



# Complexity, accuracy and practical applicability of different biogeochemical model versions

F.J. Los, M. Blaas\*

*Deltares, P.O. Box 177, 2600 MH Delft, The Netherlands*

## ARTICLE INFO

Available online 18 January 2010

### Keywords:

Eco-hydrodynamic modeling  
Validation  
Goodness of fit  
Model accuracy  
Model usability

## ABSTRACT

The construction of validated biogeochemical model applications as prognostic tools for the marine environment involves a large number of choices particularly with respect to the level of details of the physical, chemical and biological aspects. Generally speaking, enhanced complexity might enhance veracity, accuracy and credibility. However, very complex models are not necessarily effective or efficient forecast tools. In this paper, models of varying degrees of complexity are evaluated with respect to their forecast skills. In total 11 biogeochemical model variants have been considered based on four different horizontal grids. The applications vary in spatial resolution, in vertical resolution (2DH versus 3D), in nature of transport, in turbidity and in the number of phytoplankton species. Included models range from 15 year old applications with relatively simple physics up to present state of the art 3D models. With all applications the same year, 2003, has been simulated.

During the model intercomparison it has been noticed that the 'OSPAR' Goodness of Fit cost function (Villars and de Vries, 1998) leads to insufficient discrimination of different models. This results in models obtaining similar scores although closer inspection of the results reveals large differences. In this paper therefore, we have adopted the target diagram by Jolliff et al. (2008) which provides a concise and more contrasting picture of model skill on the entire model domain and for the entire period of the simulations. Correctness in prediction of the mean and the variability are separated and thus enhance insight in model functioning. Using the target diagrams it is demonstrated that recent models are more consistent and have smaller biases. Graphical inspection of time series confirms this, as the level of variability appears more realistic, also given the multi-annual background statistics of the observations. Nevertheless, whether the improvements are all genuine for the particular year cannot be judged due to the low sampling frequency of the traditional monitoring data at hand. Specifically, the overall results for chlorophyll-*a* are rather consistent throughout all models, but regionally recent models are better; resolution is crucial for the accuracy of transport and more important than the nature of the forcing of the transport; SPM strongly affects the biomass simulation and species composition, but even the most recent SPM results do not yet obtain a good overall score; coloured dissolved organic matter (CDOM) should be included in the calculation of the light regime; more complexity in the phytoplankton model improves the chlorophyll-*a* simulation, but the simulated species composition needs further improvement for some of the functional groups.

© 2010 Elsevier B.V. All rights reserved.

## 1. Introduction

The first eco-hydrodynamic models for aquatic systems were developed more than thirty years ago. Examples of these first generation models are found in Di Toro et al. (1971, 1977). At present, many models exist, some with a relatively long history, while development of others has started more recently, but of course these also include many features from older models. Many papers describe the status i.e. the present version of a model application, demonstrating its strong points and discussing some of its weaker points. While these papers are certainly

meaningful, it is often hard to determine which characteristics are of major importance and which characteristics actually do not contribute much to the quality of a particular model.

While more knowledge and computational power become available, many modellers tend to enhance the complexity of the models they develop. However, in Los et al. (2008) we have pointed out that adding more complexity does not necessarily improve the quality of the model results in terms of their ability to reproduce the measurements and hence their applicability as prognostic tools. Instead, we have argued that there should be a balance between ecological and physical resolution in relation to the specific question to be addressed. For example, an appropriate model for assessing the impacts of sand mining in a coastal area is not necessarily adequate to assess the impacts of nutrient reduction or the probability of low oxygen conditions in an

\* Corresponding author.

E-mail address: [Meinte.Blaas@deltares.nl](mailto:Meinte.Blaas@deltares.nl) (M. Blaas).

offshore area or the occurrence of undesirable blooms of *Phaeocystis* during the spring bloom.

In this paper, a comparison is made between several generations of the eco-hydrodynamic model applications developed for the North Sea at Deltares (formerly WL | Delft Hydraulics) during the last 15 years. There are many differences between these applications with respect to their forcing, resolution, biological complexity and process parameterization. In order to find out how much each modification contributes to changes in model behaviour, we need to make a systematic comparison in terms of the spatial and ecological resolutions of these models. To that purpose we have revitalized several distinguishable model versions, and run all of these imposing forcing for a single, recent year (2003). Thus, the central question is: which factors matter most and which might look important, but actually contribute less to improvements in model behaviour? A secondary question is if, and if so how, we can quantify evolution in model skill. Notice that evolution does not necessarily occur in a linear fashion, so whereas the overall skill may improve relative to the measurements, results for some variables or at some locations or in parts of the year might actually deteriorate simultaneously. The following factors were considered during this study:

- the resolution of the grid,
- the nature of transport forcing (atmospheric, density),
- attenuation of the underwater light conditions by SPM and CDOM,
- the level of detail of the phytoplankton model.

Unlike the inter-model comparisons such as by Moll and Radach (2003), Radach and Moll (2006) and Lenhart et al. (2010), all of the applications presented in the current paper belong to the same model family and their set-up and forcing has been standardized to a large extent. The main features of the models and the Goodness of Fit criteria are presented in Section 2. Results are presented in Section 3 both for the North Sea as a whole as well as for individual locations. Generalization of the results will be discussed at the end of this paper.

## 2. Main features of models

In this paper, a total of 11 biogeochemical model variants are considered. These stem from historic applications that have been developed in and applied to various research projects in the past. Each of the models was originally applied to explain observed phenomena and to predict some future conditions. Later model versions were usually run with some new processes, parameters settings and forcings for a more recent period of time. Simply comparing the existing output of previous model simulations therefore leaves many questions open on how to explain the differences between them. For this study many differences were eliminated in order to be able to concentrate on those modifications that matter most. Occasionally different combinations of forcings were run to check their impacts one at a time. For instance the latest models were also run using the suspended matter (SPM) field of the oldest models to force the underwater light climate.

In essence there are four different horizontal grids that provide the general name as identification. On these four grids 1 to 6 variants have been defined, depending on vertical resolution, the description of the transport, the SPM fields, the light extinction model and the modelling of the algal dynamics. An overview of the differences between the models applied here is given in Table 1. A key to the codes that indicate the different variants is also given. An extensive overview of historic model versions is given by Los et al. (2008). The grids are shown in Fig. 1.

Below, the general similarities are presented; particular differences are discussed in the sections thereafter.

### 2.1. General principle and similarities

#### 2.1.1. Phytoplankton dynamics

For all except one application, the algal dynamics are modelled with specific versions of the phytoplankton module BLOOM. The most

recent version is referred to as BLOOM/GEM. BLOOM is a generic model code with a long history, which in its current mode is applied to many different water systems such as the North Sea, a number of Dutch water bodies i.e. the saline lakes Grevelingen, and Veere, The Eastern Scheldt Estuary, and the future saline Lake Volkerak-Zoom. International applications include the Lagoon of Venice, the Sea of Marmara and the future saline Marina Reservoir in Singapore. An extensive description of the main features of the model is provided by Blauw et al. (2009); its application to the North Sea is described by Los et al. (2008). A more detailed description of the phytoplankton module BLOOM is presented by Los and Wijsman (2007) and in Loucks and Van Beek (2005). The application-specific details and on the usage of BLOOM will be discussed below in Section 2.4.

#### 2.1.2. Reference year 2003

Year-specific forcings i.e. nutrient loads from rivers and meteorological conditions have been adopted from data for a single, recent year, 2003, for all simulations by all models. At the time this study was performed, nearly complete data sets for forcing and monitoring were available for the entire period 1996–2003. This last year was chosen not just because it is the most recent one, but also because it is an atypical year with a wet spring and a dry, warm summer and autumn. We expected that such a year would be more suitable for finding differences between models than a more average year. We did not try to improve the performance of existing model applications, assuming their previous calibration had been done adequately. Notice that none of the models had previously been calibrated for this particular year, so the 2003 simulation may be considered as a validation case for all of the models.

#### 2.1.3. Meteorological forcing

Both the hydrodynamic and primary production models require meteorological information but not exactly the same. For instance the 2D hydrodynamic models were run with a uniform constant temperature of 20 °C, while a seasonal temperature function has been imposed on all primary production models. The day length and 2003 solar irradiance levels for each of the primary production models are adopted from historic measurements by the Royal Dutch Meteorological Institute at a single land-based station (De Kooy) near Den Helder in the north western part of the Netherlands. Some details on the meteorological forcing of the hydrodynamic models are presented in the more detailed description of each model below.

In the GENO, CSM(CDGSB) and Coastal zone models a spatially uniform, seasonally varying sea water temperature was adopted based on measurements at station Noordwijk 10 km (see Fig. 2) for 2003. In both ZUNO models temperature is specified by a spatially varying temperature field taken from the simulations by the hydrodynamic model Delft3D-Flow (Lesser et al., 2004; WL | Delft Hydraulics, 2005).

#### 2.1.4. Rivers and other nutrient sources

The nutrient loads of all models are basically the same. The model input contains the point sources of nutrients and fresh water from the main Belgian, Dutch, German, French and UK rivers in as far as they are part of the model domain. In the Coastal zone model only the Dutch rivers are explicitly included. For the Dutch rivers, substance loads were derived from measured discharges per day and concentrations in rivers at 10-day intervals for the year 2003. Data for the other main rivers is usually also available per decade (Blauw et al., 2006). Modelled substances not measured have been inferred from measured data of other substances, using stoichiometric ratios and other knowledge rules that have been developed and proven successful in previous studies (e.g., Los and Wijsman, 2007).

#### 2.1.5. Boundary conditions

There is considerable overlap between the domains included in most of the models presented here. In all but one model, whose

**Table 1**  
Description of model runs and explanation of codes.

| ID   | Application name | Grid  | Transport and dispersion  | Boundary conditions   | Turbidity field   | Sea water temperature  | Algal model  |
|--|------------------|---|---|---|---|--|--|
| <i>GDGSD</i>                                 | GENO-DYNAMO      | 16 km southern North Sea rectilinear; 1395 elements   | Repeated single tidal cycle + wind-driven residual (SW wind 4.5 m/s)<br>2DH transport from 2DH hydrodynamic model   | Climatology 1980s, Channel annual cycle, Atlantic constant in time, constant in space | Spatial pattern imposed SPM from observations, stationary annual mean   | Spatially uniform time series (2003) from station Noordwijk 10 | Michaelis–Menten kinetics, 2 functional groups ( <i>diatoms</i> and <i>'green'</i> ), 12 state variables<br>BLOOM96: 4 species, 12 types, 26 state variables |
| <i>GDGSB</i>                                 | GENO-NZB         |   |   |   | As above, but modulated with seasonal cycle   |  | BLOOM96: 4 species, 12 types, 26 state variables   |
| <i>CDGSB</i>                                 | CSM              | 8–16 km entire North Sea rectilinear; irregular; 3915 elements  | Repeated single tidal cycle + wind-driven residual (SW wind 9 m/s)<br>2DH transport from 2DH hydrodynamic model   | Channel as above; Atlantic: recomputed from GENO-NZB, constant in time and space.     | As <i>GDGSB</i> but with a different specific extinction  | Spatially uniform time series (2003) from station Noordwijk 10 | BLOOM96: 4 species, 12 types, 26 state variables   |
| <i>KSKCB</i>                                 | COAST            | 1–10 km Dutch coast < 70 km, curvilinear; 2153 elements<br>~ 1 km resolution in near-shore stations such as Walcheren 2 | As CSM but with additional, seasonal upwelling parameterization in horizontal dispersion coefficient based on 30-year average wind data<br>2DH transport from 3D hydrodynamic model | Derived from measurements at near-boundary monitoring stations                        | Spatial pattern from separate SPM model; harmonic annual cycle; Extinction due to CDOM parameterised as function of fresh water fraction<br>As <i>GDGSB</i> |  |  |
| <i>KSGCB</i><br><i>ZNZCB</i>                 | ZUNO-2D          | 2–20 km southern NS, curvilinear; 4350 elements<br>2 to 4 km resolution in near-shore stations                          | Repeated spring–neap cycle<br>2DH transport from 3D hydrodynamics model   | Channel and Atlantic: As GENO-NZB and GENO-DYNAMO                                     | Spatial pattern from separate SPM model.<br>Seasonal cycle, with wind dependent noise   | As <i>GDGSB</i>  | BLOOM with Further refined algal parameters (lab studies, see text)  |
| <i>ZRZCB</i><br><i>ZRZCD</i>                 |                  |   | Full year, variable forcing, 2DH<br>2DH transport from 3D hydrodynamic model  |   |   | Spatially varying, derived from Delft3D hydrodynamic model     | BLOOM with 2 species: diatom and flagellates<br>As <i>ZNZCB</i>  |
| <i>ZRZCB</i><br><i>Z3ZCB</i><br><i>Z3VCB</i> | ZUNO-3D          | As ZUNO 2D 10 layers; 43,500 elements   | Full year, variable forcing, 3D transport from 3D hydrodynamic model  |   | As <i>GDGSB</i><br>As <i>ZNZCB</i><br>SPM model (Van Kessel et al., 2008)   |  |  |

Loads are all consistent since they are all based on the same data set by [Blauw et al. \(2006\)](#) with data for the major rivers in the domains for 2003.

*Identifiers*: 1st letter: Grid. G = GENO, C = CSM, K = COAST, Z = ZUNO.

2nd letter: Forcing. D = average day, S = average day, seasonal correction dispersion, N = characteristic spring–neap, R = realistic (actual) 2DH, 3 = realistic (actual) 3D model.

3rd letter: SPM. G = GENO, K = COAST, Z = ZUNO steady state + seasonal harmonic + noise; V = ZUNO full SPM model.

4th letter: Extinction. S = only SPM, no CDOM, C = SPM + Salinity as proxy for CDOM.

5th letter: Algal model. D = 2 species, B = 12 species.

domain is restricted to the Dutch coastal zone, the position of the southern open boundary is at the same location in the British Channel. The concentrations of substances at this boundary have been based on the work of [Laane et al. \(1993\)](#) and are the same for each of these models. Concentrations are specified as a monthly time series, uniformly distributed over the cross section and over depth as generally water masses are vertically well mixed in this part of the North Sea. The north boundary of two of the model domains is at 57° N and the substance concentrations are assumed to be constant during the year (both vertically and horizontally). More details are presented below when the individual model is described. Notice that on an annual basis there is net import of water (and material) across the south boundary. Near the UK coast there is usually important across the north boundary as well, whereas on the Danish–Norwegian side the dominant annual flow is outward.

## 2.2. SPM forcing

Light is the main limiting factor of phytoplankton for most of the time during the winter half year in the southern North Sea. Also at many locations the onset of the spring bloom is controlled by the underwater availability of light, which depends on irradiance, the mixing depth and the turbidity (the extinction coefficient). The irradiance and physical

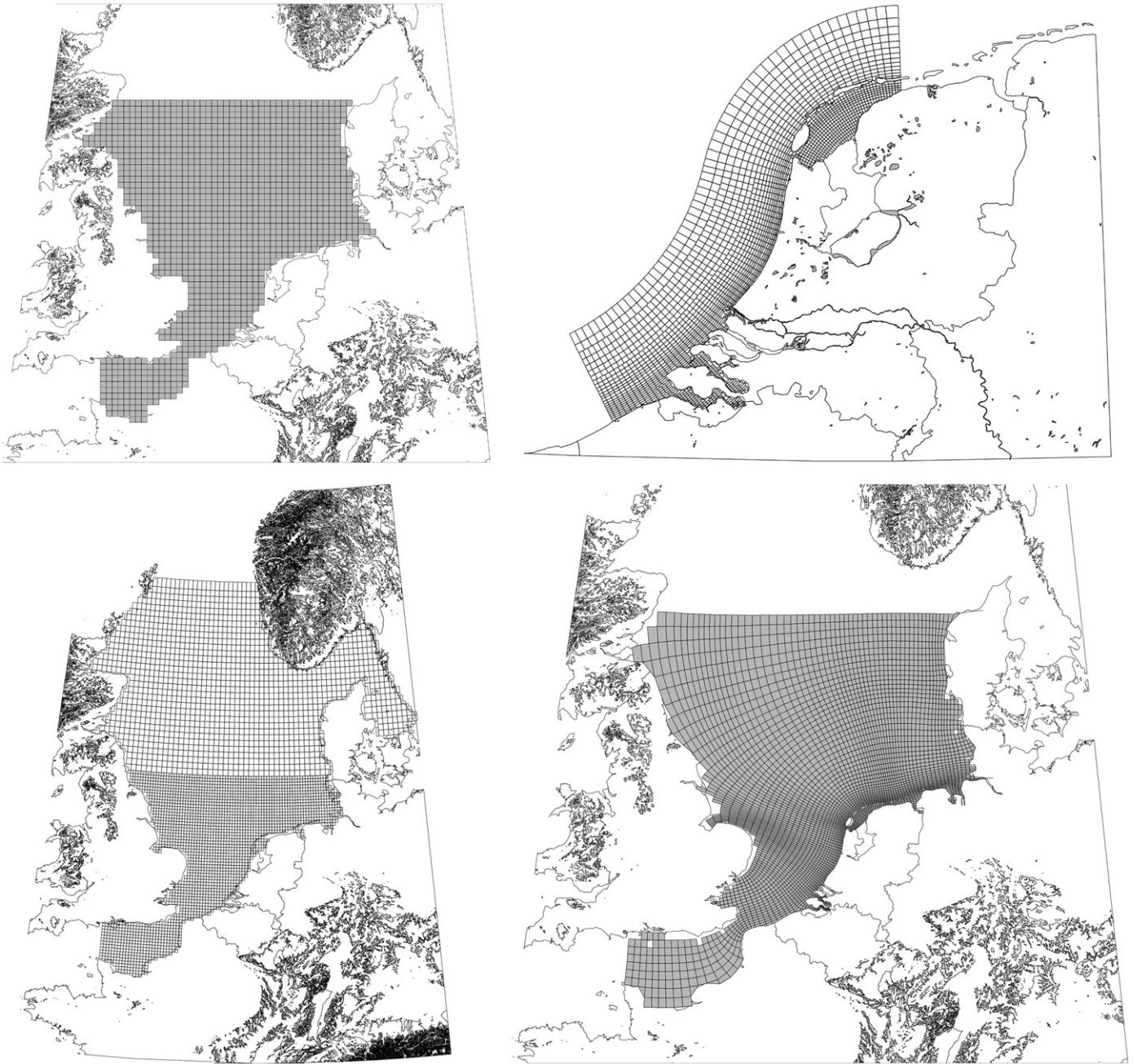
depth are the same in all of the models discussed here, apart from some minor differences in bathymetry. In the 2D models the vertical mixing equals the physical depth, but in the 3D models the mixing depth depends on the vertical diffusivities. In all models the attenuation of light is computed according to the well known Lambert–Beer equation relating the overall absorption to substance concentrations, ignoring scattering. The substance fractions taken into account and their contribution to the vertical extinction are not the same in all models, however. The following general equation is adopted for the total extinction  $K_d$ :

$$K_d = K_0 + K_{\text{SPM}} + K_{\text{CDOM}} + K_{\text{alg}} + K_{\text{det}} \quad (1)$$

where  $K_0$  is the background extinction,  $K_{\text{SPM}}$  is the contribution due to inorganic suspended matter,  $K_{\text{CDOM}}$  is the contribution of 'other' organic matter (yellow substance),  $K_{\text{alg}}$  is the total contribution of all phytoplankton species and  $K_{\text{det}}$  is the contribution of labile detritus. The first three terms are prescribed, although they may be the result of other models, the last two terms are state variables in all primary production models described here. The contribution of yellow substance (CDOM) is only explicitly taken into account in later models (see [Table 1](#)).

In offshore North Sea waters, SPM usually does not contribute much to the light attenuation, but in the continental coastal waters, typically 25 to 75% of the light extinction is caused by SPM. The





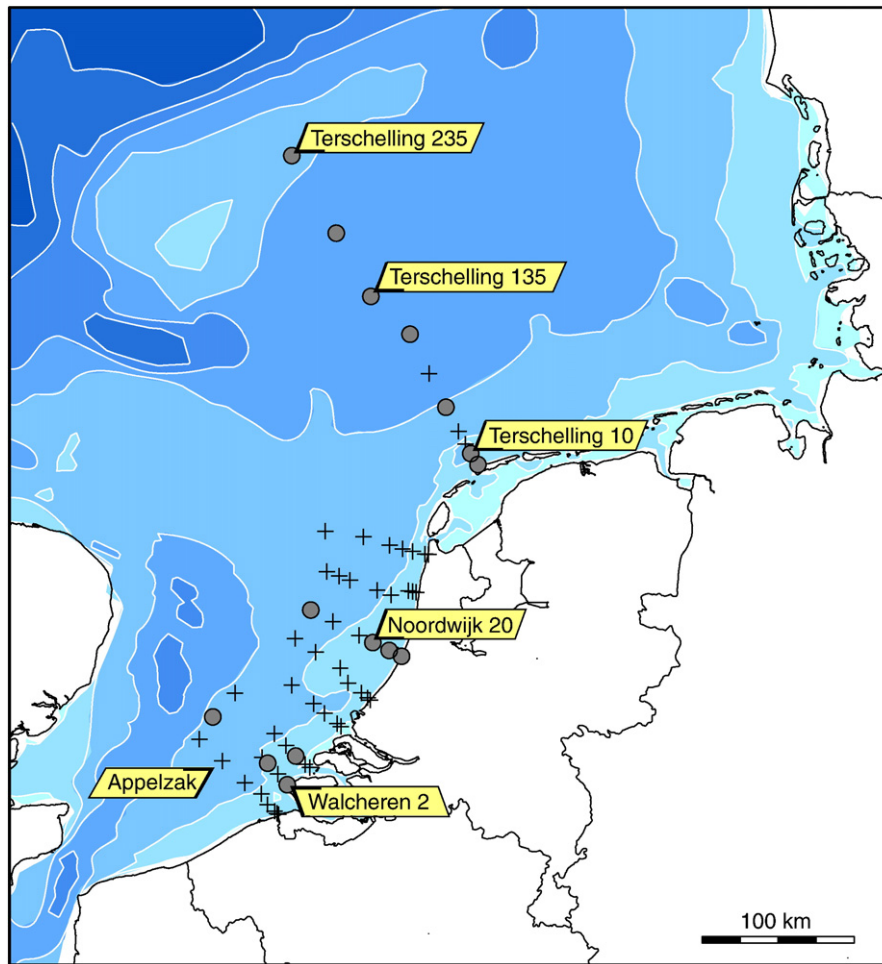
**Fig. 1.** Grid of models: GENO (upper left), Coastal zone (upper right), CSM (lower left) and ZUNO (lower right).

distribution of SPM in the North Sea is determined by coastal erosion, sea-bed resuspension by waves and tides and by the residual transport due to tidal, wind-driven, and density-driven currents (see e.g., Eisma and Kalf 1987; Eleveld et al., 2004). Because the spatial and temporal distributions of SPM vary, the SPM fields of all models discussed are adopted from SPM submodels, which are different for different versions of the primary production models. To assess the sensitivity of the modelling results to differences in SPM fields, all models have also been run using the oldest (although less accurate) SPM field. The later ones are forced by more realistic SPM values, which should be reflected in their ability to correctly model phytoplankton.

