

## ***Interactive comment on “Super-ensemble techniques applied to wave forecast: performance and limitations” by F. Lenartz et al.***

**F. Lenartz et al.**

fabian.lenartz@gmail.com

Received and published: 10 May 2010

We would like to thank referee #1 for his remarks. They have highlighted the lack of information in some sections and made the general improvement of the paper possible. We hope that our answers meet his expectations.

### **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.*

We have changed the font and increased its size in all figures. In addition, for figures

C220

4, 5, 8 and 9, we have also increased the size of the symbols and the thickness of the lines. The new figures are included in the revised version of the paper.

### **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.*

According to your comment, we have added a reference to a chapter of a book dedicated to measurement techniques:

Stewart R.H. 1980. Ocean wave measurement techniques. In: Air Sea Interaction, Instruments and Methods. Edited by L. H. F. Dobson and R. Davis. 447–470. New York: Plenum Press

and one dedicated to a comparative study using satellite altimeters and ocean buoys: Hwang, P. A., W. J. Teague, G. A. Jacobs, and D. W. Wang (1998), A statistical comparison of wind speed, wave height, and wave period derived from satellite altimeters and ocean buoys in the Gulf of Mexico region, J. Geophys. Res., 103(C5), 10,451–10,468, doi:10.1029/98JC00197

### **Section 2.2 Wave forecasting systems**

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

One introducing paragraph has been added for each model:

The WAVE Model was originally designed for modeling waves in the deep ocean or in intermediate depth water. However, in the course of time it has been adapted for simulations in shallow water by the use of a shallow-water phase speed in the

C221

expressions of wind input, a depth dependent scaling of the quadruplet wave-wave interactions, a reformulation of whitecapping in terms of wave number rather than frequency and the addition of bottom dissipation. WAM was the first wave model to use the discrete interaction approximation to calculate non-linear transfers of energy by quadruplet wave-wave interactions. Besides, it accounts for the effects of shoaling and refraction due to spatial variations in bottom and current and can also simulate blocking and reflection when waves propagate against the current. Still, WAM cannot be realistically applied to coastal regions with water depths less than 20-30 m. Regarding the numerics, WAM uses a different discretization scheme for the integration of the source functions and the calculation of the advective terms of the action balance equation: the source functions are computed with a fully implicit scheme, while two alternative explicit propagation schemes are implemented to calculate the advection terms, a first-order upwind scheme or a second-order leapfrog scheme.

The Simulating WAVes Nearshore model was developed to compute short crested waves in coastal regions with shallow water and ambient currents. Two rather important processes in coastal environment were added with respect to WAM: depth induced wave breaking and triad wave-wave interactions. Similarly to WAM, SWAN incorporates the effects of shoaling, refraction, blocking and reflection due to currents and variations in bathymetry. Concerning its numerical implementation, SWAN uses an implicit propagation scheme based on finite differences, which is unconditionally stable and more suited for small-scale, shallow-water and high-resolution computations. This scheme allows for relatively large time steps because it is only limited by accuracy. The drawback of this implicit scheme is that it is fairly diffusive for long propagation distances (oceanic scales).

*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.*

Since the fields that we have been provided with for the DART campaigns may differ from the present products, we have mentioned that such a difference may exist. In

C222

addition, we have provided a link to the webpage of these operational wave forecasting systems, so that interested people can check the details of their most recent version. We have added the links to the wave and weather forecast system websites:

WAM ATHENS

<http://forecast.uoa.gr/waminfo.php>

WAM ISMAR

[http://ricerca.ismar.cnr.it/MODELLI/ONDE\\_MED\\_ITALIA/page-html/nettur](http://ricerca.ismar.cnr.it/MODELLI/ONDE_MED_ITALIA/page-html/nettur)

SWAN ARPA

<http://www.arpa.emr.it/sim/?mare&idlivello=72>

Please also note that the operational version of SWAN NRL in the Adriatic Sea was only temporarily run in order to support the DART campaigns, thus we can not refer to a website.

*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.*

Links to the websites have been added and the horizontal resolution of the ECMWF/IFS has been precised.

SKIRON

<http://forecast.uoa.gr/forecastnewinfo.php>

ECMWF/IFS

<http://www.ecmwf.int/research/ifsdocs/CY31r1/WAVES/IFSPart7.pdf>

COSMO

<http://www.cosmo-model.org>

ALADIN CROATIA

<http://www.meteo.hr/>

*Lokal Modell is now named COSMO model. The reference Steppeler et al 2003 is*

C223

correct but the authors could also add a link to the COSMO web site ([www.cosmo-model.org](http://www.cosmo-model.org)).

The link to the webpage of COSMO has also been added.

