Ocean Sci. Discuss., https://doi.org/10.5194/os-2018-59
Manuscript under review for journal Ocean Sci.
Discussion started: 31 May 2018

Ocean Science Discussions

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

Torben Schmith, Jacob Woge Nielsen, Till Andreas Soya Rasmussen

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 in three different configurations 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 stations, the introduction of ensembles enhances the overall performance of the forecasts, whereas increasing the horizontal and spectral resolution does not. These stations 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 for one station, we find that increasing the horizontal and spectral resolution significantly improved the forecasts. This station 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, while for nearshore areas better resolution may improve results.

1 Introduction

Severe surface waves 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 development of waves is further influenced by the *fetch* (the distance, over which the wind acts), and by the *duration* of the wind. Dissipation of the wave energy occurs through internal dissipation, dissipation through bottom friction and through wave breaking over a shallow and sloping sea bed. Thus, a correct and detailed description of the bathymetry is important for correctly forecasting waves. Also refraction of waves is influenced by the bathymetry. Other factors, which potentially has an effect on the development of waves includes ocean current, varying water depth due to variations in sea level, and sea ice coverage.

Ocean Sci. Discuss., https://doi.org/10.5194/os-2018-59
Manuscript under review for journal Ocean Sci.
Discussion started: 31 May 2018

The Baltic Sea is connected to the world ocean through the Transition Area with the shallow and narrow Danish 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 shallow water areas. 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 lows 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 determines the spatial resolution of the wave model. Due to limited computer resources, only certain horizontal grid spacing can be afforded.

With additional computer resources becoming available, the horizontal spatial resolution can be increased. This allows for an improved description and forecasting of the synoptic and mesoscale atmospheric systems, including the details of the associated wind field, by the NWP model. In addition, a more detailed description of the bathymetry improves the correct description of dissipation and refraction of waves, as argued above. This is in particular true in shallow seas, such as the Baltic See. 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.

Historically, computer resources have increased through time, and this development is expected to continue in future. This has made a development towards ensemble weather forecasts 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 (Ana 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 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.

This paper is arranged as follows. Section 2 describes the model and setup, and Section 3 describes the observations. The verification methodology is described in Section 4 and applied in Section 5. Results of the verification are discussed in Section 6 and conclusions made in Section 7.

Ocean Sci. Discuss., https://doi.org/10.5194/os-2018-59
Manuscript under review for journal Ocean Sci.
Discussion started: 31 May 2018

2 Model and setup

The DMI operational wave forecasting system DMI-WAM uses the 3rd generation spectral wave model WAM Cycle4.5 (Günther et al., 1992) forced by the regional NWP model DMI-HIRLAM and the global NWP model ECMWF/GLM. WAM Cycle4.5 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 suite forecasts waves in a larger area than the Baltic Sea and therefore has a setup with two nested spatial domains of different geographical extent and spatial resolution (see Figure 1): North Atlantic and North Sea/Baltic Sea (NSB), of which forecast results from the NSB-domain are used in this study. The North Atlantic domain uses the JONSWAP spectrum for fully developed wind-sea (Hasselmann et al., 1973) along open model boundaries, while the NSB domain use modeled wave spectra at open boundaries.

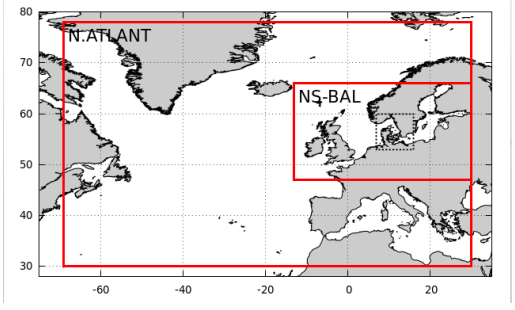


Figure 1 Nesting of domains in DMI-WAM. Outer frame is North Atlantic domain, inner frame is the North Sea/Baltic Sea(NSB)-domain. Dotted frame is the Transition Area. Only data from the NSB-domain are analyzed in this study.

