

1 Better Baltic Sea wave forecasts: Improving resolution or 2 introducing ensembles?

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

4 Danish Meteorological Institute, Copenhagen, Denmark

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

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

24

25 1 Introduction

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

29 Waves are generated by energy transfer from surface winds that act on the sea. The energy transfer is
30 determined by the *fetch* (the distance, over which the wind acts), and by the *duration* of the wind. For *deep*
31 *water waves*, defined as the wave height being much smaller than the water depth, dissipation of the wave
32 energy occurs through internal dissipation mainly. For *shallow water waves*, defined as the wave height
33 being comparable to the water depth, dissipation through bottom friction and through wave breaking over
34 a shallow and sloping sea bed becomes important. Shallow water waves may also be refracted over a
35 varying bathymetry. Therefore, a correct and detailed description of the bathymetry is important for
36 correctly forecasting waves in coastal areas and other shallow sea areas. Other factors with a potential

37 effect on the development of waves include nonlinear wave-wave interaction, ocean currents, time-varying
38 water depth due to variations in sea level, and sea ice coverage.

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

47 Short-term forecasting of surface waves is done by a wave model, forced with forecasted wind from an
48 atmospheric numerical weather prediction (NWP) model. The equations of the NWP model are discretized
49 on a horizontal grid with a certain spatial resolution, which determines the maximum spatial resolution of
50 the wave model. The available computer resources put a limit on the horizontal grid spacing, which can be
51 afforded.

52 Technical development has increased the computational resources, making possible to increase the
53 horizontal spatial resolution of the NWP and wave models. This allows for an improved description and
54 forecasting of the synoptic and mesoscale atmospheric systems, including the details of the associated
55 wind field. In addition, a more detailed description of the bathymetry improves the correct description of
56 dissipation and refraction of waves, as argued above. Additional computer resources may also be used to
57 improve the spectral resolution in the wave model. This includes the directional resolution and the number
58 of frequencies included.

59 Increasing computer resources have also made ensemble NWP possible. The purpose of ensemble
60 forecasts is to improve forecast skill by taking both the initial error of the forecast and the uncertainty of
61 the model physics into account. Furthermore, ensemble forecast allows for probabilistic forecasts,
62 identified as a priority for operational oceanography (She et al., 2016), and allows for quantifying forecast
63 uncertainty. Ensemble wave forecast systems have been implemented at global scale (Alves et al., 2013;
64 Cao et al., 2009; Saetra and Bidlot, 2002) and more regionally in the Norwegian Sea (Carrasco and Saetra,
65 2008), and in the German Bight and Western Baltic (Behrens, 2015).

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

72 It should be mentioned that improving wave forecasts is not the only driving factor in reducing the grid size
73 of the wave model. Coupling the wave model with atmosphere or ocean circulation models may give a
74 better description of vertical fluxes of heat and momentum (Cavaleri et al., 2012). For instance, Alari et
75 al.(2016) documented a significant improvement of modelled SSTs by the NEMO circulation model in the

76 Baltic Sea when a two-way coupling to the wave model WAM was introduced. Doing such couplings may
77 demand a high horizontal resolution to describe the fluxes most satisfactorily.

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

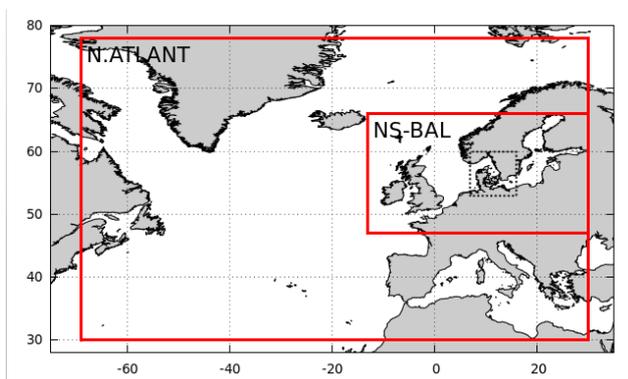
82 This paper is arranged as follows. Section 2 describes the model and setup, Section 3 describes the
83 observations used and the verification methodology is described in Section 4. Verification of DMI-HIRLAM
84 wind forecasts is in Section 5, and the SWH forecast verification is in Section 6. Results of the verification
85 are discussed in Section 7 and conclusions made in Section 8.

86 2 Model and setup

87 The DMI operational wave forecasting system DMI-WAM uses the 3rd generation spectral wave
88 model WAM Cycle4.5 (Günther et al., 1992) forced by the regional NWP model DMI-HIRLAM and the global
89 NWP model ECMWF-GLM. WAM Cycle4.5 solves the spectral wave equation, and calculates the wave
90 energy as a function of position, time, wave period and direction. Derived variables, such as the significant
91 wave height (SWH), are calculated as suitable integrals of the wave energy spectrum.

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

98



99

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

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

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

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

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

120

121 **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	<i>DMI-HIRLAM</i>
Boundary condition	JONSWAP	One-way nested
Depth-induced wave breaking	No	Yes

122

123 Each forecast runs is initialized using the sea state at analysis time, calculated by the previous run as a six
 124 hour forecast. The operational DMI-WAM suite is run four times a day to 48 h forecast range. Spatial fields
 125 of forecasted SWH and other variables are output in hourly time resolution.

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

139 All configurations are forced by winds from ECMWF-HRES in the NA domain and DMI-HIRLAM in the NSB
 140 domain. In the NSB domain, the LOW and HIGH are forced by the S03 version (3 km horizontal resolution),
 141 while LOWENS is forced by the S05 version (5 km horizontal resolution). The DMI-HIRLAM winds are
 142 interpolated to the WAM grids by bilinear interpolation. To diminish coastal effects, DMI-HIRLAM delivers a
 143 special *water-wind* to DMI-WAM, in which the surface roughness everywhere is assumed to be that of
 144 water. This enhances the wind speed in the coastal zone, most important in semi-enclosed areas (bays,
 145 fjords, etc.). It is basically a way to sharpen the land/sea boundary, reducing influence of land roughness
 146 on near-shore winds. An overview of the DMI-WAM configurations is provided in Table 2.

