Supplementary Material S1: Brief assessment of atmospheric forcing data

Swedish stations				
Station name	Latitude	Longitude	Number of observations	Covering period (years)
Almagrundet_A	59.155	19.129	100495	11,472
Holmögadd_A	63.595	20.756	62876	7,178
Skagsudde_A	63.187	19.020	198569	22,668
Holmön_A	63.807	20.864	118794	13,561
Skarpö_A	59.300	18.705	192023	21,920
Bjuröklubb_A	64.480	21.575	195433	22,310
Svenska_Högarna_A	59.442	19.502	73490	8,389
Brämön_A	62.220	17.745	195728	22,343
Järnäsklubb_A	63.437	19.677	198430	22,652
Söderarm_A	59.753	19.409	195703	22,341
Eggegrund_A	60.730	17.552	191192	21,826
Kuggören_A	61.703	17.522	200112	22,844
Trubaduren_Aut	57.596	11.635	77047	8,795
Fårösund_Ar_A	57.917	18.953	197100	22,500
Landsort_A	58.743	17.868	200005	22,832
Utklippan_A	55.956	15.705	183726	20,973
Lungö_A	62.643	18.093	193779	22,121
Vinga_A	57.632	11.606	106863	12,199
Gotska_Sandön_A	58.394	19.194	196523	22,434
Måseskär_A	58.094	11.331	195814	22,353
Väderöarna_A	58.576	11.068	189533	21,636
Gustaf_Dalen_A	58.593	17.470	97111	11,086
Nidingen_A	57.303	11.904	198337	22,641
Hallands_Väderö_A	56.450	12.545	193480	22,087
Nordkoster_A	58.894	11.009	173789	19,839
Ölands_norra_udde_A	57.367	17.095	200163	22,850
Hanö_A	56.013	14.850	195340	22,299
Ölands_södra_udde_A	56.198	16.400	197888	22,590
Pite-Rönnskär_A	65.035	21.566	197705	22,569
Örskär_A	60.526	18.376	198210	22,627
Harstena_A	58.251	17.011	199640	22,790
Rödkallen_A	65.312	22.371	182104	20,788
Östergarnsholm_A	57.441	18.984	165108	18,848

List of Swedish stations used for validation.

A comprehensive validation of the UERRA-HARMONIE data set is beyond the scope of this study. Instead we focus on the basic parameters necessary to force ocean models, here, for the Baltic Sea. The validation of atmospheric variables over open ocean areas is challenging, as independent, long-term, observational reference data sets are typically rare. We compiled data sets for near-surface air temperature and wind from various sources. Wind and temperature data were obtained from the monitoring stations of the Marine Environment Observation Network (MARNET, https://www.bsh.de/DE/THEMEN/Beobachtungssysteme/Messnetz-MARNET/messnetz-marnet_node.html) and from stations of the Deutscher Wetterdienst (DWD; the German Meteorological Service) and were used together with measurements from stations along the Swedish coast (Höglund et al., 2009; Meier et al., 2011). The quality of the UERRA-HARMONIE data set was compared to that of ERA5 (Hersbach et al. 2020), the well-known global reanalysis of ECMWF, which is also available in the CDS. ERA5 has a horizontal resolution of 31 km, which is larger than the 11 km of UERRA-HARMONIE. The comparison reinforces the choice of high-resolution forcing data for regional ocean model setups.

Wind speed

Validation data for wind speed were based on Swedish observations and were gathered from 33 automated stations along the Swedish coastline. Since 1996, data have usually been collected hourly and all considered stations have at least 7 years of data. The station names, locations, and the number of available measurements are listed in Supplementary Material S1. For comparisons of station-based observations with the gridded UERRA-HARMONIE reanalysis data set, the nearest model grid point to each observation station was selected.

The key numbers for wind speed are provided in Table 1. Whereas according to observations the average wind speed was 6.01 m/s the corresponding values were slightly higher for UERRA (6.14 m/s) and slightly lower for ERA5 (5.95 m/s). It should be noted that the average wind speed over the entire Baltic Sea is generally faster in UERRA than in ERA5, by 0.2–0.4 m/s. This difference is quite homogeneous over the open sea but the differences become much larger and in some cases of opposite sign along the coastline (not shown). For the Swedish coastal stations, a better match with observations was obtained with UERRA than with ERA5. The mean absolute error of UERRA was 0.59 m/s and that of ERA5 0.84 m/s. Also, the variability of the wind speed was better depicted by UERRA than by ERA5. Considering the difference in the standard variation at each location and computing the absolute mean over all 33 stations yielded a mean absolute standard variation of 0.30 m/s for UERRA and 0.57 m/s for ERA5 (see Table 1). In terms of the correlation of hourly values, both UERRA and ERA5 matched the observations reasonably well, evidenced by a correlation of 0.85 in each case. In terms of the root mean square error (RMSE), UERRA (1.89 m/s) performed slightly better than ERA5 (1.95 m/s).

