Setting up a water quality ensemble forecast for coastal ecosystems: A case study of the southern North Sea

Lőrinc Mészáros[1,2*], Ghada El Serafy[1,2]

ABSTRACT

Prediction systems, such as the coastal ecosystem models, often incorporate complex non-linear ecological processes. There is an increasing interest in the use of probabilistic forecasts instead of deterministic forecasts in cases where the inherent uncertainties in the prediction system are important. The primary goal of this study is to set up an operational ensemble forecasting system for the prediction of the Chlorophyll-a concentration in coastal waters, using the Generic Ecological Model (GEM). The input ensemble is generated from perturbed model process parameters and external forcings through Latin Hypercube Sampling with Dependence (LHSD). The forecast performance of the ensemble prediction is assessed using several forecast verification metrics that can describe the forecast accuracy, reliability and discrimination. The verification is performed against in-situ measurements and remote sensing data. The ensemble forecast moderately outperforms the deterministic prediction at the coastal in-situ measurement stations. The proposed ensemble forecasting system is therefore a promising tool to provide enhanced water quality prediction for coastal ecosystems which, with further inclusion of other uncertainty sources, could be used for operational forecasting.

Keywords: Coastal ecosystems, ensemble forecasting, environmental modelling, North Sea, uncertainty

*Corresponding author, email: Lorinc.Meszaros@deltares.nl;

(From production details: Lorinc.Meszaros@deltares.nl;meszaros.lorincz@gmail.com)

[1] Deltares, Boussinesqweg 1, 2629 HV Delft, The Netherlands
[2] Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands

INTRODUCTION

Water quality is a crucial factor for both coastal ecosystems and human societies. Phytoplankton blooms in the North Sea may cause mortality of mussels and other benthic organisms. Furthermore, fisheries and aquacultures are influenced by the algal primary production since it is the base of the food web. Consequently, accurate real-time phytoplankton concentration prediction is required for ecosystem and economic benefits. Timely information regarding water quality allows for early warning and adequate response such as mitigation measures and targeted monitoring.

Existing hybrid ecosystem models are powerful tools for modelling water quality; however, their reliability depends highly on the uncertainty stemming from different sources. Uncertainty originating from external forcings is further propagated and complicated by the non-linear ecological processes, with numerous intercorrelated parameters incorporated in the water quality model. Considering the high level of uncertainty in the coastal water quality forecasting process, it is assumed that a single-valued deterministic forecast may not be sufficiently reliable for decision making. Thus, a strong need arises for an ensemble forecasting system that could potentially account for the uncertainty associated with the driving forces, the model simplifications or the parameterization.

Ensemble forecasting of water quality might require considerably more effort than a deterministic approach, though in return gives added value to the forecast. Deterministic forecasts only provide a point estimate, whereas probabilistic forecasts determine the probability density function of the predictand. By performing a statistical analysis of the acquired probability density function, various measures can be derived such as the mean and standard deviation of the model output distribution, or the likelihood of a specific output value. Through these measures the uncertainty in the model output can be quantified, which serves decision support functions.

The application of ensemble prediction systems is well-known in various fields, especially in numerical weather prediction, yet few examples (Fiechter 2012; Trolle *et al.* 2014; Huang & Gao 2017) show their applicability in water quality forecasting. A recent study by Ahn *et al.* (2016) investigated the possibility of developing a water quality forecasting system for riverine ecosystems with an ensemble streamflow prediction method. In that study a historical rainfall and temperature ensemble was applied as a forcing condition in order to produce a probabilistic water quality prediction. Based on those findings it can be concluded that implementing an ensemble system for water quality forecasts by addressing the uncertainty in the input variables is advantageous. Another research in ensemble forecast accuracy in the southern North Sea was conducted by Herrmann (2015). The ensemble forecasting system was set up using perturbed meteorological forcings and the Delft3D-FLOW model. Despite expectations, the results showed that the impact of the meteorological ensemble input is not sufficient to estimate the total uncertainty in the forecasted parameters (salinity and temperature). In order to complement those findings this paper suggests the application of an extended input ensemble with the aim to provide further uncertainty estimation.

This paper aims to set up the framework for an operational water quality ensemble forecasting system, even though the proposed method is only applied in a hindcasting case study and further steps are required to reach the operational stage. The model output of interest is the Chlorophyll-a concentration which is a commonly used indicator for water quality. The simulation is carried out using the Delft3D Generic Ecological Model (GEM) with the advanced algal speciation module-BLOOM which is part of the Delft3D-WAQ software package.

The paper starts with a description of the GEM, its application to the North Sea, and the different data sources used for the forecast verification. The research methodology is presented in the following sections including the significant parameter selection, the Latin Hypercube Sampling with Dependence (LHSD) steps, and the applied forecast verification method. The performance analysis of the ensemble forecasting system applied to the southern North Sea is given together with the validation and spatial results. The conclusions and recommendations for further research are presented in the final section of the paper.

MODELLING INSTRUMENT

The modelling instrument is a comprehensive hybrid ecological model combining a three-dimensional hydrodynamic model (Delft3D-FLOW) and the GEM biogeochemical model. The GEM includes an array of modules reproducing water quality processes that are then combined with the transport model. Most importantly, the model computes primary production and Chlorophyll-a concentration while integrating dynamic process modules for dissolved oxygen and nutrient concentration calculation. Furthermore, the GEM includes a phytoplankton module (BLOOM) that simulates the growth, respiration and mortality of phytoplankton. Using this module, the species competition and their adaptation to limiting nutrients or light can be simulated. The model offers flexibility in the processes selection and provides general applicability in diverse case studies (Blauw *et al.* 2009). In recent years the GEM has already been applied to the southern North Sea. A three-dimensional GEM application was presented by Los *et al.* (2008), while a two-dimensional application was completed by Salacinska *et al.* (2010).

Model calibration and validation of the three-dimensional GEM for the southern North Sea was carried out in previous studies by Los *et al.* (2008) using an in-situ dataset for the year 1989, in Keetels *et al.* (2012) for the year 2007, and in Arentz *et al.* (2012) for years 2009-2010 also using the in-situ dataset. In general, all the reports mention acceptable Chlorophyll-a prediction accuracy, however, the results are not homogeneous across the stations and are highly dependent on proper description of the Suspended Particulate Matter field. Furthermore, it should be noted that in the Dutch coastal zone, the observed gradients of algal biomass are very steep and there is considerable natural variability in the Chlorophyll-a concentration. The peak spring bloom could be shifted ± 1 month and its magnitude could vary ±80% (Keetels *et al.* 2012).

Transport of substances

The transport of substances is described by the advection-dispersion equation below (Eq. 1), which is responsible for the change in concentration of the substances in time due to the advective transport and the diffusive and/or dispersive transport, along with the sources and sinks. The equation is solved by the hydrodynamic model (Delft3D-FLOW), where a broad selection of numerical schemes is available to compute the transport part of the equation. The selection of numerical solutions helps to cope with the 1D/2D/3D model discretization and even complex, irregular geometries for both steady and unsteady cases.

$$\frac{\partial C}{\partial t} = -u\frac{\partial C}{\partial x} - v\frac{\partial C}{\partial y} - w\frac{\partial C}{\partial z} + \frac{\partial}{\partial x}\left(D_x\frac{\partial C}{\partial x}\right) + \frac{\partial}{\partial y}\left(D_y\frac{\partial C}{\partial y}\right) + \frac{\partial}{\partial z}\left(D_z\frac{\partial C}{\partial z}\right) + S + P \qquad \text{(Eq. 1)}$$

where:

C	concentration of the state variables[g m^{-3}]
u, v, w	velocity vector components [m s^{-1}]
D_x, D_y, D_z	dispersion tensor components [m^2 s^{-1}]
x, y, z	coordinates [m]
S	source and sink term of mass due to loads and boundaries
P	source and sink term of mass due to processes
t	time [s]

Ecological processes

The concentrations of 30 state variables including algae concentrations, nutrients, and salinity are calculated through various ecological processes. The most important processes are related to nutrient cycles, oxygen dynamics, energy availability and phytoplankton processes. The GEM ecological processes are considered to be moderately complex by Los *et al.* (2008) based on the fact that the GEM excludes microbial loop, explicit grazing and higher trophic levels; also the benthic processes are relatively simple. On the other hand, full nutrient cycles and complex phytoplankton kinetics were implemented. Figure 1 contains the schematic overview of all possible variables and processes in the GEM. The reader may refer to Blauw *et al.* (2009) and Los *et al.* (2008) for further description of the ecological processes.

