PO. ID 0126

From Ensembles to Probabilistic Wind Forecasts - How Crucial is the Ensemble Size?

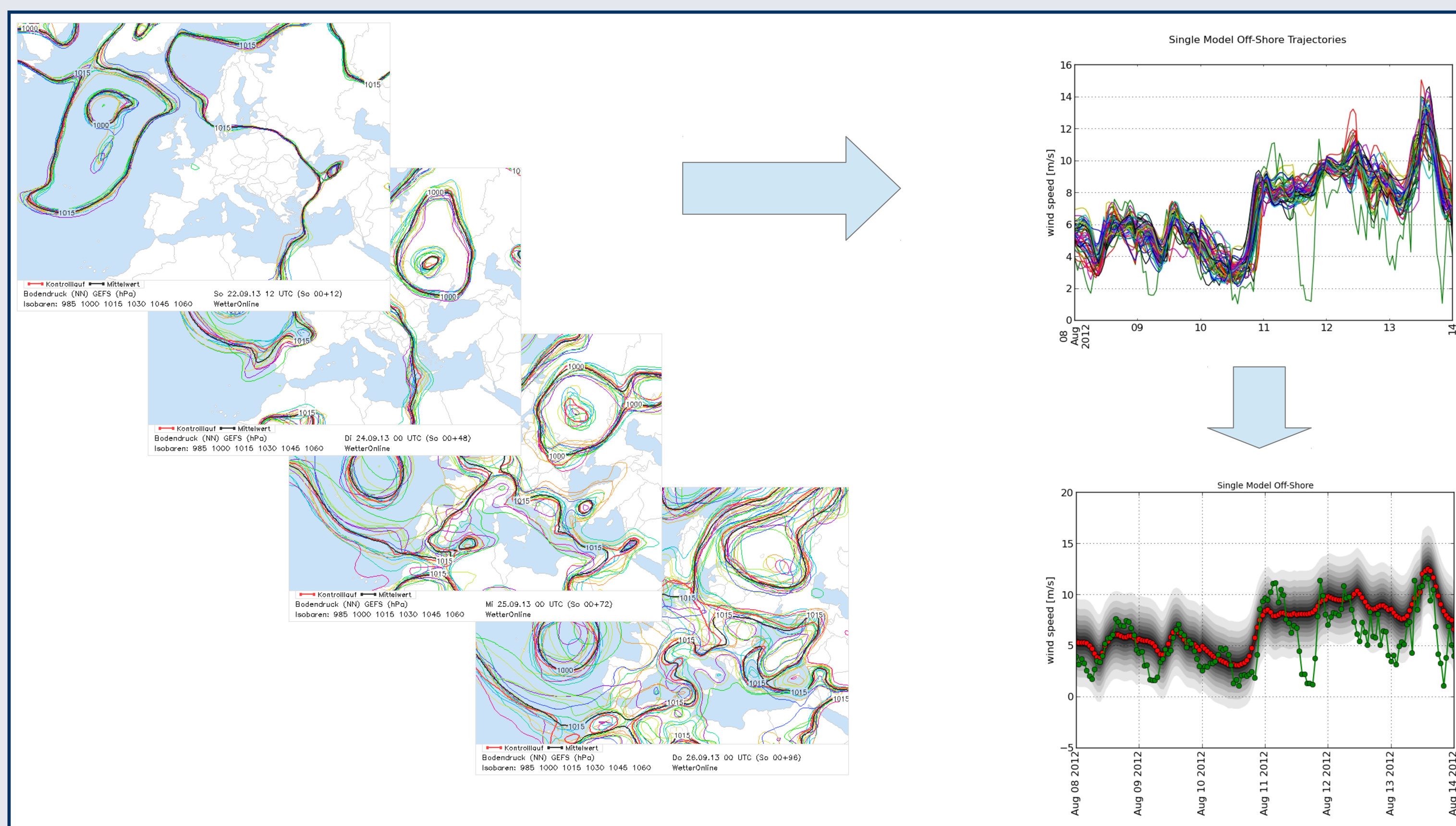
Jérémy Sack, Achim Strunk, Jon Meis, Frank Sehnke*, Martin Felder*, Anton Kaifel*
EWC Weather Consult GmbH, Karlsruhe, Germany,
+49 721 66323-0, info@weather-consult.com;
*Centre for Solar Energy and Hydrogen Research (ZSW), Stuttgart, Germany,
+49 711 7870-0, anton.kaifel@zsw-bw.de

