SafeWind

Collaborative project funded by the European Commission under the 7th Framework Program, Theme 2007-2.3.2: Energy

"Multi-scale data assimilation, advanced wind modelling & forecasting with emphasis to extreme weather situations for a safe large-scale wind power integration"

Grant Agreement N°: 213740

Deliverable Dp-7.4

"Impact of spatio-temporal dependence in forecasts and their uncertainty at the investment stage"

DOCUMENT TYPE	Deliverable
DOCUMENT NAME:	swind.deliverable-Dp-7.4.pdf
VERSION:	V2.0
DATE:	2012.08.31
CLASSIFICATION:	R0: General public
STATUS:	Approved

Abstract: Predictability aspects are most often disregarded at the time of optimally spreading wind power capacities. The potential impact of predictability and of different types of forecast products are analysed here in various market environments, i.e. zonal markets which are most common in Europe and nodal markets as existing in the US. Methodologies for forecasting, optimal participation in electricity markets in view of price dynamics and forecast uncertainty, and finally for optimal investment, are described. Simulations based on real-world datasets and simplified power systems allow us to derive a number of findings supporting the interest of fully describing forecast uncertainty and spatio-temporal dynamics at the investment stage.

AUTHORS[1], REVIEWERS				
Main Author/Editor:	P. Pinson			
Affiliation:	Technical University of Denmark, DTU Informatics			
Address:	Asmussens Allee 305, 2800 Kgs Lyngby, Denmark			
Tel.:	+45 4525 3428			
Email:	pp@imm.dtu.dk			
Further Authors:	Stefanos Delikaraoglou (DTU), Juan Miguel Morales (DTU)			
Peer Reviewers:	Robin Girard (Mines ParisTech)			
Review Approval:	Approved :	X	Rejected (improve as indicated below) :	
Suggested Improvements:	*For a long list of remarks make reference to another document*			
Approver:	Pierre Pinson (DTU), Robin Girard (Mines ParisTech)			

STATUS, CONFIDENTIALITY, ACCESSIBILITY										
Status:				Confidentiality:				Accessibility:	
S0	Approved/Released	**X**		**R0**	General public	**X**		Private web site	**X**
S1	Reviewed			**R1**	Restricted to project members			Public web site	**X**
S2	Pending for review			**R2**	Restricted to European **Commission**			Paper copy	
S3	Draft for comments			**R3**	Restricted to WP members + PL				
S4	Under preparation			**R4**	Restricted to Task members +WPL+PL				

PL: *Project leader* **WPL:** *Work package leader* **TL:** *Task leader*

[1] **The authors of this document are solely responsible for its content, which does not represent the opinion of the European Community and the European Community is not responsible for any use that might be made of data appearing therein.**

Contents

1. Document Styling Guidelines (1st-level head)

1.1 Styling Guideline and Tools to prepare your documents (Second-level head)

Text text text

1.2 How to prepare your SafeWind documents (2nd-level head)

Text text text

Title Page (3rd-level head)

The current document gives the guidelines for the front page of the report. The different types of reports defined at the moment are:

- Deliverable (reports foreseen in the contract)
- Dissemination Document (presentation, flyer, training material).
- Meeting Report (reports from project meetings)
- Memorandum (free style)
- Progress Report (reports foreseen in the contract)
- Publication (papers, contribution to books etc).
- Technical Report (technical reports).

Main text

For the main text, please respect the following styling guidelines:

Preferred File Format	PDF
Paper format	A4
Top-Margin	25mm
Bottom-Margin	25mm
Left Margin	25mm
Right Margin:	25mm
First-level head	Arial Bold 15pt
Second level head	Arial Bold 12pt
Third level head	Arial Bold 11pt
Normal Font	Arial 10pt
Text Body	1 Column; no Page-Number; no Headline
Paragraphs	5mm Indent of the first line if necessary.
Alignment	Justified
Graphics	Non text elements (graphics, pictures etc) may increase unnecessarily the file size. To avoid this, graphics can be prepared at the right size and saved in JPG, GIF or PNG format before inserting them in the document.
Additional Fonts	Please prefer standard Adobe fonts as delivered with Acrobat Reader. If you need different fonts you must ensure, that these fonts are embedded into the PDF file.

Word and .pdf examples

An example Word File (.doc) and a .pdf one, which follow the styling guidelines, are also available at the Work-Space.

Headers and footers

- If you submit a document to be included in others documents please leave headers and footers empty.
- Otherwise put in the head project name, document number and revision date. In the footer please put the type of the document, its status and page numbers as shown in the examples.
- Alternatively use the stamp-it tool. With the stamp-it tool you can add the suggested header and footer on pdf-documents.

Merging documents from different authors

If you are editing a document by merging .pdf documents from different authors you may use Adobe Acrobat. After that, you can use the stamp-it tool to stamp all pages with pages numbers and the page header (project name, document number, revision date). The stamp-it tools require Adobe Acrobat (not only the Acrobat Reader).

2. Examples of text

Accurate forecasting of the wind resource up to two days ahead is recognised as a major contribution for reliable large-scale wind power integration. Especially, in a liberalised electricity market, prediction tools enhance the position of wind energy compared to other forms of dispatchable generation.

The **SafeWind** project aims to develop advanced forecasting models that will substantially outperform current methods. Emphasis is given to situations like complex terrain, extreme weather conditions, as well as to offshore prediction for which no specific tools currently exist. The prediction models are implemented in a software platform and installed for online operation at onshore and offshore wind farms by the end-users participating in the project. The project demonstrates the economic and technical benefits from accurate wind prediction at different levels: national, regional or at single wind farm level and for time horizons ranging from minutes up to several days ahead.

2.1 More example of text

The project is structured into 3 work-packages, which address the technical objectives. Initially, the prediction requirements are defined in collaboration with end-users.

The project develops prediction models based on both a physical and an alternative statistical approach. Research on physical models gives emphasis to techniques for use in complex terrain and the development of prediction tools based on CFD techniques, advanced model output statistics or high-resolution meteorological information. Statistical models (i.e. based on artificial intelligence) are developed for downscaling, power curve representation, upscaling for prediction at regional or national level, etc.

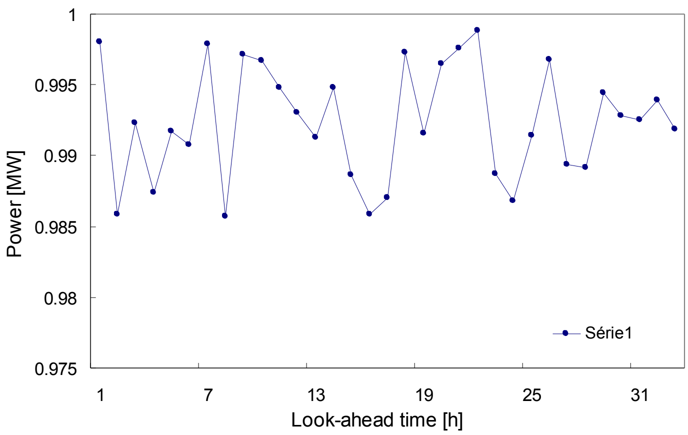


Figure 1: *This is a measurement of wind speed.*

A benchmarking process is set-up to evaluate the performance of the developed models and to compare them with existing ones using a number of case studies. The synergy between statistical and physical approaches is examined to identify promising areas for further improvement of forecasting accuracy. The performance of purely meteorological forecasts, but also long-term wind predictability up to 7 days ahead, are evaluated in detail. Appropriate physical and statistical prediction models are also developed for offshore wind farms taking into account advances in marine meteorology (interaction between wind and waves, coastal effects). The benefits from the use of satellite radar images for modelling local weather patterns are investigated.

Example continuation

The project provides an advanced technology for wind resource forecasting applicable in a large scale: at a single wind farm, regional or national level and for both interconnected and island systems. A major milestone of the project is the on-line operation of the developed software by the participating utilities for onshore and offshore wind farms.

Table 1: *These are the values of the measurements.*

Type	Value	Comment
Measurement 1	0.998	
Measurement 2	0.9858556	
Label 1	0.9922741	
Label 2	0.9874029	

The outcome of the ANEMOS.plus project will help consistently the increase of wind integration in two levels; in an operational level due to better management of wind farms, but also, it will contribute to increasing the installed capacity of wind farms. This is because accurate prediction of the resource reduces the risk of wind farm developers, who are then more willing to undertake new wind farm installations especially in a liberalised electricity market environment.

Contents

1 Introduction

Renewable energy capacities are being massively deployed throughout the world over the last few years, with a major role given to wind power. Some countries like Denmark, Germany and Spain paved the way by being the first ones to substantially integrate wind energy into their power systems and electricity markets. Since then the United States and China have taken over as leading countries in terms of deployment of new renewable energy capacities. A number of other countries are rapidly following the same path. This deployment has been heavily supported by various support schemes e.g. feed-in tariffs, quotas, market premiums, etc., some of them being discussed by Butler and Neuhoff (2008). The impact of these support schemes on investment in renewable energy installations has been analysed for example based on a real-options approach (Boomsma *et al.* 2012).

