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

Torben Schmith, Jacob Woge Nielsen, Till Andreas Soya Rasmussen, Henrik Feddersen

Danish Meteorological Institute, Copenhagen, Denmark

Correspondence to: Torben Schmith (ts@dmi.dk)

Abstract. The performance of short-range operational forecasts of significant wave height in the Baltic Sea is evaluated. Forecasts produced by a base configuration are inter-compared with forecasts from two improved configurations: one with improved horizontal and spectral resolution and one with ensembles representing uncertainties in the physics of the forcing wind field and the initial conditions of this field. Both the improved forecast classes represent an almost equal increase in computational costs. The inter-comparison therefore addresses the question: would more computer resources most favorably be spent on enhancing the spatial and spectral resolution or, alternatively, on introducing ensembles? The inter-comparison is based on comparisons with hourly observations of significant wave height from seven observation sites in the Baltic Sea during the three-year period 2015-2017. We conclude that for most wave measurement sites, the introduction of ensembles enhances the overall performance of the forecasts, whereas increasing the horizontal and spectral resolution does not. These sites represent offshore conditions, well exposed from all directions with a large distance to the nearest coast and with a large water depth. Therefore, the detailed shoreline and bathymetry is also a priori not expected to have any impact. Only at one site do we find that increasing the horizontal and spectral resolution significantly improved the forecasts. This site is situated in nearshore conditions, close to land, with a nearby island and therefore shielded from many directions. This study therefore concludes that to improve wave forecasts in offshore areas, ensembles should be introduced. For near shore areas, the study suggests that additional computational resources should be used to increase the resolution.

1 Introduction

Severe wave conditions affect ship navigation, offshore activities and risk management in coastal areas. Therefore, reliable forecasts of wave conditions are important for ship routing and planning purposes when constructing, maintaining and operating offshore facilities, such as wind farms and oil installations.

Waves are generated by energy transfer from surface winds that act on the sea. The energy transfer is determined by the *fetch* (the distance, over which the wind acts), and by the *duration* of the wind. For *deep water waves*, defined as the wave height being much smaller than the water depth, dissipation of the wave energy mainly occurs through internal processes, e.g. whitecapping.. For *shallow water waves,* defined as the wave height being comparable to the water depth, dissipation through bottom friction and through wave breaking over a shallow and sloping sea bed becomes important. Shallow water waves may also be refracted over a varying bathymetry Therefore, a correct and detailed description of the bathymetry is important for correctly forecasting waves in coastal areas and other shallow sea areas. Other factors with a potential effect on the development of waves include nonlinear wave-wave interaction, ocean currents, time-varying water depth due to variations in sea level, and sea ice coverage.

The Baltic Sea is connected to the world ocean through the Danish waters with shallow and narrow Straits (see Figure 1), and this allows virtually no external wave energy to be propagated into the area. The Baltic Sea consists of a number of basins with depths exceeding 100 m, separated by sills and water areas with more moderate water depths. Between Finland and Sweden lies an archipelago with complicated bathymetry on very small spatial scales. The wind is in general westerly over the area, and the most prominent cause for severe wind and wave conditions is low pressure systems passing eastward over central Scandinavia. Winter ice occurs in the northern and eastern parts of the Baltic Sea. There is no noticeable tidal amplitude or permanent current systems.

Short-term forecasting of surface waves is done by a wave model, forced with forecasted wind from an atmospheric numerical weather prediction (NWP) model. 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. The available computer resources limit the horizontal grid spacing, that can be afforded.

Over time, technical development has increased available computational resources, which traditionally have been used to increase the horizontal spatial resolution of the NWP and wave models. This allows for an improved description and forecasting of the synoptic and mesoscale atmospheric systems, including the details of the associated wind field. In addition, a more detailed description of the bathymetry improves the correct description of dissipation and refraction of waves, as argued above. Additional computer resources may also be used to improve the spectral resolution in the wave model. This includes the directional resolution and the number of frequencies included.

Increasing computer resources have also made ensemble NWP possible. The purpose of ensemble forecasts is to improve forecast skill by taking both the initial error of the forecast and the uncertainty of the model physics into account. Furthermore, ensemble forecast allows for probabilistic forecasts, identified as a priority for operational oceanography (She et al., 2016), and allows for quantifying forecast uncertainty. Ensemble wave forecast systems have been implemented at global scale (Alves et al., 2013; Cao et al., 2009; Saetra and Bidlot, 2002) and more regionally in the Norwegian Sea (Carrasco and Saetra, 2008), and in the German Bight and Western Baltic (Behrens, 2015).

From the above discussion it is evident that additional computer resources can be used in different ways to change the wave forecast setup, in order to increase the forecast quality. The purpose of the present study is to investigate the effect on the forecast quality of increasing the horizontal resolution and the spectral resolution vs. introducing ensemble forecasts. This will be done by verifying the DMI operational forecasting of wave conditions in the Baltic Sea in different configurations against available observations of significant wave height.

Increasing the horizontal resolution of the NWP-system may also lead to improved wind forecasts, due to in particular better descriptions of processes in extratropical cyclones. In these cases, where the wind field is strong and varying on a small spatial scale, wave forecasts may also be improved by running the wave model in a similarly high resolution.

This paper is arranged as follows. Section 2 describes the model and setup, Section 3 describes the observations used and the verification methodology is described in Section 4. Verification of DMI-HIRLAM wind forecasts is in Section 5, whilst verification of the significant wave height (SWH) is presented in Section 6. Results of the verification are discussed in Section 7 and conclusions made in Section 8.

2 Model and setup

The DMI operational wave forecasting system DMI-WAM uses the 3rd generation spectral wave model WAM Cycle4.5.1 (Günther et al., 1992), with one minor change of source term functions. To speed up wave growth from calm sea, the spectral energy has a lower limit corresponding to a wave height of 7 cm. It is forced by the regional NWP model DMI-HIRLAM and the global NWP model ECMWF-GLM. WAM solves the spectral wave equation, and calculates the wave energy as a function of position, time, wave period and direction. Derived variables, such as the significant wave height (SWH), are calculated as suitable integrals of the wave energy spectrum.

The DMI-WAM model system forecasts waves in a larger area than the Baltic Sea and therefore has a setup with two nested spatial domains of different geographical extent (see Figure 1): North Atlantic (NA) and North Sea/Baltic Sea (NSB), of which forecast results from the NSB-domain are analyzed in this study. The NA domain uses the JONSWAP wave spectrum for fully developed wind-sea (Hasselmann et al., 1973) along open model boundaries, while the NSB domain use modeled wave spectra from the NA domain at its open boundaries (one-way nesting).

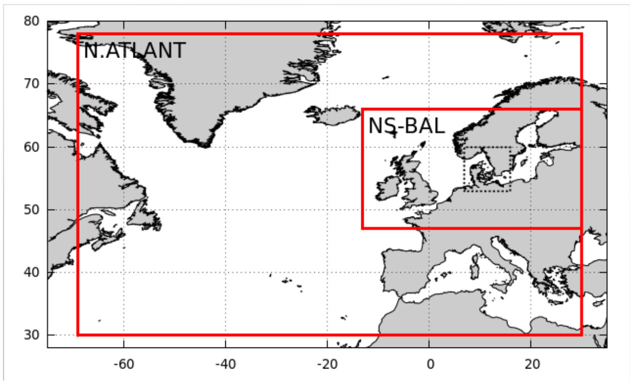

Figure 1 Nesting of domains in DMI-WAM. Outer frame is North Atlantic (NA) domain, inner frame is the North Sea/Baltic Sea(NSB)-domain. Dotted frame is the Transition Area. Only data from the NSB-domain are analyzed in this study.

The wave energy is discretized into a number of wave directions and frequencies. To facilitate wave growth from calm sea, a lower limit is applied to the spectral energy. The resulting surface roughness parameterizes the effect of capillary waves, and corresponds to a minimum significant wave height of 7 cm.

