Geosci. Model Dev. Discuss., 5, 2311–2345, 2012
www.geosci-model-dev-discuss.net/5/2311/2012/
doi:10.5194/gmdd-5-2311-2012

Geoscientific Model Development Discussions

How should sparse in situ measurements be compared to continuous model data?

L. de Mora, M. Butenschön, and J. I. Allen

Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth, PL1 3DH, UK

Received: 18 July 2012 – Accepted: 13 August 2012 – Published: 20 August 2012

Correspondence to: L. de Mora (ledm@pml.ac.uk)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

This work demonstrates the importance of an adequate method to sub-sample model results when comparing with in situ measurements. A test of model skill was performed by comparing a multi-decadal hindcast against a sparse, unevenly distributed historic in situ dataset. The comparison was performed using a point-to-point method. The point-to-point method masked out all hindcast cells that did not have a corresponding in situ measurement in order to compare each in situ measurement against its most similar cell from the model. The application of the point-to-point method showed that the model was successful at reproducing many inter-annual trends. Furthermore, this success was not immediately apparent using the previous comparison methods, which compared model and measurements aggregated to regional averages. Time series, data density and target diagrams were employed to illustrate the impact of switching from the previous method to the point-to-point method. The comparison based on regional averages gave significantly different and sometimes contradicting results that could lead to erroneous conclusions on the model performance. We therefore recommend that researchers take into account for the limitations of the in situ datasets, process the model to resemble the data as much as possible, and we advocate greater transparency in the publication of methodology.

1 Introduction

Numerical models are now used extensively in earth, climate and ocean sciences. Furthermore, numerical models are frequently used to inform policy decisions. Both policy decision and fundamental science require models to have demonstrable quality. However, the assessment of how well a model captures reality is an ongoing challenge of marine ecosystem model development.

As discussed in a meta-analysis (Arhonditsis and Brett, 2004), many models have been validated with qualitative methods exclusively. Qualitative methods are usually

GMDD
5, 2311–2345, 2012

straightforward to interpret, allowing for a simple, fast, subjective judgement of whether the model appears to be representative of the measurements. Unfortunately, a model may seem to recreate emergent properties and well known large scale features of the ecosystem, yet struggle to reproduce the historic data of a hindcast, for instance Doney et al. (2009). For these reasons, it is crucial to validate models using a variety of objective statistical tests.

In marine ecosystem modelling, quantitative descriptions of models are often based on pattern statistics, and other univariate indices, for instance in Stow et al. (2009). Pattern statistics form the axes of the popular Taylor (Taylor, 2001) and target diagrams (Jolliff et al., 2009). These univariate indices generally require equal binning of (i.e. the same number of) both model and measurement data, but the methodology used to achieve equal binning methods can introduce sampling bias. Typically, the equal binning condition is met by interpolating the data to cover the model domain. Interpolations fill under-sampled regions with information from well-sampled regions; an ideal solution for cloud coverage in satellite data, for instance Edwards et al. (2012). However, interpolating of sparse, uneven and widely distributed in situ measurements can amplify the effects of sampling bias (Robeson, 1994). This is especially true for measurements with high spatial and temporal variability. Alternatively, the equal binning condition can be achieved by taking the mean of both the model and in situ data over a sufficiently large region, as in Lewis et al. (2006). Sampling bias is introduced when the mean of measurements of a very small subset of the ocean is compared against the mean of a very large volume in the model. This problem compounded when ad hoc sampling is biased toward coastal sites that are both accessible and convenient.

In this paper, a point-to-point method is outlined to validate a marine ecosystem model hindcast in using historic in situ measurements. The point-to-point method does not introduce new uncertainties via interpolation and attempts to reduce the impact of sampling bias introduced via historic ad hoc sampling. While it may seem obvious to process the model exactly as the in situ data was produced, it seems to have rarely been done, or published with a lack of transparency. Furthermore, to the best

GMDD
5, 2311–2345, 2012

of the authors knowledge, this is the first direct comparison of the matched against unmatched methodologies.

The model used in this study is the European Regional Seas Ecosystem Model (ERSEM) coupled with the Proudman Oceanographic Laboratory Coastal Ocean Modelling System (POLCOMS) as described in Sect. 2. The in situ data is the North Sea CTD and low resolution bottle data from the International Council of the Sea (ICES) database, described in Sect. 3. A full description of the methodology of the point-to-point matching and the linear regression is in Sect. 4. The agreement of the in situ and model data, and a comparison of matched and unmatched methods are shown in Sect. 5.

2 Model

The European Regional Seas Ecosystem Model (ERSEM) is a lower-trophic level biogeochemical cycling model that uses the functional-group approach (Blackford et al., 2004). The Carbon, Nitrogen, Oxygen, Phosphorus, and Silicon cycles are explicitly resolved and the food web is composed of four phytoplankton, three zooplankton and one bacterial functional types.

ERSEM was coupled with the Proudman Oceanographic Laboratory Coastal-Ocean Modelling System (POLCOMS) hydrodynamic model (Holt et al., 2001). POLCOMS is a baroclinic three dimensional model that includes both the deep ocean and the continental shelf.

This study used POLCOMS-ERSEM in the Atlantic Margin Model (AMM) domain, which covers the area between 40° N to 65° N and 20° W to 13° E. The domain has a resolution of 1/9° by 1/6°, which equates to 12 km with a baroclinic time-step of 15 min. In terms of depth, the s-coordinates system is used, consisting of 40 wet depth layers of varying thickness. The model was run for a 40 yr hindcast between 1965 and 2004, with each state variable recorded as the daily and monthly mean. A full description of POLCOMS-ERSEM in the AMM domain is available in (Holt et al., 2012).

GMDD
5, 2311–2345, 2012

The atmospheric conditions were taken from the ERA 40 reanalysis (Uppala et al., 2005) and the open ocean boundary conditions were taken from the global model, ORCA025. The nutrients and oxygen forcing were taken from World Ocean Atlas Data (Garcia and Levitus, 2010b,a).

The freshwater fluxes in the AMM consist of 250 rivers from the Global River Discharge Database (Vörösmarty et al., 2000) and from the Centre for Ecology and Hydrology. The river nutrient content is based on measured data, as in Holt et al. (2012) and Young and Holt (2007).

3 In situ data

The in situ data used in this study were taken from the the International Council for the Exploration of the Sea (ICES) EcoSystemData Online Warehouse (ICES, 2009). Five datasets were used: nitrates, phosphates, chlorophyll *a*, temperature and salinity.

This study aimed to have good spatial and temporal coverage and consistent data quality. For these reasons, only bottle and low-resolution CTD data was used.

The region under investigation was the North Sea as defined by the ICES subdivision, IV (FAO, 2008). The North Sea ICES region is defined as the Sea between 62° N and 51° N, and 4° W. The eastern boundary of the North Sea domain passes north from Agger Tange, Jutland, Denmark, to 57° N, west to 8° E, then north to 57° 30′ N, then west to 7° E, then north to the coast of Norway.

The database was provided in a comma-separated-variable format, and contained a few data quality anomalies, such as repeated data, which were addressed during processing. The repeated measurements accounted for typically 10–20 % of the database. Repeated data were identified by searching for measurements with identical time-stamp, longitude, latitude, depth and value. In some cases, data were recorded at depths much greater than the model bathymetry at the same point, so measurements with a depth greater than 5 m below the model bathymetry at the same point were ignored. The chlorophyll dataset contained a large proportion of measurements

GMDD
5, 2311–2345, 2012

with a value of exactly 0.1, even at depths below 1 km. As no chlorophyll is expected at large depths, this was interpreted to be the detection limit of the database and all chlorophyll measurements below 100 m were removed from the study.

4 Methods

This section describes the methods that were used to test the compatibility of the model and the in situ data. The first subsection describes how the point-to-point matching was applied and the second describes the linear regression fit.

4.1 Point-to-point matching

The model and in situ datasets are intrinsically dissimilar. The model dataset is formed from a grid of continuous, evenly distributed, time-averaged cells of approximately 12 km by 12 km. The in situ dataset is a series of sporadic, unevenly distributed, instantaneous measurements from a CTD or the mean of the contents of a sampling bottle. Furthermore, the in situ measurements tended to occur in times and places which were readily accessible, convenient or well funded. These places and times do not match the uniform grid used by the model. The role of point-to-point matching was to reduce the impact of sampling bias when comparing these two distinct kinds of data.

The first step of the point-to-point matching process was performing a snap-to-grid: the in situ data was forced into the same grid as the model. Here, the full depth, daily mean, four dimensional AMM domain grid was used. This grid had 40 depth layers, 198 longitude bins and 224 latitude bins per day. In the case where multiple in situ measurements fell into the same grid cell, the mean of those measurements was used. Otherwise, the same model point would have been sampled multiple times, thus introducing a new source of sampling bias.

