https://doi.org/10.5194/bg-2023-152
Preprint. Discussion started: 21 September 2023
© Author(s) 2023. CC BY 4.0 License.

Biogeosciences Discussions

Ocean models as shallow sea oxygen deficiency assessment tools: from research to practical application

Sarah Piehl[1], René Friedland[1], Thomas Neumann[1], Gerald Schernewski[1,2]

[1]Coastal & Marine Management Group, Leibniz Institute for Baltic Sea Research Warnemünde, Rostock, 18119, Germany
[2]Marine Research Institute, Klaipeda University, Klaipeda, 92294, Lithuania

Correspondence to: Sarah Piehl (sarah.piehl@io-warnemuende.de)

Abstract.

Oxygen is a key indicator of ecosystem health and part of environmental assessments used as a tool to achieve a healthy ocean. Oxygen assessments are mostly based on monitoring data that are spatially and temporally limited, although monitoring efforts have increased. This leads to an incomplete understanding of the current state and ongoing trends of the oxygen situation in the oceans. Ocean models can be used to overcome spatial and temporal limitations and provide high-resolution 3D oxygen data but are rarely used for policy-relevant assessments. In the Baltic Sea where environmental assessments have a long history and which is known for the world's largest permanent hypoxic areas, ocean models are not routinely used for oxygen assessments. Especially for the increasingly observed seasonal oxygen deficiency in its shallower parts, current approaches cannot adequately reflect the high spatio-temporal dynamics. To develop a suitable shallow water oxygen deficiency assessment method for the western Baltic Sea, we evaluated first the benefits of a refined model resolution. Secondly, we integrated model results and observations by a retrospective fitting of the model data to the measured data using several correction functions as well as a correction factor. Despite its capability to reduce the model error, none of the retrospective correction functions applied led to consistent improvements. One reason is probably the heterogeneity of the used measurement data, which are not consistent in their temporal and vertical resolution. Using the Arkona Basin as an example, we show a potential future approach where only high temporal and/or vertical resolution station data is integrated with model data to provide a reliable and ecologically relevant assessment of oxygen depletion with a high degree of confidence and transparency. By doing so we further aim to demonstrate strengths and limitations of ocean models and to assess their applicability for policy-relevant environmental assessments.

© Author(s) 2023. CC BY 4.0 License.

1 Introduction

In recent decades, the oxygen content of coastal and marine waters has decreased worldwide (deoxygenation), with a considerable increase in the extent, severity, and duration of hypoxic areas (oxygen concentrations <2 mg/l), (Diaz and Rosenberg, 2008; Breitburg et al., 2018). In general, deoxygenation is reported to be higher in near-shore coastal areas than in the open ocean (Gilbert et al., 2010; Bograd et al., 2015). The predicted continuous increase of the human population, especially along the coast, climate change-induced temperature rise, and increased nutrient inputs will further accelerate the risk of deoxygenation (Rose et al., 2017).

Since oxygen is essential for the survival of higher organisms, it is a key indicator of the ecosystem health. Thus, deoxygenation is already subject of international and national directives (MSFD, 2008/56/EC, WFD, 2000/60/EC) and of environmental assessments conducted by regional conventions, such as OSPAR in the North East Atlantic or HELCOM in the Baltic Sea. These aim to limit the degradation of marine habitats due to human activities with the goal to achieve a Good Environmental Status (GES). For Europe's seas, the most important directive is the European Marine Strategy Framework Directive (MSFD) with ambitious targets and a well-defined circular implementation procedure. In this process, environmental assessments provide an overview of the overall health of the marine ecosystem at a given point in time. This includes both initial and reassessments and an evaluation of the success of measures taken against the environmental targets and/or thresholds which have been set. In this context, oxygen is already an integral part of existing environmental monitoring and assessments (Dos Santos Fernandes De Araujo et al., 2019; 2021).

The basis for any oxygen assessment is monitoring data used to assess the environmental status in a given area. For highly variable water parameters this means that the reliability of the assessment depends strongly on the quality of the monitoring data. As the frequency and number of measurements has increased from initial sporadic to regular measurements and research cruises, the magnitude of monitoring data and thus the reliability of environmental assessments has continuously improved. Permanent measuring stations that allow to collect higher temporal resolved data (Liblik et al., 2023) and autonomous measuring instruments may help to overcome spatio-temporal limitations. Nevertheless, the acquisition of monitoring data is limited due to the high cost of cruise time, equipment, and analysis procedures. In the Baltic Sea currently, regular national monitoring programs and assessment strategies for oxygen vary widely and requirements span from monthly measurements of depth profiles for a period of three years to monthly measurements of only a few months of near-bottom oxygen concentrations (IN-EUTROPHICATION 16-2020). Further, the spatial coverage of the monitoring data is very heterogeneous and varies widely at the local level. Considering, for example, the approximately 950 hydrographic monitoring stations reported for the entire Baltic Sea (https://metadata.helcom.fi/geonetwork/srv/eng/catalog.search#/metadata/49a98419-f049-47b5-a2e3-ce850fbe2f31): At the assessment unit level, coverage varies widely, with spatially better covered areas such as the Great Belt (~100 stations for 10,592 km^2) on the one hand and more poorly covered areas such as the Northern Baltic Proper (~40 stations for 32,832 km^2) on the other. The frequency of measurements ranges from 1 to 25 times per year. This spatio-

https://doi.org/10.5194/bg-2023-152
Preprint. Discussion started: 21 September 2023
© Author(s) 2023. CC BY 4.0 License.

temporal limitation of measurements means that neither the spatial extent nor the duration or frequency of seasonal and episodic hypoxia events can be reliably recorded.

In parallel with the developments in monitoring, there has been significant progress in coupled hydrodynamic-biogeochemical models (*here termed ocean models*) that can be used to simulate important water quality parameters relevant for environmental monitoring and assessments, such as oxygen. Thereby, ocean models have evolved from originally very simple one-dimensional (box) models that included only a few basic equations to three-dimensional models that describe a large number of ecosystem processes (Fennel et al., 2022). In addition to the improved understanding of relevant biogeochemical processes that made this progress possible, technical developments led to an increase in computing power and thus a significant improvement in the spatio-temporal resolution of the models (Robson, 2014; Fringer et al., 2019). Consequently, the quality of the models and the speed at which results are available have increased extremely in recent years. Ocean models can thus be used to overcome spatio-temporal limitations of monitoring data and to provide a complete three-dimensional picture with high temporal resolution. Thereby the possibilities of model products range from the assessment of the current state of marine ecosystems, the identification of ongoing trends and changes, and the assessment of the impacts of management actions and climate change. Nevertheless, current applications of ocean models are mostly limited to scientific questions and only few application examples where model simulations are used as an integral part of environmental assessments exist.

For deoxygenation most examples of model applications relate to scenario simulations and forecasting systems (Fennel et al., 2019, 2022) as for example in the Chesapeake Bay in the United States (https://news.maryland.gov/dnr/2022/11/16/final-chesapeake-bay-hypoxia-report-for-2022/) and the northern Gulf of Mexico (Laurent and Fennel, 2019). In the Baltic Sea where environmental assessments have a long history (Reusch et al., 2018) and which is known for the world's largest permanent hypoxic areas (Carstensen and Conley, 2019), models are not routinely used for policy-relevant oxygen assessments. One reason could be that the focus has been on permanent hypoxic areas for a long time, with only a recent shift in focus to increasingly observed seasonal hypoxia (Conley et al., 2011; Reusch et al., 2018; BLANO 2018; HELCOM 2023). For highly dynamic indicators such as seasonal oxygen deficiency modeling is already considered a promising assessment tool (BLANO 2018). This is especially the case as current data-based approaches lead to either under- or overestimated results on the scale of sub-basins (BLANO 2018) and making trend analysis hardly possible.

Other possible reasons why model products are not used for policy-relevant oxygen assessments in the Baltic Sea could be a lack of trust in the models as well as a lack of transparency and understanding of the limitations of the models. Moreover, model values will never perfectly align with measured values, i.e., absolute values can often be not obtained by models. However, it is important to recognize that a complete match between modeled and measured data is neither feasible nor desirable due to varying scales of each set of values (Bennett et al., 2013). To enhance the acceptance of model data in policy-relevant assessments where measured data are key for decisions, a method to integrate model and measurement data is essential to benefit from their respective strengths and to overcome current limitations.

In that way, model simulations can help to fill the gaps that currently exist in many areas due to limited observational data. For the western Baltic Sea, the added value of model simulations for the development of a suitable oxygen indicator could

https://doi.org/10.5194/bg-2023-152
Preprint. Discussion started: 21 September 2023
© Author(s) 2023. CC BY 4.0 License.

already be demonstrated using the coupled hydrodynamic biogeochemical model MOM-ERGOM (Piehl et al., 2023). However, due to the coarse resolution of three nautical miles, near-shore areas as well as areas with complex bottom topography could not be adequately represented. In this study, we therefore conducted a model run with a higher spatial resolution of one nautical mile to investigate I) if a higher spatial model resolution (and thus better representation of e.g. bottom structures) leads to a better agreement of observations and model predictions, II) which factors influence the error between measurements and model values and how to utilize them for a correction function and III) if a retrospective correction of the model results decreases the error between measurements and model values and if it leads to a more reliable assessment result for oxygen deficiency. The objective is to provide a high-resolution monitoring system for shallow water oxygen in the western Baltic Sea and to discuss general possibilities and limitations of ocean models for operational use in oxygen assessments, in particular reanalysis for the assessment of past and present conditions.

2 Materials and Methods

2.1 The ecosystem model approach

The integrated biogeochemical model ERGOM (www. ergom. net) coupled to a 3D circulation model (MOM) (Neumann and Schernewski, 2008; Neumann et al., 2021) was used for simulations of the oxygen dynamics in the western Baltic Sea (Fig. 1). In the past, the model was indeed used to analyze the extent of hypoxic areas in the Baltic Sea in relation to saltwater inflows from the North Sea, as wel as to analyze different nutrient input scenarios, and various climate change scenarios. (Neumann et al., 2017; Friedland et al., 2012). The biogeochemical model simulates the marine nitrogen and phosphorus cycle by representing the three nutrients dissolved in water, ammonium, nitrate, and phosphate, which form the basis for primary production realized by three functional groups of phytoplankton (large cells, small cells, and cyanobacteria). The model represents the grazing pressure on the phytoplankton by using a bulk zooplankton variable that develops dynamically. Dead organic material is considered by a detritus state variable. During sinking, part of the detritus is mineralized into dissolved ammonium and phosphate. The portion that arrives at the seafloor accumulates, becomes partially buried, or undergoes mineralization or resuspension in response to the velocity of near-bottom currents. The carbon cycle is coupled to the nitrogen and phosphorus cycles, as described in Kuznetsov and Neumann (Kuznetsov and Neumann, 2013).

In addition, under oxic conditions, part of the mineralized phosphate is bound by iron oxides and is retained in the sediment; under anoxic conditions, it is released into the water column. Oxygen production (via primary production) and consumption (especially due to respiration and the microbial degradation of organic matter) is coupled to biogeochemical processes through stoichiometric ratios. Processes such as the release or binding of phosphorus in the sediments, as well as denitrification and nitrification are controlled by oxygen levels. The physical part of the model system is based on the circulation model MOM (version 5.1; (Griffies et al., 2004; 2012)), adapted to the Baltic Sea with an open boundary condition to the North Sea and riverine freshwater input. The addition of a sea ice model also allows the estimation of ice thickness and extent (Winton, 2000).

https://doi.org/10.5194/bg-2023-152
Preprint. Discussion started: 21 September 2023
© Author(s) 2023. CC BY 4.0 License.

The model grid has a horizontal resolution of one nautical miles. The model is comprised of 152 vertically resolved layers with layer thicknesses ranging from 0.5 to 2 meters.

The model domain and bathymetry are shown in Fig. 1. The atmospheric forcing is derived from a dynamic downscaling of the coastDat dataset (Weisse et al., 2009; Geyer et al., 2013) with a grid resolution of approximately 25 × 25 km. For the western Baltic Sea, monthly nutrient loads were obtained from UBA, LUNG-MV and LfU-SH and provided by Denmark, Sweden and Poland. For the rest of the Baltic Sea, the latest available data on annual nutrient loads (1995-2019) for rivers and point sources (urban, industrial and aquaculture) were taken from the HELCOM-PLC database (https://nest.su.se/helcom_plc/) and converted to monthly river loads with an approach consistent with the Baltic Nest Institute. The atmospheric nitrogen deposition to the Baltic Sea for 1995-2019 was taken from EMEP (Gauss et al., 2021), which are harmonised with HELCOM. For the atmospheric phosphorus deposition, the constant value of 5 kg P/km^2 was used in accordance with HELCOM (HELCOM 2013; HELCOM 2021), whereas a gradient (computed out of the nitrogen deposition) was utilized to reflect the deposition differences in near- and offshore waters.