### 2.3. Initial conditions

For the sake of intercomparison of the model results, all runs of the primary production models have been spun up to statistical equilib-

rium by repeated recycling of the forcing conditions of 2003. Since 2003 is not an average year, conditions at the end of the simulation do not exactly match those at the start. This means that some discrepancies between model and observation could occur at the beginning of the time series of output, but considering the relationship between the models, there is no reason to assume that this will cause systematic differences between various models. Generally speaking, the initial conditions in the water column affect the solutions for about a month. The influences of local conditions and river input rapidly dominate the response. It should be noted however that initial conditions of the sediment bed have a longer lasting effect. These conditions are however less determined by a particular preceding year than by a longer-time history of accumulation and dispersion such as interannual climatic changes and changes in anthropogenic nutrient loads. These long-term initialisation issues have been taken into account by the chosen spin-up procedure.



**Fig. 2.** Location of coastal water quality and ecology monitoring stations from the Dutch national (MWTL) monitoring programs in the Southern North Sea. All stations indicated by a circle were monitored in 2003, results are presented for stations mentioned by name. Stations marked by + were monitored in some but not all years between 1975 and 2003 and considered during the calibration and validation of the original model applications. Bathymetry of ZUNO models is included for indication, contour lines start a 10 m depth, interval 15 m.

Most of the hydrodynamic models used here were derived for average conditions so they were already at equilibrium (but not typically for 2003). In contrast, the initial conditions for hydrodynamic models with seasonal historic forcings were derived from a simulation for the preceding year (2002). Neither the concentrations at the boundaries nor the loads were adjusted during the initialisation procedure. Hence, the resulting concentration fields reflect both the behaviour of the primary production models as well as the amount of material crossing the boundaries.

## 2.4. Set-up of individual models

### 2.4.1. GENO-DYNAMO(GDGSB)

The development of GENO-DYNAMO(GDGSB) began in the middle of the 1980s when one of the main issues was to describe, understand and predict the development and abatement of eutrophication in the Dutch coastal zone (Van Pagee et al., 1988; Glas and Nauta, 1989; Nauta et al., 1992; Los et al., 1994; Peeters et al., 1995). Transport modelling is based on 2D hydrodynamic calculations for average conditions (one representative day). The GENO (Generic North Sea) grid has a uniform  $16 \times 16$  km mesh size covering the southern North Sea (up to  $57^\circ$  N), totalling 1395 computational elements (see Fig. 1). A constant SW wind of  $4.5 \text{ m s}^{-1}$  is applied to force the transport in addition to the semidiurnal tide. To simulate the transport in the primary production model an additional horizontal dispersion of  $150 \text{ m}^2 \text{ s}^{-1}$  was added in both directions. The resulting flow field

approximates the annual average in as far as this is possible with this model set-up.

In this model a linear light model is assumed including terms for background extinction, suspended inorganic particulate matter (SPM), detritus and phytoplankton. The background extinction is constant in time and space, but SPM varies spatially. The SPM field was simulated by a steady state model (Los et al., 1994). The contributions of detritus and phytoplankton are simulated.

GENO-DYNAMO(GDGSB) considers closed nutrient cycles for N, P and Si and two functional groups of phytoplankton: 'diatoms' and 'others', the first requiring silica as a nutrient. Stoichiometry in this model, including the C/Chlorophyll ratio, is fixed. The model includes only twelve state variables and includes a simple sediment module for storage and release of nutrients and processes such as mineralization, nitrification and denitrification.

### 2.4.2. GENO-NZB(GDGSB)

This may be considered as an extended version of GENO-DYNAMO(GDGSB). There are two differences between these models. First: in GENO-NZB(GDGSB) the constant SPM field is transformed into a time variable field. The overall seasonal variation is simulated by means of a harmonic function with relative high values in winter and low values in summer. Its amplitude is based upon the observed level of variation for several years. This method was first described by Los and Bokhorst (1997). Although the harmonic function does not capture

the seasonal pattern of any particular year accurately, it does result a more realistic annual pattern compared to a constant value over time.

Second, the simple algal module was replaced by the much more advanced BLOOM module. This module considers 12 phytoplankton types grouped into 4 functional groups: diatoms, (micro)flagellates, dinoflagellates and *Phaeocystis*. BLOOM had already been applied to many fresh water systems since the end of the 1970s and as a 1Dv model to the North Sea (Peeters et al. 1995), but GENO-NZB(GDGSB) was the first extensive marine application of this model. For each species and species type, there is a different factor for converting biomass to chlorophyll concentration. This factor is depending on the limiting condition (light; nutrients) and ranges from 0.0067 to 0.0533 mg chlorophyll-*a*/mg C. The types are the state variables of the model so upon transport the composition of not only the community but also of the species is adjusted automatically. The BLOOM module then recomputes the stoichiometry and writes its output. Growth, respiration, sedimentation and mortality also vary per type; it is assumed that the nutrient adapted types have a lower maximum growth rate and higher specific loss rates compared with the energy adapted types. A complete overview is given in Los and Wijsman (2007) and Los et al. (2008).

The original set of phytoplankton model coefficients was determined using preliminary results from laboratory experiments by Roel Riegman at the Netherlands Institute of Sea Research (NIOZ), which were later published in revised form. These are not used for the simulations presented here, however. During later model applications these coefficients were adapted based on work by Riegman et al. (1992), Riegman et al. (1996), Riegman (1996), Jahnke (1989) and simulation results of the model. This set of coefficients, which was originally used for the CSM(CDGSB) and COAST(KSKCB) model applications (see below), has also been used for GENO-NZB(GDGSB) in the present study in order to allow for a direct comparison between these models. Regarding all other aspects (hydrodynamics, forcings, non-algal model coefficients etc.), GENO-DYNAMO(GDGSB) and GENO-NZB(GDGSB) are identical.

#### 2.4.3. CSM(CDGSB)

During the mid 1990s two new model applications were developed: the large area Continental Shelf Model (CSM(CDGSB)) and a near coastal application COAST(KSKCB), see Fig. 1. The domain of CSM(CDGSB) extends much further to the north in comparison to GENO-NZB(GDGSB). These models were used for the OSPAR international model intercomparison (Villars and de Vries, 1998). In CSM two grid resolutions are used. In the northern part of the domain the grid is similar to GENO-NZB(GDGSB), but in the southern part, the elements are four times smaller (i.e.  $8 \times 8$  km). Several combinations of transport fields, dispersion coefficients and wind speeds have been tested, but the most realistic results in transport of substances were obtained by simply forcing the model with a constant SW wind of  $9 \text{ m s}^{-1}$  without an additional dispersion coefficient. The total amount of computational elements of CSM(CDGSB) is 3915. The hydrodynamic model was applied in 2D mode and in addition to wind, only the tide was taken into account.

The Channel boundary of CSM(CDGSB) is at the same position as for GENO and the same concentrations are used. Notice that differences in residual flows could still lead to different fluxes of for instance nutrients though. The (new) north boundary was constructed by an iterative procedure in such a way that concentrations of total nutrients along the transect at  $57^\circ$  north are similar to those specified as north boundary for GENO-NZB(GDGSB). In the CSM model the exchange with the Baltic through the Kattegat is specified as an additional discharge with a constant flow rate of about  $14,000 \text{ m}^3 \text{ s}^{-1}$  and concentrations similar to those of the Channel boundary.

The light model is the same as in GENO-NZB(GDGSB), the same SPM field is used but the specific extinction coefficient of SPM was

increased. During the construction of the GENO models only a small number of light extinction measurements had been available. In later years, when these numbers started to be collected more regularly, it became obvious that the contribution by SPM to the extinction had been underestimated in the earlier models. Consequently: in CSM (CDGSB) a larger portion of the incident radiation is absorbed by the SPM compared to GENO-NZB(GDGSB) even though the same SPM forcing is imposed is applied.

With respect to the other model equations and parameters, differences with GENO-NZB(GDGSB) are minute. The set-up of the phytoplankton model of CSM(CDGSB) is the same as in GENO-NZB (GDGSB).

#### 2.4.4. COAST(KSKCB)

COAST(KSKCB) is the first model application using a curve-linear, fine resolution grid along the Dutch North Sea coast. This coastline-following grid allows transport to follow the coastal contours (De Kok et al., 2001). Hydrodynamic simulations were performed in 3D mode. The original 3D hydrodynamic model grid was aggregated horizontally and vertically resulting in a total number of 2153 computational elements used for the ecological simulations in 2D (Fig. 1). Seasonal variations in flow conditions are not explicitly modelled, but mimicked by correcting the dispersive flow rates of a single representative daily flow as a function of the 30 year averaged historic wind direction and speed. The horizontal dispersion is relatively small when the average wind is from the south west and largest for winds from the north west (March and April). This application, originally called 'North Sea Bloom', was extensively calibrated to obtain the optimal set of model coefficients given the objective of the modelling and the application area (Los and Bokhorst, 1997; De Vries et al., 1998; Blauw et al., 2009). It was also used for many management evaluations (e.g. Boon and Bokhorst, 1995; Los and Bokhorst, 1997; Villars and de Vries, 1998).

In comparison to all other applications presented here, the boundaries of COAST(KSKCB) are situated much closer to the continental coast. The model domain has open boundaries to the south, west and north (see Fig. 1). Since the prevailing transport direction along the Dutch coast is from south west to north east, the southern boundary is the most important. For the original model application concentration values for all substances were adopted from the long year averaged measurements at the nearby Appelzak transect (Fig. 2) between 1975 and 1985. The interannual variability at this transect was rather small, indicating that this boundary is not strongly affected by river loads or meteorological variations. Still the location of the boundaries makes the model more susceptible to the boundary conditions than any of the other models.

Data were interpolated with respect to space and time. Because chlorophyll is not a state variable of the model, we have transformed its measured concentrations into various phytoplankton types using an assumed time variable function for the annual pattern and the same C/Chl ratios adopted in the model. Likewise total nutrients were not imposed on the model but rather computed as the sum of the individual components. In the simulation results presented here, the  $\text{PO}_4$  concentrations at this boundary were reduced by 35% to account for the about 50%  $\text{PO}_4$  reduction of river discharges, which has been achieved since the mid 1980s. Using this correction factor simulated winter  $\text{PO}_4$  values along the Dutch coast agreed well with the observations for 2003 (Rijkswaterstaat, 2003).

Values for computational boundary elements in which no monitoring stations are located, were obtained by fitting an exponential curve through the observations. Little variations are observed along the western boundary. Therefore concentration values here were obtained by simply taking the average of all observations at all stations 70 km off the coast. Concentrations at the northern boundary were obtained from measurements at the Rottum transect. Notice that this is by far the least important boundary due to the



prevailing direction of the currents. Substances which are not measured directly (i.e. detritus) were computed according to same procedure that was used for the boundaries of the other models (see Los and Wijsman, 2007 for more details).

For the application of the COAST(KSKCB) model, a new SPM field was generated (De Kok et al., 1995). In general the agreement with observed SPM values has improved, although this is not the case at all stations. To distinguish between differences due to changes in hydrodynamics and those due to changes in SPM an additional simulation was performed with the COAST(KSKCB) model using the SPM field from GENO-NZB(GDGSB) (see Table 1). Time series plots of these additional runs are not shown in this paper, but the overall results are included in the Goodness of Fit scores (GOFs) presented below.

In the coastal zone coloured dissolved organic matter (CDOM) from fluvial sources contributes significantly to the attenuation of light. To first order, the CDOM concentration can be approximated as a linear function of the fresh water content (Peeters et al., 1991; Los and Bokhorst, 1997; Van Gils and Tatman, 2003; Los and Wijsman, 2007). In COAST(KSKCB) therefore a salinity-dependent term was added to the extinction model, resulting in improved simulation results of the extinction coefficient (see Los and Bokhorst, 1997 for more details). With respect to all other model equations and coefficient values, COAST(KSKCB) and CSM(CDGSB) are identical so differences in results should mainly be attributed to differences in the hydrodynamic model applications.

#### 2.4.5. ZUNO-2D(ZNZCB)

The ZUNO (*Zuidelijke Noordzee*, i.e. Southern North Sea) grid was developed as a follow-up of both the CSM and COAST grid. This is a curve-linear grid with a moderately high resolution in the Dutch coastal zone of ca.  $2 \times 2$  km and a lower resolution of up to  $20 \times 20$  km in the most north westerly part of the domain. The version of ZUNO presented here consists of 4350 active elements horizontally and 10 vertical sigma layers. The transport fields applied in this paper all stem from 3D hydrodynamic simulations using the Delft3D flow code (Lesser et al., 2004; WL | Delft Hydraulics, 2005). All hydrodynamic simulations of ZUNO have been done in 3D mode, either for a typical spring–neap cycle with a characteristic, variable wind forcing or with historic forcing for 2003. This includes air temperature, pressure, irradiance, wind speed and direction. Depending on the specific question to be addressed, the BLOOM/GEM transport model is either run in 2D or in 3D mode. The 2D transport fields are obtained by vertically averaging the results of the original 3D Delft3D flow simulation. Unless stated otherwise ZUNO-2D(ZNZCB) simulation results presented in this paper are based on a single spring–neap cycle forcing. Besides, additional results are shown in which the climatological forcing of the 2D model is the same as for the 3D results shown here. This model will be denoted by ZUNO-2DR(ZRZCB).

The Channel boundary of ZUNO-2D(ZNZCB) is the same as for GENO and CSM. Although concentration values at the north Atlantic boundary have recently been updated based on an extensive literature study (Blauw et al., 2006), these new boundary conditions have not been used during the simulations reported here for the sake of comparison to the older models. Hence, the same northern boundary was used for all simulations by all models (except COAST(KSKCB)).

The SPM forcing of ZUNO is based on simulation results of an improved steady state model (WL | Delft Hydraulics/MARE, 2001). Simulations were performed on the  $4 \times 4$  refined version of the ZUNO grid in 3D mode and projected on the coarser grid used here. In comparison to the previously used SPM fields, locally the agreement with the measurements has improved. This is particularly true in the Dutch Coastal zone due to the refined resolution. As in the other models the overall seasonal variation is simulated by means of a harmonic function with relative high values in winter and low values

in summer. However, using this function, the spring bloom in the model tends to be rather late because in reality short periods of quiet conditions with relatively low levels of suspended matter play an important role in triggering the onset of the spring bloom. To account for these short-term variations, we have assumed a relationship between the SPM concentration and the average wind speed to further adjust the harmonic signal. The amplitude of this short-term fluctuation is a multiplication factor varying between 0.3 and 1.7 depending on the difference between the actual and average wind speed ( $5.5 \text{ m s}^{-1}$ ). Again, these factors were determined empirically in such a way that the observed interannual variability could be reproduced sufficiently well (see Los et al., 2008 for more details). To distinguish between effects of changes in resolution and in SPM field, an additional simulation was also performed by which the ZUNO 2D model was forced by the same SPM field as the two GENO models and CSM(CDGSB). This simulation is denoted by ZUNO-2D(ZRZCB).

As for the previously discussed models, sea water temperature in the default ZUNO-2D(ZNZCB) applications is derived from measurements from station Noordwijk 10. In the simulations using actual transport fields (i.e. ZUNO-2DR(ZRZCB)), however, the temperatures are adopted from the Delft3D-Flow hydrodynamic model (spatially and temporarily varying). The model set-up is basically the same as in COAST(KSKCB), but some model parameters were modified according to recent insights. Some coefficients of the light model were adjusted based on an extensive data analysis (van Gils and Tatman 2003). Several phytoplankton related parameters of the BLOOM module were updated to accommodate new experimental results on the functional groups in the model. Particularly for nutrient-stressed species the stoichiometric ratios of the model were modified (less nutrients per unit of biomass) (Riegman unpublished results). The optimum light intensity of several species was reduced (Jahnke, 1989; Ferris and Christian, 1991; Garcia and Purdie, 1992) meaning enhanced growth rates at low light intensities.

#### 2.4.6. ZUNO-3D(Z3ZCB)

The main difference between ZUNO-3D(Z3ZCB) and ZUNO-2D(ZNZCB) is the vertical resolution. Furthermore historic atmospheric data force the transport in ZUNO-3D(Z3ZCB), while ZUNO-2D(ZNZCB) is forced by an average spring–neap cycle in the simulations reported here. Because the 3D model takes stratification into account, temperature adopted from the Delft3D Flow hydrodynamic model is specified as a 3D, time variable forcing to BLOOM/GEM. Other forcings and model parameters are the same as in ZUNO-2D(ZNZCB). To investigate the impact of advances in SPM modelling, an additional simulation (ZUNO-3D(Z3VCB)) was performed using results from a dynamic, fine resolution sediment transport model (Van Kessel et al., 2008).

### 2.5. Monitoring program

Halfway through the 1970s the Dutch national government initiated an extensive monitoring network covering all national waters including the North Sea which was rather unique at that time. The network covers sampling stations in river branches, estuaries and marine waters. For the present study only the marine stations are considered. These stations are visited by survey vessels every 2 to 4 weeks that collect water samples from the surface layer (nominally 1 m below sea surface level). Visits to the stations are randomly timed with respect to the tidal phase but are always carried out under relatively calm weather conditions (wind strength less than 7 Bft).

Unfortunately, the original network was stepwise minimized rather than maintained or extended in later years. Hence, for 2003, 12 variables at 17 stations have been considered for direct comparison with model results (Rijkswaterstaat, 2003). Besides, all relatively recent (less than 10 years old) data that were available for about a dozen additional stations have been used as supporting evidence to

evaluate the performance of the various model versions. An overview of the locations is shown in Fig. 2. The relevant substances for the present paper are listed in Table 2.

The list of monitored substances has not changed much over time, but there are two important exceptions within the context of this study: the vertical extinction and the species composition. Firstly, the vertical extinction coefficient has been measured since the beginning of the 1990s at most of the stations, which are still in the 2003 monitoring program (Fig. 2). Secondly, a complete data set on species composition data has become available only fairly recently (see e.g. Baretta-Bekker et al., 2009). As a consequence, only a limited amount of quantitative information on species was available during the development of the model applications and model performance in terms of species composition has up to now not been formally validated. Historically, the application development relied on expert knowledge and literature. For the Southern North Sea it is for example known that diatoms and *Phaeocystis* are usually dominant during the spring. Dinoflagellates dominate late summer and micro-flagellates are present in the summer half year, particularly under P-limiting conditions. Chlorophyll-*a* is supposed to be a useful proxy for the total biomass after conversion to units of carbon. The models have been set-up to reproduce this general pattern. Based on cell counts at about 15 monitoring stations, biovolume estimates are now available for every year since 1991 on a monthly basis. These biovolumes are converted to biomass estimates in  $\text{mg C l}^{-1}$  (Menden-Deuer and Lessard, 2000). A detailed description and trend analyses of these data are given by Baretta-Bekker et al. (2009). For future model development these species data can be used in addition to the general water quality and chlorophyll-*a* data. In general these data are in agreement with the previously used expert knowledge with the exception of the abundance of dinoflagellates, which according to the observations are mainly regionally confined to the offshore regions.

## 2.6. Calibration procedure original models

For the set-up and calibration of the original models, the following procedure has been followed in general. First, salinity and tracer simulations were performed to check the main transport characteristics. Next calculated concentrations of chlorophyll-*a* and nutrients have been compared with measurements in the following manner:

(1) The most important measure for phytoplankton biomass chlorophyll-*a* was compared graphically with measurements. Usually this was done using observations for several years plotted as a single set of data points. (2) The analysis of limiting factors and phytoplankton species and types was made. For phytoplankton only a qualitative analysis was possible. (3) The calculated dissolved nutrients were compared graphically with measurements. In this comparison it is important to know if a nutrient is (sometimes) limiting or not (see step 2). (4) The calculated total nutrients were compared graphically with measurements. In case of discrepancies with measurements, the comparison of individual terms (phytoplankton and dissolved species) also had to be reconsidered. (5) The calculated light extinction was compared graphically with measurements (only for later model versions as little data were available when the oldest models were calibrated).

