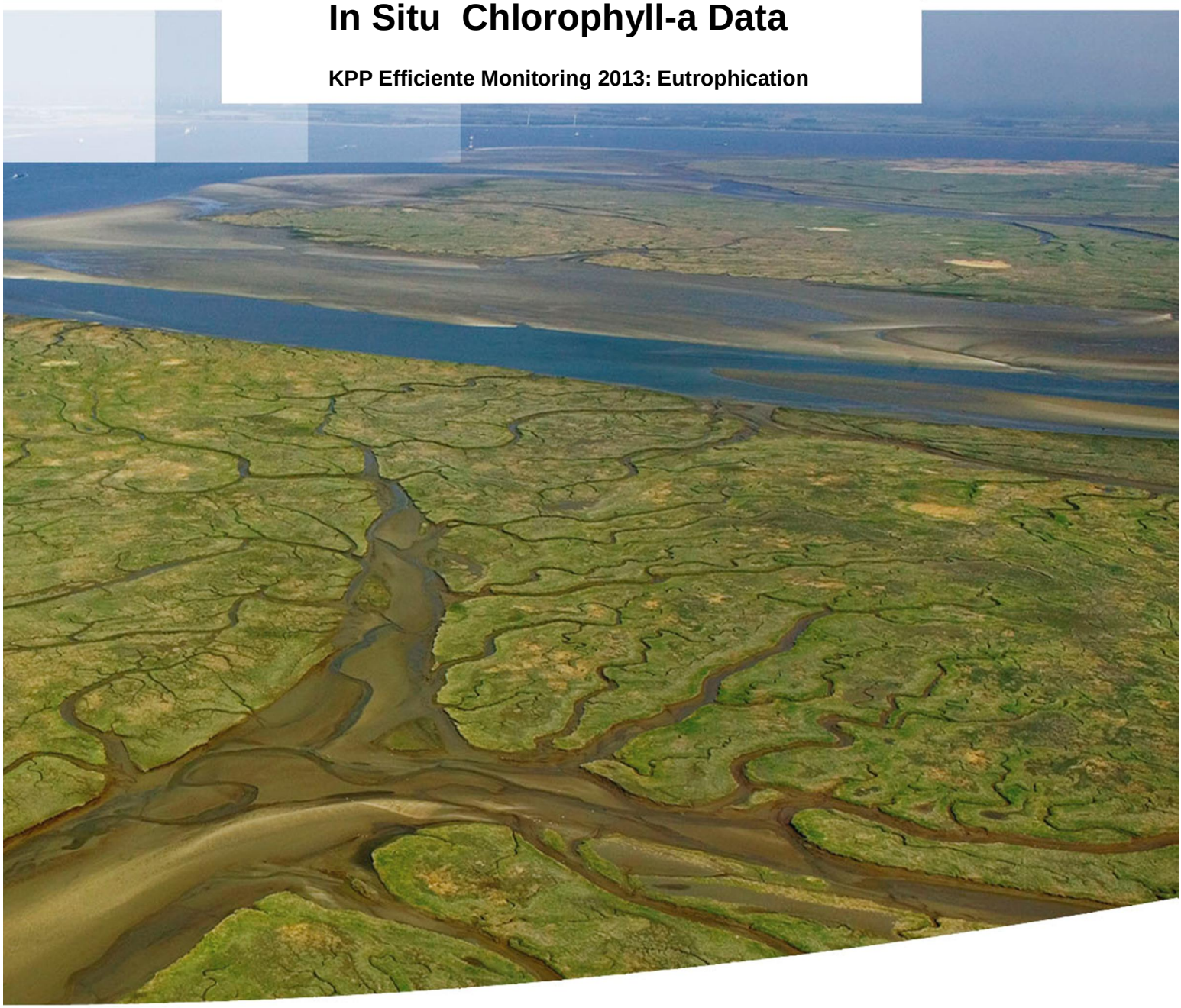
Deltares

Enabling Delta Life

Eutrophication Assessment Using Remotely Sensed and In Situ Chlorophyll-a Data

KPP Efficiente Monitoring 2013: Eutrophication



Deltares

Title
Eutrophication Assessment Using Remotely Sensed and In Situ Chlorophyll-a Data

Client	Project	Reference	Pages
RWS WVL	1207729-001	1207729-001-ZKS-0003	109

Keywords
Monitoring, eutrophication, OSPAR, MSFD, Remote Sensing, Ship-borne Observations, chlorophyll-a

Summary

National marine environmental monitoring is carried out for multiple purposes. One of these purposes is the assessment of the ecological quality of the marine waters for the Marine Strategy Framework Directive and OSPAR convention, with eutrophication as one of the descriptors. The eutrophication status is determined from a statistical analysis of several variables, amongst which chlorophyll-a concentration. This study compares the a typical chlorophyll analysis as part of an OSPAR Comprehensive Procedure based on the Dutch national *in situ* monitoring (MWTL) to one based on ocean colour remote sensing. The analysis is based on historic chlorophyll-a retrieved with the HYDROPT algorithm from MERIS reduced resolution reflectance data in comparison to the standard MWTL data by RWS.

The comparison between chlorophyll–a based on remote sensing and the *in situ* data in the context of a typical OSPAR eutrophication assessment shows that both sources of information have their strengths and limitations. In situ measurements can be taken relatively precise but are limited in spatial and temporal representation of the variability. EOF analysis shows that the MWTL network and sampling scheme is suitable for capturing the large-scale, slowly varying features of the North Sea system in two significant modes, but less the specific fluctuations that contain part of the essential information. The remote sensing data have been calibrated to a large degree towards the basin-wide characteristics of the in situ data and have much wider spatial and –due to spatio-temporal correlations- also higher effective temporal resolution. The EOF analysis shows that the data contain over 25 significant modes, of which the first 3 explain at least over 2% of variance and can be interpreted in terms of system dynamics. Still, the retrieval calibration is using global parameters and, hence, the data may suffer from local biases when optical properties of constituents in the water in certain regions deviate from the mean.

Possible biases notwithstanding, the spatial coverage of the remote sensing data is providing a more accurate way of estimating regional statistical properties of the ecosystem compared to point-based (station-wise) assessments. The use of spatially covering data leads to more stable estimates: year-to-year variations and uncertainties in the region-wise characteristics are smaller compared to the station-wise assessment. For this particular analysis, the region-wise assessment leads to a general reduction of mean and 90-percentile values in the assessment outcome. This thus leads to a fewer potentially problematic exceedances of chlorophyll levels.

The ultimate recommendation for RWS is to start implementing ocean colour remote sensing as a baseline data source in its monitoring strategy from 2014 onward. Hence the implementation should start with RWS defining the accuracy demands for their purposes. The data and service providers can then be invited to develop demand-driven services of remote sensing data with fit for purpose accuracy.

Version	Date	Author	Initials	Review	Initials	Approval	Initials
1	18 Nov.2013	M. Blaas					
2	19 Dec.2013	M. Blaas		H.F.R. van den Boogaard		F.M.J. Hoozemans	

State
final

Contents

1 Introduction

1.1 Context and aim

The Dutch Rijkswaterstaat (RWS) is responsible for monitoring a broad range of variables related to the water quality and water quantity of inland and marine surface waters. For the Dutch national government and in particular for RWS there is a need for more efficiency and effectiveness of these monitoring activities since information requirements are changing, developments in monitoring and data technology are ongoing and budgets are shrinking.

This report is part of the KPP programme *Efficiente Monitoring 2013*, which is overarching applied research by Deltares for RWS within the theme of *Monitoring & Modelling*. Within this programme, various research questions have been formulated in order to help RWS decide on and design more effective and efficient monitoring of its surface waters. This specific report aims to provide insight in the implications of incorporating ocean colour remote sensing (RS) as a source of information for the Dutch national eutrophication monitoring. In particular, it is focusing on the spatio-temporal coverage and statistical properties of the traditional observing network compared to an observation strategy extended with remote sensing. In 2014 the wider KPP programme will also evaluate other data sources relevant for eutrophication monitoring such as FerryBox, SmartMooring etc. The insights of these studies combined should contribute to a possible redesign of the marine monitoring strategy in the coming years.

It is remarked here that the RWS national monitoring programme (MWTL) is serving multiple information needs at the same time. The surveys are not only directed at eutrophication monitoring, but also at other variables that constitute compound indicators for national and international legislation. In the current project, we explore the eutrophication monitoring and even that with a specific focus on one variable, chlorophyll-a. Nevertheless, the lessons learned from this exercise are more generic and will be input into a wider evaluation and advice in marine water quality monitoring in 2014.

For the current study, key aspects are the comparison of information content, spatio-temporal representation and accuracy. In a companion study, also on behalf of RWS and carried out by Baretta-Bekker Marine Ecology, a comparison is made between chlorophyll data based on the ocean colour remote sensing and the standard Dutch *in situ* (IS) data. This comparison is done in the context of the OSPAR (and maybe in future MSFD) eutrophication assessment. The ultimate aim of this study is to indicate to what degree different chlorophyll-a observing strategies give similar or different information (with different accuracy) of the North Sea system. Specific questions for the current report are:

1) What is the quality of the information derived from ocean colour remote sensing?

2) How can ocean colour remote sensing contribute to the information required for national eutrophication monitoring?

3) What are the consequences of adopting remote sensing as part of the eutrophication monitoring in terms of the information and observation strategies?

1.2 Background

Various studies have been carried out in the last decade on the monitoring strategy for water quality. The OPTIMON studies have been broad in scope (Laane, 2013a,b &, Laane et al., *in prep.*) discussing not only eutrophication (nutrients, chlorophyll etc.) but also chemical and abiotic water quality and (pollutants, turbidity) with different means of monitoring (such as ship-borne sensors, buoys and remote sensing).

The ToRSMoN project focused on exploration of the use of ocean colour remote sensing to support the monitoring of SPM and chlorophyll-a and was based on SeaWIFS-satellite data from 1997-2004 (Roberti & Zeeberg, 2007; ARGOSS, 2007, Eleveld & Van der Woerd, 2006). Besides, related to the ToRSMoN work, an inventory of the needs for ocean colour RS data (Zeeberg & Roberti, 2007) and the quality assurance (QA) and acquisition procedure (Zeeberg, 2007) for Rijkswaterstaat had been outlined. Not only the organisational but also the methodological aspects have been explored on behalf of RWS (AGI and RIKZ at that time) in studies of validation of ocean colour remote sensing (Dury et al., 2004, Duin et al., 2005, 2006; Uhlig et al., 2007). Despite the ambitions of ToRSMoN, which covered many aspects of the marine water quality monitoring, it never reached the level of implementation. Note that already over 20 years ago Rijkswaterstaat participated in applied research to study the application of ocean colour remote sensing for its purposes (e.g. Allewijn et al., 1994). Over 20 years, the field of remote sensing research and technology has matured and acceptance and the urgency for innovation of the monitoring strategies has grown. Hence, the objective of the current overarching KPP programme mentioned above is to update the results from ToRSMoN, OPTIMON and other studies to help bridge the gap towards implementation.

In 2008, RWS Waterdienst initiated the RESMON-OK project in order to further implement the acquisition of ocean colour remote sensing by RWS for monitoring purposes (Blaas, 2008). The acquisition of these data had become a requirement for RWS because two monitoring projects related to the construction of the 2nd Maasvlakte decided to use the available, and by then well-developed, MERIS-based ocean colour data products (SPM and chlorophyll-a concentrations at 1x1 km^2 resolution) for their information requirements. Apart from these project-based information needs, the expectation was and still is, that these and similar RS data also provide a valuable information source for regular RWS monitoring. The current report aims to provide an evaluation of this in the context of eutrophication monitoring.

The series of RESMON-OK projects was aimed at the development of validation methods for ocean colour remote sensing data (Westerhoff et al., 2010; De Boer et al., 2012). It supported the acquisition of RS data on the market (Arentz, 2010; Arentz et al., 2011), but gradually it also included monitoring strategy issues. As such it resulted in a vision on the North Sea water-quality monitoring integrating remote sensing with traditional (bottle) and sensor-based ship-borne sampling and smart-moorings (Blaas et al., 2012). Comparable integrating studies have been and are also carried out abroad, such as by Mills et al. (2005) and Kröger et al. (2010).

Various approaches exist in the retrieval of ocean colour RS data. Conceptually, there are the more empirical (e.g. neural network) approaches such as by Doerffer & Schiller (2007) and approaches more directly based on the optical properties of the substances in the water such as Nechad et al. (2010) and Van der Woerd and Pasterkamp (2008). See Tilstone et al., 2012 for an overview of North Sea ocean colour retrieval developments.

The aim of the current study is not to reiterate these earlier studies on the supply side of observational data, platforms and retrieval validation, but to evaluate the main characteristics of an example set of already archived state-of-the-art ocean colour RS data purchased by RWS mostly in terms of their sampling characteristics in comparison to the standard (RWS) *in situ* data of MWTL. The effects of different sampling will be illustrated in the context of the OSPAR eutrophication assessment. In this way the implications of including ocean colour remote sensing in an updated national water-quality monitoring strategy become more specific.

For the current project, the choice was to explore the chlorophyll-a products based on existing Reduced Resolution (1x1 km^2) data from the MERIS sensor processed with HYDROPT (Van der Woerd and Pasterkamp, 2008). Reduced resolution data are the standard product provided by ESA. These have been composed of underlying Full Resolution (300x300 m^2) pixel data. Figure 1.1 illustrates the typical patterns of optically active substances in European shelf seas as recorded by MERIS in Full Resolution. It should be noted that since spring 2012 MERIS data are not supplied any more by ESA and that the currently selected historic data set above all serves as a typical example of ocean colour data becoming available from new and upcoming sensors in the near future. The ocean colour retrieval community is further developing its algorithms and preparing for the new and upcoming missions by NASA en ESA. In the context of the current research the focus therefore is not intended as a validation of the retrieval, but as an assessment of a monitoring strategy using a typical state of the art ocean colour data product of chlorophyll-a. The presumption has been that future ocean colour data products for the new missions are at least as accurate as the currently used MERIS HYDROPT data. The methodology adopted in this study is designed such as to minimize the effect of systematic differences between MERIS data and the traditional MWTL IS data but emphasize the effects of sampling characteristics.

1.3 Outline of this report

This report is organised as follows: first a general introduction is given to the monitoring and information cycle and the optimisation of monitoring strategies in the context of existing information needs, practical constraints and current knowledge of the natural system (chapter 2). In chapter 3 the information requirements are detailed further, in particular the context of the OSPAR eutrophication assessment which is one of the purposes of the current marine biogeochemical monitoring. In chapter 4 the methodology and materials are introduced and some of the general characteristics of the *in situ* network and ocean colour remote sensing coverage are discussed. Chapter 5 gives a description of the system characteristics related to the resolution in time and space of the *in situ* and remote sensing data covering 2003 to 2011. Chapter 6 reiterates the OSPAR eutrophication assessment as by default done on the *in situ* data. It illustrates the implications of using the remote sensing data for this purpose instead of (only) *in situ* data. Finally chapter 7 provides a brief discussion, conclusions and recommendations.

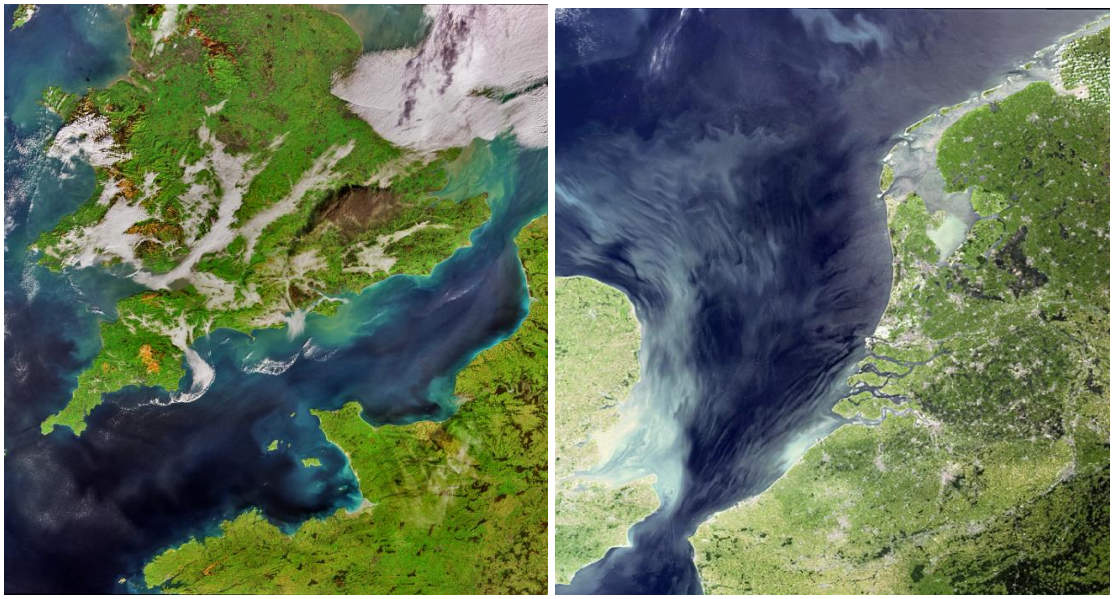


Figure 1.1 True colour images reconstructed from MERIS Full Resolution (i.e. 300x300 m^2) recordings of the English Channel (left, 2005-12-11) and Southern Bight (right 2003-07-14). Images illustrate the patterns of coloured matter and turbidity observed from space, next to clouds, haze, the sea bed and sun glint. The scales of the patterns seen in the water are typical for the shallow European shelf seas where tidal and wind driven currents and mixing interact with biogeochemical processes. (Source ESA; left: MERIS FR 11 Dec.2005; right: MERIS FR 14 July 2003.)

2 Approach to optimise monitoring

2.1 General approach KPP program

Within the framework of the KPP program, various organisations within the Dutch Rijkswaterstaat (RWS Zee en Delta, RWS CIV, RWS WVL) wish to know which monitoring techniques and strategies they should implement in the coming 5 to 10 years to address the current and upcoming information needs of the national government and its stakeholders given the physical, biogeochemical and biological characteristics of the systems observed.

The projects within the programme *Efficiente Monitoring* 2013 evaluate the applicability and feasibility of combinations of observation methods and -when relevant- modelling to generate the required information. The combination of observational methods, and samplings-resolution is referred to as monitoring strategy.

To arrive at an optimised monitoring strategy a common approach has been adopted where a combination is considered of the state and development of technology, of information and operational requirements and of the characteristics of the natural system that is being monitored. This triangle approach has been proposed e.g. by Van Bracht (2001) and has been applied already for KPP Monitoring by Laane (2013a,b), Noordhuis, (2012).

Figure 2.1 presents a sketch of the triangle approach. The triangle represents the space in which the optimal information strategy has to be found. This space is bounded by

I Information demand and operational requirements (including budgets),
W (Water) system behaviour and system knowledge (variations, interactions of state variables and indicators)
T Technological possibilities (sensors, platforms, data analysis and data management systems, computational models).

Parameters in the optimisation are the nature and number of observable state variables, the spatio-temporal resolution and coverage, accuracy characteristics of the data.

As such this may seem trivial, but the main challenge of the approach is to maintain a balance between information needs, system knowledge and technology push during the process of optimisation.

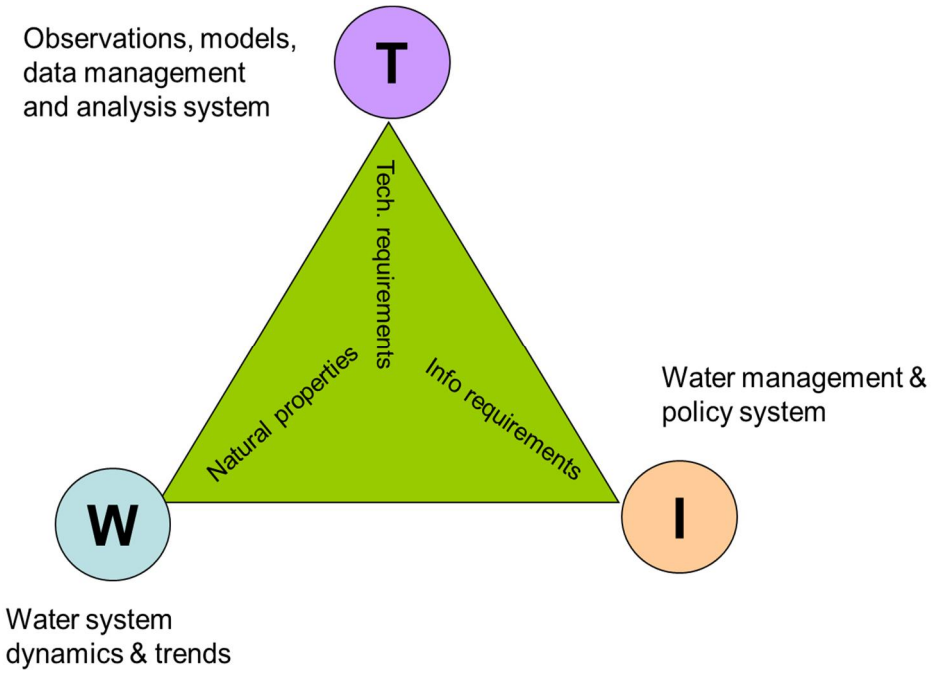


Figure 2.1 Triangle approach for development of monitoring strategies. The space of optimisation is bounded by the three angles: I is the information and operational requirements (demand); W is the current knowledge of characteristics of the natural system; T is the supply of observations and information provision technologies (after Van Bracht, 2001; Noordhuis, 2012)

Optimisation of monitoring strategies cannot be regarded independently from the monitoring and information cycle (Figure 2.2). This cycle represents the continuing evolution of information requirements based interaction of monitoring with evolutions in the domain of policy and management. As shown in Figure 2.2, the key elements in the cycle are

- Formulation of the information request;
- Design of the strategy to obtain this information (from monitoring or other data sources);
- Organisation of observing infrastructure and collection of the data;
- Analysis of the raw data, quality assurance and processing of the data into data products;
- Storage and dissemination of the data products;
- Analysis and interpretation of the data products into information

Although a monitoring cycle is conservative in order to maintain continuity and internal consistency of information, the lessons learned from the information may lead to a revision of the information demands and, through that, also to a revision of the other elements in the cycle. Apart from this top down modification of the cycle, each individual element is exposed to developments of insights, technology and operational context (including costs, relationships with other parties etc.). This may lead to bottom up opportunities for improving the cycle. The challenge of any monitoring optimisation is to benefit from bottom up developments, maintain or even improve the required information all within the bounds of the triangle of Figure 2.1.

This report focuses on the design of a strategy given two existing ways to collect data for one particular information need (marine eutrophication). It does not explicitly address issues of the

observing infrastructure, the QA, data management etc. but it is remarked that for a full optimisation eventually the entire cycle should be regarded.

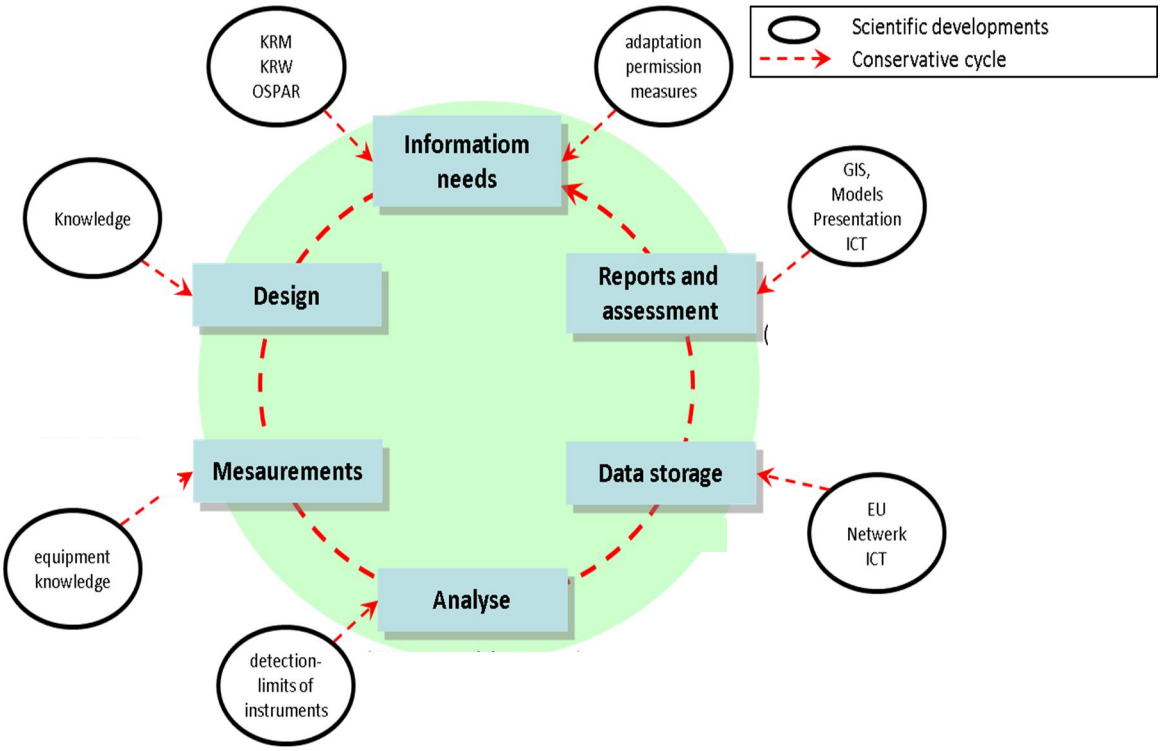


Figure 2.2 Monitoring and information cycle showing the six basic elements (see text) in squares, the external or autonomous developments in circle (after Laane 2013a,b).

2.2 Specific approach eutrophication monitoring

Within the overall triangle approach, there is a need for insight in the weighing of technological supply for eutrophication monitoring on the (Dutch) North Sea against system characteristics and information requirements. Here the three elements are briefly introduced. They are elaborated upon in the following chapters.

I: Information requirements
For the current study, the information need is mainly determined by the assessment of the eutrophication status for OSPAR and -through that- for the Marine Strategy Framework Directive. eutrophication is the effect of nutrient enrichment in the water due to anthropogenic inputs which results in 'accelerated growth of algae and higher forms of plant life to produce an undesirable disturbance to the balance of organisms present in the water and to the quality of the water' (OSPAR eutrophication Strategy, www.ospar.org).

The OSPAR assessment is based on a combination of variables and statistics of these variables applied in the so-called Comprehensive Procedure which is part of OSPAR's Common Procedure (e.g., OSPAR, 2013). It is remarked that for the Marine Strategy Framework Directive (MSFD), OSPAR has adopted determining much of the eutrophication information demands (see also Anonymous, 2012, 2013; Ferreira et al., 2010). For the near-shore waters the Water Framework Directive (WFD) is also relevant.

As remarked already above, it is important to realize that for the current study the information requirements are relatively limited, whereas in practice there are also other stakeholders and interests relevant for the information requirements for RWS (i.e., the ministries of I&M, EZ, Defence, and even parties involved and responsible for the (sustainable) use and management of the North Sea). For example, environmental Impact Assessments (EiAs) and the maintenance and operation of numerical models are other domains from which different information requirements emerge. In the few years before 2007, the Maasvlakte 2 project organisation of the Port of Rotterdam, for example, defined an information requirement for the monitoring of effects of the construction of Maasvlakte 2. Their information requirements have led to an extensive data need and monitoring strategy in which the Port of Rotterdam carried out many *in situ* measurements. However, since they acknowledged they could not provide all necessary data themselves, they also relied on ocean colour RS data, numerical models and national monitoring data as integral parts of their basic information to report to the competent authorities (RWS).
So far, EiA studies, appropriate assessments, model validation and also remote sensing retrieval development often rely heavily on national monitoring (such as MWTL) as baseline information with the implicit presumption that this national monitoring suffices for their individual information needs as well. This issue has been indicated in Blaas et al. (2012) and Laane et al. (*in prep.*) and is not elaborated upon here. The issue of addressing multiple information needs simultaneously wil, however, return when putting together the monitoring strategies for different purposes into one comprehensive operational monitoring plan in the future. In the next chapter, more details are given on the information requirements for the OSPAR assessment.

T: Technology of data supply (instruments, platforms)
For eutrophication monitoring, various methods are available to collect measurements of the marine surface waters. A collection method comprises both the choice for specific sensor techniques (e.g. optical, chemical, acoustic, chromatographic), the sampling method (immersion *in situ*, taking bottle samples, pumping and throughflow, remote) and the platform (ship-borne sampler, towed device, AUV, glider, float, FerryBox, buoy, semi-permanent post, airplane, drone, satellite). For example Osté et al. (2012), but also Kröger (2009) reviewed various methods potentially available for RWS monitoring and this will not be reiterated here. In the preceding series of projects directed at the evaluation of ocean colour remote sensing for the Dutch Rijkswaterstaat (RESMON-OK, ToRSMoN) also more extensive inventories of the technological supply for earth observation by satellites have been made (see also Westerhoff et al., 2010 and references therein).

The current study is specifically aimed at satellite earth observation data (ocean colour remote sensing) compared to the traditional *in situ* monitoring by means of bottle samples and a few mooring-based (buoy) sets. It is foreseen that other technologies, in particular automated sensor observations (from ships) and numerical models will be assessed in similar way in the future.

W: (Knowledge of) water system characteristics
The North Sea biogeochemistry is highly dynamic. Our knowledge is based on a vast collection of scientific research, based on dedicated scientific observations, but also modelling efforts from national and international institutions. Historic monitoring data are playing a key role in many of these studies. This implies that for some system characteristics our knowledge may be biased by the historical monitoring strategies. This is an inherent aspect of research in a field such as biogeochemistry: on-going observational and modelling efforts may point to system characteristics hitherto unknown. For the current study, the

system characteristics are discussed in terms of the data available within the project itself. We mainly compare scales of variation in time and space of the different sources of observational information. The North Sea biogeochemistry that is relevant for eutrophication is governed by growth and decay of phytoplankton species under conditions of light an nutrient limitation that vary spatially and in time. Seasonal variation in insolation, tidal and wind-driven mixing and resuspension, input and transport of riverine nutrients, recycling of nutrients from the sea floor and exchange of nitrogen, carbon and oxygen with the atmosphere all play partially interacting roles in this. For example, fast tidal variations haven also recently been identified to play a key role in the variability observed from CEFAS SmartBuoys in the southern North Sea (Blauw et al., 2012). Our knowledge of these system dynamics is taken implicit here. We will discuss dynamics in statistical terms, though (degree of variance). For example Laane (2013a,b) and Laane et al. (*in prep.*) and references therein review system properties in the context of observation strategies. Laane et al. (*in prep*) prepared a power analysis of the detection of trends in chlorophyll-a concentration in Liverpool Bay. They compared various time series of chlorophyll-a originating from a high-frequency (30 minutes sampling interval) CEFAS SmartMooring series. By resampling the series with lower frequencies they could diagnose what the effect of time resolution is on the estimates of the mean and 90-percentile as well as on the uncertainty in these estimates. The conclusions are more generic in the sense that also for the North Sea, a system with similar dynamics the same considerations hold: the lower the frequency of sampling the less accurate the state indicators can be estimated, and as a consequence also the less accurate trends and trend breaks in these indicators can be estimated.

After all, it should be kept in mind that the design of the eutrophication assessment procedure by OSPAR is already strongly based on knowledge of the system dynamics as it attempts to provide a descriptor of the health of the primary ecosystem (Good Ecological Status, MSFD descriptor 5).

3 Information requirements

3.1 Legislation and conventions

The information to be obtained from the national eutrophication monitoring on the North Sea is required in the context of the Marine Strategy Framework Directive (MSFD) and OSPAR convention and partially also the Water Framework Directive (WFD). These international regulations ask the Dutch national government to report regularly on the quality status of the North Sea ecosystem to OSPAR and EEA.

The OSPAR convention, formally the 'Convention for the Protection of the Marine Environment of the North-East Atlantic', comprises an agreement between 15 European nations and the European Community on the protection of the marine environment of the North-East Atlantic. Part of OSPAR is the Hazardous Substances and Eutrophication Committee (HASEC) which is responsible for the strategy to reduce eutrophication in the marine environment. OSPAR has set out a Joint Assessment and Monitoring Programme (JAMP) to assess the eutrophication status of OSPAR maritime waters. For various regions in the North Sea, thresholds have been defined for a set of water quality parameters. These thresholds are used in the Common Procedure (COMP) to assess whether eutrophication problems exist in certain areas. This assessment procedure is detailed further in the sections below.

Eutrophication monitoring is not only relevant for OSPAR. For the implementation of the Marine Strategy Framework Directive (MSFD), the OSPAR strategy for eutrophication has been adopted since it was recognized that the jurisdiction and objectives of the OSPAR Convention encompassed that of the MSFD. Also, the marine regions at which the Water Framework Directive (WFD) applies overlap with the OSPAR jurisdictional area. Hence, monitoring for OSPAR and MSFD is best integrated with monitoring for the WFD (see also Prins & Baretta-Bekker, 2013; OSPAR, 2013; Anonymous, 2012).

3.2 Common Procedure

The "Common Procedure" provides a framework to evaluate the eutrophication status of the OSPAR maritime regions and for identifying those areas for which actions are needed. The first step of the Common Procedure is a Screening Procedure to identify obvious non-problem areas. All remaining areas are periodically assessed under the second step, the Comprehensive Procedure which selects 10 parameters for harmonised application by the relevant countries to evaluate in a cause-effect relation scheme nutrient enrichment, direct and indirect eutrophication effects and other possible effects. The individual parameters are based on observed variables and are defined quantitatively. The eutrophication indicator in the end is more qualitative (i.e., a statement of 'good', 'problematic' or 'potentially problematic'). The eventual status is based on the compound of parameters as illustrated in Table 3.1.

Table 3.1. Examples of the integration of categorised assessment parameters for an initial classification (adopted from OSPAR, 2005).

	Cat. I *Nutrient enrichment* Nutrient inputs Winter DIN and DIP Winter N/P ratio	**Cat. II** *Direct effects* Chlorophyll *a* (Phytoplankton indicator species[1] Macrophytes)	**Cat. III & IV** *Indirect effects/other possible effects* Oxygen deficiency (Zoobenthos, fish kills org. matter, algal toxins)[2]	**Initial Classification**
a	+	+	+	problem area
	+	+	-	problem area
	+	-	+	problem area
b	-	+	+	problem area[3]
	-	+	-	problem area3
	-	-	+	problem area3
c	+	-	-	non-problem area [4]
	+	?	?	Potential problem area
	+	?	-	Potential problem area
	+	-	?	Potential problem area
d	-	-	-	non-problem area

(+) = Increased trends, elevated levels, shifts or changes in the respective assessment parameters.

(-) = Neither increased trends nor elevated levels nor shifts nor changes in the respective assessment parameters.

(?) = Not enough data to perform an assessment or the data available is not fit for the purpose

Note: Categories I, II and/or III/IV are scored '+' in cases where one or more of its respective assessment parameters is showing an increased trend, elevated level, shift or change.

OSPAR remarks that "when weighing data derived from the assessment process, the quality of the underlying monitoring should be taken into account. It may be appropriate to initially classify an area as potential problem area if the area shows an increased degree of nutrient enrichment (Category I). However, where data on direct, indirect/other possible effects are not sufficient to enable an assessment or are not fit for this purpose, the OSPAR eutrophication Strategy requires urgent implementation of monitoring and research in order to enable a full assessment of the eutrophication status of the area concerned within five years of its classification as potential problem area with regard to eutrophication. In addition, it calls for preventive measures to be taken in accordance with the precautionary principle." This means that the reliability of the information and fitness-for-purpose of a monitoring activity is an issue of attention. We discuss this issue in more depth in section 3.4 below.

[1] *Currently for MSFD only Phaeocystis is considered*

[2] *Currently not in scope MSFD*

[3] *For example, caused by trans-boundary transport of (toxic) algae and/or organic matter arising from adjacent/remote areas.*

[4] *The increased degree of nutrient enrichment in these areas may contribute to eutrophication problems elsewhere.*

The Comprehensive Procedure takes into account many of the known system characteristics in defining the criteria for assessment. In fact, the system characteristics aspect of the strategy optimisation is thus present also in the details of the information requirement. The ultimate information required for the Dutch government is whether or not there is a eutrophication problem in a specific part of the Dutch North Sea. In the following of this report, only the chlorophyll-a-related statistical parameters that serve as state indocatorsare discussed in detail. Still, it should be realised that the eventual compound indicator is what matters. Hence, by optimising the collection of chlorophyll-a data, the collection of other data may still be sub-optimal. An integrated strategy has been discussed by Laane 2013a and a follow up of that matter is foreseen for 2014.

3.3 Current assessment procedure

As indicated above, the assessment of the eutrophication status in the Common and Comprehensive Procedures is the outcome of a consensus-based procedure that is being developed and evaluated within OSPAR and amongst its representatives. These parties are collaborating in the Intersessional Correspondence Group on eutrophication (ICG-EUT) to define and evaluate the assessment procedure. The assessment method is frequently revised, as inherent of a monitoring and evaluation cycle.

The essence is that the assessment is based on regions that share relevant system characteristics. These regions are primarily defined based on the relative amount of riverine (fresh) water (hence nutrients input). Next to that, other physical and biogeochemical criteria are used (vertical mixing, seasonal amplitude of chlorophyll). As an illustration of one of the characteristics distinguishing the various regions, Figure 3.1 shows the annual mean SPM concentrations. This map is based on the same MERIS multispectral data that underlie the chlorophyll-a data analysed in this report. SPM-related turbidity also determines the biogeochemical dynamics, next to the stratification and salinity and nutrient characteristics. Hence, SPM concentration maps help define the boundaries of the regions.

Per region the available data are considered to determine the relevant statistics over the spatio-temporal interval of interest. For the current analysis, only the regions of the North Sea proper are of interest. The Wadden Sea and Ems and Scheldt estuarine areas will not be considered here.[5] The regions are -in a nutshell- characterised as follows:

- The **Dogger Bank (DB)** area, an offshore area, saline but shallow and hence vertically well mixed, relatively low in (abiotic) turbidity, relatively high primary production compared to directly surrounding areas
- The **Oyster Grounds (OG)**: a relatively deep and hence seasonally stratified area, relatively turbid due to the extent of the so-called East Anglia plume of SPM. Confluence of residual flow of water from the northern North Sea (‘Atlantic’ water) and form the southern Bight (‘Channel’ water), with different nutrient, temperature and salinity characteristics.
- **Southern Bight (SB)** region: in fact the region with this name is a subset of the southern Bight proper which covers the entire bight between the UK and Belgium and The Netherlands north of the Dover Straits. Region characterised by relatively low

[5] *Note that remote sensing of the Wadden Sea and Ems and Scheldt estuarine areas is in principle possible but not with similar characteristics as the remote sensing data and retrieval for the open North Sea discussed in this report.*

riverine inputs compared to the more coastal waters. Vertically well mixed throughout the year, residual transport directed towards the north/northeast.

- **Coastal Waters (CW)**: dynamic region dominated by the outflow of the Rhine and Meuse (and to lesser extent Scheldt) river waters: shallow but sloping bathymetry, relatively strong tidal currents and mixing, yet also stratification due to freshwater input and occasional summer heating. Southern part geometrically controlled by coastline hence setting up residual overturning and trapping mechanisms of water masses and particulate matter including algae. Further north (north and west of Wadden Islands) the dynamics are less determined by coastal trapping but more influenced by the exchange with the adjacent Wadden Sea (and indirectly Lake IJssel).

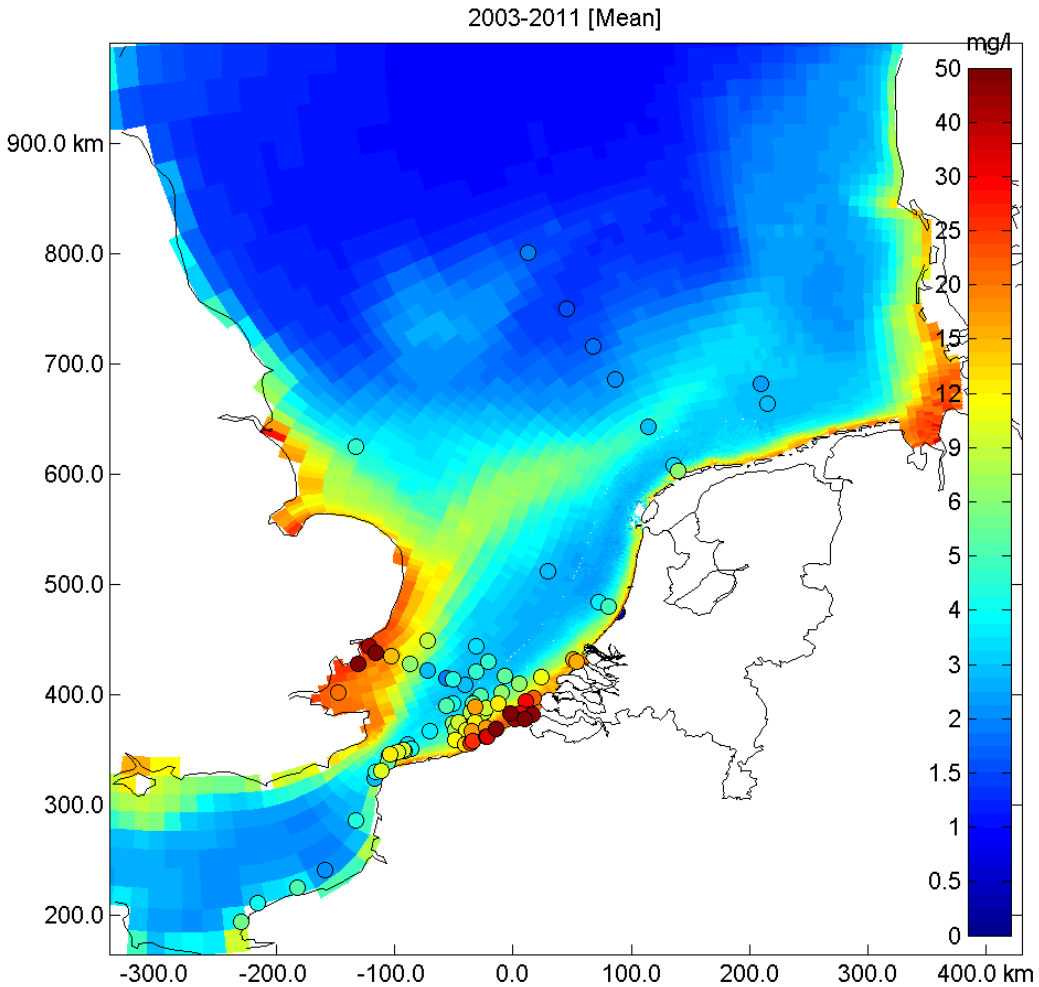


Figure 3.1 Geometric mean SPM concentration (mg/l) as derived from in situ observations (coloured dots from RWS MWTL, MUMM and CEFAS) and from MERIS remote sensing for the period 2003-2011 (coloured shades).