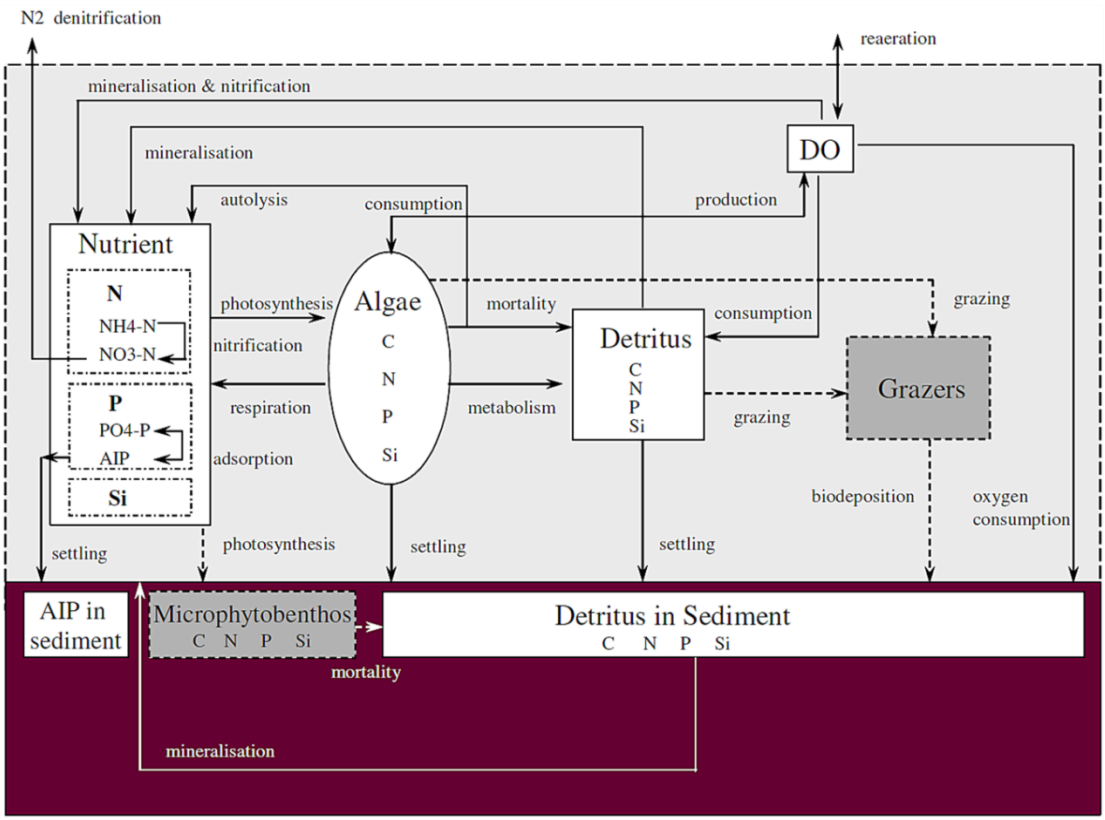

Figure 1. State variables and processes in the GEM. State variables in grey and processes indicated by dashed lines have not been included in the North Sea modelling applications. AIP is adsorbed inorganic phosphorus. *Source:* Los *et al.* (2008).

The GEM makes use of more than 400 parameters for the processes calculations. These parameters are related to algae's characteristics such as nutrient-to-carbon ratios and growth rates. Some describe light availability through extinction coefficients and settling velocities and others are connected to the nutrient cycles. The GEM is part of an integrated modelling system which contains separate modules for hydrodynamics, waves and sediment transport calculation. The water quality simulation therefore requires external forcings such as meteorological conditions, hydrodynamics, Suspended Particulate Matter (SPM) concentration field as well as nutrient loadings from atmospheric deposition and riverine loads.

Spatial discretization

The domain decomposition of the southern North Sea is a three-dimensional curvilinear grid with finer resolution along the coast as shown in Figure 2. The grid contains 12 sigma layers with unequal thicknesses (see Table 1). The distribution of the layers was designed to provide higher resolution on the top and the bottom of the water column to enable more accurate suspended matter calculation for light attenuation on the top and more detailed resuspension calculation near the bed. The resolution of the coarse segments varies from 6-by-5 km to 20-by-30 km and the resolution of the fine grid ranges from 1-by-2 km to 2.5-by-3 km.

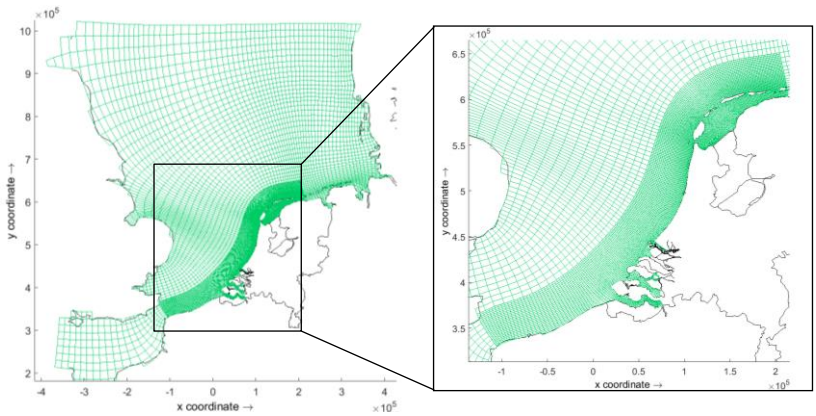

Figure 2. Spatial discretization of the GEM in the North Sea.

Table 1. Relative thickness of the sigma layers in the three-dimensional grid

Layer no.	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***11***	***12***	***Total***
Relative thickness (%)	4	5.6	7.8	10.8	10.9	10.9	10.9	10.9	10.9	7.8	5.6	4	100

DATA SOURCES

In order to validate the water quality model in this paper three types of data sources are used: a set of historical surface in-situ measurements, and remote sensing images together with their gap-filled reconstruction (see Figure 3). Fifteen stations were selected with available Chlorophyll-a in-situ measurements for the given years 2007 and 2009 in the focus area provided by Rijkswaterstaat (RWS, Dutch Ministry of Infrastructure and Environment). It should be noted that the samples are taken close to the water surface (usually the upper 3-5 metres of the water column) and therefore they can only be used to validate the first layer of the model.

The second type of dataset is remote sensing which gives a more comprehensive view of the model's success in capturing seasonal variability since the spatial and temporal resolution of the in-situ dataset is low. The remotely sensed sea colour images applied in this study are retrieved from the MEdium Resolution Imaging Spectrometer instrument (MERIS), which was on board the European Space Agency ENVironmental SATellite (ENVISAT) spacecraft. Since the ENVISAT ended its mission in 2012, the application of MERIS data in this paper is considered as a preparation for the new Copernicus Sentinel dataset usage which has recently stepped into the fully operational stage.

The recorded optical reflectance of the water surface is transformed into Chlorophyll-a or SPM concentration using the HYDROPT algorithm given in van der Woerd and Pasterkamp (2008). The algorithm not only computes the concentration of the substances but also provides a measure of error in the estimates (El Serafy *et al.* 2007). The retrieved pixel data is then interpolated onto the Southern North Sea Domain Decomposition (ZUNO-DD) grid taking into account the provided error estimate, and resulting in the so-called gridded data (Blaas 2013). It should be noted that the raw pixel data was first

interpolated using nearest-to-the-centre interpolation onto the ZUNO-COARSE gridding and then it was subsequently transposed onto the ZUNO-DD domain gridding.

Most forecast verification metrics are based on the differences between the prediction and observation; however, these differences can only be calculated if both values are available for the same time-step at the same location. Considering the incomplete temporal and spatial coverage of the samples from the MERIS measurements due to clouds, the number of matchups is fairly limited. As an attempt to tackle this issue a third dataset is used for the model validation, which is the gap-filled version of the gridded MERIS images. The gap filling is done using a data interpolating algorithm called DINEOF. The DINEOF algorithm determines the major spatial and temporal patterns of variation in the MERIS dataset and produces the gap-filled reconstruction at all segments and all time-steps (Blaas 2013). The gap-filled data applied in this study was constructed using the MERIS data only from the years of interest; 2007 and 2009.

