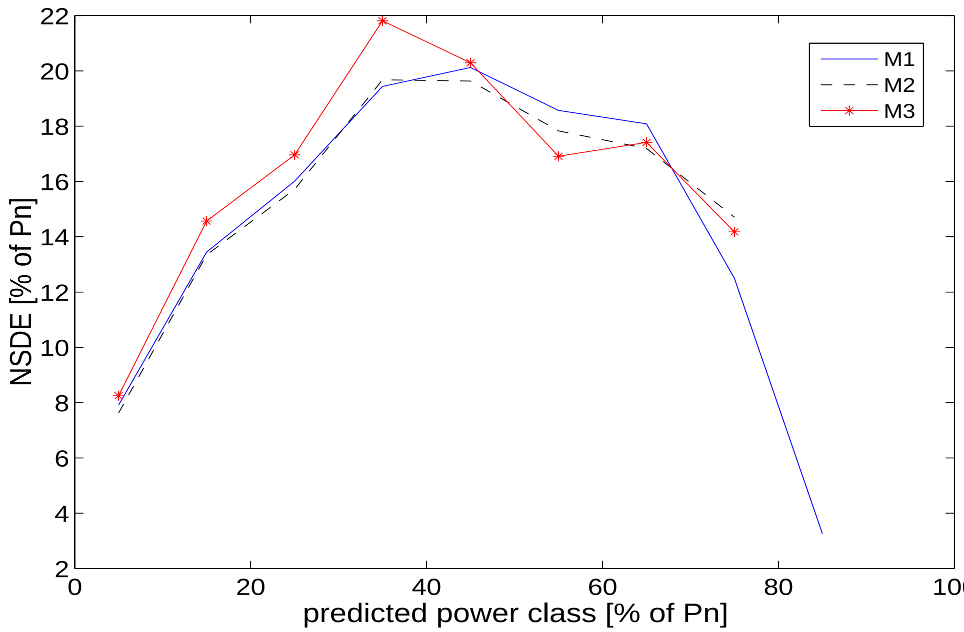


Figure E.24: *Normalized standard deviation of the forecasting error distributions depending on the predicted power range. Focus is given to 12-hour ahead predictions.*

Bibliography

[1] T. Ackermann, R. Leutz, and J. Hobolm. World-wide offshore wind potential and European projects. In *Power Engineering Society Summer Meeting, 2001 IEEE*, volume 1, pages 4–9, 2001.

[2] M. C. Alexiadis, P. S. Dokopoulos, and H. S. Sahsamanoglou. Wind speed and power forecasting based on spatial correlation models. *IEEE Trans. on Energy Conv.*, 14(3), September 1999.

[3] A. P. Alves da Silva and L. S. Moulin. Confidence intervals for neural network based short-term load forecasting. *IEEE Trans. on Power Syst.*, 15(4), November 2000.

[4] E. Anahua, F. Boettcher, S. Barth, J. Peinke, and M. Lange. Stochastic analysis of the power output for a wind turbine. In *CD-Proc. of the 2004 European Wind Energy Conference, EWEC'04, London, United Kingdom*, November 2004.

[5] F. Atger. The skill of ensemble prediction systems. *Mon. Wea. Rev.*, 127:1941–1957, September 1999.

[6] B. Bailey, M. C. Brower, and J. C. Zack. Short-term wind forecasting - Development and application of a mesoscale model. In *Proc. of the 1999 European Wind Energy Conference, EWEC'99, Nice, France*, pages 1062–1065, March 1999.

[7] R. T. Baillie and T. Bollerslev. Prediction in dynamic models with time-dependent conditional variances. *J. Econometrics*, 51:91–113, 1992.

[8] L. Balea, G. Kariniotakis, N. Siebert, and E. Peirano. Quantification of capacity credit and reserve demand from the large-scale integration of wind energy in the French power system. In *CD-Proc. of the 2004 Global Windpower Conference, Chicago, Illinois (USA)*, March 2004.

[9] T. Barker. The relationship between spread and error in extended range forecasts. *J. Climate*, 4:733–742, 1991.

[10] R. J. Barthelmie, S. C. Pryor, S. Frandsen, and S. E. Larsen. Analytical modelling of large wind farm clusters. In *Proc. of the 2004 EAWE Conference, European Academy for Wind Energy, 'The science of making torque from wind', Delft, The Netherlands*, April 2004.

[11] J. P. Barton and D. G. Ingfield. Energy storage and its use with intermittent renewable energy. *IEEE Trans. on Energy Conv.*, 19(2):441–447, June 2004.

[12] G. N. Bathurst and G. Strbac. Value of combining energy storage and wind in short-term energy and balancing markets. *Electric Power Systems Research*, 67:1–8, 2003.

[13] G. N. Bathurst, J. Weatherhill, and G. Strbac. Trading wind generation in short-term energy markets. *IEEE Trans. on Power Syst.*, 17(3):782–789, Aug. 2002.

[14] R. Belhomme. Wind power developments in France. In *2002 IEEE Power engineering society winter meeting*, volume 1, pages 357–358, 2002.

[15] K. P. Bennet and C. Campbell. Support Vector Machines: hype or hallelujah? *SIGKDD Explorations*, 2(2):1–13, December 2000.

[16] H. G. Beyer, D. Heinemann, H. Mellinghoff, K. Monnich, and H.-P. Waldl. Forecast of regional wind power output of wind turbines. In *Proc. of the 1999 European Wind Energy Conference, EWEC'99, Nice, France*, pages 1070–1073, March 1999.

[17] H. G. Beyer, T. Pahlke, W. Schmidt, H.-P. Waldl, and U. de Witt. Wake effects in a linear wind farm. *J. of Wind Eng. & Ind. Aerodyn.*, 51:303–318, 1994.

[18] H. Bindner and P. Lundsager. Integration of wind power in the power system. In *IECON 02, Industrial Electronics Society, IEEE, 28th Annual Conference on the*, volume 4, pages 3309–3316, 2002.

[19] A. Boogert and D. Y. Dupont. On the effectiveness of the anti-gaming policy between the day-ahead and real-time electricity markets in the Netherlands. *Energy Economics*, 27(5):752–770, September 2005.

[20] S. Bordignon and F. Lisi. Interval prediction for chaotic time-series. *Metron*, 59(3-4):117–140, 2001.

[21] G. E. P. Box and G. M. Jenkis. *Time-Series Analysis, Forecasting an Control.* Holden-Day, 1970.

[22] J. B. Bremnes. Probabilistic wind power forecasts using local quantile regression. *Wind Energy*, 7(1):47–54, January-March 2004.

[23] H. E. Brooks and C. A. Doswell. A comparison of measures-oriented and distributions-oriented approaches to forecast verification. *Wea. Forecasting*, 11:288–303, 1996.

[24] B. G. Brown, R. W. Watz, and A. H. Murphy. Time-series models to simulate and forecast wind speed and wind power. *J. App. Met.*, 23(8):1184–1195, 1984.

