Oceanologia, corrected, in press
Polish Academy of Sciences
Institute of Oceanology
Open access article under the CC BY license
https://doi.org/10.5697/...

Development and validation of the 3D CEMBS-PolSea biogeochemical model and its application to ecosystem analysis in the Southern Baltic Sea

Lidia Dzierzbicka-Głowacka*, Dawid Dybowski, Maciej Janecki, Artur Nowicki

Abstract

This study presents the development, implementation, and validation of the high-resolution three-dimensional biogeochemical 3D CEMBS-PolSea model for the southern Baltic Sea. The model is based on the NPZD (Nutrient–Phytoplankton–Zooplankton–Detritus) framework and simulates key ecosystem components and processes, including phytoplankton, zooplankton, chlorophyll a, dissolved oxygen, nutrients (NO_3, PO_4, SiO_3, NH_4), and organic matter dynamics. The model was developed within the CSI-POM 2 project and integrated into an operational environmental information system for Polish marine areas.

Model performance was evaluated using in situ observations from the ICES Hydrochemistry databases and satellite-derived chlorophyll-a products from the SatBałtyk system. Validation results demonstrated the highest agreement for dissolved oxygen, with correlation coefficients reaching 0.87 and 0.72 for the ICES Low Resolution and High Resolution datasets, respectively. Moderate agreement was obtained for nutrients and chlorophyll a, with correlation coefficients ranging from 0.51 to 0.68. Additional performance metrics, including bias, normalised RMSE (nRMSE), the RMSE-observation standard deviation ratio (RSR), and the coefficient of variation (CV), indicate that model errors remain comparable to or lower than the variability of the observational datasets. These results demonstrate that the model reproduces a substantial fraction of the observed environmental variability despite the high natural variability of biogeochemical conditions in the Baltic Sea.

The integration of satellite-derived chlorophyll a assimilation improves the representation of phytoplankton dynamics and enhances the operational applicability of the modelling system. The developed 3D CEMBS-PolSea model provides a valuable tool for ecosystem assessment, environmental monitoring, forecasting, and marine management in the southern Baltic Sea.

Keywords

Numerical modelling; High-resolution; Biogeochemical; Southern Baltic

Institute of Oceanology, Polish Academy of Sceinces, Sopot, Poland
***Correspondence**: dzierzb@iopan.pl (L. Dzierzbicka-Głowacka)

Received: 13 April 2026; **revised:** 23 June 2026; **accepted**: 26 August 2026

1. Introduction

The Baltic Sea ecosystem is strongly influenced by riverine inflows, episodic inflows of saline North Sea waters through the Danish Straits, and local hydrodynamic processes such as upwelling, stratification, and vertical mixing. Excessive nutrient inputs from land-based sources intensify eutrophication, stimulating phytoplankton blooms and contributing to oxygen depletion in deeper water layers. Because observational programmes provide limited spatial and temporal coverage, numerical models have become indispensable tools for analysing marine ecosystems, investigating environmental processes, and forecasting their variability (Daewel and Schrum, 2013; Dzierzbicka-Głowacka et al., 2013a,b; Eilola et al., 2009; Gräwe et al., 2015; Hordoir et al., 2019; Meier et al., 2014; Murawski et al., 2021).

Over the last two decades, numerous hydrodynamic and biogeochemical models have been developed for the Baltic Sea. Examples include the Baltic Sea Long-Term Large-Scale Eutrophication Model (BALTSEM) (Gustafsson et al., 2012), ERGOM-based ecosystem models (Eilola et al., 2009), the NEMO-Nordic modelling framework (Hordoir et al., 2019), and coupled regional ecosystem models such as ECOSMO (Daewel and Schrum, 2013). These modelling systems have substantially improved our understanding of nu-

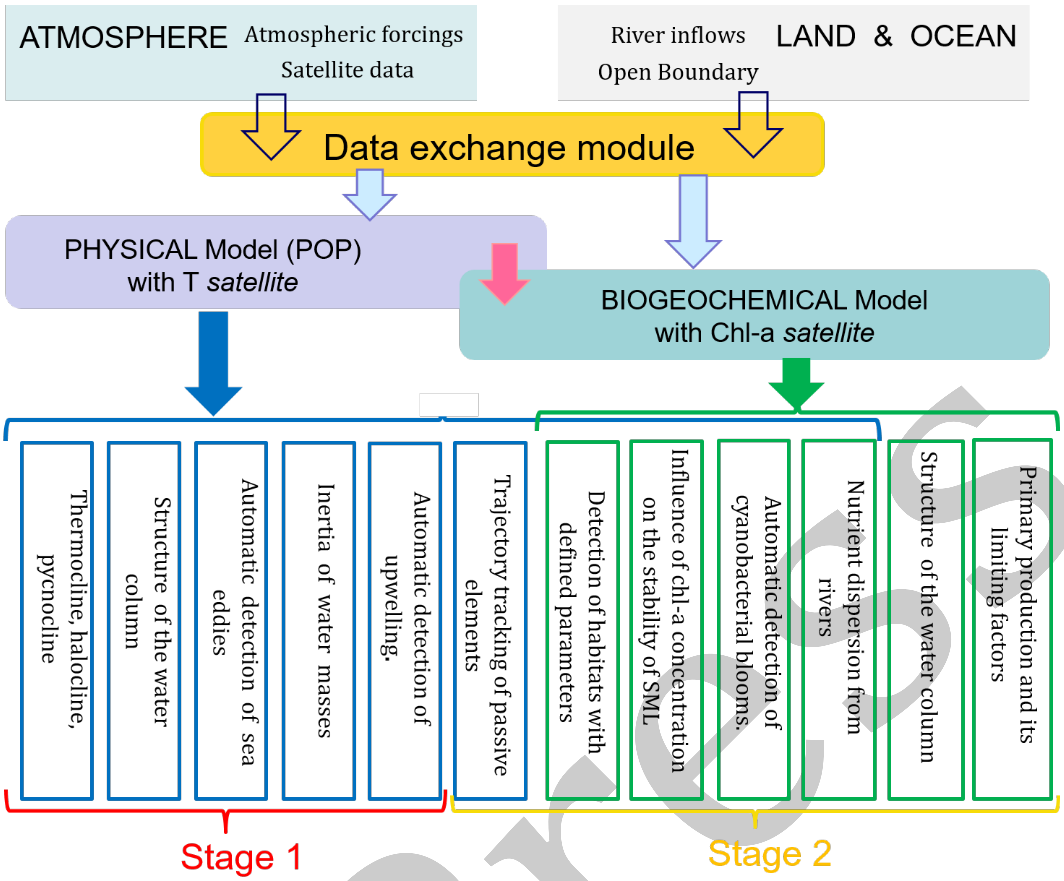

Figure 1. Conceptual architecture of the CSI-POM and CSI-POM 2 operational environmental information systems and the integration of the 3D CEMBS-PolSea ecohydrodynamic model. Stage 1 – components developed within CSI-POM while Stage 2 – components developed within CSI-POM 2.

trient cycling, oxygen dynamics, eutrophication processes, and ecosystem responses to climate variability and anthropogenic pressures. Nevertheless, the complex interactions between physical forcing, biogeochemical processes, and human activities continue to represent a major scientific challenge, particularly in coastal and shelf environments characterised by strong environmental gradients and high temporal variability.

Despite significant advances in marine monitoring and numerical modelling, comprehensive analysis of processes occurring in coastal ecosystems remains challenging. This is particularly evident in semi-enclosed basins such as the Baltic Sea, where strong vertical stratification, highly variable riverine freshwater and nutrient inputs, irregular inflows from the North Sea, and complex biogeochemical interactions control ecosystem functioning (Feistel et al., 2008; Leppäranta and Myrberg, 2009; Stramska et al., 2021). In recent decades, increasing anthropogenic pressure and eutrophication have further altered the ecological status of the Baltic Sea, leading to changes in phytoplankton biomass, water transparency, hypoxia, and the occurrence of harmful algal blooms (Gustafsson et al., 2012; HELCOM, 2023; Stramska et al., 2021). Owing to the optical complexity of Baltic waters and the strong spatial and temporal variability of environmental conditions, reliable monitoring and forecasting require the integration of in situ observations, satellite remote sensing products, and advanced numerical modelling systems (Darecki and Stramski, 2004; She et al., 2020; Stramska et al., 2021).

Although numerous studies have been conducted, uncertainties remain regarding the quantitative importance of individual ecosystem processes and their interactions. Reliable environmental assessments therefore require high-quality observational datasets together with robust numerical models capable of reproducing the observed variability of marine conditions. Model validation and quantitative evaluation of model performance are essential prerequisites for the application of numerical simulations in ecosystem research, environmental assessment, and operational forecasting.

While many Baltic Sea ecosystem models have proven successful as scientific research tools, their integration into operational environmental information systems remains relatively limited. Consequently, a gap persists between scientific advances in marine ecosystem modelling and their practical application in environmental management, marine spatial planning, and decision-support systems. Bridging this gap requires not only accurate numerical models but also efficient mechanisms for transforming model outputs into operational environmental information products accessible to end users.

The Digital System for Environmental Information of Polish Marine Areas (CSI-POM Stage 2, http://www.csipom2.pl), developed as a continuation of CSI-POM Stage 1 (http://www.csipom.pl), aims to address this challenge by integrating advanced numerical modelling with an operational environmental information system for Polish marine areas (Dybowski et al., 2025; Dzierzbicka-Głowacka et al., 2025). The overall structure of the CSI-POM and CSI-POM 2 systems, together with the role of the 3D CEMBS-PolSea model within the operational framework, is presented in Figure 1.

The three-dimensional prognostic ecohydrodynamic 3D CEMBS-PolSea model constitutes the core numerical component of the Digital System for Environmental Information of Polish Marine Areas (CSI-POM and CSI-POM 2). The system consists of two principal modules: a hydrodynamic module and a biogeochemical module, supplemented by dedicated components responsible for atmospheric forcing, riverine inputs, satellite data assimilation, operational management, and dissemination of environmental information products. The hydrodynamic component was validated independently and is described in detail in Dybowski et al. (2025).

In addition to the numerical modelling components, the system includes supervisory procedures responsible for workflow control, data exchange, quality assurance, data archiving, and user-oriented product generation. While CSI-POM Stage 1 focused primarily on the analysis and forecasting of physical marine parameters, CSI-POM Stage 2 extends these capabilities through the implementation of advanced biogeochemical modelling and ecosystem analysis tools. The development of the 3D CEMBS-PolSea biogeochemical module therefore represents a key step towards integrating high-resolution ecosystem modelling into an operational environmental information system supporting environmental monitoring, scientific research, and marine management in Polish marine areas.

The main objective of this study is the development, implementation, and validation of a high-resolution biogeochemical module of the 3D CEMBS-PolSea model for the southern Baltic Sea. The model enables simulations of phytoplankton and zooplankton biomass, chlorophyll *a* concentration, dissolved oxygen, nutrient dynamics, and organic matter transformations. Particular emphasis is placed on model validation using in situ observations from the ICES database and satellite-derived chlorophyll *a* products, as well as on the assessment of the model's ability to reproduce the spatial and temporal variability of key biogeochemical parameters in the southern Baltic Sea.

This paper focuses on the structure, implementation, and performance of the biogeochemical component of the 3D CEMBS-PolSea system and its applicability to ecosystem analysis in the southern Baltic Sea. The architecture and functionality of the operational CSI-POM 2 environmental information platform, in which model products are disseminated to end users, will be presented in a separate publication.

2. Material and methods

Achieving the objectives of the CSI-POM Stage 2 project, namely determining the dynamics of biogeochemical parameters in the southern Baltic Sea, requires the implementation of the following stages:

2.1 Model configuration

As part of model setup, a comprehensive dataset of input fields was prepared, including initial conditions, atmospheric forcing, boundary conditions at the water–water interface, as well as river inflows and nutrient deposition, all supplied through dedicated data modules.

2.1.1 Initial conditions

Initial fields were prepared on the 3D CEMBS-PolSea model grid developed during the first stage of the CSI-POM project. A horizontal resolution of approximately 575 m was applied, representing a compromise between accurate representation of physical and biogeochemical processes and computational efficiency. The vertical grid consists of 66 layers. In the upper part of the water column (down to 255 m depth), layers of 5 m thickness are used, corresponding to 26 vertical layers within the Polish marine areas. Below this depth, the layer thickness gradually increases, reaching a maximum of 83 m in the deepest parts of the southern Baltic Sea domain (Dybowski et al., 2025).

Bathymetry was developed based on EMODnet Bathymetry data, which provide consistent digital terrain models for European marine areas. The resulting bathymetry was subsequently corrected by adjusting it to a high-resolution coastline to ensure accurate representation of flows in the coastal zone (see Figure 2).

Initial conditions were prepared for 1 January 2020 and include, among others, concentrations of nitrates, phosphates, ammonia, silicates, dissolved oxygen, phytoplankton and zooplankton biomass, chlorophyll *a* concentration, and organic matter content. These data cover the entire water column at each grid point. The source data were derived from two main datasets. The first was the *Baltic Sea Biogeochemistry Reanalysis* provided by the Marine Copernicus system, based on the ERGOM model coupled with the NEMO ocean–ice model. The second source was the 3D CEMBS model with a spatial resolution of approximately 2.3 km, developed at the Institute of Oceanology, Polish Academy of Sciences.

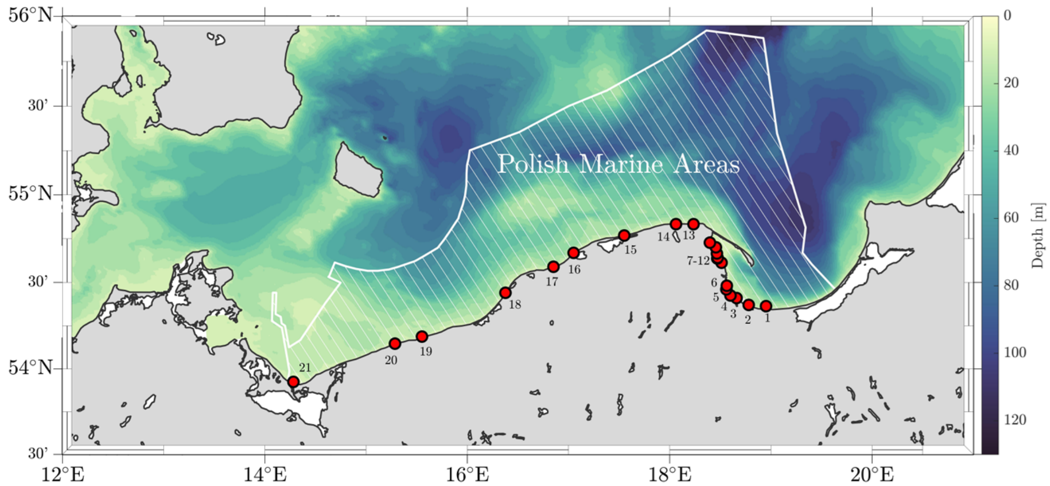

Figure 2. Bathymetry of the "effective domain" of the three-dimensional CEMBS-PolSea model, including the locations of 21 river mouths incorporated into the model and the boundaries of the Polish Marine Areas.

Due to the lower spatial resolution of the source data compared to the target model, the datasets were interpolated onto the 3D CEMBS-PolSea grid. To avoid data gaps in coastal regions, extrapolation using a Poisson Solver was applied. This was followed by vertical and horizontal interpolation using the Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) method. The final initial fields were stored in netCDF format.

2.1.2 Atmospheric forcing

Atmospheric forcing fields were prepared based on forecasts from the Unified Model (UM), developed at the Interdisciplinary Centre for Mathematical and Computational Modelling of the University of Warsaw (ICM UW). The model provides meteorological forecasts four times per day. Input data include wind components at 10 m, air temperature at 2 m, specific humidity, dew point temperature, sea-level atmospheric pressure, precipitation, and shortwave and longwave radiation components. These data are interpolated onto the 3D CEMBS-PolSea grid and stored in netCDF files with an hourly temporal resolution.

2.1.3 Open boundary conditions

The 3D CEMBS-PolSea model domain covers the southern Baltic Sea and is connected to the Danish Straits and the central Baltic. Therefore, appropriate open boundary conditions are required. These are derived from the operational 3D CEMBS model with a spatial resolution of approximately 2 km (Dzierzbicka-Głowacka et al., 2013a). The data are provided at 6-hour intervals as up-to-date model outputs and are subsequently processed through extrapolation using a Poisson Solver, followed by vertical and horizontal interpolation onto the 3D CEMBS-PolSea grid.