147 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	-	-

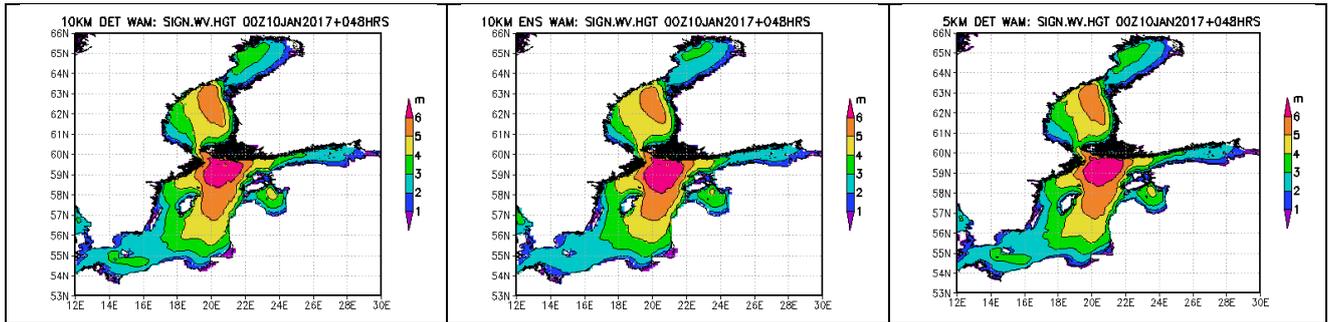
148

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

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

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

162

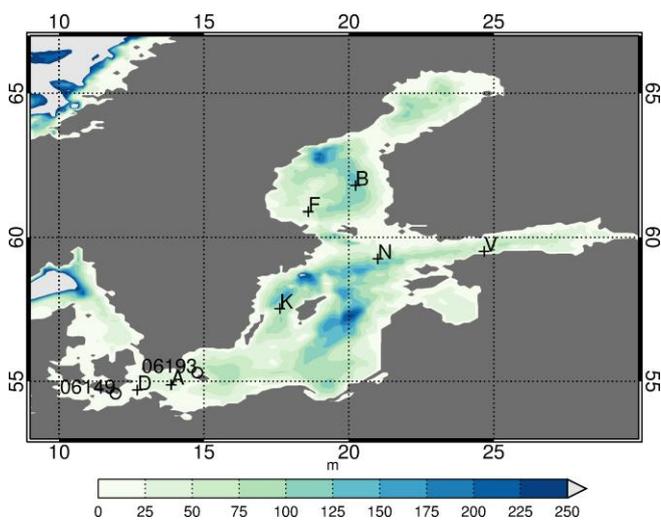


164 **Figure 2** Forecasted (48h) SWH at the peak of 'Toini' storm 10 January 2017 00z for LOW (left), LOWENSMEAN (middle) and HIGH
 165 forecasts.

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

169 3 Observations

170 Observed series of SWH from wave measurement sites in the Baltic Sea obtained from the Copernicus
 171 Marine Environmental Monitoring System (CMEMS) database are used. None of the series has a continuous
 172 record over the three-year period 2015 – 2017. Data gaps may be due to malfunction, maintenance or
 173 withdrawal of the instrument. The latter occur during winter due to the possibility of ice. We selected sites
 174 with valid observations that covered more than 40% and were distributed reasonably throughout the study
 175 period. Figure 3 and Table 3 show the positions and water depths of the wave measurement sites together
 176 with the bathymetry of the Baltic Sea. Some sites did not observe at the full hour. Observations from these
 177 sites were ascribed to the nearest full hour, if the time distance between the observation time and the full
 178 hour was less than 15 min, otherwise not used. All observation series used are shown in Figure S1. The
 179 frequency of observed SWH in different intervals for each site is given in Table 4



180

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

183 **Table 3 Details of wave measurement sites.**

Observation site	Lon	Lat	Depth [m]	
			Model	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

184

185

186 **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
Arkona WR	0.47	0.39	0.12	0.01	0.00
Bothnian Sea	0.46	0.38	0.12	0.02	0.01
Darsser Sill WR	0.67	0.31	0.02	0.00	0.00
Finngrundet WR	0.69	0.27	0.04	0.01	0.00
Knolls Grund	0.62	0.31	0.06	0.01	0.00
Northern Baltic	0.39	0.37	0.18	0.05	0.01
Vahemadal	0.78	0.20	0.02	0.00	0.00

187

188 4 Verification methodology

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

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

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

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

197 the bias

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

199 and the correlation coefficient

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

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

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

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

207
$$CRPS(\tau) = \langle \int [F^\tau(h_s) - H(h_s - h_{s,obs})]^2 dh_s \rangle,$$

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

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

217
$$BS(\tau) = \langle (p - x)^2 \rangle,$$

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

222 **4.1 Calculation of confidence bands**

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

229 Each resampled series is constructed as follows: The resampled series will contain three January's, and each
230 of these is randomly chosen, with replacement, of the three January's from the original series. A similar
231 procedure applies for February, etc. In this way, the resampled series are most likely different but the
232 annual cycle is preserved. Both the observed series and the forecast series are resampled. For each pair of
233 resampled series bootstrapped value of the performance measures are calculated. Repeating the
234 resampling procedure, we obtain 1000 resampled values of the measures, from which their approximate
235 statistical distribution and confidence bands can be calculated. As a standard, confidence bands (5/95%)
236 are calculated by the bootstrap procedure described above and this allows for a quantitative inter-

237 comparison of the performance measures for the different forecast classes: if the confidence bands do not
 238 overlap then there is a significance difference.

239 5 Verification of the wind forecasts

240 The two configurations of DMI-HIRLAM used (see Table 2) have been verified against available wind
 241 observations from Danish coastal stations, i.e. covering the western part, of the Baltic Sea, for the period 1
 242 January 2015– 31 December 2017. For the S05 configuration, the ensemble mean is verified.

