Impact of irregular sampling by MERIS on eutrophication monitoring products for WFD and MSFD applications

Dimitry Van der Zande*[1], Geneviève Lacroix[1], Xavier Desmit[1], Kevin Ruddick[1].
[1]*Royal Belgian Institute of Natural Sciences, MUMM, Gulledelle 100, B-1200 Brussels, Belgium.*
* Corresponding author, email: Dimitry.vanderzande@mumm.ac.be

Abstract

The chlorophyll-a 90 percentile product (CHL-P90) gained importance in describing the eutrophication state of the Belgian part of the North Sea when Water Framework Directive requirements were transposed into Belgian law in July 2010. A MERIS based CHL-P90 product was developed to answer this need for information as MERIS provides a better temporal and spatial coverage compared to *in situ* monitoring and thus is potentially better suited to provide a P90 estimate. However, the irregular availability of satellite chlorophyll-a observations both in space and time due to cloudiness, quality flagging, sensor malfunction, etc. has to be considered as it impacts the product accuracy. This impact is two-fold and dependent on (1) the availability of observations during the actual phytoplankton bloom and (2) a proportional distribution of observations in the bloom and non-bloom period.

This effect of irregular sampling on the accuracy of CHL-P90 was studied in detail in simulations. The MIRO&CO-3D ecosystem model was used to generate realistic time series of chlorophyll-a concentrations with high temporal resolution in the most important Belgian monitoring stations. These time series were sub-sampled using the actual observation density of the MERIS satellite in Belgian waters. Results show that a mean relative error of 25.4% on CHL-P90 estimate can be expected due to the effects of irregular sampling. The results of this study were used to improve the CHL-P90 algorithms by adding a weighing procedure taking into account irregular sampling issues to reduce these errors to 9.9%.

Keywords: chlorophyll-a P90, MERIS, Water Framework Directive, Belgium

1. Introduction

The Water Framework Directive (WFD) and the Marine Strategy Framework Directive (MSFD) are currently the most important drivers for monitoring the coastal and offshore waters in Europe with the objective of reaching a 'good environmental status' (Gohin et al., 2008). Human-induced eutrophication is one of the criteria for assessing the extent to which good environmental status is being achieved. Eutrophication can be defined as the enrichment of water by nutrients causing an 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 concerned, and therefore refers to the

undesirable effects resulting from anthropogenic enrichment by nutrients (OSPAR, 1998).

In June 2010, WFD requirements (Directive, 2000) for the Belgian Coastal Zone (BCZ) were transposed into Belgian law (Royal Decree 23.06.2010) thus structuring/guiding the water quality monitoring programs. The eutrophication status is established by monitoring the chlorophyll-a (CHL) concentration as it is a proxy of phytoplankton biomass. More specifically, the indicator of choice is the chlorophyll-a 90 percentile (CHL-P90) over the phytoplankton growing season (i.e. March – November incl.) expressed in μg/l. The CHL-P90 is the CHL concentration below which 90 percent of observations fall. While *in situ* data acquisition is still considered as the main monitoring tool there is a growing tendency to use optical remote sensing as a supporting tool to achieve the monitoring requirements because of severe resource constraints of available ship time and manpower (Sorensen et al., 2002). For example, in Belgium *in situ* measurements are collected 5-7 times in average per growing season in 10 stations distributed in the BCZ (Ruddick et al., 2008c). These numbers are too low to calculate CHL-P90 values with sufficient accuracy. Hence the need for additional information sources.

The algal1 and algal2 products from MERIS are now considered as sufficiently mature for use in applications in the turbid Belgian waters (Ruddick et al., 2008a, 2008b). The MERIS data enables the calculation of CHL-P90 pixel-by-pixel resulting in a map product which is expected to provide more accurate CHL-P90 estimates due to an increased temporal and spatial resolution compared to the *in situ* data (Park et al. 2010). Subsequently, this map product provides the policy makers the means to objectively and quantitatively report on the WFD and MSFD guidelines for reaching a 'good environmental state' by 2020.

