National Oceanography Centre, Southampton

Internal Document No. 5

A material balancing scheme for ocean colour
data assimilation

J C P Hemmings[1], R M Barciela[2] & M J Bell[2]

2007

[1]National Oceanography Centre, Southampton
University of Southampton, Waterfront Campus
European Way
Southampton
Hants SO14 3ZH
UK

Author contact details
Tel: +44 (0)23 8059 7793
Fax: +44 (0)23 8059 6400
Email: jch@noc.soton.ac.uk

[2]Meteorological Office, Fitzroy Road, Exeter EX1 3PB

DOCUMENT DATA SHEET

AUTHOR HEMMINGS, J C P, BARCIELA, R M & BELL, M J	*PUBLICATION* *DATE* 2007

TITLE
    A material balancing scheme for ocean colour data assimilation.

ABSTRACT

A material balancing scheme for assimilation of ocean colour data in the FOAM-HadOCC model (Forecasting Ocean Assimilation Model with Hadley Centre Ocean Carbon Cycle Model biogeochemistry) has been developed with the aim of exploiting satellite data to improve air-sea CO_2 flux estimates.   The balancing scheme uses surface chlorophyll increments to determine increments for the biogeochemical tracers: nutrient, phytoplankton, zooplankton, detritus, dissolved inorganic carbon (DIC) and alkalinity.

The analysis conserves carbon at each grid point and nitrogen at grid points where sufficient nitrogen is available, on the assumption that the modelling of processes transferring material between biogeochemical compartments is the main source of error. Phytoplankton increments are calculated using the model nitrogen:chlorophyll ratio. Increments to the other nitrogen pools (nutrient, zooplankton and detritus) are determined by a balancing model that responds to changes in the plankton dynamics. The nutrient balancing factor, the fraction of the phytoplankton increment to be balanced by the nutrient increment, varies according to the relative contributions of growth and loss rate errors to the phytoplankton error, as estimated from a probability model. Preliminary balancing factor values are adjusted to satisfy state-dependent restrictions on the size of the increments. Increments derived in this way are applied down to the depth of the mixed layer. Further increments are applied where necessary to avoid the creation of unrealistic sub-surface nutrient minima. Increments to DIC balance the implied carbon changes in the organic compartments and alkalinity increments are inferred from those for nutrient.

An off-line evaluation of the scheme is carried out in a 1-D test-bed in which HadOCC biogeochemistry is forced by physical data for a range of latitudes in the eastern North Atlantic. Evaluation is by twin experiments for which synthetic system trajectories are generated by perturbing model parameters during integration to provide a range of plausible truths. Assimilation of daily chlorophyll observations, with or without simulated observation error, gives major improvements in pCO_2 at the high latitudes but less improvement at low latitudes where it has a detrimental effect on summer and early autumn pCO_2 due to errors in the model nitrogen:chlorophyll ratio. Beneficial effects of nitrogen balancing are demonstrated by comparison with experiments in which only phytoplankton and DIC are updated. The sub-surface nutrient correction increments are shown to reduce, but not remove, undesirable effects of assimilation on the nutrient and phytoplankton profiles.

ISSUING ORGANISATION
National Oceanography Centre, Southampton
University of Southampton, Waterfront Campus
European Way
Southampton SO14 3ZH
UK

Not generally distributed - please refer to author

Contents

1 Introduction

The Centre for Observation of Air-sea Interactions and Fluxes (CASIX) aims to produce accurate high-resolution estimates of the air-sea flux of CO_2 by assimilating earth observation data into models of the ocean carbon cycle. Air-sea flux estimates for the open ocean are to be provided by the FOAM-HadOCC model: the Met Office's Forecasting Ocean Assimilation Model (FOAM, Bell *et al.*, 2000) with embedded biogeochemistry ported from the Hadley Centre Ocean Carbon Cycle Model (HadOCC, Palmer and Totterdell, 2001). The primary aim of assimilating ocean colour is to improve the model's surface pCO_2 estimates, thereby improving its estimates of air-sea CO_2 flux that are directly proportional to the difference in pCO_2 across the air-sea interface. This report details the design and off-line evaluation of a material balancing scheme for surface chlorophyll data assimilation in the FOAM-HadOCC model. It is intended that it will form the basis of future schemes designed to exploit new CASIX ocean colour products.

The biogeochemical model (Figure 1) is an NPZD model. Models of this type represent the flow of nitrogen between nutrient, phytoplankton, zooplankton and detritus compartments. The organic compartments (phytoplankton, zooplankton and detritus) act as pools of carbon as well as nitrogen and carbon is tracked through the system by assuming separate constant carbon:nitrogen ratios for each. Changes in total dissolved inorganic carbon (DIC) and alkalinity driven by the nitrogen-controlled plankton dynamics are thereby inferred. The version of the model used here (Totterdell, pers. comm.) differs from the Palmer and Totterdell (2001) model in having a variable phytoplankton carbon:chlorophyll ratio, reflecting acclimation of the photosynthetic apparatus to the available light and nutrients according to the model of Geider *et al.* (1997). In addition, it incorporates the light penetration and photosynthesis model of Anderson (1993), which allows for self-shading of the phytoplankton, and a slow relaxation of nutrient towards climatological profiles at depths below both the euphotic zone and the upper mixed layer to prevent excessive nutrient drift.

The annual cycle in surface pCO_2 obtained from the model is dependent on surface DIC and alkalinity, the main effect of the biota being via DIC: DIC concentration is reduced as CO_2 is taken up by phytoplankon for photosynthesis and increased as a result of plankton respiration and respiration associated with the microbial breakdown of dead organic matter. An ocean colour assimilation scheme must therefore make the best use of phytoplankton-related information for correcting not just the phytoplankton biomass itself but also the other components of the system which affect the pCO_2. The scheme described here is a material balancing scheme designed to work with the existing FOAM data assimilation scheme, to allow ocean colour data to be assimilated effectively with minimal computational cost. The ma-

Figure 1: HadOCC biogeochemistry showing carbon and nitrogen flows. Alkalinity is also affected by fluxes of OH^- ions (not shown) which are equal and opposite to the nitrogen fluxes affecting the nutrient pool. Sinking of organic detritus is modelled explicitly, while carbonate export is modelled by removing DIC and alkalinity in the euphotic zone and adding it evenly over all depths below the lysocline.

terial balancing scheme uses surface chlorophyll increments to determine balancing increments for the model's nitrogen and carbon pools (i.e. for the nutrient, phytoplankton, zooplankton, detritus and DIC tracers) and for the alkalinity tracer, in and below the surface boundary layer. Surface chlorophyll increments will be provided daily from a 2-D analysis performed by the FOAM assimilation scheme.

2   The Assimilation Scheme

The principles on which the scheme is based are introduced in Section 2.1. A complete overview is then given in Section 2.2. This section introduces two complementary nitrogen balancing models that are used to determine balancing increments for

the nitrogen tracers at individual grid points. These sub-components of the scheme are described in more detail in Sections 2.3 and 2.4.

2.1   Basic Principles of the Scheme

The scheme has been designed to correct errors arising from inadequate representation of biological processes affecting chlorophyll concentration. Errors in physical processes also have a strong effect on surface chlorophyll, especially via their effects on nutrient and light availability to the phytoplankton and redistribution of the phytoplankton biomass. These errors will inevitably be a source of interference when the scheme is tuned to real-word data. However, they are addressable by other methods and it is inappropriate for the chlorophyll assimilation scheme to attempt to compensate for the effects of errors in processes that are purely physical. The scheme should perform better as these other sources of error are reduced. Much of this improvement should come from improvements to the physical state but another important contribution will be to include biogeochemical balancing increments during physical data assimilation.

Errors in processes transferring material between different biological or chemical pools are assumed to dominate over errors in the sinking rate of particles. Balancing increments are therefore made in such a way that carbon is conserved at each grid point and that nitrogen is conserved at each grid point if sufficient nitrogen is available. Not all of the nitrogen present is made available to satisfy this conservation requirement. The amount available is subject to state-dependent restrictions on the size of the balancing increments allowed for each variable. These help to avoid excessive perturbation of the model dynamics by the assimilation. Any increase in the model's total nitrogen as a result of the non-conservative analysis should ultimately be compensated for by the nutrient relaxation to climatology, avoiding a detrimental effect on the nitrogen budget.

2.2   Scheme Overview

The core component of the chlorophyll assimilation scheme is a nitrogen balancing scheme that determines surface and sub-surface increments for the nitrogen tracers (nutrient, phytoplankton, zooplankton and detritus) at each location on the horizontal grid, based on a surface phytoplankton nitrogen increment. Section 2.2.1 describes, in brief, the chlorophyll increments from which surface phytoplankton increments are derived. Section 2.2.2 then defines the phytoplankton increment and outlines the balancing model that calculates the remaining nitrogen tracer incre-

ments for the surface boundary layer. Section 2.2.3 describes how the analysis is extended to deeper levels, introducing the second nitrogen balancing model. This model responds to nutrient increments rather than phytoplankton increments. It is required because additional nutrient increments are needed in the sub-surface analysis to maintain realistic vertical profiles. Section 2.2.4 describes how carbon and alkalinity increments are derived from the output of the nitrogen scheme and a summary of the whole assimilation procedure is given in Section 2.2.5.

2.2.1 Surface Chlorophyll Increments

Chlorophyll increments are determined by applying the FOAM analysis correction scheme to field estimates of log-transformed surface chlorophyll. The transform is used because of the tendency for chlorophyll to be log-normally distributed (Campbell, 1995): deviations in biomass concentration are typically proportional to the concentration itself due to the exponential nature of phytoplankton growth.

The chlorophyll increment is then

$$
\begin{aligned}
\Delta chl &= chl_{\mathrm{a}} - chl_{\mathrm{b}} \\
&= chl_{\mathrm{b}}.(10^{\Delta logchl} - 1),
\end{aligned}
\tag{1}
$$

where $\Delta logchl$ is the analysis correction scheme increment and the subscripts a and b denote analysis and background values respectively.

2.2.2 Increments to Nitrogen Tracers in the Surface Boundary Layer

Chlorophyll increments are converted to surface phytoplankton increments using the biogeochemical model's variable nitrogen:chlorophyll ratio. Where the resultant phytoplankton increment is less than a small threshold value, all increments are set to zero: it is assumed, given the uncertainty in the nitrogen:chlorophyll ratio, that the implied error is not sufficiently significant to justify the computational overhead of calculating balancing increments. Chlorophyll is calculated directly from phytoplankton rather than being carried as a separate tracer so chlorophyll increments are not actually realized in such cases. The threshold value is specified as an external parameter ΔP_{MIN}.

All non-zero surface phytoplankton increments are processed by the nitrogen balancing scheme. The phytoplankton-driven nitrogen balancing model calculates a time-varying factor for each of the other tracers, referred to as a balancing factor, that sets the magnitude of its surface layer increment relative to that for phytoplankton. The absolute surface increments for the nitrogen tracers (phytoplankton, nutrient, zooplankton and detritus) are

$$\Delta P = \phi \Delta chl \tag{2}$$
$$\Delta N = -b_{\mathrm{N}} \Delta P \tag{3}$$
$$\Delta Z = -b_{\mathrm{Z}} \Delta P \tag{4}$$
$$\Delta D = -b_{\mathrm{D}} \Delta P, \tag{5}$$

where ϕ is the model nitrogen:chlorophyll ratio for phytoplankton and b_{N}, b_{Z} and b_{D} are the calculated balancing factors for nutrient, zooplankton and detritus respectively. The negated balancing factors can be interpreted as estimates of the ratio of the background error covariances to the phytoplankton background error variance. However, as discussed later in this section, the balancing factors are not independent of the background errors themselves so their application is restricted to a specific background state.

For nitrogen conservation

$$b_{\mathrm{N}} + b_{\mathrm{Z}} + b_{\mathrm{D}} = 1. \tag{6}$$

Equation 6 is satisfied by the balancing factor model, subject to the total availability of nitrogen at the model grid-point.

Within the nitrogen balancing model, a probability model is used to divide the phytoplankton error into two separate components: one attributed to growth errors and the other to loss errors. This is a key step in the procedure for estimating optimal balancing factors. The reason becomes clear if we consider the biological dynamics. While growth involves the transfer of nitrogen between the nutrient and phytoplankton pools only, the main nitrogen transfer associated with loss processes (grazing, mortality and respiration) is between the phytoplankton pool and the combined zooplankton and detritus pools, a much smaller fraction being transferred to the nutrient pool. In general, this means that nutrient errors are negatively

correlated with the part of the phytoplankton error caused by errors in the growth rate, while errors in zooplankton and detritus are negatively correlated with the error attributable to errors in the loss rate.
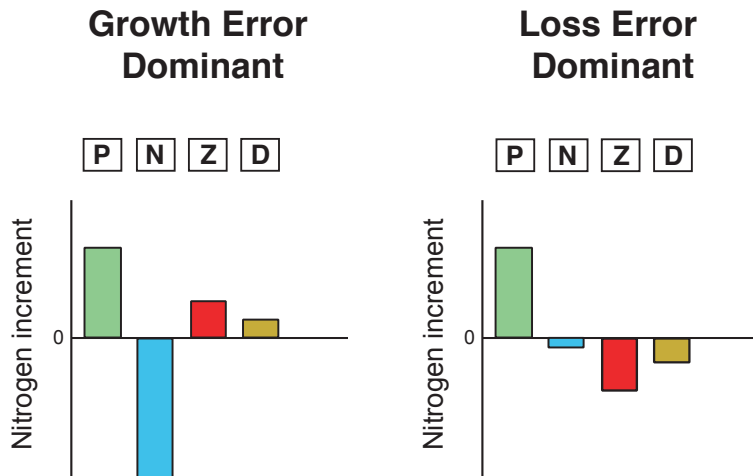

Figure 2: An example of relative nitrogen increments to nutrient, zooplankton and detritus tracers for the same phytoplankton increment under different conditions. (See text for details.)

The examples in Figure 2 show relative nitrogen increments determined by the balancing model in two different cases. In the first case, growth rate errors are assumed to dominate (on the basis of the probability model) so the positive phytoplankton increment is balanced by a negative increment to nutrient. Smaller positive increments are made to zooplankton and detritus. This reflects the fact that turnover of nitrogen from nutrient to zooplankton and detritus, via phytoplankton, can lead to zooplankton and detritus errors that are positively correlated with phytoplankton error. The extent to which this occurs depends on the phytoplankton specific turnover rate (i.e. the rate per unit phytoplankton biomass) and this rate is estimated within the balancing model. In the second case, loss rate errors are assumed to dominate and the same phytoplankton increment is balanced primarily by negative increments to zooplankton and detritus. In the presence of high turnover, the magnitudes of such increments are reduced in favour of a larger negative nutrient increment. The shift reflects the effect of the loss rate errors on nutrient uptake via their effect on the phytoplankton biomass.

The variable inputs to the phytoplankton-driven nitrogen balancing model are the model phytoplankton growth and loss rates, the nitrogen tracer concentrations and

the phytoplankton increment as a negated estimate of the phytoplankton error. The reason for the dependence on the phytoplankton error becomes clear if we consider, for example, the situation where the model phytoplankton growth rate is zero. A positive error, in this case, cannot be due to excessive growth and must therefore be wholly due to insufficient loss, whereas a negative error could be due to errors in growth or loss. Clearly, different balancing factors are appropriate for positive and negative phytoplankton errors. The balancing model is described in detail in Section 2.3.

2.2.3 Increments to Nitrogen Tracers Below the Surface Boundary Layer

The surface layer increments are applied down to the current mixed layer depth. Below this depth, compound increments for each tracer are possible. These are the sum of primary and secondary partial increments. The primary increments are based on the surface increments but take into account the sub-surface tracer concentrations. They extend the analysis updates down below the current mixed layer, reducing the sensitivity of the vertically integrated biomass changes to the phase of the diel mixed layer depth cycle at which the analysis occurs. The basis for the secondary increments is the expectation that nutrient concentrations increase monotonically with depth throughout most of the ocean, due to light-limitation of photosynthetic uptake. These increments are designed to correct the nutrient profile in the event that the surface-layer and primary increments alone would create an unrealistic sub-surface nutrient minimum. The potential exists for this to occur whenever there are positive nutrient increments in shallower layers.

