Super-ensemble techniques applied to wave forecast: performance and limitations

F. Lenartz1, J.-M. Beckers, J. Chiggiato, C. Troupin, L. Vandenbulcke, and M. Rixen

General evaluation of the paper

The scientific approach is good, some sections of the paper are well written but other sections are not accurate enough. In my opinion the paper should be accepted but only after major revisions also following my comments listed in the following.

Major comment on the figures.

- The quality of the figures is bad; too small and almost impossible to be read. In particular the quality of figures 4, 5, 8 and 9 must be substantially increased.

Section 2.1 Data

- Some references describing the characteristics of the two types of measurements (in situ or via satellite-borne remote-sensors) would help readers which are not working on this specific subject.

Section 2.2 Wave forecasting systems

- The main differences between WAM and SWAN should be mentioned when introducing the two models.

- SWAN ARPA (SA): the authors should specify that the operational output is available every hour even if it has been used with a three hour frequency.

- A reference for the atmospheric model cited in the text would be good; at least a link to web sites (SKIRON, ECMWF/IFS, ALADIN, ARPEGE,) The resolution of IFS should be written explicitly when saying *it is coarse.*

- "Lokal Modell" is now named "COSMO model". The reference Steppeler et al 2003 is correct but the authors could also add a link to the COSMO web site (www.cosmo-model.org).

- The meaning of HOTSTART should be briefly described.

Section 3 Methods

- EM - Ensemble Mean.
  Since in this case the Learning period is useless, formula 1 can be misleading.

- UEM – Unbiased ensemble mean
  The notation used to express the unbiased Ensemble Mean (equations 3 and 4), even if formally correct, can also be misleading. Using different notation during the learning and testing periods would be better.
  E.g., as regards observations, $\overline{y_h} and \overline{y_f}$ could be used instead of the same $\overline{y}$.

- KF Kalman filter
  Major comment:
  - As I know P should be defined as the "weight error covariance matrix".

Major comment:

- In formulas 11 and 12, and in the definition of Kj, I think that instead of $x_{j,i}$ the author should write $x_{j.}$

Section 4 Results

Major comment:

- The authors say:
  ……..In order to compare model outputs with observations, we performed spatial (inverse distance) and temporal (linear) interpolations. These time series are presented in Figs. 4 and 5. ..

Which is the time range of the forecasts at the different hours? Do they concatenate forecast from +?? to +??? ?

Major comments:

- The authors say:
  In order to test and validate the methods, the following procedure is applied: at each station and for each campaign, we consider the time series for which the four forecasting systems outputs are available. Then, we split them into overlapping bins of 2-day every 6 h, in order to virtually increase our dataset. The first half of each bin constitutes the learning period, the second half constitutes the testing period.

Do they take 48 hours of forecast from each model? Do these forecast have the same forecast range?
Which is the forecast ranges they use for the learning period?

The use of the forecast is not explained enough; this paragraph must be rewritten with more details. Figure captions must be also rewritten accordingly.

Again!

The authors say:
…… discarded, in order to improve the understanding of the general behaviour of our methods, we also present the results relative to a 2-day learning period and a 2-day testing period at Ortona.

How they combine each of the forecasts for this 2days+2days learning-testing?
Nothing said about this.

In the resuming comments of session 4, the authors mention the negative effect of an abrupt change in the time series of the model output but they do not explain when this happens in their application.

Section : Conclusions

Major comment:
The authors say

> Eventually, we wish to develop a SE technique that would automatically select the interesting features represented among available models, would combine them and create a physically-consistent forecast field.

I found impossible to understand what they are referring to. Authors should rephrase this sentence to give at least an idea about the approach they are referring to.

Acknowledgements.
ARPA SIMC should also be included in the acknowledgments.