Even with such support schemes, renewable energy generation has to come to electricity markets eventually, where it has to compete as equal with other forms of conventional generation. This is already the case in Scandinavian countries, Spain and the US among others. It then has a potential disadvantage stemming from limited predictability and non-dispatchability. As an example in most European electricity market, this originates from the existence of penalties for deviations between contracts in the forward market (day-ahead for most of them) and actual production, as settled through the real-time adjustment market. In a purely market-driven environment, it is potentially more difficult to see how to optimally perform renewable energy investments, in view of the combined dynamic and stochastic effects of market mechanisms and from the weather itself. This was recently looked upon also based on a real-options approach (Chen and Lu 2011), while other proposals in a hierarchical optimization framework also exist, naturally allowing to embed the market and weather aspects within the investment problem itself, e.g. Baringo and Conejo (2011) and Baringo and Conejo (2012). Similarly, this type of framework is general enough so as to also account for e.g. optimal wind energy deployment relative to transmission costs (Morales *et al.* 2012).

Optimal investments in wind power capacities in a liberalized market environment may be influenced by a number of factors of meteorological and economical nature. This was recently discussed by Girard *et al.* (2012b), who showed that for the case of Western Denmark in the Scandinavian zonal market Nord Pool, the most influential factor may certainly be the capacity factor[1] of wind farms, since market penalties are not strong enough for predictability to be a non-negligible decision factor. In contrast in a market with Locational Marginal Prices (LMPs), see Litvinov (2010) for an introduction, one has to find a right balance between optimum generation throughput and corresponding impacts on LMPs, as shown and discussed by Baringo and Conejo (2011). In the present work, it is aimed to extend the work of Girard *et al.* (2012b) by looking at the impact of spatio-temporal effects in wind generation and related production forecasts (more or less advanced) on optimal investments decisions, in both zonal and LMP markets. The input wind power and forecast data are first introduced in Section 2 while explaining how correlation effects may be modelled and accounted for. Subsequently, Sections 3 and 4 focus on the importance of these effects in zonal and LMP market environments, respectively. Finally the report ends with some general conclusions and perspectives in Section 5.

[1] The capacity factor can be defined as the ratio of actual energy generation over a period with the energy that would have been generated if power output was constantly at its nominal level. It typically ranges between 0.25 and 0.5 for European wind farms.

2 Wind power modeling and forecasting aspects

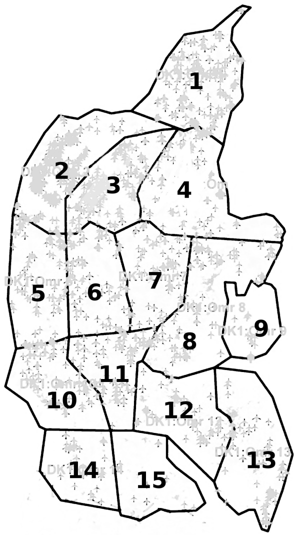
Forecasts are a necessary input to all type of market-related decision-making. One may well attempt at making decisions based on rules of thumb and expert guesses, though it is fairly sure that those based on advanced forecasting methodologies will be more optimal. The advanced forecasting products that were used as input to this study are described in the following. Based on the original measurements dataset, point forecasts, probabilistic predictions as well as spatio-temporal scenarios of short-term wind power production were generated. Some of the key methodological and application aspects are also discussed.

2.1 Original dataset: power measurements and forecasts

Energinet.dk, the Transmission System Operator (TSO) in Denmark, originally provided a dataset with power measurements for the Western Denmark area, also commonly referred to as DK-1. This control area of the Nord Pool (the electricity market for the Scandinavian region, plus areas of Northern Germany and Baltic countries) covers the Jutland peninsula and the Island of Funen. The underlying power system is connected to the UCTE (Union for the Coordination of Transmission of Electricity) system. It has approximately 70% of the entire wind power capacity installed in Denmark. Over the last few years, the yearly wind power penetration for DK-1, i.e. the share of wind energy in the electricity mix to meet demand, has been between 20 and 26%.

The control area of Western Denmark is split into 15 control zones onshore. The offshore control zone corresponds to the Horns Rev wind farms. It was not considered here due to data confidentiality issues. Hourly power measurements for these control zones were provided from a period between 1.1.2006 and 24.10.2007. The geographical set-up for these control zones may be visualized from Figure 1, along with their identification numbers (from 1 to 15, in a North-South direction). The total nominal capacity evolves during this period while being generally around 2.5 GW. In order to simplify the usage of this data in the investment problem considered, it was decided to aggregate data from these 15 zones so as to obtain 5 control zones only. For the purpose of the study, all power measurements (and subsequent forecasts) are normalized and take values in [0,1] only. This is since it will permit to model uncertainties and spatio-temporal dynamics without considering relative capacities, which may then be adjusted afterwards depending upon the investment problem set-up.

This dataset of power measurements is complemented by single-valued wind power forecasts as obtained from the power curve model of the Wind Power Prediction Tool (WPPT). Such single-valued predictions are commonly referred to as point forecasts. For a description of the forecasting approach involved, see Nielsen *et al.* (2002) and Nielsen (2002). These wind power predictions were generated with meteorological forecasts of wind speed and direction from the Danish Meteorological Institute (DMI) as input. They were originally produced for the 15 control zones individually, and then summed up to obtain those for the 5 aggregated zones, in a manner similar to the case of the power measurements. Wind power forecasts were originally issued every 6 hours, with a hourly resolution up to 48 hours ahead. It was decided to simulate having hourly update of the forecasts, by sliding the available 6-hourly updated ones. The forecast length is therefore restricted to 43 hours ahead. This is sufficient for market participation anyway, as the maximum forecast length needed for placing bids in a day-ahead market is at most 38 hours.



Agg. zone	Original zones	Share of DK-1 capacity [%]
1	1, 2, 3	31
2	5, 6, 7	18
3	4, 8, 9	17
4	10, 11, 14, 15	23
5	12, 13	10

FIGURE 1: *The 15 control zones of the original Energinet.dk dataset, then merged into 5 zones.*

Figure 2 depicts an example set of point forecasts of wind power generation issued on the 3.4.2007 at 16:00 for the 5 aggregated zones, along with the corresponding measurements. The temporal patterns in measurements have a certain level of similarity, hence showing a strong spatial correlation in wind power generation over Western Denmark. The temporal dependency is high as indicated by the smoothness of the measurement trajectories. Power forecasts also are correlated spatially thanks to the information provided by the meteorological predictions. Overall here and for the whole dataset, the quality of the wind power forecasts is at the level of the state of the art, as verified with a number of error criteria e.g. bias, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), etc. following the guidelines of Madsen *et al.* (2005).

2.2 Obtaining probabilistic forecasts and scenarios

Point forecasts do not comprise the optimal input information for an optimal participation of wind energy producers in electricity markets. They should be generalized to probabilistic forecasts instead, giving a complete picture of the uncertainty in future wind power generation for every lead time and every locations of interest. This is the case if considering both zonal (Pinson *et al.* 2007) and LMP (Botterud *et al.* 2012) types of electricity markets. In a probabilistic forecasting framework, point forecasts which aim at minimizing a quadratic error criterion correspond to the conditional expectation of wind power generation for each lead time (and every location). They are somewhat 'just the mean of whatever may happen', hence not always being so informative about what might actually happen.

2.2.1 Nonparametric predictive distributions

Probabilistic forecasts here take the form of nonparametric predictive densities, describing predictive cumulative distribution functions of wind power generation based on a set of quantile forecasts. We write $Y_{(s,t)}$ the random variable corresponding to wind power generation at location s and time t, while $y_{(s,t)}$

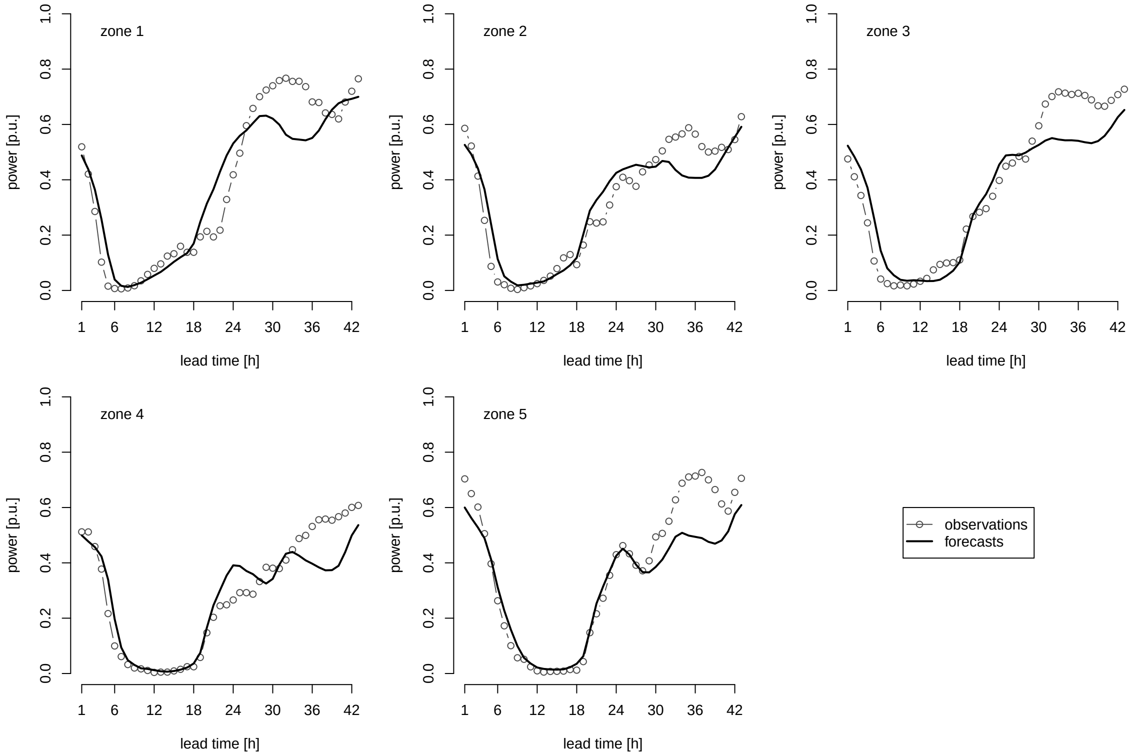


FIGURE 2: *Point forecasts of wind power generation issued on the 3.4.2007 at 16:00 for the 5 aggregated zones. These forecasts are normalized and hence take value in [0,1].*