Primary Increments The primary phytoplankton increment is the surface increment scaled to the local background phytoplankton concentration so that the relative increment is constant with depth, subject to the restriction that the size of the increments must decrease monotonically with depth. The size of the increment is therefore reduced if necessary to match that for the level above. This avoids inappropriate updates to deep phytoplankton maxima where the phytoplankton error is unlikely to be positively correlated with that for the surface layer. Balancing increments for the other nitrogen tracers are based on the balancing factors applied at the surface, taking into account the sub-surface tracer concentrations as detailed in Section 2.3.

Primary increments are applied below the current mixed layer depth, down to the maximum depth of the mixed layer over the last 24 h assimilation time-step. The basis for extending the surface layer analysis is the shared history of tracers at all levels above this depth since the last analysis, this being the period during which

the dominant phytoplankton errors are assumed to have evolved. Clearly though, the fraction of the assimilation time-step for which the sub-surface water shares its history with the surface water reduces as the time between detrainment and analysis increases. The calculations based on surface variables are therefore expected to be less reliable if the analysis is performed long after the time of deepest mixing.

In practice, the primary increments are applied to any level wholly or partially above the mixed layer depth maximum. This avoids the potential for immediate attenuation of the surface layer increments by entrainment, into the mixed layer, of water with tracer concentrations not updated in the analysis.

Secondary Increments A set of secondary increments at a point on the 3-D grid consists of a positive nutrient-profile correction increment $\Delta_2 N$ plus negative balancing increments for the organic nitrogen tracers required for nitrogen conservation. These balancing increments are calculated by the nutrient-driven nitrogen balancing model. Non-zero secondary increments occur whenever the nutrient concentration, after the addition of any primary increment, is still less than the analysis concentration for the level above, subject to the constraint that the total nutrient increment must decrease monotonically with depth. The constraint protects against inappropriate updates to pre-existing sub-surface minima which could be due to the presence of different water masses below the surface.

For the purpose of calculating the balancing increments for the organic tracers, the low sub-surface nutrient is assumed to reflect uncorrected errors due to excessive nutrient uptake by the model phytoplankton. However, it is not assumed that all of the missing nitrogen will have remained in the phytoplankton pool. The fraction will depend on the phytoplankton loss rate in the model. Appropriate balancing increments are therefore determined for phytoplankton, zooplankton and detritus on the basis of the loss rate, taking into account the tracer concentrations. This is done by the nitrogen balancing model described in Section 2.4 that takes the nutrient increment as a fixed input.

The balancing increments are:

$$\Delta_2 P = -b_{2P}\Delta_2 N \tag{7}$$
$$\Delta_2 Z = -b_{2Z}\Delta_2 N \tag{8}$$
$$\Delta_2 D = -b_{2D}\Delta_2 N, \tag{9}$$

where balancing factors b_{2P}, b_{2Z} and b_{2D} are the balancing model's negated estimates for the ratios of the background error covariances for phytoplankton, zooplankton and detritus, to the nutrient background error variance. Nitrogen is conserved where possible by attempting to satisfy the equation

$$b_P + b_Z + b_D = 1. \tag{10}$$

2.2.4   Increments to Dissolved Inorganic Carbon and Alkalinity Tracers
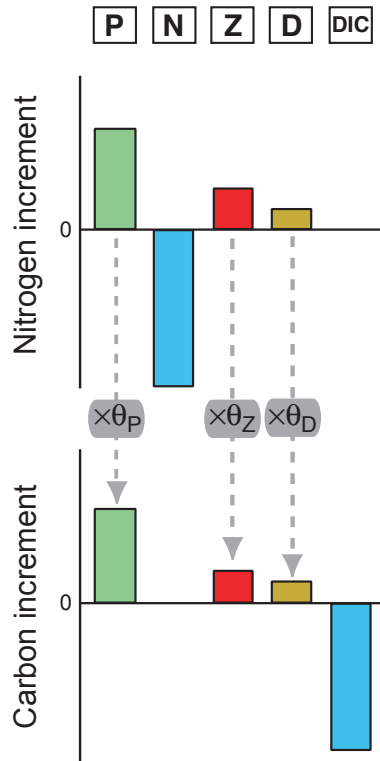

Figure 3: An example showing the derivation of carbon increments from those for nitrogen. The carbon increments include implicit increments to the organic carbon pools and an explicit carbon conserving increment to the DIC tracer. θ_P, θ_Z and θ_D are the biogeochemical model's fixed carbon:nitrogen ratios for phytoplankton, zooplankton and detritus respectively.

15

The nitrogen increments determined imply carbon changes consistent with the biogeochemical model's fixed carbon:nitrogen ratios for each of the organic compartments (phytoplankton, zooplankton and deteritus). Given these changes, balancing increments are calculated for DIC to conserve carbon. The carbon increments for the 'growth error dominant' example are shown in Figure 3.

Alkalinity is incremented in the opposite sense to nutrient, consistent with the model dynamics. No attempt is made to correct alkalinity for possible errors in carbonate precipitation. Although carbonate precipitation is present in the model as a constant fraction of phytoplankton production, this fraction is small and represents a factor that, in reality, varies greatly and will be zero in the absence of calcifiers. Chlorophyll innovations do not therefore provide sufficient information to make a sensible correction.

2.2.5 Summary

In summary, the steps in the assimilation procedure at each point on the horizontal grid are as follows.

- Determine surface chlorophyll increment from 2-D log chlorophyll analysis.

- Determine surface phytoplankton increment using model chlorophyll:nitrogen ratio.

- Determine surface balancing increments for nutrient, zooplankton and detritus using the phytoplankton-driven nitrogen balancing model.

- Assign surface increments to each level above the mixed layer depth.

- For each layer below the mixed layer depth (until done), assign compound increments equal to the sum of any primary and secondary increments calculated as follows.

  Primary increments, above maximum mixed layer depth only:

  - Determine phytoplankton increment to match relative increment at surface (subject to constraint that increment decreases with depth).
  - Determine balancing increments for nutrient, zooplankton and detritus using the phytoplankton-driven nitrogen balancing model.

  Secondary increments, where nutrient concentration after primary increment (if any) is less than in layer above:

– Determine positive nutrient increment needed to match layer above (subject to constraint that increment decreases with depth).

– Determine balancing increments for phytoplankton, zooplankton and detritus using the nutrient-driven nitrogen balancing model.

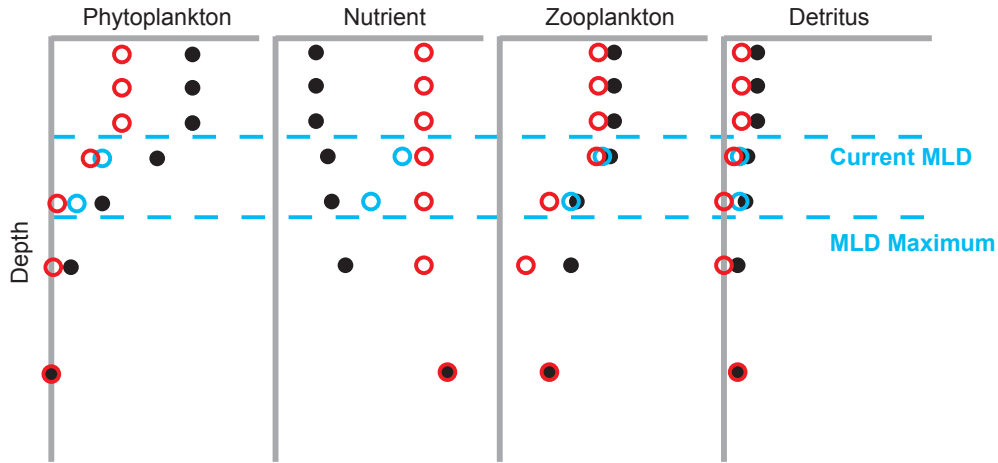• Determine DIC and alkalinity increments for all levels.

Figure 4: An example analysis for the nitrogen tracers showing the background state (black), analysis state (red) and the tracer values corresponding to primary increments only (cyan). The current mixed layer depth and the 24 h maximum mixed layer depth are shown for reference.

Figure 4 shows how the analysis update is broken down into surface, primary and secondary increments for a particular set of nitrogen tracer profiles. Note how more of the required nitrogen for the secondary nutrient increment is taken from zooplankton at the deeper levels. Here, the phytoplankton and detritus concentrations are too low to satisfy the demand.

2.3   Nitrogen Balancing Model for Phytoplankton Increments

The phytoplankton-driven nitrogen balancing model uses a two phase approach to calculate the required balancing factors. Initial approximations, referred to as pre-adjustment factors, are first determined using the model phytoplankton specific growth and loss rates and the model phytoplankton concentration, together with

17

the phytoplankton increment ΔP. Then, in the second phase, the balancing factor estimates are adjusted, taking into account the background state, to satisfy a number of state-dependent restrictions (detailed in Section 2.3.3). These restrictions ensure that the increments do not cause the tracers to go negative and prevent excessive perturbation of the model dynamics by the assimilation. The adjustment phase is particularly important given the large uncertainty associated with the initial estimates. It attempts to maintain nitrogen conservation but, in the case of a positive phytoplankton increment, conservation is only possible if there is sufficient nitrogen available in the other pools, subject to the restrictions imposed.

In the first phase, two values are calculated: the pre-adjustment nutrient balancing factor and a factor f_Z, referred to as the zooplankton loss fraction. This is the estimated zooplankton fraction of the nitrogen loss to the combined zooplankton and detritus pools. The sum of the balancing factors for zooplankton and detritus is derived from the nutrient balancing factor, according to the conservation equation (Equation 6), and the zooplankton loss fraction determines their relative magnitudes:

$$b_Z = (1 - b_N)f_Z \tag{11}$$

and

$$b_D = (1 - b_N)(1 - f_Z). \tag{12}$$

Note that the error variance is assumed to be partitioned between the two pools according to the amount of nitrogen transferred.

Equations 11 and 12 are used during the adjustment phase. Pre-adjustment values for b_Z and b_D, although defined by these equations when the pre-adjustment value for b_N is used, are never actually calculated. This is because b_N is adjusted (if adjustment is required) before b_Z and b_D are first used. The methods for determination of the pre-adjustment nutrient balancing factor and the zooplankton loss fraction are described in Sections 2.3.1 and 2.3.2. Section 2.3.3 describes the adjustment performed to get the final balancing factor values.

2.3.1 Pre-adjustment Nutrient Balancing Factor

The balancing factor for nutrient is the major factor influencing the updates to DIC and alkalinity and therefore has the most direct effect on pCO_2. In the full

scheme, the pre-adjustment nutrient balancing factor is a function of the model phytoplankton concentration P_0, the model phytoplankton specfic growth rate G_0, the model phytoplankton specific loss rate L_0 and the phytoplankton increment ΔP. The model concentration and rates are those for the surface mixed layer, averaged over the assimilation time-step Δt ($= 24$ h). As an alternative, the variable pre-adjustment nutrient balancing factor can be replaced by a constant, specified as an external parameter B_{DEF}. B_{DEF} is also used as a default value in the event that a valid nutrient balancing factor cannot be determined from the model data supplied. The same pre-adjustment nutrient balancing factor is used for surface and sub-surface increments although the final balancing factors can differ due to depth variation in tracer concentrations.

For the purposes of calculating the variable pre-adjustment nutrient balancing factor, the phytoplankton error is expressed as the sum of an error component x attributed to model errors in the phytoplankton growth rate and an error component y attributed to model errors in the phytoplankton loss rate. These unknown components of the estimated phytoplankton error are treated as particular values of random variables X and Y that represent the phytoplankton error due to growth errors and loss errors respectively.

The balancing factor is given by

$$b_{\mathrm{N}} = E\{g(X, Y, P_0, G_0, L_0) \mid X + Y = -\Delta P\}, \tag{13}$$

where E is the expectation operator and g is a function referred to as the nutrient balancing function. This is the expected value of the nutrient balancing function, conditional on the phytoplankton error implied by the increment. A 2-D probability density function for the unknown error components is constructed using the concentration P_0, the specific growth rate G_0 and the specific loss rate L_0. This prior p.d.f. is then used in conjunction with the estimate of the total phytoplankton error $-\Delta P$, to provide the required conditional probability information for calculating the expected value. Details of the p.d.f. and its application are given in the appendix (Section A).

A number of simplifying assumptions are made for the purposes of defining the nutrient balancing function and the probability model:

- Errors in phytoplankton concentration are small compared with the concentration itself.

- Variations in the phytoplankton concentration over the assimilation time-step

and over the model levels that will be within the mixed layer at the time of analysis are small compared with the concentration.

- The specific growth and loss rates within the mixed layer over the assimilation time-step are those having the dominant effect on the mixed layer phytoplankton concentration at the time of analysis; i.e. any recently entrained phytoplankta with significant histories of sub-surface growth and loss rates over the assimilation period make up a small fraction of the standing stock.

When the first assumption of small relative errors is not valid, concentration errors resulting from errors in specific growth rate can cause large errors in phytoplankton losses if the specific loss rate is high and, similarly, concentration errors due to errors in specific loss rate can cause large errors in phytoplankton growth if the specific growth rate is high. The assumption allows us to ignore this interaction between growth and loss and take x as being directly proportional to the error in the specific growth rate and y as being directly proportional to the error in the specific loss rate. All of the assumptions together lead to the following expressions for the error components.

$$x = \Delta t P_0 (G_0 - G) \tag{14}$$

and

$$y = -\Delta t P_0 (L_0 - L), \tag{15}$$

where G and L are the unknown true phytoplankton specific growth and loss rates for the mixed layer.

This is very much a first order approximation but it has two important advantages: firstly, it simplifies the specification of the nutrient balancing function and, secondly, it allows the prior probability distributions for X and Y to be modelled independently. It is not clear whether the absence of interaction in the probability model is likely to be a serious limitation. The potential for this should ideally be investigated further to determine whether a more complex probability model is desirable.

The nutrient balancing function is

$$
\begin{aligned}
g(x, y, P_0, G_0, L_0) &= u_{\mathrm{G}}(x, y) + (1 - u_{\mathrm{G}}(x, y)) B_{\mathrm{MIN}} \\
&\quad + 0.5 \Delta t (1 - B_{\mathrm{MIN}}) B_{\mathrm{T}}(x, y, P_0, G_0, L_0).
\end{aligned} \tag{16}
$$

Here, u_G is the fraction of the phytoplankton error attributed to errors in the growth rate, B_{MIN} (specified as an external parameter) is the expected fraction of phytoplankton losses returning to nutrient over the assimilation time-step (as a result of excretion and detrital breakdown) and B_T is the estimated phytoplankton specific turnover rate for the turnover of nitrogen from the nutrient pool to the combined zooplankton, detritus and nutrient pools via phytoplankton.

If, for the moment, we ignore the last term in Equation 16, the nutrient balancing function gives values of the balancing factor b_N corresponding to nutrient increments that directly balance (i.e. are equal and opposite to) the appropriate fraction of ΔP. The growth fraction u_G varies temporally and spatially reflecting the balance between likely phytoplankton growth and loss contributions. $u_G = 1$ implies bottom-up control of phytoplankton error by growth rate. $u_G = 0$ implies top-down control of phytoplankton error by loss rate. In the case of bottom-up control, we attempt to reverse the error accumulation by transferring nitrogen between phytoplankton and nutrient pools only, so the balancing factor is 1. In the case of top-down control, we want to reverse the error accumulation by transferring nitrogen between the phytoplankton pool and the other three pools in ratios consistent with the likely effect of the loss processes, so the balancing factor is B_{MIN}. Note again the assumption that the error variance is partitioned according to the amount of material transferred.

The last term in Equation 16 is a positive offset designed to correct for the effect of the phytoplankton error on the turnover from the nutrient pool to the zooplankton and detritus pools (an estimated fraction $1 - B_{MIN}$ of the total turnover). The turnover is affected directly because of its dependency on the phytoplankton concentration. The average phytoplankton error for calculating the correction is taken to be half that at the analysis time. Because turnover is uptake of nitrogen that is balanced by losses and does not cause an accumulation of phytoplankton biomass, the turnover correction does not involve any change in the phytoplankton concentration.