However, the irregular availability of satellite CHL observations both in space and time due to cloudiness, quality flagging, sensor malfunction, etc. has to be considered (Sirjacobs et al., 2011) as it impacts the accuracy of CHL-P90. This impact is two-fold and dependant on (1) the availability of observations during the actual phytoplankton bloom and (2) a proportional distribution of observations in the bloom and non-bloom periods. Figure 1 illustrates this by showing how a different timing of 50 samples on the same CHL time series can result in significantly different CHL-P90 estimates. From this example it is obvious that sufficient sampling of the bloom period is crucial for generating an accurate CHL-P90 product but can still lead to errors in case of under- or oversampling compared to the non-bloom period.

The main objective of this study is to perform a detailed sensitivity analysis on the CHL-P90 product considering irregular sampling and to use the results to develop a correction procedure to improve the CHL-P90 product. Simulation techniques were used to ensure the availability of sufficient reference data. The MIRO&CO-3D ecosystem model (Lacroix et al., 2007) provided CHL time series with a high temporal resolution (i.e. 30 minutes) in five Belgian monitoring stations for a full growing season in the arbitrarily chosen year 2006. These five time series were generated taking into account processes such as wind- and tide-induced variability, complex biological interactions etc. and were subsequently sub sampled using actual pixel specific MERIS sampling frequency during the growing season of 2003 to 2010. This resulted in more than 20 000 sub sampled CHL time series which could be directly compared to the five MIRO&CO-3D time series to perform the accuracy analysis.

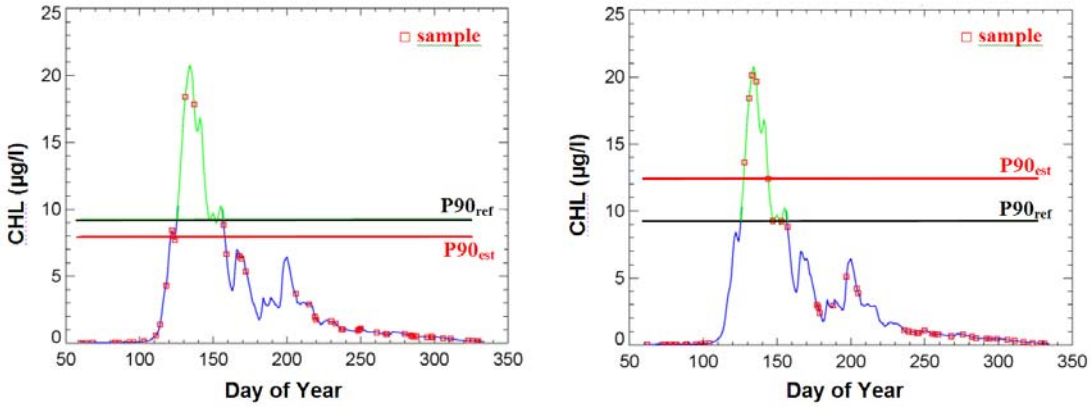

Fig 1. Illustration of the impact of two different timings of 50 samples on the same CHL time series on CHL-P90 estimates. Blue/green line: time series of CHL computed by MIRO&CO-3D model (Lacroix et al., 2007) for the year 2006 for a single Belgian monitoring station (51,27°N-2,91°W) with the phytoplankton bloom marked in green. $P90_{ref}$ (black line) and $P90_{est}$ (red line) are the percentile 90 computed respectively from the daily data and from a subsample of 50 data.

2. Materials & Methods

2.1 Belgian Coastal Zone

The Belgian Coastal Zone (BCZ) is located in the Southern Bight of the North Sea (SNS) and covers approximately 3454km² with an average depth of 20m-30m. The BCZ is influenced directly by the Scheldt river (and other small Belgian rivers) inputs as well as indirectly by waters discharges from French rivers such as the Seine and the Somme and the Dutch Rhine/Meuse system (Lacroix et al., 2004). The river nutrient input causes eutrophication in the BCZ leading to high-biomass algal blooms, mainly the Haptophycea *Phaeocystis globosa* that spreads over the entire area in spring (Lancelot et al., 1987).