GMDD
5, 2311–2345, 2012

Finally, all model pixels that did not have a corresponding in situ measurement were masked out and vice versa. In this way, no unpaired model or in situ data were used in the linear regression.

Once the matching was performed, both datasets were studied under three time granularities: full, seasonal, and annual. The full granularity compared every matched point in the given region, the seasonal granularities were the mean of the data in three month blocks, and annual was the annual mean of the data. As the AMM is a Northern Hemisphere domain, the Winter season was January, February and March, the Spring was April, May and June, the Summer was July, August and September and the Autumn was October, November and December. The mean of the entire region is also shown in the time series plots. These values were calculated the monthly mean of the variable divided by the mixed layer depth.

4.2 Linear regression

The relationship between data and measurements was plotted with the model on the x-axis and the in situ data on the y-axis, then fitted to a straight line using a least-squares linear regression. This technique minimises the sum of the square of the residual; the residual is the difference between a model value and the corresponding in situ measurement. The five output parameters of the regression were: the y-axis intersect, $\hat{\beta}_0$, the slope of the fit, $\hat{\beta}_1$, the standard error, ϵ, the correlation coefficient, R and the two tailed probability, P. The best possible outcome of linear regression, corresponding to a perfect model, would be a line of slope unity through the origin with no standard error.

5 Results

The results of the linear regression fits are shown in Tables 1 and 2. These tables consists of five columns: the full point-to-point linear regression, the Winter, Spring

GMDD
5, 2311–2345, 2012

and Summer mean, and the annual mean linear regression. Each section of these tables holds the linear regression for each of the datasets: temperature (T), salinity (Sal.), nitrates (N_3), phosphates (P_4) and chlorophyll (Chl.). The plots in this section are visualisations of some of the data used to make Tables 1 and 2.

5.1 Data density

Figures 1, 2 and 3 are two dimensional data density plots for the full time granularity temperature, salinity and nitrates matched datasets, respectively. The density plots show the matched model data on the x-axis and the in situ data on the y-axis. The shading is scaled logarithmically such that darker hue indicates higher data density. The best fit linear regression straight line is shown on each figure and the parameters of these fits are held in the "full" columns of Tables 1 and 2. The linear regression fit was calculated using a non-logarithmic scale. These figures also have a dotted black line that represents the line of slope unity that passes through the origin. The high density regions of these figures show that the model performed extremely well at reproducing the in situ salinity and temperature measurements. Temperature was especially well reproduced in the model: even with over one million matched pairs of model and in situ data, a correlation of $R > 0.9$ is observed.

The sparsely populated off-diagonal regions of Figs. 2 and 3 indicate that the model did not successfully reproduce all of the in situ data. The model tended to overestimated the extreme in situ measurements of salinity and underestimated the extreme in situ values of nitrates. In both cases, the model predicted a value less extreme than the outlying in situ measurement. This discrepancy can be explained as an effect of the high variability in salinity and nitrates in the well sampled coastal regions against the relatively low spatial resolution of the model in these regions. In addition, the in situ data was of instantaneous character, while the model data was a daily average, further enhancing the in situ measurements variability.

GMDD
5, 2311–2345, 2012

5.2 Time series

The time series plots are shown in Figs. 4–9. These figures each contain three curves: the matched model data (black line), the in situ data (dotted line), and the mixed layer depth-average of the entire model region (grey area). The entire model region plots were included to estimate whether the matched and in situ variation correspond to overall trends, or sampling biases. In all cases, the inter-annual variability of the mean of the entire region is smaller than that of the matched model data and the in situ measurements, especially in the case of nitrates, phosphates and chlorophyll in Figs. 6–9.

The results of the linear regressions of Figs. 4–9 are shown in Tables 3 and 4. This table has two columns: the results of the linear regression of the entire region against the in situ, and the linear regression of the matched data against the in situ data. In all cases shown here, the matching resulted in a higher correlation coefficient, decreased P-value, the y-intersect closer to zero. In all cases shown except Summer chlorophyll, the matching resulting in a slope closer to unity than the entire region linear regression. The fit for each dataset is discussed in more detail below.

Figure 4 is a plot of the annual mean North Sea temperature. The mean of the entire region plot, the grey area, is the depth averaged temperature, but the matched data and in situ measurements may be from any depth. For this reason, the entire region mean model data was consistently larger than the in situ and the matched data. This shift is also visible in the difference in the slope and y-intersect, $\hat{\beta}_1$ and $\hat{\beta}_0$, in Table 3. As the model temperature is forced using reanalysis based on aggregated observational data, it is not surprising that the temperature section of Table 1 shows a strong correlation between the model and the in situ data. However, the forcing dataset, ERA 40 (Uppala et al., 2005), is a meteorological sea surface temperature dataset, where as the in situ measurements and hence the matched model data may occur at any depth. As such, the success of the model is at least partially due to its own merit, instead of the similarities between the forcing and in situ datasets.

GMDD
5, 2311–2345, 2012

Figure 5 is a plot of the annual mean North Sea salinity. The linear regression results associated with this plot is shown in the "annual" column of the salinity section of Table 1. All the time granularities show a strong correlation ($R > 0.75$) between the matched model and the in situ salinity. The entire region mean model data is consistently lower than the in situ and the matched data, but displayed some skill in reproducing the overall trend. The matched model data here indicate that the model reproduced an arbitrary set of in situ measurements with moderate success. This success allows some confidence that the mean state of the model salinity is a fair representation of the mean state of the system salinity.

The mean Winter North Sea nitrates are shown in Fig. 6 and the mean Winter North Sea phosphates are shown in Fig. 7. These plots shows that the model had significant skill in reproducing the in situ nitrate and phosphate measurements, but only once unpaired model cells were masked. The Winter nitrates linear regression fit was consistent with a line of unity slope, and had a correlation coefficient of $R = 0.8191$. This correlation was not present in the unmatched depth-averaged model nitrates. The other columns of the nitrates section of Table 1 indicate that the inter-annual variability of in situ nitrates were well reproduced with all time granularities. The Winter phosphates linear regression fit was consistent with a line of unity slope and a null intersect. However, this skill was not present in the Spring and Summer months of the model.

Figures 6 and 7 both show a large peak in 1983. In the Winter of 1983, almost all North Sea nitrate and phosphate measurements in the ICES database were taken in coastal environments. The peaks are also present in the matched model data, but not in the mean of the entire region. The presence of these peaks in both model and measurement suggests that the bulk of the variability of in situ nitrates and phosphates is due to uneven coverage, rather than inter-annual variability. Due to the incongruities of historic in situ data such as these peaks, model validators should be extremely cautious to ensure that their validation compares like-datasets to each other.

Three time series figures are shown for the North Sea chlorophyll: Fig. 8 shows the mean Spring chlorophyll, Fig. 9 shows the mean Summer chlorophyll and Fig. 10 show

GMDD
5, 2311–2345, 2012

the total number of chlorophyll measurements per year grouped into high salinity and low salinity categories. The offshore high salinity region cut off of 34.5 PSU was taken from (OSPAR Commission, 2008). This figure was made by matching up the chlorophyll and salinity ICES datasets, but this process is not 100 % efficient. As such, there are years where no chlorophyll measurement has a corresponding salinity measurement. A large increase in the mean value of the in situ chlorophyll but not in the model chlorophyll can be seen after 1993 in the first two chlorophyll figures. As shown in Fig. 10, these years corresponds to years in which much of in situ data were taken in low salinity water. Furthermore, these estuarine and coastal regions have high variability in chlorophyll and salinity that the model is unable to capture due to the relatively low spatial resolution. To summarise, the bulk of the chlorophyll data after 1993 was measured where the model is less performant.

Despite these limitations, the matched model data reproduced the variability of the in situ measurements with a correlation $R = 0.82$ in the Summer. The matched model data did not produce a significant correlation with the Spring in situ data. This can be explained by the greater chlorophyll variability in the Spring, and the high sensitivity to bloom timing. A small difference in the model bloom timing relative to nature will result in a large residual. Furthermore, it is possible that some of the in situ measurements were biased towards regions that were biologically interesting.

The number of data is also shown in the row labelled *N* of Tables 1 and 2. Of the five datasets used in this study, chlorophyll is the smallest by approximately a factor of four. Although much of the in situ variability of the larger datasets (temperature, salinity, nitrates, and phosphates) can be accounted for, a more regular and diverse chlorophyll dataset is required if the model chlorophyll is to be validated with in situ measurements and the point-to-point method.