Regarding the Common Procedure applied on the regions, it should be noted that in the course of this project a slight inconsistency has been found in the definition of the regions as applied by Baretta-Bekker Marine Ecology and the definition of the regions as outlined by OSPAR[6] and the historic definition applied by the OSPAR ICG-EMO, the Intersessional Correspondence Group on eutrophication Modelling (Blaas et al., 2007; Lenhart et al., 2010). Figure 3.2 shows the currently available variants of the regions for the Dutch marine waters.

[6] *http://www.ospar.org/content/content.asp?menu=01511400000000_000000_000000*

The difference between the version applied by Baretta-Bekker (top left Fig. 3.2) and the one outlined by OSPAR (bottom left Fig. 3.2) is not critical for the analysis by Baretta-Bekker that is based on the clustered IS data alone, as the chosen clustering does not depend on the differences between these two region definitions. For the MERIS data, the exact definition of the regions is more critical. We therefore approximated the OSPAR regions by polygons on the ZUNO grid, which is shown in Fig. 3.2 bottom right panel. In order to compare our results to the results of Baretta-Bekker we used the same clustering of IS data (see also Table 3.2) next to the OSPAR regions shown in the lower right panel of Fig. 3.2.

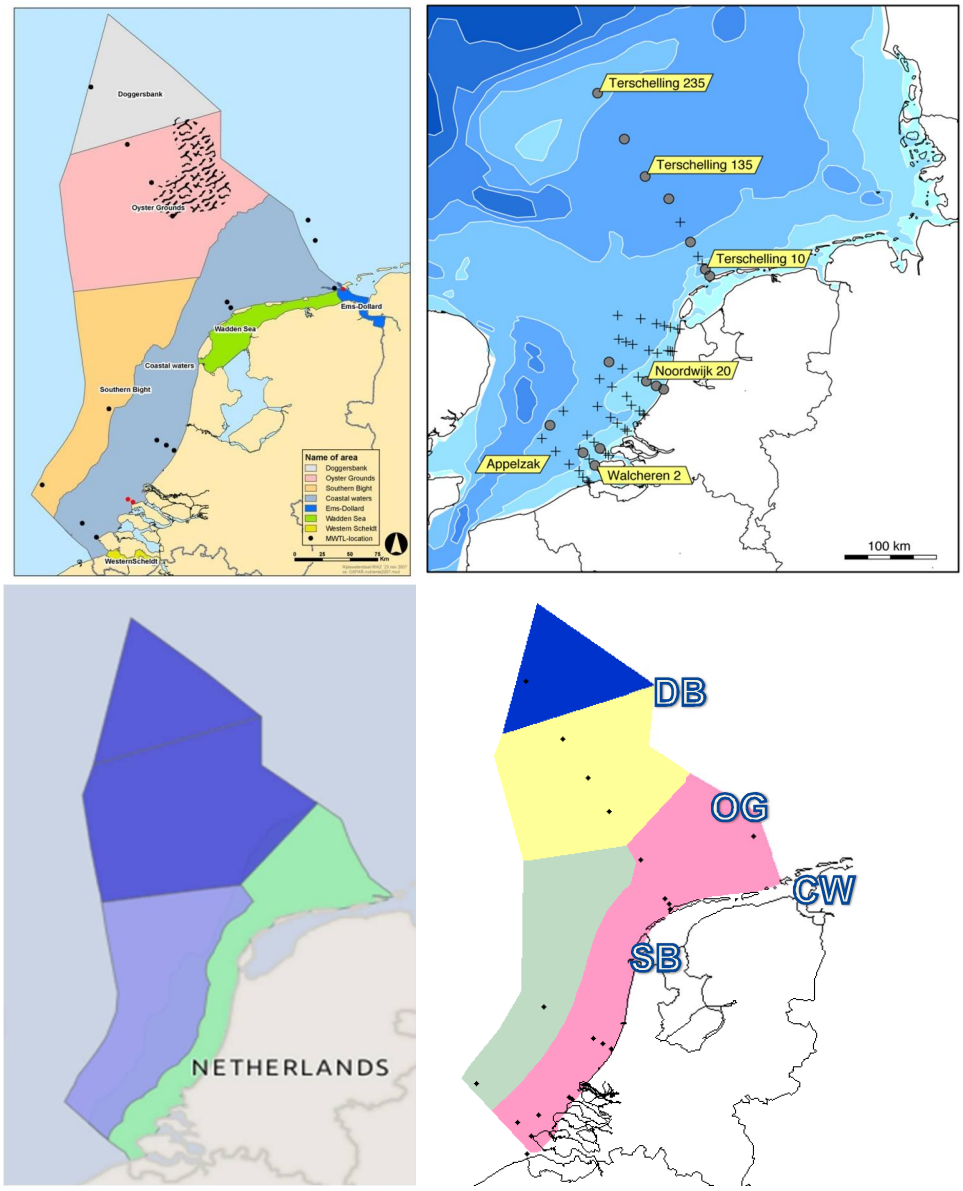


*Figure 3.2 Regions as applied in the eutrophication assessment: **Top left**: regions as applied in Prins & Baretta-Bekker (2013) and Baretta-Bekker (2013) (in black and red the considered in situ stations). **Top right**, bathymetry of the southern North Sea and the historic and currently active MWTL stations (crosses and dots, respectively, some key stations are labelled). **Bottom left:** Region lay-out based on the official OSPAR shape files; **Bottom right**: lay-out projected onto the grid used in the current study based on combination of the OSPAR shape file and the ICG-EMO definitions..*

Baretta-Bekker (2013) carries out the default analysis in a historically determined manner. The clustering of the stations per region is presented in Table 3.2 below. Here only the coverage of chlorophyll-ais shown. For the full assessment also the nutrient concentrations as total (TotN and TotP) and as dissolved (DIN and DIP) concentrations and Oxygen concentration are relevant. Moreover, for the oxygen saturation percentage also temperature and salinities are considered. These variables are not all present at all sites.

For chlorophyll-a it is noteworthy that the number and spatial layout of the stations for each region is strongly varying:

- The coastal waters are covered most extensively with 14 stations of which two (indicated with /) are conveniently considered as complementary to each other because of the proximity to each other. The spatial spread of these 14 (12) stations is fairly regular over the area, hence covering both the cross-shore and along-shore gradients that exist in this dynamic zone.

- The Southern Bight is only covered by two sites which are located in the very south and just south of the geographic center. Please note that the extent of the East Anglia (SPM) plume (see Figure 3.2) is not covered in the MWTL network. Only when the plume (occasionally) extends to the TS50 station (located in the Coastal Region) it is detected. Although the plume is mostly relevant for SPM, it is determining turbidity and hence also affecting biogeochemical dynamics.

- The Oyster grounds are covered by three regularly spaced stations on the Terschelling (TS) transect that capture not only the deepest part of Oyster Grounds proper but a larger part of the gradient at the Frisian Front.

- The Dogger Bank is represented by one station particularly at the edge of the region, located almost at the shallowest point of the bank within the Dutch EEZ.

The assessment for chlorophyll-a aims to compare the regional statistics to predefined assessment levels ('Elevated Levels'). If the regional parameters are below these thresholds this does not contribute to a problem status, if they are above, they contribute to a problem status. Table 3.3 gives the definition of the parameters and threshold levels for the Dutch marine waters. Table 3.4 outlines the steps and definitions as applied by Baretta-Bekker Marine Ecology for the current assessment (Baretta-Bekker, 2013).

Table 3.2. The OSPAR areas with the corresponding clustering of MWTL monitoring stations as also applied by Baretta Bekker (2013). The clustering is based on the regions and locations as illustrated in Figure 3.2. The underlined stations are selected as key stations shown in the time series of Chapter 5.

Area	Stations	chlorophyll-a(µg/l)
Coastal waters	GOERE06 /GOERE02[7]	+
	NOORDWK02	+
	NOORDWK10	+
	NOORDWK20	+
	ROTTMPT03	+
	ROTTMPT50	+
	ROTTMPT70	+
	SCHOUWN10	+
	TERSLG04/BOOMKDP[1]	+
	TERSLG10	+
	WALCRN02	+
	WALCRN20	+
Southern Bight	NOORDWK70	+
	WALCRN70	+
Oyster Grounds	TERSLG100	+
	TERSLG135	+
	TERSLG175	+
Dogger Bank	TERSLG235	+

Table 3.3 Definition of chlorophyll-a assessment parameters for the Dutch marine waters (adopted from Baretta-Bekker, 2013). The Elevated levels are the assessment values. These are based on the background levels, derived from longer term monitoring.

Growing season	Surface: -1m;	(µg/l)	Coastal Waters	Oyster, Dogger, Southern Bight
III- IX (incl)	Mean	Background	5	1.5
		Elevated level	7.5	2.25
	90-percentile & maximum	Background	10	3
		Elevated level	15	4.5

As indicated in Table 3.3, a combination of statistical characteristics of the chlorophyll-a concentration is used as indicator of the eutrophication: the mean over the growing season is regarded as indicative for the overall adundance consuming the nutrients available and recycling them. The 90-percentile (and, in earlier assessments, the maximum) serve as indicator of the extreme conditions during the spring peak of the plankton bloom. Both are indicators of ‘undesirable disturbance’ in the system.

Please note that, originally, OSPAR used the mean and the maximum to assess the chlorophyll-*a* concentration. In 2008 the mean and the 90-percentile have been used as it was argued that these can be estimated more robustly and thus are more representative for

[7] *Since 2007, Goeree 2 has been added and Terschelling 4 has been replaced by station Boomkesdiep, 2 km from the Terschelling coast*

the state of the ecosystem (see also Prins and Baretta-Bekker, 2013). Also note that the exact definition of the mean is not given, this will be addressed in the following chapters.

Table 3.4 the outline of the default assessment procedure for chlorophyll-aas also applied by Baretta-Bekker (2013).

Step	Parameter	Definition	Remarks
	Growing season	March to September (III- IX) (7 months)	
	Surface layer	About 1 m below water surface	In practice all samples of the top 3 to 4 m are taken as being representative for the surface mixed layer
1	Monthly, regional mean	Arithmetic mean over samples per region per month	For each region, the available data of chlorophyll-a are collated from *in situ* or sensor samples. Collect all numbers from all stations in a region per month and then take the average.
2	Growing season, regional mean	Arithmetic mean over the 7 monthly, regional means	Mean over the means differs from the mean over all samples at once when numbers of samples per month varies (weighing).
3	90-percentile per region	90 percentiles of the distribution of the (7) monthly,regional means	This differs from the 90-precentile over all underlying samples.
4	maximum	maximum values per region determined from all individual samples in a region over the entire growing season	This is without first averaging per month.

This procedure will be applied in Chapter 6, where some modifications will be applied to explore the impact of changes in the procedure on the outcome, but also in order to quantify the accuracy of the assessment.

3.4 Future assessment developments

The ICG-EUT is currently (autumn 2013) revising certain aspects of the assessment procedure (Baretta-Bekker & Zevenboom, *personal comm.*). The procedure has not been formalized but concepts are being disseminated to the member states' representatives.

One might state that historically the state of many problematic regions was well away from the target levels and significance of differences between observed and target levels was deemed not critical. Nevertheless, policy measures over the past decades are showing their effect and many systems come closer to their policy targets. Hence, significance of whether or not a state is at or close to target becomes more critical. The same holds for temporal trends in the systems: system responses and mitigation costs often follow exponential curves in opposite directions. For example the next generation of measures to further reduce nutrient inputs is increasingly expensive and the system response appears to be less sensitive (e.g. Lenhart et al., 2010). A balance between a significant trend reduction and affordable measures needs thus to be found.

According to the OSPAR agreement 2013-08, the OSPAR HASEC committee is currently reconsidering the assessment procedures based on continuing insights in the analysis methods and the developments in the natural systems under the OSPAR convention. This is

an example of evolving information requirement in the monitoring cycle. An ICG (ICG EUT) is currently considering data management and integration of data from various platforms, analysis of trends in the assessment variables in a more explicit (mathematical) formalism including issues of confidence in the data and trends. Moreover the ICG EUT is aiming for web-based visualisation and presentation product which requires standardisation and explicit treatment of data quality and analysis anyhow. (Baretta-Bekker *personal communication* 2013; OSPAR, 2013)

3.5 Considerations of uncertainty

Information obtained by monitoring is inherently uncertain. The natural system that is observed is stochastic in nature to some degree, *i.e.* some of its fluctuations are not predictable. Also the observations have stochastic components: instrument noise, fluctuations of the observing conditions etc. Moreover, it is practically impossible to observe the entire state of a natural biogeochemical marine system. Hence to get information about the state of the system, inference is required based on a sample of observations. Statements about the state of the system thus are inherently uncertain. This uncertainty needs to be considered when evaluating the quality of data and information, but also when discussing the reliability of a certain information strategy. In this section we will introduce the most frequently encountered terms related to reliability of monitoring.

A study of optimisation of monitoring strategies is in fact a study into obtaining the required information in a sufficiently reliable way, judged against costs and operational practicalities.
The trade-off between costs and what is sufficiently reliable is **a cost-benefit analysis and a risk analysis** which is not part of the current detailed study. The notion is however relevant: a piece of information is valuable because decisions (actions, judgments) need to be based on it. In the analysis of consequences of taking the wrong decision or a less optimal decision, direct costs, benefits and risks for health, safety, security of resources etc. should be taken into account. This also holds for decisions based on an assessment of the health of a marine ecosystem as given by an eutrophication assessment.

Reliability of a strategy is a common term here: it not only refers to the physical reliability (related to sustainability of the practical operation of the monitoring cycle: is information available when and where I need it? can I count on the monitoring programme in the future?) but also to the conceptual reliability, which is referred to as **information quality**. Key questions are:

- To what level does the strategy provide useful information that I need for my decision?
  - How close to the truth is this information?
  - How detailed is this information?
  - How well can this information be reproduced?

Many related terms play a role in this field, terms which are often confused, partially because their interpretation varies across fields of science, engineering, statistics and even law and philosophy. For completeness we summarize them here. Most of the definitions have been adopted from Wikipedia[8], which is regarded as a reflection of currently commonly accepted definitions beyond a single field of science or engineering etc.

[8] http://en.wikipedia.org/wiki/Information_quality, http://en.wikipedia.org/wiki/Accuracy_and_precision

Accuracy: according to the ISO 5725-1 standard (ISO 5725-1, 1994) the term "accuracy" refers to the combination of **trueness** and **precision**[9]. A measurement system or model is considered **valid** if it has both a high degree of trueness and precision. **Inaccuracy** is the lack of accuracy. Figure 3.3 shows the definitions in terms of a probability distribution of stochastic observational data relative to the (absolute, but in practice not exactly known) reference.

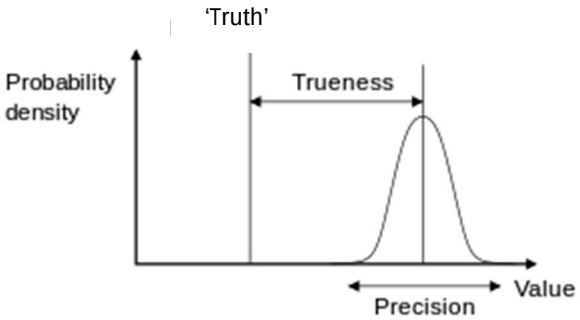


Figure 3.3 Illustration of accuracy in terms of precision and trueness visualizedby the probability distribution of a sample of observations compared to the reference value or truth. Here the truth is taken as a single value, in monitoring of marine ecosystems the truth is often also a statistical parameter (e.g. expectation value of the mean) of a varying state.

Trueness refers to the closeness of the mean of the measurement results to the actual (true, reference) value. Note that this true value is strictly speaking not known, it is inferred from analysis of observations. In the practice of lab measurements and calibration a so-called ”golden standard” is often defined to serve as reference (often based on very precise observations under standardized, controlled circumstances).

A measurement system can have high trueness but be imprecise, be precise but low in trueness, can be neither, or both. For example, if a set of measurements contains a **systematic error** (non-random effects, e.g. due to instrument offset, sampling scheme etc.), then increasing the sample size generally increases precision but does not improve trueness. The result would be a consistent yet inaccurate string of results from flawed measurements. Eliminating the only systematic error improves trueness but does not change precision.

ISO 5725-1 deliberately avoids the use of the term **bias**, because it has different connotations outside the fields of science and engineering, as in medicine and law. We will try to avoid it here as well.

Precision refers to the closeness of agreement within individual results. It is also called reproducibility or repeatability: the degree to which repeated measurements under unchanged conditions would show the same results. For models, “precision” is commonly indicating the resolution of the representation, typically defined by the number of decimal or binary digits. In

http://en.wikipedia.org/wiki/Confidence_interval
http://en.wikipedia.org/wiki/Standard_deviation
http://en.wikipedia.org/wiki/Unbiased_estimation_of_standard_deviation
http://en.wikipedia.org/wiki/Standard_error

[9] *This is different from earlier definitions also presented in the RESMON projects for example by Blaas et al, 2012, where “accuracy” was reserved for “trueness”.*

fact this is consistent with terms of reproducibility, but does not include the validity of the model formulation as such and the sensitivity of model results for model parameters.
Precision is associated to **random error** (random variability).

Figure 3.4 shows the bull's eye analogy illustrating the distinction of trueness and precision, where the reference value ('truth') is in the bull's eye and the back dots represent repeated measured (or modelled) estimates of the truth.

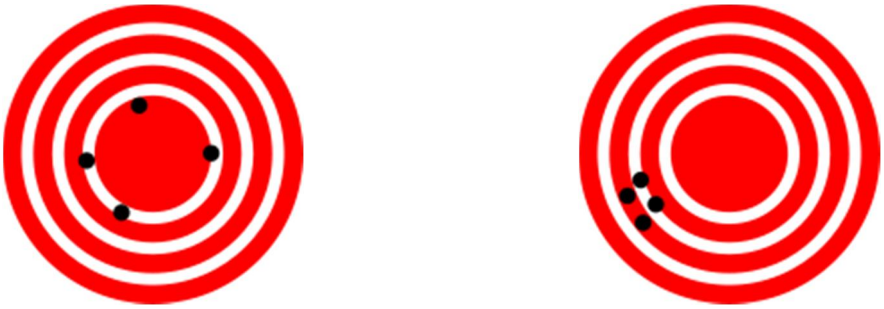


Figure 3.4: Left: low precision, relatively high trueness; right: high precision, low trueness. (Source Wikipedia)

Resolution: In addition to accuracy and precision, measurements (and models) may also have a spatial and temporal resolution. Moreover, instruments will have a detection resolution (and often also detection limits). The detection resolution is the smallest change in the underlying physical quantity that produces a response in the measurement signal (like a digit of a voltage sensor). Spatial and temporal (and e.g. optical spectral) resolution are dependent on the intervals at which data are collected or generated. Sampling volumes in the field (in time and space) are usually much smaller than the spatiotemporal resolution of computational models. Model equations are discretised such that they represent the mean conditions over a time-step or spatial step of integration.

Confidence in general can be regarded as related to reliability discussed above. It is less a property of the monitoring or model system, but more a judgement by the user of the information. In a sense it is the weight attributed to the information when evaluating it or comparing it with other information sources in making a decision. Confidence therefore can vary depending on the situation and the user. In the assessment of eutrophication and in many other instances of environmental monitoring, one wishes to know the uncertainty in a specific statistical property of the system state. To estimate this uncertainty, the practical approach is to consider a set of samples taken over a time interval or spatial range. In order to quantify uncertainties in statistical estimates, the **standard deviation** is often used. The range of uncertainty is determined by calculating the expected standard deviation in the results that would be obtained if the same set of observations were to be done multiple times on the same (stochastic) system state. For a 95 percent confidence level, the range of uncertainty is typically about twice the standard deviation for a normally distributed parameter.

To estimate the uncertainty in the mean value that we obtained from a set of measurements, the **standard error of the mean** (σ_{mean}) is a key parameter. This is usually determined from the sample standard deviation (σ) divided by the square root of the sample size (n), when we may assume statistical independence of the values in the sample[10].

[10] *Here we apply the notions of the sample (and sample mean) standard deviations (often written with an* s*), to be interchangeable with the true standard deviations of the population and population mean, commonly indicated by* σ*.*

$$\sigma_{\text{mean}} = \frac{\sigma}{\sqrt{n}} \tag{3.1}$$

In this equation, *n* is the number of statistically independent data points used to estimate the mean and standard deviation. According to Equation (3.1) an estimate of a state indicator becomes more precise, i.e. more reliable apart from systematic error, for an increasing sample size *n*. If merely only one data point is available, the spread is unknown and, strictly speaking, statements on the mean and uncertainty cannot be made. It must be noted that case of mutually correlated data points in the sample, the estimate for the standard error in the mean (as given by Equation (3.1)) provides an under estimate. The reason is that in case of correlation the total amount of information in the sample is less than for fully independent data points: the less each individual data point contains independent information, the less it can contribute to reducing the error in the estimate for the sample mean. To account for correlation the sample size *n* in Equation (3.1) must be replaced by a correction *n**. The computation of *n** involves the covariance matrix of the data sample. In the end, this correction yields an “effective” sample size *n**. It will hold that *n*<n* and thus the uncertainty, will increase when compared to the value for an uncorrelated sample. The larger the mutual correlation, the smaller the effective sample size will be. In the limit of full correlation n* will even be equal to 1.

Drawing conclusions from uncertain numbers: significance

Decisions based on monitoring data in general, and on chlorophyll-a data in particular, often relate to the questions whether or not the state of the system (oxygen level, chlorophyll-a concentration etc. or a compound indicator) is beyond or below a certain threshold and/or whether or not the state of a system is improving or deteriorating over time (trends). From a statistical point of view, threshold assessments can be regarded as one-sided hypothesis tests. Trend assessments can be either one-sided or two-sided, depending on whether the direction of the trend is relevant.

The risk of drawing the wrong conclusions from such hypothesis tests is dependent on the **statistical significance** of the outcome of the test. It is reflected in the statistical **power** of a sampling and analysis strategy: the less accurate the estimate of a statistical property, the higher the chance that one draws the wrong conclusions. For example, a chlorophyll-a mean value could be reported as too high whereas in reality it is below the threshold (‘false alarm’ or **Type I error**[11], unjust ‘problem status’). Or, vice versa, the analysis might report that the system has a non-problem status whereas in reality it is above the threshold level (**Type II error**[12]). Two concepts are closely related to the Type I and Type II errors:

Sensitivity (true positive rate (1-β), also referred to as power) measures the proportion of actual positives which are correctly identified as such (e.g. the percentage of problem statuses that is correctly identified as such).
Specificity (true negative rate (1-α)) measures the proportion of negatives which are correctly identified as such (e.g. the percentage of non-problem statuses that is correctly identified as such). The complement of the specificity is the **significance level** α (false positive or Type I error rate).

[11] *Type I error: "rejecting the null hypothesis when it is true".*
[12] *Type II error: "accepting the null hypothesis when it is false"*

A perfect predictor would be described as 100% sensitive (i.e. predicting all problem statuses as a problem status) and 100% specific (i.e. not predicting any system without a problem as a problem area); however, theoretically any predictor will possess a minimum amount of uncertainty and in practice there will be a trade-off between the specificity and sensitivity.

To summarize, the so-called confusion matrix is shown in Table 3.5. Starting from the null-hypothesis that the ecosystem is healthy, the matrix shows the following:

- True positive: *Unhealthy ecosystem correctly diagnosed as unhealthy*.
- False positive (Type I error): *Healthy ecosystem incorrectly identified as unhealthy*
- True negative: *Healthy ecosystem correctly identified as healthy*
- False negative (Type II error): *Unhealthy ecosystem incorrectly identified as healthy*

Table 3.5 Confusion matrix given the null-hypothesis H_0 that the marine ecosystem is healthy. On the top row the true states are given.

	Ecosystem is healthy	**Ecosystem is not healthy**
Reject null hypothesis	Type I error *False positive (α)*	Correct outcome *True positive (1-β)*
Fail to reject null hypothesis	Correct outcome *True negative (1-α)*	Type II error *False negative (β)*

Power analysis can be used to calculate the minimum sample size required so that one can be reasonably likely to detect an effect of a given size. Power analysis can also be used to calculate the minimum effect size that is likely to be detected in a study using a given sample size. For the current project is was not feasible to carry out a formal power analysis of the monitoring. We will show and discuss however, some aspects of significance in Chapter 6.

4 Technology & Data Supply: Materials & methods

4.1 Ocean colour remote sensing (MERIS data)

Remote sensing is an observational method that gains increasing importance in national and international monitoring activities by governments and semi-public authorities worldwide. Space-borne remote sensing in particular offers efficient ways to instantaneously cover large land or water areas, making use of existing earth observation satellites and data processing infrastructure. The increasing importance is not only reflected by the growth of the number of publications but also by the recognition of international initiatives such as the Copernicus programme (formerly GMES, http://www.copernicus.eu) of the European Community.

For monitoring of water quality parameters related to turbidity, levels of suspended particulate matter (SPM) and phytoplankton abundance and eutrophication (levels of chlorophyll, but also species specific information such as cyanobacteria and *Phaeocystis*) optical ('ocean colour') remote sensing is considered a valuable complement to *in situ* observations (see e.g. Sørensen et al., 2002; Ruddick et al., 2007; Nair et al., 2008; Eleveld et al., 2008; Pietrzak et al., 2011; Matthews et al., 2012).

Generally stated, optical RS data products and monitoring may refer to both inland waters such as Lake IJssel and Lake Marken (e.g. see also Chawira (2012)) and coastal waters such as the Wadden Sea (e.g., Hommersom, 2010), Scheldt Estuaries and Southern North Sea. The current study is focused on the open southern North Sea. For the current analysis MERIS instrument data have been used from ESA's Envisat satellite (which carries also 9 other instruments). Envisat has been launched in 2002 and is still flying, but since spring 2012 the platform lost contact with ESA control and shut itself down. The 10 years of service of the platform have been more than its expected technical lifetime. ESA is preparing a new mission (Sentinel programme) for ocean colour (OLCI sensor on Sentinel 3 satellite) of which the launch is due in autumn 2014. In the meantime the European ocean colour community is relying on NASA based ocean colour data from MODIS (Aqua and Terra satellites) and VIIRS (Suomi NPP satellite).

Although MERIS recorded the optical imagery at a full spatial resolution of about 300x300 m^2, the standard product for the North Sea is the Reduced Resolution (RR) 1x1 km^2 resolution pixel values. From the optical reflectance spectra recorded by MERIS, concentrations of chlorophyll-a and SPM (suspended particulate matter), CDOM absorption (coloured dissolved organic matter) and spectral extinction of visible light (Kd) and estimates of the standard error in these values have been computed. These data are based on retrieval by means of the HYDROPT algorithm (Van der Woerd & Pasterkamp, 2008). The current version of the data has been delivered to RWS over the past years by Water Insight B.V, based on the parameter settings from the Ovatie-2 project (Peters et al., 2008; see also Eleveld & De Reus, 2010). The Ovatie-2 project provided an update of the North Sea-wide calibration of the HYDROPT retrieval with specific focus on the Dutch MWTL data. As also indicated by Baretta-Bekker (2013), various definitions of chlorophyll as an indicator of algal biomass prevail, since it consists of a collection of pigments. Different observation, extraction and analysis techniques measure different pigments. For the current MERIS HYDROPT remote sensing data, the reference is made to the HPLC-based chlorophyll-a determined on behalf of RWS, since HYDROPT has been calibrated towards the MWTL surface concentrations.

These MERIS HYDROPT RR data for the period 2003-2011 are property of RWS and are at the moment archived by Deltares on the password-protected netCDF-CF-OPeNDAP server https://remotesensing.deltares.nl. As indicated above, the data have been provided in the context of the monitoring of the impact of the construction of Maasvlakte 2 but the same data have also been used in a pilot of an operational (Harmful) Algal Bloom forecasting system for Rijkswaterstaat in the first decade of this century (e.g. Rutten et al., 2006).

Gaytan and Blaas (2013) and references therein present the method applied to project the MERIS pixel data onto the grid of the ZUNO-DD numerical model, taking into account the uncertainty information provided with the pixel data. This method has in the current study been applied to project the chlorophyll-adata for 2003-2011 on the grid of the ZUNO Coarse model (see Fig. 4.1). The data have been evaluated for outliers and flagged pixel values have been rejected according to the criteria explored in Gaytan et al. (2013) and Eleveld & De Reus (2010). In the mapping, a mask has been applied to remove all pixel values over the areas of the Wadden Sea, Lake IJssel and Lake Marken and the lakes and estuaries in the Zuid-Holland & Zeeland Delta region because these regions have different composition of the optical active material in the water and the retrieval algorithm has not been calibrated for these conditions (see e.g. Hommersom, 2010).

The data preprocessing has a historic reason, since the data applied in the current studies have also been used for the assessment of trends in SPM during the construction of the 2nd Maasvlakte (the MoS^2 project: Model-supported Monitoring of SPM; Blaas et al., 2008, 2012; 2013, El Serafy et al., 2011, 2013). For this particular study it was required to have full control over the quality data that come with the remote sensing. These metadata consist of various quality flags by ESA and the retrieval. Next to binary flags, also estimates of standard error and goodness of fit of the HYDROPT algorithm are provided. For a standardized (operational) monitoring strategy with remote sensing this preprocessing would best be left to the providers of the remote sensing data. Then, the users need not to be bothered by all the details, but should be provided with a single, easily understandable accuracy measure.

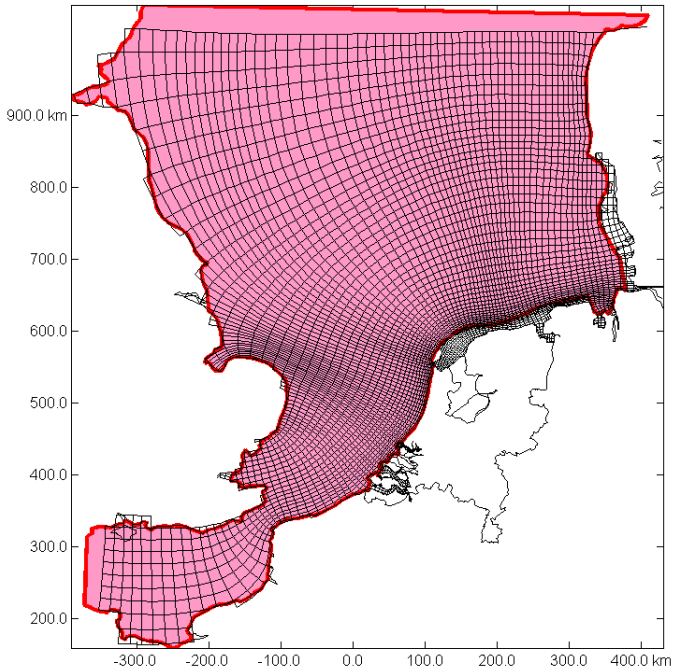
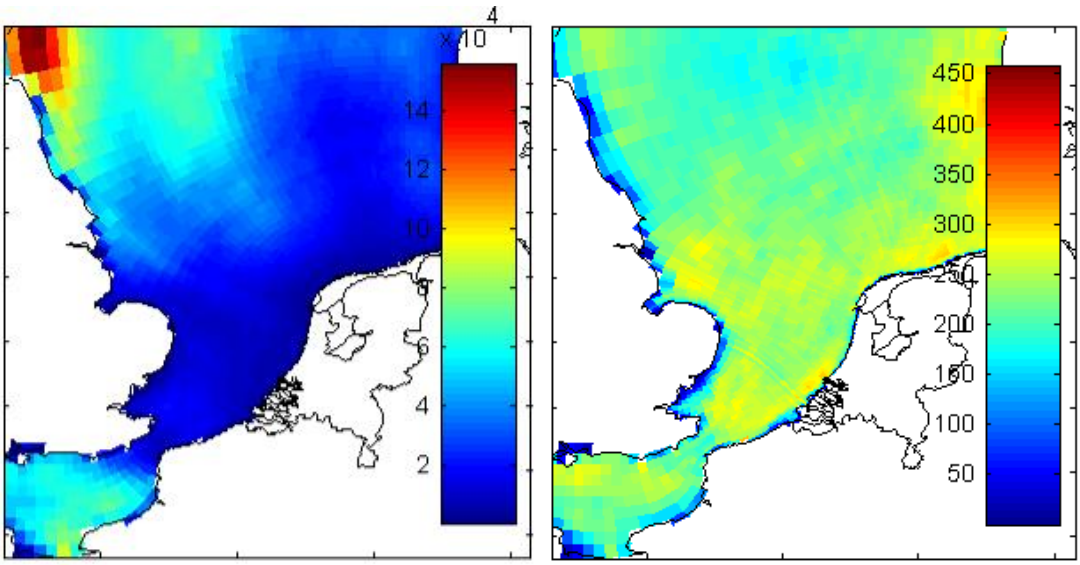


Figure 4.1 ZUNO coarse grid applied to map the pixel values on and to associate to the MWTL IS data. Pink: mask applied to accept 1x1 km^2 raw MERIS pixel values. Note that MERIS pixels in the Wadden Sea, Scheldt estuary etc. are excluded.

The spatial density of the quality-accepted pixel values within the mask is shown below in Figure 4.2 on the left. It can be seen that the total number of accepted pixel values is quite variable over time: areas close to the coast show fewer samples because the current version of the HYDROPT retrieval suffers from high SPM turbidity and, moreover, the underlying reflectance spectra are less accurate due to land-adjacency effects (increase of atmospheric aerosol, land surface reflectance). Further offshore, regions with less and more sampling are visible. These patterns are partly due to natural variations in viewing conditions (such as wave height, solar zenith angle during time of observation, cloudiness) that affect the quality of pixel values. It should be noted that the MERIS data have been collected from regular overpasses of the Envisat satellite which always occurred in the morning (roughly between 10:00 and 12:00) and that these overpasses produce swaths of several hundreds of kilometres wide. This sampling is therefore not random in space and time. On average, every square kilometre of the Dutch North Sea has valid data at least 30 times per year (about 270 samples over 2003-2011), which for a single location would compare well to the at most 26 times (bi-weekly) for the RWS MWTL programme. The major difference however is that this sampling density is achieved at every grid cell on the area.

Once the 1x1 km^2 pixel values have been screened, an aggregation is carried out to estimate the mean value (and its error) over the model grid cells. This procedure was developed for the application in the Maaslvakte-2 MoS^2 project and is outlined in detail in Gaytan et al. (2013). The current application is slightly different in the sense that the pixel data are weighted by distance but not by retrieval error to estimate the grid-cell mean and the empirical validity range applied for concentrations has not been applied here, because values with a higher retrieval error are not less likely to occur, they just have a lower precision.

The figure on the left shows the effective number of assigned model grid cell values in the time series from 2003-2011. A consequence of mapping to the particular curvilinear ZUNO model grid, is that this number is higher for larger grid cells further offshore since the chance of capturing an accepted pixel value in a grid cell increases with the cell surface area. In the Dutch coastal part of the domain the gridded sample density and its pattern is comparable to the original sampling.



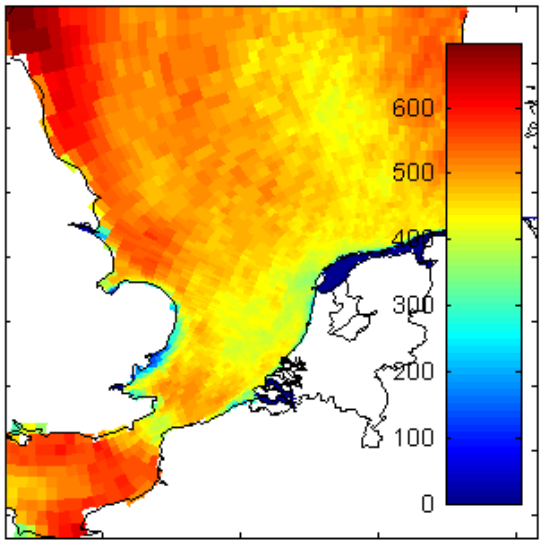


Figure 4.2 Top left: number of 1x1 km^2 pixel values per ZUNO grid cell location, top right: number of valid pixels per square km, bottom: number of valid gridded values per grid cell location over the entire period 2003-2011.

The histograms below show the distribution of log-transformed chlorophyll-a concentration values before (left) and after QC (rejection, acceptance) and gridding to the model. The flagging of certain unreliable values combined with spatial averaging introduces a rejection of values in particular in the higher ranges and a slight additional skewness around the median value. These deformations of the distribution are a consequence of the fact that unreliable values that have been removed are not uniformly distributed over the range of values which has to be regarded as an artefact of the HYDROPT algorithm. Please note that these effects are specific for the current data set produced in the years 2007-2011 by the data provider Water Insight BV. Following subsequent discussions with Water Insight on this issue, it has been concluded that improvement of the retrieval settings is possible such that the bias in rejection of values is expected to decrease. This can be verified only after reprocessing of these historic data which up to now has not been undertaken.

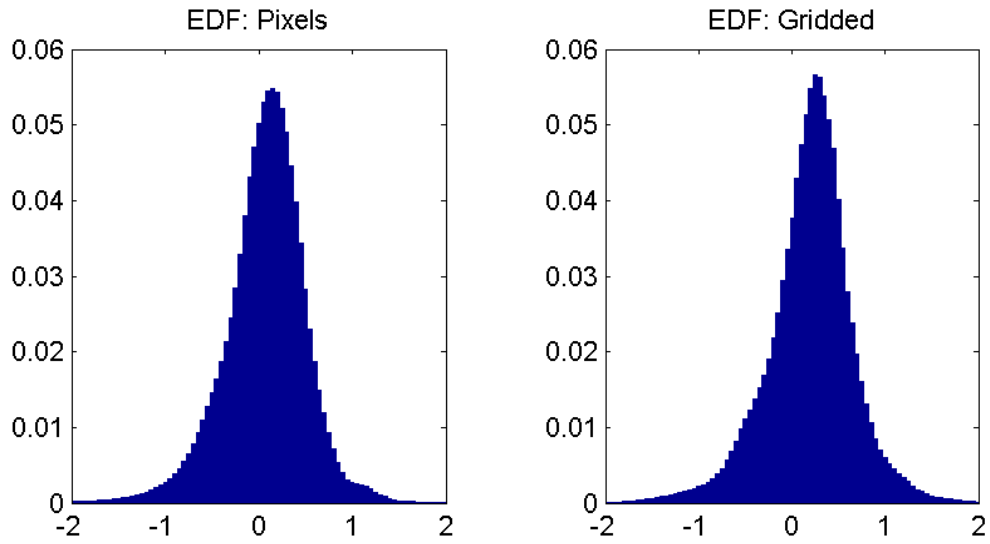


Figure 4.3 Empirical distribution function (histogram) of the chlorophyll-a values in the spatial domain of the North Sea mask in the time period 2003-2011 before mapping onto the grid (left) and after (right). Note that the

values of chlorophyll-a on the horizontal are indicated as ^{10}log of the values in µg/l (hence, -1 refers to 0.1 µg/l and 1 to 10 µg/l etc.).

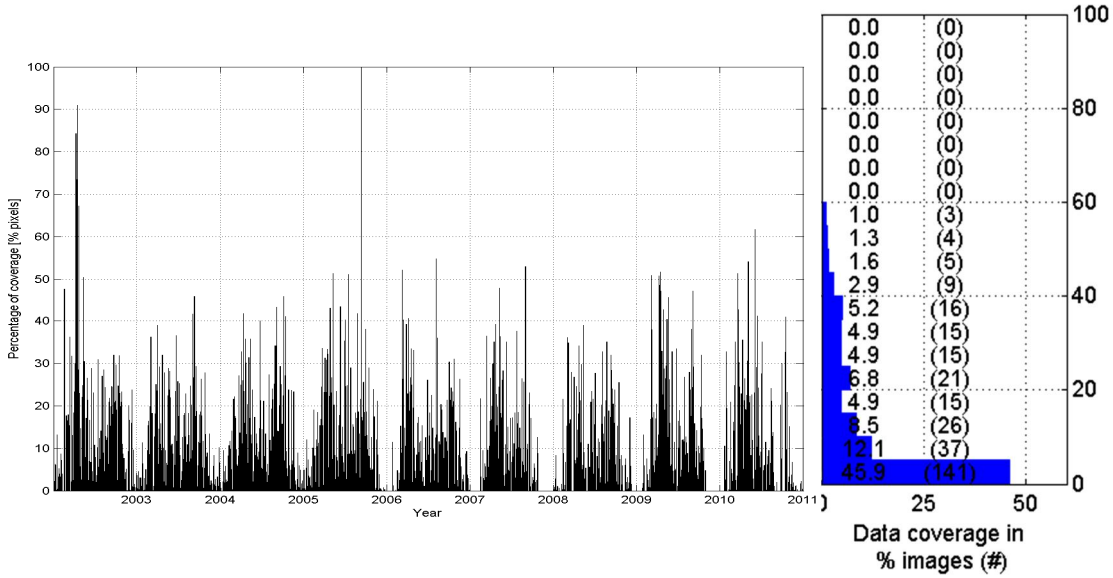


Figure 4.4 Coverage of North Sea area by MERIS sampling as percentage of the maximum number of pixels in the masked region. Left: coverage per day over time; right: histogram of coverage over all times. Note the strong seasonal variation in coverage and the fact that almost 50% of the recordings has a spatial coverage of 5% or less of the maximum.