denotes the corresponding realization. In parallel, $f_{(s,t)}$ and $F_{(s,t)}$ are the probability density function and related cumulative distribution function for that random variable. We naturally write $\hat{y}_{(s,t+k|t)}$ the point forecast issued at time t for time $t+k$, with extension to probabilistic forecasts $\hat{f}_{(s,t+k|t)}(y)$ and $\hat{F}_{(s,t+k|t)}(y)$. In a nonparametric set-up, $\hat{F}_{(s,t+k|t)}$ is characterized as

$$\hat{F}_{(s,t+k|t)} = \{\hat{q}^{(\alpha_i)}_{s,t+k|t} \mid 0 \leq \alpha_1 < \ldots < \alpha_i < \ldots < \alpha_m \leq 1\}, \tag{1}$$

that is, by quantile forecasts $\hat{q}^{(\alpha_i)}_{s,t+k|t}$ with chosen nominal proportions spread over the unit interval. A quantile forecast with nominal proportion α, issued at time t for lead time $t+k$ and location s, is a predicted power value which has a probability α to cover the observation,

$$\mathrm{P}(Y_{(s,t)} < \hat{q}^{(\alpha_i)}_{s,t+k|t}) = \alpha \tag{2}$$

or equivalently

$$\hat{q}^{(\alpha_i)}_{(s,t+k|t)} = \hat{F}^{-1}_{(s,t+k|t)}(\alpha) \tag{3}$$

There exists a number of approaches for issuing nonparametric probabilistic forecasts of wind power generation based on already issued point predictions and original meteorological forecasts, or alternatively based on meteorological ensemble predictions. Maybe the most well-documented and widely applied methods are the adapted resampling of Pinson and Kariniotakis (2010) and the time-adaptive quantile

regression of Møller *et al.* (2008). The former was chosen due to its simplicity of implementation. It is applied independently for producing probabilistic forecasts for each of the 5 aggregated zones.

Figure 3 depicts an example set of nonparametric probabilistic forecasts corresponding to the same episode as in Figure 2, i.e. issued on the 3.4.2007 at 16:00 for the 5 aggregated zones. Corresponding measurements are also shown. Due to their nonparametric and adaptive nature, these probabilistic forecasts vary with the level of predicted power, the lead time, location and time of year. Here also the probabilistic forecasts were thoroughly evaluated following the guidelines of Pinson *et al.* (2007b), in order to verify their probabilistic calibration and that their overall skill as quantified by proper skill scores (Continuous Rank Probability Score - CRPS, and Ignorance score) was at the level of the state of the art. Having probabilistic forecasts that are probabilistically calibrated is a pre-requisite for application of the methods used to generate scenarios.

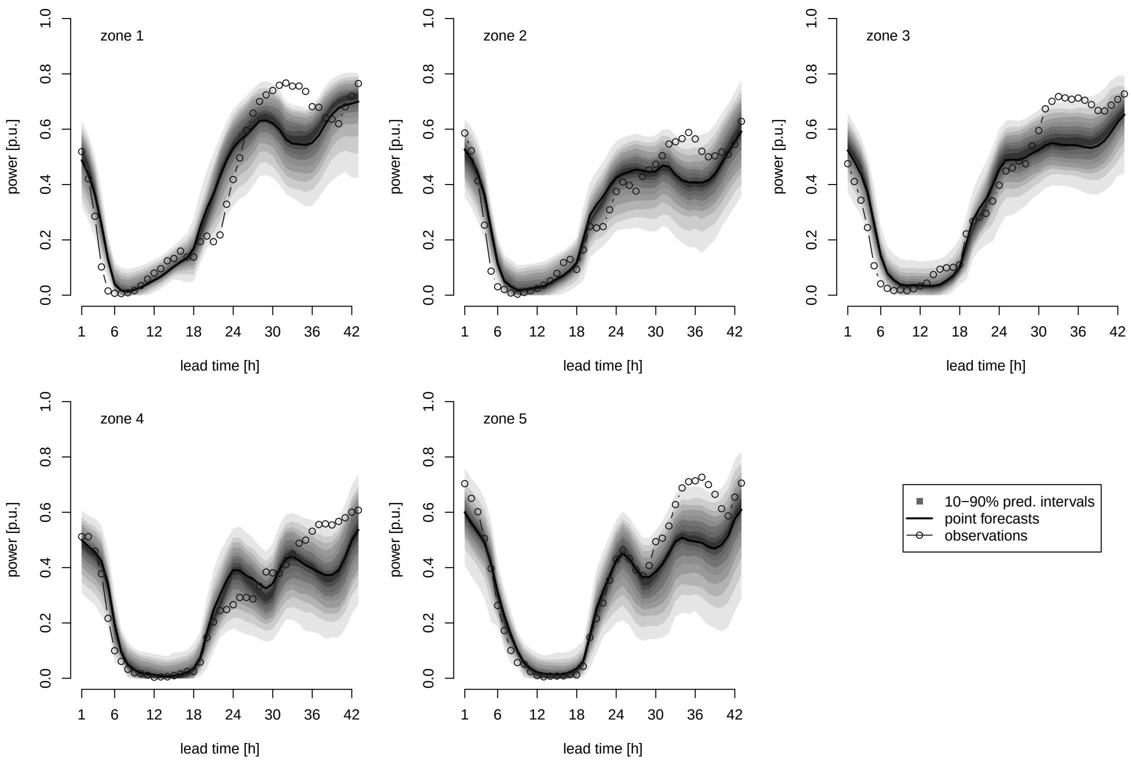


FIGURE 3: *Nonparametric probabilistic forecasts of wind power generation issued on the 3.4.2007 at 16:00 for the 5 aggregated zones, represented as a river-of-blood fan chart. Predictive densities are defined by a set of quantile forecasts with nominal proportions* $\alpha \in \{0.05, 0.1, \ldots, 0.45, 0.55, \ldots, 0.9, 0.95\}$*, yielding prediction intervals with nominal coverage rates between 10 and 90%. These forecasts are normalized and hence take value in [0,1].*

2.2.2 Spatio-temporal scenarios of wind power generation

For the generation of spatio-temporal scenarios, we extend the method described by Pinson *et al.* (2009) for the generation of scenarios of short-term wind power generation based on probabilistic forecasts. This method was presented with the temporal dependence problem in mind. It was also translated to the case of spatial dependencies in Papaefthymiou and Pinson (2008). The generalization to the space-time

case is straightforward and should still be numerically acceptable for problems of moderate size—for instance here the problem dimension is of 215 (for 5 zones and 43 lead times). We denote by p_s and p_k the number of locations and lead times, respectively. $p = p_s \times p_k$ then represents the overall dimension of the problem.

Let us assume first that one may model a spatio-temporal dependence structure by a Gaussian meta-model, also called Gaussian copula, which would translate to considering a p-dimensional standard Gaussian random variable,

$$\mathbf{Z}_t \sim \mathcal{N}_p(0, \mathbf{\Sigma}_t), \quad \mathbf{\Sigma}_t \in \mathbb{R}^{p \times p} \tag{4}$$

whose covariance matrix would fully characterize this interdependence structure, for all lead times and locations. The diagonal elements of $\mathbf{\Sigma}_t$ are all equal to 1, since being a standard Gaussian. Instead of proposing spatio-temporal models for this covariance structure, our approach is to track it based on recent data.

With that objective in mind, we assume that input probabilistic forecasts are calibrated, and this for all lead times and locations of interest. In that case, this directly translates to

$$\hat{F}_{(s,t+k|t)}(y) \sim \mathcal{U}[0,1], \quad \forall t, s, k \tag{5}$$

and by extension

$$\Phi^{-1}\left(\hat{F}_{(s,t+k|t)}(y)\right) \sim \mathcal{N}(0,1), \quad \forall t, s, k \tag{6}$$

where Φ^{-1} is the inverse Gaussian cumulative distribution function. Consequently, considering a reordering of Gaussian random variables corresponding to all locations and lead times, one obtains p standard Gaussian variables, the multivariate structure of which may be summarized by $\mathbf{\Sigma}_t$ as in Eq. (4).