[25] K. J. Brundage, S. G. Benjamin, B. Schwartz, and M. N. Schwartz. Wind energy forecasts using time lagged ensembles. In *Proc. of the 21st Conference on Weather Analysis and Forecasting/17th Conference on Numerical Weather Prediction, Amer. Meteor. Soc.*, August 2005.

[26] K. J. Brundage, S. G. Benjamin, and M. N. Schwartz. Wind energy forecasts and ensemble uncertainty from the RUC. In *Proc. of the Ninth Conference on Mesoscale Processes, Fort Lauderdale, Florida (USA), Amer. Meteor. Soc.*, pages 343–345, 2001.

[27] C. Buckley, N. Scott, S. Snodin, and P. Gardner. Review of impacts of high wind penetration in electricity networks. Technical report, Garrad Hassan, 2004. Technical report 2299/PR/01 E.

[28] R. Buizza. Chaos and weather prediction. ECMWF, Meteorological Trainng Course Lecture Series, 2002.

[29] R. Buizza, P. L. Houtekamer, Z. Toth, G. Pellerin, M. Wei, and Y. Zhu. A comparison of the ECMWF, MSC and NCEP global ensemble prediction systems. *Mon. Wea. Rev.*, 133(5):1076–1097, 2005.

[30] R. Buizza, M. Miller, and T. N. Palmer. Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quat. J. Royal Met. Soc.*, 125:2887–2908, 1999.

[31] D. W. Bunn. Forecasting loads and prices in competitive power markets. *Proc. of the IEEE*, 88(2), February 2000.

[32] D. W. Bunn and E. D. Farmer. *Comparative Models for Electrical Load Forecasting*. John Wiley & Sons, 1985.

[33] L. Butler and K. Neuhoff. Comparison of feed in tariff, quota and auction mechanisms to support wind power development. Cambridge Working Papers in Economics CWPE 0503 - CMI Working Paper 70, Cambridge-MIT Institute, 2004.

[34] D. Cabezon, I. Marti, M. J. San Isidro, and I. Perez. Comparison of methods for power curve modelling. In *CD-Proc. of the 2004 Global Windpower Conference, Chicago, Illinois (USA)*, March 2004.

[35] J. G. Carney, P. Cunningham, and U Bhagwan. Confidence and prediction intervals for neural network ensembles. In *Proc. of the International Joint Conference on Neural Networks 1999*, 1999. paper 2090.

[36] E. D. Castronuovo and J. A. Peças Lopes. On the optimization of the daily operation of a wind-hydro power plant. *IEEE Trans. on Power Syst.*, 19(3):1599–1606, August 2004.

[37] B. Chabot. Bilan du développement de l'énergie éolienne en France en fin juin 2003. ADEME report, ADEME, 2003. In French.

[38] B. Chabot. Éolien: état des lieux et perspectives. *Systèmes Solaires*, 155:66–73, May-June 2003. In French.

[39] W. S. Chan, S. H. Cheung, and K. H. Wu. Multiple forecasts with autoregressive time-series models: case-studies. *Mathematics and Computers in Simulation*, 64:421–430, 2004.

[40] J. G. Charney, R. Fjørtoft, and J. von Neumann. Numerical integration of the barotropic vorticity equation. *Tellus*, 2:237–254, 1950.

[41] C. Chatfied. *Time-Series Forecasting*. Chapman & Hall/CRC, 2000.

[42] C. Chatfied. Confessions of a pragmatic statistician. *The Statistician*, 51:1–20, 2002.

[43] C. Chevallier. Développement et évaluation de stratégies de prédiction éolienne en vue de la participation à un marché de l'électricité dérégulé. Master's thesis, École des Mines de Paris, Paris, France, 2004. in French.

[44] P. F. Christoffersen. Evaluating interval forecasts. *Intern. Econom. Rev.*, 39(4):841–862, 1998.

[45] R. T. Clemen and R. L. Winkler. Combining probability distributions from experts in risk analysis. *Risk Analysis*, 19(2):187–203, 1999.

[46] M. P. Clements. Editorial - Some possible directions for future research. *Int. J. Forecasting*, 19(1):1–3, 2003.

[47] M. P. Clements. *Evaluating Econometric Forecasts of Economic and Financial Values*. Palgrave Macmillan, 2005.

[48] M. P. Clements and D. F. Hendry. *A Companion to Economic Forecasting*. Blackwells, 1998.

[49] M. P. Clements and N. Taylor. Bootstrapping prediction intervals for autoregressive models. *Int. J. Forecasting*, 17(2):247–267, 2001.

[50] A. J. Conejo, J. Contreras, R. Espínola, and M. A. Plazas. Forecasting electricity prices for a day-ahead pool-based electricity energy market. *Int. J. Forecasting*, 21(3):435–462, 2005.

[51] A. Crespo, J. Hernández, and S. Frandsen. Survey of modelling methods for wind turbine wakes and wind farms. *Wind Energy*, 2(1):1–24, 1999.

[52] N. Cristianini and J. Shawe-Taylor. *In Introduction to Support Vector Machines and other Kernel-based Learning Methods*. Cambridge University Press, 2000.

[53] I. G. Damousis, M. C. Alexiadis, J. B. Theocharis, and P. S. Dokopoulos. A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation. *IEEE Trans. on Energy Conv.*, 19(2):352–361, June 2004.

[54] R. De Veaux, J. Schumi, J. Schweinsberg, D. Shellington, and L. H. Ungar. Prediction intervals for neural networks via nonlinear regression. *J. Amer. Stat. Ass.*, 40(4):273–282, November 1998.

[55] E. DeMeo. Implications of wind energy scheduling requirements for wind energy. NWCC Transmission Issue Briefs, National Wind Coordinating Committee (NWCC), US, available online: www.nationalwind.org, August 2003.

[56] R. Doherty, E. Denny, and M. O'Malley. System operation with a significant wind power penetration. In *Proc. of the IEEE Power Eng. Soc. General Meeting, Denver, Colorado (USA)*, volume 1, pages 1002–1007, June 2004.

[57] R. Doherty and M. O'Malley. A new approach to quantify reserve demand in systems with significant installed wind capacity. *IEEE Trans. on Power Syst.*, 20(2):587–595, May 2005.

[58] S. D. Dubey. Normal and weibull distributions. *Naval Res. Logistics Quart.*, 14:69–79, 1967.

[59] R. Dybowski and S. J. Roberts. Confidence intervals and prediction intervals for feed-forward neural networks. In R. Dybowski and V. Gant, editors, *Clinical Applications of Artificial Neural Networks*. Cambridge University Press, 2000.

[60] B. Efron. Bootstrap methods: another look at the jackknike. *Ann. Statist.*, 7:1–26, 1979.

[61] B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall/CRC, 1993.

[62] S. Engstrom, H. Ganander, and R. Lindstrom. Short-term variations in the output of wind turbines. *DEWI Magazine*, 19:27–30, 2001.