The sampling in time is illustrated in the time series and histograms above: the plots show the relative coverage of each scene recorded at a particular day as percentage of the maximum possible number of values within the masked area. Clearly a seasonal (summer-winter) cycle can be seen: in winter viewing conditions are less favourable mostly because of increased cloud cover and lower sun angle: resulting in sometimes prolonged intervals with little or no observations. A secondary mode of variation can be seen between spring, summer and autumn: a dip in coverage is often found around the middle of the summer, likely because of increased cloudiness over warmer sea water.

These seasonal variations in time may also have a spatial pattern seasonally varying cloudiness and good viewing conditions. Effectively these variations in sampling relate to weather conditions, and as such specific biases for certain weather conditions may occur. It has been explored by Blaas et al. (2008) for the Dutch data and also by Fettweis & Nechad (2011) for the Belgian waters that indeed optical remote sensing suffers from a bias towards relatively calm (low wave) conditions and sunny (unclouded) spells but that the bias of Rijkswaterstaat ship observations towards calm conditions is even larger.

Another source of sampling bias may be the regular time interval with which MERIS data are sampled. Since this interval is close to 12 hrs and hence close to the dominant M_2 and S_2 tidal periods, there may be very slowly varying patterns in space and time of observations taken at a certain location close to a certain phase of the tide (e.g. high or low water, or maximum ebb or flood currents). This may be in particular relevant for the description of the particle behaviour of plankton (settling and mixing that depend on the tidal phase) (see also Neukermans et al. (2012), and Blaas et al. (2013) for similar discussions regarding SPM and turbidity).

4.2 MWTL and other *in situ* data

The *in situ* data are used to serve as reference in this study. The key data originate from the subset of DONAR that has been released in the publicly available Waterbase archive of RWS (live.waterbase.nl, cached at opendap.deltares.nl and viewable at kml.deltares.nl). These data have been collected in the context of the MWTL programme. In addition, the Belgian national chlorophyll-amonitoring data hosted by MUMM have been included, as well as the Smartmooring time series at Oyster Grounds by CEFAS.

Figure 4.5 below indicates the sample sites of the Dutch Rijkswaterstaat monitoring (MWTL) as well as the Belgian national monitoring carried out by MUMM. Note that all these samples relate to the surface layer of the water column (usually taken at the upper 3-5 meters of the water column). Figure 4.5 shows the average number of samples per year and per growing season in the period 2003-2011 of the *in situ* data against the background of the number of MERIS gridded samples (as also shown in Figure 4.2).

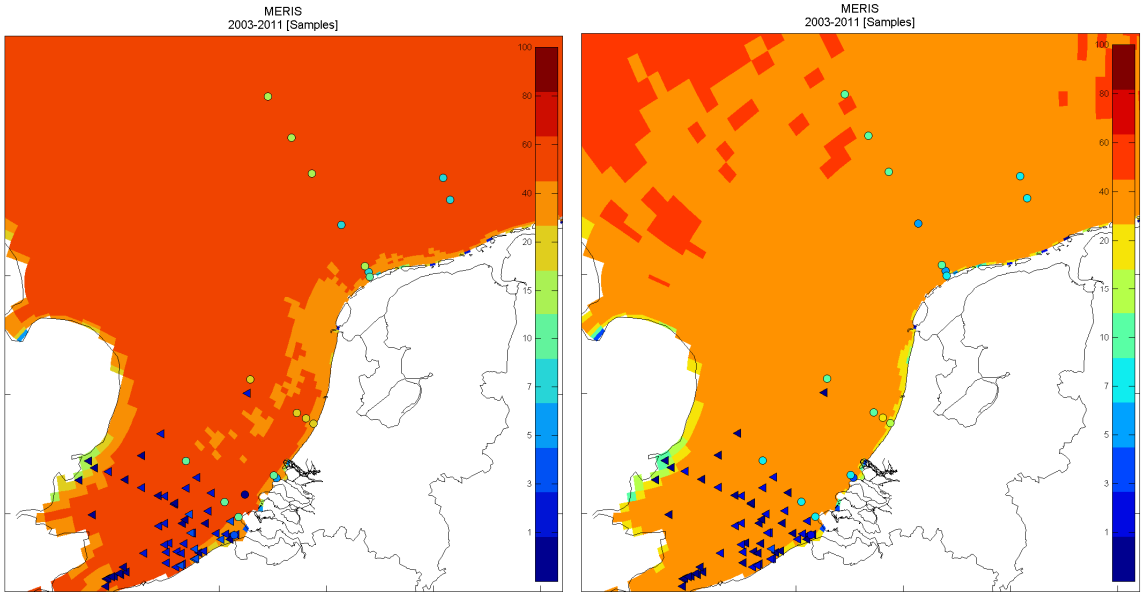


Figure 4.5 Top: map showing the mean number of samples per year between 2003-2011 for MERIS (shades) and at the locations (markers) of the IS stations available in the Deltares database. Circles indicate the MWTL stations, triangles the Belgian monitoring stations (MUMM). Not shown is the smartmooring at Oystergrounds which is a joint CEFAS, RWS mooring and not an official part of MWTL and which has a multitude of the number of samples compared to the standard stations. Bottom: as top but only for the number of samples for the growing season of all years, defined as March to September (inc.).

As mentioned above, the MERIS sampling reaches to about 50 samples per year (weekly on average on every square km), whereas the Noordwijk transect has an average of about 25 (bi-weekly visits), the Terschelling transect about 15 or less (about monthly) and the other transects even less. Note that the MUMM sampling strategy is different from the RWS strategy: the spatial lay-out is less stationary which means that some sites have only one or very few revisits. The spatial structure is much more variable though.

4.3 Methodology: DINEOF

In the section above, the basic sampling characteristics of the two data sets (MERIS and IS) have been introduced. Apart from the *in situ* and gridded MERIS data, we will make use of a data set constructed from the gridded MERIS data. This data set is based on the dominant empirical modes of variation in space and time in the MERIS data obtained with the data-interpolating EOF method (DINEOF). As will be explained in this section, DINEOF determines the dominant spatial and temporal patterns of variation in the data set, called Empirical Orthogonal Functions, EOFs (principal components). By putting together these modes of variation one can obtain a so-called gap-filled reconstruction of the MERIS data at all grid locations and sample times. This gives the opportunity to expand the matchup of *in situ* data and MERIS data and explore the bias and RMS (Root Mean Square) differences between these two sets as well.

In this section we will briefly introduce the DINEOF method. For more details the reader is referred to De Boer et al. (2012) and in particular papers like Beckers et al., (2006), Alvera-Azcárate, et al. (2012) and a similar application of matching *in situ* and MODIS RS data by Nechad et al. (2011).

Sampling coverage and resolution characteristics: The basic characteristics of the IS and RS data allow for a crude comparison in terms of spatial and temporal coverage and resolution. The spatial coverage of the MERIS data is orders of magnitude more extensive than that of ship-borne IS. On the other hand, semi-permanent IS sensors (e.g. on buoys) have a temporal resolution that is orders of magnitude higher than that of MERIS. When viewed at a single grid cell location, temporal resolution for MERIS and IS data collected by ships is of the same order of magnitude in the near coastal zone, whereas for the offshore parts MERIS has about twice as much temporal resolution (partly due to the coarser grid cell resolution chosen offshore). It has been noted already in Chapter 3 that the standard deviation of errors in data scales with the square root of the number of observations, e.g. 100 times more data results in (at most, when there is no sample correlation) a tenfold reduction of error estimates. For every individual location, the gain in accuracy between MERIS and MWTL thus may seem limited. However, the greatest difference between the MERIS sampling and the MWTL *in situ* sampling is that there are many other samples taken in the vicinity of the particular point of interest that are partially independent as well. The improved spatial coverage therefore not only results in a better representation of the spatial variability but also provides more resolution of the temporal variability. The spatiotemporal autocorrelation is exactly what is exploited in the DINEOF analysis, as explained in the next section. It also means that for error reduction, the spatial coverage of MERIS is the major added value.

Annual aggregates: After comparing MERIS and IS on the basic coverage characteristics, the next step is to compare the actual data of MERIS and IS. Due to the nature of the temporal and spatial coverage of MERIS and IS, the actual number of cases where MERIS and IS yield a value at the same time and place (matchup) is fairly limited. Comparison is therefore generally done after aggregation in time: comparing annual mean and annual standard deviation at the locations of the IS data. To emphasize the spatial structure of these statistics, these values have been presented in this report as maps, with IS dots overlying MERIS annually aggregated maps. In section 5.2 these annual aggregates are discussed.

Sophisticated temporal EOF aggregates: Annual mean and even seasonal mean maps are a limited manner of comparing IS and MERIS data. The important characteristics of seasonal

variation and in particular the spring bloom are flattened in such temporal aggregates. To remedy this flattening a more sophisticated spatio-temporal aggregation is required. Here we follow a successful approach for ocean colour Chl analysis. We used so-called Empirical Orthogonal Functions (EOF) to aggregate in time. We adopted the DINEOF software from the University of Liege (Belgium) for this (see Beckers & Rixen, 2003; Beckers et al., 2006; Alvera-Azcárate et al., 2012). EOF analysis uses simultaneous correlation in both space and time to extract dominant patterns that explain most (variance) of the signal. DINEOF is a mathematical method, it does not use any system behaviour *a priori*, such as known seasonal cycles or the known spring bloom, to fit the data. If a dataset, either IS or MERIS, is of sufficient quality, the dominant EOF modes should be representative of the system and relate to known physical and biogeochemical dynamical patterns.

EOF for MERIS: In an EOF analysis, the MERIS 3D hypercube "time series" *[x,y, time]* is decomposed into its statistically dominant patterns. This yields a total of *p* 2D maps $[x,y]_p$ with associated 1D time series $[time]_p$. Multiplying a 2D map $[x,y]_p$ with a 1D time series$[time]_p$ will yield an approximation of the original 3D dataset *[x,y, time]*. Using more modes *p* will yield a better approximation. The number of *p* used for complete reconstruction of the original spiky data, differ per physical quantity: *p* is at least order 25 for smooth SST data. For SPM data Deltares found that 6 to 9 modes can be extracted from the data (De Boer et al., 2012). However, the first few modes already contain most of the variance of the dataset. This is the variance associated with large scale patterns.

Figure 4.6 illustrates the decomposition of the MERIS data in the spatiotemporal EOF modes and de reconstruction (gap filling).

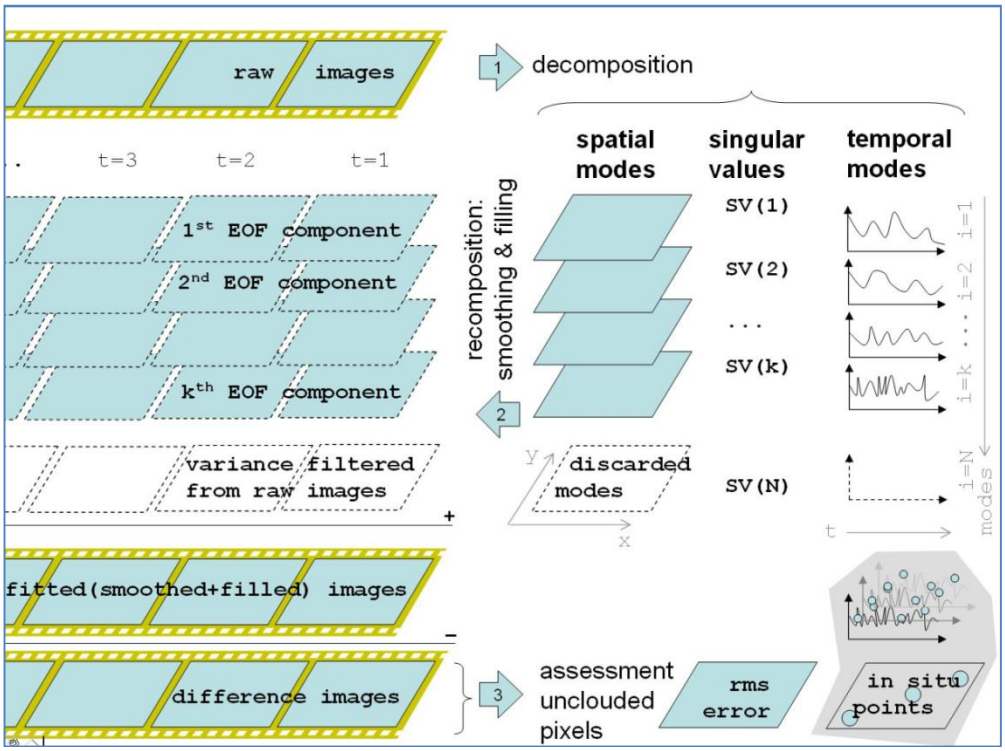


Figure 4.6 : Schematic work flow of DINEOF analysis of a time series of RS images.

EOF for IS: Same type of EOF analysis can be applied on the IS data, where the time series are inserted into one 2D "hyper time series" dataset *[station(lat,lon), time]*. Here [*station(lat,lon)]* are the 20 (present) to 80 (past) MWTL locations, and *[time]* is a common time vector for all locations. This yields *p* 1D "maps" (to be understood point collections) $[station(lat,lon)]_p$ with associated 1D time series $[time]_p$. Multiplying a 1D "map" $[station(lat,lon)]_p$ with a 1D time series $[time]_p$ will yield an approximation of the original 2D dataset *[station(lat,lon)],time]*. Using more modes *p* will yield a better approximation.

EOF requires one common time vector: Note that a common time is required for a set of spatially spread (quasi simultaneous) samples for EOF analysis to work. This has different implications for IS and MERIS data. The MERIS data of the southern North Sea that share a day-time stamp are all recorded within 10 minutes from each other during the overpass of the satellite. The MWTL IS data that we would like to regard as simultaneous are collected during a much wider time interval, though. The [time] vectors for MWTL IS data thus require a relevant temporal binning of the raw data. For MWTL De Boer et al. (2012) found the best to be 4 week bins, related to the common 4 week MWTL revisit interval. This means that EOF maps for ship collected IS are inherently somewhat smoothed and represent 4-weekly instead of daily 'maps' of data. Despite the fact that MERIS data time is identical within 10 minutes for for all pixels in a North Sea wide remote sensing, also MERIS requires one binning step: the Envisat space orbit passes over the North Sea every day one to two hours before noon local time, usually resulting in one image per day in the region of interest. However, sometimes this yield two useful images 1.5 hour apart, the first covering the eastern parts (German Bight) and the second the western part (UK side). These two images have been merged into one image. This means that EOF results for MERIS images have an effective resolution of 1 day and can capture short-lived events that a time scale of days, whereas binned IS data can only capture events with a time scale of months.

EOF for SPM extended to EOF in Chl: In the Resmon4 project (De Boer et al., 2012) we applied EOF analysis to MERIS SPM data, using one year of MERIS and the complete 30 years of *in situ* history. It was shown for both IS and MERIS SPM data that the first mode (p=1) corresponded to the annual mean with a slight seasonal modulation. It was also shown that the second mode (p=2) corresponded to the seasonal cycle. Adding these two modes proved to be a useful approximation of the bulk of the SPM dynamics in the southern North Sea. The sum of the first two modes was shown to be equivalent for IS and MERIS, and to be comparable to the generally used annual geometric mean. We adopted the basic set-up of the exercise of De Boer et al. (2012) with one major modification: we use the full 9 year MERIS data set, and limited the IS data to the same 2003-2011 period. This facilitates the interpretation. In section 5.3 the dominant modes are discussed.

EOF can fill data gaps (e.g. due to clouds): IS and MERIS data both have sampling biases due to unavailability of the data collection. Ships do not go out during storms, and less often during the night, weekends. MERIS cannot look through clouds and only passes just before noon (when light is best for optical data gathering). Hence some biases are similar others are different. EOF methods can use simultaneous temporal and spatial correlation to estimate missing data in the 3D *[x,y, time]* MERIS hypercube and the 2D *[station(lat,lon), time]* hypercube. This is a form of objective gap filling. This gapfilling requires at least 2% temporal coverage to fill a missing pixel via spatial correlations, and 2% spatial coverage to fill a missing pixel via temporal correlations. (see also Nechad et al., 2011). Although MERIS has order of magnitude more data values than IS, the number of missing MERIS data points in the hypercube still is 73% (83% when not merging the occasional twice daily images around noon). After EOF the MERIS spatio-temporal coverage has grown four-fold, thus yielding four

times as many possibilities for a match-up with IS data. The option to use gapfilled MERIS data for match-ups requires low levels of smoothing in the EOF procedure, to ensure that EOF reconstructed data are similar to the original raw data. In section 5.4 the gapfilled EOF modes with low smoothing are discussed and compared to the EOF modes obtained with high smoothing. In section 5.5 these EOF modes are used to perform an improved match-up exercise.

5 Modes of variation, system characteristics

5.1 Introduction

The analysis in this chapter will focus on the comparison of modes of variation resolved by both the IS and two RS sample sets (gridded with gaps and DINEOF gap-filled) in time and space. To this end, the basic statistics of the values in each set are analysed (median, spreads). Also, a matchup of the sets is made for direct comparison and estimates of bias and RMS differences are made. Those results will not only show the differences and similarities of the two data sets but also illustrate the system behaviour. This latter is important for the optimisation of a monitoring strategy as outlined in the introductory chapter.

Please note that both data are dependent on each other in the sense that the MERIS HYDROPT retrieval algorithm has been calibrated on the MWTL data (see. e.g. Peters 2006, Peters et al., 2008). In the calibration in particular the HYRDOPT retrieval parameters have been adopted that produced the best fit to the temporal mean for the year 2006.

For a monitoring strategy it is relevant to know what processes are relevant that need to be captured in the information and on what scales the state variables vary. The processes have been taken into account already in the formulation of the information requirements. For the current study we focus on what is following from the information requirements in terms of chlorophyll-a. The spectrum below shows the variation of temporal and spatial scales in a shelf sea such as the North Sea (see e.g. Otto et al., 1990; Mills et al., 2005). These scales are partially coupled as they are determined by physical and biogeochemical processes which are determined by external forcings such as seasons, weather and residual transport patterns (gyres), by internal dynamics of growth, decay and mineralisation, and by vertical settling and mixing proceses. The main issue is how accurately certain aspects of this variable system state can be estimated from repeated measurements.

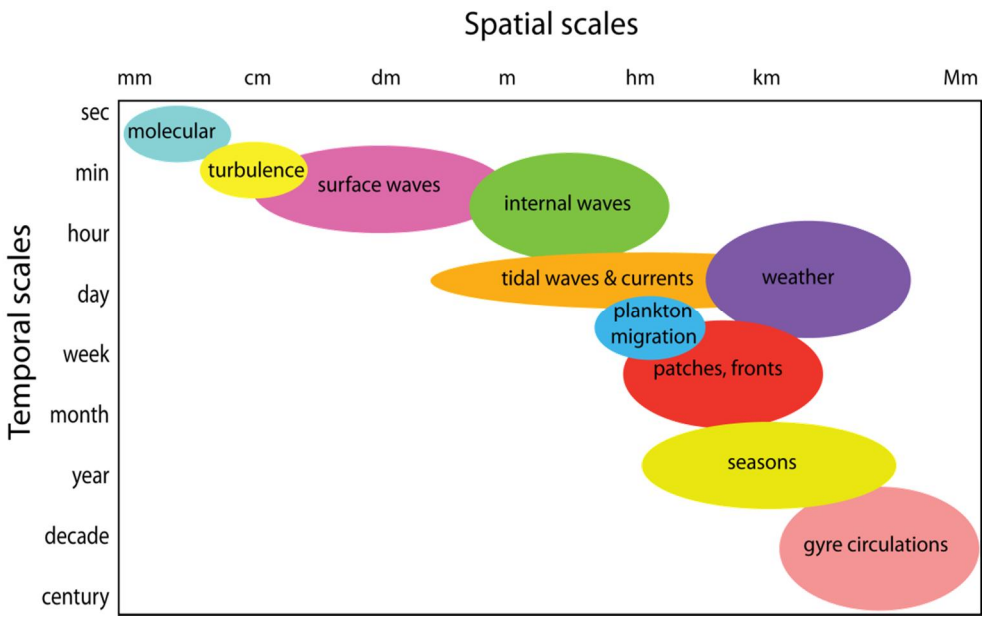


Figure 5.1 Coupled spatial and temporal scales of variation of chlorophyll-a in a typical shelf sea such as the North Sea (see also Mills et al., 2005).

When discussing system characteristics our knowledge is always limited by the historic observations. In this report we do not reiterate the standing knowledge of the biogeochemical dynamics of the southern North Sea but mostly relate to what is observed from the MERIS data. For a more extensive analysis many other data and information sources such as high-time resolution mooring time-series and also model computations need to be included. The description by MERIS is for the moment the information source with a higher resolution than the *in situ* data. For a direct comparison we also present the same statistics from the *in situ* data. Next to the MERIS data mapped onto the model grid, we show the variations of the system as inferred from the DINEOF method.

A secondary aim of this section is to validate the DINEOF method by comparing the gap-filled data to the *in situ* data next to the original data with gaps.

5.2 Basic statistics of chlorophyll-a over all samples of *in situ* and MERIS.

First the general statistical properties of both full data sets are displayed. This provides information on what the basic properties of the data are: the temporal (geometric) mean at every location, the spread in time around this mean, and the 90-percentiles. Besides, the number of samples is shown. We do not discuss the effect of possible sampling biases on the results here. We just present the characteristics as they are present in the currently available sets.

5.2.1 Maps

The map below shows the geometric mean values of both the gridded MERIS data and the IS data over the entire period 2003-2011. This gives an impression of the spatial structure and the way the mean gradients are captured. Please note that the geometric mean is used. This is done because the statistical distribution of chlorophyll data is strongly skewed. The distribution of chlorophyll-aconcentration values is approximately log-normal (see also the histograms Figure 4.3). The geometric mean of the untransformed set corresponds to the arithmetic mean value of the distribution after log-transformation. Also, the estimates of the spread around the mean will be done after log-transformation in order to get the proper statistical characterisation of the sets and avoid negative concentration values within a confidence interval.

Also note that the mean values all are based on the entire time series of the individual data sets. These sets are sampled with different time resolution and also with different sampling distribution in time. The statistics have been determined both for the entire period 2003-2011 including all samples from all seasons, and for the growing season (March-September of all years) for the period 2003-2011. The growing season in particularly relevant for the OSPAR eutrophication assessment.

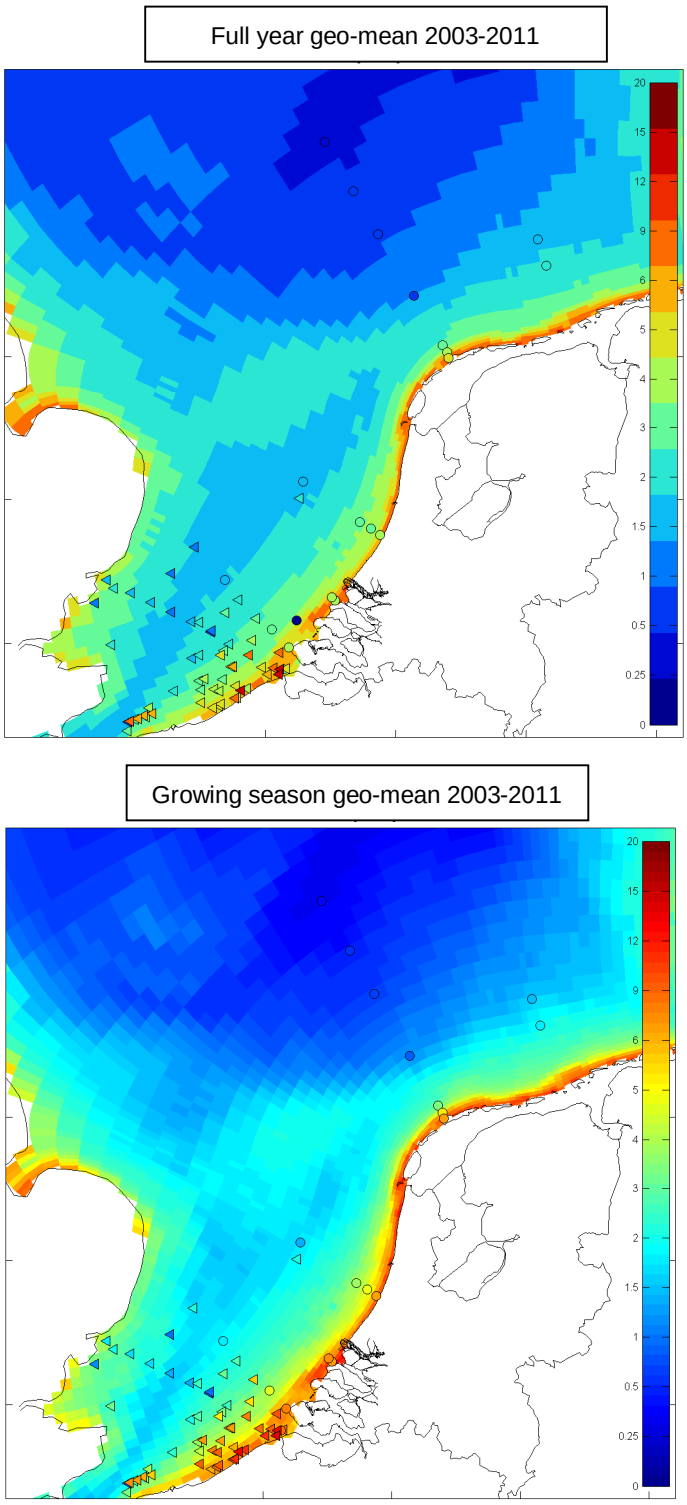


Figure 5.2 Top: multi-annual geometric mean chlorophyll-a values of the full series of gridded MERIS data and of the time series of in situ data by RWS (circles) and MUMM (triangles) for 2003-2011. Bottom: as top but only determined for the growing season, defined as March to September (inc.). Colours correspond to the values in the colourbar (µg/l).

From the maps in Figure 5.2 it can be seen that spatial gradients in the mean value of the MERIS data and the *in situ* data are mutually consistent on the large scale. Relatively high annual mean near-shore values of over 5 µg/l are in particular reported in the MERIS data and in the MUMM data off the Belgian coast. The MWTL data generally show lower annual mean values compared to the MERIS data close to the Dutch shore. The MUMM data close to the Belgian shore tend to report higher mean values than the MERIS data. Also the IS data at the UK coast show lower values than the MERIS. For the growins season these biases are less strong, but still partially visible. Baretta-Bekker (2013) also discussed the differences between the MWTL data and the MERIS data at the matching *in situ* locations. She noticed an underestimation of MERIS compared to IS data in particular in the coastal waters in terms of growing-seasonal mean values, but higher MERIS than IS values outside the growing season. This might be attributed to the retrieval and its sensitivity to adjacency of the coast, to high turbidity, and/or to seasonal variations in optical properties of the water, but it might also be due to differences in sampling with lower time resolution over the strongly variable near coastal waters. A closer analysis would be required in order to distinguish the optical from the sampling aspects.

Offshore annual mean values in the central Southern Bight are about 1.5 to 2 µg/l both in the *in situ* data and MERIS data. The local minimum between the continent and the UK appears to be captured both by the IS and MERIS data. The MUMM data, however, tend to show higher mean values in the region about 50 km offshore. Towards the north (Oyster Grounds, and Dogger Bank) annual mean concentrations drop below 1 µg/l consistently for both IS and MERIS data.

The maps in Fig 5.3 and 5.4 show an estimate of the standard error in the mean in both data sets in absolute and relative sense, respectively. As explained in chapter 3 the error in the mean is estimated taking care of number of samples in each series as in equation (3.1). This is a first approximation to the standard error in the mean where it has been assumed that the samples are statistically independent in time. This independence is partly justified as Laane et al. (*in prep.),* Blauw et al. (2012) and others have shown from high-time resolution smart mooring observations that in typical temperate climate shelf-sea conditions, the autocorrelation time of chlorophyll-a is of the order of 5-7 days, whereas the subsequent IS and RS samples on a location on average are about 5 days apart. In practice however the samples of MERIS are clustered in non-equidistant subsets for which a correction for the autocorrelation would be required (see e.g Blaas & Van den Boogaard, 2006). A fully corrected (unbiased) estimate of the error in the mean can be obtained from the MERIS data, but this requires a much more extensive analysis and knowledge of the autocorrelation at every location which is practically impossible because there are no high-frequency measurements at every location.

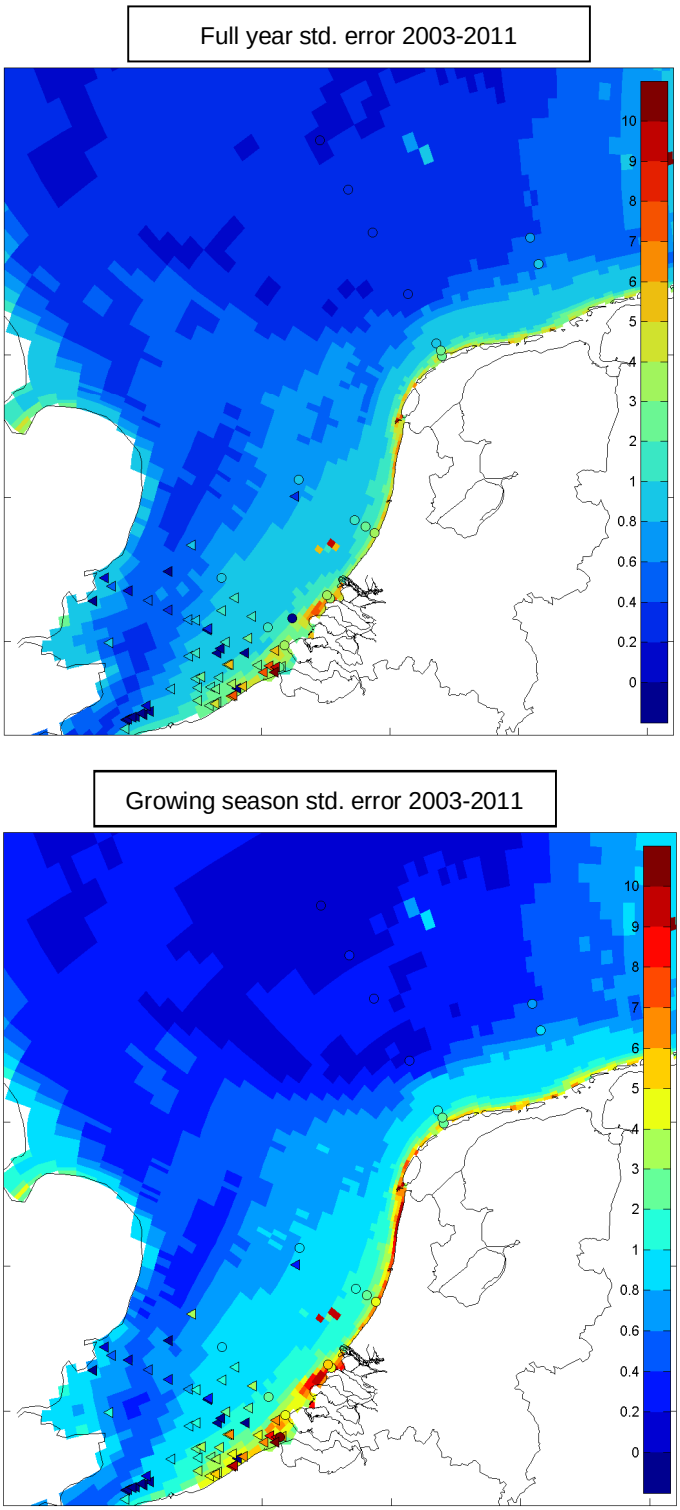


Figure 5.3 Top: Estimate of the standard error in the multi-annual mean (full period, in µg/l) estimated by the spread in the series divided by the square root of the number of samples at each location, hence presuming statistical independence in time. Despite lack of correction for temporal autocorrelation, the order of magnitude of the values is expected to be correct (see text). Bottom: as top but only determined for the growing season March-September for all years 2003-2011.

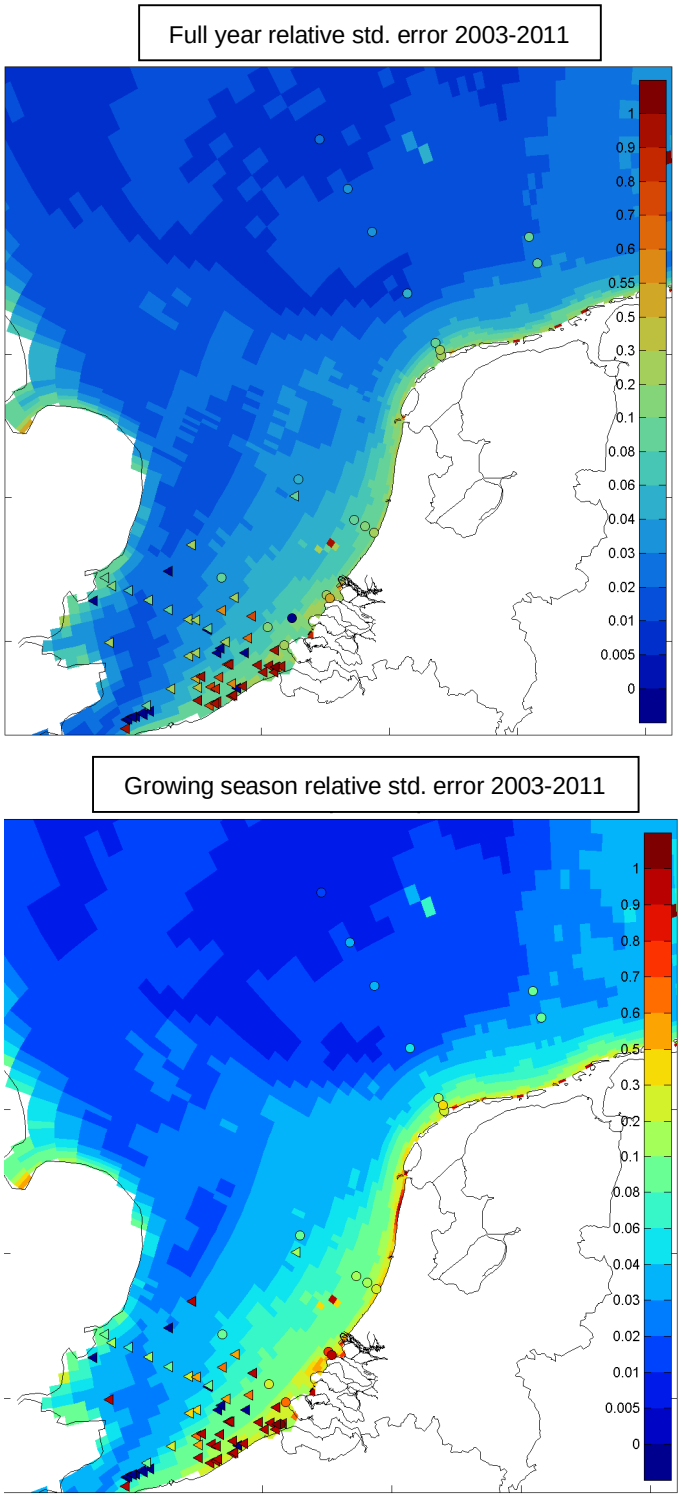


Figure 5.4 As Figure 5.3 but for the relative error in the multi-annual mean (standard error in the mean divided by the geometric mean). Dark red colours indicate a relative error of over 100%.

The distribution of the error in the mean in the figures above shows that in absolute and relative sense the RWS data show a spread in the mean very similar to the gridded MERIS data. This indicates that both sampling schemes have a similar description of the temporal variability at each individual location. The MUMM sampling, however, is much lower in time resolution, which results in higher absolute and, especially, relative errors in the mean. In the coastal water off Belgium the relative error in the estimate of the local mean chlorophyll-aconcentration reaches over 100%.

These estimates of the error in the mean will change when a regional-temporal mean is applied instead of a local temporal mean only: both MUMM and MERIS data have a wider coverage and hence more samples can be used to estimate the temporal regional mean with higher accuracy (taking into account spatial autocorrelation). The MUMM monitoring strategy is different from the RWS strategy in the sense that per location fewer samples are taken, but that the spatial coverage and resolution is higher. It is beyond the scope of this study to compare the power of both the MUMM and RWS strategies and judge the trade-off between spatial and temporal resolution. The MERIS data combine the best of both worlds: higher spatial and higher temporal resolution. The differences in regional statistics based on these MERIS data compared to the same statistics based on the RWS MWTL data will be explored in Chapter 6 as far as relevant for a typical OSPAR eutrophication assessment.

As a final comparison the multi-annual 90-percentile is shown over the entire period and growing season in Figure 5.5. This gives an impression of the upper tail of the distribution of samples, as that is relevant in the assessment in Chapter 6 for the growing season.

Figure 5.5 shows that the variability of the extreme part of the distribution behaves similar to the error in the mean in Figure 5.4. A clear pattern of high near-shore and lower offshore variability reflecting the effect of riverine nutrient inputs combined with coastal trapping mechanisms and alongshore residual transport result in a clear zonation not only of the mean but of the variability as well. For most sites the 90-percentiles of both *in situ* and remote sensing data are consistent, a few notable exceptions are the very near-shore high MERIS values off the Dutch coast not observed or captured *in situ* (in particular visible for the growing season) and a pair of MUMM stations and the Rottumerplaat 50 station that stand out with higher IS 90-percentiles than MERIS, whereas other MUMM stations show lower variaitons, in particular in the growing season. It should be kept in mind that the temporal resolution of the MUMM stations is lower than for the MWTL stations (donw to less than 3 samples per year, see also Figure 4.5) which means that the estimated percentiles thus are less reliable.

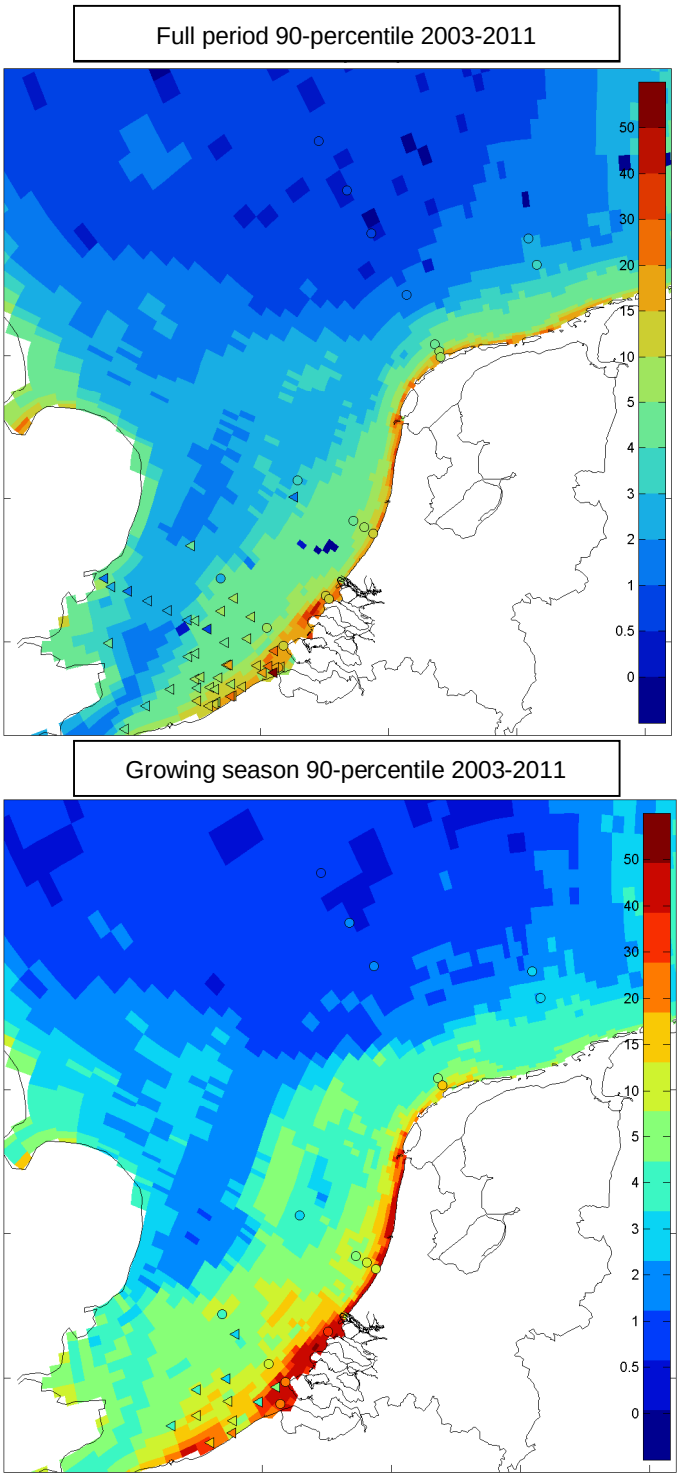


Figure 5.5 As Figure 5.2, but for the 90-percentile of the set of samples at each location.