It is proposed to adaptively track $\mathbf{\Sigma}_t$ in an exponential smoothing framework. For a given time point t, sets of probabilistic forecasts are available for the recent past period and for all locations. An estimate $\hat{\mathbf{\Sigma}}_t$ of the multivariate covariance structure is recursively updated with

$$\hat{\mathbf{\Sigma}}_t = \lambda \hat{\mathbf{\Sigma}}_{t-1} + (1-\lambda)\mathbf{z}_{t-p_k}\mathbf{z}_{t-p_k}^\top \tag{7}$$

where

$$\mathbf{z}_{t-p_k} = \begin{bmatrix} \Phi^{-1}\left(\hat{F}_{(s_1,t-p_k+1|t-p_k)}(y_{(s_1,t-p_k+1)})\right) \\ \Phi^{-1}\left(\hat{F}_{(s_1,t-p_k+2|t-p_k)}(y_{(s_1,t-p_k+2)})\right) \\ \dots \\ \Phi^{-1}\left(\hat{F}_{(s_1,t|t-p_k)}(y_{(s_1,t)})\right) \\ \Phi^{-1}\left(\hat{F}_{(s_2,t-p_k+1|t-p_k)}(y_{(s_2,t-p_k+1)})\right) \\ \Phi^{-1}\left(\hat{F}_{(s_2,t-p_k+2|t-p_k)}(y_{(s_2,t-p_k+2)})\right) \\ \dots \\ \Phi^{-1}\left(\hat{F}_{(s_2,t|t-p_k)}(y_{(s_2,t)})\right) \\ \dots \\ \Phi^{-1}\left(\hat{F}_{(s_{p_s},t-p_k+1|t-p_k)}(y_{(s_{p_s},t-p_k+1)})\right) \\ \Phi^{-1}\left(\hat{F}_{(s_{p_s},t-p_k+2|t-p_k)}(y_{(s_{p_s},t-p_k+2)})\right) \\ \dots \\ \Phi^{-1}\left(\hat{F}_{(s_{p_s},t|t-p_k)}(y_{(s_{p_s},t)})\right) \end{bmatrix} \tag{8}$$

is the vector of past observations transformed through the probabilistic forecasts series issued at time $t - p_k$, and then through the probit function Φ^{-1}. The estimated covariance structure is initialized with $\hat{\mathbf{\Sigma}}_0 = \mathbf{I}$ (i.e. a diagonal matrix of ones), while optimal values for λ were found to be around 0.995. This is somewhat equivalent to employing sliding estimation windows of approximately 8 days, consistent with the time scale for the movement of synoptic weather patterns over Northern Europe. Note that if aiming at optimizing the quality of time trajectories, more advanced covariance structures for the Gaussian copula may be envisaged, for instance combining exponential decay with horizon-dependent range parameter and kernels for representing potential seasonalities, and similarly an exponential decay in space. Parametric covariance structures in time were alternatively looked at by Pinson and Girard (2010).

Finally for issuing a set of spatio-temporal scenarios at time t, one first needs to sample a number J of realizations $\mathbf{z}_t$ from the multivariate random variable defined in Eq. (4) and with the updated space-time covariance structure $\hat{\mathbf{\Sigma}}_t$ of Eq. (7). Using the inverse probit function Φ, as well as the predictive cumulative distribution functions $\hat{F}_{(s,t+k|t)}$ for every lead time and every locations, these multivariate Gaussian realizations are transformed into trajectories of wind power generation having the appropriate marginal distributions,

$$\hat{y}^{(j)}_{(s,t+k|t)} = \hat{F}_{(s,t+k|t)}{}^{-1}\left(\Phi(z^{(j)}_{(s,t+k)})\right), \quad j = 1, \dots, J \tag{9}$$

Figure 4 depicts an example set of space-time scenarios of short-term wind power generation for the the same episode as in Figures 2 and 3. i.e. issued on the 3.4.2007 at 16:00 for the 5 aggregated zones. Corresponding measurements are also shown. The number of scenarios displayed is $J = 12$. By construction, these scenarios respect the marginal predictive densities of Figure 3 while adding information on the expected spatio-temporal dependencies. Similarly if generated a sufficiently high number of scenarios, the average of these scenarios would be equivalent to the point forecasts of Figure 2 as they represent the mean of all potential scenarios.

Depending on the decision-making problem at hand, only spatial or temporal dependencies may be needed. For the case of the market participation problem to be studied in the following, only spatial dependencies are considered since the decisions to be made for successive market time units are seen as independent. This is not the case in a more general framework, see e.g. Morales *et al.* (2010a), where the complete set of spatio-temporal scenarios may be necessary.

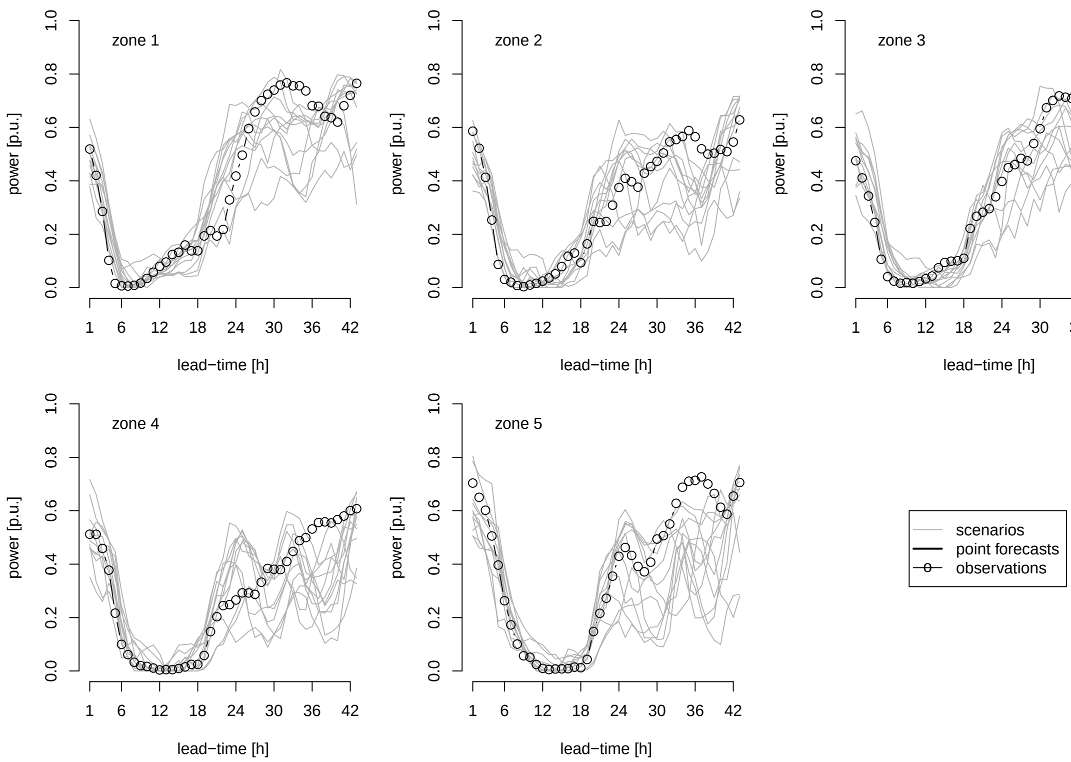


FIGURE 4: *Spatio-temporal scenarios of wind power generation issued on the 3.4.2007 at 16:00 for the 5 aggregated zones. These forecasts are normalized and hence take value in [0,1].*

The quality of these scenarios may be assessed based on a set of scores and diagnostic tools as suggested by Pinson and Girard (2010), hence complementing the verification of the marginal predictive densities performed in the above. Here they were found to be marginally calibrated and to have higher skill than if issuing scenarios with no dependence structure at all.

3 Optimal wind power investments in a zonal market environment

In a zonal market environment such at the Scandinavian Nord Pool, the clearing of the market is not associated with a systems model, and the prices cleared by the market are valid for a whole control area, if not for all of them. Only congestion between zones may prompt the need for price differentiation. The type of electricity market considered here is actually composed of two market systems which are the forward market (day-ahead) and the real-time market allowing for the settlement of imbalances, which are the deviations from original contracts.

3.1 Revenue and bidding strategy

We present some of the basic aspects of the definition of the revenue of a market participant in a zonal market like Nord Pool, then building up on the definition of more or less advanced market participation strategies, relying on the various types of forecasts previously introduced.