[63] S. Enomoto and co authors. Prediction of power output from wind farm using local meteorological analysis. In *Proc. of the 2001 European Wind Energy Conference, EWEC'01, Copenhagen, Denmark*, pages 822–825, June 2001.

[64] C. Ensslin, B. Ernst, K. Rohrig, and F. Schlogl. On-line monitoring and prediction of wind power in German transmission system operation center. In *CD-Proc. of the 2003 European Wind Energy Conference, EWEC'03, Madrid, Spain*, June 2003.

[65] B. Ernst, K. Rohrig, P. Schorn, and H. Regber. Managing 3000 MW power in a transmission system operation center. In *Proc. of the 2001 European Wind Energy Conference, EWEC'01, Copenhagen, Denmark*, pages 890–893, June 2001.

[66] European Commission. Directive 2001/77/EC of the European Parliament and of the Council, on the promotion of electricity produced from renewable energy sources in the internal electricity market. Official Journal of the European Commission, September 2001.

[67] A. Fabbri, T. G. Gómez San Román, J. R. Rivier Abbad, and V. H. Méndez Quezada. Assessment of the cost associated with wind generation prediction errors in a liberalized electricity market. *IEEE Trans. on Power Syst.*, 20(3):1440–1446, 2005.

[68] U. Focken, M. Lange, K. Monnich, H.-P. Waldl, H. G. Beyer, and A. Luig. Short-term prediction of the aggregated power output of wind farms - A statistical analysis of the reduction of the prediction error by spatial smoothing effects. *J. of Wind Eng. & Ind. Aerodyn.*, 90:231–246, 2002.

[69] U. Focken, M. Lange, and H.-P. Waldl. Previento - A wind power prediction system with an innovative upscaling algorithm. In *Proc. of the 2001 European Wind Energy Conference, EWEC'01, Copenhagen, Denmark*, pages 826–829, June 2001.

[70] U. Focken, M. Lange, and H.-P. Waldl. Previento - Regional wind power prediction with risk control. In *CD-Proc. of the 2002 Global Windpower Conference, Paris, France*, April 2002.

[71] H. Fukuda. The development of a wind velocity prediction method based on a data-mining type auto-regressive model. In *Proc. of the 2001 European Wind Energy Conference, EWEC'01, Copenhagen, Denmark*, pages 741–744, June 2001.

[72] C. García Barquero and P. de Arriba Segurado. Case study # 12: Spanish Royal Decree 2818/1998 - 'Special regime of the electricity market'. REACT - Renewable Energy Action, Altener AL/2002/157, October 2004.

[73] N. Gasset, G. J. Poitras, Y. Gagnon, and C. Brothers. Atmospheric boundary layer flow over a coastal cliff. In *CD-Proc. of the 2003 European Wind Energy Conference, EWEC'03, Madrid, Spain*, June 2003.

[74] C. Genest and K. J. McConway. Allocating the weights in the linear opinion pool. *J. Forecasting*, 9:53–73, 1990.

[75] G. Giebel. *On the benefits of distributed generation of wind energy in Europe*. PhD thesis, University Carl von Ossietzky, Oldenburg, Germany, 2000.

[76] G. Giebel, R. Brownsword, and G. Kariniotakis. State of the art on short-term wind power prediction. ANEMOS Deliverable Report D1.1, available online: http://anemos.cma.fr, June 2003.

[77] G. Giebel, L. Landberg, J. Badger, K. Sattler, H. Feddersen, T. S. Nielsen, H. Aa. Nielsen, and H. Madsen. Using ensemble forecasting for wind power. In *CD-Proc. of the 2003 European Wind Energy Conference, EWEC'03, Madrid, Spain*, June 2003.

[78] G. Giebel, L. Landberg, G. Kariniotakis, and R. Brownsword. State-of-the-art on methods and software tools for short-term prediction of wind energy production. In *CD-Proc. of the 2003 European Wind Energy Conference, EWEC'03, Madrid, Spain*, June 2003.

[79] G. Giebel, L. Landberg, and T. S. Nielsen. The ZEPHYR-project: The next generation prediction system. In *Proc. of the 2001 European Wind Energy Conference, EWEC'01, Copenhagen, Denmark*, pages 777–780, June 2001.

[80] T. Gneiting, F. Balabdaoui, and A. E. Raftery. Probabilistic forecasts, calibration and sharpness. Technical report, University of Washington, Department of Statistics, May 2005. Technical report no. 483.

[81] T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. Technical report, University of Washington, Department of Statistics, 2004. Technical report no. 463.

[82] T. Gneiting and A. E. Raftery. Weather forecasting with ensemble methods. *Science*, 310(5746):248–249, 2005.

[83] M. Grigoletto. Bootstrapping prediction intervals for autoregressions: some alternatives. *Int. J. Forecasting*, 14(4):447–456, 1998.

[84] E. P. Grimit. Redefining the ensemble spread-skill relationship from a probabilistic perspective. NCEP invited presentation, Camp Spring, Maryland (USA), April 2004.

[85] E. P. Grimit and C. F. Mass. Initial results of a mesoscale short-range ensemble forecasting system over the Pacific Northwest. *Wea. Forecasting*, 132:192–204, 2002.

[86] E. P. Grimit and C. F. Mass. Measuring the ensemble spread-skill reliationship from a probabilistic perspective: stochastic ensemble results. submitted to Monthly Weather Review, June 2005.

[87] G. J. Hahn and W. Q. Meeker. *Statistical Intervals - A Guide for Practitioners*. John Wiley & Sons, 1991.

[88] P. Hall and A. Rieck. Improving coverage accuracy of nonparametric prediction intervals. *J. Royal Stat. Soc.*, 63(4):717–725, 2001.

[89] T. M. Hamill. Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, 129, 2000. Notes and Correspondence.

[90] T. M. Hamill. Evaluating forecaster's rules of thumb: a study of $d(\text{prog})/dt$. *Wea. Forecasting*, 18:933–937, 2003. Notes and Correspondence.

[91] N. Hatziargyriou and A. Zervos. Wind power developments in Europe. *Proc. of the IEEE Special Issue*, 89(12):1765–1782, 2001.

[92] T. Heskes. Practical confidence and prediction intervals. In M. Mozer, M. Jordan, and T. Petsche, editors, *Advances in Neural Information Processing Systems 9*. MIT Press, Cambridge, 1997.

[93] M. Hibon and T. Evgeniou. To combine or not to combine: selecting among forecasts and their combinations. *Int. J. Forecasting*, 21(3):15–24, 2005.

[94] B. F. Hobbs, S. Jitprapaikulsarn, S. Konda, V. Chankong, K. A. Loparo, and D. J. Maratukulam. Analysis of the value for unit commitment of improved load forecasts. *IEEE Trans. on Power Syst.*, 14(4), November 1999.

[95] R. N. Hoffman and E. Kalnay. Lagged average forecasting, an alternative to monte-carlo forecasting. *Tellus*, 35A:100–118, 1983.