2.2 MIRO&CO-3D model

The MIRO&CO-3D model is a combination of the 3D hydrodynamical model described in Lacroix et al. (2004) based on the COHERENS model (Luyten et al., 1999) and the biogeochemical MIRO model (Lancelot et al., 2005). The MIRO&CO-3D ecosystem model was used to generate simulated time series of chlorophyll-a concentrations with high temporal resolution in five Belgian monitoring stations. The model is capable of simulating the transport and dynamics of inorganic and organic nutrients, phytoplankton, bacterioplankton and zooplankton biomass in the SNS (Lacroix et al., 2007). For this theoretical study, the BCZ was subdivided in five zones around the five *in situ* monitoring stations using the nearest neighbour interpolation. For each zone a typical CHL time series during the phytoplankton growing season was generated capturing the main variations in the BCZ (Fig.2). While these time series are theoretical,

we believe that they capture the main phytoplankton dynamics present in the BCZ with generally an intensive *Phaeocystis globasa* bloom during the spring followed by a less intensive summer Diatoms bloom (Lancelot et al., 2005). Bloom intensity decreases with increasing distance from the coast as observed from Remote Sensing observations (Rousseau et al., 2006; Ruddick et al., 2008c). The CHL time series are presented in fig 2 (right) showing a phytoplankton biomass peak around day 134 (i.e. 12th May) for the different zones.

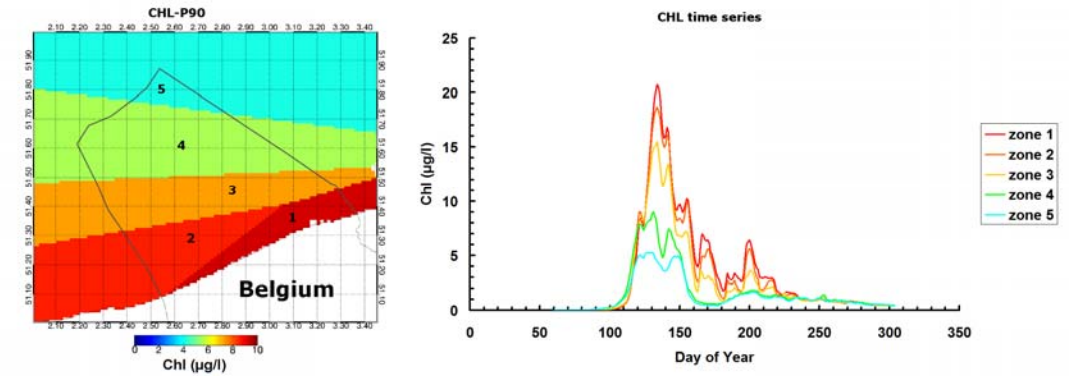

Fig.2. Left: CHL-P90 computed from daily MIRO&CO-3D data for the year 2006 five zones around the five *in situ* monitoring stations using the nearest neighbour interpolation catching the main variations in algal bloom intensities in the BCZ represented by the black line. Right: Daily time series of CHL computed by MIRO&CO-3D (Lacroix et al., 2007) at the 5 chosen stations. The colour scale referring to the 5 zones is the same for both figures.

