PSO-F&U
Interim Public Project Report

Project title	DEWEPS - *Development and Evaluation of a new wind profile theory with an Ensemble Prediction System*
Project identification	2009-1-10260
Project periode (date, year)	2009 – 2011
Responsible Institution/Company	WEPROG ApS
Reporting-periode	01/2010 - 07/2010

Publication Date: 9. August 2010

Table of Contents

1. Project Status and Introduction

The core of the DEWEPS project lies on meteorological improvements of wind power forecasting and in the application of ensemble forecasts in the energy system. In order to validate the usefulness of developments correctly, extreme care has to be taken because the energy system is almost continuously changing as a result of political pressure and possibly also pressure from stakeholders. What is the target for forecasting today may therefore not be the target in the future. This interim report will show the difficulty in evaluating what are good and bad forecasts and will conclude that the same forecast can be best or worst depending on what the target is defined.

First we will discuss an additional project deliverable, which is a hemispheric ensemble database, which can be of major value in the large-scale wind integration process. This is of interest in the future both for the project and commercially. Originally, the hemispheric forecast data were expected to be only of internal interest, but we learned at the Forsk-EL information day that such data may actually be useful and be used by third parties, if available. Therefore, we will discuss the relevance of the output data from model simulations for usage in other research projects and also in more commercial applications.

Secondly, an overview of different numerical experiments is given with emphasis on how they are verified. The detailed analysis of over 100 experiments goes beyond the scope of this report though.
Thereafter, we discuss the European SuperGrid concept and what impact and requirements it will have on the forecasting approach. We will discuss how we already today have a SuperGrid with some limitations, what benefits this provides, but also which challenges the SuperGrid does not seem to tackle. We describe in addition two project developments, which will be useful in a SuperGrid scenario with large-scale wind integration.

Finally, an extreme case from the 11th of June 2010 is evaluated. This case demonstrates that future forecasting techniques may compromise system security, because only todays and the "old" coarse resolution models performed well in this case. The event could easily have led to a 12GW imbalance in the Danish-German area, which would have been a significant challenge for the energy system in Denmark, Germany and the Netherlands. We will demonstrate that such an imbalance is a result of a too focused view on optimisation of forecasts towards good average scores rather than extreme events. In the discussion of this event we will highlight, why it is non trivial to determine what is a good or a bad forecast.

1.1 Additional Project Deliverables

The topic of large scale wind integration is getting more and more attention as it is becoming clear that wind power will remain the most cost efficient renewable energy source and that wind power has the widest application range world wide. Most regions have locations with sufficiently good wind resources. However, such locations are often distant to the load centres and this is the most common challenge in the integration of wind. A relatively low dispersion level of the wind generation reduces the correlation between demand and generation and increases additionally the forecast error and the need of reserve.

Evaluation of the expected wind power forecasting accuracy in future scenarios can therefore facilitate wind integration and this is most likely even becoming of commercial interest as the market value of the wind generated electricity and the feed-in limits may become of more interest as incentives level off.

Therefore, we decided to generate a database out of the simulation experiments that are being carried out in this project. This database can be regarded as an additional deliverable, because the output data from the project milestones on improved forecasting can be used for more than just forecast evaluation. Also planning of tenders like it is commonly used for offshore projects may benefit from the database, because the correlation to demand and other generation can be computed from a long time series of model data.
Examples of energy system technologies under development, which may link to forecasting could be storage and flexible demand for improved balancing.
Traditionally, weather forecast data is not used very much nor detailed in many studies. The approximations and assumptions usually made in a feasibility study may in fact add to the uncertainty level of a given development's feasibility. It was mentioned specifically at the ForskEL information day on the 21st of June 2010 by Copenhagen airport, that the economic feasibility was difficult to access of research project developments. The message that came across was that the industry needs to have sufficiently detailed information to evaluate all costs and benefits of a given development relative to a known technology.

New developments within wind power forecasting have traditionally been measured on the reduction of the forecast error. This is an objective measure of the success rate, but there are often additional benefits. An example from the current project is the creation of a database of hemispheric ensemble weather forecasts in hourly resolution.

Such a database is unique in the world and can in fact be used in studies, development and demonstration in cleantech areas, although it will require special effort to encourage the use of the database in the development of new applications in the dissemination phase. The hemispheric coverage means that 99% of the global wind energy market and a large fraction of other renewable energies is covered.

It is obvious that energy is a discipline, where the economic interest at stake has a major impact on any business decision. There is a trend that some companies want to have full control and ownership of what they spend resources on and not take any action that may reveal intentions to third parties. Therefore, manageable solutions seem to be preferred. In such cases, it is observed that often simplistic solutions below state of the art are applied by staff that may not even be able to access the impact and the approximations implied in their solution approach. It is impossible to keep state of the art in disciplines like forecasting without being globally represented and spending significant resources on continued development.
However there are also companies that are in the believe that the liberalised market is at the end of the day bringing most progress. Such companies have realised that they cannot keep the required staff in the long run to maintain state of the art. Instead they outsource complicated tasks, because they foresee that investments on internal development will not pay themselves back..
Given the size of stakeholders in the energy area, one can probably expect another decade, where significant resources are spent on internal developments that are not using state of the art.

As an example of a rather disappointing speed of development is the ensemble forecasting. We are approaching the 20 year anniversary of Ensemble forecasting and see a very low numbers of applications that actually use such forecasts. Is this the scientific world that cannot sell the idea or is this simply the the commercial world that refuses to adopt new or complex technology, although conceptionally well understood and acknowledged ?

The conclusion from the ForskEl information day is that new achievements have to be presented better and there should not be any risk in using them. Scientists work is to a large extend driven by devotedness to the topic. This cannot be expected to spread to everybody or any development. This in return means that it is the scientific developers task to bring new achievements much further towards the commercialisation of the applications and this is probably has not been done in the case of ensemble forecasting, starting already in 1993.
This means in fact that it is still a challenge to serve a complicated topic in a manor so it is simple to take advantage for the industry. This is an important lesson from the information day and has already received impact on the project dissemination strategy, i.e. that deliverables have to be more explicitly mentioned, described, compared and promoted. A developer cannot expect the 3rd party to get the idea from the vast information available today.

2. Scientific work

In light of the discussion in the previous section it seems to be necessary to enhance the dissemination of results.. Therefore, we aim in this project to produce a number of relevant technical documents in order to describe the work conducted during the project and to make results and relevant teachings available to the public in form of technical documents, conference proceedings and publications.

The work can be summarised in 8 parts:

- Improved mountain blocking in NWP models of moderate resolution
- Experience with a hemispheric high-resolution Ensemble
- Statistical training in NWP model space
- Benefits of exact area integrals in comparison to upscaling approaches
- Cost optimisation using pre-allocated ancillary reserves
- Benefits of coupled demand and wind power forecasting
- Optimisation in Ensemble space using a weather classification scheme
- Coupled demand and wind power forecasting in a SuperGrid context

These topics will be summarised in the final report, where we will focus on presenting the synergy of all achievements especially in a SuperGrid context. Each of the above work packages will demonstrate that significant improvements have been achieved on the quality of forecasting. We hope that by dissemination of these results, the industry will also be encouraged to advance their application systems for planning and operations in order to make use of such improvements, especially to incorporate ensemble data.

2.1 Forecasting Experiments

While we had been focusing mostly on sensitivity studies of single events in the previous period, we have now initiated a number of long-term simulations to investigate the impact of model changes on the forecast quality over longer-term periods. The goal of the longer sensitivity experiments, that contain around 100 experimental model setup, is to find a model configuration that can be regarded as optimal for wind power forecasting in a general sense.

Although it is known that the balancing costs at times are highly dominated by hours where extreme costs and high forecast error correlate, one can assume that the average error is reflected in the balancing costs after sufficiently long time.

In previous studies (e.g. The HREnsemble project) it was found that certain changes in the NWP models such as modifications related to the sea surface had rather little impact on the accuracy of forecasting. Nevertheless, it is necessary to investigate, whether many small improvements add to a bulk improvement or whether some improvements also reduce the performance of the model systems.
The many test setups contain new model formulations, where the use of climate data, analysis incrementation, area size, lateral boundary conditions, friction, condensation, handling of sea surface parameters etc. have been modified. The testing methodology for the different model formulations was setup in different stages and different length of simulation periods as follows.