3.1.1 Revenue of a market participant

The time scales of the Nord Pool here define gate closure, time units in the market, etc. Consider being at time t on a given day just before gate closure (noon at the Nord Pool). A market participant has to make a set of decisions for all time units of the following period, that is, hourly contracts for all hours of the following day. A given time unit is indexed by $t+k$. That market participant has to propose a level of contracted energy E^c_{t+k}. The revenue R_{t+k} of a market participant proposing an amount of energy E^c_{t+k} but actually generating E^*_{t+k} can be formulated as

$$\mathrm{R}_{t+k} = \pi^c_{t+k} E^c_{t+k} \; + \; T^c_{t+k} \tag{10}$$

where π^c_{t+k} is the system price for this market time unit, and T^c_{t+k} is the cost stemming from imbalance settlement on the regulation market. The imbalance d^*_{t+k} is simply defined as the deviation from the contracted energy,

$$d^*_{t+k} = E^*_{t+k} - E^c_{t+k} \tag{11}$$

and consequently the regulation cost T^c_{t+k} is given by

$$T^c_{t+k} = \begin{cases} \pi^{c,\downarrow}_{t+k} d^*_{t+k}, & d^*_{t+k} \geq 0 \\ \pi^{c,\uparrow}_{t+k} d^*_{t+k}, & d^*_{t+k} < 0 \end{cases} \tag{12}$$

with $\pi^{c,\downarrow}_{t+k}$ and $\pi^{c,\uparrow}_{t+k}$ the imbalance prices for the situations of down-regulation and up-regulation, respectively. These imbalance prices linked to the day-ahead price through potentially complex functions. Some regulation mechanisms even follow a two-settlement system as in the Nord Pool, where regulation prices depend on the sign of the system imbalance as a whole, so that only participants who offset the system are to be penalized.

Eq. (10) can be reformulated such that the revenue R_{t+k} of a participant for the market time unit $t+k$ results from the combination of the income from the selling of actual wind generation E^*_{t+k} at the spot price, minus the costs for regulation, i.e.

$$\mathrm{R}_{t+k} = \pi^c_{t+k} E^*_{t+k} \; - \; T^*_{t+k} \tag{13}$$

where

$$T^*_{t+k} = \begin{cases} \pi^{*,\downarrow}_{t+k} d^*_{t+k}, & d^*_{t+k} \geq 0 \\ -\pi^{*,\uparrow}_{t+k} d^*_{t+k}, & d^*_{t+k} < 0 \end{cases} \tag{14}$$

$\pi^{*,\downarrow}_{t+k}$ and $\pi^{*,\uparrow}_{t+k}$, which are referred to as the regulation unit costs for down- and up-regulation, respectively, are readily given by

$$\pi^{*,\downarrow}_{t+k} = \pi^{c}_{t+k} - \pi^{c,\downarrow}_{t+k} \tag{15}$$
$$\pi^{*,\uparrow}_{t+k} = \pi^{c,\uparrow}_{t+k} - \pi^{c}_{t+k} \tag{16}$$

The revenue formulation in Eq. (13) has the advantage that the first component of the revenue is a 'fatal' component, and indeed corresponds to the income one would receive if using perfect predictions. Moreover, since the contracted energy only appears in the second component of the participant's revenue, maximizing R_{t+k} now translates to minimizing the regulation costs T^*_{t+k}. In the specific case for which a market participant is not charged for his imbalance since this imbalance helps the regulation process, $T^*_{t+k} = 0$.

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 an evaluation period of N_T participation days in the electricity pool, each participation day covering N_u daily market time units (here 24), hence with $N = N_T \times N_u$ total time units, γ writes

$$\gamma = 1 - \frac{\sum_{t=1}^{N_T} \sum_{k=1}^{N_u} T^*_{t+k}}{\sum_{t=1}^{N_T} \sum_{k=1}^{N_u} \pi^c_{t+k} E^*_{t+k}} \tag{17}$$

The proposed performance ratio is such that $\gamma \in (-\infty, 1]$, and it will be expressed in percent. It is obvious that for perfect prediction $\gamma = 1$ since deviations are null. γ may be calculated for given periods, e.g. for every month, or overall for the whole period of market participation.

3.1.2 Alternative bidding strategies

It is up to the market participant to define a bidding strategy that allows him to maximize his revenue, or as a proxy his performance ratio γ. When using point forecasts only the most reasonable approach is to derive bids directly from these forecasts. The bid at market time unit $t + k$ is then

$$E^c_{t+k} = \hat{E}_{t+k|t} \tag{18}$$

These types of strategies were applied directly by Barthelmie *et al.* (2008) in the United Kingdom and by Chang *et al.* (2009) in The Netherlands for assessing the value of wind energy and forecasting systems in electricity markets. Also in practice, almost all participants in electricity markets having to deal with wind energy derive their bids directly from wind power point forecasts.

More advanced bidding strategies may be derived based on probabilistic forecasts for wind power generation, as in Pinson *et al.* (2007) for instance, or alternatively based on scenarios when aiming to make further use of wind within a generation portfolio and when participating in various markets with different

gate closures (Morales *et al.* 2010a). Such advanced bidding strategies may require additional knowledge of some of the market prices though (Zugno *et al.* 2012).

Considering probabilistic forecasts and assuming that a market participant is rational (in the sense that he focuses on maximizing his own utility) the aim of an optimal market participation strategy is to maximize the expected value of total revenue at every market time unit $t + k$. This will then lead to revenue maximization in the long run. By seeing the market participant as a price taker (that is, not influencing market outcomes with his own decisions) and considering that decisions at successive market time units are independent, the optimal bid $\tilde{E}^c_{t+k}$ writes

$$\tilde{E}^c_{t+k} = \underset{E^c_{t+k}}{\arg\min}\, \mathbb{E}\left[g(d_{t+k})\right] \tag{19}$$

where

$$\mathbb{E}\left[g(d_{t+k})\right] \;=\; \int_0^1 g(x - E^c_{t+k}) f^E_{t+k}(x)dx \tag{20}$$

The problem in Eq. (19) is a variant of the well-known linear terminal loss problem, also called the newsvendor problem. After a little algebra, which is covered by Bremnes (2004), the analytical expression for the optimal bid is

$$\tilde{E}^c_{t+k} = \hat{F}^{E\;-1}_{t+k|t} \left(\frac{\hat{\pi}^{\downarrow}_{t+k|t}}{\hat{\pi}^{\downarrow}_{t+k|t} + \hat{\pi}^{\uparrow}_{t+k|t}} \right) \tag{21}$$

with $\hat{F}^E_{t+k|t}$ the cumulative distribution function corresponding to the predictive density $\hat{f}^E_{t+k|t}$. The optimal bid hence is a specific quantile of predictive densities, whose nominal proportion is a direct function of the predicted regulation unit costs for the market time unit. In the present case, regulation unit costs were predicted with the most simple version of exponential smoothing approaches, in the spirit of Jónsson *et al.* (2012). A more extensive description of advanced market participation strategies based on forecasts for both wind power and market prices is available in Zugno *et al.* (2012).

3.2 Formulation of the investment problem

The investment problem is formulated as an optimization problem where one has to optimally spread wind power capacities over the 5 aggregated zones, $P^{(j)}_{\mathrm{n}}$, $j = 1, \ldots, 5$, for a given overall nominal capacity P_n,

$$P_{\mathrm{n}} = \sum_{i=1}^{5} P^{(j)}_n \tag{22}$$

and this given the potentially revenues to be obtained through market participation based on various types of forecasts and bidding strategies. This writes

$$\{P^{(j)*}_n\} = \underset{\{P^{(j)}_n\}}{\arg\max}\, \gamma(\{P^{(j)}_n\}; \mathcal{S}, \mathcal{F}) \tag{23}$$

where $\mathcal{F}$ and $\mathcal{S}$ generically denote the chosen type of input forecasts and market participation strategies, respectively. The optimization is performed for the whole period without accounting for the generalization value of the optimal solution. This could be considered in a multi-fold cross-validation framework.

3.3 Application results

As an illustrative example of how optimal spreading of capacities and market revenues may be influenced by the choice of forecasts and market participation strategies, let us consider the case of an investor aiming to deploy P_n= 100 MW of wind energy installations over the 5 aggregated zones of Western Denmark. There is no environmental or grid restrictions imposing constraints on the maximum capacity to be installed in any of these zones. The production from all the deployed capacity will enter as one overall wind portfolio when placing offers in the market. Market prices, i.e. day-ahead and from the regulation market, are obtained for the Nord Pool DK-1 area for the years 2010 and 2011. The potential forecasts and market participation strategies are the following:

- *case 1* Use of point forecasts for the whole portfolio directly defining bids in the electricity market,
- *case 2* Use of probabilistic forecasts for the whole portfolio, allowing for expectation-maximization bids as in Eq. (21),
- *case 3* Use of spatio-temporal scenarios, then allowing to find an optimal portfolio-wide bid in the spirit of Eq. (21).

The results from these three market participation cases, in terms of performance ratio and optimal deployment of wind energy capacities over the 5 aggregated zones are collated in Table 1.