In the North Atlantic domain, bathymetry is taken from Rtopo 30"×30" global bathymetry (Schaffer et al., 2016), while in the NSB domain, Rtopo is combined with local depth information from various sources, in order to obtain a more accurate bathymetry.

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 North Atlantic 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. The wave energy is completely dissipated in areas with sea ice cover above 30%.

The surface wind forcing is provided by different atmospheric models for the two domains. For the North Atlantic domain wind is provided by the ECMWF global weather forecast in 16 km resolution every 3 hours.

Ocean Sci. Discuss., https://doi.org/10.5194/os-2018-59
Manuscript under review for journal Ocean Sci.
Discussion started: 31 May 2018

For the NSB domain, the surface wind is provided by DMI-HIRLAM, version SKA (3 km resolution) every hour. To diminish coastal effects, DMI-WAM uses a special *water-wind*, 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. 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 GLM	Hirlam SKA ECMWF GLM
Boundary condition	JONSWAP	Nested
Bathymetry	Rtopo	Rtopo/IOW/GEO

The ice concentration originates 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 kept constant through each forecast run.

DMI-WAM is cold-started once and for all using fully developed sea with a constant fetch of 30 km based on the JONSWAP spectrum. Subsequent model runs are initialized using the sea state at analysis time, calculated by the previous run as a six hour forecast. The first two weeks after cold-start is regarded as spin-up. The operational DMI-WAM suite is run four times a day to 48 h forecast range. Spatial fields of forecasted SWH and other variables are output in hourly time resolution.

Three different configurations of the DMI-WAM setup have been applied, and data from these for the period 2015-2017 is the basis for the present verification. In the LOW configuration, the NSB-domain has approximately 10 km horizontal resolution, and the wave energy is resolved in 24 directions and at 32 frequencies, corresponding to wave periods of 1.25-23.94 s and wave lengths of 2.4-895 m (in deep water). An ensemble configuration (LOWENS) has characteristics identical to LOW, but with parallel run of 11 ensemble members forced with perturbed atmospheric fields (initial conditions and physics). Finally, in the HIGH configuration, the horizontal resolution is approximately 5 km and the wave energy resolved in 36 directions and 35 frequencies, corresponding to wave periods of 0.94-23.94 s, and wave lengths of 1.37-895 m (in deep water). An overview is provided in Table 2.

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

	Horizontal resolution [km]		# wave directions	# wave spectral frequencies	Ensemble members
	North Atlantic	NSB			
LOW	50	10	24	32	1
LOWENS	50	10	24	32	11
HIGH	25	5	36	35	1

When replacing the LOW forecast setup with the HIGH setup, the required computational resources are increased by a factor of 2^2 (increase in horizontal resolution) × 1.75 (effective decrease in time step) × 1.5

Ocean Sci. Discuss., https://doi.org/10.5194/os-2018-59
Manuscript under review for journal Ocean Sci.
Discussion started: 31 May 2018

(increase of number of directions) × 35/32 (increase of number of spectral frequencies) ≈ 11.5, while it is increased by a factor of 11 (number of ensembles) from the LOW to the LOWENS setup. Since these increases in computational effort are very similar, an intercomparison 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.

3 Observations

Observed series of SWH from stations in the Baltic Sea 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 stations with valid observations that covered more than 40% and were distributed reasonably throughout the study period. Figure 2 and Table 3 show the positions of the selected stations together with the bathymetry of the Baltic Sea. Some stations did not observe at the full hour. Observations from these stations 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 3.

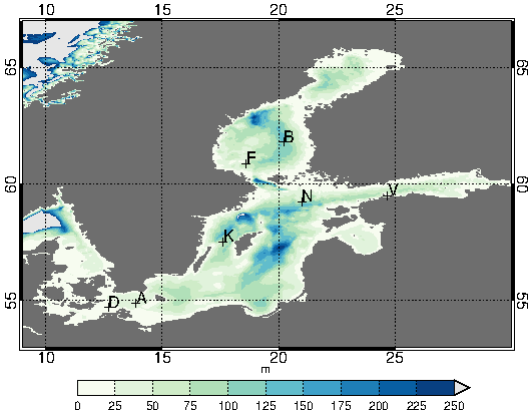


Figure 2 Map of the Baltic Sea with bathymetry and positions of station marked with crosses. For details about stations, see Table 3.