	Observations	**UERRA**	**ERA5**
Mean [m/s]	6.01	6.14	5.95
Mean bias [m/s]		0.13	-0.06
Mean abs. Error [m/s]		0.59	0.84
Average standard deviation [m/s]	3.19	3.14	2.72
Mean difference in STD [m/s]		-0.05	-0.47
Mean abs. STD [m/s]		0.30	0.57
RMSE [m/s]		1.89	1.95
Correlation		0.85	0.85

Table 1: Wind speed statistics for the 33 Swedish coastal stations for UERRA-HARMONIE and ERA5 compared to the 33 Swedish coastal stations. The statistics are based on hourly data and were computed for each individual observational side before they were averaged into a reference number.

Air temperature

Table 3 summarizes the statistics for the near-surface air temperature at the same stations. All data sets well represented the average near-surface temperature, as indicated by the low mean bias (Table 3). The RMSE indicated the performance on sub-synoptical time scales, which reflected a significant portion of stochastic noise or turbulence within the boundary layer. As such processes were not well represented, the deviation was quite large, exceeding 1 Kelvin at each station. For climate purposes, this is less important but it should be considered in high-resolution applications. The quality of UERRA-HARMONIE and ERA5 for the surface air temperature was about the same. For both data sets, the strong correlation with the hourly station data indicated that the diurnal cycle was well reproduced.

	UERRA	**ERA5**
Mean Bias [K]	-0.00	0.01
Mean abs. Bias [K]	0.15	0.14
Mean std [K]	-0.37	-0.24
Mean abs. Std [K]	0.38	0.36
RMSE [K]	1.46	1.42
Correlation	0.982	0.982

Table 2: Near-surface temperature statistics for the 33 Swedish coastal stations according to UERRA-HARMONIE and ERA5 compared to the data from 33 Swedish coastal stations. Statistics are based on hourly data and were computed for each observational site before averaging into a reference number.

Precipitation

Precipitation was compared to E-OBS version 19.0e at 0.1° horizontal resolution (Cornes et al. 2018) and HydroGFD3 (Berg et al. 2021, hereafter HGFD). Here, we considered monthly values for the period 1979–2015 as well as area averages roughly for the Baltic Sea proper and for the surrounding areas. The chosen domain for the precipitation analysis was 13°E–32°E and 52°N– 3°N. The surrounding area had to be included because E-OBS does not contain data over the ocean. For HGFD, there is a version with data from over the ocean but these data are a direct copy of ERA5, which was used to design HGFD. Total precipitation is an important contribution to the Baltic Sea's freshwater balance but, in contrast to wind speed, short-term variations are less important such that the analysis was limited to monthly data.

The main results describing precipitation are presented in Table 3. The most important variable for Baltic Sea modeling was the average precipitation, since it had a direct impact on the freshwater balance. UERRA-HARMONIE clearly overestimated the amount of total precipitation in the Baltic Sea area. With an average precipitation of 68 mm/m^2, it exaggerated the data sets based on observations. HGFD3 estimated the average precipitation as 55 mm/m^2 while in E-OBS it was 52 mm/m^2. However, in terms of variability there was good correspondence to the observation-based datasets. UERRA-HARMONIE had a correlation of 0.93 with HGFD3 and 0.96 with E-OBS (Table 3). Hence, seasonal as well as year to year variability were well depicted. For the purpose of the BMIP, the total precipitation from UERRA-HARMONIE was reduced by 20% to adjust the level to observed values. The comparison with ERA5 was equalized to the validation with the E-OBS data, since HGFD is related to ERA5. In general, ERA5 seems to outperform UERRA-HARMONIE in terms of precipitation in the Baltic Sea area. Although ERA5 also overestimated the amount of precipitation, the overestimation was not as large as that of UERRA-HARMONIE. Also, in terms of the correlation and RMSE (see Table 3), ERA5 performed better than UERRA-HARMONIE when either one was compared to E-OBS. This result was consistent with the fact that the validation was based on monthly means as well as on averages over a rather large area, and global products are designed to match the broad picture. By contrast, regional products are needed for small-scale variability. However, the latter is not important for precipitation in the context of Baltic Sea climate modeling.