The value for the growth fraction u_G depends on the signs and relative magnitudes of the error component x due to errors in growth rate and the error component y due to errors in loss rate. If x and y have the same sign then growth and loss errors are additive and

$$u_G = \frac{x}{x + y}. \tag{17}$$

Otherwise growth and loss errors partly cancel, so $u_G = 1$ if $|x| > |y|$ and $u_G = 0$ if $|x| < |y|$.

The estimated phytoplankton specific turnover rate B_T is the estimated balanced fraction of the phytoplankton specific growth rate, so

$$B_T = \min(G, L). \tag{18}$$

Estimates for G and L in terms of x and y are obtained by re-arranging Equations 14 and 15:

$$G = G_0 - \frac{x}{\Delta t P_0} \tag{19}$$

$$L = L_0 + \frac{y}{\Delta t P_0}. \tag{20}$$

Because of the turnover offset, the nutrient balancing function changes with the model dynamics. An example for particular values of phytoplankton specific growth and loss rates and phytoplankton concentration is shown in Figure 5. The discontinuity in the function occurs where growth and loss errors cancel. The highest values are found where the errors are opposite in sign and growth rate errors exceed loss rate errors. The lowest values are found when the errors are opposite in sign and loss rate errors dominate. The function is undefined beyond the upper limit of x and the lower limit of y: limits that are imposed by the model specific growth and loss rates since the true rates cannot be negative. The pre-adjustment value of the nutrient balancing factor defined by Equation 16 is a weighted average of the nutrient balancing function along the line $y = -(x + \Delta P)$, the weighting function being derived from the estimated joint probability density $p(x, y)$ defined in Section A.1.

The behaviour of the nutrient balancing function is such that values of b_N greater than 1, implying zooplankton and detritus increments that are positively correlated with the phytoplankton increments, can occur even at low turnover rates when phytoplankton growth rate errors dominate (u_G close to or equal to 1). This happens during bloom periods when the growth rate is much greater than the loss rate and is consistent with the tendency for zooplankton distribution to be positively correlated with food supply during periods of rapid phytoplankton growth. As zooplankton biomass increases, the phytoplankton specific loss rate increases due to grazing. This increases the turnover offset but also tends to increase the loss errors, reducing u_G, so the positive correlation is lost.
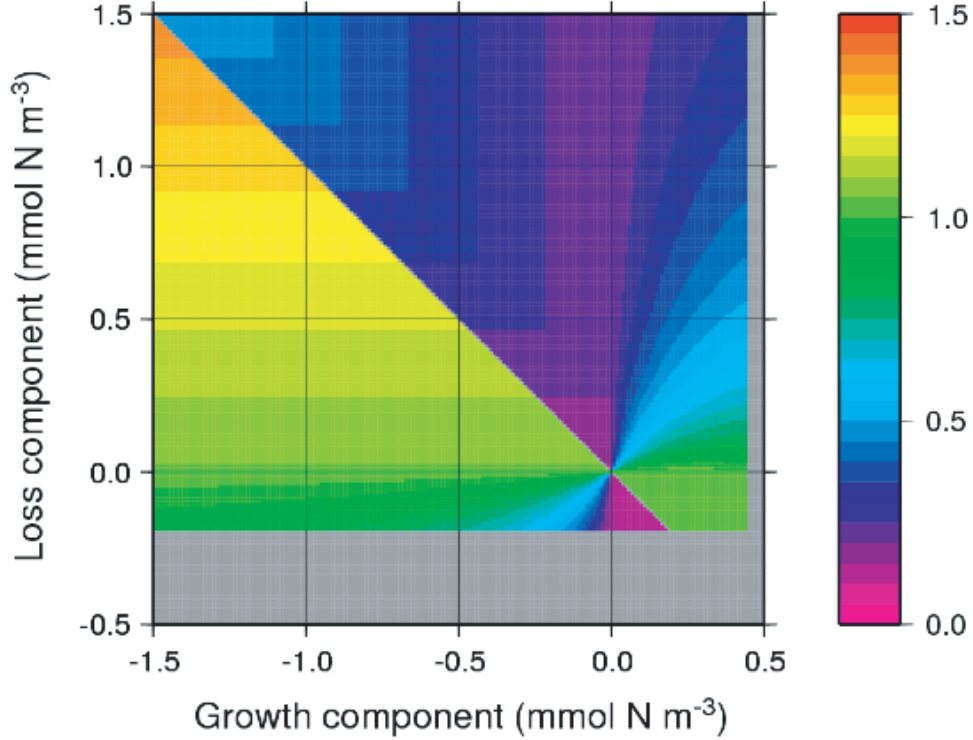
Figure 5: Variation of a nutrient balancing function with the phytoplankton error component due to growth errors x and the phytoplankton error component due to loss errors y. In this example: phytoplankton specific growth rate $G_0 = 0.45$ d^{-1}; phytoplankton specific loss rate $L_0 = 0.2$ d^{-1}; phytoplankton concentration $P_0 = 1$ mmol m^{-3}; fraction of phytoplankotn losses to nutrient $B_{\mathrm{MIN}} = 0.1$.

2.3.2 Zooplankton and Detritus Balancing Factors: the Zooplankton Loss Fraction

The zooplankton loss fraction that determines the relative magnitude of the zooplankton and detritus balancing factors is given by

$$f_{\mathrm{Z}} = \max(0, a_0 - a_1 P), \tag{21}$$

where P is the phytoplankton concentration and a_0 and a_1 are external parameters (both positive).

As defined earlier, f_{Z} is the estimated zooplankton fraction of the nitrogen loss to the combined zooplankton and detritus pools. This loss is due to grazing and mortality.

Grazing losses transfer nitrogen primarily to the zooplankton pool with smaller fractions going to detritus and nutrient, while mortality losses transfer nitrogen to primarily to detritus with a small fraction going to nutrient. High values of f_Z are therefore consistent with grazing dominated losses and low values with mortality dominated losses. Non-zero values of a_1 allow for the effect of high phytoplankton concentrations, which tend to increase phytoplankton specific mortality and decrease phytoplankton specific grazing. Increasing P reduces f_Z, transferring error variance from zooplankton to detritus to reflect the greater likelihood of mortality dominated losses. Zooplankton and, to a lesser extent, detritus concentrations also affect the balance between mortality and grazing. However, these concentrations are likely to be less reliable than phytoplankton in the assimilating model and no attempt is made to allow for this.

The value used for phytoplankton concentration in Equation 21 depends on the sign of the zooplankton and detritus increments (which is always the same for the two pools). Negative increments to these pools are designed to correct for excessive losses which have already occurred in the model, so the model concentration is used (i.e. $P = P_0$). Positive increments are designed to generate new loss to correct for insufficient loss in the model. In this case, it is more appropriate to use the best estimate of the true concentration, averaged over the assimilation time-step, i.e.

$$P = P_0 + 0.5\Delta P. \tag{22}$$

The loss fraction used for sub-surface balancing is normally the same as that at the surface (down to the 24 h mixed layer depth maximum). However, because vertical variations in nutrient concentration affect the amount of nitrogen available from nutrient, the adjustment process can cause the sign of the zooplankton and detritus increments to change with depth. When this occurs, the zooplankton loss fraction is based on the alternate surface layer phytoplankton concentration.

When correcting for excessive loss in the model, it could be argued that the partitioning of the error variance (between zooplankton, detritus *and* nutrient) should reflect the actual partitioning of nitrogen losses that has occurred in the model. However, the assumption that error variance is partitioned according to the amount of material transferred is not considered sufficiently reliable to justify the extra complexity that would be entailed in keeping account of losses to the individual pools.

Finally, because error in the phytoplankton concentration will have affected the partitioning of total phytoplankton losses in the model over the assimilation time-step, not just the loss error, It could be argued that there should be a corrective transfer of material between the zooplankton and detritus pools to allow for this. This would be analogous to the effect of the turnover offset in Equation 16 in transferring

material to or from the nutrient pool. However, once again, such a correction is not considered justified because of the inherent uncertainty involved.

2.3.3   Balancing Factor Adjustment

The balancing factor adjustment phase ensures that the following restrictions are strictly enforced, while attempting to conserve nitrogen if possible.

Firstly, to avoid excessive perturbation of the phytoplankton specific growth rate, nutrient increments are not allowed to change the nutrient limitation factor controlling growth rate in the model by more than a specified amount. The nutrient limitation factor in the HadOCC model, as a function of the nutrient concentration N, is

$$Q = \frac{N}{k_N + N},\tag{23}$$

where k_N is the model's half-saturation concentration parameter for nutrient limitation of photosynthesis. Specifically Q cannot be reduced by more than a factor of R_Q or amplified by more than a factor of A_Q as a result of the nutrient increment. (The factors R_Q and A_Q are specified as external parameters.) In addition, zooplankton concentration cannot be reduced by more than a factor of R_Z or amplified by more than a factor of A_Z by the increment. (Again, the factors R_Z and A_Z are specified as external parameters.) Finally detritus cannot be reduced below zero.

The nutrient restriction is particularly important. It only has a significant impact at low nutrient concentrations. Under these conditions, the negative correlation between errors in phytoplankton and nutrient associated with growth breaks down. This happens as the control of growth by nutrient limitation reduces its impact on the nutrient concentration. Large nutrient balancing increments are therefore inappropriate at low nutrient concentrations. Critically, in fact, they will tend to cause undesirable positive feedback by increasing the magnitude of the growth rate errors. Fortunately high values of the pre-adjustment nutrient balancing factor are less likely in this situation because of the low growth rate but it is prudent not to rely on this.

Adjustment of the balancing factors to statisfy the restrictions is carried out as follows.

Firstly, if the background nutrient is too low to allow the nutrient increment implied by the pre-adjustment value of b_N to be satisfied, given the restriction on Q, then

b_N is reduced to satisfy the restriction. b_Z and b_D are calculated from Equations 11, 12 and 21, using the new value. Note that if the original b_N is greater than 1 and is reduced below 1 then the sign of the zooplankton and detritus increments changes which affects the value of the zooplankton loss fraction estimate.

Next, if the background zooplankton is too low to allow the required zooplankton increment to be satisfied, given the restrictions, then b_Z and b_D are adjusted to transfer the remainder of the zooplankton increment to detritus, implying an increased error contribution from mortality relative to grazing. This situation is most likely to arise when zooplankton concentration is small compared with phytoplankton concentration, in which case mortality is likely to be the dominant loss process. The transfer is therefore appropriate and will tend to compensate for the absence of a zooplankton dependency in the zooplankton loss fraction model (Equation 21).

If there is insufficient detritus to satisfy a required negative detritus increment, then the remainder of the detritus increment must be transferred to the combined zooplankton and nutrient pools by reducing b_D. If there is still scope for adjusting the zooplankton (without violating the zooplankton reduction restriction) then as much of the excess increment as possible is transferred to zooplankton by increasing b_Z. By doing this, we attempt to contain the adjustment within the combined zooplankton and detritus pools and avoid affecting b_N. Any remaining component of a negative detritus increment is transferred to nutrient by increasing b_N.

If, at this point, the nutrient increment is too large then b_N is reduced to satisfy the Q restriction and nitrogen is not conserved.

2.4   Nitrogen Balancing Model for Nutrient Increments

This section describes the nutrient-driven nitrogen balancing model used to determine the secondary increments to phytoplankton, zooplankton and detritus tracers. These balance positive nutrient increments potentially required for nutrient profile correction. Once again, a two-phase approach is used. In the first phase, the pre-adjustment value of the phytoplankton balancing factor b_P and the value of the zooplankton loss fraction f_Z are calculated. The balancing factors for zooplankton and detritus are then given by

$$b_Z = (1 - b_P)f_Z \tag{24}$$

and

$$b_D = (1 - b_P)(1 - f_Z).$$ (25)

In the second phase, the values for the balancing factors are adjusted to take into account the tracer concentrations.

2.4.1 Pre-adjustment Balancing Factors

The pre-adjustment phytoplankton balancing factor is given by

$$b_P = \frac{1}{1 + 0.5\Delta t(1 - B_{\mathrm{MIN}})L_0}.$$ (26)

The second term in the denominator is the estimated amount of the excess uptake $\Delta_2 N$ that is lost to zooplankton and detritus, expressed as a fraction of that remaining as phytoplankton $\Delta_2 P$. In deriving this expression, it is assumed that the model phytoplankton error is zero after the previous analysis and evolves linearly during the assimilation time-step so that the mean error is $-0.5\Delta_2 P$. For levels below the deepest into which the surface mixed layer penetrated over the 24 h assimilation time-step, the 24 h average mixed layer loss rate L_0 is replaced, in the present version of the scheme, by the current loss rate determined from the background state. This is to avoid the need for a large amount of extra storage in the 3-D model to compute 24 h averages at all depth levels.

The zooplankton loss fraction is calculated as described in Section 2.3.2. The zooplankton and detritus increments are always negative, being corrections for excess turnover of nitrogen from nutrient via the phytoplankton. This is turnover which has occurred in the model, so the relevant phytoplankton concentration for calculating the zooplankton loss fraction is the model concentration rather than the true concentration. For levels down to the deepest into which the surface mixed layer penetrated, the 24 h average mixed layer concentration P_0 is used. For deeper levels, the current background concentration is used, again purely to avoid the storage overhead.

2.4.2 Balancing Factor Adjustment

The balancing factors are adjusted to satisfy the restrictions on zooplankton and detritus reduction described in Section 2.3.3 and a further restriction that the phy-

toplankton cannot be reduced by more than a factor R_P by the secondary increment. R_P is an external parameter. The phytoplankton restriction refers to the permissible change relative to the intermediate concentration that is obtained after applying the primary increment to the background concentration. The zooplankton and detritus restrictions apply to the size of the full increments (i.e. the sum of the primary and secondary increments) relative to the background concentrations.

Adjustment of the balancing factors to statisfy the restrictions is carried out using rules analogous to those described in Section 2.3.3. This works as follows.

Firstly, if the phytoplankton is too low to permit application of the negative phytoplankton increment implied by the pre-adjustment value of b_P, then b_P is reduced to satisfy the restriction. b_Z and b_D are calculated from Equations 24, 25 using the new value. Then, if the background zooplankton is too low for the required zooplankton increment, the remainder of the zooplankton increment is added to the detritus increment. If there is insufficient detritus for the required detritus increment, the remainder of the detritus increment is transferred to the combined zooplankton and phytoplankton pools. Ideally, changes to b_P, are to be avoided so, this transfer is contained within the zooplankton pool as far as possible. If required changes to b_P are too large they are reduced to satisfy the restriction and nitrogen is not conserved.

3 Initial Evaluation

The results of an initial off-line evaluation of the material balancing scheme's performance are presented here. These are obtained from identical twin experiments carried out in a 1-D test-bed. The test-bed comprises a 0-D model of HadOCC biogeochemistry forced by physical fields from a 3-D FOAM-HadOCC integration at 1° resolution with assimilation of sea-surface temperature and vertical temperature profiles. Output data representative of 4 different locations along the 20°W meridian (30°N, 40°N, 50°N, 60°N) are used to drive the test-bed model.

3.1 Method

3.1.1 The 1-D Test-bed

The 1-D test-bed incorporates the version of the HadOCC biogeochemistry model used in the 3-D FOAM-HadOCC model, configured with the same depth levels and the same model time-step (1 h). There are 20 unevenly spaced vertical levels with

the highest resolution (10 m) near the surface. The biogeochemistry is forced by solar irradiance, mixed layer depth, surface temperature and surface salinity and includes relaxation to climatological nutrient profiles as in the 3-D model. The temperature and salinity affect only the derivation of pCO_2 from DIC and alkalinity and the air-sea flux of CO_2. There are a number of important differences between the test-bed and the 3-D model. In particular, there are no horizontal fluxes and no vertical velocities in the present version of the test-bed and no vertical diffusion below the mixed layer. The 3-D model includes surface fluxes of fresh water, which have an important effect on DIC and alkalinity, and a variable transfer velocity for the air-sea flux of CO_2, forced by the wind-mixing energy. These effects are also not included in the present version of the test-bed.