2.3 MERIS sampling frequency

Satellite images are widely used in operational oceanography and in ecosystem dynamics studies due to their extensive spatial and temporal coverage. The Medium Resolution Imaging Spectrometer (MERIS) instrument provides level 2 algal products (CHL in µg/l) (Doerffer & Schiller, 2007) for both case 1 and case 2 waters on a daily basis with a spatial resolution of 1km and a theoretical daily temporal resolution for the BCZ. However, satellite remote sensing is subject to one major limitation: cloud presence can totally or partially cover the area of interest. For the BCZ this generally results in a high percentage of missing data in the daily images. This missing data is not evenly distributed over the year, and thus biases any simple averaging calculations. To study the impact of gaps in the observation continuity on the CHL-P90 product quality, the actual MERIS sampling frequencies per pixel for the years 2003 to 2010 were applied to the five simulated CHL time series described in section 2.2. This resulted in more than 20 000 sub sampled CHL time series accompanied with the detailed model data needed to perform the accuracy analysis.

2.4 Standard CHL-P90 product

CHL-P90 values were calculated pixel-by-pixel from the mean and standard deviation (σ) of the CHL time series over the growing season by the approximation:

$$\text{CHL-P90} = \text{Mean} + 1.28 * \sigma$$

with a minimum of 2 CHL values per pixel. The mean and standard deviation σ are computed in log transformed space. This statistical approach is an approximation for a log-normal distribution of selecting the CHL value in a time series such that 90% of the observations are equal or less than this value. It requires considerable less computing resources as no actual time series need to be stored for each pixel speeding up the processing significantly. It delivers accurate CHL-P90 values when the sampling is regularly distributed over the growing season and when the underlying distribution is log-normal.

2.5 Weighted CHL-P90 calculation

In this approach a quality assessment of the satellite-based CHL time series is taken into account when generating the CHL-P90 product to compensate for irregular sampling. In a first step a key period (KP) is defined as the period containing the main algal bloom. In this study this period is defined as the period starting 14 days before and ending 14 days after the maximum CHL observation (Fig. 3). This arbitrary definition of the KP is region specific and needs to be adjusted using prior knowledge of the local phytoplankton dynamics from model results or *in situ*/Remote Sensing (RS) data. It is important that this period is sufficiently sampled to obtain a valid CHL-P90 estimate.

Next, the key period proportion (KPP) is calculated as follows:

$$KPP = \frac{Nrdays_{KP}}{Nrdays_{GS}}$$

with $Nrdays_{KP}$ the total number of days of the KP (i.e. 28 days around the day of max CHL for the BCZ) and $Nrdays_{GS}$ the total number of days in the phytoplankton growing season (i.e. 275 days). The arbitrarily chosen KPP equals 10.18%. Knowing the KP, the KPP for the MERIS time series can be calculated is a similar fashion as well as the relative difference (RD) between KPP and KPP_{MERIS}:

$$KPP_{MERIS} = \frac{NrObservations_{KP}}{NrObservations_{GS}} \qquad RD = \frac{KPP_{MERIS} - KPP}{KPP}$$

with $NrObservations_{KP}$ the total number of CHL observations in the KP and $NrObservations_{GS}$ the total number of CHL observations in the complete phytoplankton growing season.

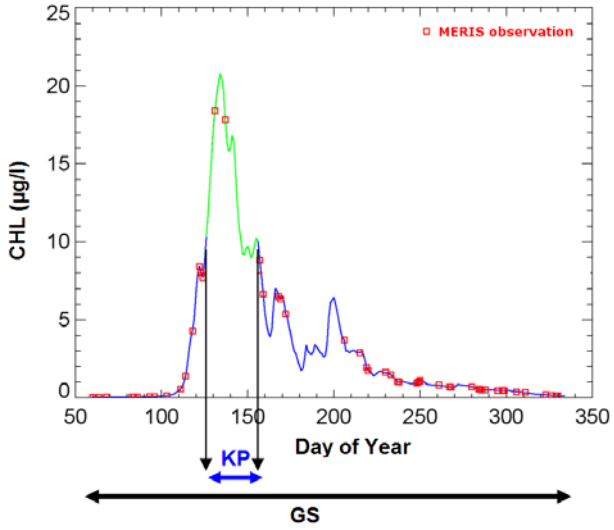

Fig.3. The region specific key period (KP) is defined as the period in the phytoplankton growing season (GS) containing the main algal bloom. Blue/green line: time series of CHL computed by MIRO&CO-3D model (Lacroix et al., 2007) for the year 2006 for a single Belgian monitoring station (51,27°N-2,91°W) with the key period marked in green. Red squares correspond to the dates where MERIS observations are available for the year 2003 in this case.