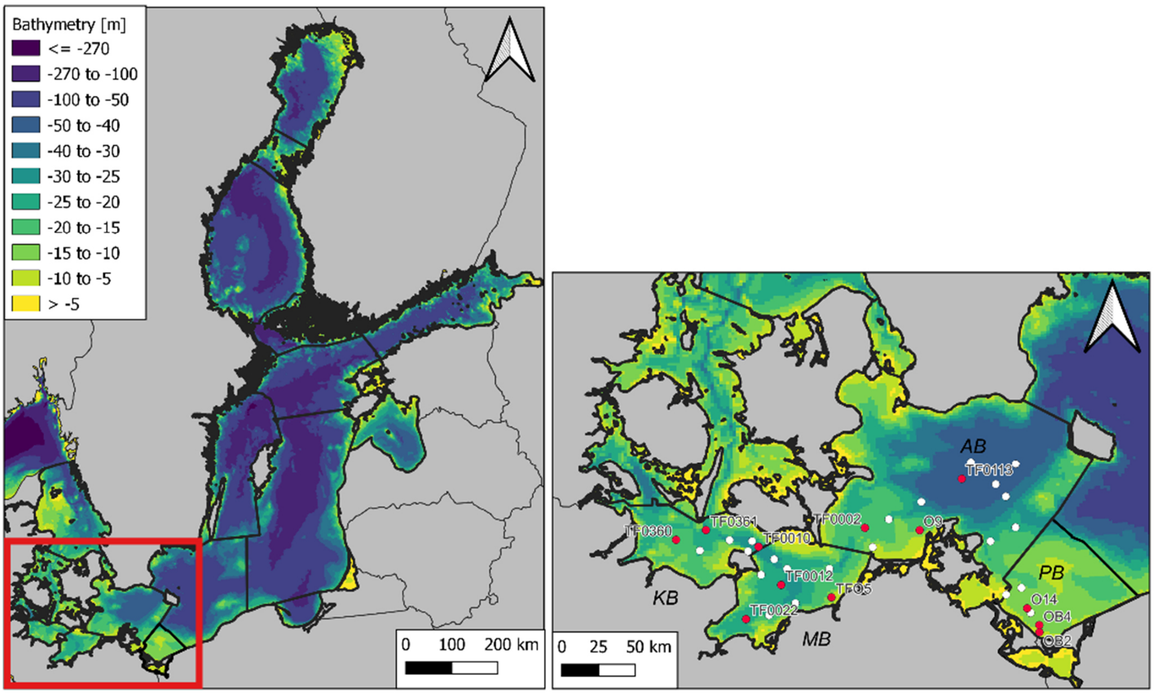


Figure 1 *The Baltic Sea in northwestern Europe (left) with sub-basin divisions according to HELCOM (black lines) and our study area (right) including the sub-basins Kiel Bay (KB), Bay of Mecklenburg (MB), Arkona Basin (AB), and Pomeranian Bay (PB; from left to right) located in the shallow western Baltic Sea including measuring stations used for model validation (red dots) and additional stations with oxygen profile data utilized for the retrospective model correction (white dots).*

https://doi.org/10.5194/bg-2023-152
Preprint. Discussion started: 21 September 2023
© Author(s) 2023. CC BY 4.0 License.

2.2 Model performance

MOM-ERGOM model results were compared with recent measurements of key parameters and stations in the Baltic Sea in several publications before (Neumann et al., 2017; Eilola et al., 2011; Neumann et al., 2020; Meier et al., 2018a). To ensure applicability for this study, the present validation emphases on oxygen dynamics in the western Baltic Sea (Fig. 1). Both, the model run with 1NM resolution and the one with 3NM resolution from Piehl et al. (2023) cover the time span 1995 to 2020 and provided the oxygen concentration as daily means. The 1995 initial conditions were taken from an earlier simulation, starting in the 1950s (see Neumann et al. 2020 for details), implying a model spin-up time of at least 45 years with realistic nutrient inputs. The period from 2011 to 2016, which covers the HELCOM assessment period of HOLAS II (HELCOM 2018), was selected to analyse the impact of the different model results on an oxygen assessment. Observational data was provided by local authorities (State Agency for Environment Schleswig-Holstein (LfU-SH, formerly LLUR-SH) and State Agency for Environment, Nature Conservation and Geology Mecklenburg-Vorpommern (LUNG-MV)) and enhanced by data from the Leibniz Institute for Baltic Sea Research Warnemünde (https://odin2.io-warnemuende.de/) and data from the ICES oceanographic database (https://www.ices.dk/data/data-portals/Pages/ocean.aspx). The oxygen depth profiles used in our study are shown in Fig. 1. The near-bottom oxygen values were derived from a harmonised dataset which combined all available near-surface oxygen values from the different sources (Vock et al. 2023; Friedland et al., 2023 accepted). For comparing the model skill of the different spatial resolutions of the model runs the four stations (TF0360, TF0022, O9, TF0113) from Piehl et al. (2023) were supplemented by eight stations to better represent the region (Fig. 1). The stations cover a depth range of 10 to 47 meters and were chosen to include vertical data.

To resolve issues related to discrepancies of different units and measuring scales among observations and models, which are compared on various temporal and spatial scales, the following steps were considered. First, oxygen values were converted to milligrams per Liter when necessary and depth measurements were aggregated to 1-meter intervals. To verify that the models are capable to reproduce the seasonal cycle and the oxygen depth profiles, skill metrics were calculated on a station basis and observations from the last years (2010–2019) were condensed to multi-annual monthly means. To analyse the spatial agreement, the average annual mean of the model results and observations were compared for the period 2010 to 2019 and averaged for each sub-basins. For the comparison of bottom oxygen values, the deepest model layers were utilized.

Three statistics most commonly used to quantify model performance (Fennel et al. 2022) were calculated:

The root mean squared error (**RMSE**) as,

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(P_i - O_i)^2}{n}} \qquad (1)$$

where O are the observations and P the model predictions.

The Pearson’s Correlation coefficient (**r**) as,

https://doi.org/10.5194/bg-2023-152
Preprint. Discussion started: 21 September 2023
© Author(s) 2023. CC BY 4.0 License.
Biogeosciences
Discussions
Open Access

$$r = \frac{\sum_{i=1}^{n} (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2 \sum_{i=1}^{n}(P_i - \bar{P})^2}} \tag{2}$$

and the average error (**AE**) as,

$$AE = \frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n} = \bar{P_i} - \bar{O_i} \tag{3}$$

In addition, the model efficiency (**ME**; also referred as to Nash-Sutcliffe Coefficient of Efficiency) was calculated as,

$$ME = 1 - \frac{\sum_{i=1}^{n}(P_i - O_i)^2}{\sum_{i=1}^{n}(\bar{O}_i - O_i)^2} \tag{4}$$

which measures how well a model predicts relative to the average of the observations (Nash and Sutcliffe, 1970). A perfect fit between model predictions and observations is indicated by a model efficiency of one. A value greater than zero means that the model is a better predictor than the observational average, while a value less than zero means that the observational average is a better predictor than the model (Lehmann et al., 2009; Stow et al., 2009; Fennel et al., 2022).

In the above equations, the sums can represent spatial or temporal averaging or both. The metrics are calculated both for the entire western Baltic Sea and for the individual sub-basins in order to detect differences in model skill among the sub-basins. The area of the bottom water layer used for calculations were computed from the model domain (3,190 km^2 for Kiel Bay, 4,229 km^2 for Bay of Mecklenburg, 16,707 km^2 for Arkona Basin, 4,956 km^2 for Pomeranian Bay). Data aggregation and analysis was performed using CDO (version 1.9.8) and R (version 4.2.2 (2022–10-31)) with the tmap package (version 3.3-3) to produce all maps.

2.3 Retrospective correction approach

To estimate the potential improvement of a posterior correction of the model results, a correction function was introduced with the aim to reduce the estimation bias of the model results. Several parameters were selected, which affected the bias (see Figure S1). As correction parameters were chosen: 1) the depth, at which an observation was made; 2) the bathymetric depth of the station; 3) the ratio of the mixed layer depth compared to the total depth (as indicator for the local stratification and the water column thickness below the halocline following (Große et al., 2016)); and 4) the distance of the station to the shoreline. All parameters were computed for each data point but were also applicable to the whole model domain to allow a correction for all simulated oxygen concentrations, independent of the single measurements. The correction function was then computed as polynoms of third degree for each parameter as:

$$corr\_func(depth, bath, MLD, dist) = \sum_{i=1}^{3}(a_i \cdot depth^i + b_i \cdot bath^i + c_i \cdot MLD^i + d_i \cdot dist^i) \tag{5}$$

where *depth* denotes the observation depth, *bath* the bathymetric depth, *MLD* the ratio of the mixed layer depth compared to total depth, and *dist* the distance to shoreline.

In order to find the best correction function, the constants a_1 to d_3 were computed for each sub-basin and month individually in order to minimize the mean Euclidian distance between the observations and the corrected model estimates, calculated as sum of the original model estimate and the correction function following Eq. (5). For the optimization procedure, the simplex search method of Lagarias et al. (1998) implemented in Matlab was used. To calibrate the correction function, different

https://doi.org/10.5194/bg-2023-152
Preprint. Discussion started: 21 September 2023
© Author(s) 2023. CC BY 4.0 License.

strategies were analysed: 1) only the profile data from the selected stations was used (see Fig. 1); 2) only the near-bottom observations were used (taken from Vock et al., 2023); and 3) a combination of 1) and 2), where all available data was used. Based on the optimized constants a_1 to d_3 for each month and sub-basin, the correction function was computed for the model grid, following Eq. (5). The values were then added to the originally modelled oxygen concentrations, accepting inconsistencies and non-continuous transitions along the sub-basin's edges as well as between single months of the posterior corrected oxygen concentrations.

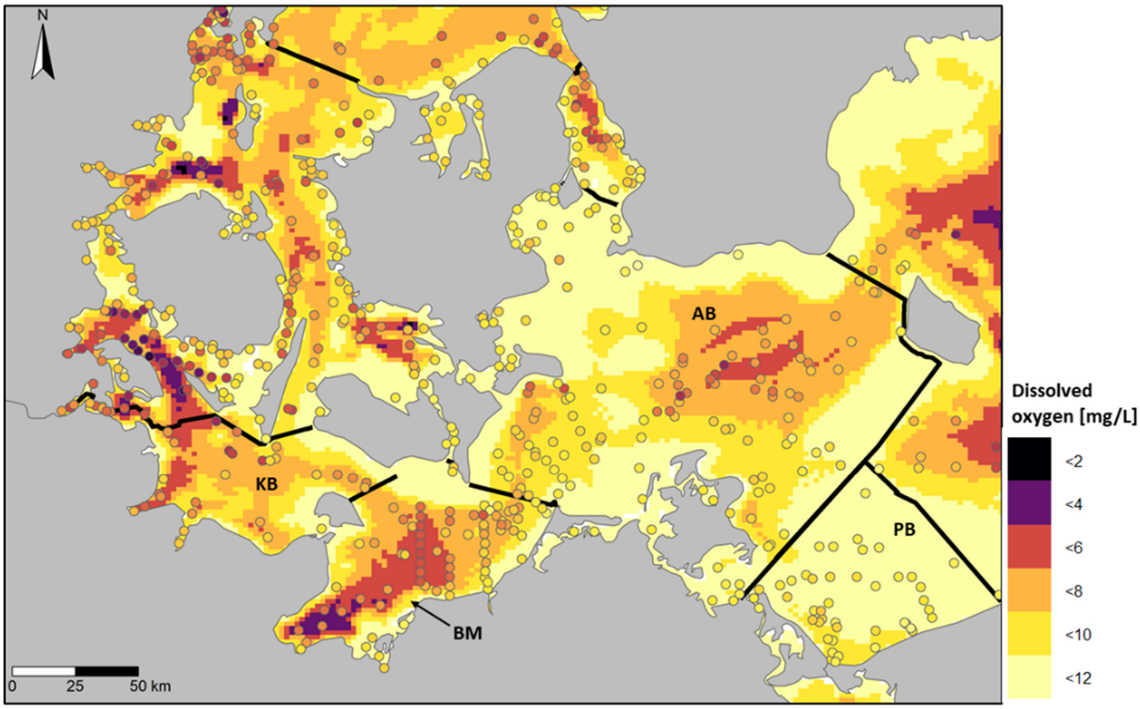
Exemplarily, two alternative approaches were tested for the Arkona Basin whereby only the temporal highly resolved data from the MARNET station Arkona Becken was used (https://www.bsh.de/EN/TOPICS/Monitoring_systems/MARNET_monitoring_network/_Module/Karussell/_documents/measuring_stations_baltic_sea_node.html). The first approach followed the one described above of applying a correction function whereas only the MARNET station data was used for the correction. Since the MARNET station provides a comprehensive temporal data set, we investigated a further method that only considers the HOLAS II assessment period of six years. The second method is based on applying a basic correction factor to the model data. Therefore, the average deviation between the observed and modeled oxygen concentrations was determined for the assessment period. The resulting correction factor was then multiplied to the entire model data of the Arkona Basin for the same period.

