Alma Mater Studiorum

UNIVERSITÀ DI BOLOGNA

Facoltà di Scienze Matematiche Fisiche e Naturali

Dottorato di Ricerca in Geofisica - Ciclo XIX

SSD GEO/12 OCEANOGRAFIA E FISICA DELL'ATMOSFERA

Numerical simulations of the coastal marine ecosystem dynamics: integration techniques and data assimilation in a complex physical-biogeochemical model

Tesi di Dottorato di

MOMME BUTENSCHÖN

Esame finale anno 2007

Relatore: Dott. Marco Zavatarelli Coordinatore: Prof. Michele Dragoni

Alma Mater Studiorum
UNIVERSITÀ DI BOLOGNA
Facoltà di Scienze Matematiche Fisiche e Naturali
Dottorato di Ricerca in Geofisica - Ciclo XIX
SSD GEO/12 OCEANOGRAFIA E FISICA DELL'ATMOSFERA

Numerical simulations of the coastal marine ecosystem dynamics: integration techniques and data assimilation in a complex physical-biogeochemical model

Tesi di Dottorato di
MOMME BUTENSCHÖN

Esame finale anno 2007

Relatore: Dott. Marco Zavatarelli Coordinatore: Prof. Michele Dragoni

Acknowledgements

I would like to take the opportunity to thank Prof. Nadia Pinardi, who has introduced me to the subject of oceanography as a whole and in particular to the arguments treated in this thesis and helped in numerous discussions throughout this work and my tutor, Marco Zavatarelli, who has accompanied me continuously in its realisation. Without them this work would not have been possible.

George Triantafyllou gave valuable contributions in the beginning of the work on data assimilation and Srdjan Dobricic gave numerous advises in the course of its realisation, while the contribution of Marcello Vichi was essential in the part on coupling methods. The help of all of them is gratefully acknowledged.

Finally, I would like to thank my parents who are at the very base of this work as of all of the things I achieved in life and my love who has supported me with her energies and positivity throughout these years.

Contents

Introduction **9**

1 Implementation of a coupling method for numerical time integration in dynamic marine ecosystem models **13**

Introduction 14

1.1 Model coupling 19

1.1.1 Coupling schemes 19

1.1.2 Integration schemes 24

1.2 The model and its sub-models 30

1.2.1 The hydrodynamic model 31

1.2.2 The biogeochemical model 34

1.2.3 The transport model 40

1.3 Implementation of the coupled model 42

1.3.1 Implementation sites 42

1.3.2 Vertical discretisation 43

1.3.3 Boundary conditions 44

1.3.4 Forcing functions 45

1.3.5 Physical parametrisation 47

1.3.6 Biogeochemical parametrisation 49

1.3.7 Initial conditions 49

1.3.8 Coupling 50

1.4 Results and validation 55

1.4.1 Reference solution 56

1.4.2 BFMStep method: sensitivity to Δt_{glob} and to integration schemes 65

1.4.3 Coupling method sensitivity: BFMStep vs. POMStep. 73
1.4.4 The POMStep-2 experiments 80
1.5 Conclusions 85
Appendix A 90
Appendix B 96

2 A data assimilation scheme for dynamic marine ecosystem models 103
Introduction - The concept of data assimilation 104
2.1 Sequential data assimilation - the Kalman Filter 109
2.1.1 Optimal Interpolation 109
2.1.2 Sequential data assimilation 111
2.1.3 The Kalman Filter 112
2.1.3.1 The Extended Kalman Filter 114
2.2 The Singular Evolutive Extended Kalman Filter - *SEEK* . . 116
2.2.1 Versions of the *SEEK* Filter 118
2.2.1.1 Version with forgetting factor (SEEK-FF) . 118
2.2.1.2 Semi-static version (SEEK-0.5) 119
2.3 Implementation 120
2.3.1 Model 120
2.3.2 Implemented filters 120
2.3.3 TWIN experiment preparation 121
2.3.4 Initialisation of the filter 121
2.3.5 Experiment overview 125
2.4 Implementation test 129
2.5 Filter comparison 136
2.6 The *SEEK-Q* Variant 162
2.7 Comparing SEEK-Q with SEEK-FF 164
2.8 Summary and conclusions 188
Appendix A 191
Appendix B 198
Appendix C 200

Bibliography **219**

Introduction

Research aimed at modelling the marine ecosystem dynamics started in the mid fifties (Riley et al., 1949). Since then the attempt to simulate the temporal and spatial variability of the marine ecosystem has progressed towards the development of progressively complex ecosystem models (Hofmann and Lascara, 1998) applied to regional (Polimene et al., 2006, 2007) and global scales (Vichi et al., 2007a) and aimed to gain insight on many different scientific issues: from the definition of the pathways of matter and energy in the marine trophic web to the quantification of the ocean biological processes in controlling levels of atmospheric CO2.

Essentially the modelling of the marine ecosystem necessarily implies the accurate modelling of the ocean physical dynamics, that plays a significant role in driving the temporal variability of the ecosystem (Sverdrup, 1953). Therefore, the first step toward sound ecological modelling is the development of a coupled modelling system, simultaneously solving for ecological/biogeochemical processes (photosynthesis, bacterial cycling of carbon and nutrients, predation etc.) and for the advective diffusive physical transport processes affecting chemical components (nutrients, oxygen etc.) and individuals/biomasses (from bacteria and phytoplankton to fishes). This involves consideration by the model of a very wide range of temporal and spatial scales as well as their translation into a numerical framework providing an accurate coupling method of the (physical and ecological) simulated rates of change and a correct numerical integration technique.

Advances in modelling of the ocean physical dynamics are now giving an accurate description of dynamical structures of the meso to large

scale general circulation. This, combined with the enhancement of the ocean observation network trough remote (satellites) and in situ (traditional sampling, automatic buoys, autonomous moving or drifting platforms) systems, has determined the decisive step towards the nowcasting/forecasting of the physical state of the ocean and led to the successful implementation of operational forecasting systems (Pinardi et al., 2005; Johannessen et al., 2006), combining simulations and observations through data assimilation techniques and designed on the general structure of the routinely operating weather forecasting systems.

The definition and the establishment of a similar system devoted to the forecasting of the biogeochemical/ecological structure of the ocean is now a major subject of today research activities (Malone and Cole, 2000; Haitvogel et al., 2000; Walstad and McGillicuddy, 2000). The assessment and the forecasting of the marine ecosystem is in fact of fundamental interest for the definition of a wide range of purely scientific to applicative problems such as impact of pollutants, harmful algal blooms or long term prognoses of the green house effect.

A strong focus in this attempt is on the coastal ocean, being at the same time the oceanic region with the most active biogeochemical processes (Wollast, 1999) and the region subject to the most intensive exploitation processes (fisheries) and to the strong anthropogenic impact. These domains represent challenging environments, as they are characterised by complex biogeochemical (land based nutrient load of natural and/or anthropogenic origin) and physical (wind stress, heat flux, land based buoyancy inputs) forcing functions. The resulting dynamics is characterised by a strong time/space variability in the circulation patterns and in the mixing/stratification characteristic of the water column, that govern an equally (or more) variable biological dynamics.

In this thesis the physical/biological coupling and the application of data assimilation techniques to biogeochemical state variables in a coupled ecological model are explored in an idealised framework with the support of a one dimensional ecological model implemented in the north-

ern Adriatic, a shallow coastal basin, subject to extremely variable physical and biogeochemical forcing functions that shape a physical and biological dynamics very challenging for the modeller. The general objective, as described in detail in the pertinent sections of the work, is twofold. First it has been attempted to gain insight on the "best" method of coupling together biogeochemical and physical advection and diffusion rates of change, separately computed by the biogeochemical and physical components of the modelling system. This implied, not only the analysis of the role of the classical numerical integration techniques in affecting the model solution, but also (and more importantly) a careful consideration of the interacting temporal scales governing the physical and biological processes. Secondly, a data assimilation technique, originally developed for application to physical state variables has been applied to the biogeochemical. The implementation occurred in the idealised framework of the so-called "twin experiments", that allow to careful evaluate the role of the data assimilation technique in keeping the model solution close to a reference solution assumed as "real" state of the system.

The thesis is organised in two chapters, each of them structured as two independent papers.

Chapter one deals with the model coupling issues. The description and the discussion of the numerical experiments performed is preceded by a general introduction defining the nature of the problem and giving an overview of the numerical aspect of the problem.

Chapter two deals with the data assimilation twin experiments. The methods used and the TWIN experiment philosophy is given in the introductory part, followed by the description discussion of the experiments performed. The chapter closes with the description of a variant of the data assimilation technique that has been originally developed in this thesis and whose results appear quite promising.

Conclusions are offered at the end of the two chapters. Appendixes to the chapters provide a wider and complementary look at the results.

1 Implementation of a coupling method for numerical time integration in dynamic marine ecosystem models

Introduction

Marine ecosystem dynamics is strongly constrained and, to some extent, governed by the ocean physics. Biogeochemical processes (primary and secondary production, nutrient cycling, etc.) are heavily dependent on physical dynamics, that determines the availability of resources (e. g. nutrients and light for primary producers) and the spatially and temporally variable distribution of individuals and biomass. Moreover, the variability of the ocean temperature field directly affects the metabolic rates of the organisms, as well as the chemical processes involved in biogeochemical cycling. Thus, adequate understanding and modelling of the marine ecosystem necessarily implies the full consideration of the coupled physical and biogeochemical dynamics.

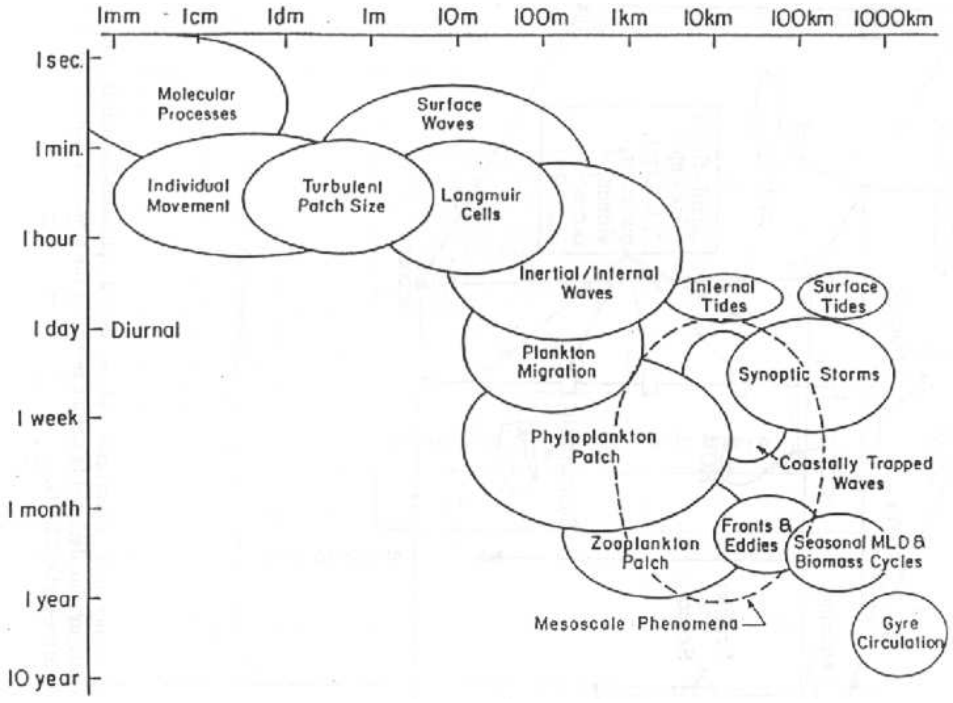


Figure 1.1: Space and time scales involved in marine ecosystem dynamics (from GLOBEC, 1993). Both physical and biogeochemical scales and their overlap are shown.

The involved interacting processes develop on a wide spectrum of temporal and spatial scales (see fig. 1.1), ranging from molecular diffusion and small scale turbulence processes to large scale circulation patterns, and, on the biological side, from individual behaviour to seasonal/decadal variability of biomass and population size.

However, due to the forcing of the system, energy is not uniformly transferred over the spectrum, but rather accumulated in patches (Nihoul and Djenidi, 1998). The energy transfer of the system is then determined not only by the interaction of the various state variables, but also by the exchanges between processes of different scales, so that the resulting annual dynamics is characterised by similar accumulative patterns determined by the alternating dominant scales, that express, broadly speaking, in a rather steady (small and slow changes) behaviour, over longer periods of the year, interrupted by relatively short periods of intense biological activity.

A classical example for temperate waters is given by the so called surface phytoplankton "spring bloom", whose dynamics, unveiled by Sverdrup (1953), is governed by the evolving physical environment (irradiance, vertical stratification, diffusive nutrient flux into the euphotic zone) that determines the conditions for a phytoplankton biomass accumulation near the surface when the seasonal thermocline starts to develop (Mann and Lazier, 1991; Miller, 2006).

These characteristics clearly illustrate the nonlinearity of the ecosystem and pose severe problems to the modeller. Since it is impossible to attempt to simulate the entire range of interacting scales, an adequate spectral window, as described by Nihoul and Djenidi (1998), has to be selected, where unresolved processes usually enter the model by a parametric background flux at the small scale interface and by the initial and boundary conditions at the large scale interface. Moreover, the numerical treatment of this complex dynamics adds further problems related to stability and stiffness that are complicated by the choice of an appropriate space and time resolution.

A common approach to face these problems is to base a numerical model of marine ecosystem dynamics on two separate sub-models each of them applicable separately:

- a Eulerian ocean circulation model that treats the physical processes and provides information on the surrounding environment to the biogeochemical model.

- an ecological/biogeochemical model that defines the ecological state variable[1] and describes dynamically the biota and its relations with the environment in terms of biogeochemical fluxes.

To these two fundamental building blocks of the marine environment modelling system, a third element, handling the advective-diffusive transport of the ecosystem state variables has to be added, as schematically depicted in fig. 1.6 on page 31.

The separate treatment of the physical and biogeochemical processes decouples the spatial dimension of the physical sub-model from the state variable dimension of the biogeochemical sub-model and hereby avoids bulky matrices, allowing for separate numerical treatment and facilitates the calibration of the two parts.

The artificial partitioning of the ecosystem into a physical and biogeochemical component implies that the modelling system separately has to provide for each ecosystem state variable the time dependent rate of change due to physical processes and the rate of change due to biogeochemical processes. The two rates are then merged to provide the total rate of change.

The coupling procedure occurs at the level of time integration where

[1] While for circulation models a broad agreement exists on the determination of the system state (by temperature, salinity, velocity, pressure and sea surface elevation), in biogeochemical models the choice of the state variables, that is to say the structure of the trophic web to be represented is a vivid field of discussion and research directly connected to the choice of the spectral window. See Hofmann and Lascara (1998) for a review on the models currently in use in the science community

the following equation has to be solved:

$$\mathbf{c}_{n+1} = \mathbf{c}_n + \int_{t_n}^{t_{n+1}} \left(-\mathbf{u} * \nabla \mathbf{c} + \nabla * (\nu \nabla \mathbf{c}) + \left. \frac{\partial \mathbf{c}}{\partial t} \right|_{bio} \right) \mathrm{d}t \,. \tag{1.1}$$

where $\mathbf{c}_n$ is any pelagic state variable $\mathbf{c}$ of the biogeochemical model at time t_n.

These equations show a striking similarity to the equations of atmospheric geochemistry dynamics. In fact, the first pioneering works on the numerical solution of these type of equations of advective-diffusive transport for active tracers where attempted in the context of atmospheric chemistry models, that faced the problem of simulating stiff chemical reactions, such as the photochemical ozone cycle in the troposphere. (see e. g. Jacobson, 2005; Graedel and Crutzen, 1995).

The schemes used for the atmospheric transport proved to be inadequate for the stiff chemical reactions, that in eq. (1.1) correspond to the biogeochemical rates $\left. \frac{\partial \mathbf{c}}{\partial t} \right|_{bio}$, and thus, the need of an adequate splitting mechanism emerged. One of the first systematic approaches to the splitting of the advective-diffusive equations for active tracers can be attributed to McRae et al. (1982) and ever since then, research on this subject has grown in particular in the context of atmospheric geochemistry dynamics (see Verwer et al., 2001 and Carmichael et al., 1996 for an overview).

In this chapter, the influence of the coupling scheme on the solution of the physical-biogeochemical equations in the regime of coastal ecosystem dynamics shall be investigated through a set of numerical simulations run with different coupling, integration schemes and time steps carried out with a complex 1D physical-biogeochemical model. It is divided in four sections:

The first section gives a general description of coupling schemes traditionally applied in dynamic atmospheric and oceanographic models, followed by the investigation of the behaviour of 4 different integration schemes in the context of a simple 0D-ecological model. In the second sec-

tion an overview of the complex 1D-model applied in the following experiments is provided and a general description of its three main building blocks of the modelling system is given. The third section describes the application of the coupling schemes to the model implemented in three different sites of the northern Adriatic Sea along with the used forcing functions and boundary conditions.

A validation of the different schemes is presented in section four, showing the results of the simulations for various configurations. A comparison to seasonal climatological data available for the 3 implementation sites is also given in appendix A.

The aim is to evaluate the influence of the coupling and integration scheme on the solution, therefore providing insight about the choice of the most appropriate integration and coupling scheme for a marine ecosystem model.

1.1 Model coupling

To propagate the model state in time, the processes represented by the various sub-models have to be taken into account and thus, a coupling strategy is needed to merge all involved processes into a global integration scheme.

As stated before, a system involving two sub-processes $\mathbf{S}_I$, $\mathbf{S}_{II}$ can be written in it's simplest form as:

$$\begin{aligned} \frac{\mathrm{d}\mathbf{c}}{\mathrm{d}t} &= \mathbf{S}_I(\mathbf{c}) + \mathbf{S}_{II}(\mathbf{c}) \\ \mathbf{c}_{n+1} &= \mathbf{c}_n + \int_{t_n}^{t_{n+1}} (\mathbf{S}_I(\mathbf{c}) + \mathbf{S}_{II}(\mathbf{c}))\,\mathrm{d}t \,. \end{aligned} \tag{1.2}$$

Eq. (1.2) applied to a marine ecosystem model indicates that the total rate of change of the non conservative variable results from the sum of two separate partial rates of change, one accounting for the conservative part of the property and one for the non-conservative behaviour.

In the following a brief account of the coupling schemes used to solve for time dependent non-conservative state variables is given along with a critical review of the accuracy problems they pose.

1.1.1 Coupling schemes

Different coupling strategies have been attempted and two main branches can be identified: the Operator Splitting method and the Source Splitting method.