## 2.7. Validation and model intercomparison: Goodness of Fit criteria

Apart from calibration, which, as outlined above, tended to focus on reproducing the multi-year mean, spatial and seasonal patterns, past validations have been carried out for particular years. Typical validation years were 1985 (GENO models), 1990 (CSM(CDGSB) and COAST(KSKCB)) and 1988–1989 (ZUNO-2D(ZNZCB) and ZUNO-3D(Z3ZCB)). In this paper, we will discuss the various model applications by comparing their results to the *in situ* monitoring data in the Dutch coastal zone by Rijkswaterstaat for 2003 (Rijkswaterstaat, 2003).

Traditionally, the validation consisted of visual inspection of time series output and spatial maps of model results. Over time, however, more formal, quantitative validation methods gained attention. This became more common practice in particular due to the 1996 OSPAR ASMO Eutrophication modelling workshop (Villars and de Vries, 1998). Since then, the so-called OSPAR cost function has been applied to quantify the performance or skill of coastal biogeochemical models, see e.g. Los et al. (2008) and Blauw et al. (2009) for a discussion on the BLOOM-GEM model, but also Radach and Moll (2006).

The OSPAR cost function  $CF$  is one of the options. It is in fact the normalized mean absolute error (MAE) between model and observation, defined as

$$CF = \frac{1}{N} \sum_{n=1}^N \frac{|M_n - D_n|}{\sigma_D} \quad (2)$$

where the average over discrete time and space  $n$  spans an annual interval and may be determined per location or span the entire region of interest.  $D_n$  is the observation at each individual location and time and  $M_n$  is the model output at its matching model grid cell;  $\sigma_D$  is the annual standard deviation of the observational data. The normalization has been chosen to express the goodness of fit in terms of multiples of the standard deviation, with  $CF < 1$  being classified as 'very good',  $1 < CF < 2$  as 'good',  $2 < CF < 3$  as 'reasonable' and any  $CF$  beyond this upper limit as 'poor'. Note that the upper limit of 3 is chosen according to Radach and Moll (2006), whereas Villars and the Vries adopted a limit of 5 to separate 'reasonable' from 'poor'.

### 2.7.1. Selection of GOF score criterion

As a first step in the analysis, cost function results have been computed for a number of substances and stations. Fig. 3 shows a typical example for chlorophyll-*a* at a number of representative stations. Based on the criteria proposed by Radach and Moll (2006), almost all model results at almost all stations could be qualified as 'very good' or 'good'. Moreover, scores for different models per station are often rather similar. In contrast, if we plot the model results against the measurements, there are sometimes clear and consistent differences between individual models which we think should be reflected by the GOF scores. Clearly, the OSPAR Cost Function has little

**Table 2**  
Monitoring information on substances used in this paper.

|                       | Substance  | Unit                | Remark  |
|-----------------------|--|---------------------|---|
| Chlorophyll- <i>a</i> | Chlorophyll- <i>a</i>  | $\text{mg m}^{-3}$  |   |
| NO <sub>3</sub>       | Nitrate NO <sub>3</sub> <sup>-</sup>                           | $\text{gN m}^{-3}$  | NO <sub>2</sub> is measured separately but ignored  |
| PO <sub>4</sub>       | Ortho phosphate PO <sub>4</sub> <sup>-</sup>                   | $\text{gP m}^{-3}$  |   |
| TotN                  | Total dissolved Nitrogen                                       | $\text{gN m}^{-3}$  | NO <sub>2</sub> + NO <sub>3</sub> + NH <sub>4</sub> + organic N<br>No refractory N in model |
| TotP                  | Total dissolved Phosphorus                                     | $\text{gP m}^{-3}$  | PO <sub>4</sub> + particulate organic and inorganic. No refractory P in model               |
| SiO <sub>2</sub>      | Dissolved silica SiO <sub>2</sub>                              | $\text{gSi m}^{-3}$ | After filtration  |
| SPM                   | Suspended particulate matter                                   | $\text{g m}^{-3}$   | Filter residue, anorganic   |
| Ext                   | Extinction coefficient of visible light                        | $\text{m}^{-1}$     |   |
| <i>Phaeocystis</i>    | Biomass of <i>Phaeocystis globosa</i>                          | $\text{gC m}^{-3}$  | Cell counts converted to biovolume converted to gC  |
| Diatoms               | Biomass of all diatom species                                  | $\text{gC m}^{-3}$  | Cell counts converted to biovolume converted to gC  |
| Micro-flagellates     | Biomass of all pico phytoplankton and micro flagellate species | $\text{gC m}^{-3}$  | Cell counts converted to biovolume converted to gC  |
| Dinoflagellates       | Biomass of all dino flagellate species                         | $\text{gC m}^{-3}$  | Cell counts converted to biovolume converted to gC  |



distinctive power. For scrutinous model performance evaluation the OSPAR Cost Function seems hardly suitable.

The issue of defining appropriate measures to quantify model skill and aid model intercomparison has received increasing attention in recent publications. Following the atmospheric modelling community (e.g. Taylor, 2001), the biogeochemical modelling community is entering a phase of growing need and possibility of quantifying model skill due to the increasing interest in operational model applications and the growing amount of observational data. Allen et al. (in press), pointed out the issue of matching the traditional visual inspection of time series to skill measures and explores a range of measures. In subsequent studies by Jolliff et al. (2008), Friedrichs et al. (2009), and Stow et al. (2008), various measures have been applied more extensively. From these, we adopted the *target diagram* as presented by Jolliff et al. (2008) as it turned out to best convey the message otherwise extracted from the comparison of model and observation time series.

The target diagram displays the difference between data sets in terms of both the bias  $B$  and the unbiased root-mean-squared difference  $RMSD'$ :

$$B = \frac{1}{N} \sum_{n=1}^N (M_n - D_n) = \bar{M} - \bar{D} \quad (3)$$

$$RMSD' = \left[ \overline{(M'_n - D'_n)^2} \right]^{1/2} \quad (4)$$

where the overbar denotes averaging in the same sense as in Eq. (1) and primes denote the residual, e.g. for  $M'_n$ :

$$M'_n = M_n - \bar{M}. \quad (5)$$

The bias  $B$  is indicative of the match between model and observations in the annual mean sense, the unbiased root-mean-squared difference  $RMSD'$  is a measure for the match between the residuals of both time series after removal of the bias. As discussed by Jolliff et al. (2008) in more detail,  $RMSD'$  is a measure of the overall agreement in both phase and amplitude of the variability of the compared time series.

Upon plotting both measures in a single graph with  $B$  and  $RMSD'$  as the Cartesian coordinates, the total root-mean-square difference ( $RMSD$ ) is indicated by the radial distance from the origin, because by definition, the following holds for the total  $RMSD$ :

$$RMSD^2 = B^2 + RMSD'^2. \quad (6)$$

Following Jolliff et al. (2008), we adopted the convention to normalize  $B$  and  $RMSD'$  by the standard deviation of the observations  $\sigma_D$  and to multiply  $RMSD'$  by the sign of the standard deviation difference  $\text{sgn}(\sigma_M - \sigma_D)$ . Due to the normalization, both  $B$  and  $RMSD'$  are non-dimensional and readily interpreted with respect to the variability within a given signal. The use of the signum ( $\text{sgn}$ ) function adds to  $RMSD'$  information on over- or underprediction of the observed variability as  $RMSD'$  itself is positive definite.

In summary, the normalized bias will be shown on the ordinate of the target diagram:

$$B^* = \frac{\bar{M} - \bar{D}}{\sigma_D}. \quad (7)$$

And the normalized, signed, unbiased root-mean-square difference on the abscissa:

$$RMSD'^* = \frac{\text{sgn}(\sigma_M - \sigma_D)}{\sigma_D} \left[ \overline{(M'_n - D'_n)^2} \right]^{1/2}. \quad (8)$$

Thus, in such a target diagram, a circle with radius 1 on the target diagram corresponds to a total normalized  $RMSD$  equal to the standard deviation of the observations. A model that would merely reproduce the annual mean of the observations would score  $RMSD'^* = 1$  and zero bias. Hence, any model result outside the circle with total normalized, signed, root-mean-square difference  $RMSD'^* = 1$  can be considered as poor. Moreover, any result on the target diagram with total  $RMSD'^* < 1$ , cannot be negatively correlated to the observations (see Jolliff et al., 2008, for details).

As discussed by Taylor (2001), and reiterated by Jolliff et al. (2008), there is a relation between the linear correlation coefficient  $R$  and the unbiased root-mean-square difference which is helpful in defining additional Goodness of Fit criteria. Because  $RMSD'^*$  is related to the correlation coefficient  $R$  by

$$RMSD'^* = \text{sgn}(\sigma_M - \sigma_D) \sqrt{1 + \sigma^{*2} - 2\sigma^*R} \quad (9)$$

where  $\sigma^* = \sigma_M / \sigma_D$  and  $R = \overline{M'D'} / \sigma_M \sigma_D$ ,  $RMSD'^*$  attains a minimum when  $\sigma^* = R$ . Since this is also the minimum total  $RMSD'^*$ , a circle with radius  $M_{R_0} = \sqrt{1 - R_0^2}$  denotes the minimum total  $RMSD'^*$  possible for a given  $R_0$  and all points between this circle and the origin correspond to  $R > R_0$ . This leads us to the definition of a second, relative criterion. Since  $R_0 = 0.67$  (i.e.  $M_{R_0} = 0.74$ ) matches the 15% percentile of all

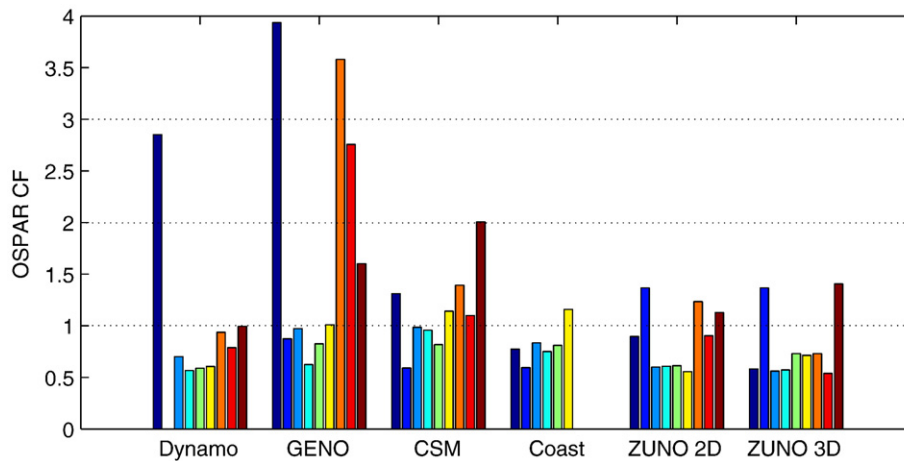


Fig. 3. OSPAR Costfunction (CF, see Eq. (1)) results all 6 base 2003 model cases at 9 representative monitoring stations for chlorophyll- $a$  using the observational data of 2003. Bars correspond to the scores for the individual stations (from left to right: Walcheren 2; Noordwijk 2, 10, 20, 70; Terschelling 10, 135, 175, 235). Not all stations are represented in all applications. Dashed lines indicate the class limits (see text).

model  $R$ -scores for chlorophyll- $a$  across all stations, we identify model results with  $RMSD^* < M_{ro}$  as 'good'. Any score with  $M_{ro} < RMSD^* < 1$  we refer to as 'reasonable'. The choice for chlorophyll- $a$  is motivated by the fact that this is the main objective variable for which the models have been developed. For the other variables the same classification has been adopted.

When discussing unbiased RMS differences ( $RMSD^*$ ) it should be kept in mind that when the model underestimates the variance ( $\sigma_M < \sigma_D$ ), i.e.  $\sigma^* < 0$ , and  $R$  is usually  $0 < R < 1$ , a decrease in  $RMSD^*$  may result (see also Eq. (9)). When correlation is less than unity, further underestimating the variance or reducing the correlation may improve the unbiased RMSD score and inadvertently lead to conclusions of better skill. However, in the analyses below this risk has been taken into consideration as the time series of the individual stations as well as other skill-metrics, i.e. the ratio of standard deviations and correlation coefficients, have been examined.

## 2.8. Illustration

Fig. 4 below illustrates the use of the Target Diagram for a particular station (*Walcheren 02*) for chlorophyll- $a$  in comparison to a time series plot. Evaluating the time series in Fig. 4A leads to the impression that for example the COAST model applications do not capture the temporal pattern which leads to overprediction in summer and autumn. On the other hand, GENO-DYNAMO (GDGSD, ●) and GENO-NZB (GDGSB, ○) show an underprediction of the signal and also a mismatch in timing. Relatively speaking, the 3D ZUNO models (★, ☆) appear to perform best, as they exhibit a distinct spring autumn blooms, albeit that the timing could still be improved.

These general statements are reflected in Fig. 4B. Clearly COAST (KSGCB) (□, with SPM from GENO) is an outlier and performs poorest in relative sense. It is the only result with  $RMSD^* > 0$ , i.e. where the model standard deviation exceeds the standard deviation of observations. Both the normalized bias and unbiased RMSD are larger than one. The bias of the base model COAST (KSKCB, ■) is remarkably smaller, but it is clear that the capture of variations is still poor. COAST (KSKCB) in this case is more or less comparable to a model merely describing the annual mean. The DYNAMO and GENO-NZB models (●, ○) also perform poorly and exhibit a relatively large negative bias (underprediction). The 3D ZUNO model (Z3VCB, ★) performs best, whereas ZUNO (Z3ZCB, ☆) suffers from a larger underprediction of the mean and of the variability.

As remarked by Jolliff et al. (2008), the choice for the current Target Diagram has the drawback that it focuses on reduction of RMS differences and is less scrutinous on phase match as expressed by the correlation coefficient  $R$ . Our choice is reflecting the ambition of the current paper in relation to the nature of the reference data set available. The present aim is to discuss model performance in terms of annual mean bias, capture of seasonality and capture of anomalies in seasonality with a timescale of several weeks. Capture of shorter-term, more localized variability would be a next step. Preliminary comparison of the low-frequency ship data to relatively recent high-frequency buoy data (e.g. Blaas and Van den Boogaard, 2006) indicates that an important part of the variability spectrum (i.e. on the tidal to weekly timescale) is unresolved and hence leads to inherent uncertainty in the low-frequency observations used here. Hence, there is a limit on the significance of skill improvement in the presented target diagrams within the innermost circle. It is beyond the scope of the current paper to address this in detail, but this will be considered in a follow-up studies.

## 3. Results

This series of model applications is intended to compare the model versions with each other and with the observations. First the overall model performance is assessed based on the Target Diagrams as a

formal goodness of fit criterion using clustered results for all monitoring stations sampled in 2003 (see also Fig. 2). Notice that by considering all stations simultaneously, discrepancies in model behaviour might be obscured for instance due to compensating errors at different locations. Findings will be illustrated by some graphs for one representative monitoring location. Next, time series simulation results are presented for some typical stations in order to demonstrate the importance of various factors i.e. resolution, historic forcing, SPM forcing etc.

### 3.1. Overall model performance

#### 3.1.1. Chlorophyll- $a$ and species composition

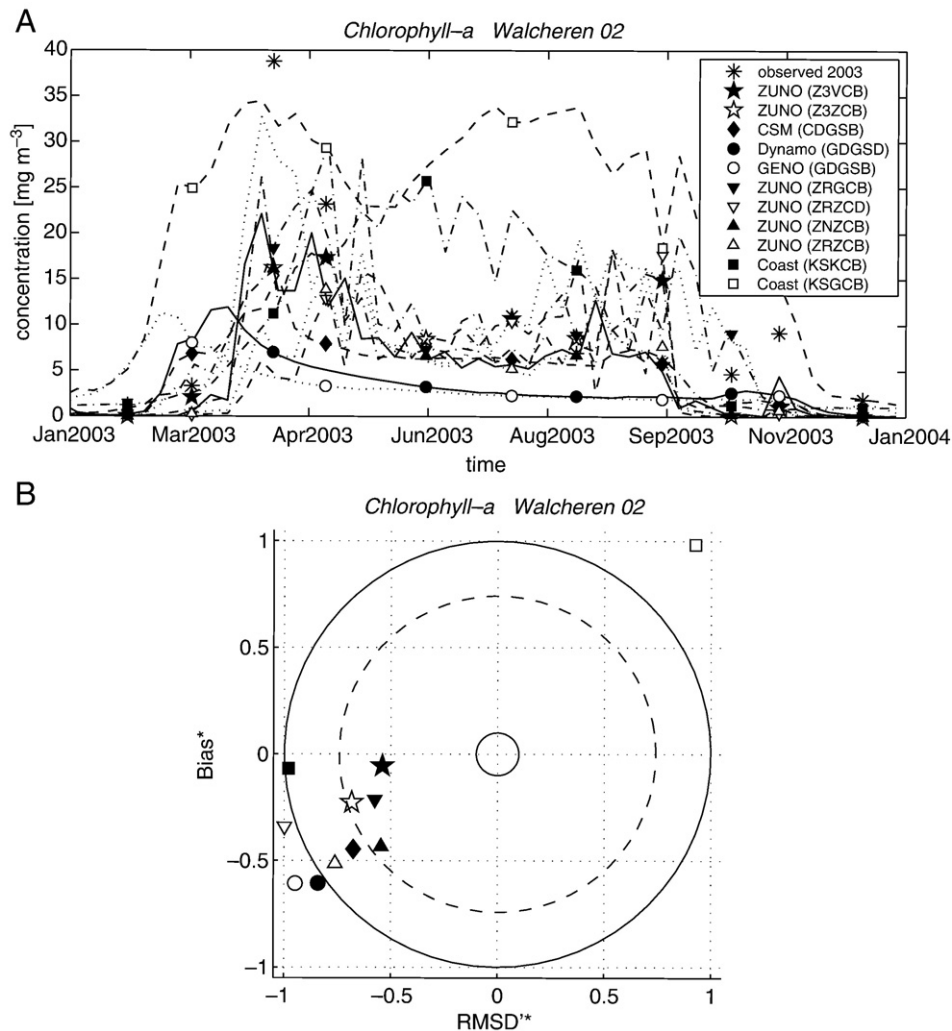
As explained in Section 2.6 during the calibration of all models an attempt was made to optimize the overall result for chlorophyll- $a$  at a number of representative stations. Fig. 5 shows the results for the present application of the models. Indeed the overall score for all models shows little variation. Almost all models have a reasonable overall RMSD score and the differences within are rather small. The typical bias  $B^*$  of the models is between  $-0.2$  and  $0.2$ , which is much smaller than the unbiased  $RMSD^*$  which is typically between  $-1$  and  $+1$ . In other words: in all models the annual average is close to the measurement, but the level of variation shows stronger discrepancies and is typically smaller than observed. This result is in agreement with the method of calibration, which focused on reproducing seasonally averaged results. Notice that even the overall RMSD score of the oldest, least advanced DYNAMO (GDGSD, ●) model is similar to the score of GENO-NZB (GDGSB, ○) and in fact does not differ much from the score of the latest ZUNO-3D (Z3ZCB, ☆). It should be pointed out, however, that the advantage of the 12 type BLOOM module is most obvious when the spatial and temporal gradients are relatively large. Hence according to Fig. 5 comparing the results of two model runs on the more refined ZUNO grid, the score for the 12 type phytoplankton version (ZRZCB, △) is clearly better than the score for the 2 species simulation (ZRZCD, ▽). Moreover for many individual stations, the overall RMSD score of chlorophyll- $a$  did improve in later model versions. We conclude that the overall ability of the model to simulate spatial and temporal chlorophyll- $a$  gradients improves if the more complex BLOOM module is adopted, but using BLOOM is not necessary in cases where only spatial averages or annual means are considered.

The singular position of the ZUNO-3D (Z3VCB, ★) model in the target diagram indicates that the results are sensitive to the forcing by SPM: its overall RMSD is comparable to its companion ZUNO-3D (Z3ZCB), but its variability exceeds the observed.

For diatoms (Fig. 6A) the magnitude of the bias between model results and biomass estimates from cell counts is typically between 0 and 1. The ability of the models to reproduce the variability of the data is considerably smaller. The absolute value of all  $RMSD^*$  are larger than 1. There are two clusters with models. The  $RMSD^*$  score is negative for all models which are forced with the SPM field from GENO and positive for all other models. So the SPM field, hence the light climate, is a crucial factor for the simulation of diatoms. Resolution and forcing of transport are less important considering the clustering of models.

The biases for *Phaeocystis* (Fig. 6B) are usually less than 0.5, which is considerably smaller than for diatoms. The  $RMSD^*$  scores are also smaller. The differences within the model results are rather small; interestingly enough the oldest and coarsest GENO-NZB (GDGSB, ○) model has the best overall  $RMSD^*$  score.

Biases for flagellates (Fig. 6C) are usually positive and range from 0 to about 2, moreover all  $RMSD^*$  scores are larger than 2. So there is a tendency to overpredict the overall biomasses and the level of variation produced by the models by far exceeds the level of variation of the data.



**Fig. 4.** (A) Time series of chlorophyll-*a* at station Walcheren O2 for all model runs discussed in this paper. Lines are shown to aid distinguishing the time series; they connect data at various intervals (models weekly, observations at least 2-weekly), but ignore actual shorter-term variability. (B) Corresponding target diagram, showing the normalized bias and signed, normalized, unbiased root-mean square difference of the model results with respect to the observations. Results within the drawn circle with overall  $RMSD^* = 1$  score at least 'reasonable', results within the dashed circle  $RMSD^* = M_{80} = 0.74$ , score 'good' (see text for further details).