243 **Table 5 Verification results for DMI-HIRLAM against Danish coastal stations for the period 1 January 2015– 31 December 2017.**
 244 **Positions of stations are marked on Figure 3 .**

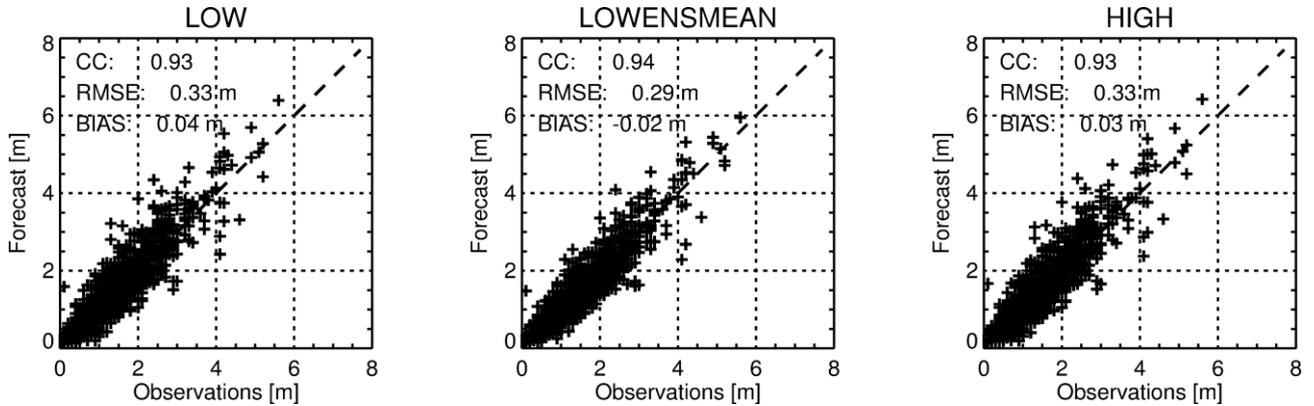
FCST RANGE	BIAS [ms^{-1}]		RMSE [ms^{-1}]		CC		Hit rate, error $\leq 2 \text{ ms}^{-1}$	
	S05 (EM)	S03	S05 (EM)	S03	S05 (EM)	S03	S05 (EM)	S03
Gedser (WMO 06149) :								
0	0.48	0.46	1.57	1.56	0.90	0.91	0.82	0.82
6	0.43	0.46	1.58	1.62	0.90	0.90	0.81	0.80
12	0.45	0.49	1.66	1.72	0.89	0.89	0.79	0.78
18	0.45	0.49	1.73	1.83	0.88	0.87	0.77	0.76
24	0.45	0.51	1.77	1.91	0.87	0.86	0.76	0.74
36	0.42	0.51	1.92	2.03	0.85	0.84	0.73	0.70
48	0.44	0.46	2.03	2.16	0.82	0.81	0.70	0.67
Hammer Odde Lighthouse (WMO 06197) :								
0	0.31	0.19	1.24	1.24	0.92	0.91	0.90	0.90
6	0.22	0.12	1.28	1.33	0.91	0.90	0.88	0.87
12	0.24	0.13	1.34	1.42	0.90	0.88	0.87	0.85
18	0.25	0.15	1.38	1.48	0.89	0.87	0.86	0.84
24	0.26	0.14	1.43	1.57	0.88	0.86	0.84	0.82
36	0.24	0.11	1.53	1.67	0.86	0.84	0.82	0.79
48	0.23	0.10	1.62	1.80	0.85	0.81	0.80	0.77

245

246 Table 5 summarizes verification results for 10m wind forecasts for the 3km-resolution S03 model and the
 247 ensemble mean of the 5km-resolution S05 model for two Danish coastal stations in the western part of the
 248 Baltic Sea. A comparison between the two model forecasts shows a small positive bias and RMS errors
 249 increasing with forecast range from approx. 1 ms^{-1} to approximately 2 ms^{-1} for 48h forecasts. The error of
 250 the ensemble mean forecasts generally increases less with forecast range than the error of the high-
 251 resolution forecasts. Similarly, the correlation and the hit rate (error $\leq 2 \text{ ms}^{-1}$) decrease with forecast
 252 range, but less so for the ensemble mean forecasts. That is, in terms of wind forcing the ensemble mean of
 253 the S05 model provides slightly more accurate forecasts than the higher resolution, deterministic S03
 254 model, especially for the longer forecast ranges.

255 **6 Verification of forecasted SWH against observations**

256 **6.1 Deterministic measures**



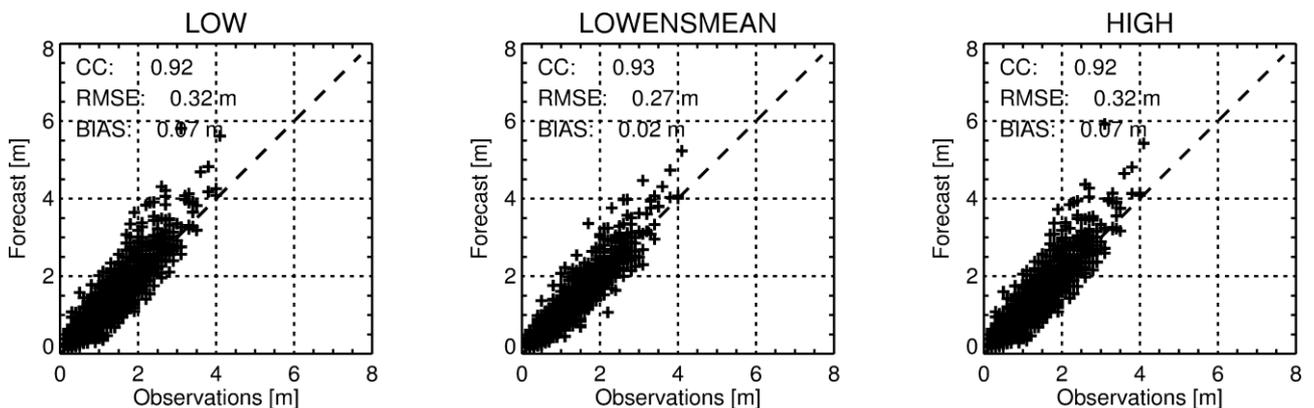
257

258 Figure 4 Scatter plot of 24 h forecasts and corresponding observations of significant wave height at site
259 Bothnian Sea for the LOW, LOWENSMEAN and HIGH forecast classes. Dotted line is the diagonal,
260 representing a 1:1 agreement between observations and model.