The KPP_{MERIS} can be calculated for each pixel and can be compared to the KPP per zone resulting in a quantification of sampling irregularities represented by the weight indices WI_{KP} and WI_{NKP} for the key period and the non-key period respectively:

$$WI_{KP} = \frac{KPP_{MERIS}}{KPP} \qquad WI_{NKP} = 1 - \frac{KPP_{MERIS}}{KPP}$$

The KP is undersampled by MERIS in the case of $WI_{KP} < 1$ which would result in an underestimation of the CHL-P90 using the standard method. Conversely, overestimation of the CHL-P90 is expected in the case of $WI_{KP} > 1$. This information can be used to improve the CHL-P90 procedure by giving all the observations a weight which is the inverse of the respective WI in order to compensate for the sampling irregularities.

Finally, the CHL values are ranked from low to high with their weights cumulated and normalized (W_{norm}). The weighted CHL-P90 value is then defined as the CHL value where W_{norm} equals 0.90 with the possibility to linearly interpolate between the two closest CHL values above and below the W_{norm} of 0.90 if this is not the case.

3. Results & Discussion

Figure 4 shows an overview of the different map products which are relevant to the generation of the CHL-P90 product. The product has been generated for the years 2003 to 2010. Only data from the years 2004 and 2005 are shown here as means of demonstration. The available number of MERIS observation per pixel and per growing season for the BCZ is shown in the first column. This illustrates the annual variations in terms of data availability due to cloud cover. Significantly more observations per pixel were available for 2004 compared to 2005. Sampling irregularities are described by the weight index WI_{KP} but are best visualized in a relative fashion using RD which is calculated from comparing the KPP_{MERIS} to KPP (Fig. 4 column 2). Red, green and blue pixels represent CHL time series where the KP is over sampled, correctly sampled and under sampled respectively compared to the KPP which equals 10.18% in this particular case. In 2004 the KP is over sampled near the coast and under sampled in the centre of the BCZ. For 2005 we mainly observe an under sampling of the KP which has the expected impact on the quality of the standard CHL-P90 product (Fig. 4 column 3) which in case of ideal sampling should be identical to the reference data (Fig. 2 left side). While for both 2004 and 2005 the same reference CHL time series were used, the standard CHL-P90 maps are significantly different as a result of different sampling frequencies. This demonstrates the need for a correction procedure as mean relative errors of up to 16% and 30% due to sampling issues were observed for 2004 and 2005 respectively (error maps not shown).

Using the weighted CHL-P90 approach, the information on sampling irregularities provided by the KPP is taken into account for each pixel separately enabling to calculate a weighing factor. The results of this approach are presented in the fourth column of figure 4 showing a closer fit of the weighted CHL-P90 product with the reference CHL-P90 map (Fig. 2 left side) compared to the standard CHL-P90 product. It can be concluded that the weighting approach works for a large number of sampling situations. Most of the over estimations are eliminated but some under estimations are still present as can be seen in the top left corner of the weighted CHL-P90 map for 2004 and in different areas for 2005. These areas are characterized by a critical unavailability of observations during the actual phytoplankton bloom. Such pixels could be identified and discarded by using a threshold as a quality control as no statistically robust CHL-P90 estimate can be made using those specific CHL time series. The artificial boundary fronts seen in these maps are present in the original (simulated) dataset defining the 5 used zones and are not generated by the weighting procedure proposed here.