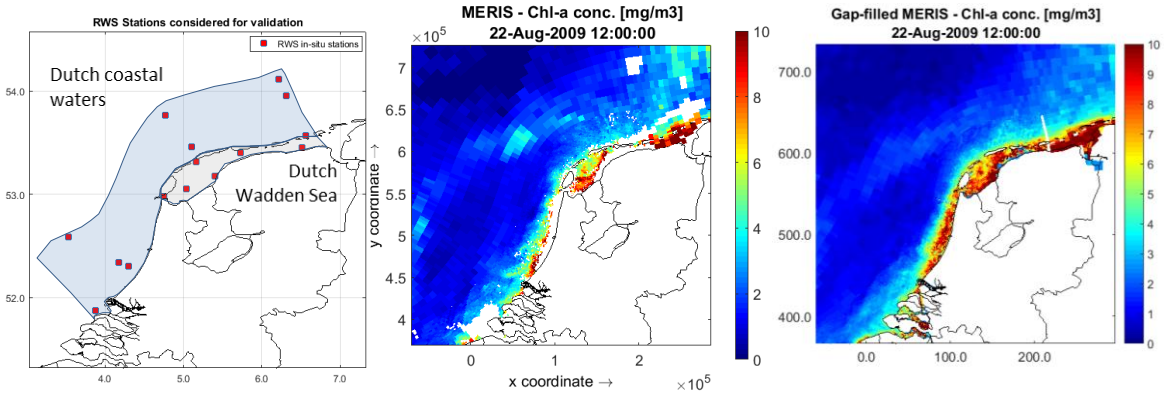

Figure 3. Validation data sources: in-situ measurements (left), MERIS (middle) and gap-filled MERIS (right) remote sensing data.

ENSEMBLE FORECASTING METHODOLOGY

The main objective of the paper is to set up a viable and generally applicable water quality ensemble forecasting system for coastal ecosystems. The proposed methodology (see Figure 4) includes the following important steps:

- Selection of the significant model process parameters and external forcings responsible for the uncertainty in the Chlorophyll-a concentration prediction.
- Statistically efficient sampling of the input ensemble (stratified sample with parameter dependency) using the selected significant parameters and forcings.
- Post-processing of the ensemble model run by fitting probability density functions to the model outputs at every time-step and all locations.
- Forecast verification to assess and compare the performance of the deterministic and ensemble predictions.

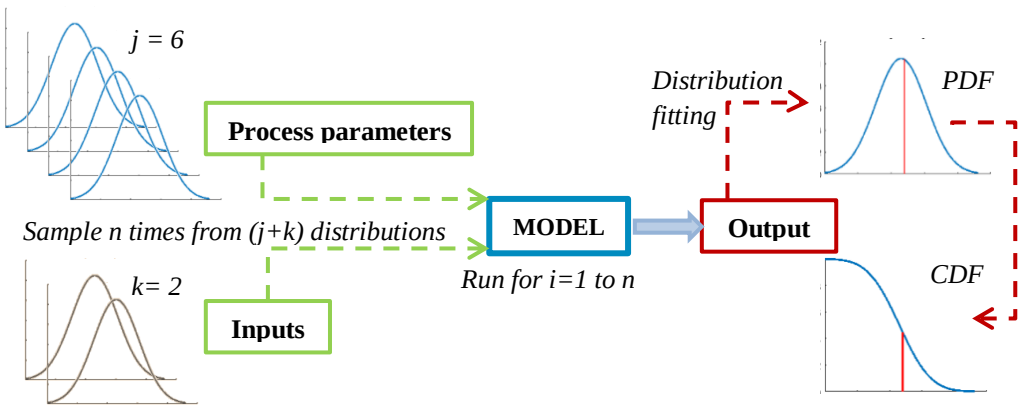

Figure 4. Schematization of the ensemble forecasting methodology. PDF is probability density function and CDF is cumulative distribution function.

SELECTION OF SIGNIFICANT PARAMETERS

The previously introduced ecological processes of the GEM make use of more than 400 model process parameters in total. Incorporating all these parameters into the ensemble generation is not feasible and would not yield a better uncertainty estimate, since most of them do not affect the Chlorophyll-a concentration.

A previous study investigated the sensitivity of the Chlorophyll-a concentration to process parameters in the North Sea application of the GEM (Salacinska *et al.* 2010) and identified 20 parameters that obtained the highest rank. These parameters consist of growth rates, extinction coefficients and nutrient-to-carbon ratios (e.g. P:C, N:C) of the different algal species. Furthermore, according to Li *et al.* (2013) primary production mainly depends on the specific rates of growth, mortality, and maintenance respiration as well as the temperature coefficient because these rates are calculated as a function of the temperature. In addition, Daggers (2013) conducted a sensitivity analysis of the nutrient concentrations influenced by the different processes such as denitrification, nitrification, mineralization and burial. Those findings suggest that in coastal regions nitrate is mainly sensitive to the denitrification rate in the top sediment layer (S1), while phosphate and all other nutrients are mainly sensitive to the burial rate. This extensive list of significant parameters was narrowed down to six parameters (see Table 2) considering the dominating algal type (Diatoms type E) and the coastal environment in the southern North Sea. Further description of the significant parameters can be found below.

Table 2. Selected significant GEM process parameters for ensemble generation

No.	*Parameter*	*Description*	*Constraint*
1	$e_{s,IM}$	Specific extinction coeff. of suspended inorganic matter	*Energy / growth*
2	$kgp^{0}_{Di,E}$	Maximum growth rate of Diatoms type E at 0°C	*Growth*
3	$ktpg_{Di,E}$	Temperature coefficient for growth of Diatoms type E	*Growth*
4	$R_{den,sed}$	Denitrification rate in the sediment	*Nutrient*
5	$R_{den,wat}$	Denitrification rate in the water column	*Nutrient*
6	b_{S1}	Burial rate for layer S1	*Nutrient*

Model process parameters

The Chlorophyll-a concentration C_{chlfa} is a function of the algal biomass concentration $C_{alg,i}$ and the stoichiometry of Chlorophyll-a in the algae $S_{chlfa,i}$ (Eq. 2), thus any parameter that alters the algal biomass will also change the Chlorophyll-a concentration.

$$C_{chlfa} = \sum_{i=1}^{n}(S_{chlfa,i} \times C_{alg,i}) \qquad \text{(Eq. 2)}$$

The algal biomass concentration is calculated using an optimization technique (linear programming) that solves a set of linear equations and constraints. The applied constraints are the energy-, growth-, nutrient- and mortality constraints. This paper focuses on the peak algae bloom prediction which is not limited by the mortality constraint which for this reason is not considered. Therefore, the selected model process parameters can be divided into three main groups depending on whether they affect the energy-, growth- or nutrient constraints as shown in Table 2.

The specific extinction coefficient $e_{s,IM}$ and concentration of suspended inorganic matter C_{IM} (external forcing) affect the light regime in the water which determines the energy availability and growth of the phytoplankton species. In shallow coastal ecosystems, like the Wadden Sea, the concentration of suspended inorganic matter is relatively high and dynamically varying. The extinction of light is due to the particulate and dissolved light absorbing substances in the water such as the algae biomass, detritus and the suspended inorganic matter. Therefore, the total extinction coefficient *et* is calculated as a linear superposition of the partial extinction coefficients of these substances (e.g. partial extinction coefficient of inorganic matter $e_{st,IM}$) and the background extinction of the water $e_{background}$ (Eq. 3). The partial extinction of these substances is expressed with their specific extinction coefficient and their concentration (Eq. 4).

$$et = e_{st,IM} + e_{st,other\ subst} + e_{background} \qquad \text{(Eq. 3)}$$

$$e_{st,IM} = e_{s,IM} \times C_{IM} \qquad \text{(Eq. 4)}$$

The visible light intensity *I* is then described by an exponential attenuation, namely the Lambert-Beer law (Eq. 5), which is a function of the total extinction coefficient et and the water depth H:

$$I_b = I_t \times e^{(-et \times H)} \qquad \text{(Eq. 5)}$$

denoting the light intensity at the bottom of the water column I_b and at the top of the water column I_t. Finally, the actual visible light intensity is converted to a light efficiency factor $E_{f,i}$ which is used in the growth constraint of the GEM (Eq. 6) to determine the maximum concentration of the algae type i $C_{algmax,i}$ for each time-step Δt.

$$C_{algmax,i} = C_{alg,i} \times e^{[(kgp_i \times E_{f,i} - krsp_i) \times \Delta t]} \qquad \text{(Eq. 6)}$$

In the growth constraint equation, kgp_i stands for specific growth rate of algae type i and $krsp_i$ is the specific maintenance respiration rate of algae type i. This specific growth is calculated using the maximum growth rate at 0°C kgp_i^0 together with the water temperature T and temperature coefficient for growth $ktpg_i$ (Eq. 7). Consequently, these parameters also have a direct influence on the algae growth. It should be noted that in the southern North Sea the spring bloom of phytoplankton is dominated by diatoms (Blauw 2015), and the beginning of the peak algal bloom is mainly influenced by the energy limited type (E-type) algae species. The maximum growth rate and the temperature coefficient for growth parameters are therefore selected for the Diatoms type E ($i = Di, E$).