261

262 To get an idea of the overall quality of the forecasts, Figure 4 shows scatter plots between 24 h forecasted
263 and observed SWH for station Bothnian Sea. The points are distributed along the diagonal in all three
264 configurations with correlation coefficients above 0.9. The RMSE is 0.33 m for both LOW and HIGH but is
265 lower at 0.29 m for the LOWENSMEAN forecasts, which also have the numerically lowest bias. Also for
266 other sites, such as Arkona WR (see Figure 5), the RMSE for LOWENSMEAN forecasts is lower than for the
267 LOW and HIGH forecasts, and similarly for the bias. However, the scatter plot appears differently for this
268 station, because there is a tendency for over-predicting high waves for all three forecast classes.

269



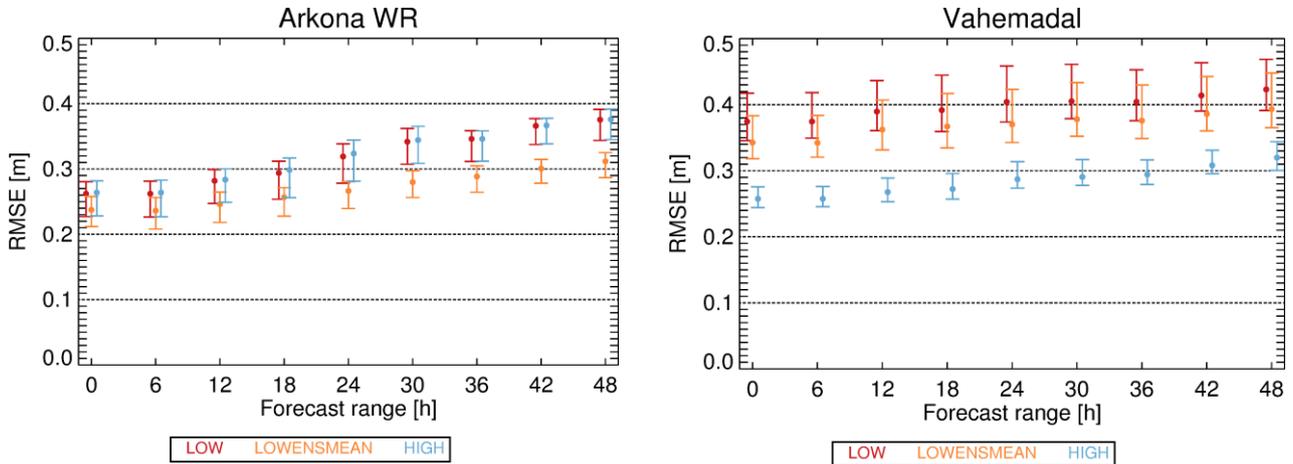
270

271 Figure 5 As Figure 4 but for site Arkona WR.

272 We now turn to the RMSE as function of forecast range, of which plots for all sites can be found in Figure
273 S2. For all sites, the RMSE increases slightly as function of forecast range. All sites except Vahemadal exhibit
274 qualitatively similar behaviour: the RMSE for the LOW and HIGH forecasts are almost similar, while it is
275 lower for the LOWENSMEAN forecasts. Thus, for Arkona WR (shown in Figure 6), Bothnian Sea and Darss
276 Sill WR, the RMSE of the LOW and the HIGH forecasts have overlapping confidence bands. The RMSE for

277 LOWENSMEAN gradually diverges to a lower value (around 5 cm) and for large forecast ranges, the
 278 confidence bands do not overlap with those for the LOW and HIGH forecast classes. The remaining sites
 279 except Vahemadal behave similarly, but with overlapping confidence bands even for the largest forecast
 280 ranges.

281



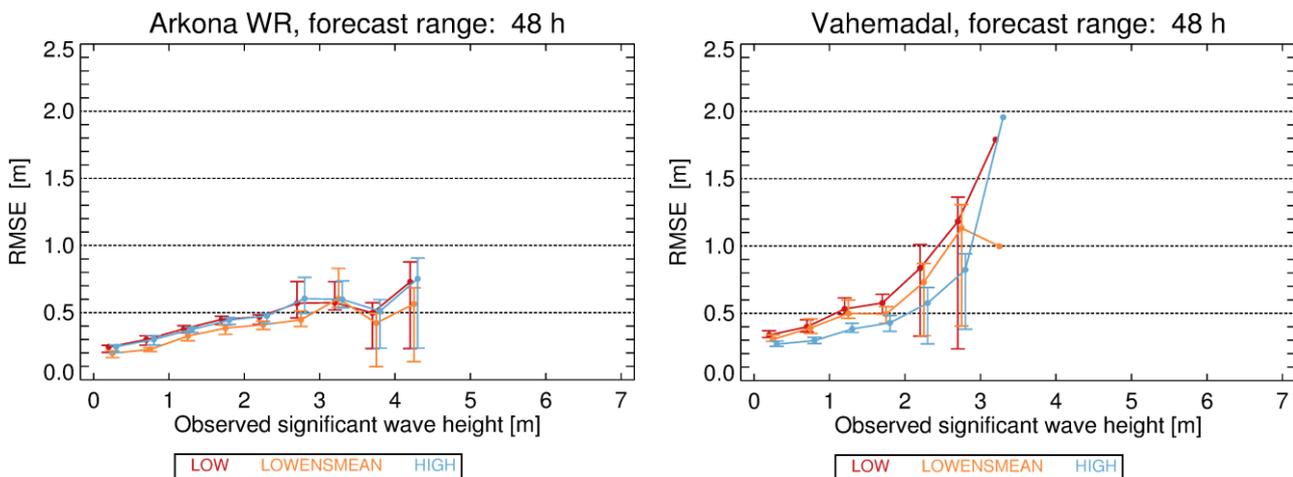
282 **Figure 6** RMSE for selected forecast ranges for Arkona WR (left panel) and Vahemadal (right panel) for LOW, LOWENSMEAN and
 283 HIGH forecasts. Error bars show 5/95% confidence bands calculated by bootstrapping.