EWC European Weather Consult

Abstract

Probabilistic forecasting has become state-of-the-art in wind energy prediction. Several studies have shown that probabilistic forecasts not only provide more extensive information about future weather development, but also allow players on the wind energy market such as grid operators, risk managers or energy traders to save valuable resources.
The use of ensemble Numerical Weather Predictions (NWP) has turned out to be an elegant way of representing the forecast uncertainties depending on prevailing weather situations. It is now common practice to derive predictive densities from the ensemble's set of trajectories and to maximize the skill through a calibration with measurements. Although expensive in terms of computing resources and money, ensembles including a large set of members have been preferred in recent studies.
The present study proposes a way of generating probabilistic forecasts obtained from a small multi-model-ensemble offering nearly similar skills as a large single-model ensemble consisting of 50 member.

Objective



Data

NWP Models:

Single Model Ensemble

- ECMWF EPS
- 50 member
- 1 control run
- Variations in initial conditions
- Interchangeable members

Multi Model Ensemble

- 2 Global Models:
  - GFS
  - CMC
- 2 Regional Models:
  - WRF
  - Hirlam

Non-interchangeable memers

Observations:

3 different locations:

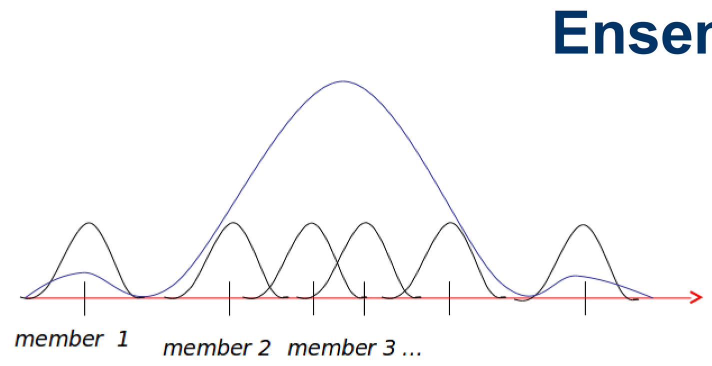
- Off-shore wind park
- On-shore wind park flat terrain
- On-shore wind park complex terrain

Time range:
1 year (2012) ...

- … including a rolling training set of 1.5 Months
- ... and a test set of 10.5 months

Method

Ensemble Calibration



The probability distribution at time t is given by

$$P_t = \frac{1}{n}\sum_{i=1...n} w_i \cdot N_i(f_k, \sigma^2)$$

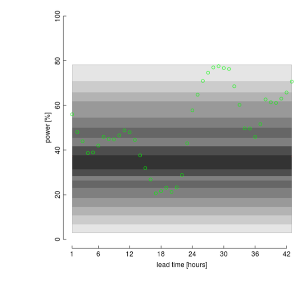
... where n is the number of ensemble members, w the weights and N the Gaussian distribution fully described by mean and standard deviation.

Probabilistic Forecasts originating from NWP Ensembles increase in skill when calibrated with observations. One possible method for this is the so called kernel dressing. For each time step t, a parametric probability distribution, in the present case a Gaussian distribution, is assigned to each member. A weighted average over all kernels then yields a non parametric distribution that can take any shape. The critical parameters within this method are the standard deviations and weights attributed to the different members. They are retrieved from a training set of approximately 1.5 months.