The energy source is the surface wind. The sink terms are wave energy dissipation through wave breaking (white capping), wave breaking in shallow areas, and friction against the sea bed. Depth-induced wave breaking (Battjes and Janssen, 1978) is used in the NSB domain only, since in the NA domain, the depth maps are not detailed enough for activation of this effect. The wave energy is redistributed spatially by wave propagation and depth refraction, and spectrally by non-linear wave-wave interaction. Interaction with ocean currents and effects due to varying sea level caused by tides or storms are not incorporated.

In addition to a land mask, we have a time-varying ice mask. Below ice 30% concentration, sea ice is assumed to have no effect. Above 30% ice concentration, no wave energy is generated or propagated, i.e. the effect is like that of land. The applied sea ice concentrations originate from OSISAF (http://osisaf.met.no/p/ice/) with a frequency of 24 hours and around 25 km true horizontal resolution, gridded to ~10 km horizontal resolution and interpolated to the WAM-grid. The ice cover is initialized every day at 00z, and kept constant throughout each forecast run.

The surface wind forcing is provided by different atmospheric models for the two domains. For the NA domain, wind is provided by the ECMWF-HRES global weather forecast every 3 hours. For the NSB domain, the surface wind is provided every hour by DMI-HIRLAM. Setup details are summarized in Table 1

Table 1 Specifications of DMI-WAM nested setup.

Domain	North Atlantic	North Sea/Baltic Sea
Longitude	69W-30E	13W-30E
Latitude	30N-78N	47N-66N
Atmospheric forcing	ECMWF-HRES	*DMI-HIRLAM*
Boundary condition	JONSWAP	One-way nested
Depth-induced wave breaking	No	Yes

Each forecast run is initialized using the sea state at analysis time, calculated by the previous run as a six hour forecast. The operational DMI-WAM suite is run four times a day to 48 h forecast range. This is also true for the North Atlantic domain, even when new forcing is available twice per day only. This is for practical reasons, since the North Atlantic domain is very cheap to run. Spatial fields of forecasted SWH and other variables are output in hourly time resolution.

Historically, three different configurations of the DMI-WAM setup have been used, and data from these for the period 2015-2017 is the basis for the present verification. In the old LOW configuration, the horizontal resolution is around 50 km in the NA domain and around 10 km in the NSB domain. The wave energy is resolved in 24 directions and at 32 frequencies, corresponding to wave periods between 1.25-23.94 s and wave lengths between 2.4-895 m (in deep water). Bathymetry is ETOPO (Amante and Eakins, 2009) in the NA domain, and the Baltic bathymetry from IOW (https://www.io-warnemuende.de/topography-of-the-baltic-sea.html) supplemented by depth data from the Danish Geodata Agency (DGA) in the NSB domain. More recently, an ensemble configuration (LOWENS) has been introduced with characteristics identical to LOW, but using a parallel run of 11 ensemble members forced with perturbed atmospheric fields (initial conditions and physics). Finally, in the recently introduced HIGH configuration, the horizontal resolution is around 25 km in the NA domain and around 5 km in the NSB domain The wave energy is resolved in 36 directions and 35 frequencies, corresponding to wave periods between 0.94-23.94 s, and wave lengths between 1.37-895 m (in deep water). Bathymetry is RTopo (Schaffer et al., 2016).

All configurations are forced by winds from ECMWF-HRES in the NA domain and DMI-HIRLAM in the NSB domain. In the NSB domain, the LOW and HIGH are forced by the S03 version (3 km horizontal resolution), while LOWENS is forced by the S05 version (5 km horizontal resolution). The S03 and S05 versions of DMI-HIRLAM were used operationally by DMI as deterministic and ensemble weather forecast models in the 2015-17 period. While the better resolution of S03 might have an impact on forecasts where orographic effects matter, the impact on wind forecasts over sea is expected to be insignificant. The DMI-HIRLAM winds are interpolated to the WAM grids by bilinear interpolation. To diminish coastal effects, DMI-HIRLAM delivers a special *water-wind* to DMI-WAM, in which the surface roughness everywhere is assumed to be that of water. This enhances the wind speed in the coastal zone, most important in semi-enclosed areas (bays, fjords, etc.). It is basically a way to sharpen the land/sea boundary, reducing influence of land roughness on near-shore winds. An overview of the DMI-WAM configurations is provided in Table 2.

Table 2 Details of DMI-WAM configuration used in this study.

	DMI-WAM Horizontal resolution [km]		# wave directions	# wave spectral frequencies	Bathymetry		Atmospheric horizontal resolution [km]		Ensemble members	
	North Atlantic	NSB			North Atlantic	NSB	North Atlantic (ECMWF)	NSB (DMI-HIRLAM)	North Atlantic	NSB
LOW	50	10	24	32	ETOPO	IOW/DGA	16	3	-	-
LOWENS	50	10	24	32	ETOPO	IOW/DGA	16	5	-	11
HIGH	25	5	36	35	RTopo	RTopo	16	3	-	-

When replacing the LOW forecast configuration with the HIGH configuration, the required computational resources for running DMI-WAM are increased by a factor of 2^2 (increase in horizontal resolution) × 1.75 (effective decrease in time step) = 7 due to higher spatial resolution, and by a factor of 1.5 (increase of number of directions) × 35/32 (increase of number of spectral frequencies) = 1.6. This gives a total factor of $7 \times 1.6 \approx 11.5$. From the LOW to the LOWENS configuration, it is increased by a factor of 11 (number of ensemble members). Since these increases in computational effort are very similar, an inter-comparison can contribute to answering the question: should additional computer resources be used for increasing the spatial and spectral resolution, or for sampling the uncertainty in meteorological conditions using ensembles.

The LOW and HIGH configurations both produce a class of deterministic forecast, which are also named LOW and HIGH, respectively. The LOWENS configuration produces a class of probabilistic forecast, called LOWENS. In addition, the ensemble mean defines a class of deterministic forecasts, called LOWENSMEAN.

To illustrate differences to be expected among the deterministic forecasts, we show 48 h forecasts of SWH valid at the peak of the ‘Toini’ storm on 10 January 2017.

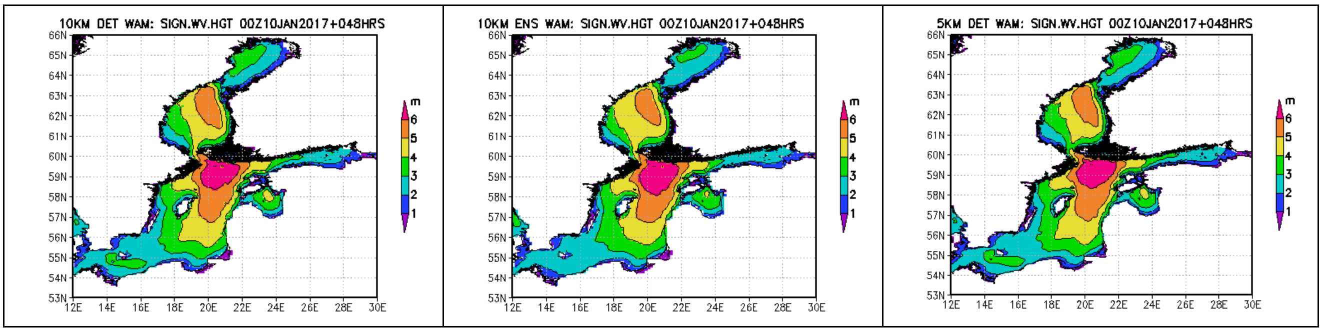

Figure 2 Forecasted (48h) SWH at the peak of 'Toini' storm 10 january 2017 00z for LOW (left), LOWENSMEAN (middle) and HIGH (right) forecasts.

All three forecasts agree in the gross features of the forecasted SWH field. However, there are differences, e.g., northeast of the island of Gotland, the area with SWH above 6 m extends further southward in the LOWNSMEAN forecast, than in the LOW and HIGH forecasts.

3 Observations

Observed series of SWH from wave measurement sites in the Baltic Sea, obtained from the Copernicus Marine Environmental Monitoring System (CMEMS) database, are used. None of the series has a continuous record over the three-year period 2015 – 2017. Data gaps may be due to malfunction, maintenance or withdrawal of the instrument. The latter occur during winter due to the possibility of ice. We selected sites with valid observations that covered more than 40% and were distributed reasonably throughout the study period. To avoid biases in the verification measures due to under- or overrepresentation of particular seasons, we also aimed at having an approximately even coverage throughout the year.

