Deltares

Enabling Delta Life

Second opinion on the JPI Oceans intercalibration of chlorophyll for NEA 3/4

Deltares

Second opinion on the JPI Oceans intercalibration of chlorophyll for NEA 3/4

dr. T.C. Prins
dr. S. Birk (University of Duisburg-Essen)

11200888-000

Deltares

Title
Second opinion on the JPI Oceans intercalibration of chlorophyll for NEA 3/4

Client	Project	Reference	Pages
RWS WVL	11200888-000	11200888-000-ZKS-0001	7

Keywords
Water Framework Directive, chlorophyll-a, coastal waters, intercalibration

Summary
In 2016, JPI Oceans carried out an analysis for the intercalibration of chlorophyll-a in the coastal waters of NEA GIG type NEA 3/4. This type is shared by the Netherlands and Germany. The analysis was carried out using data sets from monitoring in Dutch and German coastal waters, and an extensive effort was done to make optimal use of the available data.
The results of JPI Oceans and the review in this report emphasize the very heterogeneous character of the coastal waters in NEA 3/4. The water bodies in NEA 3/4 are subject to many pressures, including eutrophication among others, and to a wide range in abiotic conditions. The concentrations of chlorophyll-a are thus influenced by multiple factors. The analysis of JPI Oceans has focused on the statistical relation between chlorophyll-a and total nitrogen, but has ignored the fact that there are multiple pressures and abiotic and biotic factors influencing chlorophyll. The statistical relations between chlorophyll-a and total nitrogen that were used by JPI Oceans as basis for the intercalibration were only for two out of nineteen monitoring stations statistically significant and show that using total nitrogen as only pressure does not provide a sufficient basis for the benchmarking procedure that was applied. We cannot recommend comparing the national status boundaries on this basis.

References
Zaaknummer 31078379

Version	Date	Author	Initials	Review	Initials	Approval	Initials
	March 2017	dr. T.C. Prins		dr.ir. A.D. Buijse		F.M.J. Hoozemans MSc	
		dr. S. Birk (University of Duisburg-Essen)					

State
final

Deltares

Contents

1 Introduction

The main objective of the Water Framework Directive (WFD) intercalibration procedure is to set harmonised ecological quality criteria to meet the protection and restoration targets for all surface waters throughout the European Union. Therefore one aim is to get comparable assessment results of different Member States and a harmonized classification based on Ecological Quality Ratios (EQR).
JPI Oceans was contracted by several MSs to carry out an intercalibration in the NEA-GIG. Initially the intercalibration process was started with all distinguished coastal water types, to ensure intercomparability between all types in the NEA-GIG. As NEA-GIG type NEA 3/4 deviated from the other types, a separate report on the intercalibration of Chlorophyll for this type was finished in December 2016 (Bonne & Desmit 2016).
The NEA-GIG type NEA 3/4 consists of exposed polyhaline waters in the North Sea and Wadden Sea type polyhaline waters. This NEA GIG type is shared by the Netherlands and Germany.

The Netherlands and Germany did not reach an agreement on the procedure of intercalibration that is suited for this area [NEA 3/4]. Therefore both Member States separately formulated their views in reaction to the JPI Oceans report. Germany supported the outcomes (Grage *et al.* 2016), whereas the Netherlands had several comments, questions and a request for additional information (van den Berg & Ruiter 2016). The Joint Research Centre (JRC) organized a review of the JPI Oceans report (Anonymous 2016), but in the review some of the major concerns and questions from the Netherlands were not taken into consideration.
The Netherlands have raised their concerns about the intercalibration of Chlorophyll in NEA3/4 at both the meeting of ECOSTAT and at the meeting of the Article 21 Committee. It was agreed that the Netherlands could carry out a second opinion on the review of NEA 3/4 Chlorophyll-a.

This document gives a second opinion on the JPI Oceans intercalibration report by Bonne & Desmit (2016), looking at the scientific approach in the data analysis (Chapter 2) and the extent in which the intercalibration follows the requirements (Chapter 3).