3 Results

3.1 Comparison of model performance

The evaluation of the model performance of the 1NM model and comparison with the model performance of the 3NM model focused on the one hand on the spatial agreement of near-bottom oxygen measurements from a large number of stations (Fig. 2), and on the other hand on the agreement of the seasonal cycle of the near-bottom oxygen measurements (Fig. 3) and depth profiles (Fig. 4) for selected stations from 2010 to 2019.

https://doi.org/10.5194/bg-2023-152
Preprint. Discussion started: 21 September 2023
© Author(s) 2023. CC BY 4.0 License.



https://doi.org/10.5194/bg-2023-152
Preprint. Discussion started: 21 September 2023
© Author(s) 2023. CC BY 4.0 License.

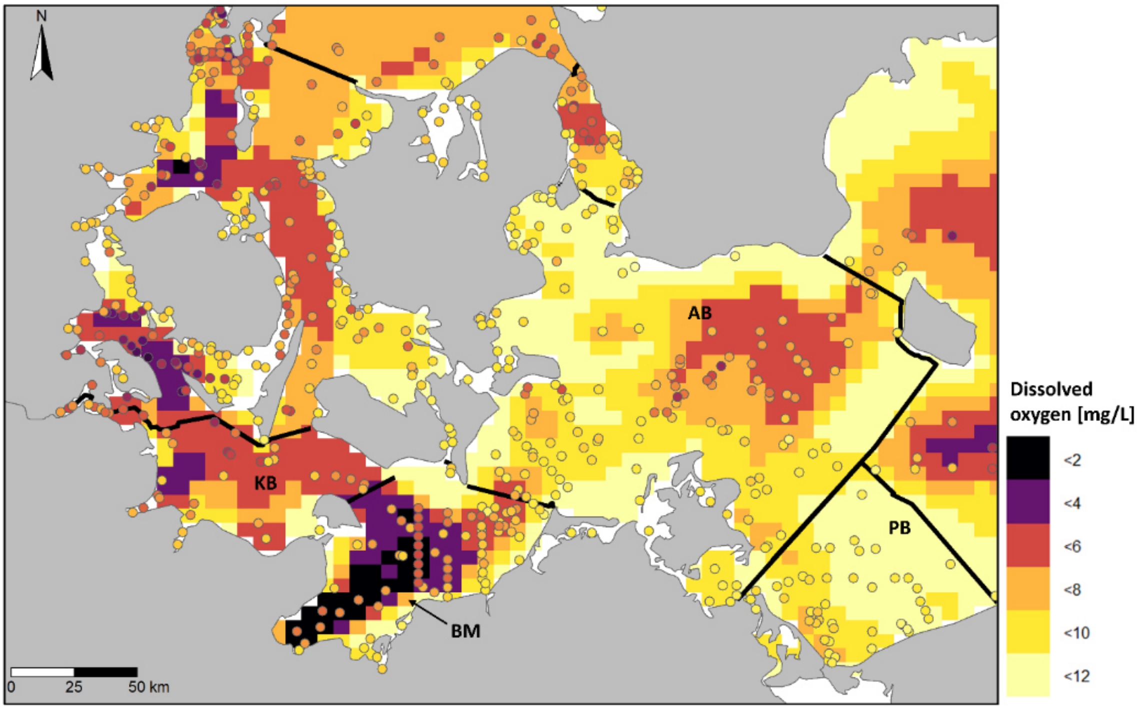


***Figure 2** Comparison of the multiannual mean near-bottom oxygen concentrations (2010-2019) from observations (circles) and model predictions with a spatial resolution of 1NM (top) and 3NM (bottom). Black lines indicate sub-basin divisions. Kiel Bay (KB), Bay of Mecklenburg (BM), Arkona Basin (AB), Pomeranian Bay (PB).*

The 1NM model system is well able to reproduce the spatial gradients in the western Baltic Sea (r=0.79, RMSE=1.71, AE=0.95, Fig. 2, Table 1), whereas it generally tends to overestimate oxygen concentrations on average (AE >0). Model predictions and measurements agree particularly well in the Bay of Mecklenburg (r=0.8, RMSE=1.24, AE=0.03) and Arkona Basin (r=0.76, RMSE=1.19, AE=0.50). While the RMSE is also in a similar range for the Pomeranian Bay (r=0.05, RMSE=1.11, AE=0.57), it is highest for the Kiel Bay (r=0.36, RMSE=1.61, AE=0.69). Both basins show only a weak linear correlation of the multiannual mean near-bottom oxygen values with measurements. The smaller number of measurements may play a role here, so that conflicting measurements from neighbouring stations have a greater effect. This becomes clearer when evaluating a MSFD assessment period of 6 years, with much fewer station data available for the spatial comparison for this period (175 as compared to 342 stations of 2010-19), leading to reduced model skill on a spatial scale of near-bottom oxygen (r = 0.68, RMSE = 1.52, ME = 0.30; Figure S3). In general, a spatial assessment of model skill with less than 10 years of observational data appears to be of limited value as most stations have limited temporal coverage. Moreover, as the spatial coverage of observational data varies across the sub-basins, a careful spatial assessment and comparison of model skill is

https://doi.org/10.5194/bg-2023-152
Preprint. Discussion started: 21 September 2023
© Author(s) 2023. CC BY 4.0 License.

required. The advantage of a spatial comparison is that it can show whether the model accurately represents general patterns as well as identify areas where also the observations show high variability.

Table 1 *Error statistics to assess model-data misfits of the multiannual mean near-bottom oxygen concentrations (2010-2019) for the two model resolutions. Pearson's correlation coefficient (r), Root Mean Squared Error (RMSE), Average Error (AE). Number of available stations per sub-basins in brackets (1NM/3NM).*

Error statistic	r		**RMSE**		**AE**	
Model resolution	**1NM**	**3NM**	**1NM**	**3NM**	**1NM**	**3NM**
Kiel Bay (34/25)	0.36	-0.25	1.61	2.41	0.69	-1.01
Bay of Mecklenburg (78/72)	0.80	0.73	1.24	3.69	0.03	-2.80
Arkona Basin (187/168)	0.76	0.61	1.19	1.42	0.50	0.00
Pomeranian Bay (43)	0.05	0.36	1.11	1.38	0.57	-0.36
SW Baltic Sea (692/570)	0.79	0.60	1.71	2.09	0.95	-0.15

Generally, when comparing the spatial model skill of the 1NM with that of the 3NM model, the correlation improves by 32% and the error (RMSE) decreases by 18% for the south-western Baltic Sea, with largest improvements in the Kiel Bay and Bay of Mecklenburg (Table 1). In contrast to the 1NM model resolution, the 3NM model tends to underestimate the near-bottom oxygen concentrations for almost all sub-basins (AE >0; Table 1) besides the Arkona Basin, where the 3NM model generally seems to have better skill as compared to the other basins.

© Author(s) 2023. CC BY 4.0 License.

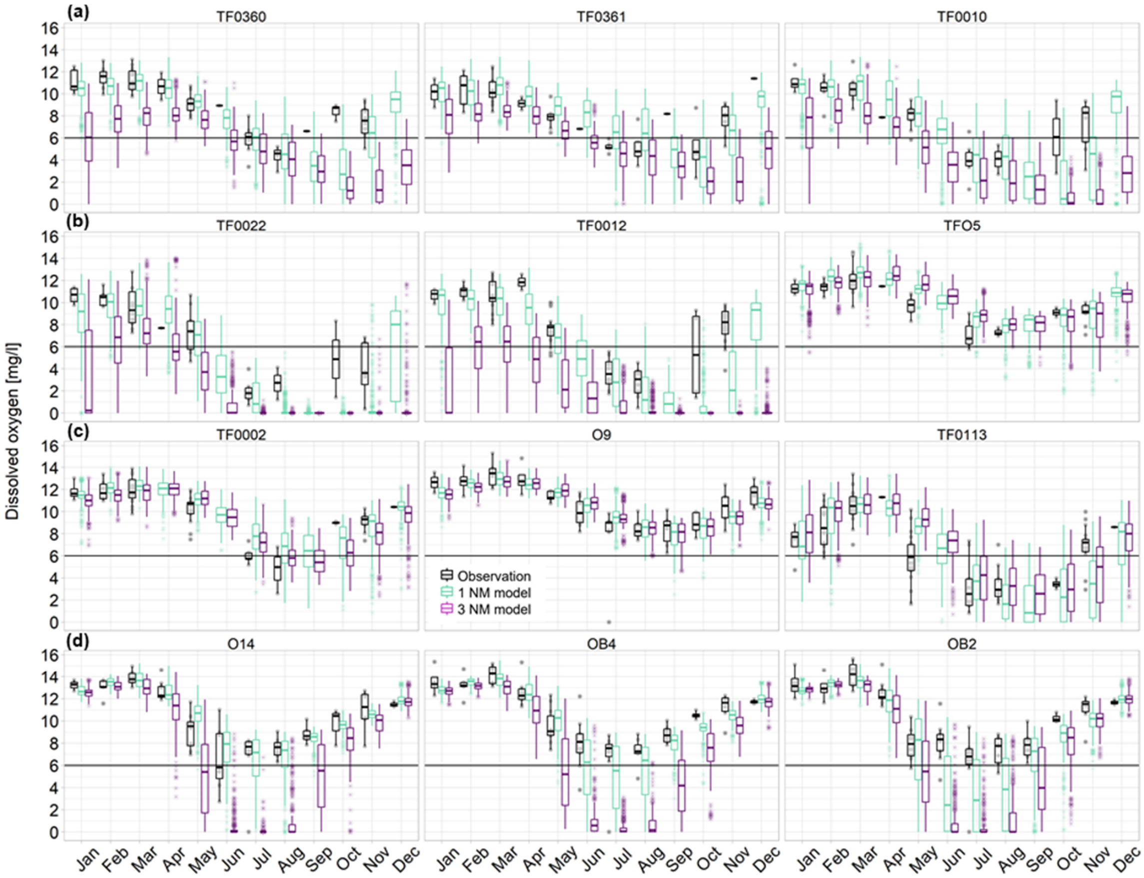


Figure 3 *Comparison of the seasonal cycle of monthly averaged near-bottom oxygen concentrations (2010-2019) from observations (black) and model predictions with two different spatial resolutions for selected stations in (a) the Kiel Bay, (b) the Bay of Mecklenburg, (c) the Arkona Basin, and (d) the Pomeranian Bay. Black line indicates uppermost critical oxygen threshold of 6 mg/l used in oxygen assessments within the Baltic Sea (HELCOM 2023).*

Moreover, the seasonal cycle of near-bottom oxygen concentrations in the Arkona Basin is already well reflected by the 3NM model and an improvement by the higher resolved 1NM model is merely visible. In contrast, the seasonal cycle of the stations of the other sub-basins is better represented by the 1NM model (Fig. 3). The 1NM model tends to underestimate near-bottom oxygen concentrations in the Kiel Bay and Bay of Mecklenburg especially during autumn and winter months and for the Pomeranian Bay during summer months (Fig. 3). It can also be seen that the scatter of the data of measured and modelled values is similar for all stations except those in the Pomeranian Bay, where model values seem to have a higher spread.

https://doi.org/10.5194/bg-2023-152
Preprint. Discussion started: 21 September 2023
© Author(s) 2023. CC BY 4.0 License.