- Stage 1: Simulation of November and December 2009 with 75 members 4 times per day 48 hours ahead.
- Stage 2: Verify on wind power in Ireland, France, Denmark and Germany
- Stage 3: Analyse and compare results to previous experiments in that period
- Stage 4: Conduct a 1.0-1.5 year simulation starting from January 2009, if the analysis in stage 3 suggests that there is a positive impact

November and December in 2009 were chosen in order to get strong response to the model changes as the error normally grows most in periods where the average wind speed is equivalent to the steepest point on the power curves.
The conclusion after a number of long-term experiments (stage 4) experiments was that the computationally more expensive setups outperformed the less expensive setups. In order to be able to run more long periods with expensive model configurations, it was decided to make a 8 member mini ensemble test environment. Because of the reduced number of members, it was then possible to double the spatial model resolution, if we also reduced the number of daily 48 hour forecasts to one instead of four. Instead, the 3 forecasts at 06, 12 and 18 UTC were then limited to 6 hours in order to keep the system giving the equivalent results of running real time system. The statistical evaluation becomes then relatively more trustworthy as each day-ahead forecast covers different weather.
Additionally, a 1 hour resolution hemispheric lateral boundary member setup was pre-computed for 1.5 years. The lateral boundary values were then common for all experiments.

The reduced test bed also allowed for faster testing of various potential improvements on the parameters indicated on Table 1.

NWP model configuration changes/upgrades
75 Members, Europe area, 8 Boundary members, stand. mean Orography, 3D analysis increment, full vert. interpol. const. Charnock, SST analysis, stand. climate file with 0.45deg. resolution
Area size and Boundary conditions
Various model domain sizes and spatial resolutions
Various boundary model sizes and spatial resolutions
Various experiments with different boundary age scheme
Analysis incrementation and initial conditions
Multiple 4D analysis increment formulations
analysis mixing of 2 analyses (CMC and NCEP)
partial vertical interpolation (only lower boundary)
Sea Surface representation
dynamic Charnock calculations
Iterative SST blending
Land surface representation/orography
modified orographic roughness representation
Steeper and taller orography
higher resolution climate file
Ensemble size
25 ensemble member on Europe area, 8 boundary member large Europe Area
8 ensemble member on ext. large Europe area, 8 boundary mem. ext. large Europe Area
8 ensemble member on ext. large Europe area, 8 boundary member northern hemisphere

Table 1: Description of the model changes for the NWP sensitivity experiments.

2.1.1 Model Improvements related to surface Roughness and Orographic Representation

One group of experiments focused on the smooth orographic representation of mountains in the models. On the one hand high numerical stability is required for reliability of the model system and on the other hand, the model should also provide a forecast based on a realistic representation of the earth's serface. The numerical stability can be compromised, if mountain summits reach far up in the atmosphere where the prevailing wind speeds are strong while the jet stream crosses over the mountains. As air parcels approach the mountains they de-accelerate and cause turbulence, which is carried further downstream with the mean flow. Very steep mountains like the Alpes reach up to the middle of the troposphere in the real world, but for numerical stability reasons they may only reach half of that altitude in the models.

The challenge has therefore been to maintain model stability for significantly tall mountains. Tests have been conducted with peak mountain heights at 4200m. This has been estimated to be a good compromise from studies of the Alpes with Google Earth. The exact formulation on which height is best suited for the models has been worked on. The tests so far have been successful and improved the forecast accuracy in Europe (measured in wind power), except for a few cases, which will be described later in this document.

Along with the choice of mountain heights, we also need to adjust roughness in the climate data, because the orographic contribution to friction from valleys reduces as the mean height increases. Similarly, we need to parameterise a new soil condition valid for the taller mountains. The snow cover increases and the surface becomes dryer and colder. However, this evaluation involves consideration of the effective land surface class, which also changes with altitude as the modified orography may change from forest to mainly rock. The overall problem is very complex, because of the interaction of physical processes at the earth's surface. Nevertheless, the numerical stability as the first challenge has been tackled and if we consider RMSE as the target for good forecasting, there is progress.

A major second challenge seems to be that a typical SYNOP measurement is located in valleys rather than at summits. However, many values may disappear during the orographic modification. This implies interpolation of the correct pressure to the actual model surface, which is dependent on the average temperature in the valley and is not forecasted explicitly, if the valley is located far under the model ground.

A major effort was spent on locating the weather events in which the new mountain formulation is superior to the old. The analysis suggests clearly that the larger the scale of the motion and the stronger the jet stream, the better the forecast skill of the tall orography. On the other hand, those events where the old system was better had all low pressure systems under development in mountainous regions. So, on the one hand the blocking of the tall orography adds value, but it can also cause severe problems for "young" low pressure systems located in mountains.
At present, we consider it dissatisfactory that the new formulation has a disadvantage of that type as we are not yet sure, whether this may be due to the vertical interpolation of SYNOP data. With mountains in all corners of Europe this work is crucial for wind power forecasting and is therefore given core attention in the project.

2.1.2 Improvements related to lateral boundary handling

Another typical error source in wind power forecasting that is in the process of investigation is the large and potentially dangerous forecast errors that arise from too small model areas of the forecasting or even the boundary generating model.

This work includes also the interpolation between the large scale boundary generating model and the choice of resolution and area size of this model area. The choice of model areas are to a large extend a matter of computational cost. As an extreme demonstration, all 75 members were run on an area covering the entire northern hemisphere for 1.5 years in 0.45 deg. resolution. The special capability of a hemispheric system is that there is very little flow across the equator for angular momentum conservation reasons. This improves the pressure field and thereby the entire forecast.

The overall result was that there was a clear benefit during the most windy months, but the spatial resolution were at times not competitive with high resolution setups, because these have a better land sea mask definition. By running the hemispheric system fully nested with 2x75 members, it was then possible to gain accuracy also in rthese periods. However, the overall pattern was that in the calm months of the year, there was little gain of the double nested system compared to the cost of running the system.

The commercial viability of the hemispheric model domain has been tested via this work and this implies ease of handling of large amounts of wind power on all continents. The combination of tall orography and hemispheric model area has been tested in single events. On this basis it is had to be concluded that there is further work required to get the climatic data and surface representation synchronised and updated. However, both developed markets in Europe and markets under strong development such as the Asian market have become very competitive on price. Therefore, the time is not yet ready for the application of this model system, especially, since Europe seems to gain more from increased spatial resolution and European end-users generally focus more on quality and only secondarily on price.

2.1.3 Objective Scores of selected Model Improvements

Table 2 provides some samples of statistical results . What characterises the results is that all experiments have a setup with an extended model area and a new formulation for the incrementation of the analysis (referred to as "4Dinc").

NUM	EXP	mean/ weight. mean	AVR of TOP50	Relative Improvment Mean/REF	Relative Improvment AVR/REF	DESCRIPTION
1	27	6.11	6.15	0.00	0.00	**Reference (75member in 45km, EU area)**
40	38	5.29	5.49	13.42	10.73	22Km ext. EU area, 4Dinc with 75 mem assim.
41	38	5.26	5.43	13.91	11.71	22km ext. EU area, 4Dinc with 24 mem assim.
37	27	5.13	5.44	16.04	11.54	45km EU, vert. Inc of analysis, bd age -6
61	27	5.37	5.50	12.11	10.57	45km, 8 bnd mem. on ext. Eu area, 75x 4Dinc

Table 2: RMSE statistics in % of inst. capacity of the experiments that were carried out in December 2009

The experiments in Table 2 have been used in stage 4 sequentially and 2 experiments have completed. Unfortunately the superior results of experiment no. 37 did not hold over longer time. The approach didn't perform as well during some of the summer months. The experiment was thereafter stopped after one year. A more detailed evaluation of the vast number of experiments will be demonstrated in a separate document.
The preliminary result suggested that none of those experiments in table Table 2 is superior in performance to the changes on tall orography described in the previous section (referred to T10 on Figure 4 later in this document). At this stage it seems to make most sense to combine this setup with the setup that seemed to be most successful on the assimilation scheme into one setup, because this implies combination of the two aspects that we found bring most overall improvement. This experiment is still in progress.

It is very promising that dedicated efforts on critical points in the model chain can lead to significant improvements. It should be noted that the model formulation changes on assimilation and mountain parameterisation do not require model retraining of power curves, because they are not systematically changing the mean wind speeds, where the wind turbines are standing.