Figure 3 and Table 3 show the positions and water depths of the wave measurement sites together with the bathymetry of the Baltic Sea. Some sites did not observe at the full hour. Observations from these sites were ascribed to the nearest full hour, if the time distance between the observation time and the full hour was less than 15 min, otherwise not used. All observation series used are shown in Figure S1. The frequency of observed SWH in different intervals for each site is given in Table 4

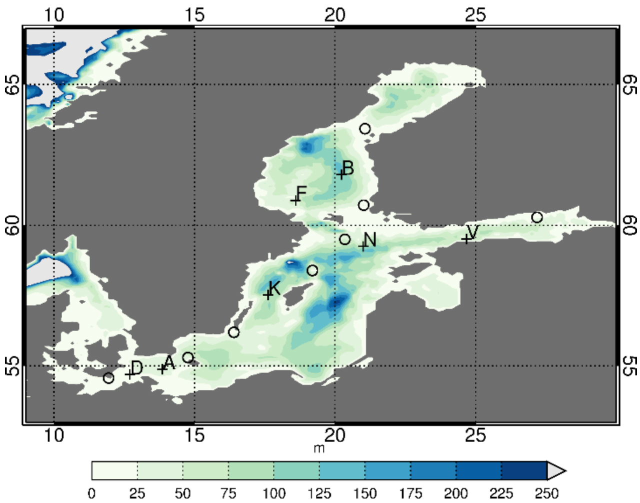

Figure 3 Map of the Baltic Sea with bathymetry and positions of wave measurement sites marked with crosses. For details about sites, see Table 3. Meterological stations used in the wind verification of DMI-HIRLAM are marked with circles.

Table 3 Details of wave measurement sites.

Observation site	Lon	Lat	Depth [m] Model	Depth [m] Actual
A Arkona WR	13.9	54.9	46	45
B Bothnian Sea	20.2	61.8	118	~120
D Darsser Sill WR	12.7	54.7	20	21
F Finngrundet WR	18.6	60.9	56	67
K Knolls Grund	17.6	57.5	63	90
N Northern Baltic	21.0	59.2	68	~100
V Vahemadal	24.7	59.5	18	5

Table 4 Observed frequency of SWH in different bins for wave measurement sites.

SWH [m]	0-1	1-2	2-3	3-4	4-5	>5
Arkona WR	0.47	0.39	0.12	0.01	<0.01	<0.01
Bothnian Sea	0.46	0.38	0.12	0.02	0.01	<0.01
Darsser Sill WR	0.67	0.31	0.02	<0.01	<0.01	<0.01
Finngrundet WR	0.69	0.27	0.04	0.01	<0.01	<0.01
Knolls Grund	0.62	0.31	0.06	0.01	<0.01	<0.01
Northern Baltic	0.39	0.37	0.18	0.05	0.01	<0.01
Vahemadal	0.78	0.20	0.02	<0.01	<0.01	<0.01

4 Verification methodology

In this section, a short overview of the verification procedure will be given. For background and more details regarding the verification measures, we refer to (Jolliffe and Stephenson, 2003)

For each measurement series of SWH, the corresponding forecast series for all forecast classes and for forecast range zero to 48 h for the grid point nearest to the position of the wave measurement site was extracted from the model output.

For the deterministic and continuous forecast classes (LOW, LOWENSMEAN and HIGH), we use the conventional performance measures *root mean square error* (RMSE), defined as the square root of the time average of the sum of squared differences between forecast and observation:

$$RMSE(\tau) = \langle \left(h^{\tau}_{s,fcst} - h_{s,obs}\right)^2 \rangle$$

the bias

$$BIAS(\tau) = \langle h^{\tau}_{s,fcst} - h_{s,obs} \rangle,$$

and the correlation coefficient

$$CC = \frac{\langle \left(h^{\tau}_{s,fcst} - \langle h^{\tau}_{s,fcst} \rangle\right)\left(h_{s,obs} - \langle h_{s,obs} \rangle\right) \rangle}{\sqrt{\langle \left(h^{\tau}_{s,fcst} - \langle h^{\tau}_{s,fcst} \rangle\right)^2 \rangle \langle \left(h_{s,obs} - \langle h_{s,obs} \rangle\right)^2 \rangle}}$$

where $h_{s,obs}$ is the observed SWH and $h^{\tau}_{s,fcst}$ is a corresponding forecast with forecast range τ .

The RMSE is a positive definite quantitative measure, and smaller values mean a better forecast. The bias can take positive and negative values, and a good forecast has a numerically small value. The averaging, indicated by $\langle \cdot \rangle$, is found based on all available values during the three-year period. Also, the RMSE and BIAS as function of $h_{s,obs}$ will be considered.

A framework for verifying probabilistic forecasts is the *continuous ranked probability score* (CRPS), defined as

$$CRPS(\tau) = \langle \int \left[F^{\tau}(h_s) - H(h_s - h_{s,obs})\right]^2 dh_s \rangle,$$

where $F^{\tau}(h_s)$ is the forecasted probability distribution, $h_{s,obs}$ is the observed value, and $H(\cdot)$ is the Heaviside step function. A small CRPS occurs when the median of the probabilistic forecasts are close to the observed values. Also a sharp probabilistic forecast with a small spread favors a small CRPS. This means that the best forecast is achieved when CRPS is small. CRPS can be applied to both the probabilistic forecast class LOWENS, as well as the deterministic forecast classes, LOW, LOWENSMEAN and HIGH, since these can be regarded as probabilistic forecasts with a step probability distribution. For the deterministic forecast classes, the CPRS equals the *mean absolute error*.

Besides the continuous and probabilistic forecasts, also the binary forecast of the SWH exceeding a specified threshold is considered. The performance measure used is the Brier Score, defined as

$$BS(\tau) = \langle (p - x)^2 \rangle,$$

where p is the forecasted probability with forecast range τ of exceeding the threshold and x takes the value of 1 or 0 dependent on whether the threshold actually was exceeded or not. The Brier Score is thus a positively definite measure, where values are between zero and one, and the lower the value, the better the forecast.

4.1 Calculation of confidence bands

All the measures described above are subject to sampling uncertainty; if they had been calculated on data from another time period than 2015-2017, they would have had different values. To estimate this sampling uncertainty and thereby obtain confidence bands, we applied a block bootstrapping procedure, where a large number of resampled series with the same length as the original series (three years) were created. A blocking length of one month was chosen. This choice takes the atmospheric decorrelation time scale of a few weeks into account and it allows a large number of different resampled series to be made.

Each resampled series is constructed as follows: The resampled series will contain three January months, and each of these is randomly chosen, with replacement, of the three January months from the original series. A similar procedure applies for February, etc. In this way, the resampled series are most likely different but the annual cycle is preserved. Both the observed series and the forecast series are resampled. For each pair of resampled series bootstrapped value of the performance measures are calculated. Repeating the resampling procedure, we obtain 1000 resampled values of the measures, from which their approximate statistical distribution and confidence bands can be calculated. As a standard, confidence bands (5/95%) are calculated by the bootstrap procedure described above and this allows for a quantitative inter-comparison of the performance measures for the different forecast classes: if the confidence bands do not overlap then there is a significance difference.

5 Verification of the wind forecasts

In order to illustrate the benefit of the meteorological ensemble on wind forecasts the S03 deterministic and S05 ensemble mean have been verified against available wind observations for eight coastal meteorological stations around the Baltic Sea (Figure 3). The RMSE of all stations for the period 1 Jan 2015 - 31 Dec 2017 is shown in Figure 4 as a function of forecast range. This reveals that the S05 ensemble mean is more accurate than S03, especially at the longer forecast ranges. Similar results are found for other verification scores, such as correlation and hit rate (not shown).