In the case of dinoflagellates (Fig. 6D) all biases are positive and always greater than 0.5. The  $RMSD^*$  scores are positive and greater than 1.5, indicating that the models strongly overestimate the level of variation for this group.

So while the overall scores for diatoms and *Phaeocystis* could still be called reasonable or good, the score for flagellates and dinoflagellates are poor. This means that for all models the overall scores for chlorophyll-*a* are much better than for individual species groups. This issue will later be addressed in the discussion of this paper.

### 3.1.2. Primary production

The concentration of chlorophyll-*a* at any place and time is the result of local processes such as primary production, respiration etc. and transport of phytoplankton. As shown in the previous sections, gradients in relevant forcing conditions for primary production such as the turbidity and the available levels of nutrients are more accurately reproduced in models with a relatively fine grid. So one may wonder if the distribution of the rate of primary production is as insensitive to the resolution as chlorophyll-*a* is. To investigate this the annual average spatial distributions of chlorophyll-*a* and the primary production are shown for two 2D models in Fig. 7: GENO(GDGSB) and ZUNO-2D (ZRZCB). In accordance with the results from the target plots, the average concentration patterns of chlorophyll-*a* are remarkably similar considering the differences between the two models. In contrast, results

for primary production are quite distinct. For the GENO-NZB model the spatial distribution resembles the chlorophyll-*a* pattern. But primary production in ZUNO-2D shows much more spatial variation; for instance production is relatively high in parts of the Dutch coastal zone, but not in the first few kilometres, and in the central North Sea (Dogger Bank region). In other areas primary production rates are lower than in GENO(GDGSB).

These results show that indeed the spatial distribution of the simulated primary production is much more sensitive to the spatial resolution of the model than the chlorophyll-*a* pattern, which due to horizontal transport is strongly flattened out in the fine resolution model.

### 3.1.3. Total nutrients

Often a comparison between the simulated and observed salinity is adopted to demonstrate the level of accuracy of the transport modelling. However, because in some of the older models salinity is not prognostically computed, we choose to show total nitrogen (TotN) and total phosphorus (TotP), which are included in all models, as proxies.

According to Fig. 8A, the models mostly differ with respect to the bias of TotN. The  $RMSD^*$  scores differ less, and for this substance correlate positively with the bias. ZUNO-3D(Z3ZCB) and COAST (KSKCB) have the best overall score. Next come ZUNO-2D (all variants)



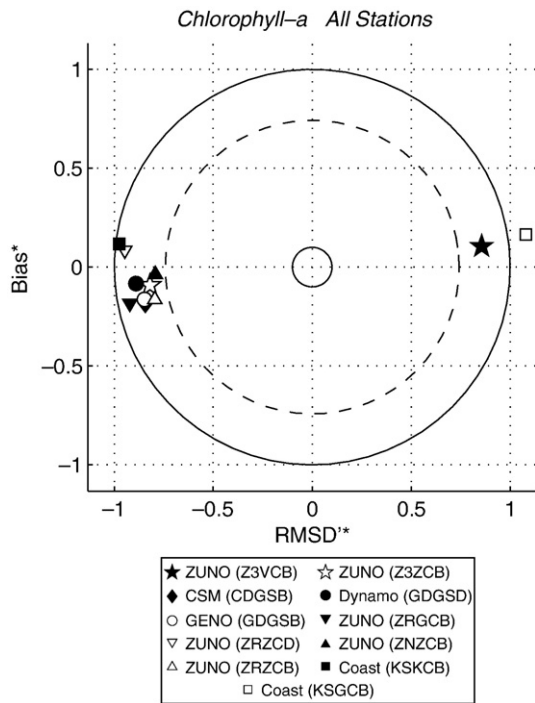


Fig. 5. Target diagram for chlorophyll-a, all stations in all applications.

and CSM(CDGSB) while the two GENO based models (GDGSD, and GDGSB) have the lowest score, and classify as ‘poor’ according to the criteria defined.

Fig. 8B shows the target plot for TotP. In comparison to TotN, the scores of individual models vary less. There is a negative bias for all models ranging from  $-0.8$  for GENO-NZB(GDGSB) to less than  $-0.5$  for COAST(KSKCB) and most of the ZUNO based models. This negative bias might be caused by an underestimation of the external P-inputs (loads and boundaries), or by an underestimation of the P release from the sediment. All models were rerun until they had achieved equilibrium for all state variables, but in reality sediments in the coastal regions might still contain significant amounts of reactive P stored during the 1970s and 1980s, when the external loads were considerably higher than in 2003. In comparison to TotN, the  $RMSD^{**}$  scores are larger and all of them are negative, indicating that there is less variation in all of these models in comparison to the data. The overall score for the two 3D ZUNO models, the COAST models, and one of the 2D ZUNO(ZRZCD) models are reasonable.

The ranking of the models clearly reflects the level of detail in resolution of the grids so this seems to be a critical factor for the accuracy of the overall transport model. Whether the model is 2D or 3D or the nature of the imposed forcing does affect the results, but seems less critical on an annual basis when all stations are taken into account. For instance the score of COAST(KSKCB) with a daily forcing is on a par with the score of ZUNO-3D(Z3ZCB). Furthermore the scores of the ZUNO 2D model with historic forcing (ZUNO-2DR(ZRZCB)) and with a representative spring–neap cycle (ZUNO-2D(ZNZCB)) are also about the same.

Notice that the  $RMSD$  score of the total nutrients is better for later models, but this has not resulted in a similar improvement of the overall score for chlorophyll-a. The improvement of the nutrient scores is partly related to the change in resolution which results in a less diffuse description of cross-shore gradients. Effectively, progressing from the older coarser to the more recent finer models, the same mass of nutrients is redistributed such that near-shore stations have increased and offshore stations have been reduced by equal amounts of mass. However, the associated near-shore nutrient concentrations increase relatively more than the offshore concentrations decrease as

a result of the offshore increase in grid cell volume (due to both increase in depth and horizontal mesh size). Since chlorophyll concentrations generally have smaller cross-shore gradients, the resolution effect is smaller in terms of chlorophyll scores.

### 3.1.4. SPM and total extinction

Fig. 9A shows the target diagram for the SPM forcing of the models. On average all models overestimate SPM and the average bias was hardly reduced during development of the SPM models except for the latest ZUNO-3D(Z3VCB) model. This application uses results from a new, dynamic SPM model (Van Kessel et al., 2008), which yields an improvement in the overall skill. With the exception of COAST (KSKCB) the modelled residuals are larger than observed for all applications. Two additional remarks need to be made. First, high-frequency OBS measurements indicate that the level a variation and the average value in SPM is underestimated by traditional monitoring (Blaas et al., 2007). Second, the SPM scores of some models, which are forced by the same SPM field, are not exactly identical as they should be in theory, but the differences in SPM score are negligible (see the clustering on the far right of the diagram). This is caused by small imperfections in the procedures to project the SPM field from one grid onto another. The differences in response in terms of chlorophyll to the underwater light conditions are much stronger, though, because the simulation of the (total) extinction has improved more than simulation of SPM. The total extinction is one of the most complex outputs from the models because it is influenced by both model input (i.e. SPM), transport (CDOM) as well as by top-end output variables (i.e. the phytoplankton biomass). Moreover, the individual terms are negatively correlated i.e. an overestimation of the contribution of SPM and CDOM may be compensated by an underestimation of the amount of phytoplankton in case of light limitation.

The bias of the total extinction (Fig. 9B) has been reduced from 0.8 in the two GENO models to less than 0.1 in the latest ZUNO models. Two factors contribute to this improvement: (1) more recent models take the salinity as an approximation for dissolved organic matter (CDOM) into account and also have a prognostic salinity field and (2) values for parameters of the light module have been revised from a fairly recent statistical analysis of all newly available data up to 2003 (Van Gils and Tatman, 2003). The  $RMSD^{**}$  has also been reduced, but less so as the bias. In the latest models the level of variation exceeds the measurements. This is caused by an overestimation of the level of variation of SPM, which has already been mentioned in the previous section. Nevertheless, the overall score has improved from poor to good over time.

### 3.1.5. Dissolved nutrients

The relative location of the models in the target diagrams for  $NO_3$  and  $SiO_2$  (Fig. 10A and B) is quite similar, but both the biases and  $RMSD^{**}$  for most models are slightly smaller for  $NO_3$ . For  $NO_3$  the score of all models except GENO-NZB is good, for  $SiO_2$ , most scores are reasonable, or marginally good. For both variables the sign of  $RMSD^{**}$  is positive for the two ZUNO-3D and the two COAST models. The 3D model and the high-resolution 2D model generate more variability by nature. This is illustrated in Fig. 11, which shows the ratio of modelled over observed standard deviation of  $NO_3$  for all eleven applications. This typical pattern is found in all output concentrations, albeit less pronounced for chlorophyll-a than for the nutrients.

There is no notable difference between the 2D ZUNO model with spring–neap forcing (ZUNO-2D(ZNZCB)) and with actual forcing (ZUNO-2DR(ZRZCB)) so we may conclude that indeed the 3D transport phenomena cause the increase in variation of ZUNO-3D (Z3ZCB). Remarkably though, the  $RMSD^{**}$  for COAST(KSKCB) is similar to the one for ZUNO-3D(Z3ZCB). Most probably this is due to the seasonal correction of the horizontal dispersion, which was intended to mimic 3D phenomena. For  $SiO_2$  one of the ZUNO 2D models has a positive  $RMSD^{**}$  as well. This is due to a different result for diatoms

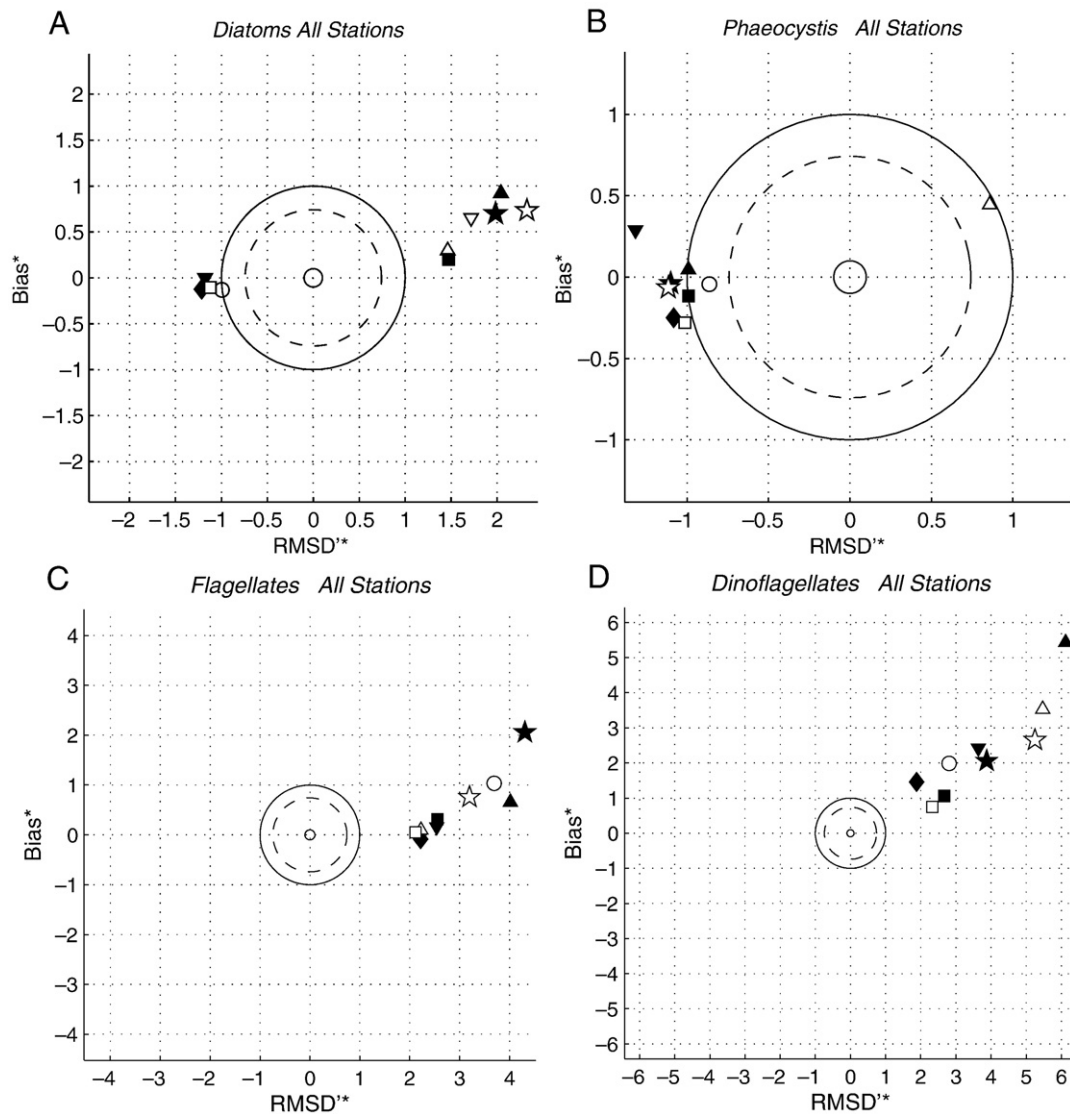


Fig. 6. Target diagram for biomass of diatoms (A), *Phaeocystis* (B), flagellates (C) and dinoflagellates (D), all stations in all applications.

caused by a different imposed SPM field which affects the development of the spring bloom.

For  $\text{PO}_4$  (Fig. 10C) the overall RMSD values are higher and only the GENO-NZB and DYNAMO models underpredict the variance. The relatively high overall RMSD values are predominantly due to relatively large unbiased RMS differences  $\text{RMSD}^*$ . It is known from all these models that they tend to exhibit a slower  $\text{PO}_4$  release from the sediment at the end of summer due to a lack of detail in the level at which inorganic processes are taken into account. Consequently simulated  $\text{PO}_4$  concentrations are often too low during summer in shallow areas (less than 20 m). Since many of the monitoring stations, for which the scores are computed, fall into this category, most of the models exhibit a poorer score for  $\text{PO}_4$  relative to those for  $\text{NO}_3$  and  $\text{SiO}_2$ . (see also Los et al., 2008 for more details). It is remarkable that GENO-DYNAMO(GDGSD) has the best overall score for  $\text{PO}_4$  while its score for TotP is the worst. A detailed analysis of the results suggests the following mechanism. In GENO-DYNAMO(GDGSD)  $\text{NO}_3$  and  $\text{SiO}_2$  rather than  $\text{PO}_4$  are limiting at many monitoring locations in summer, which is not in accordance with the measurements and results in a relatively large negative score for the bias of these two nutrients in this model (Fig. 10A and B). Consequently  $\text{PO}_4$  is not completely exhausted in this model by phytoplankton uptake during summer, resulting in a better score for this nutrient.

In spite of the obvious differences between the models, all scores for  $\text{NO}_3$  are good, for  $\text{SiO}_2$  one half ranks as good, the other half as reasonable and for  $\text{PO}_4$  only GENO-DYNAMO(GDGSD) marginally passes the good criterion, while all others are reasonable or marginally poor.

### 3.2. Overall ranking

In conclusion of this section it is desirable to summarize the overall performance of the eleven model applications for the key variables discussed above. When summarizing, we abandon the relative nuances that the target diagrams offer in terms of distinction between bias and pattern match, both in magnitude and sign. Since the general classification in the discussions above focuses on overall RMS difference, i.e. the distance of a particular result to the origin of the target diagram, we average the normalized overall RMS differences for the various variables, across all stations, in a summary diagram shown in Fig. 12. Obviously, the included parameters are not independent and also the weight that one would assign to the individual variables might depend on the objectives. In Fig. 12 all variables and stations have received equal weight, since any other choice would have been subjective.

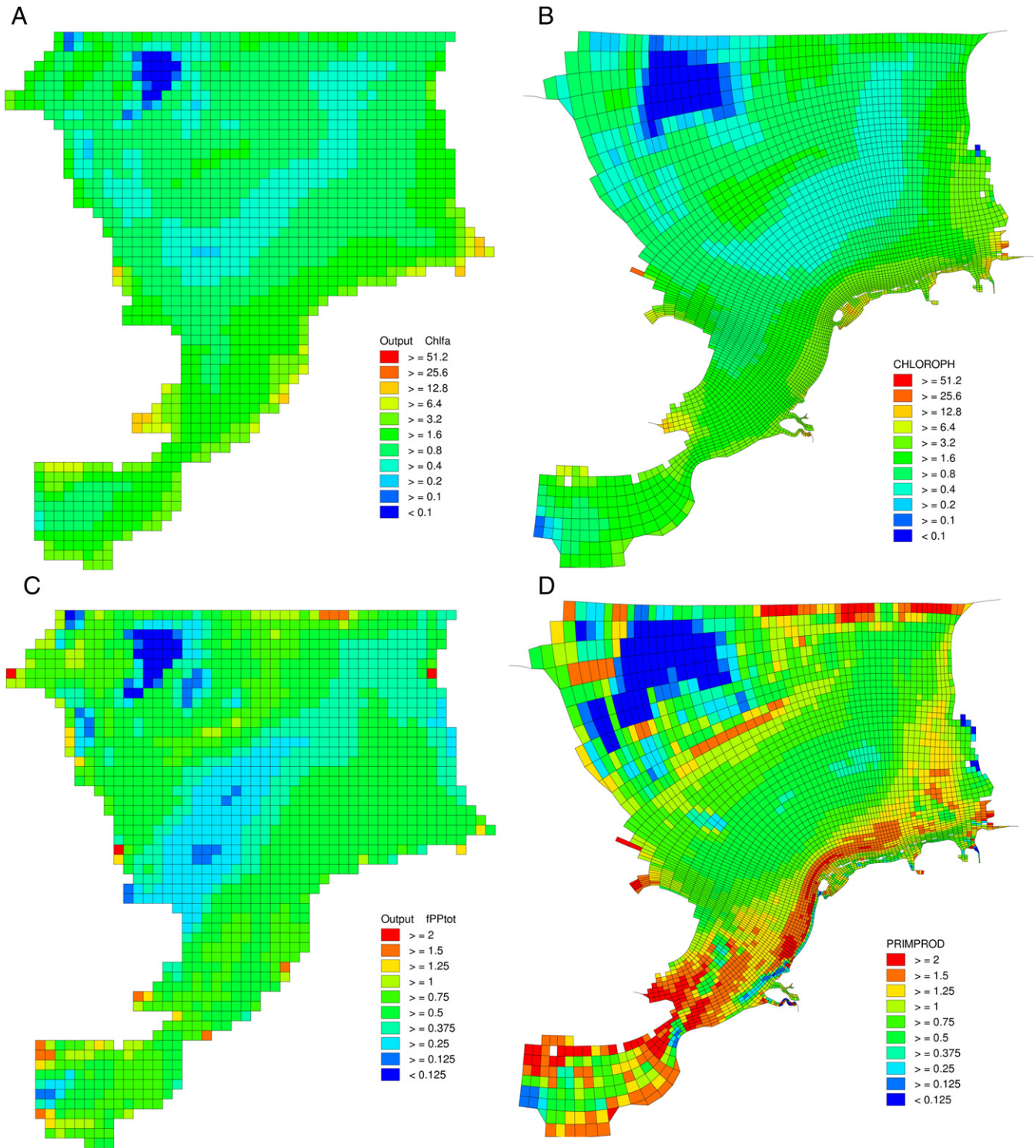


Fig. 7. Spatial distribution of annual average chlorophyll-a concentration ( $\text{mg m}^{-3}$ ) and annual average rate of primary production ( $\text{gC m}^{-2} \text{d}^{-1}$ ) in GENO (GDGSB) (A and C) and in ZUNO-2D (ZRZCB) (B and D).

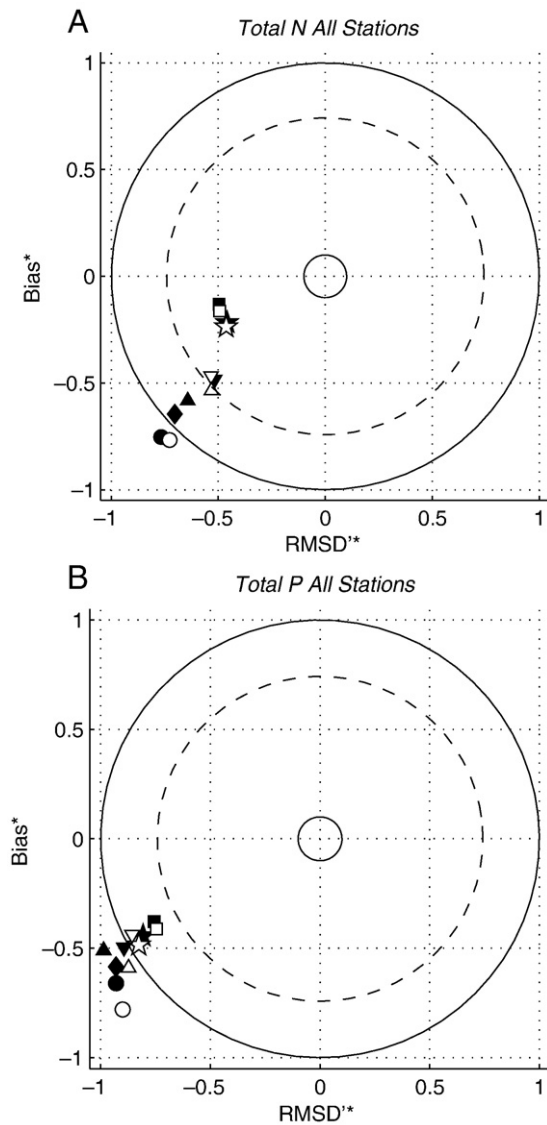
In summary, Fig. 12 confirms that the latest generation of models (ZUNO types) perform better, but also the high-resolution COAST applications perform relatively well. Moreover, the most recent 3D ZUNO models outperform in average sense their 2D counterparts.