284

285 The site Vahemadal (Figure 6) has a different behavior. For this site, the HIGH forecast class has a
 286 significantly smaller RMSE and with non-overlapping confidence bands with the RMSE of the LOW and
 287 LOWENSMEAN forecasts. This site also has a non-negligible bias of around 12 cm for the HIGH and around
 288 20 cm for the LOW and LOWENSMEAN forecasts; this bias is independent of forecast range (not shown).

289 6.1.1 Performance depending on observed SWH

290

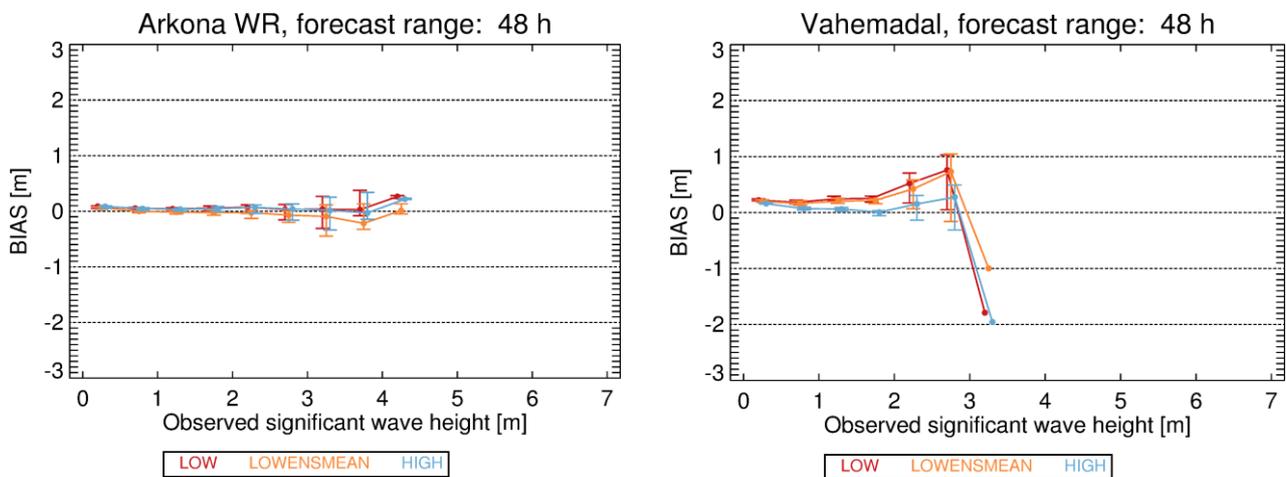


291 **Figure 7** RMSE as function of SWH for Arkona WR (left panel) and Vahemadal (right panel) for LOW, LOWENSMEAN and HIGH
 292 forecasts and forecast range 48 h. Error bars show 5/95% confidence bands calculated by bootstrapping.

293 The RMSE of the forecasts depends on the magnitude of the SWH. Plots for all sites for 24 and 48 h forecast
 294 range of RMSE as function of the SWH can be found in Figures S3 and S4. The RMSE for Arkona WR and

295 Vahemadal as function of the SWH for forecast range 48 h is shown in Figure 7. The RMSE increases as
 296 function of the observed SWH for both sites. For Arkona WR, the LOWENSMEAN forecast class has the
 297 lowest RMSE, although with confidence bands overlapping with the other forecast classes. This behavior is
 298 seen at all sites, except Vahemadal. For Vahemadal, the HIGH forecast class has the lowest RMSE, and up to
 299 a SWH of 2 m, the confidence band is well separated from the confidence bands of the other forecast
 300 classes.

301 Also the bias depends on the SWH. Plots for all sites for 24 and 48 h forecast range of the bias as function of
 302 the SWH are displayed in Figures S5 and S6. For small SWH, the bias is close to zero for most sites. For some
 303 sites, the bias remains close to zero for increasing SWH, as shown for Arkona WR in left panel of Figure 8,
 304 while for others it becomes different from zero for large values of SWH. There is no noticeable difference in
 305 the bias of the different forecast classes, except for Vahemadal, shown in right panel of Figure 8, where the
 306 HIGH forecast class has a significantly smaller bias than the other forecast classes.

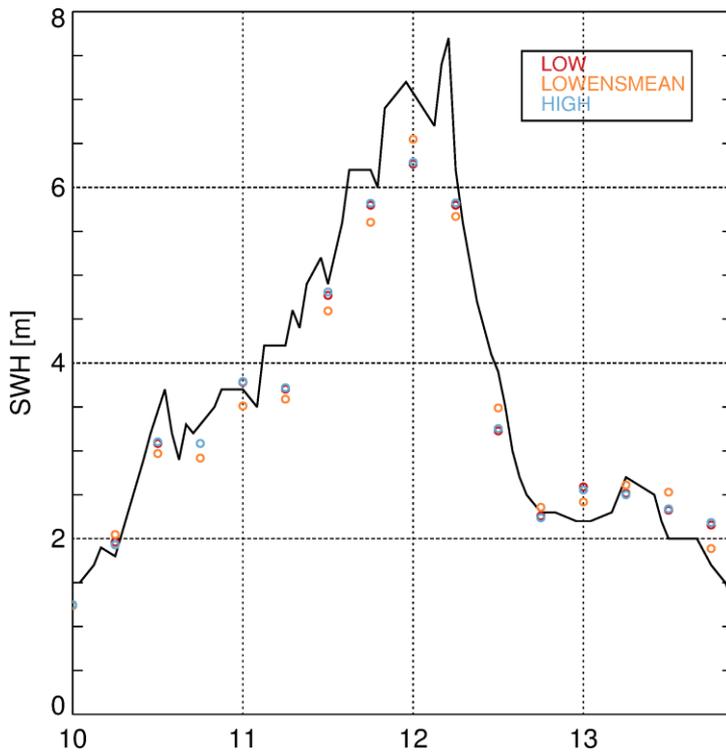


307 **Figure 8** Bias as function of SWH for Arkona WR (left panel) and Vahemadal (right panel) for LOW, LOWENSMEAN and HIGH
 308 forecasts and forecast range 24 h. Error bars show 5/95% confidence bands calculated by bootstrapping.

309 6.1.2 Forecasts during ‘Toini’ storm

310 The Toini storm on 11. January 2017, where a SWH of almost 8 m was recorded on Northern Baltic
 311 (Björkqvist et al., 2017a), is within our verification period.