The evaluation of the forecast quality on a defined set of parameters is therefore more straightforward. The power curve estimation will also benefit in the long run from improved forecasts, but for the time being, this is not critical for the evaluation. Therefore all experiments in one particular spatial resolution are using a fixed set of power curves.

2.2 Forecasting and Planing tools for at European SuperGrid

In the period towards this midterm report we found that it is more and more important to discuss the impact of the so-called European SuperGrid concept that has been mentioned a number of times recently as the solution to the European wind integration challenge. An internet platform has been established by Mainstream Renewable Power that presents themselves as "..a group of companies and organisations which have a mutual interest in promoting and influencing the policy and regulatory framework required to enable large-scale interconnection in Europe" (see http://www.friendsofthesupergrid.com). Regardless whether the European SuperGrid is a solution or not, it can be expected from such initiatives that the SuperGrid concept is most likely going to be implemented in some way or another, because of the financial crisis and the positive prospects that such large-scale projects involve.
The SuperGrid concept in itself is an idea, which can be "sold" politically and to a large audience, because it implies more centralised management, market transparency and competition. The main challenge most likely lies in the requirement for a Europe wide collaboration, where the decision process and an implementation takes longer time than on a national level. The SuperGrid may be subject to stronger influence from the parties who most actively support the idea, so that attempts to overtake the political decision process may have to be expected.
Nevertheless, we will discuss below how the European SuperGrid concept may effectively evolve with some political pressure. At present, the evolution seems to be driven by a general wish from the industry to accelerated the development. Although Germany is not fully unbundled, it seems that the central location and wind power pool size suggests that Germany may become a central point for the future energy system.

Since the project proposal was written, the discussion of large scale integration of wind power has accelerated. Eddy O'Connor from AIRTRICITY (now Mainstream Renewable Power) brought the idea of a European SuperGrid forward towards an offshore context in 2008. The idea was to combine the required transmission for large offshore wind farms and with long distance transmission between areas with low and high electricity prices. In fact, Eddy O'Connor formulated an idea that was already discussed in the system development department ("Planlaegningsafdeling") in Eltra after the initial experience of the first 160MW Horns Reef wind farm in the North Sea in 2005. Additionally, it should be mentioned that the first stage of the European SuperGrid idea in offshore context is already under planing at Krigers Flak in the Baltic Sea. The initial plan was to grid connect the offshore wind farms to all 3 countries Denmark, Germany and Sweden, although Sweden does not seem to take part in the final stage.

The basic purpose of the European SuperGrid is hence to make the total wind power generation more correlated with the demand and thereby increase the economic value of wind power. By nature, excessive wind power may lead to congestion during hours of strong wind and consequently low energy prices. The basic idea behind the SuperGrid concept is then to increase the transmission capacity and in that way enlarge the grid and at the same time establish one common European market.
One can argue that Germany and Denmark have managed to reach a high penetration of renewable energy even though wind power could not compete on the generation price for many years. Instead, the development has been driven by production incentives paid by consumers. Various studies have shown that wind power acting as price taker in the market reduces the average market price and that wind power is therefore a benefit for the consumer at the end of the day. Thus, wind power production incentives are equivalent to an insurance against high energy prices.

The question that comes to mind now is, whether it will be better to introduce a market implementation of wind power with reduced volatility from enhanced inter-connectivity or whether it will be better to use wind power as a mean to lower the market price ?

In fact, one strategy does not exclude the other and grid security will over time be enhanced by having increased flexibility in the operation, provided that the software solutions develop with technology advances. Denmark has over two decades experienced that the primary power generation became more and more dispersed partially from wind and partially from CHP plant. This process will extend further to ancillary services and flexible demand.
By October 2010, the German market has had allowance for negative prices for two years, while more than 99% of the wind power stayed in the fixed tariff scheme, although it was allowed to change payment scheme between pure market terms and fixed price (EEG) on a monthly basis. The TSO's try to recover the costs of the EEG tariff on the market. However, approximately 30% of the costs need to be recovered through tariffs. Thus, the market value of wind power is well under the actual cost, even though the penetration level of wind power is only 10% averaged over the country.

Considering that Germany is one of the most interconnected countries in Europe, it is a worry that wind power cannot compete on market terms. It is often mentioned that part of the reason is the low wind speed in Germany, but the average load factor remains stable very close to 20% of the installed capacity. It is likely that more advanced operation and maintenance concepts could bring the production over 20% of the rated capacity. And, once the offshore penetration reaches the level of the onshore capacity, the average load factor will most likely reach 30% or more.

2.2.1 Lessons learned from market study of wind power

In the following we will summarise a couple of "lessons learned" from studying the price variations of the German and Danish wind power production, as they explain why wind power is not competitive on market terms in Germany without a production bonus scheme:

- The high level of energy generation from nuclear and brown coal in Germany generates stiffness in the system, because nuclear power and brown coal plant can only be used for base load
- Fixed price incentives are required, if wind, brown coal and nuclear generation should be the main generation sources
- The need of wind to act as ancillary service increases with the amount of wind and constant nuclear power and brown coal generation
- Low average wind speeds cause additional surplus of power generation from fossil and nuclear power, whenever there is strong wind during periods of low demand
- Offshore wind power helps to reduce the number of hours with negative prices compared to a scenario of equivalent amount of generation at lower average wind speeds on land
- It is rather the total amount of wind generation than the penetration level that causes price volatility
- Germany has operated since many years with different day and night prices for the consumers in some areas to reduce the diurnal cycle of the demand.
- The practise of the so-called HOBA principle ("Horinzontaler Belastungsausgleich") in Germany can essentially be considered a predecessor of the SuperGrid principle for renewable energy, because the 4 TSO's in Germany have the responsibility to trade a predefined share of the total wind energy production in the country, rather than the production of the physically installed wind turbines in their control area.
- Germany is a large country, where the variations of the generation can be balanced well with conventional power generation.
- In Germany, there are monthly and daily bidding schemes, where significant amounts of secondary reserve can be allocated, which are under normal circumstances sufficient to balance wind power forecast errors, if the intra-day prices are unfavourable.
- Several markets centred around Germany have been coupled in the years 2008-2010 and the aggregated wind power hardly exceeds 50% of the demand in the currently coupled markets.
- The German regulator (Bundesnetzagentur) has in 2010 requested to couple the reserve markets of the 4 TSO areas in order to use synergy effects, avoid counter balancing and increase competition. Especially, the low amount of market participants for reserve in Germany suggests that there is not yet a strong enough competition.
- Despite a low wind power penetration level and strong interconnection we observe that Germany cannot recover the cost of the wind generation at the current market prices. It might be that the incentives are too favourable. However, the wind generation growth rate is not excessive and incentives will be required to meet the future targets.
- The current percentage of wind energy on the net energy consumption in Germany is at 9%. With the recent growth rate in wind power this would bring Germany as a whole to a 20% penetration approximately 20 years after Denmark, which reached this level in 2003. However, there are already 5 provinces (Bundeslaender) with a wind penetration level of more than 20% of net consumption. (Top 5 counties statistic in 2009: Sachsen-Anhalt (3.4GW) 47,08%, Mecklenburg-Vorpommern (1.5GW) 41,29%, Schleswig-Holstein (2.8GW) 39,82%, Brandenburg (4.3GW) 38,12%, Niedersachsen (6.5GW) 22,78%).

- Higher incentives and discontinuation of nuclear power generation is required to accelerate Germany's transition to renewable energy. The high offshore production incentive will accelerate the evolution and may enforce the nuclear power to stop as planned, because of significant over-capacities being reached already in 10 years, according to a recently published study initiated by HEAG Sächsische Energie AG ("Entwicklungsperspektiven des deutschen Elektrizitätsmarketes", emerit. Prof. Dr. D. Schmitt, Dept. Energy Management, University Duisburg-Essen and H.H. Forsbach, p.9ff).
- Our analysis suggests that it is fair to argue that we can already today claim to operate an onshore SuperGrid, where the bulk of Europe's wind power is installed. The incentive for a further development of this principle is that more markets are coupled and more transmission capacity is added, then more of the electrical system will develop to a SuperGrid with increased competition.

If the SuperGrid should develop further, there will be need of significant investments in reinforcement of the grid, because a centralised market is not going to eliminate congestion. Therefore, investigations of the value of enhanced transmission are valuable for the planing of the future grid. In the following two sections we will present two developments of the project, which may be used to plan and justify such investments.