2.1.4 River inflows and nutrient deposition

The inclusion of river inflows is a key component in modelling coastal processes. The 3D CEMBS-PolSea model incorporates 21 rivers discharging into the southern Baltic Sea from Polish territory. River flow data are derived from SWAT hydrological models (Dybowski et al., 2020, 2019; Dybowski and Dzierzbicka-Głowacka, 2023; Dzierzbicka-Głowacka et al., 2022), the HYPE model, and operational measurements conducted by the Institute of Meteorology and Water Management (IMGW) (Table 1 in Dybowski et al., 2025).

In the second stage of the project, the 3D CEMBS-PolSea model was extended to include riverine nutrient inputs. In situ concentration data for nutrients were obtained from the database of the Chief Inspectorate for Environmental Protection (GIOŚ), which conducts independent monitoring of riverine inputs through chemical and biological analyses of water samples collected at various locations and periods throughout the year. The GIOŚ monitoring programme is independent of measurements performed by the Institute of Meteorology and Water Management (IMGW). The collected measurements were available for all rivers implemented in the model domain; however, they did not include a complete set of biogenic substances transported by rivers. The available data included primarily concentrations of nitrates, phosphates, and dissolved oxygen, while information on silicate and chlorophyll *a* concentrations was insufficient to reconstruct their seasonal variability and was therefore supplemented using available literature sources. To obtain continuous daily variability of nutrient concentrations, the available measurements were interpolated using the PCHIP (Piecewise Cubic Hermite Interpolating Polynomial) method between observation points and extrapolated to the beginning and end of the year. This procedure resulted in continuous daily time series representing nutrient concentrations for each day of the year. Subsequently, the obtained concentrations were combined with daily freshwater discharge volumes from individual rivers to calculate daily nutrient loads entering the Baltic Sea. The calculated loads were converted into the units required by the model (kg m^{-2} s^{-1}) and prepared as input forcing files for the 3D CEMBS-PolSea model for each year of the 2018–2023 period. Additionally, based on the six-year dataset, mean daily nutrient concentrations were calculated and used to provide forcing data for the operational version of the 3D CEMBS-PolSea system. Due to the lack of real-time operational measurements of riverine nutrient concentrations, this approach provided a practical method for generating reliable nutrient deposition estimates suitable for operational forecasting applications.

2.2 Biogeochemical model

The biogeochemical component of the 3D CEMBS-PolSea model was developed to simulate the main processes governing nutrient cycling, primary production, organic matter transformation, and oxygen dynamics in the southern Baltic Sea. The model is based on the NPZD (Nutrient–Phytoplankton–Zooplankton–Detritus) concept and represents interactions among the principal biological and chemical compartments of the pelagic ecosystem.

The implemented module originates from the Baltic Sea version of the Community Earth System Model (CESM, configuration G) adapted and further developed for regional applications in the southern Baltic Sea. The model describes the coupled dynamics of three phytoplankton functional groups, zooplankton biomass, dissolved oxygen, organic matter, and the major nutrients controlling ecosystem productivity, including nitrates (NO_3), ammonia (NH_4), phosphates (PO_4), and silicates (SiO_3).

Biogeochemical processes are represented through a system of coupled advection–diffusion–reaction equations, in which transport processes are provided by the hydrodynamic model and biological transformations are represented by source–sink functions. These formulations describe nutrient uptake by phytoplankton, grazing by zooplankton, mortality, remineralisation of organic matter, oxygen production and consumption, and nutrient regeneration processes. (Dzierzbicka-Głowacka et al., 2013b; Moore et al., 2001). A conceptual representation of the model structure is presented in Figure 3.

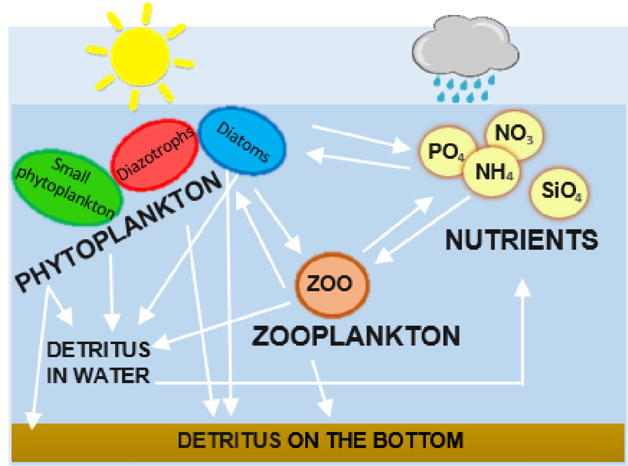

Figure 3. Simplified conceptual scheme of the NPZD biogeochemical model.

2.3 Hydrodynamic forcing of the biogeochemical module

The biogeochemical component of the 3D CEMBS-PolSea system is driven by hydrodynamic fields generated by the CSI-POM model. The physical model provides the environmental forcing required for biogeochemical simulations, including three-dimensional fields of temperature, salinity, currents, turbulent mixing, and sea-surface elevation. These variables determine the transport of nutrients and planktonic organisms, vertical exchange processes, and environmental conditions controlling biological production and biogeochemical transformations.

The hydrodynamic and biogeochemical modules operate on the same computational grid and use identical bathymetry, open boundary conditions, atmospheric forcing, and riverine inputs. Hydrodynamic fields are calculated continuously by the physical model and supplied to the biogeochemical module at each computational time step. The exchange of information is therefore unidirectional, from the hydrodynamic model to the biogeochemical component, with no feedback of biogeochemical processes on hydrodynamic calculations.

As a result, the distribution and variability of biogeochemical variables are directly controlled by physical processes such as advection, diffusion, stratification, upwelling, and vertical mixing. These processes regulate the transport and availability of nutrients, oxygen, and planktonic biomass throughout the water column and across the study area.

A detailed description and validation of the hydrodynamic model are presented by Dybowski et al. (2025). The present study focuses on the implementation and validation of the biogeochemical module, which uses the validated hydrodynamic model as its physical forcing framework.

2.4 Satellite data assimilation

2.4.1 Satellite data

The data used in the assimilation process originate from the SatBałtyk system and represent processed satellite observation products. These data are stored as daily NetCDF files with a horizontal spatial resolution of approximately 300 m. Each file contains information on total chlorophyll *a* concentration (*ChlaSAT_total*) and chlorophyll *a* concentration attributed to cyanobacteria (*ChlaSAT_diaz*), as well as a mask indicating their spatial occurrence. Cyanobacterial

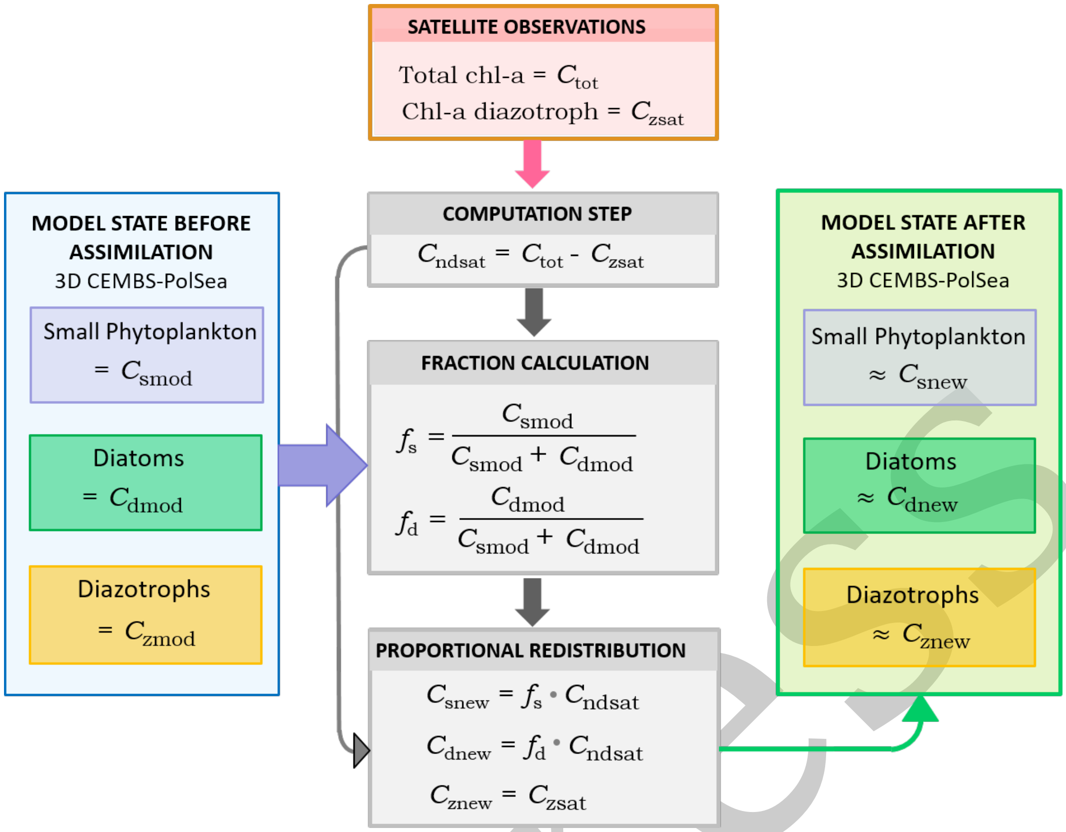

Figure 4. Schematic representation of the chlorophyll *a* data assimilation procedure in the 3D CEMBS-PolSea model.

contribution to chlorophyll *a* (*chlPC*) was estimated using a two-step approach. In the first step, phycocyanin concentration (*PC*) was derived from spectral reflectance following the formulation proposed by Woźniak et al. (2016). In the second step, *PC* concentration was converted into cyanobacteria-associated chlorophyll *a* (*chlPC*) under the assumption of cyanobacterial dominance (by Stoń-Egiert (IOPAN) – latest version: 2024-07-02). The detailed description of the satellite-derived data products and the data assimilation methodology is beyond the scope of this paper and will be presented in a separate publication.

2.4.2 Implementation of the assimilation module

The assimilation module is based on modified and extended components of the Community Earth System Model (CESM, configuration type G), a coupled global climate model developed by the National Center for Atmospheric Research (NCAR), and constitutes an integral part of the three-dimensional 3D CEMBS-PolSea model. The applied assimilation approach uses both model-derived variables and satellite observations as input data.

The assimilation process is performed daily, provided that satellite data are available, within a 24-hour assimilation window centered at 12:00 (noon). The sequence of operations carried out within the assimilation module for all model grid points is described in detail in a separate publication (in preparation). Additionally, data filtering procedures have been implemented to enhance the stability of the assimilation process by reducing or eliminating the influence of unrealistic or erroneous satellite observations on the model system.

In the three-dimensional CEMBS-PolSea model, chlorophyll-a concentration is represented by three functional groups of phytoplankton: diazotrophs, diatoms, and small phytoplankton. In contrast, satellite products provide information on total chlorophyll *a* concentration (*ChlaSAT_total*) and the fraction attributed to diazotrophs (*ChlaSAT_diaz*). Therefore, an intermediate step is required to partition satellite observations into three components corresponding to the model state variables.

The procedure consists of the following steps:

1. Direct assimilation of diazotrophs

   The chlorophyll *a* concentration of diazotrophs (*ChlaSAT_diaz*) obtained from satellite observations is directly assimilated into the corresponding model variable, in accordance with the adopted assimilation methodology.

2. Redistribution of the remaining chlorophyll *a*

   The non-diazotrophic chlorophyll *a* fraction (*ChlaSAT_nondiaz*) is calculated as the difference between the total chlorophyll *a* concentration (*ChlaSAT_total*) and *ChlaSAT_diaz*:

$$ChlaSAT\_nondiaz = ChlaSAT\_total - ChlaSAT\_diaz$$

The remaining chlorophyll *a*, representing the combined contribution of diatoms and small phytoplankton, is then redistributed between these two groups proportionally to their relative shares in the model prior to assimilation. This approach ensures that the sum of chlorophyll *a* concentrations for both groups converges to the *ChlaSAT_nondiaz* value derived from satellite data, while changes in individual components are scaled according to their pre-assimilation proportions.

This proportional redistribution method preserves the ecological consistency of the model and prevents the introduction of unrealistic changes in the functional structure of phytoplankton communities. A schematic illustration of this procedure is presented in Figure 4 (details provided in a separate publication, in preparation).

The assimilation process enables dynamic correction of model outputs, thereby increasing their consistency with the actual state of the environment.

Satellite-derived chlorophyll *a* products do not represent chlorophyll concentration in a single surface layer but rather an optically weighted signal originating from the upper illuminated part of the water column. Consequently, the satellite observations are not directly equivalent to chlorophyll concentrations simulated in the uppermost model layer. In this study, satellite chlorophyll *a* data were compared with and assimilated into the upper 5 m model layer as an approximation of the near-surface chlorophyll field. This simplification may introduce uncertainties, particularly during stratified periods when subsurface chlorophyll maxima occur below the mixed layer and contribute to the satellite signal. Therefore, part of the differences between model results and satellite observations may arise from the different vertical representations of chlorophyll *a* in the observational and modelling systems. A more rigorous comparison would require the application of an observation operator that converts modelled chlorophyll profiles into optically weighted quantities consistent with ocean-colour observations. The upper 5 m layer was selected because it represents the model layer most directly associated with the near-surface optical signal and corresponds to the vertical resolution available in the operational model configuration.

2.5 Validation methodology

Model validation consisted of comparing simulation outputs with in situ and satellite observations, quantifying the degree of agreement, and assessing whether the model is suitable for further application.

2.5.1 Observational data

To evaluate the performance of the biogeochemical module of the 3D CEMBS-PolSea model, simulation results were compared with in situ observational data. The analysis covered a four-year period from 1 January 2022 to 31 December 2025, during which a continuous numerical simulation was performed. Model outputs were then compared with available experimental measurements to assess their consistency.

During simulations, results were stored as daily output files, each containing four time fields representing mean values calculated over six-hour intervals corresponding to 00:00–06:00, 06:00–12:00, 12:00–18:00, and 18:00–24:00. This approach preserves information on diurnal variability while limiting the volume of generated output data.

Observational data used for validation were obtained from the ICES database (International Council for the Exploration of the Sea, https://data.ices.dk/). Two datasets were used: *ICES Low Resolution Hydrochemistry* and *ICES High Resolution Hydrochemistry*.

From the Low Resolution dataset (2022–2024), measurements of dissolved oxygen (O_2), chlorophyll *a* (CHL), phosphates (PO_4), silicates (SiO_3), and nitrates (NO_3) were used. The High Resolution dataset provides higher spatial and vertical sampling resolution but for a more limited number of variables; in this study, only dissolved oxygen (O_2) data were used.

Data availability for 2025 is currently limited; therefore, the validation primarily relies on observations from 2022–2024. No measurements from 2025 were present in the Low Resolution dataset, while High Resolution data from 2025 account for approximately 5% of the amount available for earlier years.

2.5.2 Quality control

Prior to the model–observation matching procedure, an initial quality control of observational data was performed. Hydrochemical data from the ICES database were filtered to remove measurements associated with high uncertainty or failing basic quality criteria.

In the first step, empty or erroneous records, as well as entries lacking information on sampling time and/or location, were removed. Subsequently, a filter based on Quality Flags provided within the ICES datasets was applied. Only high-quality measurements were retained, i.e., those for which quality flags for both depth and the analysed variable indicated valid or acceptable data. Measurements marked as doubtful or erroneous were excluded from further analysis.

Additionally, the dataset was restricted to points located within the 3D CEMBS-PolSea model domain. Observations outside the model grid or those that could not be assigned to a sufficiently close grid cell were discarded (e.g., measurements from the Vistula Lagoon or Szczecin Lagoon).