312



313

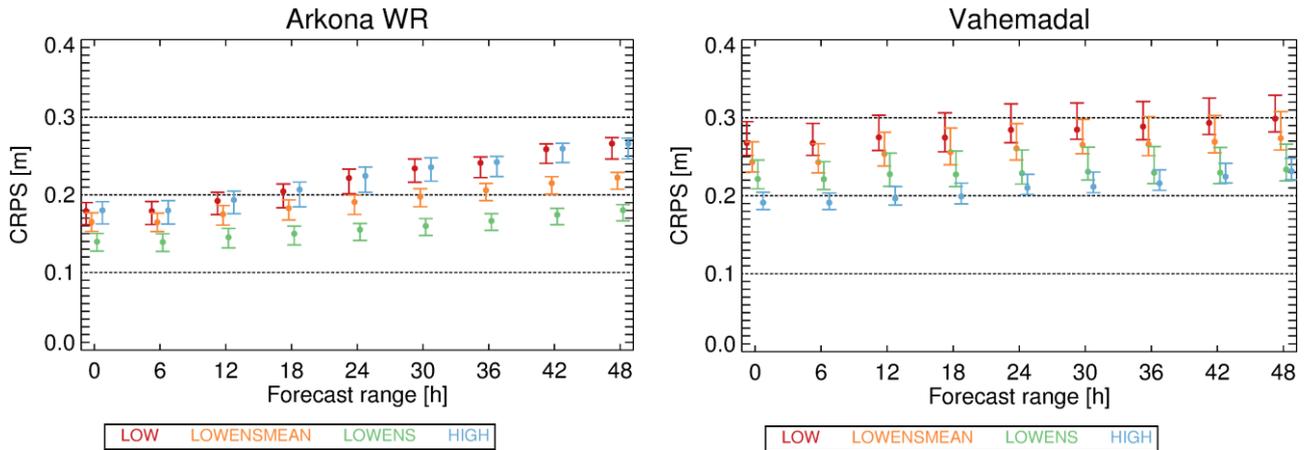
314 **Figure 9** Observed SWH for Northern Baltic during, 10-13 January 2017, including the Toini storm. Open circles are 48 h forecasts.

315 Figure 9 shows the observed SWH at Northern Baltic during 10-13 January 2017, i.e. including the Toini
316 storm, peaking in the early hours of 12 January, together with 48 h forecasts. In this case there is no
317 apparent 'best' forecast. Near the peak, LOWENSMEAN performs best, but both before and after, the
318 HIGH/LOW performs better. Further, that in most cases, the LOW and HIGH forecasts are very similar,
319 indicating that the higher resolution does not improve the forecasts.

320 **6.2 Probabilistic metrics**

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

325 As described in Section 4, we use CRPS to describe performance of probabilistic forecasts. CRPS for all sites
326 for selected forecast ranges can be found in Figure S7. As typical examples, Figure 10 displays this plot for
327 Arkona WR and Vahemadal.



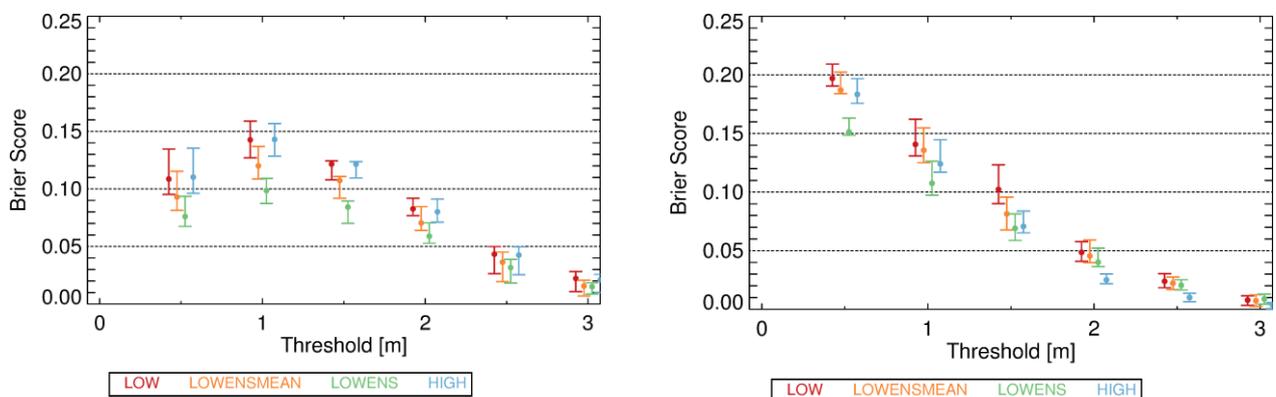
328 **Figure 10** CRPS for selected forecast ranges for Arkona WR (left panel) and Vahemadal (right panel) for LOW, LOWENSMEAN,
 329 LOWENS and HIGH forecasts. Error bars show 5/95% confidence bands calculated by bootstrapping.

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

337 6.3 Binary forecasts

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

342 The Brier Score as function of threshold is shown for all sites in Figures S7 and S8. Figure 11 shows the Brier
 343 Score as function of threshold for Arkona WR and Vahemadal for 48 h forecast range. For Arkona WR, the
 344 Brier Score for the LOWENS forecast class is the smallest, however the confidence intervals overlap with
 345 confidence intervals from the other forecasts above 2 m threshold. Also the LOWENSMEAN forecast class
 346 has low Brier Score. This behavior is common to all sites except Vahemadal. For Vahemadal, the Brier Score
 347 is smallest for the HIGH forecasts for thresholds above 1 m.



348 **Figure 11** Brier score for Arkona WR (left panel) and Vahemadal (right panel) for binary forecast for forecast range 48 h.

349 **6.4 Rank histogram**

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

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

360 **7 Discussion**

361 Our main finding in the previous section is that for most wave measurement sites included in this study, the
 362 LOWENSMEAN and the LOWENS forecast classes have a performance superior to the LOW and HIGH
 363 forecast classes. Only for one site results are different; namely that the HIGH forecast class has the superior
 364 performance. These conclusion hold, whether based on overall RMSE, CRPS or the Brier score.