$$kgp_i = kgp_i^0 \times (T - ktpg_i) \qquad \text{(Eq. 7)}$$

The last group of parameters considered for the input ensemble affects the availability of nitrate, phosphate and silicate that are essential nutrients for the algae growth. Thus, perturbation on these nutrient concentrations provides further uncertainty estimation of the Chlorophyll-a concentration. The denitrification rate and burial rate influence the nutrient availability in the water column and in the sediment for the algae species through the denitrification and burial processes. Denitrification removes nitrate NO_3^- from the water and generates elemental nitrogen N_2 which can leave the water phase and escape into the air:

$$4NO_3^- + 4H_3O^+ \Rightarrow 2N_2 + 6H_2O \qquad \text{(Eq. 8)}$$

In the GEM the denitrification process depends on the denitrification rate R_{den}, the nitrate concentration C_{NO3}, and denitrification process temperature $f_{T,den}$ (Eq. 9). It is important to mention that the denitrification process does not take place under a critical temperature T_c.

$$den = \begin{cases} 0.0 & if\ T < T_c \\ R_{den} \times C_{NO3} \times f_{T,den} \end{cases} \qquad \text{(Eq. 9)}$$

Nutrient availability for the living organisms is also influenced by the burial process. The importance of the burial process in the model is to remove the dead particulate organic matter POX_S from the active top sediment layer S1 to the deeper sediment layers. However, the burial rate represents more than just the actual burial process; it also stands for uncertainties in the nutrient loadings and unknowns in the nutrient mass balance. In the model the burial is a function of the burial rate b_{S1}, the total concentration of the dead particulate organic matter and the water depth H (Eq. 10).

$$Burial_{POX_S} = b_{S1} \times \frac{POX_S}{H} \qquad \text{(Eq. 10)}$$

Model forcings

In addition to the model process parameters, certain model forcings are selected to be perturbed based on previous research findings focusing on the southern North Sea coastal ecosystem.

The importance of the concentration of suspended inorganic matter forcing, often referenced as Suspended Particulate Matter (SPM), was previously demonstrated in (Eq. 4). In the continental coastal waters 25 to 75% of the light extinction is caused by suspended particulate matter (Los *et al.* 2010). Furthermore, Perez (2015) investigated the effect of using different types of SPM sources on the Chlorophyll-a concentration in the southern North Sea and the results suggested considerable impact. El Serafy *et al.* (2011) also indicated that the SPM data sources in the North Sea, such as the field measurements and the Delft3D-WAQ-SPM model results, are rather uncertain. Based on these findings it was decided to include the SPM concentration field in the ensemble generation. By reason of simplicity the spatial correlation in the SPM concentration field is not considered, and only a simple error range was applied to the input segment function that contains the suspended particulate matter concentration for each segment at all time-steps. Spatial correlation, however, was previously found to be important in Kamel *et al.* (2014) since the seasonal variation in fine sediment dynamics forms several ellipsoidal-shaped sedimentation traps along the Dutch coast.

Further model forcing affecting the Chlorophyll-a concentration, especially in coastal ecosystems, is the riverine nutrient load. Nutrient loads from rivers provide a significant portion of the total nitrogen load (12-17%) and total phosphorus load (8-11%) in the North Sea (Daggers 2013). Since the information about the river discharges and nutrient loads is often estimated, and the daily discharge values are interpolated from the less frequently available flow data, the riverine nutrient input is also included in the ensemble generation. For practical reasons, instead of directly perturbing the nutrient concentrations, the river discharges are altered, which linearly affects the nutrient concentrations. In this paper only those nine river discharges are perturbed that are identified to be influential in the study area based on the nutrient composition matrix derived by Los *et al.* (2014).

Sensitivity analysis

A simplified sensitivity analysis was carried out in order to comprehend the Chlorophyll-a concentration's response to the parameter value modification and to verify if previous findings can be confirmed in the present case. The sensitivity analysis was a one-factor-at-a-time method. This method allows us to gain information about the model's sensitivity to a specific parameter; however, it does not account for the dependencies between variables since the parameters interact in a non-linear way.

In the sensitivity test, model runs were executed using the baseline, lower-, and upper bound values of all identified parameters. The changes in residuals between the baseline and the scenarios were analysed at one station over one year with particular attention given to the peak concentration and timing. A brief summary of the observed correlation patterns between the selected parameters and the Chlorophyll-a concentration can be found in Table 3. The table demonstrates the type and strength of the correlation between the investigated parameters and the output parameter, also it indicates whether the time of the peak concentration was affected or not. While the results confirm previous findings, it should be

noted that the denitrification rates were found to be less significant compared to other parameters, hence, these might be negligible in deeper coastal waters.

Table 3. Chlorophyll-a concentration's sensitivity to the selected parameters and forcings (correlation patterns)

	Correlation		
Parameter	***During spring bloom***	***After spring bloom***	***Time lag in peak***
$e_{s,IM}$	(--)	(-)	Yes (→)
SPM	(--)	(-)	Yes (→)
$kgp^{0}_{Di,E}$	(++)	(+)	Yes (←)
$ktpg_{Di,E}$	(--)	0	No
$R_{den,sed}$	(-)	(-)	No
$R_{den,wat}$	(-)	(-)	No
b_{S1}	(--)	(--)	No
	Energy limited period	***Nutrient limited period***	
Riverine nutrient load*	0	(++)	No
Q_river (river discharge)*	(--)	(++)	No

Correlation: (-) Negative, (--) Strong negative; (+) Positive, (++) Strong positive; 0 - No correlation
Time lag in peak: (→) Peak later, (←) Peak earlier
** Large spatial variability, depending on distance to river mouth*

SAMPLING PROCEDURE

Parameter statistics

After selecting the significant parameters it is crucial to identify the parameter statistics, such as the possible value range, mean and standard deviations, and to specify the desired probability distributions to obtain through the sampling. The parameter statistics are given in Table 4. The mean values are assumed to be the calibrated baseline values of the deterministic model setup. The parameter ranges for process parameters were mainly found in Salacinska *et al.* (2010) and Blauw *et al.* (2009), whereas for the model inputs value ranges are set according to the estimated uncertainty level. The uncertainty level in the river discharges was assumed as 25% based on a study conducted by Lloyd *et al*. (2016) on discharge and nutrient uncertainty in streams in the UK, which reported observational discharge uncertainties ranging from ±2 to 25%. The standard deviations for four process parameters are used as described in Jiayuan (2015). For the remaining process parameters, and for the model inputs, the standard deviations are assumed in such a way that the resulting sampled values fall between the specified value ranges, given that for normal distribution 99.7% of the data are within three standard deviations of the mean. Information about the parameter dependency was available for four process parameters from Jiayuan (2015), while having no information about the rest of the parameters they are assumed to be independent.

Table 4. Parameter statistics for sampling

No	***Parameter***	***Range***	***Mean***	***Standard deviation***	***Baseline***
1	$e_{s,IM}$	0.01-0.05	0.025	0.0075	0.025
2	$R_{den,wat}$	0.00-0.20	0.006	0.01	0.006

3	b_{S1}	0.00-0.25	0.0025	0.01	0.0025
4	$kgp^{0}_{Di,E}$	0.05-0.15	0.07	0.01	0.07
5	$ktpg_{Di,E}$	(-5.00) - (-1.75)	-4.5	0.15	-4.5
6	$R_{den,sed}$	0.00-0.20	0	0.05	0
7	SPM	0.7-1.3*	1	0.1*	1
8	Q_river	0.75-1.25*	1	0.08*	1

**based on the estimated uncertainty level*

Parameter distributions

In this case study, the selected process parameters are non-measurable and therefore no prior knowledge was available to obtain empirical distributions. Consequently, the theoretical distributions were assumed according to literature (Jiayuan 2015) and general considerations. Based on the literature, exponential distributions were assumed for those parameters which can only take positive values, whereas normal distributions were applied to others. This approach was adopted as best available knowledge on the parameters' behaviour. Moreover, these distributions were adopted partly due to the fact that in the applied sampling procedure a Gaussian Copula model is used, which has the underlying assumption that the joint distributions of the random variables are elliptically symmetric. On the other hand, it is important to note that in future cases, when prior information about the parameters is insufficient, choosing uniform prior distributions might be more appropriate as demonstrated by Freni and Mannina (2010). Applying normal distribution without prior knowledge may result in wrong estimation of uncertainty as it leads to a narrower uncertainty band as compared to uniform distribution (Freni & Mannina 2010; Dotto *et al*. 2012). This may create excessive confidence in the model results especially in water quality modelling applications where the inherent uncertainties are high.

