Project n^{o}: 239456

Project acronym

OPTIMATE

Project title:

An Open Platform to Test Integration in new MArkeT designs of massive intermittent Energy sources dispersed in several regional power markets

Instrument: Collaborative project

Start date of project: 1st October 2009

Duration: 36 months

D3.1
Assumptions on accuracy of wind power to be considered at short and long term horizons

Revision: Final version

Due date of delivery: 2010-07-31

Actual submission date: 2010-07-30

Organisation name of contractor(s) for this deliverable:

RISOE

Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Document information

Identification

Deliverable number:	**D3.1**
Document name:	Wind Power accuracy and forecast
Revision version, date	Final version, 23rd July 2010
Authors	PE Morthorst, JM Coulondre, ST Schröder, P Meibom (RISOE)

General purpose

This report is deliverable 3.1 of the Optimate-project. The report describes briefly state-of-the-art forecasting of wind power and summarises the use of prediction errors in the Optimate-model. Finally, a simple and an advanced methodology of generating forecast-errors to Optimate is described.

Deliverable number:	**D3.1**
Deliverable title:	Assumptions on accuracy of wind power to be considered at short and long term horizons
Work package:	WP3
Lead contractor:	RISOE

Quality Assurance

Status	By	Date
Verified by Coordinator	Athanase Vafeas, Technofi	2010-07-23
Verified by Technical director	Jean-Marie Coulondre, RTE	2010-07-20
Submitted by Coordinator	Athanase Vafeas, Technofi	2010-07-30

Table of Content

List of figures

List of tables

1. Introduction

1.1. The Optimate project

Optimate is an **O**pen **P**latform to **T**est **I**ntegration in new **Ma**rke**T** designs of massive intermittent **E**nergy sources dispersed in several regional power markets. Optimate is coordinated by Technofi and includes in total 12 partners. Five transmission system operators are involved: RTE (France), ELIA (Belgium), EnBW TSO (Germany), 50 Hertz Transmission (Germany) and REE (Spain). Finally, six universities and research institutions are included: ARMINES (France), K.U.Leuven (Belgium), Risø-DTU (Denmark), Comillas Pontifical University (Spain), University of Manchester (UK) and SEAES (University of Paris). Optimate was started by the end of 2009 and expected to be finalised by the end of 2012. Optimate is supported by the EU 7th framework programme.

Today most of the EU-27 Member States are utilizing electricity markets in their dispatch of power. These power markets are in general designed according to the needs of the conventional electricity system, that is to take into account the characteristics of the most widely used generation units as hydro power, thermal power plants and nuclear. However, the future development points to an intensive introduction of renewable energy sources, to a certain extent relying on variable sources as wind power and photovoltaics. These renewable sources have characteristics that do not easily fit into the current electricity market frameworks, especially because of their intermittent generation. The day-ahead forecasts for the production of power from these plants are subject to some uncertainty and it is risky for them to bid into the spot market because they have to pay penalties when they do not fulfil their commitments. Thus new market designs are needed which facilitate the introduction of variable renewable power sources.

The main objective of Optimate is to develop a new tool for testing these new market designs with large introduction of variable renewable energy sources. Under the technical coordination of the French TSO (RTE) a novel network/system/market modelling approach is being developed, generating an open simulation platform able to exhibit the comparative benefits of several market design options. The Optimate-model will take as its starting point the regional differences in the EU-area emphasising the geographical dispersion of renewable energy sources, the timing and interconnections between the different power markets (day-ahead, intraday and balancing markets) will be modelled and network constraints will be taken into account.

An integral part of the project is to develop new markets designs that facilitate a massive introduction of intermittent renewable sources. Utilizing the Optimate modelling framework these new market designs will be tested and evaluated not only according to their ability to handle large amounts of renewable sources but also to see if these designs fulfil the needs of other market participants and TSO’s.

1.2. This report as part of Optimate

The accuracy of wind power forecast is part of Optimate Work package 3 on day-ahead and intraday markets and this report constitutes Deliverable 3.1: Assumptions on accuracy of wind power to be considered at short and long term horizons.

As the accuracy of wind power forecast at short term horizons (Day minus two to real time) is an important input to the Optimate model it is chosen to describe this part in more detail. The report includes three separate parts, although to a certain extent interrelated ones:

- A state-of-the-art overview on prediction errors based on existing literature is given in chapter 2.
- How prediction errors enter into the Optimate modelling framework is described in chapter 3.
- How prediction errors are to be handled in relation to Optimate is described in chapter 4, where both a simple approach and a more advanced approach are suggested.

1.3. Acknowledgement

A special acknowledgement is given to Gregor Giebel, Risø DTU, especially because of his contribution to the State-of-the-art literature overview.

2. State-of–the -art

2.1. Introduction

Defining short-time horizons as covering the time period of 0 – 48 hours ahead reflects the requirements of today's electricity markets. A number of actors, e.g. transmission system operators (TSOs), energy traders and operators of virtual power plants, depend on forecast precision for determining issues as power plant dispatch. A longer forecast horizon of up to one week would be beneficial for maintenance planning (Giebel et al., 2010). As the largest share of economic consequences is related to energy markets, this section focuses on the period of up to 48 hours ahead. It is to be seen as an introduction to short-term forecasting and obviously, the field cannot be grasped in its full extension. For greater detail, the reader might refer to Giebel et al. (2010) and the references cited therein, e.g. Lange and Focken (2005). Another in-depth state-of-the-art overview is provided by Monteiro et al. (2009).

Forecasts can be based on physical or statistical approaches, or a combination of both. Figure 1 provides a conceptual representation of forecast quality based on Numerical Weather Prediction (NWP, physical approach). The background error is due to a suboptimal representation of a single site in the model. It is the error until the first look-ahead time (note that the forecast horizon does not start at 0).

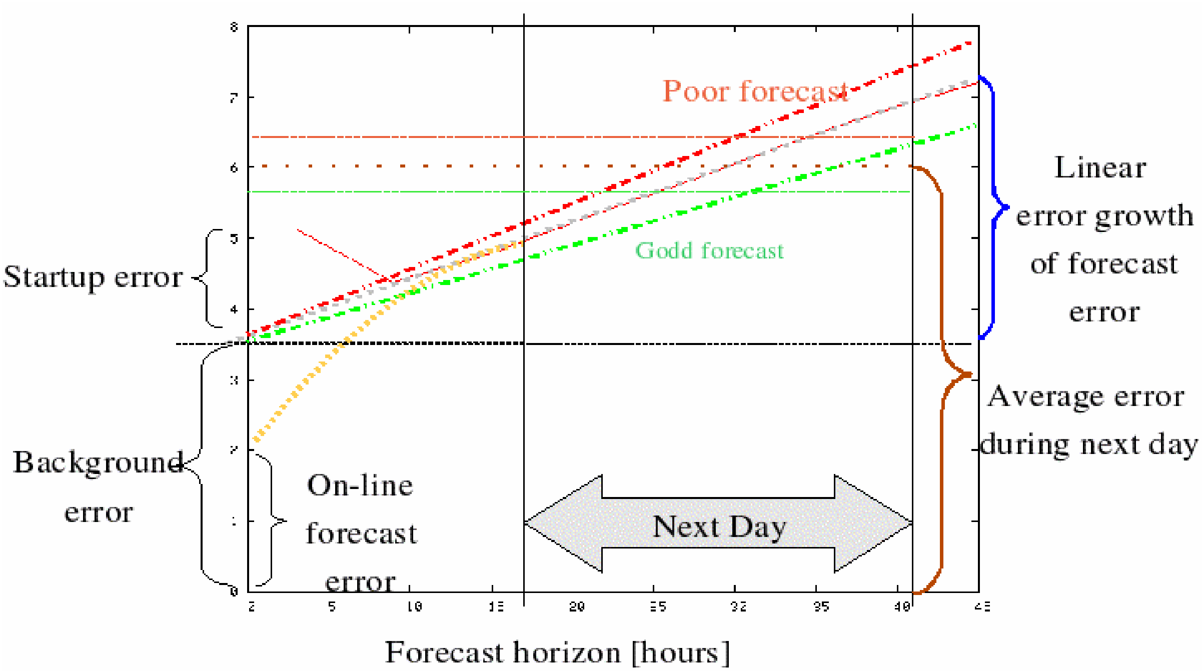


Figure 1: Typical errors introduced by the NWP. Source: WEProg.com, cited after Giebel et al. (2010)

With an increasing time horizon, the forecast error grows due to increased uncertainty about atmosphere physics conditions. The model can be improved by a better representation of physical processes in the atmosphere, but general chaotic behaviour of the atmosphere remains as an error source. Depending on the forecast quality, the average error during the next day can be reduced. The linear characteristics of error growth over time facilitate reading the figure, but is not necessarily representative.