Ocean Sci. Discuss., https://doi.org/10.5194/os-2018-59
Manuscript under review for journal Ocean Sci.
Discussion started: 31 May 2018

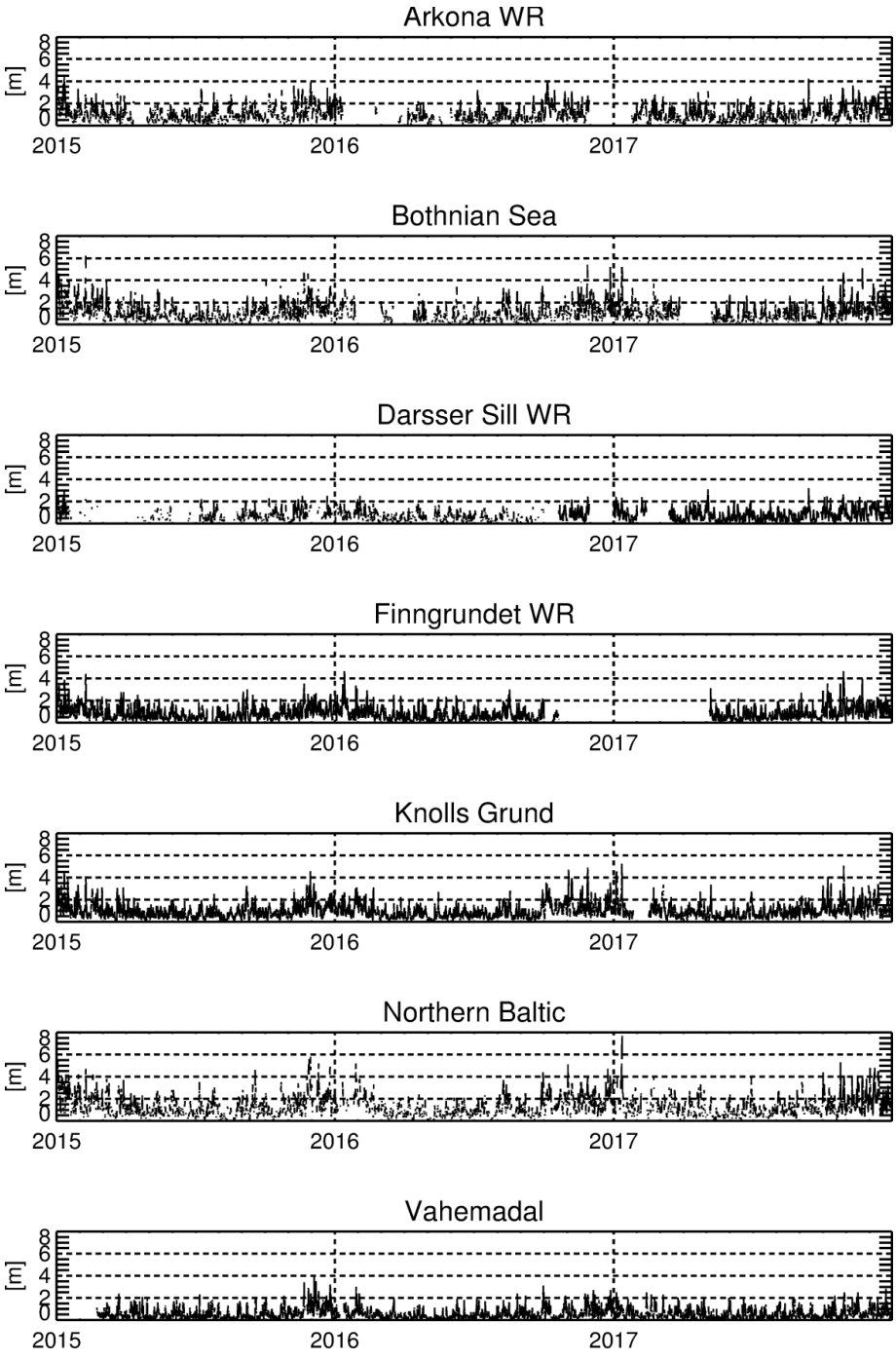
153 **Table 3 Details of stations.**

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

154

155

Ocean Sci. Discuss., https://doi.org/10.5194/os-2018-59
Manuscript under review for journal Ocean Sci.
Discussion started: 31 May 2018



156
157 **Figure 3 Observation series of SWH used in the study.**

Ocean Sci. Discuss., https://doi.org/10.5194/os-2018-59
Manuscript under review for journal Ocean Sci.
Discussion started: 31 May 2018

158

159 4 Verification methodology

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

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

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

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

168 the bias

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

170 and the correlation coefficient

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

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

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

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