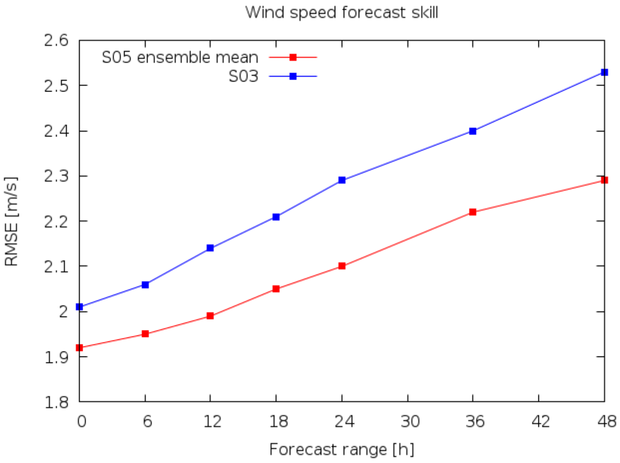

Figure 4 Verification of wind speed. Average RMSE between model and observations for eight coastal meteorological stations in the Baltic Sea area.

6 Verification of forecasted SWH against observations

6.1 Deterministic measures

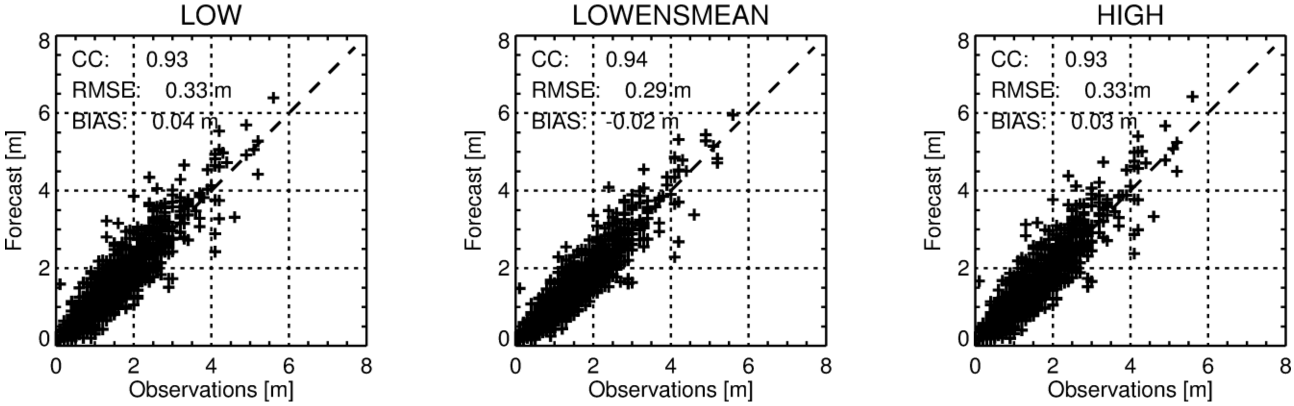

Figure 5 Scatter plot of 24 h forecasts and corresponding observations of significant wave height at site Bothnian Sea for the LOW, LOWENSMEAN and HIGH forecast classes. Dotted line is the diagonal, representing a 1:1 agreement between observations and model.

To get an idea of the overall quality of the forecasts, Figure 5 shows scatter plots between 24 h forecasted and observed SWH for station Bothnian Sea. The points are distributed along the diagonal in all three configurations with correlation coefficients above 0.9. The RMSE is 0.33 m for both LOW and HIGH but is lower at 0.29 m for the LOWENSMEAN forecasts, which also have the numerically lowest bias. Also for other sites, such as Arkona WR (see Figure 6), the RMSE for LOWENSMEAN forecasts is lower than for the LOW and HIGH forecasts, and similarly for the bias. However, the scatter plot appears differently for this station, because there is a tendency for over-predicting high waves for all three forecast classes.

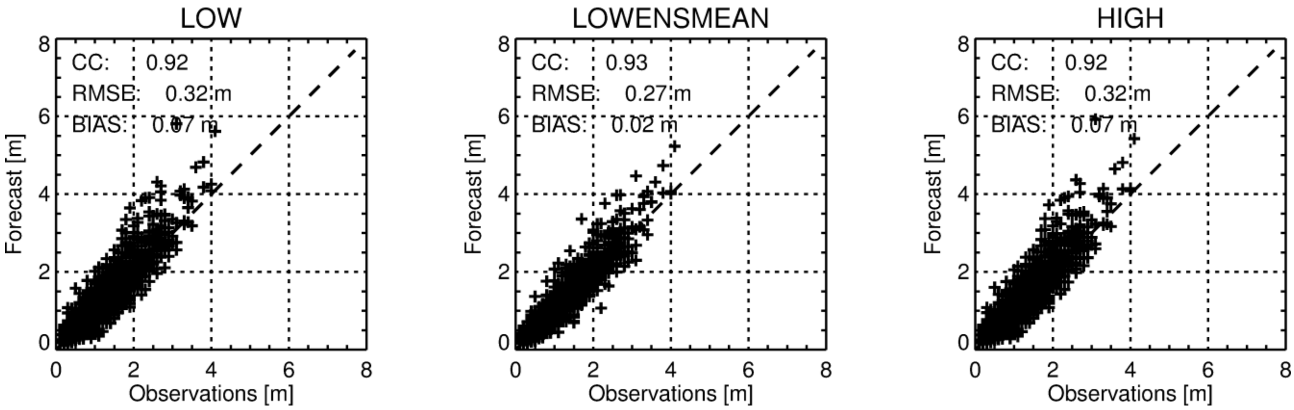

Figure 6 As Figure 5. Scatter plot of 24 h forecasts and corresponding observations of significant wave height at site Arkona WR for the LOW, LOWENSMEAN and HIGH forecast classes. Dotted line is the diagonal, representing a 1:1 agreement between observations and model.

We now turn to the RMSE as function of forecast range, of which plots for all sites can be found in Figure S2. For all sites, the RMSE increases slightly as function of forecast range. All sites except Vahemadal exhibit qualitatively similar behavior: the RMSE for the LOW and HIGH forecasts are almost similar, while it is lower for the LOWENSMEAN forecasts. Thus, for Arkona WR (shown in Figure 7), Bothnian Sea and Darss Sill WR, the RMSE of the LOW and the HIGH forecasts have overlapping confidence bands. The RMSE for LOWENSMEAN gradually diverges to a lower value (around 5 cm) and for large forecast ranges, the confidence bands do not overlap with those for the LOW and HIGH forecast classes. The remaining sites except Vahemadal behave similarly, but with overlapping confidence bands even for the largest forecast ranges.

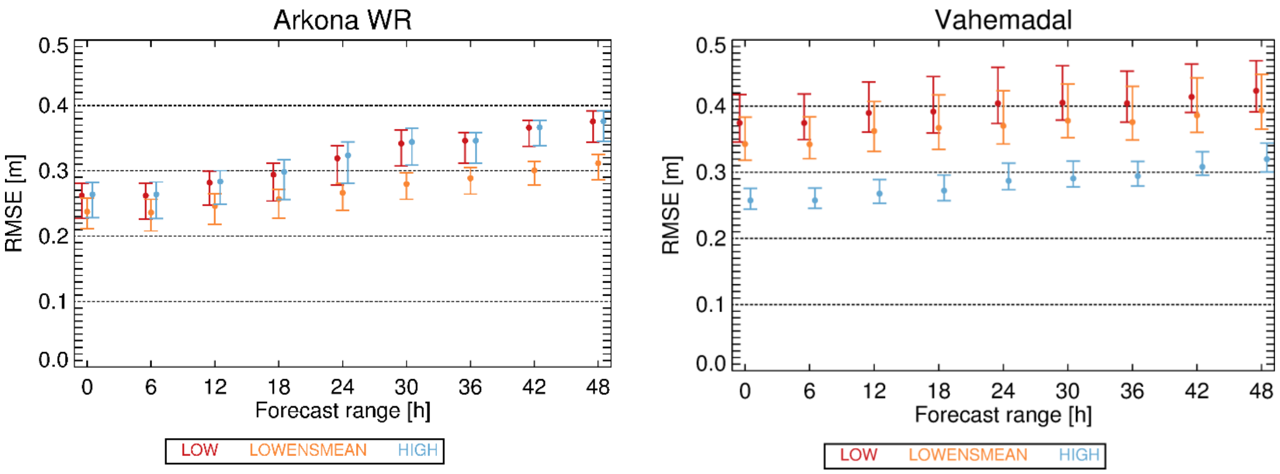

Figure 7 RMSE for selected forecast ranges for Arkona WR (left panel) and Vahemadal (right panel) for LOW, LOWENSMEAN and HIGH forecasts. Error bars show 5/95% confidence bands calculated by bootstrapping.