2.2. Model stages

In the very short term, the reference for models is the persistence approach, i.e. that the wind remains unchanged. This is due to the fact that the minute time scale cannot be represented by atmospheric models. For forecasts exceeding 3-6 hours of time horizon, models using a Numerical Weather Prediction provide commonly better results than time-series approaches. Most models combine an NWP approach with statistical models. A full procedure for a regional wind forecast could consist of the following parts:

1. The NWP is scaled to the turbines hub height. A downscaling model accounts for local specificities.
2. Using power curves or statistical estimations, wind speed is converted to power output.
3. Upscaling provides results for the region, instead of single sites.

Depending on the scope of the model, single exemplary steps can be neglected, simplified or supplemented: for a single wind farm, the regional upscaling step is not necessary.

2.3. Forecast quality criteria

Depending on the scope of the forecast, different criteria can be applied. “Common error measures are the bias, MAE [Mean Absolute Error], RMSE [Root Mean Square Error], the coefficient of determination R^2, the skill score [...] and the error distribution as a histogram” (Giebel et al., 2010, referring to Madsen et al., 2004). In Spain, the Mean Absolute Percentage Error (MAPE) stems from support scheme legislation. For single sites, the forecast error increases with more complex terrain. The RMSE is the chosen error criterion in the following part.

2.4. Exemplary results

Figure 2 shows an example of a day-ahead forecast and actual values for a German TSO. Due to a national burden-sharing mechanism, it is a share of total national wind production.
In this example, the wind forecast tool underestimated the actual generation. It is also shown that the wind forecast error is within the same magnitude as the load forecast error, though with smaller gradients. In the first hours, the two errors compensated for each other, whereas they are additive in the remainder of the regarded time period.

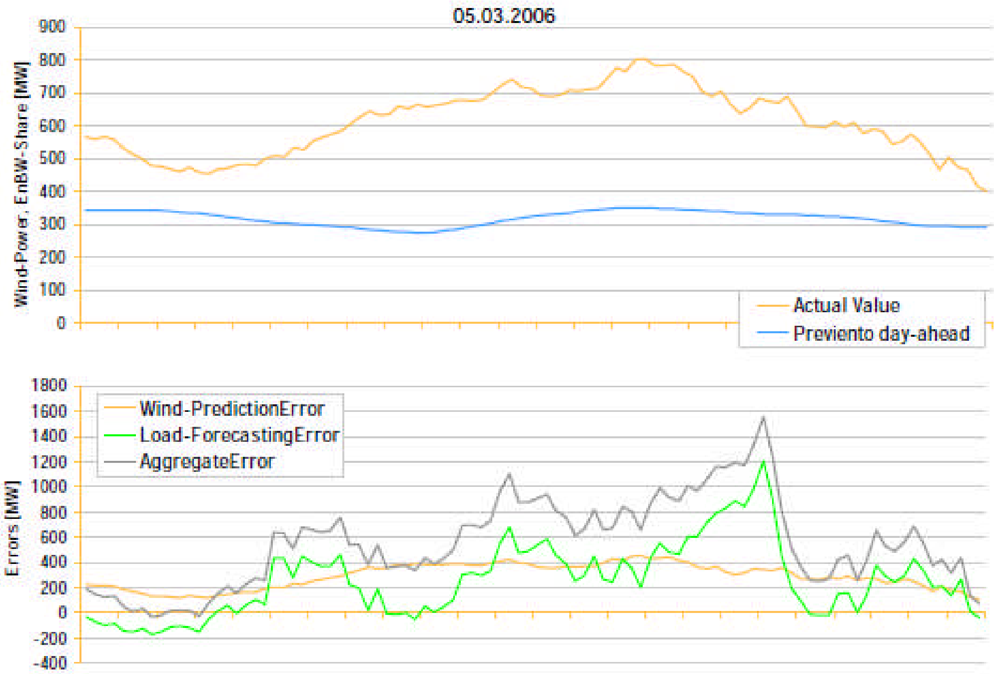


Figure 2: The forecast error for the area of EnBW for the day-ahead forecast of the Previento model (Krauß, 2006, cited after Giebel et al., 2010)

Figure 3 illustrates time-dependent forecast precision for a single site in Ireland. It can clearly be seen that the forecast error is much lower for the first look-ahead time than for the period 24 hours ahead. Furthermore, the forecast error is distributed symmetrically.

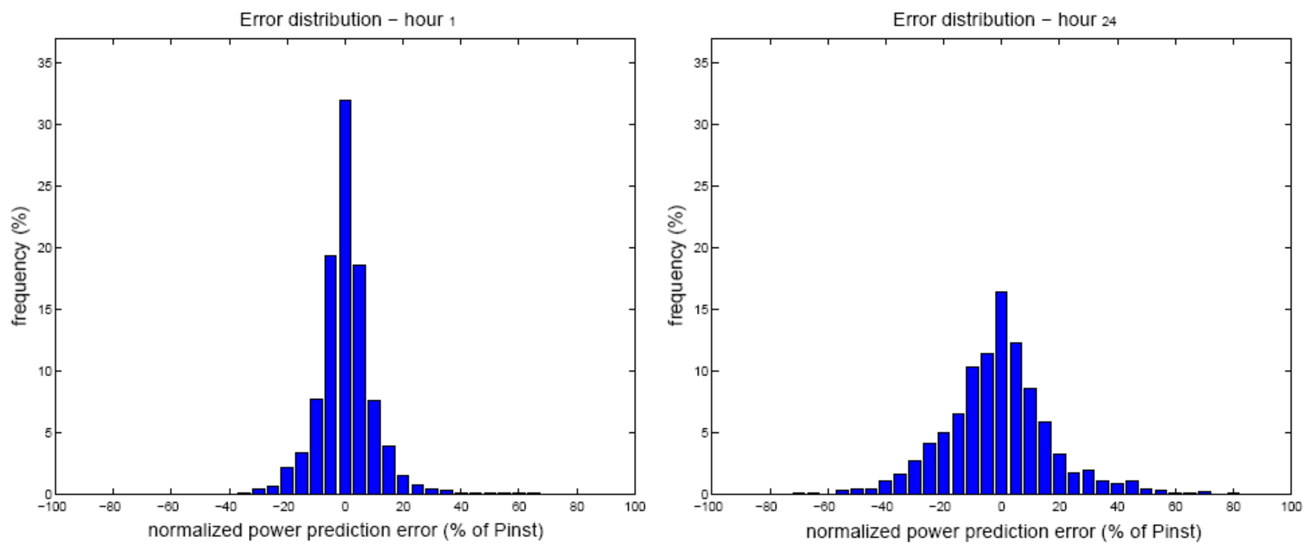


Figure 3: Normalized prediction error distributions for the first look-ahead time (left) and for lead time 24 (right) (Madsen et al., 2004)

Table 1 gives an overview of RMSE values for Germany. Over the whole area, the error is considerably smaller than for a single control zone. In addition, a few hours shorter time horizon decreases the error significantly. Note that Tastu et al. (2010) achieve better results for a number of small regions in West Denmark. Therefore, it can be concluded that RMSE values of 3% or less are state-of-the-art.