2.5.3 Match-up procedure

Comparing model outputs with observations requires matching values in space, time, and depth. A modified

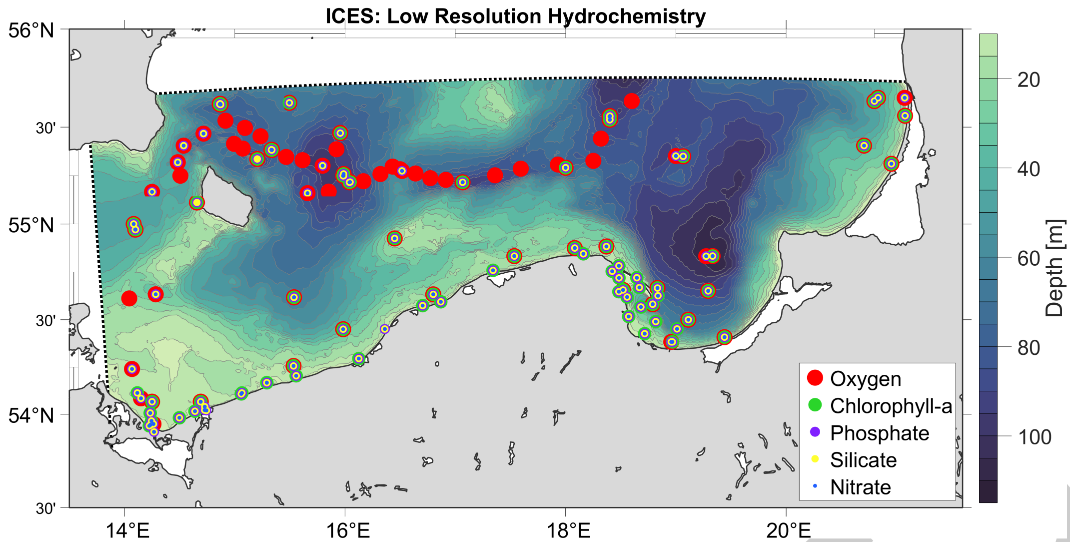

Figure 5. Spatial distribution of data from the ICES Low Resolution Hydrochemistry dataset used for the validation of the 3D CEMBS-PolSea model.

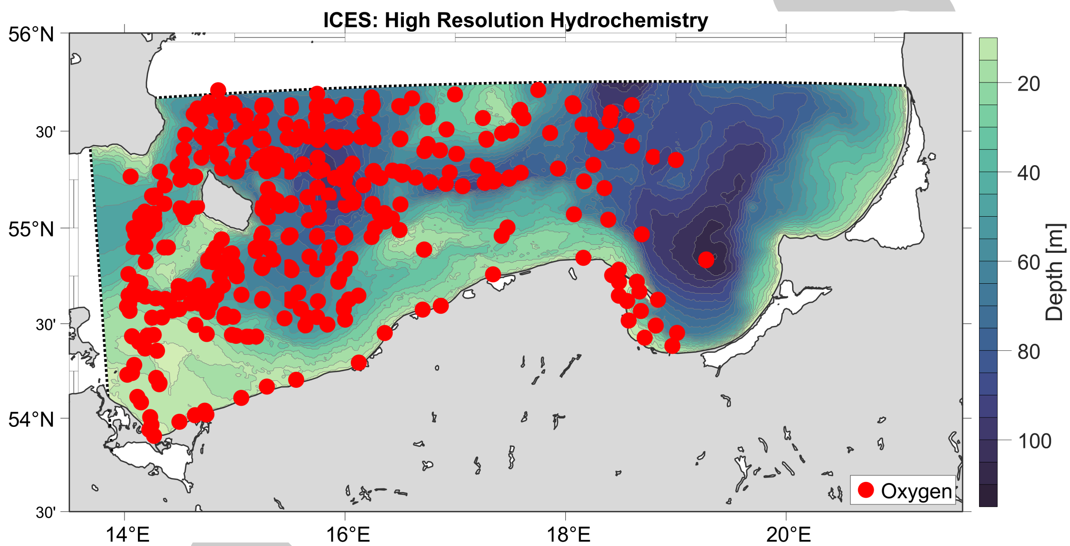

Figure 6. Spatial distribution of available data from the ICES High Resolution Hydrochemistry dataset used for the validation of the 3D CEMBS-PolSea model.

match-up procedure was applied, assigning each observational measurement a corresponding model value.

The vertical structure of the 3D CEMBS-PolSea model consists of 26 layers, each 5 m thick, within the Polish marine areas. Since in situ measurements are taken at irregular depths, they were mapped onto discrete model layers. All observations collected within the same spatiotemporal window (identical geographic coordinates and sampling day) were grouped according to depth intervals corresponding to model layers. When multiple measurements fell within the same interval, their values were averaged to produce a single representative observational value for that layer.

Geographical coordinates were then converted into model grid indices (x, y), assigning each observation to the nearest grid point. Observations outside the model domain were excluded. Model values were extracted for the corresponding (x, y, z) grid point at the time of measurement. As a result, each in situ observation was paired with a corresponding model value at the same time and in the same three-dimensional location.

This modified approach differs from the classical match-up procedure. Vertical interpolation of model results to exact measurement depths was intentionally avoided, as it introduced significant discrepancies, particularly in the surface layer. The first model layer represents an average value over the 0–5 m depth range (centered at 2.5 m), and interpolation toward the surface introduced additional uncertainty, especially since most ICES measurements are taken at or near the surface. Aggregating observations

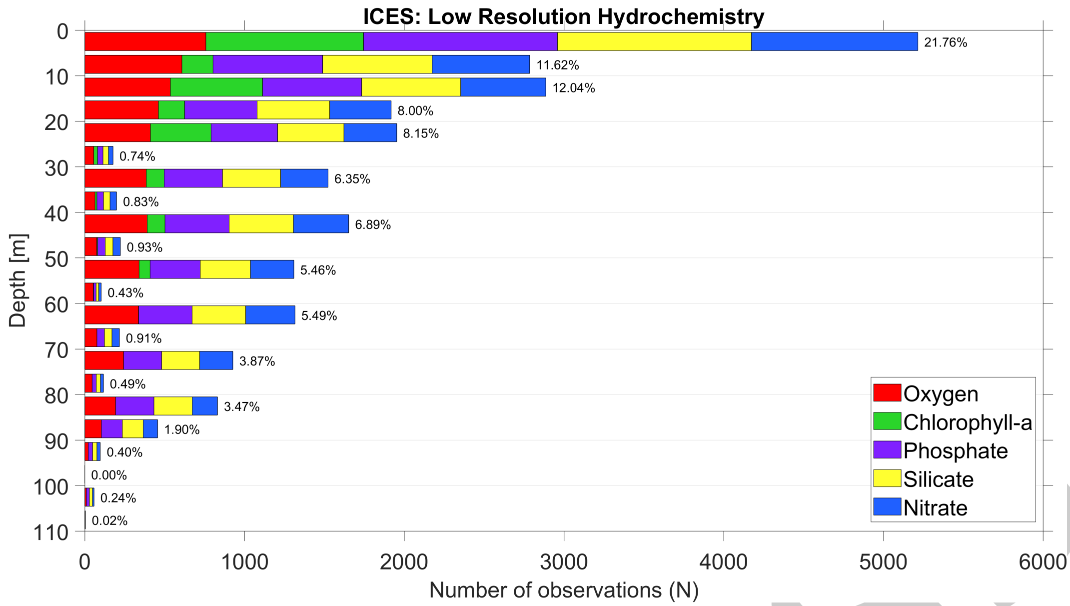

Figure 7. Vertical distribution of available data from the ICES Low Resolution Hydrochemistry dataset used for the validation of the 3D CEMBS-PolSea model.

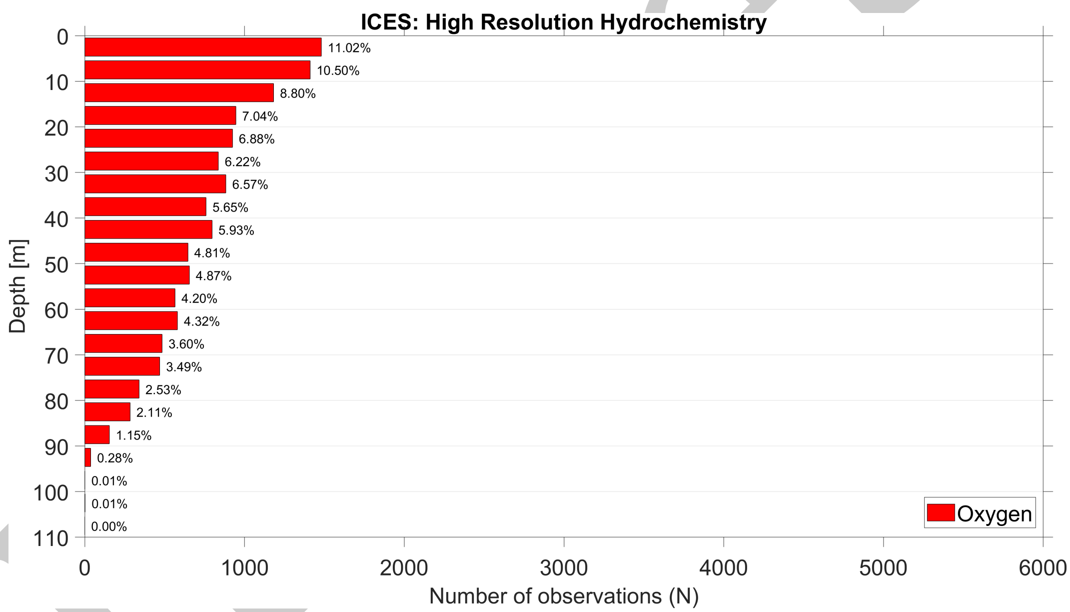

Figure 8. Vertical distribution of available data from the ICES High Resolution Hydrochemistry dataset used for the validation of the 3D CEMBS-PolSea model.

into depth intervals consistent with model layers reduces this uncertainty and enables a more robust comparison between model outputs and observations.

2.5.4 Validation metrics

To illustrate the availability of observational data used in the model validation, maps showing the spatial distribution of measurement points from both ICES databases within the 3D CEMBS-PolSea model domain were prepared (see Figures 5–6). All sampling locations used in the analysis for the years 2022–2025 are shown, covering the previously selected biogeochemical parameters. The map presents the spatial distribution of measurement points against the bathymetry of the model domain and the boundaries of the effective computational area.

The spatial distribution of the Low Resolution dataset is highly non-uniform and reflects the nature of monitoring campaigns in the Baltic Sea. The highest density of observations is found in coastal regions and along major research vessel transects, whereas offshore areas and some parts of the model domain are significantly under-sampled. Despite this spatial heterogeneity, the dataset covers a substantial portion of the study area and provides a solid basis for statistical evaluation of model performance against observations.

The number of available measurements for the analysed period and domain in the ICES Low Resolution Hydrochemistry dataset was as follows: 17691 for dissolved oxygen (O_2), 3009 for chlorophyll *a* (CHL), 7083 for phosphates (PO_4), 7081 for silicates (SiO_3), and 6174 for nitrates (NO_3). After applying the modified match-up procedure and quality-control filters, 5169 O_2 measurements, 2639 CHL measurements, 5674 PO_4 measurements, 5672 SiO_3 measurements, and 4812 NO_3 measurements were retained for analysis.

For the ICES High Resolution Hydrochemistry dataset, the number of available oxygen (O_2) measurements was 168046, which was reduced to 13437 after filtering and matching with the 3D CEMBS-PolSea model.

An important aspect of assessing the availability of observational data used in the model validation process is the analysis of their vertical distribution within the water column. Since the 3D CEMBS-PolSea model employs a discrete vertical structure with 5 m thick layers in the upper part of the water column, experimental measurements were previously assigned to corresponding model layers through aggregation within depth intervals matching the model levels. On this basis, it was possible to determine the number of observations assigned to individual vertical layers of the model.

Both ICES datasets exhibit a similar vertical profile, with a clear dominance of surface-layer observations. The analysis of the vertical distribution of data from the ICES Low Resolution Hydrochemistry dataset showed that more than 62% of all measurements originate from the upper 30 m of the water column (model layers 0–30 m). The number of observations assigned to individual model layers ranges from several thousand in the surface layers to only single observations in the deepest levels of the water column (see Figure 7).

In the case of the ICES High Resolution Hydrochemistry dataset, the vertical distribution of data is more uniform in the upper part of the water column; however, a clear dominance of surface-layer measurements is still observed. Approximately 50% of all observations are located within the upper 0–30 m model layers (see Figure 8).

Such a strong concentration of measurements in the upper part of the water column is consistent with the nature of oceanographic research in the Baltic Sea, where most observations are carried out within the euphotic zone, where the highest intensity of biological processes occurs, such as phytoplankton primary production and biogeochemical transformations. It should also be emphasised that the Low and High Resolution datasets differ in the range of available parameters. In the Low Resolution dataset, measurements of several biogeochemical variables were considered simultaneously (dissolved oxygen, chlorophyll *a*, nitrates, phosphates, and silicates); therefore, the reported numbers of observations refer to all these parameters collectively. In contrast, the High Resolution dataset contains only dissolved oxygen (O_2) data, and thus its vertical distribution refers exclusively to this variable.

3. Results

3.1 Validation Results

To assess simulation quality, modelled values were compared with corresponding observational data from the ICES databases for the period 01.01.2022–31.12.2025. The analysis included dissolved oxygen (O_2), chlorophyll *a* (CHL), phosphates (PO_4), silicates (SiO_3), and nitrates (NO_3). For each parameter, statistical metrics were calculated for paired model–observation data, including mean,

Table 1. Summary statistics from the comparison between 3D CEMBS-PolSea model results and ICES observational data (LR – Low Resolution Hydrochemistry, HR – High Resolution Hydrochemistry).

Model: 3D CEMBS-PolSea 01.01.2022–31.12.2025							
Variable	N	Mean value		STD		ICES vs Model	
		ICES	Model	ICES	Model	RMSD	r
O_2 (LR) [mmol m^{-3}]	5,169	296.74	302.34	109.85	72.61	58.82	0.87
CHL [mg m^{-3}]	2,639	3.03	1.44	3.33	1.75	2.73	0.57
PO_4 [mmol m^{-3}]	5,674	0.80	0.86	0.99	0.37	0.80	0.65
SiO_3 [mmol m^{-3}]	5,672	19.94	13.25	15.53	5.22	12.61	0.68
NO_3 [mmol m^{-3}]	4,812	3.08	6.30	9.62	9.33	8.44	0.60
O_2 (HR) [mmol m^{-3}]	13,437	286.24	302.21	133.16	69.21	95.75	0.72

Table 2. Additional skill metrics used for model evaluation of the 3D CEMBS-PolSea model. Bias represents the difference between modelled and observed mean values. nRMSE is the normalized root mean square error expressed relative to the observed mean concentration. RSR denotes the ratio of RMSE to the standard deviation of observations. CV represents the coefficient of variation of observational data.

Variable	Bias	nRMSE [%]	RSR	CV [%]
O_2 (LR) [mmol m^{-3}]	5.60	19.82	0.54	37.02
CHL [mg m^{-3}]	−1.59	90.10	0.82	109.90
PO_4 [mmol m^{-3}]	0.06	100.00	0.81	123.75
SiO_3 [mmol m^{-3}]	−6.69	63.24	0.81	77.88
NO_3 [mmol m^{-3}]	3.22	274.03	0.88	312.34
O_2 (HR) [mmol m^{-3}]	15.97	33.45	0.72	46.52

standard deviation, root mean square difference, and the Pearson correlation coefficient (Table 1). These measures evaluate both agreement in mean values and the model's ability to reproduce spatial and temporal variability. Spatial performance was assessed by calculating correlation coefficients for individual stations and mapping their distribution. Only stations with more than 20 observations (NS > 20) were included in the statistical analysis.

3.1.1 Oxygen (O_2)

Dissolved oxygen is one of the key indicators of marine ecosystem status because its distribution reflects the combined effects of biological production, respiration, air–sea gas exchange, vertical mixing, and water mass circulation.

Comparison of model simulations with ICES Low Resolution Hydrochemistry observations demonstrates very good agreement between modelled and observed oxygen concentrations. The Pearson correlation coefficient reached r = 0.87, while the mean modelled concentration (302.34 mmol m^{-3}) was close to the observed value (296.74 mmol m^{-3}), indicating only a small positive bias (+5.60 mmol m^{-3}). The corresponding RSR value of 0.54 indicates that model errors remain substantially lower than the natural variability of the observational dataset. For the ICES High Resolution Hydrochemistry dataset, the model also performed well, with a correlation coefficient of r = 0.72 and an RSR value of 0.72 (Table 2).