Another, more subtle, but nevertheless important difference between the test-bed and the 3-D model is that the biogeochemistry in the test-bed responds to a pre-defined physical state, whereas the 3-D model run includes temperature assimilation with no biogeochemical balancing increments. This can lead to a mismatch between the biogeochemical tracers and the physical structure in the 3-D model which can in turn affect the dynamics.

The solar irradiance forcing data are daily averaged data from the Met Office Numerical Weather Prediction Model. These are the same data as used in the 3-D model. The mixed layer depth, temperature and salinity data are based on daily means from the 3-D model with the values interpreted as 1200 GMT values and linearly interpolated between days. After each 1 hour time-step, the tracer values are homogenized over the mixed layer as in the 3-D model. In practice, a mixed layer mean is calculated, taking into account partial mixing of the layer below according to the ratio of the penetration of mixing into this layer to the layer thickness. This mean is applied at all levels within the mixed layer and any change to the sum over these levels is balanced by a change at the level below, conserving the total tracer in the water column. This partial mixing introduces diffusion between the mixed layer and the layer below and is a crude substitution for the vertical diffusion scheme in the 3-D model. Ideally, the test-bed would include the tracer diffusion scheme forced by vertical tracer diffusivities from the 3-D model output but this feature is not currently implemented.

Output from the test-bed is compared with the 3-D model in Figure 6. Both have the same initial tracer profiles at the beginning of February. Clearly there are significant differences in the annual cycles. In particular, the summer nutrient levels are lower in the test-bed and it is likely that much of the discrepancy is due to the absence of vertical diffusion below the mixed layer. However, the annual cycles obtained are representative of the 3-D model variation along the 20°W meridian from the eutrophic conditions in the sub-arctic to the oligotrophic conditions in the sub-tropical gyre. This is sufficient to allow experiments to be carried out in the test-
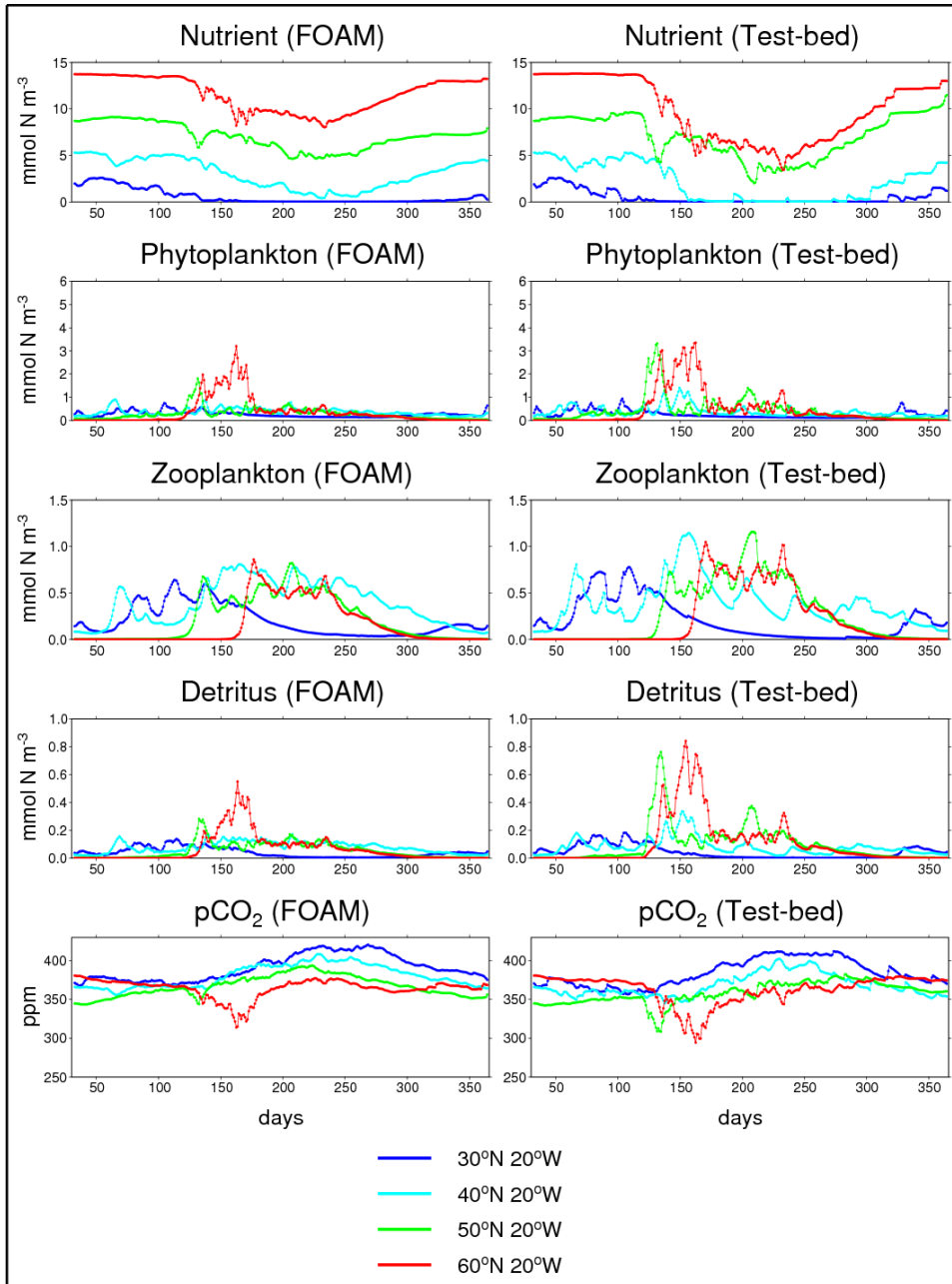
Figure 6: Comparison between 3-D model results (left) and test-bed results (right) for the year 2000. Time series shown are for the surface layer from Feb-Dec.

bed with conditions that are broadly representative of the eastern North Atlantic. At present though, the test-bed emulation of 3D FOAM-HadOCC is inadequate for meaningful tests with real-world data.

3.1.2 Assimilation Experiments

The tests are based on twin experiments in which synthetic data are assimilated. The trajectories representing the true system state evolution are generated by randomly perturbing model parameters during the integration. To allow error statistics for assimilating and non-assimilating runs to be estimated, an ensemble of 100 possible truths is generated in this way. The parameters varied are:

- maximum photosynthetic rate (i.e. maximum phytoplankton specific growth rate)

- initial slope of the photosynthesis-irradiance curve

- half-saturation concentration for nutrient limitation of photosynthesis

- phytoplankton concentration dependency of phytoplankton specific mortality rate (from zero mortality at concentrations less than 0.01 mmol N m^{-3})

- maximum zooplankton specific grazing rate

- half-saturation food concentration for zooplankton grazing

- constant offset for zooplankton specific mortality rate (dominant at low zooplankton concentrations)

- zooplankton concentration dependency of zooplankton specific mortality rate

- detrital sinking velocity

- carbonate precipitated per unit primary production (affects carbonate system only)

Each parameter was initialized from a normal distribution with a mean $\bar{p}$ equal to the nominal parameter value (i.e. that used in the standard run) and a standard deviation σ_p of 25% of its nominal value. At the start of each day of integration, each parameter value p was both randomly perturbed and weakly relaxed toward its nominal value $\bar{p}$ to get a new value

$$p' = p + 0.5 Z \sigma_p - 0.1(p - \bar{p}), \tag{27}$$

where Z is a normally distributed random variable with zero mean and unit standard deviation. The value p' was then adjusted if necessary to bring it within the range $\bar{p} \pm 2\sigma_p$.
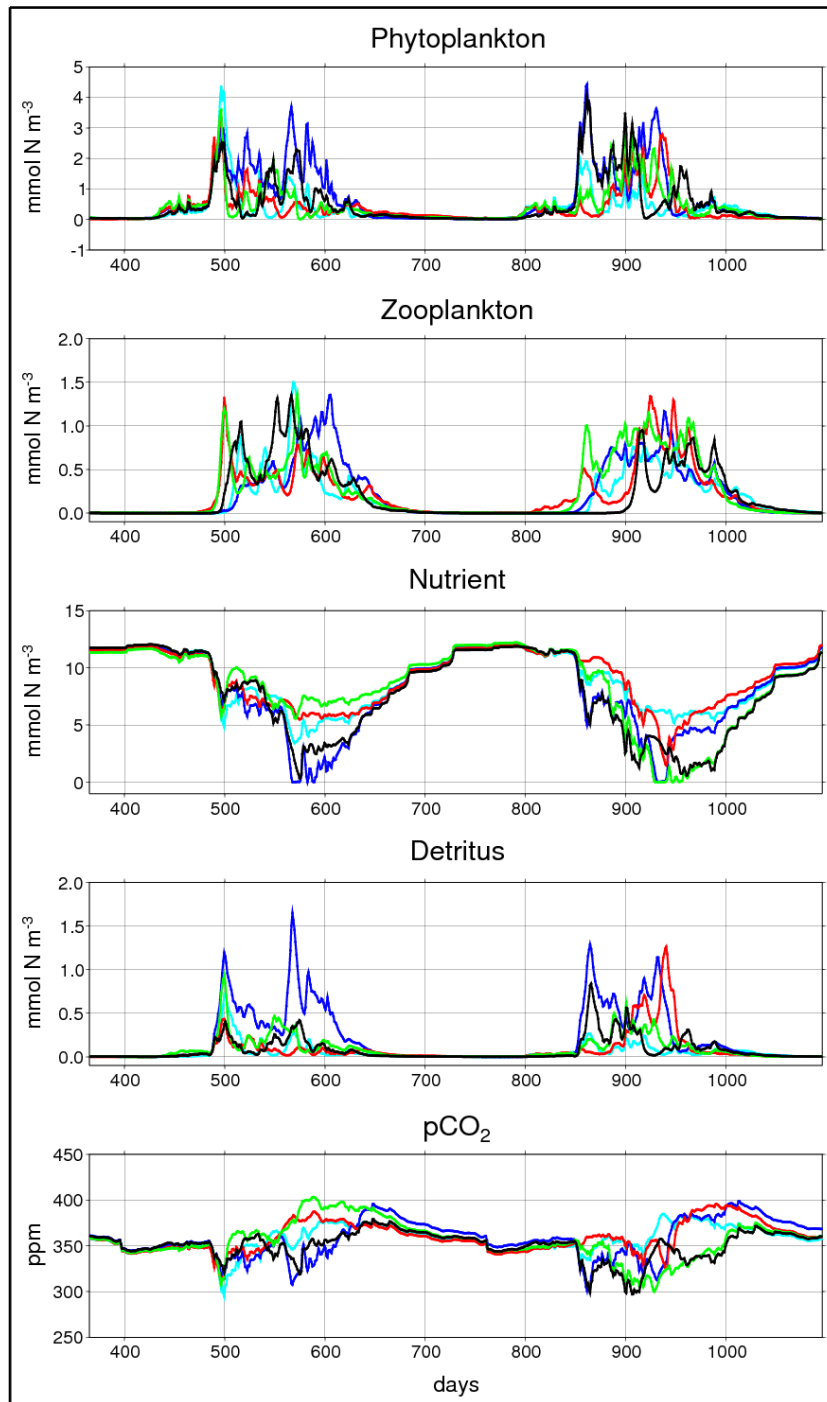
Figure 7: True system trajectories for 5 ensemble members at 50°N 20°W. The two calendar years of the assimilation experiments are shown. The time origin is the start of the previous calendar year.

Assimilation results using the parameter values specified in Table 1 are presented in Section 3.2 for two complete calendar years. The ensemble runs and the standard run, into which data from the ensemble are to be assimilated, are all started with initial conditions extracted from the 3-D model at the beginning of February and allowed to run for 11 months prior to assimilation. The same physical forcing is used for each of the 3 calendar years. Figure 7 shows examples of the true system trajectories for 5 ensemble members at 50°N 20°W.

Assimilation runs were performed with and without simulated observation error and their results were compared with the free run. To simulate measurement errors, normally distributed offsets with a standard deviation σ_o of 0.15 $\log_{10}$ units were added to the true log-transformed daily mean chlorophyll values to get the observation values. This is consistent with the ±35% target accuracy for SeaWiFS chlorophyll a which is largely achieved for open ocean (Case 1) waters (Hooker and McClain, 2000). The log chlorophyll increments were then given by

$$\Delta logchl = K(logchl_o - logchl_b),\qquad(28)$$

where $logchl_o$ is the observed value of log-transformed chlorophyll, $logchl_b$ is the background value and K is the gain. The optimal value of K which minimizes the analysis error variance is

$$K = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2},\qquad(29)$$

where σ_b is the standard deviation of the background errors. The gain was optimized by iteratively running the assimilation for all ensemble members and all stations to determine the background error statistic.

The assimilation runs are compared against the free run by comparing their r.m.s. error and bias statistics.

3.2 Results

3.2.1 Overall Performance of the Chlorophyll Assimilation Scheme

Figs. 8-11 show the performance of the scheme for synthetic daily chlorophyll observations with and without observation error (Experiments A and B respectively). In

Table 1: Nitrogen balancing scheme parameters

Parameter	Symbol	Unit	Value
Minimum phytoplankton increment	ΔP_{MIN}	mmol N m^{-3}	0.0001
Default pre-adjustment nutrient balancing factor	B_{DEF}		0.6
Fraction of phytoplankton loss to nutrient	B_{MIN}		0.1
Base zooplankton loss fraction	a_0		0.8
Phytoplankton dependency of zooplankton loss fraction	a_1	(mmol N m^{-3})$^{-1}$	0.05
Nutrient limitation max. reduction factor	R_{Q}		1.1
Nutrient limitation max. amplification factor	A_{Q}		1.1
Zooplankton max. reduction factor	R_{Z}		2
Zooplankton max. amplification factor	A_{Z}		2
Phytoplankton max. reduction factor [1]	R_{P}		10
Probability model parameters for the variable pre-adjustment nutrient balancing factor calculation (see Section A)			
Reliability of model specific loss rate [2]	r		1
Low rate bias correction for growth rate estimator	β_{G}	d^{-1}	0.05
Low rate bias correction for primary loss rate estimator	β_{L}	d^{-1}	0.05
Low rate bias correction for alternative loss rate estimator [3]	β_{M}	d^{-1}	–
Error s.d. for growth rate estimator	σ_{G}	$\log_{10}$ units	0.2
Error s.d. for primary loss rate estimator	σ_{L}	$\log_{10}$ units	0.4
Error s.d. for alternative loss rate estimator [3]	σ_{M}	$\log_{10}$ units	–

[1]Only applies to secondary increments.