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

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

Ocean Sci. Discuss., https://doi.org/10.5194/os-2018-59
Manuscript under review for journal Ocean Sci.
Discussion started: 31 May 2018

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's, and each of these is randomly chosen, with replacement, of the three January's 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 forecasted SWH against observations

5.1 Deterministic measures

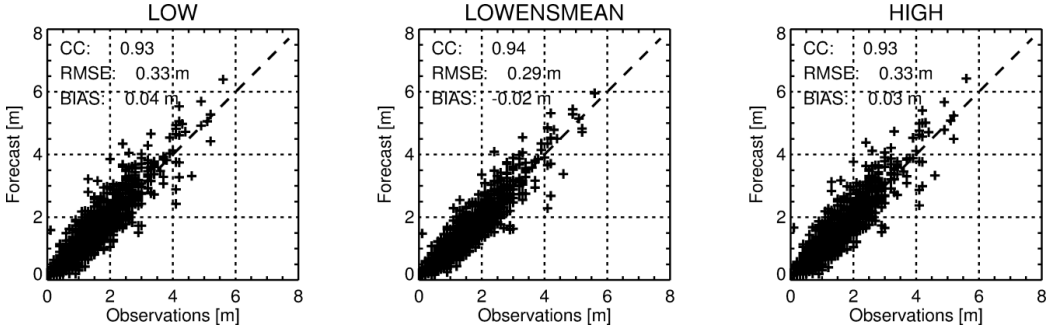


Figure 4 Scatter plot of 24 h forecasts and corresponding observations of significant wave height at station Bothnian Sea for the LOW, LOWENSMEAN and HIGH forecast classes. Dotted line is the diagonal.

Ocean Sci. Discuss., https://doi.org/10.5194/os-2018-59
Manuscript under review for journal Ocean Sci.
Discussion started: 31 May 2018

To get an idea of the overall quality of the forecasts, Figure 4 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 stations, such as Arkona WR (see Figure 5), 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.

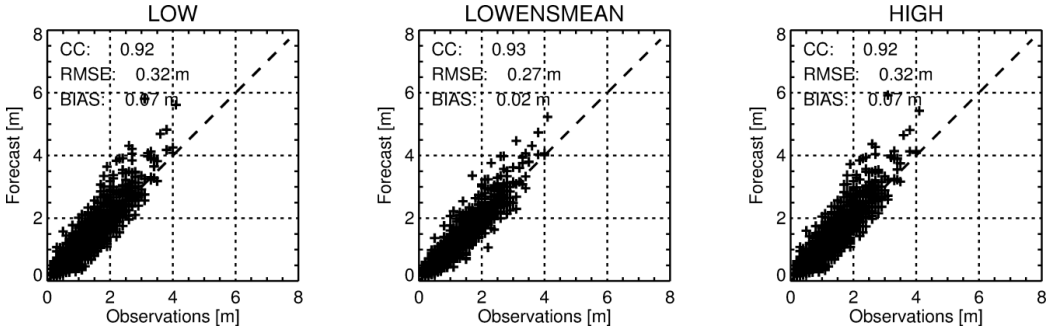


Figure 5 As Figure 4 but for station Arkona WR.

