Project title	**DEWEPS -** ***D****evelopment and* ***E****valuation of new* ***W****ind forecasting tools with an* ***E****nsemble* ***P****rediction* ***S****ystem*
Project identification	**2009-1-10260**
Project period	**2009 – 2011**
Responsible Institution/Company	**WEPROG ApS**
Reporting period	**Final Report**

Weather & wind Energy PROGnosis

WEPROG

inclusive physical uncertainty from Ensembles

Document Type: *Final Report*
Project Title: DEWEPS
Project ID: FORSKEL 2009-1-10260
Funding Authority: Energinet.dk
Submitted: February 2012
Responsible Organisation: WEPROG ApS, Denmark
Responsible Officers: Dr. Corinna Möhrlen, Jess Jørgensen

Table of Contents

1 Introduction

The topic of large scale wind and solar integration is getting more and more attention, because both types of generation increase in volume and their competitiveness are increasing. The most populated regions have today locations with sufficiently good resources to make some kind of renewable energy financially viable. In the case of wind power, the best 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.

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, care has to be taken because the energy system is frequently 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 report will show the difficulty in evaluating what are good and bad forecasts and will demonstrate that good forecast skill for one application is not the same as good forecast skill for another application.

2 Sensitivity Experiments for Forecasting Quality Improvements

With increasing requirements on forecast quality of weather and weather related forecasts for especially variable energy generation sources, there is a need for basic research to sustainably improve over time. It is no longer only the availability of computing resources that is responsible for good or increased quality, but also many other aspects in the model formulations, such as the surface representation, the horizontal resolution as well as the right amount of vertical resolution or time resolution of the original numerical processing.

The multi-scheme ensemble approach is a suitable approach to verify many of such aspects as well as theoretically new parameterisations for meteorological processes. In this project we have been carrying out a large number of sensitivity experiments on all these matters as well as a verification and testing of a wind profile theory on its suitability as a new parameterisation scheme for WEPROG's Multi-scheme Ensemble System (MSEPS). All experiments aimed to find the correct strategies to improved forecast quality.

2.1 Testing of a new wind profile theory with the multi-scheme ensemble

Considerations from the research project HREnsemble under the PSO FORSKEL Programme 2006-2009 (Project Identifier 2006-1-6387, see also www.hrensemble.net) and operational experiences with wind power forecasting over the past 8 years suggest that focus should be brought on effects of the frictional forces in the entire Planetary Boundary Layer (PBL). Any consistent improvement in the formulation of friction will reduce wind forecast errors and as a consequence wind power forecast errors.
This conclusion was in line with the work done by J.C. Bergmann at Risø in the 2nd half of the 1990ties (Bergmann 1995, 1997, 1998). Bergmann questioned the assumptions leading to the Ekman spiral and pointed out that no measurements indicate the existence of a super-geostrophic wind speed, which is one of the usual Ekman based assumptions. To verify the assumptions made by Bergmann a number of modelling tests were required, because Bergmann's theory questions fundamental assumptions common in all weather prediction models in the world (1998).

The first part of the DEWEPS experiments on the friction processes in the numerical weather prediction models was therefore constructed for the purpose of investigating the applicability and possibility of integrating the theory developed by Bergmann in 1995 into WEPROG's Multi-scheme ensemble prediction system (MSEPS). First, Bergmann's theory was investigated in detail, how it had been validated previously and under which conditions it would be applicable in the MSEPS context.

The results of this first investigation provided the necessary information to decide whether a further implementation of the theory is feasible or whether other means of improving the quality of the NWP models inside the MSEPS in the context of wind power forecasting will have to be applied.

A verification of the forecasted wind profile at the German offshore platform FINO 1 near Borkum was therefore conducted. This test was carried out for a period of 2.5 years (2004/10-2007/02) against MSEPS ensemble forecasts. The measurement data for verification is publicly available for research projects and was provided by the DEWI, the German Wind Energy Institute and BSH, the German federal maritime and hydrographic agency (http://www.fino1.de).

This overall verification showed that the wind shear bias and error between 30m and 100m altitude is small (approx. 0.09m/s bias and 1.07m/s RMS error) and that the bias does not grow with forecast horizon. This result suggests that there is no systematic error in the model friction over sea, which was assumed by Bergmann's theory. Bergmann's own non-published verification of 21 events in 2004 at a Danish coastal site indicated a much higher RMSE error of ~2m/s, when extrapolating from 30m to 120m with Bergmann's theory (personal communication Bergmann, 2010).

In general it can be stated that the wind shear over land is stronger than over sea. Thus, we cannot conclude that Bergmann's error is twice the error of the MSEPS. The 21 events chosen by Bergmann were all stationary, because the time derivatives in the theory are set to zero. One must therefore expect that the error over a year would be higher, first because the variable conditions are more difficult than the stationary conditions, and second because there is no equilibrium between the forces. Bergmann himself (personal communication, 2010) explains the error with the fact that he had to use simplifying assumptions in his theory in order to solve the problem analytically and considers especially one assumption critical. This is the fact that there are no horizontal temperature gradients, which implies that the pressure gradient is height independent. This condition is in real life mostly applicable, when there is either weak or no wind.

In addition, Bergmann's theory assumes neutral or stable static stability and stationary conditions, which means that it is practically seldom possible to find events of that type and hence it has become questionable whether Bergmann's theory could add significant value to the numerical models.

Another limitation of Bergmann's theory is the assumption of a homogeneous surface. Current and future wind power will be located in regions with good wind resources and often relatively close to the coast or located on the top of hills. The surface around the turbines will therefore rarely be homogeneous. The fact that Bergmann's theory is hardly applicable in real life is one problem. Another equally serious problem is that the theory is incomplete, because it does not prescribe the known turning of the wind direction with height. The theory is therefore missing some physical constraint or is maybe based on a wrong assumption in this matter.

2.1.1 Verification Results of the Experiment at the FINO1 Offshore Site

For the verification a total of 148.000 data values for 75 members were verified, where the wind shear was computed from 30m to 100m with the entire 75 member MSEPS ensemble prediction system. The following results were found during the study of the wind shear at FINO1:

- there is a negative bias on the TKE schemes with -0.2 - -0.4m/s
- there is no trend that the bias increases with forecast horizon
- year 2005 has a tendency to a stronger negative bias than other periods
- an RMSE error of 1.07m/s on all forecast lengths is low for a mean wind speed of 9.5m/s, but high, if we compare to the average shear of 1.3m/s
- the members that are among the best do not have a near zero BIAS
- there is a trend to a stronger negative bias in the autumn, which suggests that there is a slight over-mixing in the model in unstable cases
- the measured wind shear lies more than 85% of the time within the predicted spread of the wind shear
- the ensemble spread is better at unstable conditions than stable conditions
- the error of the wind speed is 10% higher in unstable than stable conditions

The conclusion of this experiment is that the friction process cannot be modelled systematically wrong as the Bergmann theory (1995, 1997) suggests. If this would be the case, the bias in the wind shear should develop with forecast horizon and hence also the overall error. If there is an error, such errors may only occur at specific conditions. An example of such conditions could be, when the wind and waves are out of balance. Bergmann's theory however does not cover this case, because of the constraint ot stationary and homogeneous conditions. The fact that the forecast bias is around 10% of the RMS error suggests that the error pattern is very random and changes sign rather frequently. A variable error pattern at one location suggests a low spatial correlation of the error. Thus, for aggregated wind power we should not see a major contribution of error from the shear effects and therefore we can expect a low economic loss from this kind of error.

More over, it is obvious that stationary conditions anti-correlate with price volatility, because the stationary conditions are more predictable. Thus, there can hardly be an economic value in improving modelling of stationary conditions.
From this analysis we conclude that Bergmann’s theory will most likely not be able to add value in numerical weather prediction. The analytical approach is opposite to the numerical, because assumptions are required to eliminate non-linear terms in the analytical theory and their corresponding mathematical solutions. This leads to approximate results that appear to be of worse quality than what the numerical models already provides, even when applying them in selected events, where the approximations seem to be valid.

After careful investigation of alternatives to Bergmann's theory, WEPROG has come to the conclusion that the findings from the MELTRA project (2003-2006, ELTRA internal funds), which indicated the importance of a correct description of the friction process above Norway, should replace the work that was planned in relation to the Bergmann theory. This work will bring most value to the wind energy community as Southern Norway has direct influence on approximately 20GW of wind power today, especially considering that Denmark's and a major part of Germany's wind power generation is often under influence of low pressure systems located near Southern Norway.

In fact, the forecast errors due to the surface representation in Southern Norway causes that a major fraction of the wind power in Denmark, Sweden and Germany occasionally cause correlated error .

2.2 Extreme event analysis

Although good/fair and poor forecast quality has traditionally been evaluated in terms of mean error statistics, it is nevertheless usually some specific events that are most important when studying error patterns. A daily random background error of a few percent of the capacity is not a problem for the feasibility of wind power.

However, from the point of view of the volatile behaviour of electricity prices and the need of reliability in the supply, it is evident that a reduction of the number of events of extreme high forecast error will add most value.

Therefore, emphasis will be on analysis of extreme errors with the objective to find the error sources so that some of the events can be handled better in the future.

2.2.1 Case October 15-17, 2009

In our first case, we focused on an event in October 2009, where there was a loss of around 10 mio Euro reported over 2 hours on the wind generation as a result of extreme negative prices the night between Saturday and Sunday.

We conducted the power analysis at the German power system, because the error was larger and more significant in Germany than in Denmark. However, we discovered, that the meteorological error actually took place over Norway, Sweden and Denmark, although it resulted in a wind power forecast error that was largest further south.

The left forecast on Figure 1 is our reference point, the operational forecast for the German day-ahead market. As can be seen from Figure 1, there was a 6 hour phase error of the entire German power prediction. The error was approximately 7GW during a 6 hour period, which is equivalent to twice the permanent secondary reserve. The wind started to increase rapidly in the north-western part of Germany first.

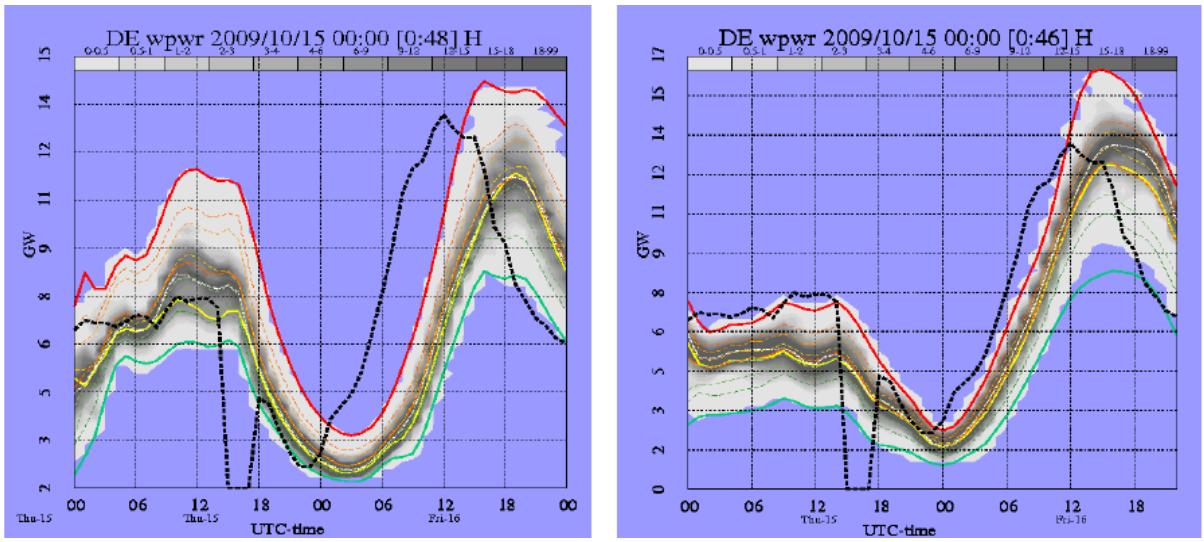

Figure 1: Probability figure of the forecast with large phase error of approximately 6 hours (left figure) and improved forecast with reduced phase error, where the orography has been changed in the model system (right figure). The published German upscaled wind power is shown as a black dotted line and is unacceptably far outside the forecast uncertainty band. The zero values between 15-17 hours on the 15th represent missing data values.

The event was a worst case scenario, because the wind generation started to ramp up in the hours of the week with the lowest demand in Germany. This is where there is already full export in the direction of the hydro generation centres north and south of Germany. This means that Germany had to balance all the forecast error locally for several hours until the planned night export stopped.

Figure 1 indicates that the error can be characterized as a kind of a phase error. This means the error originates from the weather forecast error and not from the translation of wind speed to wind power.

The estimated generation lies far outside the ensemble spread for many hours. Thus, all ensemble members failed to forecast the event correctly. Form communication with German wind balance responsible parties all forecast providers had significant errors in the event, although the errors were of different size and structure. This indicates that the error is common in all model systems and therefore relevant for a detailed study.

After a set of 11 sensitivity experiments (see Table 1), the right probability presentation in Figure 1 shows the improved simulation of 15th of October 00UTC, where the most promising changes were combined in one setup.

The error is reduced, but the uncertainty is still not large enough to cover the error as the black dotted line is still outside the ensemble spread.

The probability charts in Figure 1 show 300 power forecasts. The uncertainty is of the order 5GW between the minimum and maximum forecast during the ramp. This uncertainty level is consistent with the high error pattern reported by the German balance responsible parties.

EXP ID	Member	Analysis	Horiz. Resol. [km]	Description
24	1-75	NCEP	45	Europe reference forecast
23	1-75	CMC	45	Europe reference with CMC analysis
38	1-75	NCEP	22.5	orography adopted to europe45 grid
39	1-75	NCEP	15	Europe 15km standard
40	1-75	CMC	22.5	orography adopted to eu225 grid
50	1-75	NCEP	19	Europe 19km standard area
81	137	CMC	45	tall orography in large europe grid
82	137	CMC	22.5	vertical diffusion test on tke
83	137	NCEP	22.5	extreme tall orography
84	142	NCEP	22.5	tall orography
85	137	NCEP	22.5	NHEM boundary with tall orographgy
86	137	NCEP	22.5	short-term 6hour forecasts, low orography

Table 1: List of experiments carried out for the analysis of the influence of improved orography to prevent high forecast errors

Although Figure 1 does at first glance leave the impression of a phase error, this is nevertheless not correct. The original forecast never reached the same peak as the estimated power did. The enhanced forecast is closer to the peak, but there still appears to be a catalyst missing to trigger a sufficiently strong development of the low pressure system causing the ramp event. On top of that, it is very rare that the estimated generation is outside the ensemble spread for more than 1-2 hours in a row for aggregated generation. So there is no doubt that there is a significant additional error.

We therefore started to search for what such a catalyst could be. We were looking for instability mechanisms that could explain this development. In fact, we found an upper level front at the tropopause level moving from the Atlantic to Norway and then southward over Denmark and finally sweeping over Germany in the 48 hour period of the forecast. This movement is shown as an analysis in Figure 2 and as forecast in Figure 3, both as potential vorticity maps.

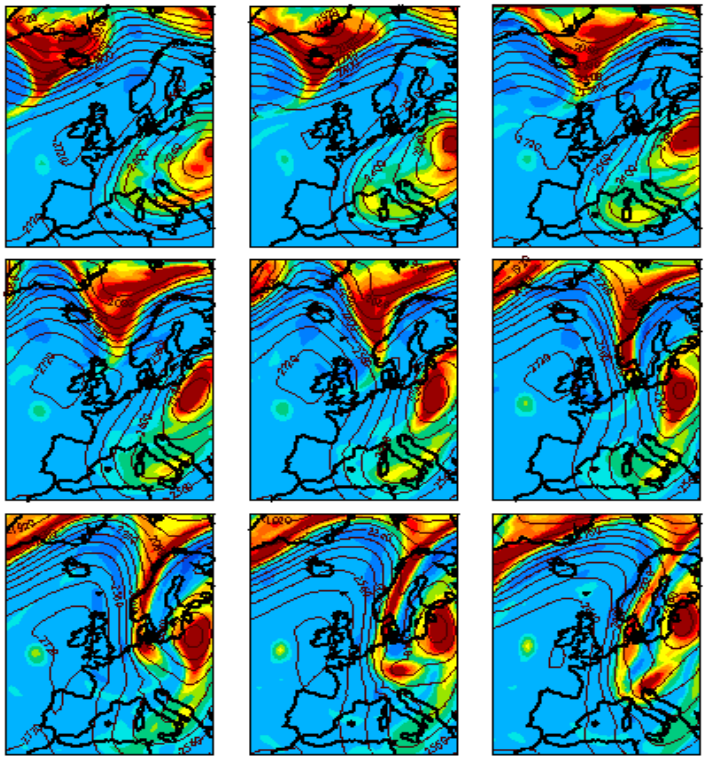

Figure 2: Analysed IPV in 6 hour steps from the 15th to 17th of October.

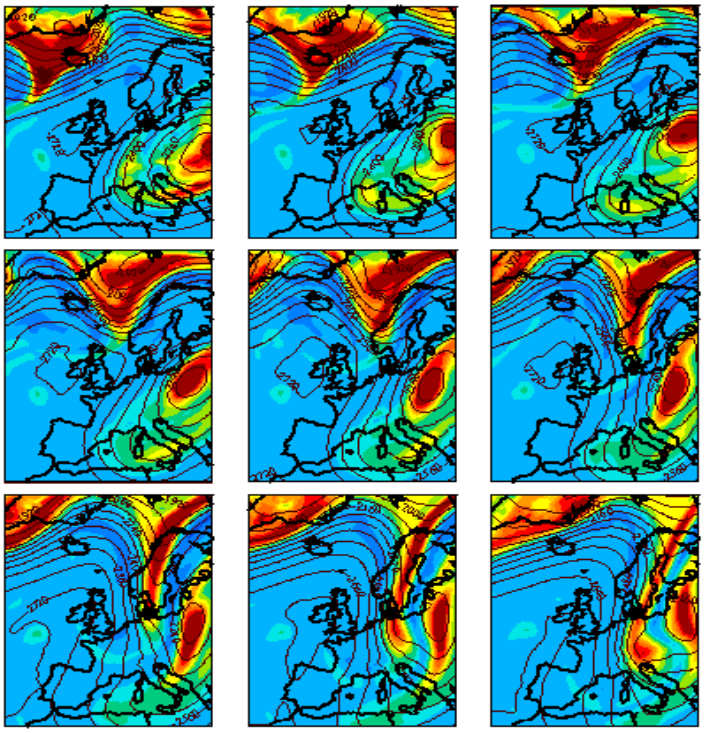

Figure 3: IPV forecast in 6 hours step from the 15th to the 17th of October.

The time stamps are running from upper left to lower right. The figures start with the state valid at the 15th Oct 00 UTC on the top left. The next figure (top middle) is 6 hours later, the following (top right) is valid at 12 UTC. The last figure (lower right) is ending the forecast on the 17th of Oct. 00 UTC. An isentropic potential vorticity (IPV) map is easy to understand, because it is shown in a plane that is a material surface in adiabatic conditions. This allows us to follow a conservative quantity over long distances.
All the contour lines on the IPV figures show the Montgomery stream function, which is an indicator of the pressure gradient force on the isentropic surfaces, thus the geostrophic wind blows parallel with the contours. The green colour correspond to a typical tropopause level IPV and red colours to stratospheric IPV values.

We focus now on the center sub-figure of Figure 2 valid at the 16th of October 00UTC. The focal point is the low south of Norway aligned with the IPV anomaly indicating a narrow upper level front heading south westward.
We now trace this point back 6 hours and locate the low north east of Scotland and further back until we reach the departure point at the 15th 00UTC, which is about 2000km west of Norway. On Figure 3 we see the same structure and timing. This is a promising result, because we know that a catalyst is required for the forecast to develop reasonably correct. Note however, that the forecast is developing very different when the front starts sweeping over Germany. This is no surprise, because the low pressure system triggered strong snowfall and therefore diabatic effects that had impact on the IPV values, i.e. the catalyst.

Nevertheless, the IPV presentation form provides precise and useful information of the weather development and in particular when the development starts to go wrong. It is directly visible that the domain of dependence for the 48 hour forecast goes 4000km back from Austria over Denmark and far out in the Atlantic. This would not be visible by looking at a low pressure system development only. A tropopause level system can move extremely fast and may change speed and direction in case interaction with another system in the low troposphere takes place.

Apart from tracing back the origin of the catalyst (in terms of IPV) it is also possible to subjectively warn about the risk in forecast mode. That means an oral warning could be given that the forecast does or does not seem to react on a catalyst that is visible and known to have the potential of developing severe weather. The warning should definitely be given, if the event was about to take place while the demand is ramping down, as this would require scheduled units to ramp double as fast down as normal, which can be critical during a weekend.

The fast movement of the upper level front also explains the 6 hour delay of power generation in the day-ahead forecast. If the low pressure system would have interacted and followed with the upper level front, then the surface winds would have been picking up timely. For this reason, the event is a very important lesson of what sudden evolutions in the weather can cause to the power system and how a warning system may be activated. It is not standard practise in forecasting centres to permanently watch out for such events. Attention is usually given to isobaric levels, where such effects may not be apparent. However, an upper level front may only be visible on an isentropic surface near the tropopause level. Therefore, it seems like subjective analysis is required, because the tropopause level changes constantly, especially when it is windy. The presence of an upper level front may not always lead to an amplification of a low troposphere system. The systems must be aligned in a certain way to amplify each other. Also, the spatial scale, static stability, wind shear and phase speed are critical parameters for the level of interaction.

We can say that not one single ensemble forecast out of many reacted correct. Multiple numerical simulations of this event also indicated that increased resolution can cause that the likelihood of a strong interaction is reduced. Thus, a too high spatial resolution may suppress strong sudden developments. Subjective evaluation can however only a be a fall-back solution for critical events or to study and improve the automatic solutions, because essentially all ensemble forecasts have to be checked. In real-time an automated solution which also computes probabilities is to prefer, because it can send out warnings well in advance. The target should therefore be that some ensemble members react on “the” catalyst while others should not.

This study is giving good hints regarding the configuration of the ensemble members. Regardless of that, it has been shown that a substantial number of experiments are required to find the best strategy for the development of a catalyst to occur and be automatically detected by the system without producing too many false warnings.
Another important lesson from this event is that the model area size should not be compromised as the sharp fast moving upper level front extends from far north and down to approximately latitude 50. The sensitivity to area size and timing of the upper front and the vertical mixing between the upper level front and the low level lee wave will therefore have to be further studied in order to find a strategy for such a warning system.

2.2.2 Case June 11, 2010

The second case is a convective summer event where Germany, the Netherlands and Denmark had correlated forecast errors. Thus, the spatial scale of the event was rather large considering that the scale of the weather systems are often smaller in the summer months.

This error is demonstrated in Figure 4, Figure 5 and Figure 6. Figure 4 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.

Figure 7 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 6hour 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.

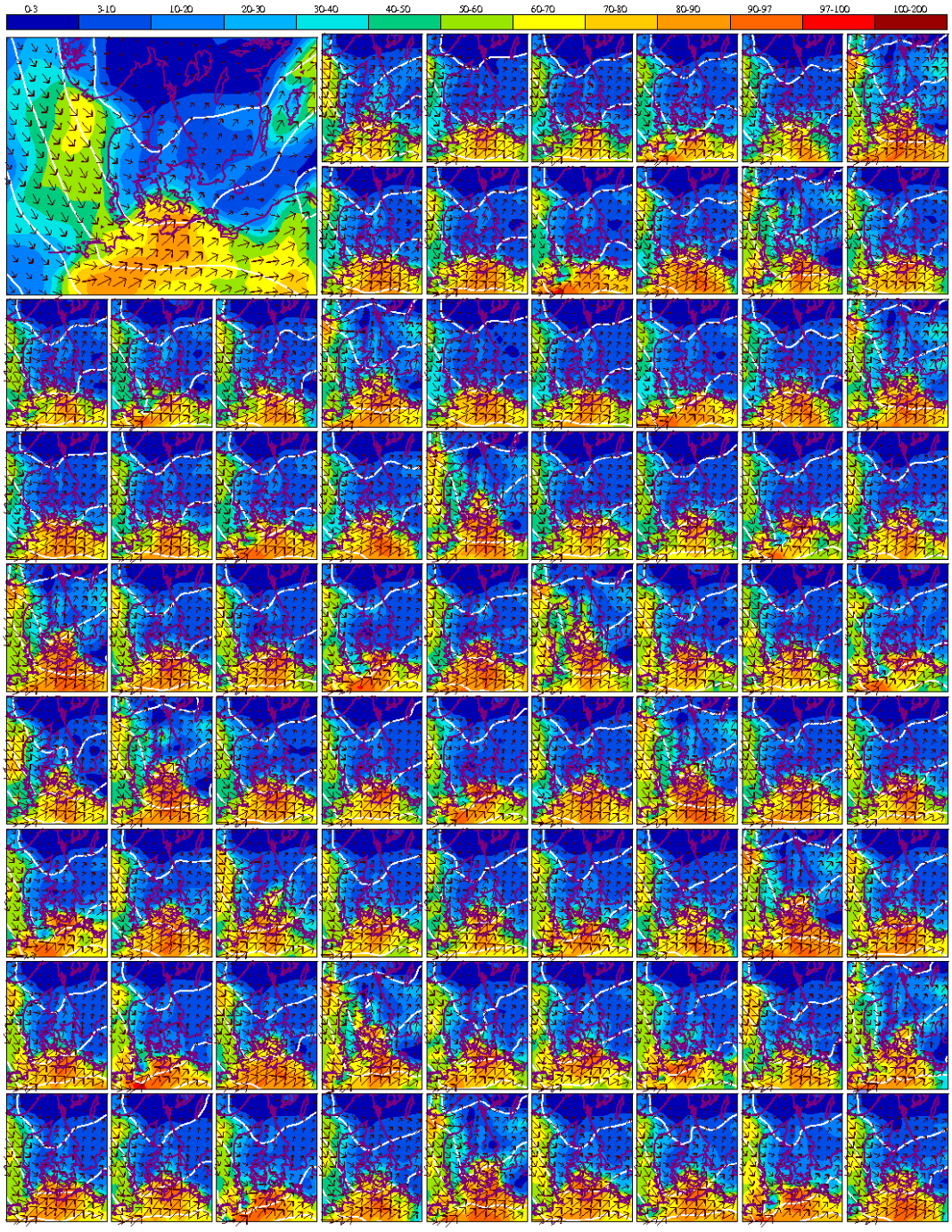

Figure 4: Short-term Ensemble Forecast (9h) from the 11th June 2010 displayed in horizontal plots of wind power load factor, inclusive wind speed arrows and isobars. The large figure is the mean of the 75 forecasts.