***Table 2** Error statistics to assess model-data misfits of oxygen observations and predictions from depth profiles of selected stations. Pearson's correlation coefficient (r), Root Mean Squared Error (RMSE), Average Error (AE).*

Basin	**Station**	r		**RMSE**		**AE**	
		1NM	**3NM**	**1NM**	**3NM**	**1NM**	**3NM**
Kiel Bay	TF0360	0.82	0.79	1.16	1.20	0.15	0.14
	TF0361	0.83	0.74	1.27	1.55	0.32	0.10
	TF0010	0.89	0.81	1.10	1.78	0.12	-0.54
Bay of Mecklenburg	TF0022	0.86	0.82	1.64	2.21	-0.03	-0.80
	TF0012	0.87	0.76	1.21	1.93	0.21	-0.47
	TFO5	0.90	0.75	0.82	1.21	0.34	0.21
Arkona Basin	TF0002	0.93	0.89	0.68	0.86	0.17	-0.03
	O9	0.80	0.79	1.53	1.50	0.03	0.00
	TF0113	0.89	0.88	1.00	1.02	0.16	0.06
Pomeranian Bay	O14	0.85	0.77	1.32	2.57	0.07	-1.07
	OB4	0.83	0.80	0.81	0.72	-0.33	-0.36
	OB2	0.77	0.76	3.74	4.69	-2.04	-3.35

A good correlation between modelled and observed values (r values >=0.77) shows that the 1NM model system can well reproduce the temporal dynamics of the vertical oxygen gradients (Table 2, Fig. 4 and Figure S2).

© Author(s) 2023. CC BY 4.0 License.

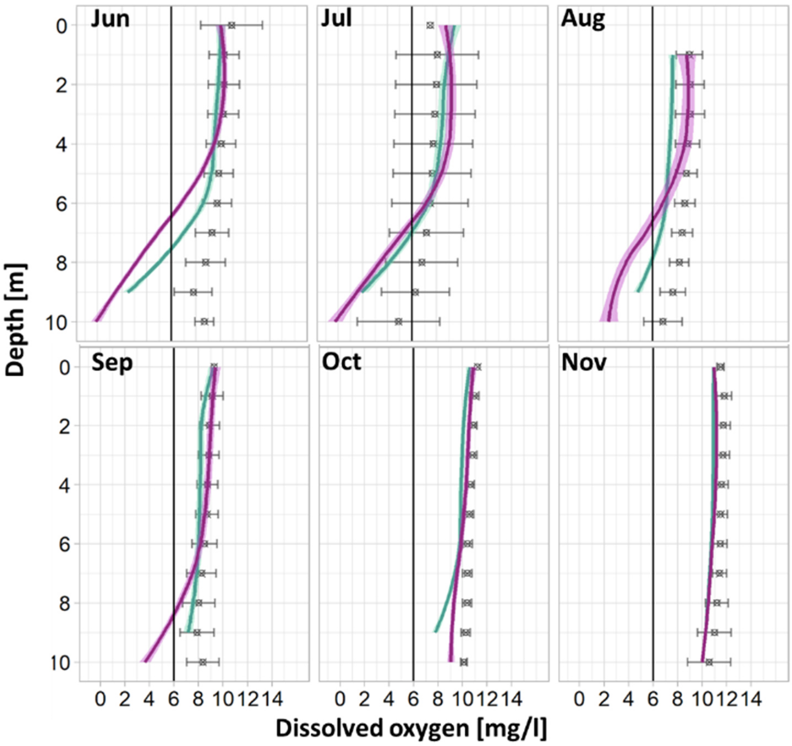


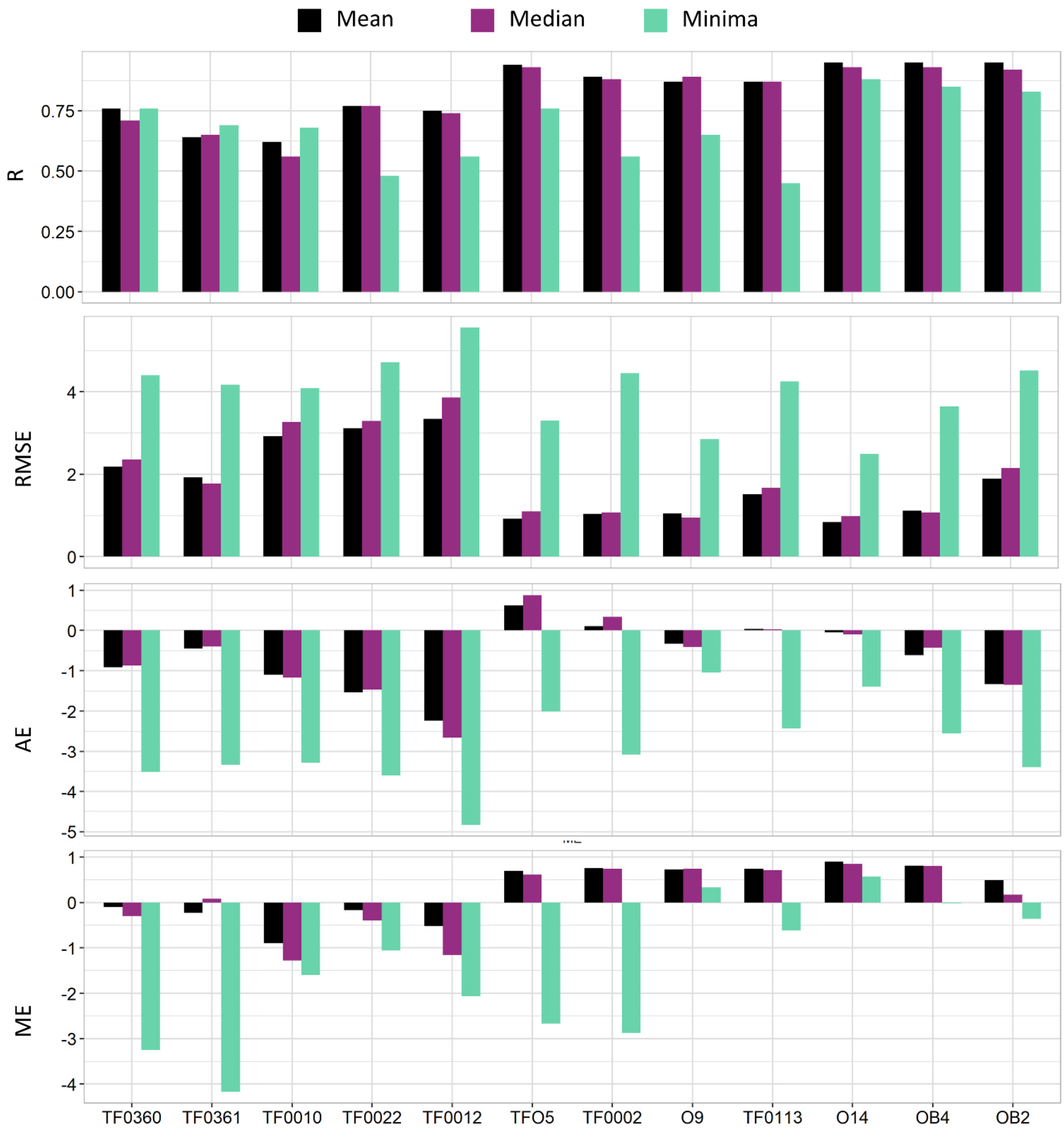
Figure 4 *Comparison of monthly averaged oxygen depth profiles (2010-2019) from observations (black) and model predictions with two different spatial resolutions (3NM purple, 1NM turquoise) for the critical months from June to November. Exemplarily for station OB2 in the Pomeranian Bay, all other depth profiles can be found in Figure S2. Black line indicates uppermost critical oxygen threshold of 6 mg/l used in oxygen assessments within the Baltic Sea (HELCOM 2023).*

As the higher horizontal resolution mainly affects the dynamics within near-bottom water layers, the improvement of the model skill for the oxygen depth profiles is not as pronounced as for the near-bottom oxygen values. For example, the improvement of RMSE ranges from 2 to 49% for the oxygen profiles of selected stations (Table 2), whereas for the spatial comparison of near-bottom oxygen values the improvement of RMSE ranges from 16 to 66% (Table 1). The highest improvement in RMSE for single stations are achieved in the Bay of Mecklenburg and the Pomeranian Bay, followed by Kiel Bay (Table 2). The comparison of the depth profiles of the individual stations shows that not only the 3NM model underestimates the oxygen concentrations in some cases, but also the 1NM model, being most pronounced for the two stations in the Pomeranian Bay (AE <0; OB2, and OB4; Table 2, Fig. 4, Figure S2). This is most likely to occur in the near-bottom layer, as the seasonal cycle of near-bottom oxygen concentrations shows that oxygen concentrations are underestimated by the models for some stations, especially in late summer and autumn (Fig. 3).

https://doi.org/10.5194/bg-2023-152
Preprint. Discussion started: 21 September 2023
© Author(s) 2023. CC BY 4.0 License.

Next, we evaluated the model performance in the context for an oxygen assessment, where we focused on the 1NM model. When evaluating model skill, one must have a clear idea of what conditions are to be modelled, i.e., the scope, objectives, and resolution of an oxygen assessment must be considered to determine what constitutes a "good" model (Bennett et al., 2013). Due to the ecological relevance for organisms, oxygen assessments are mostly carried out on the basis of critical threshold values, which are 2, 4, and 6 mg/l, according to their ecological relevance for organisms (Vaquer-Sunyer and Duarte, 2008). Both the water volumes and the near-bottom areas are assessed in the process. Accordingly, a model must reliably represent the low oxygen concentrations both at the bottom and in the water column. As some assessments focus on annual oxygen minima, a model should also reliably provide these. Due to the limited data available, the observations are usually averaged over months as well as years (e.g., 6 years are utilized for MSFD assessments), meaning that the model should reliably reproduce annual/monthly mean values. Finally, it should be taken into account that the months between June and November are mostly used for the evaluation, i.e., a model should have a high model skill especially in these months.

https://doi.org/10.5194/bg-2023-152
Preprint. Discussion started: 21 September 2023
© Author(s) 2023. CC BY 4.0 License.



305 ***Figure 5*** *Assessment of the skill of the 1NM model for the HOLAS II assessment period (2011-2016). Error statistics of the multiannual mean of near-bottom oxygen concentrations based on different data aggregation approaches either by first calculating monthly means, medians or minima before temporal averaging are shown. Pearson's correlation coefficient (r), Root Mean Squared Error (RMSE), Average Error (AE), Modelling Efficiency (ME).*

https://doi.org/10.5194/bg-2023-152
Preprint. Discussion started: 21 September 2023
© Author(s) 2023. CC BY 4.0 License.

Biogeosciences
Discussions
Open Access

Assessing model skill for temporally averaged mean or median values, besides the median of station TF0010 the correlation coefficients are above 0.6 for all stations of each sub-basin, (Fig. 5). Besides the median of station OB2 the RMSE for temporally averaged mean or median values is <2 for the stations in the Arkona Basin and Pomeranian Bay and only for one station in each, the Kiel Bay and Bay of Mecklenburg (TF0361 and TFO5 respectively; Fig. 5). Similarly, the ME values are >0 for all stations in the Arkona Basin and the Pomeranian Bay, whereas this is only the case for station the station TF0361 in Kiel Bay (here only for the temporal mean) and the station TFO5 of the Mecklenburg Bay. For all other stations in those two basins the observational average of the mean or median is a better predictor. Generally, if data is aggregated to monthly minima before averaging over the assessment time period the observations are a better predictor (Fig. 5).

3.2 Factors influencing the error between observations and model predictions

Our results show that accounting for sampling depth has the highest impact when retrospectively correcting the model's results, followed by maximum depth and mixed layer depth with distance to shore having the least impact (Fig. 6).

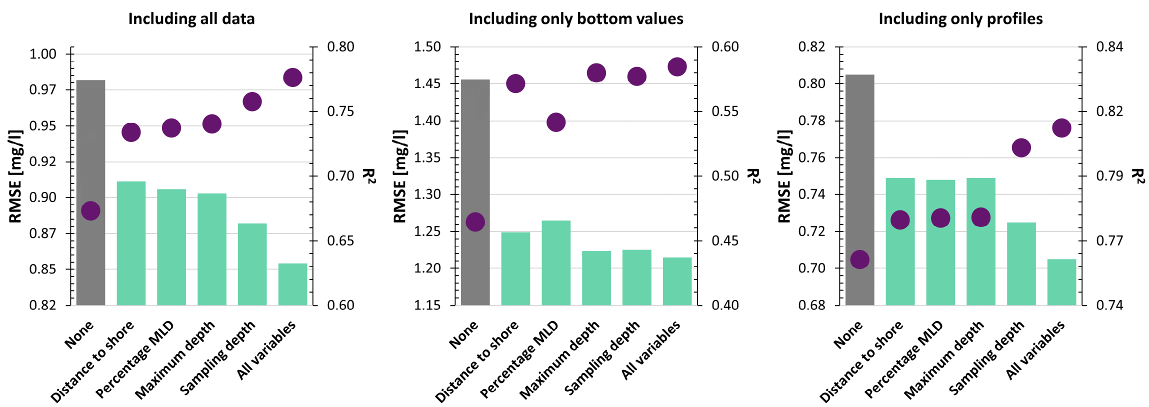


***Figure 6** Change in the Root Mean Squared Error (RMSE, bars) and the Coefficient of Determination (R-squared; purple dots) between observations and model predictions for uncorrected model results (black bar) and, based on different variables (turquoise bars) corrected model results.*

If all available observations are included (near-bottom oxygen values and profiles) the deviation between observations and model predictions is mainly influenced by sampling depth (improvement of the RMSE by about 10%) followed by maximum depth and percentage of mixed layer depth (improvement of RMSE ~8% each), and finally distance to shore (improvement of RMSE of ~7%; Fig. 6). When only taking the near-bottom oxygen measurements into account the maximum depth is equally important than sampling depth (improvement of RMSE ~16% each). When analyzing oxygen profiles alone, the most important factor is sampling depth (improvement of RMSE ~10%) followed by all other parameters (improvement of RMSE ~7% each).

https://doi.org/10.5194/bg-2023-152
Preprint. Discussion started: 21 September 2023
© Author(s) 2023. CC BY 4.0 License.