The site Vahemadal (Figure 7) has a different behavior. For this site, the HIGH forecast class has a significantly smaller RMSE and with non-overlapping confidence bands with the RMSE of the LOW and LOWENSMEAN forecasts. This site also has a non-negligible bias of around 12 cm for the HIGH and around 20 cm for the LOW and LOWENSMEAN forecasts; this bias is independent of forecast range (not shown).

6.1.1 Performance depending on observed SWH

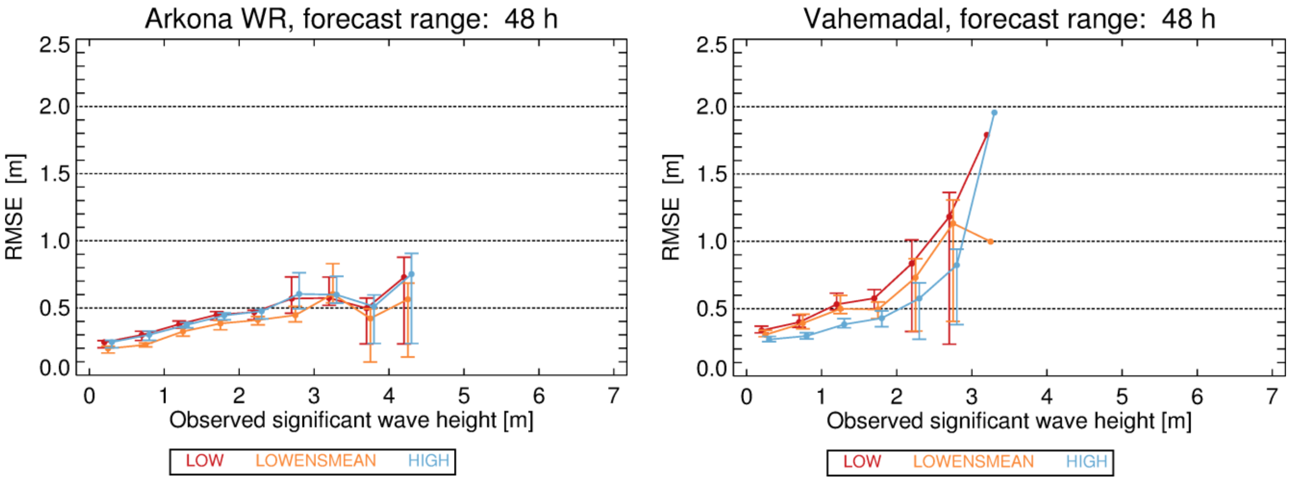

Figure 8 RMSE as function of SWH for Arkona WR (left panel) and Vahemadal (right panel) for LOW, LOWENSMEAN and HIGH forecasts and forecast range 48 h. Error bars show 5/95% confidence bands calculated by bootstrapping.

The RMSE of the forecasts depends on the magnitude of the SWH. Plots for all sites for the 24 h and 48 h forecast ranges of RMSE as function of the SWH can be found in Figures S3 and S4. The RMSE for Arkona WR and Vahemadal as a function of the SWH for the forecast range 48 h is shown in Figure 8. The RMSE increases as a function of the observed SWH for both sites. For Arkona WR, the LOWENSMEAN forecast class has the lowest RMSE, although with confidence bands overlapping with the other forecast classes. This behavior is seen at all sites, except Vahemadal. For Vahemadal, the HIGH forecast class has the lowest RMSE, and up to a SWH of 2 m, the confidence band is well separated from the confidence bands of the other forecast classes.

Also the bias depends on the SWH. Plots for all sites for 24 and 48 h forecast range of the bias as function of the SWH are displayed in Figures S5 and S6. For small SWH, the bias is close to zero for most sites. For some sites, the bias remains close to zero for increasing SWH, as shown for Arkona WR in left panel of Figure 9, while for others it becomes different from zero for large values of SWH. There is no noticeable different in the bias of the different forecast classes, except for Vahemadal, shown in right panel of Figure 9, where the HIGH forecast class has a significantly smaller under-prediction bias than the other forecast classes.

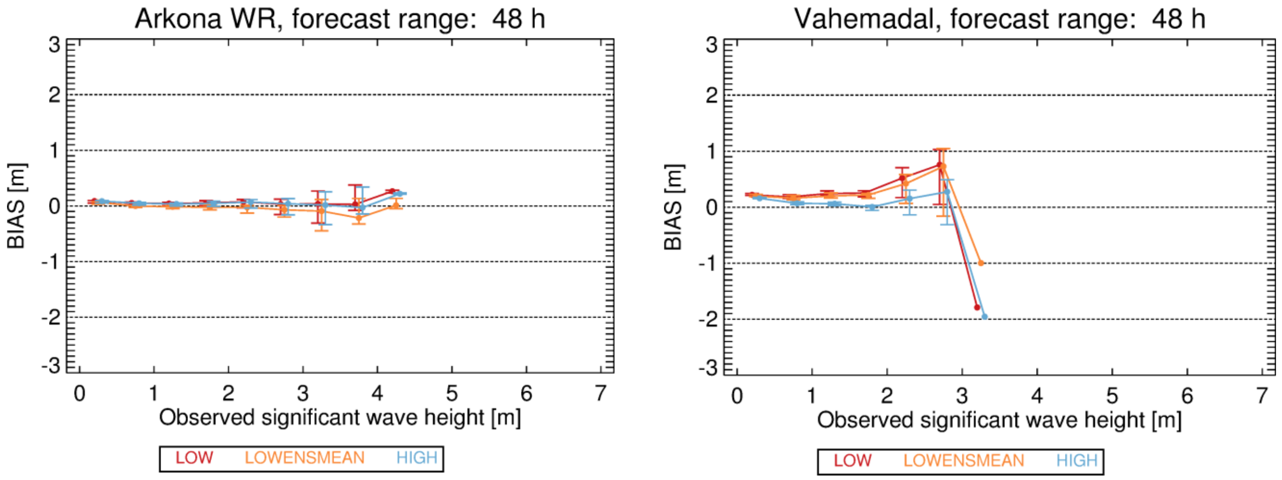

Figure 9 Bias as function of SWH for Arkona WR (left panel) and Vahemadal (right panel) for LOW, LOWENSMEAN and HIGH forecasts and forecast range 24 h. Error bars show 5/95% confidence bands calculated by bootstrapping.

6.1.2 Forecasts during 'Toini' storm

The Toini storm on 11. January 2017, where a SWH of 8.0 m was recorded at Northern Baltic (Björkqvist et al., 2017a), is within our verification period.

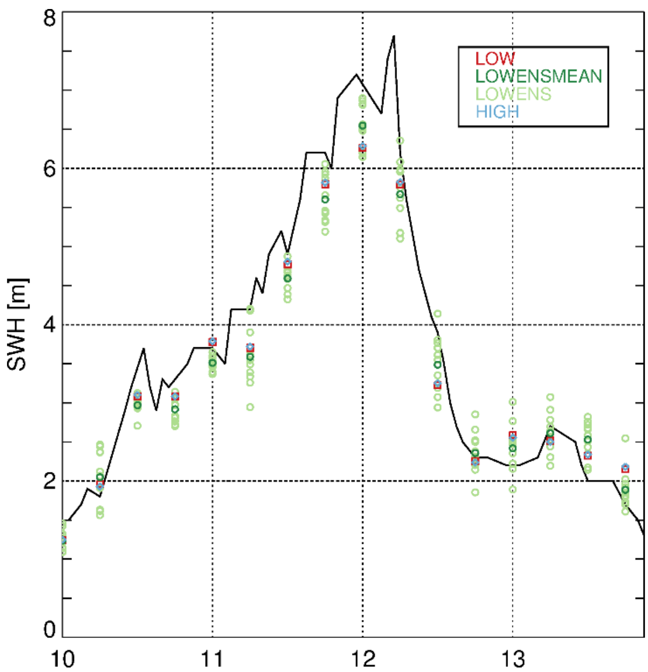

Figure 10 Observed SWH for Northern Baltic during, 10-13 January 2017, including the Toini storm. Open circles are 48 h forecasts.