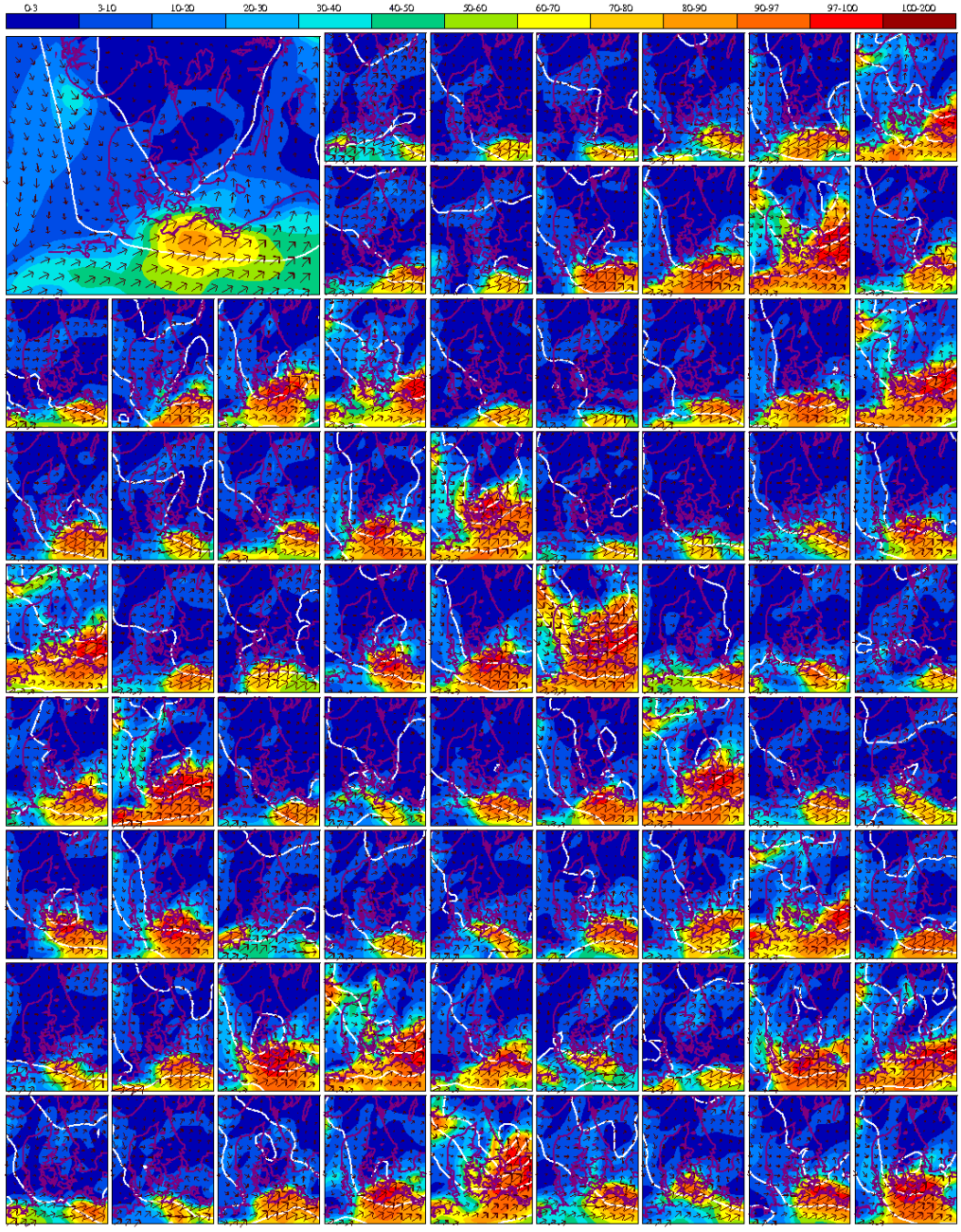

Figure 5: Ensemble Forecast (39h) from the 10th June 2010 displayed in horizontal plots of wind power load factor, inclusive wind speed arrows and isobars. The large figure is the mean of the 75 forecasts.

Figure 6 shows the probability of the German wind power. All figures are taken from the forecast setup named T04 on Figure 7, which performed considerable better than the T10, but also considerable worse than the operational/backup setups. Figure 5 shows how T04 is differed from the other systems.

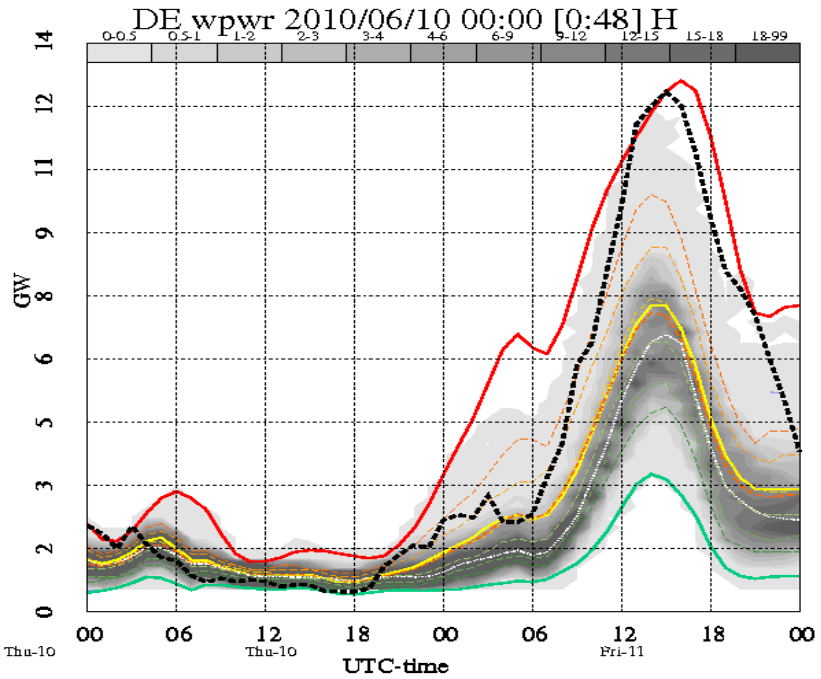

Figure 6: Probabilistic 48h wind power forecast for Germany on the 10th June 2010. The red line and the green line are the maximum and the minimum, respectively, the yellow line is the best guess, the white line the highest probability, the orange dashed line is the mean, the 60%, and 80% percentiles, the light green dashed lines are the 40% and 30% percentiles.

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. However, none of these small low pressure systems were directly located over Denmark and none of them seemed to have the strength to cause any wind in Denmark.

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 short wave radiation, which heats up the surface strongest.

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 growing low could develop. Without a large ensemble of forecasts one could not estimate the probability of the low pressure systems development.

	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 2: 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, PC_{ref} = reference power curve scaling used to increase the responsiveness of power curves)

Figure 7 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 6 on the other hand shows that the probability of the generation, which is not represented by the different meta forecasts on Figure 7. The typical pattern is that the meta forecasts are MAE or RMSE optimized, because those measures are by tradition used to evaluate, whether 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.

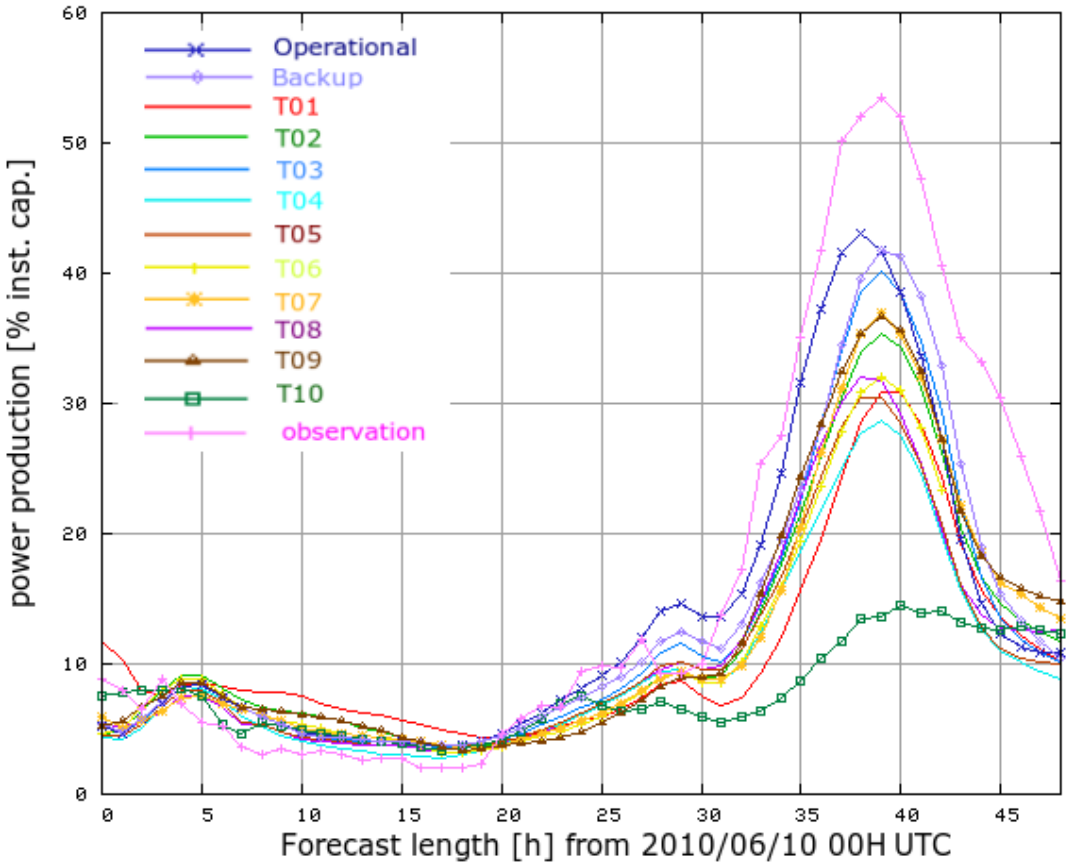

Figure 7: Wind power forecast at the 10th June 2010 and 48h ahead from different forecast systems. "Operational" stands for operational setup, "Backup" for backup setup, "T" for test setup. See Table 3 for details regarding the differences of the various systems.

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 comparison of the upper left chart of the ensemble mean on Figure 4 and Figure 5 illustrates clearly that the position of the extreme is wrong in the forecast.

The short-term forecast shows that there should be wind power generation in all TSO zones, whereas the forecast rather has the center of the power generation moved 200km towards south east. The individual day ahead ensemble forecasts differ extremely much from each other, because of the high uncertainty in this particular event, whereas the short term does not.

It is interesting that the forecast for all Germany on Figure 7 is completely hiding that the location of the power generation center is wrong. A regional evaluation would double punish the forecast with the most generation for always having the generation in the false region, whereas the All-Germany forecast is not hit by double punishment. The best forecast on regional basis would therefore be the forecast without any peak (T10), while the opposite forecasts (Operational/ Backup) are the best on all Germany. Thus, on regional level we reach the opposite conclusion about forecast quality than on the large-scale and the so-called best large-scale forecasts would be considered good for the wrong reason. They could also have had the wind extreme centred somewhere else, where there was less installed capacity and then they would not be considered good.

The fact that percentile 85 was a good forecast shows that the event was difficult, because the further out the best forecast lies within the percentile range, the more difficult the event, especially, if the distance to P50 in MW is simultaneously large. It is not unusual that the P85 forecast is the most accurate, but this particular event is special because the forecast with the best RMSE long-term statistics is T10, which in this particular case evaluated worst on the aggregated level. One can therefore extrapolate that use of a few RMSE optimised forecasts is likely to underestimate events of low probability caused by strong convection and thunder. The non-RMSE optimised forecasts were in this particular case very helpful on the meta forecast. One can say that it was pure luck that some forecasts hit the right amount of power for the event. Therefore, 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 these experiments, we then concluded 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.

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 coincidence and it is difficult to assess what the total imbalance was in this case. However, when asked, the answers of four of the balance responsible parties in Denmark and Germany indicated that all forecasters had forecasted much less than 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 7) 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 compared to the operational reference forecasts. 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 in the peak hour. The operational systems forecasted 42-44% generation of the peak that was correctly timed, but spatially shifted (see Figure 7).

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 severely 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 from the best forecast on the combination forecast was therefore 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.

At 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, and the T10 had plenty of wind. The root of the problem is however the same, i.e. the low developed behind a mountainous region. Therefore, the two events complement each other in the study of the cause of the forecast 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.

Even though the model setup T10 performs poor in these two events, the overall RMSE statistics over the 3 summer months was found to be at the normal level for summer conditions and comparable to other model configurations. Obviously the conditions causing the error pattern are rather rare. They do only lead to some spurious values in the forecast quality and can therefore easily be overlooked whether or not the events are of severe character. Only one similar event occurred in 2011. This time in July, but the event lasted for two days in a row. Again T10 under-predicted the event while the less RMSE optimised forecasts were better.

The conditions, which appear to trigger the sudden development of a low in the summer are characterized by:

1. A cold weak low located in the mountains at initial time
2. Mid tropospheric advection from the mountains in direction of wind turbines
3. Strong heating of the ground due to short wave radiation during daytime
4. High humidity in the afternoon caused by southerly winds

If all conditions are present, then convection on the small scale evolves and interacts with the weak low in the middle of the troposphere, which then causes a sudden amplification and as a result increasing wind speeds. The event will in most cases terminate a longer warm period.

An extreme example of this type of weather took also place end of August 2003, where cold air from north suddenly terminated an extremely warm summer. The August 2003 and the 10th - 11th of June 2010 event were the strongest of this type within 10 years and therefore the two most important benchmark cases.

From the study of more than 1000 test forecasts for the two events in year 2010 (see Table 2), it has in fact become apparent that typical future forecasting techniques with very high spatial resolution seem to increase the risk of failure to forecast such events on the day-ahead horizon. What happens in the high resolution modelling is that several small low pressure systems develop independently on the meso-scale. For dispersed wind power this would mean almost no generation. The models should develop a larger low pressure system on synoptic scale, but this evolution is apparently suppressed for reasons that may have to do with how the processes interact in the model system.
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 reduces 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 with an action to prevent price volatility in the intraday market.

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. This amount of uncertain generation could otherwise give rise to reliability concerns and price volatility in the intra- day market.

2.2.3 Discussion of the lessons learnt from the 11th of June event

From this experience we can therefore summarise our learnings to:

- A continued focus on finding the best deterministic forecast, where "best" is evaluated by a single error measure such as MAE or RMSE, may lead to too defensive forecasts which as a consequence increase the risk of a large volume of wind power to be sold in the intraday market or alternatively the need of large amounts of negative reserve.

- The warnings provided by ensemble forecasts are becoming more important, because the risk of a major error is increasing, especially with increasing capacity and if the frequency of such weather phenomena would increase.

- 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 ? Our results suggest that both criteria cannot be met by 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. What is good for one party is bad for another. In research we must therefore consider the overall system performance from a technical and economic point of view.
We have demonstrated two typical events in which models have difficulties simulating in a correct manner. One event is typical for autumn-winter conditions and one is typical for summer conditions. Nevertheless, it has to be emphasised that the possibility that a severe forecast error can occur at any time, because the 15th - 16th of October event can occur all year around, although it requires a strong jet stream. In fact, the forecasts for the two events fail for the same reason. Both rely on correct interaction between different weather systems in the vertical direction. The upper level system normally overtakes the lower level system without interaction, but certain states of the atmosphere allow for interaction. An exponential growth of the aggregated low pressure system can occur due to one or more instability conditions.

The conditions required for vertical interaction are proven to be difficult to find given the fact that both cases could not be forecasted fully correct on the day ahead horizon, not even with more than 1000 forecasts of each event. This suggests that large ensemble sizes with considerable spread are required to be able to compute realistic probabilities of such extreme events.

The two examples demonstrate that higher spatial resolution and deterministic modelling are two steps in the wrong direction in the context of forecasting of sudden extreme evolutions. In the future energy system, we can expect that a major fraction of the price volatility will be caused by special circumstances in the weather, while the demand is extreme.

The typical pattern will be normal balancing costs 95% of the time, while extreme costs may occur in the remainder. The total balancing costs will be dominated by the peak prices for the largest renewable energy pools, because their forecast error is likely to be the root of the price volatility.

2.3 Forecasting experiments

While we had been focusing mostly on sensitivity studies of single events in the previous paragraph, we give now attention to a number of long-term simulations to investigate the impact of model changes on the forecast quality over longer-term periods. The goals of the longer sensitivity experiments are to

- quantify the long-term benefit of the case study code developments
- verify the numerical stability
- generate training data for operational usage

In previous studies (e.g. The HREnsemble project, www.hrensemble.net) 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: Verification on wind power in Ireland, France, Denmark, Germany
- Stage 3: Analysis and comparison of results to previous experiments in that period
- Stage 4: Conduction of 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 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 input and hence results as in the 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 3.

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 3: Description of the model changes for the NWP sensitivity experiments.

2.3.1 Model Improvements related to surface roughness & 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 surface. The numerical stability may be compromised, if mountain summits reach far up into the atmosphere where the prevailing wind speeds are strong while the jet stream crosses over the mountains. Practical experience confirms this especially during periods, where the surface temperature in the mountains are much higher than the air temperature.

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. Similar considerations and tests have been applied to Norway, but only with corrections in upward direction of 500m.

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 for the success of the extended mountains 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, because steep valleys increase numerical instability risks. 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. This pattern is consistent with that SYNOP reports from the valleys do not have as strong impact on the initial conditions in the tall orography setup. On the one hand the roughness, the surface type, the direction of the valleys, the height of the sun, local inversions, clouds or fog are all factors, which may reduce the representativeness of measurement reports from mountain valleys. It is technically not feasible to ever solve this problem exact, because the model grid boxes must represent the average conditions valid for the area and not for a specific location. It is a small step process to improve on this topic by use of intelligent parametrisation. What makes the process cumbersome is that only small subjectively invisible improvements can be expected, where the wind generation is located. Thus, every experiment requires 1-2 years of model simulation to detect any impact from the changes. This is an ongoing process where a large cluster of servers run long-term simulations with four 48 hour forecasts every day.

2.3.2 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 balance between computational cost, location of the uncertainty sources and time available to deliver a real-time forecast. 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 of this simulation 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 these 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 over a large area. The combination of tall orography and hemispheric model area has been tested in single events. On this basis it has to be concluded that there is further work required to get the climatic data and surface representation synchronised and updated to reduce the risk of model instability.

A particular difficulty of the hemispheric model system is that the improvement pattern from tall orography is different. Also, the likelihood of a numerical instability is increased, because there are much more extremely tall mountains ranging far up into the troposphere. This means that the feasible model time step becomes less predictable and more time step reductions take place. This may compromise the training, because the time step may have some influence on the statistical results. Extreme wind speeds over the Rocky Mountains may therefore have impact on the applicability of the statistical training in Europe. It was found that running a nested system over Europe using lateral boundaries from the hemispheric setup produced more reliable results than the nested hemispheric model setup, because all ensemble members could run through with the designated time step opposed to the hemispheric setup. This is a significant result, which on the one hand proves the efficiency and reliability of limited area modelling, but on the other hand also illustrates the complexity in the configuration of such model systems.

2.3.3 Objective Scores of selected model improvements

Table 4 provides some samples of statistical results of the various experiments. In the beginning, the testing environment was best referred to as a "test-bed". The complexity increased because the differences in skill of the various experiments were small. A large number of statistical tests were carried out on forecast data in Denmark, Germany and Ireland in order to quantify the impact of changes. In particular, it was found that a given change seldom resulted in consistent changes in the forecast skill over the different countries. Changes in the roughness and orography over Canada and Greenland had most impact on Ireland, whereas changes in Norway had stronger impact in Germany and Denmark, but sometimes not in the same direction. The difference between Denmark and Germany can most likely be explained by the higher offshore wind power penetration in Denmark. The error pattern is different and more volatile, because of the two 360MW offshore wind farms that each correspond to two 700MW onshore wind farms in terms of power production, which is a significant fraction of the total installed capacity especially in the DK2 area. 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 4: RMSE statistics in % of inst. capacity of the experiments that were carried out in December 2009.

The experiments in Table 4 have been used sequentially in stage 4. 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 and was therefore stopped after one year.

The preliminary result suggests that none of those experiments in table Table 4 are superior in performance to the changes on tall orography described in the previous section (referred to T10 on Figure 7).

The models shown here were however added to form an optimized 115 member ensemble, which was set up for the demonstration phase of the project.

2.4 Hemispheric ensemble forecast database from model simulations

A database has been generated out of the simulation experiments described in the previous sections. This database can be regarded as an additional project deliverable, because the output data from the project milestones on improved forecasting can be used for more than just forecast evaluation. Also the 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 in many of these 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. New projects require better justification in order to reduce technical and financial risk.

The database is unique in several ways, because it is 75 independent forecasts in hourly resolution. Such a database is unique in the world and can in fact be used in studies, development and demonstration in clean-tech areas, although it will most likely 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.
The database specifically consists of 1.5 years of high quality hemispheric ensemble forecasts with 4 forecasts per day and 48 hours look-ahead time. This database in fact contains 7.8mio forecast hours of the weather on the northern Hemisphere (48h x 75member x 365days x 4 x 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 the hourly time resolution. Moreover, 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 pros and cons for the model setup used have been discussed in Section 2.2.1 . The experience was mostly positive. For the time being the focus in Europe on forecast accuracy is highest and as long as Europe is mostly concerned about the day-ahead horizon, there is a 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 found in our experiments. As we shall discuss later, there are however other useful targets besides RMSE and we can therefore state that the database has a wide application area.

Wind power forecasts generated from the database have been verified in Europe, Canada and China. The quality is similar to other systems in the same spatial resolution in the first forecast day, while 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 at present.

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.

3 Large Scale Integration of Wind Power

The targets set by most European countries is based on the assumption that a large fraction of new wind power production units will be located offshore. This will extend the geographical area covered by wind farms and reduce the periods of extremely low and high generation. Prices will become less volatile compared to the same amount of wind capacity positioned on land. The coupling of electricity markets through-out Europe and the interconnection of offshore wind power in the North Sea, Baltic sea and parts of the Mediterranean sea will be a natural step in the process of increasing the renewable energy penetration. Similarly a centralized management of a European SuperGrid connecting all European counties is likely to develop. This chapter will focus on the various benefits of large-scale wind integration and take advantage of the ensemble data mentioned in the previous chapter.

An integrated European market will require large-scale forecasting of wind power to ensure efficient trading of electricity. By using ensemble forecasts of the large-scale wind power generation over entire Europe, the total reliability of the energy system can be maintained and further increased. The use of ensemble forecasts in this context in fact will reduce price volatility, as excess wind generation from one region can balance missing wind generation in other regions. Additionally, the total wind generation can be estimated more accurately with the use of large forecast ensembles.

For this reason, large-scale forecasting gives smoother error patterns and lower costs than regional forecasting, because a weather forecast error can only count once at a particular point in time when the same forecast is used over the entire area. The purpose of this study is therefore to quantify the benefit and consequently the error reduction of an enlarged area and implicit exploitation of inter-connector capacity for balancing internal area errors.