Table 1: Level of accuracy of wind power predictions in Germany (NRMSE = normalized root mean square error, % of installed wind capacity; Rohrig, 2005, cited after Giebel et al., 2010)

NRMSE [%]	Germany (all 4 control zones) ~1000 km	1 control zone ~ 350 km
day-ahead	5.7	6.8
4h ahead	3.6	4.7
2h ahead	2.6	3.5

The concept of the background error introduced in figure 1 could be applied within the OPTIMATE model such that it separates the forecast error for the day-ahead market into two categories: a basic background error term and a further additive term for additional hours, possibly even linear. It is reasonable to argue that such an error division has a simpler mathematical formulation than a forecast error model covering the whole forecast horizon. Analogously, the intraday market might be approached with a similar concept and a smaller background error. Depending on the modelling approach, calculation times and desired precision for forecast errors, these aspects need to be weighted carefully.

3. The use of prediction errors in Optimate

3.1. How prediction errors enter into the modeling framework of Optimate

A central assumption of Optimate is that all stakeholders have access to the same forecast: either they use the same tools, or the forecast of the most accurate tool is published and used by everybody.

For each current day, Optimate links together iteratively two processes, the DA process and the ID process.

3.1.1. The DA process

In the DA process, the whole 24 hours of the current day are dealt considering a unique DA forecasts of load and intermittent generation:

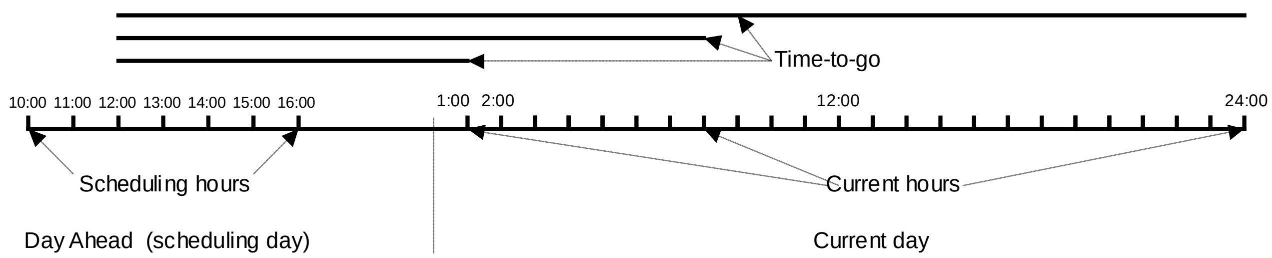


Figure 4: Day-ahead schedule

The forecast is assumed to be the one made at DA 12:00, on the scheduling day, for the 24 hours of the current day. In other words, the time-to-goes range from 13 to 36 hours. This DA 12:00 forecast is used in each module of the DA process, whatever scheduling hour corresponds to the module in the real world (the modules' real scheduling hour range between DA 10:00 to DA 16:00).

But the decision simulated at DA takes into account not only the DA forecast average values, but also the standard deviation of their errors. Each stakeholder has indeed a certain degree of risk aversion, which means for instance that she wants to make the decision that hedges her against x% of possible case due to forecast errors.

In terms of congestion management, a TSO stakeholder is looking at the distortion between nodal effect and zonal effect. To assume the risk of deviating from the DA forecast, each TSO assesses a mixed standard deviation of each cluster[1] combining those of wind power forecast error, photovoltaic forecast error, load forecast error and generation sudden outage risk error. This combination can be done straightforwardly assuming that wind, P, load and generation outage errors are not correlated.

[1] In Optimate, cluster of nodes are used to mimic the nodal effects

In terms of balancing requirements, a TSO stakeholder is looking at its Control block[2]. To assume the risk of deviating from the DA forecast, each TSO assesses a mixed standard deviation of her Control block combining those of wind power forecast error, photovoltaic forecast error, load forecast error and generation sudden outage risk error. This combination can be done straightforwardly assuming that wind, P, load and generation outage errors are not correlated.

This information is key for Control block Reserve requirements. Assuming that this information is either published, or known through experience, it is also key for Marketers: it provides an indication of volume of the physical IntraDay upward / downward[3] market they should partly[4] anticipate when making their DA unit commitment.

But each of those DA standard deviations (for wind, PV, load and generation) per Control Block cannot be straightforwardly assessed from standard deviations per cluster, as they are related by geographical cross-correlations. Therefore DA forecast error standard deviations must be assessed both per cluster and per Control Block to feed in Optimate Simulator.

DA forecast error standard deviations plays also a role when aggregated by Balance Responsible Party: depending on its Balance Responsible Perimeter (BRP) a Marketer tends to prefer to be slightly imbalanced in proportion of the imbalance prices asymmetry. In theory this effect would need to assess DA forecast error standard deviations also per BRP, which once again could not be straightforwardly assessed from standard deviations per cluster, as they are related by geographical cross-correlations. In practice, the BRP definition itself is already a proxy in Optimate[5], then some proxy will be used to assess DA forecast error standard deviations per BRP out of the other simulation input data. Some guidelines for building this proxy are given in this document.

3.1.2. The ID process

In the ID process, the next 8 hours from any scheduling hour are dealt considering successive hourly forecasts of load and intermittent generation.

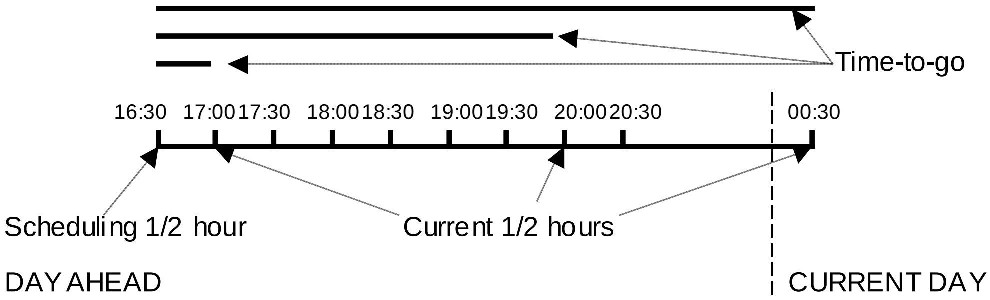
---

[2] A Control block currently correspond approximately to a country. Some evolutions are under discussion but Optimate first version will stick to the current situation

[3] Optimate simulates only the upward anticipation of marketers, the downward one being basically fulfilled due to other assumptions made

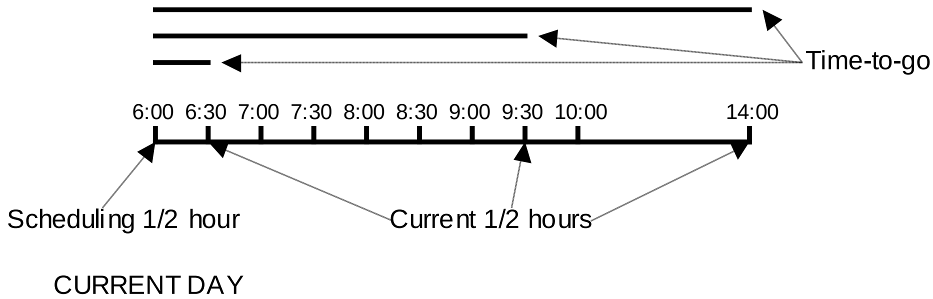
[4] Assuming a non strategic behaviour, they will anticipate the part of it which corresponds to the quantity they can expect to supply.

[5] The BRP definition is commercial data out of reach of the Project. Moreover, it is not a persistent data over time and its anticipation at year 2015 would anyhow be questionable.



The first scheduling hour dealt by the Intra-Day process is 16:30 at DA, as some starting up /shutting down decisions can be made then for 00:30 on the current day.

Then the ID process iterates on the scheduling hour per ½ hourly steps. The time-to-go period analysed at a scheduling hour of 6 am would be for instance:



Each current half hour benefits from its own Forecast, which is becoming more and more accurate when the time-to-go decreases (in other words, its error standard deviation decreases). The time consistency of those successive forecasts is to be ensured in between them, with the DA forecast previously used, and with the RT output used at the very end of the ID process.