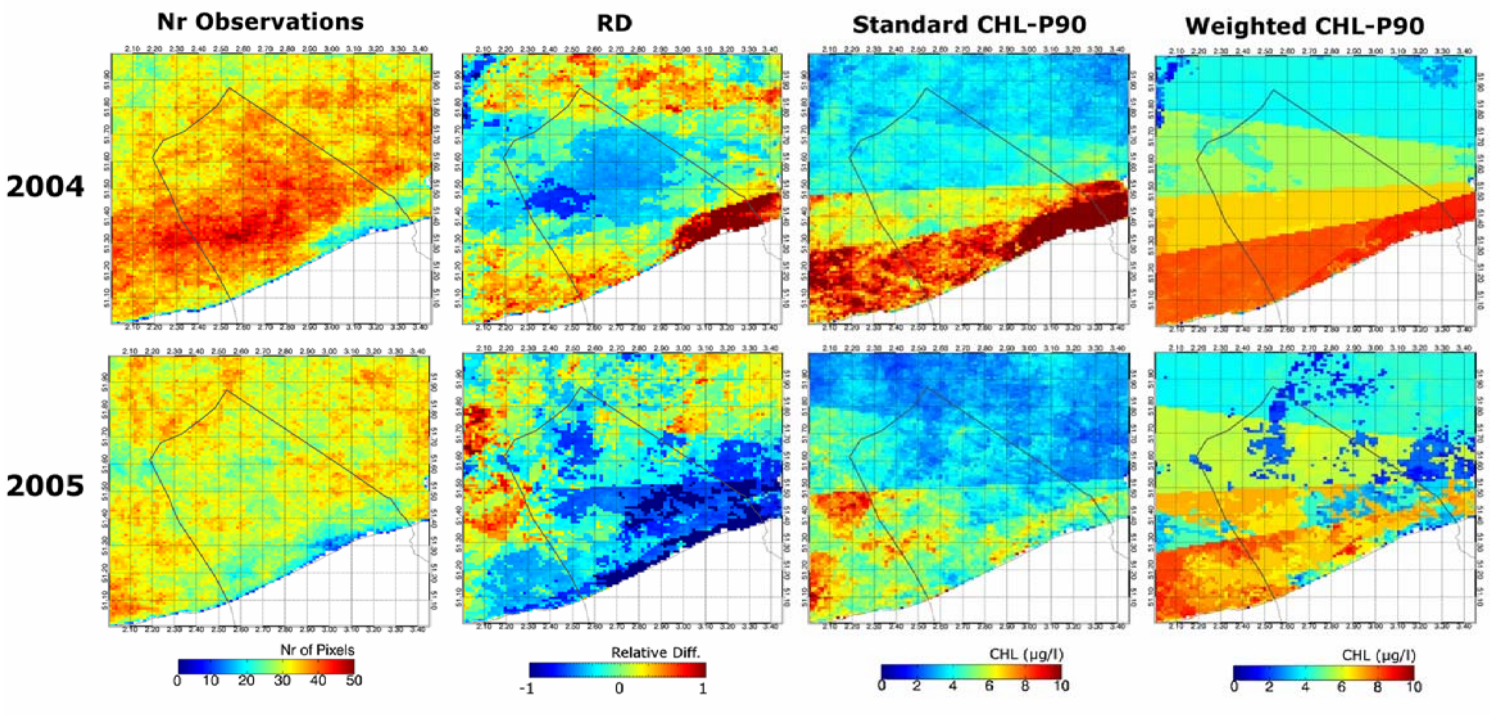

Fig.4. Overview of different map products relevant
Column 1 presents the data availability in terms of n
years. In the second column the relative differenc
irregularities considering the key period. Column 3 an

Figure 5 shows the mean relative errors of the standard and weighted CHL-P90 products which are calculated for the years 2003 to 2010 by directly comparing the estimated CHL-P90 values to the values obtained from the MIRO&CO-3D model and subsequently calculating the average for the BCZ. For the standard CHL-P90 product this error ranges from 16% to 35% with a mean error of 25.4% and for the weighted CHL-P90 product this ranges from 6% to 19% with a mean error of 9.9%. This results in a decrease in relative errors of 15.5%.