[96] J. R. Holton. *An Introduction to Dynamic Meteorology*. International Geophysics Series. Academic Press, 1992.

[97] H. Holttinen. *The impact of large scale wind power production on the Nordic power system*. PhD thesis, Helsinki University of Technology, Espoo, Finland, 2004.

[98] H. Holttinen. Optimal electricity market for wind power. *Energy Policy*, 33(16):2052–2063, November 2004.

[99] H. Holttinen and R. Hirvonen. Power system requirements for wind power. In T. Ackermann, editor, *Wind Power in Power Systems*. John Wiley & Sons, 2005.

[100] P. L. Houtekamer. Global and local skill forecasts. *Mon. Wea. Rev.*, 121:1834–1846, 1993.

[101] J. J. G. Hwang and A. A. Ding. Prediction intervals for artificial neural networks. *J. Amer. Stat. Ass.*, 92:748–757, 1997.

[102] R. J. Hyndman. Highest-density forecast regions for non-linear and non-normal time-series models. *J. Forecasting*, 14:431–441, 1995.

[103] IEA. Variability of wind power and other renewables. Technical report, IEA (International Energy Agency), 2005. Paris: IEA/OCDE.

[104] J. U. Jørgensen and C. Möhrlen. Verification of ensemble prediction systems for a new market: wind energy. Technical report, Sustainable Energy Research Group, University College Cork, Ireland, June 2003. ECMWF special project interim report 2.

[105] J. U. Jørgensen, C. Möhrlen, B. O'Gallachoir, K. Sattler, and E. J. McKeogh. HIRPOM: description of an operational numerical wind power prediction model for large-scale integration of on- and offshore wind power in Denmark. In *CD-Proc. of the 2002 Global Windpower Conference, Paris, France*, April 2002.

[106] M. Jørgensen and D. I. K. Sjøberg. An effort prediction interval approach based on the empirical distribution of previous estimation accuracy. *Information & Software Technology*, 45:123–136, 2003.

[107] C. G. Justus, W. R. Hargraves, A. Mikhail, and D. Graber. Methods for estimating wind speed frequency distribution. *J. App. Met.*, 17(3):350–353, 1978.

[108] R. E. Kalman. A new approach to linear filtering and prediction problems. *Trans. ASME, J. Basic Eng.*, 82(Ser. D):35–45, 1960.

[109] E. Kalnay, S. J. Lord, and R. D. McPherson. Maturity of operational numerical weather prediction: medium range. *Bull. Am. Met. Soc.*, 79(12):2753–2769, 1998.

[110] G. Kariniotakis. *Contribution to the development of an advanced control system for autonomous wind-diesel power systems.* PhD thesis, Ecole des Mines de Paris, France, 1996.

[111] G. Kariniotakis and co authors. What performance can be expected by short-term wind power prediction models depending on site characteristics? In *CD-Proc. of the 2004 European Wind Energy Conference, EWEC'04, London, United Kingdom*, November 2004.

[112] G. Kariniotakis and D. Mayer. An advanced on-line wind resource prediction system for the optimal management of wind parks. In *CD-Proceedings of the MedPower Conference, Athens, Greece*, November 2002.

[113] G. Kariniotakis, E. Nogaret, and A. G. Dutton. Wind power forecasting using advanced neural networks models. *IEEE Trans. on Energy Conv.*, 11(4):762–767, December 1996.

[114] G. Kariniotakis, E. Nogaret, A. G. Dutton, J. A. Halliday, and A. Androutsos. Evaluation of advanced wind power and load forecasting methods for the optimal management of isolated power systems. In *Proc. of the 1999 European Wind Energy Conference, EWEC'99, Nice, France*, pages 1082–1085, March 1999.

[115] G. Kariniotakis and P. Pinson. Evaluation of the MORE-CARE wind power prediction platform. Performance of the fuzzy logic based models. In *CD-Proc. of the 2003 European Wind Energy Conference, EWEC'03, Madrid, Spain*, June 2003.

[116] G. Kariniotakis and P. Pinson. Uncertainty of short-term wind power forcasts - A methodology for on-line assessment. In *Proc. of the 2004 PMAPS Conference, 'Probabilistic Methods Applied to Power Systems', IEEE PMAPS 2004, Ames, Iowa (USA)*, pages 729–736, September 2004.

[117] G. Kariniotakis, P. Pinson, and A. G. Dutton. MORE-CARE evaluation report - Evaluation results for the case study of Ireland. MORE-CARE European Project, Deliverable report D13, May 2003.

[118] N. K. Kasabov and Q. Song. DENFIS: Dynamic Evolving Neural-Fuzzy Inference System and its application for time-series prediction. *IEEE Trans. on Fuzzy Systems*, 10(2), April 2002.

[119] I. Katic, J. Højstrup, and N. O. Jensen. A simple model for cluster efficiency. In W. Palz and E. Sesto, editors, *Proc. of the European Wind Energy Conference and Exhibition*, volume 1, pages 407–410. A. Raguzzi, 1986.

[120] A. B. Koehler. An inappropriate prediction interval. *Int. J. Forecasting*, 6(4):557–558, 1990.

[121] M. Korpaas, A. T. Holen, and R. Hildrum. Operation and sizing of energy storage for wind power plants in a market system. *Electrical Power and Energy Systems*, 25:599–606, 2003.

[122] B. B. Kristoffersen, B. Donslun, and P. Børre Eriksen. Impacts of large-scale wind power on the power market. In *Proc. of the Fourth International Workshop on Large-Scale Integration of Wind Power and Transmission Networks for Offshore Wind Farms, Billund, Denmark*, Oct. 2003.

[123] J.-P. Lam and M. R. Veall. Bootstrap prediction intervals for single period regression forecasts. *Int. J. Forecasting*, 18(1):125–130, 2002.

[124] L. Landberg. A mathematical look at a physical power prediction model. *Wind Energy*, 1:23–28, 1998.

[125] L. Landberg. Short-term prediction of the power production from wind farms. *J. of Wind Eng. & Ind. Aerodyn.*, 80:207–220, 1999.

[126] L. Landberg, G. Giebel, L. Myllerup, J. Badger, H. Madsen, and T. S. Nielsen. Poor-man's ensemble forecasting for error estimation. In *Proc. of the 2002 American Wind Energy Association conference, Portland, Oregon*, June 2002.

[127] L. Landberg and S. J. Watson. Short-term prediction of local wind conditions. *Boundary-Layer Meteorology*, 70:171–195, 1994.

[128] B. Lange, S. E. Larsen, J. Højstrup, and R. J. Barthelmie. Importance of thermal effects and sea surface roughness for offshore wind ressource assessment. *J. of Wind Eng. & Ind. Aerodyn.*, 92:959–988, 2004.

[129] M. Lange. *Analysis of the uncertainty of wind power predictions*. PhD thesis, University Carl von Ossietzky, Oldenburg, Germany, 2003.