Operator Splitting

In the method of *Operator Splitting* , widely used in atmospheric chemistry (McRae et al., 1982; Blom and Verwer, 2000; Verwer et al., 2001), as well as in general ocean circulation models (e. g. Blumberg and Mellor (1987)),

the original system (1.2) is divided into two separate subsystems that are solved sequentially:

$$\begin{aligned} \gamma &= \mathbf{c}_n + \int_{t_n}^{t_{n+1}} \mathbf{S}_I(\mathbf{c}) \mathrm{d}t \\ \mathbf{c}_{n+1} &= \gamma + \int_{t_n}^{t_{n+1}} \mathbf{S}_{II}(\gamma) \mathrm{d}t \; . \end{aligned} \tag{1.3}$$

Independently of the sub-integration method used, a conceptual error is introduced to the system, as the integration is treated in two separate processes, that are internally not influenced by each other. This so called splitting error can be quantified in a certain sense by a Taylor expansion of eq. (1.2) (shown for an ordinary differential equation (ODE) for simplicity), that yields

$$\begin{aligned} c_{n+1} = c_n &+ S_I\left(c_n\right) \Delta t + S_{II}\left(c_n\right) \Delta t + S_I\left(c_n\right) S_I^{'}\left(c_n\right) \frac{\Delta t^2}{2} \\ &+ S_I\left(c_n\right) S_{II}^{'}\left(c_n\right) \frac{\Delta t^2}{2} + S_{II}\left(c_n\right) S_I^{'}\left(c_n\right) \frac{\Delta t^2}{2} \\ &+ S_{II}\left(c_n\right) S_{II}^{'}\left(c_n\right) \frac{\Delta t^2}{2} \\ &+ O\left(\Delta t^3\right) \end{aligned} \tag{1.4}$$

where $\frac{\mathrm{d}S_{I,II}}{\mathrm{d}t} = \frac{\partial S_{I,II}}{\partial c} \frac{\mathrm{d}c}{\mathrm{d}t} = S_{I,II}^{'}\left(S_I + S_{II}\right)$, while the expansion of the split system (1.3) yields

$$\begin{aligned} c_{n+1} = c_n &+ S_I\left(c_n\right) \Delta t + S_{II}\left(c_n\right) \Delta t + S_I\left(c_n\right) S_I^{'}\left(c_n\right) \frac{\Delta t^2}{2} \\ &+ 3 S_I\left(c_n\right) S_{II}^{'}\left(c_n\right) \frac{\Delta t^2}{2} + S_{II}\left(c_n\right) S_I^{'}\left(c_n\right) \frac{\Delta t^2}{2} \\ &+ S_{II}\left(c_n\right) S_{II}^{'}\left(c_n\right) \frac{\Delta t^2}{2} \\ &+ O\left(\Delta t^3\right) \; . \end{aligned} \tag{1.5}$$

It can easily be seen, that this mechanism involves a splitting error of first order as the right hand side of eq. (1.5) does not match eq. (1.4) in the 2nd order term for $S_I\left(c_n\right) S_{II}^{'}\left(c_n\right)$.

Nevertheless, the advantage of a process subdivision of this kind is that the subprocesses can be treated independently with their appropriate time step and the integration scheme of choice in implicit or explicit form. Also the computational effort might be reduced significantly.

Much more than the low order, a major disadvantage of this mechanism is the sequential treatment of the processes: an integration step of one functional process is followed by an integration step of a process with completely different functional relations. Thus, discontinuities are introduced to the system resulting in stiff transients. In relation to this problem, an investigation of several experiments presented by (Sportisse, 2000) implied, that it appears preferable to treat the stiffer process last.

Another disadvantage is the fact that this mechanism is relying on synchronised methods, as the time interval covered by the two sub-integrations steps should be identical in order to avoid further inconsistencies. This implies, for instance, a leap-frog scheme should not be followed by an Euler scheme.

Source Splitting

A mechanism that overcomes the problem of discontinuity is the so called *Source Splitting* mechanism (also described in the atmospheric chemistry context by Sun, 1996 and Blom and Verwer, 2000).

Here, instead of directly subdividing the integration operator the source term is estimated. Usually the adoption of the global time step Δt_{glob} is dictated by the faster evolving process, while the source term for the slower evolving process is estimated on a equal or coarser time interval $\Delta t_{\tilde{S}} \geqslant \Delta t_{glob}$ and kept constant for the intermediate steps, as variations

over this period are assumed to be small:

$$\mathbf{c}_{n+1} = \mathbf{c}_n + \int_{t_n}^{t_{n+1}} \mathbf{S}_{fast}(\mathbf{c})\mathrm{d}t + \tilde{\mathbf{S}}_{slow} \underbrace{(t_{n+1} - t_n)}_{\Delta t_{glob}} . \tag{1.6}$$

The piecewise constant source term $\tilde{\mathbf{S}}_{slow}$ is usually obtained from a sub-integration that considers only the slower process:

$$\tilde{\mathbf{S}}_{slow} = \frac{\mathbf{\Phi}_{n+s} - \mathbf{c}_n}{\Delta t_{\tilde{S}}} , \tag{1.7}$$

where

$$\mathbf{\Phi}_{n+s} = \mathbf{c}_n + \int_{t_n}^{t_{n+s}} \mathbf{S}_{slow}\left(\mathbf{\Phi}\right) \mathrm{d}\tau \tag{1.8}$$

and $\Delta t_{\tilde{S}} = t_{n+s} - t_n \geqslant \Delta t_{glob}$ is the time frame in which $\mathbf{S}_{slow}$ is considered constant. The conceptual error here lies in ignoring the effects of the faster evolving process on the slower one. Due to the piecewise constant treatment of the slower process this mechanism is clearly of first order, as a Taylor expansion over the splitting interval $\Delta t_{\tilde{S}}$ shows (compare to eq. (1.4)):

$$\begin{aligned} c_{n+1} = c_n &+ S_{fast}\left(c_n\right) \Delta t_{\tilde{S}} + S_{slow}\left(c_n\right) \Delta t_{\tilde{S}} + S_{fast}\left(c_n\right) S'_{fast}\left(c_n\right) \frac{\Delta t_{\tilde{S}}^2}{2} \\ &+ S_{slow}\left(c_n\right) S'_{fast}\left(c_n\right) \frac{\Delta t_{\tilde{S}}^2}{2} \\ &+ O\left(\Delta t_{\tilde{S}}^3\right) . \end{aligned} \tag{1.9}$$

The splitting error is determined by the missing of the 2nd order terms that involve the derivative of the slower evolving process $S'_{slow}\left(c_n\right)$ and thus, by the quality of the initial assumption of constant Source S_{slow}.

However, the problem of the stiff transients is removed and the scheme promises a much more stable behaviour. In fact, in Verwer et al. (2001) the case of a stiff linear system of ODE's with analytical solution is presented

for three schemes: a first and second order *Operator Splitting* scheme and a *Source Splitting* Scheme, with the result, that while the loss of precision of the *Source Splitting* scheme with respect to the 2nd order *Operator Splitting (Strang, 1968)* scheme was small, the stiff part was resolved much more accurately through the *Source Splitting* mechanism with respect to both *Operator Splitting* versions.

Note further, that for a constant Source term $\mathbf{S}_{slow}$the *Source Splitting* technique solves the equations correctly. (The terms for S'_{II} vanish, so that eq. (1.4) and (1.9) match.)

Consistency of the slow process estimation $\tilde{S}_{slow}$

The estimation term in equation (1.7) works correctly for most integration schemes like the various one-step methods and the schemes of Adams type. However, there are some exceptions.

Consider for example a leap-frog scheme for the sub-integration of the estimated process with $\Delta t_{\tilde{S}} = \Delta t_{phys} = \Delta t$:

$$\mathbf{\Phi}_{n+1} = \mathbf{c}_{n-1} + 2\mathbf{S}_{slow}\left(\mathbf{c}_n\right)\Delta t\ . \tag{1.10}$$

Using the above estimation scheme (1.7), the estimation becomes

$$\tilde{\mathbf{S}}_{slow} = 2\mathbf{S}_{slow}\left(\mathbf{c}_n\right) - \frac{\mathbf{c}_n - \mathbf{c}_{n-1}}{\Delta t}\ . \tag{1.11}$$

For an explicit Euler scheme in the overall integration this leads to

$$\mathbf{c}_{n+1} = \mathbf{c}_{n-1} + \left(2\mathbf{S}_{slow}\left(\mathbf{c}_n\right) + \mathbf{S}_{fast}\left(\mathbf{c}_n\right)\right)\Delta t\ . \tag{1.12}$$

A consequence of eq. (1.12) is that the slower process damps the effect of the fast process significantly by taking out the previous integration step, transforming eq. (1.12) from an Euler step into an incomplete leap-frog step.

However, using an estimation term adapted to the sub-integration scheme,

as in

$$\tilde{\mathbf{S}}_{slow} = \frac{\mathbf{\Phi}_{n+1} - \mathbf{c}_{n-1}}{2\Delta t} = \mathbf{S}_{slow}\left(\mathbf{c}_n\right) \tag{1.13}$$

one obtains the much more reasonable result

$$\mathbf{c}_{n+1} = \mathbf{c}_n + \left(\mathbf{S}_{slow}\left(\mathbf{c}_n\right) + \mathbf{S}_{fast}\left(\mathbf{c}_n\right)\right)\Delta t\ . \tag{1.14}$$

This simple example highlights the necessity to adapt the estimation of the slower process source $\mathbf{S}_{slow}$to the schemes used in the sub-integration of the actual implementation.

Mass conservation

In the context of dynamic ecosystem models mass conservation is a major concern for a reasonable integration mechanism.

The mechanisms described above are easily shown to conserve the total mass of the system as long as the sub-integration steps do. In case of the *Operator Splitting* method, mass conservation follows directly from the successive application of the two sub-integrations. For the Source Splitting mechanism mass conservation is ensured for $\sum \tilde{S}_{slow,i} = 0$ by the linear treatment.

1.1.2 Integration schemes

The coupling methods described can in principle utilise any numerical integration scheme. In order to evaluate possible numerical interactions between the coupling and the integration schemes all the numerical experiments described in the following have been carried out adopting the following four numerical schemes, that are representative of different families of integration schemes widely used in meteorological and oceanographic modelling. They are:

- The explicit Euler scheme:

$$\mathbf{c}_{n+1} = \mathbf{c}_n + \mathbf{S}\left(\mathbf{c}_n\right)\Delta t \tag{1.15}$$

the simplest of all three schemes, a one-step scheme of first order.

- The explicit Runge-Kutta scheme of second order:

$$\mathbf{c}_{n+1} = \mathbf{c}_n + \frac{\mathbf{S}(\mathbf{c}_n) + \mathbf{S}(\mathbf{c}_n + \mathbf{S}(\mathbf{c}_n)\Delta t)}{2}\Delta t \tag{1.16}$$

- The leap-frog scheme:

$$\mathbf{c}_{n+1} = \mathbf{c}_{n-1} + \mathbf{S}(\mathbf{c}_n)\Delta t \tag{1.17}$$

  an explicit two-step scheme of second order. Well known stability issues due to uncoupling of odd and even integration steps impose the application of a numerical filter, here the well-documented Asselin-filter (Asselin, 1972) was implemented:

$$\mathcal{F}(\mathbf{c}_n) = \mathbf{c}_n + \frac{\alpha}{2}(\mathbf{c}_{n+1} - 2\mathbf{c}_n + \mathbf{c}_{n-1}) \tag{1.18}$$

- The Adams-Bashforth scheme of second order:

$$\mathbf{c}_{n+1} = \mathbf{c}_n + (3\mathbf{S}(\mathbf{c}_n) - \mathbf{S}(\mathbf{c}_{n-1}))\frac{\Delta t}{2} \tag{1.19}$$

  an explicit, two-step scheme.

A simple test case

Before proceeding with the numerical experiments aimed to evaluate the coupling scheme an evaluation of the integration schemes by means of an application to a simple 0D ecological model has been carried out with the aim of isolating potential numerical problems arising from the application of a specific integration scheme to the biogeochemical state variable. The model is a three constituent non-dimensional N-P-D (nutrient, phytoplankton and detritus) model, such as the one used by Burchard et al. (2003).

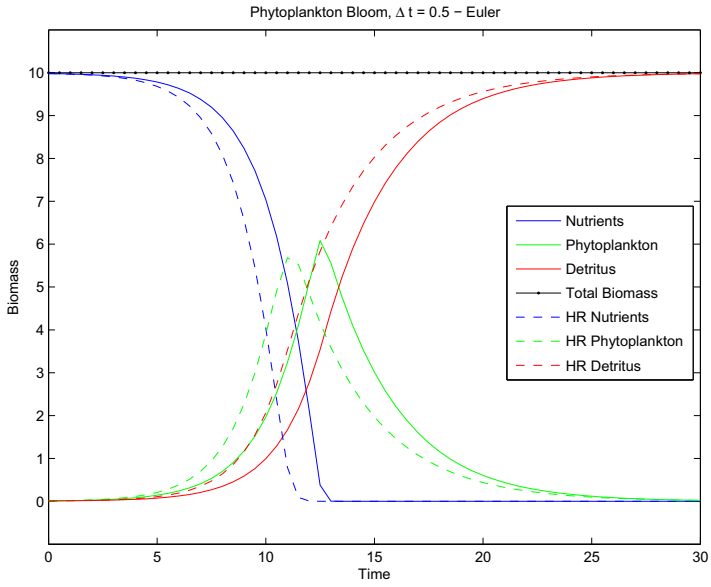


Figure 1.2: Euler scheme. Continuous lines refer to the simulations carried out with a Δt of 0.5.; dashed lines refer to simulations with a Δt of $6.25 * 10^{-4}$.*Blue:* nutrients, *green:* phytoplankton, *red:* detritus.

The governing equations for the three non-dimensional state variables are:

$$\begin{aligned} \frac{\mathrm{d}N}{\mathrm{d}t} &= -\frac{PN}{N+1} \\ \frac{\mathrm{d}P}{\mathrm{d}t} &= \frac{PN}{N+1} - \alpha P \\ \frac{\mathrm{d}D}{\mathrm{d}t} &= \alpha P \,. \end{aligned} \tag{1.20}$$

This system can be interpreted as a simplified model of the dynamics of a phytoplankton bloom: the phytoplankton takes up nutrients following a Michaelis-Menten law, mortality turns it into detritus at a fixed rate α. No detritus cycling is considered.

The initial conditions were set to $N_0 = 9.98$ and $P_0 = D_0 = 0.01$ giving

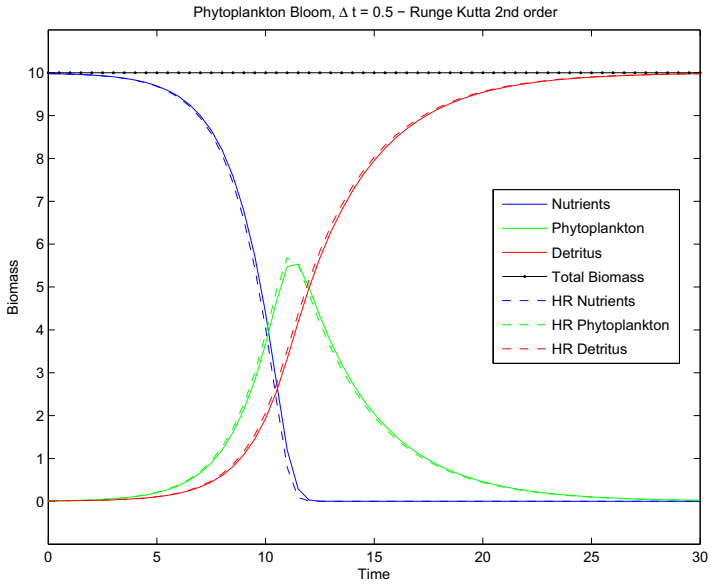


Figure 1.3: As in fig. 1.2, but for the Runge-Kutta scheme.

the system a total mass of 10. All computations used a loss rate of $\alpha = 0.3$, while the time step was set to $\Delta t = 0.5$.

The same simulation has been repeated by using the four integration schemes described above. For all integration schemes time-step cutting is applied to guarantee positive results, i.e. the time step is halved whenever the computed new state would be negative. To accomplish this the terms $\mathbf{c}_{n-1}$ and $\mathbf{S}_{n-1}$ for the leap-frog scheme and the Adams Bashforth scheme respectively are (quadratically and linearly) interpolated to the intermediate values at the new Δt without changing the order of the applied scheme.

Figures 1.2, 1.3, 1.4 and 1.5 show the results of the four investigated integration methods in comparison to a high resolution solution performed at a time step of $\Delta t = 6.25 * 10^{-4}$ where the four schemes do not show significant differences anymore.

It can be seen from the results that all four schemes conserve the total mass of the system, as the sum over the source terms of the three state

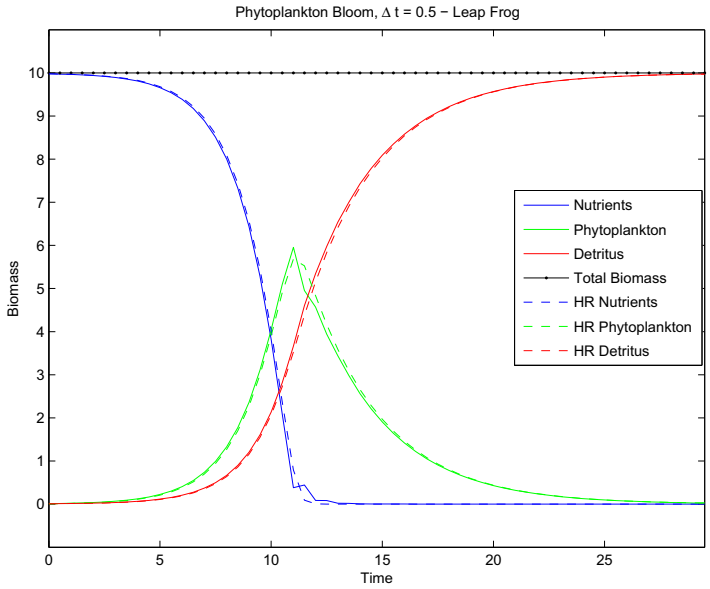


Figure 1.4: As in fig. 1.2, but for the Leap-frog scheme.

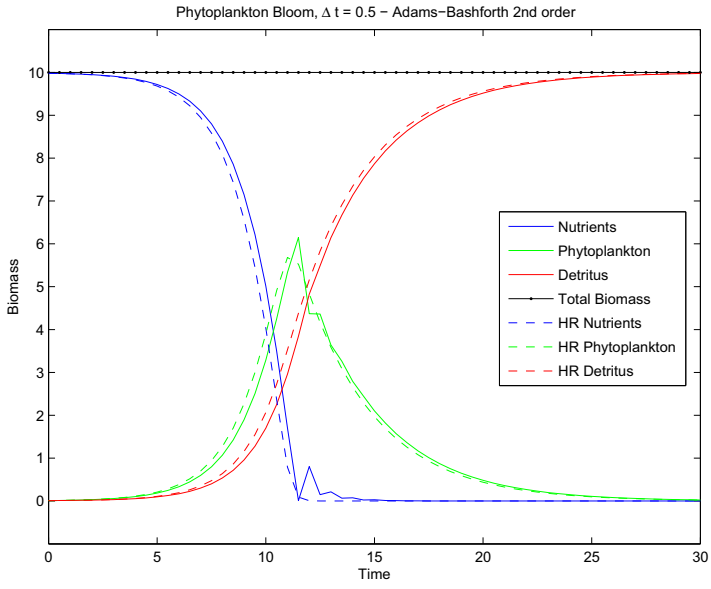


Figure 1.5: As in fig. 1.2, but for the Adams-Bashforth scheme.

variables cancel out and the source terms in eq. (1.15), (1.16), (1.17) and (1.19) are evaluated at the same time level for the various state variables, so that the total sum $N_{n+1} + P_{n+1} + D_{n+1}$- $N_n - P_n - D_n$ gives 0.

In terms of precision the three schemes of second order perform significantly better than the Euler scheme, which is delaying the bloom process by about 2 units of time with respect to the reference solution and the other schemes, that show similar results close to the reference solution.

In terms of stability the two multi-step methods (leap-frog and Adams-Bashforth) show significant oscillations in phytoplankton and nutrients when the bloom starts to decay, while the two one-step methods (Euler and Runge-Kutta) still behave completely stable.

The computational cost is lowest for the Euler method, the difference to the leap-frog and the Adams Bashforth method consists mainly in the need to keep memory of a previous state or source vector and in the filter operation for the leap-frog scheme in eq. (1.18). The Runge-Kutta method has a considerable higher cost as the Source term has to be re-evaluated after an Euler integration and the system has to be re-integrated. This higher cost might prove to be expensive in the context of a three-dimensional ecosystem model as the state vector will be of much higher dimension and the source-functions will be increasingly complex.

1.2 The model and its sub-models

A model of marine ecosystem dynamics has to solve the advection-diffusion equations for (non-conservative) active tracers (1.1). In one-dimensional form these can be written as:

$$\frac{\partial \mathbf{c}}{\partial t} + W\frac{\partial \mathbf{c}}{\partial z} = \frac{\partial}{\partial z}\left(\nu_{turb}\frac{\partial \mathbf{c}}{\partial z}\right) + \left.\frac{\partial \mathbf{c}}{\partial t}\right|_{bio}, \tag{1.21}$$

where $\mathbf{c}$ is the pelagic state vector (depending on temperature, salinity and solar radiation), ν_{turb} is the turbulent viscosity,.The vector $\left.\frac{\partial \mathbf{c}}{\partial t}\right|_{bio}$contains the changing rates of the state vector due to biogeochemical processes and W is a sinking velocity introduced for specific state variables.