2 Review of the data analysis

Common dataset

For the intercalibration, a large common dataset was constructed based on monitoring data provided by Germany and the Netherlands. This common dataset had a good spatial coverage of the NEA 3/4 type and covered a period of ca. 15 years (2000-2015) of data of salinity, concentrations of nitrogen (N) and phosphorus (P) and chlorophyll-a (CHLa). If more historic data (i.e. years before 2000) had been added, this would have included a longer period with a larger difference in nutrient loads to coastal waters. This could have resulted in a larger range for the pressure window of nutrient concentrations. However, it is uncertain that this would have improved the statistical relations (see Prins *et al.* 2017).
Conclusion: the common dataset had a sufficient coverage of the spatial diversity in the coastal waters of NEA 3/4, and covered a sufficient number of years to account for interannual variability.

Pressure-response relations

In the statistical analysis of the pressure-response relations, several assumptions were made:

- Only the relation of chlorophyll with nitrogen was considered
- This ignores the fact that nitrogen concentrations are strongly correlated with phosphorus concentrations
    - o in space as they share the same source (river discharges) and both are highly correlated with salinity
    - o in time as both riverine P- and N-loadings have decreased since 1990
- Consequently, on the basis of monitoring data it is not possible to decide whether N or P is determining CHLa concentrations
- In addition, biogeographical differences between water bodies/monitoring stations are not only caused by differences in N or P, but also by other factors determining chlorophyll-a levels such as light limitation, mixing depth, flushing, grazing
- The relative importance of the nutrients N and P and the other factors differs between waterbodies/monitoring stations

Conclusion: while statistical correlations between chlorophyll-a and nitrogen may be found, these correlations are not necessarily indicative of a causal relation because of the confounding effect of the other factors that have an impact on phytoplankton growth

Quality of correlations

Table 5 in Bonne & Desmit (2016) summarises the attempts to find pressure-response relations, using various metrics for chlorophyll-a and nitrogen concentrations. Generally the results were very poor. The choice was made to use “spring TN” (mean for January-April) and “summer CHLa” (mean for May-September) in the further benchmarking approach.

- No attention was paid to covariation with other factors and the potential effect of covariation on the quality of the CHLa-TN regressions
- Figure 7 in Bonne & Desmit (2016) shows the results of the regression results of summer CHLa versus spring TN:
    - o For only two stations a statistically significant correlation was found (Noordwijk and Huibertgat oost)
- Figure 9 in Bonne & Desmit (2016) shows all water bodies/stations combined: the conclusion is drawn that CHL-a reaches a maximum of 45 µg/l at a TN concentration around 100 µM
    - o This conclusion is questionable: it is the effect of different stations at different points in the TN range, each showing their own specific pattern as shown in

Figure 7, with highest CHLa concentrations at the stations in the Dutch Wadden Sea east (Dantziggat and Zoutkamperlaag Zeegat).
   - One of the few relationships with a statistically significant regression (Huibertgat oost) shows a linear increase of CHLa over the range from 50-175 µM TN and no sigh of levelling off at 100 µM TN.

Conclusion: the regressions of CHLa on TN show very poor results and do not support the conclusions on CHLa-TN relations and the levelling off of CHLa concentrations above TN concentrations of 100 µM.

Alternative benchmarking; construction of common metric

In the approach for alternative benchmarking, a pressure window was selected for TN concentrations between 75-125 µM TN

- It is questionable that CHLa concentrations reach a maximum around this pressure window (see above)
- Water bodies/stations are divided in three categories (green: positive correlation CHLa-TN; blue: for stations without a correlation, data points within the pressure window; red: for stations without a correlation, average of data points at "observed plateau"). This part of the analysis lacks transparency, as it is not clear in the report which water bodies /stations belong to which category. Note that only two stations in Figure 7 in Bonne & Desmit (2016) show a significant correlation.
- Table 22 in Bonne & Desmit (2016) indicates for which stations summer mean in the alternative benchmark window was calculated using linear regression (Walcheren, Boomkensdiep, Terschelling), but the regressions for those stations (Figure 7) were not statistically significant
- Figure 12b in Bonne & Desmit (2016) with the real data shows clearly that the distinction between the three groups of stations (green, blue, red) suggested in their Figure 12a is not valid