### 3.3. Simulation results characteristic stations

As an illustration of the performances of the individual models at a typical location, results are presented here for three typical stations:

Terschelling 10 km, Walcheren 2 km and Terschelling 235 km (Dogger bank) (see Fig. 2). In all graphs the simulation results of some models are plotted against the measurements for 2003. Also included is the monthly mean, median and 90 percentile of the measurements for the years 1996–2002. These give an impression of the long-term typical values for a particular station. Given the advances in the models, one might expect a gradual shift in model behaviour with an improved fit of later models with respect to the actual 2003 measurements.



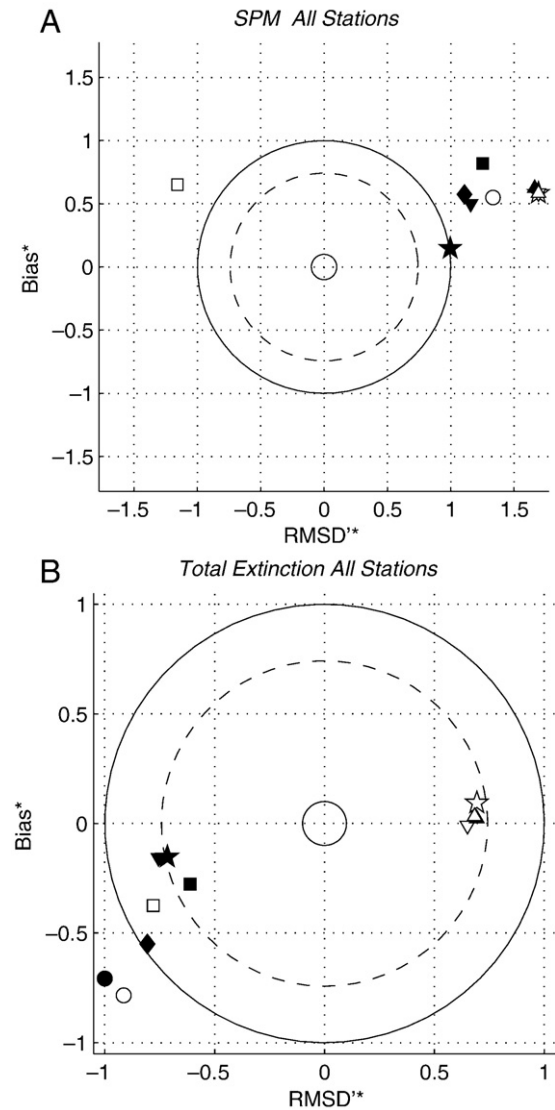


**Fig. 8.** Target diagram for SPM (A) and vertical extinction (B), all stations in all applications.

### 3.3.1. Terschelling 10 km

Terschelling 10 km was selected for two reasons. First it is situated 10 km offshore hence also the coarser models could still produce reasonable results here. Second it is more than 100 km north of the river Rhine outflow so observed nutrient levels are still clearly elevated in comparison to offshore locations, but the water is nearly always vertically mixed and horizontal gradients are not as strong as along the Noordwijk transect.

Fig. 13 shows the results of chlorophyll-*a* for all models. Although simulated chlorophyll-*a* levels of the two GENO based models are mostly below the 2003 observations, all models results are usually within the range of the long-term observations. So if the purpose of the model application is to produce a multi-year mean result, chlorophyll-*a* simulations from all models should be regarded as reasonable or good. A closer inspection of the results does show some differences in the ability of the models to reproduce seasonality. The variability of the results tends to increase in later model versions. The spring peak gets more pronounced and while in the GENO based models chlorophyll-*a* declines almost monotonously following the spring bloom, elevated chlorophyll-*a* levels later in the year are simulated by some of the other models in particular by COAST (KSKCB) and the ZUNO models. The anomalous late autumn peak at



**Fig. 9.** Target diagram for total  $\text{NO}_3$  (A),  $\text{SiO}_2$  (B) and  $\text{PO}_4$  (C) concentration, all stations in all applications.

Terschelling 10 (November) is due to exceptionally low SPM concentrations in late autumn 2003 which are only captured by the ZUNO models. Both the ZUNO models and COAST capture the more regular late summer (August–September) bloom.

Fig. 14 shows the results for total nitrogen and total phosphorus for all models at this station. Obviously the January observation differs considerably from the one in December. Remembering that all models were run with initial conditions taken from the end of the simulation, so the January measurements can hardly be reproduced by any of the models and should be ignored.

In the case of total nitrogen, simulation results for the two GENO models and for CSM(CDGSB) are consistently below the measurements (Fig. 14A). Results of ZUNO-2D(ZNZCB) are also below the measurements, but the difference is smaller. ZUNO-3D(Z3ZCB) and COAST(KSKCB) clearly show the best performance (Fig. 14C). Notice that there is an exceptionally large river outflow at the beginning of the year, which is accounted for in the load of all models, but only in the residual transport due to density distribution and atmospheric conditions of ZUNO-3D(Z3ZCB). This explains why ZUNO-2D(ZNZCB) with its average spring-neap forcing does not match the high observations in March and April, whereas ZUNO-3D(Z3ZCB) does. In spite of its daily forcing, COAST(KSKCB) does reproduce these peaks

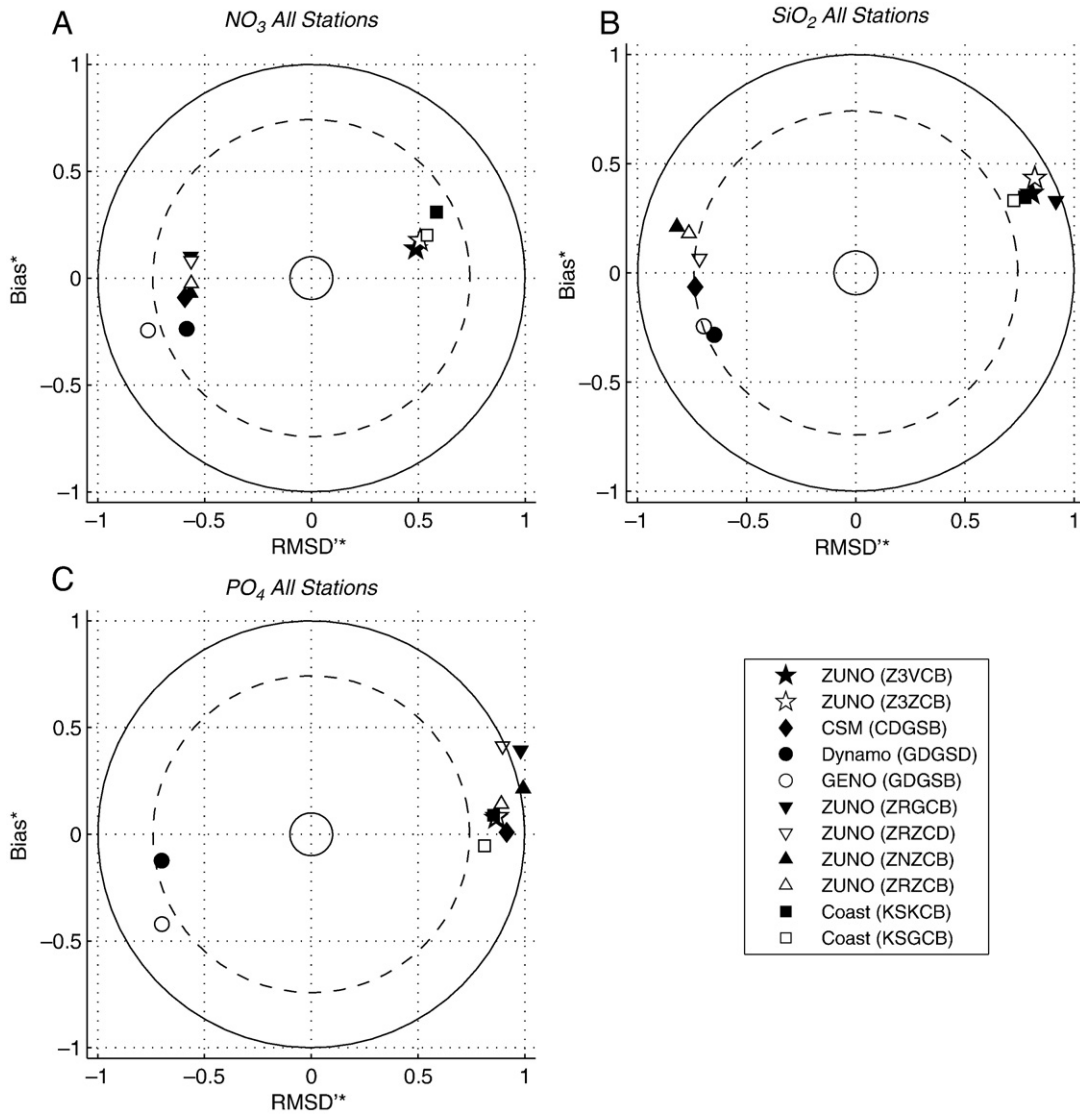


Fig. 10. Ratio of modelled ( $\sigma_M$ ) over observed ( $\sigma_D$ ) standard deviation of  $\text{NO}_3$  for all eleven applications determined across all stations.

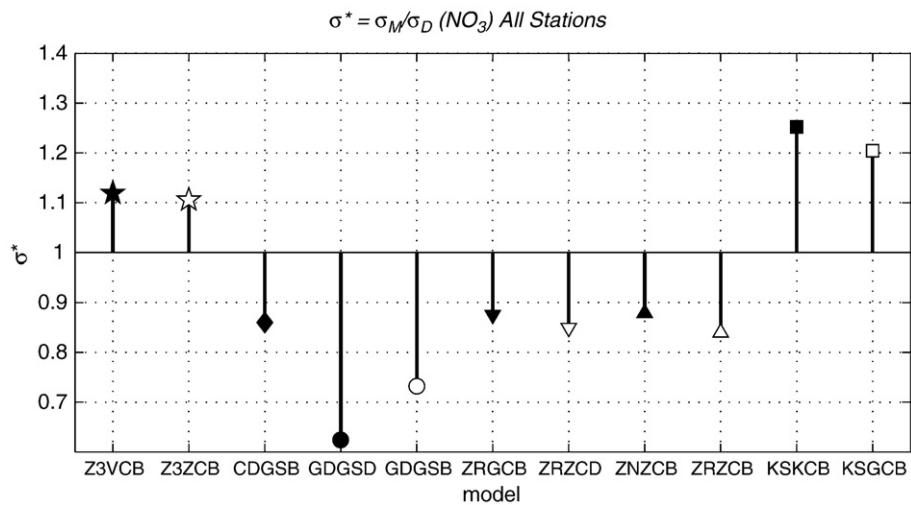


Fig. 11. Target diagram for total Nitrogen (A) and total Phosphorus (B) concentration, all stations in all applications.

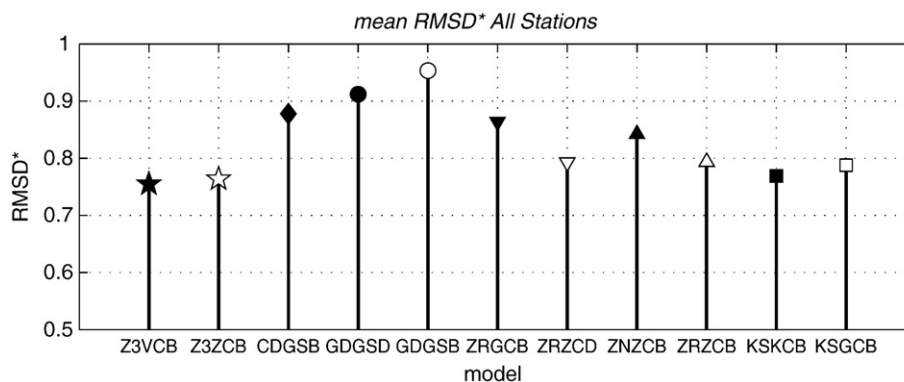


Fig. 12. Overall normalized RMS difference for all stations, averaged over the key variables, chlorophyll-*a*, Total-N, Total-P, NO<sub>3</sub>, PO<sub>4</sub>, SiO<sub>2</sub> and Total extinction.

values and as a matter of fact even seems to do better than the ZUNO-3D(Z3ZCB) model. Apparently the effects of the anomalous river discharge are well represented within the high-resolution domain of COAST; the proximity of the open boundaries does not contribute to the better match of the event in this case as the boundary conditions have been based on average climatology.

For total phosphorus differences between models are smaller than for total nitrogen. The most plausible explanation for this is a difference in the relative importance of the nutrient sources. In the case of nitrogen, there is a dominant contribution by the rivers. In contrast, phosphorus reductions in the river basin of the Rhine since the end of the 1980s have diminished the concentration differences with the Channel boundary. Hence TotN, with its larger spatial gradients, will be more susceptible to imperfections in the transport modelling along the Dutch coast than TotP. The underestimation of phosphorus by all models during the summer is caused by imperfections of the relatively simple sediment-bed model for nutrients adopted in all these simulations. The sediments are playing an important role in the phosphorus cycle, mostly in the shallow areas. In the deeper areas, most of the mineralization occurs in the water phase rather than in the sediment. Because the validation stations are predominantly located near shore and hence have a general bias for shallower areas, the differences in phosphorus stand out in the current results.

Fig. 15 shows the SPM forcing and the simulated vertical extinction coefficient. In the non-ZUNO based models, the SPM forcing at this location systematically exceeds the observations. Nevertheless the simulated extinction coefficients of these models are typically lower than observed. The main reason is that the contribution by CDOM (approximated by salinity) was not taken into account in these models. Average results for the ZUNO based models agree much better with the observations. The wind-based seasonal variation (see Section 2) is similar in amplitude to the long-term observed variation at this station. The simulated vertical extinction coefficients by COAST (KSKCB) and the two ZUNO models agree better with the observations compared to those of the previous models. The enhanced variability of SPM is also reflected in an enhanced variability of the extinction relative to COAST(KSKCB). Notice that the relatively high value of SPM of the COAST(KSKCB) model does not result in an overprediction of the extinction coefficient. This is because during the calibration of this model a lower specific extinction was adopted which compensates for high SPM values. The specific extinction coefficient of the ZUNO models is based on a statistical analysis (Van Gils and Tatman, 2003) and should be considered as more realistic.

From a visual inspection of the results for this station it may be concluded that they are in line with the general results of the target plots presented previously (Fig. 9A, B) where only ZUNO-3D(Z3VCB) has a reasonable skill for SPM. Resolution affects the transport, and hence the total nutrients, but even the coarsest model does not appear to be far off. The same holds for chlorophyll-*a*. There have been clear

advances in the simulation of the vertical extinction coefficient due to improvements in the light module.

### 3.3.2. Walcheren 2 km

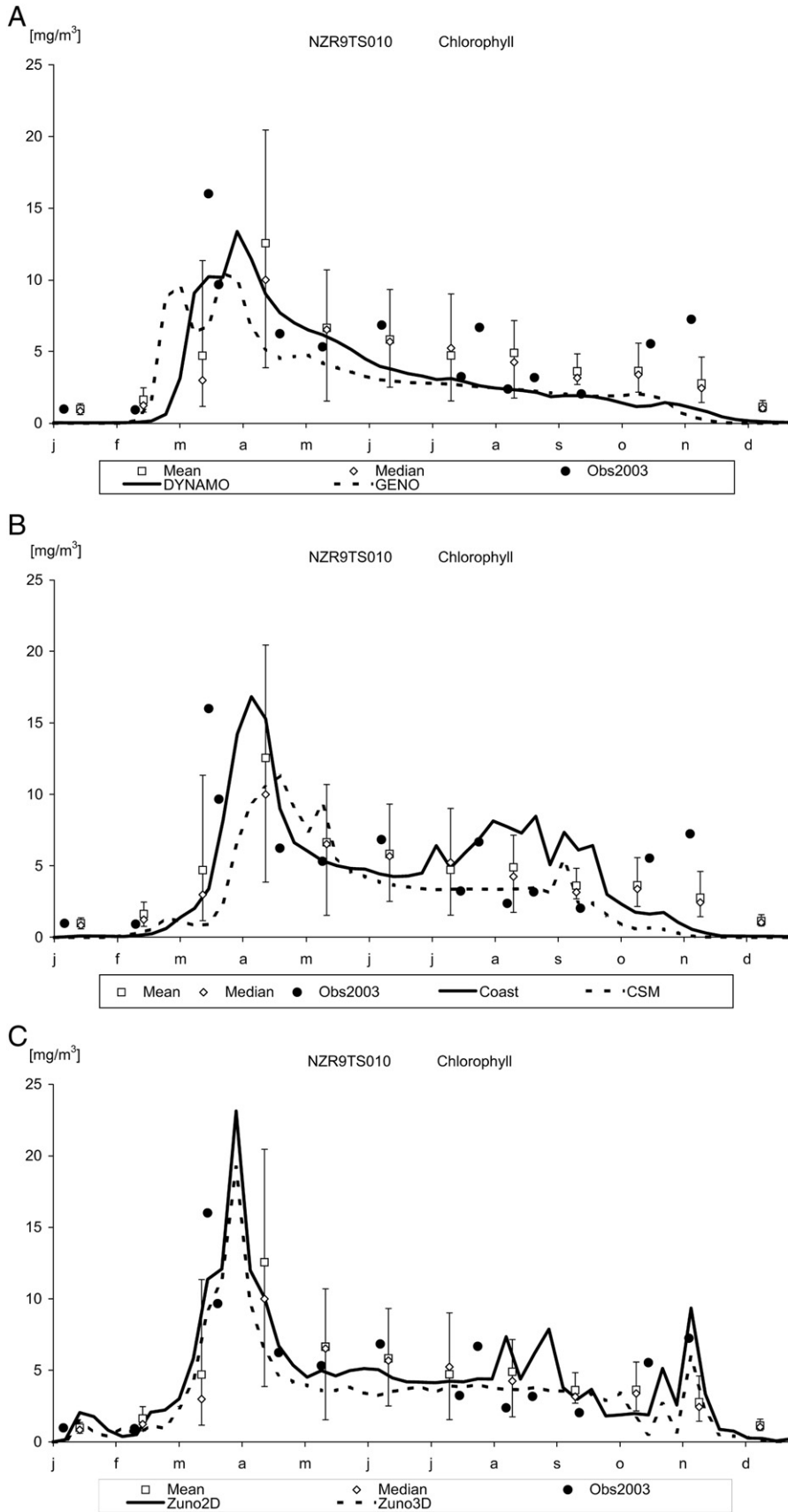
In an area with horizontal gradients, one might expect that the grid resolution has a clear impact on the results of the models. To investigate the importance of resolution at a coastal station, results are shown for the location Walcheren 2. This station is situated 2 km offshore, just north of the Western Scheldt estuary (Fig. 2). It is relatively shallow with an average depth of about 12 m and is characterized by high SPM levels. Because the residual current is towards the north east, the typical salinity is about 31 ppt, which is high in comparison to the coastal stations on the Noordwijk transect north of the Rhine-Meuse river mouth. Occasionally high fresh water discharges from the Rhine-Meuse system do protrude southward, resulting in a reduction of the salinity by about 1–3 ppt.

Fig. 16 shows the results for the coarse models (DYNAMO (GDGSD), GENO (GDGSB), CSM(CDGSB)), Fig. 17 or the fine resolution models (COAST(KSKCB), ZUNO-2D(ZNZCB) and ZUNO-3D(Z3ZCB)) for NO<sub>3</sub>, SiO<sub>2</sub>, chlorophyll-*a* and the extinction coefficient. In the coarse models, winter levels of the two nutrients are far below the measurements and should be qualified as reasonable or poor. Summer levels of both NO<sub>3</sub> and SiO<sub>2</sub> are limiting for a long period in all models. Consequently chlorophyll-*a* levels are strongly underpredicted by the two GENO models and so is the extinction coefficient. This is not just because chlorophyll-*a* (and detritus) are too low, but also because the forcing generated by the SPM sub-model on the same coarse grid is also far below the measurements. The chlorophyll-*a* result for CSM (CDGSB) is clearly better in terms of phasing, but still systematically below the measurements.

In all three fine resolution models shown here, results for NO<sub>3</sub>, SiO<sub>2</sub> agree much better to the observations (Fig. 17). In combination with a better forcing of the light climate by SPM, this results in a better simulation result of chlorophyll-*a* in comparison to the coarse models as well. The overall result of the ZUNO based models is better than for COAST(KSKCB) because in the latter summer levels are overpredicted. So while for all stations together, results of chlorophyll-*a* do not vary much between the different models, the differences at this coastal station are considerable. Results for other coastal stations consistently show the same pattern. Hence we conclude that a fine resolution (in the order of 3 × 3 km or less) is a necessary condition to adequately describe conditions at coastal stations. With respect to COAST(KSKCB) the proximity of the boundary somewhat complicates the interpretation of its results at this station. It is obvious though that steep gradients perpendicular to the coast are well maintained in this model.

