

Better Baltic Sea Wave forecasts: Improving resolution or introducing ensembles?

Reviewed at iteration: Revised Submission

Recommendation: paper is accepted subject to minor/technical revisions.

Generic remarks: Thanks a lot for the authors for taking on board the previous review comments. In particular the improved documentation of the atmospheric forcing used by the wave models has made the paper and its conclusions a lot clearer. I would be happy to see this published without a further review subject to some technical (written English) corrections suggested below. However, I'd also like the authors to consider some potential minor revisions to the manuscript.

Potential minor revisions:

Section 2: Please confirm in the text whether any of the source term tuning parameters are different or the same in the various wave configurations. Previous review comments had also asked for the WAM version(s) to be documented.

Section 2: Is it possible to add some justification as to why the authors might consider the 3km atmospheric system to offer a step change in wind forcing skill over the 5km configuration that underpins the ensemble – this is not an enormous resolution change, so which processes are believed to be improved? Ideally provide references for the atmosphere systems in question.

Section 5: Not suggesting that additional results/tables are presented – but is it possible to comment on the performance of the S05 atmosphere model's control member performance? Is it similar to, better or worse than S03? This would help put the effect of the ensemble into context.

Section 6: As per section 5, if it is possible to comment on performance of S05's control member relative to the LOW wave model that will help contextualise the ensemble.

Figure 9: Is it possible to show the range of LOWENSMEAN forecasts – it would be good to see if any of the ensemble members predicted the peak in SWH.

Section 7 (or 8): I'd still be keen to see a little more discussion on the systems, what has been verified and what that means in terms of application. Now that the source of wind forcing has been stated more explicitly, what we have are: LOW – a system with 3km winds integrated onto a 10km wave grid; HIGH – 3km winds integrated onto a 5km grid; and LOWENS – a system with 5km winds integrated on a 10km grid but run as an ensemble in order to sample uncertainty in the atmosphere model and evolution of the meteorological conditions. Validation is carried out on a site specific basis. So, the important points to consider are:

1. Whilst the validation is a very standard approach for wave models, it is actually documenting site-specific forecast performance for the model. This is an important consideration when testing high resolution systems, where 'double-counting errors' start to get introduced and site specific verification does not improve as a result. On the other hand, the majority of forecast use for these models are site-specific, so the results are absolutely valid when one considers the products that get generated from these models.
2. The comparison between LOW and HIGH suggest that integrating the 3km winds onto a coarser wave model grid has no impact in terms of the verification at offshore sites. Why might this be? Speculating, this is probably because wave development is a function of winds, but integrated over a longer fetch area – so the resolution of the atmosphere model is (within reason) perhaps less important than the model's ability to properly place major synoptic features. Figure 2 hints at this, since there appears to be little additional structure in the HIGH model field offshore compared to LOW.
3. The ensemble mean gives the best site-specific forecasts offshore, in terms of limiting the overall error. This is despite using a coarser resolved atmospheric model and the coarser wave model. This is consistent with the argument above that uncertainty large scale feature development is perhaps (generally) more important to wave forecasts in open waters than getting the momentum exchange associated with small scale atmospheric features correct. I'd suggest that argument is further supported if the LOWENS control member verification at the offshore locations is not significantly worse than for LOW and HIGH.

I appreciate it if points 2 and 3 feel too speculative for the authors to consider including in their discussion, but I would advocate some comment along the lines of point 1.

Section 8: The under-spread in the ensemble suggests that there is scope for improving that system. I'd suggest this is a valid conclusion.

Technical revisions:

Line 31-32: dissipation of the wave energy mainly occurs through internal processes, e.g. whitecapping.

Line 48-49: The equations of the NWP model are discretized on a horizontal grid with a certain spatial resolution, which influences the maximum spatial resolution of the wave model.

Line 52: Over time, technical development has increased available computational resources, making it possible to increase...

Line 75: ...modelled sea-surface temperatures (SSTs) by the NEMO...

Line 76-77: Introducing such coupling may demand a high horizontal resolution, in atmosphere, wave and ocean models, in order to describe the fluxes most satisfactorily.

Line 84: ...wind forecasts is in Section 5, whilst verification of a principle wave forecast variable, significant wave height (SWH), is presented in Section 6.

Line 123: Each forecast run...

Line 133: ...with characteristics identical to LOW, but using a parallel...

Line 167: ...the area with SWH above 6m extends further southward...

Line 170-171: Observed series of SWH from wave measurement sites in the Baltic Sea, obtained from the Copernicus Marine Environment Monitoring Service (CMEMS) database, are used.

Line 306: ...HIGH forecast has a significantly smaller under-prediction bias than the other forecast classes.

Line 364: The conclusions hold,...

Line 410: ...field approaches an ice-covered area,...

Lines 412-413: ...when dense enough, acts as a solid shield that effectively removes all local wave energy...

Line 414: ...thick enough for this to be approximately correct.

Line 445-446: ...there are no indications that a further increase of the WAM model will result in enhanced site-specific forecast performance.