Figure 10 shows the observed SWH at Northern Baltic during 10-13 January 2017, i.e. including the Toini storm, peaking in the early hours of 12 January, together with 48 h forecasts. In this case there is no apparent 'best' forecast. Near the peak, LOWENSMEAN performs best, but both before and after, the HIGH/LOW performs better. Furthermore, the LOW and HIGH forecasts are very similar in most cases, indicating that the higher resolution does not improve the forecasts. Finally, we note that the observations generally are within or just a little outside the range of the ensemble forecast.

6.2 Probabilistic metrics

The 11 ensemble members of the LOWENS forecast class defines a statistical distribution function, which is a probabilistic forecast of the wave conditions. The deterministic forecast classes LOW, LOWENSMEAN and HIGH may be regarded as probabilistic forecasts with probability one for the deterministically forecasted future state and probability zero for all other states.

As described in Section 4, we use CRPS to describe performance of probabilistic forecasts. CRPS for all sites for selected forecast ranges can be found in Figure S7. As typical examples, Figure 11 displays this plot for Arkona WR and Vahemadal.

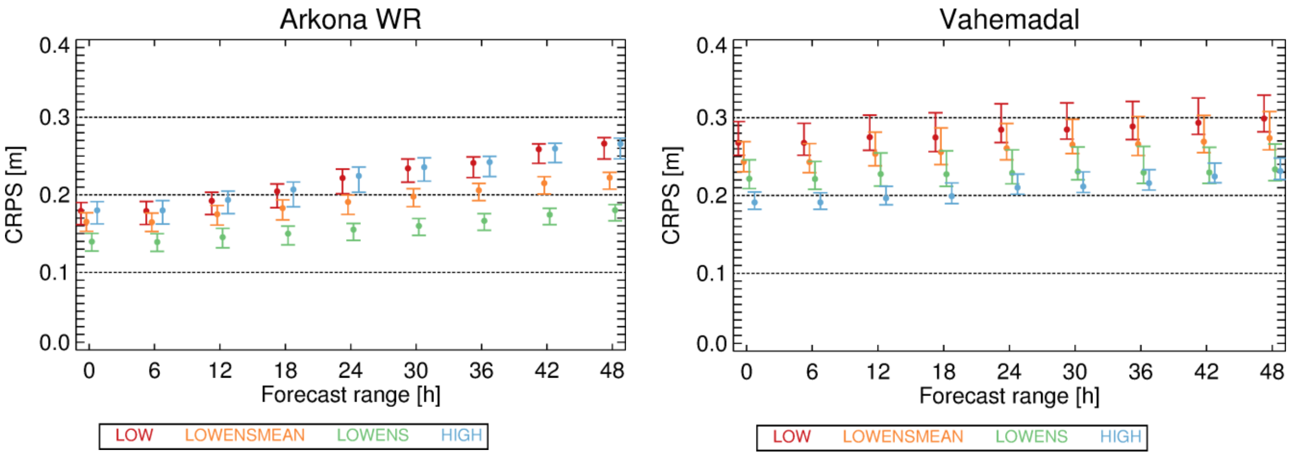

Figure 11 CRPS for selected forecast ranges for Arkona WR (left panel) and Vahemadal (right panel) for LOW, LOWENSMEAN, LOWENS and HIGH forecasts. Error bars show 5/95% confidence bands calculated by bootstrapping.

All sites except Vahemadal behave qualitatively as Arkona WR: the LOWENSMEAN forecast class has a lower CRPS compared to both the HIGH and LOW classes, although the difference is significant (non-overlapping confidence bands) for Arkona WR, Bothnian Sea and Darsser Sill WR only, and only for the largest forecast ranges. Furthermore, for all these sites, the LOWENS forecast class has an even lower CRPS, with confidence bands separated from those of all other forecasts classes. Again, Vahemadal behaves differently; here the HIGH forecast class has the best performance in terms of CRPS. However, for large forecast ranges, the LOWENS forecast class tends to perform equally well.

6.3 Binary forecasts

For the probabilistic LOWENS forecast class, a binary forecast can be derived as the probability of exceeding a defined threshold of SWH. For the deterministic forecast classes: LOW, LOWENSMEAN and HIGH, this probability of exceedance is either zero or one. As described in Section 4, the Brier Score is used as performance measure for probabilistic, binary forecasts.

The Brier Score as a function of threshold is shown for all sites in Figures S7 and S8. Figure 12 shows the Brier Score as a function of threshold for Arkona WR and Vahemadal for 48 h forecast range. For Arkona WR, the Brier Score for the LOWENS forecast class is the smallest, however the confidence intervals overlap with confidence intervals from the other forecasts above the 2 m threshold. Also the LOWENSMEAN forecast class has a low Brier Score. This behavior is common to all sites except Vahemadal. For Vahemadal, the Brier Score is smallest for the HIGH forecasts for thresholds above 1 m.

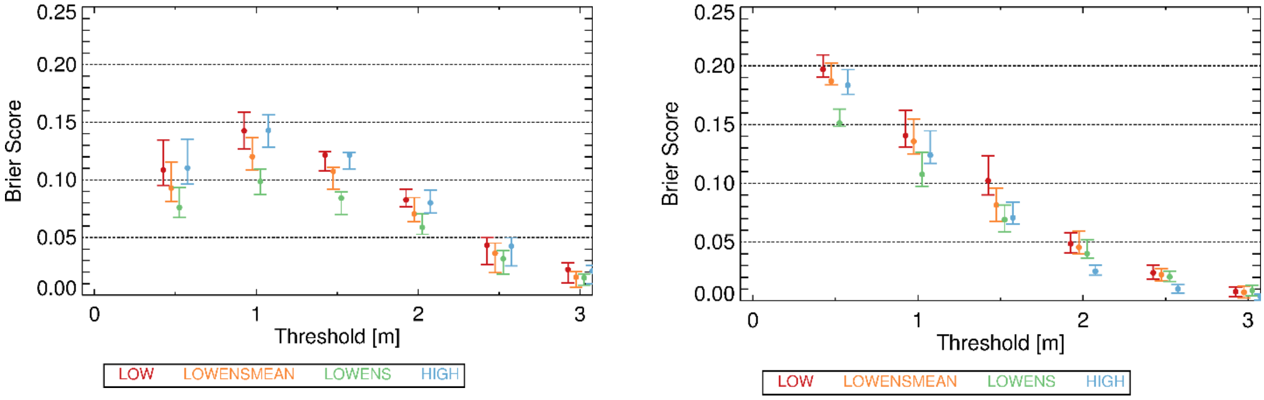

Figure 12 Brier score for Arkona WR (left panel) and Vahemadal (right panel) for binary forecast for forecast range 48 h.

6.4 Rank histogram

Rank histograms serve the purpose of illustrating the reliability of probabilistic ensemble forecasts. It is a histogram of the rank of the observation, when the observation and all ensemble members of the corresponding forecast are pooled together. If the observations and the ensemble members belong to the same distribution, then the rank histogram will be flat, while a U-shaped histogram indicates too small variance within the ensemble members. For more discussion, see Jolliffe and Stephenson (2003).

Rank histograms for all wave measurement sites for forecast range 24 and 48 h are shown in Figure S10 and S11 for forecast range 24 h and 48 h respectively. We note that all histograms show the U-shape, indicating an unrealistically small variance within the ensembles. For most sites the U-shape is symmetric, except for Vahemadal, where the U-shape is strongly asymmetrical. This corresponds well with the bias mentioned in Section 6.1.

7 Discussion

Our main finding in the previous section is that for most wave measurement sites included in this study, the LOWENSMEAN and the LOWENS forecast classes in many cases have a better performance than the LOW and HIGH forecast classes. Only for one site results are different; namely that the HIGH forecast class has the superior performance. The conclusions hold, whether based on overall RMSE, CRPS or the Brier score.