The spatial distribution of correlation coefficients (Figures 9–10) indicates consistently high agreement throughout most of the study area. The highest correlations are observed in offshore waters, whereas slightly lower values occur in some coastal regions and near-bottom environments characterised by stronger local variability. The model reproduces the seasonal cycle and large-scale distribution of oxygen particularly well in surface and intermediate water layers.

A tendency to overestimate oxygen concentrations is observed under low-oxygen conditions, mainly in deeper waters. This behaviour may be related to uncertainties in the representation of sediment oxygen demand, benthic remineralisation processes, and local circulation patterns affecting deep-water ventilation. Nevertheless, the high correlation coefficients and favourable RSR values confirm that the model successfully reproduces the dominant processes controlling oxygen dynamics in the southern Baltic Sea.

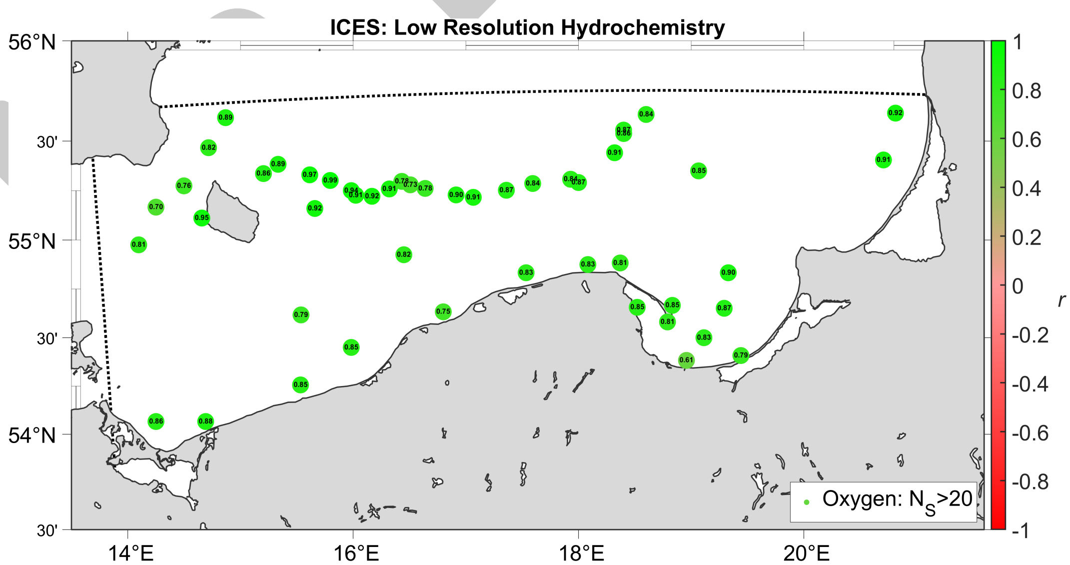

Figure 9. Spatial distribution map of the Pearson correlation coefficient (r) at individual locations for dissolved oxygen (O_2) concentration, based on the ICES Low Resolution Hydrochemistry dataset.

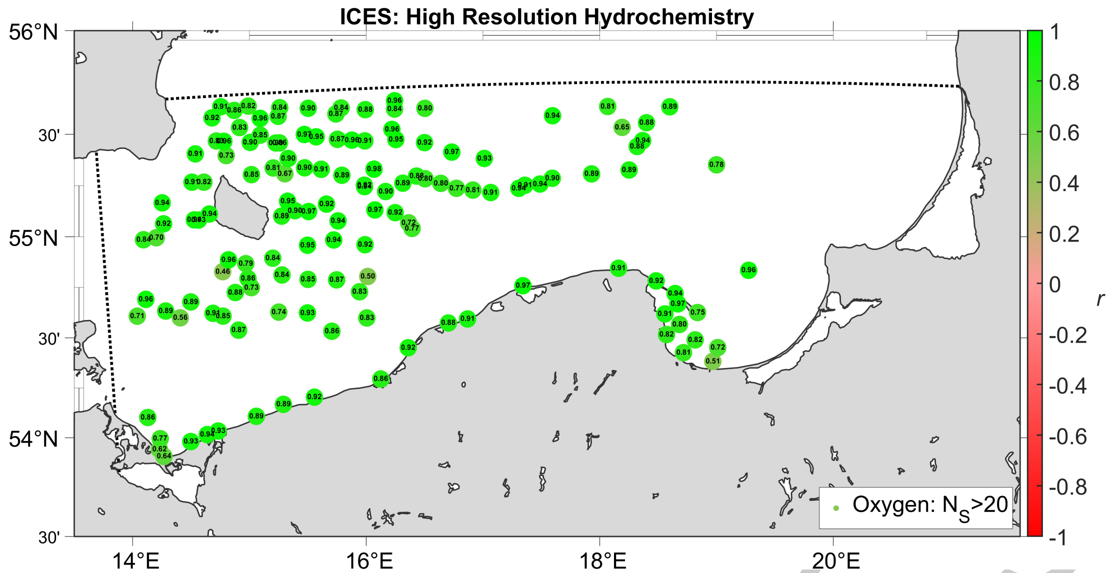

Figure 10. Spatial distribution map of the Pearson correlation coefficient (r) at individual station locations for dissolved oxygen (O_2) concentration, based on the ICES High Resolution Hydrochemistry dataset.

3.1.2 Chlorophyll *a* (CHL)

Chlorophyll *a* concentration is commonly used as an indicator of phytoplankton biomass and primary production intensity. Its variability is strongly controlled by nutrient availability, light conditions, species composition, and short-term hydrodynamic forcing.

Comparison of model simulations with ICES observations yielded a correlation coefficient of r = 0.57 (Table 1).

Although this is the lowest correlation among the analysed variables, chlorophyll *a* also exhibits exceptionally high natural variability. The coefficient of variation of the observational dataset reaches 109.9%, reflecting the highly episodic and localised nature of phytoplankton blooms. The model slightly underestimates chlorophyll *a* concentrations, with a bias of -1.59 mg m^{-3}. The RSR value of 0.82 indicates that model errors remain lower than the overall variability of the observations.

The spatial distribution of correlation coefficients (Figure 11) shows generally better agreement in offshore waters than in coastal regions. Lower correlations are observed in areas influenced by riverine inputs and rapid environmental changes, where short-lived bloom events and local ecological processes are difficult to reproduce accurately. Such spatial patterns are typical for regional biogeochemical ecosystem models.

Part of the discrepancy between model results and observations may result from differences between the model representation and the observation system itself. Satellite-derived chlorophyll *a* products represent an optically weighted signal integrated over a certain depth range rather than the concentration occurring exclusively within the uppermost model layer. During periods of strong stratification, chlorophyll maxima may occur below the surface mixed layer, contributing to differences between modelled and observed chlorophyll *a* concentrations. In addition, the high spatial and temporal variability of phytoplankton blooms, particularly in coastal waters, remains challenging for regional ecosystem models.

3.1.3 Phosphates (PO_4)

Phosphates constitute one of the primary nutrients controlling phytoplankton production and eutrophication processes in the Baltic Sea. Their distribution is influenced by biological uptake, remineralisation, sediment-water interactions, and riverine nutrient inputs.

The comparison between model results and ICES observations indicates good agreement, with a Pearson correlation coefficient of r = 0.65 (Table 1). Mean modelled phosphate concentrations closely match observations, resulting in a very small positive bias (+0.06 mmol m^{-3}). Although the RMSD reaches 0.80 mmol m^{-3}, the observational dataset itself exhibits substantial variability (CV = 123.8%). The RSR value of 0.81 indicates that model errors remain lower than the standard deviation of the observations.

The spatial distribution of correlation coefficients Figure 12) reveals generally higher agreement in offshore regions than in coastal waters. Lower correlations occur in the Gulf of Gdańsk, Puck Bay, the Pomeranian Bay, and near river mouths. These areas are strongly influenced by freshwater inflows and nutrient inputs, which introduce significant short-term variability and local gradients that are difficult to reproduce precisely.

Overall, the results indicate that the model captures the principal mechanisms controlling phosphate transport and cycling in the southern Baltic Sea, while remaining discrepancies are largely associated with the complexity

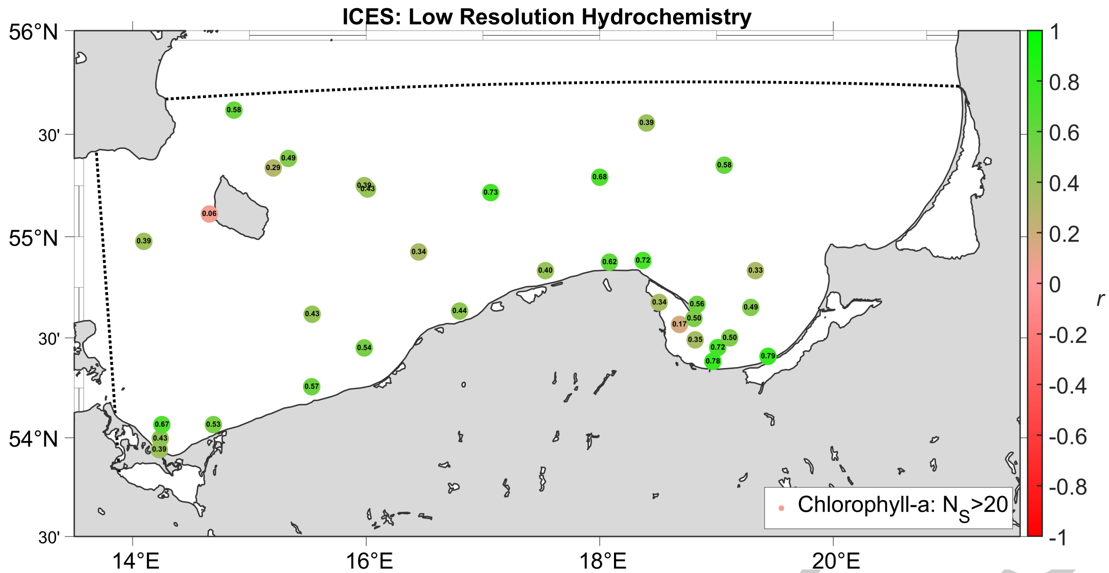

Figure 11. Spatial distribution map of the Pearson correlation coefficient (r) at individual locations for chlorophyll *a* (CHL) concentration, based on the ICES Low Resolution Hydrochemistry dataset.

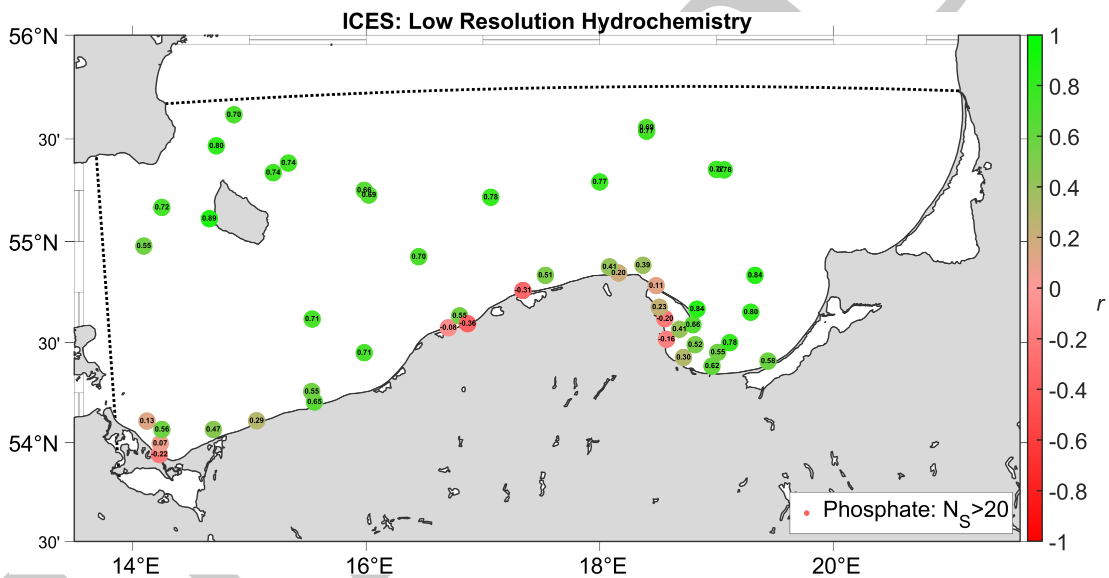

Figure 12. Spatial distribution map of the Pearson correlation coefficient (r) at individual locations for phosphate (PO_4) concentration, based on the ICES Low Resolution Hydrochemistry dataset.

of coastal processes and uncertainties in riverine nutrient load estimates.

3.1.4 Silicates (SiO_3)

Silicates play a fundamental role in the Baltic Sea ecosystem because they constitute an essential nutrient for diatoms, one of the dominant phytoplankton groups. Consequently, their availability directly affects primary production and the structure of phytoplankton communities.

Comparison of model simulations with ICES observations indicates good agreement, with a Pearson correlation coefficient of r = 0.68 (Table 1). The mean modelled concentration (13.25 mmol m^{-3}) is lower than the observed value (19.94 mmol m^{-3}), resulting in a negative bias of −6.69 mmol m^{-3}. The RMSD reaches 12.61 mmol m^{-3}; however, interpretation of this value should consider the substantial variability of the observational dataset (CV = 77.9%). The calculated RSR value of 0.81 indicates that model errors remain lower than the standard deviation of the observations.

Modelled and observed concentrations show good agreement for silicate concentrations up to approximately 25 mmol m^{-3}. For higher concentrations, the model tends to underestimate observed values. This behaviour may

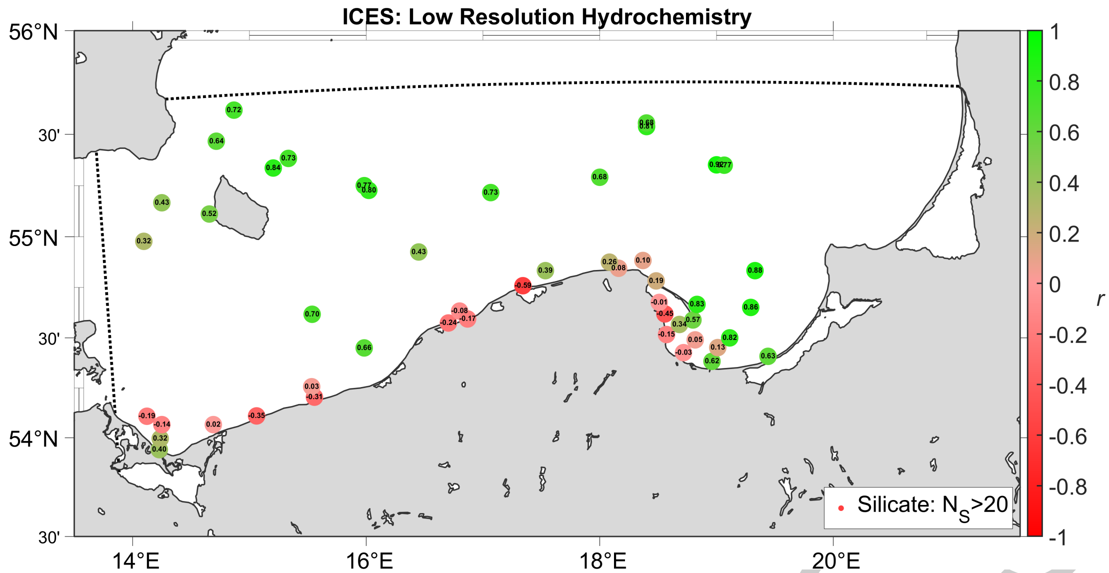

Figure 13. Spatial distribution map of the Pearson correlation coefficient (r) at individual locations for silicate (SiO_3) concentration, based on the ICES Low Resolution Hydrochemistry dataset.