[130] M. Lange. On the uncertainty of wind power predictions - Analysis of the forecast accuracy and statistical distribution of errors. *Trans. of the ASME, J. Solar Energy Eng.*, 127(2):177–194, May 2005.

[131] M. Lange and D. Heinemann. Relating the uncertainty of short-term wind speed prediction to meterological situations with methods from synoptic climatology. In *CD-Proc. of the 2003 European Wind Energy Conference, EWEC'03, Madrid, Spain*, June 2003.

[132] M. Lange, J. Tambke, U. Focken, J.-O. Wolff, and J. A. T. Bye. Forecasting offshore wind power. In *Proc. of the 2004 EAWE Conference, European Academy for Wind Energy, 'The science of making torque from wind', Delft, The Netherlands*, April 2004.

[133] M. Lange and H.-P. Waldl. Accuracy of short-term wind power prediction depending on meteorological conditions. In *CD-Proc. of the 2002 Global Windpower Conference, Paris, France*, April 2002.

[134] K. Larson and T. Gneiting. Advanced short-range wind energy forecasting technologies - Challenges, solutions and validation. In *CD-Proc. of the 2004 Global Windpower Conference, Chicago, Illinois (USA)*, March 2004.

[135] K. Larson and K. Westrick. Short-term wind power forecasting using off-site observations. *Wind Energy*, 9((1-2)):55–62, 2005.

[136] C. E. Leith. Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, 102:409–418, 1974.

[137] S. Li, D. C. Wunsch, and E. A. O'Hair. Using neural networks to estimate wind power turbine generation. *IEEE Trans. on Energy Conv.*, 16(3):276–282, September 2001.

[138] U. Linnet. Tools supporting wind energy trade in deregulated markets. Master's thesis, Informatics and Mathematical Modeling, Technical University of Denmark, Kgs. Lyngby, Denmark, 2005. IMM-Thesis-2005-56.

[139] E. N. Lorenz. A study of the predictability of a 28-variable atmospheric model. *Tellus*, 17:321–333, 1965.

[140] E. N. Lorenz. The predictability of a flow which possesses many scales of motion. *Tellus*, 21:289–307, 1968.

[141] P. Louka, G. Galanis, N. Siebert, G. Kariniotakis, P. Katsafados, G. Kallos, and I. Pytharoulis. Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering. In *Proc. of the 5^{th} Conference on Mathematical Methods in Science & Engineering, 2005 CMMSE*, June 2005.

[142] A. Luig, S. Bofinger, and H. G. Beyer. Analysis of confidence intervals for the prediction of regional wind power output. In *Proc. of the 2001 European Wind Energy Conference, EWEC'01, Copenhagen, Denmark*, pages 725–728, June 2001.

[143] H. Lund and E. Munster. Management of surplus electricity-production from a fluctuating renewable-energy source. *Applied Energy*, 76:65–74, 2003.

[144] R. Madlener and M. Kaufmann. Power exchange spot market trading in europe: theoretical considerations and empirical evidence. OSCOGEN (Optimisation of Cogeneration Systems in a Competitive Market Environment)- Project Deliverable 5.1b, March 2002.

[145] H. Madsen, P. Pinson, H. Aa. Nielsen, T. S. Nielsen, and G. Kariniotakis. Standardizing the performance evaluation of short-term wind power prediction models. *Wind Engineering*, 29(6):475–489, 2005.

[146] M. Magnusson and L. Wern. Wind energy predictions using CFD and HIRLAM forecast. In *Proc. of the 2001 European Wind Energy Conference, EWEC'01, Copenhagen, Denmark*, pages 861–863, June 2001.

[147] S. Makridakis, M. Hibon, E. Lusk, and M. Belhadjali. Confidence intervals - An empirical investigation of the series in the M-competitions. *Int. J. Forecasting*, 3(3-4):489–508, 1987.

[148] S. Makridakis, S. C. Wheelwright, and V. E. McGee. *Forecasting: Methods and Applications*. John Wiley & Sons, 2nd edition, 1983.

[149] I. Marti and H. Madsen. Prediction models in complex terrain. In *Proc. of the 2001 European Wind Energy Conference, EWEC'01, Copenhagen, Denmark*, pages 875–878, June 2001.

[150] J. Mason. Definition of technical terms in forecast verification and examples of forecast verification scores. In *Proc. of the IRI Workshop on Forecast Quality, New York*, October 2000.

[151] J. McCann. Operating reserve requirements as wind power penetration increases in the Irish electricity system. SEI report 04-RERDD-011-R-01, Sustainable Energy Ireland (SEI), 2004. available online: www.irish-energy.ie.

[152] E. McCarthy. Wind speed forecasting in the central California wind resource area. In *EPRI-DOE-NREL WInd Energy Forecasting Meeting, Burlingame, California (USA)*, March 1998.

[153] S. McNees. Forecast uncertainty: can it be measured? mimeo, Federal Reserve Bank of Boston, Boston, Massachusetts (USA), 1995.

[154] K. Methaprayoon, W. J. Lee, C. Yingvivatanapong, and J. Liao. An integration of ANN wind power estimation into UC considering forecast uncertainty. In *Proc. of the Industrial and Commercial Power Systems Technical Conference, 2005 IEEE*, pages 116–124, May 2005.

[155] D. Milborrow. Penalties for intermittent energy sources. Working paper for the PIU Energy Review - available online: www.cabinet-office.gov.uk/innovation/2002/energy/workingpapers, December 2001.

[156] V. Miranda and L. M.. Proenca. Probabilistic choice vs. risk analysis - Conflicts and synthesis in power system planning. *IEEE Trans. on Power Syst.*, 13(3):1038–1043, 1998.

[157] C. Möhrlen. On the accuracy of land cover data in NWP forecasts for high resolution wind energy prediction. In *Proc. of the 2001 European Wind Energy Conference, EWEC'01, Copenhagen, Denmark*, pages 854–857, June 2001.

[158] C. Möhrlen. *Uncertainty in wind energy forecasting*. PhD thesis, University College Cork, National University of Ireland, Cork, Ireland, 2004.

[159] C. Möhrlen and J. U. Jørgensen. Verification of ensemble prediction systems for a new market: wind energy. Technical report, Sustainable Energy Research Group, University College Cork, Ireland, June 2005. ECMWF special project interim report 3.

[160] C. Möhrlen, J. U. Jørgensen, and E. J. McKeogh. Power predictions in complex terrain with an operational numerical weather prediction model in Ireland including ensemble forecasting. In *Proc. of the 2002 World Wind Energy Conference, Berlin, Germany*, June 2002.

[161] F. Molteni, R. Buizza, T. N. Palmer, and T. Petroliagis. The ECMWF ensemble prediction system: methodology and validation. *Quat. J. Royal Met. Soc.*, 122:73–119, 1996.

[162] D. Moon, D. Christenson, and R. Chevallaz-Perrier. Support Vector Machines technology coupled with physics-based modelling for wind facility power production forecasting. In *CD-Proc. of the 2004 Global Windpower Conference, Chicago, Illinois (USA)*, March 2004.