365 **7.1 Comparison with other operational forecast systems**

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

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

373 **Table 6 Comparison of RMSE for SWH of 6h forecast runs for selected sites. FIMR values are from (Tuomi et al., 2008) and FMI**
 374 **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

375

376 From Table 6 one can see that for the sites considered, the LOWENSMEAN has the lowest RMSE. This
377 supports the finding of this study that for offshore conditions, there is no reason to improve the resolution
378 further than that of the LOW configuration. In addition, the results emphasize the value of describing the
379 uncertainties of in the atmospheric forcing by introducing ensembles, as this leads to a lower RMSE of the
380 forecasts. This is also in line with our findings in the previous section.

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

384 Fore completeness, we remind the reader that the cases compared in Table 6 have different wind forcing
385 and probably also different version of WAM. Therefore the differences seen cannot with certainty be
386 attributed to differences in horizontal resolution.

387 **7.2 Limitations of the study**

388 **7.2.1 Length of verification period**

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

394 Thus, the operational forecasts performed on the present system, supplemented by delayed-mode
395 forecasts has determined the three-year verification period used in our study. A longer verification period
396 could evidently have reduced the sampling uncertainty in the analyses and thereby sharpened the
397 conclusions. On the other hand, the three-year verification is not short compared to other studies, e.g.
398 Bunney and Saulter (2015) or Tuomi et al.(2017)

399 **7.2.2 Choice of observational base**

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

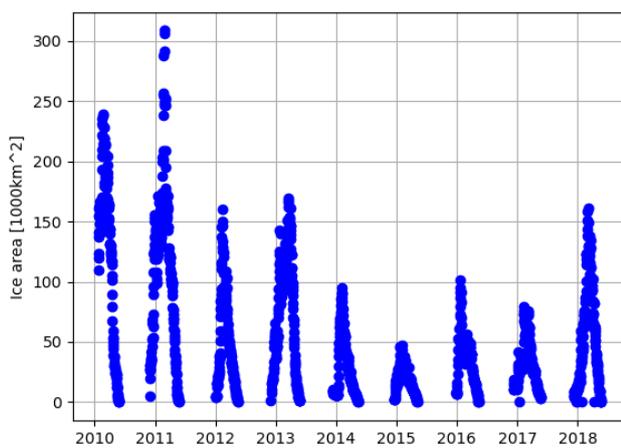
403 SWH derived from satellite-borne altimeters (Kudryavtseva and Soomere, 2016) offers an alternative,
404 which could be pursued in a future study. These data has a fair spatial data coverage but at the cost of a
405 temporal resolution of one day or less. This means that maximum wave heights connected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

408 **7.3 Effect of sea ice coverage**

409 The main effect of sea ice on formation of waves is to limit the fetch. Furthermore, when a developed wave
410 field approach an ice-covered area, the wind and the waves decouple, so that the waves act more like
411 swell, propagating through ice-covered areas while losing energy by breaking up the ice cover. The WAM
412 model does not account for such interactions, and sea ice, when dense enough, act as a solid shield that
413 effectively remove all local wave energy in the model. It is implicitly assumed that dense ice will also be

414 thick enough for this to approximately correct. In the Baltic Sea, that may not always be the case, and
415 therefore sea ice occurrence may represent a systematic error source in the present study. Another effect
416 of sea ice in the Baltic is that the wave observing systems are withdrawn, when ice is expected. This may
417 cause a systematic bias in the verification analysis, if strong winds during winter are left out.

418 Based on Copernicus sea ice charts produced by the Finnish Meteorological institute the ice conditions for
419 the Baltic have been evaluated. The Finnish ice charts are produced on a grid of approximately 1 km² with a
420 temporal resolution of approximately one day in the ice season. Data is available from 2010 onwards. The
421 average ice conditions for February for all years and the three years in focus can be found in Figure S12. All
422 three years 2015-2017, and in particular 2015, have a smaller ice cover relative to the period 2010-2018.



423
424 **Figure 12 Integrated sea ice area of the Baltic Sea based on Finnish ice charts**

425 Another way to illustrate this is considering the Baltic Sea integrated sea ice area, depicted in Figure 12,
426 which shows that the years 2015-2017 have the lowest sea ice area over the whole period 2010-2018.
427 Therefore, we may anticipate that systematic errors arise from the occurrence of sea ice are relatively
428 small.

429 **8 Conclusion**

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

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

440 For high wave heights, there are significant systematic biases for most sites shared among all three forecast
441 configurations. These are most probably to be ascribed to model deficiencies and act to mask any
442 differences in performance between the different forecast classes. Also the RMSE becomes large for large
443 observed SWH. This is expected since small timing errors in the predicted wave time series will have larger
444 impacts on the model-observation match-up when the SWH is large. The present study therefore suggests
445 that for offshore conditions, there are no indications of further increase of the resolution of the WAM
446 model will result in enhanced forecast performance. In addition, the results show that introducing
447 ensembles increases the performances. This is both true for deterministic forecast in the form of ensemble
448 mean and for probabilistic forecast.

449 For nearshore conditions conclusions are based on only one site, but results from this indicates that
450 increasing the resolution gives better forecasts, while introducing ensembles does not. This can be due to
451 both enhanced spatial resolution, allowing a better representation of shadow and shallow water effects,
452 and/or spectral resolution.

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

461

462

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

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

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

470

471 **References**

472 Alari, V., Staneva, J., Breivik, Ø., Bidlot, J.-R., Mogensen, K. and Janssen, P.: Surface wave effects on water
473 temperature in the Baltic Sea: simulations with the coupled NEMO-WAM model, *Ocean Dyn.*, 66(8), 917–
474 930, 2016.