Applying the ecosystem partitioning described in the beginning of the chapter, these equations can be formally divided as:

$$\frac{\partial \mathbf{c}}{\partial t} = \left.\frac{\partial \mathbf{c}}{\partial t}\right|_{phys} + \left.\frac{\partial \mathbf{c}}{\partial t}\right|_{bio} \tag{1.22a}$$

$$\left.\frac{\partial \mathbf{c}}{\partial t}\right|_{phys} = -W\frac{\partial \mathbf{c}}{\partial z} + \frac{\partial}{\partial z}\left(\nu_{turb}\frac{\partial \mathbf{c}}{\partial z}\right). \tag{1.22b}$$

The complete model is constituted by the three fundamental building blocks described in the introduction to this chapter, complemented by a numerical coupler adopting an appropriate coupling method (section 1.1.1). A schematic of the sub-model interaction in the 1-dimensional model is given in fig. 1.6. The physical model is forced by heat ($Q_s + Q_b + Q_h + Q_e$), water ($E - P - R$; E: evaporation, P: precipitation, R: riverine input) and momentum (wind stress, τ_w) fluxes, and computes vertical profiles of temperature (T), salinity (S) and turbulent diffusivity (ν). The temperature and salinity fields are passed to the biogeochemical model for the computation of the metabolic response of the biota (T) and the oxygen saturation concentration (S). The turbulent diffusivity is passed to the transport model (which handles also external nutrients and particulate matter input) for the computation of $\left.\frac{\partial \mathbf{c}}{\partial t}\right|_{phys}$. The biogeochem-

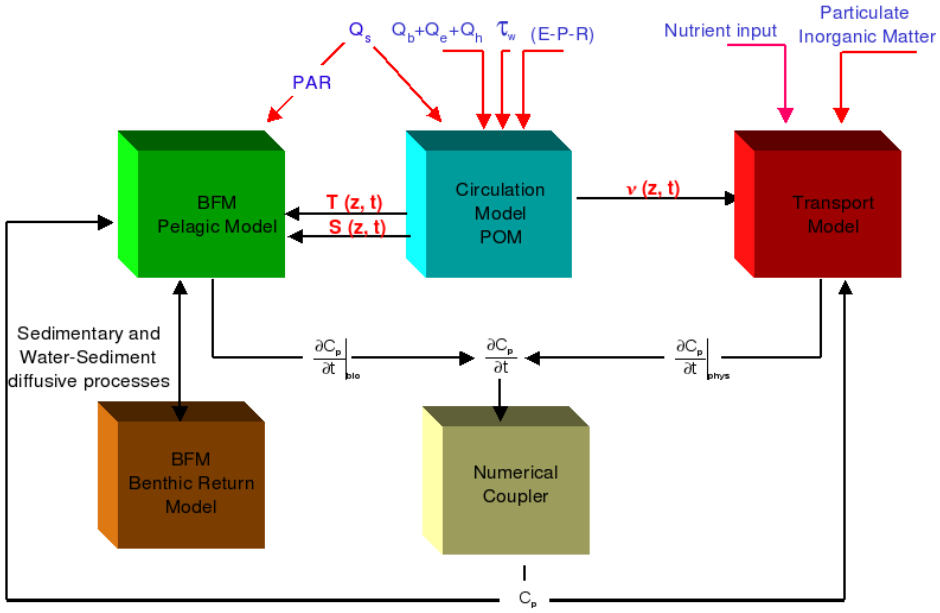


Figure 1.6: Schematic representation of the coupled physical biogeochemical model component and connections. Specific sub-models components for hydrodynamics and biogeochemistry are referred with the name of the models used in this work.

ical model is forced by solar radiation (Q_s), expressed as photosynthetically available radiation (PAR, about 50% of the incoming solar radiations flux) and computes $\frac{\partial \mathbf{c}}{\partial t}\big|_{bio}$. The coupling of the system occurs at the level of the numerical coupler, which merges $\frac{\partial \mathbf{c}}{\partial t}\big|_{bio}$ and $\frac{\partial \mathbf{c}}{\partial t}\big|_{phys}$ according to the characteristics of the coupling method used.

In the following, details of the specific sub-models used are presented, while the description of the coupling techniques adopted is given when describing the specific experiments.

1.2.1 The hydrodynamic model

The physical model applied in this implementation is the 1D-version of the Princeton Ocean Model (Blumberg and Mellor, 1987). It is based on the turbulent Navier-Stokes equations for geophysical flow in a rotating

coordinate system. The pressure is assumed to be in hydrostatic equilibrium and density differences are only considered in gravitational forces (Boussinesq approximation).

Due to the reduction to the vertical dimension, the inner physical dynamics of the system is determined exclusively by vertical turbulent transport, which is modelled by the two equation turbulence closure model developed by Mellor and Yamada (1982), that models the turbulent fluxes of the system by computing the turbulent diffusion coefficient through two additional state variables: the turbulent kinetic energy q and the turbulence length scale l.

Thus, the equations of motion to be solved are:

$$\begin{aligned} \frac{\partial U}{\partial t} - fV &= \frac{\partial}{\partial z}\left[(\nu_M)\,\frac{\partial U}{\partial z}\right] \\ \frac{\partial V}{\partial t} + fU &= \frac{\partial}{\partial z}\left[(\nu_M)\,\frac{\partial V}{\partial z}\right] \\ \frac{\partial P}{\partial z} &= -\rho g \end{aligned} \tag{1.23}$$

In these equations U and V are the horizontal velocity components, f is the Coriolis parameter and ν_M is the turbulent diffusivity for the velocity components, defined as

$$\nu_M = K_M + \chi_1 \,, \tag{1.24}$$

where χ_1 is a constant background diffusivities to adjust for unresolved effects, and the turbulent diffusion coefficient K_M is defined as

$$K_M = qlS_M \,. \tag{1.25}$$

where parameter S_M is a stability function of the Richardson Number.

The turbulent kinetic energy q and the turbulent length scale l are de-

termined by the balance equations:

$$\begin{aligned}\frac{\partial q^2}{\partial t} &= \frac{\partial}{\partial z}\left[K_q \frac{\partial q^2}{\partial z}\right] + P_s + P_b + \epsilon \\ \frac{\partial q^2 l}{\partial t} &= \frac{\partial}{\partial z}\left[K_q \frac{\partial q^2 l}{\partial z}\right] + E_1 l \left(\frac{P_s}{2} + E_2 \frac{P_b}{2}\right)\tilde{W} - \epsilon l\end{aligned} \tag{1.26}$$

Again K_q is the turbulent diffusion coefficient for the turbulence state variables analogous to K_M.

$$P_B = 2K_M \left[\left(\frac{\partial U}{\partial x}\right)^2 + \left(\frac{\partial V}{\partial x}\right)^2\right] \tag{1.27}$$

gives the production of turbulent energy due to velocity shear, while

$$P_B = \frac{2g}{\rho_0} K_H \frac{\partial \rho}{\partial z} \tag{1.28}$$

is the buoyancy production. $\tilde{W}$ is a wall proximity function depending on l.

With this set of closed equations the turbulent diffusivity coefficient for tracers K_H can be determined analog to K_q and K_M, again combined with a background diffusivity parameter χ_H to allow the compensation for unresolved small scale processes, such that the turbulent viscosity in equation (1.21) becomes

$$\nu_{turb} = K_H + \chi_H \ . \tag{1.29}$$

Therefore, the dynamics of the temperature and salinity field is completely described by the equations

$$\begin{aligned}\frac{\partial T}{\partial t} &= \nu_{turb} \frac{\partial}{\partial z}\left[\frac{\partial T}{\partial z}\right] + \frac{\partial I}{\partial z} \\ \frac{\partial S}{\partial t} &= \nu_{turb} \frac{\partial}{\partial z}\left[\frac{\partial S}{\partial z}\right] - \omega_S \ ,\end{aligned} \tag{1.30}$$

where

$$I = \frac{Q_s}{\rho_W c_p} Tr \, \exp(\lambda z) \tag{1.31}$$

is the fraction Tr of solar radiation Q_s penetrating the water column exponentially attenuated by the coefficient λ for infrared light. A compensation technique for the missing horizontal advection of salinity is realised by ω_S, that will be given in section 1.3 along with a description of the boundary and initial conditions, as well as the forcing functions applied to the model.

The original numerical scheme by *POM* (refer to Blumberg and Mellor, 1987; Mellor, 2004 for details), applied here to integrate in time the 1D partial differential equation system composed of eq. (1.23), (1.26) and (1.30), is a semi-implicit leap-frog scheme:

$$\phi_{n+1} = \phi_{n-1} + 2\frac{\partial}{\partial z}\left(\nu_{turb}\frac{\partial \phi_{n+1}}{\partial z}\right)\Delta t + 2S_\phi(\phi_n)\,\Delta t \tag{1.32}$$

where S_ϕ is a generic production term for the variable ϕ. Note, that the only remainders of inertial terms are the Coriolis terms for the horizontal velocity components in eq. (1.23).

1.2.2 The biogeochemical model

The ecosystem model used in this work is based on the pelagic part of the Biogeochemical Flux Model (*BFM* - Vichi et al., 2006, 2007b; `http://www.bo.ingv.it/bfm`), an advancement of the European Sea Regional Ecosystem Model (*ERSEM* - Baretta et al., 1995). In this model, the ecosystem is represented by a selection of interacting chemical and biological processes that are supposed to reproduce the system behaviour coherently.

The BFM State variables are considered in their basic biogeochemical components: carbon (C), nitrogen (N), phosphorus (P), silicon (Si) etc., and can be divided into two broad functional classes (Vichi et al., 2007b): the living functional groups (LFG) and the chemical functional families

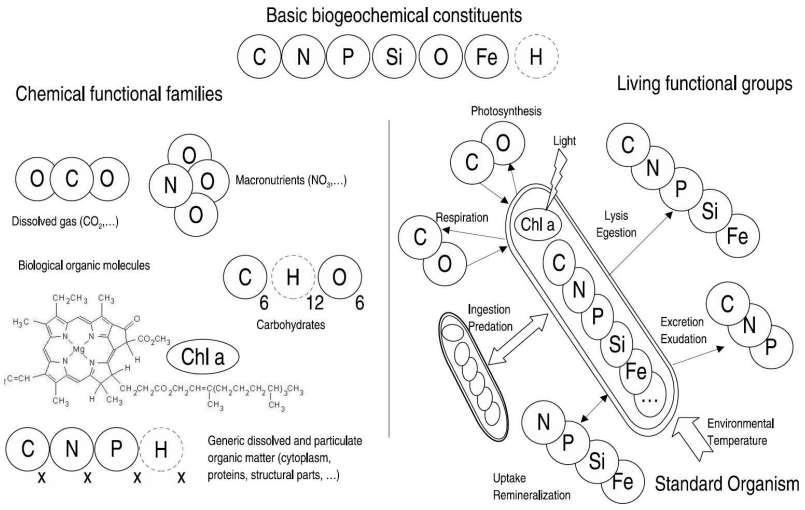


Figure 1.7: Scheme of the three major constituents of the three major components of the pelagic biogeochemical model: basic constituents, chemical functional families and the standard organism, which is the prototype of any living functional group. Hydrogen is not considered as biogeochemical constituent, but indicated for completeness of the chemical compound formulations. (Vichi et al., 2007b)

(CFF), depicted in fig. 1.7. LFG dynamics is described by population dynamics (growth, mortality, predation) and physiological processes (uptake, ingestion, respiration, excretion, egestion). LFG represent the whole biota and can be divided into three functional types: producers (phytoplankton), decomposers (bacteria) and consumers (zooplankton). Each LFG is designed on the basis of the standard organism approach (Baretta et al., 1995): all organisms merged in the same functional group (e. g. diatoms, nanoflagellates, picophytoplankton etc.) share the same functional properties and have the same trophic interactions, differing only in the specific parametrisation of each functional group. This kind of approach has been confirmed to be very valid for the description of population and community dynamics of unicellular organisms, while it shows several problems when dealing with size structured populations such as mesozooplankton. The dynamical interactions between LFG and CFF in terms of basic chemical constituent flux is schematised in Fig. 1.8.

It should be stressed, that the *BFM* ecosystem structure explicitly resolves the two major pathways of the marine ecosystem: the so called "herbivorous food web" (connection between primary and secondary producers is directly established between phytoplankton and mesozooplankton through grazing) and the so called "microbial food web" (mass and energy flow between primary and secondary producers mediated by phytoplankton production of dissolved organic matter, bacteria utilisation and microzooplankton predation of bacteria). Previous modelling studies (Vichi et al., 2003) have shown that the shift between the two configurations of the trophic web is naturally reproduced by the *BFM* in response to the changing environmental conditions. Moreover, the *BFM* allows also for the intermediate configuration of the trophic web (the "Multivorous" food web proposed by Legendre and Rassoulzadegan (1995)). A complete listing of the *BFM*'s state variables is given in table 1.1. With respect to the *ERSEM* model the *BFM* differs mostly in the description of the primary production processes. The *BFM*, in fact, explicitly considers chlorophyll as a basic biogeochemical constituent and the chlorophyll synthesis has

Variable	Constituent	Description	Type
$O2$	O	Dissolved oxygen	CFF
$N1$	P	Phosphate	CFF
$N3$	N	Nitrate	CFF
$N4$	N	Ammonium	CFF
$N5$	Si	Silicate	CFF
$N6$	R	Reduction Equivalents	CFF
$P1$	C, N, P, Si, L	Diatoms	LFG
$P2$	C, N, P, L	Nanoflagellates	LFG
$P3$	C, N, P, L	Picophytoplankton	LFG
$Z3$	C, N, P	Carnivorous Mesozooplankton	LFG
$Z4$	C, N, P	Omnivorous Mesozooplankton	LFG
$Z5$	C, N, P	Microzooplankton	LFG
$Z6$	C, N, P	Heterotrophic Flagellates	LFG
$R1$	C, N, P	Dissolved Organic Detritus	CFF
$R2$	C, N, P	Carbohydrates	CFF
$R6$	C, N, P, Si	Particulate Organic Detritus	CFF

Table 1.1: The ecosystem state vector. The units used for the various constituents are $\left[\frac{mg}{m^3}\right]$ for carbon and chlorophyll and $\left[\frac{mmol}{m^3}\right]$ for all others. (C: Carbon, N: Nitrogen, P: Phosphorus, S: Silicon, L: Chlorophyll-a; CFF: chemical functional family, LFF: Living Functional Group)

been parametrised according to Geider et al. (1997).

Phytoplankton uptake of macronutrients (nitrate, ammonia and phosphate) has been decoupled from the carbon assimilation processes by including dynamic nutrient kinetics according to Droop (1973) and Nyholm (1977), depending on both, the level of intra-cellular nutrient storage. The only exception is the silicate uptake that is regulated by the external concentration.

The microbial food web includes bacteria, heterotrophic flagellates and microzooplankton, each with dynamically varying C:N:P ratios (Baretta-Bekker et al., 1995, 1998). Bacteria act to consume dissolved organic carbon and decompose detritus and, in addition to that, they can compete for inorganic nutrients with phytoplankton under conditions of severe

nutrient depletion in the organic substrate.

The biogeochemical changing rate of each state vector component (given in table 1.1) is given by an equation of the type

$$\left.\frac{\partial c}{\partial t}\right|_{bio} = \left.\frac{\partial c}{\partial t}\right|_{v1}^{e1} + \left.\frac{\partial c}{\partial t}\right|_{v2}^{e2} - \left.\frac{\partial c}{\partial t}\right|_{v3}^{e3} \cdots \left.\frac{\partial c}{\partial t}\right|_{vn}^{en} \tag{1.33}$$

where the superscripts e indicate the process and the subscripts v refer to the state variable playing the counterpart in the process. Mass conservation obviously dictates $\left.\frac{\partial c}{\partial t}\right|_{v1}^{e1} = -\left.\frac{\partial v}{\partial t}\right|_{c1}^{e1}$.

For the three living functional groups phytoplankton (P), Zooplankton (Z) and bacteria (B) the rates are composed of:

$$\begin{aligned}
\left.\frac{\partial P}{\partial t}\right|_{bio} &= \left.\frac{\partial P}{\partial t}\right|_{upt} - \left.\frac{\partial P}{\partial t}\right|_{exu} - \left.\frac{\partial P}{\partial t}\right|_{lys} - \left.\frac{\partial P}{\partial t}\right|_{resp} - \left.\frac{\partial P}{\partial t}\right|_{graz} \\
\left.\frac{\partial Z}{\partial t}\right|_{bio} &= \left.\frac{\partial Z}{\partial t}\right|_{ingest} - \left.\frac{\partial Z}{\partial t}\right|_{egest} - \left.\frac{\partial Z}{\partial t}\right|_{resp} - \left.\frac{\partial Z}{\partial t}\right|_{pred} \\
\left.\frac{\partial B}{\partial t}\right|_{bio} &= \left.\frac{\partial B}{\partial t}\right|_{upt} - \left.\frac{\partial B}{\partial t}\right|_{remin} - \left.\frac{\partial B}{\partial t}\right|_{resp} - \left.\frac{\partial B}{\partial t}\right|_{pred}
\end{aligned} \tag{1.34}$$

where the subscripts indicate the processes of uptake (*upt*), exudation (*exu*), lysis (*lys*), respiration (*resp*), grazing(*graz*), ingestion (*ingest*), egestion (*egest*), predation (*pred*) and remineralisation (*remin*).

Virtually all processes in the ecosystem depend on the water temperature $T\,[°C]$. This is modelled by a regulating factor

$$f^t = Q_{10}^{\frac{T-10}{10}} \tag{1.35}$$

with Q_{10} as a parameter specific to each LFG.

Moreover, primary production is heavily depending on ambient light, or better photosynthetically available radiation PAR, given by the Lambert-

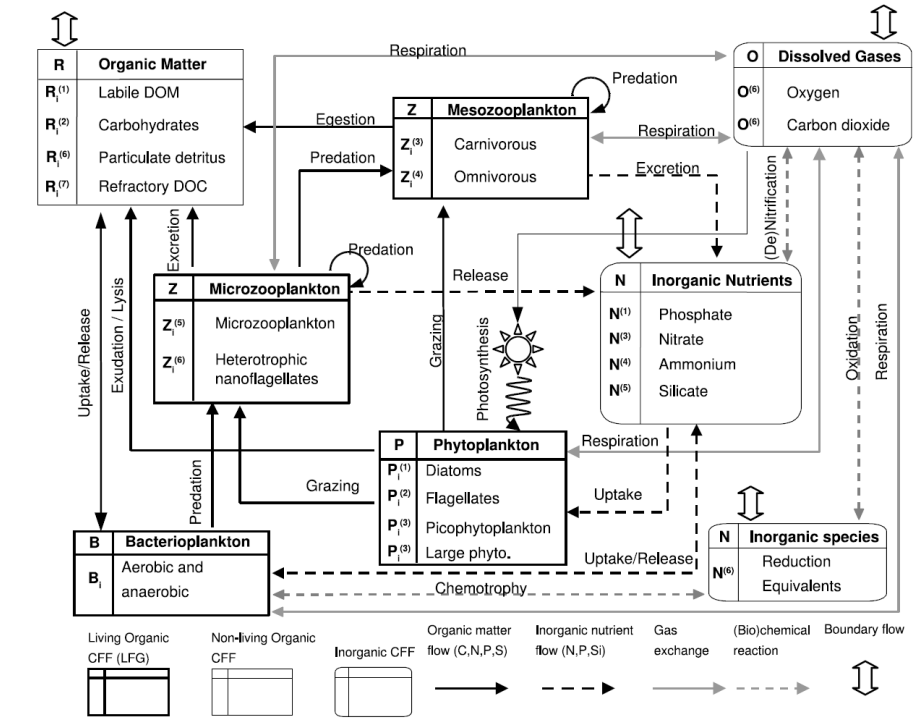


Figure 1.8: General overview of the pelagic food web of the *BFM*. Squared bold boxes correspond to $LFGs$, squared thin boxes to non-living organic $CFFs$ and rounded boxes to non-organic $CFFs$. Dashed black arrows represent nutrient fluxes, Continuous black arrows nutrient *and* carbon fluxes, dashed grey arrows indicate biogeochemical reactions and continuous grey arrows gas exchange. Boundary flows are given by double arrows.