5.2.2 Time series

In order to illustrate not only the spatial pattern of the mean and spread in the data, but also the temporal behaviour a comparison between the matches of the gridded time series of MERIS and the IS series is made here. Matching here implies that only directly corresponding samples in space (grid cell) and time are considered which results in a limited number of matching pairs. For the MUMM and MWTL data a time window of 24 hours (+/- 12 hours around the moment of MERIS observation) has been applied. For the 15 minutes time-resolution CEFAS data a window of 12 minutes has been considered such that every sample has a unique match. From the matching series also the differences in the mean (bias) and variance (unbiased RMS error) are determined. The limited number of matches however means that the significance of estimates of bias and RMS error may be limited. These numbers have not been included here.

Also please note that the chosen time window of 24 hours for the low-frequency samples is wider than applied in standard RS retrieval pixel validation studies where commonly a window of +/- 1 hour is taken (e.g. Peters et al., 2008; Tilstone et al., 2012). A narrow window is relevant when the data vary strongly in a 24 hr interval e.g. due to tides and other short-term dynamics. In the current studies we however already aggregated the MERIS pixel data onto the model grid, hence, implying some spatial smoothing of the underlying spatial variations. The aim of the current studies is not to do a precise pixel validation but to discuss the representation of features in both data sets on a time scale of a day or longer. Therefore the current matchup method is deemed an acceptable compromise between precision and getting sufficient number of matching samples. Note that using the gap-filled data will partially remediate this lack of matchups, but still the ideal validation approach would be to apply data from semi-permanent moorings such as the CEFAS SmartBuoys on key locations, as for example done by Nechad et al. (2011).

The panels below show a selection of time series of the gridded MERIS data matched to the IS data in space. The local temporal matches in the time window are indicated as well by red markers. Also the number of samples and matches is given. The goodness of fit of the matching pairs is summarized in scatter diagrams at the end.

In the figures below, for each region a key station has been selected, the other stations are shown in the appendix. Since Station TS135 is the only station that also is monitored by the shared CEFAS-RWS SmartBuoy Oyster Grounds, we include the time series and matchup also for that high frequency series. Since the sampling frequency of the SmartBuoy is 15 minutes, the matching has been done to the nearest sample value within that window around the MERIS data. At the end of this section the scatter diagrams containing all matching sets are shown.

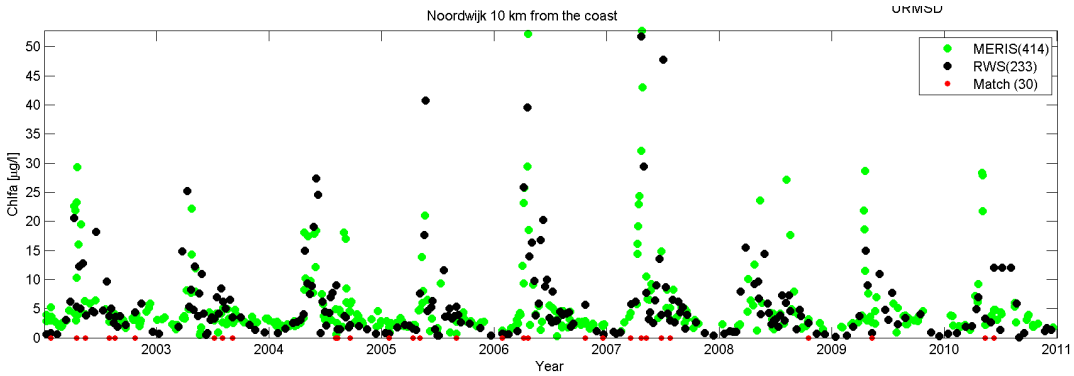


Figure 5.6 Time series of observed chlorophyll-a at Noordwijk 10 km offshore (taken as representative for Coastal Waters) by RWS MWTL (black squares), from the gridded MERIS (green circles). The red dots indicate when pairs of squares and circles match in a window of 24 hrs. The legend indicates the number of samples and matches, respectively.

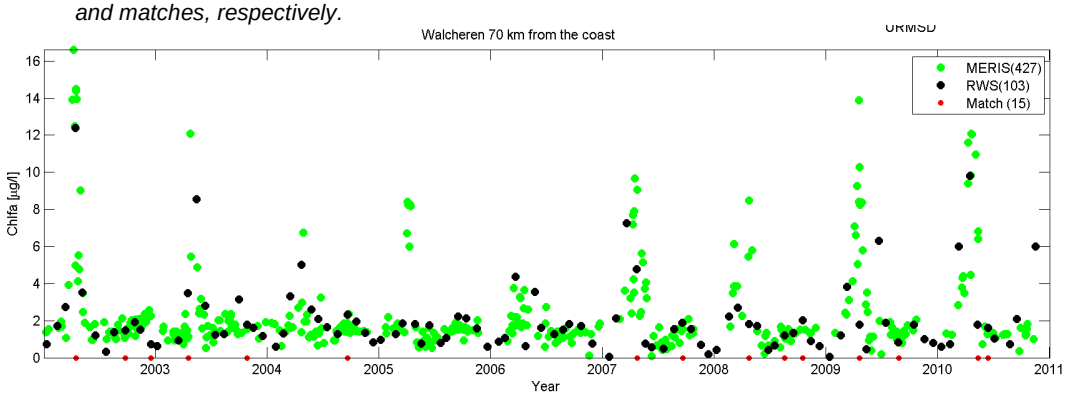


Figure 5.7 As Figure 5.6 but for Walcheren 70 km offshore, representative for the Southern Bight.

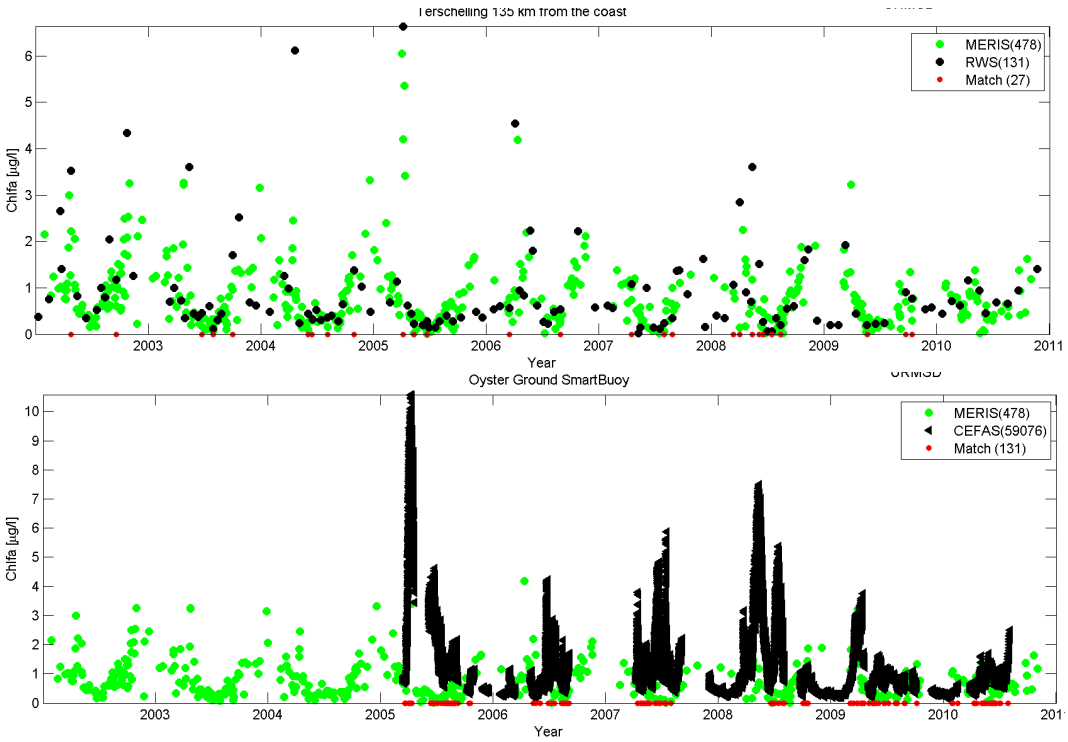


Figure 5.8 As Figure 5.7 but for the RWS in situ samples (top) and CEFAS SmartBuoy data (bottom) at Terschelling 135 km offshore, representative for Oyster Grounds.

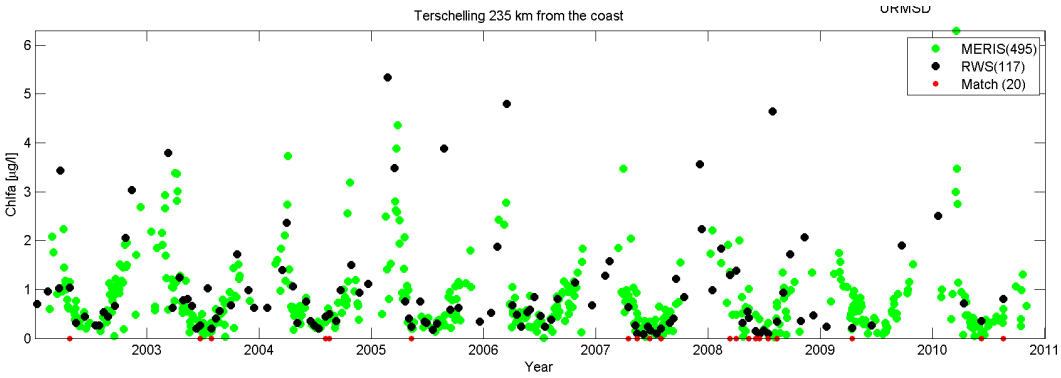


Figure 5.9 As Figure 5.6 but for Terschelling 235 km offshore, representative for Dogger Bank.

The time series illustrate that on all locations the general seasonal variations are captured in both series but that the MERIS data show two to four times higher temporal resolution and hence better describe the spring bloom peaks. Since the peak is relatively narrow with respect to the *in situ* sampling frequency the higher RS sampling results is less variation per year of the estimated seasonal peak than the IS data. This means that the distribution of MERIS data will contain more samples from the peaks and hence the higher percentiles may be higher in the MERIS data than in the IS data. This will be discussed more extensively in Chapter 6.

Differences in dynamics are visible from the various series: the timing of the spring bloom, the relative and absolute magnitude of the bloom and the occurrence of autumn blooms all vary in space depending on the physical and biogeochemical conditions in the regions.

It is also important to notice that the matchup pairs of MWTL data and MERIS data are limited in number (usually 30 or less, i.e. less than 20% of the MWTL samples). Moreover, the matchups are in many cases clustered in time such that spells of more than a half year or even a year without matchups occur. The goodness of fit between the matches thus is biased by these clustered intervals. It is therefore advised not to draw strong conclusions from the matchups at individual locations, but instead to look at the larger scale goodness of fit. The CEFAS mooring provides many more samples and matches, and thus is more suitable for a local goodness of fit assessment. The notable characteristic of the SmartBuoy series is that the summer values of chlorophyll-a by the mooring are relatively higher and are showing multiple subsequent peaks, not present in the MWTL and MERIS data. It is yet unknown if this is a genuine feature or an artefact of the sensor measurements e.g. due to bio-fouling. For a more extensive discussion on the analysis of the SmartBuoy chlorophyll-a data (which are based on fluorescence) see for example Blauw et al. (2012); Kröger et al. (2009) and Mills et al. (2005).

As summary of the goodness of fit over all MWTL stations and for the CEFAS SmartBuoy at Oyster Grounds is given in the scatter diagrams below which include also the linear regression coefficients.

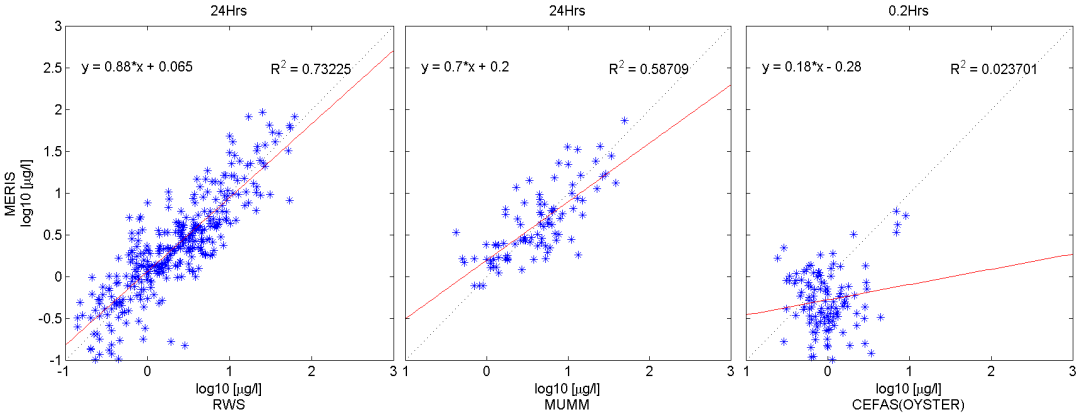


Figure 5.10 Scatter diagrams of all matches between gridded MERIS samples and RWS MWTL in situ samples (left), MUMM samples (middle) and CEFAS smartmooring data at Oyster grounds (right). All values are ^{10}log µg/l chlorophyll-a. Plot title indicates matching time window: 24 hrs for the RWS and MUMM data, 0.2 hrs for the CEFAs smartmooring data. Please note that the RWS and MUMM diagrams show both spatial and temporal variations in one plot with relative few temporally compared to spatially distributed data, whereas the CEFAS plot only shows temporal variations. The black dotted line indicates the 1:1 relation of both variables.

5.3 EOF analysis

5.3.1 Smooth DINEOF on MERIS and IS

As indicated above, the DINEOF method makes a linear decomposition of the variations in the data in time and space. It produces a series of modes of variation. These modes are pairs of a spatial pattern (map) linked to a temporal pattern (time series) for the entire domain of the model grid and all time instances where there has been more than 2% coverage. Every pair of spatial maps and series is referred to as empirical orthogonal functions (EOFs), or principal components in space and time. For each pair of map and series (mode) a movie is obtained when multiplying the map with the time series.

Each EOF mode captures a certain amount of the total variance present in the total dataset. By ranking the EOFs from high to low variance, we get an impression of the dominant pieces of information in the data. In Table 5.1 the explained variance is shown per mode for an EOF analysis with high smoothing (α=0.1) and for an EOF analysis with low smoothing (α=0.01). The first mode clearly captures most of the variance, about 2/3 to 3/4. The second mode contains 6 to 14 % of the explained variance, while the third mode is only 2% of the Chlfa variance. DINEOF was performed in log-space, analysis in linear space failed for Chlfa. In following we will analyse the dominant modes. Here we discuss the smooth modes (α=0.1) , and try to relate them to known biogeochemical dynamics. In the time series analysis and OSPAR assessment we will apply the less smoothed modes, and assess how useful they are for gap-filling data to be used for extra artificial match-ups with IS data. For completeness and comparison the time vector and spatial pattern of the less smooth modes are also shown in Appendix B.

Table 5.1. Explained variance (%) by each of the EOF modes determined for the MERIS HYDROPT griodded chlorophyll-a data set for relatively strong (α=0.1) and weak (α=0.01) temporal smoothing in the DINEOF analysis.

Smoothing (α)	0.1	0.01
Parameter	Chlfa	Chlfa
Mode 1	72.67	67.33
Mode 2	8.37	10.37
Mode 3	2.33	2.01
Mode 4	1.27	1.88
Mode 5	1.23	1.78
Mode 6	1.08	1.76
Mode 7	0.90	1.73
Mode 8	0.57	1.28
Mode 9	0.54	0.83
Mode 10	0.50	0.83
Mode 11	0.45	0.72
Mode 12	0.40	0.56
Mode 13	0.36	0.52
Mode 14	0.36	0.47
Mode 15	0.33	0.47
Mode 16	0.30	0.41
Mode 17	0.26	0.40
Mode 18	0.26	0.39
Mode 19	0.24	0.37
Mode 20	0.23	0.32
Mode 21	0.20	0.31
Mode 22	0.20	0.27
Mode 23	0.19	0.26
Mode 24		0.21
Mode 25		0.19
Total (%)	93.24	95.67

EOF mode spatio-temporal mean: EOF analysis is an empirical method that does not know anything about the physical or biological mechanisms underlying the variability in the data. EOF analysis works best on a normalised dataset. For this reason we transform the data by applying a ^{10}log on the concentration values. The DINEOF toolbox removes the mean value of the entire dataset, one scalar value. This is simply the mean of all active pixel values in the hypercube *[x,y,time]*. For interpretation of the EOF modes the prior removal of this value should be taken into account.

EOF mode sign and normalization: EOF analysis is a mathematical method that is neutral to the sign of the modes. To facilitate analysis in terms of physics we have normalized the spatial and temporal modes. The temporal mode is divided by its RMS (Root Mean Square) value, such that the temporal mode is a dimensionless signal that varies roughly between -1 and 1. A perfect harmonic sine wave would oscillate exactly between -1 and 1. A skewed signal, such as the annual algal bloom, could lead to amplitudes larger than 1, with the RMS still being one. The spatial mode is multiplied by the RMS value of the temporal mode and with the singular value (the scaling value of the mode) such that is presented as a North Sea map with the physical quantities of chlorophyll-a. This map can be compared to the well-known maps e.g. of the annual mean. As a last step, we swapped the sign (if needed) to make the temporal mode match known system properties: chlorophyll-a is higher in the spring/summer, while SPM is higher in the autumn/winter.

EOF mode and seasonal cycle The dominant signal of SPM and Chlf-a in the North Sea is known to be the seasonal cycle. We accommodate this *a priori* knowledge by chopping up the 9-year MERIS temporal mode into annual subsets. We plot these subsets in the same reference year to visually assess the seasonality of the temporal modes and its interannual variations.

EOF mode figure explanation The upper left panel of all Figures 5.11, 5.12 and 5.13 contains the spatial mode (map) in Chlfa units, obtained by multiplying the spatial EOF mode U[x,y]p with the singular values Sp and the RMS of the temporal mode V[time]p. The lower panel contains the normalized first EOF temporal mode V[time]p in red, with all higher modes in black (all modes are listed in Table 5.1). The upper right shows the first temporal EOF mode once more but now chopped up into annual subsets. Note the symmetric vertical limits of the temporal mode panels because of the normalisation. This figure is repeated for both the MERIS data (on a daily time basis) and the MWTL data (on a 4-weekly time basis as explained in section 4.3), and for all dominant modes. Please note that only the 19 currently visited MWTL chl-a stations on the North Sea have been used.

First chlorophyll-a EOF mode The first EOF mode in Figure 5.11 reflects a seasonal variation for both MERIS and IS which is fluctuating slightly on the interannual scale. The spatial MERIS pattern is close to the regular geometric mean shown earlier (Figure 5.2), the difference being that the pattern is moving up and down from winter to summer. The chlorophyll-a map and points are modulated by one order of magnitude between winter and summer, as seen by the temporal mode undulating between 0 and 1. The seasonal variation is quite symmetrical with a maximum in the summer-., half of the year and a minimum in the winter-half of the year especially for MERIS. Note that the spring bloom, responsible for most of the chlorophyll -a maxima relevant for environmental monitoring, is absent in the first EOF mode for MERIS but is partially visible in the first mode of MWTL data. The spatial gradients are obviously better discernible in the MERIS results than in the MWTL results. In the Dutch coastal area the dominant gradients are mostly cross-shore. Overall Chfa patterns seem to correlate with depth (Dogger Bank partly), turbidity and nutrient availability (river water distribution). The IS results show the lower values in the central North Sea (blue), and some higher values near the coast (orange). However, the IS values in the coastal zone are generally quite flat from near coastal to the central Southern Bight (green), whereas MERIS clearly produces a strong gradient from high near the coast (reds) to moderate in the central Southern Bight (green).

The dominant temporal pattern in both MERIS and IS is the overall spring to summer elevation and winter low that explains 73% of the total variance over all years and the entire domain. This pattern is indeed known to dominate the basic dynamics of chlorophyll-a but contain yet little specific information as to where and when the spring bloom is exactly, how long it lasts and wheter is changes from year to year.. Effectively this first mode is reflecting more spatial contrast and variance than temporal variance. Apart from the scalar mean value that was removed beforehand, it reflects the a basin wide mean pattern with strong spatial gradients moving up and down over the seasons.

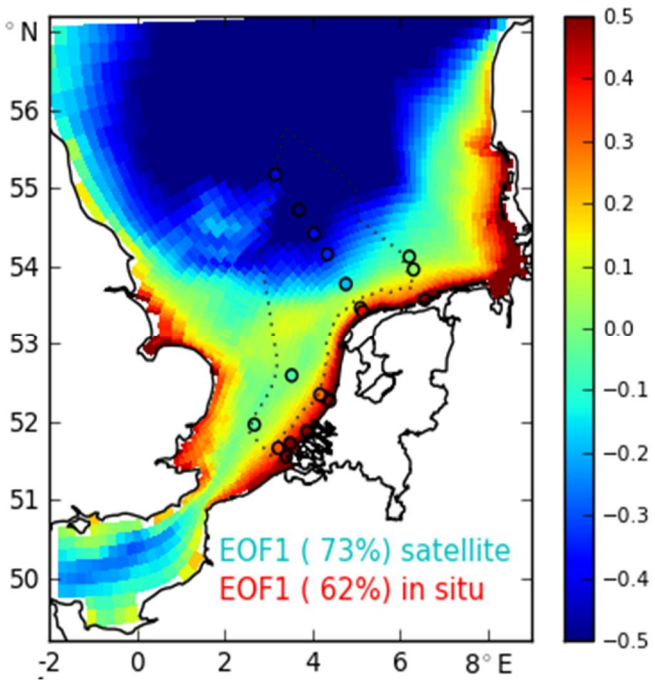


Figure 5.11a. Spatial pattern of first EOF mode of MERIS (shades) and in situ MWTL (dots) for Chlfa, obtained with high smoothing (α=0.1).

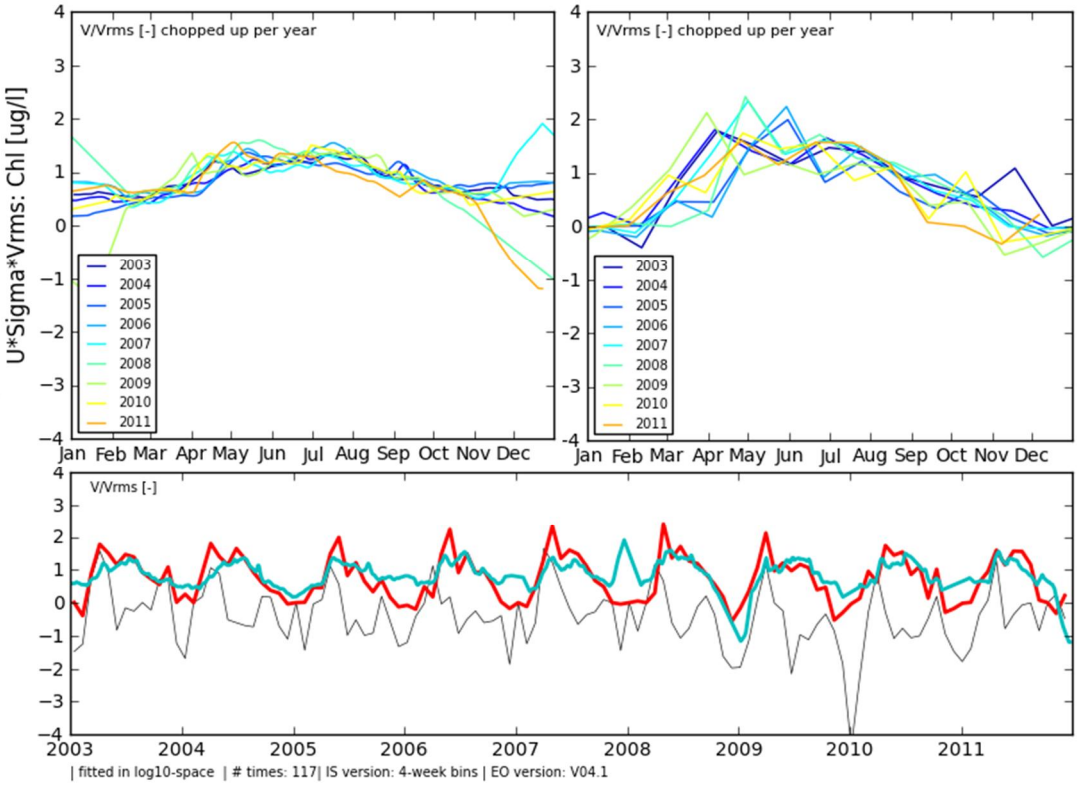


Figure 5.11b. Time series of first EOF mode of MERIS and MWTL Chlfa, shown in Fig. 11a. The top panels show MERIS (left) and MWTL (right) temporal mode chopped into annual pieces and overplotted to illustrate the interannual variations. The bottom panel shows the full temporal pattern: the blue line is the MERIS mode, red is the MWTL mode, the black line is the other significant mode obtained for MWTL. Remark that the MWTL modes are based on time series with 4-weekly resolution, the MERIS modes are based on daily resolution data.

Second chlorophyll-a EOF mode: The second chlorophyll-a EOF mode in Figure 5.12 reflects more details of the typical spring bloom events for both IS and MERIS. The overall amplitude of the temporal mode ranges over several orders of magnitude, from -2 to +2 (on the ^{10}log scale) and even +4 for MERIS (consistent with what can be seen in the time series comparison). Next to a peak in spring, a second peak is visible in the autumn for both IS and MERIS. Both IS and MERIS also show the relative lows in February and in mid-summer. The second mode clearly captures the spring bloom that is responsible for the high chlorophyll-a values that environmental indicators are concerned with. Despite the importance of the spring bloom, this mode only contains 13% of the variance for MERIS and 30% for IS, while the seasonal variation in mode 1 contains respectively 73% and 55%. This difference in variance is explained by the fact that most variance in the MERIS and MWTL data is spatial. The first mode mostly reflects that. The seasonal pattern is superimposed on top of the spatial pattern: it has relatively less variation in space but more variation in time.

The second mode is significant in statistical sense (as determined by the EOF tools), but also important in terms of monitoring. For OSPAR it is relevant what the year-to-year deviations from the typical pattern are, not so much the typical pattern itself For environmental parameters that are concerned with extremes, such as the growing season mean and 90% percentile, the temporal variability of this mode is more relevant than the mean spatial variation. In the OSPAR assessment the spatial variation is already accounted for by the definitions of the regions and the aggregation. (A chlorophyll-a-data based classification of regions could thus be done based on the first mode.)

Note that the second MERIS EOF mode contains more variation than the corresponding IS mode. The MERIS modes show modulation from 0 to 4 in spring, whereas IS only shown modulations from 0 to 2. As indicated in section 5.2 this is due to the higher temporal resolution of MERIS that allows a more detailed pinpointing of the spring bloom. Apart from the peak statistics, also the timing of the onset of the spring bloom is better resolved by MERIS. The MERIS mode shows a full month starting range from March to April. The 4 week bins in IS, related to the approx. 4 week revisit interval, do not allow to pinpoint the spring bloom at that detail. The one month variation in onset of the spring bloom apparent from MERIS data shows that 4 week IS revisit intervals can lead to missing the spring bloom altogether, catching only the flanks or accidentally catching the a maximum with no extra measurements on duration or to determine a reliable 90% percentile value within certain limits.

The MERIS map show that the overall spatial contrast of this mode is lower than that of the first mode. The high values for this mode as shown by MERIS occur in the central North Sea, in the mid Southern Bight and in the parts offshore of Zeeland. The gradient along the Terschellig transect towards the central North Sea high is captured by the IS data, but as with the first mode, the near coastal IS pattern is quite flat and too low in resolution to match the pattern resolved with the MERIS data.

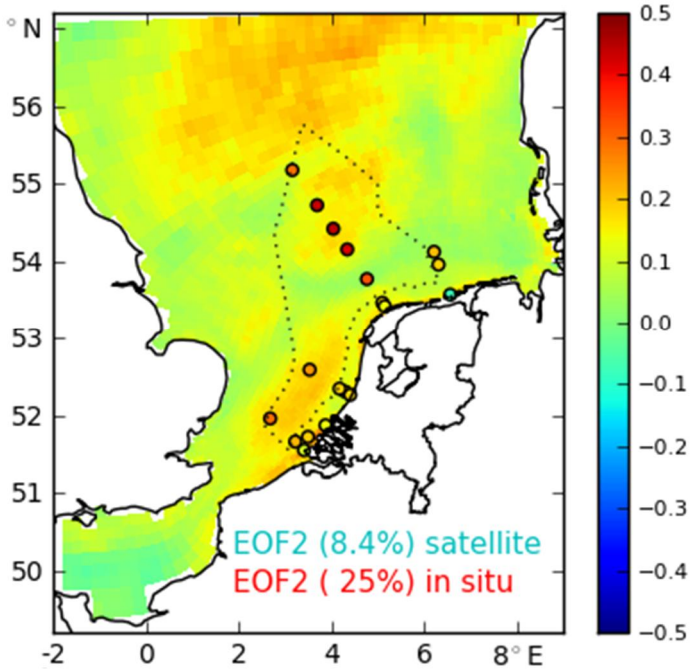


Figure 5.12a. Spatial pattern of second EOF mode of MERIS (shades) and in situ MWTL (dots) for Chlfa, obtained with high smoothing (α=0.1).

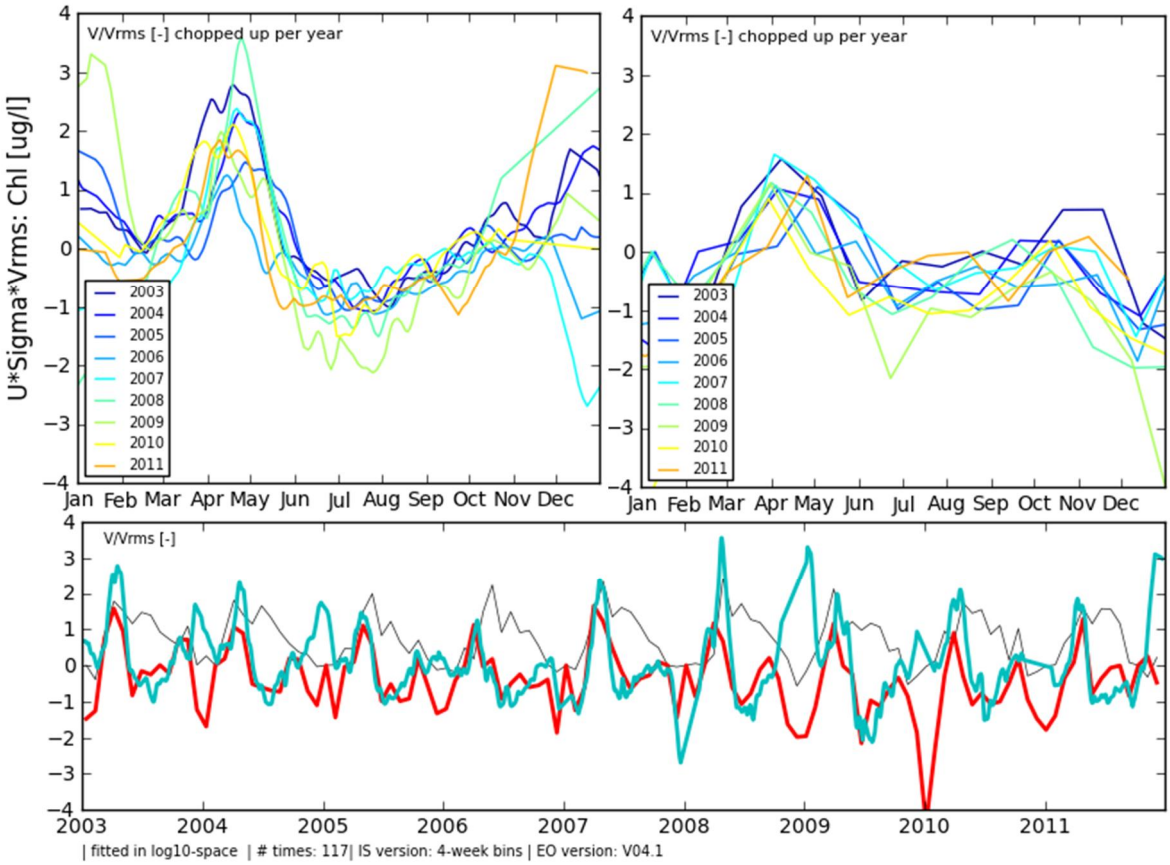


Figure 5.12b. As Figure 5.11b, but for the second EOF mode of MERIS and IS chlorophyll-a obtained with high smoothing (α=0.1).

Third chlorophyll-a EOF mode: The third chlorophyll-a EOF mode in Figure 5.13 captures 2.3 % of the variance. Still, this is about a quarter of the seasonal (second) mode and as such is relevant as the deformation of the seasonal cycle. The DINEOF analysis of the IS data only produced two significant modes, whereas MERIS resulted in more than 25. Hence, the IS resolution is not sufficient to reliably capture this pattern from the data, the individual samples are not correlated strongly enough to distinguish dynamical patterns from smaller scale noise. We can therefore not compare this mode between IS and MERIS.

Still it is worthwhile to discuss the mode for MERIS. This pattern has a strong seasonal signal just like mode 1, but has a distinctly different spatial pattern. Because the information content of the IS data is not sufficient to capture more modes, it is difficult to decide whether this mode has a physical nature. We performed a consistency check on this third mode by doing a DINEOF analysis on the full hypercube with > 200 images, including all images with less than 2% coverage. The resulting pattern in Figure 5.13 is identical to the one with only ~1500 proper images. Adding extra low-quality data apparently does not influence the mode. We also modified the temporal smoothing parameter α from 0.1 to 0.01 which allows for more temporal variations in the EOFs (and thus results in more 'noisy' patterns). The modification of parameter α affects all 3 dominant modes to some extent. The first two modes are relatively less affected, the third one is more strongly affected (see Figures 5.13). This indicates that this third mode is less robust as might be expected from a mode explaining about 2% of the total variance. Still it might be interpreted in terms of the known system dynamics: it can be regarded as fine tuning the shape of the spring bloom with its timing varying from south to north.

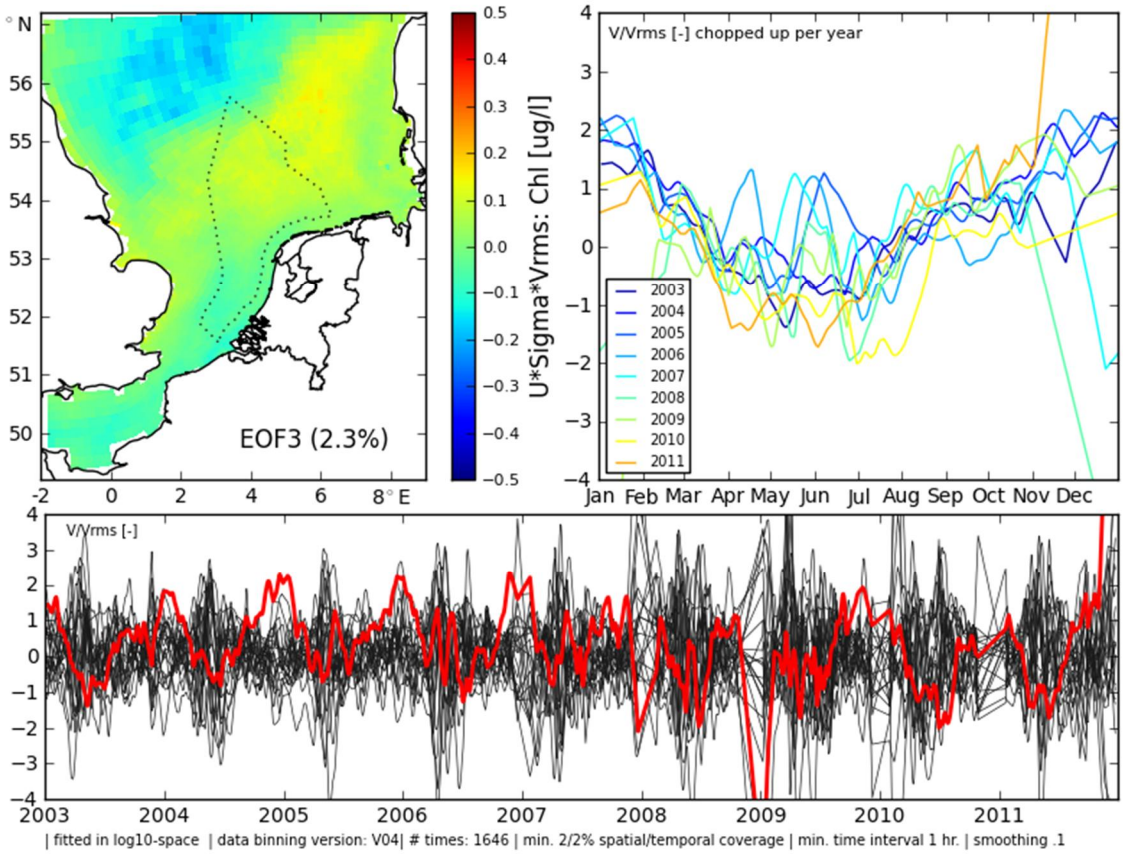


Figure 5.13. Third EOF mode of MERIS chlorophyll-a obtained with low smoothing (α=0.01).

Using Chlfa EOF modes for reduced order representation: In the study of De Boer et al. (2012) the EOF modes of SPM were used to conclude that the combination of the spatio-temporal mean the first mode are a robust combination to act as reduced reality representation with the same information content as the annual geometricmean. From this study we see that this mode even captures more variance using a 9-year MERIS dataset than a one-year MERIS dataset. For environmental monitoring, the acceptable changes of SPM are usually determined in terms of background concentrations, e.g. for the Maasvlakte 2 constructionimpact monitporing. The annual mean and the first EOF mode appear good proxies for this.
For chlorophyll-a we can draw similar conclusions. The first mode can act as a robust estimate of the mean for all locations, even better than for SPM. However, for an OSPAR eutrophication assessment, different statistical properties than the background concentration or annual mean are required. For chlorophyll-a the 90% percentile or other criteria that capture extreme events are used. For these the first chlorophyll-a EOF mode is not sufficient. The second mode contains the significant amount of variance, related to the seasonal variation with its spring peaks.

5.3.2 Comparison of reconstructed MERIS time series with IS data (gap filling)

As indicated above, the DINEOF method produces estimates of the modes of variation. In case of the MERIS data over 22 significant modes were obtained (*i.e.* adding to the reduction of the RMS error in the residual, hence not fitting noise in the data). The modes resulting from DINEOF can be put together to make a reconstruction of the total variation in space and time thus effectively filling the gaps in the images (at space-time locations with at least 2% coverage) due to clouds and other disturbances. Because the modes are limited in number the most fine-scaled structures cannot be estimated accurately. Hence, the reconstruction is to be considered as a smoothed time-space interpolated data set. With the current set of modes obtained with the low-smoothing setting about 90% of the variance is captured in the reconstruction by the 25 modes of Table 5.1. It is worthwhile to compare this gap-filled dataset to the MWTL data in order to diagnose the differences between this data set and the IS data as was done in section 5.2 for the MERIS data with gaps. The main benefit of the gap filled data set is that we have more matchups to the IS data and hence get more robust estimates of the goodness of fit.

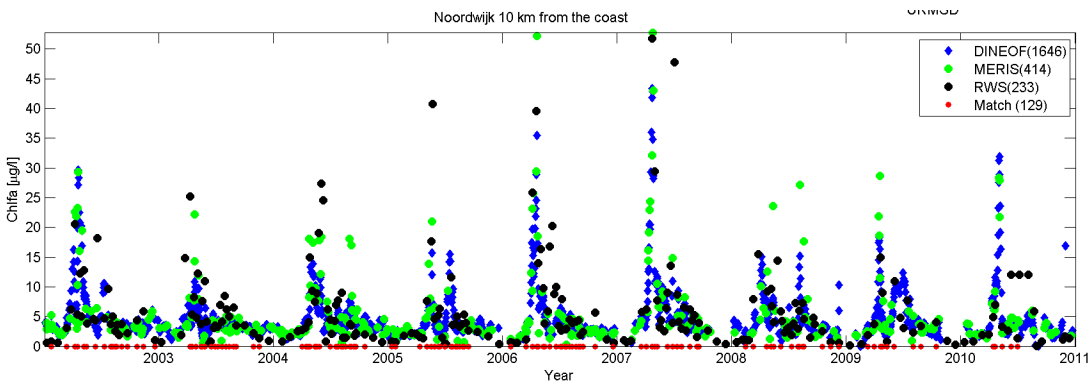


Figure 5.14 Time series of chlorophyll-a at Noordwijk 10 km offshore (as representative for Coastal Waters) by RWS MWTL (black), from the gridded MERIS (green) and the reconstruction ('gap filling') by DINEOF of the MERIS data (blue). The red dots indicate when pairs of gap-filled and in situ data match in a window of 24 hours. The legend indicates the number of samples and matches, respectively. (The number of matches increased from 30 to 129, when comparing the match with the non-gapfilled MERIS to the match with the gap-filled MERIS).

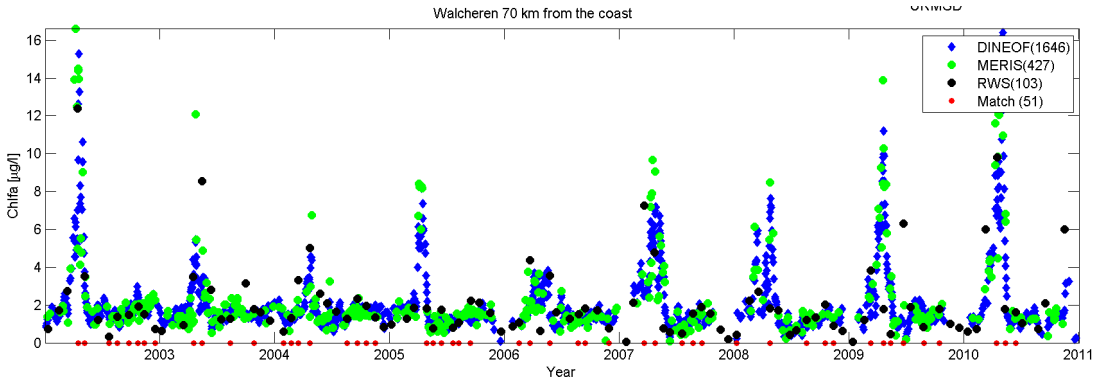


Figure 5.15 As Figure 5.14 but for Walcheren 70 km offshore, (Southern Bight). (Number of matches increased from 15 to 51 after gap fillling.)