[2]This is the fractional weight given to the primary loss rate estimator: the model phytoplankton specific loss rate. The remainder is given to the alternative loss rate estimator: the model phytoplankton specific growth rate.
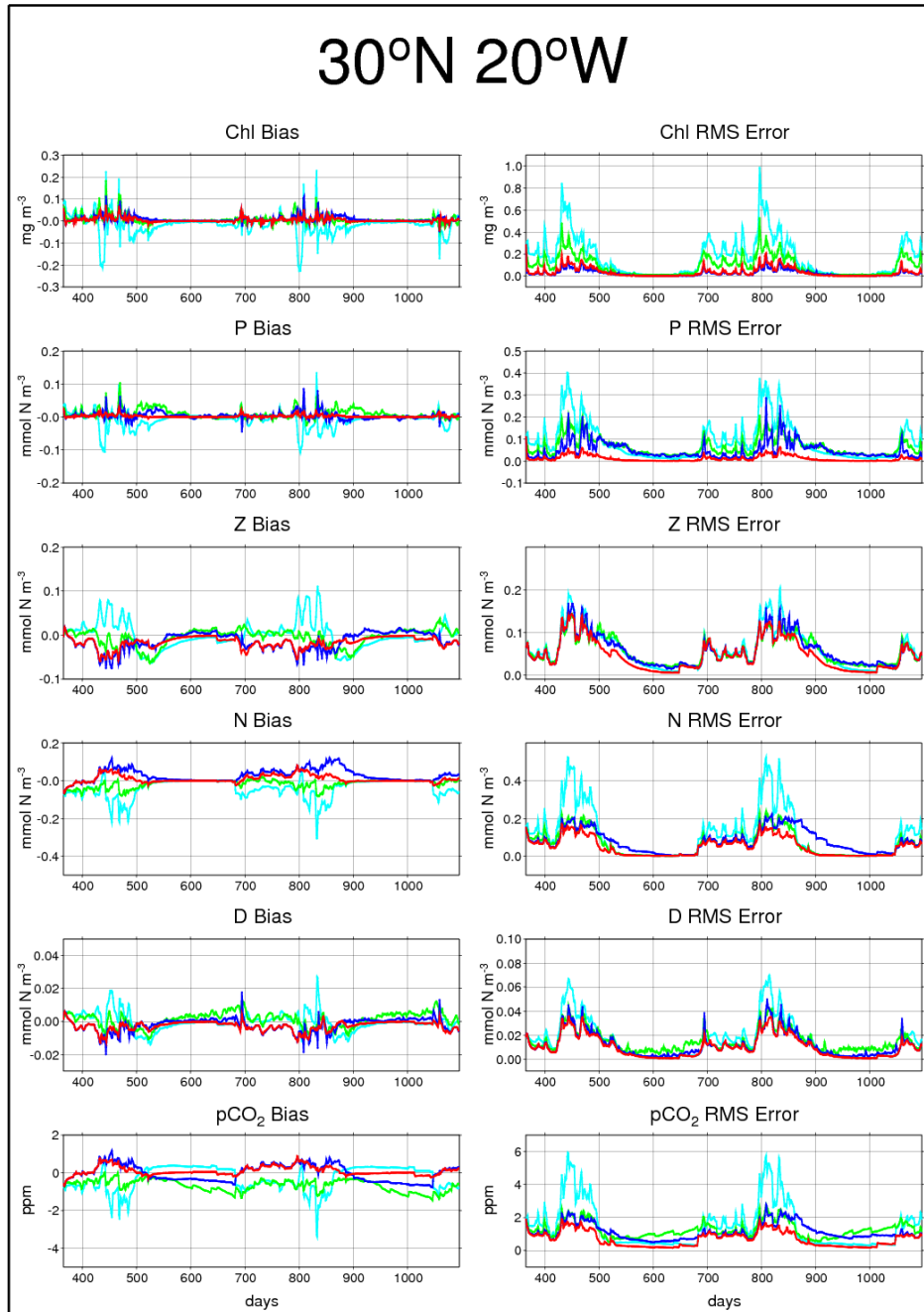
[3]Not applicable when $r = 1$.

Figure 8: Bias and r.m.s. errors for daily mean surface concentrations at 30°N 20°W for Experiment A: assimilating chlorophyll with observation errors (green), Experiment B: assimilating chlorophyll with no observation error (blue) and Experiment C: assimilating chlorophyll and phytoplankton with no observation error (red). The free run is shown in cyan.
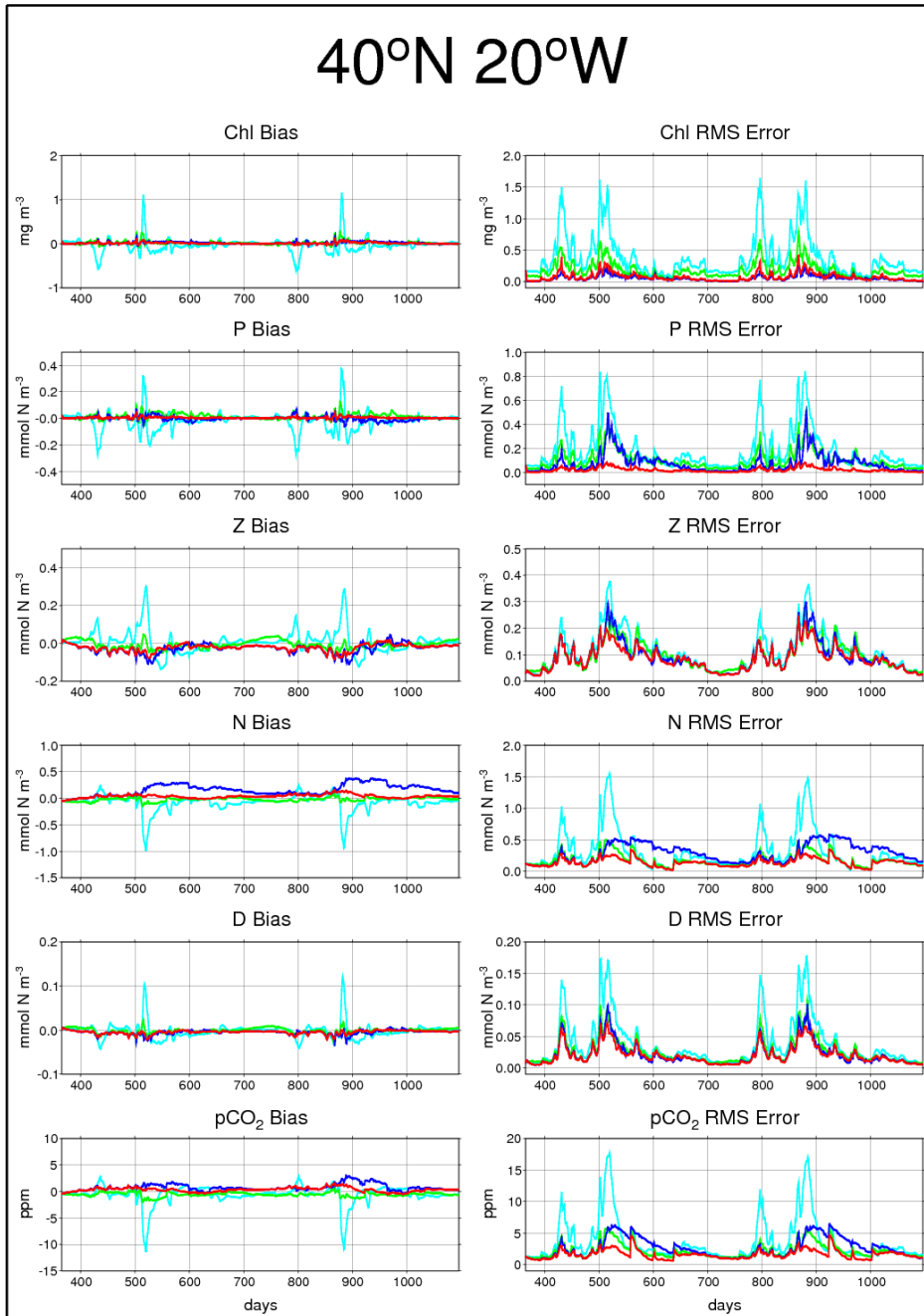
Figure 9: Bias and r.m.s. errors for daily mean surface concentrations at 40°N 20°W for Experiment A: assimilating chlorophyll with observation errors (green), Experiment B: assimilating chlorophyll with no observation error (blue) and Experiment C: assimilating chlorophyll and phytoplankton with no observation error (red). The free run is shown in cyan.
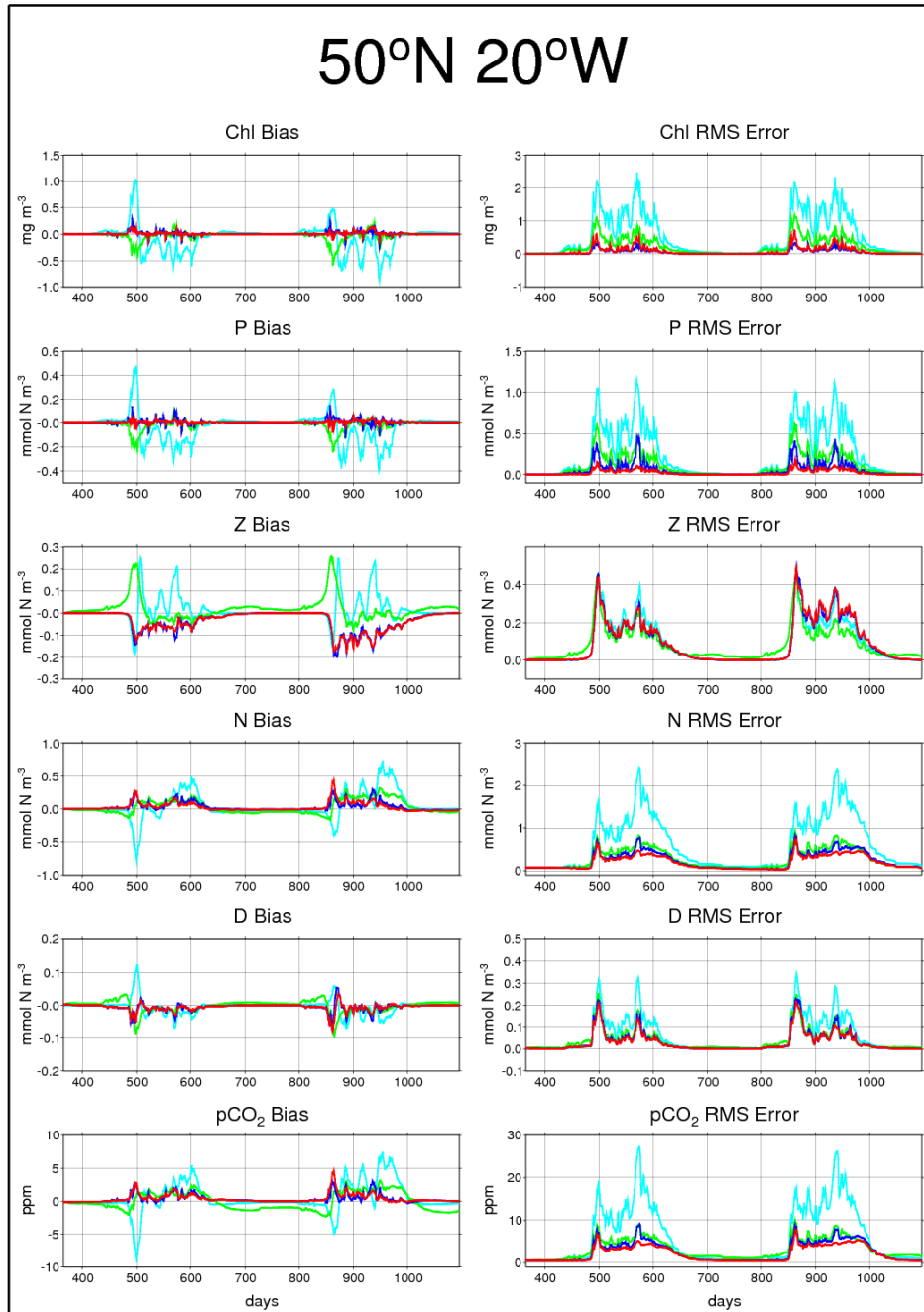
Figure 10: Bias and r.m.s. errors for daily mean surface concentrations at 50°N 20°W for Experiment A: assimilating chlorophyll with observation errors (green), Experiment B: assimilating chlorophyll with no observation error (blue) and Experiment C: assimilating chlorophyll and phytoplankton with no observation error (red). The free run is shown in cyan.
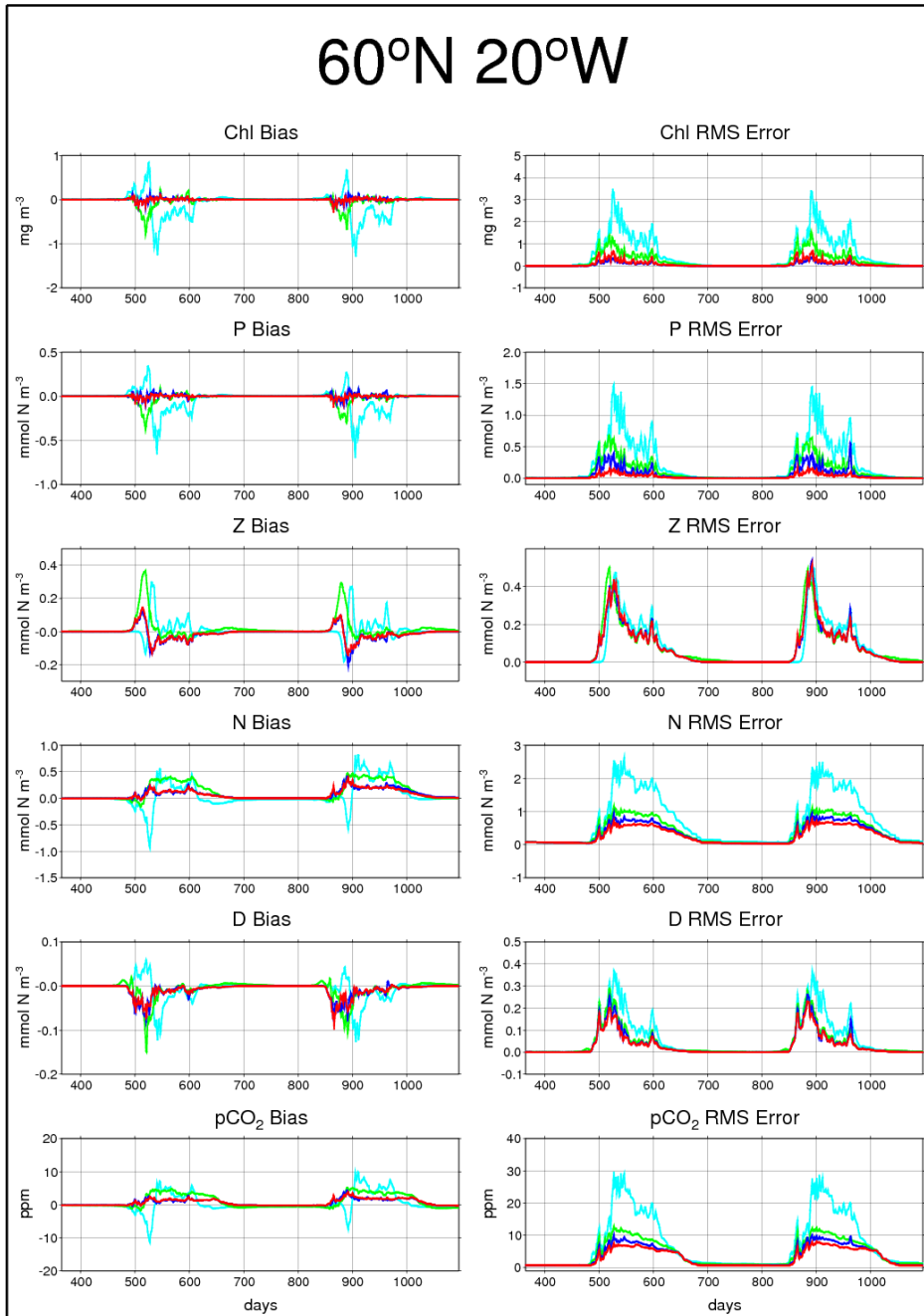
Figure 11: Bias and r.m.s. errors for daily mean surface concentrations at 60°N 20°W for Experiment A: assimilating chlorophyll with observation errors (green), Experiment B: assimilating chlorophyll with no observation error (blue) and Experiment C: assimilating chlorophyll and phytoplankton with no observation error (red). The free run is shown in cyan.

the former case the optimal gain was found to be 0.4. In the latter case, the gain is set to 1 so that there is no analysis error in chlorophyll. Also shown, for comparison, are the results for an assimilation run where both chlorophyll and phytoplankton nitrogen are assimilated with no measurement error such that the analysis errors for both variables are zero in the surface layer (Experiment C). The comparison allows the direct effect of errors in the model nitrogen:chlorophyll ratio during the analysis to be seen. The first calendar year in each figure (time = 365-730 days) is preceeded by 11 months with no assimilation and the second is preceeded by a years worth of assimilation. There are only minor differences in the results between years.