2.3 The Hemispheric Ensemble Database

As previously mentioned we have generated a 1.5 year database of high quality hemispheric ensemble forecasts with 4 forecasts per day and 48 hours in the optimisation work package. This database contains 7.8mio forecast hours of the weather on the northern Hemisphere (48hx75memberx365days*4*1.5). These hours can be interpreted as perturbations within the current climate. And, since we do not know how the climate will develop in the future more than it will not revert to the past, we believe that they are in fact more suitable for future scenario studies than traditional re-analysis data of the past 50 years. Such data neither have the same spatial resolution nor time resolution. More over, those datasets never include wind speeds in an appropriate height for modern wind turbines. Instead the tradition is to provide 10m wind speed as the only boundary layer wind speed. With this database it is also possible to study frequency distributions of arbitrary capacity distributions within the northern hemisphere

The pro and contras for this model setup has already been discussed in Section 7.1 and the experience was mostly positive. For the time being the focus on forecast accuracy is highest in Europe and as long as Europe is mostly concerned about the day-ahead horizon, there is clear preference to limit the model area and increase the spatial resolution. This means that forecasts from the database will have a 10-20% reduced quality on RMSE compared to the best system configuration at present. As we shall discuss later, there are however other useful targets than RMSE and we can therefore state that the database has a wide application area despite the higher RMSE.

Wind power forecasts generated from the database has been verified in Europe, Canada and China. The quality is similar to other systems in the same spatial resolution in the first day, but the quality is higher in the second half of the day ahead horizon.

Because the database has a higher resolution in time and space than any other ensemble system on this spatial scale, the database can be considered the best that is available. It should especially be noted that the ensemble spread in a hemispheric model system is as good as it can be, because it is model generated and is not a result of approximate lateral boundary handling, which is sometimes the case in a limited area ensemble. The correlation between forecast error and ensemble spread is therefore the best possible.

It can be expected that this type of forecasting methodology will become feasible at some stage in the future, when processing capacity delivers more performance per unit price, or there is a larger share of costs for the high-end forecasting product among end-users in other areas.

2.4 A new approach to Estimate Consistent Local Wind Generation

The studies mentioned in the previous section require a new approach for estimation of valid power curves or rather local generation, because the power curve estimation can not be changed from region to region.

State of the art is to quality check power measurements and discard incorrect values. However, this is not possible in large-scale wind integration, as only aggregated production numbers will be published. A quality check will only eliminate part of the errors, because reduced availability within a small wind farm may not be detected from typical settlement data. Thus, a new approach has been designed to circumvent the quality check completely by inherent distribution of uncertainties on all wind turbines.

The new approach is developed to fit the existing training and forecast methodologies using least square fitted power curves. The purpose of the new approach is therefore to estimate consistent wind power production numbers for each area of interest from published aggregated generation. The wording consistent refers to consistency with the forecast and not consistency with the measured local generation. For the purpose of model training and forecast accuracy, this definition is a clear benefit, because this prevents that the training misinterprets seriously wrong information as true information.

A typical problem in the handling of aggregated generation is that there are many uncertainty factors, which may have some local impact, but less impact over areas. It is seldom possible to separate the uncertainty factors related to weather forecasts and the operating status of the wind turbines, because the true turbine availability is seldom reported.

Negative influence on the power curve estimation is reduced by aggregating over larger areas. This approach has been adopted for dispersed wind power for many years. What is new is the ability to estimate a consistent local power generation which results in that the aggregated generation is consistent with the local and that the local power curve is responsive. The primary goal with this is that the same forecasting methodology can be used for everything from local grid congestion management to optimized bidding into the market for aggregated generation. In this way we attack the problem addressed in the first section, which is that it seems to be the ensemble forecast producers responsibility to simplify and tailor the system to a level where it is meeting the requirements of the industry.

This approach is already now suitable for forecasting for large volumes of wind power, but will become an even more useful tool if balancing of wind power is shared over even larger areas as it is currently practised in Germany.

2.4.1 Practical Application of the approach

This approach first downscales the aggregated generation according to short-term forecasts with an iterative approach. In the first iteration, a reference power curve for the entire area is generated. In the next iterations local power curves are generated and the total error is distributed according to the uncertainty of the ensemble. Thus, regions with small ensemble spread are hardly influenced by the mismatch between forecast and actual generation. This results in an accurate power curve, which may reach the true lower and upper extremes of the power curve. If a certain area has systematically more forecast uncertainty, then the power curve will be less responsive to reflect the difficulty in forecasting at that location.

The total error in the training is in this way reduced compared to training on the true local generation, because the forecast error for aggregated generation is relatively lower than for localized generation.

This type of power curve training also takes places directly in the weather model grid for dispersed wind power or at individual wind farms. In this way the so-called upscaling principle is eliminated and large volumes of wind power can be handled more consistent, because all units are explicitly forecasted for with the forecasted wind speed valid at the physical location. This is distinct different than using up-scaling where off-site information is used to estimate the generation.

The fact that estimated local measurements fit the weather forecasts better and the power curves therefore become more responsive solve two challenges in one step. It is with one approach possible to forecast the upper and lower extreme with an ensemble for dispersed generation, while still using a least square fit. Previous approaches implied an either or and that had impact on either the accuracy or the validity of the ensemble spread.

It is a great benefit that one forecast can be used for uncertainty forecasting even on the local scale and at the same time be RMSE optimized for aggregated generation. This has increased the applicability of an ensemble technique and reduced the maintenance costs.

With centralized European publication of all wind farms, it is possible to apply model grid-scale forecasting for all wind power. With some assumptions, it is also possible to extend the forecast to include areas where no measurements are published.

By using the hemispheric ensemble database, it is in addition possible to create scenarios not only of the entire European market, but also of the North American market, the Asian market or combinations of them in something that could be a GigaGrid. In this way nearly all wind power can be simulated consistent and the benefit of a grid reinforcement can be computed and simulated.

The generated data and generalised power curve fitting approach provides the possibility to develop important expertise in large and very large-scale wind integration and will be of significant value in the globalized future, as well as centralised forecasting and publication of relevant data, as probabilistic forecasts will be necessary to enhance transparency and competition.

2.5 An Extreme Forecast Example

In the previous sections we discussed the trend on a very general level and concluded that there already exist a predecessor of a European SuperGrid environment in central Europe. One of the ideas behind the SuperGrid is that the predictability of wind power increases by increasing the region, i.e. aggregating over more dispersed wind turbines in the forecast. This argument is valid, but we will below present a case, where area aggregation is insufficient and that should demonstrate that even in very large areas there will be errors in the day-ahead forecasts. The error is demonstrated by Figure 1, Figure 2 and Figure 3. The first figure shows the day ahead forecast valid at 5pm over Denmark. It shows that a large group of ensemble members have almost no generation from wind symbolized by dark blue everywhere. A few forecasts have almost full generation in large regions of Denmark. From this forecast it is evident that the uncertainty at this time is extremely high.
The second figure confirms that the uncertainty is rather high even on the 9 hour horizon. In fact this would be the last weather forecast before the event in a 6 hour schedule. The figures therefore illustrate that there is extreme uncertainty in the weather, but high likelihood of no generation in Denmark. The likelihood is higher for high generation in Germany than in Denmark.
The third figure shows the probability of the German wind power. All figures are taken from the forecast setup named T04 on Figure 4, which performed considerable better than the T10, but also considerable worse than the operational/backup setups. Table 3 shows how T04 differed from the other systems.

The 11th of June 2010 was an unusual day seen from a wind power perspective, because there was no sign of a significant low pressure system at the beginning of the forecast on the 10th of June. The low pressure system developed almost instantly during the forecast as a result of an instability in the large scale flow. The large scale weather pattern was dominated by two stronger low pressure systems in the Atlantic, respectively north of Scotland and west of France. There were several other very weak low pressure systems, some of these were located around Denmark in a northerly flow of colder air. None of these small low pressure systems were located over Denmark and none seemed to cause any wind in Denmark, because they were small and insignificant. The weather had been warm up to that day in both Denmark and Germany. This meant that the northerly flow from Norway in the middle troposphere would lead to a situation where cold air would lie on top of warm air with continued heating from below in the afternoon. One could therefore expect strong rain in the late afternoon, but the formation of a low pressure system with strong wind would still be an unlikely evolution. However, if the afternoon convection would be simultaneous over a large region, then a low could develop. Without a large ensemble of forecasts one could not estimate the probability of the low pressure systems development.