TABLE 1: *Results in terms of optimal capacity deployment and corresponding performance ratio in the market for the 3 investment cases.*

case n°	γ	capacity deployment $\{P_n^{(j)*}\}$
1	0.969	$\{0, 0, 0, 0, 100\}$
2	0.971	$\{0, 0, 0, 0, 100\}$
3	0.977	$\{29.80, 2.32, 0.03, 23.88, 43.97\}$

First of all, these illustrative results confirm that the Nord Pool is not a highly penalizing market as recently confirmed by Girard *et al.* (2012b) when looking at the value of predictability in that same market. They also reveal that following intuition, the performance ratio increases when using probabilistic forecasts instead of point predictions, in line with previous work on participation in electricity markets with probabilistic forecasts of wind power generation e.g. Zugno *et al.* (2012). Note that the differences between performance ratio values for the various cases can be seen as significant in view of the length of the test cases and of previous studies. What is maybe more interesting is that in the case of not considering spatio-temporal dependencies, the optimal investment strategy then consists in putting all capacities in the same zone, which certainly is that which provide the best compromise between capacity factor and predictability, hence ensuring maximum revenues from the market. In contrast when space-time dependencies are accounted for, there is a tendency to spread capacities—one could say in order to diversify

exposure to market penalties. The three aggregated zones where the capacities are mostly spread (1, 4 and 5) corresponds to the Northern tip of Denmark, as well as its South-West and South-East corners, see Figure 1. The zone with the maximum capacity is the same than for the cases 1 and 2, i.e. zone 5.

4 Optimal wind power investments in a LMP market environment

The case of the zonal market in the above was to be seen as an introduction to the more complex case of the LMP market environment. In such an environment, the market clearing relies on a system model and the resulting market prices are not obviously the same for all locations of the system, depending on the power flows, transmission capacities and marginal costs of the participating units. A nice introduction to LMP markets is given in Litvinov (2010).

In the following, the experimental setup consisting of a thorough description of the system considered is first introduced. The methodology for the derivation of market participation strategy and optimal investment is described subsequently. Finally, applications results are presented and discussed.

4.1 Experimental setup

4.1.1 The PJM 5-bus modified system

A core part of this experimental setup consists in the power system model that is used for the following investment problem in an LMP market environment. Power system models can be quite large and therefore leading to high-computational costs when it comes to simulating LMPs over long periods and a large number of times (for optimization purposes). Therefore it was chosen to employ some of the smallest yet realistic power system model, which is that from the PJM training sessions originally published in 1999 and recently analysed by Li and Bo (2010). PJM stands for Pennsylvania, Jersey, Maryland, and is an abbreviation designating their common electricity market. It is represented in Figure 5. Some modifications were performed on the generation side to make the market more dynamic and competitive for all producers when wind power is present. More precisely, a few units were added at buses D and E. The basic set-up for that PJM 5-bus modified system in terms of is described in Table 2. Wind power generation in zones 1 to 5 in the above is allocated to buses A to E, with nominal capacities to be determined through the optimal investment problem. Since working a simplified version of an American system, let us assume that the monetary units are US dollars here.

Regarding the load, the configuration of the system is respected, with 30% of the total demand at bus B, 30% of the total demand at bus C and finally 40% of that total demand at bus D. The load profiles (both predicted at the moment of bidding and observed a posteriori) are obtained by scaling demand data obtained from Energinet.dk for the years 2010 and 2011, similarly to the market data used in the zonal market experiment in the above. The maximum load considered for this scaling is of 1432MW, in line with the generation mix and the systems constraints.

Based on this system set-up, LMPs are calculated using the DC optimal power flow approach, as discussed by Li and Bo (2007). So as to initialize the dynamical models for the LMPs described, sub-

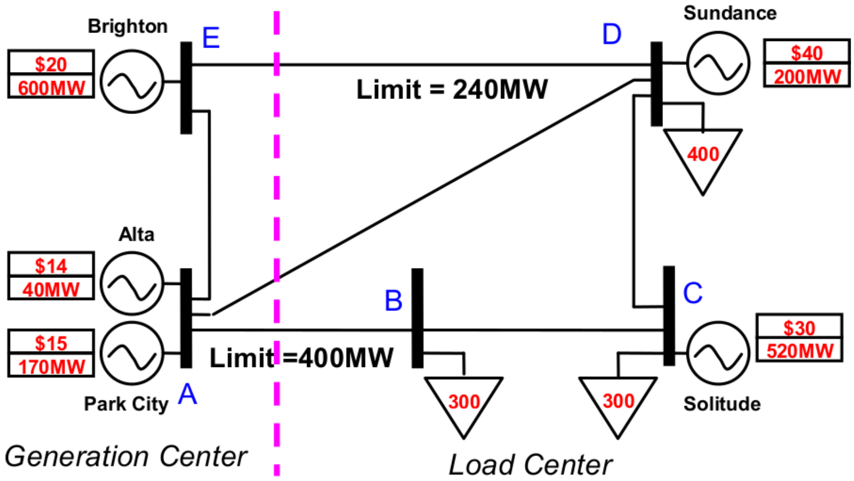


FIGURE 5: *The PJM 5-bus system as taken from Li and Bo (2010), though originally published as part of PJM training sessions documents.*

TABLE 2: *Set-up for the PJM 5-bus system.*

Bus	P_{min}	P_{max}	Marginal cost [$]
A	0	40	14
A	0	170	15
C	0	520	30
D	0	50	17
D	0	30	20
D	0	120	40
E	0	300	10
E	0	300	25

sequently used for forecasting at the time of bidding, hourly LMP calculations are performed over a 3-month prior period.

4.1.2 Dynamical model for LMP prices

Dynamical models are required for the prediction of LMPs at the time of proposing production offers in the market. For that purpose, the dynamical models of LMP prices, for both forward and real-time market, propose by Botterud *et al.* (2012) are employed. We employ notations similar to the case of the zonal market in the above, i.e. with π^c_{t+k} denoting the day-ahead price for the market time unit $t+k$, and with $\pi^{\uparrow\downarrow}_{t+k}$ the real-time LMP price. In contrast to zonal markets like Nord Pool, this LMP market environment is based on single pricing for the real-time market, hence explaining the change in notation.

The model for these prices write

$$\pi^c_{t+k} = \varepsilon_t \sigma_{\varepsilon,k} + \mu_{\varepsilon,k} \tag{24}$$

$$\pi^{\uparrow\downarrow}_{t+k} = \left(\rho_k \varepsilon_t + (1-\rho_k^2)^{\frac{1}{2}} \nu_t\right) \sigma_{\nu,k} + \mu_{\nu,k} \tag{25}$$

with ε and ν independent standard Gaussian random variables, $\varepsilon, \nu \sim \mathcal{N}(0,1)$. In the above, the means and variances of ε and ν are simply tracked in time by being estimating over a window in the past (1 month) prior to the day of interest. This may be seen as a form of exponential smoothing for the characteristics of a bivariate Gaussian process. This somehow comprise a naive model for the determination of future LMPs, though it will be sufficient for the purpose of this investment problem study.

It is to be noted that this dynamic model for the LMPs is only to be used by the wind power producer at the time of bidding in the day-ahead LMP market. After that, actual day-ahead and real-time LMPs are determined based on a classical methodology using a DC optimal power flow (DCOPF), as presented by Ott (2003) among others. For the day-ahead LMPs, load forecasts are to be used, while for the real-time ones the realized load is employed instead. A more extensive description of these aspects is given in Delikaraoglou (2012).

4.2 Methodology

In a LMP market environment, the revenue formulation for a participant in the market is different from that in the zonal market case, since prices may be directly linked to where your production unit are located and how they may impact power flows over the system. We therefore introduce below the formulation of the revenue of a participant in an LMP market. Subsequently, the design of alternative bidding strategies is considered. Finally, the overall methodology for the investment problem analysis is presented in detail.

4.2.1 Revenue of a participant in an LMP market

Using the same notations than in Section 3.1 and following the formulation of Botterud *et al.* (2012), the revenue of a participant in an LMP market environment, for a market time unit $t+k$, is

$$\mathrm{R}_{t+k} = \pi^c_{t+k} E^c_{t+k} \; + \; \pi^{\uparrow\downarrow}_{t+k} (E^*_{t+k} - E^c_{t+k}) - \tau |E^*_{t+k} - E^c_{t+k}| \tag{26}$$

It is therefore composed by a revenue from the day-ahead market, where the energy quantity E^c_{t+k} is contracted at a unit price π^c_{t+k}. Then the difference between contracted and actual production, $E^*_{t+k} - E^c_{t+k}$, is settled through the real-time market at a unit price $\pi^{\uparrow\downarrow}_{t+k}$. Finally, there is an additional penalty τ for unit deviation between day-ahead and real-time contracts.

Similarly to the case of the zonal market, this may be reformulated in a more useful manner for subsequent use in a stochastic optimization framework. Based on the imbalance (d^*_{t+k}) definition of Eq. (11), the revenue of Eq. (26) for market time unit $t+k$ may rewritten as

$$\mathrm{R}_{t+k} = \pi^c_{t+k} E^*_{t+k} \; + \; (\pi^{\uparrow\downarrow}_{t+k} - \pi^c_{t+k}) d^*_{t+k} - \tau |d^*_{t+k}| \tag{27}$$

This revenue has a fatal part from selling actual wind generation in the day-ahead market, and a stochastic part linked to market prices and deviations between contracted and actual delivery of energy.

It is not possible here to introduce a performance ratio as in Eq. (17) for zonal market case, since even for perfect forecasts, the overall revenues may be affected by the wind power penetration in the system itself. In a similar vain, we introduced an overall revenue criterion, which we similarly denote γ_{lmp},

$$\gamma_{\mathrm{lmp}} = \sum_{t} \sum_{k} \mathrm{R}_{t+k} \tag{28}$$

over the set of market times unit for every day, and over the number of days considered in the market participation simulation.