Unsurprisingly, the results for chlorophyll from Experiment B, where there is no observation error, are better than those from Experiment A at all latitudes. While the same is true for phytoplankton, the improvement is much less. The impact of observation error on the scheme's performance for the other variables is more complicated and difficult to generalize. For nutrient and pCO_2 the high latitude results (Figs. 10-11) are slightly better when there is no observation error but the 40°N results are actually worse, with long periods of positive bias in spring and summer. A similar pattern is seen for nutrient at 30°N. The removal of observation error also leads to worse performance for zooplankton at 50°N in the summer, following a period of improvement in the spring. This is especially noticeable in the second year of assimilation. The degraded summer performance in Experiment B appears to be associated with a reversal of the late spring zooplankton bias between the two experiments, suggesting an over-correction for positive errors. Such an over-correction might be reduced by tightening the restriction on the size of the zooplankton increments, although the possibility of undesirable side effects cannot be discounted.

Comparing the assimilation results with those for the free run, it can be seen at the higher latitudes (Figures 10 and 11) that all variables with the exception of zooplankton show major improvements after chlorophyll assimilation throughout most of the year when errors are significant, with little sign of any detrimental effect. This is true for both experiments (Experiments A and B). For zooplankton, the r.m.s. errors are broadly similar to the free run. The situation is less satisfactory at the lower latitudes. At 30°N, summer and early autumn errors in pCO_2 are greater than those for the free run in both experiments due to negative biases introduced by the assimilation. In Experiment B, where there is no observation error, the degraded performance is extended to 40°N as well (although here the pCO_2 bias is positive) and the low latitude summer nutrient errors are also greater than those for the free run. Also notable, in both experiments, is that there is much less improvement in phytoplankton over the free run at the lower latitude stations. At 30°N there are actually clear periods in early summer when phytoplankton error is greater than that for the free run.

The difference between the results at high and low latitudes is largely due to the effect of nutrient limitation of phytoplankton growth on the nitrogen:chlorophyll ratio, which increases its variance dramatically in the model at low latitudes: the standard deviation over the ensemble at 30°N is 0.68 mmol N (mg Chl)$^{-1}$, with values ranging from 0.28 to 2.51, compared with a standard deviation of 0.10 mmol N (mg Chl)$^{-1}$ at 60°N, with values from 0.26 to just 1.26. In the experiment with combined chlorophyll and phytoplankton assimilation (Experiment C), in which model errors in nitrogen:chlorophyll do not affect the analysis, there are improvements in pCO$_2$ and nutrient at all stations and the low latitude pCO$_2$ biases are much reduced. From this result it is clear that, in the twin experiment at least, model drift in nitrogen:chlorophyll is a major source of error.

In Experiment C, there are clear improvements over the free run in all variables except zooplankton at all locations over almost all parts of the annual cycle where significant errors occur. The performance for zooplankton remains less satisfactory than for the other variables. The only r.m.s. errors which are notably worse than for the free run in Experiment C are the summer zooplankton errors at 50°N and the zooplankton errors for a short period in spring at 60°N. The encouraging results for Experiment C show that the scheme has the potential to make major improvements in surface pCO$_2$ estimates over a wide range of oceanic conditions if an effective method can be found for correcting errors in the nitrogen:chlorophyll ratio.

3.2.2   Impact of the Nitrogen Balancing Scheme

To examine the impact of the nitrogen balancing scheme, the combined chlorophyll and phytoplankton assimilation (without observation errors) was repeated with nitrogen balancing switched off (Experiment D). In this experiment, only the phytoplankton and DIC tracers are updated in the analysis; carbon is conserved but nitrogen is not. A further set of runs (Experiment E) was performed with nitrogen balancing using a constant pre-adjustment nutrient balancing factor of 0.6 in place of the variable factor from the probability model for growth and loss error contributions used in Experiment C. Given $B_{\mathrm{MIN}} = 0.1$ (Table 1), a nutrient balancing factor of 0.6 is approximately equivalent to the value from Equation 16 with a growth fraction of phytoplankton error $u_{\mathrm{G}} = 0.5$ and a typical phytoplankton specific turnover value $B_{\mathrm{T}} = 0.1$. The results for the two experiments are compared with those for Experiment C in Figures 12 to 15.

The most obvious point to note, in relation to Experiment D, is that without nitrogen balancing the nutrient errors are much worse than the free run at all stations. These errors are mostly associated with positive biases in the nutrient concentration. Moreover, unlike the other experiments where there are only relatively minor
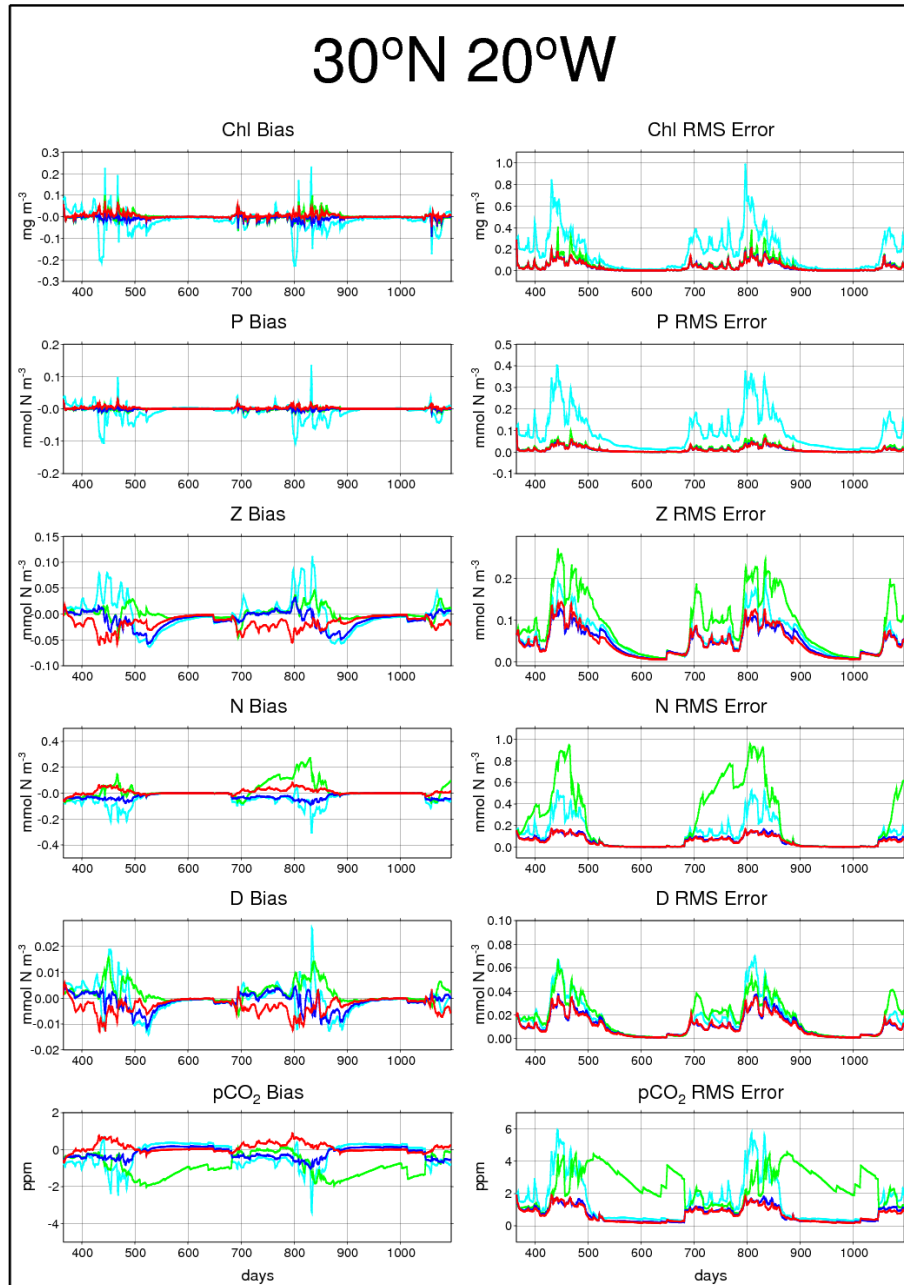
Figure 12: Bias and r.m.s. errors for daily mean surface concentrations at 30°N 20°W for Experiment D: no nitrogen balancing (green), Experiment E: nitrogen balancing with constant pre-adjustment nutrient balancing factor (blue) and Experiment C: nitrogen balancing with variable pre-adjustment nutrient balancing factor (red). All experiments are for chlorophyll and phytoplankton assimilation with no observation error. The free run is shown in cyan.
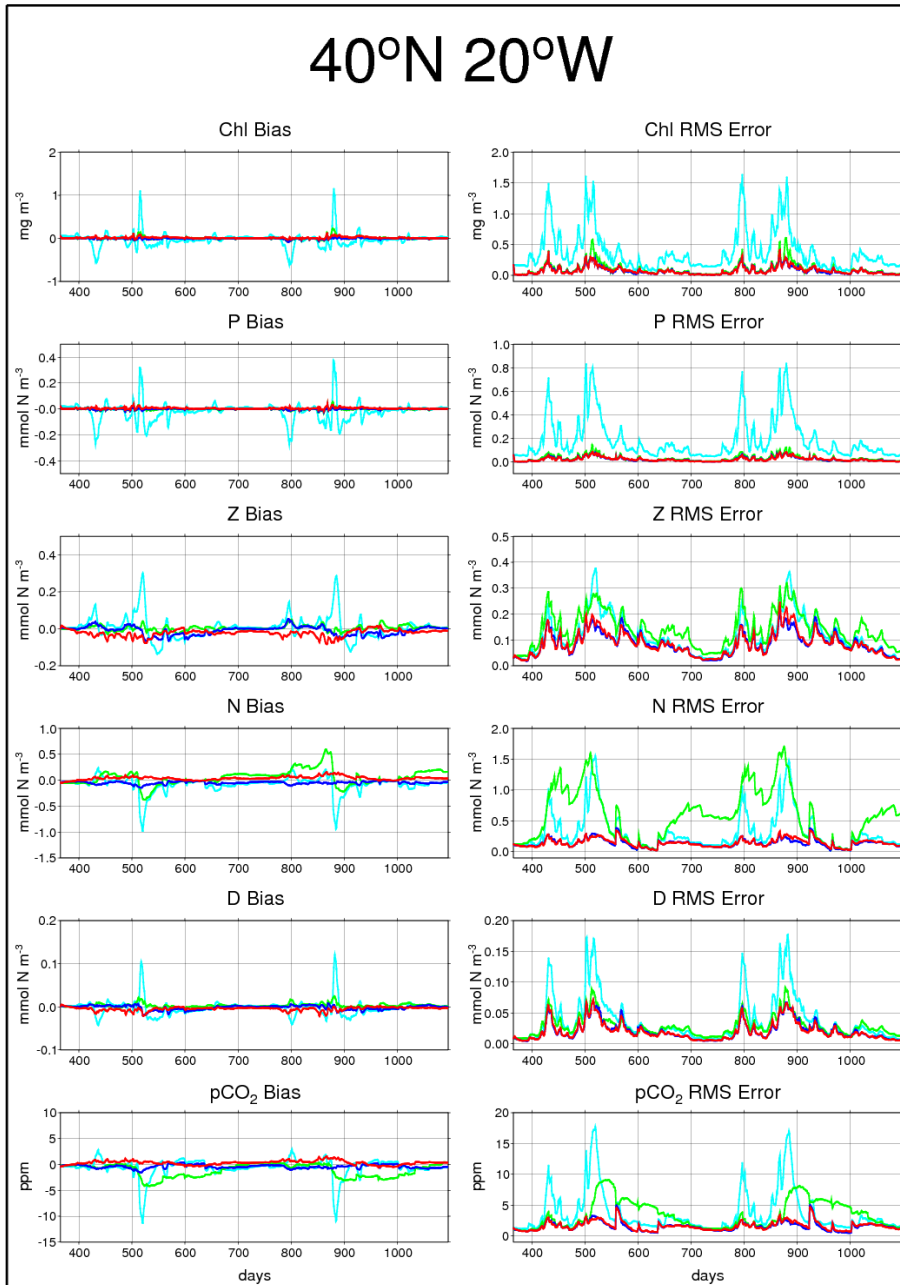
Figure 13: Bias and r.m.s. errors for daily mean surface concentrations at 40°N 20°W for Experiment D: no nitrogen balancing (green), Experiment E: nitrogen balancing with constant pre-adjustment nutrient balancing factor (blue) and Experiment C: nitrogen balancing with variable pre-adjustment nutrient balancing factor (red). All experiments are for chlorophyll and phytoplankton assimilation with no observation error. The free run is shown in cyan.