5.3 Target diagrams

Figures 11–15 are target diagrams (Jolliff et al., 2009) showing the pattern statistics for each of the five datasets. The x-axis shows the normalised *unbiased* Root Mean

GMDD
5, 2311–2345, 2012

Square Difference (RMSD′) and the y-axis shows the normalised bias. Normalisation is performed by dividing by the reference standard deviation, σ_{ref}, which is the standard deviation of the in situ data. The diagrams two large circles correspond to lines of constant root mean square difference (RMSD), the outer has of RMSD = 1.0 and the dashed inner circle has an RMSD = 0.71. In these plots, the square markers described the comparison of the entire mean of the region to the mean of the in situ data (unmatched). The round markers are the pattern statistics of the matched model data against the mean of the in situ data (matched). The grey arrows indicate the change due to moving the unmatched to the matched methods. The colour scale of the markers shows the correlation coefficient. As with most targets, the best outcomes occur closer to the centre of the target.

Figures 11 and 12 are concise plots showing change due to the application of the matching method to temperature and salinity data. Reflecting the conclusions of Fig. 4 and 5, the matching significantly improved the correlation, and the normalised Bias and RMSD′, moving all temperature and salinity markers closer to the centre.

Although Winter nitrogen in Fig. 13 shows the best improvement in bias and RMSD of that figure, all time selections shows unambiguous increases in correlation. The others time selections show substantial shifts; the markers move across the diagram while maintaining an approximately constant RMSD.

In Fig. 14, the Autumn and Winter phosphates time selections both moved inside the RMSD = 1.0 circle and increased in correlation. The other phosphate time selections maintained similar unbiased RMSD′ while decreasing their normalised bias.

In terms of the chlorophyll *a*, Fig. 15 shows that matching does not produce the dramatic shifts seen in the other measurements. However, it is clear from the legend that the match increased the chlorophyll correlation, except for in the Winter. The normalised bias decreased in all time selections, and the unbiased RMSD′ decreased in all time selections but Winter.

These figures illustrate the importance of the matching method in at least two ways. Firstly, a model may seemingly fail to reproduce the mean state of the system, when it

GMDD
5, 2311–2345, 2012

is the in situ data that is not representative of the mean state of a system. For instance, in Fig. 12, the unmatched comparison barely reaches a correlation of $R = 0.4$, while the matched comparison has high correlation and RMSD > 0.71. Secondly, the mean state of the model may appear to underestimate or overestimate an in situ dataset, even when the opposite is true. For instance, in Fig. 13, all unmatched points have a negative normalised bias and the unmatched model appears to underestimate the in situ nitrates. However, the matched comparisons all have a normalised bias greater than zero and the model appears to overestimate the in situ nitrate.

6 Conclusions

A simple point-to-point method was outlined as a tool to validate a marine ecosystem model hindcast, POLCOMS-ERSEM, using sparse historic CTD and low resolution bottle in situ measurements from ICES. To demonstrate the method, in situ temperature, salinity, nitrates, phosphates, and chlorophyll *a* from the North Sea were compared against their model counterparts.

Firstly, it was shown in all cases that the point-to-point comparison produced linear regressions with higher correlation, decreased P-value and a y-intersect closer to zero than in the previous method: the entire region method. In the case of matched points with large residuals, the model typically predicted a value closer to the bulk of the data than the outlying in situ measurement.

Secondly, using time series plots, it was shown that POLCOMS-ERSEM has significant skill in reproducing the inter-annual variability of the in situ datasets. It also became apparent that the bulk of the variability in the in situ measurements may be due to uneven and low coverage, rather than inter-annual variability of the mean state of the system. For these reasons, we recommend that datasets such as these should be used with caution in trends and inter-annual variability studies.

Thirdly, target diagrams were used to identify some of the strengths and weaknesses of the matching method. It was found that the matching method is not guaranteed to

GMDD
5, 2311–2345, 2012

produce simultaneous improvements to bias, root mean square difference and correlation. Furthermore, the model was not always capable of reproducing the in situ data, even after matching. However, improvements in bias, RMSD′ and correlation were observed in most cases studied here. Additionally, the matching method can be used to identify hidden biases.

While the ICES datasets have been shown to be useful for the model validation, there is a need for more large and long-term non-coastal datasets. In addition to validating the models ability, these datasets are required to understand model behaviour and consequently, to plan next generation model development and validation. Future in situ datasets should strive for consistent multi-decadal coverage and good representation of both coastal and offshore environments.

Finally, it is important to remember that historic datasets were not recorded for the purpose of model validation; they have limits. As such, it is crucial to account for these restrictions when validating hindcasts. When performing a model validation using a direct comparison, it is necessary to process the model data to resemble the in situ dataset as much as possible. If a direct comparison validation is performed without some kind of matching, the predictive power of the model could be seriously misjudged.

Acknowledgements. This work is supported by the NERC National Capability in Modelling programme at Plymouth Marine Laboratory and Theme 6 of the EC seventh framework program through the Marine Ecosystem Evolution in a Changing Environment (MEECE No. 212085) Collaborative Project.

References

Arhonditsis, G. B. and Brett, M. T.: Evaluation of the current state of mechanistic aquatic biogeochemical modeling, Mar. Ecol. Prog. Ser., 271, 13–26, doi:10.3354/meps271013, 2004. 2312

Blackford, J. C., Allen, J. I., and Gilbert, F. J.: Ecosystem dynamics at six contrasting sites: a generic modelling study, J. Marine Syst., 52, 191–215, doi:10.1016/j.jmarsys.2004.02.004, 2004. 2314

GMDD

5, 2311–2345, 2012

Doney, S. C., Lima, I., Moore, J. K., Lindsay, K., Behrenfeld, M. J., Westberry, T. K., Mahowald, N., Glover, D. M., and Takahashi, T.: Skill metrics for confronting global upper ocean ecosystem-biogeochemistry models against field and remote sensing data, J. Marine Syst., 76, 95–112, doi:10.1016/j.jmarsys.2008.05.015, 2009. 2313

Edwards, K. P., Barciela, R., and Butenschön, M.: Validation of the NEMO-ERSEM operational ecosystem model for the North West European Continental Shelf, Ocean Sci. Discuss., 9, 745–786, doi:10.5194/osd-9-745-2012, 2012. 2313

FAO: FAO Major Fishing Areas, (Major Fishing Area 27), CWP Data Collection, available at: http://www.fao.org/fishery/area/Area27/en#NB04F5, (last access: 17 August 2012), 2008. 2315

Garcia, H. E. and Levitus, S.: World Ocean Atlas 2005, vol. 4, Nutrients (phosphate, nitrate, silicate), Tech. Rep. 64, National Oceanographic Data Center (US), Ocean Climate Laboratory, Washington, D.C., 2010a. 2315

Garcia, H. E. and Levitus, S.: World Ocean Atlas 2005, vol. 3, Dissolved oxygen, apparent oxygen utilization, and oxygen saturation, Tech. Rep. 63, National Oceanographic Data Center (US), Ocean Climate Laboratory, Washington, D.C., 2010b. 2315

Holt, J., James, I. D., and Jones, J. E.: An s coordinate density evolving model of the northwest European continental shelf 2, Seasonal currents and tides, J. Geophys. Res., 106, 14035–14053, doi:10.1029/2000JC000303, 2001. 2314

Holt, J., Butenschön, M., Wakelin, S. L., Artioli, Y., and Allen, J. I.: Oceanic controls on the primary production of the northwest European continental shelf: model experiments under recent past conditions and a potential future scenario, Biogeosciences, 9, 97–117, doi:10.5194/bg-9-97-2012, 2012. 2314, 2315

ICES: ICES Dataset on Ocean Hydrography, The International Council for the Exploration of the SeaCopenhagen, Copenhagen, 2009. 2315

Jolliff, J. K., Kindle, J. C., Shulman, I., Penta, B., Friedrichs, M. A., Helber, R., and Arnone, R. A.: Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment, J. Marine Syst., 76, 64–82, doi:10.1016/j.jmarsys.2008.05.014, 2009. 2313, 2321

Lewis, K., Allen, J., Richardson, A. J., and Holt, J.: Error quantification of a high resolution coupled hydrodynamic-ecosystem coastal-ocean model: Part 3, validation with continuous plankton recorder data, J. Marine Syst., 63, 209–224, doi:10.1016/j.jmarsys.2006.08.001, 2006. 2313

GMDD
5, 2311–2345, 2012

OSPAR Commission: Common Procedure for the Identification of the Eutrophication Status of the OSPAR Maritime Area, UK National Report, London, UK, 2008. 2321

Robeson, S.: Influence of spatial sampling and interpolation on estimates of air temperature change, Clim. Res., 4, 119–126, 1994. 2313