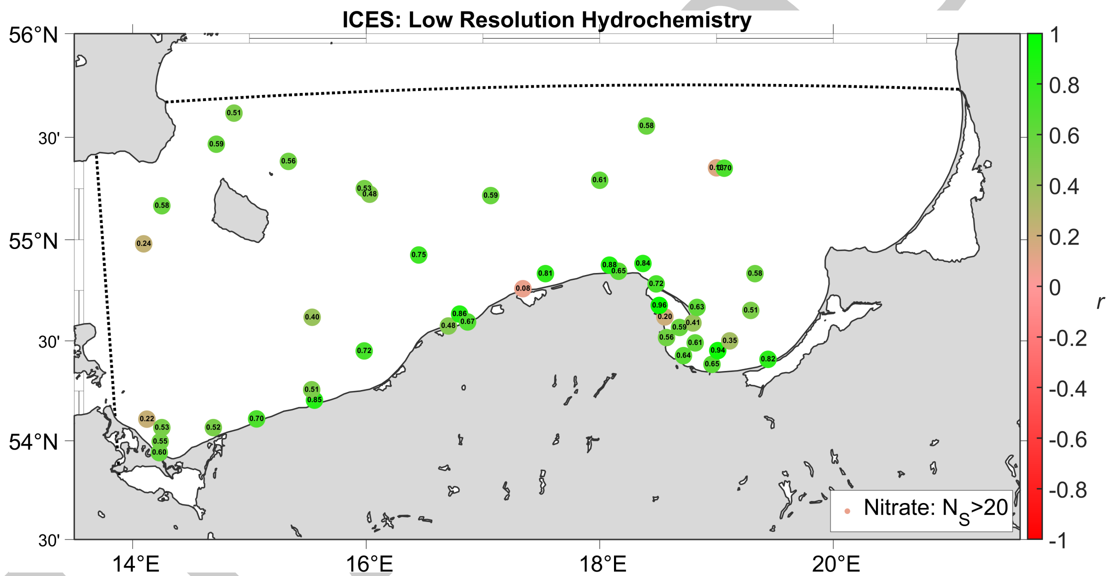

Figure 14. Spatial distribution map of the Pearson correlation coefficient (r) at individual locations for nitrate (NO_3) concentration, based on the ICES Low Resolution Hydrochemistry dataset.

be associated with uncertainties in riverine silicon inputs, local sediment-water interactions, or resuspension processes that are particularly important in shallow coastal environments.

The spatial distribution of correlation coefficients (Figure 13) reveals a clear contrast between offshore and coastal regions. Higher correlations are observed throughout most open-sea areas, whereas lower values occur in coastal waters influenced by river discharge and local hydrodynamic processes. The spatial pattern suggests that the model adequately reproduces large-scale silicate variability while local discrepancies are primarily associated with the complexity of coastal nutrient dynamics.

Overall, the obtained results indicate that the model successfully captures the dominant mechanisms controlling silicate distribution in the southern Baltic Sea and provides a realistic representation of silicon cycling at the regional scale.

3.1.5 Nitrates (NO_3)

Nitrates constitute one of the principal forms of bioavailable nitrogen and play a key role in regulating primary production and ecosystem functioning in the Baltic Sea. Their concentrations are controlled by riverine inputs, biological uptake, remineralisation processes, and physical transport.

For nitrates, the model achieved a Pearson correlation coefficient of r = 0.60, indicating moderate agreement with observations (Table 1). The mean modelled concentration (6.30 mmol m^{-3}) exceeds the observed value (3.08 mmol m^{-3}), resulting in a positive bias of +3.22 mmol m^{-3}. Although the RMSD reaches 8.44 mmol m^{-3}, the observational dataset is characterised by exceptionally high variability, with a coefficient of variation of 312.3%. The corresponding RSR value of 0.88 demonstrates that model errors remain lower than the overall variability of the observations.

The spatial distribution of correlation coefficients (Figure 14) indicates relatively homogeneous model performance throughout the study area. In contrast to phosphates and silicates, no extensive regions of systematically low correlation are observed. Most monitoring stations exhibit moderate to good agreement between modelled and observed nitrate concentrations, suggesting that the model successfully reproduces the dominant processes controlling nitrate transport and transformation.

The relatively high variability of nitrate concentrations reflects strong seasonal cycles associated with winter accumulation, spring phytoplankton uptake, and local riverine inputs. These processes generate large concentration gradients that are difficult to reproduce precisely in regional ecosystem models. Nevertheless, the spatial distribution of correlation coefficients confirms that the model captures the principal temporal and spatial patterns of nitrate variability across the southern Baltic Sea.

Taken together, the results indicate that the 3D CEMBS-PolSea model provides a realistic representation of nitrate dynamics and successfully reproduces the major mechanisms governing nitrogen cycling in the analysed region.

3.1.6 Validation summary

The validation results demonstrate that the 3D CEMBS-PolSea model successfully reproduces the principal spatial and temporal patterns of variability of the analysed biogeochemical parameters in the southern Baltic Sea. The assessment was based on comparisons with ICES observational datasets and included both statistical indicators (Table 1) and the spatial distribution of correlation coefficients (Figures 9–14). To support the interpretation of RMSD values, additional normalised statistics were calculated, including bias, normalised RMSE (nRMSE), the RMSE-observation standard deviation ratio (RSR), and the coefficient of variation (CV). Additional normalised performance metrics are presented in Table 2.

The highest model performance was obtained for dissolved oxygen. Pearson correlation coefficients reached 0.87 for the ICES Low Resolution dataset and 0.72 for the ICES High Resolution dataset (Table 1). The corresponding RSR values were 0.54 and 0.72, indicating that model errors remained substantially lower than the natural variability of the observations (Table 2). Spatial correlation maps (Figures 9–10) show consistently high agreement across most of the model domain, confirming that the model adequately reproduces the physical and biogeochemical processes controlling oxygen dynamics in the southern Baltic Sea.

For nutrients, correlation coefficients ranged between 0.60 and 0.68 (Table 1). Phosphates and silicates exhibited the highest agreement among the analysed nutrients (r = 0.65 and r = 0.68, respectively). The spatial distribution of correlation coefficients (Figures 12–13) reveals a clear contrast between offshore and coastal regions. Higher correlations were generally observed in open-sea areas, whereas lower correlations occurred in coastal waters affected by riverine inflows, particularly in the Gulf of Gdańsk, Puck Bay, the Pomeranian Bay, and near the mouths of the Łeba and Słupia rivers. These discrepancies most likely reflect uncertainties in riverine nutrient loads and the strong short-term variability characteristic of estuarine and coastal environments.

For nitrate concentrations, the model achieved a correlation coefficient of r = 0.60 (Table 1) despite the exceptionally high variability of the observational dataset (CV = 312.3%) (Table 2). The spatial distribution of correlations (Figure 14) indicates relatively homogeneous model performance across the study area, without extensive regions of systematically poor agreement. This suggests that the model captures the dominant mechanisms governing nitrate transport and transformation despite difficulties associated with reproducing extreme concentration values.

The lowest correlation was obtained for chlorophyll *a* (r = 0.57) (Table 1). However, chlorophyll *a* was also characterised by very high observational variability (CV = 109.9%) (Table 2), reflecting the episodic and highly localised nature of phytoplankton blooms. The spatial distribution of correlation coefficients (Figure 11) indicates generally better model performance in offshore waters than in coastal regions. Lower correlations observed in nearshore areas are likely related to rapidly changing environmental conditions, local nutrient inputs, and short-term bloom events that remain difficult to resolve even in high-resolution ecosystem models.

The interpretation of RMSD values should be considered in relation to the natural variability of the analysed parameters. Additional normalised skill metrics indicate that the coefficients of variation of the observational datasets reached 109.9% for chlorophyll *a*, 123.8% for phosphate, and 312.3% for nitrate concentrations (Table 2). Consequently, relatively large RMSD values (Table 1) partly reflect the broad range of environmental conditions occurring within the study area rather than solely model deficiencies. The calculated RSR values range from 0.54 to 0.88 (Table 2), indicating that model errors remain lower than the standard deviation of observations for all anal-

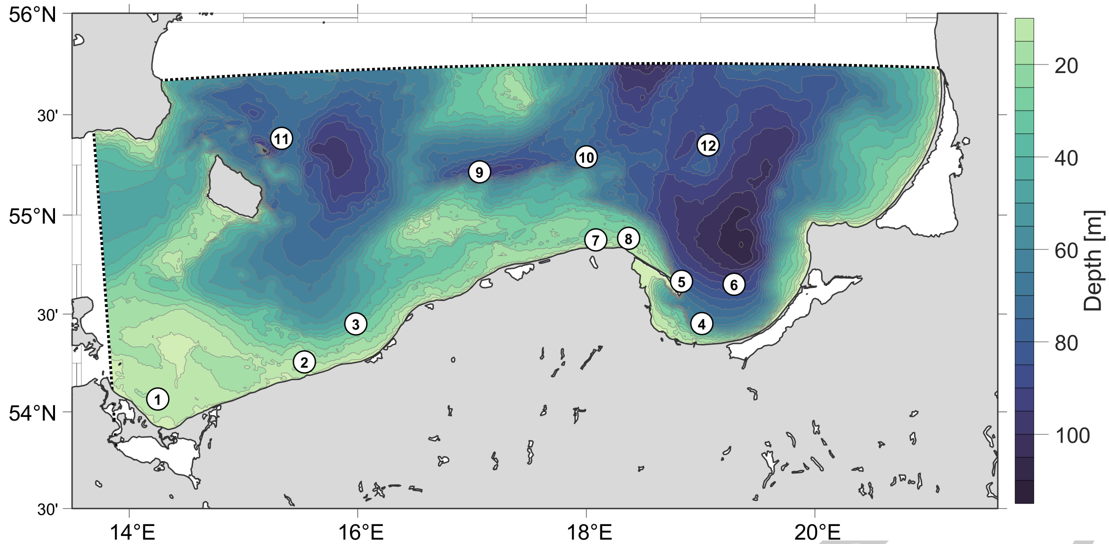

Figure 15. Locations of stations selected for comparison of 3D CEMBS-PolSea model results with satellite data for chlorophyll *a* concentration.

ysed variables.

Overall, the validation results indicate that the 3D CEMBS-PolSea model successfully reproduces the major spatial and temporal patterns of variability of dissolved oxygen, chlorophyll *a*, phosphates, silicates, and nitrates in the southern Baltic Sea. Although larger discrepancies occur in some coastal regions and for highly variable biological parameters, the obtained performance is comparable with that reported for regional biogeochemical ecosystem models operating in the Baltic Sea and other coastal shelf-sea environments. The model therefore provides a suitable tool for ecosystem analysis, environmental assessment, and operational forecasting applications.

3.2 Method of comparison with satellite data

To assess the agreement between model results and satellite observations, a comparative analysis was conducted for the period 01.01.2022–31.12.2024. For this purpose, 12 control points were selected within the model domain in a manner ensuring representation of both coastal areas and more open-sea regions of the southern Baltic Sea (see Figure 15).

The 12 control points selected for the analysis were distributed across the entire model domain in order to represent different types of hydrological environments in the southern Baltic Sea. Some stations are located in coastal regions, particularly near the Pomeranian coast and in areas influenced by riverine inputs, where chlorophyll *a* concentrations exhibit high spatial and temporal variability. The remaining points are located in more open-sea regions, where environmental conditions are more homogeneous and variability of biogeochemical parameters is generally lower.

Such a distribution of points allows for the assessment of model performance both in dynamic coastal environments and in more stable open-sea conditions. In particular, nearshore stations enable the analysis of local processes such as riverine nutrient inputs and enhanced hydrodynamic variability, while offshore stations allow evaluation of the model's ability to reproduce chlorophyll *a* variability at a regional scale.

At the selected points, total chlorophyll *a* concentrations from satellite data were compared with modelled values computed as the sum of chlorophyll associated with three phytoplankton groups. It should be emphasised that, unlike the model, which provides daily output, satellite data are not available for all days of the analysed period. This limitation is mainly due to cloud cover, which prevents ocean colour observations in the visible spectrum, as well as other factors affecting data quality. As a result, the number of available satellite scenes at the analysed points over the three-year period ranges from 129 to 213 days, depending on the location.

To provide a more comprehensive assessment of model performance, two sets of statistical indicators were calculated. Table 3 presents the basic descriptive and comparative statistics, including the number of observations (N), mean values, standard deviations (STD), root mean square deviation (RMSD), and Pearson correlation coefficients (r). Table 4 complements this analysis with additional performance metrics, including bias, normalised root mean square error (nRMSE), the RMSE-observation standard deviation ratio (RSR), and coefficients of variation (CV) for both satellite observations and model results. These indicators allow model performance to be evaluated in relation to the natural variability of chlorophyll *a* concentrations and provide a more robust interpretation of the agreement between simulations and observations.

Table 3. Statistical comparison of surface chlorophyll *a* (CHL) concentrations from the 3D CEMBS-PolSea model with available satellite observations.

Station number	N	Mean value [mg m^{-3}]		STD [mg m^{-3}]		Satellite vs model	
		Satellite	Model	Satellite	Model	RMSD [mg m^{-3}]	r
1	213	2.18	2.63	1.74	1.61	1.61	0.54
2	165	1.28	1.75	1.12	1.38	1.14	0.60
3	129	1.14	1.43	1.01	0.85	1.02	0.41
4	185	2.43	2.60	3.39	1.71	2.64	0.64
5	201	1.71	2.05	1.17	1.01	1.14	0.46
6	189	1.72	1.94	1.34	1.02	1.35	0.37
7	172	1.44	1.54	1.37	0.97	1.14	0.57
8	170	1.52	1.81	1.33	1.15	1.16	0.57
9	143	1.31	1.57	0.93	0.79	1.01	0.32
10	137	1.66	1.85	1.25	0.87	1.30	0.29
11	146	1.24	1.50	0.89	0.80	1.01	0.29
12	132	1.74	1.85	1.34	0.97	1.26	0.45
All stations 2022–2024	1981	1.63	1.92	1.35	1.20	1.27	0.51

Table 4. Additional statistical performance metrics for the comparison of surface chlorophyll *a* concentrations from the 3D CEMBS-PolSea model and satellite observations, including bias, normalized RMSE (nRMSE), RMSE-observation standard deviation ratio (RSR), and coefficients of variation (CV).

Station number	Bias	CVsat [%]	CVmod [%]	nRMSE [%]	RSR
All stations 2022–2024	0.29	82.8	62.5	77.9	0.94

Comparative analysis of all measurement points (1–12) for the years 2022–2024, based on 1981 observations, showed that the mean chlorophyll *a* concentration in satellite data was 1.63 mg m^{-3}, while the mean simulated by the 3D CEMBS-PolSea model was 1.92 mg m^{-3}. The resulting bias of 0.29 mg m^{-3} indicates a moderate tendency of the model to overestimate chlorophyll *a* concentrations relative to satellite observations.

The coefficient of variation (CV) reached 82.8% for the satellite observations and 62.5% for the model results, indicating a high level of natural spatial and temporal variability in chlorophyll *a* concentrations across the southern Baltic Sea. The lower CV obtained for the model suggests that the simulated chlorophyll *a* field is smoother and exhibits slightly reduced variability compared to the observations.

The root mean square deviation (RMSD) between satellite and model data was 1.27 mg m^{-3}. When normalised by the mean observed concentration, the nRMSE reached 77.9%. However, the RSR value was 0.94, indicating that the model error remains lower than the standard deviation of the observations. This suggests that the model retains predictive skill and reproduces a substantial part of the observed variability despite the large natural variability of chlorophyll *a* concentrations.

The Pearson correlation coefficient (r) reached 0.51, indicating a moderate positive relationship between model results and satellite observations. Correlation coefficients at individual stations ranged from 0.29 to 0.64, demonstrating that the model captures the general temporal variability of chlorophyll *a*, although its performance differs among locations.

Overall, the results indicate that the 3D CEMBS-PolSea model is capable of reproducing the main spatial and temporal patterns of surface chlorophyll *a* variability in the southern Baltic Sea. The remaining discrepancies are likely related to the high natural variability of chlorophyll *a* concentrations, limitations of satellite-derived products, and differences between model representation and observation sampling.

3.2.1 Interpretation of results

The statistical comparison between the 3D CEMBS-PolSea model and satellite-derived chlorophyll *a* concentrations indicates a moderate agreement between the two datasets. The model exhibits a positive bias of 0.29 mg m^{-3}, suggesting a tendency to slightly overestimate chlorophyll *a* concentrations relative to satellite observations. The coefficient of variation of the satellite observations reaches 82.8%, confirming the high natural temporal and spatial variability of chlorophyll *a* in the southern Baltic Sea. The corresponding value for the model is lower (62.5%), indicating that the simulated chlorophyll *a* field is smoother and exhibits reduced variability compared to the observations.