Beer formulation

$$I = \frac{Q_s}{\rho_W c_p} \varepsilon_{PAR} \exp\left(\int_z^0 \left(\lambda_w(\zeta) + \lambda_{eco}(\zeta) \right) \mathrm{d}\zeta \right) , \tag{1.36}$$

where ε_{PAR} gives the fraction of photosynthetically available radiation, λ_w is the background extinction of water particles and λ_{bio} is the extinction due to phytoplankton, particulate detritus and suspended inorganic matter.

A simple benthic return model has been used for the benthic closure. In order to parameterise the benthic re-mineralisation, a fixed quota of each detritus component (C,N,P,Si) reaching the bottom is returned to the water column as nutrients.

1.2.3 The transport model

The transport equation of the ecosystem state variables is given by eq. (1.22b). To solve this equation *POM's* semi-implicit integration scheme applied to the physical state variables, as in eq. (1.32), was used. The transport of the biological state vector in discrete form is then given by the equation:

$$\tilde{\mathbf{c}}_k = \mathbf{c}_{n-1,k} + \frac{1}{2\Delta\sigma} \left(\left(\mathbf{c}_{n,k-1} + \mathbf{c}_{n,k} \right) W_k - \left(\mathbf{c}_{n,k} + \mathbf{c}_{n,k+1} \right) \right) W_{k+1} 2\Delta t , \tag{1.37}$$

$$\mathbf{c}_{n+1} = \tilde{\mathbf{c}} + \mathit{Diff}\left(\mathbf{c}_{n+1} \right) 2\Delta t . \tag{1.38}$$

where k refers to the vertical levels, n to time levels, $\Delta\sigma$ is the cell thickness in the σ-coordinate, Δt represents the time step and Diff is a shorthand for the spatial discretisation of the implicit diffusion term, that is not of interest here and therefore omitted

As for the physical state variables, the reader is referred to section 1.3 for the description of boundary conditions and forcing.

The whole model parametrisation specific to this implementation is described in section 1.3.

1.3 Implementation of the coupled model

Here a summary of the implementation characteristics is given. The implementation is based on the previous one-dimensional modelling effort (for the same locations) of Vichi et al. (2003) that carried out one dimensional simulations of the marine ecosystem dynamics using the same physical model (*POM*), but the ecological model implemented was the *ERSEM* model (Baretta et al., 1995), from which the *BFM* originated.

1.3.1 Implementation sites

The coupled model described in section 1.2 has been implemented at the three different locations, (named S1, S3 and AA1) in the northern Adriatic Sea depicted in fig. 1.9

The rationale for the choice of these sites lies in the fact that (although belonging to the coastal domain) they show distinct hydrological and ecological characteristics and, at the same time, they correspond to sampling sites frequently visited in the framework on national research projects that yielded observational data useful for model validation.

Site S1 is directly influenced by the Po river runoff providing large amounts of nutrients and organic/inorganic, particulate/dissolved matter (Cozzi et al., 1999; Pettine et al., 1998; Giani et al., 2000, 2001). It is located about 5 km offshore from the Po delta and has a bottom depth of 20 m.

Site S3 (37 km offshore of the Western Adriatic coast) has characteristics more typical of open sea areas, being less affected by land based inputs. It is approximately located in the central part of the northern Adriatic central gyre (Artegiani et al., 1997a). The bottom depth is 30 m.

The hydrological seasonal cycle at S1 and S3 is similar, but the density stratification at S1 is present for most of the year due to the Po freshwater input. On the contrary S3 is well mixed during autumn-winter and its seasonal characteristics are more influenced by the variability of the local cyclonic circulation.

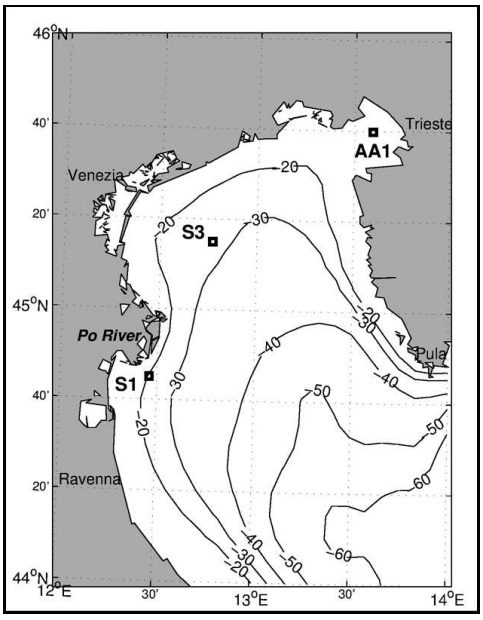


Figure 1.9: Location of the implementation sites in the northern Adriatic Sea.

Site AA1 (with a bottom depth of 20 m) is located in the centre of the Gulf of Trieste, a relatively small semi-enclosed basin at the north-eastern corner of the northern Adriatic basin. The fresh water input (Isonzo river) is significant, particularly in view of the relatively small volume of the basin (Malej et al., 1995). Its hydrology is characterised by a strong inter-annual variability, mainly related to the strong easterly winds (Bora) affecting the area. The lower trophic levels cycle is strongly influenced by the variability of the abiotic factors (Fonda Umani et al., 1995; Cataletto et al., 1995; Mozetic et al., 1998).

1.3.2 Vertical discretisation

The water column at the implementation sites has been vertically discretised by applying a logarithmic distribution of layer depths at the surface and at the bottom, and a uniform depth distribution in the water column

interior. The S1 and AA1 vertical grid is defined by 30 levels with 6 levels logarithmically distributed at both the surface and the bottom. The S3 water column has been resolved by 40 levels with an analogous logarithmic distribution at surface and at the bottom. In this way at all implementation sites the interior of the water column has a vertical resolution of about 1 m.

This discretisation was considered sufficiently for the scope of representing the annual cycle in the ecosystem dynamics at the given sites. As mentioned in the introduction each model of fluid dynamics is restricted to a certain resolution, that covers the representation of the processes under investigation. The present model is designed to give a qualitatively and quantitatively reasonable description of annual vertical cycle of the water column.

1.3.3 Boundary conditions

For the momentum equation (1.23) the boundary fluxes are given by the shear stresses at the interfaces, i. e.

$$\nu_M \left. \frac{\partial \sqrt{U^2 + V^2}}{\partial z} \right|_{z=0,-H} = \frac{\tau_{w,b}}{\rho_W} \,, \tag{1.39}$$

where τ_w is the wind stress, τ_b is the bottom stress and ρ_W is a reference water density. To compute the bottom stress a logarithmic drag law is used with a bottom roughness length of $0.01m$.

The surface boundary condition for temperature results from the heat budget:

$$\nu \left. \frac{\partial T}{\partial z} \right| = \frac{(1 - Tr)\, Q_s - Q_b - Q_h - Q_e + Q_{corr}}{\rho_W c_p} \tag{1.40}$$

where Q_s is the solar radiation, Q_b is the long wave radiative flux, Q_h is the sensible heat flux, Q_e the latent heat flux and Q_{corr} a correction term to compensate for advective heat fluxes into the water column (see Vichi et al. (1998)). Only a fraction of solar radiation acts as boundary heat flux,

as the other part is considered to enter the water column and appears in eq. (1.30),(1.31) as heat source. The bottom boundary layer is assumed to be adiabatic.

Surface salinity is relaxed to a climatological value $\hat{S}$:

$$\nu_H \left.\frac{\partial S}{\partial z}\right|_{z=0} = -\alpha_S \left(S\left(t\right)|_{z=0} - \hat{S}\left(t\right)\right) , \tag{1.41}$$

where α_S (see table 1.2) is the relaxation velocity. Again, at the bottom boundary a no flux condition is imposed. The turbulent kinetic energy at surface and bottom is assumed to be proportional to the friction velocity by an empirical constant, while the turbulent length scale is set to 0 on the boundaries.

For the biogeochemical state variables, the surface fluxes for the nutrients are given by a relaxation to seasonally varying climatological data $\hat{N}$ (from ABCD, Zavatarelli et al., 1998), analogous to the salinity surface flux in eq.(1.41)

$$\nu_H \left.\frac{\partial N}{\partial z}\right|_{z=0} = -\alpha_N \left(N\left(t\right)|_{z=0} - \hat{N}\left(t\right)\right) , \tag{1.42}$$

while at the bottom the flux is given by the nutrient return ω_N from the benthic closure model, as described in section 1.2.2, multiplied by the cell thickness of the last layer of the water columnΔz_{bot}:

$$\nu_H \left.\frac{\partial N}{\partial z}\right|_{z=-H} = -\omega_N \Delta z_{bot} , \tag{1.43}$$

For all other state variables no-flux conditions are applied at surface and bottom.

1.3.4 Forcing functions

The wind stress forcing function τ_w used to evaluate boundary conditions (eq. (1.39)) and the heat flux terms Q_b, Q_h, Q_e for eq. (1.40), plus shortwave incoming radiation flux Q_s in eq. (1.31 and 1.40), were all calculated

from the 6-hours ECMWF (European Center for Medium-range Weather Forecast) surface reanalysis of atmospheric properties relative to the 1982-93 period (Gibson et al., 1997). The surface fluxes are computed following the procedures described in Maggiore et al. (1998) and Zavatarelli et al. (2002). Surface salinity data $S(t)$ in eq. (1.41) for the S1 and S3 site are taken from the Adriatic BiogeoChemical Data set (ABCD, Zavatarelli et al. 1998) as monthly mean climatological time series considering an adequate area centred around the implementation sites (fig. 1.9). Climatological values of salinity at the AA1 site were calculated from a time series collected by the Marine Biology laboratory (Trieste University) between 1986 and 1997.

Another important external forcing function that has been shown to largely affect the simulated biological properties is the Inorganic Suspended Matter (ISM) in the water column (Vichi et al., 1998). The concentration of suspended sediments modifies the ambient light through extinction processes and consequently constrains phytoplankton productivity. The ISM profiles are applied in the model as external forcing functions affecting the light extinction coefficient for biology as in eq. (1.36); the ISM concentrations at the three sites are computed as linear interpolation of the seasonal profiles obtained by Vichi et al. (2003) from direct measurements at the implementation sites. Unfortunately, there is scarce climatological information concerning the seasonal mean concentrations of ISM in the northern Adriatic. At the AA1 site observations collected monthly over the period 1997-2000 are used, from which seasonal mean concentration profiles have been calculated. For S1 and S3, the PRISMA-I (Progetto di Ricerca per la salvaguardia del Mare Adriatico) project data collected during only four seasonal surveys is applied. Therefore, those data are not climatologically representative.

Perpetual time series of nutrients at surface ($\hat{N}$ in eq. (1.42)) are climatological mean seasonal values extracted from the ABCD data set in the case of S1 and S3, and seasonal mean values from the Laboratorio di Biologia marina monitoring program for the Gulf of Trieste. The sea-

sonal frequency was the only possible due to the systematic lack of observations in some months. The main differences in the three time series (not shown) are that both S1 and AA1 show high nitrate and phosphate surface concentrations, typical of river-affected coastal stations, while S3 presents lower surface values, indicating oligotrophic characteristics that are proper of open-sea areas. Besides that, at S1, all nutrients show a distinct, strong peak in autumn (concentrations are about 10 times more than at the other sites), while S3 and AA1 do not show such a significant autumn increase.

1.3.5 Physical parametrisation

The application of 1-D vertical models for the representation of hydrodynamic processes usually requires the use of additional parameterisations to account for the missing of horizontal dynamics. In the specific case of the northern Adriatic Sea, vertical processes alone (which are determined by the specification of surface heat, momentum and water fluxes) are not completely sufficient for an appropriate description of the seasonal evolution of the water column structure. Therefore, it is necessary to include a closure of the annual heat and water budgets, in order to let the model reproduce a perpetual climatological annual cycle in the hydrodynamics.

The annual heat flux budget is negative in the northern basin (Supic and Orlic, 1999), and is likely to be compensated (on a climatological time scale) by the advection of warm waters from the south (Artegiani et al., 1997b). This has been specified in the model, as proposed in Vichi et al. (1998), by calculating the annual surface heat loss in the forcing functions and distributing this bulk value along the year in the form of an empirical heat correction function added as a surface boundary condition (Q_{corr} in eq. (1.40)). The most suitable shape of this empirical function at the three sites was established by Vichi et al. (2003) by analysing the simulated seasonal profiles of T with respect to the observed means and their range of variability.

Parameter description	AA1	S1	S3
Background turbulent diffusivity T (and tracers) $\left[\frac{\mathrm{m}^2}{\mathrm{s}}\right]$	$2.5 * 10^{-5}$	$1.3 * 10^{-5}$	$1.0 * 10^{-5}$
Background turbulent diffusivity S $\left[\frac{\mathrm{m}^2}{\mathrm{s}}\right]$	$0.9 * 10^{-5}$	$0.9 * 10^{-5}$	$0.9 * 10^{-5}$
Apportioning coeff. for infrared light [%]	90.0	50.0	80.0
Attenuation coefficient for visible light $\left[\frac{1}{\mathrm{m}}\right]$	0.17	0.17	0.17
Attenuation coefficient for infrared light $\left[\frac{1}{\mathrm{m}}\right]$	5.0	5.0	5.0
Relaxation constant at surface for S $\left[\frac{\mathrm{m}}{\mathrm{d}}\right]$	0.5	0.5	5.68
Relaxation constant at surface for nutrients $\left[\frac{\mathrm{m}}{\mathrm{d}}\right]$	0.22	0.566	0.568
Relaxation constant to compensate for hor. adv. $\left[\frac{1}{\mathrm{d}}\right]$	$\frac{1}{60}$	$\frac{1}{60}$	$\frac{1}{10}$
Starting Depth of hor. adv. compensation $[\mathrm{m}]$	1.0	1.0	1.0
Percentage of photosynth. available radiation [%]	50.0	50.0	50.0
R6 contribution coefficient to light shading $\left[\frac{\mathrm{m}^2}{\mathrm{mg\,C}}\right]$	$1.0 * 10^{-4}$	$1.0 * 10^{-4}$	$1.0 * 10^{-4}$
ISM contribution coefficient to light shading $\left[\frac{\mathrm{m}^2}{\mathrm{mg\,C}}\right]$	$3.9 * 10^{-4}$	$1.3 * 10^{-4}$	$3.9 * 10^{-4}$

Table 1.2: Parameter values for the three implementation sites.

Concerning the closure of the water flux, a crucial problem when dealing with 1-D models is that the local buoyancy losses are not compensated by long-term, basin-wide lateral advection of buoyancy, in order to maintain a perpetual dynamic equilibrium in the water column. The imposition of local net positive or negative heat and water fluxes at the surface produces a model drift; in reality, the water column budget is closed by the horizontal advection processes. In a purely one-dimensional model the horizontal advection terms are neglected, and so it is necessary to parameterise the lateral advective adjustment process. This has been achieved by applying a closure of the water cycle based on the imposition of a surface salt flux correction (Vichi et al., 1998), as done for the heat flux in eq. (1.40).

In this work we use a parameterisation of local lateral advection representing the basin-scale contribution to the vertical water column stability. The method consists of the introduction of a climatological time-varying vertical profile of salinity, to which the dynamics of the water column has to adjust within a given time scale. The source term added to the salinity

balance in eq. (1.30) is parameterised as

$$\omega_S = \alpha_{adv}(z)\left(S - \hat{S}\right) \tag{1.44}$$

where α_{adv} is the depth-varying relaxation frequency, defined to be 0 from surface down to a given depth.

The various physical parameters applied are given in table 1.2

1.3.6 Biogeochemical parametrisation

The *ERSEM* parametrisation defined by Vichi et al. (2003) for the numerical implementation at the three northern Adriatic sites has been essentially maintained for the *BFM* implementation. However, the explicit consideration of chlorophyll-a involves for the three phytoplankton groups (diatoms, nanoflagellates and picophytoplankton) the addition of two new parameters: the maximal allowable chlorophyll-carbon ratio (ϑ_{max}) and the initial slope of the P-I (production-irradiance) curve (P_α). The values adopted for each phytoplankton functional group are given in table 1.3 and have been identically applied at all the implementation sites.

Parameter description	Diatoms	Nanoflag.	Picophytopl.
Max. carbon-chlorophyll ratio ϑ_{max} $[-]$	0.03	0.02	0.02
Init. slope of P-I curve α $\left[\frac{mgC}{m^2}\frac{\mu}{E}\frac{1}{mgChl}\right]$	$1.38 * 10^{-5}$	$1.1 * 10^{-5}$	$1.1 * 10^{-5}$

Table 1.3: Parameter values for the chlorophyll dynamics.

1.3.7 Initial conditions

The physical model is initialised from winter profiles of the climatological data sets (Zavatarelli et al., 1998) for the three implementation sites, that are described in section 1.3.4. Unfortunately, this kind of data is not available for the biogeochemical states. Some sensitivity studies to varying initial conditions have been performed by Vichi et al. (2003), indicating that

for the given sites after an appropriate spin-up the ecosystem converges to the same perpetual year cycle. In this implementation, corresponding initial conditions were applied, and a spin-up interval of 4 years was allowed.

1.3.8 Coupling

Of the two coupling schemes presented in section 1.1.1 after initial tests the Source Splitting mechanism was preferred for this implementation. In fact, the non-continuous, sequential character of the *Operator Splitting* method caused severe stiff transients to the system and introduced serious instabilities to the system, a problem that is even complicated by the fact, that the leap-frog integration used by *POM* requires the value of the state variable on the level of the previous time step which enhances the instability.

As stated above, the *Source Splitting* mechanism is based on the approximation that the slower of the two involved processes is considered constant on a certain coarse time interval and then their process rates are kept constant, while the finer global integration time step is imposed by the faster process.

While in atmospheric chemistry the chemical processes can be generally considered as the potentially faster evolving processes, this is not clear for models of marine ecosystems. A glimpse on figure 1.1 illustrates, that the scales involved are widely overlapping, and no clear a priori distinction can be done with respect to the hierarchy of the physical and biogeochemical time scales, but will depend on the choice of the spectral window and the discretisation and parametrisation of the processes, and might even depend on the local characteristics of the ecosystem. Concerning the resolution of the physical processes, the behaviour of the hydrodynamic sub-model has already been studied intensively not only in the already mentioned paper of Vichi et al. (2003), but also in several other works concerning the three-dimensional simulation of the Adriatic Sea

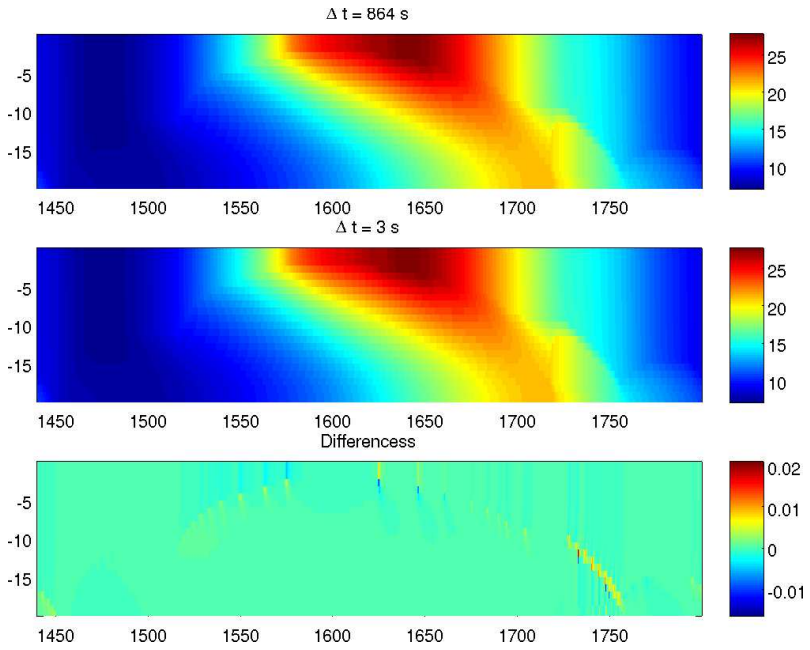


Figure 1.10: The temperature T in $[°C]$ at station AA1, 5th year of integration *On top:* the standard resolution of $\Delta t = 864s$. *In the middle:* high resolution solution at a time step of $\Delta t = 3s$. *At the bottom:* Differences of standard resolution to high resolution.

general circulation (e. g. (Zavatarelli et al., 2002; Zavatarelli and Pinardi, 2003; Oddo et al., 2005)), so that it was applied here without considerable modifications. In particular, the time step for the integration of the physical processes has been stabilised at the fixed step size of $\Delta t_{phys} = 864s$ (further on referred to as standard time step), which was confirmed in the present configuration by running the model adopting also a $\Delta t = 3s$ and comparing the solution for the identical run adopting $\Delta t = 864s$ (figures 1.10 and 1.11 for the station AA1). It can be seen, that the differences in between the two runs are limited to some noise around the pycnocline. Errors are at maximum 1% of the local state value for temperature and less than 0.2% for salinity and can therefore be safely neglected. [2]

Thus, in principle there are two possible scenarios for the implemen-

[2] A comparison of the seasonal profiles to climatological data is given in the Appendix.