In the discussion below, it should be mentioned that improving wave forecasts is not the only driving factor in reducing the grid size of the wave model. Coupling the wave model with atmosphere or ocean circulation models may give a better description of vertical fluxes of heat and momentum (Cavaleri et al., 2012). For instance, Alari et al.(2016) documented a significant improvement of modelled sea-surface temperatures by the NEMO circulation model in the Baltic Sea when a two-way coupling to the wave model WAM was introduced. Introducing such coupling may demand a high horizontal resolution, in atmosphere, wave and ocean models, in order to describe the fluxes satisfactorily. Note also that the methodology applied in this study is a site-specific verification and inter-comparison of the different forecast families. This is a valid approach, since most uses of the wave forecasts are site-specific. However, it must be remembered, that the approach has a risk of under-estimating the overall performance due to *double-counting errors* in both space and time. We have made no attempt to assess the magnitude of this potential effect.

7.1 Comparison with other operational forecast systems

Multi-year verification results from two operational deterministic wave forecast systems that covers the region in focus have been published, and can be compared to results from the present study. Both these systems are based on the third generation WAM; the system described in (Tuomi et al., 2008) has about 22 km horizontal resolution, while the system described in (Tuomi et al., 2017) has 1 naut. mile horizontal resolution.

For certain sites, the RMSE of the 6 hour forecasts of SWH are available for at least one of the aforementioned forecast systems in addition to the DMI-WAM forecasts; thus comparison of the systems is possible. All sites have a water depth of more than 46 m and therefore represent offshore conditions.

Table 5 Comparison of RMSE for SWH of 6h forecast runs for selected sites. FIMR values are from (Tuomi et al., 2008) and FMI values are from (Tuomi et al., 2017)

	FIMR	FMI	DMI LOW	DMI LOWENSMEAN	DMI HIGH
Horizontal resolution WAM	~ 22 km	1 naut. mile	10 km	10 km	5 km
Horizontal resolution NWP	~ 22 km	2.5 km	3 km	5 km	3 km
Arkona WR	-	0.28	0.26	0.24	0.26
Bothnian Sea	-	0.28	0.25	0.23	0.25
Finngrundet WR	-	0.27	0.24	0.22	0.23
Helsinki Buoy	0.25	0.26	-	-	-
Northern Baltic	0.31	0.26	0.24	0.23	0.24

We remind the reader that the cases compared in Table 5 have different wind forcing and probably also different version of WAM. Therefore the figures cannot be directly compared and differences cannot with certainty be attributed to differences in horizontal resolution.

From Table 5 one can see that for the sites considered, the LOWENSMEAN has the lowest RMSE. This supports the finding of this study that for offshore conditions there is no reason to improve the resolution further than that of the LOW configuration. In addition, the results emphasize the value of describing the uncertainties of in the atmospheric forcing by introducing ensembles, as this leads to a lower RMSE of the forecasts. This is also in line with our findings in the previous section.

Test runs of a few months duration of deterministic and ensemble wave forecasts of SWH for the Baltic Sea (Behrens, 2015) also show slight improvement of ensemble mean forecasts, compared to deterministic forecasts, and thus support our findings.

7.2 Limitations of the study

7.2.1 Length of verification period

Operational centers typically renew their computer installations every 5-6 years with about an order of magnitude increase in performance. At DMI, a new installation was introduced early 2016, allowing the HIGH and LOWENS configurations to replace the LOW configuration. Presently (mid-2018) the system is mid-term upgraded and this makes it appropriate to do the inter-comparison now as a guidance for any changes in the operational setup.

For this reason, the operational forecasts performed on the present system, supplemented by delayed-mode forecasts has determined the three-year verification period used in our study. A longer verification period could evidently have reduced the sampling uncertainty in the analyses and thereby sharpened the conclusions. On the other hand, the three-year verification is not short compared to the study by Bunney and Saulter (2015) or the CMEMS verification report by Tuomi et al.(2017)

7.2.2 Choice of observational base

The present verification is based on observations at near-hourly resolution from a number of sites in the Baltic Sea. Therefore, in the majority of the Baltic Sea, verification is not possible, which limits the firmness of our conclusions.

SWH derived from satellite-borne altimeters (Kudryavtseva and Soomere, 2016) offers an alternative, which could be pursued in a future study. These data have a fair spatial coverage but at the cost of a temporal resolution of one day or less. This means that maximum wave heights connected to severe storms may easily be missed. Nevertheless, these data has proven useful for verification in the Baltic Sea by (Tuomi et al., 2011)

7.3 Effect of sea ice coverage

The main effect of sea ice on formation of waves is to limit the fetch. Furthermore, when a developed wave field approaches an ice-covered area, the wind and the waves decouple, so that the waves act more like swell, propagating through ice-covered areas while losing energy by breaking up the ice cover. The WAM model does not account for such interactions, and sea ice, when dense enough, acts as a solid shield that effectively removes all local wave energy in the model. It is implicitly assumed that dense ice will also be thick enough for this to be approximately correct. In the Baltic Sea, that may not always be the case, and therefore sea ice occurrence may represent a systematic error source in the present study. Another effect of sea ice in the Baltic is that the wave observing systems are withdrawn when ice is expected. This may cause a systematic bias in the verification analysis if strong winds during winter are left out.

Based on Copernicus sea ice charts produced by the Finnish Meteorological Institute the ice conditions for the Baltic have been evaluated. The Finnish ice charts are produced on a grid of approximately 1 km^2 with a temporal resolution of approximately one day in the ice season. Data is available from 2010 onwards. The average ice conditions for February for all years and the three years in focus can be found in Figure S12. All three years 2015-2017, and in particular 2015, have a smaller ice cover relative to the period 2010-2018.

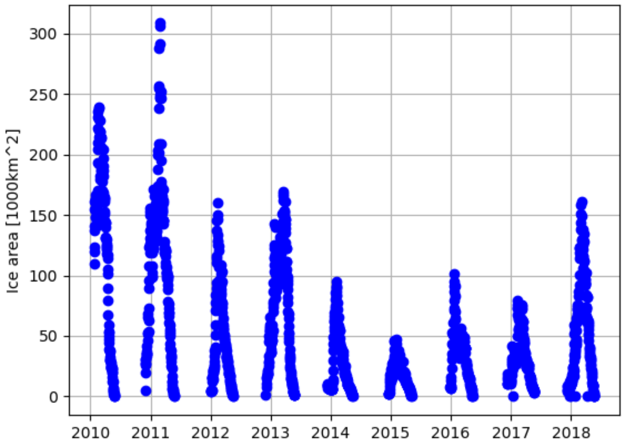

Figure 13 Integrated sea ice area of the Baltic Sea based on Finnish ice charts

Another way to illustrate this is considering the Baltic Sea integrated sea ice area, depicted in Figure 13, which shows that the years 2015-2017 have the lowest sea ice area over the whole period 2010-2018. Therefore, we may anticipate that systematic errors arise from the occurrence of sea ice are relatively small.

8 Conclusion

For most sites, we find that the HIGH forecast class does not perform better than the LOW forecast class in forecasting SWH. These sites are all positioned well away from coasts in deep water and are thus freely exposed from all directions. This suggests that the resolution of the bathymetry and the spectral resolution are adequate. For these offshore sites, introducing ensembles increases the performance of the forecasts, whether as in the LOWENSMEAN deterministic forecasts and the LOWENS probabilistic forecasts. A similar conclusion generally holds for the binary forecast of exceeding a threshold.

For one site, Vahemedal just outside Tallin, the HIGH forecast class performs better than the other classes. The bathymetry near Vahemedal is complex and relatively shallow, thus the bathymetry affects the wave field and an improved description will therefore improve the modelled wave field. Further verification with near-coast stations may reveal whether this conclusion is general for coastal areas.