Latin Hypercube Sampling with Dependence (LHSD)

The generated sample has two specific requirements to fulfil: (1) the dependence structure between the variables must be represented, and (2) a good coverage of the parameter space should be achieved with only a limited sample number (variance reduction). The water quality model makes use of numerous model parameters which are in many cases correlated. If the sampling technique ignores those correlations the simulation could result in unrealistic outputs. One way to describe dependency between ecological process variables is to use copulas (Jiayuan 2015). In addition to the parameter dependency, choosing a sampling technique with variance reduction gives a greater precision of the output random variables for a given number of iterations, this way providing a statistically efficient sampling.

Packham and Schmidt (2010) proposed an extension of the Latin Hypercube Sampling (LHS) from independent random vectors to random vectors with dependence (Eq. 11). This extension allows us to make use of the important variance reduction property of the LHS, while maintaining the dependence structure between the random variables. The main idea behind the modification is to choose a specific permutation in each dimension instead of

randomly taking them. The specific permutation depends on the rank of the random variables calculated by rank statistics. If the components of the random vectors are linked by a copula (e.g. Gaussian copula) using the rank statistics the permutation can be constrained, and the sample generated by the Latin Hypercube Sampling with Dependence (LHSD) is given as:

$$V_{i,n}^{j} = \frac{r_{i,n}^{j}-1}{n} + \frac{\eta_{i,n}^{j}}{n}, \; i = 1, \dots, n; \; j = 1, \dots, d \qquad \text{(Eq. 11)}$$

Where $V_{i,n}^{j}$ is the generated LHSD sample, $r_{i,n}^{j}$ is the *i*-th rank statistic of the random vectors and $\eta_{i,n}^{j}$ is a random variable between [0,1] (e.g. $\eta_{i,n}^{j}$= 0.5 means that each sample is in the middle of its stratum). For a thorough description of LHSD the reader should refer to Packham and Schmidt (2010).

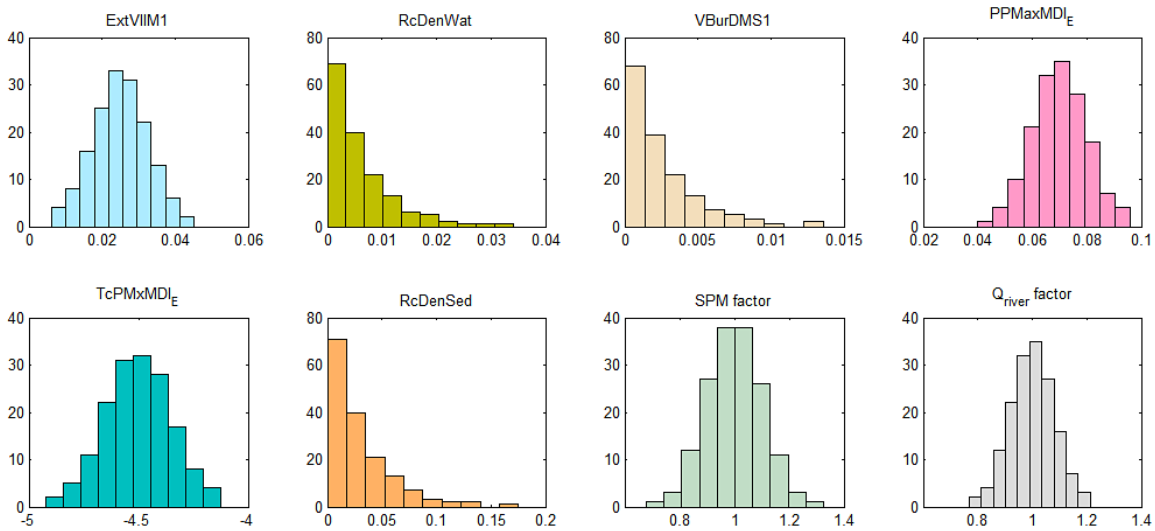

Figure 5. Sample distribution of the selected significant parameters from Latin Hypercube Sampling with Dependence, *n* = 160.

The parameters are transformed to the above-mentioned prior distributions. The sample number *n = 160* was based on Jiayuan (2015) who conducted a simplified analysis to identify the optimal sample number. Applying the LHSD, smooth and evenly distributed sample distributions could be achieved (see Figure 5) while maintaining the parameter dependency. Samples that fall outside of the predefined range were removed in order to avoid values with non-physical meaning.

Distribution fitting

The results of the ensemble simulation runs are transformed into probabilities by fitting probability density functions (PDF) to the ensemble members at each time-step. In the applied distribution fitting method, continuous probability distributions were fitted and ranked in order to obtain the best fit. This fitting procedure does not allow multimodal fitting meaning that the PDF cannot have multiple peaks; this simplification might result in inaccuracies.

FORECAST VERIFICATION AND PERFORMANCE ANALYSIS

With the help of multiple verification metrics, different forecast attributes can be assessed and quantified. In this paper the Mean Absolute Error (MAE), Root Mean Square Error (RMSE) (Harmel & Smith 2007) and Percent Bias (PBIAS) (Gupta *et al.* 1999) are applied to quantify the error between the observation and forecast. In addition, further accuracy measures such as the Brier Score (BS) (Brier 1950) and the Continuous Ranked Probability Score (CRPS) (Brocker 2012) are included in the verification. The CRPS can be calculated for single-valued and probabilistic predictions, allowing a direct comparison between the two types of forecasts. The forecast reliability is assessed with the False Alarm Ratio (FAR), whereas the discrimination attribute is evaluated with the Probability Of Detection (POD) (McGovern *et al.* 2011). Finally, the model's goodness-of-fit is determined by the Index of Agreement (IoA), Coefficient of determination (R^2) and Nash-Sutcliffe (NS) model efficiency (Harmel & Smith 2007).

All the metrics are collected in Table 5 together with their perfect scores. The formulas of the deterministic verification metrics are given in (Eq. 12)-(Eq. 19), whereas the probabilistic verification measures are calculated as shown in (Eq. 20)-(Eq. 21). The computed verification metrics are then used to assess the improvement in forecast performance. Table 6 shows the relationship between the changes in verification metrics, grouped by forecast attributes, and the corresponding changes in forecast performance.

Table 5. Verification metrics used for performance analysis

Verification metrics	***Formulas***		***Perfect score***
Index of Agreement	$d = 1 - \frac{\sum_{i=1}^{n}(o_i - f_i)^2}{\sum_{i=1}^{n}(\lvert f_i - \bar{o}\rvert + \lvert o_i - \bar{o}\rvert)^2}$	(Eq. 12)	1
Coefficient of determination	$r^2 = \frac{[cov(f,o)]^2}{[var(f)][var(o)]}$	(Eq. 13)	1
Nash-Sutcliffe	$E = 1 - \frac{\sum_{i=1}^{n}(o_i - f_i)^2}{\sum_{i=1}^{n}(o_i - \bar{o})^2}$	(Eq. 14)	1
Mean Absolute Error	$MAE = \overline{(\lvert f - o\rvert)}$	(Eq. 15)	0
Root Mean Square Error	$RMSE = \sqrt{\overline{(f - o)^2}}$	(Eq. 16)	0
Percent Bias	$PB = 100 \times \left[\sum_{i=1}^{n}(f_i - o_i) \Big/ \sum_{i=1}^{n} o_i\right]$	(Eq. 17)	0
False Alarm Ratio (FAR)	$FAR = \frac{False\ alarms}{Hits + False\ alarms}$	(Eq. 18)	0
Probability of Detection (POD)	$POD = \frac{Hits}{Hits + Misses}$	(Eq. 19)	1
Continuous Ranked Probability Score (CRPS)	$\int_{x=-\infty}^{x=\infty}\left(F_i^f(x) - F_i^o(x)\right)^2 dx$ F_i^f - CDF of forecast, F_i^o- CDF of observation	(Eq. 20)	0
Brier Score (BS)	$BS = \overline{(f - o)^2}$, where $f = [0,1], o = 0\ or\ 1$	(Eq. 21)	0

Some of the above-mentioned metrics such as the POD and FAR use the comparison of the measured and forecast occurrence frequency of a predefined event. In this study an event is defined when the Chlorophyll-a concentration in the top layer of the water column exceeds the elevated assessment levels. The mean elevated assessment level during the growth season (March-September) is 7.5 mg/cm^3 in the Dutch coastal waters and 12 mg/cm^3 in the Dutch Wadden Sea. These levels were previously specified by Baretta-Bekker *et al.* (2008) for the application of the OSPAR Comprehensive procedure.