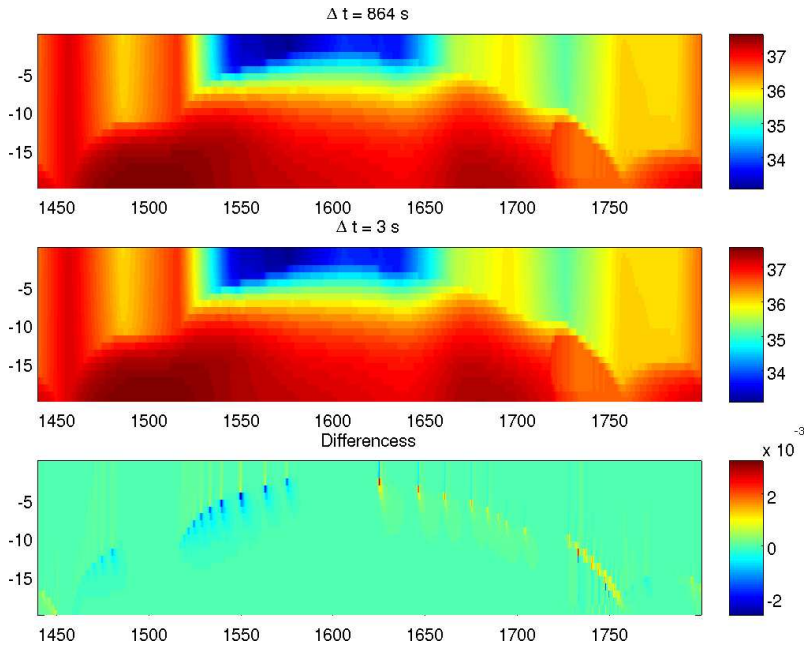


Figure 1.11: As fig. , but for the salinity S in $[psu]$.

tation of the coupling scheme into the model, which will be investigated below:

- The biogeochemical processes and their modelling require a time step finer than the one adopted for the transport processes of $\Delta t_{phys} = 864s$ and thus, the global time step is imposed by the biogeochemical processes. Numerical experiments carried out within this general framework will in the following be referred to as BFMStep.

- The biogeochemical processes and their modelling can be computed on a coarser time step than the Δt_{phys} of $864s$ stabilised for the transport processes. In this case the global integration scheme applies the time step of the transport model. Numerical experiments carried out within this general framework will in the following be referred to as POMStep.

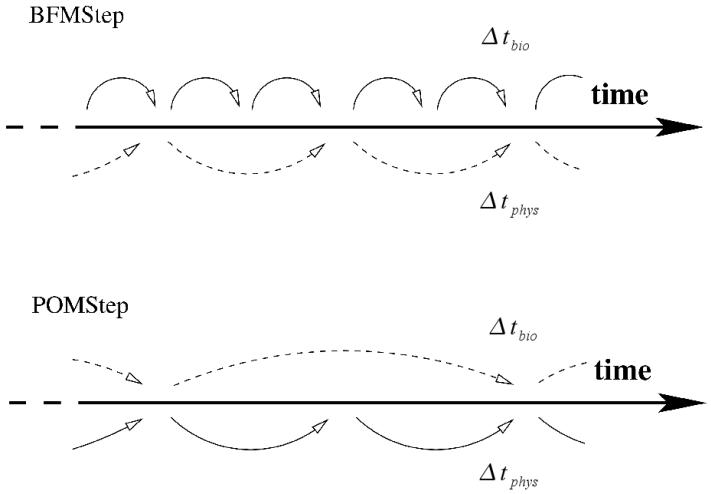


Figure 1.12: BFMStep and POMStep method. Full lines indicate the global integration and dashed lines the process estimation.

Time step imposed by biogeochemistry (BFMStep)

In this case, the transport rates $\frac{\partial \mathbf{c}}{\partial t}\big|_{phys}$ are assumed to be the slower evolving process rates $\tilde{\mathbf{S}}_{slow}$ and equation 1.1 takes the form:

$$\mathbf{c}_{n+1} = \mathbf{c}_n + \int_{\Delta t_{glob}} \frac{\partial \mathbf{c}}{\partial t}\bigg|_{bio} dt + \widetilde{\frac{\partial \mathbf{c}}{\partial t}\bigg|_{phys}} \Delta t_{glob} \,. \tag{1.45}$$

However, the transport model by construction updates the state variable, rather than computing rates and thus, the transport rates have to be estimated from a sub-integration considering the transport process only. This generates a splitting error. As the respective sub-model is based on the leap frog algorithm of *POM* (eq. (1.32)), the estimation term derived in section 1.1.1 is used:

$$\widetilde{\frac{\partial \mathbf{c}}{\partial t}\bigg|_{phys}} = \frac{\mathbf{\Phi}_{n+1} - \mathbf{c}_{n-1}}{2\Delta t} \,. \tag{1.46}$$

The integration scheme for the global integration step can be chosen independently, so that the four integration schemes described in section

1.1.2 were applied. As it was decided to keep the time step for the physical processes fixed at $\Delta t = 864s$, the BFMStep mode applies to global time steps of $\Delta t \leqslant 864s$.

Time step imposed by physics (POMStep)

In the opposite case, where the time step is dictated by the physical processes, biogeochemical processes are assumed to be the slower evolving processes and the global integration step is expressed as:

$$\mathbf{c}_{n+1} = \mathbf{c}_n + \int\limits_{\Delta t_{glob}} \left.\frac{\partial \mathbf{c}}{\partial t}\right|_{phys} dt + \left.\widetilde{\frac{\partial \mathbf{c}}{\partial t}}\right|_{bio} \Delta t_{glob} \; . \tag{1.47}$$

Here, the changing rates are computed directly by the biogeochemical sub-model and the splitting error is reduced to the piecewise constant treatment on the coarser estimation interval.[3] To reduce this error further, a modification to the original Source Splitting scheme is introduced by estimating the biogeochemical rates on the intermediate steps by linear interpolation of the two preceding rates:

$$\left.\widetilde{\frac{\partial \mathbf{c}\,(t_{n+s})}{\partial t}}\right|_{bio} = \left.\frac{\partial \mathbf{c}\,(t_n)}{\partial t}\right|_{bio} + \frac{\left.\frac{\partial \mathbf{c}(t_n)}{\partial t}\right|_{bio} - \left.\frac{\partial \mathbf{c}(t_n - \Delta t)}{\partial t}\right|_{bio}}{\Delta t_{est}} \left(t_{n+s} - t_n\right) \; , \tag{1.48}$$

where Δt_{est} is the estimation interval of the biogeochemical processes.

The scheme for the global integration in this case is limited to the scheme used in *POM* (eq. (1.32)), as a modification of the integration scheme would interfere with the scheme of spatial discretisation.

With the time step imposed at the fixed step size of $\Delta t_{glob} = 864s$, the estimation time step may vary as $\Delta t_{est} \geqslant 864s$.

[3] Note, that when the two processes groups are evaluated with the same time resolution (with a time step of $\Delta t_{glob} = \Delta t_{est} = 864s$), this formulation (BFMStep) does not involve any splitting error, as the leap-frog scheme of *POM* evaluates the Source terms at the actual time level only. Thus, in this case the error is restricted to the error of the integration scheme only.

1.4 Results and validation

In table 1.4 an overview is given over the experiments presented below, conducted to evaluate the coupling schemes implemented. All experiments were carried out in all 3 locations to include the particularities of the various local ecosystems. The discussions below focus on the results for the two variables chlorophyll-a and phosphate as the key variables of the involved ecosystem dynamics. A broader overview of the various elements of the trophic web is given in appendix B. All solutions are given for the 5th year of integration.

Experiment	Integration schemes				Applied time steps
BFMStep-Ref	LF	AB2	RK2	E	$\Delta t_{bio,glob} = \Delta t_{phys}$
BFMStep-1	LF	AB2	RK2	E	$\Delta t_{bio,glob} = \Delta t_{phys}$
BFMStep-2	LF	AB2	RK2	E	$\Delta t_{bio,glob} < \Delta t_{phys}$
Experiment	Rate estimation				Applied time steps
POMStep-Ref	(con=int)				$\Delta t_{phys,glob} = \Delta t_{bio}$
POMStep-1	(con=int)				$\Delta t_{phys,glob} = \Delta t_{bio}$
POMStep-2	con		int		$\Delta t_{phys,glob} < \Delta t_{bio}$

Table 1.4: Matrix of experiments conducted with the BFMStep and the POMStep scheme. The used abbreviations are: LF - *leap frog*, AB2 - *Adams-Bashforth 2nd order*, RK2 - *Runge-Kutta 2nd order*, E - *Euler;* Δt_{glob} - global time step, Δt_{bio} - time step for biogeochemical processes, Δt_{phys} - time step for physical processes; con - constant rate estimation, int - interpolated rate estimation.

Experiments BFMStep-Ref and POMStep-Ref are the reference experiments described below. In experiments BFMStep the solution of the system is achieved according to eq. (1.45), i. e. the biogeochemical rates are assumed to be the fastest rates, therefore the time step applied to the biogeochemical processes is (Δt_{bio}) is assumed to be the global time step Δt_{glob} for the solution of the whole system and has been kept equal (experiment BFMStep-1) or smaller (BFMStep-2) than the time step applied to the computation of the physical processes (Δt_{phys}). Each BFMStep ex-

periment has been replicated four times by utilising the four integration schemes described in section 1.1.2.

The experiments POMStep solve the system by applying eq. (1.47), where the biogeochemical rates are assumed to be the slowest and Δt_{phys} is assumed to coincide with Δt_{glob}, while Δt_{bio} has been kept equal (POMStep-1) or longer than Δt_{glob} (POMStep-2).

As stated in the previous section, in this case the numerical integration scheme for the global integration is limited to the use of the physical model scheme (which for *POM* is the semi-implicit leap-frog scheme). However, POMStep-2 allows for two possible ways to estimate $\left.\frac{\partial \mathbf{c}(t_n)}{\partial t}\right|_{bio}$ by holding the rate constant or by linearly interpolating it using eq. (1.48).

1.4.1 Reference solution

For an objective comparison of the various experiments some reference solution is needed. In principle, this could be provided solving analytically the system equations or by data. An analytical solution is not known, given the nature of the equation forming the coupled system, while reference from data has some other obstacles. The eventual errors found in the simulation with respect to the data might have numerous sources, from measurement errors to modelling errors, that risk to cover the differences in the numerical schemes completely. In addition to that, the climatological forcing does not necessarily lead to the climatological mean as true solution, as the involved processes are non-linear. Moreover, availability of data is rather scarce, so that reliable climatological means are averaged over large periods in time and differences in the numerical schemes might be smoothed out before the actual comparison can be done.

The reference solution was therefore obtained from the model itself choosing $\Delta t_{glob} = \Delta t_{phys} = \Delta t_{bio}$ appropriately small, such that physical and biogeochemical processes are accurately resolved and the differences among coupling methods, as well as among integration schemes can be

Experiment	Integration scheme	Δt_{glob}
BFMStep-Ref-E	Euler	3 s
BFMSTEP-Ref-LF	Leap-frog	3 s
BFMSTEP-Ref-AB2	Adams-Bashforth	3 s
BFMSTEP-Ref-RK2	Runge-Kutta	3 s
POMStep-Ref	POM	3 s

Table 1.5: The high resolution experiments.
Δt_{glob}: time step of global integration, Δt estimation interval.

considered negligible. The Δt adopted was $3s$ and reference experiments were run adopting the numerical schemes listed in table 1.5. In order to provide an assessment of the model performance in replicating the observed seasonal cycle of hydrological and biogeochemical characteristics a comparison of the reference run model results with the observational data extracted from the ABCD data set is offered in appendix A.

Figures 1.13 to 1.18 compare the experiments BFMStep-Ref-LF and POMStep-Ref performed at a time step of $\Delta t = 3s$. The simulated state variables shown are chlorophyll-a (fig.1.13, 1.15, 1.17) and phosphate (fig. 1.14, 1.16, 1.18) concentrations. Differences are shown in percentage with respect to the POMStep-Ref field. They are generally of the order of 10^{-4} in all sites, either for chlorophyll-a either for the phosphates, as well as for the other state variables (not shown). In general the BFMStep mode seems to give a higher chlorophyll-a level at the surface in winter, when the column is well mixed and a higher level of phosphate in the bottom layer during stratification periods from summer to autumn for S1). However, differences are sufficiently small to be considered negligible.

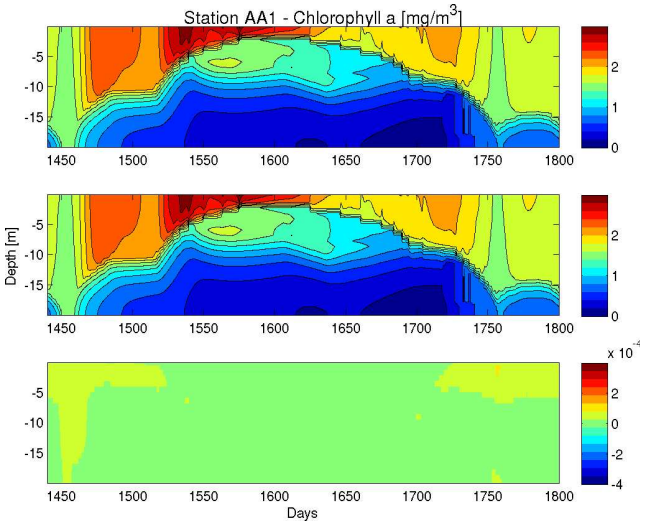


Figure 1.13: Reference solutions of the two coupling schemes for chlorophyll-a content $\left[\frac{mg}{m^3}\right]$ at site AA1. *Top:* BFMStep-Ref-LF, *middle:* POMStep-Ref, *bottom:* Fractional differences between BFMStep-Ref-LF and POMStep-Ref, using POMStep-Ref as normalisation field.

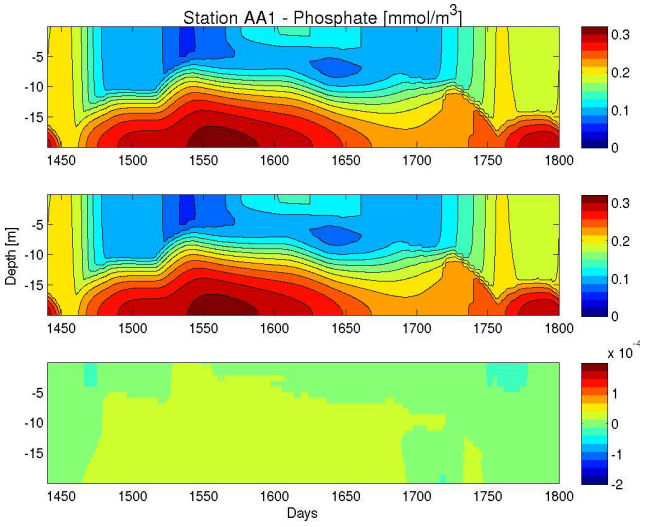


Figure 1.14: As fig. 1.13, but for phosphate $\left[\frac{mmol}{m^3}\right]$.

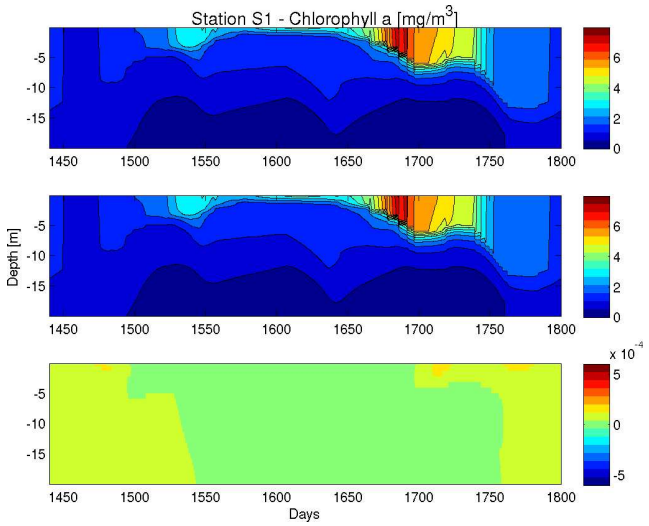


Figure 1.15: Reference solutions of the two coupling schemes for chlorophyll-a content $\left[\frac{mg}{m^3}\right]$ at site S1. *Top:* BFMStep-Ref-LF, *middle:* POMStep-Ref, *bottom:* Fractional differences between BFMStep-Ref-LF and POMStep-Ref, using POMStep-Ref as normalisation field.

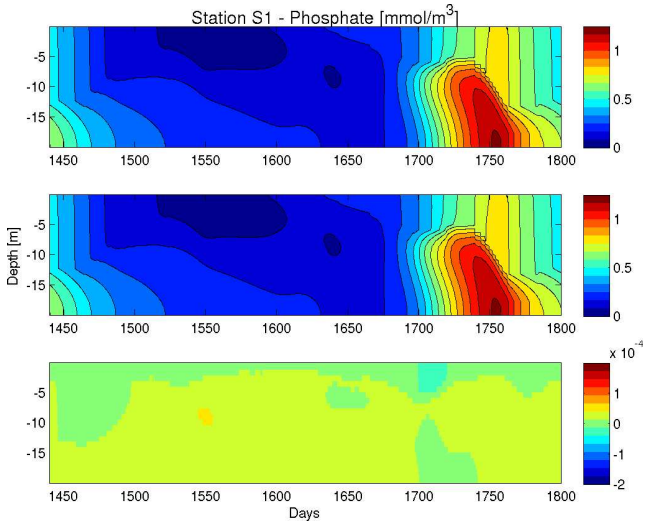


Figure 1.16: As fig. 1.15, but for phosphate $\left[\frac{mmol}{m^3}\right]$.