But how important is the degree of realism in the transport? The three fine resolution models differ strongly with respect to their transport. The COAST(KSKCB) model is basically a residual current-driven model with a seasonal correction term of the dispersion for the





**Fig. 13.** Comparison of model results for chlorophyll ( $\text{mg m}^{-3}$ ) for different model versions at station Terschelling 10 km. Circles are measurements for 2003, bars indicate 90 percentile of measurements for the years 1996–2002.

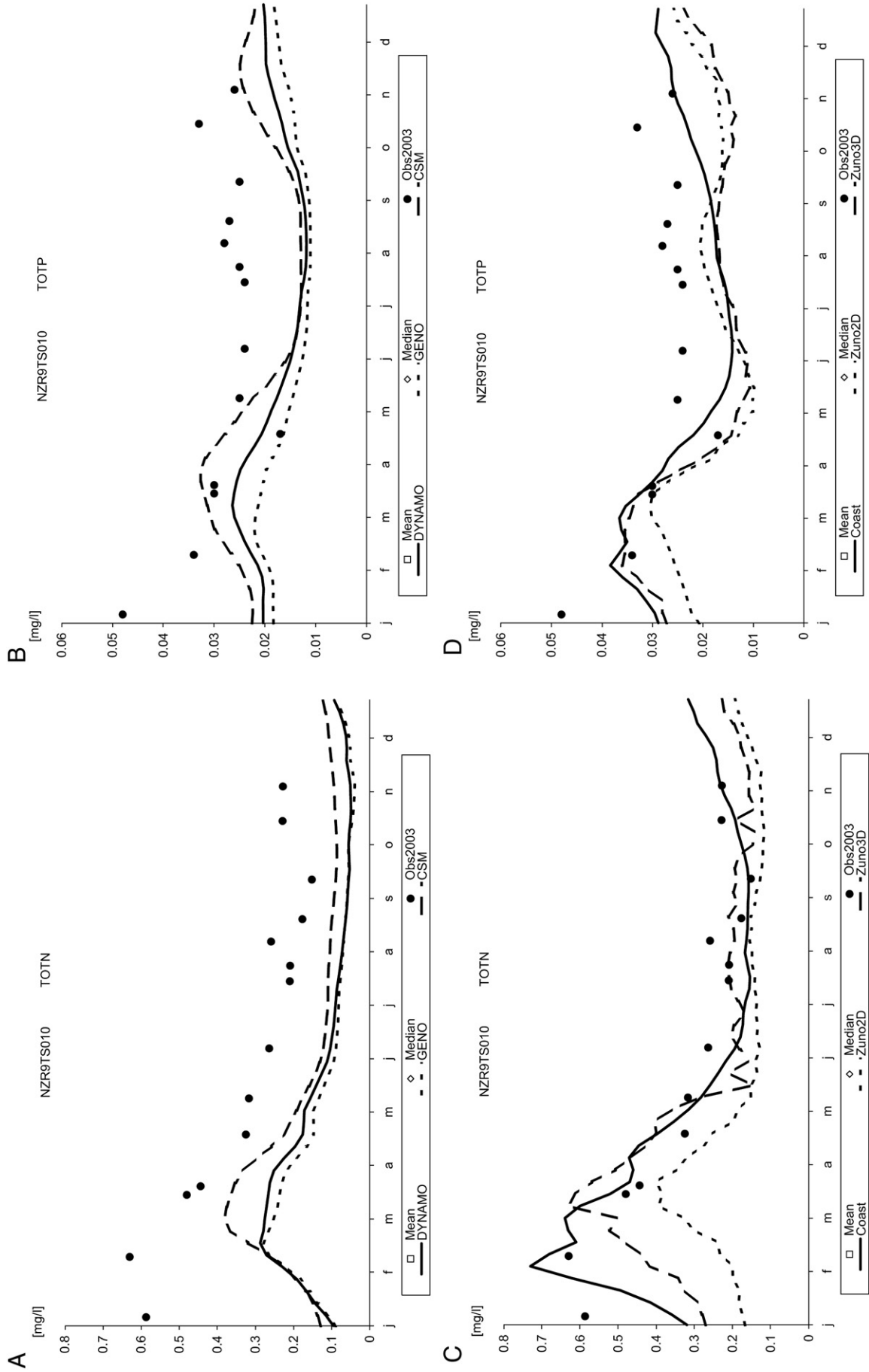
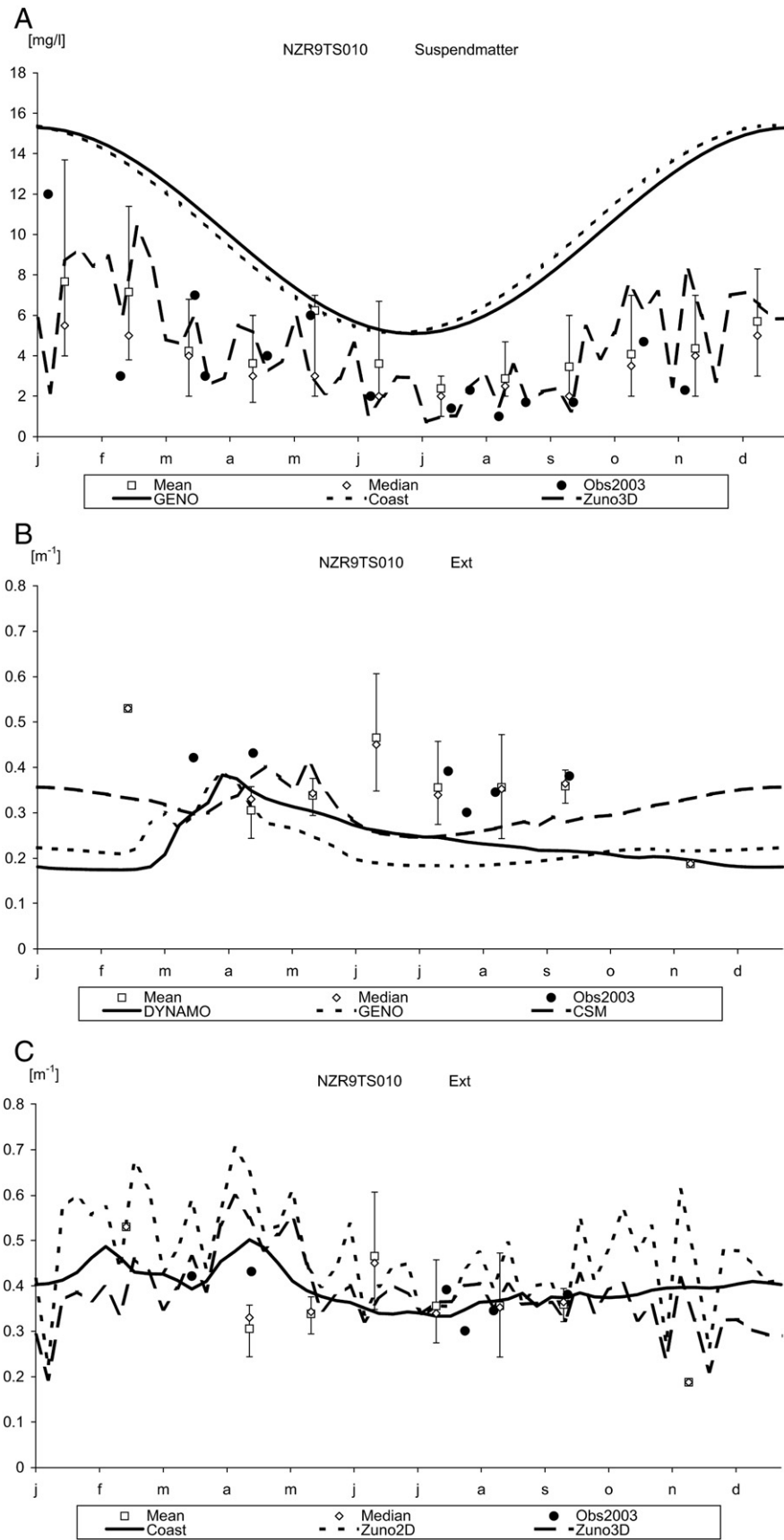
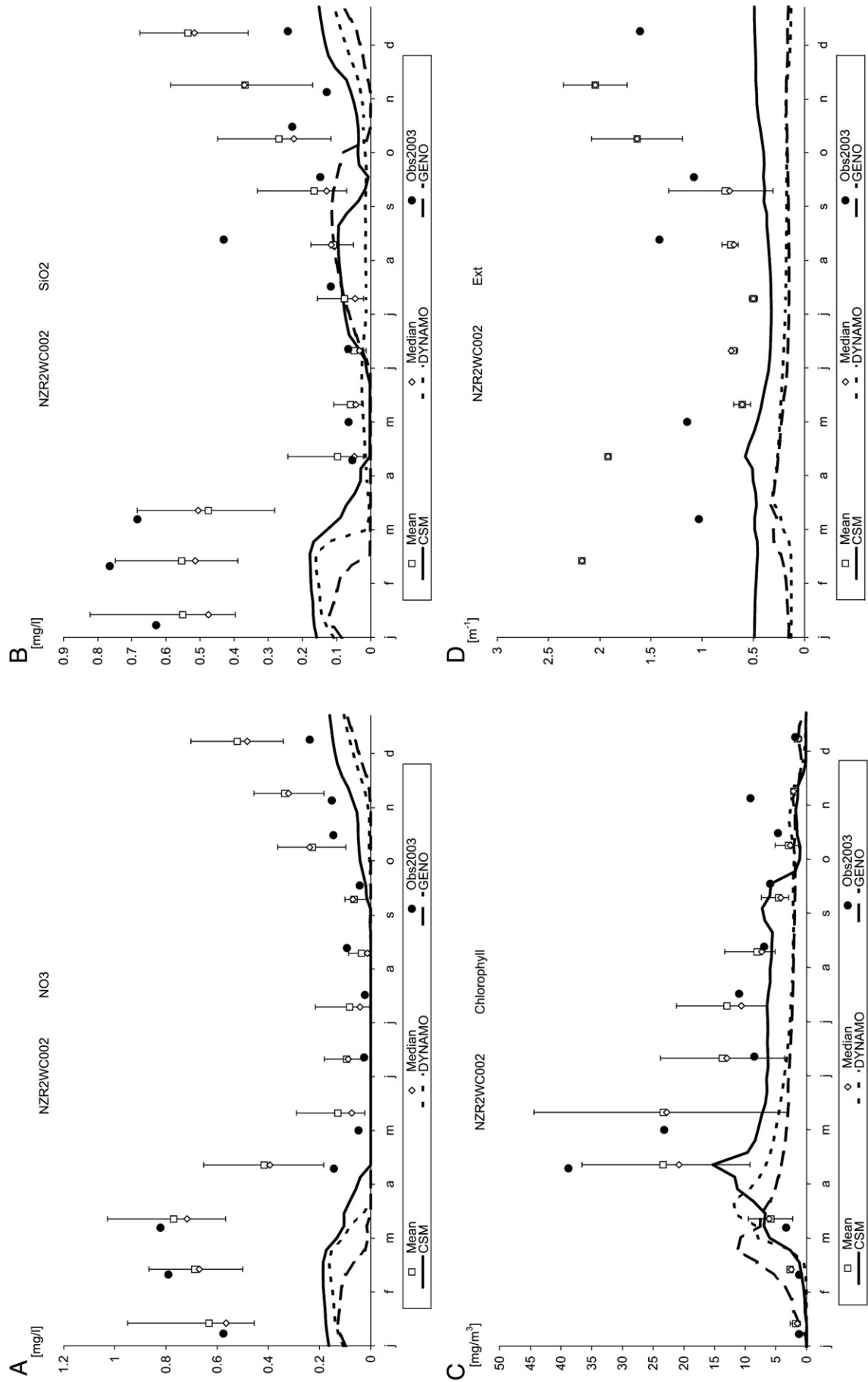


Fig. 14. Comparison of model results for TOTN (mg l<sup>-1</sup>) and TOTP (mg l<sup>-1</sup>) for different model versions at station Terschelling 10 km. Dots are measurements for 2003. Notice that the first measurement should be disregarded due to the spin-up procedure.

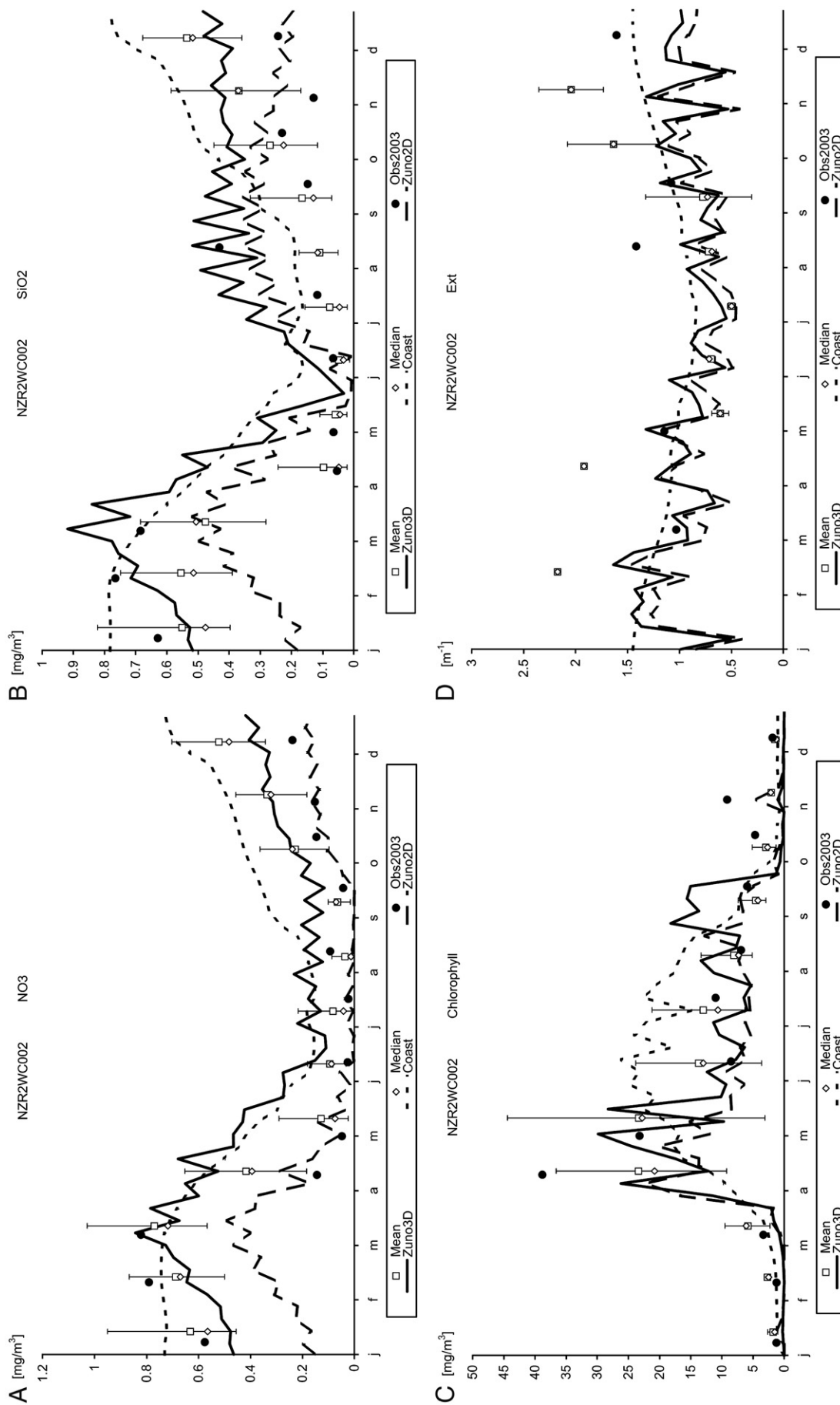


**Fig. 15.** Comparison of model forcing by suspended matter ( $\text{mg l}^{-1}$ ) (A) and the resulting vertical extinction coefficient ( $\text{m}^{-1}$ ) for the two GENO based models and CSM (B) and for the Coastal zone model and the two ZUNO based models (C) at station Terschelling 10 km. Forcings for DYNAMO and CSM (not shown) are the same as for GENO. Circles are measurements for 2003, bars indicate 90 percentile of measurements for the years 1996–2002.





**Fig. 16.** Comparison of NO<sub>3</sub>(A), SiO<sub>2</sub> (B) (mg l<sup>-1</sup>), chlorophyll-*a* (C) (mg m<sup>-3</sup>) and vertical extinction coefficient (D) (m<sup>-1</sup>) for coarse resolution models at station Walcheren 2 km. Circles are measurements for 2003, bars indicate 90 percentile of measurements for the years 1996–2002.



**Fig. 17.** Comparison of  $\text{NO}_3$  (A),  $\text{SiO}_2$  (B) ( $\text{mg}/\text{m}^3$ ), chlorophyll-*a* (C) ( $\text{mg}/\text{m}^3$ ) and vertical extinction coefficient (D) ( $\text{m}^{-1}$ ) for fine resolution models at station Walcheren 2 km. Circles are measurements for 2003, bars indicate 90 percentile of measurements for the years 1996–2002.

climatological wind speed and direction, ZUNO-2D(ZNZCB) is forced by a representative spring–neap cycle and ZUNO-3D(Z3ZCB) is driven by time and space varying historic forcing for 2003 (see the Section 2 for more details). The overall RMSD scores at the station Walcheren 02 (not shown) for  $\text{NO}_3$  and  $\text{SiO}_2$  for these three models are almost the same and qualify as good. On a seasonal basis ZUNO-3D(Z3ZCB) and COAST(KSKCB) give the best results for dissolved nutrients in spring during the period when the Rhine–Meuse discharges are extremely high, but during the rest of the year, nutrient levels exceed the measurement and the best fit is obtained by ZUNO-2D(ZNZCB). In the target diagrams the 3D ZUNO applications score at least as good or better than their 2D counterparts. Hence, when averaged over the entire year at this location, historic forcing results in a better skill. When viewed against the differences in experimental design, the improvements may seem marginal, still, especially when expressed in terms of chlorophyll-*a*.

### 3.3.3. Dogger Bank (Terschelling 235 km)

In the central North Sea, the residence times are relatively long. In particular in the vicinity of the Dogger Bank, represented by the station 235 km offshore along the Terschelling transect (Fig. 2). Tracer simulations indicate that after a year, between 50 and 80% of the water in this area was already within the southern and central North Sea domain when the simulation started. Hence, external sources are less important here in comparison to the coastal zone. The Dogger Bank area is relatively shallow (about 23 m deep) and usually well mixed. Turbidity is low and, hence, the spring bloom at this location may occur quite early in the season and even in winter it is not uncommon to find chlorophyll-*a* levels of about  $1 \mu\text{g l}^{-1}$ , which in this area is a typical summer value.

Unlike in the case of Terschelling 10 km or Walcheren 2 km, the resolution of all model applications is similar here. Hence, differences should be mostly attributed to either differences in model formulation and parameters such as the stoichiometric ratios or differences in forcing. In Fig. 18 results are shown for three different models, one from each model generation: GENO-NZB(GDGSB), CSM(CDGSB) and ZUNO-3D(Z3ZCB). In GENO-NZB(GDGSB) winter levels of the two most important nutrients,  $\text{NO}_3$  and  $\text{SiO}_2$  are low in comparison to the measurements. Winter levels in CSM(CDGSB) are highest, ZUNO-3D(Z3ZCB) is in between. In spite of the limited number of data points in the winter season, it seems that winter levels of the two nutrients are too low in GENO and of the correct order of magnitude in the other two models. All three models correctly indicate that  $\text{NO}_3$  is limiting for a very large part of the year. All models indicate correctly that  $\text{SiO}_2$  is also an important limiting nutrient.

ZUNO-3D(Z3ZCB) gives the most accurate result for chlorophyll-*a* and although its curve is much smoother, levels simulated by GENO-NZB(GDGSB) are usually close to the measurements as well. Results by CSM(CDGSB) systematically exceed the measurements. ZUNO-3D(Z3ZCB) gives the best simulation result for the vertical extinction coefficient; simulated levels by the other two models are usually too low.

Nevertheless, closer inspection of the algal bloom and its population composition shows that the GENO and CSM models mismatch the growth in spring, which is dominated by diatoms. As reflected in chlorophyll-*a*, the bloom starts too early and is too high. As a result there is a large build-up of the sediment Si pool, which is remineralised in summer resulting in an overestimation of  $\text{SiO}_2$  in the second half of summer and autumn. Not only the amount of remineralised  $\text{SiO}_2$ , but also the remineralisation rate turns out to be too high. It turns out that the two major steering factors of these processes are the light availability and sea water temperature at the Dogger Bank. The older models, forced with spatially uniform temperature time series derived from the near-shore station Noordwijk 10 off the Dutch coast (see Table 1), suffer from a too high temperature by 1 to  $2^\circ\text{C}$  and an underprediction of the light

extinction. The net effects on the algal growth and mineralization are obvious and are mostly notable in the  $\text{SiO}_2$  evolution in the second half of the year.

The importance of temperature in explaining this response is confirmed by the comparison of two ZUNO-2D simulations, one with historic transports and spatially varying temperature ZUNO-2D(ZRZCB) and one with a spring–neap tidal cycle and uniform near-shore temperature ZUNO-2D(ZNZCB) (not shown). The results of ZRZCB closely resemble those of the 3D ZUNO model shown in Fig. 18, whereas the ZNZCB results again show the same notable peak in late summer  $\text{SiO}_2$  also simulated by CSM and GENO-NZB.