We now turn to the RMSE as function of forecast range, of which plots for all stations can be found in Figure S1. Plots for stations with qualitatively different behavior in this respect are shown in Figure 6. For Arkona WR, all three forecasts have a non-zero RMSE for forecast range zero (the analysis). The reason for this is that this 'analysis' is a forecast with forecast range 6 h, made six hours before. The RMSE increases slightly as function of forecast range. The RMSE of the LOW and the HIGH forecasts coincide to a large degree, while the RMSE for LOWENSMEAN gradually diverges to a value of 5 cm lower and for forecast ranges larger than 24 h, the confidence bands do not overlap with those for the LOW and HIGH forecast classes.

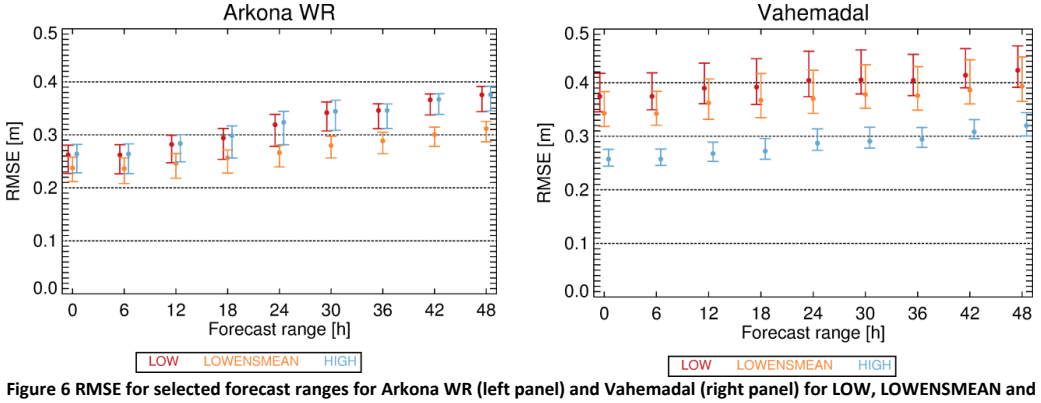


Figure 6 RMSE for selected forecast ranges for Arkona WR (left panel) and Vahemadal (right panel) for LOW, LOWENSMEAN and HIGH forecasts. Error bars show 5/95% confidence bands calculated by bootstrapping.

Ocean Sci. Discuss., https://doi.org/10.5194/os-2018-59
Manuscript under review for journal Ocean Sci.
Discussion started: 31 May 2018

Qualitatively the same picture is found for most other station: RMSE and BIAS for LOW and HIGH forecasts are almost similar, while they are lower for LOWENSMEAN forecasts. However, the RMSE values of the LOWENSMEAN forecasts are not for all stations well separated by non-overlapping confidence bands from RMSE of the other forecast classes.

The station Vahemadal has a different behavior. For this station, the HIGH forecast class has a significantly smaller RMSE than the LOW and LOWENSMEAN forecasts, which have overlapping confidence bands. This station 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).