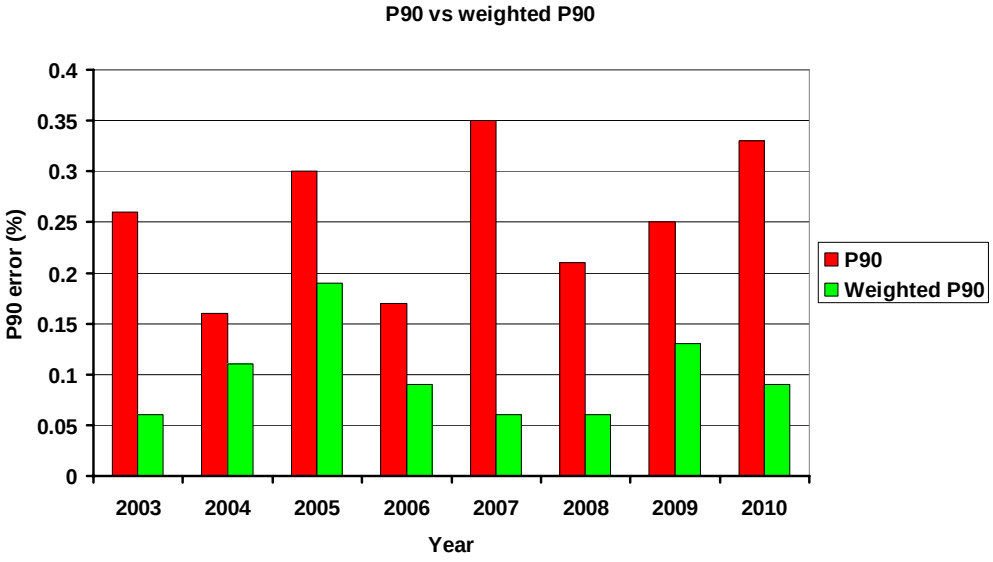

Fig. 5. Mean relative errors of the standard (red) and weighted (green) CHL-P90 products for the years 2003-2010.

4. Conclusions

With the CHL-P90 product gaining importance in describing the eutrophication state of the European seas for WFD and MSFD applications, product quality has become an essential concern of potential users. While most studies have considered only the accuracy of individual satellite CHL measurements, this study has focussed on the additional errors generated in multitemporal products. The sensitivity study performed in this work allowed more insight into the impact of sampling irregularities on the final product which should be considered carefully as significant errors were found. The proposed weighing approach provides a tool to take into account sampling issues resolving a significant part of this problem without the need for additional data except for the identification of the key period. Obviously a CHL-P90 product can only be generated if sufficient observations during the algal bloom period or key period are available.

This theoretical study was performed in a simulated environment raising the question of how the developed methodology can be transformed into an operational status. This

only requires the knowledge of the region specific key period which could be obtained from *in situ* data sets (e.g. SmartBuoys, traditional point sampling, etc.), ecological models (e.g. MIRO&CO-3D) or the Remote Sensing data itself by using pixels with well-defined time series. Next, theoretical look-up-tables need to be generated for estimated CHL-P90 errors in terms of the number of observations and KPP for different phytoplankton dynamics (i.e. bloom intensity, length of bloom, etc.). These look-up-tables can then be used to generate an error map to accompany the CHL-P90 map product. The combination of the weighted CHL-P90 map and a (theoretical) error map could then be considered as a more complete product needed by policy makers for WFD and MSFD reporting requirements.

References

Directive (2000), Directive 2000/60/EC of the European Parliament and of the council of 23 October 2000 establishing a framework for community action in the field of water policy. Official Journal 22 December L327/1, Brussels: European Commission.

Doerffer R., Schiller H. (2007) The MERIS Case 2 water algorithm. International Journal of Remote Sensing, Vol. 28, pp. 517-535.