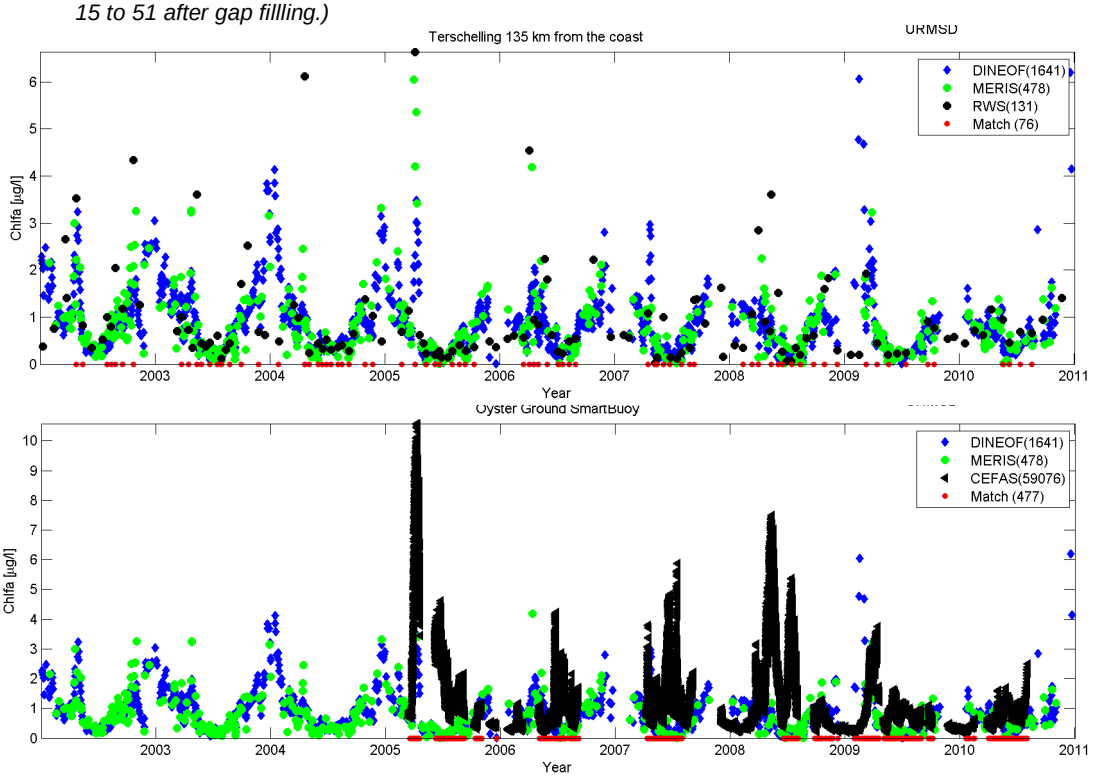


Figure 5.16 As Figure 5.15 but for the RWS in situ samples (top) and CEFAS SmartBuoy data (bottom) at Terschelling 135 km offshore, representative for Oyster Grounds. (Number of matches increased from 27 to 76 and from 131 to 477, respectively, after gap filling).

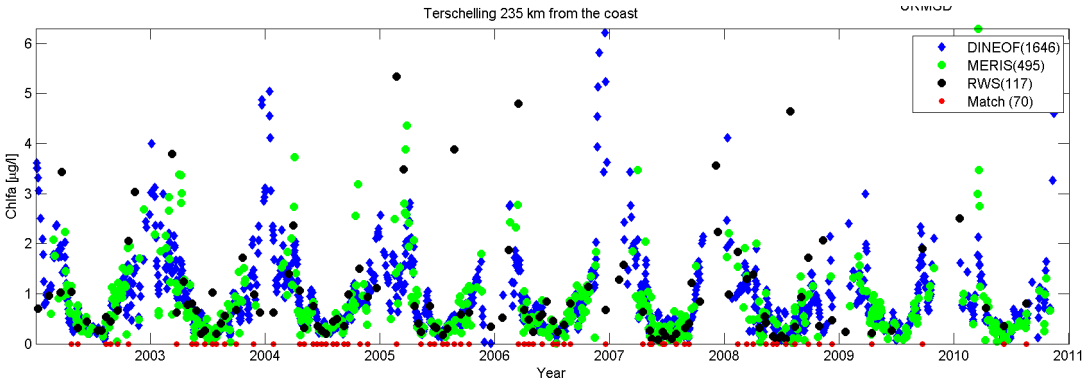


Figure 5.17 As Figure 5.16 but for Terschelling 235 km offshore, representative for Dogger Bank. (Number of matches increased from 20 to 70 after gap filling.).

From the reconstruction with the DINEOF modes in the figures above, it can be seen that local gaps are not only smoothly interpolated but that entire spring peaks (e.g. TS235 2007) can be reconstructed. This is due to the information in the surroundings of the locations shown. Other features that shows up more clearly than in the individual IS and MERIS data are the different shape and timing of the spring peak at the various locations. At Noordwijk 10 the peak is relatively narrow and occurs relatively early. At Walcheren 70 the spring bloom is lasting longer than close to shore. Further offshore, at Oyster Grounds and Dogger Bank, the peak reproduced by the DINEOF MERIS data is less a 'spring' peak but more a, relatively wide peak reaching across the winter. Visually this seasonal cycle appears to be consistent with the low-frequency MWTL sampling, but it can hardly be validated because the ship-borne sampling scheme has been tuned such as to have higher time resolution in the summer half year than in the winter half year. The semi-permanent and high-resolution CEFAS Smartmooring series may provide more grounds for understanding and validating these dynamics. Those series partially reproduce the winter peaks, but they also show frequent summer peaks that are not seen in neither the MWTL nor MERIS data. Before drawing firm conclusions it should be noted that the current Smartmooring data set may suffer from other issues such as instrument fouling and challenges relating fluorescence to chlorophyll-a without quenching artefacts (see e.g. Blauw et al., 2012; Mills et al., 2005). It is beyond the scope of the current study to explore the nature of these winter blooms. Still, it is recommended to take these issues into account in the assessment of eutrophication and the monitoring programme. If the winter variability indeed is much stronger than the summer variability in these areas and if these peaks are relevant for the eutrophication status, the sampling needs to be adjusted to resolve it. In that case, Smartmooring and Remote Sensing sampling can be complementary to the ship-borne sampling.

As summary of the goodness of fit over all MWTL stations and for the CEFAS SmartBuoy at Oyster Grounds is again given in the scatter diagrams below.

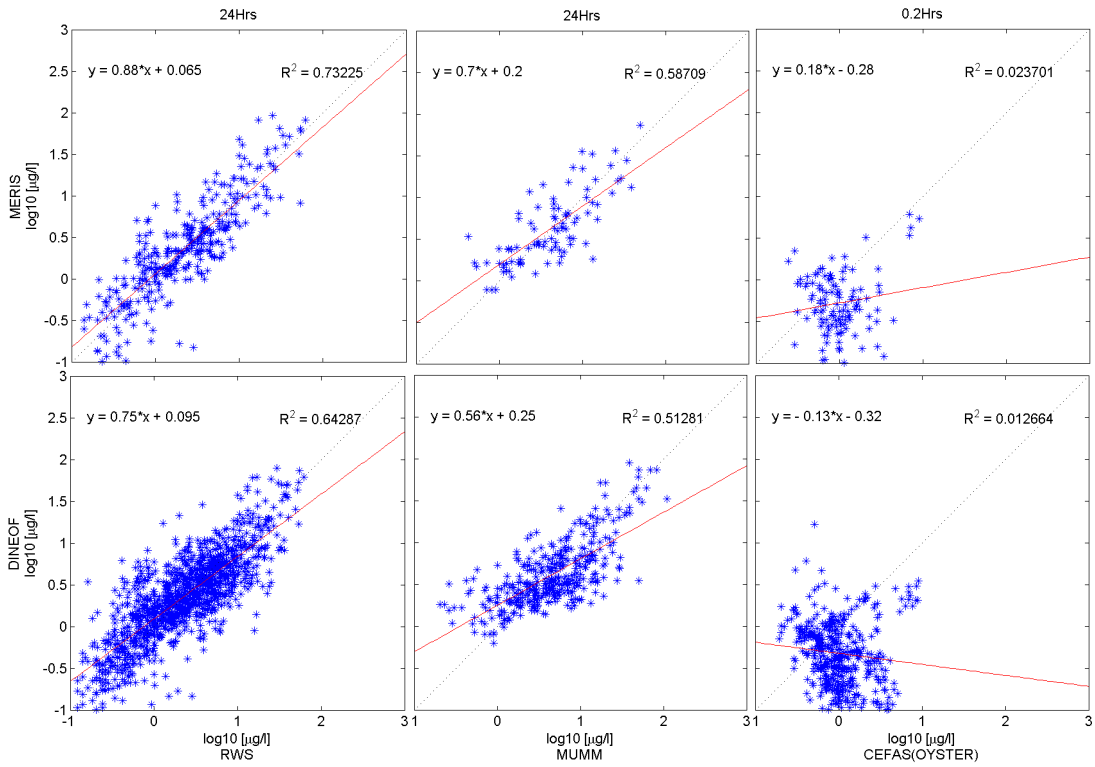


Figure 5.18 Scatter diagrams of all matches between gridded MERIS samples and RWS MWTL in situ samples (left), MUMM samples (middle) and CEFAS mooring data at Oyster grounds (right). All values are ^{10}log µg/l Chl-a. Top row is the same as Figure 5.10, included to facilitate comparison, matching gridded MERIS with gaps to the IS data, bottom row are the DINEOF gap-filled, smoothed MERIS data matched to the in situ data (more samples).

When judged from the scatter diagrams, the characteristics of the DINEOF gap-filled data (bottom row) appear similarly consistent with the MWTL and MUMM *in situ* data as the original gridded MERIS data (top row). However, the clouds of dots have not only increased in size because of the extra matches, but also the R^2 values have decreased. This may be explained by the smoothing effect of the DINEOF, which reduces the goodness of fit to individual temporal variations at the IS locations. Besides, the tendency to underestimate the chlorophyll-values has been aggravated. Apparently, the DINEOF interpolation is not only smoothing but also introducing a bias towards lower values into the data. Partially, this may be explained by the fact that a truncated and EOF set has been used in these regions, but it may also be due to the characteristics of the gappy MERIS data themselves that may contain biases to lower chlorophyll-a data. This is to be kept in mind whenever applying the DINEOF data as such. It is beyond the scope of the current project to explore the characteristics of the DINEOF method in relation to bias. The DINEOF data are not the eventual monitoring data to be applied in the assessment etc. they mostly serve as a diagnostic tool to show the effective resolution of the various monitoring strategies.

In contrast to the reasonably good fit between MERIS and bottle sample (MUMM, RWS) IS data, the comparison of MERIS data to the Oyster Grounds Smartmooring data is problematic. The gridded data with gaps already suffer from an extremely poor correlation, but the gapfilled data even appear to show a poorer correlation. As indicated above, the time series of the Oyster Grounds mooring have been applied as is without further quality check

(such as removal of outliers or flagging for fouling other than by default done by CEFAS). Most probably, the limited QC is hampering quantitative analysis for now, because it is know that other CEFAS smartmoorings are capable of producing well validated time series when compared to both bottle samples and (MODIS) remote sensing (see in particular Nechad et al., 2011). Blauw et al. (2012) already indicated that a considerable QC on the fluorescence based mooring data was required which was beyond the scope of this current KPP project. However, once such QC is in place, the Smartmooring may provide an unprecedented time resolution at key locations ideal for short-term process resolution and remote sensing matchups.

6 chlorophyll-a assessments

6.1 Introduction

In this chapter we present the part of the OSPAR Comprehensive Procedure (eutrophication assessment) that is based on the chlorophyll-a data. The aim is to explore the sensitivity of the outcome for different approaches. We compare the current standard approach for the Dutch waters as e.g. followed by Baretta-Bekker (2013) and which is based on the RWS (MWTL) *in situ* samples with a set of variants in which either station-wise time series or area-covering MERIS data are included. In this way we can discuss the accuracy of an assessment and show to what degree remote sensing data can add to the accuracy of an assessment nex to or instead of the IS data.

First we introduce the annual statistical properties of the chlorophyll-a data per region for the different assessment variants. Then, we compare our results to the 'default' assessment provided by Baretta-Bekker (2013). We will explore 7 variants (in sets) as summarized in Table 6.1 below. In each variant, either the methodology and/or definitions of statistical parameters are different, or the data used as input into the methodology are different. The methodologies range from taking only time series of data from the MWTL station location and estimate regional mean and monthly mean values from these to taking area covering data. The input data themselves are either the station-wise MWTL IS data, the gridded MERIS data without DINEOF gapfilling, and -as an exercise- the MERIS data after DINEOF gapfilling,

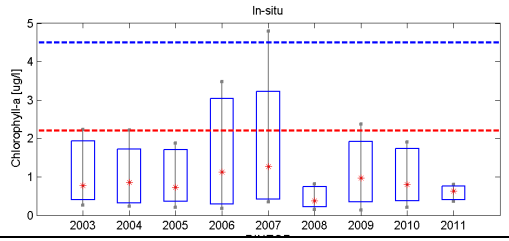
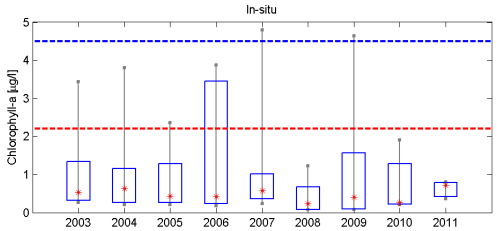
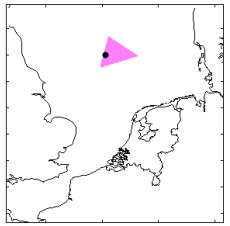
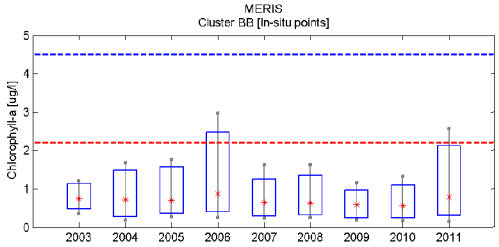
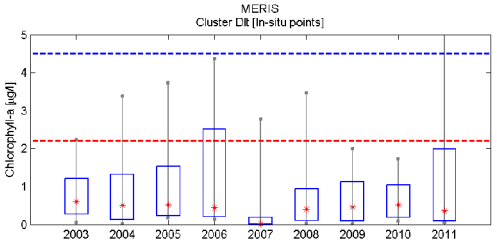
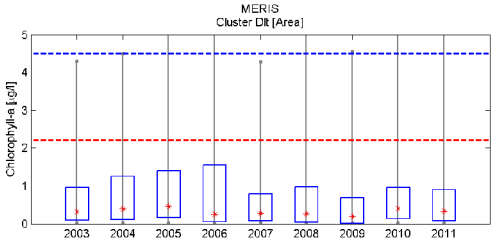
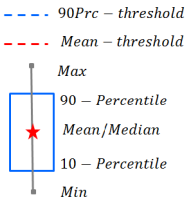
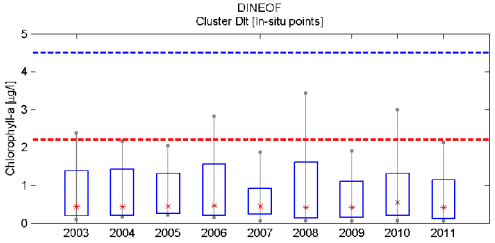
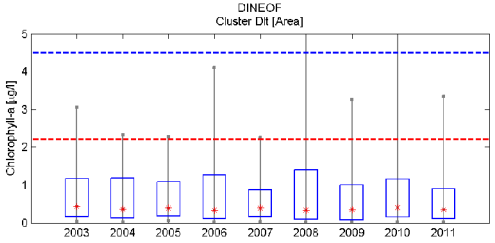
The figures below shows the box-whisker plots of the different analyses. For every region, the annual characteristics over the growing season (March-September) are shown in terms of the mean (red asterisk), 10 and 90-percentiles (box bounds) and extremes (min & max, endpoints of the grey line segments). The assessment levels (thresholds) are also shown both for the mean and 90-percentile (red and blue dashed lines). The little inserted maps show the region layout and the location of the MWTL stations used.

Table 6.1 Summary of the different variants of the chlorophyll-a analysis in the context of the Comprehensive Procedure. The second column refers to the clustering of station time series and analysis method as applied by Baretta-Bekker (2013). The third column refers to the same spatial clustering of the station time series but with modified statistical analysis. The fourth column is the assessment based on the region-covering MERIS data, again applying the modified analysis. The list of stations is given in Table 3.2.

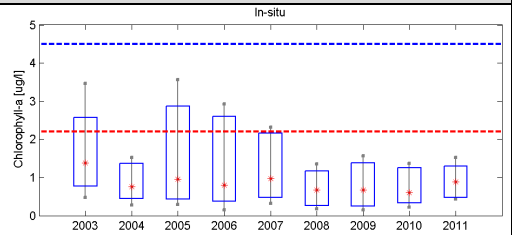
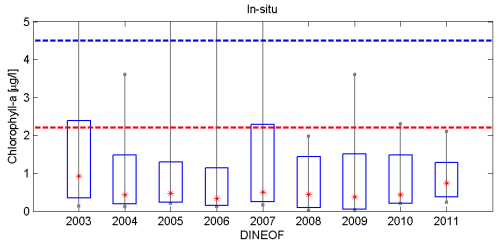
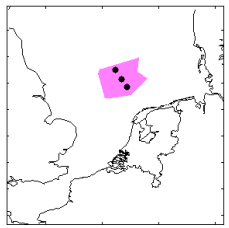
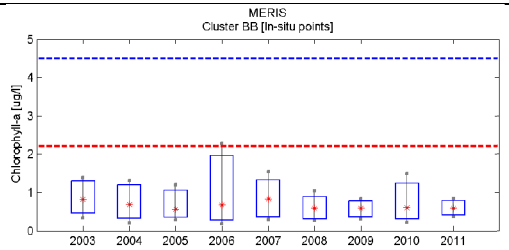
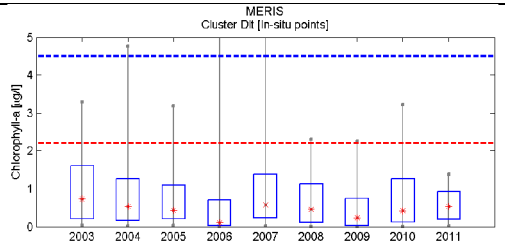
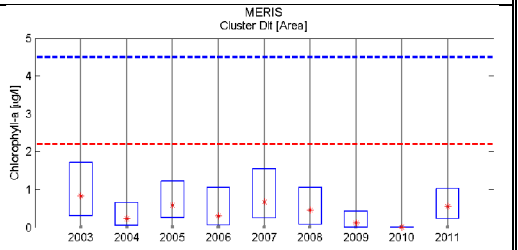
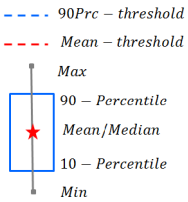
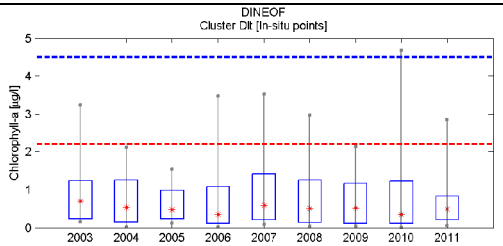
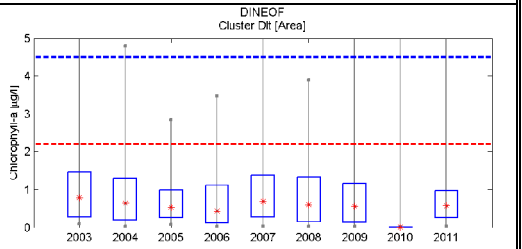
	Cluster 'Default' Baretta-Bekker (2013)	**Cluster Deltares**	**Area Deltares**
MWTL only	Only the current set of stations clustered in OSPAR regions.	Only the current set of stations clustered in OSPAR regions.	
MERIS (with gaps)	Subset of MERIS samples only on the current set of stations clustered in OSPAR regions	Subset of MERIS samples only on the current set of stations clustered in OSPAR regions	Clustered in (gridded) OSPAR regions
MERIS (DINEOF gapfilled)		Subset of MERIS samples only on the current set of stations clustered in OSPAR regions	Clustered in (gridded) OSPAR regions
Approach	First monthly arithmetic means per region, growing season (III-IX) means based on monthly means, 90% values based on monthly regional means	Statistics over growing season (III-IX) and regions without intermediate monthly averaging and with log transformation (hence geometric means instead of arithmetic means).	

6.2 Results

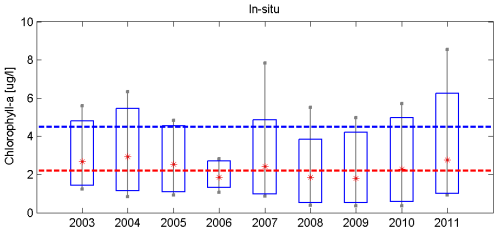
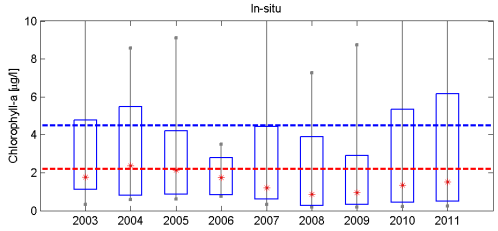
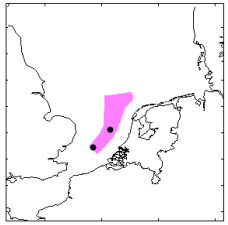
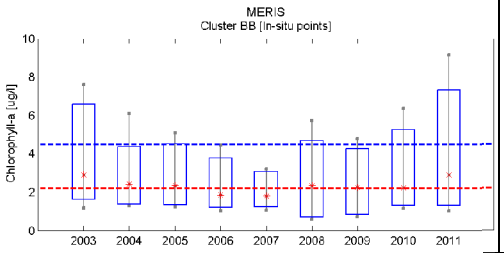
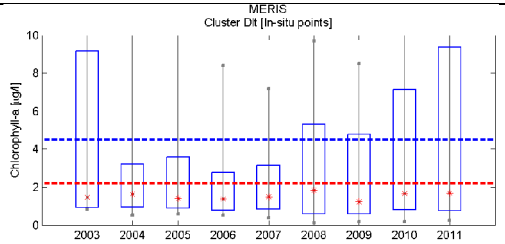
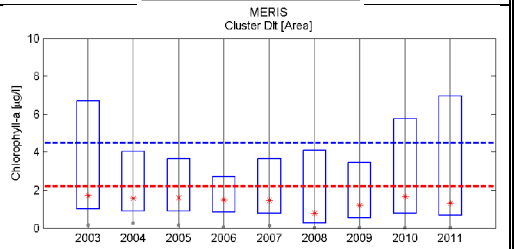
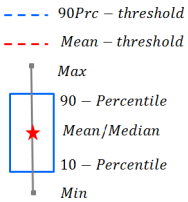
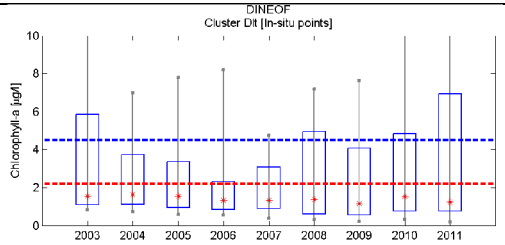
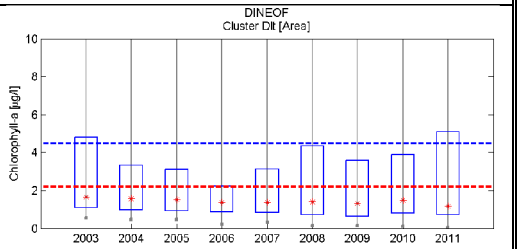
6.2.1 Dogger Bank

	Cluster Default	Cluster Deltares	Area Deltares
MWTL only			
MERIS with gaps			
MERIS without gaps			

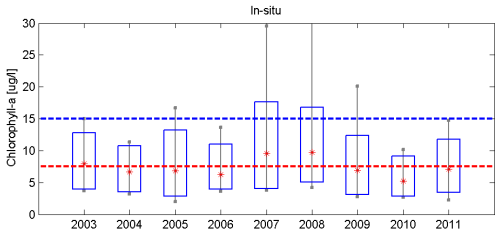
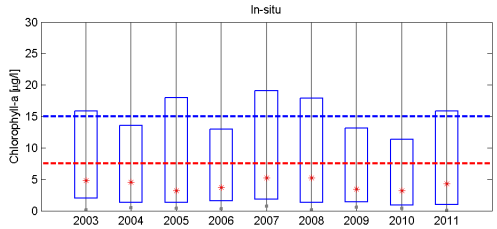
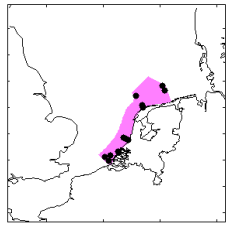
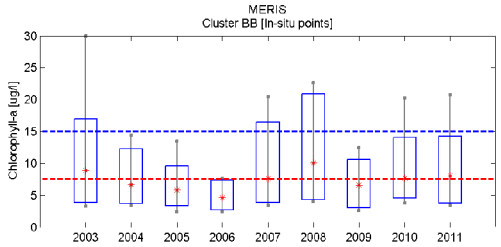
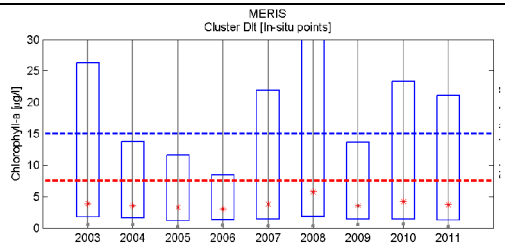
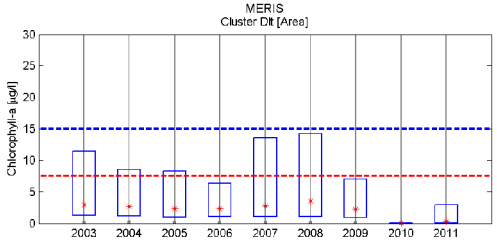
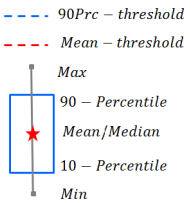
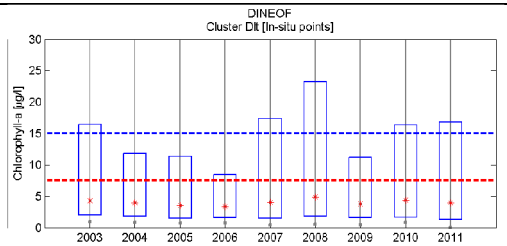
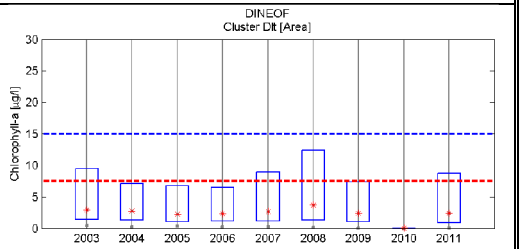
6.2.2

	Cluster Default	Cluster Deltares	Area Deltares
MWTL only			
MERIS with gaps			
MERIS without gaps			

6.2.3 Southern Bight

	Cluster Default	Cluster Deltares	Area Deltares
MWTL only			
MERIS with gaps			
MERIS without gaps			

6.2.4 Coastal Waters

	Cluster Default	Cluster Deltares	Area Deltares
MWTL only			
MERIS with gaps			
MERIS without gaps			

The results above show that the outcome of the typical assessment depends (strongly, occasionally) on the method and (subset of) data used.

The impact of the “default method” vs. the “Deltares method” for both the IS series as well as the MERIS series can be seen by comparing left to the central column for the top 2 rows of the matrices above. On average, the “Deltares method” of taking the geometric mean instead of the arithmetic mean reduces the mean values as may be expected for the skewed chl-a distribution. Likewise, the Deltares method increases the extremes, because the intermediate (smoothing) averaging to monthly mean values is skipped. The 90-percentiles are not always increasing, however, because the “default method’ is determining the percentiles from 7 monthly values, the 90-precentile value will be more sensitive to random fluctuations in the numbers and sometimes may end up higher than the 90-percentile of the distribution of all samples over the region and period. For the Southern Bight, the exceedance of the thresholds (ignoring significance issues for now) is lower in terms of the mean values and about equal or a bit lower in terms of the 90-precentile when applying the Deltares method instead of the default method. For the other potentially problematic area, the Coastal Waters, the mean values do not exceed the thresholds for neither method, but the exceedance of the 90-percentiles increases from about 2 to 3, to 5 instances.

The consistency between MWTL and the MERIS time series can be assessed by comparing the top row with the second row, for left and central column. The interannual fluctuations of the statistics show a high degree of consistency, although occasional deviations are found: sometimes the MERIS data show higher means and/or 90-precentiles, sometimes the MWTL data. For the two potentially problematic areas mentioned above, the number of exceedances of the thresholds is roughly equal for both the default method on one hand and the Deltares method on the other, although the particular years that contribute to these exceedances are different depending on the method and data set. This is a clear indication that neither of the 4 sets in the top left corner of the matrix is really robust. All of these station-wise analyses suffer from the lack of (mostly spatial) coverage and thus are influenced by occasional events observed in particular locations during particular years that may not be representative for the entire region or growing season.

The difference between clustered station-wise and area-wise statistics (central column vs. right column) shows the general tendency of smoothing the mean and 90-percentile values, as might be expected from taking more samples over areas instead of one or just a hand full of locations inside the area. When the full areas are taken into account, the interannual variations become less. This effect is strongest in the Dogger Bank region where only one station is located and in the coastal waters that contain the strongest spatial gradients in Chl-a. The Southern Bight shows an intermediate response to the switch from station-wise to area-wise analysis whereas the results from Oyster Grounds appear most robust for the spatial variations. The potentially problematic area of the Southern Bight now may appear less problematic since the number of exceedances of the 90-percentiles has dropped from 5 to 3 and the exceedance of the mean has also become less likely. For the Coastal Waters the impact is even stronger since all exceedances of the 90-percentiles would disappear from the analysis results.

As an extra step, it is interesting to compare the statistics of the gridded MERIS with gaps to the gapfilled data by DINEOF (central row to bottom row, central and left columns). Yet again, it should be realized that the DINEOF method produces statistically dependent and smoothed samples. It is an interpolation method based on the internal autocorrelation of the samples in space and time and the reconstruction is based on the superposition of the modes that

together described about 90% of the variance in the data. Still, this reconstruction may introduce local extremes in time series as seen in Chapter 5. On average however, the distribution of samples is smoothed as can be seen from the more moderate interannual variations in the mean and 90-percentiles and also in the Southern Bight and Coastal Waters by the further reduction of the exceedances of the 90-percentiles over the thresholds.

As a final step, the original “default” MWTL-based assessment (upper left) is compared to the most extensive MERIS-based area-wise assessment (central row, right column). For all areas the alternative assessment results in less pronounced increases e.g. found in 2006 and 2007 for the Dogger Bank., and in 2003 and 2005-2007 for the Oyster Grounds and in 2007 and 2008 for the Southern Bight and Coastal Waters. The general interannual pattern is largely preserved per region but the extremes are flattened. Finally it is remarked that the interannual pattern of the three areas excluding the Dogger Bank, more closely resemble each other in the region-wise analysis than in the point-wise analysis. They all show a temporal decrease and subsequent increase in chlorophyll-awhen going from 2003 to 2008, followed by another dip in 2009 and 2010 (most pronounced for Oyster Grounds and Coastal Waters). The data from Dogger Bank are showing a more opposite behaviour. The increased resemblance is a consequence of the fact that more spatially adjacent samples have been used so that spatial correlations increase. Similarly, this coherence is also interpreted as showing more the general tendency of the system and less the tendency of accidental samples.

In conclusion, a flattening of the patterns and the decrease in sensitivity for difference in method and region is observed when switching from the default MWTL station-wise assessment to the MERIS-based region-wide assessment. The temporarily more stable and robust statistical properties when switching from station-wise to regional approach within the same dataset is interpreted as an indication of gained accuracy of the results. This presumes that the MERIS data produce a realistic representation of the spatial and temporal variations, a possible systematic error (bias) in the results notwithstanding.

6.3 Significance

Another aspect of accuracy is the error in the estimated mean values and percentile values. As outlined in Chapter 3, section 3.4 a statement on whether or not certain statistical parameters are above or below a certain threshold should always be considered in relation to the confidence put in that statement. In other words: how significant is the obtained difference from the threshold level? What are the chances that a result of e.g. the mean is not the actual state of the system but flawed by a relative outlier?
The standard error in the mean is often applied to determine confidence limits. For a 90-percentile, also an estimate of the spread in its value can be obtained, but this requires a more complicated analysis (e.g. via a bootstrapping method). For the current example we limit ourselves to the assessment in terms of the mean and the estimate or the error in the mean. We show the annual growing season mean values for Dogger Bank and Southern Bight, to illustrate the two most extreme cases in terms of below and at or above threshold level in the mean, but also in terms of number of *in situ* stations available per region. The means and their errors are estimated in the same ways as presented in section 6.2.
Please note that the current estimates of the error in the mean are not accounting for spatial and temporal autocorrelation and thus are optimistic (i.e. low) estimates in cases where spatially correlated samples are applied. At the current stage in this study, the correction for autocorrelation is not feasible, and the results mainly serve as qualitative illustrations of the differences in significance resulting from the various approaches.

Dogger Bank

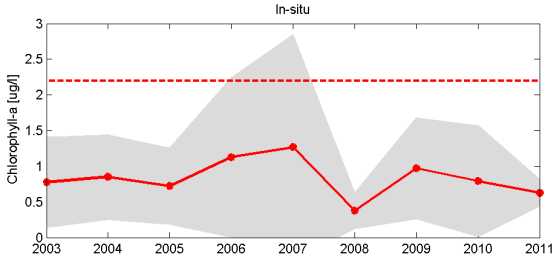
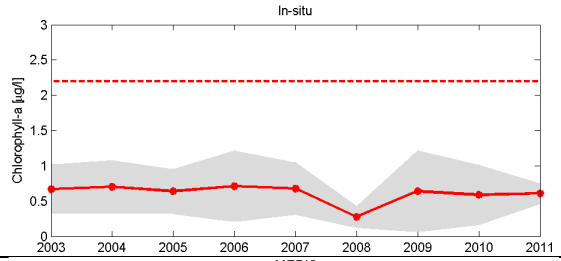
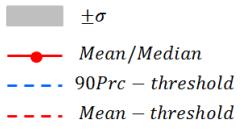
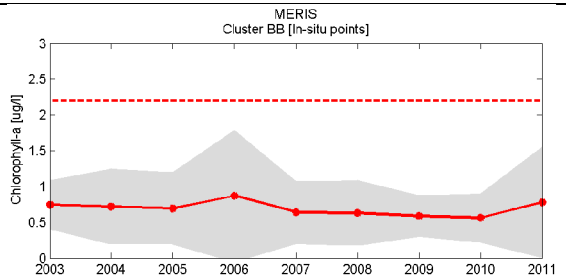
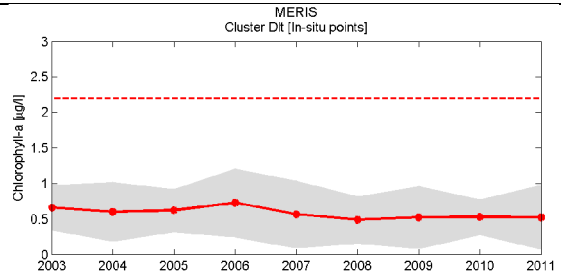
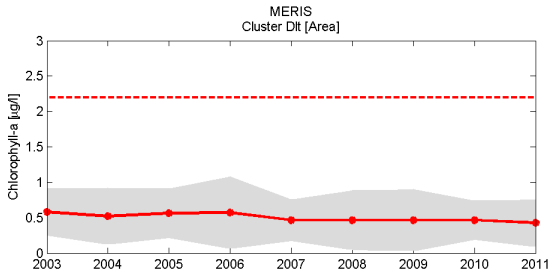
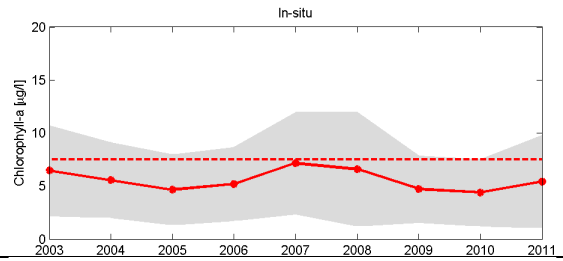
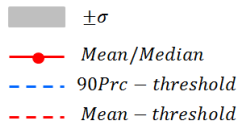
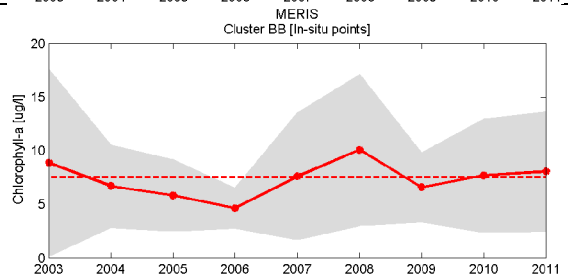
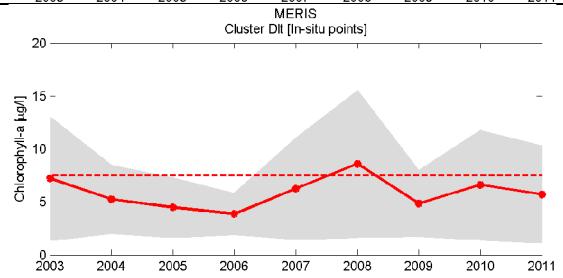
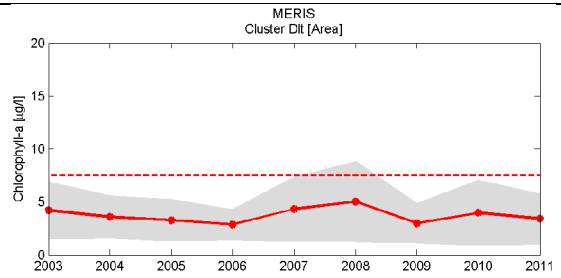
	Cluster Default	Cluster Deltares	Area Deltares
MWTL only			
MERIS with gaps			

Table 6.2 Similar to the figures in section 6.2, but now showing the assessment results for the various methods and data sets but only for the mean and its error. The red dots and line indicate the mean value as also shown by the red markers in section 6.2. The grey shades are the interval of +/- one standard error in the mean value. The dashed lines indicate the mean assessment level. The mean shown is the geometric mean, the standard error is estimated in accordance with the log-normal distribution.

Coastal Waters

	Cluster Default	Cluster Deltares	Area Deltares
MWTL only			
MERIS with gaps			

The examples above clearly illustrate the impact of the method on significance of the results. For the **Dogger Bank** the default assessment results in terms of the mean may appear well-away from the problematic status when only considering the mean values, but when including the spread in the mean values, it can be seen that the threshold is occasionally approached within one standard error from the mean and significance of such distance from threshold appears low. Adopting the revised strategy of estimating the geometric mean and not applying monthly means, already reduces the standard error strongly. This may imply a different interpretation of distance from assessment level. When comparing stationwise results for MWTL vs. MERIS within the Deltares method, it can be seen that the behaviour of both data sets is consistent: both means and standard errors are mostly similar and significantly below the assessment level. Introducing the area-wise assessment instead of station-wise modifies the means and its errors only slightly at Dogger Bank. Hence the interpretation of the assessment is consistent showing a significant non-problem status throughout time for the area.

For the **coastal waters** the 'default' assessment based on *in situ* samples shows mean values just varying around the assessment level, whereas including the standard errors of that method shows that the interannual variations are practically insignificant: the error bounds are up to 100%. Also, the differences between MWTL and MERIS data in this region within the default method are not significantly different. Applying the clustered Deltares method again reduces the means and the errors, but relatively less than for the Dogger Bank. This is thanks to the larger number of MWTL data points (stations and time resolution) used in this region. Still, both the MWTL and station-wise MERIS data are fluctuation close to the assessment level almost always within one standard error. Only when the area covering MERIS data are applied, the distance from the threshold becomes more than one standard error in most years. This implies that the conclusion on the problem status of the Coastal Waters either based on only the *in situ* stations (with various statistical steps) or the area covering data may be totally different, i.e. ranging from the naive '"problematic" to "not significantly non-problematic", to "significantly non-problematic". The risk of drawing the wrong conclusion in this case (Type I error, "false alarm") appears not negligible.

6.4 Conclusions eutrophication assessment

The analysis above has shown that the use of MERIS-based chlorophyll-a data can be fitted into a typical eutrophication assessment with much added benefit. The basic properties of the IS and RS data are largely consistent, hence the outcome of the assessment as well given a certain assessment methodology. Nevertheless, the application of just the time series at the station locations per month and per growing season shows differences in mean and spread, but it should be concluded that these differences fall within the range of uncertainty of these parameters. Both MERIS and IS sample those particular locations with limited time-resolution at different times and the question whether the differences between the means and percentiles are systematic or due to chance cannot be addressed.