*The meaning of HOTSTART should be briefly described.*

The sentence is now:

Each simulation starts with a *hotstart field*, an initial wave field derived from a previous run, and is run twice a day (...).

### Section 3 Methods

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

Since the method does not require any learning period, Eq. 1 might be misleading but we would prefer to keep it in order to be systematic in our presentation.

*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.*

We have considered adding superscript  $^h$  and  $^f$  for denoting the hindcast and forecast of  $x$ , however it also seems to be confusing for some people, so that we prefer to keep our original notation. Regarding the  $y$  vector, a clarification is required: actually the method can be qualified "unbiased" only at hindcast. Since in an operational framework we do not have the observations for the forecast period, the method implicitly assumes that there is no radical change of the system between both periods and the same averaged value of observations available during the learning period is used for the forecast.

C224

*KF Kalman filter. As I know  $P$  should be defined as the "weight error covariance matrix".*

You are right, we have changed it.

*KF Kalman filter. In formulas 11 and 12, and in the definition of  $K_j$ , I think that instead of  $x_{j,i}$  the author should write  $x_j$ .*

Again you are right, we have made modifications accordingly.

### Section 4 Results

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

We have modified the last sentence of the 1<sup>st</sup> paragraph of Section 4 - as well as a part of the 2<sup>nd</sup> and the 4<sup>th</sup> paragraphs, see next two remarks - in order to better explain our experiments:

The time series of the concatenated daily first 24-h of forecast of each model are presented in Figs. 4 and 5.

*(...) 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.*

As suggested, we have re-written the paragraph:

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 days every 6 h, in order to virtually increase our dataset. The first half of each bin

C225

constitutes the learning period, the second half constitutes the testing period. Since three models provide a forecast starting at midnight of each day, while the 4<sup>th</sup> one provides a forecast starting at noon of each day, a 12-h time-shift exists in the forecast range between them. Hence, in our experiments, at midnight of a hypothetical day 4, the 24-h learning period consists of the 0-24 h forecast of day 3 for SA, SN and WA, while for WI it consists of the second 12-h range of forecast of day 2 followed by the first 12 h of forecast of day 3. The corresponding 24-h testing period consists of the 0-24 h forecast of day 4 for SA, SN and WA, while for WI it consists of the second 12-h range of forecast of day 3, followed by the first 12 h of forecast of day 4. For both the learning and testing periods the bias, the linear correlation coefficient and the root-mean-square difference (RMSD) of the forecasting systems outputs, as well as of the SE techniques outputs, are computed.

We don't know how to change the figure captions in an elegant way, but think that the details provided in the text might be sufficient.

*(...) 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.*

One sentence has been added in the 4<sup>th</sup> paragraph of Section 4:

*(...) Due to the small number of measurements during a 1-day learning period, at least at the GS1, GS2 and A20 stations, previous conclusions are somewhat biased. Though they should not be discarded, in order to improve the understanding of the general behavior of our methods, we also present the results relative to a 2-day learning period and a 2-day testing period at Ortona. For this experiment the learning period consists of two successive one day learning periods, as presented previously, whereas the testing period consists of the 48-h forecast of the considered day for SA, SN and WA, and of the second 12-h range of forecast of the previous day followed by the first 36 h of forecast of the considered day for WI.*

C226

Regarding your second point, we actually met the problem of an abrupt change when we tried to use the previous day model outputs to fill in the gaps in the time series when model outputs were not available, but not really in the experiments presented here. However we think the remark is still worth being mentioned.

## Section Conclusions

*(...) 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. We know that this sentence may sound a bit like science-fiction but let us explain our point of view - or at least mine, since I am not speaking on behalf of the co-authors for this particular remark. Presently, the SE method adapts one weight per model in order to best fit the local observations available during the hindcast period and assumes that the behaviour of the system will not significantly change at short-term. However, nothing prevents it to predict negative values for instance (though we haven't faced such a situation in all our test cases). Hence, to make sure of the physical-consistency of the forecast, an a posteriori constraint would be desirable. In future applications, one could have a network of weights, one per model at each buoy and along each satellite track, allowing for a correction over a larger area and hopefully, pointing out the weakness and strength of each model. This is what I meant by *select the interesting features represented among available models*. I have modified this sentence as follows:*

Eventually, we wish to develop a SE technique with an increased number of weights per model, which would allow an automatic selection of the best features represented by the available models, and would combine their outputs to create a physically-consistent forecast field presenting a lower RMS error than any individual model.

If you consider this sentence useless I may delete it, for I don't want it to take too much importance.

C227

## **Acknowledgements**

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

The reference to ARPA SIMC was removed by mistake from the acknowledgements section, when Jacopo Chigiato earned is co-authorship. Thanks to your comment this error has been corrected.