In conclusion with respect to TS235, accurateness of the temperature forcing in addition to improvements in the parameterization of BLOOM and the extinction module are the major factors leading to overall model improvement. The overall gain in skill is confirmed in the Target Diagrams in Fig. 19 for chlorophyll-*a*.

## 4. Discussion

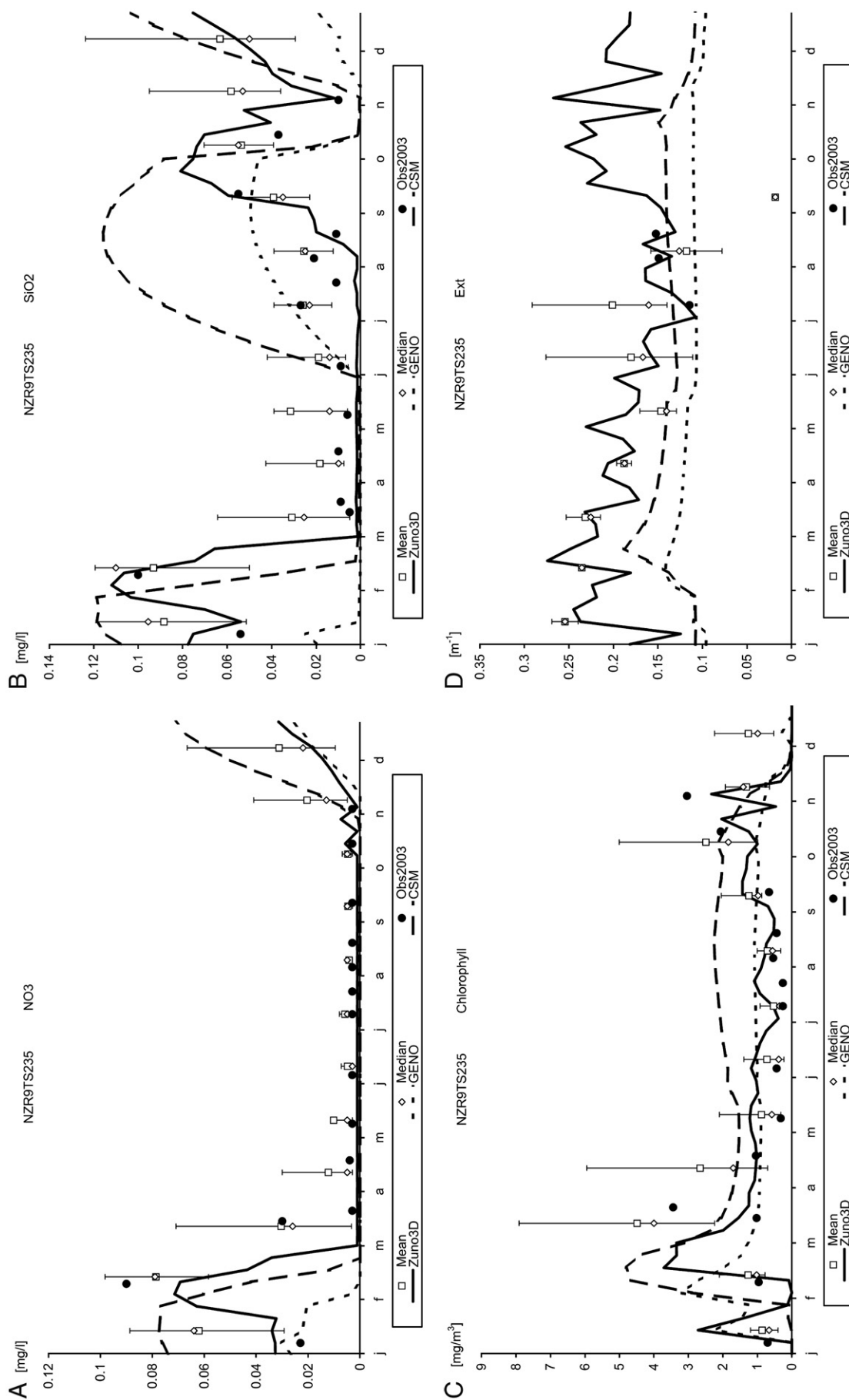
### 4.1. GOF criteria

At the beginning of our analysis, we have applied the OSPAR cost function (Villars and de Vries, 1998; Radach and Moll, 2006) to compare the scores of individual models. Previously, we had used this function to assess the performance of the ZUNO-3D(Z3ZCB) model as applied for the year 1989 (Los et al., 2008). However, during the current model intercomparison we notice that the discrimination of different model results by the cost function is rather poor. On several occasions models obtained a similar score for a particular output, although a visual inspection of the results revealed rather large differences which we felt should be reflected in their score. In particular for chlorophyll-*a*, almost all scores for all models at all considered monitoring locations pass the ‘good’ criterion (Fig. 4B), in spite of obvious differences noticed in the time series graphs. In this paper therefore, we have adopted the target diagram by Jolliff et al. (2008) which provided a concise but more contrasting picture of model performance on the entire model domain and for the entire period of the simulations. As with any method that aggregates results, local nuances in time and space are lost as flaws may obscure each other or overshadow good local performances. The evaluation of the bias in the target diagram focused on the annual mean, whereas the unbiased RMS difference is mostly related to capturing the seasonal cycle. Given the limited number of observations in the present validation data set, we judged that it was not sensible to much further refine the evaluation of the results on sub-season or regional scale, although in principle this would be advisable. For the present paper, the focus is on the overall performance given the present data supported by a discussion of a selection of time series of local stations in the traditional sense. A follow-up of this study might include additional mooring and remote sensing data such that a spatial and temporal breakdown of the results is feasible.

### 4.2. General observation

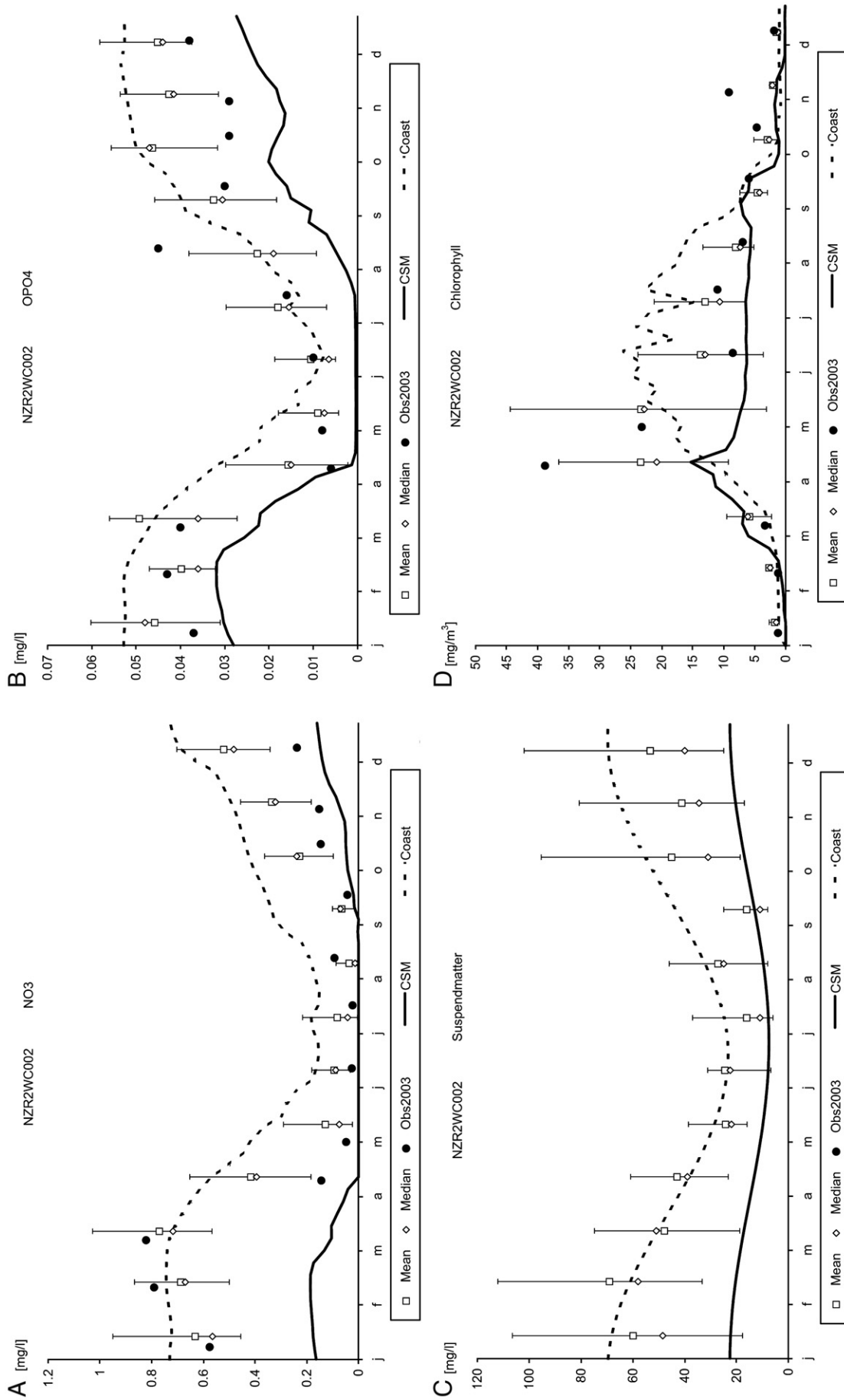
During model development every modeller attempts to improve the overall performance in such a way that obvious shortcomings are corrected while maintaining the quality of those results that already qualify as good according to the criteria adopted by the modeller. For relatively complex models with many interactions, parameters and forcings this proves to be difficult. The most recent and most advanced model that we have used during this comparison (ZUNO-3D(Z3ZCB)) indeed obtains the best overall scores (see Fig. 12), but it is easy to find exceptions, locally or temporarily, where some other model showed a better match with the observations. Several of those examples are included in this paper.





**Fig. 18.** Comparison of three model generations GENO-NZB(GDCSB), CSM(CDCSB), ZUNO-3D(Z3ZCB) at the Dogger Bank (Terschelling 235 km). Panels: (A) NO<sub>3</sub>, (B) PO<sub>4</sub>(C) chlorophyll-*a* (mg m<sup>-3</sup>), (D) vertical extinction coefficient (m<sup>-1</sup>). Circles are measurements for 2003; bars indicate 90 percentile of measurements for the years 1996–2002.





**Fig. 20.** Comparison of  $\text{NO}_3$  and  $\text{PO}_4$ , suspended matter ( $\text{mg l}^{-1}$ ) and chlorophyll- $a$  ( $\text{mg m}^{-3}$ ) for COAST-KSKCB and CSM-CDGSB model at station Walcheren 2 km. Circles are measurements for 2003, bars indicate 90 percentile of measurements for the years 1996–2002.

**Table 3**  
Impact of nutrient reduction scenario on NO<sub>3</sub> and chlorophyll-*a* at station Walcheren 2 km for Coastal zone and for CSM model.

| Case  | Base            |                 |               | Nutrient reduction scenario |                 |               | Percent effect scenario |                 |               |
|-------|-----------------|-----------------|---------------|-----------------------------|-----------------|---------------|-------------------------|-----------------|---------------|
|       | NO <sub>3</sub> | NO <sub>3</sub> | Chl- <i>a</i> | NO <sub>3</sub>             | NO <sub>3</sub> | Chl- <i>a</i> | NO <sub>3</sub>         | NO <sub>3</sub> | Chl- <i>a</i> |
| Model | Winter          | Summer          | Summer        | Winter                      | Summer          | Summer        | Winter                  | Summer          | Summer        |
| COAST | 0.46            | 0.3             | 16.7          | 0.33                        | 0.19            | 15.9          | –28.26                  | –36.67          | –4.79         |
| CSM   | 0.07            | 0.01            | 7.3           | 0.05                        | 0               | 5.3           | –28.57                  | –100            | –27.39        |

consistency also the scores for controlling factors such as SPM, CDOM and nutrients should be taken into account. A low score for a factor, which is actually controlling the model behaviour to a large extent, means that scenarios affecting that factor will not be simulated correctly.

#### 4.5. Total inorganic matter forcing of light

As explained previously, the SPM forcing is very important, particularly in determining the onset of the spring bloom. Locally, particularly in the parts of the Dutch coastal zone, light remains limiting to phytoplankton all year round due to high values of SPM, but the area where this is the case has decreased recently. Due to fluvial reductions in phosphorus since the 1980s, PO<sub>4</sub> is now often limiting during summer.

So how sensitive are the modelling results to the forcing by SPM? This question cannot easily be answered by comparing the results of the base simulations considered here because changes in SPM forcings coincide with other modifications notably in resolution and transport. For the 3D model two results are available that differ only with respect to the forcing by SPM. From Figs. 9A and 5 it may be concluded that the better RMSD score of the ZUNO-3D(Z3VCB) simulation for SPM also results in a better score of chlorophyll-*a*.

The effect of SPM variations on the timing of the spring phytoplankton bloom can also be demonstrated differently as was done in Los et al. (2008). Sensitivity studies presented in that paper showed that any type of short-term fluctuation in the suspended matter concentration, even if it was purely random, was sufficient to trigger an improvement in the timing of the spring bloom in the model at those location where it is controlled by light.

We conclude that improvements in SPM modelling clearly contribute to an improvement of the chlorophyll-*a* simulation.

#### 4.6. Complex versus simple phytoplankton modelling

Development and application of the fresh water implementation of BLOOM had already started in the 1970s and 1980s (Los, 1982; Los et al., 1984; Los and Brinkman, 1988). Modelling several functional groups, species and types is considered interesting for two reasons: (1) species dominance is an important ecosystem characteristic, not in the least since a number of species have been denoted objectionable for one reason or another and (2) it was demonstrated by for instance Zevenboom et al. (1982) that even if one and the same species is dominant in a particular water body for a prolonged period of time, its characteristics in terms of stoichiometry, chlorophyll-*a* contents and maximum growth rate vary considerably in response to changes in external conditions. For management of the North Sea, *Phaeocystis* is considered to be an important species, for which separate target values have been defined e.g., for the EU Water Framework Directive.

With respect to the species composition it was previously concluded in this paper that the overall, seasonally averaged performance of the models is reasonable or good for diatoms and *Phaeocystis*, but poor for the other two groups. Also there is little difference between model generations: there has not been much progress in this domain during development of the models.

In contrast, Los (1991) demonstrated that the fresh water implementation of BLOOM reproduced the observed phytoplankton species

groups adequately in most of 30 lakes that were modelled. Van der Molen et al. (1994) showed that the model is capable of reproducing a major shift in species dominance in the well investigated Lake Veluwe in the Netherlands after a series of management measures.

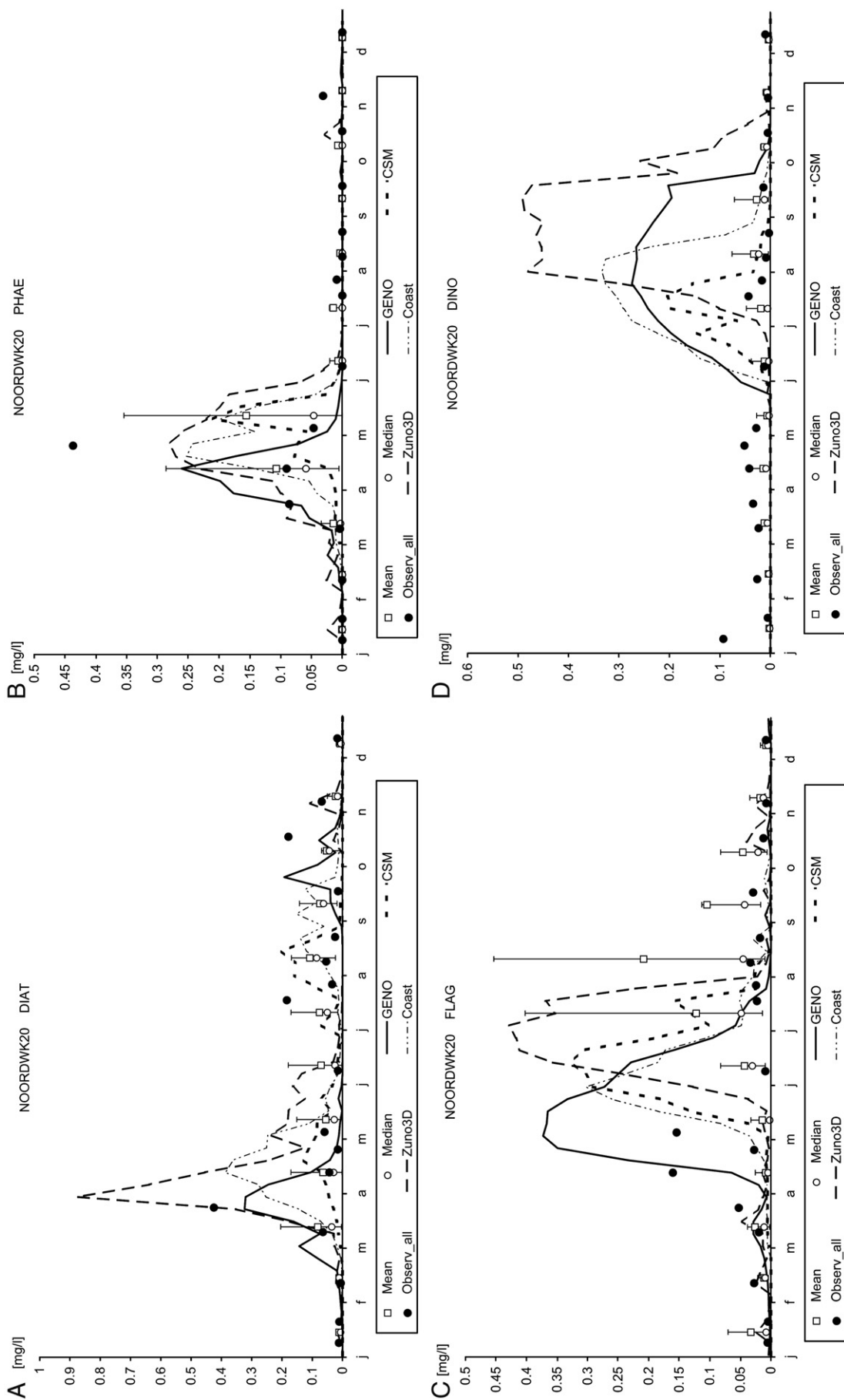
A seasonal example of the species composition simulated by the models GENO-NZB(GDGSB), COAST(KSKCB), CSM(CDGSB) and ZUNO-3D(Z3ZCB) and of the biomass estimates from cell counts at the coastal station Noordwijk 20 km is shown in Fig. 21. As for the RMSD scores of all station, the agreement between model results and observations look reasonable or good for diatoms and *Phaeocystis*, results are mixed for flagellates and timing and size of the simulated dinoflagellates biomass is rather poor. As was previously remarked, results of different model versions show some obvious differences, but there is not much progression.

There are several reasons why the scores for the simulated species biomasses in the North Sea models do not yet equalize those by the fresh water model. Firstly, monitoring data were lacking during the development of the marine models. In fact quantitative data have only become available after the calibration of the models presented here. Secondly, for the fresh water species more information is available on species growth and nutrient characteristics, both from chemostats and from field data. This is particularly true for the cyanobacteria species, which dominate in many Dutch lakes. Thirdly, biomass estimates for different years and locations show a lot of scatter. This in combination with a low (nominally monthly) sampling frequency, makes it difficult for any model to faithfully reproduce these numbers. In the example presented here, the 2003 observations for flagellates are highest in April and May, but these values are far out of the range of observations during all other years. In contrast the low values observed in July and August are at the very low end of the range obtained from all other years. Fourthly, Fig. 21E shows that there is no systematic difference between the measured and simulated total algal biomass during winter and spring, but there is a clear difference in summer. Instead, a corresponding discrepancy is not found for chlorophyll-*a* (Fig. 21F). This means that either the stoichiometric ratios adopted in all model versions for the typical summer species are incorrect, or that biomass estimates from cell counts and the chlorophyll-*a* measurements are inconsistent during the summer. This could be due to errors in the conversions from cell numbers to biomass, or because a significant part of the actual biomass consists of (small) species which are not observed under the microscope. Clearly this issue needs to be further investigated to improve the BLOOM model.

Notice that with respect to the usefulness of the model, the best species scores are obtained for what is from a management point of view considered to be the most important species: *Phaeocystis*.

The complexity of the phytoplankton module also affects the chlorophyll-*a* simulation of the model. The 12-type BLOOM phytoplankton module has the ability to display a wider range of variation both spatially as well as temporarily compared to a 2 species model because it selects its composition from a wider range of possible stoichiometric ratios. Differences are most obvious in case there are strong gradients in environmental forcings. Hence the differences in results between GENO-DYNAMO(GDGSD) and GENO-NZB(GDGSB) are not very obvious because the spatial gradients of the GENO grid are rather weak. In contrast the ZUNO grid is better suited for such a comparison. Indeed according to Fig. 5 the RMSD score of a 12 type simulation (ZRZCB) is clearly better than the score for the species





**Fig. 21.** Comparison of simulated and measured biomasses of diatoms, Phaeocystis, microflagellates and dinoflagellates (mg C l<sup>-1</sup>) and for total biomass (mg C l<sup>-1</sup>) and chlorophyll-*a* (mg m<sup>-3</sup>) of four different models at station Noordwijk 20 km. Circles are measurements for 2003. bars indicate 90 percentile of measurements for the years 1996–2002.