Table 6. Relationship between changes in verification metrics and forecast performance

Performance improvement	***Verification metrics***
Improved goodness-of-fit (regression line of the forecast better approximates the observations)	(↑) Index of Agreement (IoA) (↑) Coefficient of determination (R^2) (↑) Nash-Sutcliffe (NS)
Improved accuracy (reduced error measures)	(↓) Mean Absolute Error (MAE) (↓) Root Mean Square Error (RMSE) (↓) Percent Bias (PBIAS) (↓) Continuous Ranked Probability Score (CRPS) (↓) Brier Score (BS)
Improved reliability	(↓) False Alarm Ratio (FAR)
Improved discrimination attribute	(↑) Probability of Detection (POD)

(↑) - Increase, (↓) - decrease

RESULTS AND DISCUSSION

The above-presented ensemble method was tested for hindcasting the Chlorophyll-a concentration in the southern North Sea for different hydrodynamic years, first for year 2009 and then for year 2007. The ensemble method's performance is analysed separately in the Dutch Wadden Sea and Dutch coastal waters using three observation types with varying temporal coverage and accuracy. In-situ measurements are considered as the most reliable data but their temporal coverage is low, only 15 measurements per year on average. In comparison, the remote sensing data has increased temporal coverage, 30 MERIS observations per year and 200 gap-filled MERIS observations per year on average, allowing better verification of the model but introducing more uncertainty.

Ensemble forecasts are often assessed by verifying how well the ensemble band could capture the observations. Figure 6 and Table 7 show the percentage of measurements captured by the ensemble band and the confidence intervals (50%, 75%, 95%) at the in-situ stations. Overall, it can be concluded that the percentage of measurements lying within the ensemble band is relatively low, 60% on average considering all stations and both years. The percentage results for the 95% and 75% confidence intervals are even lower, but the accuracy of the former is still acceptable for the research. Nevertheless, particular differences can be observed in the ensemble forecast's skill once the two areas are analysed separately. In the Wadden Sea only 35 to 57% of the measurements could be

captured in the ensemble band while in the coastal waters this percentage is as high as 67 to 80%.

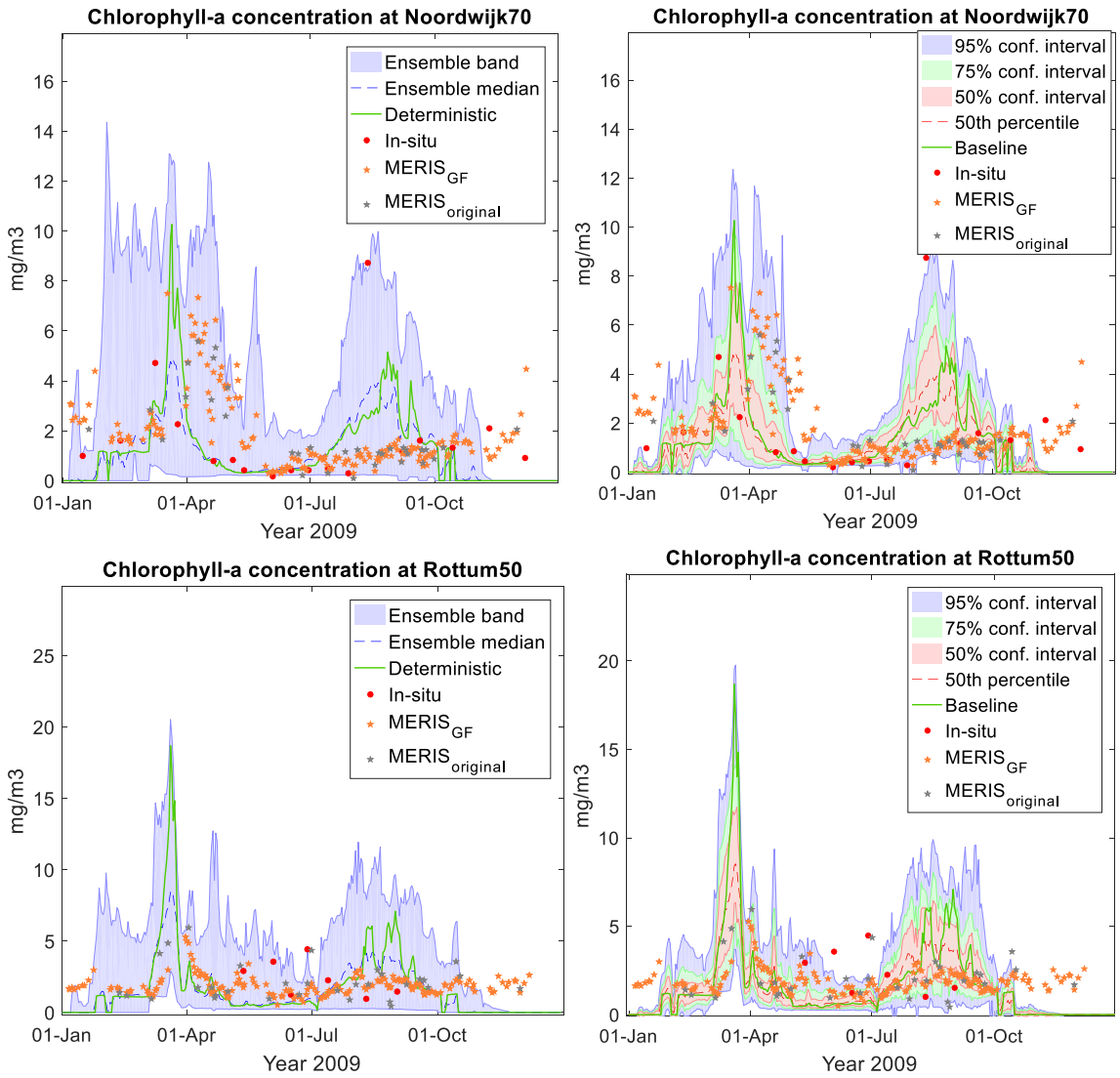

Figure 6. Ensemble band (left) and confidence intervals (right) with the measurements at station Noordwijk 70 and Rottum 50, year 2009.

In this paper, processed remote sensing images are also applied as a verification dataset. Given the high degree of uncertainty in this dataset, it is advised to provide an error estimate together with the results. The MERIS data is compared with the in-situ measurements to obtain the error estimate at the stations where a sufficient number of matchups are available. Considering all stations, the MERIS dataset has an average Mean Absolute Error (MAE) of 41% in 2009 and 48% in 2007, whereas the DINEOF gap-filled MERIS data's average error is 52% and 48% for the same years. Thus, applying averaged error bands for the remote sensing dataset instead of the uncertain single point values might provide a better picture of the ensemble forecast's performance. The reader may refer to Harmel and Smith (2007) for further information about considering measurement uncertainty in the evaluation of goodness-of-fit metrics in water quality modelling.

Table 7. Percentage of measurements captured in the ensemble band at the stations, and average results for the 95%, 75% confidence intervals, years 2009 and 2007

No.	*Station*	*In-situ [%]*	*MERIS [%]*	*MERIS_GF [%]*	*In-situ [%]*	*MERIS [%]*	*MERIS_GF [%]*
			2009			*2007*	
				Dutch Wadden Sea			
1	Marsdiep	45.0	28.6	45.9	25.0	58.8	56.1
2	DoovBW	33.3	70.4	61.3	n/a	n/a	n/a
3	Vlies	58.3	51.6	51.9	36.4	50.0	44.9
4	Dantzig	19.0	50.0	41.4	26.3	30.8	30.7
5	ZuidOlWot	42.9	41.7	21.0	57.9	70.6	75.6
6	Harling	11.1	0.0	16.6	50.0	74.1	74.1
Average (ensemble band)		**34.9**	**40.4**	**39.7**	**39.1**	**56.9**	**56.3**
Average (95% conf. int.)		**23.8**	**35.5**	**30.3**	**24.8**	**45.3**	**43.4**
Average (75% conf. int.)		**13.9**	**21.7**	**16.6**	**16.3**	**29.4**	**27.0**
				Dutch coastal waters			
7	Rottum3	66.7	80.0	70.7	70.0	76.1	87.3
8	Rottum50	85.7	90.6	86.7	100.0	74.5	82.9
9	Rottum70	57.1	75.0	76.2	100.0	72.3	80.0
10	Terschelling10	61.1	87.5	81.8	64.3	64.5	62.4
11	Terschelling50	82.4	88.6	84.0	n/a	n/a	n/a
12	Noordwijk10	58.6	72.2	69.1	27.6	35.5	42.9
13	Noordwijk20	47.1	70.8	56.9	38.9	44.7	40.0
14	Noordwijk70	77.8	90.9	87.8	78.9	76.0	85.9
15	Goeree6	66.7	65.2	69.1	63.6	66.7	78.5
Average (ensemble band)		**67.0**	**80.1**	**75.8**	**67.9**	**63.8**	**70.0**
Average (95% conf. int.)		**54.1**	**68.2**	**62.8**	**51.3**	**53.6**	**57.5**
Average (75% conf. int.)		**33.6**	**43.7**	**42.2**	**38.8**	**32.0**	**36.0**

Taking into account the measurement uncertainty in the remote sensing data, the percentage of captured measurements would increase considerably up to 84% in the Wadden Sea, and up to 89% in the coastal waters, however, 100% coverage could not be achieved. A possible reason for this might be that the coastal zone and the Wadden Sea are shallow, dynamically varying ecosystems with high turbidity and therefore the deterministic simulation, which serves as the base for the ensemble forecast, should be re-calibrated here to achieve improved results. This model uncertainty should be considered when evaluating the results. Moreover, the fact that not all uncertainty sources were taken into account may explain the additional uncertainty stemming from the model structure and from the meteorological- and hydrodynamic inputs. Finally, the presumed level of uncertainty in the inputs (e.g. SPM concentration field), might be underestimated or their assumed sample distribution may not be appropriate.