Figure 14: Bias and r.m.s. errors for daily mean surface concentrations at 50°N 20°W for Experiment D: no nitrogen balancing (green), Experiment E: nitrogen balancing with constant pre-adjustment nutrient balancing factor (blue) and Experiment C: nitrogen balancing with variable pre-adjustment nutrient balancing factor (red). All experiments are for chlorophyll and phytoplankton assimilation with no observation error. The free run is shown in cyan.
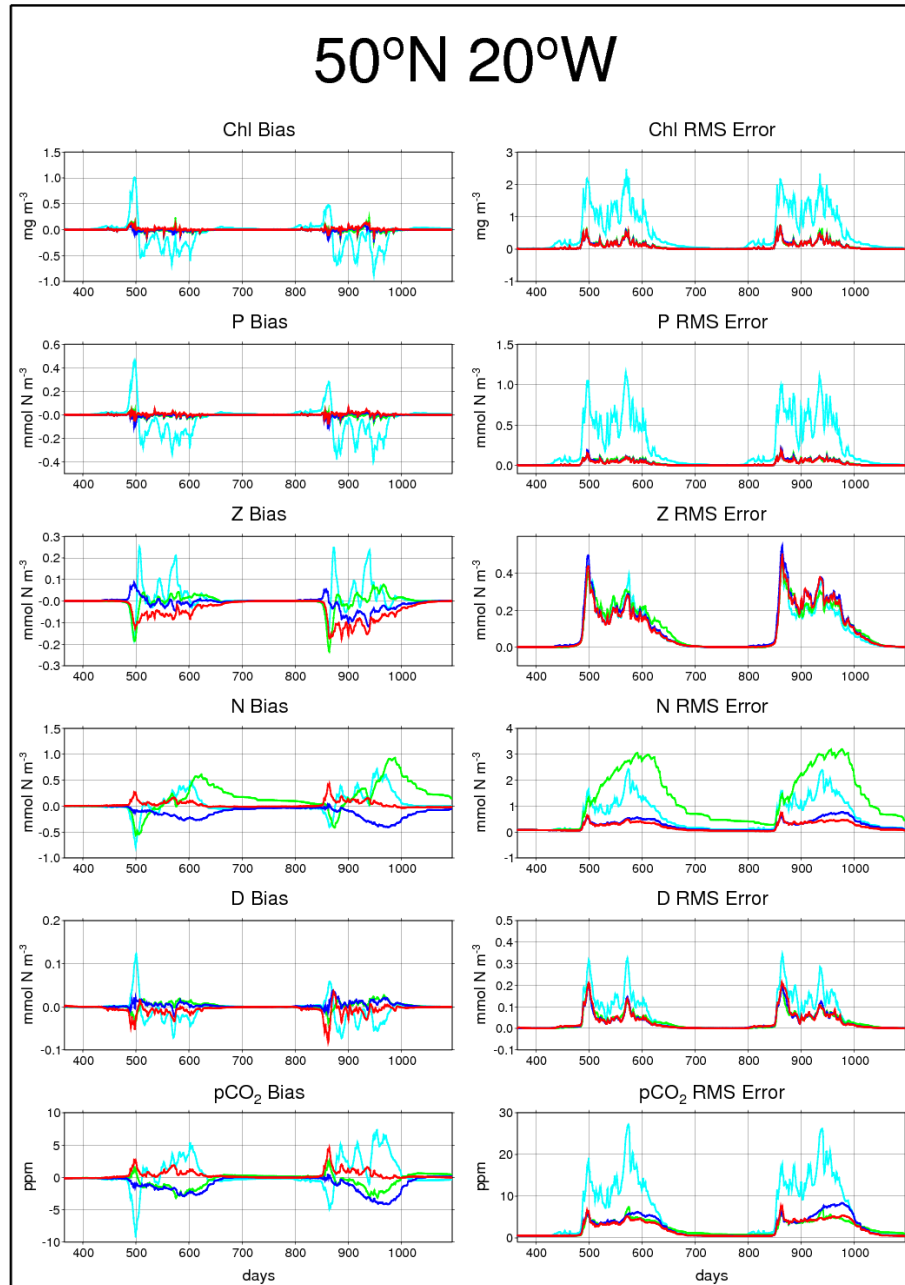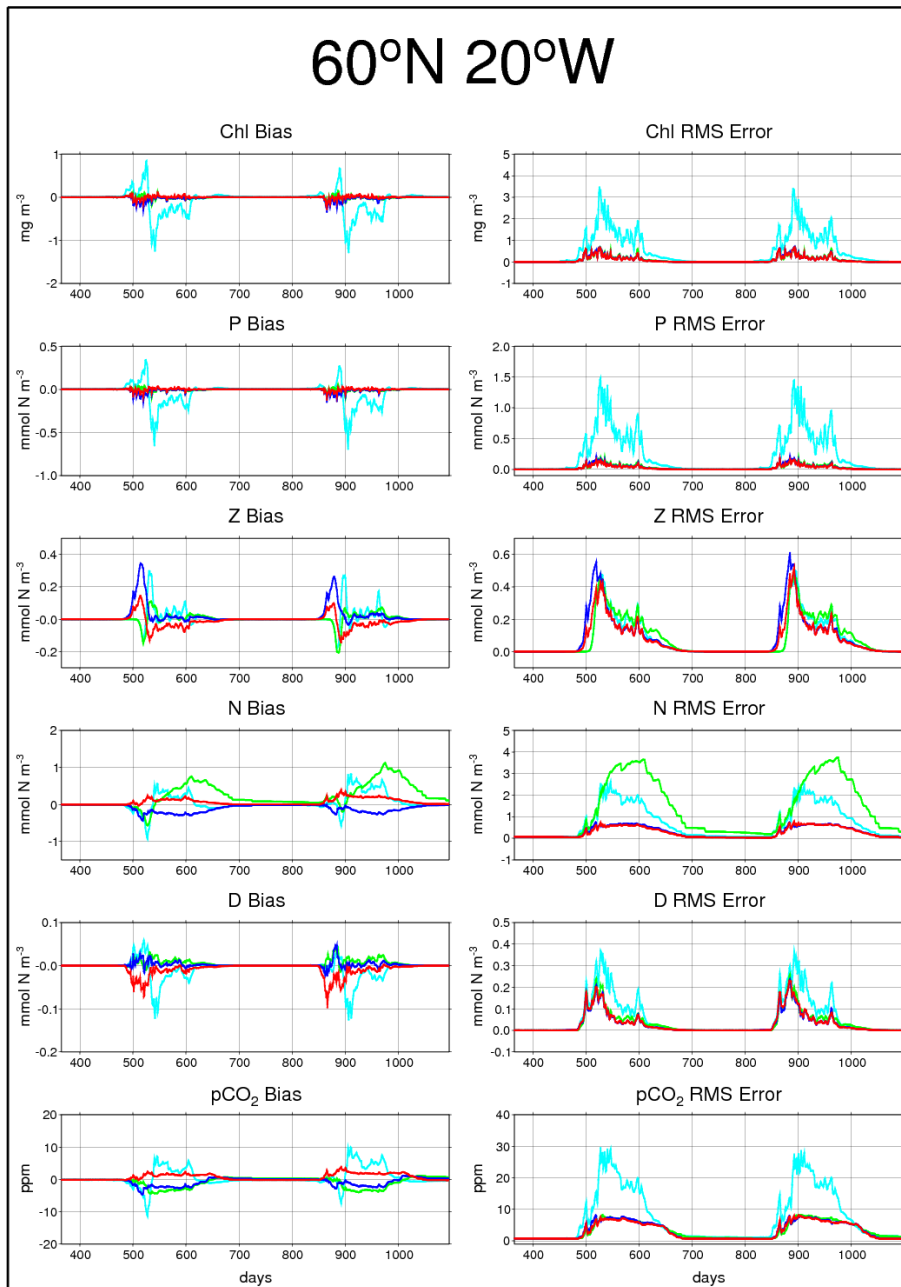
Figure 15: Bias and r.m.s. errors for daily mean surface concentrations at 60°N 20°W for Experiment D: no nitrogen balancing (green), Experiment E: nitrogen balancing with constant pre-adjustment nutrient balancing factor (blue) and Experiment C: nitrogen balancing with variable pre-adjustment nutrient balancing factor (red). All experiments are for chlorophyll and phytoplankton assimilation with no observation error. The free run is shown in cyan.

differences between the first and second years of assimilation, there are much higher errors in nutrient at the beginning of the second calendar year at all latitudes. The system does not seem to recover over the winter from the errors introduced by assimilation in the previous year. Importantly for air-sea CO_2 flux estimation, there are also serious negative biases in pCO_2 at the lower latitudes (Figs. 12 and 13) that are not present in Experiments C or E. These seem to be caused by excess production fueled by excess nutrient in early summer (just before the summer period of low nutrient error). It can be seen from these results that, although much of the improvement in pCO_2 comes as a consequence of improvements in the phytoplankton estimates, especially at high latitudes, the nitrogen balancing scheme is clearly having a beneficial effect and appears to be particularly important for pCO_2 at low latitudes.

Comparison of results for the constant and variable pre-adjustment nutrient balancing factor experiments (Experiments E and C respectively) reveals small improvements in the r.m.s. errors at high latitudes when the more sophisticated balancing model is applied (Experiment C). Specifically, there are reductions in pCO_2 and nutrient errors at 50°N and a reduction in the spring zooplankton error at 60°N. There is a tendency for reversal of pCO_2 biases from negative to positive at all latitudes. In most cases, the magnitude of the bias is similar but at 50°N the positive biases in Experiment C are notably smaller than the negative biases in Experiment E, which explains the reduction in the r.m.s. error at this latitude. The pattern is similar for nutrient. There is a clear tendency for the high latitude zooplankton and detritus biases to be more negative in Experiment C (Figs. 14 and 15). In the case of detritus, this causes the overall magnitude of the bias to be increased at both 50°N and 60°N. These biases are generally smaller than those for zooplankton though. In the case of zooplankton, the magnitude of the bias is increased relative to Experiment E at 50°N but reduced at 60°N where Experiment E is dominated by positive bias.

3.2.3   Effects of Assimilation on the Sub-surface Structure

To examine the effects of assimilation on the sub-surface structure two further assimilation experiments were carried out: one in which no increments were applied at levels below the fully mixed layer (Experiment F) and one in which only primary increments were applied (Experiment G). The results for errors in the sub-surface nutrient minima and phytoplankton maxima are shown in Figures 16 to 19. The minima and maxima are defined by their differences from the surface concentrations.

Under most conditions, phytoplankton deplete nutrient fastest in the surface layer due to the availability of light. It is therefore unusual for sub-surface nutrient minima to occur as a result of local processes. The lowest value for the sub-surface
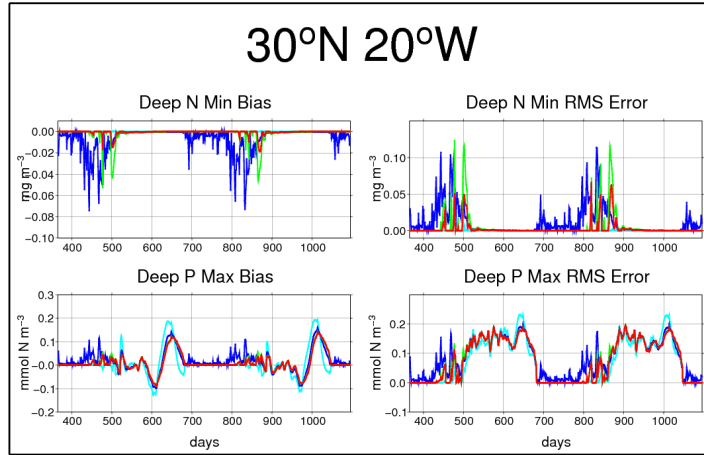
Figure 16: Bias and r.m.s. analysis errors for differences of sub-surface nutrient minimum and phytoplankton maximum from their respective surface concentrations at 30°N 20°W. Results are shown for Experiment F: no increments below fully mixed layer (blue), Experiment G: no secondary increments (green) and Experiment C: full increments at all depths (red). All experiments are for chlorophyll and phytoplankton assimilation with no observation error. The free run is shown in cyan.

Figure 17: Bias and r.m.s. analysis errors for differences of sub-surface nutrient minimum and phytoplankton maximum from their respective surface concentrations at 40°N 20°W. Results are shown for Experiment F: no increments below fully mixed layer (blue), Experiment G: no secondary increments (green) and Experiment C: full increments at all depths (red). All experiments are for chlorophyll and phytoplankton assimilation with no observation error. The free run is shown in cyan.

Figure 18: Bias and r.m.s. analysis errors for differences of sub-surface nutrient minimum and phytoplankton maximum from their respective surface concentrations at 50°N 20°W. Results are shown for Experiment F: no increments below fully mixed layer (blue), Experiment G: no secondary increments (green) and Experiment C: full increments at all depths (red). All experiments are for chlorophyll and phytoplankton assimilation with no observation error. The free run is shown in cyan.
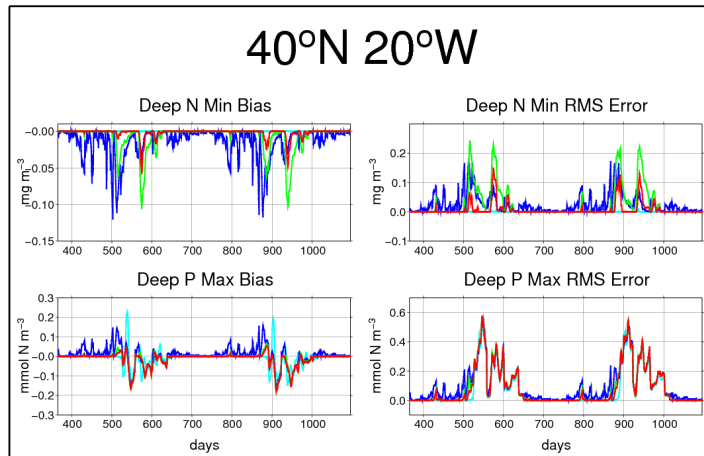

Figure 19: Bias and r.m.s. analysis errors for differences of sub-surface nutrient minimum and phytoplankton maximum from their respective surface concentrations at 60°N 20°W. Results are shown for Experiment F: no increments below fully mixed layer (blue), Experiment G: no secondary increments (green) and Experiment C: full increments at all depths (red). All experiments are for chlorophyll and phytoplankton assimilation with no observation error. The free run is shown in cyan.
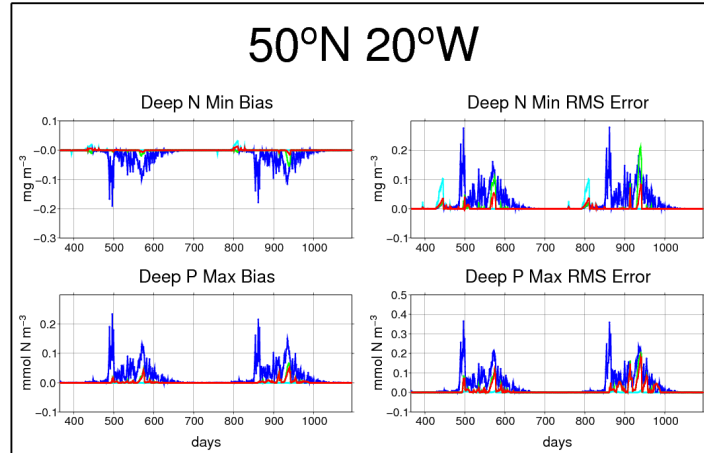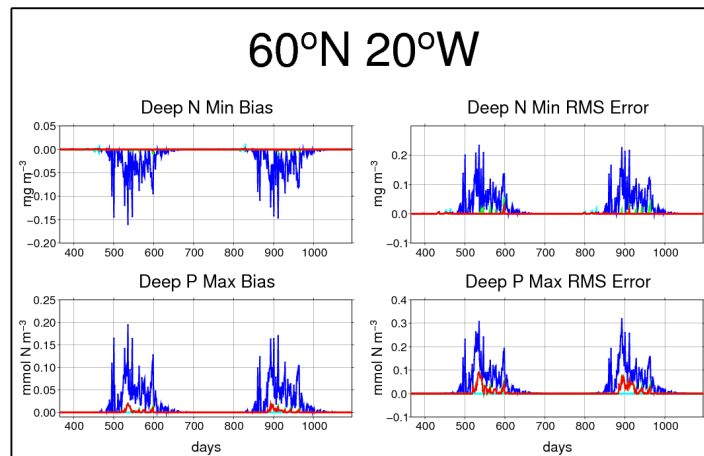
minimum, relative to the surface concentration, in all truth ensembles is -0.56 mmol N m^{-3}. This occurs at 50°N, largely as a result of the sub-surface nutrient relaxation to climatology. At the other latitudes, this concentration difference remains small (above -0.11 mmol N m^{-3}) throughout.

It is clear, from comparison with the free run in Figures 16 and 17, that assimilation is causing the development of unwanted nutrient minima at low latitudes in all of the experiments. Without sub-surface updates (Experiment F) the minima are introduced earlier in the annual cycle. However, under certain conditions the sub-surface updates cause larger errors to occur. Fortunately, these are much reduced in Experiment C where the secondary increments are included.

The benefits of sub-surface increments are clearer at higher latitudes (Figures 18 and 19). At 50°N and 60°N, the sub-surface increments in Experiments C and G reduce the nutrient errors throughout the annual cycle (where significant error occurs), with the notable exception of a short period in late summer at 50°N. Here, the errors in Experiment G are greater than those obtained without sub-surface updates in Experiment F. With the addition of secondary increments in Experiment C all nutrient errors are reduced. In general, the performance in Experiment C at these latitudes is similar to the free run.

The deep phytoplankon maximum errors in Experiment C are similar to the free run at low latitudes (Figures 16 and 17) but there is some tendency for bias reduction so, on the whole, the impact of assimilation seems favorable. The performance is clearly worse than the free run at high latitudes (Figures 18 and 19). Here, assimilation tends to introduce deep phytoplankton maxima where they would not otherwise occur. Despite these problems, comparison with Experiment F shows that the deep phytoplankton maximum errors are reduced at all stations by including the sub-surface increments.

Overall, with regard to sub-surface structure in nutrient and phytoplankton fields, the best assimilation results are obtained in Experiment C (i.e. when sub-surface increments, including secondary increments, are applied). While the detrimental effects of assimilation are reduced by the sub-surface updates, the remaining errors are a cause of concern and should be investigated further.

3.3   Discussion and Conclusions

Improving the state estimates for the biogeochemical tracers on the basis of chlorophyll innovations only is challenging. The phytoplankton nitrogen:chlorophyll ratio can vary by an order of magnitude as the phytoplankton adapt to their environment.

This effect is modelled but model error is inevitable and any error in the nitrogen:chlorophyll ratio used for assimilation has a direct effect on the phytoplankton increment and therefore on all of the other biogeochemical tracer increments. It is even possible for errors in this ratio to cause the phytoplankton increment to have the wrong sign. This occurs if the error in the background chlorophyll, calculated from phytoplankton nitrogen using the model ratio, is of the opposite sign to the background error in phytoplankton. Even perfect knowledge of the phytoplankton error is of limited use for determining corrections to the non-phytoplankton variables. This is because much of the model error affecting these variables does not affect phytoplankton directly and some of the errors that do affect phytoplankton cancel out leaving no signal in the phytoplankton concentration. In the idealized case of a perfect phytoplankton analysis, some analysis error will always remain in the other variables. Because the phytoplankton error then only gives information relating to the last 24 h assimilation time-step, these errors can accumulate over long time scales, with the consequence that increments correcting for recent errors can increase the overall error at certain times. Fortunately though, the seasonal nature of the errors means that large errors tend not to persist for multiple annual cycles.