Standard deviation of errors being used to build consistent sets of successive forecasts, their consideration is enough in Optimate simulation to capture the anticipative behaviour of stakeholder: at each scheduling hour, Optimate simulates that each stakeholder considers the successive forecast of each next eight hours to make her decisions.

3.2. Intermittent generation data need in Optimate

3.2.1. Input data for simulation

The following input data are needed per cluster and per (half) hours over the whole year:

- DA forecast of Wind power generated
- DA forecast of Photovoltaic power generated
- DA standard deviation of Wind power forecast error
- DA standard deviation of Photovoltaic power forecast error
- ID forecasts of Wind generation for time-to-go ranging from ½ hour to 8 hours
- ID forecasts of Photovoltaic generation for time-to-go ranging from ½ hour to 8 hours

* RT outputs of Wind generation
* RT outputs of Photovoltaic generation

The following input data are needed per Control Block and per (half) hours over the whole year:

* DA standard deviation of Wind power forecast error
* DA standard deviation of Photovoltaic power forecast error

The following input data are needed per Balance Responsible Party and per (half) hours over the whole year:

* DA standard deviation of Wind power forecast error
* DA standard deviation of Photovoltaic power forecast error

Those last inputs are assumed to be assessed from the previous ones considering parameters such as installed capacities of each BRP per clusters, in each Control area. This calculation is part of the data pre-processing.

4. Methodology to handle prediction errors to be used in Optimate

In Optimate forecasts and standard deviations of these forecasts are needed for DA and ID for wind power and photovoltaics. In this chapter only wind power will be considered while photovoltaics will be reported separately[6]. The shown methodologies will be capable of handling both the DA and ID forecasts, although emphasis is on DA.

4.1. A simple methodology for generating forecast errors to Optimate

4.1.1. Parameters to be taken into account in the simple approach

Optimate is subdivided into a number of clusters and for each of these clusters predictions for wind power production is needed alongside the actual realised wind power production. To generate these predictions the starting point in this section is a generation of prediction errors. When these errors are added to the realised time-series for wind power production a total time-series for prediction is achieved.

Within the clusters as defined by Optimate the prediction error will depend on a number of different issues:

- The installed capacity of wind power within the cluster (MW)
- The area size of the cluster (km2)
- The wind conditions within the cluster.

Moreover, in a simple model for generating prediction errors it has to be taken into account that a timely correlation exists (wind speed in hour t+1 will be correlated to wind speed in hour t), but also a spatial correlation between the clusters might be relevant to take into account.

As the starting point we will take an actual time-series for wind power production (P). This might not exist for the single cluster, however it will in many cases exist on the control block level (e.g. on a country level). This time-series will be adjusted to cluster by the installed wind power capacity within the cluster (C_c). Thus we get:

$P_c = P * C_c/C$

that is P_c is defined by the share of installed wind power capacity within the cluster (C_c) as part of total installed capacity (C). This implicitly assumes an even distribution of wind

[6] See deliverable D2.1 Assumptions on accuracy of photovoltaic power to be considered at short-term horizons

power on all clusters in the given control block. If needed it will be possible by aggregation to go from the cluster level to the Balance Responsible Party level[7].

4.1.2. Prediction errors for Optimate

In Optimate we will need a prediction error forecast for a time period of 0 – 36 hours ahead. The size of the prediction error will depend on following issues:

- The amount of wind power in the cluster
- The area size of the cluster – the larger the area the more dispersed will the wind power production be.
- The climatic conditions for wind power in the area.
- The time period of prediction. Typically we observe that a longer prediction period increases the probability for a higher prediction error.

The distribution of prediction errors will be calculated given the total time-series for wind power production, this is characterised by the root mean square error (RMSE), in this section given by the coefficient of variance (σ). How σ might develop over the prediction period is illustrated in Figure 5. The size and prediction-time dependence of σ is to be determined by available figures from existing literature.

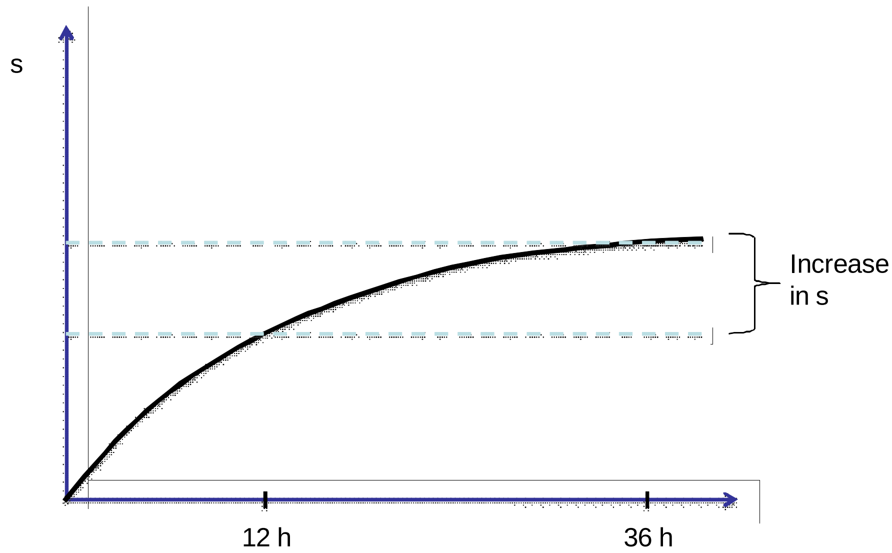


Figure 5: Dependence of σ on time period of prediction.

When we are looking at a series of prediction errors to be used as input to Optimate, we can find this series by a number of drawings in the general distribution of prediction errors:

PD = f(σ) = g(P, C, A), where C is wind power capacity and A is area size.

[7] This will depend on the definition of the Balance Responsible Parties in comparison to clusters.

As input to Optimate we have to find the distribution by cluster:

$PD_{c,t} = PD/(c*a)$

where c is a capacity parameter and a is an area parameter both to be defined as an index for the individual cluster (value between one and zero). In general we find that the smaller the size of the area (A) and the lower the capacity installed in the area (C) the more the distribution at the cluster level will deviate from the overall distribution and have larger fluctuations. Correspondingly, the larger the area and the capacity installed the closer the distribution of the cluster gets to the overall distribution. Thus we assume that:

$c \rightarrow 1$ for $C \rightarrow \infty$ and

$a \rightarrow 1$ for $A \rightarrow \infty$

where the values of c and a are to be determined according to the characteristics of the individual cluster.

Results for Germany as stated in Table 1 (p. 10) show clearly that the more dispersed the wind power installed capacity is, the lower the normalized root mean square error (NRMSE) of the wind power predictions. In this table a separate control zone (one out of four) is compared to overall Germany and at all the chosen time-horizons the prediction error is larger for the shown control zone than for total Germany. The ratio of NRMSE of Germany/Control zone range between 0.74 and 0.84 indicating that the adjustment factor for the simple approach (c*a) should be within the same range in this case. However, the results in Table 1 also indicate a time-dependence of the adjustment factor: The longer the prediction period the smaller the discrepancy between total Germany and the control block. Thus, for 2 hours ahead the ratio is 0.74, 4 hours ahead 0.77, whilst for the day-ahead period (24 hours) the ratio is 0.84 – the difference between total Germany and the control block gets smaller the longer the prediction period. This last-mentioned relationship is at present not included into the simple approach, but this could eventually be considered.

To include the timeliness of the prediction errors over the prediction period (correlation between hours), we will use:

$$PD_{c,t} = (1 - \alpha) * PD_{c,t-1} + \alpha * PD_{c,t} = (1 - \alpha) * PD_{c,t-1} + \alpha * f(\sigma_t)/(c*a)$$

Where the last-mentioned parameter ($f(\sigma_t)$) represents a drawing in the wind power variance-distribution for time-period t and α is the correlation-coefficient between prediction hours. α is to be determined by figures in existing literature.