Conclusion: the construction of the alternative benchmark window is not supported by the underlying data

Alternative benchmarking; calculation of benchmark standardized reference

The calculation of the benchmark standardized reference values is based on the differences (ratios) found in the alternative benchmark window

- This implicitly assumes linear relations between CHLa and TN for all water bodies/stations, which is not supported by the data (Fig.7 in Bonne & Desmit (2016))

Conclusion: the construction of the benchmark standardized reference is not supported by the underlying data

Overall confidence in results of benchmarking

- The final result of the calculation of the benchmark standardized references and H/G and G/M boundaries is based on many assumptions that are not substantiated by the data (see above)
- In the data analysis many steps have been taken using calculations and regression results which were not always statistically significant, and even in the case of significant correlations, all have a range of uncertainty. Some of these steps are, e.g. CHLa-TN correlation, calculation of common metric, correlation of CHLa summer mean - CHLa 90 percentile to recalculate references.
- The uncertainty in these calculation steps is not taken into account but is quite large and adds uncertainty to the suggested new class boundaries

Conclusion: The confidence in the results of the data analysis is low

3 Review of the *WFD intercalibration of biological quality elements for coastal water bodies – North-East Atlantic Geographical Intercalibration Group (NEA-GIG) NEA 3/4 Wadden Sea – North Sea in coherence with NEA 1/26b & 1/26c – Chlorophyll a*

3.1 Introduction

The intercalibration report of Bonne & Desmit (2016) represents a valuable contribution to the quest of harmonising the national classifications of coastal phytoplankton in IC type NEA 3/4. The procedural steps conducted by the authors are complex and not easy to grasp, this is why we repeat these steps in the following. The review of this intercalibration exercise addresses specific issues along the sequence of the performed work-steps.

- Step 1. Choosing IC Option 2 and selecting ‘summer mean Chlorophyll a concentration’ as intercalibration common metric (ICM).
- Step 2. Establishing pressure-response relationships between spring Total Nitrogen (TN) and ICM.
- Step 3. *Excursus*: Exploring biogeographical differences by investigating into Chlorophyll a dynamics using cluster/ordination analysis.
- Step 4. Defining ‘window of pressure’ for alternative benchmarking to identify the average ICM value per sampling station within this ‘window of pressure’.
- Step 5. Benchmark standardising the ICM per sampling station by dividing average ICM (converted into station-specific reference concentrations) by observed ICM (summer average over six years).
- Step 6. Regressing national EQR (i.e. P90 seasonal mean Chlorophyll a concentration) against benchmark standardised ICM.
- Step 7. Boundary comparison.

3.2 Opening remarks

At first sight, the setting of this intercalibration exercise looks straightforward: Two national classifications are compared that use the same assessment metric (i.e. P90 seasonal mean Chlorophyll a concentration). Despite the good reasons for Bonne & Desmit (2016) not to refer to IC Option 1 (see next section), this feature nevertheless allows for informative first insights into similarities and differences of the two national classifications. While the definition of ecological status boundaries does not differ, the discrepancy between near-natural reference conditions is very obvious. Dutch water bodies feature reference Chlorophyll a concentrations up to 4.5 µg/l higher than German water bodies. This accounts for almost a full status class in difference.

This discrepancy reflects (i) methodological differences in terms of how Chlorophyll a is measured in both countries (see next section), (ii) biogeographical differences due to natural water body features such as turbidity, grazing activity and background nutrient concentrations, and/or (iii) systematic differences in the models to derive Chlorophyll a reference conditions. While (i) and (ii) represent important factors to be considered the intercalibration analysis, any comparison of national class boundaries would be flawed when based on (iii).

The reference values were indeed modelled within two different projects (DE: Topcu *et al.* 2011; NL: Baptist & Jagtman 1997). Without judging the quality and related uncertainties of these works, it is clear that given the dynamic (and partly stochastic) nature of coastal water bodies, two discrete models of a similar area may come up with different model results (e.g. depending on input data and model design). In such a case, reference definition (and subsequent status boundary harmonisation) of adjacent water bodies would benefit from selecting a single model and agreeing on the same reference conditions. Such an agreement was obviously possible for the shared water body Ems-Dollard, but not envisaged for the other national water bodies.