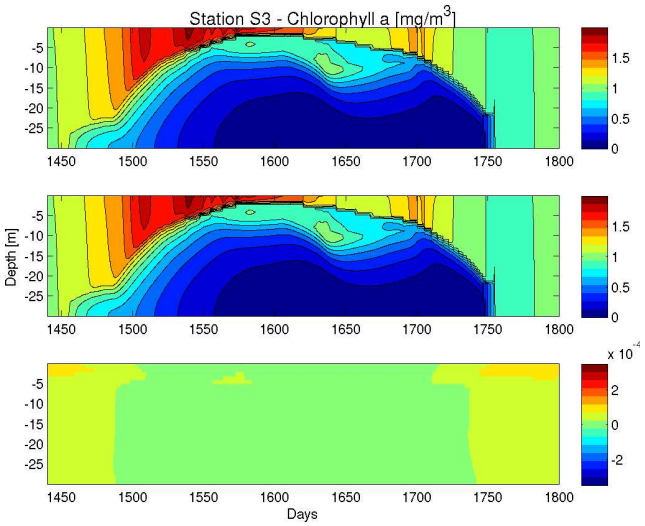


Figure 1.17: Reference solutions of the two coupling schemes for chlorophyll-a content $\left[\frac{mg}{m^3}\right]$ at site S3. *Top:* BFMStep-Ref-LF, *middle:* POMStep-Ref, *bottom:* Fractional differences between BFMStep-Ref-LF and POMStep-Ref, using POMStep-Ref as normalisation field.

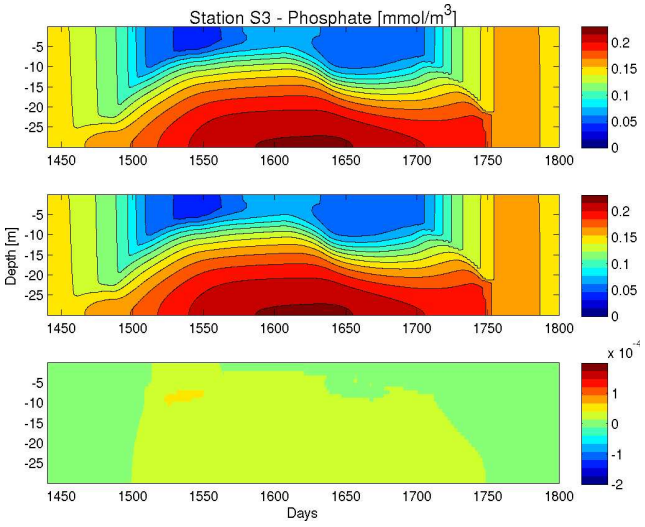


Figure 1.18: As fig. 1.17, but for phosphate $\left[\frac{mmol}{m^3}\right]$.

For the BFMStep scheme a further analysis was carried out with respect to the integration scheme used in the global integration. Results for the same state variables shown before can be seen in fig. 1.19 to 1.24. It can be seen that the differences in this case are even smaller and rather localised. Their position is associated to the strong gradients connected to the fast deepening and to the breaking up of the pycnocline by turbulent mixing.

It results, that the variations of the solutions with respect to both coupling and integration schemes at this step size are of the order 10^{-4} or smaller in all cases and thus, the fields can be considered close enough to the true solution of the model equations to provide a reference solution. Further on in this work experiment BFMStep-Ref-LF will be referred to as reference case.

In appendix A a comparison of the results obtained from the reference solution with seasonally averaged observations is given in order to evaluate the quality of the simulation.

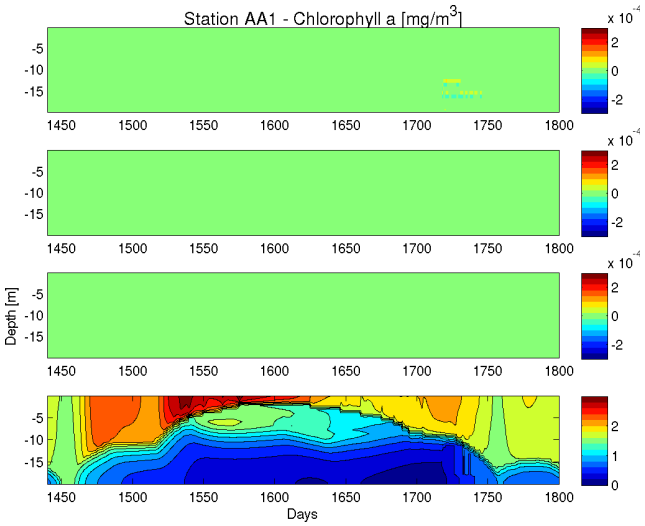


Figure 1.19: Reference solutions of the four integration schemes for chlorophyll-a content $\left[\frac{mg}{m^3}\right]$ at site AA1. *From bottom to top:* BFMStep-Ref-RK2, differences of BFMStep-Ref-E, BFMStep-Ref-AB2, BFM-Step-Ref-LF with respect to BFM-Step-RK2 normalised by the BFM-Step-RK2.

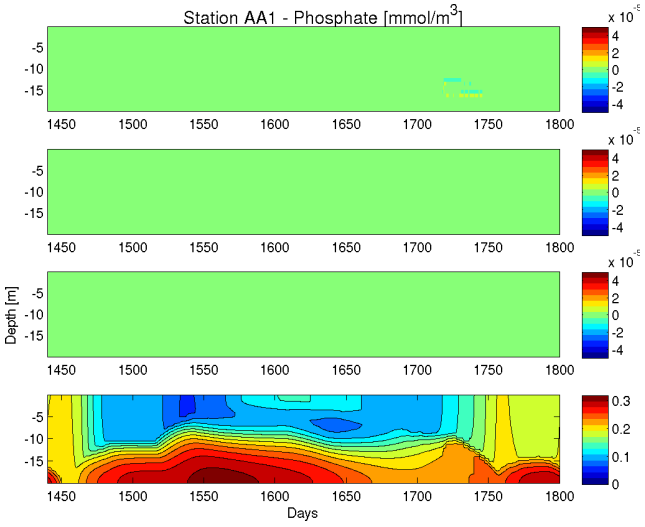


Figure 1.20: As fig. 1.19, but for phosphate $\left[\frac{mmol}{m^3}\right]$.

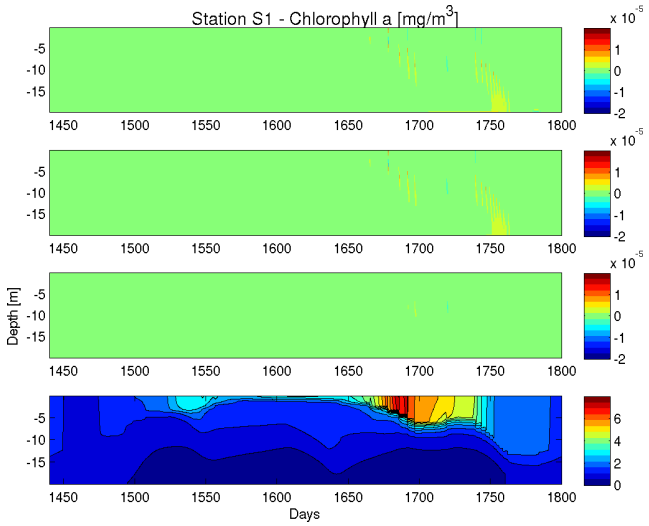


Figure 1.21: Reference solutions of the two four integration chlorophyll-a content $\left[\frac{mg}{m^3}\right]$ at site S1. *From bottom to top:* BFMStep-Ref-RK2, differences of BFMStep-Ref-E, BFMStep-Ref-AB2, BFM-Step-Ref-LF with respect to BFM-Step-RK2 normalised by the BFM-Step-RK2.

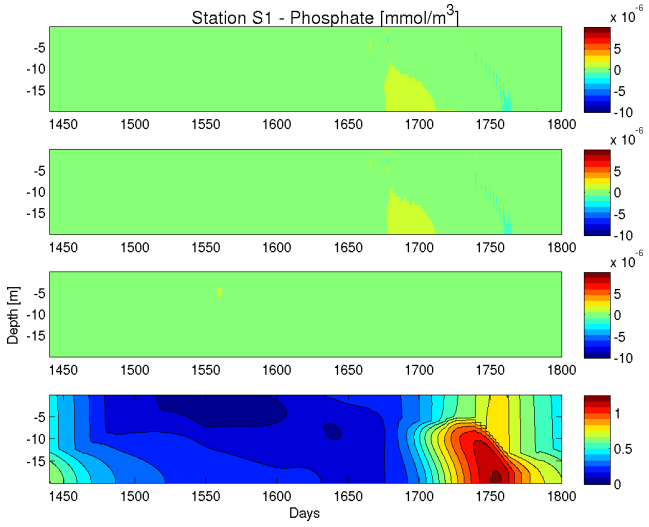


Figure 1.22: As fig. 1.21, but for phosphate $\left[\frac{mmol}{m^3}\right]$.

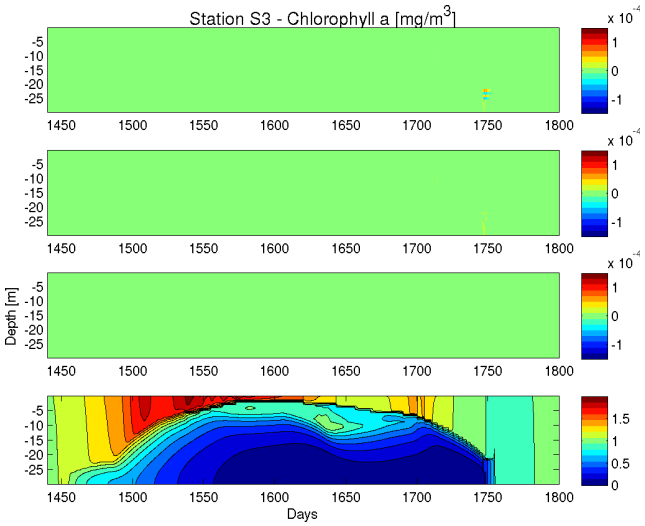


Figure 1.23: Reference solutions of the two four integration for chlorophyll-a content $\left[\frac{mg}{m^3}\right]$ at site S3. *From bottom to top:* BFMStep-Ref-RK2, differences of BFMStep-Ref-E, BFMStep-Ref-AB2, BFM-Step-Ref-LF with respect to BFM-Step-RK2 normalised by the BFM-Step-RK2.

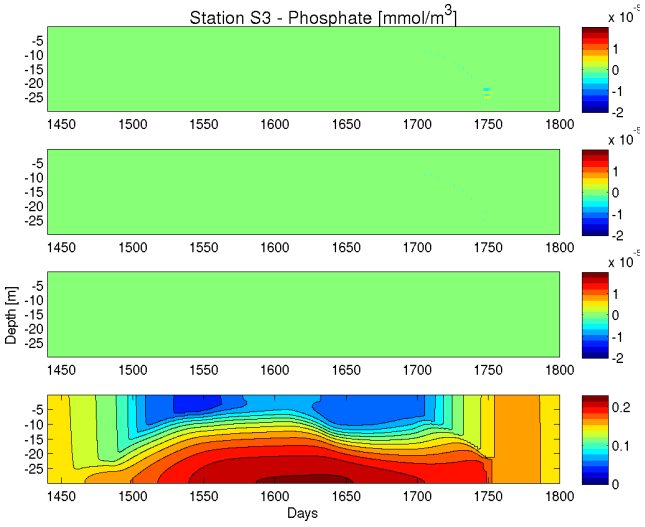


Figure 1.24: As fig. 1.23, but for phosphate $\left[\frac{mmol}{m^3}\right]$.

1.4.2 BFMStep method: sensitivity to Δt_{glob} and to integration schemes

Before proceeding with the comparison of the coupling method, a BFMStep scheme to be compared with the POMStep scheme has to be established. The test case presented in section 1.1.2 indicated a significant sensitivity of biogeochemical processes to the integration scheme applied, so that the sensitivity of the model with respect to the scheme used in the global integration of the BFMStep scheme is investigated here.

To accomplish this, the four experiments BFMStep-1 of table 1.4 were carried out using the four implemented integration schemes at the standard time step of $864s$ in global integration and physical rate estimation (see table 1.6).

The respective errors of the four experiment runs are shown in figures 1.25 to 1.30 (for chlorophyll-a and phosphate) with respect to the reference solution defined in section 1.4.1.

Experiment	**Integration scheme**	Δt_{glob}	Δt_{est}
BFMStep-1-E	Euler	864 s	864 s
BFMStep-1-LF	Leap-frog	864 s	864 s
BFMStep-1-AB2	Adams-Bashforth	864 s	864 s
BFMStep-1-RK2	Runge-Kutta	864 s	864 s

Table 1.6: BFMStep experiments using different integration schemes.

The four integration schemes yielded very similar simulated fields indicating, that the range $\Delta t_{bio} \leqslant \Delta t_{phys}$ can be considered irrelevant with respect to order. Moreover, the experiments revealed notable errors for all schemes. In particular during late autumn and winter the simulations yield levels of phytoplankton biomass (fig. 1.25, 1.27 and 1.29 at the surface that are significantly high for all schemes in all three sites. Also the growth in phosphate concentration in the bottom layer in summer is consistently overestimated (fig. 1.26, 1.28, 1.29).

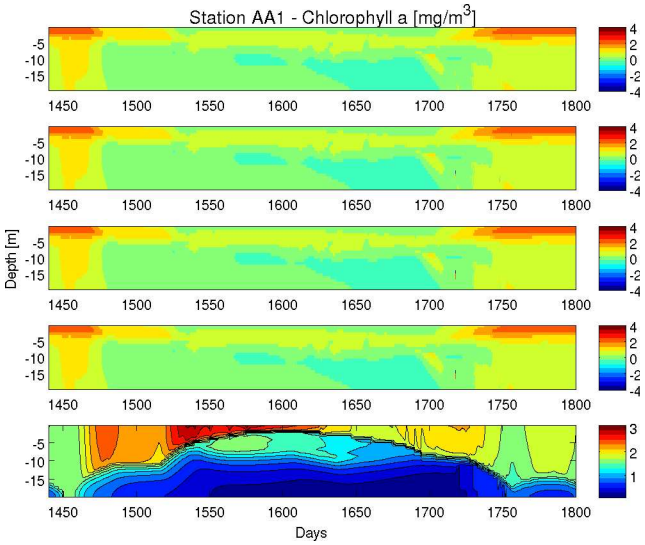


Figure 1.25: Chlorophyll-a content $\left[\frac{mg}{m^3}\right]$ at site AA1 for different integration schemes. *From bottom to top:* Reference field, error in % of BFM-Step-1-RK2, BFM-Step-1-E, BFM-Step-1-AB2, BFM-Step-1-LF.

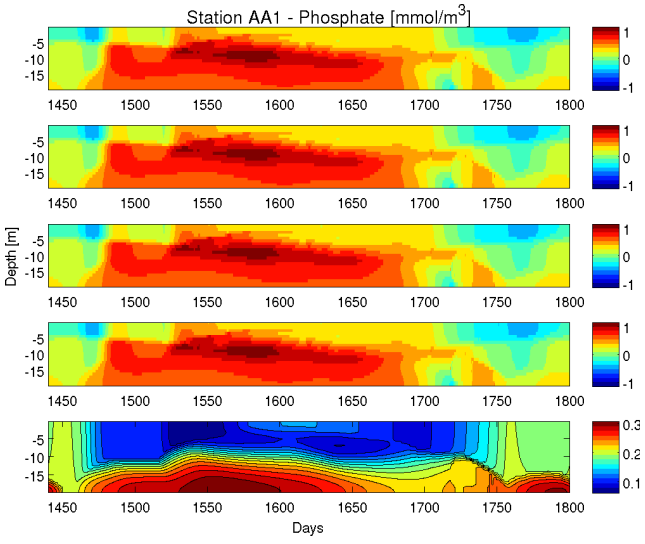


Figure 1.26: As fig. 1.25, but for phosphate $\left[\frac{mmol}{m^3}\right]$.

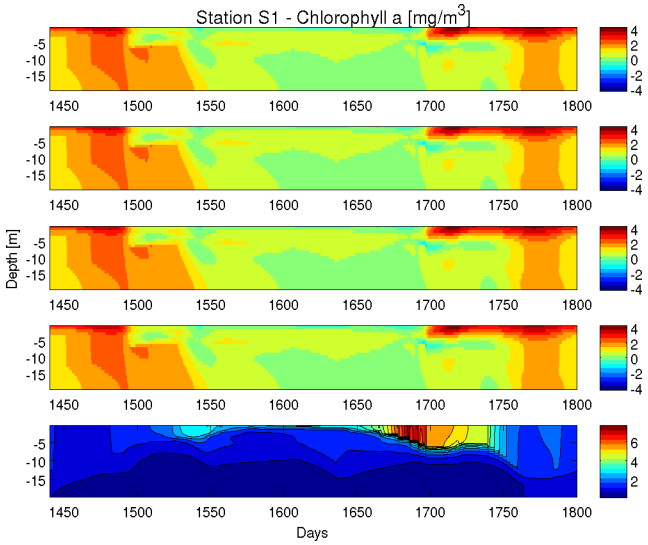


Figure 1.27: Chlorophyll-a content $\left[\frac{mg}{m^3}\right]$ at site S1 for different integration schemes. *From bottom to top:* Reference field, error in % of BFM-Step-1-RK2, BFM-Step-1-E, BFM-Step-1-AB2, BFM-Step-1-LF.

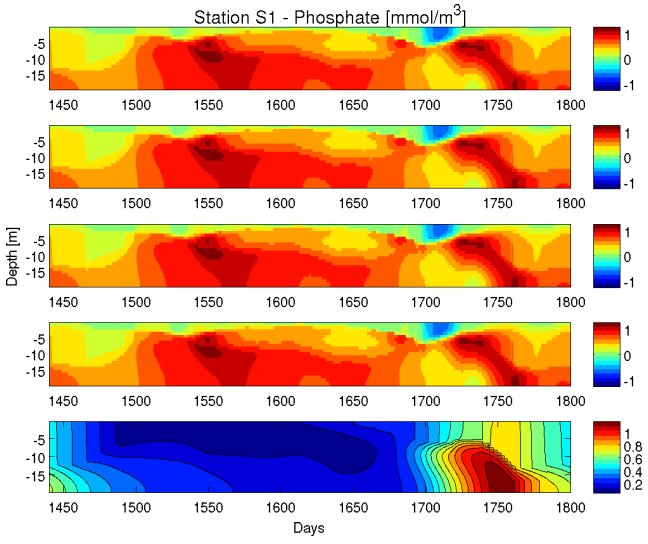


Figure 1.28: As fig. 1.27, but for phosphate $\left[\frac{mmol}{m^3}\right]$.

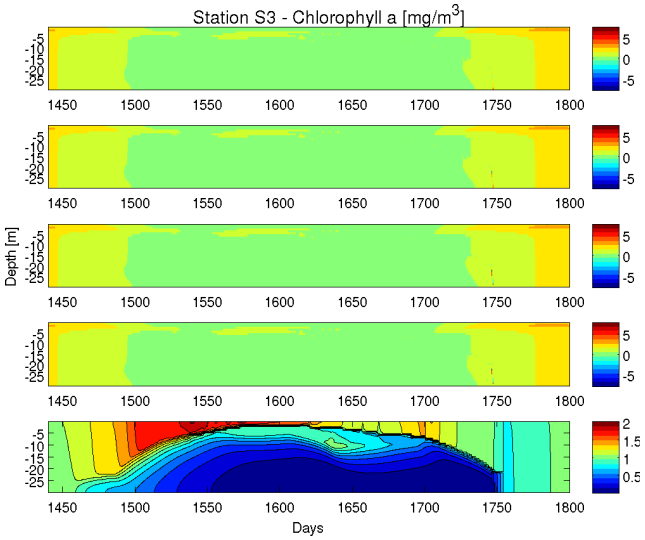


Figure 1.29: Chlorophyll-a content $\left[\frac{mg}{m^3}\right]$ at site S3 for different integration schemes. *From bottom to top:* Reference field, error in % of BFM-Step-1-RK2, BFM-Step-1-E, BFM-Step-1-AB2, BFM-Step-1-LF.