4.1.3. Generating prediction error trajectories

How the simple approach generates a trajectory of prediction errors for the considered time-period is shown in Figure 6. In this figure the starting point is at the time t. In this hour a drawing from the distribution of prediction errors is carried out and the outcome shown as the round dot. In the next period (t+1) a new drawing is carried out (with a $\sigma_{t+1} > \sigma_t$) and again shown by the dot but because of the correlation to the previous period (α is assumed to be 0.5) the outcome is as shown by the x. And this process is continued with a still increasing σ. The higher the correlation coefficient (α) the more the drawing of the given time-period is modified by the previous period.

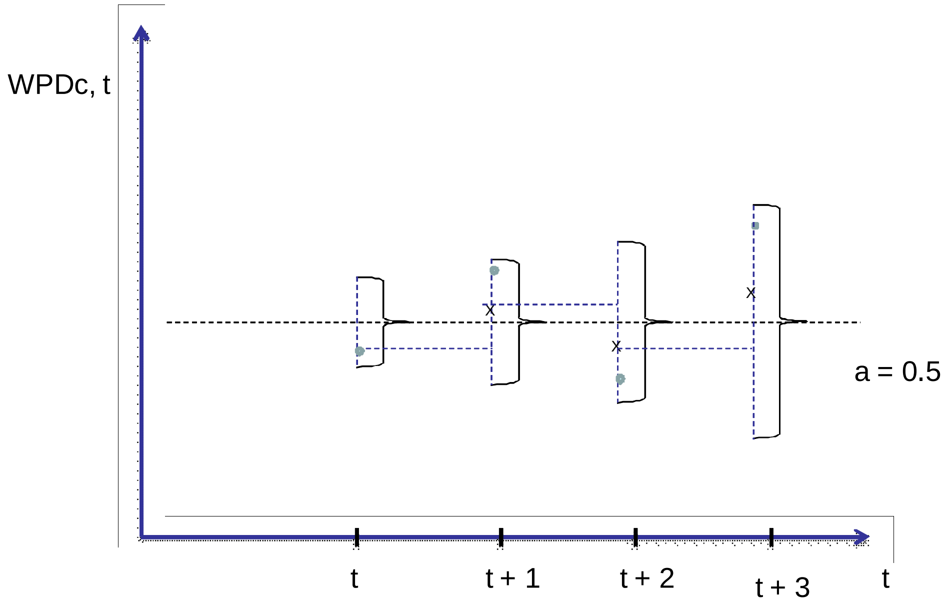


Figure 6: Trajectory of prediction error forecast as determined by the simple approach

This simple approach does not take into account the correlation between different geographical areas. Including such correlations will increase the complexity of equations and calculations, requiring a simultaneous computing of prediction trajectories compared to the partial calculations suggested by the above-mentioned simple approach.

However, a fairly simple remedy to this weakness could be to define a number of consistent wind areas within Europe (e.g. the North and Baltic sea, Inland Continental Europe, Spain etc.) and use the same drawing (or slightly modified drawing) within these areas.

4.2. An advanced methodology for handling forecast errors in Optimate.

This section of the report corresponds to page 81-91 in (Meibom et al 2008). The author of these sections in (Meibom et al 2008) is Rüdiger Barth from IER – University of Stuttgart.

The methodology is developed for the Wilmar-model and the following section describes the simulation of wind and load forecast errors. For the load forecast only the first ("wind and load forecast scenarios") and the third ("simulate isolated forecast errors") sections are relevant.

4.2.1. *Wind and load forecast scenarios*

The approach to simulate forecast errors is quite similar for wind and load forecast errors. This section explains the difference between the two methods. For this a brief outline of the wind forecast simulation module is presented.

The module simulates for each hour a set of wind prediction scenarios on hourly basis up to 36 hours days ahead. The development of this method is based on (Söder 2004). The simulated wind prediction scenarios include:

a) The wind forecast errors over the forecast length for a specific wind measurement station (standard deviations of forecast errors).
b) The correlations of the wind forecast errors between individual wind measurement stations for the individual forecast hours (spatial correlation of forecast errors).

The approach to handle point a) is described in the section “Simulate isolated forecast errors” (section 4.2.3). It applies also to the simulation of load forecast errors. In the subsequent section “Simulate correlated forecast errors” (section 4.2.4), the method to simulate spatial correlations of different regional wind forecasts is presented. This section does not apply to the simulation of load forecasting as only one load forecast for the whole power system is given.

The forecast error is always simulated by Auto Regressive Moving Average (ARMA) series that are established by tuning their statistical characteristics to those of real forecasts. Many sample paths of the ARMA series, that are drawn randomly, represent many different possible outcomes of forecasting. So, for example, i sample paths (or scenarios) of wind forecasts and j scenarios of load forecasts are derived. The scenarios of wind forecasts are aggregated with the load scenarios. It is not necessary to combine every wind scenario with every load scenario and to apply the scenario reduction module, see section 4.2.5, to $j \cdot i$ scenarios in this example. It is sufficient to allocate one load scenario for each wind scenario in a random way and to apply the scenario reduction module to a large number of scenarios (for example $i = 1000$). Statistically this leads to the same result.
The following section describes how the necessary statistical wind characteristics can be derived from the data provided to adapt the proposed method.

4.2.2. *Adapting wind data*

Within the approach it is assumed that data concerning the accuracy of wind speed forecasts in different regions and the correlations of the wind speed prediction errors are known. This data can be derived from the wind series and wind forecast series provided. Both typical empiric standard deviations for every forecast hour and typical empiric spatial correlations for every forecast hour have to be calculated by averaging the corresponding values for single wind stations and pairs of wind stations respectively.

The simulation of wind power forecast errors and future wind power series is based on wind speed series. If only wind power series are available as metered data, these power series have to be transformed to speed series by the use of an appropriate power curve. The speed series derived in this manner are added to the simulated wind speed forecast error scenarios (see next section). The resulting speed forecast scenarios are transformed to power forecast scenarios following (Norgard, Holtinnen 2004). The power output from a single wind power turbine is depending on the short-term variation of wind speed at the location of the wind power turbine. Due to the spatial distribution of the individual wind turbines within a region in combination with the stochastic behaviour of wind speeds, the power outputs at a given time from different wind power turbines vary. The simultaneous power outputs from the individual wind turbines are assumed to be distributed around an average value and the deviation of the spatial distribution depends on the extent of the considered region. Thus the aggregated power generation from more wind power units in a certain area will smooth out the short-term fluctuations of wind speed, as the power generation from the individual units are not fully correlated.

Typically the information of the instantaneous wind resource for an area is available in terms of only one time-series of the wind speed, valid only for the specific site, but representative for the entire area. A time-series of the aggregated power generation from a cluster of wind turbines in a region on the basis of the time-series of the wind speed in a single point or alternatively on the basis of the time-series of power generation from a single wind power turbine or a smaller wind farm is derived. Thereby a standard wind power curve representative for all wind turbine units in question (it is assumed that all wind power turbines within the regarded area are similar in size and control principle) and the smoothing effects both in time and space is considered.

The methodology is described in the following step by step:

1. Specification of a representative dimension of the regarded region describing the extent of the region in the North-South and West-East direction. This parameter is called “AreaSize”.

2. Specification of the wind speed distribution representative for the regarded regions by defining the two Weibull distribution parameters (scale factor A and form factor k).

3. The various wind speeds at the individual wind turbine units are assumed to be distributed around the block-average wind speed according to a normal distribution, compare Figure 7. Thereby the appropriate normalised standard deviation of the spatial wind speed distribution has to be identified in dependence of the spatial dimension “AreaSize”.