4.2.2 Derivation of bidding strategies

In a LMP-market environment, the design of optimal bidding strategies is of similar nature than in a zonal market, though the actual optimization could become more complex owing to the interdependence between wind and prices. Disregarding any potential risk aversion, the expected utility maximization bid is formulated as

$$\tilde{E}^c_{t+k} = \underset{E^c_{t+k}}{\arg\min}\, \mathbb{E}\left[(\pi^c_{t+k} - \pi^{\uparrow\downarrow}_{t+k})d^*_{t+k} + \tau|d^*_{t+k}|\right] \tag{29}$$

which is somewhat equivalent to

$$\tilde{E}^c_{t+k} = \underset{E^c_{t+k}}{\arg\max}\, \mathbb{E}\left[\mathrm{R}_{t+k}\right] = \underset{E^c_{t+k}}{\arg\max}\, \mathbb{E}\left[\pi^c_{t+k}E^*_{t+k} \;+\; (\pi^{\uparrow\downarrow}_{t+k} - \pi^c_{t+k})d^*_{t+k} - \tau|d^*_{t+k}|\right] \tag{30}$$

since here again the first part of the revenue is fatal for the wind power producer participating in the LMP market. The optimization problem in Eq. (29) may be solved analytically (similarly to the case of the zonal market, though the unit regulation penalties for up and down-regulation are more sophisticated) or numerically based on an optimizer and a set of scenarios for representing the uncertainty in wind power production. Remember that here, such scenarios for the total wind portfolio are to be obtained as a summation of scenarios for each of the 5 zones.

4.2.3 Formulation of the investment problem

The investment problem is seen in a manner similar to the case of zonal markets. Here however, the computational burden is increased significantly owing to the necessity to run a DCOPF calculation every time the day-ahead and real-time LMPs are to be determined, based on contracted and actual wind power production. The overall investment problem methodology is illustrated in Figure 6. What is of importance here is the outer loop in this flowchart, consisting of optimizing the nominal capacities at the various nodes of the system. In parallel the inner loop corresponds to the daily cycling with market participation consisting of optimal bidding and determination of day-ahead quantities, followed by actual operations and determination of actual dispatch and real-time prices.

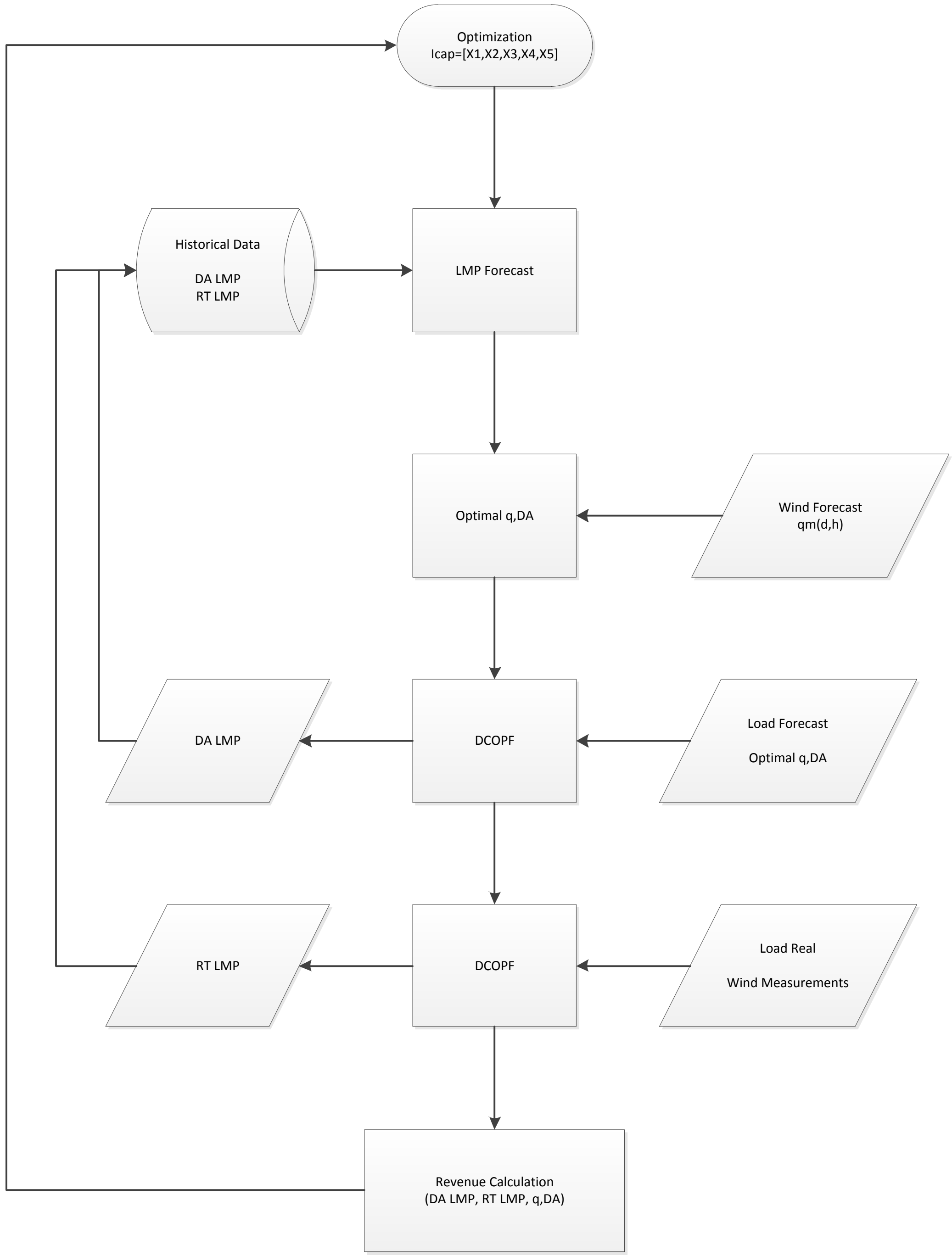


FIGURE 6: *Flowchart of the methodology used for the investment problem in the LMP market environment.*

The overall optimization problem is as in Eq. (23), i.e. consisting in optimally spreading wind power capacities over the 5 aggregated zones, $P_{\mathrm{n}}^{(j)}$, $j = 1, \ldots, 5$, for a given overall nominal capacity P_n, while accounting for daily market participation over the period,

$$\{P_n^{(j)*}\} = \underset{\{P_n^{(j)}\}}{\arg\max}\, \gamma_{\mathrm{lmp}}(\{P_n^{(j)}\}; \mathcal{S}, \mathcal{F}) \tag{31}$$

with $\mathcal{F}$ and $\mathcal{S}$ generically denote the chosen type of input forecasts and market participation strategies, respectively. Here only the full set of spatially correlated scenarios are considered, and the expected utility maximization market participation strategies. Other types of forecasts and market participation strategies could be employed and used as a basis for the investment problem in the future.

The optimization problem is solved numerically also looking at different initial points for the optimizer, in order to look at a variety of potentially local maxima, and decide upon an overall maximum accordingly. More precisely, the Matlab built-in function fmincon, which is designed for minimization of a constrained nonlinear multivariable objective function, is employed. Different approaches for the optimization to be performed by this function can be chosen. The interior-pint approach was chosen, after comparison with the active-set one (and an alternative SPSA approach - Simultaneous Perturbation Stochastic Approximation). Since these various methods are not designed for global nonlinear optimization, a heuristic was employed consisting in solving the optimization problem with a set of different starting points (15), to find a 'best' optimum, while assessing the robustness of the optimization results.

4.3 Application results

The results are given for penalty 10 and for a period of 30 days (day 125 to 154 starting from the beginning of January 2006). The short length of the optimization period is due to the computational burden involved. It could be extended in the future if simplifying the inner loop calculations (for instance by avoiding the DCOPF part). Three levels of overall nominal capacity P_n are considered: 100, 300 and 500 MW, hence gradually increasing the stress on the system. Results covering the geographical distributions of these nominal capacities are gathered in Table 3.

TABLE 3: *Optimal portfolio spreading for various nominal capacities. The unit revenue corresponds to the total revenue normalized by the nominal capacity P_n of the wind portfolio*

P_n	Bus A	Bus B	Bus C	Bus D	Bus E	Total Revenue [k$]	Unit Revenue [k$]
100	0	0	0	100	0	351	3.51
300	91.16	0.25	0.31	206.31	1.96	923	3.08
500	406.19	0	0.03	93.72	0.06	1401	2.80

In the case of limited nominal capacity of the wind portfolio (100MW) the solution is quite trivial and easy to obtain whatever the initial point of the optimizer. This optimal portfolio distribution relates to the fact that placing all capacity at bus D allows to get the highest LMP thanks to the other generators connected to this node while being also directly connected to a load center. Somehow the other unit sets the price at this node while the wind power generation is ensured to be dispatched locally. This also

yields the highest unit revenue, corresponding to the revenue obtained over that month normalized by the nominal capacity P_n.