3.1 European SuperGrid Study

The goal of the investigation is to describe a framework in which the economic and technical feasibility of present and future renewable power generation can be compared with objective measures. The framework is build on WEPROG's MSEPS ensemble forecasting system, a system built upon the so-called multi-scheme ensemble approach. We included most of the currently operational wind farms in 13 European countries (distributed in 1220 regions) with a total installed capacity of 75GW. With the MSEPS system we are able to simulate forecast errors, locate extreme events and predict reserve requirements for all countries. Particularly, the benefit or need of new power lines and inter-connectors can be estimated for a given capacity distribution. The framework is therefore highly relevant for evaluating the impact of a SuperGrid for renewable energy.

3.1.1 Framework of the SuperGrid Study and Setup of the Experiment

The MSEPS system differs from other wind power prediction systems in many ways. First of all, the core of the MSEPS is a 75 member ensemble weather prediction system, which simulates the physical uncertainty of the weather forecasts well already after forecast hour six. The uncertainty predicted by the MSEPS is therefore equally relevant for intra-day forecasts than for multi-day forecasting.

The MSEPS system handles single wind farms with interpolation to the real location and dispersed wind power by integration into the NWP model's grid. What is most suitable depends on the distribution of wind farms and how detailed wind farm location information was available. For the present study all wind power was considered as dispersed and accumulated in 1220 regions coinciding with model grid points. Confidential measurements could also not be used for the validation. Instead short-term forecasts were used. This means that the wind power generation was approximated by essentially one wind farm per model grid point and centred in the middle of the grid point in a 0.45° grid all over Europe. The summation in grid point space pre-serves the border of the countries and electrical grid in order to separate the energy production by grid.

The distribution of wind farms in Europe has been calculated by special software developed to process publicly available data. Some countries have significantly more detailed data than other. The "WindPower.net"-database [1] has been used in a number of countries. However, a considerable amount of adjustments were required to adjust the capacity to a level, which is consistent with the publicly known capacity. The power curves are generated from published aggregated data using a special technique to localise the generation to the *virtual* wind farms in each of the grid points. The localisation is iterative and distributes power generation onto the grid points according to the forecast. The process involves the following steps:

- Physical wind power prediction for 1220 grid points with all ensemble members
- Area integration of the power generation
- Computation of a time series of percentiles and ensemble average
- Computation of a bias correction as a function of predicted generation
- Adjustment of the percentiles and ensemble average with the bias correction
- Finding the so-called *true percentile value* corresponding to the actual generation to get the consistent local generation.

Finally, it is assumed that the same *true percentile* applies at any location and this assumption uniquely defines the *virtual* local wind farm generation. Thereafter follows a normal least square training. This training is done with the standard MSEPS approach, where direction dependent power curves are made for different stability classes. In this way, power curves for all regions were produced, where the aggregated wind power generation was published. In other regions the most relevant power curves are used. This is feasible, because Germany has as an example wind turbines located in nearly all types of terrain, where the effective power curves are an average estimate of more than 25000 turbines. Reduced availability is therefore built into the power curves.

3.1.1.1 Forecasts and Observations in the SuperGrid study

Wind power forecasts were produced for a period of 2 years from July 2008 to June 2010 for each of the 1220 *virtua*l wind farms (grid points). Accumulation is then done for each ensemble member over each region/country of interest.

To make the area integral of each ensemble member individually is fundamental, because the errors partially cancel each other out. This means that the larger the area, the lower the forecast error measured relatively to the total generation or installed capacity.

We then needed to define what should be considered the true generation, referred to later on as "estimated generation". This is done by using a weighted average of the ensemble members using a scheme, where the long-term verification scores are used to give weight to the individual ensemble members dependent on the weather situation. The most accurate prediction horizon from the NWP's perspective is the 6-11 hour horizon. This is after the noise caused by the new initial conditions in the weather forecast has levelled out. It allows us to define a complete time series, because new forecasts are generated every 6 hours.

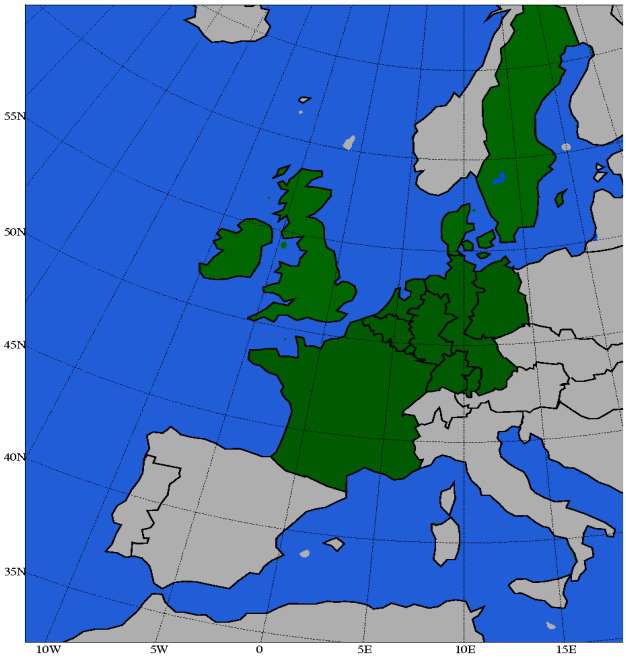

Figure 8: The selected 8 countries (green) for the SuperGrid study.

Although this time series is strictly speaking itself a forecast, it is possible to utilise the time series as independent measurements. Although the estimated error level for a large pool of wind power is not significantly different from the true error, it should be pointed out that the error is an estimated number.

The magnitude of the error is not of significance for the result, because we use the same technique for validation in each area of interest and the scope is to evaluate the relative error reduction of the error by looking at a large scale instead of on the regional level.
It should also be pointed out that we are using a two year period for validation. Thus, single events of extreme error of the short-term forecast can be regarded as a missing noise component not present neither in the forecasts nor in the estimated generation. Thus, the results shall be valid as long as we restrict the comparison to a relative comparison between areas of different sizes for the two year data set.

3.1.2 Test period and Area selection

The primary purpose of the SuperGrid is to establish a large market to facilitate sale of intermittent energy. In this SuperGrid study we focus on the various countries' energy generation pattern and on how the countries would interface with each other when connected in a SuperGrid.
The selection of countries was made on the basis of several factors such as

- weather pattern
- wind power potential
- geographical extent
- inter-connector plans

The purpose of the selection is to create one coherent area, which often experiences the same type of weather, but slightly phase shifted. With this strategy, less fast acting and expensive capacity (fossil fuel) will be required and parts of the wind power can be balanced over inter-connectors. It is suggested that major parts of the wind generation peaks should always be exported as the peaks always move from one area to the next. The export is maximised to the region, where the wind extreme is moving to the next area in order to reduce the peak and instead rather prolonging the peak. Using this strategy, the flow reverts as the extreme enters the new area. All regions will in this way experience smoother generation changes due to wind.

The approach in fact cuts off the top of the generation peak in the small countries and increases the efficiency of the entire energy generation. One can imagine this process as a weather extreme moving through the British Channel to the Netherlands. Instead of ramping fossil fuel down heavily within the Netherlands, part of the wind energy is first exported to Germany and as the extreme arrives in Germany, part of the energy is send back to the Netherlands. The result is a softer generation profile and less disturbance from wind in both countries. Reduced extreme generation eliminates the use of the very most expensive balancing power units and thereby reduces the cost of balancing ramps.

An objective measure for how well areas fit to each other can mathematically be expressed in terms of correlation between the individual generation pattern. We selected only countries with a correlation above 0.4 to at least one neighbour country for our study (see Table 5).

country	ec	sg	ie	de	dk	at	be	es	fi	fr	it	nl	no	se	uk
ec	1	0.94	0.29	0.88	0.67	0.23	0.72	0.45	0.21	0.61	0.24	0.79	0.22	0.51	0.66
sg	0.94	1	0.32	0.95	0.74	0.14	0.77	0.13	0.19	0.58	0.02	0.86	0.17	0.52	0.72
ie	0.29	0.32	1	0.17	0.21	0	0.28	0.04	0.09	0.2	-0.06	0.29	0.15	0.17	0.42
de	0.88	0.95	0.17	1	0.68	0.2	0.63	0.1	0.14	0.45	0.01	0.74	0.09	0.45	0.49
dk	0.67	0.74	0.21	0.68	1	0.03	0.34	0.06	0.21	0.16	-0.07	0.51	0.2	0.68	0.53
at	0.23	0.14	0.00	0.2	0.03	1	0.04	0.19	0.02	0.1	0.39	0.03	0.04	0.06	-0.04
be	0.72	0.77	0.28	0.63	0.34	0.04	1	0.09	0.1	0.82	0.05	0.86	0.08	0.25	0.59
es	0.45	0.13	0.04	0.1	0.06	0.19	0.09	1	0.09	0.24	0.39	0.06	0.13	0.12	0.08
fi	0.21	0.19	0.09	0.14	0.21	0.02	0.1	0.09	1	0.12	0.04	0.13	0.46	0.56	0.14
fr	0.61	0.58	0.2	0.45	0.16	0.1	0.82	0.24	0.12	1	0.21	0.6	0.14	0.17	0.4
it	0.24	0.02	-0.06	0.01	-0.07	0.39	0.05	0.39	0.04	0.21	1	0.01	0.12	0.02	-0.03
nl	0.79	0.86	0.29	0.74	0.51	0.03	0.86	0.06	0.13	0.6	0.01	1	0.11	0.34	0.74
no	0.22	0.17	0.15	0.09	0.20	0.04	0.08	0.13	0.46	0.14	0.12	0.11	1	0.47	0.17
se	0.51	0.52	0.17	0.45	0.68	0.06	0.25	0.12	0.56	0.17	0.02	0.34	0.47	1	0.34
uk	0.66	0.72	0.42	0.49	0.53	-0.04	0.59	0.08	0.14	0.4	-0.03	0.74	0.17	0.34	1

Table 5: Results of the weather correlations of the original 13 countries. 8 countries with a correlation >0.4 were selected for the study. Here, "sg" is the SuperGrid and average of the 8 selected countries and "ec" of all countries.

A lower correlation effectively means that the countries can exchange positive and negative imbalances in wind power too seldom. However, export of energy will be maximized at low correlation to the benefit of the wind generators on market terms.

For the sales process a lower correlation is clearly a benefit in order to avoid simultaneous generation. However, balancing and fitness to the demand are also key objectives for the integration of wind.

The correlations were computed on the basis of a 2-year time series of generation for each country. The best correlating countries were: Ireland, the United Kingdom, the Netherlands, Belgium, France, Germany, Denmark and Sweden with a combined installed capacity in July 2010 of 44GW.

3.1.3 Results of the SuperGrid study

Figure 9 shows the frequency distribution of the 8 countries and the SuperGrid. The first apparent result is that there exist little hours, where the generation exceeds 60%. Such conditions can in practise only occur during a few hours of storm events. In comparison, all individual countries have peak generation above 80% except France and Germany, which has only few hours above 70% of peak.

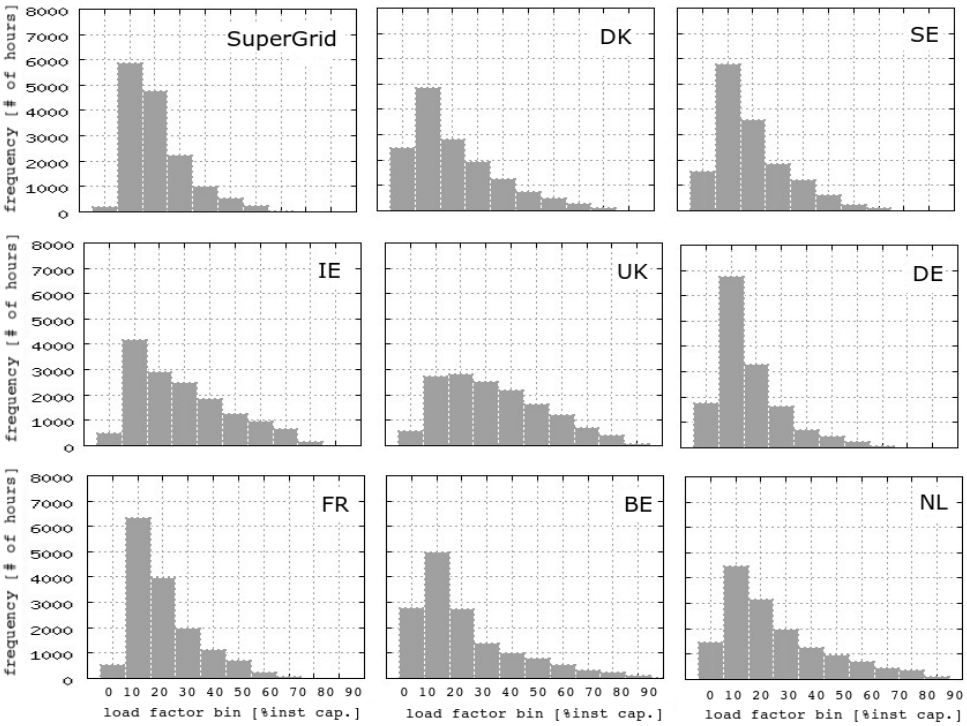

Figure 9: Frequency distribution of the 8 countries and for the SuperGrid.

It should be noted that neither of the two years had significant storm events and that has impact on the frequency of high wind events.

The statistical results for the day-ahead horizon is presented by country and for the entire SuperGrid in Table 6. The most interesting result is the relative difference between AVR and the SuperGrid. It can be seen that the SuperGrid is nearly 30% less than AVR. Here, AVR is the capacity weighted error for the 8 countries assuming that no forecast errors are exchanged across inter-connectors between countries. SuperGrid is the result of exchanging errors over the inter-connectors assuming infinite transmission capacity.

Another interesting result is that the error in the Netherlands is reduced by more than 50% in the SuperGrid. This is so, because the Netherlands is lying geographically centralized and can easily exchange errors in all directions. This possibility will even be expanded as the North Sea offshore capacity grows. In fact, the result for the Netherlands shows the core of the approach and why high correlations between the countries are required to make the balancing feasible.

The gain from the SuperGrid in percent is in general highest, where there is wind power on both sides of a common border. The results indicate that a consistent and intelligent handling of the generation near borders will reduce the effective error. Increased offshore wind power will over time also provide a similar smoothing on the coastal side.

Country	BIAS [% inst. cap]	MAE [% inst. cap]	RMSE [% inst. cap]	STDV [% inst. cap]	CAP [MW]	RMSE [MW]	Gain [MW]
BE	-0.09	4.28	6.8	6.8	642	44	22
DE	0.37	2.54	4.08	4.06	25500	1040	189
DK	-0.21	3.79	5.74	5.74	3200	184	77
FR	-0.06	2.65	4.02	4.02	4709	189	32
IE	-0.36	4.06	5.83	5.82	1412	82	35
NL	-0.09	4.48	6.86	6.86	2775	190	98
SE	0.61	2.76	4.12	4.07	1537	63	12
UK	-0.49	4.25	5.97	5.95	5089	304	134
AVR	0.14	3.03	4.67	4.66	44864	2097	598
SuperGrid	0.14	2.16	3.34	3.34	44864	1498	-
ratio	1.0	0.7	0.7	0.7	-	-	-
saving	-2.8	28.8	28.5	28.3	-	-	-

Table 6: Statistical Results of the 2-year SuperGrid simulation for 8 European countries, their average (AVR) and the SuperGrid.

France has several high wind power concentration areas along borders. In the SuperGrid, they are suitable for imbalance exchange with Belgium and Germany. Increased interconnection is therefore crucial for France for nuclear energy export and for balancing wind power ramps.

The underlying data material has shown that Germany gains internally approximately 30% in terms of RMSE error by using the load balancing Super Grid-principle (in German context also called "Horizontaler Belastungsausgleich", or "HOBA") across the four TSO zones and it is therefore interesting to note that another 20% of the error could be exchanged with other countries. So, even though Germany is very dominant with 56% of the total capacity of this study area, it is still possible to reduce the forecast error considerably by using interconnections with much smaller countries and less capacity. This is so, even though Poland is not among the 8 selected countries. A closer study on Germany has indicated high uncertainty on Germany's eastern border. Thus, the 20% gain for Germany is a conservative estimate of what Germany could gain from the SuperGrid.

The overall average result is that 600-700MW of balancing power can be saved by enlarging the grid, where Germany alone saves 180MW. Until year 2011 there was most likely no attempt to exchange any of these errors except between Norway, Sweden and Denmark. Instead, reserve capacity is bought on the market. There is no central management system telling that wind is out of balance in one control area nor are there parties seeking partners for such imbalances. At the time of writing this report primary reserve is shared across borders, and 2 initiatives within the "international grid control corporation" IGCC on the Germany border to the Netherlands and Denmark are in a testing phase with secondary and tertiary reserve exchange (TenneT, 2012).

3.1.4 Conclusions from the SuperGrid Study and future Outlook

The presented statistical results show that enlarging the grid intelligently opens the possibility to save nearly 30% in balancing capacity, equivalent to 600-700MW for a combined capacity of 44GW of wind power.

The results indicate that further improvements could be achieved from offshore wind power in the North Sea, Baltic Sea and inclusion of Poland. All countries would have a gain and all power plant would be able to operate more efficient than at present.

Although a market coupling has a built-in capability to adjust the export/import of primary power, it is expected that the above mentioned 600MW saving is an additional benefit on top of the competition benefits of the market coupling. The results are extremely promising for the existing and planned offshore wind power, where more connections across the North Sea will open the possibility to get a much higher utilization of the North Sea wind power.

From these encouraging results, it is nevertheless tempting to think ahead and consider how such a SuperGrid can be brought into operation. Can the system work with the current market coupling, national TSO's, different national incentive schemes and different ways of recovering balance costs ?

One way forward would be to expand the German load balancing principle (HOBA) to the SuperGrid, but with centralized forecasting managed by a Market Operator (MO). The obligation of the MO would then be to sell all non-scheduled wind power on the basis of conditional meta-forecasts.

The conditional meta-forecasts would cover three scenarios:

- Full import => Lower uncertainty band active
- Not congested => Use standard forecast
- Full export => Upper uncertainty band active

This scheme would be grid secured with a N-1 control check on the inter-connectors to ensure that wind power will be distributed by the market system. The scheme would also ensure maximum competition on reserve internally, as well as externally in each control area.

The modification of the forecasts added by the MO cannot alter the flow direction, but the prices are likely to change whenever either pure import or export conditions occur. The import case is likely to increase the price and the export case will lower the local price, dependent on the prices in other price zones.

The MO could also run short-term forecasts to balance wind on the entire grid and deduct the cost of that from what is recovered from the sales on the day-ahead market.

The presented results encouraged to carry out a more in depth study to quantify how much of the error can be exchanged in practise with the current transmission capacity and the planned (offshore) future capacity. In that way, it could also be found out, where the bottlenecks are in wind power context and, where on the grid additional inter-connectors should be added.

3.2 The extended Grid Control Corporation Study (eGCC)

The experience from Denmark shows that the use of inter-connectors increases the competition and therefore lower the price of balancing wind power. Following the results from the SuperGrid study, we wanted to further investigate how such large-scale integration of wind power can help to reduce the total costs for balancing wind power and other Renewables by coupling the German and Danish area to an extended Grid Control Corporation (eGCC) after the German "Netzregelverbund" principle (Zolotarev et al, 2009, Bundesnetzagentur, 2010).

The primary goal has been to make use of inter-connectors for the exchange of imbalances and to quantify the benefits. We have used a 32GW wind power subsystem of Europe and our analysis showed that an average balancing power saving of 25% is achieved from a combination of balancing via inter-connectors, increasing the area and amount of wind turbines, consideration of forecast uncertainty and deployment of two types of balancing power.

The study differs from the SuperGrid study on the spatial scale. The benefit is that much better and independent estimates of the actual generation is published for the selected areas. Also less uncertainty on the installed capacity is achieved. The availability of estimated generation from the TSO's also opens the possibility to study the short-term forecast accuracy and intra-day trading aspects.

3.2.1 Introduction of the eGCC study

The goal of this study is to show how efficiency drops of wind power can be prevented and the balancing requirements significantly reduced independent of whether the type of responsible balancing party is a TSO or private party. In 2009 and 2010, the German regulator Bundesnetzagentur (BNA) introduced a grid control corporation for the TSO's to prevent counter balancing and thereby reduce balancing costs (Bundesnetzagentur, 2010, see also Online summary http://de.wikipedia.org/wiki/Netzregelverbund).
In 2011, the BNA has started a new consultation together with the German power exchange (EPEX), stakeholders from the industry, research institutions and associations to investigate, whether and how costs for balancing wind power and other Renewables can be lowered further (Bundesnetzagentur, 2011). One of the discussed suggestions is to no longer let the bulk of the trading of wind power and solar power be carried out by the German TSO's but instead by one or multiple third parties.
Additionally, the EPEX is suggesting the introduction of a "green" spot market, which could be introduced in a 2nd step auction, where the first step is the traditional energy auction and the second step is a green proof auction (Paulun, 2011).

In the following we have been investigating and designed an alternative trading mechanism that could be carried out on the same basis as this green energy trading in a second step scenario by the market operator. In our study, the balancing area is extended to a Danish-German grid but intended to be sliced between different forecast providers for maximum competition and minimal risk. Although the equations and balancing technique that will be introduced in this paper could be applied to smaller pools, we will show that the benefit of a large pool lies in the inertia and internal cancellation of errors. Therefore, all wind power is included, regardless of ownership and incentive scheme.

The weakness of the strategy is that confidential information on the currently active capacity would never reach the forecast and confidential production data would have to be shared wider than owners naturally prefer.

We also justify this strategy by claiming that only offshore wind farms in a windy January may on their own be competitive to such a trading approach in terms of balancing costs per produced MWh. Another issue is that the work-overhead per MW to manage many small pools is much higher than centralized balancing.
What is proposed in this study can be formulated as an attempt to start the balancing process 2 hours in advance and avoid excessive balancing through use of uncertainty forecasts with a minimal of counter balancing and maximized utilization of inter-connectors.

The balancing process of the wind power generation starts shortly after gate closure by using a combination of:

- a day-ahead summation of contracted wind power (DFC)
- a short-term forecast for the wind power pool (SFC)
- an aggregated Pool Forecast Uncertainty (PFU)

The traditional approach is to either run pure day-ahead spot market and not participate in the intra-day market. That is to use DFC only. The more efficient approach is to run a 24-hour trading policy and correct the spot market forecast with the short-term forecast, that is to use SFC always. We shall investigate the error and effective trading volume of these and develop an alternative strategy. From the 3 basic forecasts we will compute how much energy can be balanced by the inter-connectors, how much should be traded in the intra-day market and how much should be balanced by shared balancing power. This is done by a sign evaluation of the expected balance:

$$EB = SFC - DFC \quad (1)$$

and the absolute balance

$$AB = | SFC - DFC | - PFU \quad (2)$$

where (2) expresses, whether we trust the sign of equation (1). From (1) and (2) we can derive a decision table (see Table 7) for the forecast update process (FUP). The column “a,b,c” will be used to generated a FUP increment later.

CASE	EB	AB	FUP	a,b,c
1	<0	<0	DFC	0,0,0
2	>=0	>0	SFC-PFU	1,-1,1
3	<0	>0	SFC+PFU	1,1,1
4	>=0	<0	DFC	0,0,0

Table 7: Decision table for the forecast update process (FUP).

When EB < 0 we are in a risk state, because a large power plant can suddenly fail and wind power is charged for expensive balancing power. Thus, we prefer keeping the error small. For EB >=0 the risk of a very high price is low with the suggested correction. In real-time operation, intra-day trading is likely to be competitive, if the error remains the same over several hours (i.e. start-up once). Thus, for EB < 0 we can often choose SFC and benefit, if we look ahead. When AB < 0 we have small DFC errors, which could be balanced by primary and secondary reserve.

The need of the PFU term has been identified by studying the intra-day sales and corresponding costs of the German EEG-based wind power (see 50Hertz, 2009 and EEG-KWK, 2010)). It was found that so far there is a significant loss on trading wind power on the intra-day market. This loss is due to the fact that it is more expensive to buy additional power than selling excess wind power in the intra-day market combined with an unbiased day-ahead forecast (DFC).

When studying this pattern it becomes obvious that an efficient trading scheme has to ensure to not re-trade the imbalance nor be charged balancing costs for the same MW multiple times. This is in fact what is achieved by acting in accordance with forecasted uncertainty (FUP) given in Table 7 and by limiting the trading by applying the PFU uncertainty band.

3.2.1.1 The Short-term Forecasting (SFC) with Uncertainty considerations

It is not very important for the approach how the short-term forecast is generated. The forecast could be a meta forecast put together from a number of deterministic forecasts for all generation or come from an ensemble forecasting system.