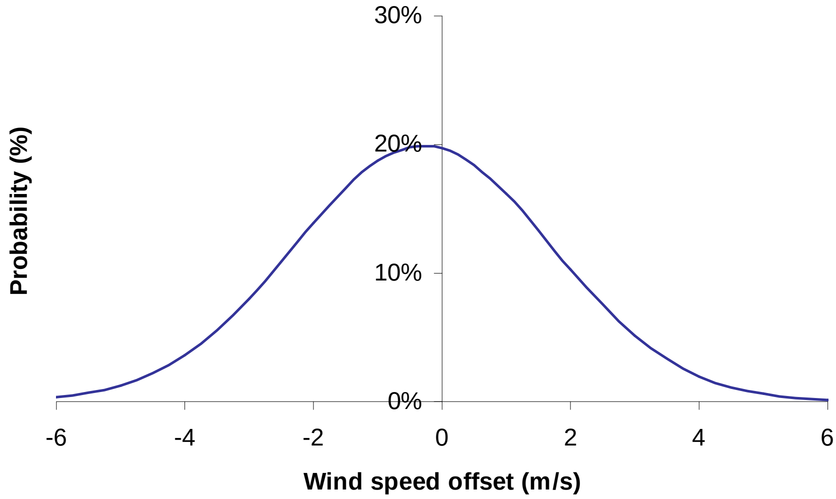


Figure 7. Example of the probability function for the block-averaged wind speeds for the individual wind turbines in an area at a given time. An offset adjustment of –0.15 m/s results in an unchanged accumulated production for the aggregated multi-turbine power curve and the given wind speed distribution.

4. Generation of the normalised aggregated multi-turbine curve by applying the normal distribution of the spatial wind speed distribution on the standard single-turbine power curve. The smoothed normalised multi-turbine power curve is representative for the aggregated power curve of the wind turbines within the regarded region. The aggregated power curve will result at lower wind speed levels in a higher average power generation per unit than for the single unit and at higher wind speed levels in a lower average power generation, compare Figure 8.

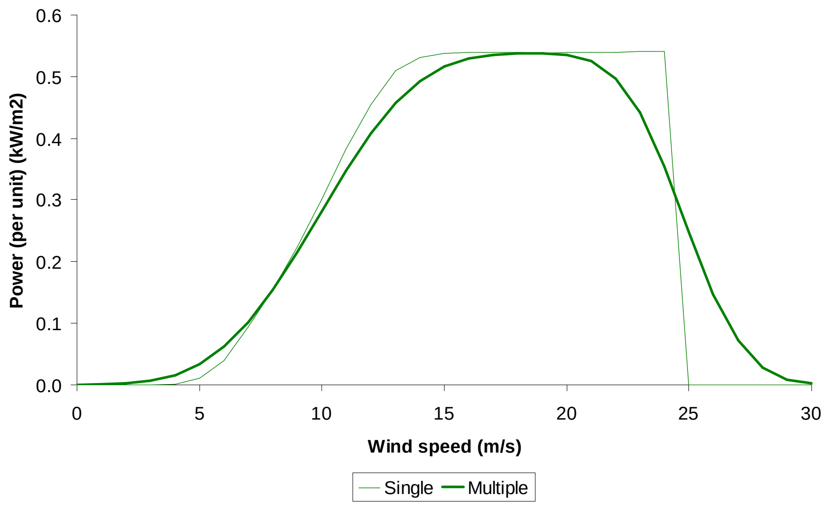


Figure 8. Example for normalised wind power curves corresponding to single and aggregated multi turbines.

5. The estimated normalised annual energy productions for a given wind speed distribution in time (Weibull distribution) should be equal for the single- and multi-turbine power curve. This is obtained by comparing the normalised annual energy production and adjusting the offset of the spatial wind speed distribution found in step 4 until the energy productions of both power curves are equal.

6. Generation of the aggregated power curve for the considered region by up scaling the normalised aggregated power curve appropriately to the corresponding installed wind power capacity.

7. Generation of wind power time-series for the considered region by applying the aggregated wind power curve to the block-averaged wind speed time-series.

With the described approach the wind speed series can be transformed to wind power series to get typical wind series that can also be applied in the future.

4.2.3. Simulate isolated forecast errors

The transformation of wind data as described above results in the generation of wind power time-series for the regions considered. In order to simulate forecast errors a simulation method has to generate realistic possible forecast error outcomes considering the historic statistical behavior of wind power. This is done using an ARMA approach, i.e. Auto Regressive Moving Average series, following (Söder 2004). For example by using an ARMA(1,1) approach, this series is defined as

$$\begin{aligned} X(0) &= 0 \\ Z(0) &= 0 \\ X(k) &= \alpha X(k-1) + Z(k) + \beta Z(k-1) \end{aligned} \tag{1}$$

Where

X(k) = forecast error in forecast hour k $\in$ N

Z(k) = random Gaussian variable with standard deviation σ_Z in forecast hour k $\in$ N

α, β = parameter of the ARMA-series.

Here the wind speed forecast errors are simulated with this approach, compare Figure 9. The assumed wind speed forecasts for each hour can then be calculated as the sum of the measured wind speed time-series and the wind speed forecast error scenarios.

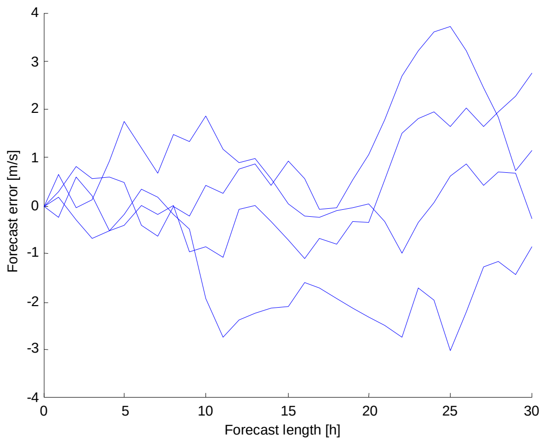


Figure 9. Four examples of ARMA(1,1)-outcomes of wind speed forecast errors with assumed ARMA-parameters α=0.95, β=0.02 and σ_Z=0.5.

The variance of the exemplarily ARMA(1,1) model, i.e. the variance of X(k), can be calculated in the following way:

$$\begin{aligned} &V(0)=0 \\ &V(1)=\sigma_Z^2 \\ &V(k)=\alpha^2 V(k-1)+(1+\beta^2+2\alpha\beta)\sigma_Z^2 \end{aligned} \qquad (2)$$

For $k \geq 2$, this equation can be rewritten as

$$V(k)=\sigma_Z^2\left(\alpha^{2(k-1)}+(1+\beta^2+2\alpha\beta)\sum_{i=1}^{k}\alpha^{(i-1)}\right) \qquad (3)$$

The standard deviation of the forecast error is then calculated as

$$\sigma\left(X(k)\right)=\sqrt{V(k)} \qquad (4)$$

To estimate the parameters of the ARMA series, the standard deviations of the ARMA series (that can be calculated theoretically) are compared to empiric standard deviations for every forecast hour that can be estimated analyzing the historic forecasts. By comparing the empiric and ARMA standard deviations and trying to have a minimal deviation between the two values one get a typical optimization problem that allows the estimation of the parameters of the ARMA time series.

4.2.4. Simulate correlated forecast errors

The preceding approach enables the simulation of wind forecast errors and generation of representative scenarios for single wind stations. If the system considered is a region with several wind stations or wind areas spatial correlations between these single stations have to be taken into account.

When wind speeds are forecasted for the same time period but for different locations, the forecast errors will be correlated because unpredicted wind conditions will affect both sites. The short time forecast errors of two measurement stations that are far from each

other are assumed to be less correlated, since the unpredictable wind situations are not the same for the two sites. For longer forecasts the unpredictable wind conditions are, however, similar for the two stations, so the forecast errors become more correlated.
In Figure 10 three examples of correlations between wind speed forecast errors are shown. As no real wind speed forecasts have been available for these measurement stations, it has been assumed that persistence forecasts have been used.