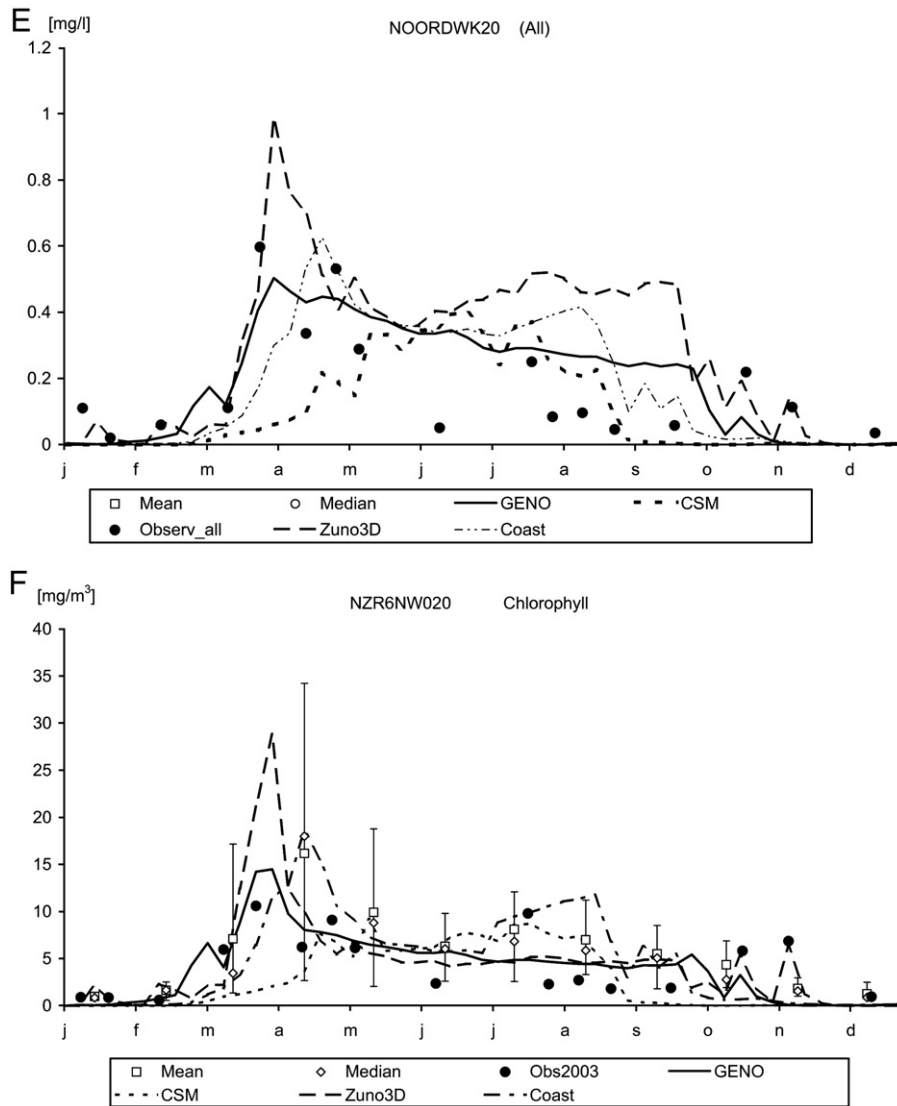


Fig. 21 (continued).

simulation (ZRZCD). We conclude that the overall ability of the model to simulate spatial and temporal chlorophyll-*a* gradients improves if the more complex BLOOM module is adopted; using BLOOM is not necessary in cases where only spatial averages or annual means are considered.

#### 4.7. One model or several related models?

Modellers usually concentrate their activities on their latest, most advanced and best performing model version. In our case ZUNO-3D (Z3ZCB) and ZUNO-3D(Z3VCB) with improved SPM forcing have the best overall score and hence will be the basis for further model development. But this model intercomparison has also demonstrated that the GOF scores of other model versions are sometimes equally good, sometimes even better and at least accurate enough to answer specific questions. So there are a number of reasons why we intend to keep more than one model version operational. (1) Ensemble modelling generates insight even if the models are as closely related as those used during this study. (2) For long forecast simulations, realistic forcings are not available so using a schematic, repetitive forcing of the transport is often necessary. (3) Quick scans on for instance nutrient reductions in decision support systems in the Netherlands are now usually based on expert rules. The fastest

performing deterministic model application shown here, simulates an entire year in a few minutes and one might argue that they are more accurate than the knowledge rules. (4) Related to the previous point, in many management studies the total number of simulations requested by the clients makes it more practical to adopt a somewhat simplified approach (i.e. 2D instead of 3D; limited model domain etc.). (5) A formal parameter sensitivity analysis of the complete 3D model is not very practical due to the necessity to perform a very large number of simulations. Instead, recently an extensive analysis was completed for ZUNO-2D(ZNZCB) (Salacinska, 2008), the results of which are also meaningful for ZUNO-3D(Z3ZCB).

## 5. Overall conclusions

During this study the following factors were considered:

- the resolution of the grid,
- the nature of transport forcing,
- attenuation of the underwater light conditions,
- the level of detail of the phytoplankton model.

With respect to resolution we concluded that refinement has a clear and positive effect on the transport of substances, not just locally as might be expected but also on a global scale, taking all measurements

into account simultaneously. Results for the forcing of the transports are less conclusive. In the Dutch coastal zone it appears that if the residual flow is correct, 2D models with daily tidal forcing, or with spring–neap forcing and a steady wind forcing do not differ much in performance compared to the 3D model with actual wind and density forcing. It should be noted though, that validation of models with real time forcing is hampered by the lack of high-frequency observation data. Based on modelling exercises in other areas and considering the importance of the forcing to model performances, we expect that the ZUNO-3D model will show better skills in reproducing high-frequency variability in comparison to the other models.

With respect to the underwater light conditions, it is obvious that the results of the primary production models are sensitive to this forcing in particular with respect to the timing, size and species composition of the spring bloom. Unfortunately neither of the SPM models considered here performs very well. The SPM modelling therefore should be improved. The contribution of CDOM to the attenuation of light is significant. The accuracy with which models simulate the extinction coefficient, clearly improves when CDOM (parameterized in terms of salinity) is taken into account.

Modelling 12 rather than 2 types of phytoplankton improves the ability of the model to correctly simulate chlorophyll-*a* because a wider range in nutrient and biomass to chlorophyll-*a* ratios is considered. This range is less well described by a 2 species model with fixed coefficients. Using the BLOOM module, results for *Phaeocystis* are good, but the match between simulation results for micro-flagellates and dinoflagellates should be improved. There is an inconsistency between the skill for modelling chlorophyll-*a* and the individual species. Consequently, not only the model should be improved, but more extensive and accurate data on species compositions are also needed.

The use of objective Goodness of Fit criteria depends on the nature of the available validation data. If only low-frequency data are available, obviously a certain degree of spatial and temporal aggregation is required in order to arrive at statistically sound results. The use of target diagrams facilitates to discern improvements in the mean and in the variability and is preferred over a cost function of the mean absolute error type. Classification still remains a partially subjective issue depending on the nature of the observations (resolution and uncertainty) and the purpose for which a model application is developed (i.e., time and spatial scale, bulk quantities or species composition etc.). In that context, we conclude that the model applications discussed here have been developing from representing multi-annual mean seasonal signals to seasonal patterns within a particular year, but validation of further refinement is only possible with further refined observations.

## Acknowledgements

We thank the two anonymous reviewers for their very constructive comments and suggestions to improve this paper. The research activities presented here have been partly funded by the Deltares Applied Research Programme BTO43 (contract Z4575) funded by Rijkswaterstaat and partly by the Deltares Strategic Research Programme SO9.3 (contract Z4753) and SO12.1 (contract Z4672).

## References

- Allen, J.I., Holt J.T., Blackford, J. and Proctor, R., 2007. Error quantification of a high-resolution coupled hydrodynamic ecosystem coastal-ocean model: Part 2. Chlorophyll-*a*, nutrients and SPM, *Journal of Marine Systems* 68 (3–4), 381–404.
- Baretta-Bekker, J.G., J.W. Baretta, M.J. Latuhihin, X. Desmit and T.C. Prins, 2009. Description of the long-term temporal and spatial distribution of phytoplankton carbon biomass in the Dutch North Sea, *Journal of Sea Research* 61, 50–59.
- Blaas, M., El Serafy, G.Y.H., van Kessel, T., Eleveld, M.A., 2007. Data Model Integration of SPM transport in the Dutch coastal zone. Proceedings of the Joint 2007 EUMETSAT/AMS Conference, Sep. 2007, Amsterdam, The Netherlands.
- Blaas, M., van den Boogaard, H., 2006. Statistical methods to assess the impact of MV2 on SPM along the Dutch coast, WL | Delft Hydraulics Report Z4046, November 2006, Delft The Netherlands.
- Blauw, A.N., Van de Wolfhaar, K., Meuwese, H., 2006. Transboundary nutrient transports in the North Sea: model study. WL | Delft Hydraulics Report, Z4188, December 2006, WL | Delft Hydraulics, Delft.
- Blauw, Anouk N., Hans, F.J., Los, Marinus Bokhorst, Paul, L., Erfemeijer, A., 2009. GEM: a generic ecological model for estuaries and coastal waters. *Hydrobiologia* 618, 175–198.
- Boon, J.G., Bokhorst, M., 1995. KSENOS, adjustment and extension of the modelling suite for toxic substances and eutrophication in the North Sea and Dutch coastal waters. WL | Delft Hydraulics report T1236 (in Dutch).
- De Kok, J.M., de Valk, C., van Kester, J.H.Th.M., de Goede, E., Uittenbogaard, R.E., 2001. Salinity and temperature stratification in the Rhine plume. A model study. *Estuarine, Coastal and Shelf Science* 53, 467–475.
- De Kok, J.M., Salden, R., Rozendaal, I., Blokland, P., Lander, J., 1995. Transport paths of suspended matter along the Dutch coast. Computer Modelling of seas and Coastal Regions. Computational Mechanics Publications, Southampton, pp. 75–86.
- De Vries, I., Duin, R.N.M., Peeters, J.C.H., Los, F.J., Bokhorst, M., Laane, R.W.P.M., 1998. Patterns and trends in nutrients and phytoplankton in Dutch coastal waters: comparison of time-series analysis, ecological model simulation and mesocosm experiments. In *ICES Journal of Marine Science* 55, 620–634.
- Di Toro, D.M., Fitzpatrick, J.J., Thomann, R.V., 1971. A dynamic model of the phytoplankton population in the Sacramento San Joaquin Delta. *Advances in Chemistry Series* 106, 131–180.
- Di Toro, D.M., Thomann, R.V., O'Connor, D.J., Mancini, J.L., 1977. Estuarine phytoplankton models—verification analysis and preliminary applications. In: Goldberg, E.D. (Ed.), *The Sea. Marine Modeling*. New York.
- Eisma, D., Kalf, J., 1987. Dispersal, concentration and deposition of suspended matter in the North Sea. *Journal of the Geological Society* 144, 161–178.
- Eleveld, M.A., Pasterkamp, R., van der Woerd, H.J., 2004. A survey of total suspended matter in the southern North Sea based on the 2001 SeaWiFS data. *EARSeL Proceedings* 3 (2/2004), 166.
- Ferris, J.M., Christian, R., 1991. Aquatic primary production in relation to microalgal responses to changing light: a review. *Aquatic Sciences* 53 (2/3), 1991.
- Friedrichs, M.A.M., Carr, M.-E., Barber, R.T., Scardi, M., Antoine, D., Armstrong, R.A., Asanuma, I., Behrenfeld, M.J., Buitenhuis, E., Chai, F., Christian, J.R., Ciotti, A.M., Doney, S.C., Dowell, M., Dunne, J., Gentili, B., Gregg, W., Hoepffner, N., Ishizaka, J., Kameda, T., Lima, I., Marra, J., Mélin, F., Moore, J.K., Morel, A., O'Malley, R.T., O'Reilly, J., Saba, V.S., Schmelz, M., Smyth, T.J., Tjiputra, J., Waters, K., Westberry, T.K., Winguth, A., 2009. Assessing the uncertainties of model estimates of primary productivity in the tropical Pacific Ocean. *Journal of Marine Systems* 76, 113–133.
- Garcia, V.M.T., Purdie, D.A., 1992. The influence of irradiance on growth, photosynthesis and respiration of *Gyrodinium cf. aureolum*. Marine Microbiology Group, Department of Oceanography, The University, Southampton. *Journal of Plankton Research* 14 (9), 1251–1265.
- Glas, P.C.G., Nauta, T.A., 1989. A North Sea computational framework for environmental and management studies: an application for eutrophication and nutrient cycles. Contribution to the International Symposium on Integrated Approaches to Water Pollution Problems (SISIPPA 89), Lisbon.
- Jahnke, J., 1989. The light and temperature dependence of growth rate and elemental composition of *Phaeocystis globosa* Scherffel and *P. Pouchetti* (Har.) Lagerh. in batch cultures. *Netherlands Journal of Sea Research* 23 (1), 15–21.
- Jolliff, J.K., Kindle, J.C., Shulman, I., Penta, B., Friedrichs, M.A.M., Helber, R., Arnone, R.A., 2008. Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. *J. Mar. Sys.* doi:10.1016/j.jmarsys.2008.05.014.
- Laane, R.W.P.M., Groeneveld, G., De Vries, A., Van Bennekom, A.J., Sydow, J.S., 1993. Nutrients (N, P, Si) in the Channel and the Dover Strait: seasonal and year-to-year variation and fluxes to the North Sea. *Oceanologia Acta* 16, 607–616.
- Lenhart, H.-J., Mills, D.K., Baretta-Bekker, J.G., van Leeuwen, S.M., van der Molen, J., Baretta, J.W., Blaas, M., Desmit, X., Kühn, W., Lacroix, G., Los, F.J., Ménesguen, A., Neves, R., Proctor, R., Ruardij, P., Skogen, M.D., Vanhoute-Brunier, A., Villars, M.T., Wakelin, S.L., 2010. Predicting the consequences of nutrient reduction on the eutrophication status of the North Sea. *Journal of Marine Systems* 81, 148–170.
- Lesser, G.R., Roelvink, J.A., Van Kester, J.A.T.M., Stelling, G.S., 2004. Development and validation of a three-dimensional morphological model. *Coastal Engineering* 51, 883–915.
- Los, F.J. An Algal Bloom model as a tool to simulate management measures. In: Barica, J. and Mur, L.R. (eds.), *Hypertrophic Ecosystems*, Junk, Dr. W. BV Publishers, The Hague-Boston-London, 1982.
- Los, F.J., 1991. Mathematical Simulation of Algae Blooms by the Model BLOOM II Version 2. WL | Delft Hydraulics Report T68.
- Los, F.J., Bokhorst, M., 1997. Trend analysis Dutch coastal zone. New Challenges for North Sea Research. Zentrum für Meeres- und Klimaforschung. University of Hamburg, pp. 161–175.
- Los, F.J., Brinkman, J.J., 1988. Phytoplankton modelling by means of optimization: a 10-year experience with BLOOM II. *Verhandlungen - Internationale Vereinigung für Theoretische und Angewandte Limnologie* 23, 790–795.
- Los, Hans, Jansen, Rene, Cramer, Sandra, 1994. MANS eutrophication modelling system. National Institute for Coastal and Marine Management (RIKZ).
- Los, F.J., Smits, J.G.C., De Rooij, N.M., 1984. Application of an Algal Bloom Model (BLOOM II) to combat eutrophication. *Verhandlungen - Internationale Vereinigung für Theoretische und Angewandte Limnologie* 22, 917–923.
- Los, F.J., Villars, M.T., Van der Tol, M.W.M., 2008. A 3-dimensional primary production model (BLOOM/GEM) and its applications to the (southern) North Sea (coupled physical–chemical–ecological model). *Journal of Marine Systems* 74, 259–294.
- Los, F.J., Wijsman, J.W.M., 2007. Application of a validated primary production model (BLOOM) as a screening tool for marine, coastal and transitional waters. *Journal of Marine Systems* 64, 201–215.

- Loucks, D.P., Van Beek, E. (Eds.), 2005. Water Resources Systems Planning and Management — an introduction to methods, models and applications; Chapter 12 'Water quality modelling and prediction'. Studies and Reports in Hydrology. UNESCO publishing. ISBN: 92-3-103998-9.
- Menden-Deuer, S., Lessard, E., 2000. Carbon to volume relationships for dinoflagellates, diatoms and other protist plankton. *Limnology and Oceanography* 45, 569–579.
- Moll, A., Radach, G., 2003. Review of three-dimensional ecological modelling related to the North Sea shelf system. Part I: models and their results. *Progress in Oceanography* 57, 175–217.
- Nauta, T.A., De Vries, I., Markus, A.A., De Groot, E.G., 1992. An integral approach to assess cause–effect relationships in eutrophication of marine systems. *Science of the Total Environment, Supplement*, 1992. Elsevier Science Publishers B.V, Amsterdam, pp. 1133–1147.
- Peeters, J.C.H., Haas, H.A., Peperzak, L., 1991. Eutrofiering, primaire produktie en zuurstofhuishouding in de Noordzee. National Institute of Coastal and Marine Management. Report GWAO-91.083 (in Dutch).
- Peeters, J.C.H., Los, F.J., Jansen, R., Haas, H.A., Peperzak, L., De Vries, I., 1995. The oxygen dynamics of the Oyster Ground, North Sea. Impact of eutrophication and environmental conditions. *Ophelia* 42, 257–288.
- Radach, G., Moll, A., 2006. Review of three-dimensional ecological modelling related to the North Sea shelf system. Part II: model validation and data needs. *Oceanography and Marine Biology: an Annual Review* 44, 1–60.
- Riegman, R., Noordeloos, A.A.M., Cadee, G., 1992. *Phaeocystis* blooms and eutrophication of the continental coastal zones of the North Sea. *Marine Biology* 112, 479–484.
- Riegman, R., De Boer, M., De Senerpont Domis, M., 1996. Growth of harmful marine algae in multispecies cultures. *Journal of Plankton Research* 18 (10), 1851–1866.
- Riegman, R., 1996. Species composition of harmful algal blooms in relation to macronutrient dynamics, In: Allan D. Cembella and Gustaaf M. Hallegraeff, *Physiological Ecology of Harmful Algal Blooms*, Donald M. Anderson, NATO ASI Series, Vol. 41., Springer Verlag, 1996.
- Rijkswaterstaat, 2003. Waterbase, Database of the Dutch National Surface Water Monitoring Network. <http://www.waterbase.nl>.
- Salacinska, K., 2008. Sensitivity analysis of the 2D application of the Generic Ecological Model to the North Sea. MSc thesis, Delft University of Technology, Delft, The Netherlands, 84 pp.
- Stow, C.A., Jolliff, J., McGillicuddy Jr, D.J., Doney, S.C., Allen, J.I., Friedrichs, M.A.M., Rose, K.A., Wallhead, Ph., 2008. Skill assessment for coupled biological/physical models of marine systems. *Journal of Marine Systems*. doi:10.1016/j.jmarsys.2008.03.011.
- Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research* 106 (D7), 7183–7192.
- Van der Molen, D.T., Los, F.J., Van Ballegoijen, L., Van der Vat, M.P., 1994. Mathematical modelling as a tool for management in eutrophication control of shallow lakes. *Hydrobiologia* 275/276, 479–492.
- Van Gils, J., Tatman, S., 2003. Light penetration in the water column. MARE Report, WL2003001 Z3379 WL | Delft Hydraulics, Delft, The Netherlands.
- Van Pagee, J.A., Glas, P.C.G., Markus, A.A., Postma, L., 1988. Mathematical modelling as a tool for assessment of North Sea Pollution. In: Salomons, W., Bayne, B.L., Duursma, E.K., Forstner, U. (Eds.), *Pollution of the North Sea, an Assessment*. Springer-Verlag, London.
- Van Kessel, T., Winterwerp, J.C., van Prooijen, B., van Ledden, M., Borst, W.G., 2008. modelling the seasonal dynamics of SPM with a simple algorithm for the buffering of fines in a sandy seabed, Proc. INTERCOH 2007, Brest.
- Villars, M., de Vries, I. (eds.), 1998. Villars, M., de Vries, I. Bokhorst, M., Ferreira, J., Gellers-Barkman, S., Kelly-Gerreyn, B., Lancelot, C., Menesguen, A., Moll, A., Patsch, J., Radach, G., Skogen, M., Soiland, H., Svendsen, E. and Vested, H.J., 1998. Report of the ASMO modelling workshop on eutrophication issues, 5–8 November 1996, The Hague, The Netherlands. OSPAR Commission Report, Netherlands Institute for Coastal and Marine Management, RIKZ, The Hague, The Netherlands.
- WL | Delft Hydraulics, 2005. Delft3D-FLOW users manual, v 3.12. WL | Delft Hydraulics, Delft, The Netherlands.
- WL | Delft Hydraulics/MARE., 2001. Description and model representation T0 situation. Part 2: the transport of fine-grained sediments in the southern North Sea. Delft Hydraulics Report WL2001003 Z3030.10.
- Zevenboom, W., Bij De Vaate, A., Mur, L.R., 1982. Assessment of factors limiting growth rate of *Oscillatoria agarhii* in hypertrophic Lake Wolderwijd, 1978, by use of physiological indicators. *Limnology and Oceanography* 27 (1), 39–52.