3.3 Choice of intercalibration option

The NEA 3/4 Wadden Sea – North Sea Phytoplankton intercalibration exercise compares the good status boundaries of the same assessment method (i.e. P90 Chlorophyll a seasonal mean) between DE and NL. This setting would suggest an IC Option 1 (same method), but Bonne & Desmit (2016) chose to perform an IC Option 2 (using a common metric) for good reasons: (i) Methodological differences between DE and NL in analysing the *Chlorophyll a* pigments (spectrophotometry versus HPLC), (ii) the different data aggregation used in the analysis (summer instead of seasonal mean) improving the pressure-response relationship, and (iii) the need to average *Chlorophyll a* concentrations per sampling station to obtain an alternative benchmark. This resulted in selecting the 'summer mean *Chlorophyll a* concentration' as the ICM.

Compared to IC Option 1, Option 2 has the disadvantage of introducing uncertainty when translating the national status boundaries into the ICM scale using regression analysis. Depending on the quality of the regression (r^2, slope), this translation may account for significant differences in the boundary comparison. Bonne & Desmit (2016) address this issue in Chapter 7.2 and highlight major difficulties in establishing convincing regressions. They perform various steps towards most appropriate relationships, e.g. aggregating sampling stations and selecting P90 Chlorophyll a values. However, we do not fully understand what the authors have done exactly and why. This lack of clarity may give rise to reservations regarding the intercalibration outcomes.

3.4 Benchmark standardisation

Bonne & Desmit (2016) demonstrated water body-specific phytoplankton dynamics referring to the patterns of variability in Chlorophyll a long-time series (see Chapter 6.3 therein). Although their preparatory analysis of the plankton variability was also influenced by anthropogenic pressures (i.e. nutrient enrichment), distinct patterns were discernible that point at biogeographical differences between the sampling stations. To intercalibrate the national classifications, this analysis was repeated by referring to data corresponding to a limited 'window of pressure' (TN concentrations ranging from 75 to 125 µmol/l). This is an important procedural step in the intercalibration analysis (called 'alternative benchmarking'), required when the national reference conditions are defined differently and stations in near-natural reference conditions do not exist.

Alternative benchmarking requires significant pressure-response relationships to be demonstrated between an anthropogenic pressure (here: TN) and the biological response (here: Chlorophyll a) (see Böhmer *et al.* 2016). Key is to identify differences in the biological conditions that are not influenced by the pressure, but by other factors such as biogeographical or methodological differences. Alternative benchmarking 'controls' for the pressure effect on the biology. In this way, the effects of other influences on the biological data (i.e. residual variance) are revealed to the analyst. But if a significant pressure-impact relationship cannot be identified, the pressure effect cannot be controlled for. Then any benchmark standardisation may become baseless.

According to Figure 7 in Bonne & Desmit (2016), two out of 19 TN–Chlorophyll a relationships are significantly correlated. All other relationships are not significant. This means that the variance of TN concentration cannot explain the variance of Chlorophyll-a for most sampling stations. From a strict analytical perspective, selecting TN to perform benchmark standardisation is thus only acceptable for the two stations with significant pressure-response relationships (i.e. Noordwijk and Hulbertgat Oost). But judging from the plots in Figure 7, these two relationships are even significant within the 'window of pressure', revealing that the pressure level still influences the variance of Chlorophyll a for the benchmark samples. Here, continuous benchmarking (Böhmer *et al.* 2016) represents a viable alternative.

In some completed intercalibration exercises, benchmark standardisation has been applied to datasets showing very small pressure gradients and, for this reason, lacking significant pressure-response relationships (e.g. invertebrates in very large rivers, see Birk *et al.* 2016). This supports the approach of Bonne & Desmit (2016) to consider sampling stations with small TN gradients in benchmark standardisation (e.g. Dantziggat, Goeree)[1]. However, various stations feature long TN gradients but no significant relationships. In these cases, it is questionable whether the selected benchmarking approach is appropriate. We doubt that referring to the explanation that the Chlorophyll a concentrations have "reached a plateau over increasing TN concentrations" (Bonne & Desmit 2016, pages 42ff.) provides adequate reason to perform benchmarking. It seems that in such cases TN is not a driving factor of the Chlorophyll a variance. Hence, it cannot be used for benchmarking (as the only pressure factor).