	Monthly mean [mm]	Correlation with HGFD	RMSE with HGFD [mm]	Correlation with E-OBS	RMSE with E-OBS [mm]
HGFD3	54,5	---	---	0,985	4,573
E-OBS	52,2	0,985	4,573	---	---
UERRA-HARMONIE	68,1	0,927	16,516	0,956	17,607
ERA5	61,4	0,990	7,603	0,987	9,923

Table 3: Precipitation statistics comparing UERRA-HARMONIE and ERA5 data with observations from HGFD3 (Berg et al., 2021) and E-OBS (Cornes et al., 2018). Statistics were computed for each observational site and the values then averaged into a reference number.

Supplementary Material S2: Brief model descriptions

GETM

The General Estuarine and Transport Model (GETM, Burchard and Bolding, 2002; Klingbeil & Burchard, 2013) was run in two different configurations and with a horizontal resolution of 250 m, 1 nm, and 2 nm. GETM computes on an Arakawa-C-grid and utilizes mode-splitting between the barotropic and baroclinic modes. In the vertical, both setups employ 60 terrain-following vertical adaptive layers (Hofmeister et al., 2010), with a minimum layer thickness of 30 cm. The thickness of the surface layer is limited to a maximum thickness of 25 cm to resolve the surface-layer physics correctly. The SST and surface currents were fed into the bulk formulae of Kara et al. (2005) to compute all necessary surface fluxes. To account for ice coverage, the thermodynamic sea-ice-model of Winton (2000) was used. Horizontal viscosity was parameterized according to Smagorinsky (1963) using a Smagorinsky constant of 0.2 and a turbulent Prandtl number of 2.0 for tracers. To account for unresolved submesoscale processes, the minimum horizontal viscosity was set at 1.5 m^2/s in the 1-nm setup and 3.2 m^2/s in the 2-nm setup. Vertical mixing was parameterized by means of a two-equation k-epsilon turbulence model with an algebraic second-moment closure (Umlauf & Burchard, 2005). A background turbulent kinetic energy level of k_min = 5*10-8 m^2 s^2 was defined (see also Holtermann et al., 2014), with a stratification limitation of the turbulent length scale by the Ozmidov scale as proposed by Galperin (1988). To account for wind-wave induced turbulence, we implement the Langmuir turbulence parameterization of Axell (2002), which is directly coupled with the production term in the turbulence closure. For further details of model setup and validation, the reader is referred to Gräwe et al. (2019) or Radtke et al. (2020).

MOM

The Modular Ocean Model (MOM; version 5.1) uses an open boundary condition connecting the Baltic Sea and the North Sea with an explicit free surface. In addition, mode splitting between the barotropic and baroclinic modes is implemented. MOM uses an Arakawa-B-grid, which places tracer variables in the middle of each cell with the corresponding velocities at the northeast corner. To calculate transports at least two adjacent cells are needed, since a no-slip boundary condition is used. Vertical diffusion is implemented using the KPP boundary layer mixing scheme [Large et al., 1994], and horizontal diffusion is parameterized using a Smagorinsky scheme (Smagorinsky, 1963).

MOM is used in the BMIP with horizontal resolutions of the model grid of three and one nautical miles. The vertical resolution consists of 152 z-layers with a layer thickness of 0.25 m at the top, which gradually increases towards the bottom, resulting in a maximum layer thickness of 2 m.

The integrated ice model is directly coupled to the ocean model and converts the fluxes at the atmosphere-ocean interface (Winton, 2000; Hunke and Dukowicz (1997). The initial conditions for the BMIP run were those in Neumann et al. (2017).

HBM

The Danish Meteorological Institute (DMI) utilizes the regional three-dimensional ocean model HBM (the HIROMB-BOOS Model) for the North and Baltic Seas in order to provide 5-day forecasts of the physical state of Danish and adjacent waters (Berg and Poulsen, 2012). In the BMIP, this operational model system is used to simulate the spatial and temporal evolution of ocean states. The use of operational models in climate studies enables long-term simulations with the same level of detail as in ocean forecasting and ensures a well-tested and validated system. The HBM model describes a hydrostatic, free surface, baroclinic ocean circulation, with a sea-ice model optimized for two-way nesting in complex bathymetries. The model is capable of multi-decadal hindcast simulations (Madsen et al., 2015; Tian et al., 2016; Andree et al., 2020) as well as operational forecasting (Huess and Woge Nielsen, 2019); different versions of the model code are currently used for operational water-level and storm-surge forecasting in several countries bordering the Baltic Sea (Capet et al., 2020). The model code and setup used in this study were from the DKSS2013 operational version launched in October 2013 at the DMI, with 17 tidal constituents and sea-level heights at the North Sea open boundary provided by a 2D model covering the northeastern North Atlantic (NOAMOD, She et al., 2007). In the German Bight and Inner Danish Waters, the DKSS2013 version features two finer nests with three and six times the resolution of the previous version. Daily average 3D fields and hourly 2D fields are archived on their native grids. The Hydrological Predictions for the Environment model for Europe (E-HYPE) provides data on river runoff (Donnelly et al., 2016). The simulation was run in parallel mode on the DMI's high-performance computing system using eight nodes (32 CPUs per node). For a year, the integration wall-clock time is approximately 24 hours, excluding the queue waiting time.