When increasing the nominal capacity P_n to be geographically distributed, the situation gets more complex both in terms of optimization and of intuitive explanation of the results. Buses A and D are privileged buses for both $P_n = 300$ MW and $P_n = 500$ MW, even though the optimal share to allocate to one or the other depends on the overall wind production. The fact that these buses are preferred location can be explained by the fact the correlation in generation and forecast errors for these two locations (zones 1 and 4) is minimal. The unit revenue per MW installed decreases as P_n gets greater, as the result of a downward pressure of wind power generation on market prices.

5 Conclusions and perspectives

The issue of an optimal geographical distribution of wind power capacities was looked at with particular focus on aspects of market participation. Both zonal and LMP markets were considered, in order to underline the differences in the methodology of defining market participation strategies, simulation of market participation and finally portfolio optimization. In contrast with other studies defining an optimization problem with a static view of the stochastic processes involved, it was chosen here to fully simulate the dynamics and operational aspects based on real-world wind power forecasts and observations. In the zonal market case, prices from the Nord Pool were employed, while in the LMP market case, it was necessary to build and run our own market since the market-clearing and resulting prices is to be system-dependent. Our main objective was to study the importance of various types of forecast information at the investment stage, therefore going further than considering the wind resource only. Aspects of transmission capabilities, permissions, etc. were disregarded so as to focus on the wind characteristics themselves.

A number of interesting findings were gathered. First of all in a zonal market environment, not considering the spatio-temporal dynamics of wind generation and their uncertainty at the investment stage will most certainly result in placing all capacities in a most favourable location. That location indeed provides the most appropriate balance between capacity credit and imbalance penalties to be settled through the real-time market (hence linked to forecast quality). This is in line with recent findings of Girard *et al.* (2012b) who showed that capacity credit, and to a lesser extent forecast accuracy, was a main driver for optimal wind power investment in a zonal market like the Nord Pool. But when going into more detail with the spatio-temporal dynamics of wind power generation, it appears that the spatial pattern of the optimal portfolio is diversified, also allowing for an increased revenue from the market. Note that wind was considered as a price taker and that this assumption may be refined in the future so as to make the investment results more realistic. The impact of wind power generation on prices could be simulated along the lines of the findings of Jónsson *et al.* (2010). The general qualitative conclusions are expected to be the same though: wind resource alone is not the optimal input information at the investment stage in a zonal market environment.

In parallel in an LMP market environment, emphasis was only placed on the most detailed type of wind power dynamics i.e. with spatio-temporal scenarios, since the base case of not considering the spatial correlation in wind dynamics would obviously lead to worse results in terms of revenues. This is since

the determination of LMP prices is based on power flow calculations, hence being directly impacted by the joint influence of wind power infeed at various buses. Wind power generation is therefore naturally considered as a price-maker. This comes at the cost of greater computational needs. The analysis performed showed that the optimal portfolio results may be less intuitive even though capacity credit and (predicted) correlation would clearly contribute to the optimal results. Additionally, the topology of the system itself, as well as the locations and characteristics of the other generation units, plays a key role since setting the prices that will drive the revenues of the wind power producer. For larger nominal capacities, all these aspects combined are to influence the optimal geographical distribution of wind power portfolio. It appeared here it would be very difficult to optimally spread a wind power portfolio while disregarding advanced forecast and predictability information.

Overall for both market environments, the appraisal and modeling of forecast uncertainty and spatio-temporal dynamics were shown to be of interest, since forming a good basis as input to optimal wind portfolio investment problems. Perspectives for future work should focus on finding ways to alleviate computational costs in order to perform longer and larger simulations in the LMP case. Also the formulation and set-up of the optimization problem may be revised so as to better ensure to obtain an optimal portfolio solution without having to use any form of heuristics. Finally, different forms of market clearing may be looked at in the future, in order to see how sensitive optimal investment decisions may be.

References

Baringo L, Conejo AJ (2011). Windpower investment within a market environment. *Applied Energy* **88**(9): 3239-3247.

Baringo L, Conejo AJ (2012). Wind power investment: a Benders decomposition approach. *IEEE Transactions on Power Systems* **27**(1): 433–441.

Barthelmie RJ, Murray F, Pryor SC (2008). The economic benefit of short-term forecasting for wind energy in the UK electricity market. *Energy Policy* **36**(5): 1687–1696.

Boomsma TK, Meade N, Fleten S-E (2012). Renewable energy investments under different support schemes: A real options approach. *European Journal of Operational Research* **220**(1): 225–237.

Botterud A, Zhou Z, Wang J, Bessa RJ, Keko H, Sumai J, Miranda V (2012). Wind power trading under Uncertainty in LMP markets. *IEEE Transactions on Power Systems* **27**(2): 894–903.

Bremnes JB (2004). Probabilistic wind power forecasts using local quantile regression. *Wind Energy* **7**(1): 47–54.

Butler L, Neuhoff K (2008). Comparison of feed-in tariff, quota and auction mechanisms to support wind power development. *Renewable Energy* **33**(8): 1854-1867.

Chang J, Ummels BC, van Sark Wilfried GJHM, den Rooijen HPGM, Kling WL (2009). Economic evaluation of offshore wind power in the liberalized Dutch power market. *Wind Energy* **12**(5): 507–523.

Chen C, Lu Z (2011). Analysis on the strategy of wind power investment under uncertainty. *International Journal of Nonlinear Science* **12**(1): 112–116.

Delikaraoglou S (2002). Optimal wind portfolio investment in view of uncertainties in power predictions and market dynamics. M.Sc. thesis, Technical University of Denmark, Department of Informatics and Mathematical Modelling.

Girard R, Allard D (2012). Spatio-temporal analysis of wind power prediction errors. *Wind Energy*, in press.

Girard R, Laquaine K, Kariniotakis G (2012). Assessment of wind power predictability as a decision factor in the investment phase of wind farms. *Applied Energy*, in press.

Jónsson T, Pinson P, Madsen (2010). On the market impact of wind energy forecasts. *Energy Economics* **32**(2): 313–320.

Jónsson T, Pinson P, Nielsen HAa, Madsen H (2012). Exponential smooting approaches for prediction in real-time

electricity markets. Working paper, under review.

Li F, Bo R (2007). DCOPF-based LMP simulation: algorithm, comparison with ACOPF, and sensitivity. *IEEE Transactions on Power Systems* **22**(4): 1475–1485.

Li F, Bo R (2010). Small test systems for power system economic studies. *Proc. of the 2010 IEEE Power and Energy Society General Meeting*, Minneapolis, Minnesota (US).

Litvinov E (2010). Design and operation of the locational marginal prices-based electricity markets. *IET Generation, Transmission & Distribution* **4**(2): 315–323.

Madsen H, Pinson P, Nielsen TS, Nielsen HAa, Kariniotakis G (2005). Standardizing the performance evaluation of short-term wind power prediction models. *Wind Engineering* **29**: 475–489.

Morales JM, Conejo AJ, Mínguez R (2010). Short-term trading for a wind power producer. **25**(1): 554-564.

Morales JM, Mínguez R, Conejo AJ (2010). A methodology to generate statistically dependent wind speed scenarios. *Applied Energy* **87**(3): 843-855.

Morales JM, Pinson P, Madsen H (2012). A transmission-cost-based model to estimate the amount of market-integrable wind resources. *IEEE Transactions on Power Systems* **27**(2): 1060–1069.

Møller JK, Nielsen HAa, Madsen H (2008). Time-adaptive quantile regression. *Computational Statistics and Data Analysis* **52**(3): 1292–1303.

Nielsen TS, Madsen H, Nielsen HAa (2002). Prediction of wind power using time-varying coefficient functions. in *Proceedings of the IFAC 2002, 15th World Congress on Automatic Control*, Barcelona, Spain.

Nielsen TS (2002). Online prediction and control in nonlinear stochastic systems. Ph.D. thesis, Technical University of Denmark, Department of Informatics and Mathematical Modelling.

Ott AJ (2003). Experience with PJM market operation, system design, and implementation. *IEEE Transactions on Power Systems* **18**(2): 528–534.

Papaefthymiou P, Pinson P (2008). Modeling of spatial dependence in wind power forecast uncertainty. Proc. of PMAPS08, IEEE Conference on Probabilistic Methods Applied to Power Systems, Rincon, Puerto Rico.

Pinson P, Chevallier C, Kariniotakis G (2007). Trading wind generation from short-term probabilistic forecasts of wind power. *IEEE Transactions on Power Systems* **22**(3): 1148–1156.

Pinson P, Girard R (2012). Evaluating the quality of scenarios of short-term wind power generation. *Applied Energy* **96**: 12–20.

Pinson P, Kariniotakis G (2010). Conditional prediction intervals of wind power generation. *IEEE Transactions on Power Systems* **25**(4): 1845-1856.

Pinson P, Nielsen HAa, Møller JK, Madsen H, Kariniotakis G (2007). Nonparametric probabilistic forecasts of wind power: required properties and evaluation. *Wind Energy* **10**(6): 497–516.

Pinson P, Papaefthymiou G, Klöckl B, Nielsen HAa, Madsen H (2009). From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy* **12**(1): 51-62.

Zugno M, Jónsson T, Pinson P (2012) Trading wind energy on the basis of probabilistic forecasts of both wind generation and market quantities. *Wind Energy*, available online.