Stow, C. A., Jolliff, J., Mcgillicuddy, D. J., Doney, S. C., Allen, J. I., Friedrichs, M. A. M., Rose, K. A., and Wallhead, P.: Skill assessment for coupled biological/physical models of marine systems, J. Marine Syst., 76, 4–15, doi:10.1016/j.jmarsys.2008.03.011, 2009. 2313

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res., 106, 7183–7192, 2001. 2313

Uppala, S. M., KÅllberg, P. W., Simmons, A. J., Andrae, U., Bechtold Da Costa, V., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, L. Van De, Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., Mcnally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P. and Woollen, J.: The ERA-40 re-analysis, Q. J. Roy. Meteorol. Soc., 131, 2961–3012, doi:10.1256/qj.04.176, 2005. 2315, 2319

Vörösmarty, C. J., Fekete, B. M., Meybeck, M., and Lammers, R. B.: Global system of rivers: its role in organizing continental land mass and defining land-to-ocean linkages, Global Biogeochem. Cy., 14, 599–621, doi:10.1029/1999GB900092, 2000. 2315

Young, E. F. and Holt, J. T.: Prediction and analysis of long-term variability of temperature and salinity in the Irish Sea, J. Geophys. Res., 112, C01008, doi:10.1029/2005JC003386, 2007. 2315

GMDD

5, 2311–2345, 2012

Table 1. Linear regression output parameters for temperature (T), salinity (Sal.) and nitrates (N_3). The parameters shown are the slope of the line, $\hat{\beta}_1$, the y-axis intersect, $\hat{\beta}_0$, the correlation coefficient, R, the two tailed probability, P, the standard error, ϵ, and the Number of data, N.

		Full	Winter	Spring	Summer	Annual
T	$\hat{\beta}_1$	0.9083	1.192	0.9183	0.9101	0.9565
	$\hat{\beta}_0$	1.131	−0.8403	0.8538	1.215	0.6754
	R	0.9363	0.9602	0.942	0.9814	0.9572
	P	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$
	ϵ	0.0003	0.0563	0.0531	0.0289	0.0469
	N	1 191 530	271 599	334 676	308 946	1 191 530
Sal.	$\hat{\beta}_1$	1.245	1.176	1.31	1.096	1.268
	$\hat{\beta}_0$	−8.305	−6.023	−10.63	−3.298	−9.221
	R	0.7681	0.7742	0.9212	0.8115	0.7558
	P	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$
	ϵ	0.001	0.1559	0.0897	0.128	0.1783
	N	1 176 225	264 732	333 987	303 397	1 176 225
N_3	$\hat{\beta}_1$	1.05	1.336	0.6463	0.6168	0.6872
	$\hat{\beta}_0$	−4.248	−10.05	0.9477	−0.8275	0.2063
	R	0.5928	0.8191	0.7484	0.5798	0.8132
	P	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	0.0001	$< 10^{-4}$
	ϵ	0.0042	0.1518	0.0929	0.1406	0.0798
	N	116 933	31 575	30 243	22 892	116 933

GMDD
5, 2311–2345, 2012

Table 2. Linear regression output parameters for phosphates (P_4) and chlorophyll (Chl.). The parameters shown are the slope of the line, $\hat{\beta}_1$, the y-axis intersect, $\hat{\beta}_0$, the correlation coefficient, R, the two tailed probability, P, the standard error, ϵ, and the Number of data, N.

		Full	Winter	Spring	Summer	Annual
P_4	$\hat{\beta}_1$	0.6823	0.938	0.2687	0.0152	0.1814
	$\hat{\beta}_0$	0.1802	−0.0255	0.3015	0.532	0.4929
	R	0.4153	0.5799	0.2179	0.0082	0.1283
	P	$< 10^{-4}$	0.0001	0.1769	0.9598	0.43
	ϵ	0.0043	0.2167	0.1952	0.2991	0.2274
	N	121 860	32 957	31 231	24 153	121 860
Chl.	$\hat{\beta}_1$	0.7479	0.1448	1.62	3.031	1.098
	$\hat{\beta}_0$	2.052	1.839	0.2675	−0.0887	1.139
	R	0.2379	0.0632	0.37	0.7083	0.2844
	P	$< 10^{-4}$	0.7492	0.0263	$< 10^{-4}$	0.0793
	ϵ	0.0171	0.4483	0.6975	0.5515	0.6085
	N	32 019	6552	11 406	10 121	32 019

GMDD
5, 2311–2345, 2012

Table 3. Linear regression output parameters for temperature (T), salinity (Sal.) and nitrates (N_3). The parameters shown are the slope of the line, $\hat{\beta}_1$, the y-axis intersect, $\hat{\beta}_0$, the correlation coefficient, R, the two tailed probability, P, and the standard error, ϵ.

		Entire Region	Matched
Annual T	$\hat{\beta}_1$	0.4495	0.9565
	$\hat{\beta}_0$	6.02	0.6754
	R	0.643	0.9572
	P	$< 10^{-4}$	$< 10^{-4}$
	ϵ	0.0932	0.0469
Annual Sal.	$\hat{\beta}_1$	0.2063	1.268
	$\hat{\beta}_0$	27.11	−9.221
	R	0.3401	0.7558
	P	0.0456	$< 10^{-4}$
	ϵ	0.0993	0.1783
Winter N_3	$\hat{\beta}_1$	0.0295	1.336
	$\hat{\beta}_0$	13.64	−10.05
	R	0.3781	0.8191
	P	0.0251	$< 10^{-4}$
	ϵ	0.0126	0.1518

GMDD
5, 2311–2345, 2012

Table 4. Linear regression output parameters for phosphates (P_4) and chlorophyll (Chl.). The parameters shown are the slope of the line, $\hat{\beta}_1$, the y-axis intersect, $\hat{\beta}_0$, the correlation coefficient, R, the two tailed probability, P, and the standard error, ϵ.

		Entire Region	Matched
Winter P_4	$\hat{\beta}_1$	0.0229	0.938
	$\hat{\beta}_0$	0.7217	−0.0255
	R	0.1602	0.5799
	P	0.3654	0.0001
	ϵ	0.0249	0.2167
Spring Chl.	$\hat{\beta}_1$	0.007	1.62
	$\hat{\beta}_0$	1.966	0.2675
	R	0.1473	0.37
	P	0.4212	0.0263
	ϵ	0.0085	0.6975
Summer Chl.	$\hat{\beta}_1$	−0.0113	3.031
	$\hat{\beta}_0$	0.7444	−0.0887
	R	−0.5414	0.7083
	P	0.002	$< 10^{-4}$
	ϵ	0.0033	0.5515

GMDD
5, 2311–2345, 2012

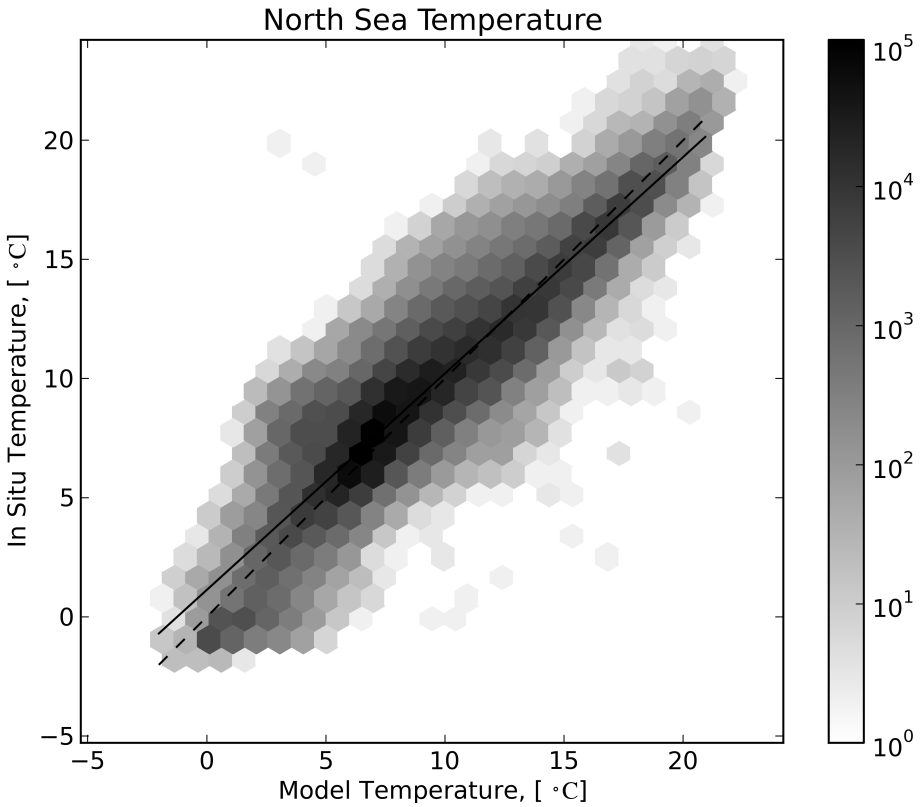

Fig. 1. Model vs. in situ data density for the North Sea temperature. The black line shows the linear regression fit and the dashed line shows the line of unity slope that passes through the origin.