Figure 4 shows the typical output of combined forecasts (also referred to as meta forecasts). They are all dampened to take account for the uncertainty in the evolution and suggest therefore values in the middle of the possible range of values, only partially influenced by the probability. Figure 3 on the other hand shows that the probability of the generation, which is not represented by the different meta forecasts on Figure 4. The typical pattern is that the meta forecasts are MAE or RMSE optimized, because those measures are by tradition used to evaluate if forecasts are good or not. A collection of meta forecast do therefore not represent a likely physical outcome, but a defensive guess on what is expected to give the least error. Their spread or rather difference to each other is therefore suppressed, because of RMSE optimisation.

One could now argue that the forecast with the best MAE/RMSE measured over 18 months (T10) turned out to be be a good minimum forecast even though it is a combined forecast. However, there is no combined forecast, which represents the actual generation. Only the percentiles of the ensemble and the individual members above P80 provide a good estimate for the maximum.

A study of the operational day-ahead forecasts showed that they had the center of the power generation 100-200km too much to the East. Because Germany is regarded as a SuperGrid, such an error is not counted even though it is an error. If we would verify all forecasts in smaller regions on DSO level, then T10 on Figure 4 could probably be defined to be the best forecast, because no meta forecast would have the power generation at the right location. That is because T10 is not double punished anywhere for having generation at the wrong time. In this way we can argue for that the absolutely worst SuperGrid forecast may be the best DSO forecast, if RMSE is used as an error measure on both.

The fact that Percentile 85 was a good forecast shows that the event was difficult, because the further out in the percentile range we find the best forecast, the more difficult the event, if the distance to P50 in MW is simultaneously large.

It is not unusual that the P85 forecast is the most accurate. Thus, such an event could easily be overlooked as being extreme, because there have been events with more imbalance. What is interesting in this event is that the bulk of the ensemble forecasts were extremely different and the ensemble spread was large. This indicates, that if the final forecast coming out would have been wrong, whether as a meta forecast of different deterministic forecasts or the best guess of an ensemble, the event could have been causing a major issue on the grid.

Because of the potential security issue on the grid, we decided to carry out forecasts in high spatial resolution, as this is the expected trend for forecasting in the future. From the results of the experiments, we conclude that the forecasts could have resulted in an imbalance of 12GW of wind power, if the Netherlands, Denmark and Germany would be counted together in a future forecasting scenario.

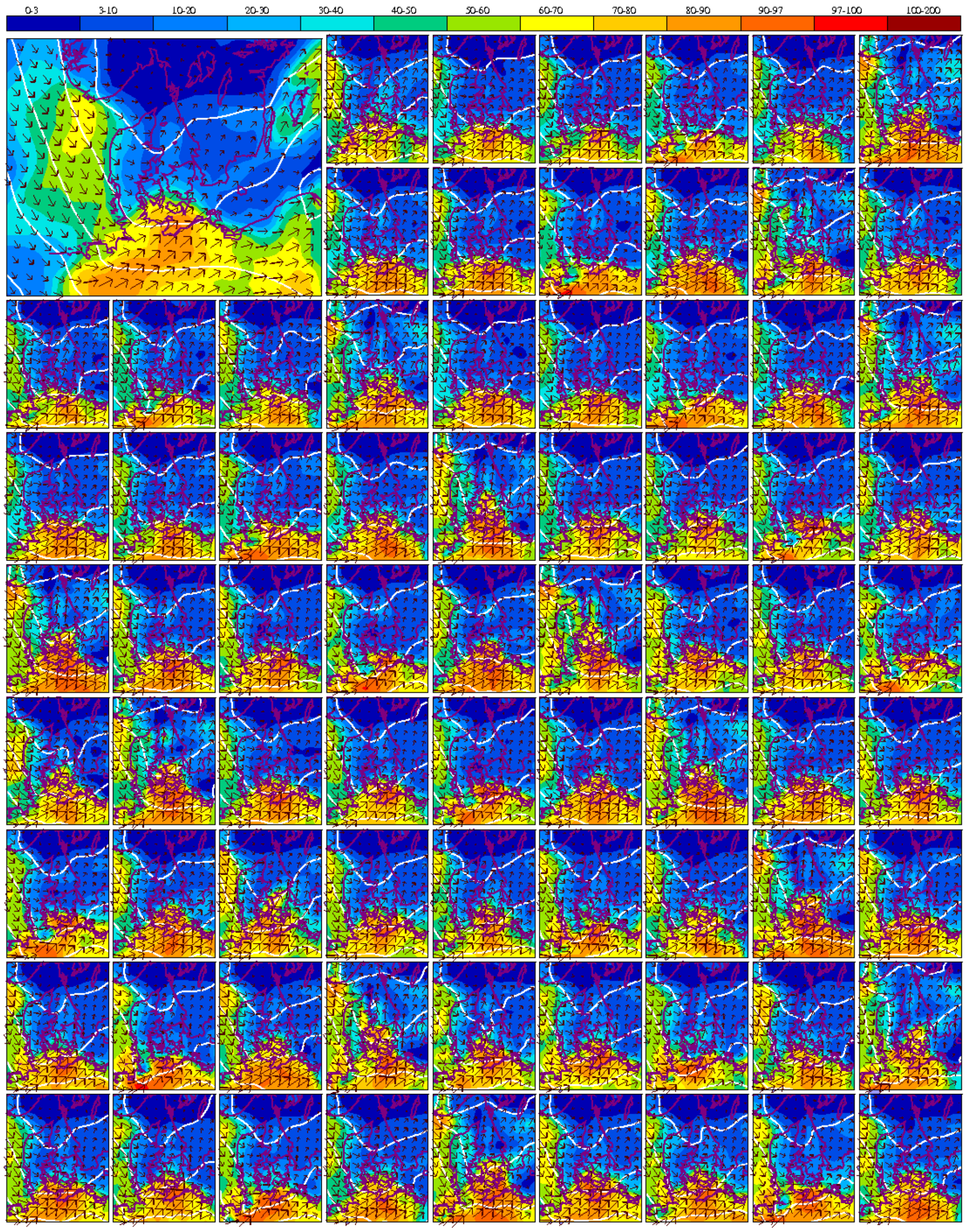

Figure 1: Short-term Ensemble Forecast (9h) from the 11th June 2010 displayed in horizontal plots of wind power load factor, inclusive wind speed arrows and isobars. The large figure is the mean of the 75 forecasts.