However, the forecast must cover the entire area and be based on consistent weather forecasts valid for the full area. In other words, the "meta" forecast has to be a sum of consistent forecasts for the entire pool. For our study we used the SFC forecast generated by the iEnKF algorithm described in Möhrlen et al (2009) and published online estimates of each TSO region. This meant that 6 regional measurements were used per time stamp to produce one forecast for the entire area. This does not provide an ideal SFC compared to using all online data. In such a system there would be more feed back across borders than what can be achieved with 6 regional online numbers.

3.2.1.2 The Pool Forecast Uncertainty (PFU)

The probabilistic forecast uncertainty (PFU) is independent of the day-ahead (DFC) and short-term forecast (SFC) in the sense that both forecasts can be generated by any method. The PFU needs to be calibrated with live historical data. This means, forecast data from a real-time system or historical data produced in real-time like conditions. Thus, if SFC is a meta forecast, then hourly forecast values need to be generated with the same meta forecast combination scheme. The first step to determine the PFU is to compute the average error (EA) of the uncorrected short-term forecast for the total pool over a year with live data:

$$EA=\int(SFC-OBS)\,dt \qquad (3)$$

A fraction of this error cannot be explained by weather uncertainty. This is represented as a constant base uncertainty. The remainder of the uncertainty varies with the weather. A number of tests that have been carried out have shown that the weather related uncertainty is best modelled via the ensemble spread in wind power (ESP) using its correlation with the forecast error (ERR). The PFU can then be expressed as a sum of the weather dependent uncertainty and a random uncertainty

$$PFU=C\cdot ESP(\,EA/ESPM\,)+(1.0-C)\cdot EA \qquad (4)$$

where ESPM is the time integral of ESP and C is the correlation of forecast error and ensemble spread (ESP,ERR). There is some freedom to choose ESP, but experience has shown that the distance between two percentiles centred around P50 works well. The percentile pair with the highest correlation (C value) is therefore chosen. An example could be to represent ESP as P75 – P25. The magnitude of C, ESPM and EA depend on each other, but they are constants for a given pool and a fixed look ahead time. All variability lies in the ESP term.

Generally, the C, EA and ESPM terms increase with forecast horizon, which means that the PFU will improve with forecast horizon for almost any pool. This is because the very short forecast horizon is dominated by non weather related issues and unpredictable weather on a very local scale and also short time scale. Much of this error would be captured by the short term forecast, if detailed reports from each individual unit would be available regarding availability and local weather conditions.

3.2.1.3 The Refinement Process

Spatial smoothing over regions is commonly used to lower the error and improve the value of wind power. What is special in our study is that we extend the spatial smoothing over 6 areas with very different levels of installed capacity. Spatial smoothing has often been applied, but it is often forgotten that there is an implicit assumption of no congestion.

This assumption is not valid when discussing large-scale wind integration using inter-connectors between different price areas, which are likely to be congested every so often. Thus, it is not sufficient to look at the apparently low forecast error of the big dispersed pool. We will therefore formulate an extension to the aggregated forecast approach, which will provide overview and control of the internal imbalances in a large area. In order to explain the value of the enlarged problem and also to illustrate the hidden error in the system better, we decompose generation into:

- Foreground generation (FG) as the mean of all areas
- Background generation (BG) for each specific area

The total generation is then:

$$TG=FG+\Sigma_{area}\ (BG) \qquad (5)$$

where the second term is zero for the total area. The BG term for a single region is:

$$BG=\Sigma_{region}\ (P_{turbine})- FG \qquad (6)$$

where we remove the mean generation from each turbine's generation $P_{turbine}$. This results in non-zero region BG summation, representing the excess or missing generation in each area. Thus, it is exactly that part of the generation that needs to be exchanged over the inter-connectors to maintain balance everywhere. A constant value of zero for the 2nd term of (5) expresses that the pool balances this term itself over the inter-connectors.

Thus, the BG integral is balanced by inter-connector flow of wind power from other regions and is not visible in the market. The BG generation occupies transmission capacity and would be valueless for the market unless the grid is congested. In the non-congested case trading or balancing with third parties is only relevant for the FG Generation. What is traded in the intra-day market is then the difference between the forecast update process (FUP) and the day-ahead forecast (DFC) according to Table 7. We can formulate the FG Correction Forecast (FGCF) with the help of the 3 constants from Table 7:

$$FGCF= a\cdot SFC + b\cdot PFU - c\cdot DFC \qquad (7)$$

The magnitude of FGCF changes smoothly compared to the BG elements on region basis of (6), because all 3 terms in (7) have a rather slow variation due to the inertia of the large pool. The smoothing of generation and therefore also the trading expresses exactly that we have conducted a refinement of the wind power. We have separated the valueless part from the valuable part and hidden the valueless part from the market by balancing over the inter-connectors.

3.2.1.4 The Verification Technique

In wind power forecast quality studies it is normally not taken into account that there is a price asymmetry, that a fraction of the small errors do not result in costs nor that some errors count double when forecasts have been applied to be traded multiple times. We therefore need to consider new error measures to reach a fair comparison on different trading and balancing strategies.

We have for this reason defined a quantity named "Total Balancing Volume" (TBV) as the sum of balancing in the intra-day, secondary and primary reserve. The TBV is similar to the MAE measure, but contains a discount and a penalty term. The discount is activated with 50% whenever the error is under 2%. That is, whenever wind power causes an error of less than 2% of the installed wind power capacity, then it is random, whether or not this is a cost or an income. The 2% limit is chosen because it corresponds to the average error of the demand forecast in MW. The penalty in TBV on the other hand lies in that the absolute value of the intra-day trade volume is added to the absolute value of the primary and secondary balance volume, because each volume represents a cost, also if the sign would cancel out. Thus, TBV is a verification parameter, which is capable of penalizing errors that count twice due to incorrect SFC and discount for insignificant errors where wind power is not imbalance causer.

We have tested 5 different balancing scenarios each with two different accuracy levels of the online estimates. The different scenarios are targeted to show with objective criteria how much intra-day trading should be activated to bring the pool in near balance. The scenarios differ essentially in their trust regarding the correctness of the online estimate and the accuracy of the short-term forecast (SFC), which is done with an objective uncertainty estimate. They can either be fully trusted, partially trusted or ignored. The purpose of this type of validation is to test how much double trading a given methodology leads to.

3.2.2 The eGCC area and experiment description

The forecast update process (FUP) has been applied over one year (07/2010 – 06/2011). This was possible by using public online estimates of the total generation in each of the 6 zones (2 in Denmark and 4 in Germany) with a combined system capacity of 32GW distributed over an area of 1200x600km

(see Figure 10). The raw online data covers close to 5GW of installed capacity in a rather inhomogeneous distribution. It is only the published estimated generation for each of the 6 zones, which is used for all computations.

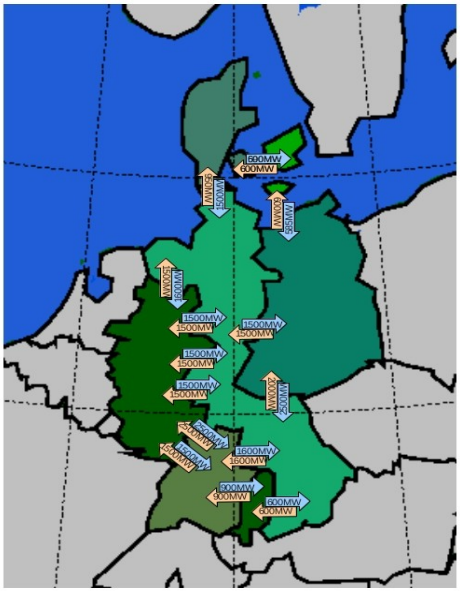

Figure 10: Area and inter-connectors between the 6 TSO areas.

It is assumed that the difference between actual and estimated values varies rather random, fast and unstructured when 6 different results are combined into the one final measure, which is used for verification.

For the one year study (07/2010-06/2011) we used the following datasets and assumptions in our simulation:

1. The capacity of the forecast was kept constant over the year and set as per January 2011. The only exception is that the offshore wind farm Hornsrev2 was included at 1st October 2010.
2. Published measurements for the 6 regions were scaled to January 2011 capacities with some approximations on Denmark West.
3. 75 member weather forecasts from WEPROG's MSEPS low resolution system were used to generate the DFC power forecast in each sub area. This step also produces the so-called raw forecast input (rSFC) which is input to the iEnKF forecast
4. WEPROG's iEnKF short-term forecast algorithm [8] was used to generate 2 hour short-term forecasts (SFC)
5. The pool forecast uncertainty (PFU) was computed for each hour

3.2.3 Results of the eGCC study

Table 8 shows the errors in different regions and for different forecast types as well as different traditional statistical parameters. It was found that there was an improvement in RMSE for all regions on the 2H forecast horizon. For total Germany, the error improved 0.34% of installed capacity in this specific test compared to the scenario, where Denmark is not present. However, the gain in accuracy is very different from region to region.

Area	**DFC** [%inst.cap]	**rSFC** [%inst.cap]	**SFC** [%inst.cap]	**Persistence** [%inst.cap]	**SFC + PFU** [%inst.cap]	**Hidden Error (BG)** [%inst.cap]
DK1	7.98	5.74	5.39	5.35	-	7.19
DK2	10.75	7.87	7.73	7.23	-	9.59
DE_50H	6.51	4.89	3.91	4.32	-	3.57
DE_TTG	5.97	4.49	3.69	4.00	-	3.13
DE_AMP	5.54	4.33	3.98	4.35	-	4.77
DE_ENBW	5.14	4.60	4.67	3.81	-	6.29
DE	4.96	3.54	2.53	3.28	-	0.95
DK+DE	4.79	3.39	2.29	3.13	3.69	0
Imbalance Reduction in [MW]	704	622	787			

Table 8: RMSE errors in different regions and for different forecast types. all values are in % of installed capacity. the MW saving is scaled to a installed capacity of 43GW.

In particular the SFC forecast improvement is very high compared to persistence on the total pool. Although the SFC can easily outperform persistence, we noted that the uncertainty is so high that the SFC+PFU cannot match persistence in raw error measures for DK-DE. Table 9 shows the total volume (TBV) measured for a number of trading scenarios representing different options. The effective TBV shows the double trading, which is expected to occur with and without PFU consideration.

Figure 11 shows the frequency distribution of errors for the various forecasts. This shows that DFC has some large errors, which would mean high costs for secondary reserve. The forecasts using PFU are hybrids between SFC and DFC in the frequency distribution.

Forecast type	Intra-day Trading Vol	% of time with small errors	Case 1 (0.5% RMSE) DFC Volume	Case 1 (0.5% RMSE) Total Volume	Case 2 (0.7% RMSE) DFC Volume	Case 2 (0.7% RMSE) Total Volume	Effect. Trading Volume	Double Trading
Case	1+2	100	1	1	2	2	1+2	1+2
Upscaled Obs	0.00	100	0.77	0.77	0.50	0.50	0.00	0.00
SFC	2.56	91	1.85	4.41	1.73	4.29	3.51	
SFC+PFU	0.61	77	2.84	3.45	2.74	3.35	2.69	0.03
DFC	0.00	71	3.41	3.41	3.31	3.31	2.70	0.00
rSFC	2.20	81	2.60	4.80	2.47	4.67	3.99	1.30
rSFC+PFU	0.22	73	3.21	3.43	3.11	3.33	2.70	

Table 9: Effective balancing volume and amount of double trading required for the various forecasts as average of 2 cases, where we assume an upscaling error of 0.5% of inst. capacity in case 1 and of 0.7% of inst. cap. in case 2. All numbers are percentages of the installed capacity.

In order to compare our results with the 600MW improvement estimated from 8 EU countries with the Netherlands as centre of the grid (Möhrlen et al, 2010), we have scaled the improvements achieved in this study from 32GW to 43GW and obtained a permanent saving of 704MW with DFC, 622MW using intra-day forecast without online data (rSFC) and 787 MW with SFC (see Table 8).

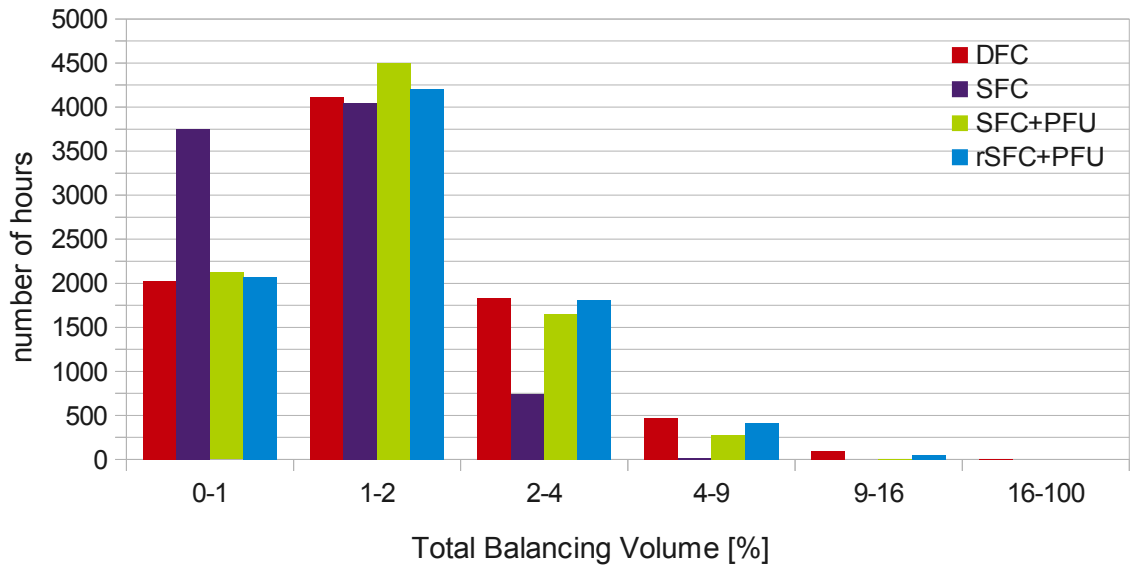

Figure 11: Frequency distribution of the Total Balancing Volume (TBV) for the different forecast types.

The region results of the short-term forecast SFC indicate that there are many hours, where the sign of the error is random and that the amplitude of the error must be significant. Another way to express this is that a considerable part of the error lies in the online estimation, otherwise SFC could not show the highest improvement of all measures from the enlarged problem. This gives a partial cancellation of error, when multiple regions are added together.

The last column in Table 8 shows explicitly the hidden error (BG) of the day-ahead error (DFC), which is less than the total error for each region, because part of the error lies in the mean area error (FG).

Other important results to note are the correlation of the 2 hour forecast error with the ensemble spread. We have two relevant correlations. One is the DFC's ESP correlation of 0.43 and the other is rSFC's ESP correlation of 0.53.

3.2.4 Qualitative analysis of Pool Forecast Uncertainty (PFU)

A 2-hour RMSE error of 2.29% of inst. capacity sounds like a high error, but the persistence forecast error is 36% higher (3.13%). This is the most relevant comparison for this horizon. The correlation between ensemble spread and forecast error gives insight of the source of the error, which is otherwise very difficult to access, because the frequency distribution indicates that there exist very few large errors. The 0-2% error gives a RMSE of 1.07% and represents 92% of all hours. Errors above 4% take place only 0.6% of the year.

The predictability of a high error on the day-ahead horizon is a useful parameter, because a major fraction of the secondary reserve is allocated day-ahead. For this reason, it is worthwhile to note that the short-term forecast (SFC) error has a 0.43 correlation to the ensemble spread and is available before the spot market closure. This means that the planing of the amount of reserve can be predicted with a reasonable good approximation on the day-ahead horizon, also before the spot market gate closure.

This is a major saving of costs, when the amount of wind power increases to a level, where the secondary reserve could be fully utilised by wind power, but very rarely is (maybe only 0.6% of time given the fact that SFC error is less than 4% 99.4% of the hours).

The correlation numbers and forecast error ranges achieved in this test can most likely be improved, as they were influenced by:

1. inconsistency in capacity and MW
2. errors in the measurement handling
3. insufficient accuracy of online estimation approach
4. a moderate resolution of the MSEPS ensemble data
5. a better PFU predictor than the selected percentiles may exist

An operational setup would not need to be limited by using only 6 online estimates, but could use all available measurements. This would most likely increase the correlation by 0.1-0.2, which would mean a day-ahead correlation of more than 0.5 for allocation of secondary reserve.

The tested system is influenced by temporary high errors in the individual 6 areas. Those errors are most likely the most important to solve as they are responsible for part of the double trading in the intra-day market. In the tested system, the pool forecast uncertainty (PFU) is probably not only predicting the forecast uncertainty, but also some uncertainty on the measurement side, because the online estimation process is more uncertain, whenever the forecast process is most uncertain. This is when there are strong horizontal gradients in the wind speed in regions of strong gradients in the installed capacity.

These gradients are what generates most uncertainty on both the forecast and online estimation process and are the reason why the PFU works well regardless of how the SFC is generated. As mentioned above, the uncertainty does not depend on the type of SFC. The uncertainty of a given forecast is a function of the PFU at initial time and PFU at 2 hours ahead. There can be an increase or reduction in uncertainty depending on how the ensemble spread changes over the 2 hours relative to the starting point.
This difficulty applies to SFC regardless of how it is generated. It is the uncertainty in the physical conditions of the weather and the installed capacity regardless of which forecast technology is being used.

The forecast uncertainty measure is a result of an evaluation of many ensemble forecasts on the complete generation and there is no disturbance from few recent measurements with random errors. The ensemble forecasts are not influenced by inherent errors in the SFC input and the difference between the ensemble forecasts describes how difficult it is to forecast the present moment.

The double trading is also visible in terms of RMSE error. Although the DFC RMSE error is 2.5% (4.79%-2.29%) higher than the SFC RMSE, we have found a squared average difference between DFC and SFC of 3.79%. This is equivalent to an effective RMSE of 6.08% (2.29%+3.79%) in terms of physical balancing. Thus, the DFC+SFC has increased the DFC error with 1.29% (6.08%-4.79%). The result confirms the double trading is a significant factor. The lesson is the 3.79% difference, which is critical and can fool the trader using short term forecasts, thus trading on the basis of short-term forecasts is likely to result in a loss unless the intraday prices are favourable.

3.2.4.1 Cost profile comparison

We showed in the previous sections that a simple balance volume consideration would suggest that uncertainty corrected short-term forecasting does not pay off, but a cost analysis suggests the opposite. The trading has to be seen as a pre-balancing process, which has a MWh price and where the rule applies that the more MWh there are traded, the higher the loss per MW of error. The cost in the intra-day market is expected to follow the secondary reserve costs for the same amount of MW or is moderately cheaper. Otherwise there would not be made contracts. Thus, an offer of 1GW should cost approximately the same in the intra-day and secondary reserve market.

The scheduled generators know that the wind generator will pay that price. This is because another party could have a correlated error and cause the balancing costs to go even higher. Intra-day trading can therefore be regarded as a kind of hedging of costs of large errors. The willingness of an intra-day trader to accept a given price or not depends on the expectation to the balancing costs. This expectation can in theory only be based on uncertainty forecasts on “demand-intermittent generation”.

Thus, in a competitive market, intraday prices are expected to lie at the same level as the pre-allocated reserve price for the same amount of MW with some noise due to false expectations to the need of secondary reserve. The wind generator would have no reason to not pay according to the cost profile of the secondary reserve, if the sign of the balance is certain, which is the case for a large error over 2% of the installed capacity.

We can assume reserve prices go with a power of 2, because large amounts of secondary reserve is less often activated than small amounts and the marginal costs are higher for doing any work. With a simple 50% split of the error between intra-day and secondary reserve, it is therefore possible to balance half the amount twice and each at 25% of the cost and thereby save 50% compared to using an either-or-solution for balancing.

Reality is however more complicated, because we do not know the error, so the shared balancing is so to speak only idealized theory.

3.2.4.2 Quantitative analysis of PFU

The total effective balancing volume (TBV) is very similar for the different scenarios (see Table 9). It is obvious that the two scenarios, where the short-term forecasts are fully trusted, result in significantly more TBV. This is a result of trading power in the intra-day market first and thereafter balancing a portion of the same energy with secondary reserve.
That is, the poorer the short-term forecast, the more double trading is carried out. However, because the error can be predicted with a reasonable success rate we achieve a considerable saving in TBV by using the pool forecast uncertainty (PFU) correction defensively. For two similar TBV values, it is the frequency distribution of the large errors that make the difference.
The SFC with PFU has the least large errors and seems therefore to be the best combination in terms of volume. The weak point of this analysis is that the frequency distribution of the SFC for large errors seems much better than with PFU included. The final choice of strategy therefore depends on the competitiveness of the involved markets.

3.2.4.3 Impact of Full Congestion

Full import or export conditions caused by extreme intermittent generation relatively to the demand are normally predictable. An optimal day-ahead forecast should be tuned to keep maximum competition on reserve. A full export case suggests to choose the P90 or Maximum forecast, while full import should suggest selection of the P10 or Minimum forecast. In this way reserve prices for the expected reserve requirement will remain competitive.

3.2.5 Summary and findings of the extended grid control corporation study

We have studied the impacts of an extended grid control corporation between Denmark and Germany with a capacity of 32GW wind power (see also Jørgensen et al, 2011). A previous study (Section 3.1 and Möhrlen et al, 2010) indicated a sustained reduced balancing requirement of 600MW by pooling 43GW in 8 countries, which we are now able to confirm on different forecast horizons down to 2 hours with a fully independent data set.

In fact, we have found a significant value increase of wind power and increased predictability with more constant generation. In addition, we found a methodology, where uncertainty forecasts are an extremely useful tool, even on the 2 hour forecast horizon and in a system with a considerable inherent uncertainty due to the geographical scale of the area and a rather inhomogeneous concentration of generation capacity.

The core approach is the application of a multi-scheme ensemble (MSEPS) and an iEnKF short-term forecast (Möhrlen et al 2009) applied using public data from 6 regions. This approach in combination with an objective uncertainty forecast provides an intra-day trading scheme suitable for automation. The scheme assumes that all wind power is going in as one pool at a gate closure of 2 hours, regardless of how the power is sold and by whom. In that way, the suggested trading scheme could eliminate all of the 0.79% MAE double trading, which would be the effect of not considering uncertainty. This was achieved by not correcting the day-ahead forecast more than to the boundary of a uncertainty band around the short-term forecast. The frequency distribution of the error did also confirm that the large errors were reduced to smaller errors.

Our analysis also showed that the saving of the enlarged area is achieved from a combination of balancing via interconnectors, an increased area and more wind turbines, consideration of forecast uncertainty and deployment of two types of balancing power. Altogether, we found a reduction of 25% in the day-ahead forecasts (DFC) and 44% in the short-term forecasts (SFC) of the balancing power for a 32GW pool of wind power distributed over Denmark and Germany.

This can be regarded a technology shift that enables the handling of significantly more wind power, because such a system will be safer to operate and will ensure that wind power rather gains than looses market value at higher penetration levels.

4 Trading strategies and Tools for variable Energies using Ensemble Forecasts

The experience has shown that it takes a dedicated effort to bring probabilities computed on the basis of ensemble forecasts into the energy markets beyond the level of subjective evaluation of risks. Nevertheless one of the targets in the project has been to develop a dynamic method, which is simple to use and yet adaptive to the varying predictability of renewable energy. In other words, it is suggested that by using predicted uncertainty and present market prices for reserve, it is possible to change the role of wind power in the current market structure towards a more pro-active role compared to the way wind power is handled today in most jurisdiction (Möhrlen et al, 2012).

4.1 A new approach to Estimate Consistent Local Wind Generation

The studies mentioned in the previous section require a new approach for estimation of representative power by region without knowing the detailed production of the 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 in most cases only aggregated production numbers are published.

A quality check will therefore 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 existing training and forecast methodologies using least square fitted power curves. In that sense, the purpose of the new approach is more focusing on the estimation of 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 "unknown" 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 on the local scale. Instead seriously misleading information, which is still likely to be inherent in the aggregated data is being spread out as a smaller error on each unit.

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 when considered 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 in Denmark and Germany. What is new is the ability to estimate a consistent local power generation, which results in the aggregated generation being consistent with the local generation and a local responsive power curve.

The primary goal with this methodology 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.

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 and planned on a European level (see also Möhrlen et al, 2012).

4.1.1 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. One can estimate that the accuracy of the power curve fitting increases almost proportional with the error reduction from local scale to total area. A typical local scale day-ahead error is 9-10%, while the error for the aggregated generation is at 3-4%. The result is a more realistic local power curve.

This type of power curve training also takes places directly in the weather model grid for dispersed wind power or at individual wind farms. The methodology is opposite of the most commonly used upscaling algorithm. This is because total generation is used to estimate local generation while the traditional approach is to estimate total generation from selected samples with proper weight coefficients. The fact that estimated local measurements fit the weather forecasts better and the power curves therefore become more responsive solves 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 fitting optimisation method. Previous approaches implied an either or and that had impact on either the accuracy or the validity of the ensemble spread. The typical pattern was that least square damping took place at every wind farm with systematic under prediction in strong wind. After application of least square optimization on every wind farm and aggregation, the final result is of course under dispersive.