The robustness of the estimated mean and percentiles is further explored by modifying the method to estimate the mean and percentiles. By adopting the geometric mean and determining percentiles over the underlying samples instead of over the monthly means, the results become less sensitive to incidental outliers. This leads to the recommendation to more precisely define the statistical approach in the OSPAR Comprehensive Procedure so that all

member states adopt the most robust and representative statistical parameters in the same fashion.

The remote sensing data thus can be used to complement or partially replace the *in situ* data at the station locations. The biggest gain for the eutrophication assessment however is the increased accuracy obtained when using the area covering information instead of only the station-wise subset of the data. Remote sensing by nature provides area covering information and this can be directly used to determine area-wise properties such as growing season mean or 90-percentile chlorophyll-a. Despite the fact that the regions have been chosen such to behave more or less homogeneously in the biogeochemical sense, there are considerable spatial gradients. Gradients are observed in the maps of mean and spreads, but they are also present in even stronger sense in the instantaneous patches of algal blooms that grow, decay and are transported. After all, a remote sensing spatio-temporal sampling scheme like that by MERIS is better capable than the MWTL data of capturing the properties of these variations and hence produces more accurate estimates of the mean and spread per region. When the entire MERIS coverage of an area is used, the statistics clearly show less year-to-year variation and the 90-perecentiles decrease with respect to the station-wise determination. These all are indicators of a more accurate set of estimates of the actual state of the system.

7 Conclusions, discussion, recommendations

7.1 Specific conclusions

As indicated in the introduction the specific questions for the current report have been:

1) What is the quality of the information derived from ocean colour remote sensing?
2) How can ocean colour remote sensing contribute to the information required for national eutrophication monitoring?
3) What are the consequences of adopting remote sensing as part of the eutrophication monitoring in terms of the information and observation strategies?

In brief:

1) The quality of the remote sensing data consists of two parts: the information provided on spatial and temporal variations of the North Sea system and the information provided on numerical values of surface chlorophyll-a. The spatiotemporal information is consistent with the MWTL data as far as the MWLT data resolve the scales of variation, The rempote sensing data however resolve many more scales of variation than the MWTL data, which in this study thus have been taken as accurate as well. The numerical values of e.g. instantaneous local concentrations, but also estimates of temporal and regional means and percentiles are largely consistent with those from MWTL. There are local signs of differences, and in particular in the coastal waters there appear to be biases. Statistically speaking these can hardly be determined as significant, however. Still, knowledge of the natural system and retrieval methods tells that accuracy of remote sensing data will be less in the coastal waters compared to offshore waters on the Dutch continental shelf.

2) Ocean colour remote sensing can contribute in two ways to the national eutrophication monitoring: adding increased resolution and coverage of chlorophyll-a and hence increased accuracy of the determination of characteristics such as the mean and percentiles over regions and growing season. Secondly it may serve as a replacement of part of the standard bottle samples, provided that a matching strategy of validation measurements is implemented.

3) When remote sensing is adopted as part of the eutrophication monitoring a complement observation strategy is required to maintain the basis for validation and calibration of the retrieved data. Such strategy is best composed of a combination of various sensors and platforms such as FerryBoxes and other ship-borne sensors and possibly smartmoorings or other semi-permanent stations. Retrieval validation requires matchups of representative samples. A low frequency single spot MWTL measurement is less suited for this than track cruises with ship sensors. These ship sensors again need to be validated with in situ bottle samples. In the information strategy RWS is recommended to approach the remote sensing data providers more proactively and specify the RWS specific information requirements and priorities, such that data products can be tailored to the needs of RWS.

Below we address these aspects and more in more detail.

7.2 General conclusions

Ocean colour remote sensing is a well-established method to obtain water-surface information from marine ecosystems. Also for optically complex, turbid waters (so-called Case II waters) like the North Sea, retrieval algorithms and data processing have developed to provide a stable basis of near-real time information provision (e.g. for forecasting) but also for monitoring purposes. Water quality parameters such as concentrations of chlorophyll-a and SPM, but also extinction coefficient of visible light (PAR), can be derived sufficiently accurately from multispectral sensors such as MERIS, MODIS, VIIRS and (upcoming) OLCI to be used not only qualitatively (patterns in space and time) but also quantitatively in an integrated marine monitoring strategy.

The resolutions of these sensors in space is typically between 200x200 m^2 and 1x1 km^2, which is sufficient to describe the most dominant spatial features in the open North Sea. At a given geographic location (i.e. within a pixel of a square kilometer), the average temporal resolution of ocean colour sampling is comparable to or slightly higher than that of currently common MWTL ship-borne sampling, where it should be kept in mind that both ship-borne and space-borne sampling suffer from (partially similar) sampling biases such as for weather and wave conditions, time of day, the surface layer. The spatial bias however is different, if not opposite, with more remote sensing samples further offshore, whereas the MWTL sampling is relatively more frequent in the coastal waters.

The biggest advantage of remote sensing is that it provides area coverage combined with a daily revisit over the North Sea. The use of area-covering data gives additional accuracy over the use of single *in situ* stations when determining statistical properties of the system. The autocorrelation within the data is such that the most dominant patterns (over 90% of the total variance) can be reconstructed with about 25 empirical orthogonal functions (EOFs) of which the first 3 modes already capture about 80% of the variance. In other words there are some very strong global patterns of chl-a variation in the system resolved by remote sensing. This is confirmed by the EOF analysis of the lower resolution MWTL data that resulted in a comparable first EOF mode in space and time and a partially comparable second EOF mode. The two MWTL EOFs capture in total 87% of the variance in the monthly aggregated MWTL data. Interestingly, the EOF analysis of the MWTL data did not manage to produce statistically robust modes beyond the second mode. In other words, the MWTL network and sampling scheme is suitable for capturing the large-scale, slowly varying features of the system, but less the specific fluctuations that contain part of the essential information: such as particular shifts in the spring peak in time or magnitude in a certain region and certain year. This specific information may be relevant for the assessment of trends in eutrophication status: often it is not the global behaviour that is making the difference between a good or problematic quality status, but the local and temporal deviations from this behaviour.

The area coverage and daily revisit of the remote sensing data is providing a more accurate way of estimating regional and temporal statistical properties of the ecosystem when compared to point-based (station-wise) assessment. Although an individual pixel may be revisited not more frequently than an individual MWTL station, the surroundings of the pixel location are all sampled in comparable density but partially on different days. As a consequence of spatiotemporal correlation, a more spatially representative estimate of the state of the system can be derived. The use of spatially covering data leads to more stable estimates of the statistics as well: year-to-year variations in the statistics of the region-wise analysis are lower compared to the station-wise assessment. In the region-wise assessment, outliers influence the results of mean and 90-precentiles less. For this particular case, the

consequence is that the number of potentially problematic exceedances of elevated chlorophyll levels has decreased. Hence, a more accurate monitoring and assessment method may pay off in terms of policy.

Despite the expected autocorrelation in both the MWTL and MERIS data values, yet unknown in detail, we chose to provide a tentative estimate of the errors in the mean values for the various assessment methods. Using more samples reduces the error in the mean and thus provides more reliable information. The gain in confidence depends on the variability in the region and the distribution of the samples. As a logical consequence of a reduced error in the mean, it is concluded that the significance of the assessment results improved when applying regionally aggregated remote sensing data compared to station-wise remote sensing or *in situ* data. The degree to which the significance improved or deteriorated also depends on the region.

It can be concluded that a monitoring programme incorporating remote sensing data, adding to or partially replacing the low-frequency ship-borne chlorophyll measurements, can easily achieve the same accuracy as the current MWTL programme for chlorophyll-a. The spatial analysis even shows that more accurate assessments can be obtained.

7.3 Discussion

The current study has been based on historic MERIS data of a certain quality level achieved a couple of years ago (Peters et al. 2008). Nevertheless, the developments in the ocean colour remote sensing have continued ever since, and most importantly, the MERIS data are no longer collected. Hence, the conclusions and recommendations from this analysis should be regarded as based on the presumption that the quality and sampling scheme of these MERIS data is typical for any state-of-the-art ocean colour remote sensing data product for the North Sea for the coming years, independent on which retrieval algorithm or sensor will be used. This also implies that whenever remote sensing data products of chlorophyll-a (and other related variables) are to be acquired from the market, the accuracy demands should be specified for the purposes of national marine water quality monitoring. The general approach to this has been outlined in Roberti & Zeeberg (2007_ and the subsequent RESMON-OK projects (summarized in Blaas et al., 2012 and De Boer et al., 2012).

Another important issue to remark is that we did not carry out a formal power analysis. Partially this was because of the practical reason that the current work had to be limited in time and budget and the current analysis -a first step towards any future power analysis- was as much as what was feasible within the scope of the project. A full power analysis takes more effort: it requires a more precise assessment of the error in the spatially aggregated mean and percentiles, accounting for autocorrelation (see e.g. Eleveld & Van der Woerd, 2006), but first of all, it requires a more precise definition of the desired detectable trends and desired significance of the difference between assessment level and thresholds. This definition is something to be done collectively by the stakeholders involved in the assessment and policy and also by the providers and analysts of data. Fortunately, the issue of confidence (significance level) is now appearing on the agenda of the OSPAR eutrophication working group. Also other issues appear on the OSPAR agenda such as how to deal with spatial heterogeneity of the data and how to apply regional analysis methods (e.g. like we applied here with the MERIS data). Also methods of formal statistical testing on (compounds

of) variables (nutrients, oxygen, chlorophyll) and parameters (means, percentiles) that go into an ecosystem indicator are subject to evaluation.
Hence, OSPAR information requirements are developing. Even if descriptors are currently getting established, the accuracy needs to be quantified to define acceptable and feasible confidence levels. New insights in accuracy of trend analysis and estimation of confidence will lead to more specific quantitative criteria in the near future that can be applied to better test whether a monitoring strategy is meeting the requirements. The current studies thus aims to provide a first semi-quantitative insight in the sensitivity of the outcome of a basic eutrophication assessment only based on chlorophyll-a for different sampling and analysis. Eventually, a full OSPAR assessment has to be based on a set of compound quantities.

As with any indirect observing method, being remote sensing, or automated sensors on moorings or ship, calibration and validation observations are required. Partially replacing the low-frequency chlorophyll observations with automated observations is providing a gain in accuracy building on an already existing infrastructure of sensors, algorithms and expertise, but it also requires a different approach to monitoring. The underlying expertise and algorithms have often been based (at least for part) on the low-frequency national monitoring programmes. If these programmes are revised to contain more sensor observations and less *in situ* measurements, independent calibration and validation measurements become a requirement that has to be fulfilled in a wider information strategy. These measurements need not be collected in the way the current national monitoring data are collected. Preferably, these data are collected in occasional dedicated ground-truthing campaigns that have a much shorter time span but higher spatiotemporal resolution. In such campaigns not only the surface chlorophyll is then collected but also information on the underlying optical properties of the water and its constituents. The wider strategy thus not only addresses the information need for one or more legal requirements of today, but encompasses the underlying information needs. Such approach supports to sustainably maintain and exploit the national marine ecosystem services and physical resources for tomorrow. The information needs of the underlying strategy should even include those of computational models that also require regular (but not yearly) specific data for calibration and validation.

7.4 Strategic recommendations:

Deltares recommends RWS to start incorporating ocean colour remote sensing as a source of information for its monitoring as soon as possible.

A first step is to take the current historic MERIS data into archive and consider them as part of the standard RWS data pool. This means that the management and dissemination of these data needs to be organized on behalf of RWS. At present, Deltares is managing these data and providing limited support with the use and interpretation to RWS users. Transferring this to RWS may require a few documentation and ICT steps to filter unreliable data values from the user and supply a brief description on how to interpret and use the meta data. After all, this should make the current data up to RWS data quality standards and make them more directly usable for users.

Secondly, RWS should start the process of acquiring RS data for the future. This will start with the formulation of the accuracy requirements and organisational aspects of contracting data service providers. It should be kept in mind that the ocean colour community is currently in transition of moving from MERIS and MODIS products to VIIRS and OLCI products (the latter expected to be operational not before late 2014, early 2015). With these new sources of data the existing retrieval algorithms are being and will be updated and tuned. This process can be left partially to the external parties working on the ocean colour market, but RWS

should also be more actively involved and co-determining these developments via a community of practice around ocean colour (or more general) satellite remote sensing. In such community RWS is one of the parties defining an information need, next to suppliers of so called upstream and downstream remote sensing data and services. Deltares recommends to approach the market more in search for service providers and not only in search for a data provider. Service providers can provide data, but also take care of the organisation around it: the interpretation of the information requirements of RWS (or I&M) into choices for platforms, sensors and algorithms available. According to Deltares such a service should encompass also the dissemination in near-real time and delayed (archive) mode of the data products with standardized meta data and file formats on the internet. The foundation of these services should be published, peer-reviewed, well-documented retrieval algorithms, stable ICT infrastructure (24/7 accessibility), data version control and an open data policy.

As soon as new data are coming in, RWS should determine how many of the traditional monthly to 2-weekly ship-borne samples can be replaced at which locations by the remote-sensing based samples. This cannot be fully detailed beforehand, because the characteristics of the data for the future (2015 and later) are not yet known and the information accuracy requirements need to be made more quantitative first. Based on the current MERIS data and still loosely defined accuracy in the light of the system variability, one could consider replacing ship-borne chl-a samples on the Terschelling and Rottum transect by remote sensing for example.

A couple of things should be kept in mind though before removing or revising sampling sites:

(1) Sampling is not only done for chlorophyll-a but also for SPM, turbidity etc. Hence, sampling is not only done for the purpose of eutrophication monitoring. Also, remote sensing cannot observe all variables required for an eutrophication assessment (such as nutrient and oxygen concentrations). The costs and benefits of revising chlorophyll-a sampling needs to be evaluated in terms of the total monitoring programme.
(2) In any revision of a monitoring programme a transition phase is required in which both the ship-borne and remote sensing will run in parallel. Every monitoring cycle has a conservative character to provide a way to maintain consistent trend analysis over multiple years with varying sensors or observation programmes. The current analysis provides an insight in terms of matching series, but this needs to be re-established with any new sensor or algorithm. The task of providing cross validation of different strategies can partially be put to service providers. In this transition the above-mentioned strategy of a dedicated ground truthing campaign in several periods during one year is strongly recommended. In such wider strategy also semi-permanent moorings and ship-borne automated sensors can play an important role.
(3) Any update of information requirements will not be definite overnight. This is a process involving many parties. Hence the recommendation is to take both the new way of sampling and the formulation of the accuracy needs into the monitoring cycle. A community of practice will help to keep all parties involved in the monitoring cycle in touch and up to date.

While making the following steps towards a more efficient monitoring strategy, the costs and benefits of the implementations need to be evaluated as well. Many costs and benefits are difficult to assess beforehand, since desired gains in accuracy and justifiable costs per sample (“more or better information per Euro”) are all too often dependent on how often and in what context the results are used further down into the Dutch national government and its stakeholders. Many costs and benefits are thus hidden from direct view.

7.5 Scientific recommendations

Following from the discussion, a few scientific recommendations are given. Some aspects of the analysis presented here are open for further improvement or extension. This will most likely not affect the main conclusions presented above, but it is recommended to take them into the development steps to follow since they relate to a fine tuning of the analysis in conjunction with the information requirements, a process that should always be part of a monitoring cycle. ("Monitoring the monitoring".)

First, the assessment procedures are subject to change: this is an inevitable aspect of working in a cyclic context such as the (coupled) information cycle and the policy cycle. It is recommended to evaluate the changes in the assessment protocol not only in organisational terms but also statistical terms. In the current project it was only feasible to explore the sensitivity of the outcome of the basic assessment parameters to a different data source and to illustrate issues of confidence. Confidence is however becoming a more explicit part of the assessment in the near-future. A power analysis of a certain strategy can be made either starting from a certain set of target requirements (e.g. on minimal detectable trend with certain confidence, and/or minimal accuracy of assessment of near-threshold values) hence resulting in a required if at all feasible sampling scheme, or starting from a given sampling scheme and then showing what is the maximum power that can be attained.

- Spatial Autocorrelation needs to be determined before a proper power analysis is possible. Now the errors in the spatial means are flattered because spatial and temporal autocorrelation is ignored. Still the vast increase of samples in a region provides a reduction of the error in the spatial mean and a spatially more representative estimate when compared to an estimate based on just a few sites.
- The error in the 90-%-tile can be estimated using a bootstrapping method from the empirical distribution of the samples.

Second, a more extended power analysis might be carried out to in relation to not only the remote sensed variables of chlorophyll but also SPM and extinction coefficient and algal species (such as *Phaeocystis* or cyanobacteria) and primary production rates. These quantities are all based on the same underlying optical reflectance spectra. The questions then are if the individual quantities from the same source are up to the required levels of accuracy and, if not, where the ocean colour community should best put its effort in development. Guidelines as to where to further improve the suitability of remote sensing for eutrophication monitoring depend not only on the accuracy requirements for the individual variables and the underlying algorithms, but also on the total information strategy.
Thirdly, numerical biogeochemical or water quality models can help to validate disparate data that cannot or hardly be matched directly. As an example of the use of numerical models as linking pin, the MoS^2 project is mentioned (Blaas et al., 2008, 2012, 2013; El Serafy et al., 2011, 2013). In the MoS^2 project for the Port of Rotterdam Maasvlakte 2 monitoring, for example, a numerical model of surface water SPM has been used to serve as a reference to various *in situ* and remote sensing data so that estimates systematic and random differences between the data sets could be made. This concept of model-supported monitoring can also be applied for chlorophyll-a for example.

Finally, the apparent development of winter blooms in the northern part of the Dutch North Sea may require attention. Not only from a scientific point of view but also from a monitoring stategy point of view. If these blooms are real and are conveying relevant information on the health of the ecosystem, they may need to be monitored more accurately that currently done by the biased schemes for the ship cruises (few cruises in winter) and also the remote sensing (fewer samples in winter).

Acknowledgements

Deltares would like to acknowledge Rijkswaterstaat Dienst Water Verkeer & Leefomgeving (WVL) for financing this research in the context of the *KPP Programma Efficiente Monitoring 2013*, and in particular Andrea Houben and Albert Mulder as coordinators on behalf of RWS. Niels Kinneging (also RWS WVL) is thanked for supporting the preceding RESMON OK work without which the current analysis would not have been feasible either. CEFAS (Naomi Greenwood and Dave Mills) are gratefully thanked for the smartmooring data. Hanneke Baretta-Bekker is highly appreciated for sharing her information of and knowledge about the OSPAR assessments and her review of the current report. The GHER team of the University of Liege (Aida Alvera-Azcárate, Alexander Barth and Jean-Marie Beckers) is kindly thanked for developing and publishing the DINEOF method and software. Henk van den Boogaard and Remi Laane are thanked for their feedback as internal reviewers.

References

Allewijn R., J.M.M. Kokke, A.G. Dekker, and R.J. Vos (1994), Remote Sensing and water quality in the Netherlands; *Proceedings of the EARSEL symposium* Gotenburg, July 1994.

Alvera-Azcárate, A., A. Barth, D. Sirjacobs, J.-M. Beckers, (2012), Outlier detection in satellite data using spatial coherence *Remote Sensing of Environment*, Volume 119, 16 April 2012, Pages 84-91

Anonymous (2012), Ontwerp Mariene Strategie voor het Nederlandse deel van de Noordzee 2012-2020, Deel I, Report Ministry I&M. May 2012.

Anonymous (2013,. Draft Mariene Strategie Deel II Deel Eutrofiering. Report Ministry I&M 1-10-2013

Arentz, L. (2010), Onderbouwing selectie sample Argoss tbv pilot project.Deltares Report RESMON-OK 3, Deltares report 1202695. Delft

Arentz, L., M. Blaas, (2011), RESMON-OK 3, Ondersteuning Validatie Ocean Color Remote Sensing Deltares report 1202695, Delft

Arentz, L., S. Gaytan Aguilar, M. Blaas (2011), Pre-analysis of MERIS Remote Sensing water quality products (Part of RESMON-OK 3), Deltares report 1202695, Delft

ARGOSS (2007), Does satellite observation of chlorophyll-a and sediment complement Rijkswaterstaat monitoring of the North Sea? ARGOSS contribution TORSMON - WP2 May 24, 2007

Baretta-Bekker, J.G. (2013), Guidelines to carry out the OSPAR Comprehensive Procedure for the Dutch OSPAR areas. Technical Report BarettaBekker Marine Ecology

Baretta-Bekker, J.G., and T.C. Prins (2013), Assessments of phytoplankton in the Netherlands and neighbouring countries according to OSPAR and WFD. Report BarettaBekker- mariene ecologie & Deltares Version: 20130815

Beckers J.-M., A. Barth and A. Alvera-Azcarate (2006), DINEOF reconstruction of clouded images including error maps. Application to the Sea Surface Temperature around Corsican Island. *Ocean Science*, 2(2):183-199, 2006. (doi:10.5194/osd-3-735-2006)

Beckers, J.-M., and M. Rixen (2003), EOF calculations and data filling from incomplete oceanographic data sets. *Journal of Atmospheric and Oceanic Technology, 20(12):1839--1856, 2003. (10.1175/1520-0426(2003)020%3C1839:ECADFF%3E2.0.CO;2)*

Blaas et al. (2013) MoS^2: Model-supported Monitoring of SPM in the Dutch coastal zone. Summary MoS2 phase II. Deltares 2011-2013. Deltares report 1204561. March 2013.

Blaas M., Cronin K., G.Y. El Serafy, Y.F. Friocourt, I.D.T.F. Garcia Triana, S. Gaytan-Aguilar, G.H. Keetels, (2012), MoS^2: Model-supported Monitoring of SPM in the Dutch coastal zone: Model setup, data assimilation and skill assessment. Deltares Report 1002611, Deltares Delft.

Blaas, M, L. Arentz, G.J. de Boer (2012), Aardobservatie in optimalisatie monitoringsstrategie: combinatie ocean colour remote sensing en *in situ* waarnemingen. Report 1206180.000. December 2012, Deltares Delft.

Blaas, M. (2008), Concept plan van aanpak ondersteuning RIJKSWATERSTAAT t.a.v. Remote Sensing (RESMON-OK), Deltares Memo Z4750_191208_MB

Blaas, M., X. Desmit, and F. J. Los (2007), OSPAR eutrophication Modelling Workshop 2007. WL | Delft Hydraulics report Z4351, Delft, The Netherlands.

Blaas, M., H.F.P. van den Boogaard, (2006). Statistical methods to assess the impact of MV2 on SPM along the Dutch coast. Results of WP3 & 5 (Baseline Silt PMR, RIKZ 1661), WL | Delft Hydraulics Report Z4046. Delft 2006.

Blaas, M., J. Dijkstra, S. Gaytan Aguilar, J. Sumihar, M.T. Villars (2013), Trend analysis using *in situ* and remote sensing data: MoS2-III: model-supported monitoring of SPM in the Southern North Sea. Deltares report 1208242, July 2013:

Blaas, M., M.A. Eleveld, G.Y. El Serafy, H.J. van der Woerd, G.J. de Boer, T. van Kessel (2008), TnulTSM, Integration of Remote Sensing *and Modelling of Total Suspended Matter in the Dutch coastal zone.* Deltares report with VU IVM. (2008).

Blauw, A.N., E. Beninca, R.W.P.M. Laane, N. Greenwood, J. Huisman (2012), Dancing with the Tides: Fluctuations of Coastal Phytoplankton Orchestrated by Different Oscillatory Modes of the Tidal Cycle. *PLoS ONE* 7(11): e49319. doi:10.1371/journal.pone.0049319

Chawira, M. (2012). Monitoring blue-green algae in the IJsselmeer using remote sensing and in-situ measurements. *ITC Thesis TU Twente*, Feb, 2012.

De Boer, G.J., M. Blaas, L. Arentz (2012), Analysis of RS validation methods. Part of RESMON-OK4 (1204620).

Doerffer, R. and H. Schiller (2007), The MERIS case2 water algorithm. Internat. *J. Remote Sensing*, 28: 517-535

Duin, R.N.M., S.J. Dury, J.R. Roberti (2006), Towards operational use of MERIS and SeaWiFS data for water quality monitoring: challenges for the end-user. *Proceedings 26th EARSeL Symposium*, Warsaw, Poland, May 29 - June 2, 2006.

Duin, R.N.M., H.R. Roberti, R.V.M. Pasterkamp and S. Dury (2005), Chlorofyl en Zwevend stof waarnemingen van de Noordzee vanaf satellieten en schepen. *RIZK Rapport*, 2005.

Dury, S., H. Hakvoort and R. Jordans (2004), Water quality monitoring in the North Sea by optical remote sensing: validation of satellite data. *AGI rapport* 0406GAR001, 2004.

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

El Serafy, G.Y., S. Gayan Aguilar, K. Cronin, J. van Beek, J. Dijkstra, D. Stuparu, M. Blaas (2013), SPM Model optimisation and forecast analysis: MoS2-III: model-supported monitoring of SPM in the Southern North Sea. *Deltares report* 1208242

Eleveld, M. A., R. Pasterkamp, H. J. van der Woerd and J.D. Pietrzak (2008), Remotely sensed seasonality in the spatial distribution of sea-surface suspended particulate matter in the southern North Sea. *Estuarine, Coastal and Shelf Science* 80, 103–113 (2008).

Eleveld, M.A., H.J. van der Woerd (2006), Patterns in water quality products of the North Sea: variogram analyses of single and compound SeaWiFS CHL & SPM grids. *In: Kerle, A.K. Skidmore (eds). Remote Sensing: From Pixels To Processes. ISPRS Proceedings Technical Commission VII.* Publisher: ITC, The Netherlands.

Ferreira, J.G. , J.H. Andersen, A. Borja, S.B. Bricker, J. Camp, M. Cardoso da Silva, E. Garcés, A.S. Heiskanen, C. Humborg, L. Ignatiades, C. Lancelot, A. Menesguen, P. Tett, N. Hoepffner & U. Claussen (2010), Marine Strategy Framework Directive Task Group 5: eutrophication, *JRC Report*, APRIL 2010, Ispra Italy.

Fettweis M.P., B. Nechad (2011), Evaluation of *in situ* and remote sensing sampling methods for SPM concentrations, Belgian continental shelf (southern North Sea). *Ocean Dynamics* 61(2-3), 157-171. doi: 10.1007/s10236-010-0310-6

Garcia, I.D.T.F. and M. Blaas (2013), Rijkswaterstaat sensor data on the North Sea: CTD, FerryBox and ScanFish: Assessment of their potential for SPM model validation of the North Sea. Deltares Report MoS^2 (1204561.003). Deltares, Delft

Gaytan, S., M. Blaas (2013) MoS2-II: Analysis of MERIS Remote Sensing data: Preparing assimilation of SPM data into a North Sea transport model. Deltares report 1204561, March 2013.

Hommersom, A., M. R. Wernand, S.W.M. Peters and J. Boer (2010) A review on substances and processes relevant for optical remote sensing of extremely turbid marine areas, with a focus on the Wadden Sea. *Helgoland Marine Research* 64, 75–92.

Houseago-Stokes, R. (2000), Using optimal interpolation and EOF analysis on North Atlantic satellite data. *International WOCE Newsletter*, No.38, 26-28.

ISO 5725-1 (1994): Accuracy (trueness and precision) of measurement methods and results - Part 1: General principles and definitions, pp.1 (1994)

Kröger, S., E. R. Parker, J. D. Metcalfe, N. Greenwood, R. M. Forster, D. B. Sivyer and D. J. Pearce (2009), Sensors for observing ecosystem status. *Ocean Sci.,* 5, 523–535, 2009

Laane, R.W.P.M. (2013a) Roadmap for innovative monitoring "the optimal monitoring network does not exist" (OPTIMON), Deltares Report 1206240 Delft, 2013 (Apr.), 34 p

Laane, R.W.P.M (2013b), Innovative monitoring of the water quality in the Dutch coastal part of the North Sea: background document. Deltares report 1204469, Delft, 2013 (Apr.) 105 p.

Laane, R.W.P.M. A.N. Blauw, H.F.P. van den Boogaard, N.Greenwood (in prep.) Green Monitoring Strategy for Chlorophyll in coastal and marine waters. *Manuscript in prep for publication*

Lenhart, H.J., D.K. Mills, H. Baretta-Bekker, S.M. van Leeuwen, J. van der Molen, J.W. Baretta, M. Blaas, X. Desmit, W. Kühn, G. Lacroix, F.J. Los, A. Ménesguen, R. Neves, R. Proctor, P. Ruardij, M.D. Skogen, A. Vanhoutte-Brunier, M.T. Villars, S.L. Wakelin, (2010), Predicting the consequences of nutrient reduction on the eutrophication status of the North Sea. *J. Mar. Syst.* 81, 148–170.

Los, F. J. , M. Blaas (2010), Complexity, accuracy and practical applicability of different biogeochemical model versions. *Journal of Marine Systems* 81, 44–74 (2010).

Matthews, M. W., S. Bernard and L. Robertson (2012), An algorithm for detecting trophic status (chlorophyll-a), cyanobacterial-dominance, surface scums and floating vegetation in inland and coastal waters. *Remote Sensing of Environment* 124, 637–652.

Mills, D. K., N. Greenwood, S. Kröger, M. Devlin, D.B. Sivyer, D. Pearce, S.J. Malcolm (2005), New Approaches to Improve the Detection of eutrophication in UK Coastal Waters. *Environmental research, engineering and management* 2, 36–42 (2005).

Mills, D. K., P. B. Tett and G. Novarino (1994), The spring bloom in the south western North Sea in 1989. *Netherlands Journal of Sea Research* 33, 65–80 (1994).

Müller, W.G. (2007) Collecting spatial data: optimum design of experiments for random fields. Springer Berlin, New York, 2007.

Nair, A., S. Sathyendranath, T. Platt, J. Morales, V. Stuart, M.-H. Forget, E. Devred, H. Bouman (2008), Remote sensing of phytoplankton functional types. *Remote Sensing of Environment* 112, 3366–3375 (2008).

Nechad, B., A. Alvera-Azcárate, K. Ruddick, and N. Greenwood, (2011), Reconstruction of MODIS total suspended matter time series maps by DINEOF and validation with autonomous platform data. *Ocean Dynamics*, 61:1205-1214, 2011. (DOI: 10.1007/s10236-011-0425-4).

Nechad, B., K.G. Ruddick, and Y. Park (2010), Calibration and validation of a generic multisensor algorithm for mapping of total suspended matter in turbid waters. *Remote Sensing of Environment* 114, 854–866

Neukermans, G., K. G. Ruddick and N. Greenwood (2012). Diurnal variability of turbidity and light attenuation in the southern North Sea from the SEVIRI geostationary sensor. *Remote Sensing of Environment* 124, 564–580 (2012).

Noordhuis, R. (2012). Stappenplan Informatiestrategie. Deltares report KPP 2012, Delft, The Netherlands, 2012.

OSPAR (2005), Revised Common Procedure for the Identification of the eutrophication Status of the OSPAR Maritime Area. EUC(2) 05/2/Info.1-E, Dec. 2005, (reference number: 2005-3)

OSPAR (2010), Quality Status Report 2010. OSPAR Commission. London. 176 pp.

OSPAR (2013), Common Procedure for the Identification of the eutrophication Status of the OSPAR Maritime Area (Reference number: 2013-8)

Osté, L.A., L.M. Dionisio Pires, M. Schroevers, G. Rutten, R.S. Westerhoff (2012). Verkenning ontwikkelingen natte informatievoorziening, Deltares rapport 1206432-001, Delft, The Netherlands, 2012.

Otto, L. et al. (1990), Review of the Physical Oceanography of the North Sea. *Netherlands Journal of Sea Research* 26, 161–238 (1990).

Peters, S.W.M. (2006) MERIS reflectance and Algal-2 validation at the North Sea, *Proceedings ESA MAVT,* Frascati, Rome 2006

Peters, S.W.M., H.J. van der Woerd, M.A. Eleveld (2008). Ovatie-2 Final report. *IVM report*, 99 pp + appendices.

Pietrzak, J. D., G.J. de Boer, M.A. Eleveld (2011), Mechanisms controlling the intra-annual mesoscale variability of SST and SPM in the southern North Sea. Continental Shelf Research 31, 594–610 2011.

Prins, T.C., J.G. Baretta-Bekker (2013), Assessments of phytoplankton in the Netherlands and neighbouring countries according to OSPAR and WFD. *Report Baretta-Bekker Marine Ecology & Deltares*. Version 20130815

Roberti, J.R., J.J. Zeeberg, (2007), Does satellite observation of chlorophyll-a and suspended sediment complement Rijkswaterstaat monitoring of the North Sea? *Report AGI 'ToRSMoN'*, AGI-2007-GPMP-017, 50 pp.

Ruddick K., G. Lacroix, Y. Park, V. Rousseau, V. De Cauwer, S. Sterckx (2007), Overview of Ocean Colour: theoretical background, sensors and applicability for the detection and monitoring of harmful algae blooms (capabilities and limitations). *In: Babin M, Roesler C, Cullen J (eds) Real-time coastal observ- ing systems for ecosystem dynamics and harmful algal blooms.* UNESCO Monographs on Oceanographic Methodology Series. UNESCO publishing, Paris

Rutten T., H. Roberti, M. Ohm (2006) Algenplagen beter voorspelbaar, *Zoutkrant RWS* 2:5 2006

Sirjacobs, D., A. Alvera-Azcárate, A. Barth, G. Lacroix, Y. Park, B. Nechad, K. Ruddick, J.-M. Beckers, (2011). Cloud filling of ocean color and sea surface temperature remote sensing products over the Southern North Sea by the Data Interpolating Empirical Orthogonal Functions methodology. *Journal of Sea Research*, 65(1):114-130. 2011. (10.1016/j.seares.2010.08.002).

Sørensen K, G.G. Severinsen, G. Aertebjerg, V. Barale, C. Schiller (2002), Remote sensing's contribution to evaluating eutrophication in marine and coastal wa- ters. *European Environment Agency*

Suijlen, J.M. and R.N.M. Duin (2002), Atlas of near-surface Total Suspended Matter concentrations in the Dutch coastal zone of the North Sea. *Rijkswaterstaat report* RIKZ/2002.059.

Tilstone, G. H., G. F. Moore, K. Sørensen, R. Doerffer, R. Røttgers, K.G. Ruddick, R. Pasterkamp, and P.V. Jørgensen (2004), Protocols for the validation of MERIS products in Case 2 waters. *Proceedings of the ENVISAT validation workshop, European Space Agency*. 6(10):875-889.

Tilstone, G., S.W.M. Peters, H.J. Van der Woerd, M.A. Eleveld, K. Ruddick, W. Schoenfeld, H. Krasemann, V. Martinez-Vicente, D. Blondeau-Patissier, R. Røttgers, K. Sørensen, P.V. Jørgensen, J.D. Shutler (2012). Variability in specific-absorption properties and their use in a semi-analytical Ocean Colour algorithm for MERIS in North Sea and Western English Channel Coastal Waters, *Remote Sensing of Environment* 118, 320–338. doi: 10.1016/j.rse.2011.11.019

Uhlig, S., K. Simon, S. Antoni, K. Kunath (2007), Incorporation of Remote Sensing Data in Monitoring Programmes, *Technical Report QuoData Dresden*, December 2007.

Van Bracht, M.J. (2001). Made to Measure. Information requirements and Groundwater Level Monitoring Networks. *PhD thesis, Free University of Amsterdam*, The Netherlands.

Van der Woerd, H.J., R. Pasterkamp (2008), HYDROPT: A fast and flexible method to retrieve chlorophyll-a from multi-spectral satellite observation of optical-complex coastal waters. *Rem. Sens. Env*. 112, 1795-1807

Von Storch, H., and F.W. Zwiers, (1999), Statistical Analysis in Climate Research, Cambridge University Press, ISBN 0 521 45071 3.

Westerhoff, R., M. Kleuskens, T. van de Ven, M. Blaas (2010), RESMON-OK 2 - Ondersteuning Validatie Ocean Color Remote Sensing. Deltares report.

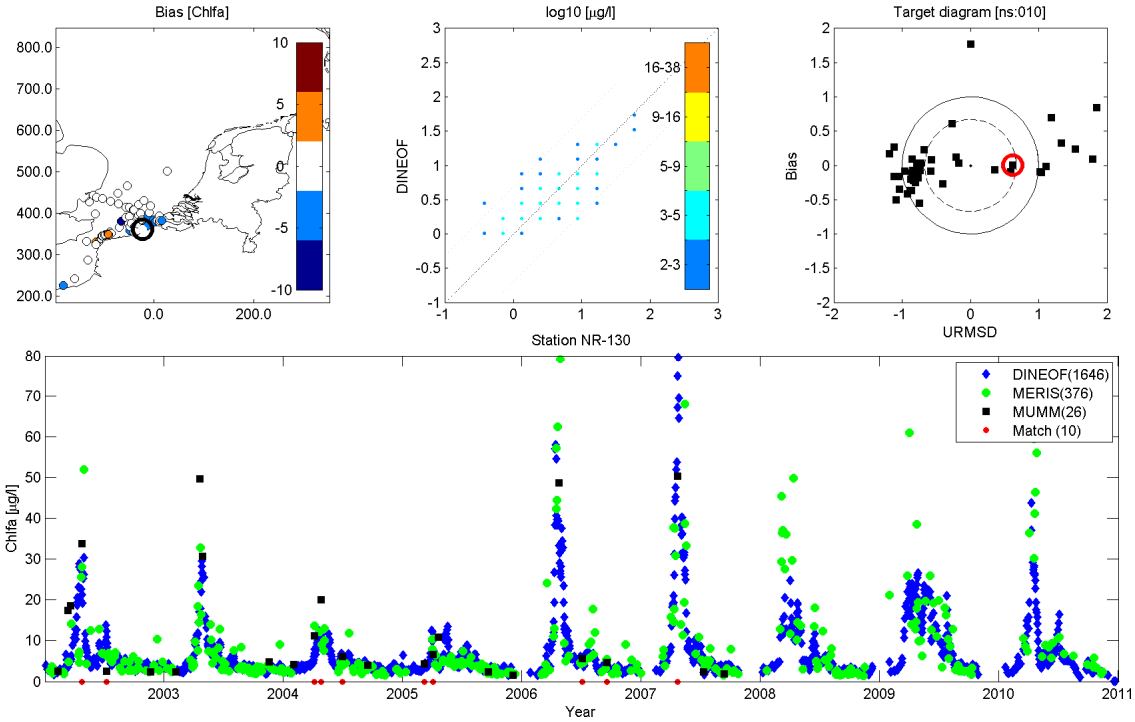
Zeeberg, J.J. & H.R. Roberti (2007), Productspecificaties Aardobservatie Waterkwaliteit, AGI concept-rapport, December 2007.

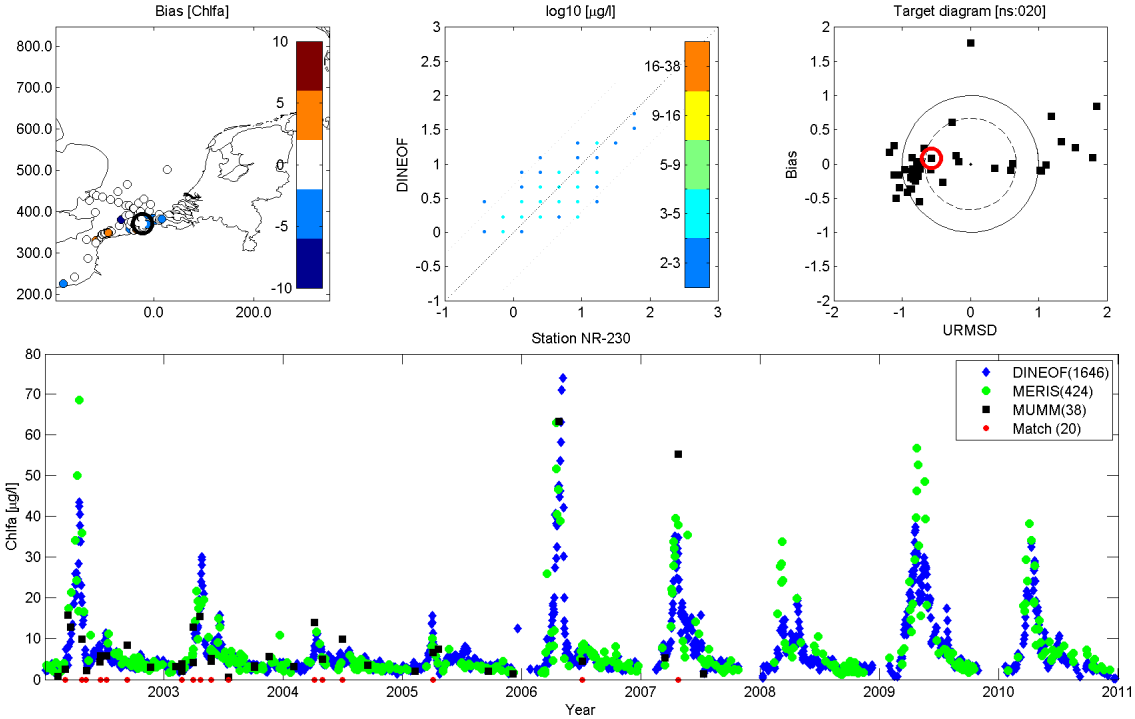
Zeeberg, J.J. (2007), Kwaliteitstoetsingsplan Aardobservatie Waterkwaliteit, AGI concept rapport, November 2007.