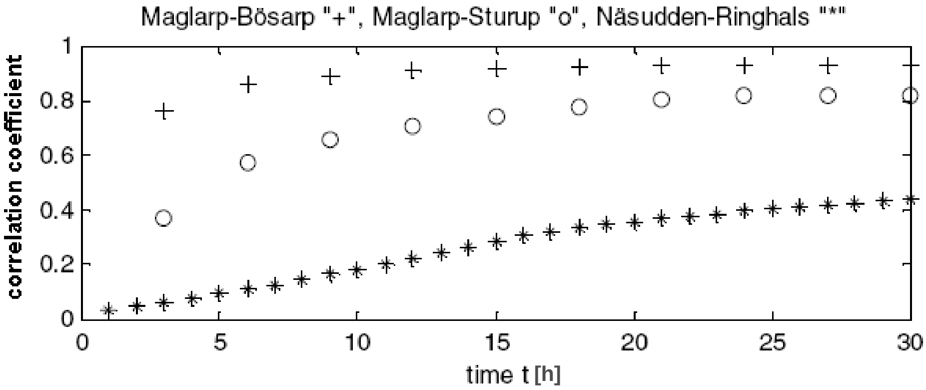


Figure 10. Correlation between forecast errors for different pairs of stations (Söder 2004). The distances between the stations are Maglarp-Bösarp (15 km), Maglarp-Sturup (26 km), Näsudden-Ringhals (370 km).

In Figure 10 it is obvious that the closer the stations are, the higher the correlation between forecast errors becomes. The correlation increases with the forecast length. In our case the empiric correlations should be derived by the delivered wind forecast series. The following approach is based on the approach in (Söder 2004) but has been changed primarily to consider different correlations for different forecast hours.

The used method simulates the correlations with a multidimensional ARMA-model. Since the correlation increases with time, the added uncertainty at different sites has to be more similar when the forecast horizon increases. Therefore the Z-variables in the ARMA-series should have an increased correlation if the correlations between the resulting X-variables increase.

The method adds a correlated Gaussian matrix C_{ZZ} to the individual ARMA-series X_k considering the assumption that the standard deviation of the common Gaussian variable Z(k) is constant. The derivation of the correlated Gaussian matrix C_{ZZ} works as follows.
The covariance between for example two wind speed measurement stations is calculated with:

$$C_{12}(k) = \rho_{12}(k) * \sqrt{V_{X1}(k) \cdot V_{X2}(k)} \qquad (5)$$

Where

$C_{12}(k)$ = covariance for the forecast hour $k \in N$

$\rho_{12}(k)$ = given correlation between the individual measurement stations for the forecast hour $k \in N$

$V_x(k)$ = variance for the forecast hour $k \in N$

The correlated Gaussian matrix C_{ZZ} can now be calculated with:

$$
\begin{aligned}
&C_{ZZ}(0) = 0 \\
&C_{ZZ}(1) = C_{12}(1) \\
&C_{ZZ}(k) = C_{12}(k) - \hat{\alpha}(C_{12}(k-1) - C_{ZZ}(k-1))\hat{\alpha} - (\hat{\alpha} + \hat{\beta})C_{ZZ}(k-1)(\hat{\alpha} + \hat{\beta})
\end{aligned}
\tag{6}
$$

Where

C_{ZZ} = correlated Gaussian matrix

$C_{12}(k)$ = covariance for the forecast hour $k \in N$

$\hat{\alpha}, \hat{\beta}$ = diagonal matrix containing the elements of α and β

The correlation between the individual ARMA-series X(k) is constant, equal to the correlation between the Gaussian variables Z and independent of the regarded hour when the forecast is made. The standard deviations of the Gaussian variables do not have to be the same, thus the variances of the individual ARMA-series X(k) do not have to be the same.

For the generation of the wind speed forecast error scenarios, the eigenvalues (D) and eigenvectors (V) of the correlated Gaussian matrix C_{ZZ} are determined. In the style of the Cholesky decomposition the matrix M is derived with:

$$M = V\sqrt{D} \tag{7}$$

Historical wind power forecast data shows that the expected wind power forecast errors corresponding to different forecast horizons are not zero. As our wind speed error scenarios are linear combinations of normally distributed random values, the expected wind power forecast errors calculated from these wind speed error scenarios would in fact be close to zero. Hence, when generating the scenarios for each individual forecast hour there is the option to draw firstly one scenario of the Gaussian variable by multiplying the matrix M with a normally distributed random value. This scenario is then treated as a simulation of the expected forecast error. Secondly a defined number of scenarios are generated by multiplying the M-matrix with the defined number of drawings of the normally distributed random values. This Monte-Carlo-simulation represents the uncertainty in the forecast. Finally the single drawing is added to these drawings.

4.2.5. Scenario Reduction

The generation of forecasts for wind power and load is based on a Monte-Carlo-simulation of a significant number of scenarios, see section 4.2.1. For very large numbers of scenarios it is impractical to obtain numerically a solution for the multi-stage optimisation problem. Moreover, the scenario tree consisting of these scenarios is only a one-stage tree. Thus, strategies for reducing the number of scenarios have to be studied to find a numerical solution of the problem as well as algorithms for constructing a multi-stage scenario tree out of a given set of scenarios. Simply generating a very small number of scenarios by

Monte Carlo simulations is not desirable since less scenarios give less information. Indeed, the aim is to lose only a minimum of information by the reduction process applied to the whole set of scenarios.

Actually, two steps are necessary: first, the pure number of scenarios is reduced. Afterwards, based on the remaining scenarios that still form a one-stage tree, a multi-stage scenario tree is constructed by deleting inner forecasts and creating branching within the scenario tree. The principle chosen for the setup of the scenario tree used in the Scheduling Model is shown in Figure 11. It consists of a three stage tree with 6 leaves.

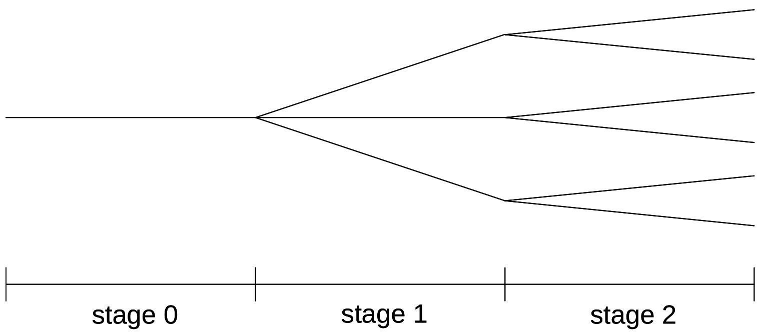


Figure 11. Principle setup of the scenario tree used in the Scheduling Model.

In the mathematical literature some algorithms are proposed for reducing a given set of scenarios and constructing a scenario tree based on the idea that the reduced scenario tree in a given sense is still a sufficient approximation of the original one (Dupacova et al. 2003). For this purpose, the Kantorovich distance $D_{KA}(P,Q)$ between a probability distribution P of a given number of scenarios and a distribution of scenarios Q with given probabilities for each scenario is considered (Rachev 1991). In the special case, that for Q a subset of all scenarios is chosen together with their probabilities, i.e. Q is a reduced probability distribution for P, an optimal probability distribution Q* based on these scenarios can be constructed possessing a minimal Kantorovich distance to P. A heuristic approach is used for finding the scenarios to be deleted from all scenarios (Dupacova et al. 2003). The resulting reduction algorithm is described in detail in (Barth et al. 2006a).

4.3. Concluding remarks

Undoubtedly, the advanced approach is to be preferred from a theoretical viewpoint. This approach tries to get as close as possible to a "real" forecast for wind power production, which is highly important, especially when considering a use within the Wilmar model whose main objective is to describe accurately the integration of wind power in Europe.

However, the main objective of the Optimate model is not to test the Power System integration of wind power, but to test new market designs assuming a strong growth in wind power production. For that reason a more simplified approach for describing wind power forecasts should be sufficient. Moreover the simplified approach should mitigate an extensive use of detailed data and avoid interference between the use of a Monte-Carlo process associated to a scenario tree in Wilmar Scheduling Model (see 4.2.5) and the need to keep track of deterministic results in Optimate to allow comparisons between different designs.