Despite these inherent problems, the tests show that the material balancing scheme performs well in the sense that, on average, it improves estimates for most of the biogeochemical variables in the upper mixed layer under most conditions in a way that cannot be achieved by a simple scheme without nitrogen balancing. The improvement seems to be attributable in part to the variable pre-adjustment nutrient balancing factor. A more rigorous comparison, in which the constant value of the pre-adjustment nutrient balancing factor used in the alternative experiment was optimized, would be required to confirm this. The major area of concern remaining is the susceptability of the assimilation system to errors in the model nitrogen:chlorophyll ratio. It is shown that these can lead to excessively low pCO_2 estimates in areas where nutrient limitation is important, which would tend to result in insufficient outgassing in the sub-tropics. Degradation of the vertical structure by assimilation, although reduced by the sub-surface increments, is also still an issue.

In this initial evaluation, standard or 'first guess' parameter values are used throughout. Further investigations are required to determine the sensitivity of the scheme's performance to its parameter values and to find an optimal parameter set. The fact that the scheme appears to be fairly successful before optimization is encouraging. It is likely that periods where the scheme appears to perform less well than the free run for certain variables can be reduced by tuning. However, it seems unlikely that the issue of drift in the nitrogen:chlorophyll ratio will be solved in this way.

The nitrogen:chlorophyll ratio is determined by a phytoplankton acclimation model in which the biomass specific rate of change of chlorophyll is dependent on the

biomass specific growth rate and the ambient light. While improvements in the nutrient concentration will have some beneficial effect via the growth rate, a more direct means of correction is desirable. Ocean colour products other than chlorophyll may provide an effective way to address the problem. The most direct solution would be to use ocean colour-based estimates of phytoplankton carbon (Behrenfeld *et al.*, 2005) in conjunction with the chlorophyll product, although the accuracy of these estimates is not well quantified at present. In addition, ocean colour information relating to the light field or more directly to the biomass specific growth rate of the phytoplankton could potentially be used to improve the performance of the acclimation model.

While the tests described here go some way to evaluating the scheme a more thorough evaluation will be required in further twin experiments and experiments with real-world data, some of which can be practically performed in the 3-D model and some of which are more suited to the test-bed environment. A high priority will be to test the effects of poor data coverage on the scheme's performance since daily data will rarely be available in the target system: coverage at high latitudes especially will be very much reduced by cloud cover.

Twin experiments are extremely valuable because they allow all aspects of an assimilation systems performance to be quantified with reference to known truths. Their main disadvantage is their reliance on assumptions about the likely errors in the model and observations. The results presented here for the 100 member ensemble of plausible truths might be considered reasonably robust. However, there are other ways that model error could be introduced and the relative importance of different types of error is unknown. This can be mitigated to some extent by exploring the sensitivity of the results to the assumptions, although errors due to missing process are difficult to simulate. Ideally the experiments would be repeated with a variety of different assumptions about the dominant errors.

Real-world experiments involve no such assumptions but their interpretation is made more complicated by other errors in the system. These include both errors in the physical model and errors due to the absence of biogeochemical balancing increments in the assimilation of physical data. It is not the role of a biogeochemical assimilation scheme to correct for these errors: they can and should be addressed by other means. The material balancing scheme does not allow for them and they are likely to degrade its performance. Separate balancing schemes will be required to correct for physical data assimilation. Material conservation rules may or may not be involved: where errors in temperature and/or salinity are associated with improper representation of a front between water masses, then there are likely to be errors in the total concentrations of carbon and nitrogen at individual grid points and a local conservation constraint would be undesirable. If the impacts of errors in the physical simulation are significant, then assessments of the material balanc-

ing scheme with respect to it's value in the long-term, when such errors have been reduced, might actually be less meaningful in real-world experiments than in twin experiments. In this way, twin experiments do provide a valuable analytical tool for examining the strengths and weaknesses of individual parts of an assimilation system in isolation, avoiding the need to address all problems simultaneously.

The issue of tuning is not straightforward. Some progress towards finding optimal parameters can and should be made in twin experiments to avoid parameters being forced to compensate for errors in other parts of the system. It is also appropriate for pragmatic reasons, since the scheme almost certainly has more tunable parameters than can be independently constrained by the available real-world data. Parameter optimization is computationally intensive and so it is desirable to pre-tune the scheme in the test-bed, leaving some flexibility in a small number of parameters for final tuning in the 3-D model. The facility to optimize parameters with respect to real-world data in the test-bed would be desirable but would require the test bed's emulation of the 3-D model to be improved. While the difficulty of including horizontal fluxes and the effects of physical data assimilation in the test-bed may make sufficiently accurate emulation impractical, the possibility seems worth exploring.

Finally, sequential assimilation of ocean colour data should not be considered as a substitute for improvements in the biogeochemical model itself. Ocean colour data are equally valuable in this context too: steps towards improving the model should include optimization of the model parameters to fit available ocean colour and *in situ* data. Such parameter optimization is computationally intensive and is another activity that would be facilitated by improved emulation of the 3-D model in the test-bed.

Acknowledgements

A   Probability Model of Growth and Loss Errors

The 2-D prior probability density function used to model uncertainty in the phytoplankton error due to growth and loss rate errors is defined in Section A.1. The

function is parameterized in terms of a phytoplankton specific growth rate estimate and two alternative estimates of phytoplankton specific loss rate. The rate estimators used are described in Section A.2. Section A.3 then describes how the prior probability distribution is used in conjunction with the estimated total phytoplankton error from the phytoplankton analysis to calculate the expected value of the nutrient balancing function.

A.1 Probability Density Function

The approximate relationships defined by Equations 14 and 15 change the problem of modelling uncertainty in the phytoplankton error components x and y, attributed to growth and loss errors respectively, to one of modelling uncertainty in the phytoplankton specific growth and loss rate errors themselves. The magnitude of the rate errors is expected to increase as the rates increase. This is consistent with the model dynamics in which both rates are products of uncertain factors. We therefore assume a constant coefficient of variation for the errors and so use log-normal distributions for X and Y. The probability density functions $p(x)$ and $p(y)$ describing these distributions vary in time as functions of the model phytoplankton concentration P_0 and the model phytoplankton specific growth rate G_0 or the model phytoplankton specific loss rate L_0 as appropriate. (By convention, the function p is used to refer generically to different p.d.f.s with the functional form being determined contextually by the function's argument. Subscripts are used here only where there would otherwise be ambiguity).

Positive errors in specific growth rate cannot be greater than the model rate itself so the value of x has a maximum, dependent on the time-series of the model specific growth rate and model phytoplankton concentration during the assimilation time-step. An estimate for this is given by

$$x_{\text{MAX}} = \Delta t P_0 G_0, \tag{30}$$

obtained by setting the true growth rate to zero in Equation 14. Similarly, positive errors in specific loss rate cannot be greater than the model specific loss rate. y therefore has a minimum value, occurring when the true loss rate is zero, approximated by

$$y_{\text{MIN}} = -\Delta t P_0 L_0, \tag{31}$$

derived from Equation 15.

The p.d.f. for the phytoplankton error due to errors in growth rate is

$$p(x) = \frac{1}{x_{\mathrm{MAX}} - x} N\left(\frac{\log(x_{\mathrm{MAX}} - x) - \mu_{\mathrm{G}}}{\sigma_{\mathrm{G}}}\right), \tag{32}$$

where the function N is the normal distribution with zero mean and unit standard deviation and σ_{G} is the standard deviation for the log-transformed specific growth rate estimator, supplied as an external parameter. μ_{G} is the mean of the log-transformed specific growth rate estimator, scaled by $\frac{1}{\Delta t P_0}$ to convert to log phytoplankton units. μ_{G} is chosen such that the expected value of the growth rate estimator is equal to a best estimate G_1 of the true specific growth rate. i.e.

$$\frac{x_{\mathrm{MAX}} - E(X)}{\Delta t P_0} = G_1, \tag{33}$$

where the expression on the left hand side for the expected value of the growth rate estimator follows from Equations 14 and 30. The appropriate mean value is

$$\mu_{\mathrm{G}} = \log(\Delta t P_0 G_1) - \frac{\sigma_{\mathrm{G}}^2}{2}. \tag{34}$$

The estimate G_1 is derived from the model rate G_0 as described in the next section.

The p.d.f. for the phytoplankton error due to errors in loss rate is modelled as the weighted average of two distributions derived from different loss rate estimates (although, in the test-bed experiments, the weighting is set so that only one is used). The primary estimate L_1 is based on the model specific loss rate L_0. This is likely to be less reliable than the model specific growth rate G_0. However, because growth and loss rates are equal when the system is in steady state, the more reliable value G_0 provides another estimate of the specific loss rate. This is the basis for the alternative specific loss rate estimate M.

The p.d.f.s for the two distributions are

$$p_{\mathrm{L}}(y) = \frac{1}{y - y_{\mathrm{MIN}}} N\left(\frac{\log(y - y_{\mathrm{MIN}}) - \mu_{\mathrm{L}}}{\sigma_{\mathrm{L}}}\right), \tag{35}$$

$$\mu_{\mathrm{L}} = \log(\Delta t P_0 L_1) - \frac{\sigma_{\mathrm{L}}^2}{2} \tag{36}$$

and

$$p_{\text{M}}(y) \quad = \quad \frac{1}{y - y_{\text{MIN}}} N\left(\frac{\log(y - y_{\text{MIN}}) - \mu_{\text{M}}}{\sigma_{\text{M}}}\right), \tag{37}$$

$$\mu_{\text{M}} \quad = \quad \log(\Delta t P_0 M) - \frac{\sigma_{\text{M}}^2}{2}, \tag{38}$$

where σ_{L} and σ_{M} are external parameters giving the standard deviations for the log-transformed primary loss rate estimator and the log-transformed alternative loss rate estimator respectively. The values of μ_{L} and μ_{M} above are chosen so that the expected values of the two loss rate estimators are L_1 and M respectively. The p.d.f. for Y is then

$$p(y) = r p_{\text{L}}(y) + (1 - r) p_{\text{M}}(y), \tag{39}$$

where r is a model loss rate reliability parameter. This is a measure of confidence in the primary estimator, relative to the alternative estimator.

Given that X and Y are independent, the joint p.d.f. $p(x, y)$ is simply the product of $p(x)$ and $p(y)$, so

$$p(x, y) = p(x)\{r p_{\text{L}}(y) + (1 - r) p_{\text{M}}(y)\}. \tag{40}$$

A.2 Growth and Loss Rate Estimators

The estimate G_1 of the true specific growth rate is based on the model specific growth rate G_0 but diverges as G_0 gets small. The divergence is introduced to allow for an expected negative bias in the model estimate as the latter tends to zero. Such a bias must be present if there is a non-vanishing error variance since growth rate cannot be negative. The estimate is given by

$$G_1 \quad = \quad \beta_{\text{G}} + \frac{G_0^2}{4\beta_{\text{G}}}, \quad G_0 < 2\beta_{\text{G}} \tag{41}$$

$$G_1 \quad = \quad G_0, \quad G_0 >= 2\beta_{\text{G}} \tag{42}$$

where β_G is the low growth bias correction at $G_0 = 0$, specified as an external parameter. Similarly the primary specific loss rate estimate is

$$L_1 = \beta_L + \frac{L_0^2}{4\beta_L}, \quad L_0 < 2\beta_L \tag{43}$$

$$L_1 = L_0, \quad L_0 >= 2\beta_L \tag{44}$$

where β_L is the low loss bias correction at $L_0 = 0$. (This value is never actually reached in the HadOCC model since the minimum value of the specific loss rate is fixed by the model's respiration rate constant).

The alternative specific loss rate estimate, based on the specific growth rate, is also bias-corrected but differs from the bias-corrected growth rate estimate. The appropriate bias correction is likely to be larger when the model specific growth rate is used as a loss rate estimator than when it is used as a growth rate estimator because of the increased uncertainty. The alternative specific loss rate estimate is

$$M = \beta_M + \frac{G_0^2}{4\beta_M}, \quad G_0 < 2\beta_M \tag{45}$$

$$M = G_0, \quad G_0 >= 2\beta_M \tag{46}$$

where β_M is the low loss bias correction at $G_0 = 0$. M is therefore identical to the bias-corrected growth rate estimate G_1 at high growth but diverges at low growth rates.

A.3 Calculating the Expected Value of the Nutrient Balancing Function

We assume $-\Delta P$ to be an accurate estimate of the total error $x + y$. The loss rate error component is therefore given by

$$y = -(\Delta P + x) \tag{47}$$

and, for particular values of P_0, G_0, L_0 and ΔP, the nutrient balancing function g becomes a function of x only. Its expected value is

$$E(g(x) \mid X + Y = -\Delta P) = \int_{-\infty}^{\infty} p(x \mid X + Y = -\Delta P)g(x)dx, \tag{48}$$

where $p(x \mid X + Y = -\Delta P)$ is the probability density of x conditional on the total error $X + Y$ being equal and opposite to the phytoplankton increment ΔP. The right hand side of Equation 48 can be interpreted geometrically in (x,y) space as a weighted average of the function $g(x, y)$ along the line $y = -(x + \Delta P)$.

From Bayes' theorem

$$p(x \mid X + Y = -\Delta P) = \frac{p(X + Y = -\Delta P \mid x)p(x)}{p(X + Y = -\Delta P)}. \tag{49}$$

The p.d.f. $p(X + Y = -\Delta P \mid x)$ is simply $p(y)$ where $y = -(x + \Delta P)$. The denominator is a constant. For a particular value $X = x$, the function $p(X + Y = -\Delta P)$ becomes $p(y)$ so the denominator is equal to the integral of the joint probability density $p(x, y(x))$ (Equation 40) over all x. i.e.

$$p(X + Y = -\Delta P) = \int_{-\infty}^{\infty} p(x, y(x))dx. \tag{50}$$

Substituting in Equation 48, gives the equation

$$E(g(x)) = \frac{1}{\int_{-\infty}^{\infty} p(x, y(x))dx} \int_{-\infty}^{\infty} p(x)p(y(x))g(x)dx \tag{51}$$

for calculating the expected value of the balancing function. Note that since X and Y are independent $p(x)p(y) = p(x, y)$.

In practice the integrals must be evaluated numerically for each surface grid point at each assimilation time-step. This is potentially expensive but the number of steps in the integration can be specified externally to trade accuracy for speed or *vice versa*. 20 steps were used in the test-bed experiments presented here. Finite integration bounds are determined for the p.d.f. $p(x, y(x))$ at each grid point which ensure that integration at least covers all probability densities greater than a certain fraction of the peak. This fraction is set by an externally specified threshold (=0.01 in the test-bed experiments).

References

Anderson, T. R. 1993. A spectrally averaged model of light penetration and photosynthesis. Limnology and Oceanography 38, 1403-1419.

Behrenfeld, M. J., E. Boss, D. A. Siegel and D. M. Shea, 2005. Carbon-based ocean productivity and phytoplankton physiology from space. Global Biogeochemical Cycles 19, GB1006, doi:10.1029/2004GB002299.

Bell, M. J., R. M. Forbes and A. Hines, 2000. Assessment of the FOAM global data assimilation system for real-time operational ocean forecasting. Journal of Marine Systems 25, 1-22.

Campbell, J. W. 1995. The lognormal distribution as a model for bio-optical variability in the sea. Journal of Geophysical Research 100, 13237-13254.

Geider, R. J., H. L. MacIntyre and T. M. Kana. 1997. Dynamic model of phytoplankton growth and acclimation: responses of the balanced growth rate and the chlorophyll a:carbon ratio to light, nutrient-limitation and temperature. Marine Ecology Progress Series 148, 187-200.

Hooker, S. B. and C.R. McClain. 2000. The calibration and validation of SeaWiFS data. Progress In Oceanography 45, 427-465.

Palmer, J. R. and I. J. Totterdell, 2001. Production and export in a global ocean ecosystem model. Deep-Sea Research I 48, 1169-1198.