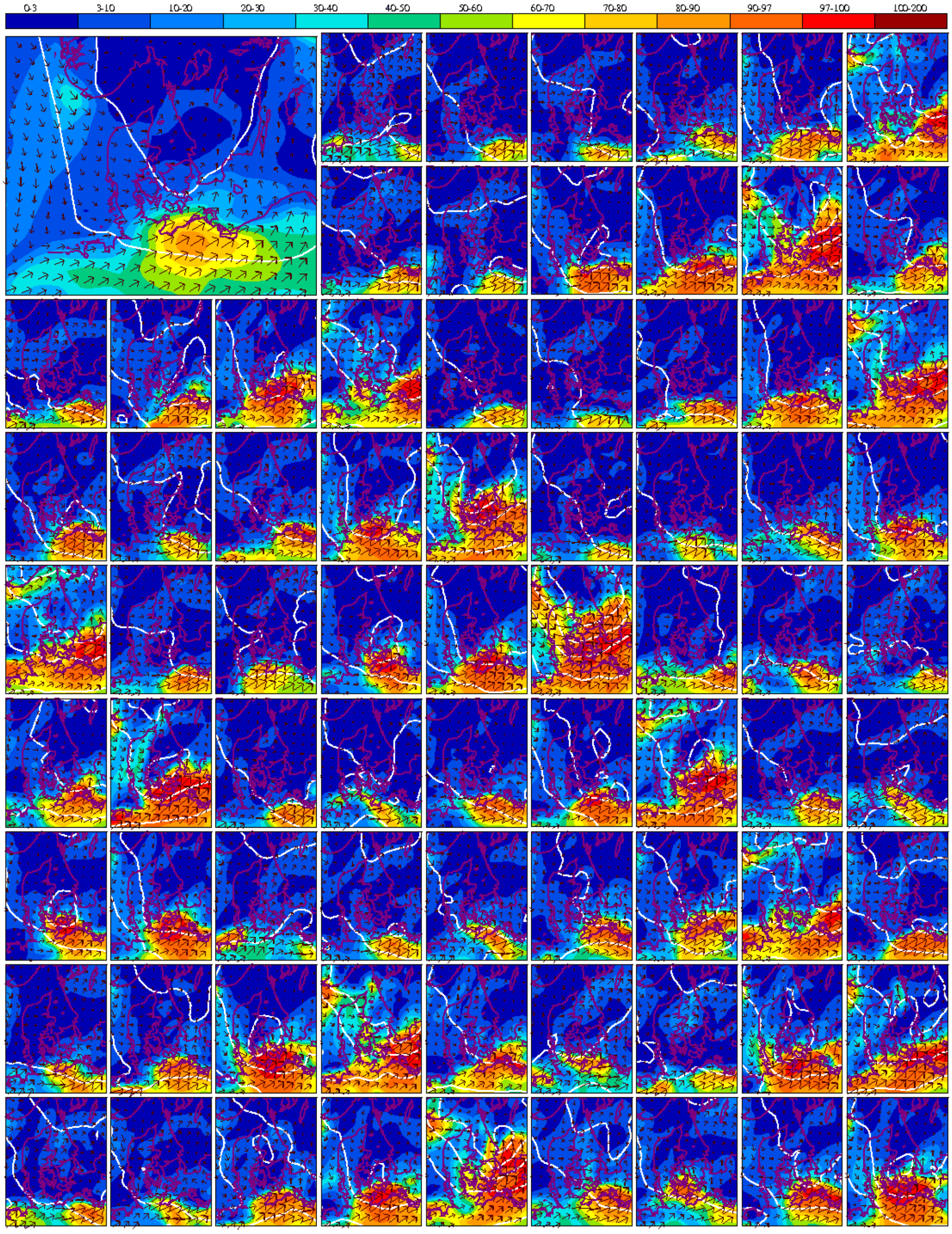

Figure 2: Ensemble Forecast (39h) from the 10th June 2010 displayed in horizontal plots of wind power load factor, inclusive wind speed arrows and isobars. The large figure is the mean of the 75 forecasts.

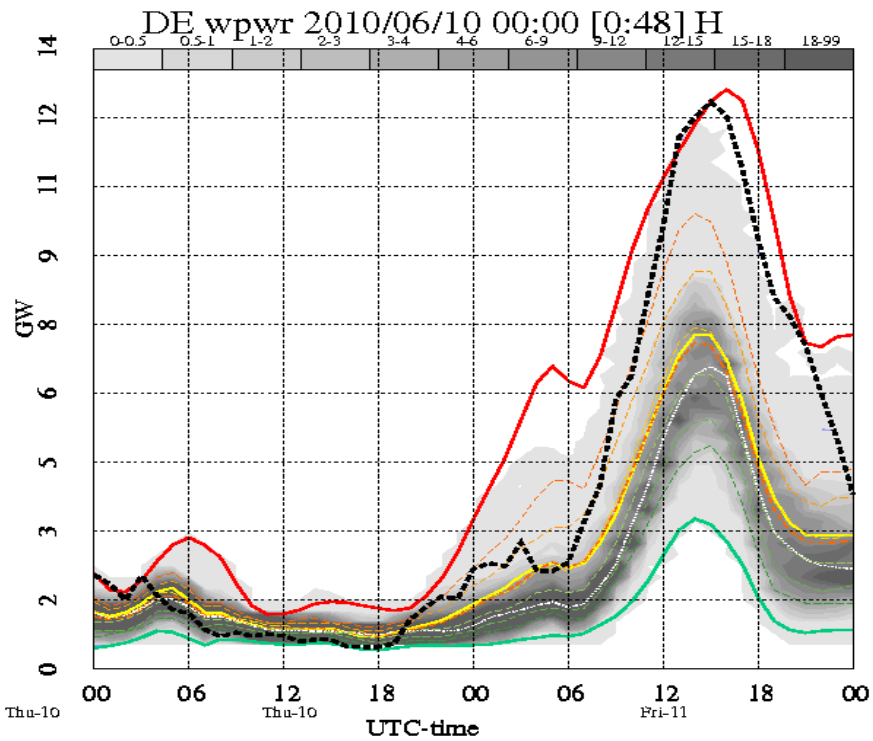

Figure 3: Probabilistic 48h wind power forecast for Germany on the 10th June 2010. The red line and the green line are the maximum and the minimum, respectively, the yellow line is the best guess, the white line the highest probebility, the orange dashed line is the mean, the 60%, and 80% percentiles, the light green dashed lines are the 40% and 30% percentiles.

An interesting aspect was that the more we increased the spacial resolution of the model system, the worse the result became regarding the likelihood of high power generation.
The fact that the power system did not experience an imbalance of 12GW could be regarded as luck. It is difficult to assess what the total imbalance was in this case. However, the answers of four of the balance responsible parties in Denmark and Germany indicated that all forecasters had forecasted well under the actual produced power. It was only WEPROG's P90 and maximum forecast that over-predicted the production.

One of the test setups (model "T10" in Figure 4) had been showing very promising results over entire 2009 and the first 5 months of 2010 with improvements of 25% of the total error in Germany and 15% in Denmark. This is an exceptionally strong improvement for this type of work even though the model configuration is fundamentally different than the other setups in several stages of the model chain. However, in this particular case, we noticed a complete failure to produce the correct power production. The average load factor was 10% in the forecast, while it was estimated to reach 55% of the installed capacity at the peak. The operational systems forecasted 42-44% generation at the peak correctly timed but spatially shifted (see Figure 4).

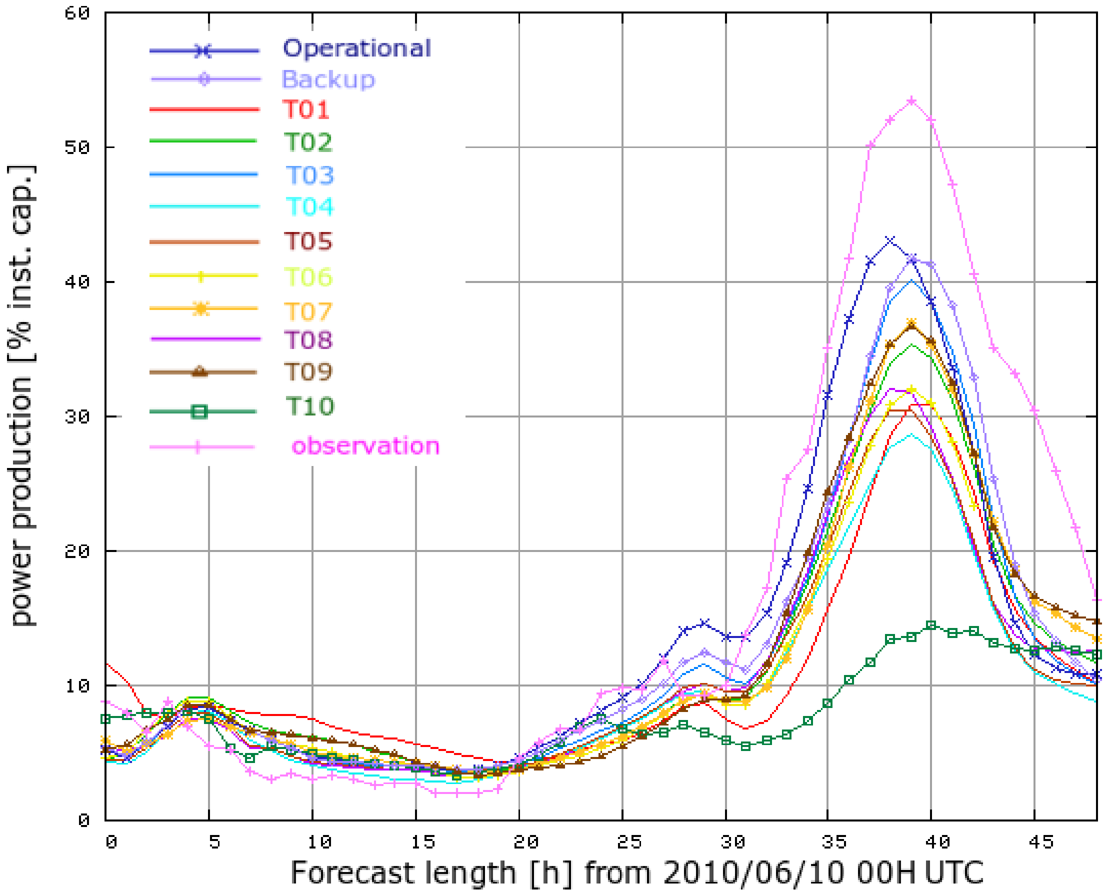

Figure 4: wind power forecast at the 10th June 2010 and 48h ahead from different forecast systems. "Operational" stands for operational setup, "Backup" for backup setup, "T" for test setup. See Table 3 for details regarding the differences of the various systems.