NEMO-Nordic

NEMO-Nordic is used for operational services and research applications (Hordoir et al., 2019). The model is based on the Nucleus for European Modeling of the Ocean (NEMO) framework (Madec, 2015). Within the BMIP, NEMO-Nordic is set up for the North Sea and Baltic Sea, with a horizontal resolution of 2 nautical miles. NEMO has a free surface and the water column is divided into 56 z* levels. Hence, local layer thicknesses are re-scaled at every time integration step (180 seconds) due to sea-surface undulations (see NEMO reference manual, Madec, 2015). The turbulence closure is based on a $k-\epsilon$ turbulence scheme (Hordoir et al., 2019). Like HBM, NEMO-Nordic includes the North Sea, and tidal harmonics are defined on the sea surface elevation, along the lateral open model boundaries of the English Channel at ~4°W and the northern North Sea at ~66°N. Nemo-Nordic includes the LIM3 multi-class sea ice model (Vancoppenolle et al., 2009), which was validated in the Baltic Sea by Pemberton et al. (2017).

Supplementary Material S3: same as Figure 3 in the main text but using a Baltic Sea reanalysis (Liu et al., 2017) as the reference data set.

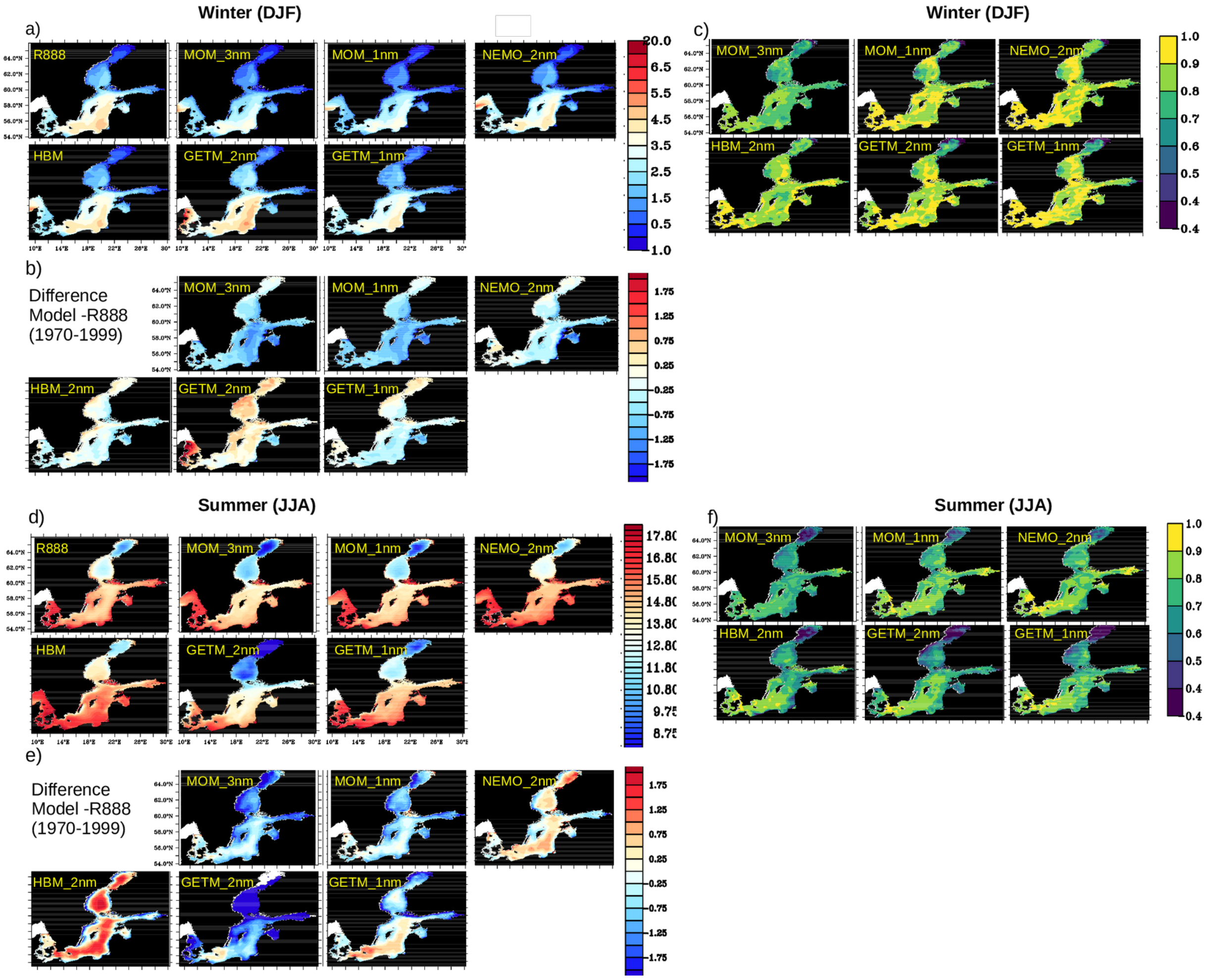

Figure: a) Comparison of modeled winter SST with an oceanographic reanalysis (Liu et al. 2017). b) Difference between the models and the reanalysis for winter. c) Inter-annual correlation of winter sea surface temperature between models and the reanalysis product. d-f) same as a-c) but for summer climatology.

Supplementary Material S4: Comparison of modeled SST Cycle and BED data

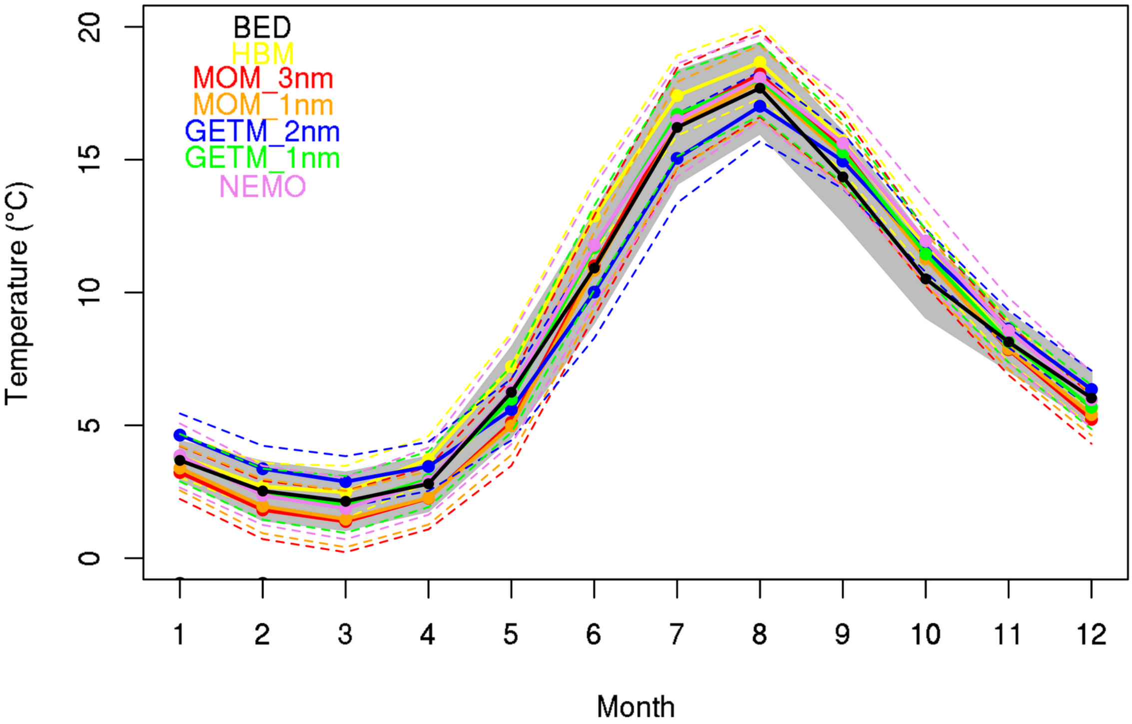

Figure: Thirty-year (1970–1999) average mean seasonal cycle of sea surface temperature at monitoring station BY15. Dashed colored lines indicate the multi-year standard deviation. The black line indicates the observational data from the Baltic Environmental Database (BED) of the Baltic NEST Institute, and the shaded area the multi-year standard deviation of the BED data.

Supplementary Material S5: Comparison of modeled deep-water salinity (colored lines) at station BY15, using data from the Baltic Environmental Database (black line).