5.1.1 Performance depending on observed SWH

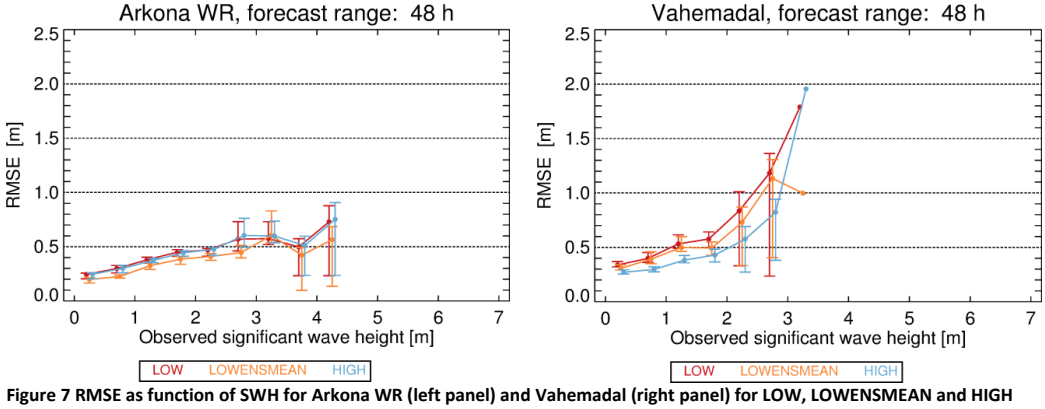


Figure 7 RMSE as function of SWH for Arkona WR (left panel) and Vahemadal (right panel) for LOW, LOWENSMEAN and HIGH forecasts and forecast range 48 h. Error bars show 5/95% confidence bands calculated by bootstrapping.

The RMSE of the forecasts depends on the magnitude of the SWH. Plots for all stations for 24 and 48 h forecast range of RMSE as function of the SWH can be found in Figures S2 and S3. The RMSE for Arkona WR and Vahemadal as function of the SWH for forecast range 48 h is shown in Figure 7. The RMSE increases as function of the observed SWH for both stations. 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 in all stations, 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 stations for 24 and 48 h forecast range of BIAS as function of the SWH are displayed in Figures S4 and S5. For small SWH, the BIAS is close to zero for most stations. For some stations, the bias remains close to zero for increasing SWH, as shown for Arkona WR in left panel of Figure 8, while for others it becomes different from zero for large values of SWH. The is no noticeable different in the bias of the different forecast classes, except for Vahema, shown in right panel of Figure 8, where the HIGH forecast class has a significantly smaller bias than the other forecast classes.

Ocean Sci. Discuss., https://doi.org/10.5194/os-2018-59
Manuscript under review for journal Ocean Sci.
Discussion started: 31 May 2018

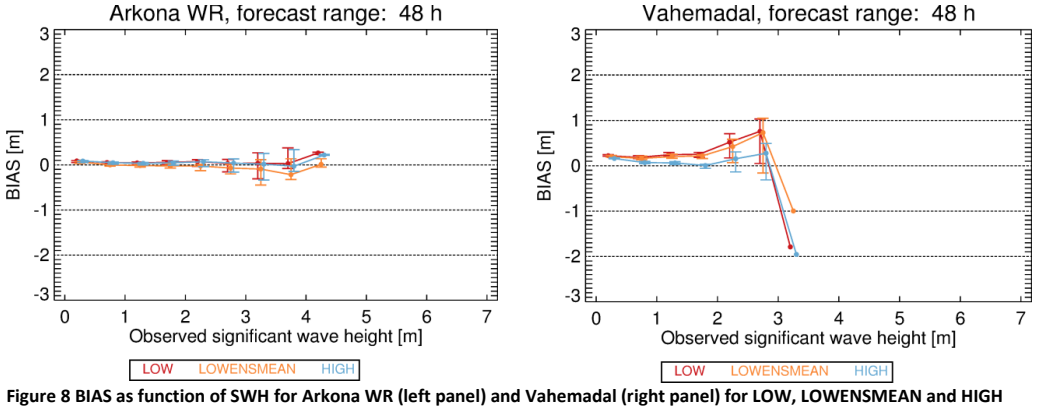


Figure 8 BIAS as function of SWH for Arkona WR (left panel) and Vahemadal (right panel) for LOW, LOWENSMEAN and HIGH forecasts and forecast range 24 h. Error bars show 5/95% confidence bands calculated by bootstrapping.

5.2 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. Besides, 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 stations for selected forecast ranges can be found in Figure S6. As typical examples, Figure 9 displays this plot for Arkona WR and Vahemadal.