Thus, under-dispersiveness for aggregated generation has most likely been an artefact of the statistical handling of the problem rather than the ensemble data themselves.

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 significantly.
With centralized European publication of all wind farms, it is also 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.

So far, two challenges have been experienced with the approach. One challenge occurs when a large wind farm has operational problems over long time. The methodology distributes these operational problems evenly in the power curves of all wind farms. The phenomena is not relevant for dispersed wind power, but it is so in growing markets where the newest wind farm size is often bigger than all the existing and possibly not operating reliable. A second challenge occurs when two conditions apply: The terrain characteristics of the model differ significantly from reality and the model wind speeds are biased by the terrain type.

The typical problem is that there is no or little bias in non-complex terrain, but a negative wind speed bias in complex terrain, because wind farms are located in terrain with a higher wind resource than the average of the region. This issue can result in that wind farms in non-complex regions get a too steep power curve and vice versa for wind farms in regions with complex terrain. The error in the power curve is invisible on the end result, if the pressure gradient is homogeneous, but that is not always the case. One can say that there is a latent error, which is mostly visible at variable weather. Both issues can in general be circumvented, if detailed measurement data, e.g. on region basis, is available. Otherwise it is difficult to prevent some degradation of the predictability of the region.

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 for the harmonisation of the European electricity grid and other large areas, as well as centralised forecasting and publication of relevant data, as probabilistic forecasts will be necessary to enhance transparency and competition.

4.2 Ensemble based Optimization Techniques

Nearly all decision making in business is associated with a cost or a loss. Many of these decisions are at the end of day weather related. If a decision is based on a deterministic forecast, then there is a risk of that the decision will be random, because the forecast used for the decision is likely to have a random error. The longer the forecast horizon, the more likely it is that a given decision turns out wrong, because the error grows with lead time and it is getting more unlikely that there is more than 50% chance of success. If a user of forecast detect that there is only 50% chance of success, then the user gives up and starts guessing instead, but guessing is not leading to any progress.

Continued economic growth occurs, because humans act more clever by means of new and enhanced tools and because it is believed that "efficiency makes money grow". The alternative is recession in the economy and unemployment. It is therefore our obligation to assure that what we present as an improvement also can contribute to economic growth.

It is not a sufficient condition to postulate that a given approach is better than another for a measure, which is not sufficiently hard linked to the economic value of the problem at hand. We need to conduct a proof in terms of resource usage and cost space. In fact it is the latter that counts in a market economy. Legislation should then ensure that economic optimisation implicitly imply that resource usage is optimised as well, if necessary via fees and taxes.

4.2.1 The Cost-Loss Analysis

Ensemble weather forecasts have become the established method to determine the uncertainty of the weather development since their development in the early 90's (e.g. Brankovic et al. 1990). The main reasons for developing ensembles has been to provide an objective value of the forecast uncertainty and also to warn better about extreme events. In fact, these two aspects are closely connected, because extreme weather traditionally has an implication on economic loss. The loss can be related to damages by not protecting something, which is too fragile to withstand extreme weather or by taking action that is inappropriate in extreme weather conditions. Alternatively, the loss could be caused by invoking a protection mechanism based on a false alarm.
The pioneer work in ensemble forecasting of the 90'ies (e.g. Palmer et al. 1993; Toth and Kalnay 1993; Molteni et al. 1996; Houtekamer and Mitchell, 1998, Stensrud et al. 2000) therefore plays a central role in the society on risk management and trading and has been a funding resource over many years for the development of ensembles forecasting methods.

Two decades after the introduction of the first operational ensemble forecasting systems from the European Center for Medium Range Forecasting (ECMWF) and the National Center for Environmental Prediction (NCEP) in 1993, it is still not common knowledge that there exist simple cost based algorithms on how to exploit probabilities (e.g. Murphy, 1976, Katz et al, 2006, Lee et al, 2007, Hagedorn et al, 2009). The cost-loss method is a classical example of such an algorithm, but has apparently not been applied much in the energy market, most likely because the economic sensitivity related to marginal costs of certain actions is considered commercially confidential. Part of the reason is also lack of short-range applicability of many of the ensembles generated world wide due to the approach used to generate them (see e.g. Palmer et al. 2005) .

The cost-loss method results in a critical probability even for a binary decision process. Because it is probability based and the decision is binary, there is a certain likelihood of failure. One or two failures in a row to begin with may cause that the end user looses confidence in the approach and gives up.

The application of the cost-loss method requires therefore repeated decisions and tolerance of failure in the decision. The economic value can in fact only be recovered by applying the cost-loss evaluation in many events. Nevertheless, if sufficiently many parties would spent efforts on evaluating their costs and losses of certain processes related to weather, then this would make a positive effect on the global economy, because what is not lost from wrong decisions can be invested again into improving the algorithms.

4.2.2 Growing System Complexity

Forecasting of expected wind energy (or solar energy) can be considered additionally complicated on top of a already complex meteorological problem. Wind power forecasting is complicated, because the wind resource of specific locations often differs from average conditions, there exist strong concentration of capacity in specific regions, sensitivity of data, a near cubic relationship between wind speed and power, operational issues at wind farms, some of these even weather related and last but not least restrictions on the transmission grid. Many of these factors are even unknown to both the forecast supplier and the user of the wind power forecast. Therefore, a combination of physical and statistical methods is in general superior to a pure physical translation of meteorological variables into power generation. The more volume and the more dispersed distribution, the lower the error of the aggregated generation, also because the assumptions behind the statistical method become gradually more valid with increasing volume.

The traditional approach to evaluate forecast quality is to compare simple statistical error measures such as RMSE or MAE on physical quantities. We have demonstrated via the TBV (total balancing volume) calculation in section 3.2.1.4 that for the combined day-ahead forecast and permanent intra-day update, the total balancing requirements increased compared to not using the intra-day forecast at all. Several of the verified strategies resulted in very similar TBV values despite significant differences in the frequency distribution of the error.

These results suggest in fact that the traditional measures of forecast quality have become insufficient to also reflect the value of the forecasts for the trader on the one hand and the total system on the other side. The obvious next step is therefore to conduct an economic analysis of forecasts on a real problem, which is the target for the following sections.

4.2.3 Need of Short Term Forecasting

Figure 12 shows a native example, which demonstrates that forecast uncertainty is variable and not always growing with forecast horizon. The so-called spaghetti plot of the 75 ensemble members from a Multi-Scheme Ensemble System (MSEPS) (e.g. Möhrlen and Jørgensen, 2006, Lang et al. 2006,) is in itself not useful for showing more than that variability and uncertainty is a fundamental challenge. The uncertainty remains high for the first 18 hours of the forecast and many forecasts ramp up while other ramp down.

It can be said that the weather forecasts alone have insufficient accuracy and that a better forecast can be generated with combined use of forecasts and measurements by giving weight to the forecasts according to their fit to the measurements over longer time.

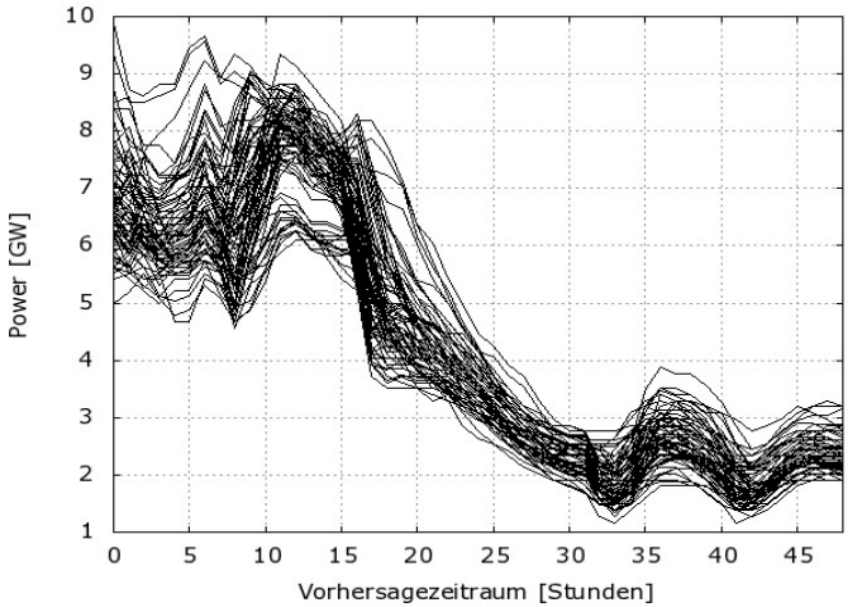

Figure 12: Example for the spread of the 75 physically based ensemble members of hte MSEPS for an arbitrarily chosen forecast horizon of 48 hours.

4.2.4 Introduction of the Incremental Ensemble Imbalance Correction scheme (IEIC)

Our goal is to use our knowledge and experience on ensemble forecasting to develop an enhanced trading strategy for managing wind power with focus on the day-ahead and intraday horizon. More precisely, we wanted to develop an algorithm that reduces the cost of the imbalance between the spot market and the actual generation. Thus, our target is not to alter the spot market bid, but to reduce the cost of the balancing process by more intelligent usage of forecasts for the trading practices at the intraday market. To achieve this, we used the basic forecasting principles described in Section 3.2 but extended the approach to cover the time from spot market gate closure to gate closure of the intra-day, especially because we have already demonstrated the approach's ability to minimize the trading volume (see section 3.2.3 and 3.2.4).

At this point we should clarify why we use this approach, which seems complicated compared to a simple dynamic pre-allocation of reserves according to the ensemble spread, which could be tendered daily and in that way maximum competition on daily markets could be exploited. Any centralized solution is complicated to get in place and one would also consider demand, because demand and wind generation uncertainty are correlated.

Our goal has therefore been to develop a balancing methodology for stakeholders using the market instead of centralized TSO managed balancing, but with a completely different solution than for a centralized system. The reason is that it will not be cost efficient to buy reserve for a portion of a pool, because there is a latent "cost free" value in sharing uncertainty with other parties, which should be utilized. If each party would allocate reserve, there would be over-allocation and the result would be too high costs.

The basic assumptions for the new approach are that:

- price volatility occurs when larger amounts are traded simultaneously
- reserve has been tendered and is competitive for smaller amounts
- the startup costs of reserve are essentially shared
- the marginal cost per energy unit is higher for balancing power than constant generation

- the strength of the intraday market lies in that it is running and can respond to new information quickly
- the intraday market is suitable to swapping contracts (deals at spot price)
- the intraday market is suitable for balancing with price limits
- the forecast error is considerably higher during the last part of the day than during the first half

These assumptions suggest an iterative trading scheme comprising a combination of contract swap attempts based on a optimal forecast and limited loss offers based on an uncertainty bounded forecast.
Such a scheme will minimize the loss by sharing a fraction of the imbalance with others and still keep the imbalance small also in events where the day-ahead forecast had a high error. The small increment method is optimized to not itself trigger price volatility.

With the proposed scheme it is in fact possible to use the spot market to get rid of additional uncertain generation that has been sold on the day-ahead market. This is because one can expect that there will be matching offers from scheduled generators at the opening of the intraday market. In this way the spot market can be used to hedge the price of uncertain wind and solar power generation above percentile 50. At low generation, competition considerations suggest to sell less than percentile 50, if the demand is also high. Hedging of prices is one aspect.

The other aspect is to keep maximum competition during full import/export conditions (see 3.2.4.3). Those conditions suggest to go to the extreme and stay at the lower/upper extreme.

It is now essential to start after opening of the intra-day market, because:

- the market volume is most likely highest and thereby most competitive
- at this time start-up of high inertia generation can still be scheduled, if the new forecast suggests much less non-scheduled generation

- there is an accuracy gain, because the weather forecast generated in the afternoon is based on many more recent aircraft data in the upstream direction in Europe
- the result of the spot auction is known and it is expected that some generators may at times wish to change the schedule to run more continuous

One can therefore expect that the afternoon trading will become more and more useful for all market participants over time, because one can argue that the market can be used as a system wide adjustment process for unexpected site effects of the spot market auction. An incremental forecast approach for the intraday based on a price profile seems therefore to be a solution which can increase the efficiency of the entire market considerably. This trading scheme will hereafter be referred to as the *Incremental Ensemble Imbalance Correction scheme* (IEIC). Figure 13 illustrates the principle of this methodology.

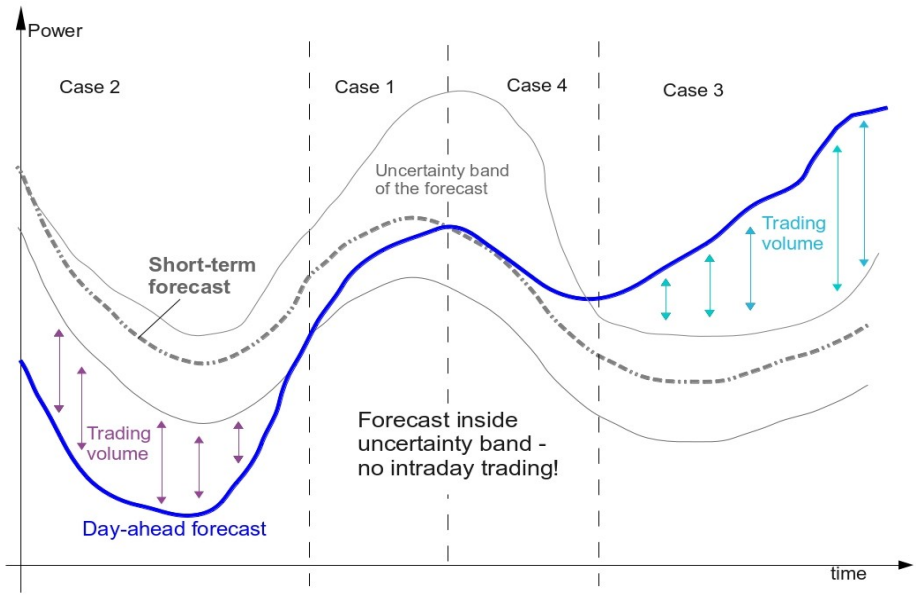

Figure 13: Trading principle of IEIC when the uncertainty band is used to determine the volume that is to be traded on the intra-day market. The figure shows the first correction to the day-ahead forecast. The dashed grey line is the new short-term forecast (NFC), the black line is the contracted day-ahead forecast (CFC) and the grey lines are the uncertainty forecast with the upper and lower limit valid for NFC.

Given the 3 constants a,b,c, from Table 7 we can formulate the correction forecast as a linear equation:

$$CF = a \cdot NFC + b \cdot PFU - c \cdot CFC \quad (8)$$

based on the terms described in equation (7) and Table 7 in Section 3.2 , where the coefficients a, b and c all have values of either -1, 0 and 1. The values can change from one trading interval to another. The CFC is the forecast that belongs to the currently contracted amount. This is rather a result of an auction than a forecast, even in the very first iteration. The NFC is a newer forecast, which should ideally replace the contracted forecast as a "swap". If this fails, an offer is made with the probability forecast uncertainty (PFU) that will be bound to a bid with limited loss. The next correction will then apply to what is contracted. In each iteration new values of a, b and c are computed with a new sign evaluation of equation 1 and 2 where the current contracted amount CFC replaces the day-ahead contracted amount DFC and the new forecast NFC replaces the "old" short-term forecast SFC.

In cases 1 and 4 on Figure 13, where the NFC lies inside the uncertainty band around CFC, the CFC forecast should not be corrected (all coefficients in equation (7) a,b,c = 0). This is because the difference of the forecasts lies within the uncertainty band bounds (+/- UP). In case 2 and 3 we correct the forecast to the closest uncertainty band. Hence, it is either attempted to sell (case 2) or buy (case 3) the determined difference in power between CFC and NFC-/+PFU. In case 2, where the NFC lies above CFC and outside of the uncertainty band, it is important to get in balance, since there is a risk of failure of a large power plant and hence the imbalance of the control area of the wind energy pool may be high and expensive reserve power is required.

In practice, the price for reserve, apart from the convenience to ascertain the best possible balance, is an incentive in itself for keeping the error as small as possible. The risk for a high balance cost is substantially less for case 3 in a future market where intra-day balancing for non-scheduled is becoming a standard.

The need of a limit on the volume to put into the intra-day market has been identified clearly by a detailed study[1,2].

4.2.5 Setup and Simulation of the IEIC Test case

Due to the fact that the Nordic and German market differ in their calculation of balance costs, it was found that we would have had to make too many unrealistic assumptions to conduct a cost analysis on the combined German/Danish area. A price model has been defined by the international grid control corporation (IGCC), but one year of historic prices were not yet available at the time of conducting this analysis.
Therefore, it was an either or choice between Denmark and Germany. The choice fell on Germany, because 4 different public online estimates of the wind generation is available here, whereas in Denmark there are only 2 estimates, one for each zone, which are at times dominated by offshore wind power, which has a different variability and forecast error pattern than the dispersed power on land. For demonstration of the IEIC approach, it was therefore more suitable to focus on Germany alone.

A simulation of a period of one year, from July 2010 to June 2011, has been set up in order to test the IEIC approach. The evaluation takes place on all Germany and not by control zone. Hereby power prices on the spot market, prices on the intra-day market as well as prices for control reserve have been calculated for the analysis.

The analysis comprised daily day-ahead forecasts with a forecast horizon of 48 hours, 6-hourly forecast updates with hourly short-term forecasts looking 8 hours ahead and hourly intra-day forecasts. The following formula 2*(8+forecast horizon)/6 provides a good estimate of the potential amount of hourly increments with the forecast horizon counted from the spot market gate closure.

[1] Transparenz der Vermarktungstätigkeiten gemäß § 2 AusglMechAV, http://www.eeg-kwk.net/cps/rde/xchg/eeg_kwk/hs.xsl/525.htm

[2] Regionenmodell des Stromtransports 2009, German TSOs. Online: http://www.50hertz-transmission.net/de/1388.htm

For the longest horizon there will be 26 increments, where new relevant information have increased the chance of successful balancing forecast errors.

For the purpose of illustration we compared a number of forecast scenarios ranging from the pure Day-ahead forecast (DFC) without use of intraday to the other extreme where the DFC is corrected every hour by a 2 hour forecast:

1. Day-ahead forecast without intra-day balance (DFC)
2. Day-ahead forecast with continuous hourly intra-day balance by means of a short-term forecast (SFC)
3. Day-ahead forecast with 6-hourly intra-day updates, which includes a raw short-term forecast with a forecast horizon of 13 hours without consideration of measurements (rSFC)
4. Same as (2), but with additional uncertainty forecast (SFC+UP)
5. Same as (3), but with additional uncertainty forecast (rSFC+UP)
6. A 36-hour forecast, generated from the 12UTC-weather forecast cycle (see section cost analysis of the forecast scenarios)

We refer to these forecasts from their forecast type number, because neither forecast horizon nor increment number will give a unique naming convention. In Figure 14 it is shown how forecast 2 and 4 are computed every hour. Forecast 5 is generated every 6 hours once a new ensemble is available, while forecast 6 is only generated once and covers the entire next day.

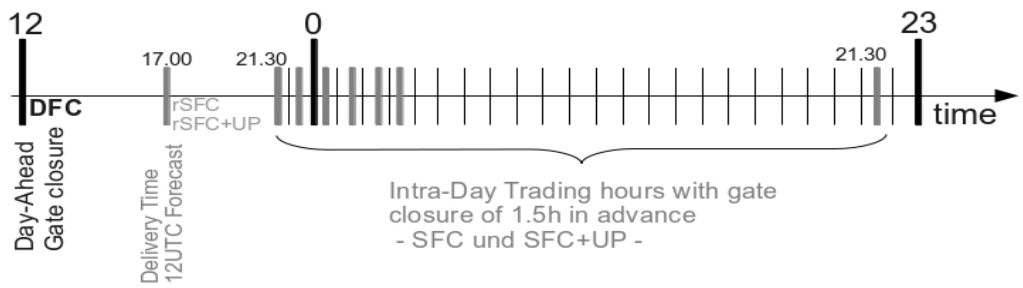

Figure 14: Schematic showing the trading times for different forecast scenarios

4.2.6 Analysis of the IEIC forecast scenarios

A statistical analysis in power has been carried out for the time span that was considered in this study. The analysis comprises a traditional statistical analysis in % of capacity, which is cost independent. A second part follows, where real prices for the same period have been used.

The purpose of this analysis is to show the frequency distributions of errors and how often a given forecast type actually will suggest an increment for trading. For the longest horizon there are up to 26 increments issued to the market. How many depends on the forecast error pattern and the market volume, which we neither simulate objective nor fair. Thus, the forecasts selected for closer analysis are selected samples from the long list of increments for illustration purposes.

This analysis indicates that some of the increments will only very rarely result in corrections. How often depends on the current and past volume in the market, but also how the predictability has changed since the spot market closure. Low market volume and constant predictability may not allow for any increments, whereas high market volume combined with increasing predictability may allow for many increments. Also, a very well centred spot market forecast may prevent generation of new increments.

The IEIC scheme therefore couples market and forecasting close to each other and it is very difficult to verify the impact of changes in the procedure in a real-time environment, because setup changes may interact rather complex with other market participants. Economic verification of development and tuning can only really be conducted in a controlled environment over a longer time, whereas a traditional statistical analysis is always possible. Nevertheless, the following analysis still reveals that even the interpretation of the physical error is non-trivial.

4.2.6.1 Traditional Statistical Analysis

In Table 10 and Table 11 the RMSE and the BIAS values are summarized for different forecast types and areas. It can be seen, that the forecast type no. 2 is the superior forecast measured in RMSE without consideration of the volume difference between DFC and SFC.

Forecast Number	1	3	2		4	5	6
Forecast type/area	Day-ahead forecast 00UTC	Raw short-term forecast	iEnKF short-term FC	Persistence (2 hours)	iEnKF short-term FC corrected	Raw short-term FC corrected	Day-ahead Forecast 12UTC
DE_50H	6.81	5.40	4.61	5.34	-	-	-
DE_TTG	5.39	4.14	3.42	3.96	-	-	-
DE_AMP	5.13	4.36	4.03	4.03	-	-	-
DE_ENBW	5.84	5.14	5.25	3.29	-	-	-
DE	4.69	3.55	2.74	3.11	3.65	4.26	4.27

Table 10: RMSE for different forecast types and areas. All values are % of installed capacity.

Forecast Number	1	3	2		4	5	6
Forecast type/area	Day-ahead forecast 00UTC	Raw short-term forecast	iEnKF short-term FC	Persistence (2 hours)	iEnKF short-term FC corrected	Raw short-term FC corrected	Day-ahead Forecast 12UTC
DE_VET	0.57	-0.79	-0.43	0	-	-	-
DE_EON	-0.12	-1.12	-0.87	0	-	-	-
DE_RWE	0.32	-0.94	-0.7	0.01	-	-	-
DE_ENBW	1.75	0.33	0.55	0	-	-	-
DE	0.28	-0.87	-0.63	0	-0.08	0.08	-0.29

Table 11: BIAS in different regions and for different forecast types. All values are % of installed capacity.

However forecast type 2 and 4 are measured on the 2-hour horizon and are not directly comparable with forecast type 6, which is generated in the afternoon around 7 hours before the day starts. Forecast type 1 appears much worse because of the longer horizon.

Figure 15 shows a frequency distribution of the intraday trading. For forecast type 4 and 5 (SFC+UP and rSFC+UP) nearly all hours in bin 0-1 are equivalent to that no trading has taken place, i.e. the DFC was considered good enough or the uncertainty of the short-term forecast was too high to justify any trading.

Forecast type 2 and 3 (SFC and rSFC) has many more hours with 1-9% corrections, which increases the required trading volume considerable.

Figure 16 shows the distribution of the required balancing volume for different forecast types. Obviously, for the SFC with continuous intra-day trading, a large fraction of the error accumulates in the lowest error bin.