In T10 only one ensemble member produced a forecast, which was close to the actual. However, the other forecasts had so little wind power generation that the final forecast was seriously under-predicting the event. It has been found that the forecast, which performed best, was one of those with the poorest long term statistics. The influence on the combination forecast was then minor.

The event emphasizes that extreme care is required when evaluating forecasts, because T10 could match the other systems measured over the month on MAE/RMSE, but made one extreme error during the month, which is much more serious than any of the other system's smaller errors during the month.

A similar event without impact on Denmark took place only 30 days later, this time on the border between Germany and Poland. The relative performance of all systems were similar to the 11th of June case. In this time, a low pressure system developed over Poland in southerly flow, where again only 1 forecast caught the event in the T10 setup.

The two events were similar in one way, but different in many other ways. This time the critical spot was in Poland, just outside the SuperGrid and the high density of the wind generation, and the T10 had plenty of wind. The root of the problem is however the same, i.e. the low came from a mountainous region. Therefore, the two events complement each other in the study of the cause of the error.

Given the success of T10 it would be extremely convenient for the future development, if those two events could be handled with just the same quality as the average of all the other systems. Therefore extremely many model formulation changes have been applied to T10 to search for the crucial parameters. So far, only softer Norwegian mountains and coarser model resolution seem to have positive impact, but such changes make MAE/RMSE scores worse measured over long time.

From the study of more than 1000 test forecasts for the two events (see Table 3), it has in fact become apparent that typical future forecasting techniques with very high spatial resolution seem to have a potential risk to fail to forecast such events on the day-ahead horizon. What happens is that several small low pressure systems develop independently in a region of 400x400 km. For dispersed wind power this would mean almost no generation, but in fact one large low pressure system developed with significant power generation as the result.

	OPR	Backup	T01	T02	T03	T04	T05	T06	T07	T08	T09	T10
Description of boundary/ outer model system												
Number of nesting levels	2	2	2	2	2	1	2	2	2	2	2	2
Area of boundary generating model	Std	Ext	VL	Ext	Ext	Glb	Ext	Ext	Ext	Ext	Ext	Hem
Number of outer boundary generating models	8	10	8	8	11	1	8	8	11	8	8	1
Advection scheme in boundary models	Seml	Euler	Seml	Seml	Multi	Seml	Seml	Seml	Multi	Seml	Seml	Seml
Number of increments per run on boundaries	1	1	1	2	1	1	2	2	1	2	2	2
Resolution of boundaries	0.45	0.45	0.45	0.45	0.45	0.4	0.45	0.45	0.45	0.45	0.45	0.6
Boundary update frequency of Boundary models	6	6	6	6	6	n/a	6	6	6	6	6	6
Boundary update frequency of inner model	1	1	1	1	1	n/a	1	1	1	1	1	1
Number of members in assimilation of boundary models	8	10	8	8	11	n/a	8	25	11	8	8	8
Description of nested/inner model system												
Number of members	75	75	75	75	75	75	75	75	75	75	75	8
Area	EU	EU	EU	EU	EU	VL	EU	EU	Small	Small	Small	Ext
Number of increments per run	1	1	1	2	1	1	2	2	1	2	2	2
Number of assimilation members	75	75	75	25	75	75	75	0	75	75	25	8
Spatial resolution	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.22	0.22	0.22
Initialization hour	0	0	0	6	0	0	0	0	0	0	6	0
Sea surface roughness in coastal regions	Std	Std	Std	Std	Std	Std	Std	Std	Std	Std	Std	Increased
Iterations of SST analysis scheme	1	1	1	3	3	3	3	3	3	3	3	3
Analysis Increment in PBL	Std	Std	Std	Std	Std	Std	Std	Std	Std	Std	Std	Reduced
Power curve method	Std	Std	Std	Pcref	Std	Pcref	Pcref	Pcref	Std	Pcref	Pcref	Pcref

Table 3: Description of model configurations for the test case on the 10th June 2010. (Abbrev. Meaning: Std = standard, Ext. = extended, Glb = global, VL = very large, SemL = Semi-Lagrangian, Multi = use of either Euler, Euler Upstream, Semi Lagrangian advection, Pcref = reference power curve scaling used to increase the responsiveness of power curves)

Even if we consider the operational forecasts in the two events, it was found that they under-predicted strongly, while the percentile P85 was in both cases nearly perfect. The fact that the operational systems did warn about the event was positive, but it is not positive that increased spatial model resolution reduce the forecast skill in such events. Knowing that the distance between P10 and P90 was extremely high, it would be desirable to already on the day-ahead horizon trade this uncertainty into the market in competition.

The benefit of such a possibility becomes evident, when we note that the low pressure system seemed to give concurrent uncertainty corresponding to 40% of full capacity in Denmark, Germany and the Netherlands, which is equivalent to 12GW. This amount of uncertain generation could otherwise give rise to concerns and price volatility considering the relative small volume in the intra- day market.

2.5.1 Discussion of the lessons from the 11th of June event

From this experience we can therefore summarise our learnings as:

- A continued focus and evaluation criteria being to find the best deterministic forecast, where "best" is evaluated by a single error measure such as MAE or RMSE, may lead to severe problems on the grid now, if a forecast fails during a low demand period and even more so in the future

- The warnings provided by ensemble forecasts are becoming more important, because the risk of a major error is increasing, especially with increasing capacity in combination with changing predictability of the weather.

- It is not trivial to state what a good forecast actually is. Is this the one that rarely or never makes serious mistakes or is this the one that makes the lowest average error ? The results suggest that this is not the same model formulation.

What we want to emphasise with this analysis, is that extremely much care has to be taken, when evaluating what is a good and a bad forecast. Over several decades, a core experience that developed in meteorology was that all information is good information, also if it is uncertain or wrong information.
This is state of the art practice particularly in data assimilation and will also need to be adopted in all other applications using weather forecasts, if the weather forecast is the key input to the application. Users of weather forecasts automatically experience that finding sooner or later, if they analyse the performance of their forecasting system.

The wind power forecast is more or less a direct output of weather forecasts and there is no other possibility to forecast wind power than by using weather forecasts. Demand forecasting, solar energy and hydro energy are not in the same degree driven by weather forecasts. In wind power context it is therefore required to adopt the experience from meteorology, which is the ability to use information that is only maybe correct. The 11th of June 2010 is a proof that there are events that can only be predicted by using uncertainty forecasts. By not doing so, there is a risk of extreme price volatility and possibly also grid security issues. We have demonstrated that future forecasting techniques enhance the risk and the increasing installed capacity of wind power will further increase the risk.

The example has been chosen to demonstrate that the evolution on the weather depends on factors that cannot always be predicted explicitly. Small disturbances in a complicated system can lead to instabilities that may have major impact on human life, both direct and indirect. If we accept that wind power should be a major contributor to the energy generation in the future, then it may also be a good idea to adopt the electricity market to what is optimal for handling wind power. The feasibility of wind power will level out, if wind power is always operating in competition, while many other generators feel very little competition at times where the wind is not blowing and the demand is high.

2.6 Best practises on the use of Ensemble Forecasts

It is soon 20 years ago, where it has been scientifically demonstrated that ensemble forecasting is the state of the art method to find the instabilities in the atmosphere that causes most forecast uncertainty and error. For an ensemble forecast provider like WEPROG it is then questionable, whether it is actually best practise to search for the best RMSE optimised forecast, because it is has become apparent that optimisation on MAE/RMSE is against the interest of the TSO and the consumers.