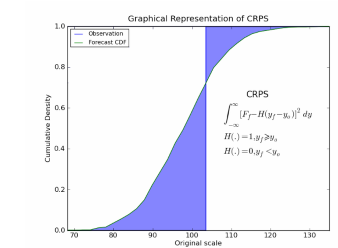
Evaluation of probabilistic forecasts

Since classic evaluation methods for deterministic forecasts can't be applied to probabilistic forecasts, the latter require particular evaluation tools:

- Reliability: Comparison between forecast quantiles and observed Quantiles. Although this is a very obvious and intuitive measure, it can easily be cheated on as indicated in the graph to the right.



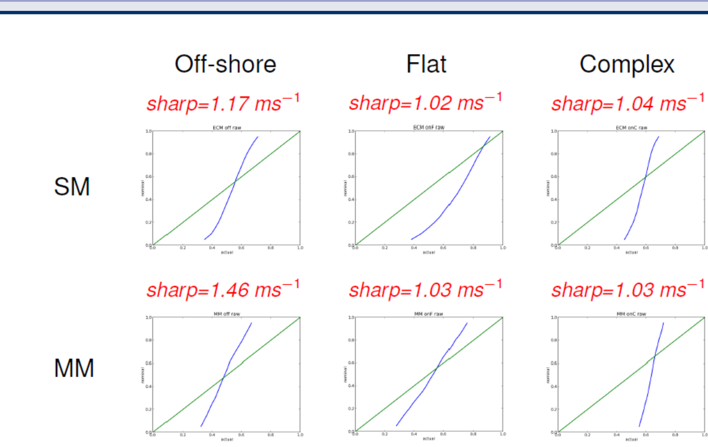
- Sharpness: The mean width of certain prediction intervals. In the present case, P75 – P25 was used
- Continuous Ranked Probability Score (CRPS): A robust statistical score for probabilistic forecasts that can be interpreted as the shaded area between the heaviside step function at the observation value and the cumulative distribution function of the probabilistic forecast as indicated in the graph to the right.

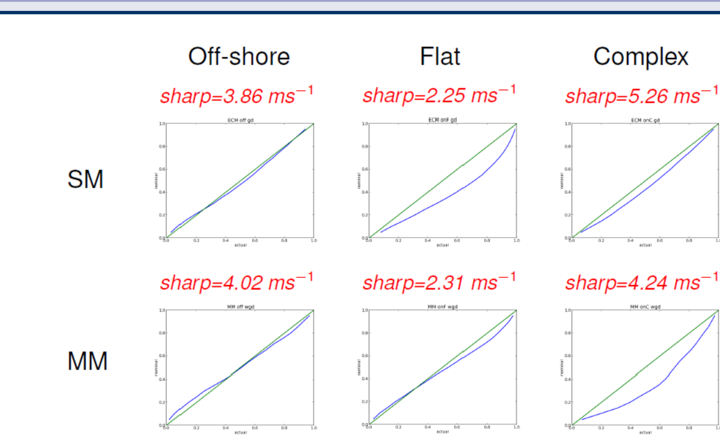


Optimization

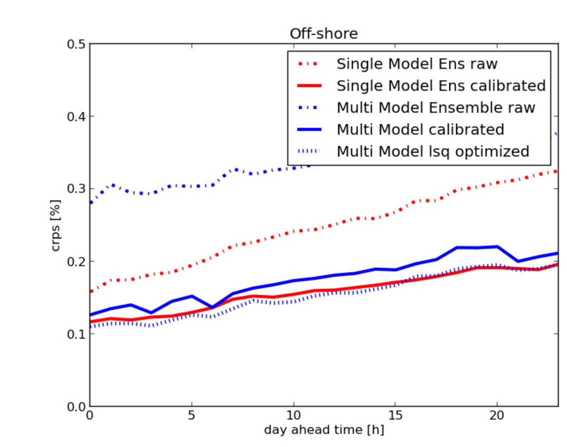
The weights used within the ensemble calibration method are optimized within a least squares framework in order to reduce the CRPS. As the single model's members are interchangeable, this is only possible for the non-interchangeable members of the multi model ensemble. In this way, the specific characteristics with respect to model physics, initial conditions and seasonal/local performance of each different model are accounted for.