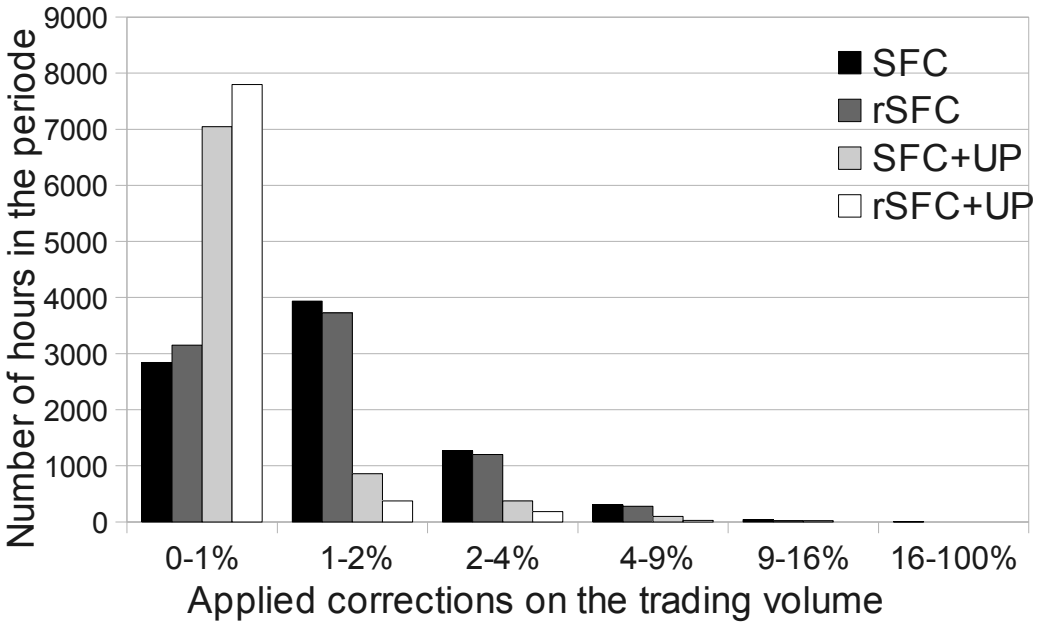

Figure 15: Number of hours for the individual marketing strategies, where the day-ahead trading volume has been corrected.

The high trading volume of forecast type 2 and 3 need to be considered separately, because all this trading must result in 50% counter corrections (>1600 hours), because the total system uncertainty certainly exceeds 1% of wind power capacity. As a consequence, 50% of the small corrections lead to an unnecessary loss, if traded on the market.

For the purpose of creating a penalty scheme for counter regulation, publicly available upscaling data were used. Therefore, the upscaling error had to be simulated in order to include the error of the upscaling relative to the actual measurements. Based on experience, a random error of 0.50% and 0.76% of the installed capacity has therefore been added to the upscaling values in two different scenarios.

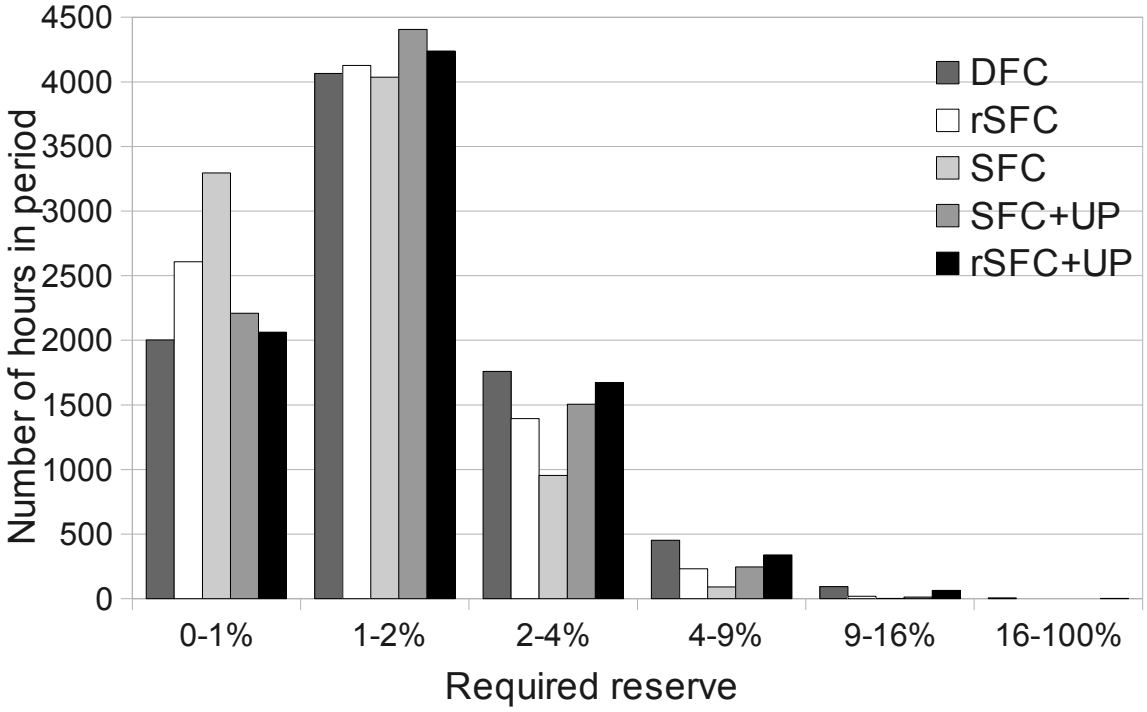

Figure 16: Frequency distribution of the required control reserve for the individual trading strategies

Table 12 provides information in terms of mean absolute error for the different forecast types on total volume of intraday trading and automatic balancing via reserve. We use the wording "double trading" whenever automatic balancing counter corrects the intra-day order made on the basis of this forecast. The different forecast types all suggest different intraday orders.

The two scenarios (1+2) differ in their simulated upscaling error and the final result is averaged. A "discount" of 50% of the small errors has been added to the last column named "effective traded volume". It is considered that 50% of the small errors in reality help the system, thus for the SFC 0.5*87.52% of the hours with small errors do not count as error. The forecast type SFC gains most on the 50% small error discount, but still has one of the highest effective trading volumes. The result shows that both SFC and rSFC oscillate around the true value, because there is a limit for how close the forecast can get to the true value due to the variable nature of the problem and the fact that there are hidden errors from not knowing the actual generation of all units in real-time.

Forecst type	% fraction small errors	Intra-Day Tra-ding	Upscaling error 1 (0.5%) Required reserve RE1	Upscaling error 1 (0.5%) Traded volume 1 ID+RE1	Upscaling error 2 (0.7%) Required reserve RE2	Upscaling error 2 (0.7%) Traded volume 2 ID+RE2	effect. traded volume*
Upscaled OBS	-	-	0.76	-	0.50	-	-
DFC	72.42	0.00	3.31	3.31	3.22	3.22	2.465
SFC	87.52	2.56	2.21	4.77	2.07	4.63	3.733
rSFC	80.39	2.37	2.77	5.15	2.61	4.98	4.172
SFC+UP	78.95	0.64	2.75	3.39	2.65	3.29	2.468
rSFC+UP	75.21	0.28	3.07	3.34	2.97	3.25	2.469

*Table 12: Break down of the required total balancing volume for the day-ahead forecast error. All values are given in % of installed capacity (*total volume minus 50% of the error < 2%).*

The effective trading volume demonstrates that care must be taken to avoid a loss compared to not doing anything. Our results suggest that plain intra-day trading is likely to be a more expensive approach than not trading on the intra-day at all. Table 13 shows the double traded volume with and without artificial oscillations on measurements.

Forecast type	Volume with„real measure-ments“	Volume with upscaling	Fraction trading hours per year
rSFC+UP	0.01	0.035	11
SFC+UP	0.04	0.075	25
SFC	1.24	1.435	100
rSFC	1.61	1.795	100

Table 13: Volume that has been double traded. Shown are results for the various investigated forecast types with erroneous upscaling and the simulated real measurements, as well as percentage of trading hours, where some corrections have been applied. All values are given as % of installed capacity.

From this we conclude that the oscillations of SFC and rSFC are not an artifact of the verification methodology, but a problem in the forecast in its dependency on the use of measurements with inherent random errors. What is most significant in Table 13 is the percentage of time, where the forecast type suggests changes, where the uncertainty bounded forecasts have only 11% and 25%, respectively.

In a real time IEIC system, the percentage will be even lower at the 2-hour horizon, because some of the earlier increments may actually match an order. Also, the opposite can occur, that is that no order is matched at +2 hour horizon. The fractions of the required balancing power and of the used intra-day trading are broken down in Table 14, where there is no discount for 50% of the small errors. The accumulated error in the last column therefore is equivalent to the average of the two scenarios in Table 12 without 50% discount.

The most significant result of Table 14 is the low amount of double traded volume for the uncertainty bounded forecasts. It is impressing that the number of strong oscillations are so small that we can state that counter correction hardly occurs. Of course, one can argue, that the algorithm has not balanced all errors, so the forecast is far from perfect. In Table 12 we note that the SFC is after all 0.56 better on reserve requirements than SFC+UP without small error discount. It should also be noted that SFC+UP would in a IEIC real-time environment not correct the day-ahead DFC forecast, but for one of the previous increments. This would reduce both the reserve and intraday error contribution considerable.

Forecast type	Double traded volume	Fraction reserve	Fraction intra-day trading	Total error forecast
rSFC+UP	0.03	3.02	0.28	3.30
SFC+UP	0.08	2.70	0.64	3.34
SFC	1.44	2.14	2.56	4.70
rSFC	1.80	2.69	2.37	5.06
DFC	0.00	3.27	0.00	3.27

Table 14: Break down of the traded volume for the different forecast types. All values are given in % of installed capacity.

In terms of total error without the 50% discount effect we can see that the pure DFC is actually the one with the lowest error. However, the DFC is the forecast with the lowest discount for small errors, because the longer forecast horizon allow oscillations to grow.

4.2.6.2 Cost analysis of the forecast scenarios

In the next step the different forecast scenarios have been analyzed according to their costs for balancing of day-ahead forecast errors. The analysis in power space has illustrated the characteristics of the different forecast types. However, there are still some questions to be answered:

- Is it beneficial to have many small errors ?
- Does the permanent correction scheme imply too much volume ?
- What is the maximum loss to accept on the intraday market ?

These questions will be answered by use of a small fraction of the total wind power pool using actual reserve prices over a period of one year. The assumption of a small pool or more precisely a small share of the balance of all wind turbines is required for the calculation of costs. This is because error of the total German wind power would influence the reserve cost in reality. This effect is negligible, if the actual volume is assumed to be small. The assumption will underestimate the costs for all forecast types and in fact most for the DFC, because it has the largest errors.

Several methodologies for calculating the balance costs have been tested. However, it turned out that intraday prices are too volatile for use of regression methods being able to generate a realistic price function. Therefore, it was decided to combine actual reserve prices and conduct the full calculation with fixed losses on the intraday market for all forecast types separate. Thus, we compute the average balancing costs for each forecast type as a function loss on the intraday market. Cost computations have been made for all forecast types with fixed losses between 0.1-20 Euro/MWh. The result of these scenarios are shown on Figure 13. One forecast type has been added compared to the verification in power. This is a 12UTC uncertainty bounded forecast. This would in reality be increment number 2 in a real-time IEIC environment.
The resolution of the y-axis in Figure 17 of the balancing costs are 0.125 EUR/MWh between 0 to 3.5 EUR, while it is 1 EUR above 3.5 EUR/MWh.

The loss expresses that the intraday price is always related to the spot market price as:

$$P_{ID} = P_{SP} - K(i) \tag{9}$$

where P_{ID} is the intra-day price, P_{SP} is the spot market price and $K(i)$ represent the loss. $K(i)$ is positive, if power amounts are sold and negative, for the amount of power that has to be bought.

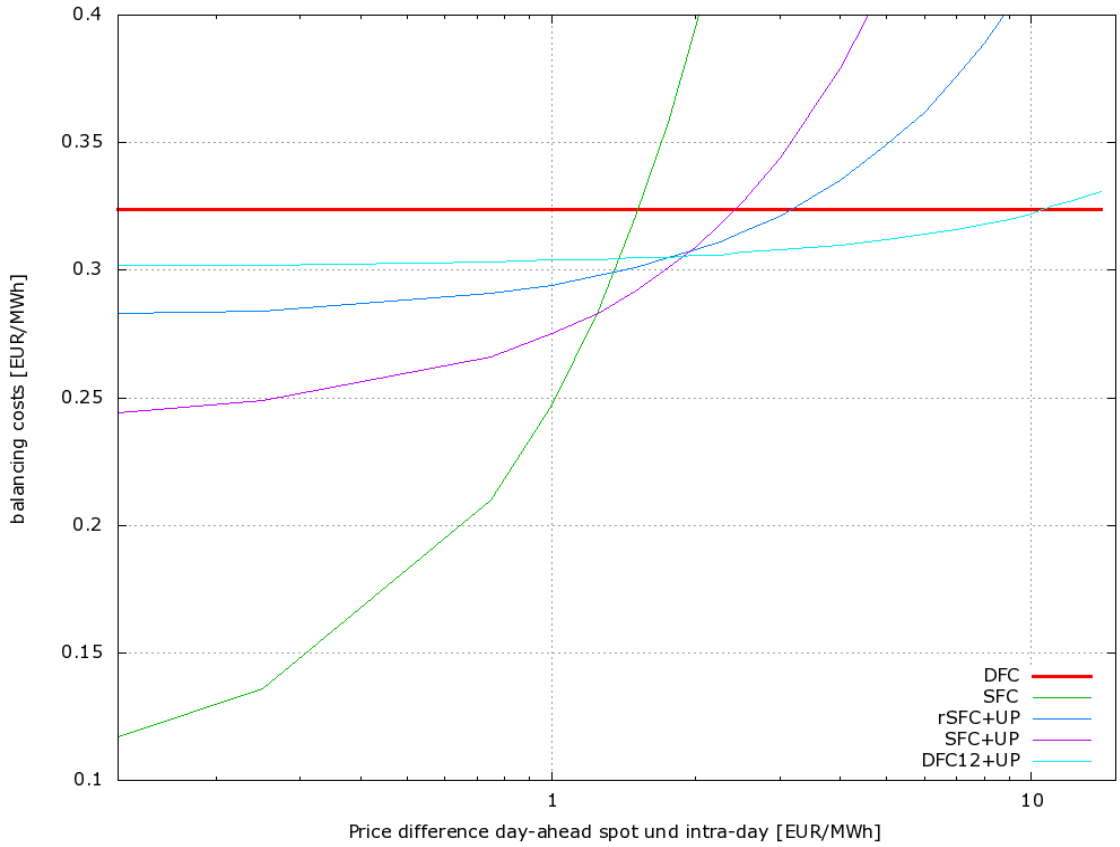

Figure 17: Results of the cost analysis for the intra-day trade with different price categories and forecasts.

In addition the reserve costs for the trading on the day-ahead market without intraday corrections are shown in Figure 17 as the red horizontal line, which is independent of the price loss in the intraday market. The forecasts on Figure 17 provides an overview of the cost of balancing. The best choice of forecast is always the lowest line. The best forecast is DFC for very high loss while SFC is the best for very low loss. In between these boundaries, the uncertainty bounded forecasts result in cheaper balancing.

Because of the LOG scale on the x-axis, the SFC seems to cover a wide range of price differences, but even the maximum loss of 1.3 EUR/MWh is not likely to cover startup costs, so the price loss bounded SFC will not result in a contract with a scheduled unit. Thus, a price loss limit would mean that all the large expensive DFC errors would count and then SFC would be a really poor forecast choice, because it reduces small insignificant errors and fails on the serious errors. One can therefore say that the SFC is an impractical forecast in the market and there will not be the expected gain in cost, which the RMSE in power would indicate.

In the following analysis we refer to the price ranges as strict as they are listed, but in real life, it would be desirable to forecast the reserve price, because higher reserve prices make the intraday market more economic and vice versa for smaller reserve prices. Market participants will possibly gain from forecasting reserve prices to the extend that this is possible on the horizon in question. Therefore, Figure 17 shall be considered valid for average conditions. For real-life applications, time of the day should at least be used to create more detailed price ranges for the recommended forecast type such that higher loss are accepted during ramps in the demand, wind and solar generation.

The DFC is the "first guess" to the IEIC scheme. This forecast can be a traditional RMSE optimized forecast. Under the assumptions that the pool is a small share of the total wind power, we achieve balance costs of 0.32 EUR/MWh by a daily spot market bid without correcting the errors in the intraday market.
Below, typical IEIC iterations are listed with information on how much volume a given forecast update would result in relative to the produced power under the assumption that the offer is always send into the market and accepted. The higher the amount and the stronger the loss limit, the less likely it is that there will be a matching offer in the market. The loss limit is the core of the incremental trading approach and in a real-time application it is realistic to perform up to 26 increments for the last hours of the day based on newer and more accurate forecast information covering these hours. How many is needed depends on the market volume and the changes in predictability from spot market closure to gate closure.

The following paragraphs illustrate the incremental principle in detail:

1. Once the 12UTC forecast (DFC12) is ready, we put an offer into the market with zero loss. This is the difference between DFC12 and DFC. This implies a trading volume of 11.3% of the produced power. The resulting forecast in the market is now named R1A for later reference, which is a mixture of DFC and DFC12 depending of the amount of matching offers.

2. During periods of lower predictability and little market volume we expect R1A=DFC, thus, there will be additional imbalance to trade unless R1A=DFC12. We offer the difference between DFC12+UP and DFC in those hours with a maximum loss of 12 Euro/MWh. The resulting forecast is now R1B, which is a linear combination of DFC, DFC12 and DFC12+UP. The difference between between DFC and DFC12+UP is 1.94% of the produced power. This means R1B and R1A differ less or equal to 1.94% of the produced power. Less than 2% of the produced power sounds little, but this covers over many hours with too high uncertainty, resulting in no increment.

3. We jump to 6-7 hours before the event. For the 2nd half of the day there would in reality be two increments per 6 hours between step 2 and 3. The next increment is based on rSFC with a lead time of 6-7 hours offered as contract swap typically for the rest of day. This update is repeated every 6 hours. The resulting forecast is now R2A.

4. Again, during low predictability the rSFC at spot price is not likely to meet matching offers in all hours of the day. Thus, we need rSFC+UP to get rid of large imbalances with allowed losses up to 3.5EUR/MWh. This may result in a trading volume of up to 6.11% of the produced power activated in 577 hours of the year, where we assume that R2A still equals DFC. In a market a DFC would often already have been incremented. Our resulting forecast is now named R2B and is a linear combination of DFC, DFC12, DFC12+UP, rSFC and rSFC+UP.

5. We now step over some intermediate increments of the same type of iteration 3 and 4. For illustration purposes, the important horizon is shortly before gate closure and we offer SFC with a maximum loss of 1.3 EUR/MWh. This is not likely to lead to much more than contract swaps, because of the short notice and relatively large volume. Without an attempt to utilize the intra-day market before gate closure, this offer correspond to 20.5% of the produced power. This is an unrealistic amount of power to trade with losses under 1.3 Euro/MWh at gate closure. So this increment can only bring balance between non-scheduled generators. The total imbalance of the non-scheduled generation will remain. The resulting forecast in the market is now R3A.

6. Because of the short notice and our price condition, there is high likelihood that R3A needs correction with SFC+UP with a maximum loss of 3.5 EUR/MWh. Thus, we must offer SFC+UP in each trading interval, where R3A is outside the uncertainty range of SFC. Our final forecast is now R3B and is a linear combination of DFC, DFC12, DFC12+UP, rSFC, rSFC+UP, SFC and SFC+UP.

In case of little volume in the market and/or no increase in predictability, we risk that R3B equals DFC and we risk high balance costs due to forecast errors. On the other hand, if the volume is good, R3B equals SFC as the other extreme and we have minimum loss. The trading strategy provides a high likelihood of some success, because we have several attempts to get rid of expected imbalance in due time. The imbalance, which is left after the various attempts is therefore much smaller than the total error of the DFC forecast. The incremental approach makes maximum use of the intra-day market with a number of positive side effects on the competitiveness.

Other cases to consider are the full export/import case. Competition factors suggest to stay at the lower/upper boundary of the uncertainty interval (see 3.2.4.3), because intraday prices for balancing power will otherwise become volatile.

We cannot estimate what the resulting balance cost is over the year, because this is entirely dependent on the offers in the intraday market. The important achievement is that we know that participation in the intra-day market is likely to pay off, because there are several opportunities to get rid of part of the imbalance. This will equivalence a much improved DFC forecast at the end of the day.

The entire IEIC strategy is targeted for minimum loss of each step. Alone the fact of splitting up the imbalance in small steps levels out the risk of price volatility, if all parties act in this way, but also the risk for an individual party to be hit hard by price volatility is reduced by the incremental approach. The IEIC is a market optimized approach to balance pools of intermittent energy like wind power. The disadvantage is extra maintenance of various parameters related to uncertainty and prices. Evaluation of the approach is not trivial and there are plenty of optimization possibilities especially linked to reserve prediction, because the reserve volume is extremely volatile especially compared to the generation of a single wind farm.

4.2.7 Practical Application of the IEIC tool

Although the approach seems rather technical with respect to the high number of increments, it should in fact be considered a strength, because the IEIC is nothing else than a fully market based explicit balancing approach deploying uncertainty and thereby increasing system security and reducing price volatility. The more volume administrated by IEIC, the better the system performance. We should also emphasize that there is a deeper scientific value of IEIC, because the approach can be generalised to other applications. An ensemble of economic models can be applied in the same way.

We have demonstrated that meteorology as science with its introduction to ensemble forecasting has shown the way forward to a fundamental and widespread effect. What meteorological science has identified is that the effect of instabilities (in the atmosphere) can only be realistically simulated with ensembles. The tailoring of ensembles has mostly been practised by the cost-loss method for decision making on days and weeks ahead.

The new algorithm is constructed to keep a running system in balance at a minimum cost. The design circumvents two challenges: (1) every action is likely to cause a loss, such that we should not act more than required. (2) the stronger we act, the higher the likelihood that we contribute to the oscillation around the true value, which is often not desired. The typical situation is that we do not know the true state of the system. Implied in this is that the forcing terms in partial difference equations describing the motion are not accurate either. On top, small waves in a system near the balance point may not shake the system, while they can shake the system when the system is out of balance. This is what IEIC is developed to handle intelligently and thereby provide a robust algorithm, which will act when things are getting critical.
An instability in meteorological context occurs when a wave caused by a small disturbance is growing rapidly. This is essentially the same phenomena that occurs in financial markets. The waves are in that case triggered by the release of positive or negative information. The IEIC principle will therefore be a helpful and defensive strategy to keep systems near balance.

Real ensembles in other markets will not develop from day to day, but the so-called poor-man ensemble technique may be utilized. This is also still applied in meteorology as the German weather service (DWD) is the leader of the so-called poor-man ensemble (PEPS) project. It is a collection of output from deterministic models from the European national met-centres. Following the principle of combined national models a poor-man model system can be constructed in various disciplines. Probabilistic forecasts of the evolution of key economic parameters will in itself be interesting for the public. However, the strength lies in the possibility to much better steer the economy following the IEIC technique with small adjustments as responses to the level of imbalance. One may even ask the question, whether this is a required approach to stabilize and tune a market economy under evolution.

Together the MSEPS ensemble approach, the Ensemble Kalman Filter (iEnKF) based short term forecast technique (Möhrlen et al, 2009) and the IEIC provide a complete system targeted for a market economy using advanced IT solutions, yet fast enough to provide immediate response to the observed pattern.

Together, they are in fact an instrument to keep a complex system in balance. They provide optimised probabilistic results that can technically replace the deterministic systems results. There is no need of a significant change in the way data is processed. New forecasts are though generated more often. However, all external communication does not need structural changes. The advanced probabilistic information stays inside the "MSEPS+iEnKF+IEIC"-box. In fact the graphical user interface called ELFI (ELectricity Forecast Interface), which has been developed and tested in the demonstration phase complements the system, because the more advanced internal probabilistic information can be accessed there either in graphical form or as tables. In that sense, we can state that the developments in this project have led to an increased application level of the MSEPS ensemble system.

The applicability of MSEPS with neural networks were examined, because of the need of enlarging the system to simulate prices on the large scale. It was found that the MSEPS is especially suited to provide data for this purpose. The reason is that the ensemble mean forecast contains much less random noise than individual forecasts. This prevents that the neural network accidentally learns something wrong in the training. Also the fact that smoothed behaviour of the ensemble mean forecast of highly uncertain and variable parameters is a benefit for improved neural network forecasting. This is an important capability, which allows use of the MSEPS+neural network+iEnKF+IEIC in new and complex disciplines with a standardized model concept.

4.2.7.1 Communication aspects for the application of the IEIC

Secure data communication via web-services over encrypted connections is another important part of bringing the IEIC system to the market. A number of web-services were developed in this respect to manage standing data, receive SCADA data and deliver forecasts.

The benefit of web-services, or more general HTTPS applications is not only that it is secure, but the request/response type is comprehensive, because the request can deliver data enveloped with SOAP and receive the forecast as one request and response (see e.g. http://www.w3.org/2002/ws/).

This increases efficiency. By use of XML format for the body of the request and response, it is possible to schema validate the data flow on both sides. In this way automated solutions between different systems can communicate based on standardized IT solutions. Web-services can also communicate directly with EXCEL. This feature has also been utilized in the demonstration phase. This is a comprehensive solution to get the latest forecast fast and directly to the user.

For certain type of communication SFTP plays a role. There is a benefit of the SFTP process in the sense that the migration from FTP to SFTP is simpler. The SFTP solution lacks a number of capabilities for advanced communication compared to HTTPS web-services, but the SFTP solution suffices for scheduled data communication and is in that sense compatible with interactive tools such as FileZilla, which is available for all common systems. MS windows systems communicate also well via SFTP to Linux systems with various tools. This has also been found a reliable way of communication in the demonstration and data are encrypted to prevent man in the middle access of confidential data.