However, since measurement depth and maximum depth are strongly auto-correlated for the station measurements of only near-bottom oxygen, the overall improvements of the different correction functions should be viewed with caution. Although the RMSE difference is minor (e.g., RMSE improvement of 12% to 17% when including all variables), the correction function based solely on profile data shows the lowest RMSE in direct comparison.

3.3 Evaluation of retrospective model corrections

3.3.1 Correction based on multiple stations

Considering the four sub-basins of the western Baltic Sea and including all data in the correction function the corrected 1NM run (incl. all data) has a better spatial fit of near-bottom oxygen values (decrease of RMSE by ~10% as compared to uncorrected run) and leads to improvements of the model skill in almost all basins (besides correlation coefficient r, Table 3). A correction only based on bottom values leads to poorer model skill in the Arkona Basin (increase of RMSE by ~12%) and a correction only based on profiles to a poorer model skill in the Pomeranian Bay (increase of RMSE by ~40%) as compared to the non-corrected model results (Table 3).

Table 3 *Error statistics to assess model-data misfits of the multiannual mean near-bottom oxygen concentrations (2010-2019) for the corrected model results either based on near-bottom oxygen values (Bottom values), oxygen profiles (Profiles) or both data sources (All data). Spearman Rank correlation coefficient (r), Root Mean Squared Error (RMSE), Average Error (AE) and Modelling Efficiency (ME).*

Basin	Correction	R	RMSE	AE	ME
Kiel Bay	**None**	**0.28**	**1.61**	**0.69**	**-0.41**
	All data	0.23	1.42	0.22	-0.10
	Bottom values	0.26	1.41	0.15	-0.09
	Profiles	0.27	1.57	0.49	-0.35
Bay of Mecklenburg	**None**	**0.78**	**1.24**	**0.03**	**0.38**
	All data	0.78	0.99	-0.01	0.60
	Bottom values	0.79	1.04	-0.39	0.56
	Profiles	0.78	1.14	0.36	0.48
Arkona Basin	**None**	**0.72**	**1.19**	**0.50**	**0.39**
	All data	0.72	1.11	-0.23	0.47
	Bottom values	0.72	1.34	-0.73	0.23
	Profiles	0.72	1.17	0.24	0.41
Pomeranian Bay	**None**	**0.19**	**1.11**	**0.57**	**-1.51**
	All data	0.30	1.05	0.57	-1.27
	Bottom values	0.34	0.88	-0.47	-0.58
	Profiles	0.12	1.55	1.11	-3.91
All basins	**None**	**0.78**	**1.23**	**0.41**	**0.46**
	All data	0.79	1.11	-0.04	0.57

https://doi.org/10.5194/bg-2023-152
Preprint. Discussion started: 21 September 2023
© Author(s) 2023. CC BY 4.0 License.

Bottom values	0.77	1.23	-0.54	0.46
Profiles	0.77	1.25	0.39	0.45

With changes of r less than 0.1 the improvement of the linear correlation between observations and model predictions is only marginal considering only near-bottom oxygen concentrations or oxygen profiles of single stations (Table S1 and Table S2). Highest increase of the linear correlation is seen in the Pomeranian Bay for the spatial comparison of mean annual oxygen concentrations with an improvement in r of 0.28 if considering only the bottom values for correction (Table 3). In contrast this improvement is not seen for the considered single stations in the same basin (Table S1). When considering either near-bottom oxygen or oxygen profiles for correction, the linear correlation indeed decreases slightly for some basins (Table 3).

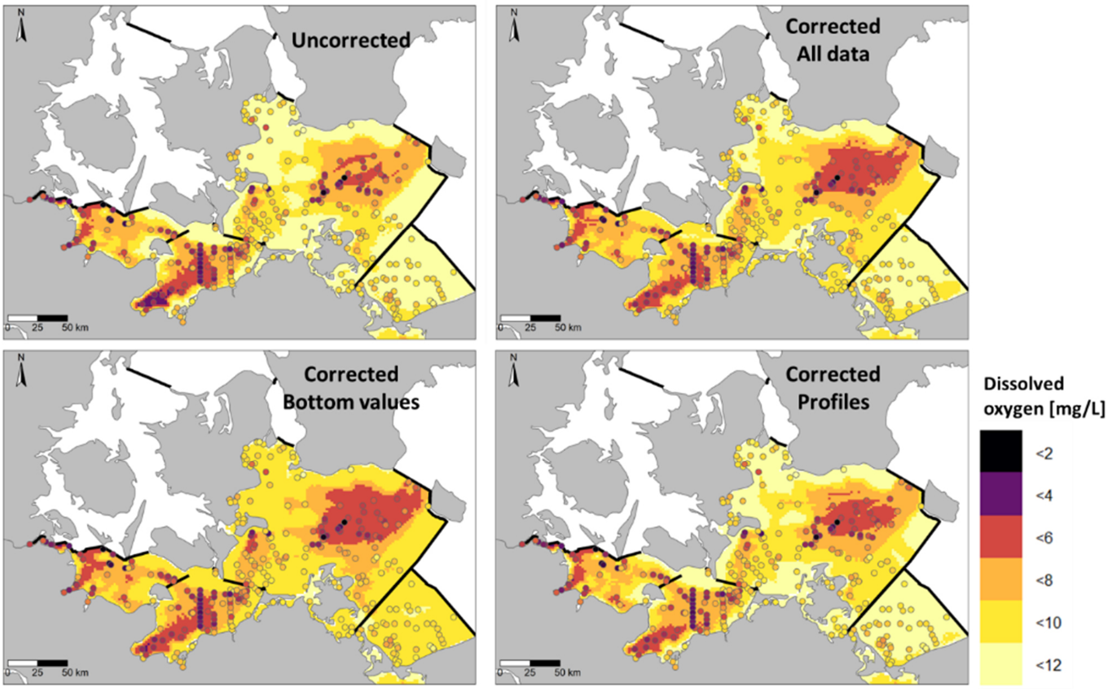


***Figure** 7 Comparison of the multiannual mean near-bottom oxygen concentrations (2010-2019) from observations (circles) and model predictions with a spatial resolution of 1NM for the uncorrected and retrospective corrected model results. Black lines indicate sub-basin divisions.*

Major influences of retrospective model corrections are seen for the Bay of Mecklenburg and the Arkona Basin as compared to the uncorrected model results (Fig. 7). Looking at all three aspects of model performance (spatial near-bottom oxygen, near-bottom oxygen and oxygen profiles of selected stations), the correction of the model results mostly improves the RMSE in the Bay of Mecklenburg and Kiel Bay, followed by the Arkona Basin and Pomeranian Bay (Table 3, Table S1 and Table S2).

https://doi.org/10.5194/bg-2023-152
Preprint. Discussion started: 21 September 2023
© Author(s) 2023. CC BY 4.0 License.
Biogeosciences Discussions

Considering the near-bottom oxygen concentrations of selected stations alone, a 13% improvement in RMSE can be achieved by correcting model results based on all data, a 10% improvement by correcting using bottom oxygen measurements alone, and a 14% improvement by correcting using oxygen profile data only (Table S1). Considering the oxygen profiles of selected stations, an improvement in RMSE of 9% can be achieved by correcting model results based on all oxygen data and of 8% by correcting only by oxygen profile data (Table S2).

Evaluating the average error (AE), a correction based on all data would change the tendency of the 1NM model from an overestimation of near-bottom oxygen values to an underestimation of near-bottom oxygen values for the Bay of Mecklenburg and the Arkona Basin (Table 3). This is similar to a correction based on bottom values alone whereas here additionally the Pomeranian Bay values are underestimated.

The modelling efficiency (ME) indicates that mainly the average near-bottom oxygen concentrations in the Kiel Bay and Bay of Mecklenburg would be better represented by the retrospective corrected model results if the correction is based either on all data or on oxygen profile data alone (Table S1 and Table S2). A correction based on near-bottom oxygen data alone would lead in several cases to a worsening of model skill as compared to the uncorrected 1NM run (Table S1 and Table S2) and generally be not beneficial if the vertical dimension needs to be considered. The improvement of the ME for the comparison of the oxygen profiles is rather low with an average of ~ 0.04 (Table S2).

3.3.2 Correction based on one temporal highly resolved station

As can be seen by the comparison of the standard error of the temporally high resolved MARNET station and e.g., the station TF0113 in the Arkona Basin (Fig. 8) the precision of the estimate of the mean increases (i.e., decreasing SE) drastically with increasing observations. The high temporal resolution of the MARNET station is especially advantageous in the critical month of the development of seasonal oxygen deficiency from July to November, resulting in lower standard errors as compared to the uncorrected model results.

https://doi.org/10.5194/bg-2023-152
Preprint. Discussion started: 21 September 2023
© Author(s) 2023. CC BY 4.0 License.

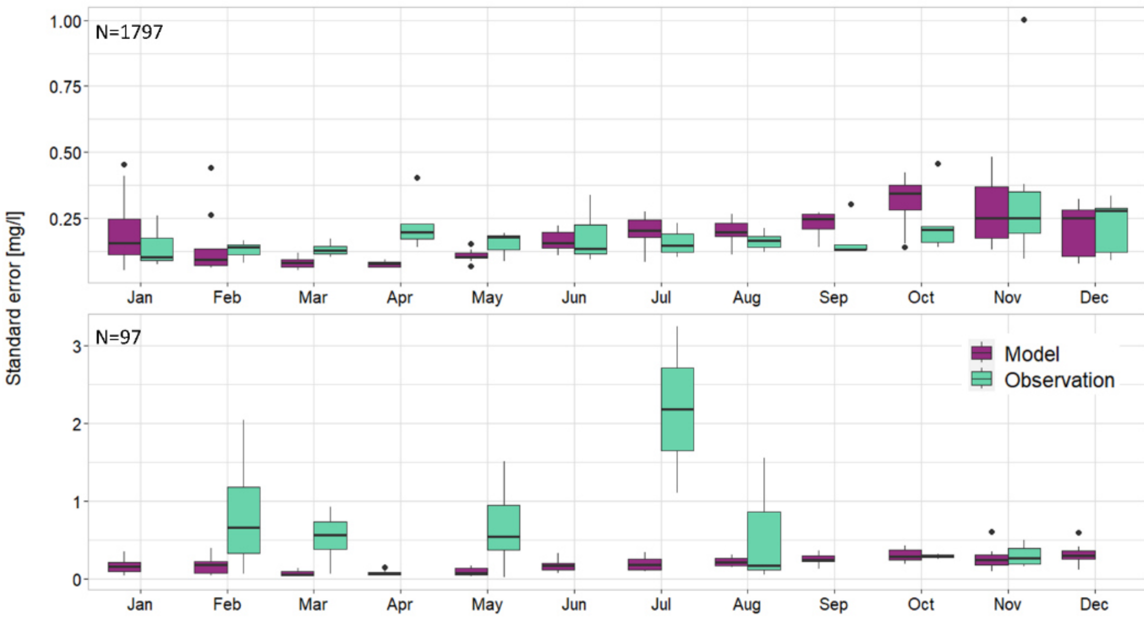


Figure 8 *Average monthly standard error (2010-2019) of observed and modelled oxygen concentrations at 40 m depth for the MARNET station (top) and the station TF0113 (bottom) in the Arkona Basin.*

The analyse of the model quality based on our previous approach, leads initially to poorer skill metrics. A spatial comparison of near-bottom oxygen values leads to a decrease of the linear correlation from 0.76 to 0.67 as compared to the uncorrected results in the Arkona Basin. Further, the RMSE increases from 1.19 to 1.69 for the spatial comparison in the Arkona Basin. In addition, the RMSE for near-bottom oxygen and oxygen profiles of selected stations in the Arkona Basin (i.e., stations O9, TF0113, and TF0002) would on average increase by 0.35 and by 0.48, respectively. Reasonably, this approach does not necessarily lead to better skill metrics when compared to other stations. If compared to the MARNET station data the linear correlation increases from 0.79 to 0.82 and the RMSE decreases from 2.19 to 1.67 for the corrected model results based on the same station. The tendency of the uncorrected model result to overestimate oxygen concentrations (AE = 1.34) changes for the corrected model results to a slight underestimation (AE = -0.18). A significant improvement is seen in model efficiency, which nearly doubles from 0.32 to 0.6.