The case has also shown that model changes that lead to better average deterministic forecasts lack internal perturbations that however seem to be required to forecast instabilities. Such instabilities are difficult to forecast and therefore it is often better to not predict them at all, if only a single value can be used and if low RMSE is the target. The likelihood of a forecast being hit by a double punishment is high, because of an incorrect phase. Thus, the stiff model system that may develop a small insignificant low may not be punished as much as the one that tries to develop a bigger and stronger low, which may be more realistic.
In the 11th of June case the instability did not take place in the RMSE optimised forecast and all evolution was suppressed.

2.6.1 The Role of Stakeholders

It is a challenge to avoid price volatility if dispersed wind power is handled in a liberalized market. The wind farm owners are not directly participants and would need to agree on a solidarity principle to protect themselves against other stakeholders.
Nobody can expect that the balance responsible parties (BRP) have incentives to avoid high volatility events, unless they are penalised by the balancing cost or from the wind farm owners. The BRP may earn on one account and loose on another when prices change. Thus, the incentive on reduced volatility lies exclusively at the TSO among the direct participants in the market. Therefore, it is more than surprising that to our knowledge only two TSO's in Europe request probabilistic forecasts inclusive maximum and minimum from WEPROG, where it seems that no other provider is offering ensemble forecasts in hourly or quarter hourly resolution. There are certain characteristics of the two receivers. One is balance responsible for the largest pool of wind in Europe and the other is poorly interconnected. Common is that both have received such forecasts for more than five years. Thus, there has been a relative recession in the interest of probabilistic forecasts in the time, where the wind power capacity has been growing the most Europe wide.

The TSO and consumers should essentially both prefer a stable energy price and a competitive market as this is the best basis to operate the grid in a stable manor. Price volatility is the result by not trying to tackle forecast uncertainty in time. Typical market stakeholders see volatility as an opportunity, thus it is only regulators and system responsible parties that have an interest to prevent that the price volatility grows to a level where the system becomes inefficient.

A direct parallel can be drawn to the economy leading to the finance crisis in 2008/2009. The crisis was triggered by volatility and resulted in years of artificial low interest, where the flow of money almost stopped through the banks. A similar situation could occur in the electricity market, which can for example be triggered by a physical handling problem or a market problem. Therefore it is important to find methods to prevent volatility in a market compatible manor.

2.6.2 How to define a Forecast Optimization Criteria

We have raised the question on the feasibility of MAE/RMSE optimisation on the basis of high forecast uncertainty. How should we take account for a minority of ensemble forecasts , which is very far away from other forecasts ? It seems like the most appropriate step beyond the pure deterministic philosophy is to consider high and low uncertainty separate. If the uncertainty is low, then RMSE/MAE is a fair target. If the uncertainty is high, then the forecast should rather try to avoid the bigger mistakes and at them same time suggest an additional pre-allocation of reserve, which will prevent price volatility and increase system security.
Our operational experience suggests that a good forecast lies approximately in the middle between the minimum and maximum of the ensemble, also referred to as the mean or the weighted mean of the ensemble. A skewness adjustment may be considered, if the available reserve is asymmetric. The forecast should follow a linear trend over several hours, unless it is very certain that the actual value will not follow this trend.

In this way the choice of wind power forecast can contribute positively to increased efficiency of the energy system and because of the reduced volatility, there is reason to expect that the balancing costs will be lower, even if the MW deviation between forecast and actual generation will increase.

2.6.3 No Severe Storms in Europe for 5 years

The installed capacity has been growing and there may already be overcapacity in some regions. Consequently, the amount of large generation units may reduce, which in return reduces the capability to balance severe wind power forecast errors. Grid security may therefore need to get more attention again over the next years, because there has not been severe storms that brought attention to extreme weather forecasting for 5 years.
It also seems that in both countries, Denmark and Germany, the amount of available reserve compared to the increase of installed wind capacity has increased until now, and imbalances have not been balanced over regional borders (Germany), but rather treated separate and thereby prevented competition and the evolution of new techniques to forecast reserve requirement and enhanced competition on the reserve market.

Finally, it should be pointed out that the wind power capacity is increasing in new growing markets (especially in the eastern European counties), where there is not yet much experience with wind power. Although the European wind energy association is usually initiating a dialogue between new markets and fully established countries, the local situations can never be transferred one to one.

The global production capacity of wind turbines also continues to grow despite the crisis in the economy. Therefore, we have to expect that the next stormy January will be a challenge for both the Danish and the German systems. It is in fact unusual that the weather has been relatively calm over almost 5 years. However, there is now much more wind power on the grid, new staff is responsible, new companies are involved and there is also little experience on how the offshore and newer larger turbines work in extreme weather.
A parallel can be drawn to the economy in the year 2005-2007. Things worked extremely well on the surface, but there was a latent problem building up, which became apparent very sudden and required dramatic changes and new solutions.

3. Summary and Conclusions

In the light of the above described experience, we will continue the analysis of the 11th of June and other similar events in this project and give more emphasis to the understanding of the fact that an RMSE optimised forecast, which is significantly improved over longer periods can cause severe errors. It is important to get a deeper understanding of the "trap" that lies in the RMSE optimisation, because an ensemble forecasting system must not develop in this way with improved technology.
WEPROG has over many years solely produced ensemble forecasts, while other providers of ensemble forecasts always separated between deterministic and ensemble forecasts, where the two systems have been kept separate and the corresponding ensemble also performed very different for that reason. The success in the application of ensembles in applications, where statistical methods are required, have therefore been rather rare. WEPROG's philosophy has been to deliver the same output as deterministic forecasts in terms of time resolution, but in slightly lower spatial resolution for cost and handling reasons. The 11th of June case actually suggests that it may be required to keep the ensemble responsive, while optimising on RMSE in a deterministic context. Nevertheless, this case also shows that the combined meta forecasts of the ensemble and the percentiles have skills that deterministic forecasts do not have. It is therefore important to use a forecast that lies reasonably well in the middle of the range of likely values. This will reduce the big errors and thereby enhance grid security, although the RMSE/MAE may be slightly higher than they could be with one-sided optimisation.

This recommendation is consistent with the results from the demonstration period carried out in the HRensemble project (see final report at http://www.hrensemble.net/reports.html). Here, it was found that the ratio between RMSE and correlation penalises errors most that are inconvenient for the power system.
The 11th of June case indicates on top of that, that for high uncertainty events, it is most important to take precautions to not make a short lasting extreme error. Regardless of how we optimise, the near 50% weight of the minimum and maximum in extreme uncertainty has to be enforced explicitly in the construction of the final forecast.

Given these criteria it is non trivial to state that a forecast is good or bad by only looking at a RMSE measure. However, we have also learned that all information adds value and therefore it is also not important to be able to measure the relative quality of forecasts, but rather to be able to drag out the important information from a large number (ensemble) of forecasts.
There is little doubt that the use of forecasts optimised with the above mentioned targets need more MWh balancing than an RMSE optimised forecast. However, we have come to the conclusion that the cost of those reserve hours or short-term trading will be less, because of the possibility of pre-allocation with fixed prices and/or adjustments in the intra-day and hence reduced price volatility. This postulate requires assumptions to be made and hence is very difficult to prove objectively, because the choice of wind power forecast determines the market clearing price and the dispatch, which again have influence on the prices of the intra-day market and vice versa. Nevertheless, we will further investigate this topic in the cost optimisation.

From an end-user perspective, it may be difficult to assess the amount of work spend on wind power forecasting. However, it is our observation that the technology advances are not being explained sufficiently well, often because of competition reasons among forecasters. Although it is rare, we still occasionally observe that parties have difficulties in providing historical production data for forecasters for apparently commercial reasons. Publication of production data enhances competition and transparency and publication is therefore to the benefit of wind power producers at the end of the day. The information becomes only commercially sensitive, if both wind speed and power generation is published. Considering today's volume of wind power production data, the value of an individual wind farm is however very limited anyway.

However, if advances are explained better, then it will also be understood that state of the art is moving ahead and potentials to use the technology in new applications can develop, e.g. in the EMS systems or the operation and maintenance area. Wind power forecasting and the way it is handled is very different from region to region and this is a benefit for the international forecast providers. It is today not possible to keep a wind power forecasting system as state of the art by looking at one region only, not even if this is 25GW as is the case for Germany. It is necessary to work with observational data in different climates and surroundings, as we experience climate changes and extreme weather with long return periods.