As part of increasing the reliability and availability of the system, fail-over tools have been developed. This applies to both HTTPS and SFTP solutions. The challenges have been to make multiple systems look exactly the same to the user, to manage many users and to create a secure CHROOT jail for SFTP users. Such a jail is essential to protect the confidential data coming in and going out. The customer interface is with these enhancements future compatible and capable of communicating with different IT platforms.

4.3 Application of IEIC on longer Forecast Horizons

The optimization process for high inertia scheduled systems must spawn over a multiple of days. This applies on CHP level and for thermal generation, because prices can become unfavorable for several days in a row. This is where it becomes feasible to power some generation off. In order to illustrate that the IEIC is not limited to day-ahead and intraday trading, we have constructed Figure 18, which shows all ensemble forecasts generated during 6 days valid at a particular time.

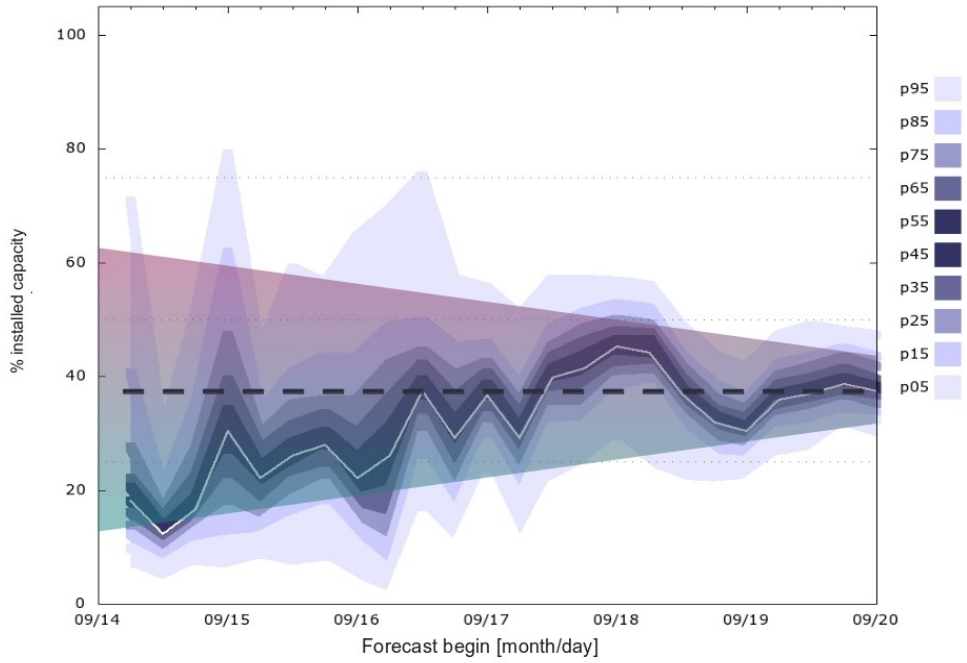

Figure 18: Schematic depiction of the change in uncertainty spread for different forecast horizons, starting with 144 hours in 6 hour intervals, up to the point in time when the forecast is valid. (2011/09/20 at 3:00UTC). The black dashed line depicts the measurements at 2011/09/20 at 3:00UTC, the white line is the so called optimum forecast, the blue shaded areas are percentiles.

The example shows that the IEIC technique can be started several days in advance to assess how much capacity is feasible to run "warm". The example shows the development of the uncertainty distribution of a forecast evolution over the course of six days. However, not as a forecast with a forecast horizon of six days, but instead from 24 different forecast runs. They exhibit the same point in time for which the forecast is valid (here: 2011/09/20-03:00 UTC). Therefore, forecasts with 144, 136, 130...24, 18, 12, 6 hour forecast horizon are combined in one graph. It is intended to illustrate with this figure, in which respect forecasts differ from one generation time to the next generation time for the same point in time. The superimposed cone clearly shows, that the spread generally becomes smaller with closeness to the point of interest, and that the forecasts oscillate around the measured value – shown as the black dashed line – even though the forecast lies always within the spread.

This is what is required for IEIC to work efficient. If the black dotted line would lie outside the spread, then the imbalance have to be dealt with in the last moment at higher loss rates. It is also fundamental that the ensemble spread is sufficient to cover the distance from the initial guess to the yet unknown true value already when it is recommended to start trading after the IEIC scheme. Over the course of six days, different percentiles were closest to the measured value, whereby the optimum forecast (white line) came closer and closer to the true value with time and with smaller positive and negative deviations. It becomes clear, that the ensemble spread represents a certain result with each time step with increasing accuracy. However, it is also important to note, that by virtue of the crossing of the measured line of values, one and the same amount would be traded multiple times, if the proposed uncertainty band is not considered. Therefore trading based on the optimal forecast need to take place without loss otherwise the uncertainty bound should be used.

Additionally, it is intended to illustrate with Figure 18, how variations in the atmosphere alter the forecasts from cycle to cycle and that it is only possible within a few hours to generate forecasts that are a nearly 100% correct in a sustained way. Most of the time, however one has to account for forecast errors. Therefore, it is critical to recognize, that for weather sensitive production units such as wind and solar power units, it cannot be expected that there will come a day, where the perfect forecast can be provided. The paradigm has to rather be that by using uncertainty estimates, the correct direction can be given and small and insignificant variations are centrally balanced in the last moment. The core of the problem is to understand, which fraction of this variability should be traded in order to accomplish a well functioning power grid with lots of variable generation units and at the same time achieve a competitive power trading with input from these fluctuating renewable energies.

4.3.1 Summary and Outlook of the applicability of the IEIC approach

Until we had developed the IEIC concept, it seemed difficult to imagine how forecast optimisation could be applied to a commercialized wind power pool or rather several pools. Centralized optimisation is more straightforward, because of the size of the pool and access right to measured generation.

From a forecast perspective, significantly more work is involved in the management of the smaller pools. Nevertheless, it was found that the IEIC is equally useful in centralized and decentralized applications. Thus, it is an algorithm, which is valid on all markets.

Some of the other developments in the project, such as the large-scale short-term forecasting after gate closure and the conditional bidding scheme are likely to be future solutions at much higher penetration levels. Therefore, it is of significant value that IEIC opens the possibility to use the spot market auction to sell power that does not yet have a very high likelihood to actually be produced. In this case, the spot market can be used for hedging to avoid a very poor value of additional wind power on the intra-day market. The market approach actually does provide opportunities to optimise the system, although the kind of optimisation, which will take place is somewhat indirect.

As an outlook it is probably of relevance to point out that large amounts of solar power and wind power in combination with the varying demand will in itself contribute to price volatility. As a consequence, there is a risk that pure financial institutions will invade the power market and help to increase the price volatility further via speculative activity. A valid concern is that the price of electricity paid by the consumer will be above the physical price of the electricity. That means we will see the spot market price lower, but the consumer will not feel this. This will essentially mean that already high prices in regions of high renewable energy capacities will get even higher with negative side effects on the competitiveness of the industry.

One can therefore conclude that the market economy is in itself a challenge. However, the purely speculative financial behaviour is a challenge for a sustainable growth, because the weather triggers volatility, which is accelerated by the market unless precaution is taken. The IEIC approach can therefore provide an answer to a number of these challenges ahead.

5 Demonstration of the developed tools and methodologies

Some of the tools and methodologies regarding the optimisation of the wind power and solar power forecasts have been brought into operation already, while some have been run over 3-6 months in a daily pre-production environment.

To be able to visualise and provide access to market participants and customers for feedback reasons, a graphical platform has been developed and built into WEPROG's web page, which has been running in a beta-version on a non-official site. The so-called "ELectricity Forecast Interface" or short ELFI is located at www.weprog.org. We have provided access to interested parties for 6 months, from June – December 2011.

Some examples of the page can be seen in Figure 19, Figure 20 and Figure 21 .

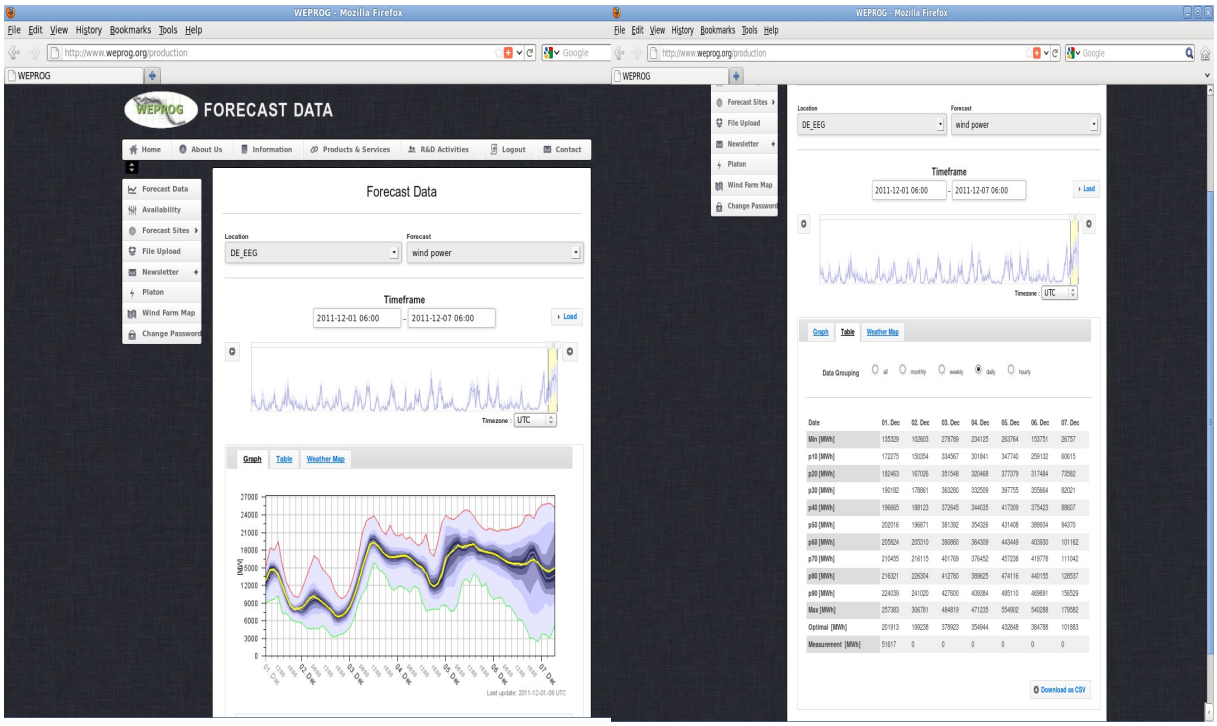

Figure 19: Example of a time series of wind power forecast and the corresponding table with values of optimal forecast, min, max and 9 percentiles

Access to ELFI has been provided as part of the demonstration. Two CHP plants have also been participating. With ELFI it is possible to view probability charts of wind power, solar power and heat-demand on the local scale as well as aggregated by country and over the entire market.

Figure 20: Example of an ELFI time series of temperature and wind speed for Nordsjæland that was available for the combined heat and power plants in the demonstration phase.

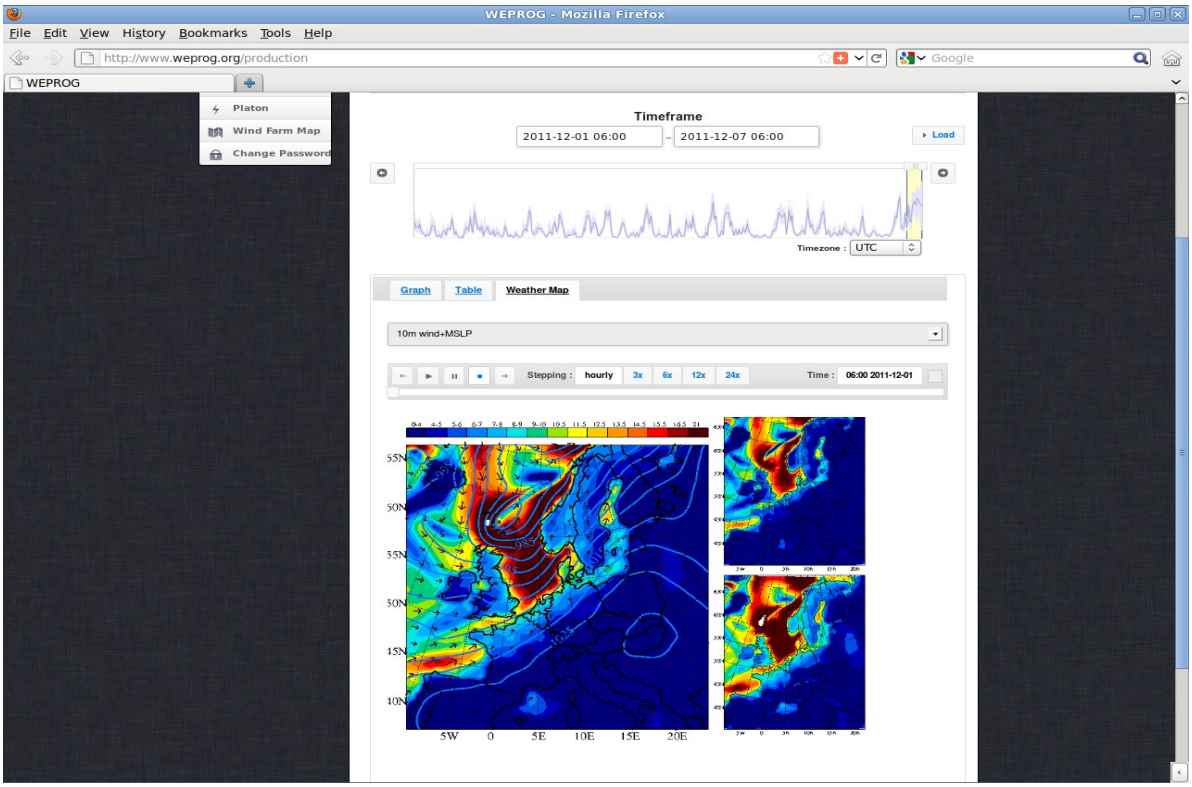

Figure 21: Example of an ELFI weather map containing a plot of mean, min and max of the selected parameter (here wind speed). The figure can be animated with frequencies of 1h, 3h, 6h.

A special example is wind power + solar power for Denmark and Germany aggregated together. This type of forecast is extremely easy to use especially for the indirect market participants like a CHP plant. With a minimum of time requirements, it is possible to evaluate periods of increased risk. A typical example for a risk situation lies in the choice of providing primary reserve or primary power. The Danish legislation prescribes that this is a either or process, which in practise has to be planned day-ahead. Primary reserve is mainly economically feasible at low spot prices, that is when there is a high amount of wind power generation during low demand periods. Because ELFI shows the probability of the basic relevant sources and sinks of energy, it is possible to evaluate the risk of extreme prices, which can be expected to occur. The strength of ELFI lies not only in what is currently visible, but also in the ease of use and maintenance. It has been designed to meet a transaction time of 0.1 seconds with 12 concurrent users and 1 second with active 100 users. In fact this is conservative, because much of the interest will be concentrated on the same information, which means faster response due to the inherent caching.

Users also spent time on analysing. Currently ELFI runs redundant, there is yet no need for or attempt to load share.

One strength of ELFI is that the underlying data can be of arbitrary type unit. At present, focus has been put on wind power, solar power, heat demand and plain meteorological variables, all represented as percentiles for either locations or aggregated areas. Emphasis has also been on ensuring low transaction time for each request regardless of whether it is historic or real-time related data. With its historic data window below the selection part, the user also gets a fast long term overview of the trend of the parameter under consideration.

5.1 Demo Setup of historical meteorological MSEPS systems

The results of many historic experiments for 2009-2010 indicated that events with low predictability are forecasted better with 75 members than 8 members. It was found that the small ensemble size of 8 members with high resolution over a large model domain performed best whenever the relevant weather system is on a very large-scale and the members do not differ too much.

In other words, the higher the predictability, the relative better performance of the mini-ensemble. What the mini-ensemble averaging does in these cases, is to ensure that the final forecast lies in the middle of a rather certain evolution. It is so to speak creating a softer solution between ensemble forecasts that differ rather little from each other. This results in a "meta-forecast" rather than an ensemble forecast. A more intelligent averaging or cluster analysis procedure could probably drag more performance out of a mini-ensemble in the low predictability events and this is a study that remains to be carried out.

The large ensemble size of 75 members however seems to have the capability to produce a much better ensemble mean in low predictability events than the mini-ensemble. One should understand the word better in the sense that the ensemble mean lies relatively close to most of the members. There are so many members in the ensemble that outliers tend to lie homogeneously around the mean with the density of member inverse proportional to the distance to the ensemble mean. For subjective analysis, it is not essential to have so many members, but for automatic systems it is convenient to make a good basic forecast of the ensemble mean, which is not too far from most of the members. Potentially, the members can form clusters and this is where the mean can end up as a poor approximation.

Clustering are maybe seen up to a couple of times per month over 6-12 hours in a row. Nevertheless, these events can potentially trigger price volatility, because many deterministic models fall into the same cluster. Whenever this cluster of forecasts is wrong, then there is a large-scale concurrent error and need of urgent adjustments. Cluster events can be identified by the trader via dense percentiles from probabilistic forecasts from large ensembles such as the 75 member MSEPS ensemble. Knowing that the bulk of the market believes in the largest cluster, then the trader is encouraged to select another cluster to not cause a correlated error, which is more difficult to balance in a cost efficient manor. The large ensemble can in that case be used defensive to protect against excessive balance costs. In high wind events without cluster formation, the large ensemble is rather used offensive to oversell compared to the MW optimized forecast.

The overall conclusion is that the mini-ensemble can only serve very specific tasks such as being a good MW optimized forecast in high predictability events. The larger ensembles offers more possibilities for strategic trading based on risk evaluations.

Previous 75 member MSEPS setups have either been nested directly with global forecast data or were setup as double nested MSEPS ensembles. For the demonstration, the results of the historic experiments suggested hybrid setups consisting of:

A) Large scale 8 member mini-ensemble used as system generating lateral boundary conditions in 0.45° for system C.
B) Large scale 8 member mini-ensemble for wind power forecasting in 0.225° resolution
C) European scale 75 member ensemble for wind power forecasting in 0.225° resolution
D) Detailed 24 member ensemble for wind power forecasting in DK1/DK2 in 0.05° resolution

For all system setups the focus was on the forecast horizon 6-48 hours. This implies 107 weather forecasts split on large scale and local scale in 3 different spatial resolutions. The demonstration setup shall be seen as a 75 member basic ensemble (C) with 32 additional forecasts from system B and D to test whether an approximation that is common in all 75 C members leads to a solution that is incorrect and not reflecting the true uncertainty. This issue has been identified as a common problem in ensemble forecasting, because every ensemble prediction system is perturbed from a kind of reference or kernel setup (see e.g. Buizza et al, 2005). Therefore, our design of the demonstration setups ensured especially through the different domain sizes and resolutions, that no event should be forecasted with too high likelihood, because of single common approximations in the ensemble system design. In our test environment, the smallest scale of motion differs with nearly one order of magnitude between the members. This alone ensures that the numerical solutions of the members are likely to turn out different, unless the predictability is very high.

At first glance one could get the impression that there is no trust in the model result, because of the differences between A, B, C and D. However, this setup shall be understood as an attempt to gain spread when there is a possibility that the solution is sensitive to a common approximation of the core ensemble in setup C. So, the goal is rather to trigger different solutions whenever possible.

Fundamental in this design is a sufficiently large and nearly independent model domain behind each member to allow for uncertainty sources far away from Europe. The domain of dependence goes 4500km west of Europe counted for a phase speed of 90km/hour. Although wind speeds can be 3-4 times that amount at jet level, it is unlikely that a low pressure system that reaches ground level can move faster than 70-80km/hour except in very shortly time periods.
Due to high uncertainty over the Arctic region, all ensemble members should actually cover that region even though the weather systems from this direction develop only rapidly over the northern Atlantic and are very young and yet small when they pass over Denmark.

The phase speeds for weather systems south and east of Europe are smaller, but we nevertheless use the worst case and choose two large scale ensembles of 9000x9000 km. The purpose of this ensemble is to provide lateral boundaries for the two independent high resolution systems and pair A with C and system B with D for maximum independence.

5.1.1 Technical Considerations with respect to Real-time Operation

The high resolution ensembles were dimensioned according to what is technically feasible to conduct within reasonable time before spot market closure for the day-ahead horizon to make the demonstration system not only an academic exercise.

Setup C in 0.225° ensemble can be integrated stable with 4000x4000km with 75 members in due time and the 0.05° ensemble was dimensioned to 1000x1000km. This means that only a few of the very fast low pressure systems will develop in A and reach Denmark within 48 hours.

It is with current technology not economically feasible to increase the area size of system D further. Therefore, we have to accept that many systems will cross the boundary during the forecast, but the systems crossing the boundary are then resolved in 0.225°.
An attempt to further increase the 5km area seems to increase the risk of fictive numerical instability that triggers small scale low pressure systems, which may develop to low pressure systems and consequently false alarms. A high resolution model domain of 1000x1000km is still sufficient to capture meso-scale systems well.

5.1.2 Capacity handling, training and wind power forecasting

The translation of weather to wind power takes place fully independent in the four different setups. This applies to the training as well as the forecast phase. As an example, setup D generates separate power curves for 960 grid points in DK1/DK2 against 47 and 126 grid points in setup A and B/C. A factor of 7.5 and 20 in the level of detail versus setup A and B/C should give setup D the possibility to better reproduce a correct diurnal cycle on DSO level, where setup A-C smooths out the differences and hides the actual error under the umbrella of area aggregation.

5.1.3 Verification from July to December 2011

As can be seen from the results in Figure 22 of B, C and D, there are hardly any differences in the error statistics on aggregated wind power. However, the simple combination of the 3 setups provides lower error, regardless of whether we verify on the specific day-ahead or at all four forecasts per day from 24-48 hours. This indeed confirms that area aggregation can hide errors quite well.

One question that sticks out even stronger is why we cannot demonstrate setup D to be better ? The answer lies in that more responsiveness to the weather in the form of sharper ramps and more extreme local conditions leads to more double punishment, especially for phase errors.

That is, first the setup is punished for predicting an event that did not take place and later it is punished for not having forecasted the real event. The softer moderate resolution setups are punished less in each event seen in RMSE scores. The results confirm this theory, because the MAE indicates that B/C/D are in scores close to each other, while the RMSE evaluation suggests that B/C are better than D. Thus, setup D generates more large errors than the other 2 setups. The same, but more pronounced pattern can be found, when D is compared against the plain average of B/C/D.

AREA System/ Parameter	DK1				DK2			
	BIAS	MAE	RMSE	RSC	BIAS	MAE	RMSE	RSC
06UTC runs								
Reference	1.49	5	7.17	7.41	-0.27	5.35	7.82	8.47
B	0.55	4.55	6.46	6.75	-0.66	4.99	7.3	7.76
C	0.64	4.48	6.43	6.71	-0.33	5.04	7.32	7.8
D	-0.63	4.58	6.56	6.9	0.06	5.24	7.65	8.2
BCD	0.51	4.4	6.29	6.59	-0.3	4.91	7.16	7.62
00,06,12,18UTC runs								
Reference	1.53	5.41	7.73	8.15	-0.3	5.68	8.28	9.03
B	0.53	5.02	7.11	7.56	-0.73	5.3	7.81	8.4
C	0.65	4.97	7.05	7.47	-0.39	5.34	7.82	8.43
D	-0.6	5.02	7.17	7.65	0.13	5.67	8.27	9.01
BCD	0.2	4.87	6.9	7.34	-0.33	5.24	7.7	8.3

Figure 22: Statistical Results for 3 MSEPS Test setups, a Reference Setup and a combination setup of the Test setups A, B and C.

Consequently, we conclude that setup D is mostly relevant for further physical computation of derived variables or alternatively rather simplistic statistical translations of weather forecasts to end-user products, because the local data needs less statistical corrections. A sufficiently advanced statistical translation of setup B/C gains so much on area aggregation relative to setup D, that setup D can only be justified, when it is combined with setup B/C. This is not a bad result considering that there is some added value from the more detailed localized forecasts. For system security purposes, setup D is always to prefer, especially on regional level.

Fundamental is also that the training in this case starts from aggregated published generation. There is no detailed information other than the location. The turbine availability is constant and only published aggregated generation is used. The training therefore distributes any mismatch between the actual generation and the aggregated value over all grid points.

Thus, we can conclude that it is possible to generate detailed forecasts, although it is not known, whether detailed measurements would result in more accurate forecasts or not. It is also not known whether the choice of running 24 members is compromising accuracy compared to 75 members. However, the fact that setup B and C have approximately similar errors suggests that the number of members does not matter too much for this result.