GMDD
5, 2311–2345, 2012

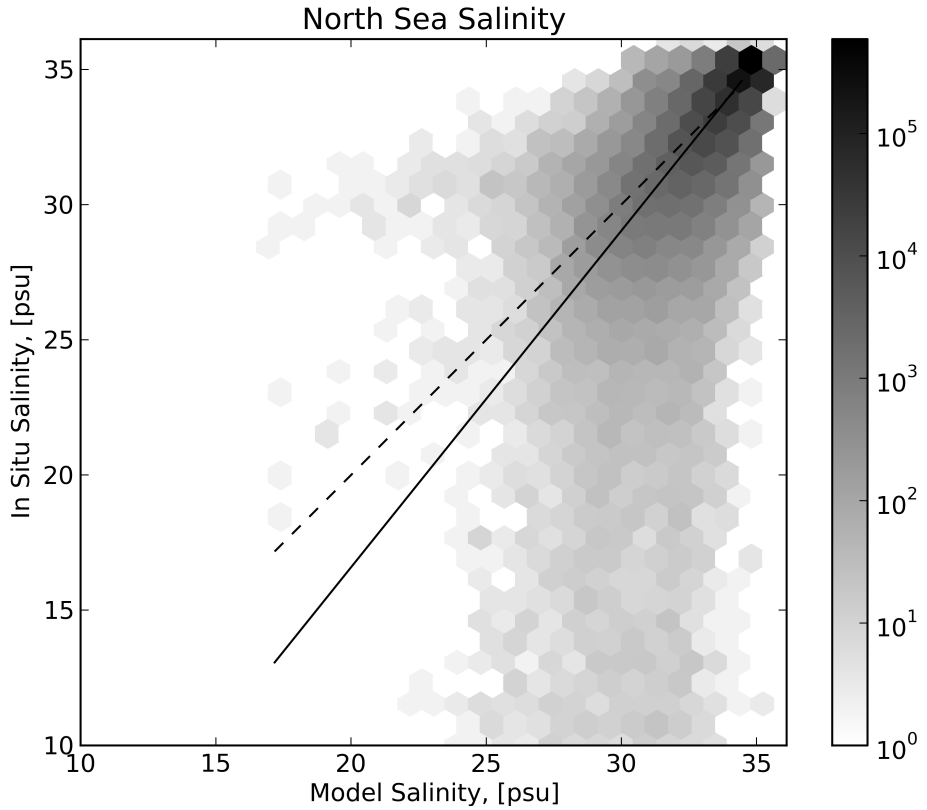

Fig. 2. Model vs. in situ data density for the North Sea CTD salinity. The black line shows the linear regression fit and the dashed line shows the line of unity slope that passes through the origin.

GMDD
5, 2311–2345, 2012

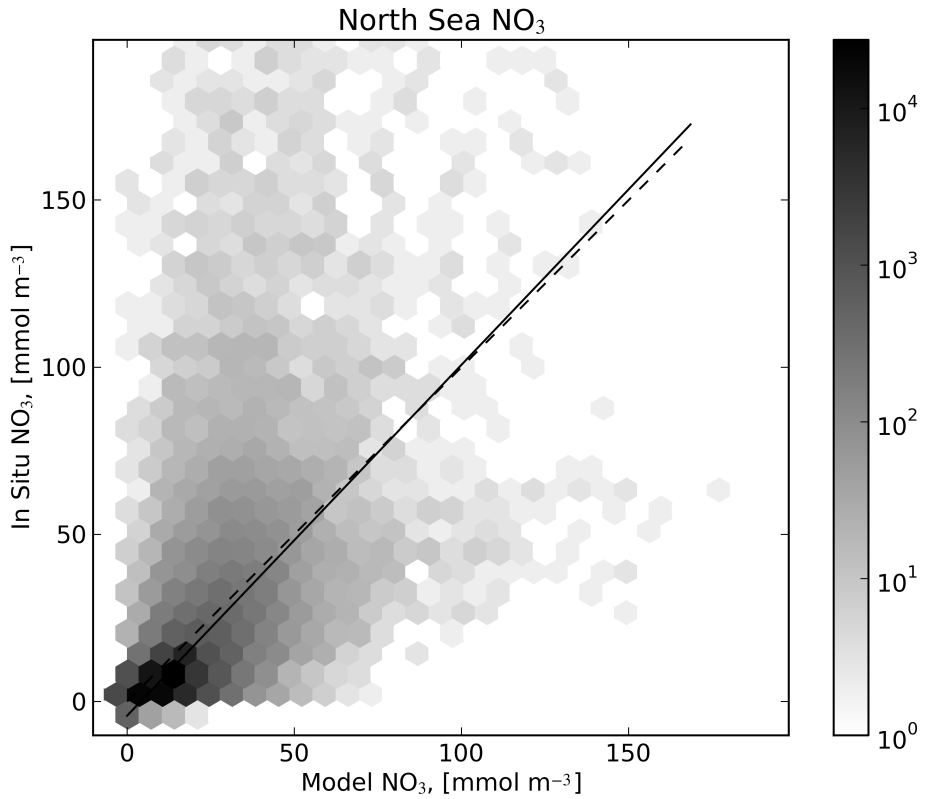

Fig. 3. Model vs. in situ data density for the North Sea Bottle and low resolution CTD nitrates. The black line shows the linear regression fit and the dashed line shows the line of unity slope that passes through the origin.

GMDD
5, 2311–2345, 2012

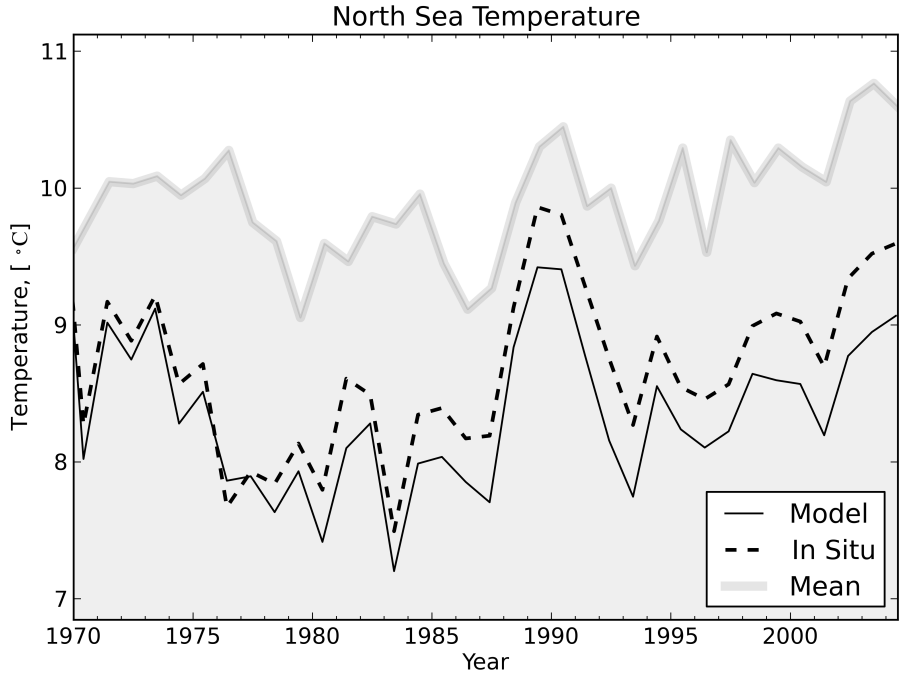

Fig. 4. North Sea mean annual temperature time series; the black line is the matched model mean annual temperature, the dashed line is the mean annual in situ temperature, and the grey area shows the depth integrated mean of the entire model region.