The simple approach described above constitutes a fairly easy way of generating prediction error trajectories to the Optimate model.

The needed input data[8] will include:

- time-series for wind power production for a control block (e.g. a country)
- time-series for prediction errors to generate the variance (alternatively this can be based on existing literature)
- assumptions on cluster dependent parameters such as area and capacity dependence and, finally
- assumptions on correlation coefficients (spatial and timeliness).

The results[9] to be used in Optimate simulator will be:

- for the DA-market a trajectory of correlated prediction errors will be calculated when added to time-series for wind power production will form a trajectory for wind power forecasts
- taking as starting point the DA-forecasts the same approach can be utilized for the ID-market ensuring consistency between forecasts in the two markets
- the standard deviation of the wind power forecast errors will be made available from the statistical analyses.

[8] Entry of Optimate future preprocessor

[9] Output of Optimate preprocessor and input data of Optimate simulator

5. Conclusions - summary

The future development of the energy system points to an intensive introduction of renewable energy sources, to a certain extent relying on variable sources as wind power and photovoltaics. These renewable sources have characteristics that do not easily fit into the current electricity market frameworks, especially because of their intermittent generation. The day-ahead forecasts for the production of power from these plants are subject to some uncertainty and it is risky for them to bid into the spot market because they have to pay penalties when they do not fulfil their commitments. New market designs are needed which facilitate the introduction of variable renewable power sources.

Thus, the main objective of Optimate is to develop a new tool for testing these new market designs with large introduction of variable renewable energy sources. In Optimate a novel network/system/market modelling approach is being developed, generating an open simulation platform able to exhibit the comparative benefits of several market design options.

This report constitutes delivery 3.1 on the assumptions on accuracy of wind power to be considered at short and long term horizons. The report handles the issues of state-of-the-art prediction, how predictions for wind power enter into the Optimate model and a simple and a more advanced methodology of how to generate trajectories of prediction errors to be used in Optimate.

The main conclusion is that undoubtedly, the advanced approach is to be preferred to the simple one seen from a theoretical viewpoint. However, the advanced approach was developed to the Wilmar-model with the purpose of describing the integration of large-scale wind power in Europe. As the main purpose of the Optimate model is not to test the integration of wind power, but to test new market designs assuming a strong growth in wind power production, a more simplified approach for describing wind power forecasts should be sufficient. Thus a further development of the simple approach is suggested, eventually including correlations between geographical areas.

In this report the general methodologies for generating trajectories for wind power forecasts are outlined. However, the methods are not yet implemented. In the next phase of Optimate, the clusters will be defined and the needed data collected. Following this phase actual results will be generated to be used in Optimate.

6. References

Anderson, C.L.; Davidson, M (2005).: An aggregate Weibull approach for modelling short-term system generating capacity. IEEE Transactions on power systems, Vol. 20 (2005), No. 4, pages 1783 – 1789

Barth, R., Söder, L., Weber, C., Brand, H., Swider, D., (2006a), Deliverable D6.2 (b) – Documentation Methodology of the Scenario Tree Tool, Institute of Energy Economics and the Rational Use of Energy (IER), University of Stuttgart, available from www.wilmar.risoe.dk.

Barth, R., Brand, H., Swider, D.J., Weber, C., Meibom, P., (2006b), Regional electricity price differences due to intermittent wind power in Germany – Impact of extended transmission and storage capacities. *International Journal of Global Energy Issues*, vol. 25 (2006), issue 3/4, pp. 276 – 297.

Doherty, R.; O'Malley, M. (2005): A new approach to quantify reserve demand in systems with significant installed wind capacity. IEEE Transactions on power systems, Vol. 20, No 2, 2005

Dupacova, J.; Gröwe-Kuska, N.; Römisch, W. (2003): Scenario reduction in stochastic programming: An approach using probability metrics. Mathematical programming. Volume 95, number 3, 2003. Pages 493 – 511

Endrenyi, J. (1978): Reliability modelling in electric power systems. John Wiley & Sons. Chichester, New York, Brisbane, Toronto, 1978

Giebel, G., R. Brownsword, G. Kariniotakis, C. Draxl. (2010). "The State-Of-The-Art in Short-Term Prediction of Wind Power. A literature overview, 2nd edition." Unpublished draft of the ANEMOS.plus project. Earlier version available under: http://anemos.cma.fr/modules.php?name=Downloads&d_op=getit&lid=8

Krauß, C. (2006). *Integration of wind power in the German electricity market - A trader's view*. EnBW Trading (DE). Talk held on the Workshop on Best Practice in the Use of Short-Term Prediction of Wind Power, Delft (NL), 25. 10. 2006.

Lange, M., U. Focken. (2005). *Physical Approach to Short-Term Wind Power Prediction*. Berlin: Springer.

Madsen, H., G. Kariniotakis, H.Aa. Nielsen, T.S. Nielsen, P. Pinson. (2004). *A Protocol for Standardizing the Performance Evaluation of Short-term Wind Power Prediction Models*. Project report for the Anemos project, on anemos.cma.fr.

Meibom,P; R. Barth, H. Brand, D. Swider, H. Ravn, and C. Weber (2008), "All island renewable grid study workstream 2b - wind variability management studies," *http://www.dcmnr.gov.ie*, 2008.

Monteiro, C., Bessa, R., Miranda, V., Botterud, A., Wang, J., Conzelmann, G. (2009). Wind power forecasting: State-of-the-Art 2009. Argonne National Laboratory, ANL/DIS-10-1.

Norgaard, P.; Holtinnen, H. (2004): A multi-turbine power curve approach. Contribution to the Nordic Wind Power Conference 2004. Göteborg, 2004

Perman, M.; Senegacnik, A.; Tuma, M. (1997): Semi-Markov models with an application to power-plant reliability analysis. IEEE Transactions on reliability, Vol. 46 (1997), No. 4, pages 526 – 532

Pievatolo, A.; Tironi, E.; Valade, I. (2004): Semi-Markov processes for power system reliability assessment with application to uninterruptible power supply. IEEE Transactions on power systems, Vol. 19 (2004), No. 3, pages 1326 – 1333

Rachev, S. T. (1991): Probability metrics and the stability of stochastic models, J. Wiley

Rohrig, K. (ed). (2005). Entwicklung eines Rechenmodells zur Windleistungsprognose für das Gebiet des deutschen Verbundnetzes, Abschlussbericht Forschungsvorhaben Nr. 0329915A, gefördert durch Bundesministeriums für Umwelt, Naturschutz und Reaktorsicherheit (BMU). Kassel, Germany.

Söder, L. (2004): Simulation of Wind Speed Forecast Errors for Operation Planning of Multi-Area Power Systems. Contribution to the 8th International Conference on Probabilistic Methods Applied to Power Systems. Ames (Iowa), 2004

Tastu, J., H. Madsen, P. Pinson. (2010). Multivariate Conditional Parametric models for a spatio-temporal analysis of short-term wind power forecast errors. European Wind Energy Conference2010, Warsaw.

Van Casteren, J.F.L.; Bollen, M.H.J.; Schmieg, E. (2000): Reliability assessment in electrical power systems: The Weibull-Markov stochastic model. IEEE Transactions on power systems. Vol. 36, No. 3, 2000

7. Appendix – Glossary

ARMA
Auto regressive moving average

BRP
Balance Responsible Party or Balance Responsible Perimeter

DA markets
Day-ahead markets. Market participants bid into the DA-market for a certain hours in advance.

ID markets
Intraday markets. The ID-market gives the market participants a possibility to adjust the DA-bid before entering the balancing market.

Forecast scenarios
A trajectory of interconnected predictions covering a certain period of time

MAE
Mean Absolute Error

MAPE
Mean Absolute Percentage Error

NRMSE
Normalized root mean square error

NWP
Numerical weather prediction

Prediction (forecast) error
The difference between forecast and actual observation

RMSE
Root mean square error

RT
Real time

TSO
Transmission system operator