Although the normalised root mean square error (nRMSE) reaches 77.9%, the RMSE-observation standard deviation ratio (RSR) equals 0.94, indicating that the model error remains lower than the standard deviation of the observations. This suggests that the model retains predictive skill and reproduces a substantial fraction of the observed

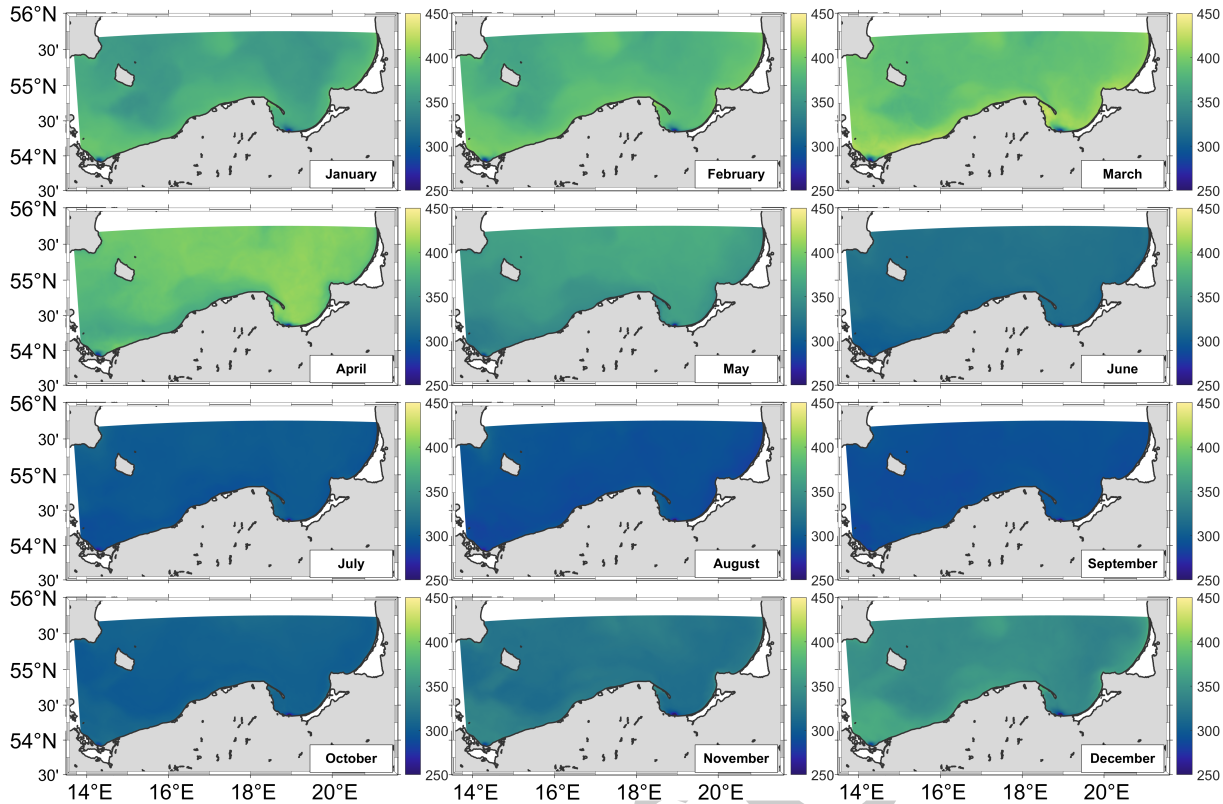

Figure 16. Mean monthly concentrations of dissolved oxygen [mmol m^{-3}] at the surface derived from the 3D CEMBS-PolSea model simulations for the years 2022–2025.

variability despite the complexity of the ecosystem processes and the limitations of satellite observations. The Pearson correlation coefficient (r = 0.51) further demonstrates that the model captures the general temporal variability of chlorophyll *a* concentrations, although differences remain in the magnitude and timing of individual bloom events.

The analysis of chlorophyll *a* time series at selected locations indicates that the model generally tends to converge toward values derived from satellite observations. The results demonstrate a progressive adjustment of the simulated chlorophyll *a* field toward the observed patterns, particularly during periods of enhanced phytoplankton bloom activity. Overall, the model reproduces the seasonal variability of chlorophyll *a* concentrations reasonably well, although discrepancies between simulated and observed values remain at individual time steps.

It should be emphasised that the data assimilation approach implemented in the 3D CEMBS-PolSea model does not involve direct fitting of model outputs to observations. Instead, observational information is incorporated gradually through small corrections distributed over multiple time steps. This approach preserves the internal consistency of the model equations and maintains the balance between physical and biogeochemical processes. Consequently, the model is not expected to reproduce satellite-derived chlorophyll *a* values exactly at every time step, but rather to progressively adjust its state toward the observed conditions.

Additional uncertainty is associated with the comparison between model simulations and satellite-derived chlorophyll *a* observations. As discussed in Section 2.4.2, satellite products represent an optically weighted signal originating from the upper illuminated part of the water column rather than chlorophyll concentrations within a single model layer. Consequently, part of the discrepancy between model results and observations may arise from differences in the vertical representation of chlorophyll *a* in the observational and modelling systems. This effect is likely to be most pronounced during stratified periods, when subsurface chlorophyll maxima may contribute to the satellite signal while remaining only partially represented in the near-surface model layer. Furthermore, satellite-derived chlorophyll *a* products are affected by cloud cover, uncertainties in atmospheric correction, water turbidity, and suspended particulate matter, all of which contribute additional uncertainty to the comparison.

3.3 Mean monthly values of modelled biogeochemical parameters

This chapter presents mean monthly concentrations of the analysed biogeochemical parameters, i.e. dissolved oxy-

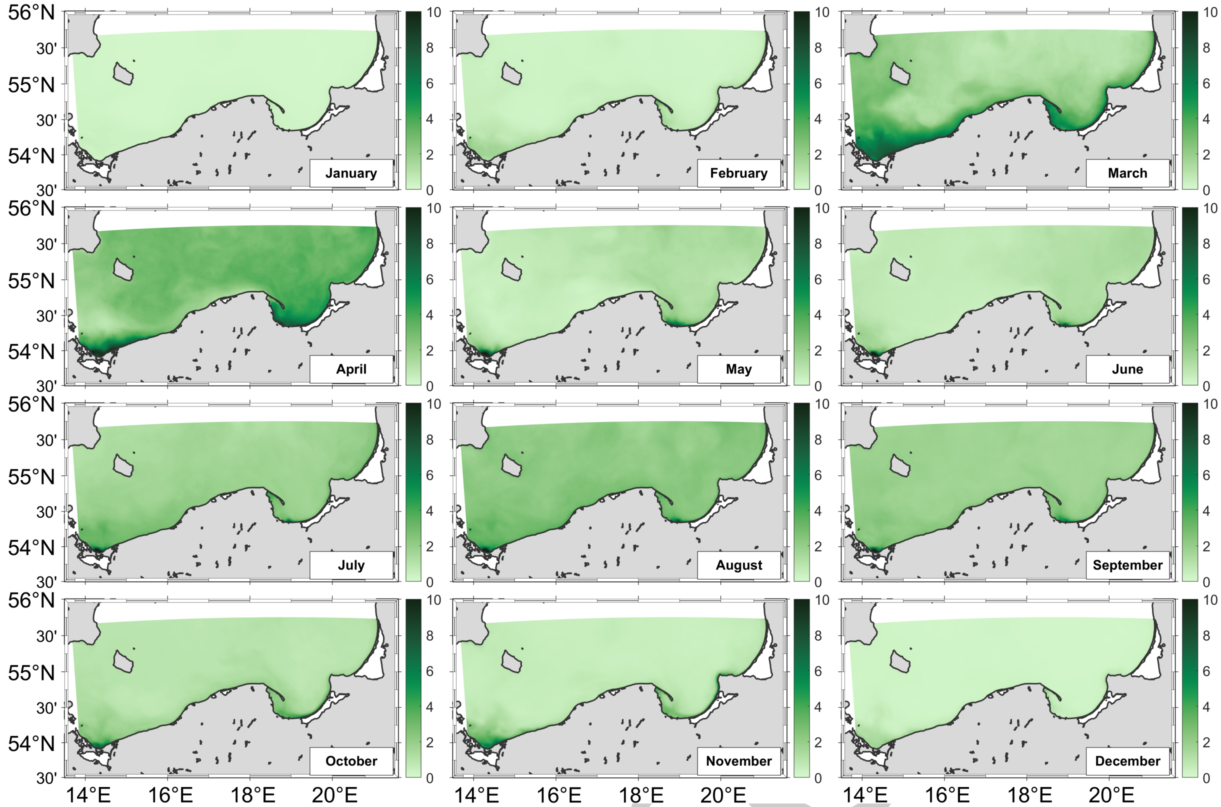

Figure 17. Mean monthly chlorophyll *a* concentrations [mg m^{-3}] at the surface derived from the 3D CEMBS-PolSea model simulations for the years 2022–2025.

gen (O_2), chlorophyll *a* (CHL), phosphates (PO_4), silicates (SiO_3), and nitrates (NO_3). The data reflect seasonal variability in the surface layer of the 3D CEMBS-PolSea model over a four-year simulation period from 1 January 2022 to 31 December 2025.

3.3.1 Oxygen (O_2)

The seasonal variability of dissolved oxygen concentration in the surface layer is controlled by both physical and biological factors. The most important include water temperature, gas exchange with the atmosphere, and primary production processes associated with phytoplankton activity. During the analysed period (2022–2025), the highest mean monthly oxygen concentrations in the surface layer occur in the winter–spring period. This is associated with lower water temperatures, which increase oxygen solubility, as well as the onset of the phytoplankton growth season (see Figure 16).

The highest mean oxygen concentrations in the modelled domain occur in the spring months. In March, the mean value reaches 392.79 mmol m^{-3}, while in April it is 392.74 mmol m^{-3}. During this period, maximum spatial values reach approximately 429.91 mmol m^{-3} and 411.50 mmol m^{-3}, respectively.

In the following months, as water temperature increases during the summer period, a gradual decrease in mean oxygen concentration in the surface layer is observed. The lowest values occur in late summer and early autumn. The minimum mean monthly value was recorded in September and amounted to approximately 290.86 mmol m^{-3}, with minimum spatial values reaching about 214.49 mmol m^{-3}.

During the autumn and winter months, a renewed increase in oxygen concentration in the surface layer is observed, associated with decreasing water temperature and intensified vertical mixing of the water column. This enhanced mixing is driven by surface cooling and increased wind stress. In December, the mean oxygen concentration reaches 346.67 mmol m^{-3}, while maximum values in the model domain reach approximately 383.17 mmol m^{-3}.

The obtained seasonal oxygen pattern is consistent with the general processes observed in the Baltic Sea and reflects the influence of water temperature and the seasonal dynamics of primary production on the variability of oxygen concentrations in the surface layer.

Detailed monthly data are presented in Table 5.

3.3.2 Chlorophyll *a* (CHL)

Chlorophyll *a* is widely used as an indicator of phytoplankton biomass and the intensity of primary production in marine environments. In the 3D CEMBS-PolSea model, phytoplankton is represented by three functional groups:

Table 5. Summary of mean monthly dissolved oxygen (O_2) concentrations in the surface layer. Model data from the 3D CEMBS-PolSea for the years 2022–2025.

[mmol m^{-3}]	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
MAX	404.91	412.66	429.91	411.50	374.25	328.34	319.04	321.13	305.55	322.31	346.78	383.17
MEAN	366.18	379.60	392.79	392.74	357.40	316.88	298.78	291.91	290.86	304.93	322.95	346.67
MIN	252.34	251.30	240.12	234.37	237.14	237.99	231.43	223.84	214.49	221.44	216.48	211.55

Table 6. Summary of mean monthly chlorophyll *a* (CHL) concentrations in the surface layer. Model data from the 3D CEMBS-PolSea for the years 2022–2025.

[mmol m^{-3}]	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
MAX	1.43	3.18	10.23	14.55	13.49	17.31	17.33	16.18	14.30	10.86	8.34	3.20
MEAN	0.23	0.45	2.18	3.10	0.98	0.94	1.84	2.42	1.86	1.15	0.76	0.32
MIN	0.15	0.21	0.68	0.80	0.26	0.33	1.01	1.60	1.45	0.72	0.38	0.19

small phytoplankton (nano- and picoplankton), large phytoplankton mainly represented by diatoms, and nitrogen-fixing cyanobacteria. The growth of each group is constrained by nutrient availability (nitrogen, phosphorus, and silicon – only for diatoms), temperature, and light conditions. The seasonal variability of chlorophyll *a* concentration in the model surface layer reflects the typical cycle of primary production in the Baltic Sea (see Figure 17). The lowest mean values occur during winter months, when solar radiation is weak and light conditions limit phytoplankton growth.

In the analysed period, the lowest mean monthly chlorophyll *a* concentration was observed in January, with an average domain value of 0.23 mg m^{-3} and minimum spatial values of approximately 0.15 mg m^{-3}. With the onset of spring, a rapid increase in chlorophyll *a* concentration is observed, associated with the spring phytoplankton bloom dominated by diatoms. The highest mean values occur in April, when chlorophyll *a* reaches 3.10 mg m^{-3} on average, while maximum domain values reach approximately 14.55 mg m^{-3}. High concentrations are also observed in March, with a mean value of about 2.18 mg m^{-3}.

After the spring bloom, chlorophyll *a* concentrations decrease due to nutrient depletion in the euphotic zone. During summer, a secondary increase in chlorophyll *a* is observed, driven by the development of cyanobacteria, which can grow under low nitrate conditions due to their ability to fix atmospheric nitrogen. In the analysed period, the highest maximum spatial values in summer months exceed 17 mg m^{-3}, indicating localised cyanobacterial blooms in combination with nitrate limitation. In the following autumn months, chlorophyll *a* concentrations gradually decrease as temperature declines and light availability diminishes. In December, the mean domain value is 0.32 mg m^{-3}, marking a return to winter conditions.

The spatial distribution of chlorophyll *a* also indicates variability within the model domain. The highest concentrations typically occur in coastal zones, where the availability of biogenic substances is enhanced due to riverine inputs and intensified mixing processes. In the more open sea areas, chlorophyll *a* concentrations are generally lower and exhibit reduced spatial variability. The obtained seasonal pattern of chlorophyll *a* variability is consistent with the typical primary production cycle observed in the Baltic Sea, including the spring diatom bloom and summer cyanobacterial development.

Detailed monthly data are presented in Table 6.

3.3.3 Phosphates (PO_4)

Phosphates are one of the fundamental nutrients regulating primary production in the Baltic Sea. Unlike nitrogen, which under certain conditions can be supplemented through atmospheric nitrogen fixation by cyanobacteria, phosphorus remains a limiting element for the growth of all phytoplankton groups. As a result, its concentration in the euphotic layer is strongly controlled by primary production intensity and the seasonal dynamics of the phytoplankton biological cycle.

The analysis of mean monthly phosphate values (see Figure 18) in the surface layer of the 3D CEMBS-PolSea model for the period 2022–2025 shows a distinct seasonal variability of this parameter. The highest mean values occur in winter months, when biological activity is limited and vertical mixing processes transport nutrients from deeper layers into the surface waters.