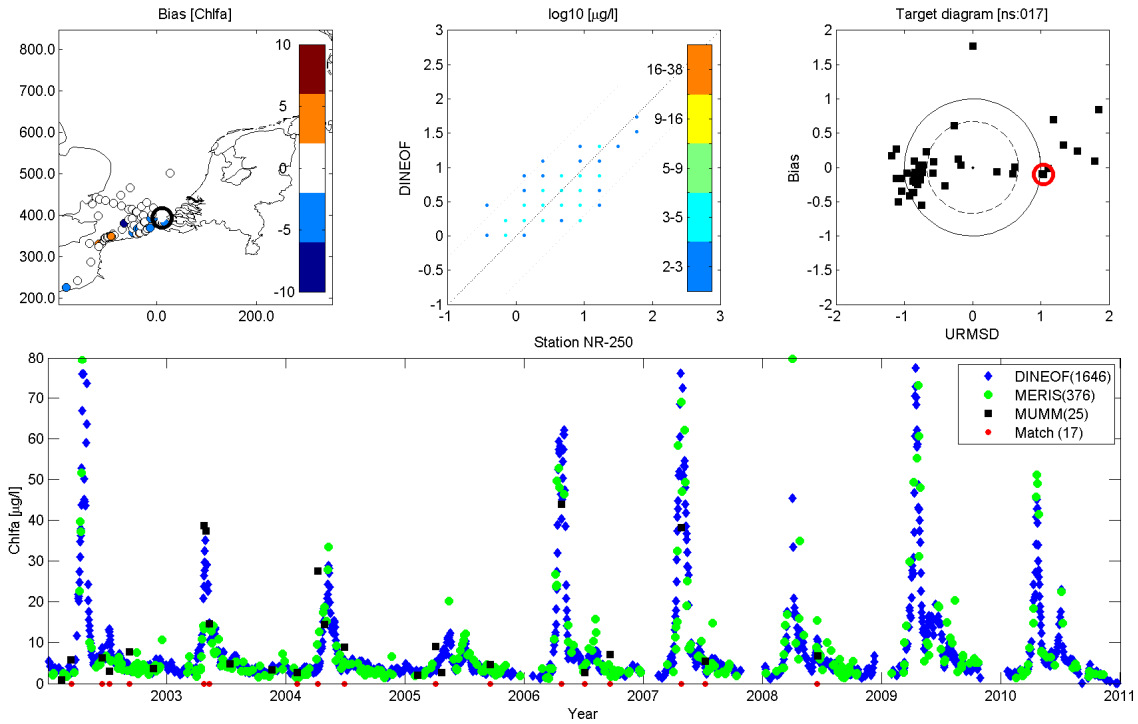
A Appendix with time series

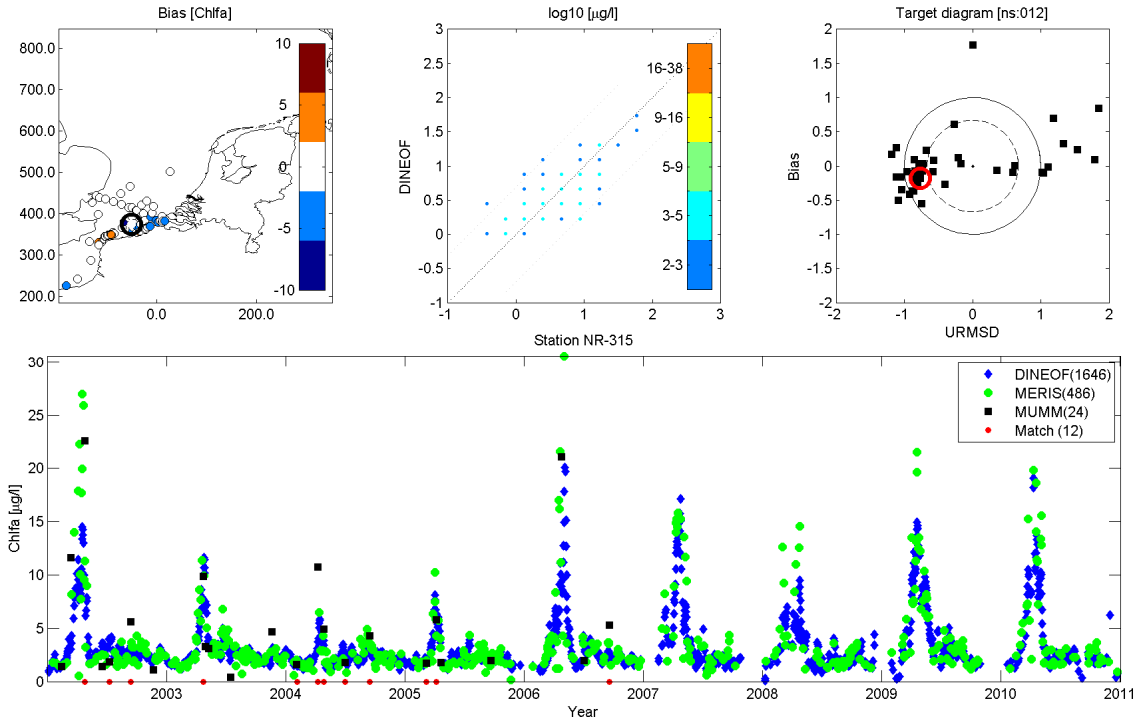
This appendix shows the timeseries of the *in situ* (IS), gridded MERIS, and DINEOF gap-filled gridded MERIS data at all available IS stations (alphabetically). All stations with less than 10 matches between gapfilled MERIS and IS have been left out (i.e. about 10 MUMM stations and the station Schouwen 10) because less than 10 samples provide unreliable statistics in the matchup. This information is provided as background to the selected time series shown in Chapter 5 (which are repeated here).

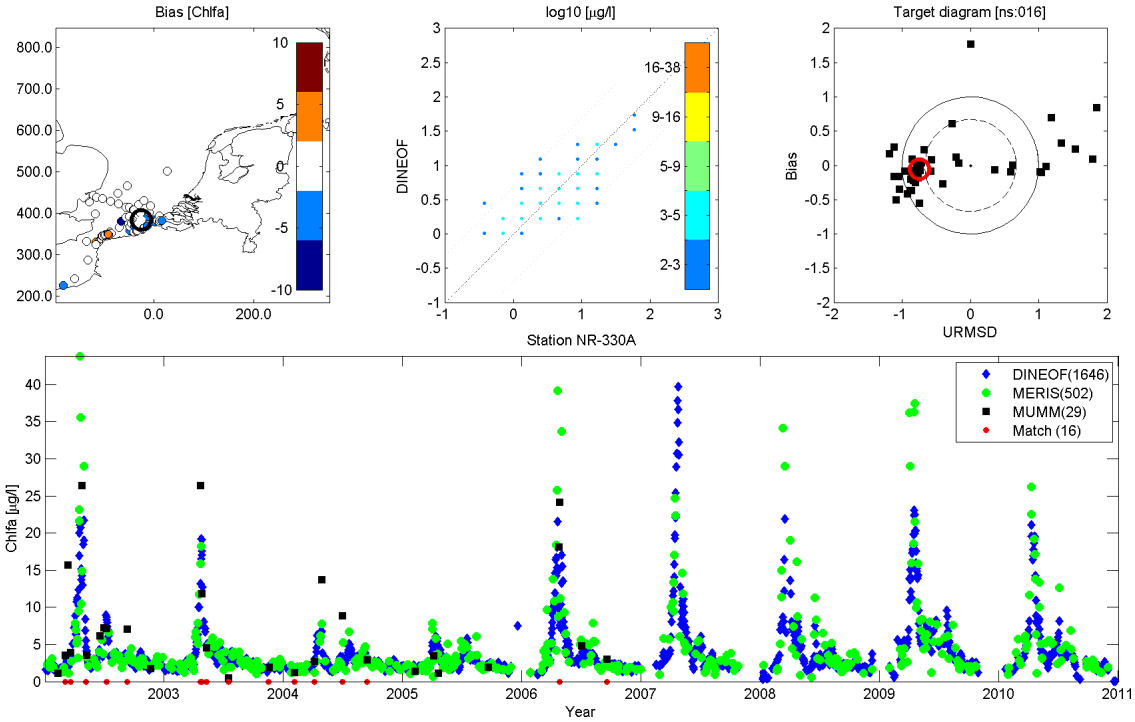
The upper panels show a map with the normalized bias (left) the binned scatter diagram of DINEOF vs. in situ data in a matching window of 24 hours (centre) and a target diagram (right) indicating the goodness of fit of the matching series. The target diagram shows the combined normalized bias on the vertical axis and the unbiased normalized root-mean-square difference (URMSD) on the horizontal axis for each station as represented by a dot. For more explanation on the target diagrams, see e.g. Los & Blaas (2010). The red circle indicates the dot corresponding to the station for which the timeseries is shown. The bias maps show spatial structure of the systematic error: if significant, this may point to ertain spatial bias in the calibration of the retrieval. Unfortunately, the maps of the bias based only on the direct matches are hardly significant. Instead, one can examine the difference in the mean and standard error for both full series, of which the full maps have been shown in chapter 4, or one can take external information that represents the main characteristics of the dynamics such as well-validated numerical model results to serve as a reference. The bias map is also shown as a reference for the geographic location of the station shown in the time series. This station is indicated by a bold black circle round the marker on the map.

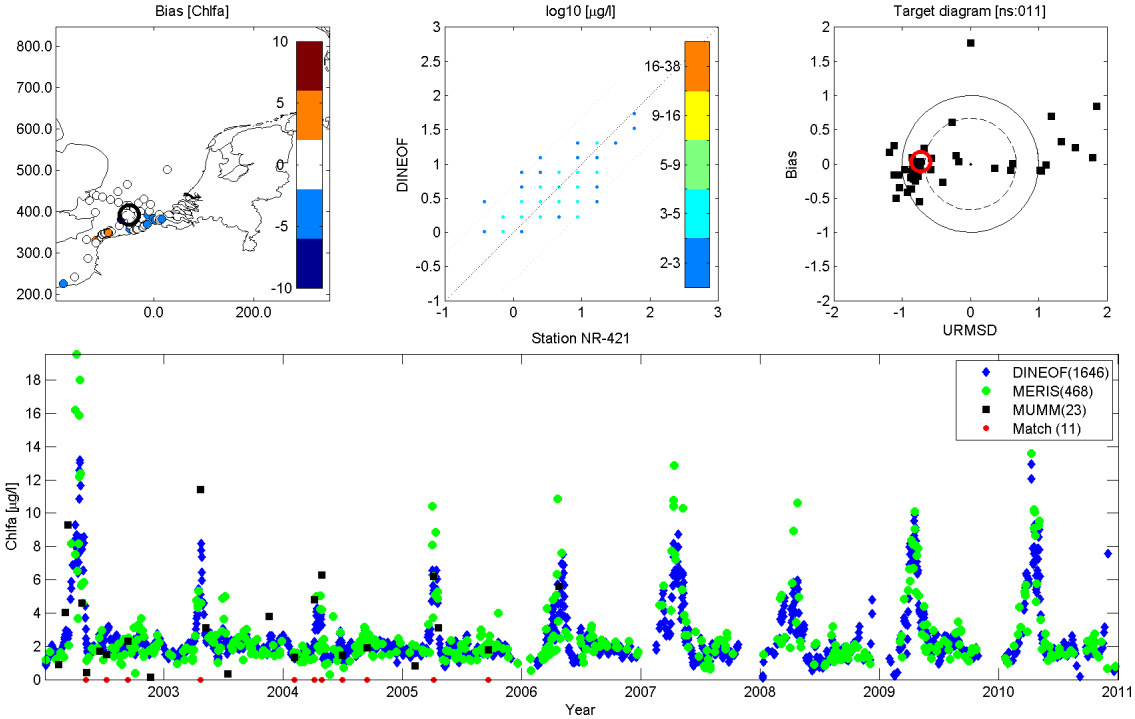

Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:010]
DINEOF
Bias
URMSD
16-38
9-16
5-9
3-5
2-3
Station NR-130
DINEOF(1646)
MERIS(376)
MUMM(26)
Match (10)
Chlfa [µg/l]
Year

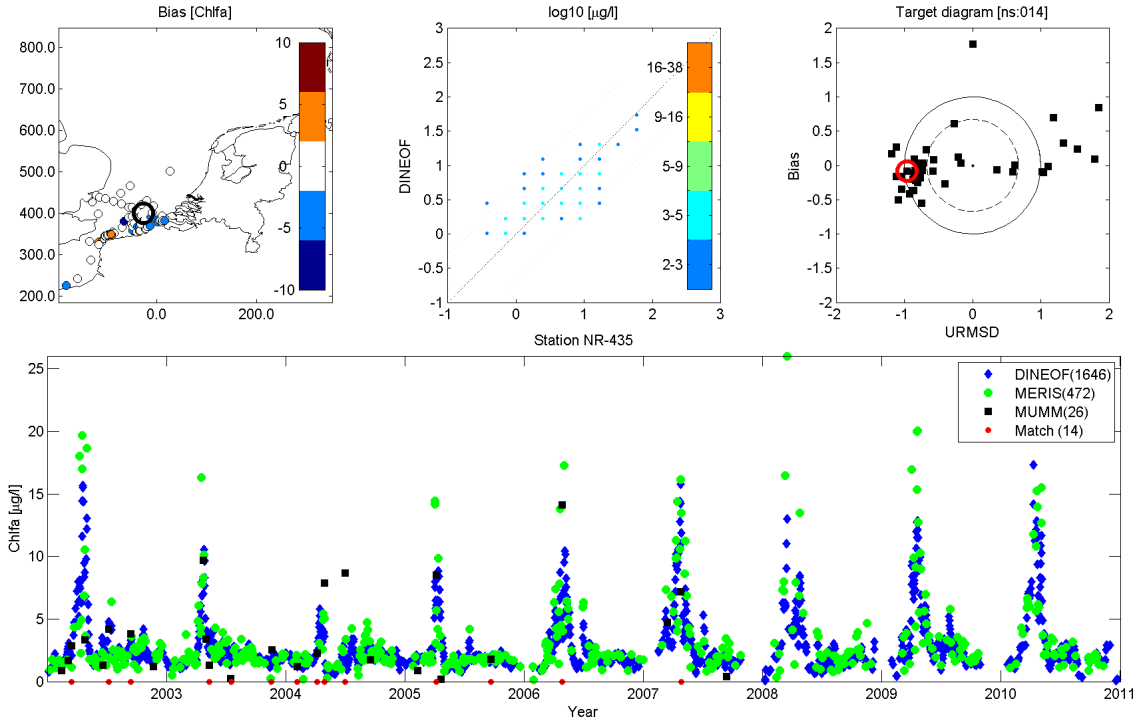

Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:020]
DINEOF
Bias
URMSD
16-38
9-16
5-9
3-5
2-3
Station NR-230
DINEOF(1646)
MERIS(424)
MUMM(38)
Match (20)
Chlfa [µg/l]
Year

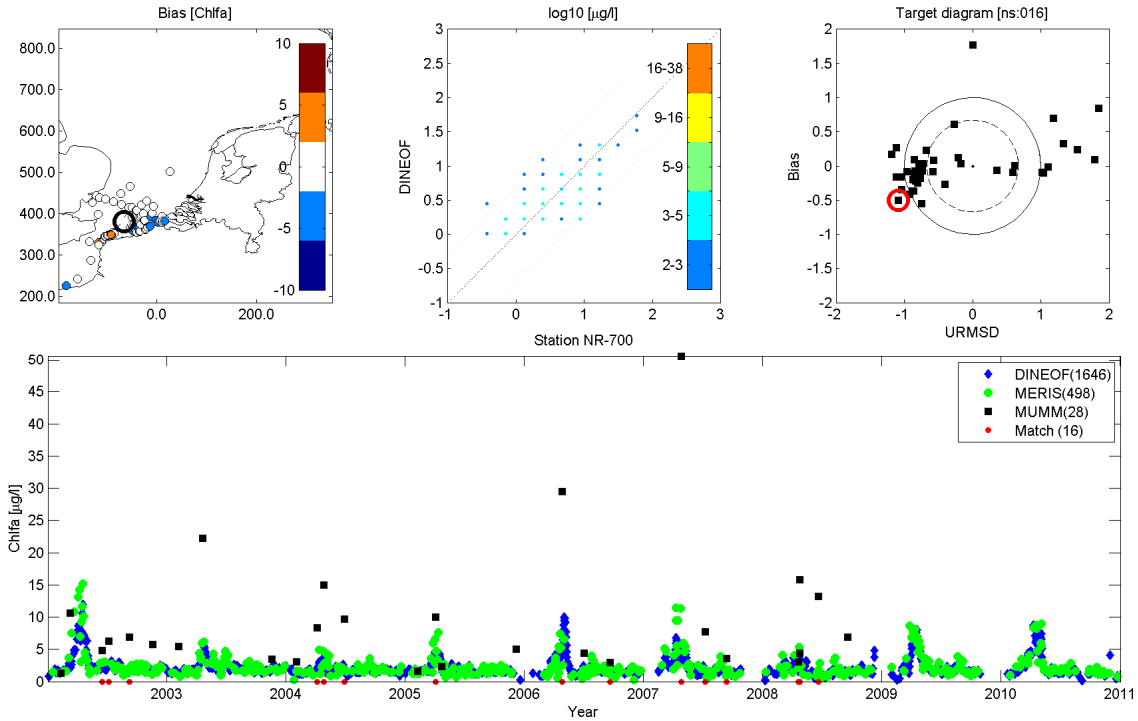

Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:017]
DINEOF
Bias
URMSD
16-38
9-16
5-9
3-5
2-3
Station NR-250
Chlfa [µg/l]
Year
DINEOF(1646)
MERIS(376)
MUMM(25)
Match (17)

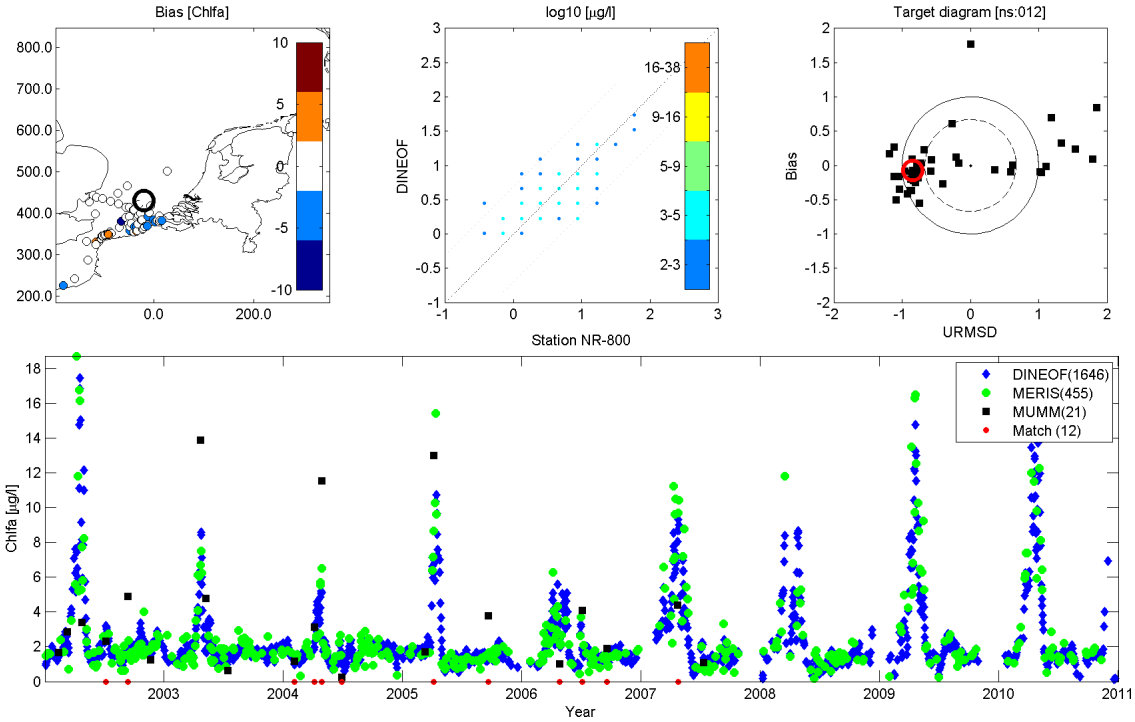

Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:012]
DINEOF
Bias
URMSD
16-38
9-16
5-9
3-5
2-3
Station NR-315
Chlfa [µg/l]
Year
DINEOF(1646)
MERIS(486)
MUMM(24)
Match (12)

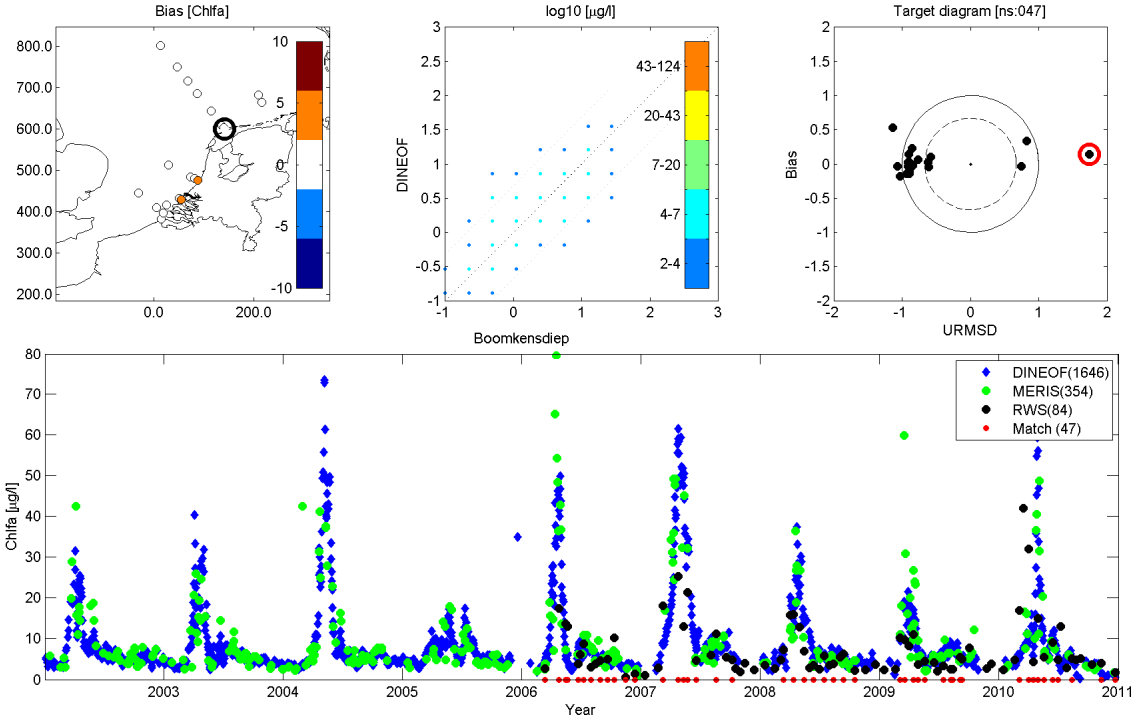

Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:016]
DINEOF
Bias
URMSD
16-38
9-16
5-9
3-5
2-3
Station NR-330A
DINEOF(1646)
MERIS(502)
MUMM(29)
Match (16)
Chlfa [µg/l]
Year

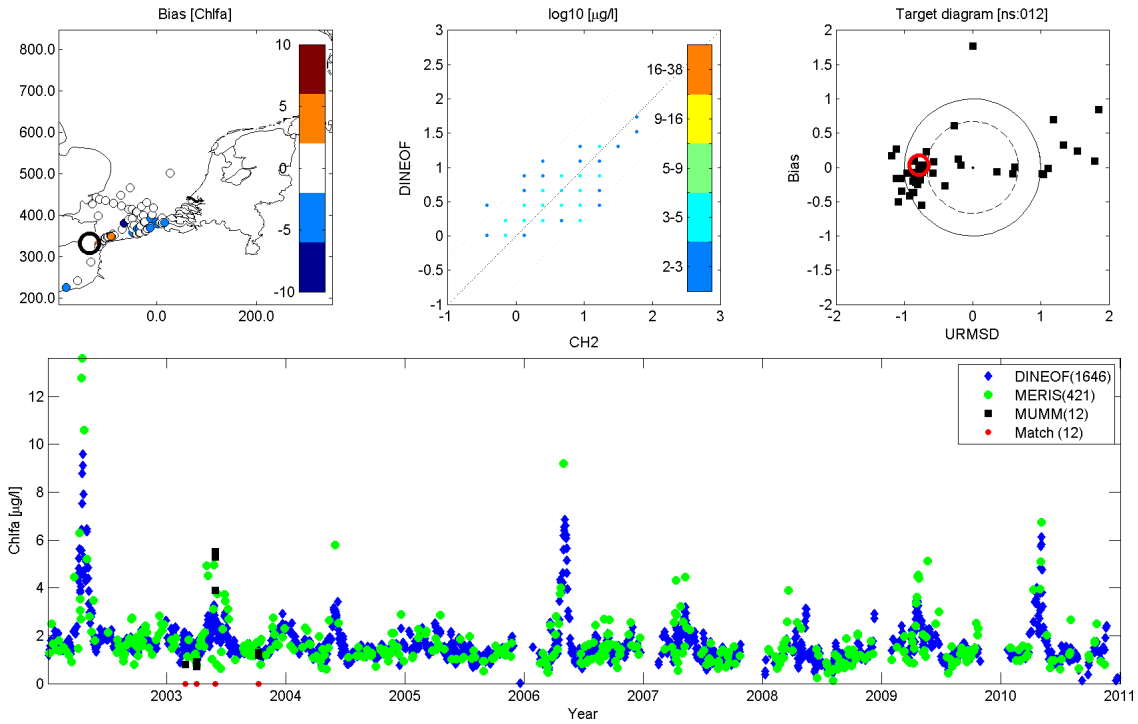

Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:011]
DINEOF
Bias
URMSD
16-38
9-16
5-9
3-5
2-3
Station NR-421
DINEOF(1646)
MERIS(468)
MUMM(23)
Match (11)
Chlfa [µg/l]
Year

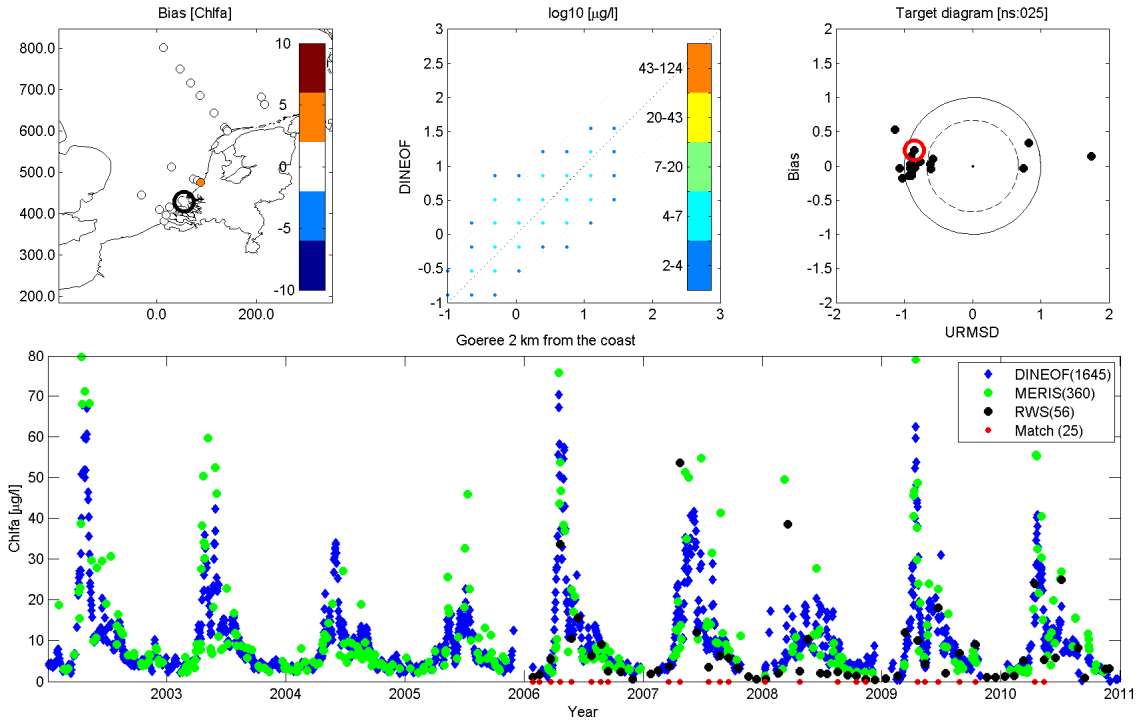

Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:014]
DINEOF
Bias
URMSD
16-38
9-16
5-9
3-5
2-3
Station NR-435
DINEOF(1646)
MERIS(472)
MUMM(26)
Match (14)
Chlfa [µg/l]
Year

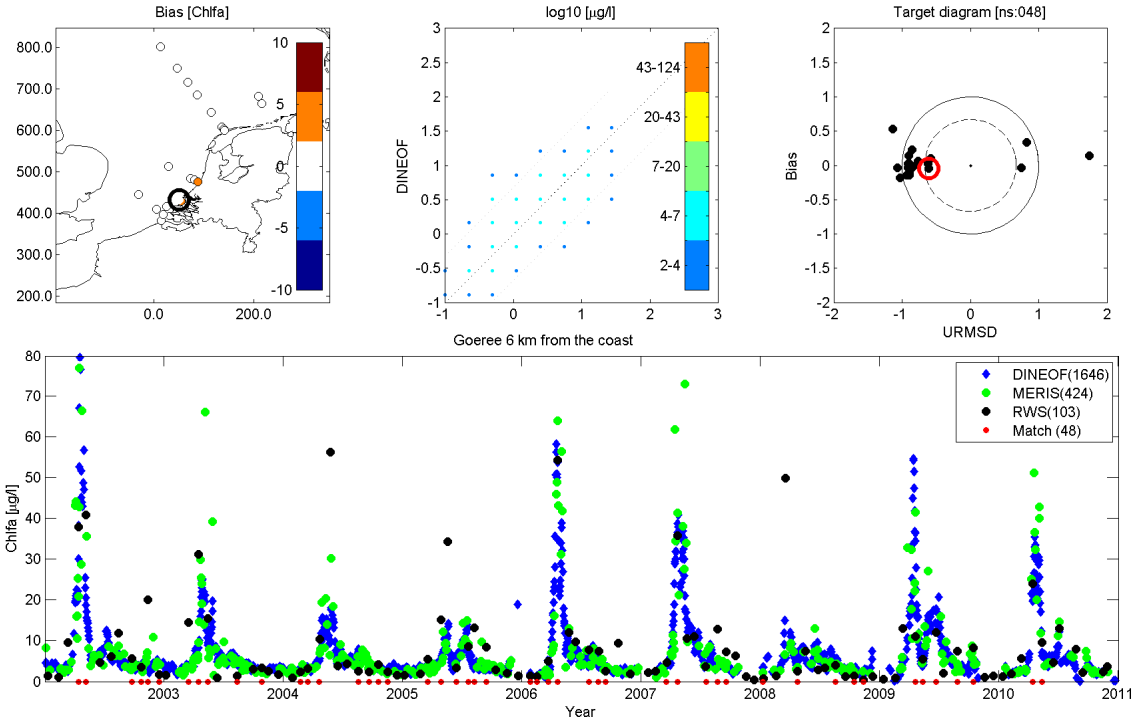

Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:016]
DINEOF
Bias
URMSD
16-38
9-16
5-9
3-5
2-3
Station NR-700
DINEOF(1646)
MERIS(498)
MUMM(28)
Match (16)
Chlfa [µg/l]
Year

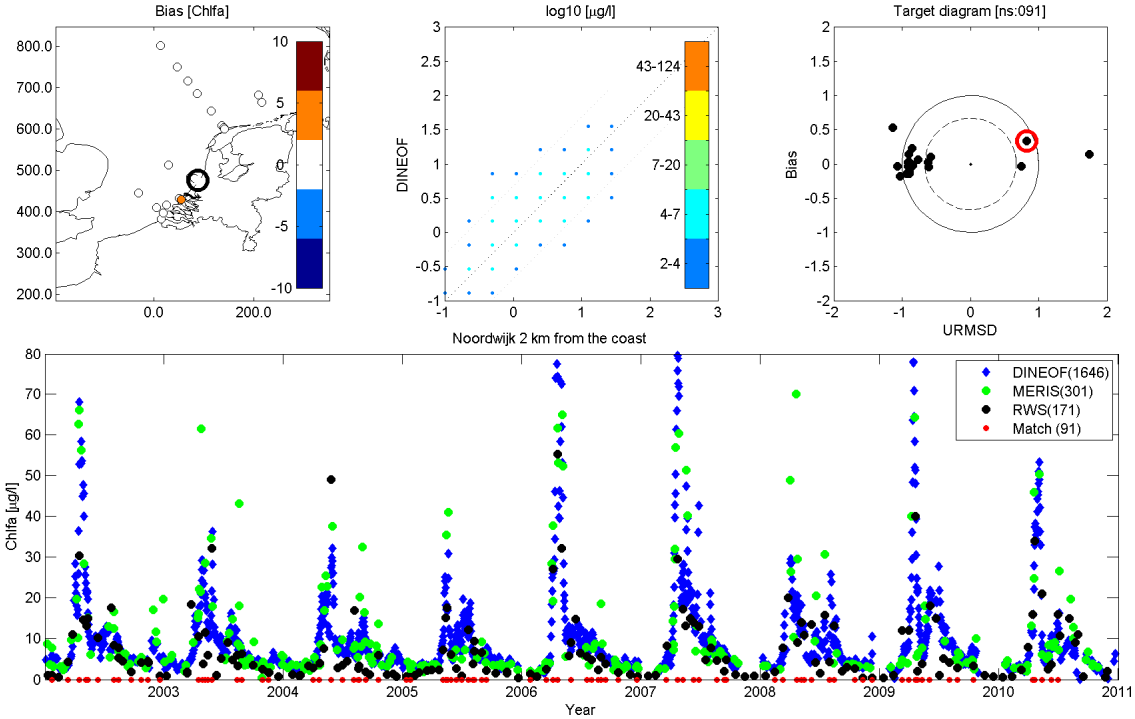

Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:012]
DINEOF
Bias
URMSD
16-38
9-16
5-9
3-5
2-3
Station NR-800
DINEOF(1646)
MERIS(455)
MUMM(21)
Match (12)
Chlfa [µg/l]
Year

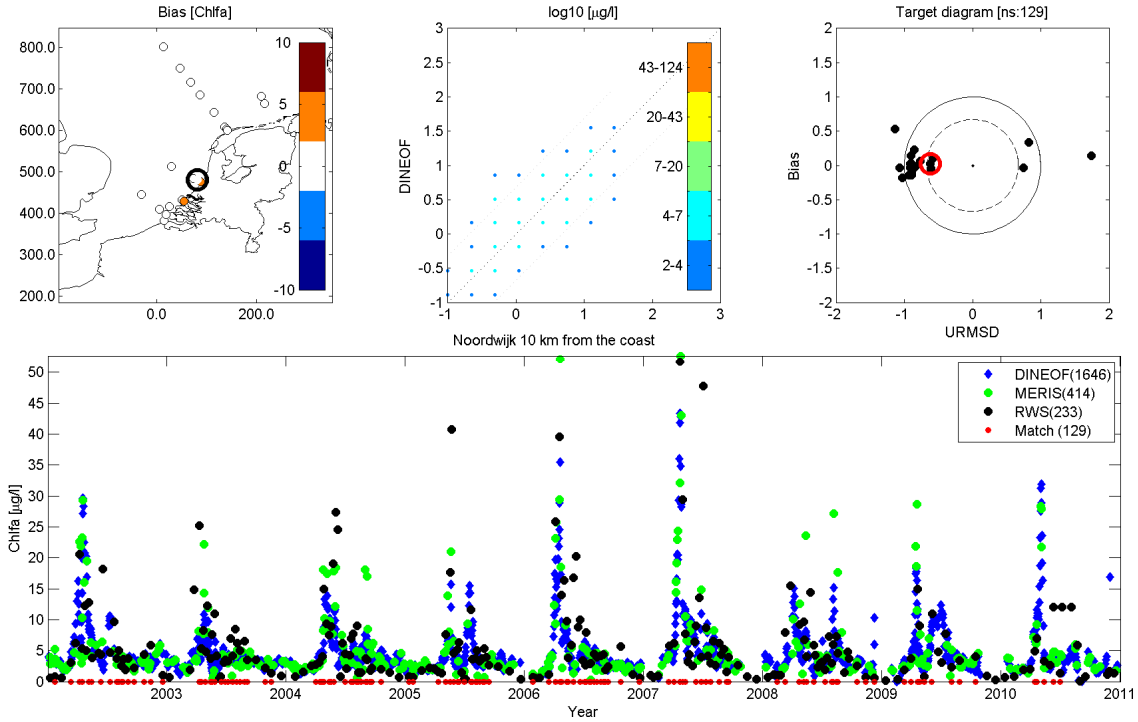

Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:047]
DINEOF
Bias
URMSD
43-124
20-43
7-20
4-7
2-4
Boomkensdiep
DINEOF(1646)
MERIS(354)
RWS(84)
Match (47)
Chlfa [µg/l]
Year

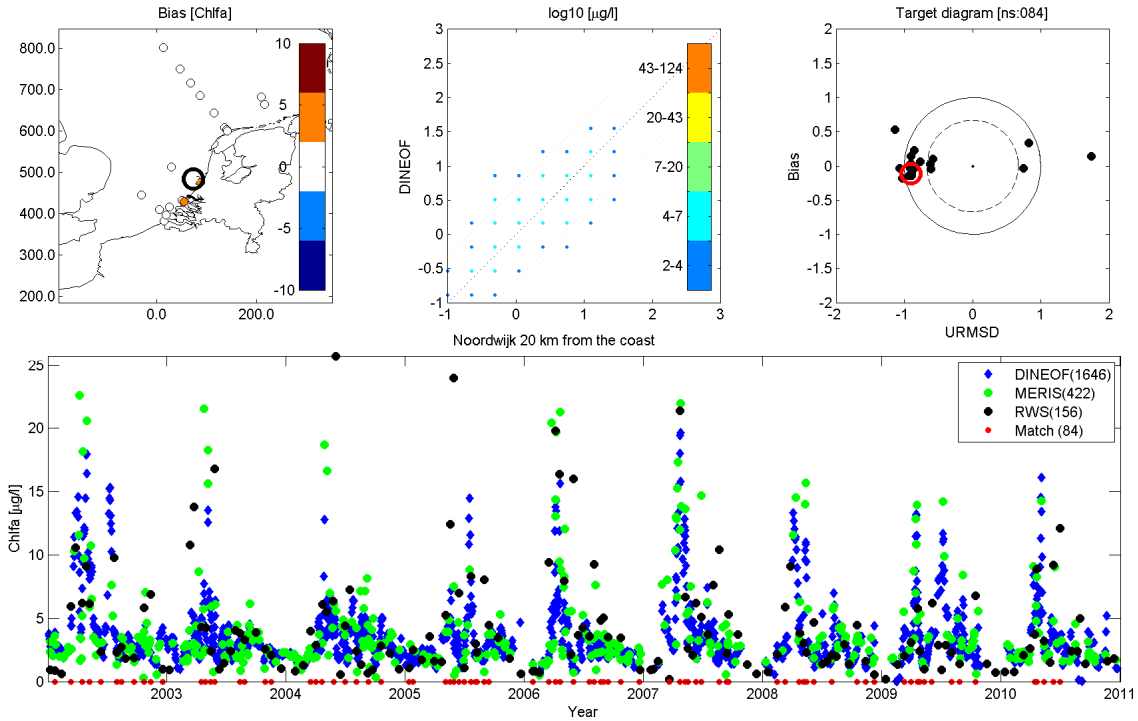

Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:012]
DINEOF
CH2
Bias
URMSD
16-38
9-16
5-9
3-5
2-3
DINEOF(1646)
MERIS(421)
MUMM(12)
Match (12)
Chlfa [µg/l]
Year

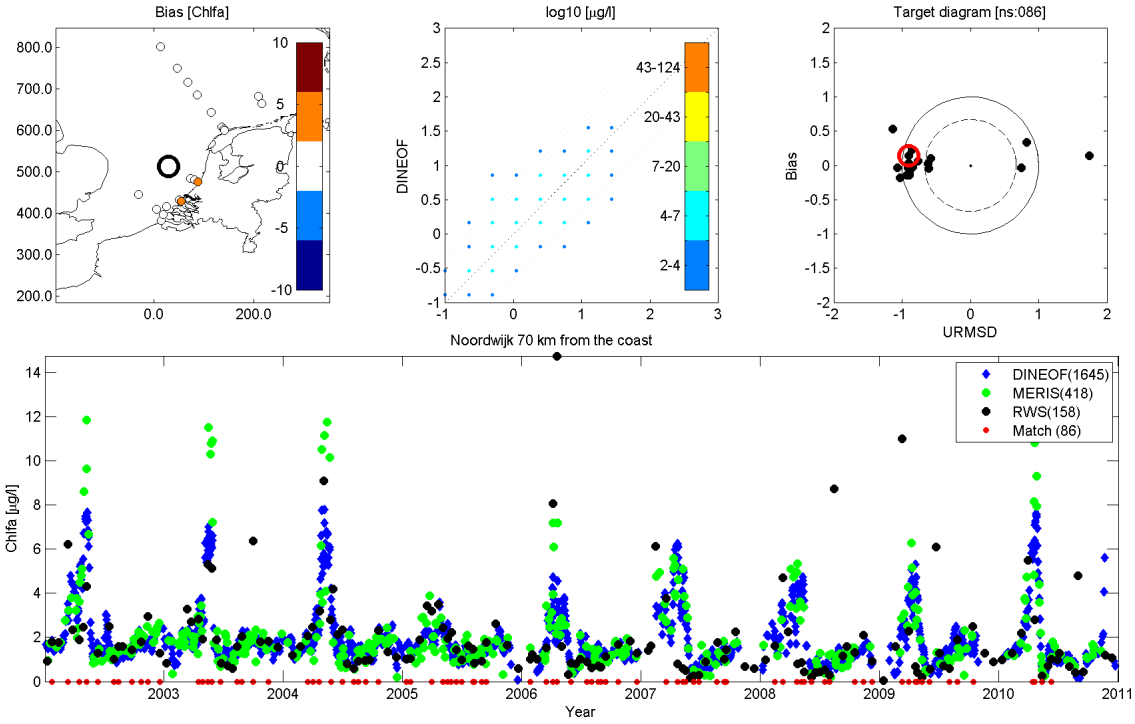

Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:025]
DINEOF
Goeree 2 km from the coast
Bias
URMSD
43-124
20-43
7-20
4-7
2-4
DINEOF(1645)
MERIS(360)
RWS(56)
Match (25)
Chlfa [µg/l]
Year