[163] A. Moore and R. Kleeman. Skill assessment for ENSO using ensemble prediction. *Quat. J. Royal Met. Soc.*, 124:557–584, 1998.

[164] P. E. Morthorst. Wind power and the conditions at a liberalized power market. *Wind Energy*, 6:297–308, July-September 2003.

[165] A. H. Murphy. What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, 8(2):281–293, 1993.

[166] A. H. Murphy and E. S. Epstein. Skill scores and correlation coefficients in model verification. *Mon. Wea. Rev.*, 117:784–794, 1989.

[167] A. H. Murphy and R. L. Winkler. A general framework for forecast verification. *Mon. Wea. Rev.*, 115:1330–1338, 1987.

[168] J. M. Murphy. The impact of ensemble forecasts on predictability. *Quat. J. Royal Met. Soc.*, 114:463–493, 1988.

[169] C. S. Nielsen and H. F. Ravn. Criteria in short-term wind power prognosis. In *CD-Proc. of the 2003 European Wind Energy Conference, EWEC'03, Madrid, Spain*, June 2003.

[170] H. Aa. Nielsen. Analysis and simulation of prediction errors for wind power productions reported to nord pool. IMM-Eltra project report, Informatics and Mathematical Modeling, Technical University of Denmark, Denmark, 2002. in Danish.

[171] H. Aa. Nielsen, T. S. Nielsen, and H. Madsen. Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. In *Proc. of the 2004 European Wind Energy Conference, EWEC'04, Scientific Track, London, United Kingdom*, pages 34–38, November 2004.

[172] H. Aa. Nielsen, T. S. Nielsen, H. Madsen, and K. Sattler. Wind power ensemble forecasting. In *CD-Proc. of the 2004 Global Windpower Conference, Chicago, Illinois (USA)*, March 2004.

[173] H. Aa. Nielsen, D. Yates, H. Madsen, T. S. Nielsen, J. Badger, G. Giebel, L. Landberg, K. Sattler, and H. Feddersen. Analysis of the results of an on-line wind power quantile forecasting system. Technical report, Informatics and Mathematical Modeling, Danish Technical University, October 2005. PSO (FU 2101) project report.

[174] L. H. Nielsen and co authors. Wind power and a liberalised North European electricity exchange. In *Proc. of the 1999 European Wind Energy Conference, EWEC'99, Nice, France*, pages 379–382, March 1999.

[175] T. S. Nielsen, A. Joensen, H. Madsen, L. Landberg, and G. Giebel. A new reference model for wind power forecasting. *Wind Energy*, 1:29–34, 1998.

[176] T. S. Nielsen and H. Madsen. ZEPHYR - The prediction models. In *Proc. of the 2001 European Wind Energy Conference, EWEC'01, Copenhagen, Denmark*, pages 868–871, June 2001.

[177] T. S. Nielsen and H. Madsen. Prediction of regional wind power. In *CD-Proc. of the 2002 Global Windpower Conference, Paris, France*, April 2002.

[178] T. S. Nielsen, H. Aa. Nielsen, and H. Madsen. Prediction of wind power using time-varying coefficient-functions. In *Proc. of the 15th IFAC World Congress, Barcelona, Spain*, July 2002. Elsevier, ISBN: 0-08-044295-1.

[179] J.-M. Noel and R. Chevallaz-Perrier. Wind resource assessment with AriaWind in a complex site. In *Proc. of the 2001 European Wind Energy Conference, EWEC'01, Copenhagen, Denmark*, pages 872–874, June 2001.

[180] M. Olsson and L. Söder. Generation of regulating power pice scenarios. In *Proc. of the 2004 PMAPS Conference, 'Probabilistic Methods Applied to Power Systems', IEEE PMAPS 2004, Ames, Iowa (USA)*, pages 26–31, September 2004.

[181] D. Orrell, L. A. Smith, J. Barkmeijer, and T. N. Palmer. Model error and operational weather forecasts. *Nonlin. Proc. Geophys.*, 8(6):357–371, November 2001.

[182] T. N. Palmer. Predicting uncertainty in forecasts of weather and climate. *Rep. Prog. Phys.*, 63(2):71–116, 2000.

[183] P. Pinson and G. Kariniotakis. Wind power forecasting using fuzzy neural networks enhanced with on-line prediction risk assessment. In *CD-Proc. of the 2003 Bologna Power Tech Conference, IEEE Power Tech 2003, Bologna, Italy*, June 2003.

[184] P. Pinson and G. Kariniotakis. On-line adaptation of confidence intervals based on weather stability for wind power forecasting. In *CD-Proc. of the 2004 Global Windpower Conference, Chicago, Illinois (USA)*, March 2004.

[185] P. Pinson, G. Kariniotakis, and D. Mayer. Uncertainty and prediction risk assessment of short-term wind power forecasts. In *Proc. of the 2004 EAWE Conference, European Academy for Wind Energy, 'The science of making torque from wind', Delft, The Netherlands*, April 2004.

[186] P. Pinson, H. Aa. Nielsen, T. S. Nielsen, H. Madsen, and G. Kariniotakis. Properties of quantile and interval forecasts of wind generation and their evaluation. In *Proc. of the 2006 European Wind Energy Conference, EWEC'06, Scientific Track, Athens, Greece*, February 2006.

[187] P. Pinson, T. Ranchin, and G. Kariniotakis. Short-term wind power prediction for offshore wind farms - Evaluation of fuzzy-neural network based models. In *CD-Proc. of the 2004 Global Windpower Conference, Chicago, Illinois (USA)*, March 2004.

[188] P. Pinson, N. Siebert, and G. Kariniotakis. Forecasting of regional wind generation by a dynamic fuzzy-neural networks based upscaling approach. In *CD-Proc. of the 2003 European Wind Energy Conference, EWEC'03, Madrid, Spain*, June 2003.

[189] S. C. Pryor and R. J. Barthelmie. Persistence of offshore winds: implications for power quality. In *Proc. of the 2001 European Wind Energy Conference, EWEC'01, Copenhagen, Denmark*, pages 717–720, June 2001.

[190] A. Rauh and J. Peinke. A phenomenological model for the dynamic response of wind turbines to turbulent wind. *J. of Wind Eng. & Ind. Aerodyn.*, 92:159–183, 2004.

[191] N. Ravishankar, L. Shiao-Yen Wu, and J. Glaz. Multiple prediction intervals for time-series: comparison of simultaneous and marginal intervals. *J. Forecasting*, 10:445–463, 1991.

[192] J. J. Reeves. Bootstrap prediction intervals for ARCH models. *Int. J. Forecasting*, 20(2):237–248, 2004.

[193] P. J. Roebber. Variability in successive operational model forecasts of maritime cyclogenesis. *Wea. Forecasting*, 5:586–595, 1990.

[194] J.-S. Roger Jang. ANFIS: Adaptive-Network-based Fuzzy Inference System. *IEEE Trans. on systems, man, and cybernetics*, 23(3), May-June 1993.