GMDD
5, 2311–2345, 2012

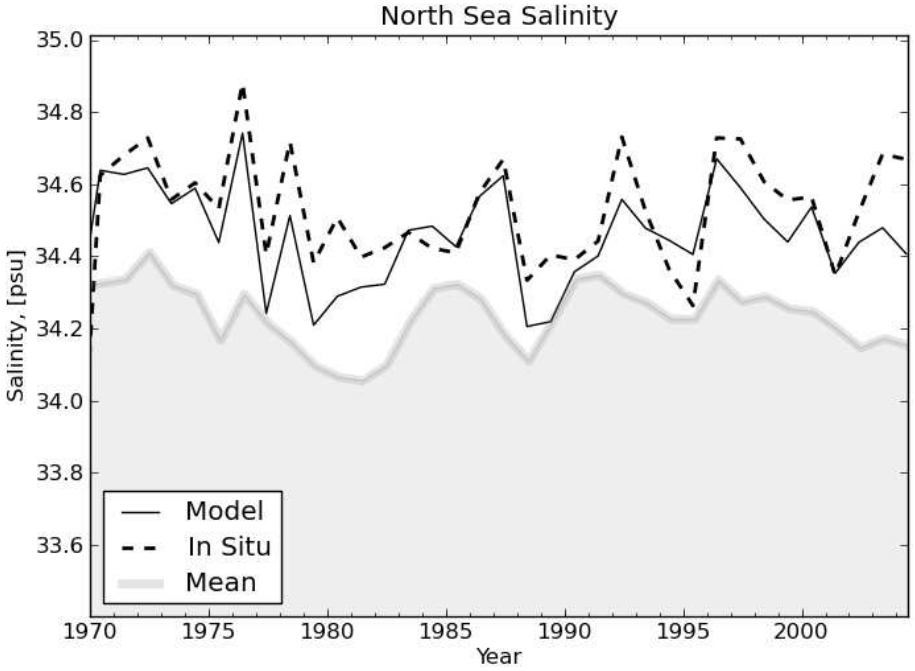

Fig. 5. North Sea mean annual salinity time series: the black line is the matched model mean annual salinity, the dashed line is the mean annual in situ salinity, and the grey area shows the depth integrated mean salinity of the entire model region.

GMDD
5, 2311–2345, 2012

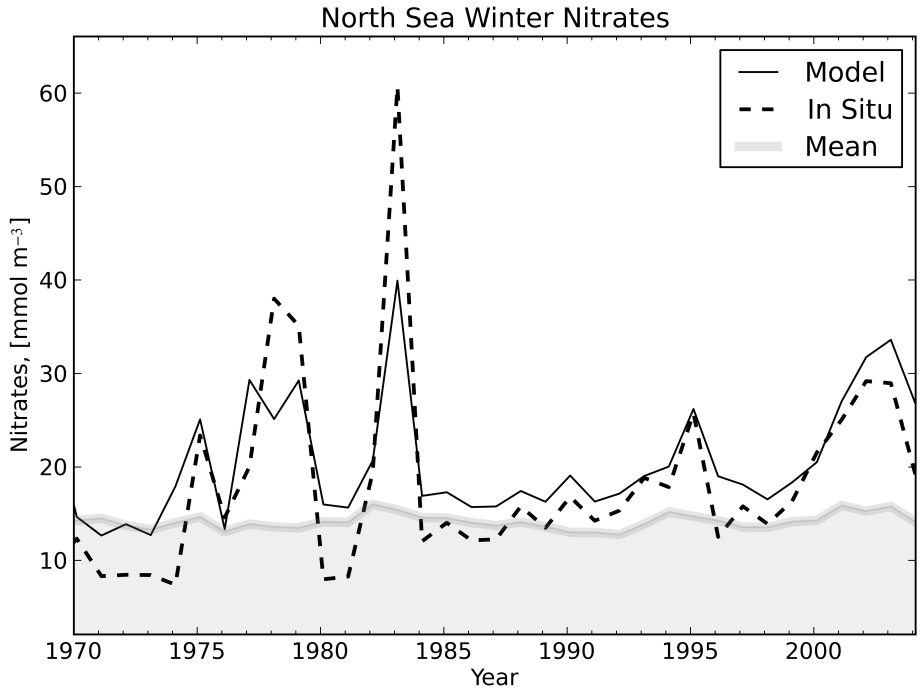

Fig. 6. North Sea mean Winter nitrates time series: the black line is the matched model mean Winter nitrates, the dashed line is the mean Winter in situ nitrates, and the grey area shows the depth integrated mean nitrate of the entire model region.

GMDD
5, 2311–2345, 2012

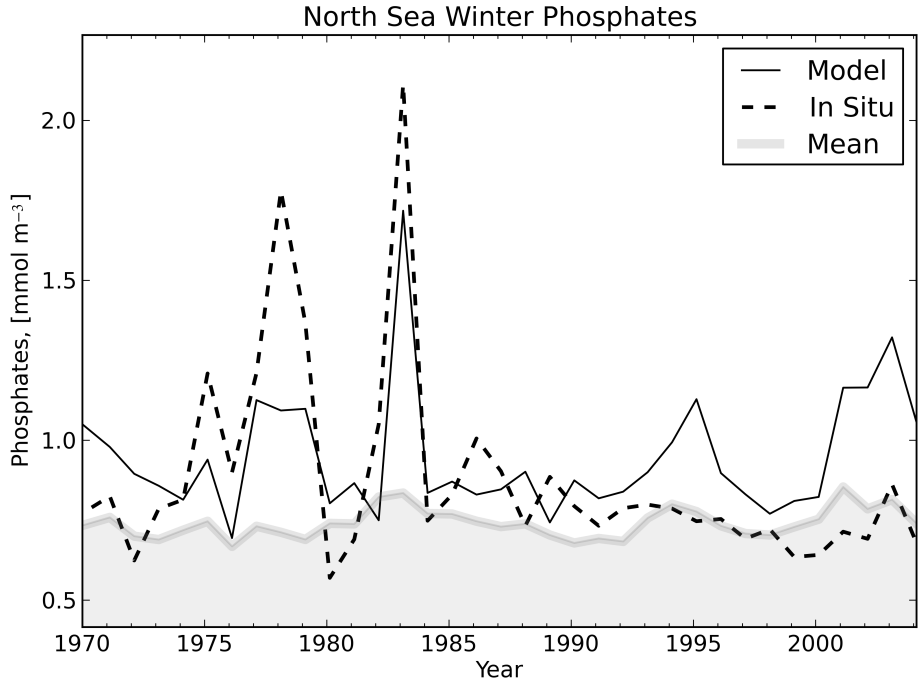

Fig. 7. North Sea mean Winter phosphates time series: the black line is the matched model mean Winter phosphates, the dashed line is the mean Winter in situ phosphates, and the grey area shows the depth integrated mean phosphates of the entire model region.

GMDD
5, 2311–2345, 2012

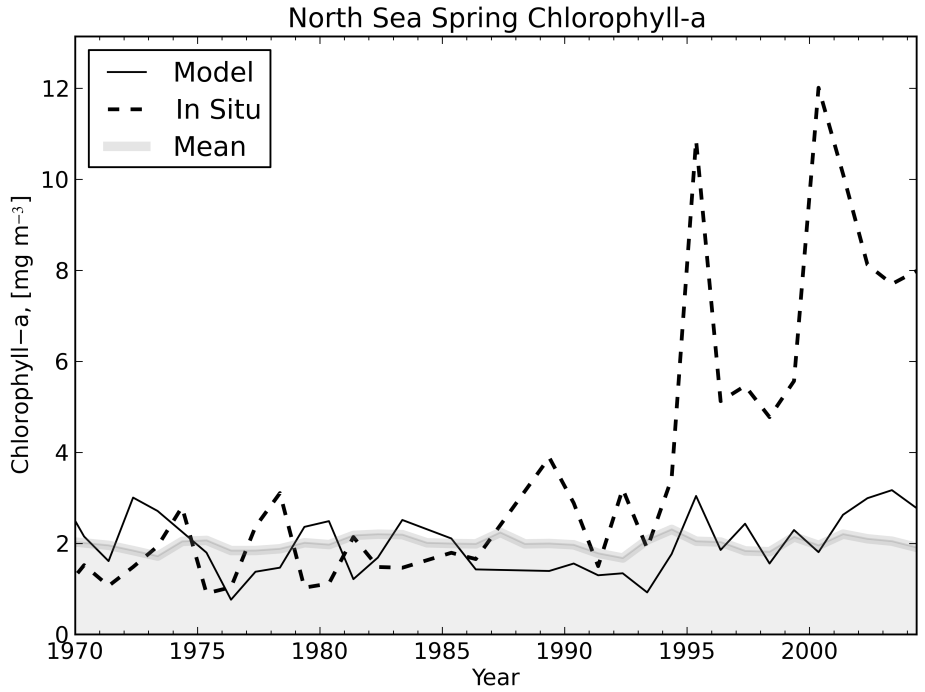

Fig. 8. North Sea mean Spring chlorophyll time series: the black line is the matched model mean Spring chlorophyll, the dashed line is the mean Spring in situ chlorophyll, and the grey area shows the depth integrated mean chlorophyll of the entire model region.