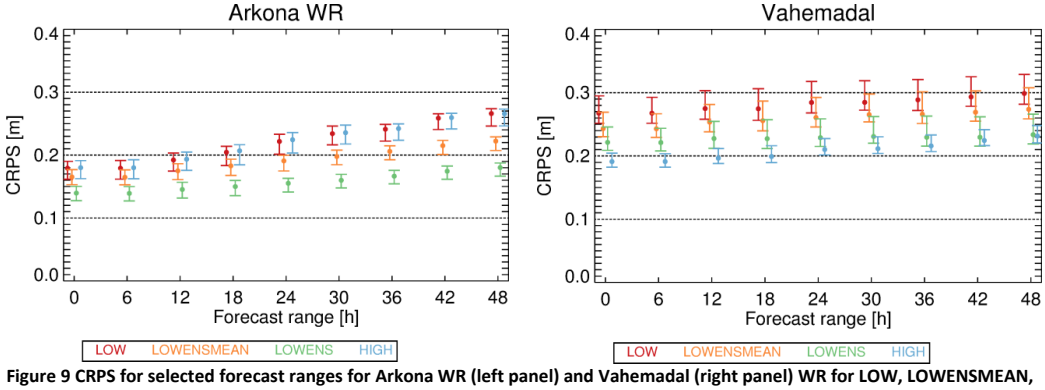


Figure 9 CRPS for selected forecast ranges for Arkona WR (left panel) and Vahemadal (right panel) WR for LOW, LOWENSMEAN, LOW and HIGH forecasts. Error bars show 5/95% confidence bands calculated by bootstrapping.

This plot reveals that for Arkona WR, the LOWENSMEAN forecast class has a significantly lower CRPS compared to both the HIGH and LOW classes. This is most prominent for the large forecast ranges, where its confidence band is non-overlapping with the confidence band for other forecast classes. Furthermore, the LOWENS forecast class has an even lower CRPS, with confidence bands separated from those of all other forecasts classes. This behavior is common among almost all stations. Vahemadal is the exception, where 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.

Ocean Sci. Discuss., https://doi.org/10.5194/os-2018-59
Manuscript under review for journal Ocean Sci.
Discussion started: 31 May 2018

5.3 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 function of threshold is shown for all stations in Figures S7 and S8. Figure 10 shows the Brier Score as 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 2 m threshold. Also the LOWENSMEAN forecast class has low Brier Score. This behavior is common to almost all stations. The exception is again Vahemadal, where the Brier Score is smallest for the HIGH forecasts for thresholds above 1 m.

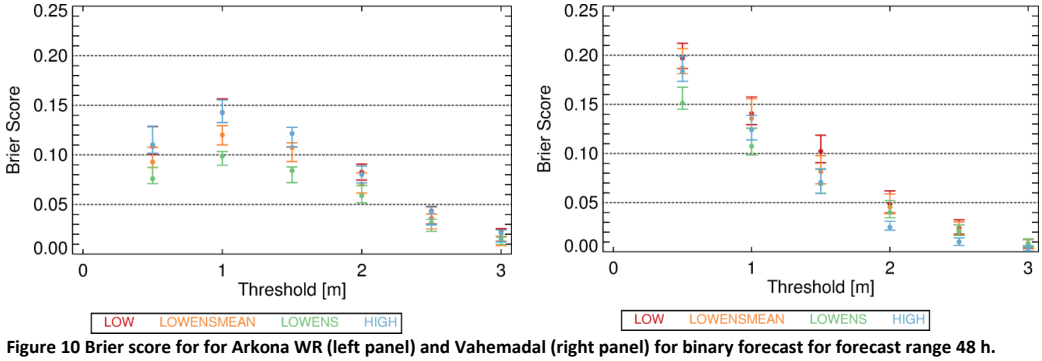


Figure 10 Brier score for for Arkona WR (left panel) and Vahemadal (right panel) for binary forecast for forecast range 48 h.

6 Discussion

Our main finding in the previous section is that for most stations, the LOWENSMEAN and the LOWENS forecast classes have a performance superior to the LOW and HIGH forecast classes. Only for one station results are different; namely that the HIGH forecast class has the superior performance. These conclusion hold, whether based on overall RMSE, CRPS or the Brier score.