3.5 Conclusions

The intercalibration report of Bonne & Desmit (2016) describes an ambitious attempt to intercalibrate the national classifications using Chlorophyll a in type NEA 3/4. The approach is inventive and cleverly devised by the authors – especially the analyses to group the water bodies according to their phytoplankton biomass dynamics, demonstrating the heterogeneous biogeographical conditions in this area, are forward-looking.

However, the key-aspect of benchmark standardisation reveals a central shortcoming: The lack of significant TN–Chlorophyll a relationships (even for stations with long TN gradients) renders the definition of alternative benchmarks highly uncertain. Benchmark standardisation is an indispensible element in this intercalibration exercise, but it currently lacks sound implementation. Although the IC outcomes may confirm the seemingly obvious differences in classifications between DE and NL, we cannot recommend comparing the national status boundaries on this basis. Additional concern is raised by the lack of clarity on how the regressions were established to translate the national EQRs into ICM units.

The report of Bonne & Desmit (2016) represents an important contribution towards the harmonisation of ecological status classifications in coastal waters using Chlorophyll a. Against this background, we suggest further efforts to intercalibrate this phytoplankton component in the near future, either based on an extended monitoring dataset including multiple regression analyses with more explanatory variables (e.g. phosphorus concentration, turbidity, grazing activity), or joint modelling exercises to devise harmonised reference conditions, covering all relevant German and Dutch water bodies.

[1] *But continuous benchmarking would be more appropriate than defining a 'window of pressure'.*

4 References

Anonymous (2016). Evaluation Report on the Chla harmonization. Coastal Waters NEA 3/4 for the Joint Research Centre - European Commission, Institute for Environment and Sustainability. 2 pp.

Baptist, H.J.M. & E. Jagtman (1997). Watersysteemverkenningen 1996. De AMOEBES van de zoute wateren. RIKZ, Den Haag, RIKZ 97.027, 149 pp.

Birk, S., J. Böhmer & F. Schöll (2016). XGIG Large River Intercalibration Exercise – Milestone 6 Report. Intercalibrating the national classifications of ecological status for very large rivers in Europe. Biological Quality Element: Benthic Invertebrates. Version 2, November 2016.

Böhmer, J., S. Birk, N. Willby, G. Phillips & S. Poikane (2016). Approaches for metric standardisation in intercalibration: Reference benchmarking, alternative benchmarking and continuous benchmarking in comparison. In: Birk, S., J. Böhmer & F. Schöll. XGIG Large River Intercalibration Exercise – Milestone 6.

Bonne, W. & X. Desmit (2016). WFD Intercalibration of biological quality elements for coastal water bodies. North-East Atlantic Geographical Intercalibration Group (NEA-GIG) NEA 3/4 Wadden Sea – North Sea in coherence with NEA 1/26b & 1/26c (Dutch and German coastal zone North Sea and Wadden Sea). Chlorophyll a JPI Oceans, Report submitted to WG Ecostat, 94 pp.

Grage, A., G. Petri, J. Witt & J. Knaack (2016). German comments to the technical report and for submission to JRC – Review Panel. WFD intercalibration NEA 3/4 Chlorophyll a (QE phytoplankton) – MS Netherlands and Germany – Third phase of Intercalibration.

Prins, T.C., T.A. Troost & S. Birk (2017). Phytoplankton in NEA 3/4 coastal waters. WFD Class boundary values for chlorophyll-a. Deltares, Delft, 11200888-000-ZKS-0002, 28 pp.

Topcu, D., H. Behrendt, U. Brockmann & U. Claussen (2011). Natural background concentrations of nutrients in the German Bight area (North Sea). *Environmental Monitoring and Assessment* 174: 361-388.

van den Berg, M. & H. Ruiter (2016). NL recommendations for the intercalibration results of chlorophyll-a in NEA3/4 (Waddensea-type). Rijkswaterstaat.