Gohin F., Saulquin B., Oger-Jeanneret H., Lozac' h L., Lambert L., Lefebvre A., Riou P., Bruchon F. (2008) Towards a better assessment of the ecological status of coastal waters using satellite-derived chlorophyll-a concentrations.Remote Sensing of Environment, Vol. 112, Issue 8, pp. 3329-3340.

Lacroix G., Ruddick K., Ozer J., Lancelot C. (2004). Modelling the impact of the Scheldt and Rhine/Meuse plumes on the salinity distribution in Belgian waters (Southern North Sea). Journal of Sea Research, 52: 149-163.

Lacroix G., Ruddick K., Gypens N., Lancelot C. (2007). Modelling the relative impact of rivers (Scheldt/Rhine/Seine) and Channel water on the nutrient and diatoms/*Phaeocystis* distributions in Belgian waters (Southern North Sea)**.** Continental Shelf Research, Vol. 27, Issue 10-11, pp. 1422–1446.

Lancelot C., Billen G., Sournia A., Weisse T., Colijin F., Veldhuis M. J. W., Davies A., Wassman P. (1987). *Phaeocystis* blooms and nutrient enrichment in the continental coastal zones of the North Sea. Ambio. Vol. 16, Issue. 1, pp. 38-46.

Lancelot, C., Spitz, Y., Gypens, N., Ruddick, K.G., Becquevort, S., Rousseau, V., Lacroix, G., Billen, G. (2005). Modelling diatom and *Phaeocystis* blooms and nutrient cycles in the Southern Bight of the North Sea: The MIRO model. Marine Ecology Progress Series 289, pp. 63-78

Luyten, P.J., Jones, J.E., Proctor, R., Tabor, A., Tett, P., Wild-Allen, K. (1999), COHERENES documentation: a coupled hydrodynamical-ecological model for regional and shelf seas: user documentation. MUMM internal document, MUMM, Brussels, P. 903.

OSPAR (1998), OSPAR Agreement 1998-18. OSPAR Strategy to Combat Eutrophication. www.ospar.org

Park, Y., Ruddick, K., Lacroix, G. (2010), Detection of algal blooms in European waters based on satellite chlorophyll data from MERIS and MODIS. International Journal of Remote Sensing , Vol. 31, Issue 24, pp 6567-6583.

Rousseau V., Park Y., Ruddick K., Vyverman W., Parent J.-Y. & Lancelot C. (2006). Phytoplankton blooms in response to nutrient enrichment. In: Current status of Eutrophication in the Belgian Coastal Zone. Editor(s): V. Rousseau, C. Lancelot, D. Cox.

Ruddick K., Lacroix G., Lancelot C., Nechad B., Park Y., Peters S. & Van Mol B. (2008c). Optical remote sensing of the North Sea. In: Remote sensing of the European Seas. Springer-Verlag. Editor(s): V. Barale and M. Gade

Ruddick K., Park Y., Astoreca R., Borges A., Lacroix G., Lancelot C., Rousseau V. (2008b), Applications of the MERIS algal2 product in Belgian waters. Proceedings of the 2nd MERIS-(A)ATSR workshop, 2008, Frascati, Italy. ESA Special Publication SP-666. 4p.

Ruddick K., Park Y., Astoreca R., Neukermans G., Van Mol B. (2008a), Validation of MERIS water products in the Southern North Sea: 2002-2008. Proceedings of the 2nd MERIS-(A)ATSR workshop, 2008, Frascati, Italy. ESA Special Publication SP-666. 8p.

Sirjacobs D., Alvera-Azcarate A., Barth A., Lacroix G., Park Y., Nechad B., Ruddick K., Beckers J.-M. (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, Vol. 65, pp. 114–130.

Sorensen K, Severinsen GG, Aertebjerg G, Barale V, Schiller C (2002), Remote Sensing's contribution to evaluating eutrophication in marine and coastal waters. European Environment Agency, 41p.