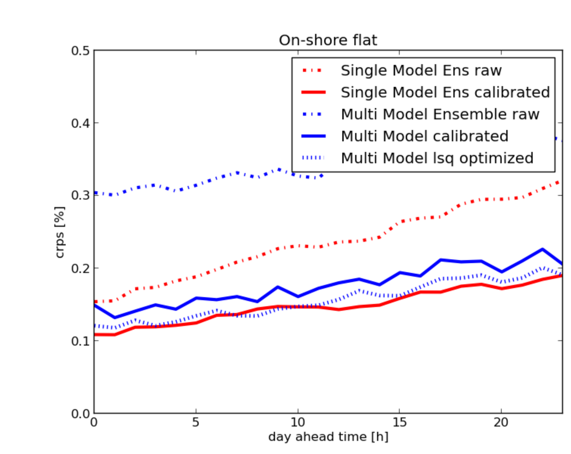
Results

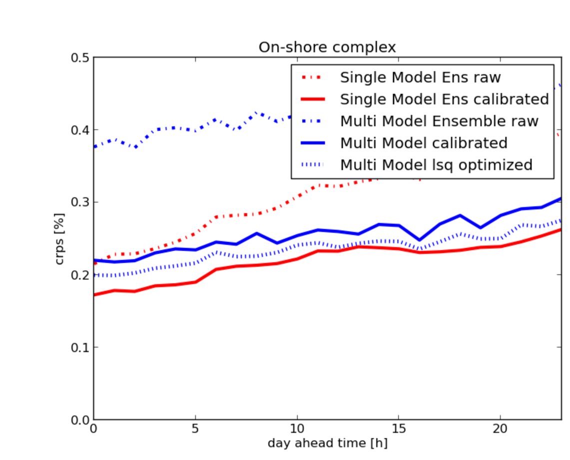




The top figures show reliability diagrams and sharpness for uncalibrated (left) and calibrated (right) forecasts for all locations. While in all cases, the calibration drastically ameliorates the reliability (the diagonal would be a perfect reliability), it also increases the sharpness values due to the addition of spread originating from the kernels assigned to each member – which is a well known trade off.
The CRPS graphs to the right offer the possibility to evaluate whether or not this trade off signifies a gain in quality. As the multi model ensemble consists of only 4 members, its raw (uncalibrated) forecasts show rather low quality (high CRPS values) compared to the single model ensemble.
After calibration, both ensembles significantly gain in quality, whereas the single model ensemble outperforms the multi model ensemble for all forecast horizons.
While after the calibration, the single model ensemble can't be further optimized due to the interchangeability of its members, the least squares optimized weighted multi model probabilistic forecast (lsq optimized) quality comes remarkably close to the single model one. For the off-shore location, it even outperforms the single model ensemble for a large portion of forecast horizons.
Also note here that the probabilistic forecast quality is higher in the off-shore and on-shore (lowest CRPS value ~ 10%) flat terrain locations than in complex terrain (~18%).







Conclusions

Two different NWP ensembles, an interchangeable single model ensemble and a multi model ensemble have been used to produce day ahead wind forecasts for 3 different wind park locations. The quality of both forecasts could be increased through a Gaussian kernel dressing technique. In a first step, the parameters used for the calibration have been retrieved from a rolling training window of 1.5 months. This significantly increased the probabilistic forecast quality for both ensembles.
In a next step, the weights accorded to the different members of the multi model ensemble have been adapted through a least squares optimization, increasing the forecast quality of the multi model ensemble consisting of 4 members nearly to the level of the single model ensemble consisting of 50 members.
Seen the difference in computational cost between both methods, we find this a very promising result and will continue carrying out research on this topic.

EWEA THE EUROPEAN WIND ENERGY ASSOCIATION

EWEA Wind Power Forecasting Technology Workshop, Rotterdam, 3-4 December 2013

Call4Posters