[195] C.-G. Rossby and co authors. Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacements of the semi-permanent centers of action. *J. Mar. Res*, 2:38–55, 1939.

[196] M. S. Roulston, D. T. Kaplan, J. Hardenberg, and L. A. Smith. Using medium-range weather forecasts to improve the value of wind energy production. *Renewable Energy*, 28:585–602, 2003.

[197] M. S. Roulston and L. A. Smith. Evaluating probabilistic forecasts using information theory. *Mon. Wea. Rev.*, 130:1653–1660, June 2002. Notes and Correspondence.

[198] Y.-M. Saint Drenant. Wind power predictions analysis - Part 1: TenneT imbalance price system. Development of a model for TenneT imbalance price system. ECN report - ECN-I-02-010, Energy research Center of the Netherlands, 2002.

[199] I. Sanchez. Short-term prediction of wind energy production. *Int. J. Forecasting*, 22(1):43–56, 2006.

[200] D. W. Scott. *Multivariate Density Estimation: Theory, Practice and Visualization.* John Wiley & Sons, New York, 1992.

[201] M. Sewalt and C. de Jong. Negative prices in electricity markets. Commodities Now - Online journal: www.commodities-now.com, June 2003.

[202] A. Sfetsos. *Time-series forecasting of wind speed and solar radiation for renewable energy sources.* PhD thesis, Imperial College, London, UK, 1999.

[203] A. Sfetsos. A comparison of various forecasting techniques applied to mean hourly wind speed time series. *Renewable Energy*, 21:23–35, 2000.

[204] K. Skytte. The regulating power market on the Nordic power exchange Nord Pool: an econometric analysis. *Energy Economics*, 21:295–308, 1999.

[205] T. A. B. Snijders. On cross validation for predictor evaluation in time-series. In T. K. Dijstra, editor, *On Model Uncertainty and its Statistical Implications. Lecture Notes in Economics and Mathematical Systems 307.* Oxford University Press, Oxford, 1988.

[206] L. Söder. Reserve margin planning in a wind-hydro-thermal power system. *IEEE Trans. on Power Syst.*, 8(2), May 1993.

[207] T. R. Stewart. Uncertainty, judgement, and error in prediction. In D. Sarewitz, R. A. Pielke, and R. Byerly, editors, *Prediction: Science, Decision Making, and the Future of Nature*, pages 41–57. Island Press, 2000.

[208] M. Stone. The opinion pool. *Annals of Mathematical statistics*, 32:1339–1342, 1961.

[209] M. Stone. Cross validation and assessment of statistical predictions (with discussion). *J. Royal Stat. Soc., B*, 36:111–147, 1974.

[210] J. Sveca and L. Söder. Wind power integration in power systems with bottleneck problems. In *CD-Proc. of the 2003 Bologna Power Tech Conference, IEEE Power Tech 2003, Bologna, Italy*, June 2003.

[211] N. R. Swanson and H. White. Forecasting economic time-series using flexible versus fixed specification and linear versus nonlinear econometric models. *Int. J. Forecasting*, 13(4):439–461, 1997.

[212] W. Sweet. Reap the wild wind. *IEEE Spectrum*, 39:34–39, October 2002.

[213] L. Takacs. A two-step scheme for the advection equation with minimized dissipation and dispersion errors. *Mon. Wea. Rev.*, 113:1050–1065, 1985.

[214] J. Tambke, M. Lange, and U. Focken. Forecasting offshore wind speeds above the North Sea. *Wind Energy*, 8(1):3–16, January-March 2005.

[215] J. W. Taylor and R. Buizza. Neural network load forecasting with weather ensemble predictions. *IEEE Trans. on Power Syst.*, 17(3):626–632, August 2002.

[216] J. W. Taylor and R. Buizza. Using weather ensemble predictions in electricity demand forecasting. *Int. J. Forecasting*, 19(1):57–70, 2003.

[217] J. W. Taylor and D. W. Bunn. Investigating improvements in the accuracy of prediction intervals for combination of forecasts: a simulation study. *Int. J. Forecasting*, 15(3):325–339, 1999.

[218] K. E. Taylor. Summarizing several aspects of model performance in a single diagram. *J. Geophys. Res.*, 106:7183–7192, 2001.

[219] H. Tennekes. Karl Popper and the accountability of numerical forecasting. In *New Developments in Predictability*, 1991. ECMWF Workshop Proceedings, ECMWF, Reading, United Kingdom.

[220] S.-E. Thor and P. Weis-Taylor. Long-term research and development needs for wind energy for the time frame 2000-2020. *Wind Energy*, 5:73–75, April-June 2003.

[221] H. Tong. A personal overview of non-linear time-series analysis from a chaos point of view. *Scand. J. Stats*, 22:399–445, 1995.

[222] Z. Toth and E. Kalnay. Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, 125:3297–3319, 1997.

[223] Z. Toth, O. Tallagrand, G. Candille, and Y. Zhu. Probability and ensemble forecasts. In I.T. Jolliffe and D.B. Stephenson, editors, *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley & Sons, Ltd, 2003.

[224] Z. Toth, Y. Zhu, and T. Marchock. The use of ensembles to identify forecasts with small and large uncertainty. *Wea. Forecasting*, 16(4):463–477, 2001.

[225] I. Troen and E. L. Petersen. *European Wind Atlas*. Risø National Laboratory, 1989.

[226] J. Usaola and P. de Arriba Segurado. Wind prediction in electricity markets. ANEMOS Deliverable Report D8.2, available soon, November 2005.

[227] J. Usaola, O. Ravelo, G. Gonzalez, F. Soto, C. Davila, and B. Diaz-Guerra. Benefits for wind energy in electricity markets from using short term wind power prediction tools - A simulation study. *Wind Engineering*, 28(1):119–128, 2004.

[228] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, 1998.

[229] J. W. Verkaik. A method for the geographical interpolation of wind speed over heterogenous terrain. KNMI HYDRA project report, Royal Meteorological Institute of the Netherlands (KNMI), 2001. available online: www.knmi.nl/samenw/hydra/documents/geograph/.

[230] L. J. Vermeer, J. N. Sørensen, and A. Crespo. Wind turbine wake aerodynamics. *Progress in Aerospace Sciences*, 39:467–510, 2003.

[231] K. F. Wallis. Forecast uncertainty, its representation and evaluation. Tutorial lecture, IMS Singapore, May 2004.

[232] L.-W. Wang. *Adaptive fuzzy systems and control*. Prentice Hall, 1994.

[233] R. Watson and L. Landberg. Evaluation of the Prediktor wind power forecasting system in Ireland. In *CD-Proc. of the 2003 European Wind Energy Conference, EWEC'03, Madrid, Spain*, June 2003.

[234] J. S. Whitaker and A. F. Loughe. The relationship between ensemble spread and ensemble mean skill. *Mon. Wea. Rev.*, 126:3292–3302, 1998.

