Journal of Marine Systems 81 (2010) 44–74

Contents lists available at ScienceDirect

Journal of Marine Systems

journal homepage: www.elsevier.com/locate/jmarsys

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.

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 improves 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 (M. Blaas).

0924-7963/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jmarsys.2009.12.011

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 a-typical 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