During the analysed period, the highest mean monthly phosphate concentration was recorded in February, with a domain-averaged value of 0.92 mmol m^{-3} and maximum spatial values reaching approximately 1.43 mmol m^{-3}. With the onset of spring and the intensification of phytoplankton growth, a clear decline in phosphate concentrations in the surface layer is observed. This process begins in March and reaches its minimum during summer. The lowest mean phosphate values occur in August, when the domain-averaged concentration decreases to approximately 0.45 mmol m^{-3}, with minimum spatial values around 0.12 mmol m^{-3}. Low phosphate concentrations during this period are directly associated with its intensive

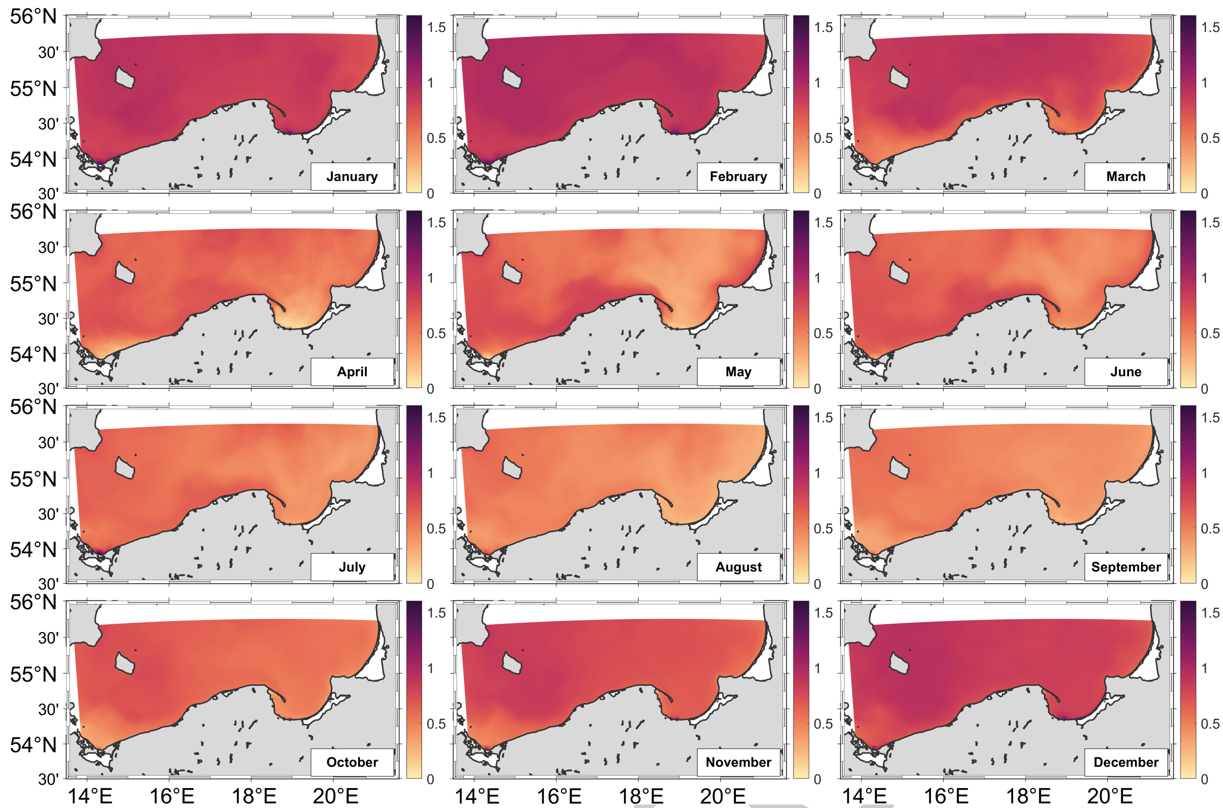

Figure 18. Mean monthly phosphate concentrations [mmol m^{-3}] at the surface derived from the 3D CEMBS-PolSea model simulations for the years 2022–2025.

uptake by phytoplankton (mainly cyanobacteria) during bloom events.

In the autumn months, phosphate concentrations in the surface layer increase again. This is linked to decreasing water temperature and intensified vertical mixing processes, which transport nutrients from deeper layers of the water column into the surface layer. In November, the mean phosphate concentration increases to approximately 0.74 mmol m^{-3}, reaching 0.84 mmol m^{-3} in December. The seasonal amplitude of mean phosphate values during the analysed period is approximately 0.47 mmol m^{-3}, reflecting the strong influence of primary production dynamics on the phosphorus balance in the euphotic zone.

The spatial distribution analysis also indicates considerable variability of phosphates within the model domain. In some months, local maximum values exceed 1.7 mmol m^{-3}, indicating areas of elevated concentrations of this nutrient. Higher phosphate levels are typically observed in coastal regions and in areas influenced by riverine inputs, where the transport of biogenic substances from land constitutes an important source of phosphorus.

The obtained seasonal pattern of phosphate variability is consistent with general processes observed in the Baltic Sea. The decline in concentrations during the spring–summer period is associated with intense primary production and phytoplankton uptake of phosphorus, whereas the increase during the autumn–winter period results from vertical mixing of the water column and regeneration of biogenic substances through organic matter mineralisation.

Detailed monthly data are presented in Table 7.

3.3.4 Silicates (SiO_3)

Silicates are one of the key nutrients in the Baltic Sea ecosystem, as they are an essential element for the construction of silica shells of diatoms. Unlike many other phytoplankton groups, whose growth is mainly limited by nitrogen and phosphorus availability, diatom development also depends on the availability of dissolved silica in seawater. As a result, the seasonal variability of silicate concentration in the euphotic zone is strongly linked to the primary production cycle and the intensity of the spring diatom bloom.

The analysis of mean monthly silicate values in the surface layer of the 3D CEMBS-PolSea model for the period 2022–2025 shows a clear seasonal variability of this parameter (see Figure 19). The highest mean values occur in winter and early spring, when biological activity is still limited and vertical mixing of the water column supplies nutrients accumulated in deeper layers to the surface. During the analysed period, the highest mean silicate concentrations are observed in February and December, when

Table 7. Summary of mean monthly phosphate (PO_4) concentrations in the surface layer. Model data from the 3D CEMBS-PolSea for the years 2022–2025.

[mmol m^{-3}]	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
MAX	1.45	1.43	0.95	0.83	0.96	1.05	1.71	0.95	0.76	0.83	1.12	1.22
MEAN	0.86	0.92	0.81	0.60	0.57	0.59	0.54	0.45	0.47	0.61	0.74	0.84
MIN	0.59	0.66	0.42	0.08	0.12	0.26	0.30	0.12	0.09	0.18	0.26	0.39

Table 8. Summary of mean monthly silicate (SiO_3) concentrations in the surface layer. Model data from the 3D CEMBS-PolSea for the years 2022–2025.

[mmol m^{-3}]	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
MAX	32.55	34.13	29.12	27.34	24.95	25.67	23.62	25.03	24.14	24.81	25.55	24.87
MEAN	12.14	12.45	11.06	8.23	8.17	8.85	9.18	9.32	9.32	10.29	11.33	12.15
MIN	8.84	9.51	7.83	3.70	3.16	4.10	5.57	6.73	7.15	7.87	8.25	8.84

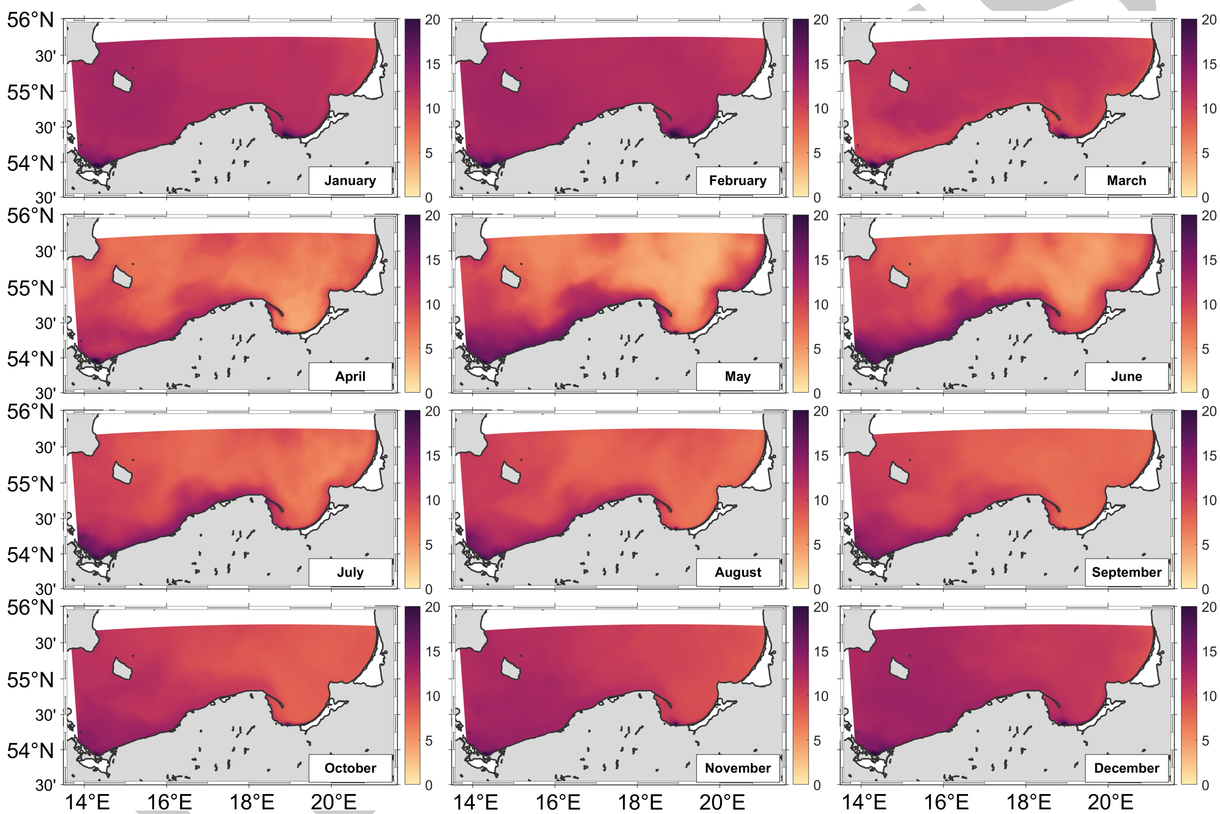

Figure 19. Mean monthly silicate concentrations [mmol m^{-3}] at the surface derived from the 3D CEMBS-PolSea model simulations for the years 2022–2025.

the domain-averaged values reach approximately 12.45 mmol m^{-3} and 12.15 mmol m^{-3}, respectively.

With the onset of the phytoplankton growth season, a gradual decline in surface silicate concentrations is observed. This process begins in March, when the spring diatom bloom develops and actively consumes silica for the formation of their cellular structures. In the following spring months, silicate concentrations remain at a lower level. The lowest mean values are recorded in April, when the average concentration decreases to approximately 8.23 mmol m^{-3}, with minimum spatial values reaching about 3.70 mmol m^{-3}.

During summer, silicate concentrations remain relatively stable, with mean values around 9–9.3 mmol m^{-3}. This is associated with reduced diatom productivity and the dominance of other phytoplankton groups, such as cyanobacteria. In autumn, a gradual increase in silicate concentrations in the surface layer is observed, driven by vertical mixing processes and the regeneration of nutrients within the water column.

Table 9. Summary of mean monthly nitrate (NO_3) concentrations in the surface layer. Model data from the 3D CEMBS-PolSea for the years 2022–2025.

[mmol m^{-3}]	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
MAX	121.88	159.01	133.15	67.79	16.94	9.92	9.15	9.42	12.25	24.69	47.4	62.73
MEAN	7.97	9.60	8.03	2.19	0.08	0.03	0.06	0.33	1.11	2.80	4.75	6.65
MIN	6.28	6.52	1.62	0	0	0	0	0	0	0.04	0.84	1.65

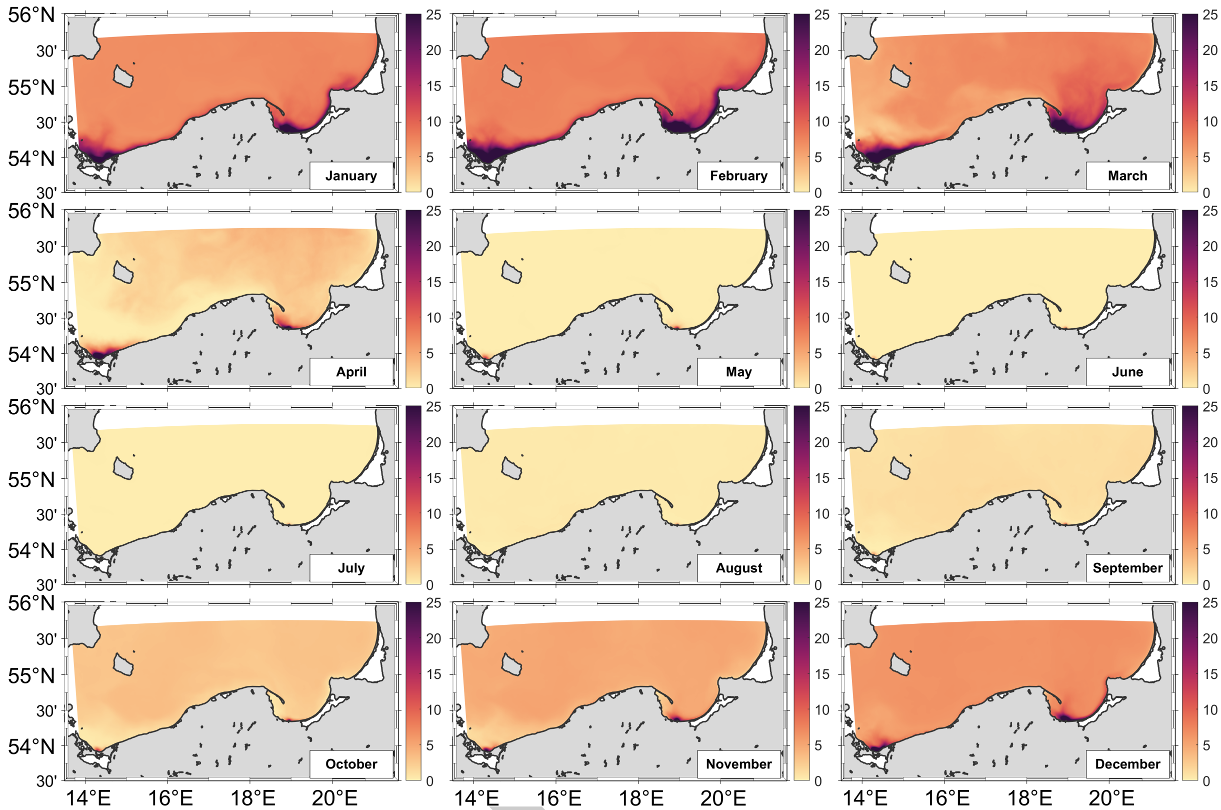

Figure 20. Mean monthly nitrate concentrations [mmol m^{-3}] at the surface derived from the 3D CEMBS-PolSea model simulations for the years 2022–2025.

The seasonal amplitude of mean silicate values during the analysed period is approximately 4.2 mmol m^{-3}, indicating a moderate seasonal variability of this parameter in the surface layer. The analysis of the spatial distribution of silicate concentrations also reveals considerable variability within the model domain. In some months, local maximum values exceed 30 mmol m^{-3}, indicating areas with significantly elevated nutrient concentrations. Such high values may be associated with riverine inputs as well as regeneration processes of silica resulting from the decomposition of organic matter.

The obtained seasonal pattern of silicate variability is consistent with the typical biogeochemical cycle observed in the Baltic Sea. Higher winter and early spring concentrations result from limited phytoplankton uptake and vertical mixing of the water column, whereas the decline in spring is directly linked to the intensive consumption of silicates during the diatom bloom. Detailed monthly data are presented in Table 8.

3.3.5 Nitrates (NO_3)

Nitrates are one of the main nitrogen sources for phytoplankton and play a key role in regulating primary production in the Baltic Sea. In the euphotic layer, their concentration is strongly controlled by the seasonal dynamics of phytoplankton development and vertical mixing processes in the water column. During periods of intense primary production, nitrates are rapidly consumed by phytoplankton, whereas in the autumn–winter period their surface-layer stocks are replenished through vertical mixing and the input of nutrients from deeper water layers. The analysis of mean monthly nitrate values in the surface layer of the 3D CEMBS-PolSea model for the period 2022–2025 shows a very pronounced seasonal variability of this parameter. The highest mean values occur in winter and early spring, before the onset of the intensive phytoplankton growth season (see Figure 20). During the analysed period, the highest mean monthly nitrate concentration was recorded in February, with a domain-averaged value of 9.60 mmol m^{-3} and maximum spatial values exceeding 159 mmol m^{-3}. High mean values are also observed in January and March, when concentrations reach approximately 7.97 mmol m^{-3} and 8.03 mmol m^{-3}, respectively. With the onset of the spring phytoplankton bloom, a rapid decline in nitrate concentrations in the surface layer is observed. Intensive nitrogen uptake by developing phytoplankton leads to near-complete depletion of nitrate stocks in the euphotic zone during late spring and summer. The lowest mean values occur in June, when the domain-averaged concentration drops to approximately 0.03 mmol m^{-3}. During summer months, minimum values in many parts of the model domain reach 0 mmol m^{-3}, indicating complete nitrate exhaustion in the surface layer. In the second half of the year, a gradual increase in nitrate concentrations in the surface layer is observed. This process begins in September, when autumn storms and intensified vertical mixing transport nitrates from deeper layers toward the surface. In October, the mean concentration increases to approximately 2.80 mmol m^{-3}, followed by a continued systematic increase throughout the autumn–winter period. The seasonal amplitude of mean nitrate values during the analysed period is very large, at approximately 9.57 mmol m^{-3}, reflecting the strong influence of seasonal primary production on the nitrogen balance in the euphotic layer.