Forecast verification metrics

The direct comparison of the deterministic and ensemble forecast was achieved through calculating a selection of verification metrics, (introduced above) for both types of predictions. In order to be able to compute deterministic scores for the ensemble forecast, the ensemble median (50-th percentile) prediction was selected in this study, although further options are available such as the ensemble mean or any other percentile. The ensemble median (or ensemble mean) usually verifies better than the deterministic forecast by most verification scores, because it presents the most predictable elements of the

forecast and smooths out the extreme unpredictable elements. It indicates the future values of the model output variable that can be predicted with confidence, but it will rarely capture the extreme events, and therefore should not be relied upon on its own.

The percentage improvement in the groups of verification metrics if the ensemble median prediction is used instead of the deterministic forecast is presented in Table 8. It is important to note that all metrics are equally weighted within the groups. Nevertheless, it might be possible that in other studies the metrics would be weighted or some specific metrics would be omitted according to the forecasters' need. The group members are shown in Table 6. The verification metrics are averaged values of six stations in the Dutch Wadden Sea and nine stations in the coastal waters. Furthermore, the percentage improvements in the metrics are computed as an average of the results using all measurement types.

Table 8. Averaged % improvement in the groups of verification metrics if the *ensemble median* prediction is used instead of *deterministic forecast*

	% improvement*			
	2009		***2007***	
Group of verification metrics	***Dutch coastal waters***	***Dutch Wadden Sea***	***Dutch coastal waters***	***Dutch Wadden Sea***
Goodness-of-fit	28%	4%	18%	12%
Accuracy	16%	8%	12%	7%
Reliability	7%	-13%	0%	-4%
Discrimination	6%	-26%	1%	-12%

** Group average using the three validation datasets. The metrics within the groups are not weighted*

The goodness-of-fit and accuracy metrics show moderate improvement in both years and both areas but the magnitude of the improvement is markedly higher in the coastal waters nonetheless. The Brier Score is only calculated for the ensemble forecast and thus it cannot be compared to the deterministic prediction. Considering all measurement types the average Brier Score in the Wadden Sea is 0.13 in 2009 and 0.12 in 2007, whereas at the Dutch Wadden Sea stations the Brier Scores for the same years are 0.05 and 0.09. Reliability and discrimination measures only experience a minor improvement in the coastal waters and no improvement at all in the Wadden Sea. The verification results confirm the previous finding that the ensemble forecast performs better in the Dutch coastal waters than in the Wadden Sea. In the problematic Wadden Sea area the ensemble method's efficiency is less convincing; on the other hand, in the coastal waters the preliminary results are promising.

Spatial results

Making use of the three-dimensional water quality model the forecast can be visualized as a Chlorophyll-a concentration map. As mentioned above, the ensemble median prediction should not be relied upon on its own but with the indication of the degree of certitude. The ensemble spread can provide a measure of the level of uncertainty in the output parameter; hence, it is appropriate to complement the ensemble median. Figure 7 illustrates the

ensemble median prediction map at a specific time-step together with the ensemble spread map calculated as the standard deviation of the Chlorophyll-a concentration.

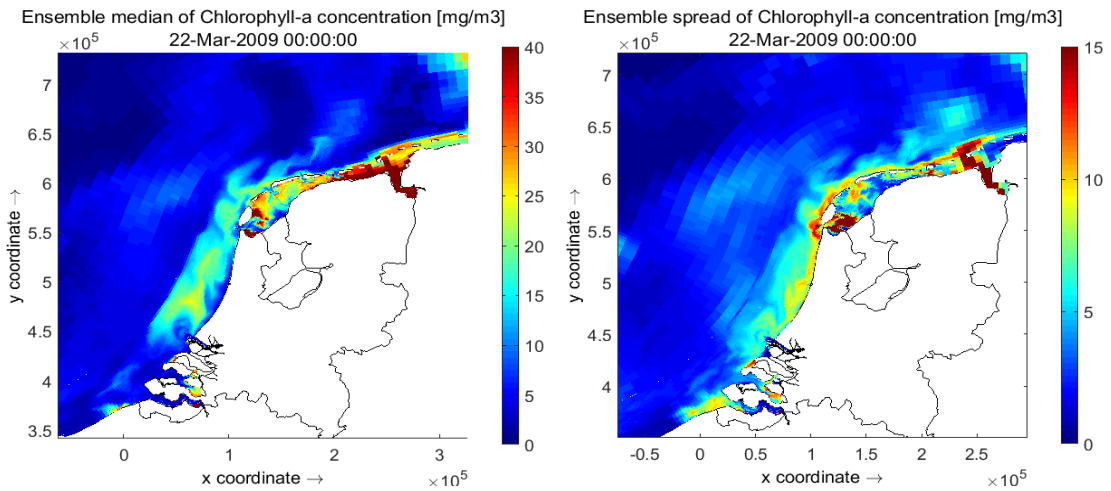

Figure 7. Ensemble median prediction (left) and corresponding ensemble spread (right) of Chlorophyll-a concentration along the Dutch coast during the peak spring algal bloom in 2009.

Observing the ensemble spread map allows us to draw conclusions on the prediction's spatial uncertainty. In late March, during the peak spring algal bloom, the prediction in most off-shore areas has low standard deviation, while significant spread can be observed in the near shore, shallow areas of the North Sea and in the Dutch Wadden Sea. This might be partly due to the direct effect of perturbations on the river loads in the focus area and the high spread may indicate the rivers' zone of reach. Nonetheless, it could be also explained with the specific regional system dynamics, since in the near shore shallower zones the water is low-dynamic and the primary production levels are already elevated.

CONCLUSIONS

This paper can contribute to existing knowledge in probabilistic forecasting by providing an application to water quality prediction with a three-dimensional ecosystem model. The promising results indicate that ensemble prediction techniques can produce enhanced forecast of water quality indicators due to their ability to account for the error in the model output variable. Therefore, it is suggested that in the future, by further developing the proposed forecast set-up, the implementation of operational ensemble forecasting systems for coastal water quality prediction is viable. Nevertheless, the potential of the ensemble forecasting system ultimately depends on the input ensemble. Moreover, the sampling technique that is used to generate the input ensemble should be tailored to the specific requirements of the application.

In the presented case study, the ensemble forecast moderately outperforms the deterministic forecast at the coastal stations, which might be advantageous in decision making if the underlying baseline (deterministic) model is sufficiently well calibrated and validated. Moreover, the degree of certitude and the likelihood of an event, which could be expressed through ensemble forecasting, provide opportunity to set risk-based criteria for

the response measures. In other words, the uncertainty estimate produced by the proposed ensemble forecast promotes rational decision making, and offers potential for additional ecosystem and economic benefits.

Recommendations

Due to their fundamental importance the identified significant parameters should be revised based on a comprehensive sensitivity analysis. In addition, the parameter value ranges and prior distributions should be re-evaluated for future applications. Furthermore, the uncertainty from the meteorological- and hydrodynamic inputs could also be included in the ensemble generation in order to better estimate the model input uncertainty. The spatial correlation of the SPM concentration field should be also considered to achieve a more realistic SPM perturbation.

Further necessary advancement of the presented ensemble prediction system would be to develop the method into an operational forecasting system. This could be achieved by implementing the proposed technique in a platform (e.g. Delft-FEWS) which helps to organize the integrated modelling system elements and the various types of observations from data importing and processing through the ensemble simulation with the GEM to the model output post-processing. Finally, a statistical post-processing technique could complement the ensemble forecast in order to estimate the total predictive uncertainty. This approach includes statistical post-processing uncertainty estimation for each ensemble member independently for each lead time. The product is a separated probability density function for each ensemble member which can then be averaged to obtain the total predictive uncertainty.