5.1.4 Discussion and Further Improvements

The way forward is to try to understand how different factors influence the result and especially how to make the ensemble forecasts differ from each other to ensure that the final forecast lies close to the middle of the uncertainty interval. Automatic smoothing via increased spread while the weather is uncertain and increased resolution beyond grid point scale of the average forecast during certain weather events are two key factors of ensemble forecasting. The real-time challenge lies in how to make the simulations so reliable that spurious noise does not arise from technical errors or extreme weather causing some ensemble members to not perform as intended. There are many pit falls when running 107 different forecasts in real-time.

The weather from beginning of August 2011 to the end of the year was highly variable. Farmers will remember the harvest as one of the most difficult in the past decades. Additionally, there were only few dry periods in October and November and the December weather became extremely windy in the end. From a demonstration perspective, the period from June to December 2011 could not have been much more challenging.

The relative low errors achieved for the period up to 15th of December confirm that the system has been running technically correct. The last half of December revealed some numerical instabilities in setup D. However, the weather was extreme and revealed that this period should be rerun a couple of times to tune the setup with respect to numerical stability. The challenges were related to the combination of steep orography and high resolution, because it was possible to recreate numerical stability in each instability event by manually tuning of these parameters.

5.1.5 Dynamic Meta Forecast Calculation versus a Forecast Composer

The strength of setup B, C and D lies in that they are fundamentally different and they differ on parameters that are common between the members in C. This means that whenever a common approximation between the members in C cause a common error, then it is likely that this will show up with some sensitivity in either setup B or D. Two typical examples where C could have weaknesses would be in the case of:

- a small scale low pressure system developing in a size smaller than the grid scale of C not far from Denmark. This is likely to occur in late summer after one or more days of flow from North

- a small scale low developing exactly on the border of C in upstream direction and during the first forecast hours, which is poorly balanced and not present at all in the lateral boundaries due to a lack of resolution

Setup C is constructed to capture nearly all events while B and D are constructed to capture exactly those events where C has a weakness. Since we understand when B and D are superior to C, then it is also possible to calculate weight coefficients based on simple parameters. Two possible functions could be:

1. The higher the average production measured over 48 hours, the more weight on B.

2. The more the forecast correlates to the average daily generation, the higher the weight of D.

Instead of deciding the final meta-forecast for the user, it was decided to create a forecast composer application in the form of a web-service. This was developed for the demonstration phase, where it is possible to combine forecast A, B, C and D to an improved meta-forecast with time dependent weight coefficients. The web-service produces a result of native forecast values and weight coefficients for A, B, C and D. The entire sheet loads into Excel, where the user can then change the weight of the forecasts and in that way compose the final forecast locally and submit this forecast further to the market.

The forecast composer strategy allows the end-user to choose the final forecast and consider the risks. In fact, the composer was configured to load 3 sheets. Two for DK1/DK2, respectively and one for all Germany to allow for more strategic bids and especially to provide the possibility to consider the risk of negative prices. The combined use of the forecast composer and IEIC will provide the highest possible level of flexibility to the balance responsible party to reduce balancing costs. Low market volume in the weekends will probably encourage to bid into the spot market with minimum error, while high wind events during higher demand periods will probably be utilized to sell more than the RMSE optimized forecast suggests.

Trading of wind power will therefore most likely develop from being balance optimized to be price optimized. For the electricity prices this will probably mean lower prices, because more non-scheduled cheap generation will be bid into the market.

6 Overall Conclusions and Outlook

"Unfortunately when you most need predictability, that's usually when the atmosphere is most unpredictable." This comment made by McElroy from the American National Weather Service is equally valid in an energy context. Here, one can argue that a certain amount of predictability is always required to operate the electrical transmission grid in a safe manner. We nevertheless experience regularly issues that are directly connected to a lack of predictability of the generating units.

In the energy markets this lack of predictability is reflected in price volatilities, which are often triggered by uncommon weather phenomena that are difficult to predict, and/or some other abnormal conditions related to transmission or/and sudden failure of scheduled generation plant. These are moments where the market participants experience the highest losses and profits. So, there is an underlying need of forecasting tools that increase predictability and reduce risks.

This is what ensemble forecasts have been developed for since the early 1990's, where the first two operational ensemble prediction system were launched by ECMWF and NCEP using the singular vector and the breeding method, respectively (see e.g. Mureau et al, 1993, Buizza et al, 1995, Toth et al, 1993, Tracton, 1993).

In fact, this has been the starting point of scientific demonstrations proofing that ensemble forecasting is the state-of-the-art methodology to find the instabilities in the atmosphere that cause most forecast uncertainty and error. Ehrendorfer (1997) and Toth (1998) expressed this by stating that "Ensemble forecasting is the only practical technique to assess the flow dependent predictability of the atmosphere, and to create probabilistic forecasts reflecting it".

Though statistical techniques can also produce probabilistic forecasts based on traditional single control Numerical Weather Prediction (NWP) model forecasts (see e.g. Leith et al. 1974, Stephenson et al. 2000), it is to expect that the quality of ensemble technique will lap that of the deterministic approaches soon, because the relative gain from increased resolution becomes less and less.

This is because there is no detailed information available about the true state of the atmosphere, which matches the resolution. Alone from this single fact it is likely that the deterministic approaches cannot deliver much more accuracy by increased resolution than what is achieved today, whereas the ensemble technique can benefit of resolution for another decade. The defender of the deterministic approach will claim that if more detailed observations are fed into the deterministic model it will increase the accuracy. The opponent will say that this is of no help, because the local measurements are contaminated with energy on scales that have no significance for the evolution of the weather. Thus the measurements provide only a collective value, so at the end of the day the local scale accuracy does not increase except in moments of calm and predictable weather, which is where the improvement is not needed.

From a practical point of view users of weather forecasts should focus on ensembles sooner than later, because the deterministic model results will only sustained improve during weather events where the forecasts are already good enough and possibly on the very local scale in semi complex terrain.

6.1 Investigations of the weather relevant aspects on predictability

The numerical experiments we have been carrying out within this project confirm this expectation. In fact, we carried out a large number of experiments that allowed us to get a deeper understanding on these fundamental and underlying processes that affect the forecast quality of the weather models and the predictability of the weather at hand. The following list provides an overview of the aspects we investigated:

1. Test of a new wind profile theory with the multi-scheme ensemble
2. Extreme event analysis, where all models failed
3. 1.5 year simulation of a 75 member hemispheric ensemble
4. Long-term simulations with focus on and model changes of:
   - Different orography formulations
   - Variation of Initial conditions
   - Various different resolution considerations of the NWP Models
   - The isentropic potential vorticity (IPV) as catalyst to trigger strong developments of low pressure systems causing steep ramp event.
   - typical weather conditions/regimes with risk for large error

The results from these experiments have been described in Sections 2.2.3 , 2.3.1 , 2.3.2 , 2.3.3 , 2.4 and are manifold. The complexity increased almost with each sensitivity experiment, because the differences in skill of the various experiments were often small. A large number of statistical tests were carried out on forecast data in Denmark, Germany and Ireland in order to quantify the impact of changes. In particular, it was found that a given change seldom resulted in consistent changes in the forecast skill over all of the different countries. Changes in the roughness and orography over Canada and Greenland for example had most impact on Ireland, whereas changes in Norway had stronger impact in Germany and Denmark, but sometimes not in the same direction.

A major effort was also spent on locating the weather events in which the new mountain formulation is superior to the old. The analysis suggested that the larger the scale of motion and the stronger the jet stream, the better the forecast skill of the model setups with tall orography.

The overall result and conclusion that can be drawn from our experiments is that the essential part of weather forecasting in an energy context is to be able to reflect the uncertainties that lie in the weather conditions in a realistic way. Only then will it be possible to meet the challenges that come with the massive integration of variable energy sources such as wind and solar power. The consequences of that is the need for tools to understand the uncertainty and to be able to make decisions on the basis of the uncertainties.
For this reason, we have developed some essential tools that enable end-users of weather- and/or wind- and solar power forecasts to take decisions on the basis of the uncertainties that we compute with the ensemble forecasts.

6.2 Forecast Optimization Criterion

We have raised the question on the feasibility of MAE/RMSE optimisation on the basis of high forecast uncertainty in Section 2.2.2 and 2.2.3 . One of the questions was how we should take account for a minority of ensemble forecasts, that are 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 the same time suggest an additional pre-allocation of reserve, which will prevent price volatility and increase system security.
Our nearly 10 years of 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 cost profile of the available reserve is asymmetric. The forecast should then follow a linear trend over several hours, unless it is very certain that the actual value will not follow this trend. Typically, the percentiles of ensembles have the capability of suppressing small uncertain ramps and keep the trend over hours of high unpredictable variability.

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.

Our studying of the market integration of wind power in Denmark and Germany, the SuperGrid simulations (Section 3.1) and extended Grid control corporation study (Section 3.2) resulted in a new understanding of the needs and requirements of large-scale integration of variable generation units and eventually to a paradigm change (Möhrlen et al, 2012). The results pointed strongly to a situation, where not the forecast with the lowest RMSE or MAE is the best forecast, but that forecast which fits best to the market situation, i.e. the forecast that requires the least reserve capacity and produces the highest revenue. When these new parameters were taken into account in the forecasting step, it became clear that new techniques were required to solve these targets.

We found that when trading and balancing volume is taken into consideration that care must be taken when evaluating what is a good and a poor forecast. We also found that the when using the best possible short-term forecast to correct errors from the day-ahead, the result was highest amount of effective volume to balance and creation of the highest costs. The required balancing volume was significantly reduced with a simple uncertainty band consideration even though the plain MAE/RMSE of the uncertainty bounded forecast is higher. The very simple explanation of why the RMSE tuned forecast is the most expensive in cost space is that the forecast over-adjust and hits at times to the wrong side, because it is fed with some information which is not representative. The result is that a large amount of energy is traded multiple times. The uncertainty bound forecast has a quality assessment on top of a trading volume consideration. Consequently less loss can be expected.

6.3 Ensemble Forecast based Tools and Techniques for optimal trading

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 in the market and need to agree on a solidarity principle to protect themselves against other stakeholders. Sharing of imbalance in pools is a simple solidarity example.

Secondly their balance responsible parties (BRP) need to consider the risk factors based on ensemble forecasts and other relevant technical information on the operating status of the power system.

A number of new methods and optimisation to prevent price volatility has been presented in Section 4.2 . The first part of this work was focusing on cost optimisation and was split into two slightly different applications. One approach was developed that works with current market and ancillary service structures along with a second approach, which aimed to develop more future compatible practises on how to handle more renewable energy in a market without compromising grid security. Both branches are targeted to reduce the costs of wind power generation in order to reduce the costs that are today mainly carried by the consumer. It is justified that renewable energy today receives some kind of consumer paid production incentive, which works as a matter of fact like an insurance against increasing energy prices. The idea is that this improved handling makes wind power more competitive and will therefore contribute positively to lower the average energy price in exchange for the incentive payment. For this purpose, we introduced a so-called price optimised forecast, where we compute the uncertain part of the forecast to additionally base an optimisation criterion on the market situation and the expected price structure. By doing this, we achieve a bidding strategy for the wind power, where the risks of over production or under-production are additionally evaluated upon the expected market development.

In the next step, we showed how the use of uncertainty forecasts can prevent counter regulation of imbalances caused by errors in forecasts used for the intra-day market. The use of uncertainty bands around an optimised single forecast ensures additionally that imbalances are only traded, if they are certain. Because the certain parts of errors can be forecasted long in advance. This methodology in fact reduces the very short-term requirements of volume in the intra-day and thereby reduces the risk of failure of trading imbalances and thereby increases the overall volume and competition on the intra-day market.

In Section 4.2.6.2 , we have additionally to our in depth cost analysis provided a recipe on how to use such uncertainty forecasts for trading purposes and how imbalances can be traded long in advance while the market volume is larger.

With the introduction of what we call an I*ncremental Ensemble-generated Imbalance Correction* technique, or short IEIC technique, we have developed a tool that can be considered a milestone in wind power forecasting. The IEIC is a major step forward in reducing actual costs of balancing wind power through it's iterative and uncertainty bounded characteristics. The IEIC is very suitable for a fully automated system, because uncertainty is the core element in this algorithm and automated data quality check algorithms play an important role in a real-time system.

The tools that we have developed will help to further increase the volume of the intra-day market at the opening hours such that this market over time will develop to serve as a system-wide adjustment platform to level out undesirable effects from the spot market auction. Such ongoing iterative auctions in the intra-day market environment can therefore lead to an important efficiency increase of the energy system.

We have also demonstrated that the non-scheduled generator can increase the revenue by utilizing the intra-day market to achieve a better balance between what is traded on the market and what is produced. Fundamental for the success is the use of uncertainty forecasts to limit the volume to be traded on the intra-day market. Otherwise the intra-day market will not be a competitive platform for balancing in the future.

We have demonstrated that the plain and permanent use of short-term forecasts cause that a significant amount of energy is first traded on the intra-day market and secondly counter corrected by the TSO, because the short-term forecast had errors from inadequate measurements or up-scaling techniques. Trading too much is not in itself a problem, but on average it is known that the non-scheduled generator is loosing revenue on the intra-day market relative to the spot price, reducing the economic value of wind power and in the longer run making investments less feasible.

The algorithm we have developed ensures minimum loss of revenue, as it continuously estimates how far the non-scheduled pool is for sure out of balance. The amount that is for sure out of balance can be traded with a slightly higher loss factor, when it is computed and compared to the expected balance cost.

When trading the uncertain part of the imbalance, it is crucial that this part should only be bid into the market at the day-ahead spot price for the hour, i.e. as a type of contract swaps. Such a swap will reduce the risk for both parties. Thus, there should be an incentive for all non-scheduled generators to participate fully in the intra-day market with bids bound by uncertainties at prices close to the spot market price. This is a fundamental step in the direction of a competitive market.

Our verification has shown that a RMSE evaluation shows the fit of the forecast to the measurements, but the economic analysis demonstrates that the RMSE optimised forecast is not competitive unless intra-day trading can take place up to the gate closure with no or little loss. That should always work for swapping contracts with other non-scheduled generators, but they may have correlated imbalance. The prices required to balance with scheduled generators increases towards the gate closure due to urgency and increasing start-up costs. This is where the major losses occur and where uncertainty limits on the bid pay off.

Although the development of the correction tool IEIC is one out of many in this project, it is the one development that will have immediate influence, because it is a practical solution to the challenge that the balance of non-scheduled units experiences with the increasing commercialization. When the balance is split out on many shoulders, then it is getting more difficult to make centralized optimisation.

Outlook

Although we have shown that it is non-trivial to improve wind power forecasting in power space and even more difficult in cost space, we have shown techniques that have a potential to change the balancing of renewable energy systems to the better. We have discussed how the incremental ensemble based IEIC correction approach can be catalyst to increased intra-day volume, which will further increase the applicability of IEIC. A second benefit is improved hedging possibilities on the spot market for renewable energy, which in turn will contribute to lower prices and thereby push more volume into the intra-day market. The IEIC can therefore be regarded as a catalyst to a transition process in the direction of a more dynamic market.

Acknowledgement

This work was supported by the Danish Public Service Obligation (PSO) fund project under contract PSO-2009-1-10260, which is gratefully acknowledged. The measurement data for verification of the wind profile theory was provided by the DEWI, the German Wind Energy Institute and BSH, the German federal maritime and hydrographic agency (http://www.fino1.de).

References

50Hertz (2009): Regionenmodell des Stromtransports 2009, German TSOs. http://www.50hertz-transmission.net/de/1388.htm

Amprion (2011a): Ausgleichsenergieabrechnung gegenüber der Bilanzkreisverant-wortlichen. http://www.amprion.net/ausgleichsenergiepreis#

Amprion (2011b): Am Intra-Day-Markt beschaffte bzw. veräußerte Strommenge. http://www.amprion.net/bilanzkreis-eeg#

Bergmann, J. (1998): A Physical Interpretation of von Karman's Constant Based on Asymptotic Considerations - A New Value, J. Atmos. Sci., 55, pp 3403-3405.

Bergmann, J. (1997): On the Quantification of Vertical Exchange in an Idealized Neutral PBL A Physical Derivation of the Exchange Coefficient, Contr. Atmosph. Phys. 70, 65-77.

Bergmann, J. (1995): Zur Quantifizierung des vertikalen Austausches in einer idealisierten neutralen oder stabilen Planetarischen Grenzschicht - Ein Konzept zur Herleitung des vertikalen Austauschkoeffizienten, PhD thesis, accepted by the faculty of geo-sciences, University of Hamburg.

BMJ (2002): Kraft-Wärme-Kopplungsgesetz vom 19. März 2002 (BGBl. I S. 1092), das zuletzt durch Artikel 11 des Gesetzes vom 28. Juli 2011 (BGBl. I S. 1634) geändert worden ist, http://www.gesetze-im-internet.de/kwkg_2002/index.html

BMU (2011): Gesetz zur Neuregelung des Rechtsrahmens für die Förderung der Stromerzeugung aus erneuerbaren Energien.
http://www.bmu.de/erneuerbare_energien/downloads/doc/47585.php

BMU (2008): Gesetz für den Vorrang Erneuerbarer Energien (Erneuerbare-Energien-Gesetz – EEG). http://www.bmu.de/erneuerbare_energien/downloads/doc/40508.php

Brankovic C, Palmer TN, Molteni F, Tibaldi S, Cubasch U (1990) Extended-range predictions with ECMWF models: time-lagged ensemble forecasting, Q. J. R. Meteorol. Soc., 116: 867-912

Buizza, R. and T.N. Palmer (1995): The singular vector structure of the atmosphere global circulation. J. Atmos. Sci., 52, 1434-1456.

Buizza, R., Houtekamer, P.L., Pellerin, G., Toth, Z., Zhu, Y., Wei, M. (2005): A comparison of the ecmwf, msc, and ncep global ensemble prediction systems. Mon. Wea. Rev., 133, 1076–1097.

Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen (2010): Beschluss BK6-08-111.

Bundesnetzagentur (2011): EEG-Strom an der Boerse, Press release from 12.07.2011. Online:
http://www.bundesnetzagentur.de/cln_1931/SharedDocs/Pressemitteilungen/DE/2011/110712_Berlin_EEGStromAnDerBoerse.html

EEG-KWK (2010): Transparenz der Vermarktungstätigkeiten gemäß § 2 AusglMechAV, http://www.eeg-kwk.net/cps/rde/xchg/eeg_kwk/hs.xsl/525.htm

EEG Konto (2010): http://www.eeg-kwk.net/de/file/UeNB_EEGKontostand-2010-12-31.pdf

Ehrendorfer, M., (1997): Predicting the uncertainty of numerical weather forecasts: a review. Meteorol. Z., N.F., 6, 147–183.

EPEX SPOT (2011a): http://www.epexspot.com/de/presse/nachrichten/details/news/Konferenz_EPEX_SPOT_BNetzA_zu_Erneuerbaren_Energien

EPEXSPOT (2011b): EPEX Spot Operational Rules, 08-08-2011.

Houtekamer, P.L., and Herschel L. Mitchell, (1998): Data Assimilation Using an Ensemble Kalman Filter Technique. Monthly Weather Review: Vol. 126, No. 3, pp. 796-811.

Jørgensen JU, Möhrlen C (2011): Increasing the competition on reserve for balancing wind power with the help of ensemble forecasts, Proceedings 10th International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plants, Aarhus, Denmark

Lang S, Möhrlen C, Jørgensen J, ÓGallachóir B, McKeogh E (2006): Application of a multi-scheme ensemble prediction system for wind power forecasting in Ireland and comparison with validation results from Denmark, Scientific Proceedings European Wind Energy Conference, Greece

Leith C.E. (1974): Theoretcal skill of Monte Carlo forecasts, Mon Wea Rev 102, 409-418.

Möhrlen C (2004): Uncertainty in Wind Energy Forecasting, Thesis (PhD) Department of Civil and Environmental Engineering, University College Cork, Ireland, DP2004 MOHR

Möhrlen C, Jørgensen JU (2006): Forecasting wind power in high wind penetration markets using multi-scheme ensemble prediction methods, Proceedings German Wind Energy Conference DEWEK, Bremen, Germany

Möhrlen C, Jørgensen JU, Pinson P, Madsen H, Runge Kristoffersen J (2007): HRENSEMBLEHR - HIGH RESOLUTION ENSEMBLE FOR HORNS REV, Proceedings European Offshore Wind Energy Conference, Berlin, 2007.

Möhrlen C, Jørgensen JU (2009): A new algorithm for Upscaling and Short-term forecasting of wind power using Ensemble forecasts, Proceedings 8th Int. Workshop on Large-Scale Integr. of Wind Power.

Möhrlen,C., Jørgensen, J.U. (2010): Using Ensembles for Large-scale Forecasting of Wind Power in a European SuperGrid context, Proc. of the German Wind Energy Conference DEWEK, Bremen.

Möhrlen,C., Pahlow, M., Jørgensen, J.U. (2012): Untersuchung verschiedener Handelsstrategien für Wind-und Solarenergie unter Berücksichtigung der EEG 2012 Novellierung, Zeitschrift f. Energiewirtschaft, No. 1/2012, Vol.36, No.1.

Molteni F, Buizza R, Palmer TN, Petrollagis T (1996): The ECMWF Ensemble Prediction System: Methodology and Validation, Q.J. R. Meteorol. Soc., 122:73-119

Mureau, F., Molteni, F. and Palmer, T.M., (1993): Ensemble prediction using dynamically conditioned perturbations. Quart. J. Roy. Meteor. Soc., 199, 299-298.

Nanahara T, Asari M, Maejima T, Sato T, Yamaguchi K, Shibata M (2004): Smoothing effects of distributed wind turbines. Part 2. Coherence among power output of distant wind turbines. Wind Energy, 7:75–85, 2004.

Paulun, T., (2011) EPEX Spot Study: Green Power Trading on the Exchange, Presentation at the EEG conference, Berlin. Available Online:
http://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/BNetzA/Sachgebiete/Energie/VortraegeVeranstaltungen/EEGStromAnDerBoerse/PaulunEPEXSPOTpdf.pdf

Palmer TN, Molteni F, Mureau R, Buizza R, Chapelet P, Tribbia J (1993): Ensemble Prediction, ECMWF seminar proceedings „Validation of models over Europe", Vol 1, ECMWF, Shinfield Park, Reading, UK.

Stensrud, DJ, Bao, JW, Warner, TT (2000): Using Initial Condition and Model Physics Perturbations in Short-Range Ensemble Simulations of Mesoscale Convective Systems , Month. Weath. Rev., 128: 2077-2107.

Stephenson, D. B. and F. J., Doblasreyes, (2000): Statistical methods for interpreting Monte Carlo ensemble forecasts, Tellus 52(3), 300-322.

TenneT TSO (2012), IGCC Price Model: Market Information from the Transmission System Operators, Online available at:
http://www.tennet.org/images/Market%20information%20IGCC_tcm41-20521.pdf

The WindPower.net (2005-2012): www.thewindpower.net

Toth Z, Kalnay E (1993): Ensemble forecasting at NMC: the generation of perturbations. Bull. Am. Meteorol. Soc., 74: 2317-2330

Toth, Z., and E. Kalnay, (1997): Ensemble forecasting at NCEP and the breeding method. Mon. Wea. Rev, 125, 3297-3319.

Toth, Z., Y. Zhu, T. Marchok, Tracton, and E. Kalnay, (1998): Verification of the NCEP global ensemble forecasts. Preprints of the 12th Conference on Numerical Weather Prediction, 11-16 January 1998, Phoenix, Arizona, 286-289.

Tracton, M. S. and E. Kalnay, (1993): Ensemble forecasting at NMC: Operational implementation. Wea. Forecasting, 8, 379-398.

Tastu J, Pinson P, Kotwa E, Madsen H, Nielsen HA (2011), Spatio-temporal analysis and modeling of short-term wind power forecast errors. Wind Energy, 14:43–60.

Vincent C, Giebel G, Pinson P, Madsen, H (2010) Resolving nonstationary spectral information in wind speed time series using the Hilbert–Huang Transform. J. Appl. Meteor. Climatol., 49:253–267

Vincent, C. L., Pinson, P. and Giebela, G. (2011), Wind fluctuations over the North Sea. International Journal of Climatology, 31: 1584–1595

Zolotarev P, Treuer M, Weißbach T, Gökeler M (2009) Netzregelverbund, Koordinierter Onsets von Sekundärregelleistung. VDI-Berichte 2008, VDI-Verlag GmbH, ISBN:9783180920801

Zolotarev P, Gökeler M (2011), Netzregelverbund – Koordinierter Einsatz von Sekundärregelleistung, in VDE Proc. „Die Dynamik des Netzes - Konsequenzen für Erzeugung, Übertragung und Verteilung im sich verändernden Umfeld" - 10th ETG/GMM-Fachtagung "Netzregelung und Systemführung", VDE ETG-Fachbericht Band 127 , ISBN 978-3-8007-3336-1.