The spatial distribution analysis of nitrate concentrations indicates a very high variability within the model domain. In some regions, maximum values exceed 100 mmol m^{-3}, which is associated with high nitrate concentrations in deeper water layers as well as local hydrodynamic processes. The highest nitrate concentrations in the water column are typically observed at greater depths (below approximately 60 m), where dissolved inorganic nitrogen accumulates as a result of organic matter decomposition.

The obtained seasonal pattern of nitrate variability is consistent with the typical biogeochemical cycle observed in the Baltic Sea. High winter and early spring concentrations result from limited biological production and intensified vertical mixing, whereas the sharp decline in spring is directly linked to intensive nitrogen uptake during phytoplankton blooms.

Detailed monthly data are presented in Table 9.

4. Discussion

The validation results demonstrate that the 3D CEMBS-PolSea biogeochemical model is capable of reproducing the major spatial and temporal patterns of biogeochemical variability in the southern Baltic Sea. The highest agreement between simulations and observations was obtained for dissolved oxygen, with correlation coefficients reaching 0.87 for the ICES Low Resolution dataset and 0.72 for the ICES High Resolution dataset. These results indicate that the model realistically represents the combined influence of physical transport, vertical mixing, biological production, and respiration processes that control oxygen dynamics in the Baltic Sea.

The validation of nutrient concentrations revealed moderate to good agreement between model results and observations. Correlation coefficients for phosphates, silicates, and nitrates ranged from 0.60 to 0.68, indicating that the model captures the dominant mechanisms controlling nutrient cycling and transport within the study area. Lower performance observed in coastal regions is likely associated with uncertainties in riverine nutrient loads, highly variable hydrological conditions, and local-scale processes that are difficult to resolve even with high-resolution numerical models.

The interpretation of model skill should also consider the substantial natural variability of biogeochemical parameters in the Baltic Sea. While some variables exhibit relatively high RMSD values, the observational datasets themselves are characterised by very large coefficients of variation, particularly for nitrate, phosphate, and chlorophyll *a* concentrations. The coefficient of variation reaches 109.9% for chlorophyll *a*, 123.8% for phosphate, and 312.3% for nitrate observations. Such variability reflects strong seasonal cycles, vertical gradients, riverine influences, and local biological processes characteristic of the southern Baltic Sea.

To provide a more robust assessment of model performance, additional normalised skill metrics were calculated, including bias, normalised RMSE (nRMSE), the RMSE-observation standard deviation ratio (RSR), and the coefficient of variation (CV). The RSR values range from 0.54 for dissolved oxygen to 0.88 for nitrate in the ICES dataset and reach 0.94 for the comparison with satellite-derived chlorophyll *a* concentrations. According to the model evaluation guidelines proposed by Moriasi et al. (2007), the obtained RSR values are higher than those typically considered satisfactory in hydrological applications. However, these criteria were developed for watershed models and may be overly restrictive when applied to highly variable marine biogeochemical variables.

Therefore, the RSR values should be interpreted in the context of the variability of the observational datasets. In all analysed cases, model errors remain comparable to or lower than the standard deviation of the observations, indicating that the model reproduces a substantial fraction of the observed variability. Consequently, RMSD values should not be interpreted independently of observational variability. The obtained results indicate that the model retains predictive skill despite the complexity of the analysed biogeochemical processes and the uncertainties associated with both in situ and satellite observations.

The lowest correlation was obtained for chlorophyll *a* concentration (r = 0.51). However, this result remains within the range commonly reported for regional biogeochemical ecosystem models operating in coastal and shelf-sea environments (Daewel and Schrum, 2013; Eilola et al., 2009; Meier et al., 2014). Chlorophyll *a* is characterised by strong short-term and local variability driven by nutrient availability, light conditions, species composition, and physical forcing. In coastal waters, rapid bloom development and small-scale processes may further increase discrepancies between model simulations and observations.

Additional uncertainty is associated with the comparison between model simulations and satellite-derived chlorophyll *a* observations. As discussed in Section 3.2, satellite products represent an optically weighted signal originating from the upper illuminated part of the water column rather than chlorophyll concentrations within a single model layer. Consequently, part of the discrepancy between model results and observations may arise from differences in the vertical representation of chlorophyll *a* in the observational and modelling systems, particularly during stratified conditions when subsurface chlorophyll maxima occur.

The assimilation experiment demonstrated that the incorporation of satellite-derived chlorophyll *a* information improves the representation of surface phytoplankton dynamics while maintaining consistency with the internal physical and biogeochemical structure of the model. Because the assimilation procedure applies gradual corrections rather than direct replacement of model values, the model state evolves toward the observed conditions while preserving the dynamical relationships among biological, chemical, and physical variables. This approach reduces the risk of introducing artificial discontinuities and maintains the stability of the modelling system.

The spatial distribution of correlation coefficients indicates generally better model performance in offshore regions than in coastal waters. This pattern is typical of ecosystem models and reflects the greater complexity of nearshore environments, where riverine inputs, sediment–water interactions, local circulation patterns, and anthropogenic pressures exert a stronger influence on biogeochemical conditions. Despite these challenges, the model reproduces the observed variability of all analysed parameters across most of the southern Baltic Sea domain.

Particular attention should be paid to the role of silicates in shaping ecosystem functioning in the Baltic Sea. Silicon availability directly influences the development of diatoms, one of the dominant phytoplankton groups in the region. Consequently, silicate concentrations affect not only the magnitude of primary production but also the structure of phytoplankton communities and nutrient cycling pathways. The tendency of the model to underestimate higher silicate concentrations in some coastal regions may indicate uncertainties in riverine silicon loads or insufficient representation of local sediment processes. Further investigations of silicon dynamics may therefore contribute to improving future model performance.

An important advantage of the developed system is its operational implementation within the CSI-POM 2 platform. Unlike many ecosystem models that remain primarily research tools, the 3D CEMBS-PolSea model provides continuously updated information that can support environmental monitoring, ecosystem assessment, marine spatial planning, and decision-making processes. The integration of numerical modelling with satellite observations and operational environmental information services bridges the gap between scientific research and practical applications in marine management.

Future work will focus on further validation and refinement of the biogeochemical module, improved estimates of riverine nutrient loads, better representation of coastal processes, and the implementation of advanced data assimilation methods. These efforts are expected to enhance both the predictive performance and the operational applicability of the 3D CEMBS-PolSea modelling framework.

5. Conclusions

The main conclusions of this study are as follows:

1. A high-resolution three-dimensional biogeochemical model (3D CEMBS-PolSea) was successfully developed and implemented for the southern Baltic Sea within the CSI-POM 2 operational framework.

2. The model reproduces the major spatial and temporal patterns of dissolved oxygen, chlorophyll *a*, phosphate, silicate, and nitrate variability observed in the southern Baltic Sea.

3. The highest validation performance was obtained for dissolved oxygen, with correlation coefficients reaching 0.87 (ICES Low Resolution) and 0.72 (ICES High Resolution).

4. Nutrient concentrations were reproduced with moderate to good agreement ($r = 0.60 - 0.68$), indicating that the model captures the dominant mechanisms governing nutrient cycling and transport.

5. Chlorophyll *a* simulations showed moderate agreement with observations ($r = 0.51$), demonstrating the capability of the model to reproduce the main features of seasonal phytoplankton dynamics despite the high natural variability of chlorophyll *a* concentrations and the limitations associated with satellite-derived products.

6. Additional performance metrics, including bias, nRMSE, RSR, and the coefficient of variation, indicate that model errors remain comparable to or lower than the variability of the observational datasets.

These results confirm that the model retains predictive skill and reproduces a substantial fraction of the observed environmental variability.

7. The assimilation of satellite-derived chlorophyll *a* observations improves the representation of surface phytoplankton dynamics and enhances the operational applicability of the modelling system.

8. The developed modelling framework provides a valuable tool for environmental monitoring, ecosystem assessment, forecasting, and marine management in Polish marine areas.

Acknowledgements

Publication financed from the state budget under the programme of the Minister of Education and Science (Poland) entitled "Science for Society II" No. NdS-II/SP/0003/2023/01, amount of funding PLN 1,996,763.77, total value of the project PLN 1,996,763.77. Some elements of the 3D CEMBS-PolSEA model (i.e., river runoff data) are based on the solutions developed during the WaterPUCK project funded by the National Centre for Research and Development of Poland within the BIOSTRATEG III program BIOSTRATEG3/ 343927/3/NCBR/2017. Calculations were carried out at the Academic Computer Centre in Gdańsk.

This study has been conducted using E.U. Copernicus Marine Service Information; https://doi.org/10.48670/moi-00013

Conflict of interest

None declared

References

Daewel, U., Schrum, C., 2013. *Simulating long-term dynamics of the coupled North Sea and Baltic Sea ecosystem with ECOSMO II: Model description and validation*. J. Marine Syst. 119–120, 30–49. https://doi.org/10.1016/j.jmarsys.2013.03.008

Darecki, M., Stramski, D., 2004. *An evaluation of MODIS and SeaWiFS bio-optical algorithms in the Baltic Sea.* Remote Sens. Environ. 89, 326–350. https://doi.org/10.1016/j.rse.2003.10.012

Dybowski, D., Dzierzbicka-Głowacka, L., 2023. *Analysis of the impact of nutrients deposited from the land side on the waters of Puck Lagoon (Gdańsk Basin, Southern Baltic): A model study*. Oceanologia 65 (2), 386–397. https://doi.org/10.1016/j.oceano.2022.11.005

Dybowski, D., Jakacki, J., Janecki, M., Nowicki, A., Rak, D., Dzierzbicka-Glowacka, L., 2019. *High-Resolution Ecosystem Model of the Puck Bay (Southern Baltic Sea) – Hydrodynamic Component Evaluation*. Water 11, 2057. https://doi.org/10.3390/w11102057

Dybowski, D., Janecki, M., Nowicki, A., Dzierzbicka-Glowacka, L.A., 2020. *Assessing the Impact of Chemical Loads from Agriculture Holdings on the Puck Bay Environment with the High-Resolution Ecosystem Model of the Puck Bay, Southern Baltic Sea*. Water 12, 2068. https://doi.org/10.3390/w12072068

Dybowski, D., Janecki, M., Nowicki, A., Jakacki, J., Dzierzbicka-Głowacka, L., 2025. *Development and validation of a high-resolution hydrodynamic model for Polish Marine Areas*. Oceanologia 67 (1), 67105. https://doi.org/10.5697/EWGU8323

Dzierzbicka-Glowacka, L., Dybowski, D., Janecki, M., Wojciechowska, E., Szymczycha, B., Potrykus, D., Nowicki, A., Szymkiewicz, A., Zima, P., Jaworska-Szulc, B., Pietrzak, S., Pazikowska-Sapota, G., Kalinowska, D., Nawrot, N., Wielgat, P., Dembska, G., Matej-Łukowicz, K., Szczepańska, K., Puszkarczuk, T., 2022. *Modelling the impact of the agricultural holdings and land-use structure on the quality of inland and coastal waters with an innovative and interdisciplinary toolkit*. Agr. Water Manage. 263, 107438. https://doi.org/10.1016/j.agwat.2021.107438

Dzierzbicka-Głowacka, L., Jakacki, J., Janecki, M., Nowicki, A., 2013a. *Activation of the operational ecohydrodynamic model (3D CEMBS) – the hydrodynamic part*. Oceanologia 55 (3), 519–541. http://dx.doi.org/10.5697/oc.55-3.519

Dzierzbicka-Głowacka, L., Janecki, M., Nowicki, A., Jakacki, J., 2013b. *Activation of the operational ecohydrodynamic model (3D CEMBS) – the ecosystem module*. Oceanologia 55 (3), 543–572. https://doi.org/10.5697/oc.55-3.543

Dzierzbicka-Głowacka, L., Dybowski, D., Janecki, M., Nowicki, A., 2025. *Digital Information System for the Polish Marine Areas – Modelling of Structures and Dynamics of Physical Processes in the Southern Baltic*. Oceanologia 67 (1), 67104. https://doi.org/10.5697/CUKW4719

Eilola, K., Meier, H.E.M., Almroth, E., 2009. *On the dynamics of oxygen, phosphorus and cyanobacteria in the Baltic Sea; A model study*. J. Marine Syst. 75, 163–184. https://doi.org/10.1016/j.jmarsys.2008.08.009

Feistel, R., Nausch, G., Wasmund, N. (Eds.). 2008. *State and Evolution of the Baltic Sea, 1952–2005: A Detailed 50-Year Survey of Meteorology and Climate, Physics, Chemistry, Biology, and Marine Environment*. John Wiley & Sons, Inc., Hoboken, 703 pp. https://doi.org/10.1002/9780470283134

Gräwe, U., Holtermann, P., Klingbeil, K., Burchard, H., 2015. *Advantages of vertically adaptive coordinates in numerical models of stratified shelf seas*. Ocean Model. 92, 56–68. https://doi.org/10.1016/j.ocemod.2015.05.008

Gustafsson, B.G., Schenk, F., Blenckner, T., Eilola, K., Meier, H.E.M., Müller-Karulis, B., Neumann, T., Ruoho-Airola, T., Savchuk, O.P., Zorita, E., 2012. *Reconstructing the development of Baltic Sea eutrophication 1850–2006*. AMBIO, 41, 534–548. https://doi.org/10.1007/s13280-012-0318-x

HELCOM, 2023. *State of the Baltic Sea 2023*. Baltic Marine Environment Protection Commission – Helsinki Commission. https://stateofthebalticsea.helcom.fi

Hordoir, R., Axell, L., Höglund, A., Dieterich, C., Fransner, F., Gröger, M., Liu, Y., Pemberton, P., Schimanke, S., Andersson, H., Ljungemyr, P., Nygren, P., Falahat, S., Nord, A., Jönsson, A., Lake, I., Döös, K., Hieronymus, M., Dietze, H., Löptien, U., Kuznetsov, I., Westerlund, A., Tuomi, L., Haapala, J., 2019. *Nemo-Nordic 1.0: A NEMO-based ocean model for the Baltic and North seas – Research and operational applications*. Geosci. Model Dev. 12, 363–386. https://doi.org/10.5194/gmd-12-363-2019

Leppäranta, M., Myrberg, K., 2009. *Physical Oceanography of the Baltic Sea*. Springer-Praxis Books, Dordrecht, 401 pp. https://doi.org/10.1007/978-3-540-79703-6

Meier, H.E.M., Andersson, H.C., Arheimer, B., Donnelly, C., Eilola, K., Gustafsson, B.G., Kotwicki, L., Neset, T.-S., Niiranen, S., Piwowarczyk, J., Savchuk, O.P., Schenk, F., Murawski, J., She, J., Mohn, C., Frishfelds, V., Nielsen, J.W., 2021. *Ocean Circulation Model Applications for the Estuary-Coastal-Open Sea Continuum*. Front. Mar. Sci. 8. https://doi.org/10.3389/fmars.2021.657720

She, J., Berg, P., Berg-Leonardsson, L., 2020. *Operational forecasting of marine environmental conditions in the Baltic Sea*. Earth Syst. Sci. Data 12, 693–716. https://doi.org/10.5194/essd-12-693-2020

Stramska, M., Konik, M., Aniskiewicz, P., Jakacki, J., Darecki, M., 2021. *Comparisons of Satellite and Modeled Surface Temperature and Chlorophyll Concentrations in the Baltic Sea with In Situ Data*. Remote Sens. 13, 3049. https://doi.org/10.3390/rs13153049

Węsławski, J.M., Zorita, E., 2014. *Ensemble Modeling of the Baltic Sea Ecosystem to Provide Scenarios for Management*. AMBIO, 43, 37–48. https://doi.org/10.1007/s13280-013-0475-6

Woźniak, M., Bradtke, K.M., Darecki, M., Krężel, A., 2016. *Empirical Model for Phycocyanin Concentration Estimation as an Indicator of Cyanobacterial Bloom in the Optically Complex Coastal Waters of the Baltic Sea*. Remote Sens. 8, 212. https://doi.org/10.3390/rs8030212