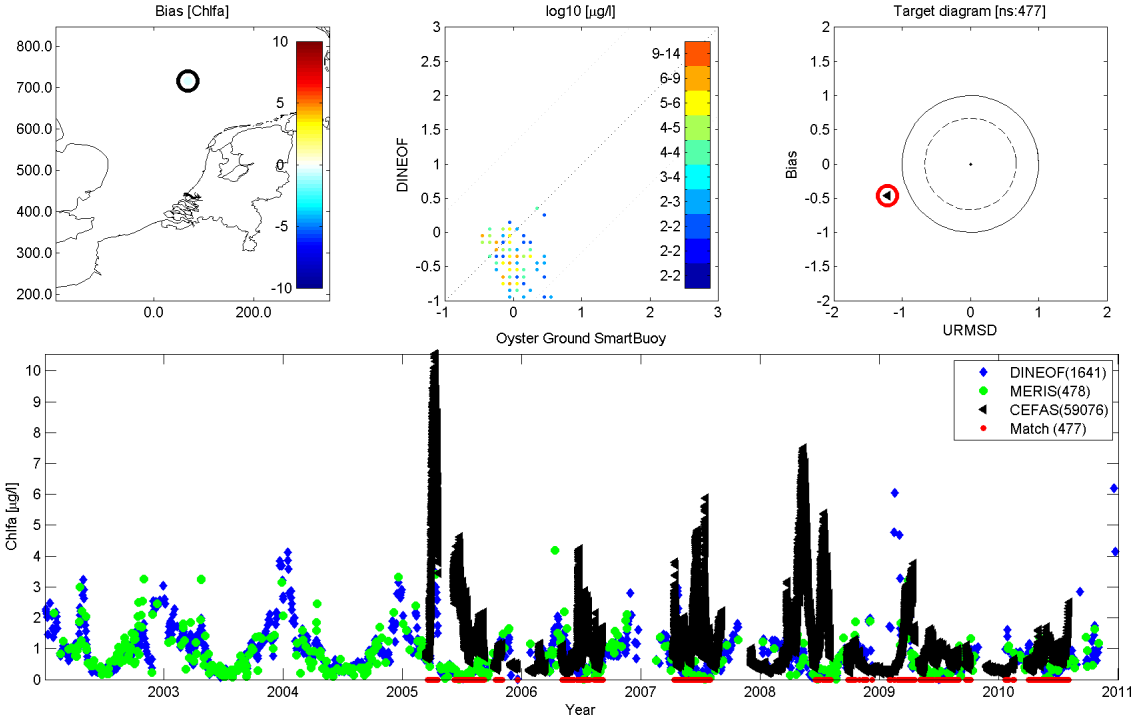

Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:048]
DINEOF
Bias
URMSD
43-124
20-43
7-20
4-7
2-4
Goeree 6 km from the coast
DINEOF(1646)
MERIS(424)
RWS(103)
Match (48)
Chlfa [µg/l]
Year

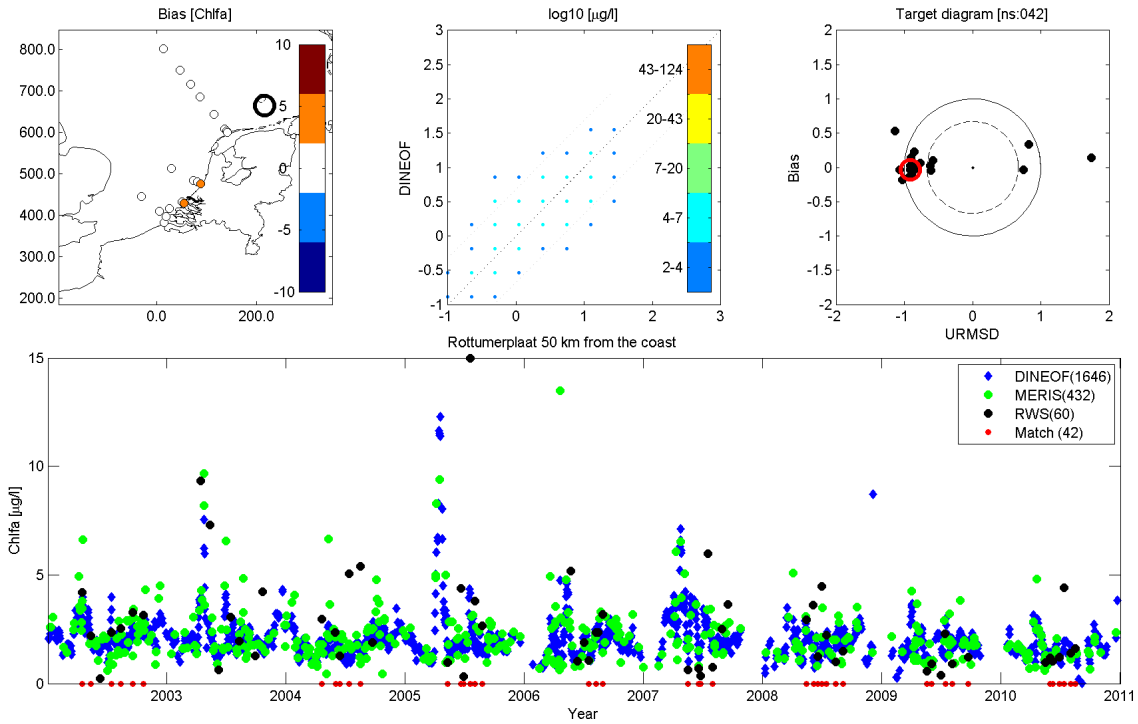

Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:091]
DINEOF
Bias
URMSD
43-124
20-43
7-20
4-7
2-4
Noordwijk 2 km from the coast
DINEOF(1646)
MERIS(301)
RWS(171)
Match (91)
Chlfa [µg/l]
Year

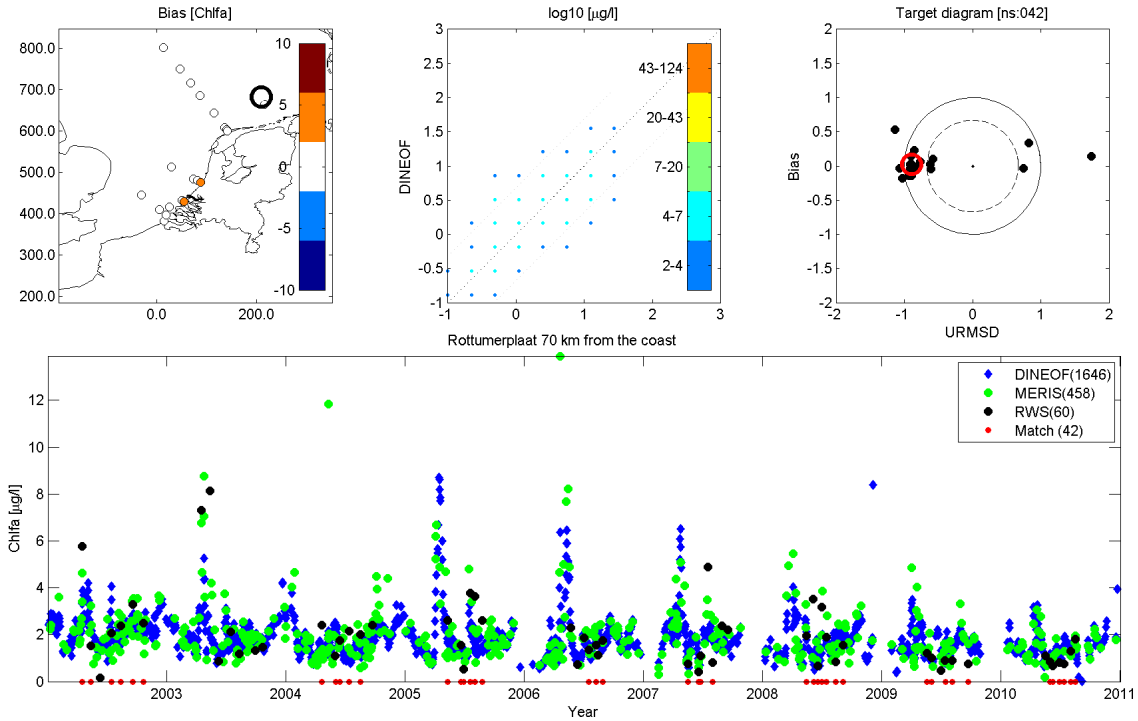

Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:129]
DINEOF
Bias
URMSD
43-124
20-43
7-20
4-7
2-4
Noordwijk 10 km from the coast
DINEOF(1646)
MERIS(414)
RWS(233)
Match (129)
Chlfa [µg/l]
Year

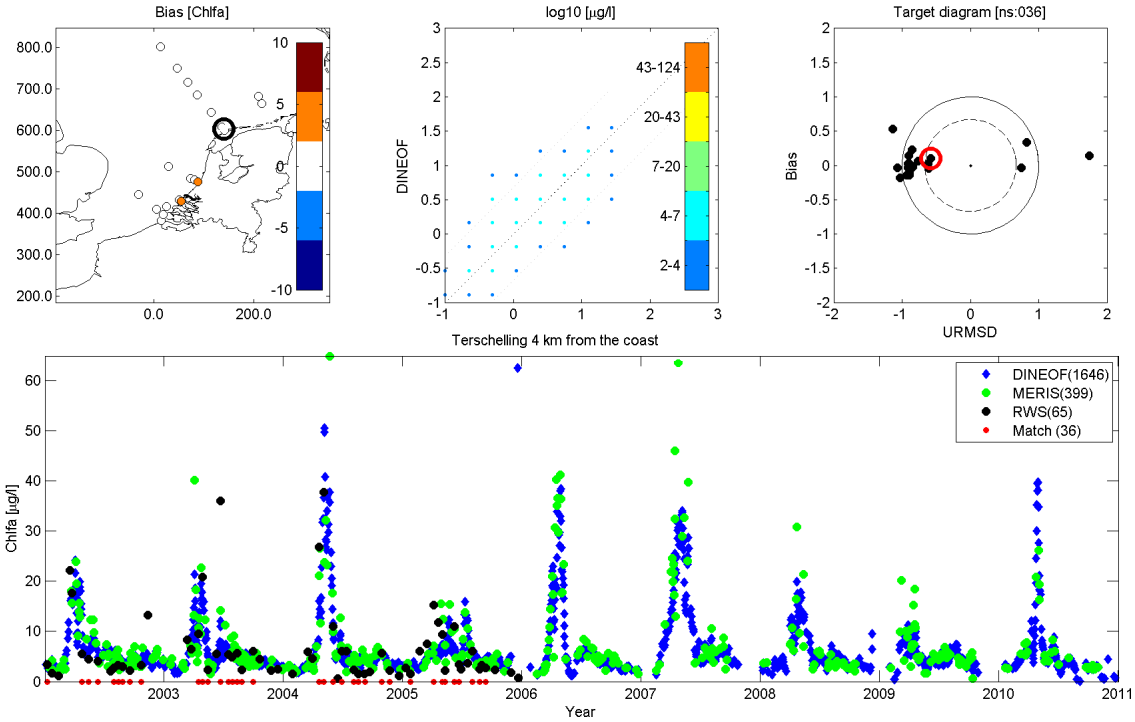

Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:084]
DINEOF
Bias
URMSD
43-124
20-43
7-20
4-7
2-4
Noordwijk 20 km from the coast
DINEOF(1646)
MERIS(422)
RWS(156)
Match (84)
Chlfa [µg/l]
Year

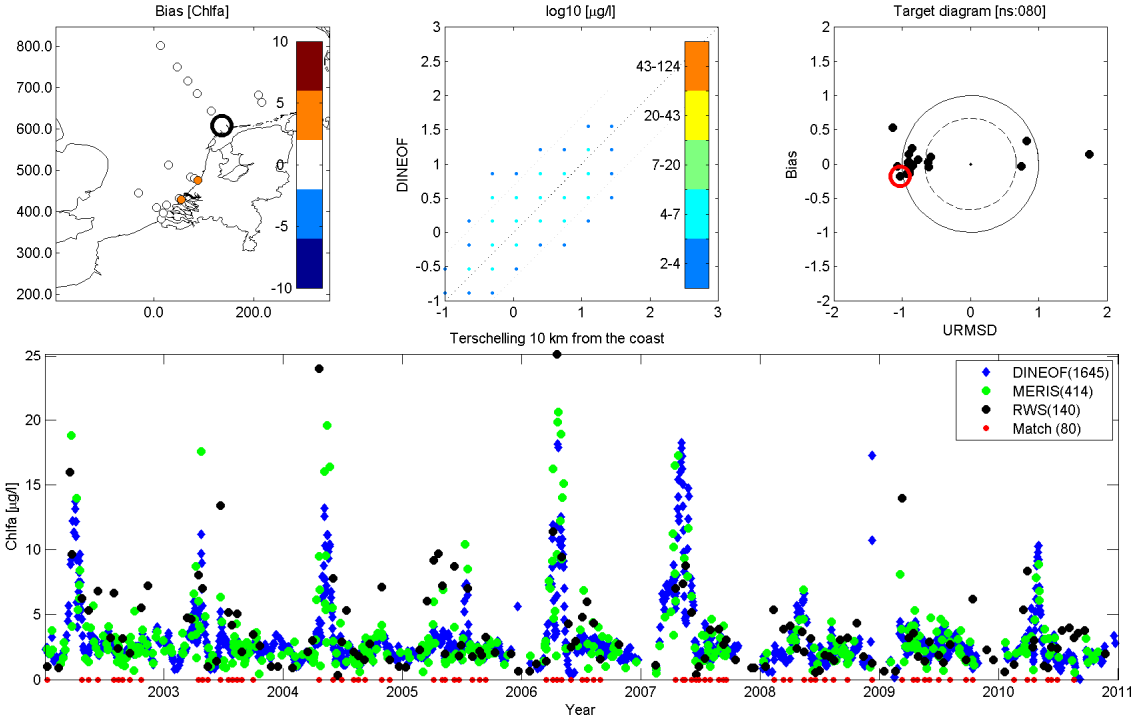

Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:086]
DINEOF
Bias
URMSD
43-124
20-43
7-20
4-7
2-4
Noordwijk 70 km from the coast
DINEOF(1645)
MERIS(418)
RWS(158)
Match (86)
Chlfa [µg/l]
Year

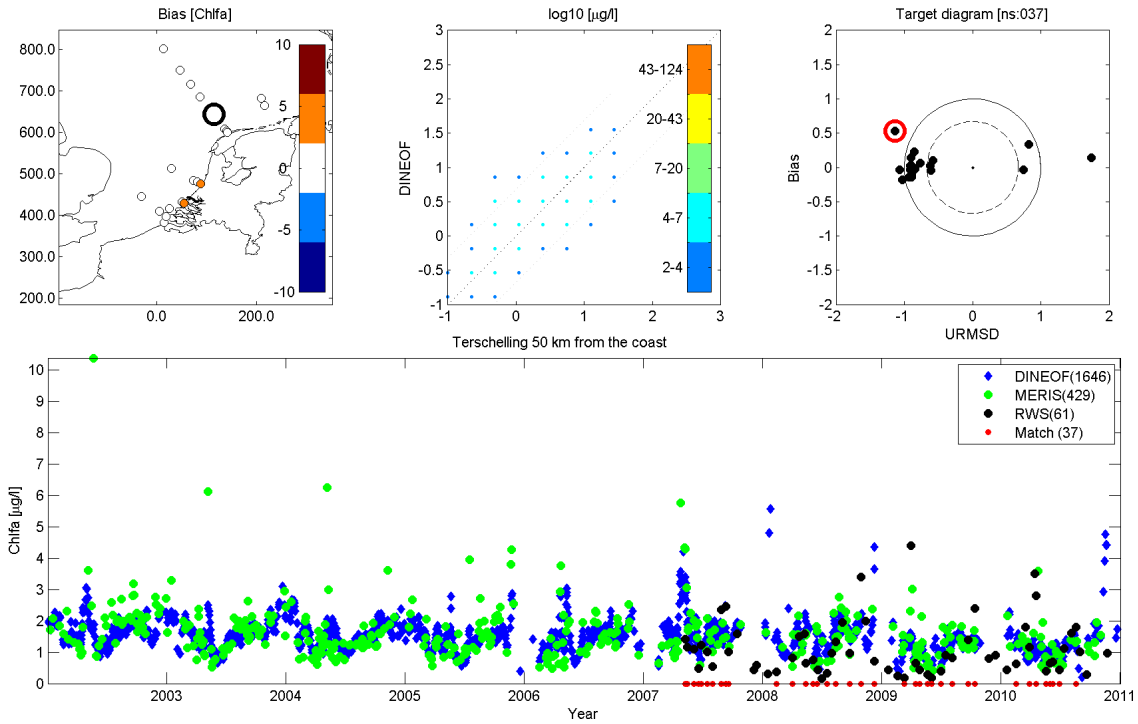

Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:477]
DINEOF
Bias
URMSD
9-14
6-9
5-6
4-5
4-4
3-4
2-3
2-2
2-2
2-2
Oyster Ground SmartBuoy
DINEOF(1641)
MERIS(478)
CEFAS(59076)
Match (477)
Chlfa [µg/l]
Year

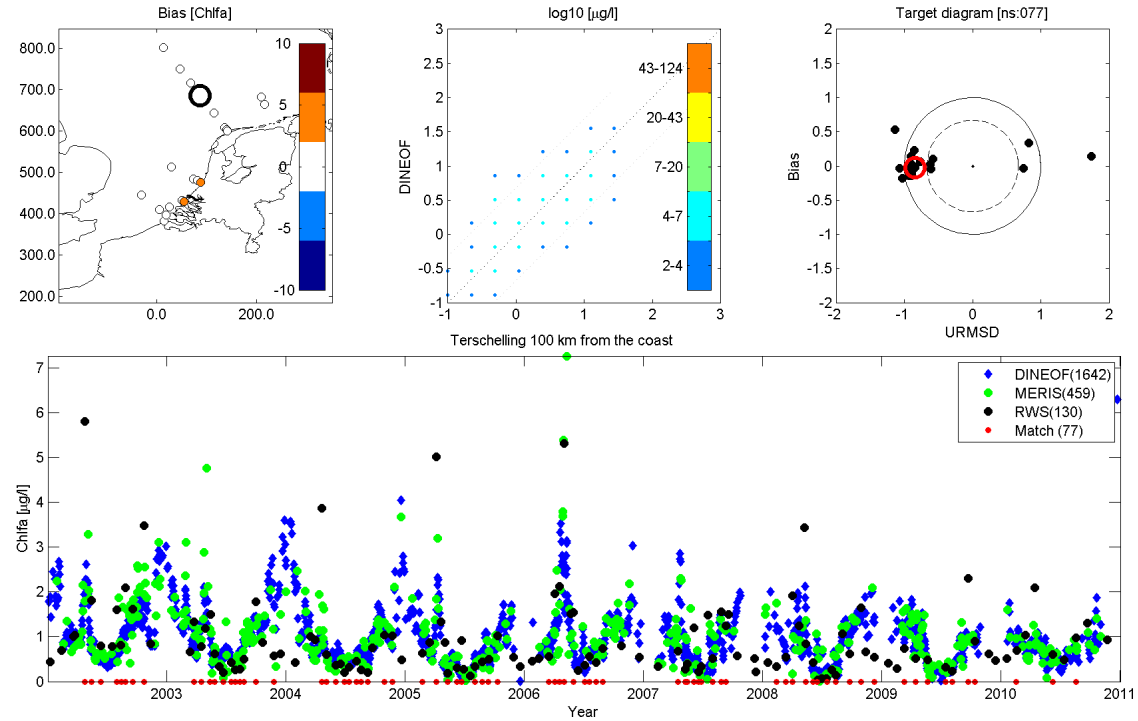

Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:042]
DINEOF
Bias
URMSD
43-124
20-43
7-20
4-7
2-4
Rottumerplaat 50 km from the coast
DINEOF(1646)
MERIS(432)
RWS(60)
Match (42)
Chlfa [µg/l]
Year

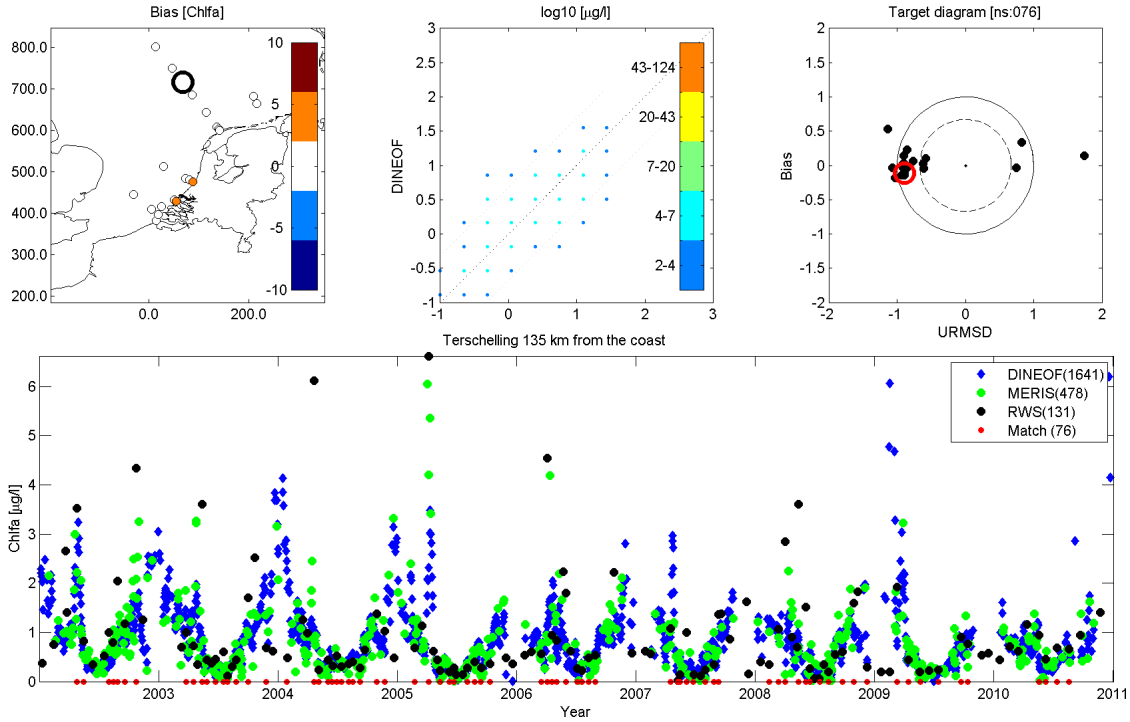

Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:042]
DINEOF
Bias
URMSD
43-124
20-43
7-20
4-7
2-4
Rottumerplaat 70 km from the coast
DINEOF(1646)
MERIS(458)
RWS(60)
Match (42)
Chlfa [µg/l]
Year

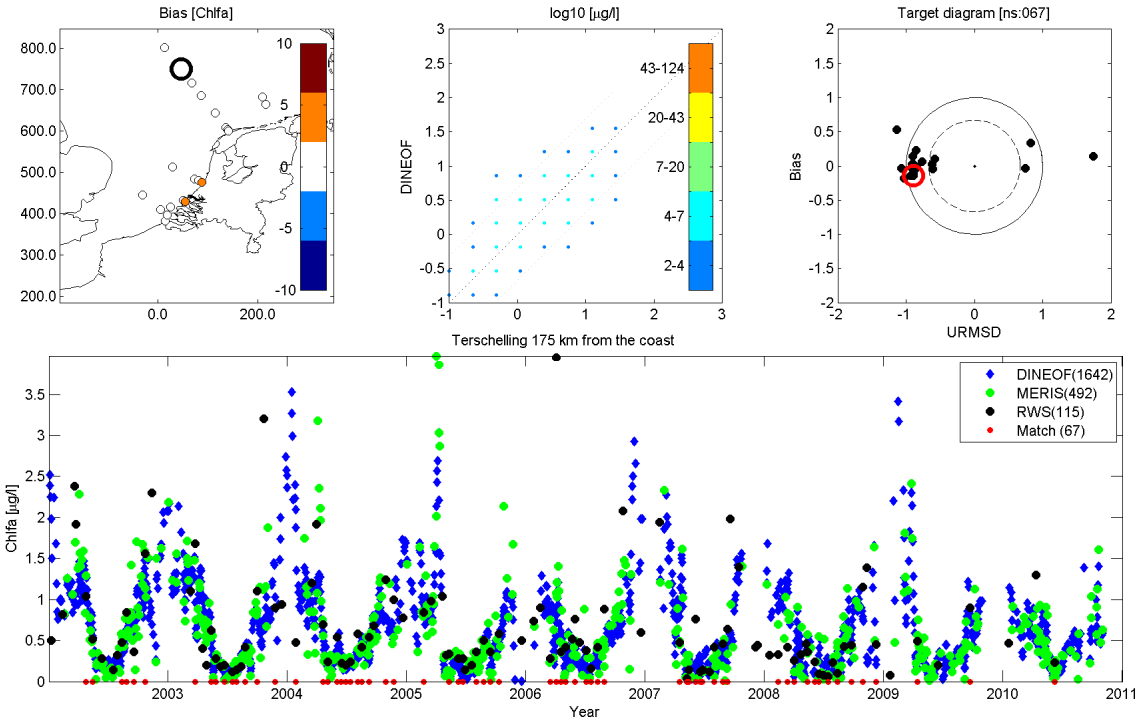

Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:036]
DINEOF
Bias
URMSD
43-124
20-43
7-20
4-7
2-4
Terschelling 4 km from the coast
DINEOF(1646)
MERIS(399)
RWS(65)
Match (36)
Chlfa [µg/l]
Year

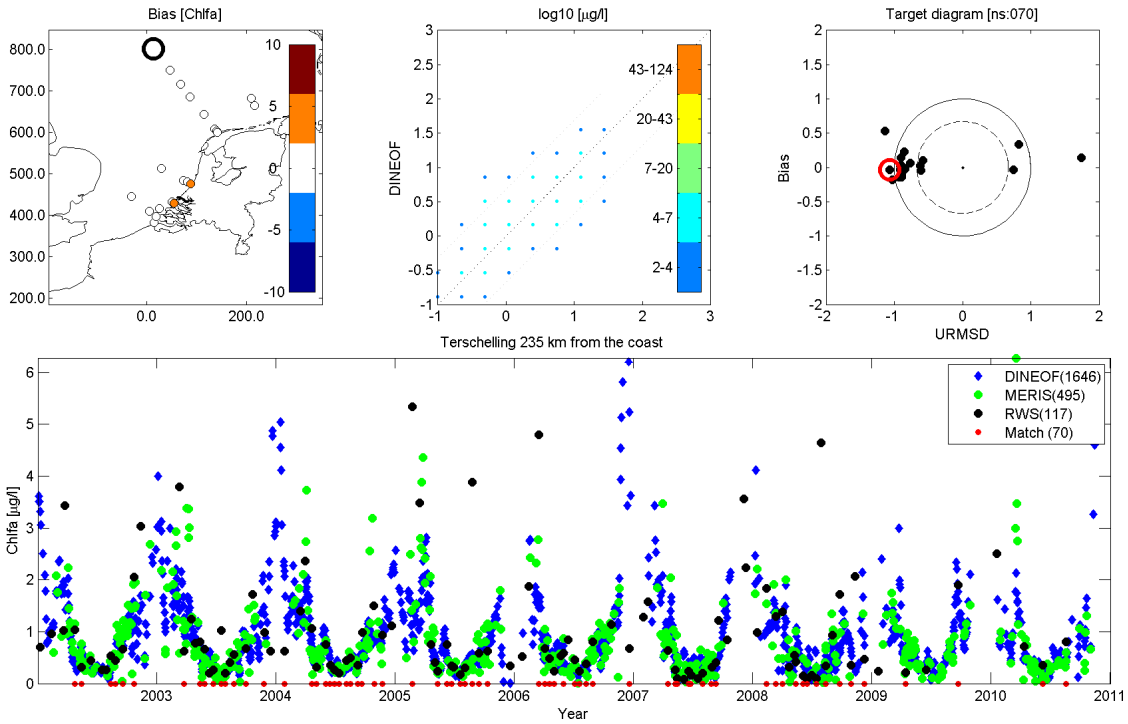

Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:080]
DINEOF
Bias
URMSD
43-124
20-43
7-20
4-7
2-4
Terschelling 10 km from the coast
DINEOF(1645)
MERIS(414)
RWS(140)
Match (80)
Chlfa [µg/l]
Year


Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:037]
DINEOF
Bias
URMSD
43-124
20-43
7-20
4-7
2-4
Terschelling 50 km from the coast
DINEOF(1646)
MERIS(429)
RWS(61)
Match (37)
Chlfa [µg/l]
Year


Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:077]
DINEOF
Bias
URMSD
43-124
20-43
7-20
4-7
2-4
Terschelling 100 km from the coast
DINEOF(1642)
MERIS(459)
RWS(130)
Match (77)
Chlfa [µg/l]
Year


Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:076]
DINEOF
Bias
URMSD
43-124
20-43
7-20
4-7
2-4
Terschelling 135 km from the coast
DINEOF(1641)
MERIS(478)
RWS(131)
Match (76)
Chlfa [µg/l]
Year


Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:067]
DINEOF
Bias
URMSD
43-124
20-43
7-20
4-7
2-4
Terschelling 175 km from the coast
DINEOF(1642)
MERIS(492)
RWS(115)
Match (67)
Chlfa [µg/l]
Year


Bias [Chlfa]
log10 [µg/l]
Target diagram [ns:070]
DINEOF
43-124
20-43
7-20
4-7
2-4
Bias
URMSD
Terschelling 235 km from the coast
DINEOF(1646)
MERIS(495)
RWS(117)
Match (70)
Chlfa [µg/l]
Year

B Fine-grained DINEOF on MERIS

Below, the three dominant DINEOF modes resulting from the analysis of the MERIS chl-a data are shown for low smoothing parameter α (α=0.01). These are the modes eventually used to reconstruct the so-called gap-filled fields used in the analysis of time series and OSPAR eutrophication levels. The plots below are similar to what has been shown in Figure 5.11 and 5.12 but only for the MERIS modes, since the IS data are smooth by nature of their 4-weekly time resolution. The high smoothing (α=0.1) was applied in Figure 5.11 and 5.12 to enable comparison to the smooth IS modes.

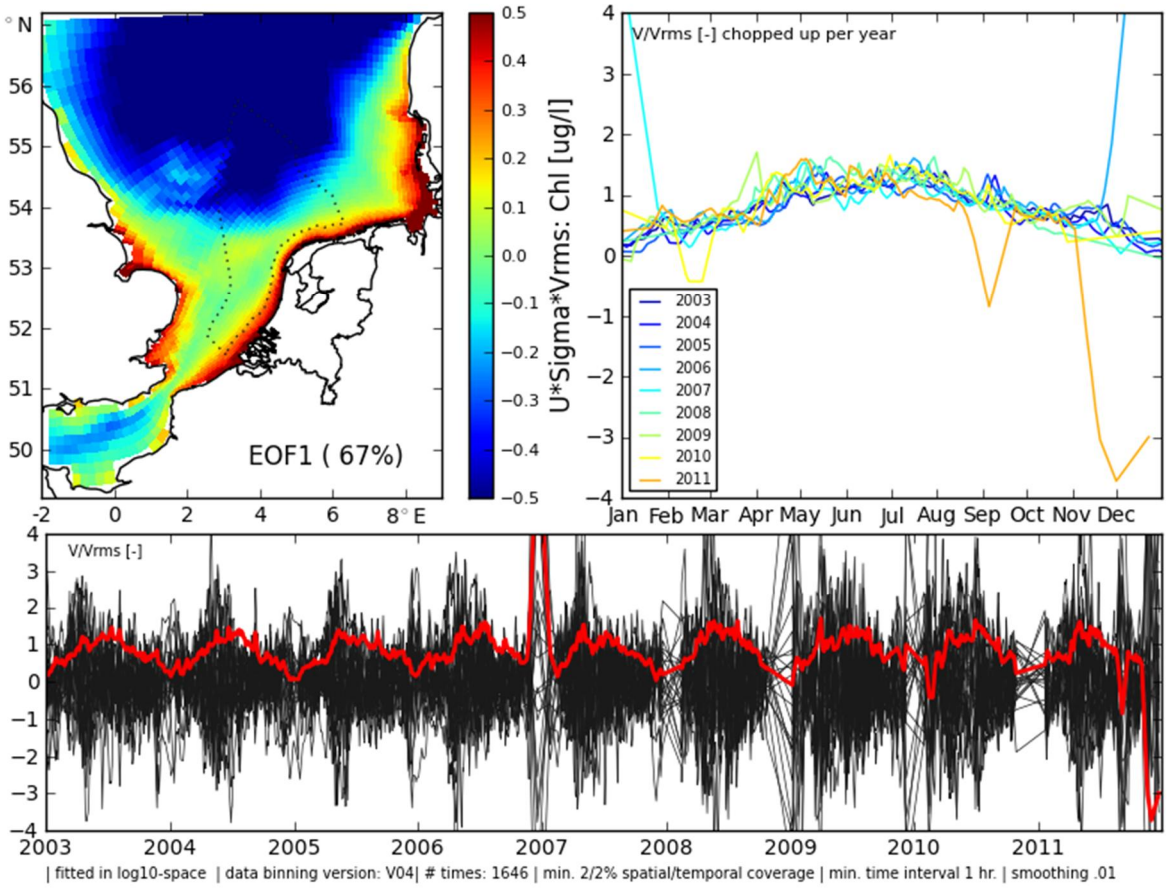


Figure B.1. First EOF mode of MERIS Chlfa obtained with high smoothing. Top left spatial pattern, bottom: temporal pattern, top right the temporal pattern chopped in calendar years.

Note that the first mode for less smoothing is nearly identical to the smoothed one in Chapter 5. This indicates that this pattern is a robust and global feature of the data.

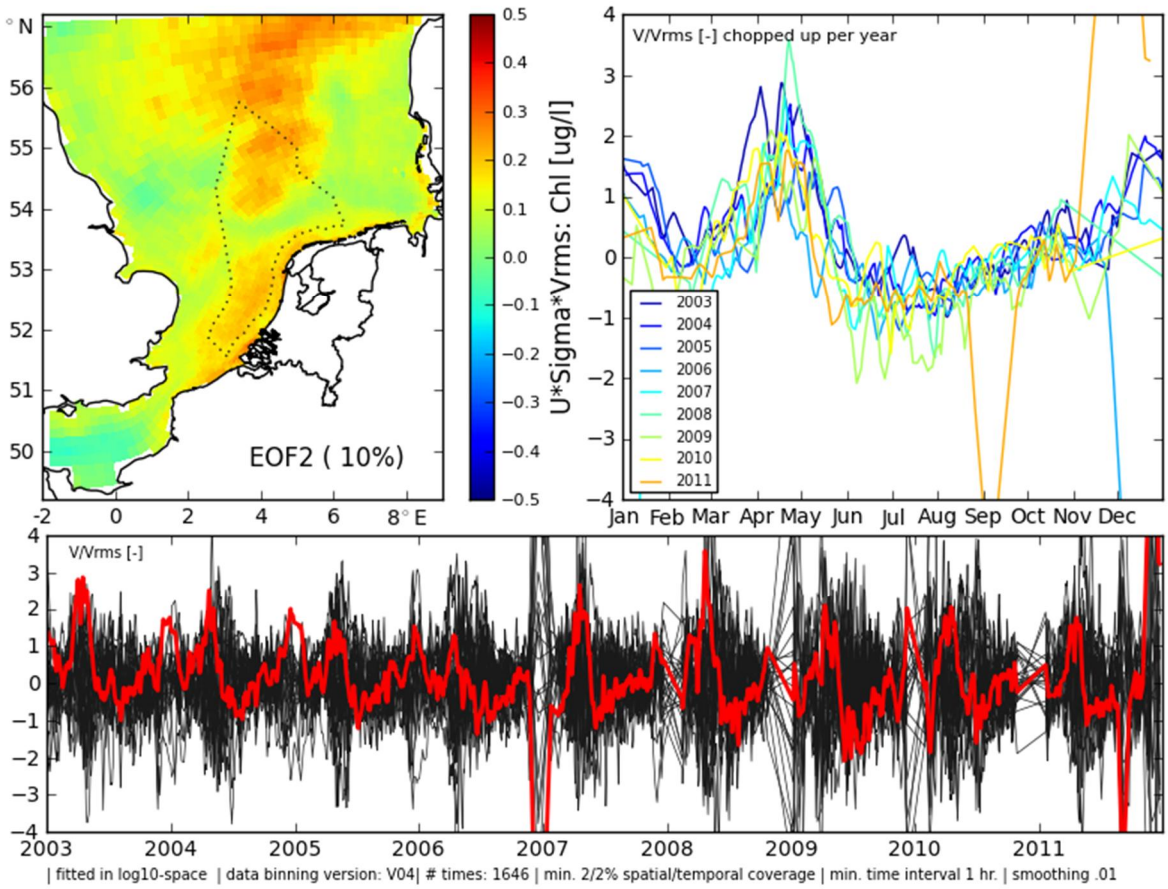


Figure B.2. Second EOF mode of MERIS Chlfa obtained with high smoothing.

Also the second mode for less smoothing is highly comparable to the smoothed one in Chapter 5. This indicates that also this pattern is a robust and global feature of the data.

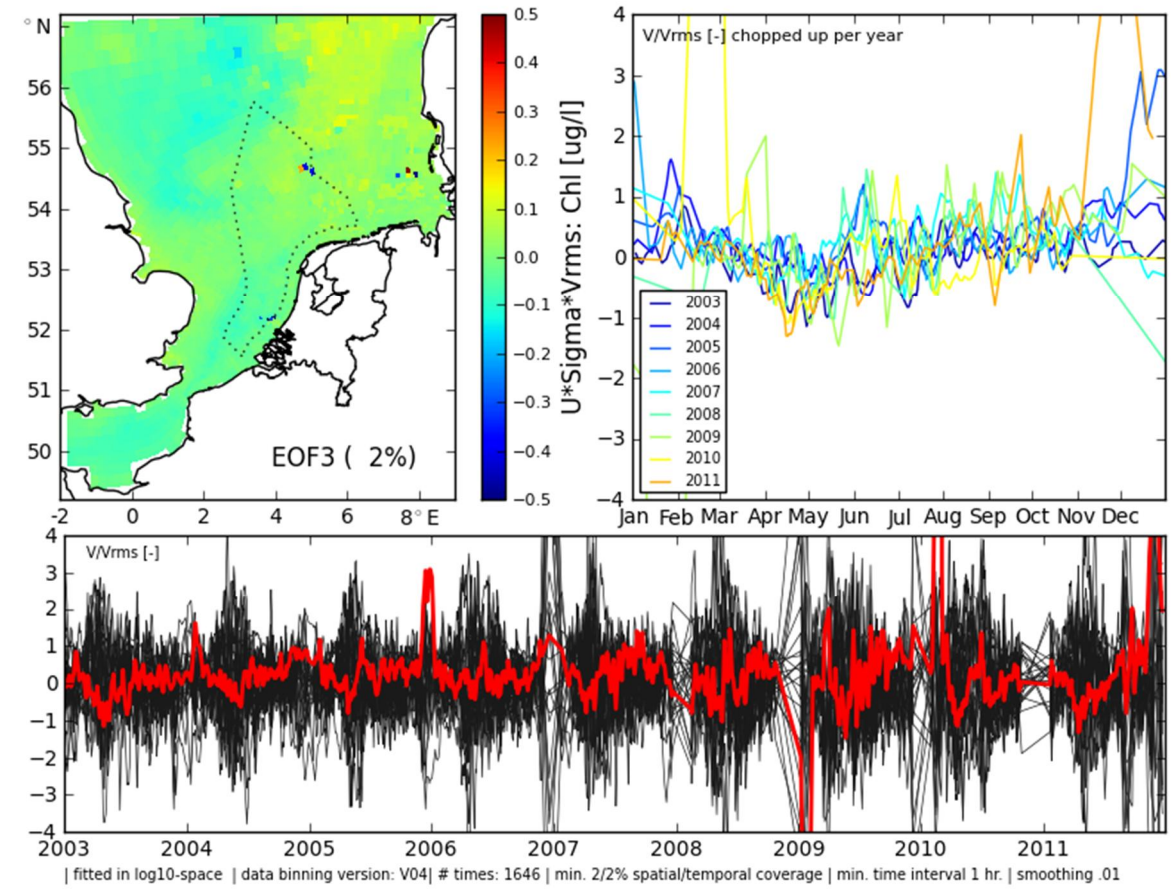


Figure B.3 Third EOF mode of MERIS Chlfa obtained with high smoothing.

The third EOF mode with low temporal smoothing is quite different from the third mode presented in Chapter 5 with higher smoothing. This indicates that these smaller scale patterns are more sensitive to the temporal smoothing and hence their estimate may be less reliable when smoothing is too strong.