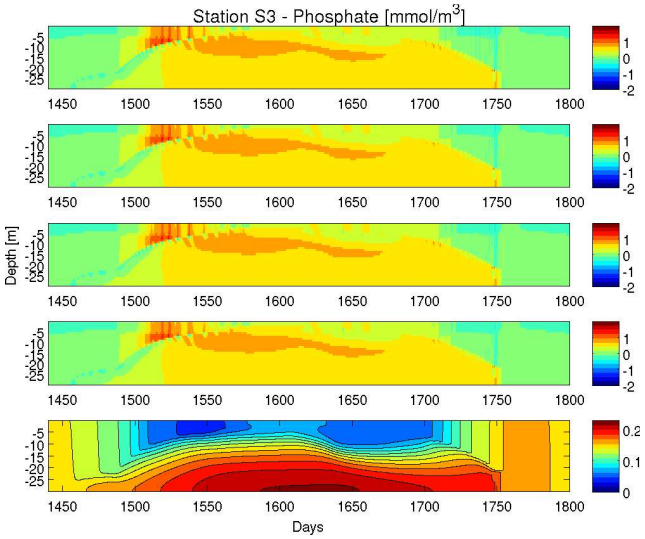


Figure 1.30: As fig. 1.29, but for phosphate $\left[\frac{mmol}{m^3}\right]$.

At this point, a too low time resolution in the biogeochemical processes might be suspected as a cause, so that a series of experiments (BFMStep-2) was carried out to test the role of the time step applied to the biogeochemical rates. The simulations in table 1.7 were carried out at a decreased time step of $\Delta t_{bio} = \Delta t_{glob} = 27s$ for the global integration applied to the biogeochemical rates, while the standard time step of $\Delta t_{phys} = 864s$ for the physical rate estimation was left unchanged. Their errors are compared to the errors of the BFMStep-1 experiments.

Experiment	**Integration scheme**				Δt_{glob}	Δt_{est}
BFMStep-1	LF	AB2	RK2	E	864 s	864s
BFMStep-2	LF	AB2	RK2	E	27 s	864s

Table 1.7: BFMStep experiments on time step sensitivity.

The comparison for the leap-frog scheme and for the chlorophyll-a and phosphate fields is shown in figures1.31 to 1.36. (The description of the results obtained with the other integration schemes is omitted from this point on, as the previous considerations revealed, that the model is not sensitive to the integration scheme applied to the biogeochemical processes in the range $\Delta t_{bio} \leqslant \Delta t_{phys}$.)

The plots show, that the reduction in time step did not give an improvement of the solution. The structure of the errors already evident in the standard resolution persists, their magnitude remains unchanged, so that virtually no difference is visible between the BFMStep-1 and BFMStep-2 errors. Hence, it can be concluded, that the errors are not caused by an insufficient resolution of the biogeochemical processes and therefore, the time step applied to the biogeochemical processes can be kept at the same level as the one for the physical processes, so that we can proceed to the comparison of the actual coupling schemes.

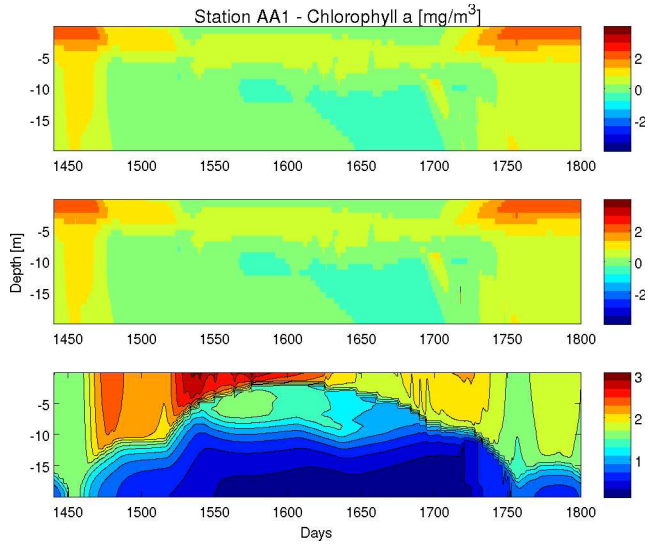


Figure 1.31: Chlorophyll-a content $\left[\frac{mg}{m^3}\right]$ at site AA1 - influence of time step in BFMStep scheme. *Top:* BFMStep-2-LF ($\Delta t_{bio} = 27s$) - error in %; *middle:* BFMStep-1-LF ($\Delta t_{bio} = 864s$) - error in %; *bottom:* reference case.

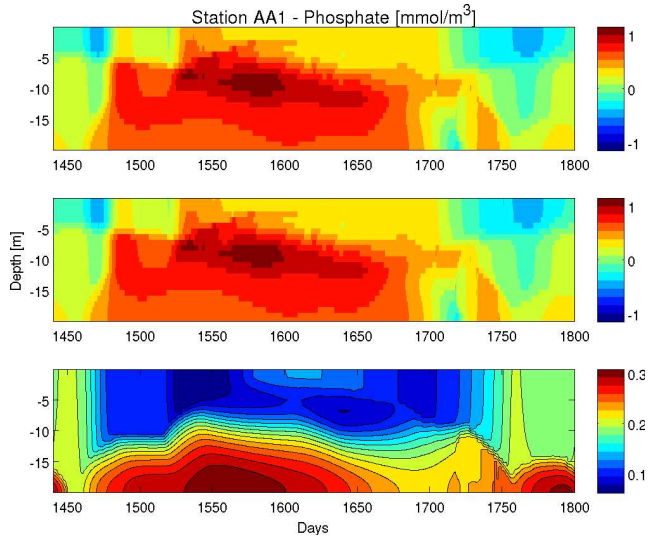


Figure 1.32: As fig. 1.31, but for phosphate $\left[\frac{mmol}{m^3}\right]$.

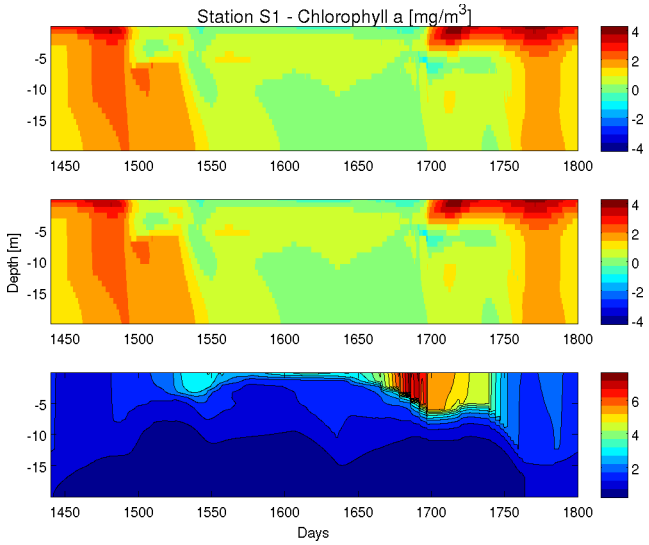


Figure 1.33: Chlorophyll-a content $\left[\frac{mg}{m^3}\right]$ at site S1 - influence of time step in BFMStep scheme. *Top:* BFMStep-2-LF ($\Delta t_{bio} = 27s$) - error in %; *middle:* BFMStep-1-LF ($\Delta t_{bio} = 864s$) - error in %; *bottom:* reference case.

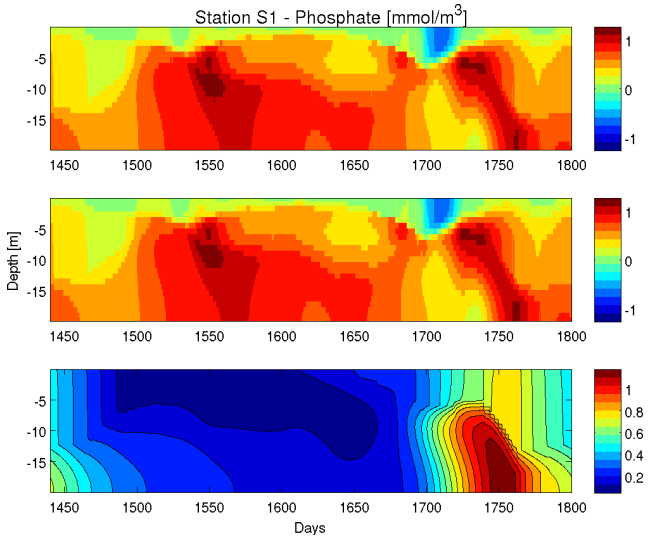


Figure 1.34: As fig. 1.33, but for phosphate $\left[\frac{mmol}{m^3}\right]$.

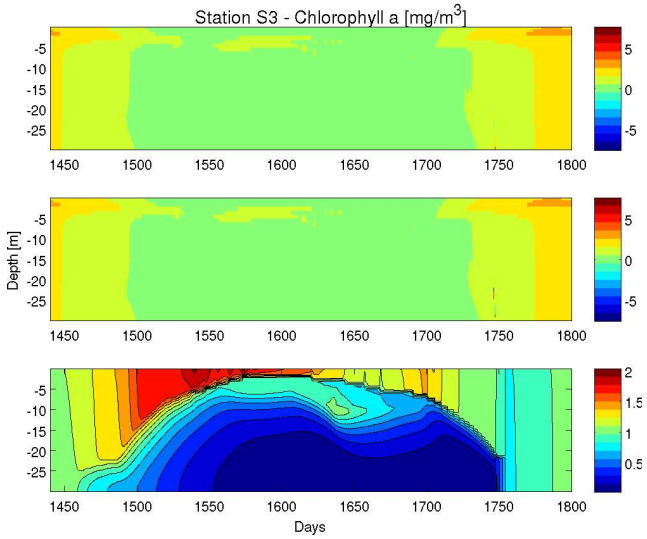


Figure 1.35: Chlorophyll-a content $\left[\frac{mg}{m^3}\right]$ at site S3 - influence of time step in BFMStep scheme. *Top:* BFMStep-2-LF ($\Delta t_{bio} = 27s$) - error in %; *middle:* BFMStep-1-LF ($\Delta t_{bio} = 864s$) - error in %; *bottom:* reference case.

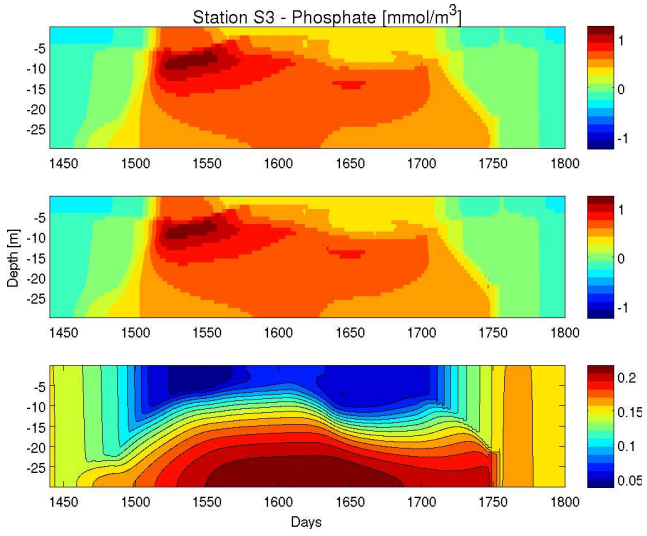


Figure 1.36: As fig. 1.35, but for phosphate $\left[\frac{mmol}{m^3}\right]$.

1.4.3 Coupling method sensitivity: BFMStep vs. POMStep.

To obtain a consistent comparison of the two coupling methods described above, experiments were performed with global time step and estimation interval set synchronously to the same time resolution for both schemes (BFMStep and POMStep). In the previous section it was shown for the BFMStep scheme, that the solution of the model does not show a dependence on the resolution of the biogeochemical processes in the given range of time steps, so that the standard time step of $864s$ was applied to physical and biogeochemical processes. (see table 1.8).

Experiment	**Coupling Scheme**	Δt_{glob}	Δt_{est}
BFMStep-1	BFMStep	864 s	864 s
POMStep-1	POMStep	864 s	864 s

Table 1.8: Experiments considered in the coupling scheme comparison.

The results of these experiments are shown in figures 1.37 to 1.42. Rather different solutions are obtained for the two coupling methods at all the three implementation sites. Generally speaking BFMStep results show (with respect to POMStep) an overestimation of the surface phytoplankton biomass and of the phosphate concentration in the mid and lower part of the water column. More precisely, these significant errors are persisting in all three sites for several month in both variables, but appear in different periods and locations of the water column. For chlorophyll-a the error appears in autumn and is particularly high in the first five meters of the water column, although it is mixed downwards in weaker form through the not stratified water column and lasts until mid-winter. In all three sites the highest errors are found at the surface at around 4%. These errors are completely removed in the POMStep solution.

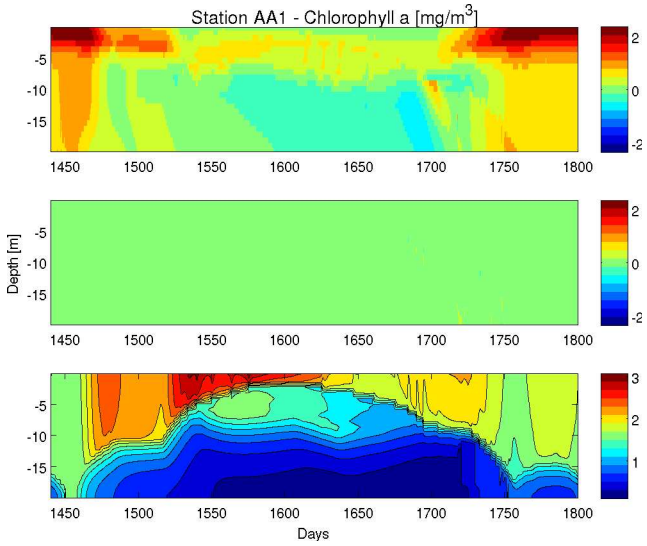


Figure 1.37: Chlorophyll-a content $\left[\frac{mg}{m^3}\right]$ at site AA1 - comparison between coupling methods. *Top:* BFMStep-1-LF error, in %; *Middle:* POMStep-1 error in %; *Bottom:* Reference solution.

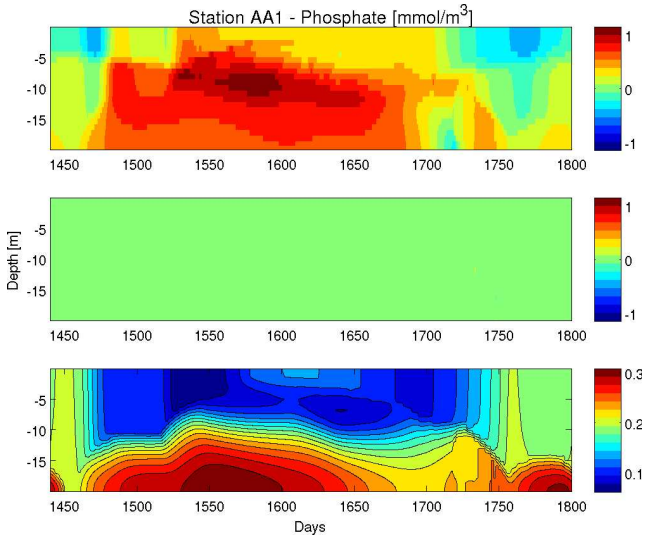


Figure 1.38: As fig. 1.37, but for phosphate $\left[\frac{mmol}{m^3}\right]$.

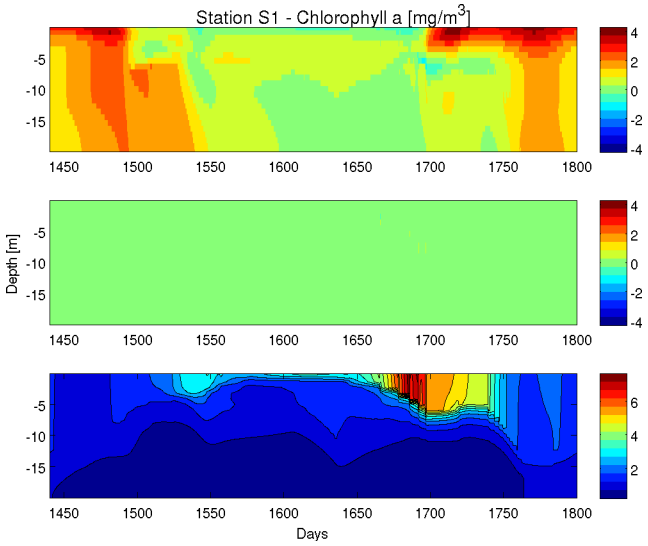


Figure 1.39: Chlorophyll-a content $\left[\frac{mg}{m^3}\right]$ at site S1 - comparison between coupling methods. *Top:* BFMStep-1-LF error, in %; *Middle:* POMStep-1 error in %; *Bottom:* Reference solution.

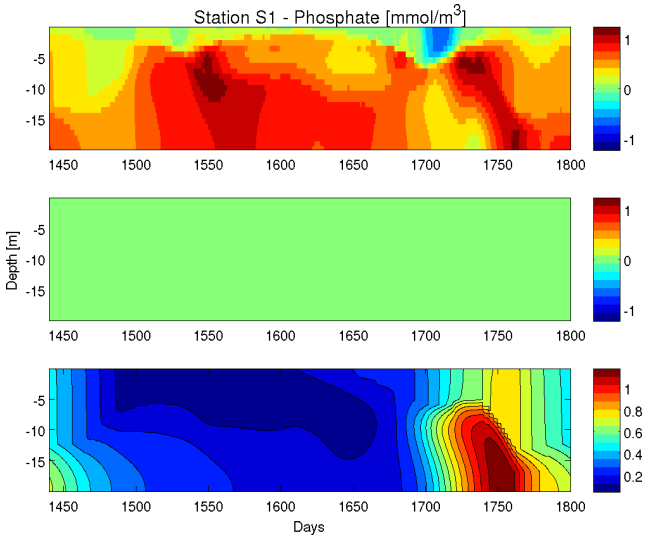


Figure 1.40: As fig. 1.39, but for phosphate $\left[\frac{mmol}{m^3}\right]$.

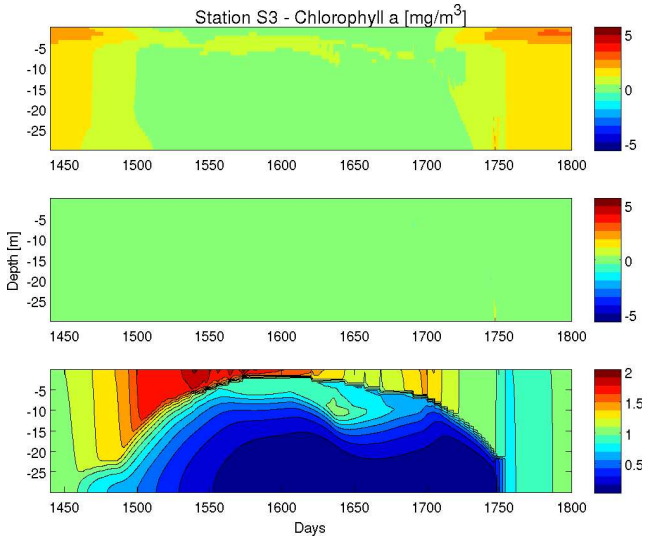


Figure 1.41: Chlorophyll-a content $\left[\frac{mg}{m^3}\right]$ at site S3 - comparison between coupling methods. *Top:* BFMStep-1-LF error, in %; *Middle:* POMStep-1 error in %; *Bottom:* Reference solution.

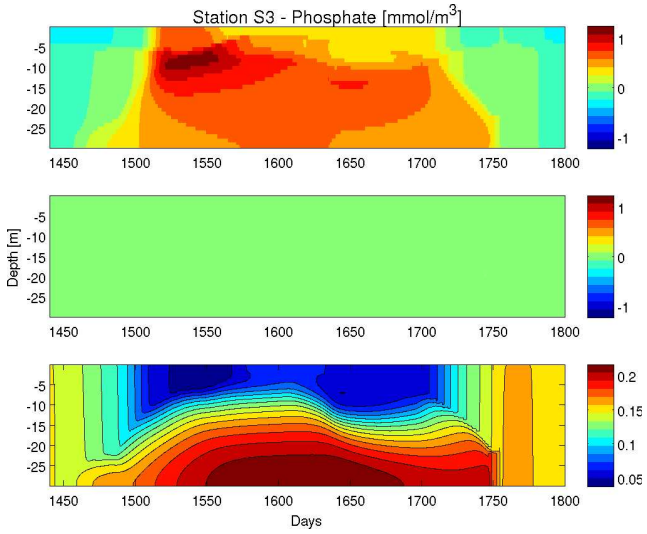


Figure 1.42: As fig. 1.41, but for phosphate $\left[\frac{mmol}{m^3}\right]$.