For high wave heights, there are significant systematic biases for most sites shared among all three forecast configurations. These are most probably to be ascribed to model deficiencies and act to mask any differences in performance between the different forecast classes. Also the RMSE becomes large for large observed SWH. This is expected since small timing errors in the predicted wave time series will have larger impacts on the model-observation match-up when the SWH is large.The present study therefore suggests that for offshore conditions, there are no indications that a further increase of the resolution of the WAM model will result in enhanced forecast performance. In addition, the results show that introducing ensembles increases the performances. This is both true for deterministic forecast in the form of ensemble mean and for probabilistic forecast. For nearshore conditions conclusions are based on only one site, but results from this indicates that increasing the resolution gives better forecasts, while introducing ensembles does not. This can be due to both enhanced spatial resolution, allowing a better representation of shadow and shallow water effects, and/or spectral resolution.

The results of the present study thus underpins that a wave model setup with an equidistant grid cannot deliver optimal wave forecasts for both coastal and offshore conditions. This is particularly true for the Baltic Sea, where very small spatial scales are found in the archipelago near the coasts of Sweden and Finland (Björkqvist et al., 2017b). Besides implementing a 0.1 naut. miles model, these authors improved forecasts by introducing semi-empirical modifications to the wave model. Cavaleri et al. (2018) also write about this and discuss other approaches. These include one-way nesting, used in the present study (see Section 2), multi-cell grids (Bunney and Saulter, 2015), and triangular unstructured grids (e.g. Zijlema, 2010). These techniques may be worth testing for the Baltic Sea.

Finally, we note the under-spread in the ensemble forecasts demonstrated in Section 6.4. This points to a potential for improving the combined weather-wave system.

Data availability. Model data is available from the authors upon request, whereas wave observations can be found on the CMEMS server.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work was carried out under the EfficienSea2 project and supported by European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 636329. Observational data were kindly provided by Copernicus Marine Environmental Monitoring System (CMEMS). We thank two anonymous referees for valuable suggestions and dr. Ruth Mottram for help with the English language.

References

Alari, V., Staneva, J., Breivik, Ø., Bidlot, J.-R., Mogensen, K. and Janssen, P.: Surface wave effects on water temperature in the Baltic Sea: simulations with the coupled NEMO-WAM model, Ocean Dyn., 66(8), 917–930, 2016.

Alves, J.-H. G., Wittmann, P., Sestak, M., Schauer, J., Stripling, S., Bernier, N. B., McLean, J., Chao, Y., Chawla, A., Tolman, H. and others: The NCEP–FNMOC combined wave ensemble product: Expanding benefits of interagency probabilistic forecasts to the oceanic environment, Bull. Am. Meteorol. Soc., 94(12), 1893–1905, 2013.

Amante, C. and Eakins, B. W.: ETOPO1 1 ARC-MINUTE GLOBAL RELIEF MODEL: PROCEDURES, DATA SOURCES AND ANALYSIS., 2009.

Battjes, J. A. and Janssen, J. P. F. M.: Energy Loss and Set-Up Due to Breaking of Random Waves, in Coastal Engineering 1978., 1978.

Behrens, A.: Development of an ensemble prediction system for ocean surface waves in a coastal area, Ocean Dyn., 65(4), 469–486, doi:10.1007/s10236-015-0825-y, 2015.

Björkqvist, J.-V., Tuomi, L., Tollman, N., Kangas, A., Pettersson, H., Marjamaa, R., Jokinen, H. and Fortelius, C.: Brief communication: Characteristic properties of extreme wave events observed in the northern Baltic Proper, Baltic Sea, Nat. Hazards Earth Syst. Sci., 17(9), 1653–1658, doi:10.5194/nhess-17-1653-2017, 2017a.

Björkqvist, J.-V., Tuomi, L., Fortelius, C., Pettersson, H., Tikka, K. and Kahma, K. K.: Improved estimates of nearshore wave conditions in the Gulf of Finland, J. Mar. Syst., 171, 43–53, doi:10.1016/j.jmarsys.2016.07.005, 2017b.

Bunney, C. and Saulter, A.: An ensemble forecast system for prediction of Atlantic–UK wind waves, Ocean Model., 96, 103–116, doi:10.1016/j.ocemod.2015.07.005, 2015.

Cao, D., Tolman, H. L., Chen, H. S., Chawla, A. and Gerald, V. M.: Performance of the ocean wave ensemble forecast system at NCEP, in The 11th International Workshop on Wave Hindcasting & Forecasting and 2nd Coastal Hazards Symposium. [online] Available from: http://nopp.ncep.noaa.gov/mmab/papers/tn279/mmab279.pdf (Accessed 22 September 2017), 2009.

Carrasco, A. and Saetra, Ø.: A limited-area wave ensemble prediction system for the Nordic Seas and the North Sea, Norwegian Meteorological Institute., 2008.

Cavaleri, L., Fox-Kemper, B. and Hemer, M.: Wind Waves in the Coupled Climate System, Bull. Am. Meteorol. Soc., 93(11), 1651–1661, doi:10.1175/BAMS-D-11-00170.1, 2012.

Cavaleri, L., Abdalla, S., Benetazzo, A., Bertotti, L., Bidlot, J.-R., Breivik, ø., Carniel, S., Jensen, R. E., Portilla-Yandun, J., Rogers, W. E., Roland, A., Sanchez-Arcilla, A., Smith, J. M., Staneva, J., Toledo, Y., van Vledder, G. P. and van der Westhuysen, A. J.: Wave modelling in coastal and inner seas, Prog. Oceanogr., doi:10.1016/j.pocean.2018.03.010, 2018.

Günther, H., Hasselmann, S. and Janssen, P. A. E. M.: The WAM Model cycle 4, World Data Center for Climate (WDCC) at DKRZ., 1992.

Hasselmann, K., Barnett, T., Bouws, E., Carlson, H., Cartwright, D., Enke, K., Ewing, J., Gienapp, H., Hasselmann, D., Kruseman, P., Meerburg, A., M{ü}ller, P., Olbers, D., Richter, K., Sell, W. and Walden, H.: Measurements of wind-wave growth and swell decay during the {Joint North Sea Wave Project}, Deut Hydrogr Z, 8(12), 1–95, 1973.

Jolliffe, I. T. and Stephenson, D. B.: Forecast verification: a practitioner’s guide in atmospheric science, John Wiley & Sons., 2003.

Kudryavtseva, N. A. and Soomere, T.: Validation of the multi-mission altimeter wave height data for the Baltic Sea region, Est. J. Earth Sci., 65(3), 161, doi:10.3176/earth.2016.13, 2016.

Saetra, Ø. and Bidlot, J.-R.: Assessment of the ECMWF Ensemble Prediction Sytem for Waves and Marine Winds, European Centre for Medium-Range Weather Forecasts., 2002.

Schaffer, J., Timmermann, R., Arndt, J. E., Kristensen, S. S., Mayer, C., Morlighem, M. and Steinhage, D.: A global, high-resolution data set of ice sheet topography, cavity geometry, and ocean bathymetry, Earth Syst. Sci. Data, 8(2), 543–557, doi:10.5194/essd-8-543-2016, 2016.

She, J., Allen, I., Buch, E., Crise, A., Johannessen, J. A., Le Traon, P.-Y., Lips, U., Nolan, G., Pinardi, N., Reißmann, J. H., Siddorn, J., Stanev, E. and Wehde, H.: Developing European operational oceanography for Blue Growth, climate change adaptation and mitigation, and ecosystem-based management, Ocean Sci., 12(4), 953–976, doi:10.5194/os-12-953-2016, 2016.

Tuomi, L., Kangas, A., Leinonen, J. and Boman, H.: The Accuracy of FIMR Wave Forecasts in 2002-2005., 2008.

Tuomi, L., Kahma, K. K. and Pettersson, H.: Wave hindcast statistics in the seasonally ice-covered Baltic sea., Boreal Environ. Res., 16, 2011.

Tuomi, L., Vähä-Piikkiö, O. and Alari, V.: Baltic Sea Wave Analysis and Forecasting Product BALTICSEA_ANALYSIS_FORECAST_WAV_003_010, 2017.

Zijlema, M.: Computation of wind-wave spectra in coastal waters with SWAN on unstructured grids, Coast. Eng., 57(3), 267–277, doi:10.1016/j.coastaleng.2009.10.011, 2010.