GMDD
5, 2311–2345, 2012

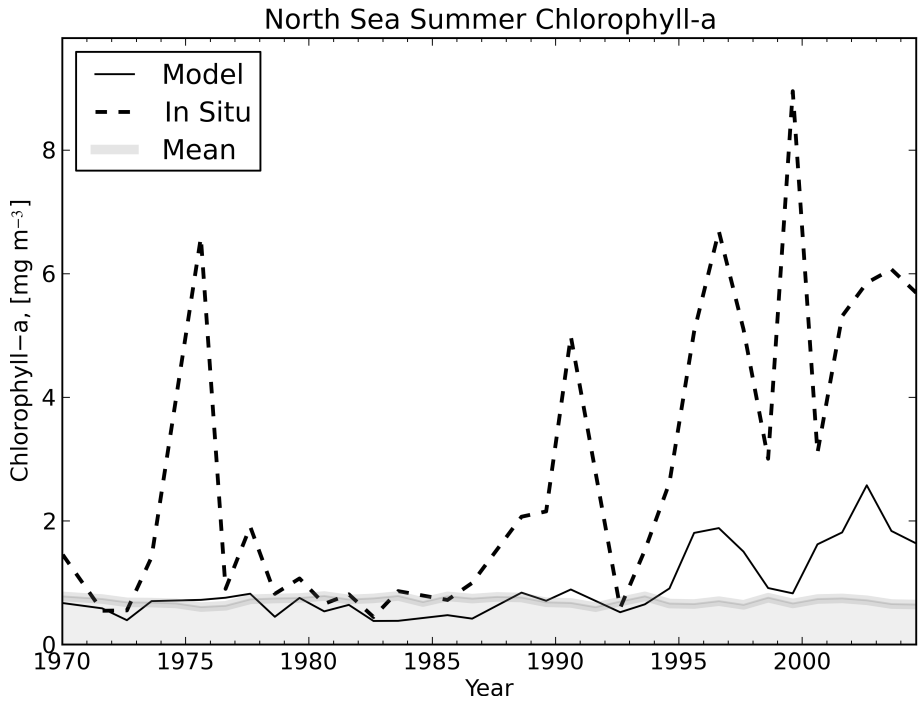

Fig. 9. North Sea mean Summer chlorophyll time series: the black line is the matched model mean Summer chlorophyll, the dashed line is the mean Summer in situ chlorophyll, and the grey area shows the depth integrated mean chlorophyll of the entire model region.

GMDD
5, 2311–2345, 2012

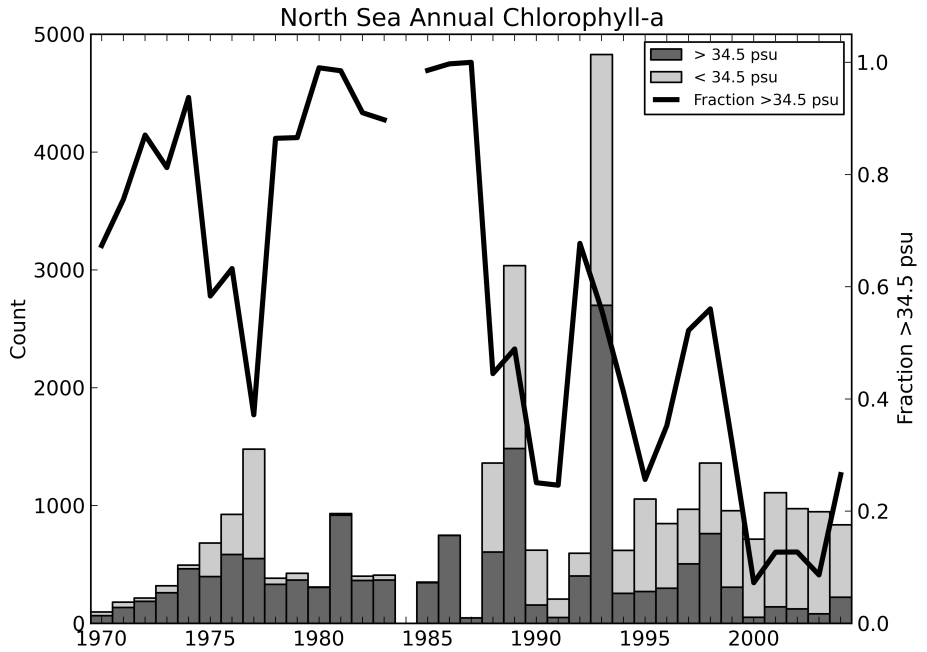

Fig. 10. The number of North Sea in situ chlorophyll measurements, grouped into a high salinity (Sal. > 34.5 PSU) in dark grey and low salinity (Sal. < 34.5 PSU) in light grey. The dark line shows the fraction of high salinity.

GMDD
5, 2311–2345, 2012

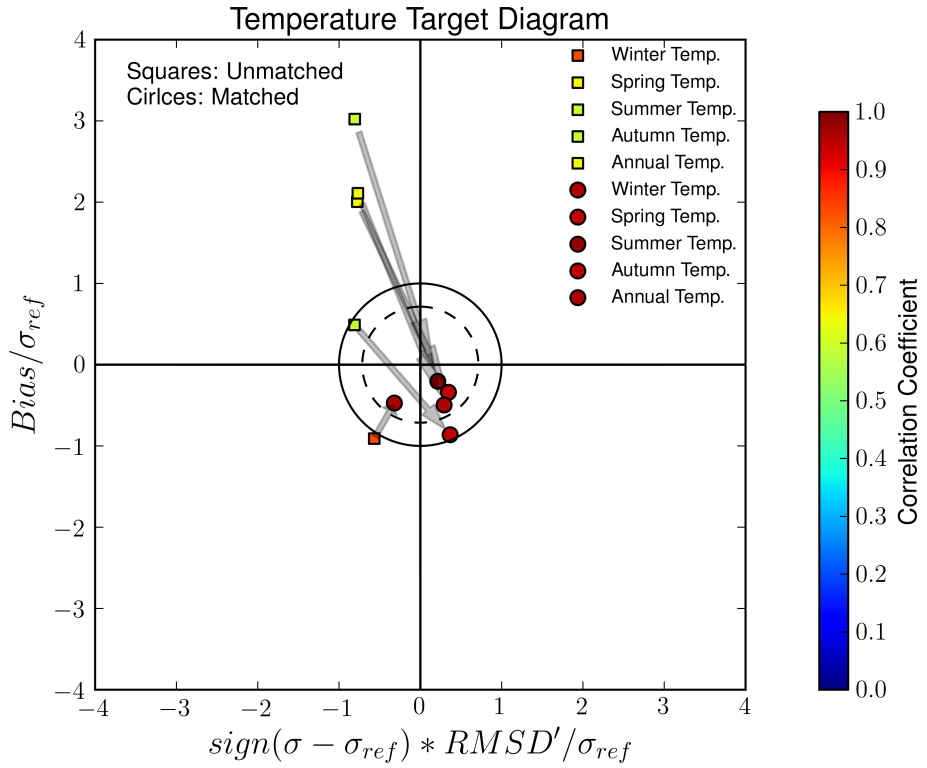

Fig. 11. Target diagram showing the impact of the transition from the unmatched to the matched methods on temperature.

GMDD
5, 2311–2345, 2012

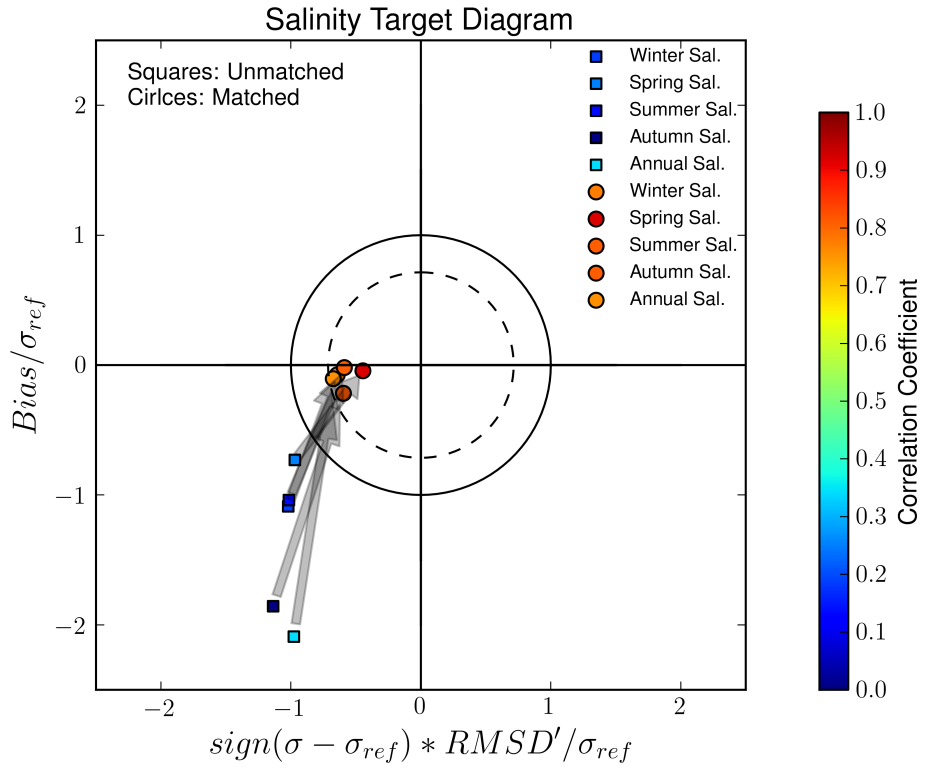

Fig. 12. Target diagram showing the impact of the transition from the unmatched to the matched methods on salinity.