The average observed and modelled oxygen concentration at 40 m depth for the assessment period (2011-2016) are 7.40 mg/l and 8.68 mg/l, respectively. The corresponding correction factor applied to the model data is 0.85 accordingly. After application of the correction factor the one to one comparison of corrected model values and observations from the MARNET stations leads overall to improved model skill metrics for the station (r = 0.79, RMSE = 1.64, AE = 0.01, ME = 0.62).

https://doi.org/10.5194/bg-2023-152
Preprint. Discussion started: 21 September 2023
© Author(s) 2023. CC BY 4.0 License.

3.4 Impact of integrated model results on near-bottom oxygen assessment

A comparison of the average bottom area affected by hypoxia for the HOLASII period (2011-2016) shows that for the Kiel Bay and Pomeranian Bay the difference among the uncorrected and various corrected model results for the assessment is minor (<= 2%; Table 4). However, the profile corrected data led to a 13% less affected area by hypoxia in the Bay of Mecklenburg while the data corrected with all available data led to a 25% increase of the hypoxic area in the Arkona Basin, as compared to the uncorrected model run (Table 4).

Table 4 *Average hypoxic area [km²] and corresponding percentage from total area of a sub-basin within the HOLASII assessment period (2011-2016; month from June to November) for the different 1NM corrected model results as compared to the uncorrected model run.*

Basin	Model	Area [km²]	Area [%]	Difference [%]
Kiel Bay	**None**	716	22	
	All data	658	21	-2
	Bottom values	638	20	-2
	Profiles	643	20	-2
Bay of Mecklenburg	**None**	1483	35	
	All data	1184	28	-7
	Bottom values	1309	31	-4
	Profiles	916	22	-13
Arkona Basin	**None**	1850	11	
	All data	5973	36	25
	Bottom values	2121	13	2
	Profiles	2107	13	2
	MARNET Function	2933	18	6
	MARNET Factor	2077	12	1
Pomeranian Bay	**None**	130	3	
	All data	119	2	0
	Bottom values	116	2	0
	Profiles	103	2	1

In contrast, almost no difference of the average hypoxic area is seen when applying a correction factor based solely on the MARNET station data. If the correction function is solely based on the MARNET station, the resulting hypoxic area lies with about 6% somewhere within the range of the other estimates.

https://doi.org/10.5194/bg-2023-152
Preprint. Discussion started: 21 September 2023
© Author(s) 2023. CC BY 4.0 License.

4 Discussion

With our case study we aimed to develop a method for a high-resolution monitoring and assessment system for shallow water oxygen deficiency in the western Baltic Sea. Based on our findings we will discuss in the following general possibilities and limitations of ocean models as an integral part of oxygen assessments.

4.1 Influence of a higher horizontal model resolution on modelled oxygen concentrations

Despite allowing to model more complex shorelines and bathymetries, refining purely the spatial resolution does not automatically lead to better model results. Gröger et al. (2022) have shown, how key features of the ocean model (like upwelling dynamics or thermocline depth) have changed due to refining the horizontal model resolution from three to one nautical miles of our model system MOM-ERGOM. Our analysis showed that the higher horizontal model resolution improves the oxygen dynamics with respect to the seasonal cycle, depth profiles, and spatial agreement between observations and model predictions (Fig. 2, 3, and 4; Tables 1 and 2). Generally, measured and modelled oxygen values match very well in upper, well-mixed layers. The agreement decreases notable around and below the pycnocline for most stations (Fig. 3 /Figure S2). The higher horizontal resolution has a lower impact on the model quality of the generally deeper Arkona Basin, where oxygen deficiency seems to develop further offshore as compared to the other three basins (Fig. 2). In general, the higher resolution of 1NM is advantageous for representing small-scale structures of the bottom topography, influencing the spread of oxygen deficiency (Virtanen et al., 2019). Despite, a higher horizontal model resolution generally allows near-coastal features to be better represented (Fennel et al., 2022). The higher resolution of our model already allows transitional waters such as lagoons and estuaries to be resolved to a certain degree, allowing a better representation of the dynamics in adjacent nearshore areas. For example, in the Pomeranian Bay area the Szczecin lagoon is mapped through the 1NM model (Fig. 2) resulting in a better representation of the Odra/Oder river plume influencing this area. Despite improving the results substantially, even in the fine resolution model the simulated oxygen concentrations have still a noticeable deviation from the observations (Table 2), indicating that the ecosystem model and its included processes need further developments and improvements.
Considering the area being assessed by the EU-MSFD, i.e., all waters one nautical mile from the baseline, and the kilometre scale considered (area of square kilometres), we do not expect that an additional higher resolution of the physical dynamics will result in a drastic improvement in model capabilities for reproducing oxygen dynamics. However, further improvements in model resolution would additionally allow a reliable representation of the spatio-temporal dynamics of hypoxia in inner coastal waters such as lagoons or estuaries, which are required to be assessed for oxygen in accordance with the EU-WFD.

4.2 Integration of observations and model results

Since not for all sub-basins model skill was equally good with the application of a higher horizontal model resolution of 1NM, an attempt was made to further improve model skill by a retrospective correction of the model results based on observational

https://doi.org/10.5194/bg-2023-152
Preprint. Discussion started: 21 September 2023
© Author(s) 2023. CC BY 4.0 License.

data. Therefore, several factors influencing the error between observations and model predications were analysed and used for correction.

Does a correction of model results based on observations increase model skill?

The comparison of the retrospective corrected model results with the uncorrected ones of the 1NM model run has shown that the correction led to only minor improvements in the linear correlation (r) of observations and model predictions (Table 3, Table S1 and Table S2). Moreover, the highest improvement seen spatially for the Pomeranian Bay in the linear correlation is not reflected by the analysis of selected stations (Table S1 and Table S2). Overall, the RMSE for the corrected model results improves compared to the uncorrected 1NM run. However, the improvements of up to 26% (or ~0.3 mg/l, TFO5 correction with all data; Table S1) for individual stations must be contrasted with a deterioration of up to 91% in the RMSE of other stations (>1 mg/l, TF0360 correction with profiles; Table S2). Although the ME improves when a correction is based on either all oxygen data or oxygen profile data alone, the improvement is <0.2 when considering near-bottom oxygen, and marginal when comparing oxygen profiles (~0.04 of the ME value). Considering the Kiel Bay and the Bay of Mecklenburg, the results of the different correction functions do not show a consistent pattern (Table S1 and Table S2).

Most of the limitations of a retrospective correction and a subsequent assessment of the model skill can probably be explained by the spatially and temporally heterogeneous observational data, as well as a non-homogeneous bias between measurements and model values among the available observational data (Fig. 2, Figure S1). As can be seen in Figure 2, the 2010-2019 period already contains some inconsistent observations, where, for example, neighbouring stations have very different mean oxygen values. An inconsistent bias between model and observations for different locations leads to the problem that corrected model results derived from correction functions that include all stations do not always lead to improved model performance for individual stations. In addition, the bias between observations and model predictions is unlikely to be constant and may be influenced by various factors over time, e.g., through technical adjustments (changes in measurement instruments, protocols, or sampling designs) or changes in the ecosystem (e.g., hydrodynamic changes). This would require a corresponding adjustment of the correction function and would lead to a lack of transparency for environmental trend analysis. In general, correcting model data retrospectively means sacrificing transparency and is very difficult to evaluate externally. For the application of model data in the field of environmental assessment with political implications, transparency is a key criterion for the adaption of ocean models as tool (Bennett et al., 2013). Concurrently, a lack of transparency is often cited as a reason for the failure of model-based approaches (Alexandrov et al., 2011). In our case, none of the correction functions was superior in the sense that it gave significantly better results for all sub-basins. The use of different correction functions (based on different databases) might improve the model results but would reduce the comprehensibility and the transparency of the method further.

On the other hand, the outcome of correcting model data using observational data heavily relies on the quality of the observational data utilized for the correction. During the compilation of the observational data, it was discovered that although the data set initially seemed large, a vast amount of unusable data was excluded due to selection based on specific regions, years, and/or months as well as removing unplausible data. This can be seen by the comparison of the seasonal cycle of the near-bottom oxygen values (Fig. 3) and the oxygen depth profiles (Fig. 4, Figure S2). It should be noted that, at some stations, data for monthly measurements were available only once over a ten-year period, and for some months data were not available at all.

Thus, our second approach was focusing exemplarily on the Arkona Basin and only considering high temporal resolution data from one representative station located in the centre of the sub-basin. Comparison of modelled and observed values after applying the correction function as well as the correction factor resulted in the expected improved model skill metrics for that station. A reliable improvement in the model's results can only be attained if high reliable measurement data is utilized, i.e., with a high precision of the estimate of the mean as it is the case for the MARNET station. Nevertheless, it should be noted that the station observations are not taken at the maximum depth, but five meter above bottom, which may has influenced the derived correction function as well as correction factor. Moreover, if the hypoxic water volume needs to be assessed, a correction factor preferably should also be derived from a station which combines both, a high temporal as well as vertical resolution.

To sum it up, although for some aspects the retrospective correction function based on multiple station data resulted in partly better skill metrics, this was not consistent across all sub-basins and/or stations considered, and aspects of model skill analysed. One reason for this is likely the heterogenous measurement data utilized to derive the correction functions, which are not consistent in their temporal and vertical resolution. Using the Arkona Basin as an example, we show a future path in which only high temporal and/or vertical resolution observations are integrated with model data to achieve the highest possible confidence in the assessment.

4.3 Possibilities and limitations of ocean models as oxygen deficiency assessment tool

Over the last decades, ocean models have evolved from simple box models to highly complex 3D models (for a review see Fennel et al., 2022). Together with technical progress (Robson, 2014), this has led to ocean model products being available today in adequate time and quality to be used for environmental management, e.g., for operational forecasts or environmental assessments. In our study we show that our well validated ocean model MOM-ERGOM can be used as a routine tool for oxygen assessments, in particular for past and present conditions in the southwestern Baltic Sea.

Comparing the average hypoxic areas determined by our model with those obtained in other studies proves to be difficult. Utilizing the uncorrected model results, we calculated the areal percentages by monthly averaging from the critical period from June to November before averaging over the assessment period and calculated a hypoxic area of about 22% of the Kiel Bay, 35% of the Bay of Mecklenburg, 11% of the Arkona Basin and 3% of the Pomeranian for the HOLASII assessment period (Table 4). For example, Friedland et al. 2023 (accepted) utilized two variants of the machine learning algorithm Quantile Regression Forest, among others, for interpolation of observational data. By first identifying the annual minima before averaging over the same period (HOLASII) they found that about 25-27% of the Kiel Bay, ~17% of the Bay of Mecklenburg,

https://doi.org/10.5194/bg-2023-152
Preprint. Discussion started: 21 September 2023
© Author(s) 2023. CC BY 4.0 License.

about 8-9% of the Arkona Basin and 0% of the Pomeranian Bay were affected by hypoxia. For the Arkona Basin and the Pomeranian Bay the results are close to our findings, whereas for the Kiel Bay and Bay of Mecklenburg the results are somewhat reversed. The latest HOLASIII assessment (2016-2021) by HELCOM reported the poorest condition for the Bay of Mecklenburg, followed by the Arkona Basin and Kiel Bay, which would support the result of our model for the Bay of Mecklenburg. Nevertheless, the somehow better result for the Kiel Bay in contrast to the Arkona Basin may be explained by the fact, that for the calculation of the area three different thresholds (2, 4, and 6 mg/l) were utilized and averaged (HELCOM 2023).