ACKNOWLEDGEMENT

This study was conducted within the framework of the ECOPOTENTIAL project. This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762. The authors would like to express their gratitude for the contribution that was provided by Alexander Ziemba.

REFERENCES

Ahn, J. M., Lee, S. J. & Kang, T. 2016 Development of water quality forecasting system with ensemble stream prediction method in the Geum River Basin, Korea. *Desalination and Water Treatment*, **57**(2), 670-683. doi:10.1080/19443994.2014.996010

Arentz, L., Harezlak, V., van Kessel, T. & van der Kaaij, T. 2012 *Kalibratie slibtransport- en GEM model (Calibration of the sediment transport- and GEM model)*, Report 1205620-000-ZKS-0014, Deltares, Delft, The Netherlands.

Baretta-Bekker, H., Bot, P., Princ, P., Prins, T. & Zevenboom, W. 2008 *Report on the second application of the OSPAR Comprehensive Procedure to the Dutch marine waters*, OSPAR Commission, The Netherlands.

Blaas, M. 2013 *Eutrophication Assessment Using Remotely Sensed and In Situ Chlorophyll-a Data*, Report 1207729-001-ZKS-0003, Deltares, Delft, The Netherlands.

Blauw, A. N. 2015 *Monitoring and prediction of phytoplankton dynamics in the North Sea.* PhD thesis, Faculty of Science (FNWI), Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, Amsterdam, The Netherlands.

Blauw, A. N., Los, H. F., Bokhorst, M. & Erftemeijer, P. L. 2009 GEM: a generic ecological model for estuaries and coastal waters. *Hydrobiologia*, **618**(1), 175-198. doi:10.1007/s10750-008-9575-x

Brier, G. W. 1950 Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, **78**(1), 1-3. doi:10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

Brocker, J. 2012 Evaluating raw ensembles with the continuous ranked probability score. *Quarterly Journal of the Royal Meteorological Society*, **138**(667), 1611-1617. doi:10.1002/qj.1891

Daggers, T. 2013 *Validation of marine primary production model for the North Sea using in-situ data.* Msc thesis, Joint Programme Sustainable Development, Utrecht University, Utrecht, the Netherlands.

Dotto, C. B. S., Mannina, G., Kleidorfer, M., Vezzaro, L., Henrichs, M., McCarthy, D. T., Freni, G., Rauch, W., Deletic, A. 2012 Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling. *Water Research*, **46**(8), 2545-2558. doi:10.1016/j.watres.2012.02.009

El Serafy, G. Y., Eleveld, M. A., Blaas, M., van Kessel, T., Gaytan Aguilar, S. & Van der Woerd, H. J. 2011 Improving the description of the suspended particulate matter concentrations in the southern North Sea through assimilating remotely sensed data. *Ocean Science Journal*, **46**(3), 179-204. doi:10.1007/s12601-011-0015-x

El Serafy, G., Blaas, M., Eleveld, M. A. & van del Woerd, H. J. 2007 Data assimilation of satellite data of suspended particulate matter in Delft3D-WAQ for the North Sea. *EUMETSAT/AMS Conference*, 1-8. Darmstadt, Germany.

Fiechter, J. 2012 Assessing marine ecosystem model properties from ensemble calculations. *Ecological Modelling*, **242**(2012), 164-179. doi:10.1016/j.ecolmodel.2012.05.016

Freni, G. & Mannina, G. 2010 Bayesian approach for uncertainty quantification in water quality modelling: The influence of prior distribution. *Journal of Hydrology*, **392**(1-2), 31-39. doi:10.1016/j.jhydrol.2010.07.043

Gupta, H. V., Sorooshian, S. & Yapo, P. O. 1999 Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *Journal of Hydrologic Engineering*, **4**(2), 135-143. doi:10.1061/(ASCE)1084-0699(1999)4:2(135)

Harmel, R. D. & Smith, P. K. 2007 Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling. *Journal of Hydrology,* **337**(3-4), 326-336. doi:10.1016/j.jhydrol.2007.01.043

Herrmann, O.-C. 2015 *Validation of Ensemble Forecast Accuracy in Integrated Modelling in the North Sea. Identification of potential improvements of the ensemble forecast accuracy for hydrodynamics and water quality applications.* Msc thesis, EUROAQUAE Hydroinformatics and Water Management, Cottbus, Germany.

Huang, J. & Gao, J. 2017 An ensemble simulation approach for artificial neural network: An example from chlorophyll a simulation in Lake Poyang, China. *Ecological Informatics*, **37**(2017), 52-58. doi:10.1016/j.ecoinf.2016.11.012

Jiayuan, C. 2015 *Framework for Assessing Uncertainty in Ecological Risk Mapping. A Case Study of North Sea in ECOSTRESS Project.* MSc thesis, UNESCO-IHE Institute for Water Education, Delft, The Netherlands.

Kamel, A. M., El Serafy, G. Y., Bhattacharya, B., van Kessel, T. & Solomatine, D. P. 2014 Using remote sensing to enhance modelling of fine sediment dynamics in the Dutch coastal zone. *Journal of Hydroinformatics,* **16**(2), 458-476. doi:10.2166/hydro.2013.211

Keetels, G., Harezlak, V., van Kessel, T., van Rooijen, A., Friocourt, Y. & van der Kaaij, T. H. 2012 *Winning suppletiezand Noordzee 2013-2017 (Sand mining North Sea 2013-2017),* Report 1204963-000-ZKS-0038, Deltares, Delft, The Netherlands.

Li, Z., Chen, Q., Xu, Q. & Blanckaert, K. 2013 Generalized likelihood uncertainty estimation method in uncertainty analysis of numerical eutrophication models: Take BLOOM as an example. *Mathematical Problems in Engineering,* **2013**(4). doi:10.1155/2013/701923

Lloyd, C. E., Freer, J. E., Johnes, P. J., Coxon, G. & Collins, A. L. 2016 Discharge and nutrient uncertainty: implications for nutrient flux estimation in small streams. *Hydrological Processes*, **30**(1), 135-152. doi: 10.1002/hyp.10574

Los, F. J. & Blaas, M. 2010 Complexity, accuracy and practical applicability of different biogeochemical model versions. *Journal of Marine Systems,* **81**(1-2), 44-74. doi:10.1016/j.jmarsys.2009.12.011

Los, F. J., Troost, T. A. & van Beek, J. K. 2014 Finding the optimal reduction to meet all targets—Applying linear programming with a nutrient tracer model of the North Sea. *Journal of Marine Systems,* **131**, 91-101. doi:10.1016/j.jmarsys.2013.12.001

Los, F. J., Villars, M. T. & Van der Tol, M. W. 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**(1-2), 259-294. doi:10.1016/j.jmarsys.2008.01.002

McGovern, A., Rosendahl, D. H., Brown, R. A. & Droegemeier, K. K. 2011 Identifying predictive multi-dimensional time series motifs: an application to severe weather prediction. *Data Mining and Knowledge Discovery,* **22**(1), 232-258. doi:10.1007/s10618-010-0193-7

Packham, N. & Schmidt, W. 2010 Latin hypercube sampling with dependence and applications in finance. *Journal of Computational Finance,* **13**(3), 81-111. doi:10.21314/JCF.2010.224

Perez, N. C. 2015 *Validation of ecological models: DELFT3d-GEM.* Msc thesis, EUROAQUAE Hydroinformatics and Water Management, Cottbus, Germany.

Salacinska, K., El Serafy, G. Y., Los, F. J. & Blauw, A. 2010 Sensitivity analysis of the two dimensional application of the Generic Ecological Model (GEM) to algal bloom prediction in the North Sea. *Ecological Modelling*, **221**(2), 178-190. doi:10.1016/j.ecolmodel.2009.10.001

Trolle, D., Elliott, A., Mooij, W. M., Janse, J. H., Bolding, K., Hamilton, D. P. & Jeppesen, E. 2014 Advancing projections of phytoplankton responses to climate change through ensemble modelling. *Environmental Modelling & Software*, **61**(2014), 371-379. doi:10.1016/j.envsoft.2014.01.032

van der Woerd, H. J. & Pasterkamp, R. 2008 HYDROPT: A fast and flexible method to retrieve chlorophyll-a from multispectral satellite observations of optically complex coastal waters. *Remote Sensing of Environment*, **112**(4), 1795-1807. doi:10.1016/j.rse.2007.09.001