6.1 Comparison with other operational forecasts

Multi-year verification results from two operational deterministic wave forecast systems 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 stations, 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 stations have a water depth of more than 46 m and therefore represent offshore conditions.

Ocean Sci. Discuss., https://doi.org/10.5194/os-2018-59
Manuscript under review for journal Ocean Sci.
Discussion started: 31 May 2018

307 **Table 4 Comparison of RMSE for SWH of 6h forecast runs for selected stations. FIMR values are from (Tuomi et al., 2008) and FMI**
308 **values are from (Tuomi et al., 2017)**

	FIMR	FMI	LOW	LOWENSMEAN	HIGH
Horizontal resolution	~ 22 km	1 naut. mile	10 km	10 km	5 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

309

310 From Table 4 one can see that for the stations considered, the LOWENSMEAN has the lowest RMSE. This
311 supports the finding of this study that for offshore conditions, there is no reason to improve the resolution
312 further than that of the LOW configuration. In addition, the results emphasize the value of describing the
313 uncertainties of in the atmospheric forcing by introducing ensembles, as this leads to a lower RMSE of the
314 forecasts. This is also in line with our findings in the previous section.

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

318 ## 6.2 Choice of observational base

319 The present verification is based on observation in near-hourly resolution from a number of stations in the
320 Baltic Sea. Therefore, in the major parts of the Baltic Sea, verification is not possible, which puts a limit on
321 how strong conclusions can be made.

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

327 ## 6.3 Effect of sea ice coverage

328 The main effect of sea ice on formation of waves is to limit the fetch. Furthermore, when a developed wave
329 field approach an ice-covered area, the wind and the waves decouple, so that the waves act more like
330 swell, propagating through ice-covered areas while losing energy by breaking up the ice cover. The WAM
331 model does not account for such interactions, and sea ice, when dense enough, act as a solid shield that
332 effectively remove all local wave energy in the model. It is implicitly assumed that dense ice will also be
333 thick enough for this to approximately correct. In the Baltic Sea, that may not always be the case, and
334 therefore sea ice occurrence may represent a systematic error source in the present study.

Ocean Sci. Discuss., https://doi.org/10.5194/os-2018-59
Manuscript under review for journal Ocean Sci.
Discussion started: 31 May 2018

7 Conclusion

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

For one station, Vahemadal, the HIGH forecast class performs better than the other classes. Vahemadal is a coastal station, situated just outside Tallinn in a complicated bathymetry with an island nearby and therefore shielded from many directions. This can explain that better description of bathymetry and better description of short waves improves the forecast.

For high wave heights, there are significant systematic biases for most stations 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.

Based on the above, we hypothesize that for offshore conditions, there are no indications of 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 station, 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.

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).

365 **References**

366 Alves, J.-H. G., Wittmann, P., Sestak, M., Schauer, J., Stripling, S., Bernier, N. B., McLean, J., Chao, Y.,
367 Chawla, A., Tolman, H. and others: The NCEP–FNMOC combined wave ensemble product: Expanding
368 benefits of interagency probabilistic forecasts to the oceanic environment, Bull. Am. Meteorol. Soc., 94(12),
369 1893–1905, 2013.

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

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

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

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

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

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

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

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

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

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

395 She, J., Allen, I., Buch, E., Crise, A., Johannessen, J. A., Le Traon, P.-Y., Lips, U., Nolan, G., Pinardi, N.,
396 Reißmann, J. H., Siddorn, J., Stanev, E. and Wehde, H.: Developing European operational oceanography for
397 Blue Growth, climate change adaptation and mitigation, and ecosystem-based management, Ocean Sci.,
398 12(4), 953–976, doi:10.5194/os-12-953-2016, 2016.

399 Tuomi, L., Kangas, A., Leinonen, J. and Boman, H.: The Accuracy of FIMR Wave Forecasts in 2002-2005.,
400 2008.

Ocean Sci. Discuss., https://doi.org/10.5194/os-2018-59
Manuscript under review for journal Ocean Sci.
Discussion started: 31 May 2018

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

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

405