GMDD
5, 2311–2345, 2012

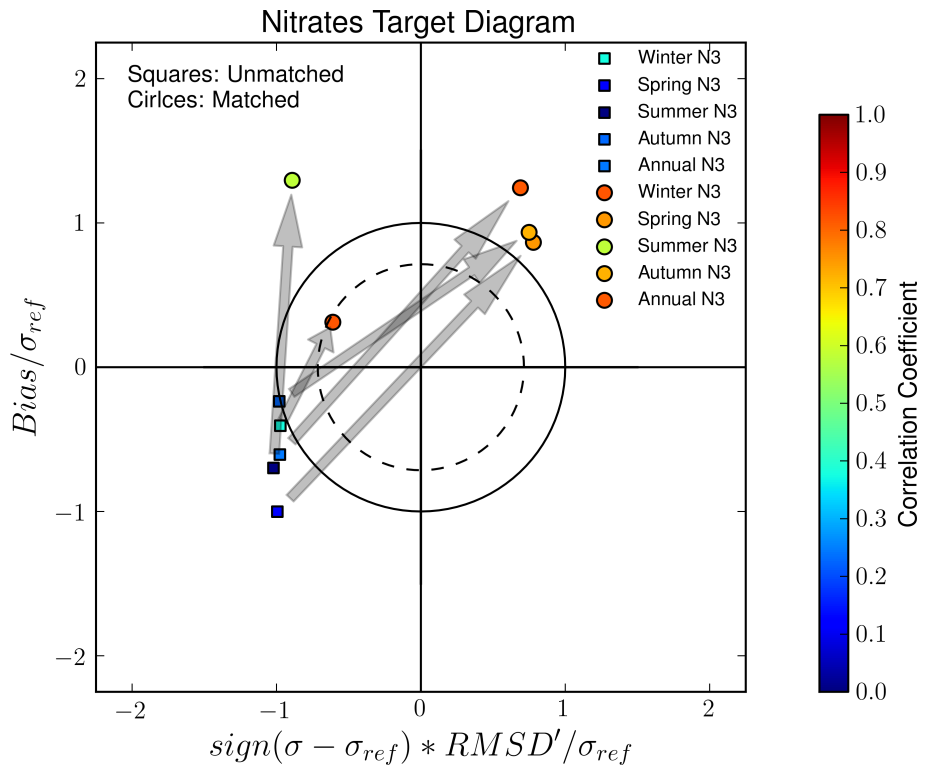

Fig. 13. Target diagram showing the impact of the transition from the unmatched to the matched methods on nitrates.

GMDD
5, 2311–2345, 2012

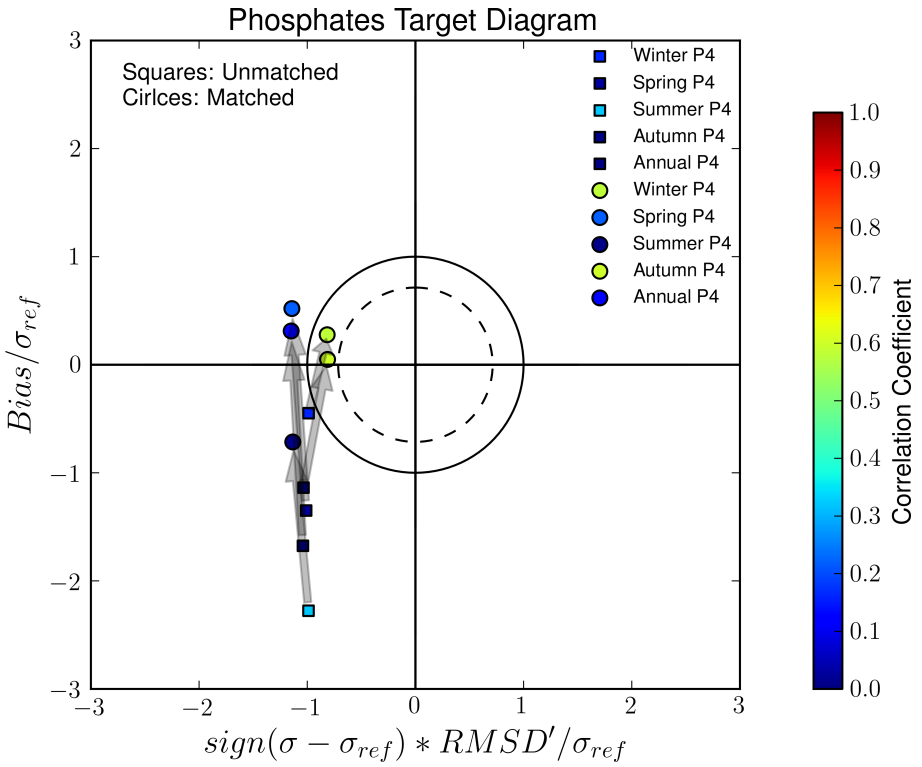

Fig. 14. Target diagram showing the impact of the transition from the unmatched to the matched methods on phosphates.

GMDD
5, 2311–2345, 2012

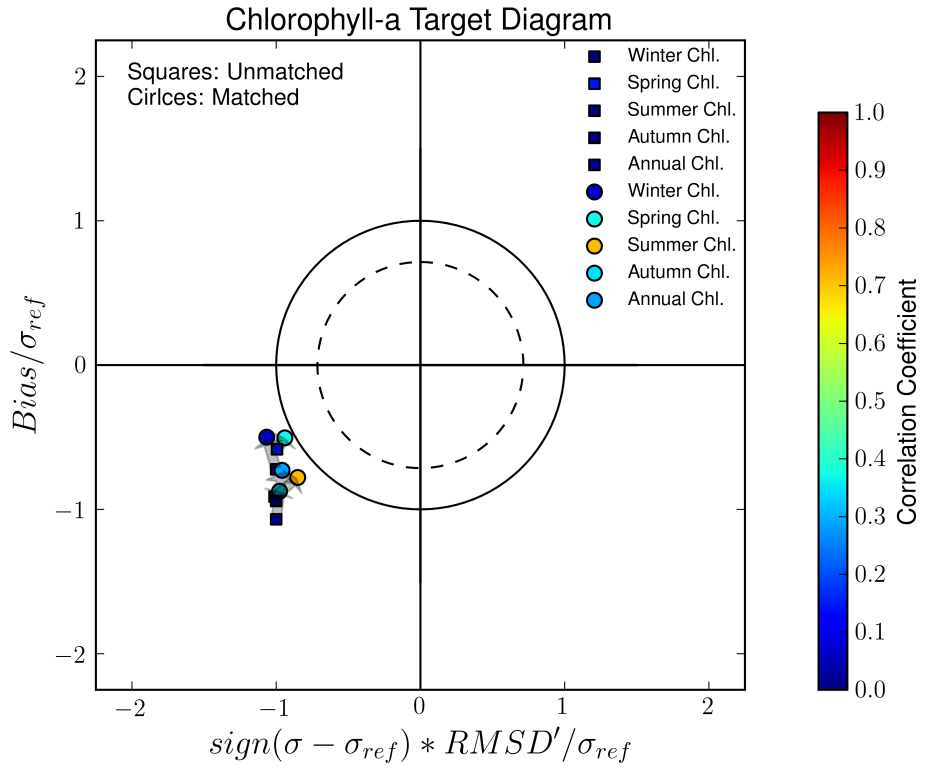

Fig. 15. Target diagram showing the impact of the transition from the unmatched to the matched methods on chlorophyll *a*.

GMDD
5, 2311–2345, 2012