In general, our results show larger areas affected by hypoxia as e.g., compared to the latest HELCOM assessment report on shallow-water oxygen. In contrast to most other methods, which rely on measurements mostly taken about 1 to 4 m above the bottom, our model reflects the deepest water layer above the bottom which could explain the generally lower near-bottom oxygen concentrations. Furthermore, due to the daily resolution of our model, all of the episodic and seasonal hypoxia events are accounted for, which is reflected in our results. The recording of these very dynamically occurring events is not possible by monthly observational data alone. Still, seasonal hypoxia has been observed in the Kiel Bay since the 1960s covering an area of about 890 km^2 (~28% of the sub-basin) and in the Bay of Mecklenburg since the 1980s over an area of about 1,860 km^2 (~44% of the sub-basin) as well as in small areas in the Pomeranian Bay (Karlson et al., 2002; Dias et al., 2011). Although not directly comparable due to differences in time periods, the reported hypoxic areas align well with our results obtained from 2011 to 2016 for those sub-basins. Specifically, the Kiel Bay had an average hypoxic area of 716 km^2, while the Bay of Mecklenburg had an area of 1,483 km^2.

Despite the limitations of our model, we consider it as an effective instrument to evaluate the spatial and temporal trends of oxygen deficiency for the analysed sub-basins, particularly when monthly mean or median values are used. The model data discrepancies were generally less than the variability of the observations for near-bottom oxygen and profiles for single stations (ME>0; Table S1 and Table S2), indicating a performance level acceptable for utilisation in policy-relevant situations (Allen et al., 2007). A reason for the observed poorer modelling efficiency (ME<0) for the spatial comparison on a sub-basin scale for the Kiel Bay and the Pomeranian Bay (Table 3) may be the lower spatio-temporal resolution of observational data compared to the other two sub-basins. In addition, we see great potential in the exemplary method of a retrospective adaption of model results to observational data as presented for the Arkona Basin. The example demonstrates the value of station data with high temporal resolution which is essential for a meaningful integration of model and observational data for an improved oxygen assessment. With available high resolution observational data, the approach could get exploited further in future, e.g., by incorporating the short-term variability. Such an approach can overcome the current limitations of single point measurements and, by using measured data as a basis, can provide greater confidence in integrated model products while maintaining a high level of transparency.

From major campaigns to abate coastal eutrophication and its symptoms such as deoxygenation, globally (Boesch, 2019), we only could find two examples were modelling approaches are specifically utilized for the environmental assessment of past

© Author(s) 2023. CC BY 4.0 License.

and current oxygen conditions (Fennel et al., 2019). In the Chesapeake Bay a regional ocean model is used to provide forecasts as well as seasonal analysis of the severity of hypoxia (https://www.vims.edu/research/topics/dead_zones/forecasts/report_card/index.php) to inform managers and the public about the water-quality of Chesapeake Bay. Another example where model products are used for environmental management is the Northern Gulf of Mexico (Fennel et al., 2019). Here the regional ocean model provides the information base for the Hypoxia Taskforce, which is responsible for devising strategies to mitigate the hypoxic area and assessing advances made in achieving this objective (https://www.epa.gov/ms-htf/hypoxia-task-force-action-plans-and-goal-framework). As oxygen deficiency is being reported more frequently in the Baltic Sea (Conley et al., 2011; Carstensen et al., 2014; Breitburg et al., 2018;) and will continue to increase due to projected warmer water temperatures in many areas (Meier et al., 2018b; Friedland et al., 2012; Naumov et al., 2023), there is highly likely a growing societal need for monitoring and assessment of oxygen deficiency in coastal areas, including transitional waters such as estuaries and lagoons. Since computational power is no longer the bottleneck for the use of biogeochemical models as a management tool, in particular for the reanalysis of past and present conditions, it is probably mainly limited by trust in the models. To increase the reliability in the model results, major efforts are undertaken, e.g., by using best available data on water- and airborne nutrient inputs, atmospheric forcing, and boundary conditions to adjacent seas. Here, the Baltic Sea has the advantage that it has a long history of environmental monitoring and freely available data, allowing comprehensive model validations (Eilola et al., 2011; Gröger et al., 2022). Yet, contrary to expectations, the number of measurements seems to have been decreasing in our area since the early 2000s (Friedland et al., 2023; accepted). But merely extending the observational network spatially also seems not to be a determining factor for reliable environmental monitoring and assessment if it contains different temporal and/or vertical resolutions as we see in our spatial comparison of oxygen (Fig. 2). Rather, stations that have both high temporal and vertical resolution are needed for a reliable oxygen monitoring and assessment using model products. Therefore, decisions on expanding ocean observing systems should prioritize the temporal and vertical expansion of existing representative stations for oxygen and other parameters that require 3D observation. In this way, the advantages of both methods are utilized and can be combined for a high-resolution spatio-temporal assessment of the environment.

Conclusion

The demand for information on oxygen levels is rising globally and locally. Nevertheless, there are limitations on measurement programs, due to e.g., cost constraints. In the future, supplementing these programs with model data could be a cost-effective way to monitor oxygen deficiency. Refining the horizontal resolution of our ocean model to one nautical mile, improved the oxygen dynamics in the coastal and open sea areas (EU-MSFD) of the southwestern Baltic Sea. Refining the model resolution further is essential in the future to reliably represent also the inner coastal waters (EU-WFD). Especially for the assessment of the duration and extent of highly dynamic seasonal hypoxia, our approach is currently superior to methods that are based exclusively on observational data, which cannot reliably reflect the spatio-temporal dimensions of oxygen deficiency. Several purely data-based oxygen assessment approaches are thus associated with a high degree of uncertainty and do not provide

© Author(s) 2023. CC BY 4.0 License.

information relevant for assessing the impacts of hypoxia on organisms. One option to provide a more reliable and ecologically relevant assessment of oxygen deficiency in the future, with a high level of confidence and transparency, is a combined approach where temporally and/or vertically high-resolution observations are integrated with the model data. An adapted monitoring and assessment approach could bring together the advantages of both methods by extrapolating the monitoring data important for policy decisions into space using model products. To become truly ready for use, end-user need to be involved in the next steps in order to develop a strategy for confidence assessment (considering the scope, aims and resolution of the application domain of the model) meeting the requirements for management decisions.

Code availability

Not applicable.

Data availability

The simulation datasets analysed during the current study are available in the IOW THREDDS repository, https://thredds-iow.io-warnemuende.de/thredds/catalogs/projects/integral/catalog_integral.html. Other datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Author contribution

All authors contributed to the study conception and design. Model simulations were carried out by TN and RF. Data collection and analysis were performed by SP and RF. The first draft of the manuscript was written by SP and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

Supercomputing power was provided HLRN (North-German Supercomputing Alliance). The authors thank LUNG-MV and LfU-SH for providing all necessary observational data.

https://doi.org/10.5194/bg-2023-152
Preprint. Discussion started: 21 September 2023
© Author(s) 2023. CC BY 4.0 License.

References

Alexandrov, G. A., Ames, D., Bellocchi, G., Bruen, M., Crout, N., Erechtchoukova, M., Hildebrandt, A., Hoffman, F., Jackisch, C., Khaiter, P., Mannina, G., Matsunaga, T., Purucker, S. T., Rivington, M., and Samaniego, L.: Technical assessment and evaluation of environmental models and software: Letter to the Editor, Environ. Model. Softw., 26, 328–336, https://doi.org/10.1016/j.envsoft.2010.08.004, 2011.

Allen, J.I., Somerfield, P.J., and Gilbert, F.J.: Quantifying uncertainty in high-resolution coupled hydrodynamic-ecosystem models, J. Mar. Sys., 64, 3-14, https://doi.org/10.1016/j.jmarsys.2006.02.010, 2007.

Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. A., Hamilton, S. H., Jakeman, A. J., Marsili-Libelli, S., Newham, L. T. H., Norton, J. P., Perrin, C., Pierce, S. A., Robson, B., Seppelt, R., Voinov, A. A., Fath, B. D., and Andreassian, V.: Characterising performance of environmental models, Environ. Model. Softw., 40, 1–20, https://doi.org/10.1016/j.envsoft.2012.09.011, 2013.

BLANO: Zustand der deutschen Ostseegewässer 2018. Aktualisierung der Anfangsbewertung nach § 45c, der Beschreibung des guten Zustands der Meeresgewässer nach § 45d und der Festlegung von Zielen nach § 45e des Wasserhaushaltsgesetzes zur Umsetzung der Meeresstrategie-Rahmenrichtlinie. Anlage 1: Ergänzende nationale Indikatorblätter, 2018. Retrieved August 30, 2023, from https://mitglieder.meeresschutz.info/files/meeresschutz/berichte/art8910/zyklus18/Nationale_Indikatoren_Ostsee_Anlage_1.pdf.

Boesch, D. F.: Barriers and Bridges in Abating Coastal Eutrophication, Front. Mar. Sci., 6, https://doi.org/10.3389/fmars.2019.00123, 2019.

Bograd, S. J., Buil, M. P., Lorenzo, E. Di, Castro, C. G., Schroeder, I. D., Goericke, R., Anderson, C. R., Benitez-Nelson, C., and Whitney, F. A.: Changes in source waters to the Southern California Bight, Deep. Res. Part II Top. Stud. Oceanogr., 112, 42–52, https://doi.org/10.1016/j.dsr2.2014.04.009, 2015.

Breitburg, D., Levin, L. A., Oschlies, A., Grégoire, M., Chavez, F. P., Conley, D. J., Garçon, V., Gilbert, D., Gutiérrez, D., Isensee, K., Jacinto, G. S., Limburg, K. E., Montes, I., Naqvi, S. W. A., Pitcher, G. C., Rabalais, N. N., Roman, M. R., Rose, K. A., Seibel, B. A., Telszewski, M., Yasuhara, M., and Zhang, J.: Declining oxygen in the global ocean and coastal waters, Science (80-.)., 359, eaam7240, https://doi.org/10.1126/science.aam7240, 2018.

Carstensen, J., Andersen, J.H., Gustafsson, B.G., and Conley, D.J.: Deoxygenation of the Baltic Sea during the last century, PNAS, 111 (15), 5628-5633, https://doi.org/10.1073/pnas.1323156111, 2014.

Carstensen, J. and Conley, D. J.: Baltic Sea Hypoxia Takes Many Shapes and Sizes, Limnol. Oceanogr. Bull., 28, 125–129, https://doi.org/10.1002/lob.10350, 2019.

Conley, D. J., Carstensen, J., Aigars, J., Axe, P., Bonsdorff, E., Eremina, T., Haahti, B.-M., Humborg, C., Jonsson, P., Kotta, J., Lännegren, C., Larsson, U., Maximov, A., Medina, M. R., Lysiak-Pastuszak, E., Remeikaitė-Nikienė, N., Walve, J., Wilhelms, S., and Zillén, L.: Hypoxia Is Increasing in the Coastal Zone of the Baltic Sea, Environ. Sci. Technol., 45, 6777–

https://doi.org/10.5194/bg-2023-152
Preprint. Discussion started: 21 September 2023
© Author(s) 2023. CC BY 4.0 License.
Biogeosciences Discussions Open Access EGU

6783, https://doi.org/10.1021/es201212r, 2011.

Diaz, R. J. and Rosenberg, R.: Spreading Dead Zones and Consequences for Marine Ecosystems, Science (80-.)., 321, 926–929, https://doi.org/10.1126/science.1156401, 2008.

Diaz, R., Selman, M., and Chique, C.: Global Eutrophic and Hypoxic Coastal Systems. Washington, DC: World Resources Institute. Eutrophication and Hypoxia: Nutrient Pollution in Coastal Waters, 2011. Available online at: https://datasets.wri.org/dataset/eutrophication-hypoxia-map-data-set.

Dos Santos Fernandes De Araujo, R. and Boschetti, S.: Marine Strategy Framework Directive Review and analysis of EU Member States' 2018 reports - Descriptor 5: Eutrophication, EUR 30677 EN, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-36246-3, doi:10.2760/180642, JRC124915.

Dos Santos Fernandes De Araujo, R., Somma, F., Aigars, J., Axe, P., Bartolo, A., De Cauwer, K., Devreker, D., Fleming-Lehtinen, V., Garcia-Denche, L., Knudsen-Leerbeck, H., Krasniewski, W., Lazar, L., Lefebvre, A., Lips, I., Magaletti, E., Moncheva, S., Pagou, P., Precali, R., Ruiter, H. and Wilkes, R.: Eutrophication in marine waters: harmonization of MSFD methodological standards at EU level, EUR 29854 EN, Publications Office of the European Union, Luxembourg, 2019, ISBN 978-92-76-11224-2, doi:10.2760/437291, JRC117109.

Eilola, K., Gustafsson, B. G., Kuznetsov, I., Meier, H. E. M., Neumann, T., and Savchuk, O. P.: Evaluation of biogeochemical cycles in an ensemble of three state-of-the-art numerical models of the Baltic Sea, J. Mar. Syst., 88, 267–284, https://doi.org/10.1016/j.jmarsys.2011.05.004, 2011.