- 475 Alves, J.-H. G., Wittmann, P., Sestak, M., Schauer, J., Stripling, S., Bernier, N. B., McLean, J., Chao, Y.,
 476 Chawla, A., Tolman, H. and others: The NCEP–FNMOOC combined wave ensemble product: Expanding
 477 benefits of interagency probabilistic forecasts to the oceanic environment, *Bull. Am. Meteorol. Soc.*, 94(12),
 478 1893–1905, 2013.
- 479 Amante, C. and Eakins, B. W.: ETOPO1 1 ARC-MINUTE GLOBAL RELIEF MODEL: PROCEDURES, DATA
 480 SOURCES AND ANALYSIS., 2009.
- 481 Battjes, J. A. and Janssen, J. P. F. M.: Energy Loss and Set-Up Due to Breaking of Random Waves, in *Coastal
 482 Engineering 1978.*, 1978.
- 483 Behrens, A.: Development of an ensemble prediction system for ocean surface waves in a coastal area,
 484 *Ocean Dyn.*, 65(4), 469–486, doi:10.1007/s10236-015-0825-y, 2015.
- 485 Björkqvist, J.-V., Tuomi, L., Tollman, N., Kangas, A., Pettersson, H., Marjamaa, R., Jokinen, H. and Fortelius,
 486 C.: Brief communication: Characteristic properties of extreme wave events observed in the northern Baltic
 487 Proper, Baltic Sea, *Nat. Hazards Earth Syst. Sci.*, 17(9), 1653–1658, doi:10.5194/nhess-17-1653-2017, 2017a.
- 488 Björkqvist, J.-V., Tuomi, L., Fortelius, C., Pettersson, H., Tikka, K. and Kahma, K. K.: Improved estimates of
 489 nearshore wave conditions in the Gulf of Finland, *J. Mar. Syst.*, 171, 43–53,
 490 doi:10.1016/j.jmarsys.2016.07.005, 2017b.
- 491 Bunney, C. and Saulter, A.: An ensemble forecast system for prediction of Atlantic–UK wind waves, *Ocean
 492 Model.*, 96, 103–116, doi:10.1016/j.ocemod.2015.07.005, 2015.
- 493 Cao, D., Tolman, H. L., Chen, H. S., Chawla, A. and Gerald, V. M.: Performance of the ocean wave ensemble
 494 forecast system at NCEP, in *The 11th International Workshop on Wave Hindcasting & Forecasting and 2nd
 495 Coastal Hazards Symposium*. [online] Available from:
 496 <http://nopp.ncep.noaa.gov/mmab/papers/tn279/mmab279.pdf> (Accessed 22 September 2017), 2009.
- 497 Carrasco, A. and Saetra, Ø.: A limited-area wave ensemble prediction system for the Nordic Seas and the
 498 North Sea, *Norwegian Meteorological Institute.*, 2008.
- 499 Cavaleri, L., Fox-Kemper, B. and Hemer, M.: Wind Waves in the Coupled Climate System, *Bull. Am.
 500 Meteorol. Soc.*, 93(11), 1651–1661, doi:10.1175/BAMS-D-11-00170.1, 2012.
- 501 Cavaleri, L., Abdalla, S., Benetazzo, A., Bertotti, L., Bidlot, J.-R., Breivik, Ø., Carniel, S., Jensen, R. E., Portilla-
 502 Yandun, J., Rogers, W. E., Roland, A., Sanchez-Arcilla, A., Smith, J. M., Staneva, J., Toledo, Y., van Vledder, G.
 503 P. and van der Westhuysen, A. J.: Wave modelling in coastal and inner seas, *Prog. Oceanogr.*,
 504 doi:10.1016/j.pocean.2018.03.010, 2018.
- 505 Günther, H., Hasselmann, S. and Janssen, P. A. E. M.: The WAM Model cycle 4, World Data Center for
 506 Climate (WDCC) at DKRZ., 1992.
- 507 Hasselmann, K., Barnett, T., Bouws, E., Carlson, H., Cartwright, D., Enke, K., Ewing, J., Gienapp, H.,
 508 Hasselmann, D., Kruseman, P., Meerburg, A., Müller, P., Olbers, D., Richter, K., Sell, W. and Walden, H.:
 509 Measurements of wind-wave growth and swell decay during the {Joint North Sea Wave Project}, *Deut
 510 Hydrogr Z*, 8(12), 1–95, 1973.
- 511 Jolliffe, I. T. and Stephenson, D. B.: *Forecast verification: a practitioner’s guide in atmospheric science*, John
 512 Wiley & Sons., 2003.

- 513 Kudryavtseva, N. A. and Soomere, T.: Validation of the multi-mission altimeter wave height data for the
514 Baltic Sea region, *Est. J. Earth Sci.*, 65(3), 161, doi:10.3176/earth.2016.13, 2016.
- 515 Saetra, Ø. and Bidlot, J.-R.: Assessment of the ECMWF Ensemble Prediction System for Waves and Marine
516 Winds, European Centre for Medium-Range Weather Forecasts., 2002.
- 517 Schaffer, J., Timmermann, R., Arndt, J. E., Kristensen, S. S., Mayer, C., Morlighem, M. and Steinhage, D.: A
518 global, high-resolution data set of ice sheet topography, cavity geometry, and ocean bathymetry, *Earth
519 Syst. Sci. Data*, 8(2), 543–557, doi:10.5194/essd-8-543-2016, 2016.
- 520 She, J., Allen, I., Buch, E., Crise, A., Johannessen, J. A., Le Traon, P.-Y., Lips, U., Nolan, G., Pinardi, N.,
521 Reißmann, J. H., Siddorn, J., Stanev, E. and Wehde, H.: Developing European operational oceanography for
522 Blue Growth, climate change adaptation and mitigation, and ecosystem-based management, *Ocean Sci.*,
523 12(4), 953–976, doi:10.5194/os-12-953-2016, 2016.
- 524 Tuomi, L., Kangas, A., Leinonen, J. and Boman, H.: The Accuracy of FIMR Wave Forecasts in 2002-2005.,
525 2008.
- 526 Tuomi, L., Kahma, K. K. and Pettersson, H.: Wave hindcast statistics in the seasonally ice-covered Baltic sea.,
527 *Boreal Environ. Res.*, 16, 2011.
- 528 Tuomi, L., Vähä-Piikkiö, O. and Alari, V.: Baltic Sea Wave Analysis and Forecasting Product
529 BALTICSEA_ANALYSIS_FORECAST_WAV_003_010, 2017.
- 530 Zijlema, M.: Computation of wind-wave spectra in coastal waters with SWAN on unstructured grids, *Coast.
531 Eng.*, 57(3), 267–277, doi:10.1016/j.coastaleng.2009.10.011, 2010.
- 532