[235] W. H. Williams and M. L. Goodman. A simple method for the construction of empirical confidence limits for economic forecasts. *J. Amer. Stat. Ass.*, 66(336):752–754, 1971. Applications Section.

[236] Wind Focus. A summary of opinion surveys on wind power. *Wind Directions*, pages 16–33, September-October 2003. available at www.ewea.org/documents/WD22vi_public.pdf.

[237] R. L. Winkler. A decision-theoretic approach to interval estimation. *J. Amer. Stat. Ass.*, 67:187–191, 1972.

[238] R. L. Wobus and E. Kalnay. Three years of operational prediction of forecast skill at NMC. *Mon. Wea. Rev.*, 123:2132–2147, 1995.

[239] A. Yamaguchi, T. Ishihara, and Y. Fujino. Experimental study of the wind flow in a coastal region of Japan. *J. of Wind Eng. & Ind. Aerodyn.*, 91:247–264, 2003.

[240] L. Yuancheng, F. Tingjian, and E. Yu. Short-term electrical load forecasting using least squares Support Vector Machines. *IEEE Trans. on Power Syst.*, 17(3):626–632, August 2002.

[241] J. C. Zack. A new technique for short-term wind energy forecasting: a rapid update cycle with a physics-based atmospheric model. In *CD-Proc. of the 2004 Global Windpower Conference, Chicago, Illinois (USA)*, March 2004.

[242] A. Zervos. Developing wind energy to meet the Kyoto targets in the European Union. *Wind Energy*, 6(3):309–319, July-September 2003.

[243] G. Zhang, Patuwo B. E., and M. Y. Hu. Forecasting with artificial neural networks: the state of the art. *Int. J. Forecasting*, 14(1):35–62, 1998.

[244] C. Ziehmann. Skill prediction of local weather forecasts based on the ECMWF ensemble. *Nonlin. Proc. Geophys.*, 8:419–428, 2001.

ESTIMATION DE L'INCERTITUDE DES PREDICTIONS DE PRODUCTION EOLIENNE

Résumé

L'énergie éolienne connaît un développement considérable en Europe. Pourtant, le caractère intermittent de cette énergie renouvelable introduit des difficultés pour la gestion du réseau électrique. De plus, dans le cadre de la dérégulation des marchés de l'électricité, l'énergie éolienne est pénalisée par rapport aux moyens de production contrôlables. La prédiction de la production éolienne à des horizons de 2-3 jours aide l'intégration de cette énergie. Ces prédictions consistent en une seule valeur par horizon, qui correspond à la production la plus probable. Cette information n'est pas suffisante pour définir des stratégies de commerce ou de gestion optimales. C'est pour cela que notre travail se concentre sur l'incertitude des prédictions éoliennes. Les caractéristiques de cette incertitude sont décrites à travers une analyse des performances de certains modèles de l'état de l'art, et en soulignant l'influence de certaines variables sur les moments des distributions d'erreurs de prédiction. Ensuite, nous décrivons une méthode générique pour l'estimation d'intervalles de prédiction. Il s'agit d'une méthode statistique non-paramétrique qui utilise des concepts de logique floue pour intégrer l'expertise acquise concernant les caractéristiques de cette incertitude. En estimant plusieurs intervalles à la fois, on obtient alors des prédictions probabilistes sous forme de densité de probabilité de production éolienne pour chaque horizon. La méthode est évaluée en terme de fiabilité, finesse et résolution. En parallèle, nous explorons la possibilité d'utiliser des prédictions ensemblistes pour fournir des 'prévisions d'erreur'. Ces prédictions ensemblistes sont obtenues soit en convertissant des prévisions météorologiques ensemblistes (fournies par ECMWF ou NCEP), soit en appliquant une approche de décalage temporel. Nous proposons une définition d'indices de risque, qui reflètent la dispersion des ensembles pour un ou plusieurs horizons consécutifs. Une relation probabiliste entre ces indices de risque et le niveau d'erreur de prédiction est établie. Dans une dernière partie, nous considérons la participation de l'énergie éolienne dans les marchés de l'électricité afin de démontrer la valeur de l'information 'incertitude'. Nous expliquons comment définir des stratégies de participation à ces bourses de l'électricité avec des prédictions déterministes ou probabilistes. Les bénéfices résultant d'une estimation de l'incertitude des prédictions éoliennes sont clairement démontrés.

Mots clés : énergie éolienne, prédiction, incertitude, intervalles de prédiction, indices de risque, prédictions ensemblistes, prise de décision, prédiction probabiliste

ESTIMATION OF THE UNCERTAINTY IN WIND POWER FORECASTING

Abstract

Wind power experiences a tremendous development of its installed capacities in Europe. Though, the intermittence of wind generation causes difficulties in the management of power systems. Also, in the context of the deregulation of electricity markets, wind energy is penalized by its intermittent nature. It is recognized today that the forecasting of wind power for horizons up to 2/3-day ahead eases the integration of wind generation. Wind power forecasts are traditionally provided in the form of point predictions, which correspond to the most-likely power production for a given horizon. That sole information is not sufficient for developing optimal management or trading strategies. Therefore, we investigate on possible ways for estimating the uncertainty of wind power forecasts. The characteristics of the prediction uncertainty are described by a thorough study of the performance of some of the state-of-the-art approaches, and by underlining the influence of some variables e.g. level of predicted power on distributions of prediction errors. Then, a generic method for the estimation of prediction intervals is introduced. This statistical method is non-parametric and utilizes fuzzy logic concepts for integrating expertise on the prediction uncertainty characteristics. By estimating several prediction intervals at once, one obtains predictive distributions of wind power output. The proposed method is evaluated in terms of its reliability, sharpness and resolution. In parallel, we explore the potential use of ensemble predictions for skill forecasting. Wind power ensemble forecasts are obtained either by converting meteorological ensembles (from ECMWF and NCEP) to power or by applying a poor man's temporal approach. A proposal for the definition of prediction risk indices is given, reflecting the disagreement between ensemble members over a set of successive look-ahead times. Such prediction risk indices may comprise a more comprehensive signal on the expected level of uncertainty in an operational environment. A probabilistic relation between classes of risk indices and the level of forecast error is shown. In a final part, the trading application is considered for demonstrating the value of uncertainty estimation when predicting wind generation. It is explained how to integrate that uncertainty information in a decision-making process accounting for the sensitivity of end-users to regulation costs. The benefits of having a probabilistic view of wind power forecasting are clearly shown.

Keywords: wind power, forecasting, uncertainty estimation, prediction intervals, ensemble prediction, skill forecasting, decision-making processes, probabilistic forecasting.

Laboratoire d'accueil :	Centre Energétique et Procédés - Ecole des Mines de Paris Rue Claude Daunesse - B.P. 207 - F-06904 Sophia Antipolis Cedex
Thèse présentée par :	PINSON Pierre, **le :** 23 mars 2006
Discipline :	« Energétique » - Ecole des Mines de Paris