Fennel, K., Gehlen, M., Brasseur, P., Brown, C. W., Ciavatta, S., Cossarini, G., Crise, A., Edwards, C. A., Ford, D., Friedrichs, M. A. M., Gregoire, M., Jones, E., Kim, H.-C., Lamouroux, J., Murtugudde, R., and Perruche, C.: Advancing Marine Biogeochemical and Ecosystem Reanalyses and Forecasts as Tools for Monitoring and Managing Ecosystem Health, Front. Mar. Sci., 6, https://doi.org/10.3389/fmars.2019.00089, 2019.

Fennel, K., Mattern, J. P., Doney, S. C., Bopp, L., Moore, A. M., Wang, B., and Yu, L.: Ocean biogeochemical modelling, Nat. Rev. Methods Prim., 2, 76, https://doi.org/10.1038/s43586-022-00154-2, 2022.

Friedland, R., Neumann, T., and Schernewski, G.: Climate change and the Baltic Sea action plan : Model simulations on the future of the western Baltic Sea, J. Mar. Syst., 105–108, 175–186, https://doi.org/10.1016/j.jmarsys.2012.08.002, 2012.

Fringer, O. B., Dawson, C. N., He, R., Ralston, D. K., and Zhang, Y. J.: The future of coastal and estuarine modeling: Findings from a workshop, Ocean Model., 143, https://doi.org/10.1016/J.OCEMOD.2019.101458, 2019.

Gauss, M., Bartnicki, J., Jalkanen, J., Nyiri, A., Klein, H., Fagerli, H., and Klimont, Z.: Airborne nitrogen deposition to the Baltic Sea: Past trends, source allocation and future projections, Atmos. Environ., 253, 118377, https://doi.org/10.1016/j.atmosenv.2021.118377, 2021.

Geyer, B., and Rockel, B.: coastDat-2 COSMO-CLM atmospheric reconstruction. World Data Center for Climate (WDCC) at DKRZ. Available at: https://doi.org/10.1594/WDCC/coastDat-2_COSMO-CLM, 2013.

Gilbert, D., Rabalais, N. N., Díaz, R. J., and Zhang, J.: Evidence for greater oxygen decline rates in the coastal ocean than in the open ocean, Biogeosciences, 7, 2283–2296, https://doi.org/10.5194/bg-7-2283-2010, 2010.

https://doi.org/10.5194/bg-2023-152
Preprint. Discussion started: 21 September 2023
© Author(s) 2023. CC BY 4.0 License.

Griffies, S. M., Harrison, M. J., Pacanowski, R. C., and Rosati, a: A technical guide to MOM4, GFDL Ocean Gr. Tech. Rep, 5, 371, 2004.

Griffies, S. M.: Elements of the Modular Ocean Model (MOM) - 2012 release with updates. GFDL Ocean Group technical report no. 7. NOAA/Geophysical Fluid Dynamics Laboratory, 2012.

Gröger, M., Placke, M., Meier, H. E. M., Börgel, F., Brunnabend, S., Dutheil, C., Gräwe, U., Hieronymus, M., Neumann, T., Radtke, H., Su, J., and Väli, G.: The Baltic Sea model inter-comparison project BMIP – a platform for model development , evaluation , and uncertainty assessment, 1–34, https://doi.org/10.5194/gmd-15-8613-2022, 2022.

Große, F., Greenwood, N., Kreus, M., Lenhart, H.-J., Machoczek, D., Pätsch, J., Salt, L., and Thomas, H.: Looking beyond stratification: a model-based analysis of the biological drivers of oxygen deficiency in the North Sea, Biogeosciences, 13, 2511–2535, https://doi.org/10.5194/bg-13-2511-2016, 2016.

HELCOM: Approaches and methods for eutrophication target setting in the Baltic Sea region. Balt. Sea Environ. Proc. No. 133, 2013. Retrieved August 20, 2023, from https://helcom.fi/wp-content/uploads/2019/10/BSEP133.pdf

HELCOM: HELCOM Thematic assessment of eutrophication 2011-2016. Baltic Sea Environment Proceedings No. 156, 2018. Retrieved August 30, 2023, from https://helcom.fi/wp-content/uploads/2019/12/BSEP156.pdf

HELCOM: HELCOM Baltic Sea Action Plan – 2021 update. ELCOM 2021, 2021. Retrieved August 30, 2023, from https://helcom.fi/wp-content/uploads/2021/10/Baltic-Sea-Action-Plan-2021-update.pdf

HELCOM: Shallow water oxygen. HELCOM core indicator report, 2023. Retrieved August 20, 2023, https://indicators.helcom.fi/indicator/shallow-water-oxygen/.

I N-EUTROPHICATION 16-2020: HELCOM indicator on near-bottom oxygen concentrations in shallow waters - towards a Baltic Sea wide application. Code 5-2; Category CMNT; Agenda Item 5– Further work on indicators and assessments, 2020.

Karlson, K., Rosenberg, R., and Bonsdorff, E.: Temporal and spatial large-scale effects of eutrophication and oxygen deficiency on benthic fauna in Scandinavian and Baltic waters: a review, in: Gibson, R.N. et al. Oceanogr. Mar. Biol. Ann. Rev. 40. Oceanography and Marine Biology: An Annual Review, 40: pp. 427-489, 2002.

Kuznetsov, I. and Neumann, T.: Simulation of carbon dynamics in the Baltic Sea with a 3D model, J. Mar. Syst., 111–112, 167–174, https://doi.org/10.1016/j.jmarsys.2012.10.011, 2013.

Lagarias, J. C., Reeds, J. A., Wright, M. H., and Wright, P. E.: Convergence properties of the Nelder-Mead simplex method in low dimensions, SIAM J. Optim., 9, 112–147, https://doi.org/10.1137/S1052623496303470, 1998.

Laurent, A. and Fennel, K.: Time-Evolving, Spatially Explicit Forecasts of the Northern Gulf of Mexico Hypoxic Zone, Environ. Sci. Technol., 53, 14449–14458, https://doi.org/10.1021/ACS.EST.9B05790/SUPPL_FILE/ES9B05790_SI_001.PDF, 2019.

Lehmann, M. K., Fennel, K., and He, R.: Statistical validation of a 3-D bio-physical model of the western North Atlantic, Biogeosciences, 6, 1961–1974, https://doi.org/10.5194/bg-6-1961-2009, 2009.

Liblik, T., Stoicescu, S. T., Buschmann, F., Lilover, M. J., and Lips, U.: High-resolution characterization of the development and decay of seasonal hypoxia in the Gulf of Riga, Baltic Sea, Front. Mar. Sci., 10, 1119515, https://doi.org/10.3389/FMARS.2023.1119515/BIBTEX, 2023.

Meier, H. E. M., Edman, M. K., Eilola, K. J., Placke, M., Neumann, T., Andersson, H. C., Brunnabend, S. E., Dieterich, C., Frauen, C., Friedland, R., Gröger, M., Gustafsson, B. G., Gustafsson, E., Isaev, A., Kniebusch, M., Kuznetsov, I., Müller-Karulis, B., Omstedt, A., Ryabchenko, V., Saraiva, S., and Savchuk, O. P.: Assessment of eutrophication abatement scenarios for the baltic sea by multi-model ensemble simulations, Front. Mar. Sci., 5, 1–24, https://doi.org/10.3389/fmars.2018.00440, 2018a.

Meier, H. E. M., Väli, G., Naumann, M., Eilola, K., and Frauen, C.: Recently Accelerated Oxygen Consumption Rates Amplify Deoxygenation in the Baltic Sea, J. Geophys. Res. Ocean., 123, 3227–3240, https://doi.org/10.1029/2017JC013686, 2018b.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I — A discussion of principles, J. Hydrol., 10, 282–290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

Naumov, L., Meier, H. E. M., and Neumann, T.: Dynamics of oxygen sources and sinks in the Baltic Sea under different nutrient inputs, 1–9, https://doi.org/10.3389/fmars.2023.1233324, 2023.

Neumann, T. and Schernewski, G.: Eutrophication in the Baltic Sea and shifts in nitrogen fixation analyzed with a 3D ecosystem model, J. Mar. Syst., 74, 592–602, https://doi.org/10.1016/j.jmarsys.2008.05.003, 2008.

Neumann, T., Radtke, H., and Seifert, T.: On the importance of Major Baltic Inflows for oxygenation of the central Baltic Sea, J. Geophys. Res. Ocean., 122, 1090–1101, https://doi.org/10.1002/2016JC012525, 2017.

Neumann, T., Siegel, H., Moros, M., Gerth, M., Kniebusch, M., and Heydebreck, D.: Ventilation of the northern Baltic Sea, Ocean Sci., 16, 767–780, https://doi.org/10.5194/os-16-767-2020, 2020.

Neumann, T., Koponen, S., Attila, J., Brockmann, C., Kallio, K., Kervinen, M., Mazeran, C., Müller, D., Philipson, P., Thulin, S., Väkevä, S., and Ylöstalo, P.: Optical model for the Baltic Sea with an explicit CDOM state variable: A case study with Model ERGOM (version 1.2), Geosci. Model Dev., 14, 5049–5062, https://doi.org/10.5194/gmd-14-5049-2021, 2021.

Piehl, S., Friedland, R., Heyden, B., Leujak, W., Neumann, T., and Schernewski, G.: Modeling of Water Quality Indicators in the Western Baltic Sea: Seasonal Oxygen Deficiency, Environ. Model. Assess., 28, 429–446, https://doi.org/10.1007/s10666-022-09866-x, 2023.

Reusch, T. B. H., Dierking, J., Andersson, H. C., Bonsdorff, E., Carstensen, J., Casini, M., Czajkowski, M., Hasler, B., Hinsby, K., Hyytiäinen, K., Johannesson, K., Jomaa, S., Jormalainen, V., Kuosa, H., Kurland, S., Laikre, L., MacKenzie, B. R., Margonski, P., Melzner, F., Oesterwind, D., Ojaveer, H., Refsgaard, J. C., Sandström, A., Schwarz, G., Tonderski, K., Winder, M., and Zandersen, M.: The Baltic Sea as a time machine for the future coastal ocean, Sci. Adv., 4, eaar8195, https://doi.org/10.1126/sciadv.aar8195, 2018.

Robson, B. J.: When do aquatic systems models provide useful predictions, what is changing, and what is next?, Environ. Model. Softw., 61, 287–296, https://doi.org/10.1016/j.envsoft.2014.01.009, 2014.

Rose, K.A., Justic, D., Fennel, K., Hetland, R.D.: Numerical Modeling of Hypoxia and Its Effects: Synthesis and Going Forward. In: Justic, D., Rose, K., Hetland, R., Fennel, K. (eds) Modeling Coastal Hypoxia. Springer, Cham. https://doi.org/10.1007/978-3-319-54571-4_15, 2017.

https://doi.org/10.5194/bg-2023-152
Preprint. Discussion started: 21 September 2023
© Author(s) 2023. CC BY 4.0 License.

Stow, C. A., Jolliff, J., McGillicuddy, D. J., Doney, S. C., Allen, J. I., Friedrichs, M. A. M., Rose, K. A., and Wallhead, P.: Skill assessment for coupled biological/physical models of marine systems, J. Mar. Syst., 76, 4–15, https://doi.org/10.1016/j.jmarsys.2008.03.011, 2009.

Vaquer-Sunyer, R. and Duarte, C. M.: Thresholds of hypoxia for marine biodiversity, Proc. Natl. Acad. Sci., 105, 15452–15457, https://doi.org/10.1073/pnas.0803833105, 2008.

Virtanen, E. A., Norkko, A., Nyström Sandman, A., and Viitasalo, M.: Identifying areas prone to coastal hypoxia – the role of topography, Biogeosciences, 16, 3183–3195, https://doi.org/10.5194/bg-16-3183-2019, 2019.

Vock, C., Friedland, R., and Piehl, S.: Near-bottom Oxygen observations from the Western Baltic Sea. 2023. DOI:10.5281/zenodo.8116592, 2023.

Weisse, R., von Storch, H., Callies, U., Chrastansky, A., Feser, F., Grabemann, I., Günther, H., Pluess, A., Stoye, T., Tellkamp, J., Winterfeldt, J., and Woth, K.: Regional Meteorological–Marine Reanalyses and Climate Change Projections, Bull. Am. Meteorol. Soc., 90, 849–860, https://doi.org/10.1175/2008BAMS2713.1, 2009.

Winton, M.: A reformulated three-layer sea ice model, J. Atmos. Ocean. Technol., 17, 525–531, https://doi.org/10.1175/1520-0426(2000)017<0525:ARTLSI>2.0.CO;2, 2000.