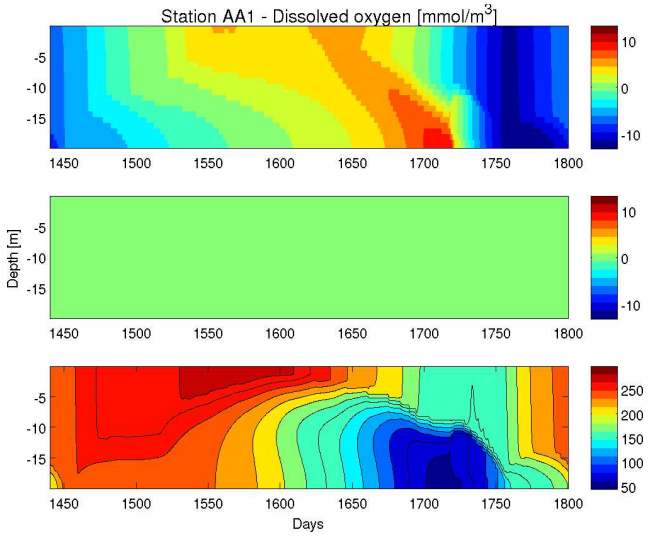


Figure 1.43: Dissolved oxygen $\frac{mmol}{m^3}$ at the site AA1 - comparison between coupling methods.*Top:*BFMStep-1-LF error , in %; middle: POMStep-1 error in %; *bottom:* reference solution.

This applies similarly to most of the other state variables with the exception of nitrate and dissolved oxygen for the BFMStep experiments: during winter in all three sites, the oxygen content (fig. 1.43,1.44; S1 results not shown, as very similar to AA1) of the mixed water column is significantly underestimated (by 5% for site S3 and by 10% for sites AA1 and S1). During summer, when the water column is stratified, the oxygen content is overestimated and all three sites show a maximum error of around 5% in late summer located in the surface layer in site S3, while in the bottom layer for sites S1 and AA1. .

Nitrate concentration in sites AA1 and S1 (shows errors reaching up to 50% in late autumn coincident with nitrate depletion, so that the solution appears completely unreliable with respect to nitrate concentration with the BFMStep method (again only site AA1 is shown here in fig. 1.45 as the error field is very similar to site S1). This effect is absent in station S3 (fig. 1.46), where on the contrary nitrates are overestimated by up to 4-5% in the surface layer in spring.

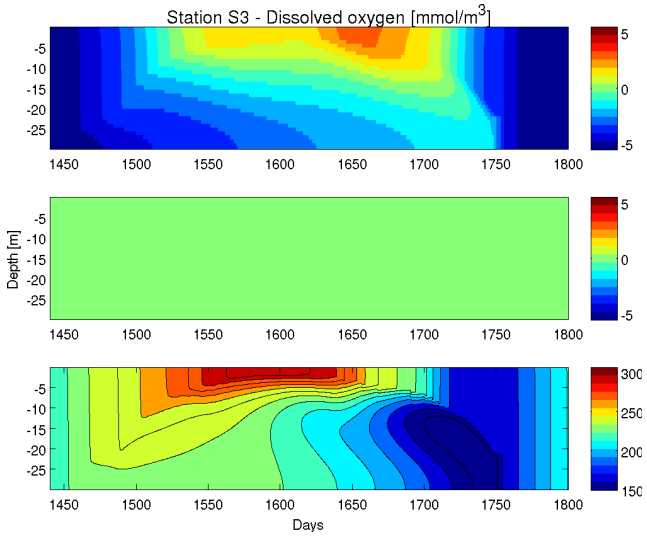


Figure 1.44: As fig. 1.43, but for site S3.

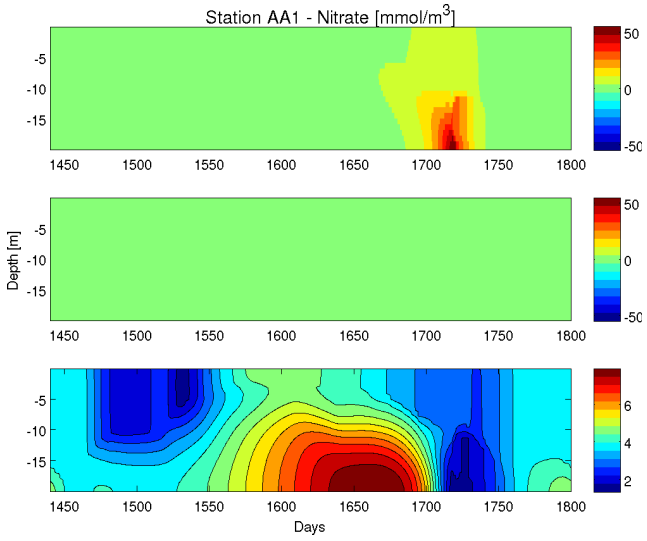


Figure 1.45: Nitrate $\frac{mmol}{m^3}$ at the site AA1 - comparison between coupling methods.*Top:*BFMStep-1-LF error , in %; middle: POMStep-1 error in %; *bottom:* reference solution.

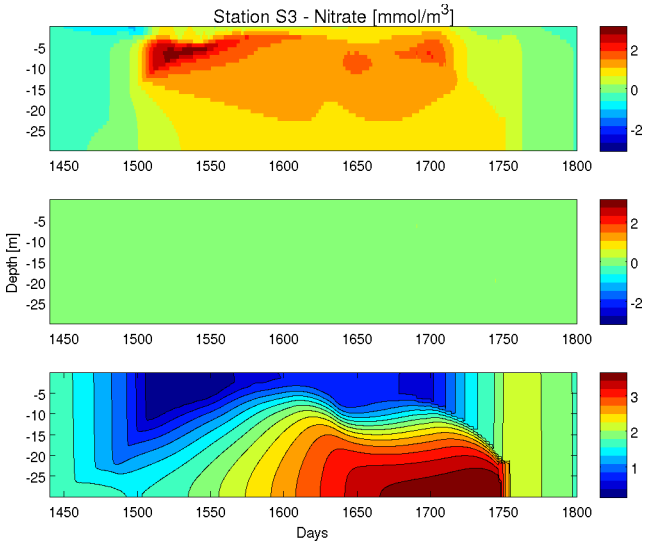


Figure 1.46: As fig. 1.45, but for site S3.

The overall quality of the solution is clearly superior for the POM-Step scheme and this applies to all state variables as demonstrated by the POM-Step results for selected states variables shown in appendix B . The absence of the significant error structures in this scheme (errors are < 1%), in contrast to their persistence over all the experiments BFMStep-1 and BFMStep-2 implies, that they are caused by the splitting of the physical transport processes from the global integration, so that it can be concluded that the leading error term in the simulation is the splitting error connected to the sub-integration of the physics.

1.4.4 The POMStep-2 experiments

The consistency of the various experiments BFM-Step (section 1.4.2), with respect to integration scheme and time resolution of the biogeochemical processes, hint, that biogeochemical processes are sufficiently resolved at the time step of $864s$ and that it is possible to perform experiments with the POMStep scheme using an estimation interval for the biogeochemical processes, that is higher, than the integration step imposed by the transport rate.

Experiment	**Coupling Scheme**	**Integration scheme**	Δt_{glob}	Δt_{est}
POMStep-2-con	POMStep	POM, constant	864 s	6912 s
POMStep-2-int	POMStep	POM, interpolated	864 s	6912 s

Table 1.9: Experiments on the POMStep estimation scheme.

In the experiments POMStep-2 (see table1.9) the estimation interval is increased up to eight times the integration step of the global integration and the two implemented estimation strategies described in section 1.3.8 are validated.

The results in fig. 1.47 to 1.52 show a much better behaviour for the interpolated estimation (POMStep-2-int). Errors are mostly generated by the dynamics of the pycnocline, that is slightly shifted with respect to the reference case and spread by turbulence to the surface and bottom layer being much higher in magnitude for the constant rate estimation (POMStep-con). The latter also seems to generate smaller but persisting errors in the bottom layer when stratification starts to build up.

However, in both cases the low resolution of biogeochemical processes of 8 times the size of the standard step creates only limited errors of less than 1% with the only exception of chlorophyll in site S3, that shows a isolated error of 5% connected to the retardation of the breaking up of the pycnocline, that should be seen rather as a small shift in the dynamics than a high local error.

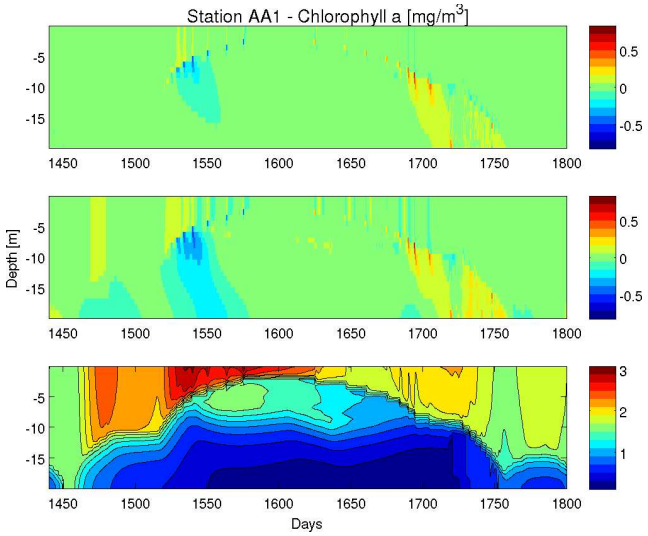


Figure 1.47: Chlorophyll-a content $\left[\frac{mg}{m^3}\right]$ at site AA1 - Differences of the two versions of rate estimation for the POMStep scheme. *Top:* POMStep-2-int error in %; *Middle:* POMStep-2-con, relative differences to POMStep-1; *Bottom:* POMStep-1.

Overall the solutions can be considered as reasonably accurate with respect to the general precision of the model.

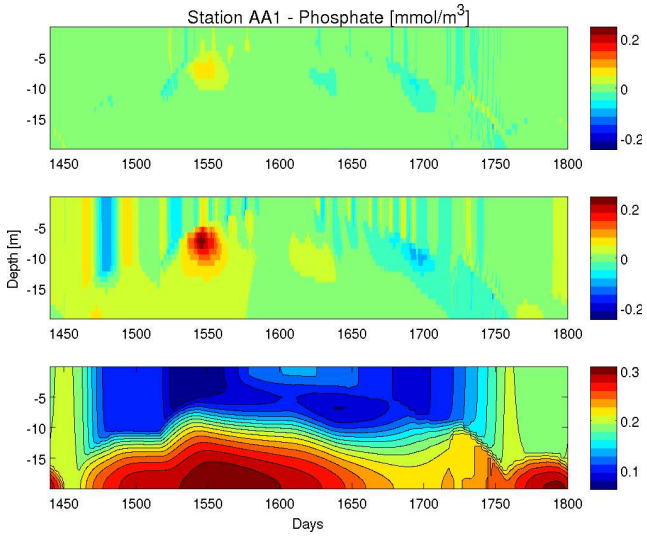


Figure 1.48: As fig. 1.47, but for phosphate $\left[\frac{mmol}{m^3}\right]$.

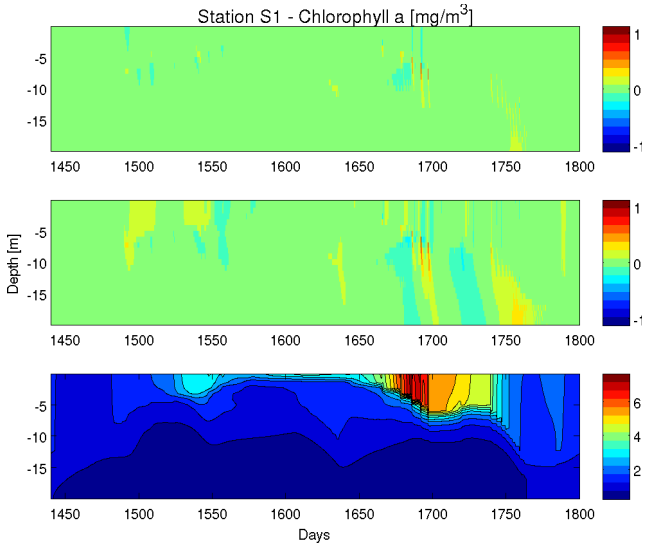


Figure 1.49: Chlorophyll-a content $\left[\frac{mg}{m^3}\right]$ at site S1 - Differences of the versions of rate estimation for the POMStep scheme. *Top:* POMStep-2-int error in %; *Middle:* POMStep-2-con error in %; *Bottom:* POMStep-1.

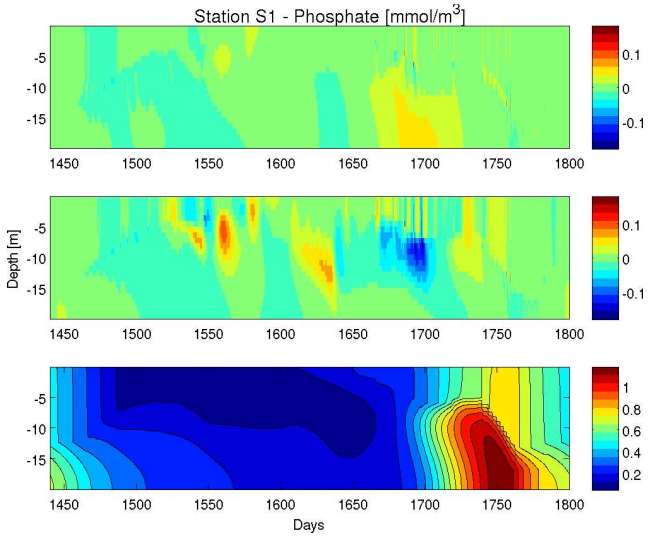


Figure 1.50: As fig. 1.49, but for phosphate $\left[\frac{mmol}{m^3}\right]$.

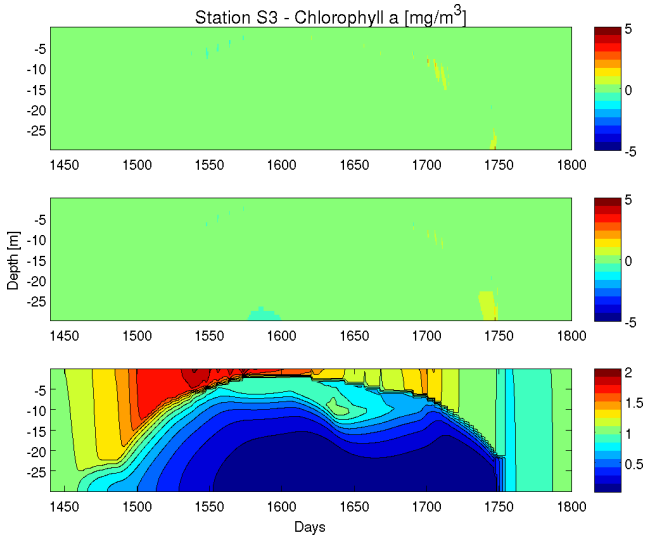


Figure 1.51: Chlorophyll-a content $\left[\frac{mg}{m^3}\right]$at site S3 - Differences of the two versions of rate estimation for the POMStep scheme. *Top:* POMStep-2-int error in %; *Middle:* POMStep-2-con error in %; *Bottom:* POMStep-1.

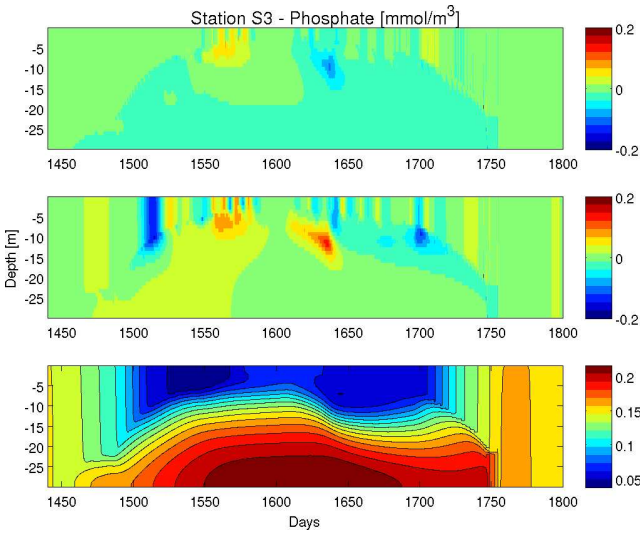


Figure 1.52: As fig. 1.51, but for phosphate $\left[\frac{mmol}{m^3}\right]$.

1.5 Conclusions

In this chapter the implementation of a coupling method for dynamic marine ecosystem models has been analysed and validated against a reference solution.

The applied *Source Splitting* method is based on the assumption of processes evolving on different time scales, so that the slower evolving process can be evaluated on larger intervals and is assumed constant on intermediate steps. As in marine ecosystem dynamics the general relation of the physical and biogeochemical processes with respect to their characteristic time scale is not clear (see section 1.3.8), the *Source Splitting* method was considered alternatively with the biogeochemical $\left(\frac{\partial \mathbf{c}(t_n)}{\partial t} \Big|_{bio} \right)$ or the physical $\left(\frac{\partial \mathbf{c}(t_n)}{\partial t} \Big|_{phys} \right)$ rate of change as the slower evolving one.

In the range of time steps $\Delta t_{bio} \leqslant \Delta t_{phys}$ the model (with BFMStep scheme) did not show significant sensibility to the resolution of the biogeochemical processes. Also, with respect to the numerical scheme used in the time integration step no differences in performance were detected for this range, so that the biogeochemical processes can be considered sufficiently resolved at $\Delta t_{bio} = \Delta t_{phys}$. However, as physical models usually directly update the state variables, an artificial rate had to be calculated a priori to provide the changing rate due to physical processes, yielding a splitting error. This splitting error came out to be rather significant, in particular in singular state variables (see section 1.4.3).

On the contrary, a convenience of the case, where the time scale is dictated by the physics and thus, the biogeochemical rates are estimated (POMStep), is given by the fact, that biogeochemical models usually directly compute rates, that can be inserted in the integration step of the physical model without any estimation involved.

A comparison of the two methods POMStep and BFMStep at synchronous time steps showed much higher errors for the experiment BFMStep, where the physical rates where estimated. As the integration scheme used in POMStep for the global integration was the same as the one for the sub-

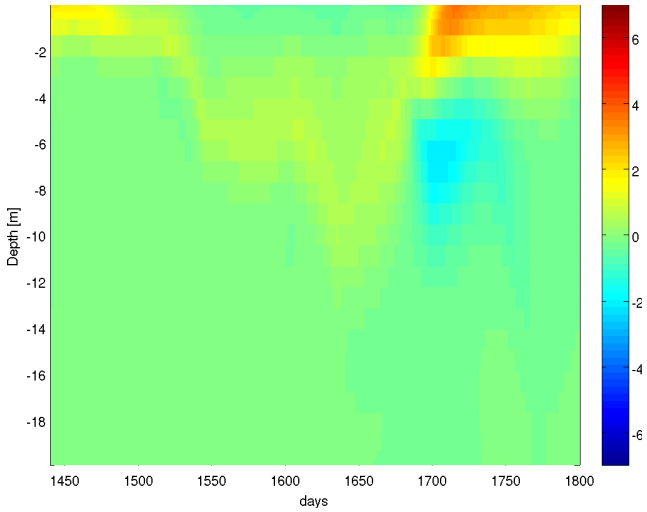


Figure 1.53: The chlorophyll-a rates $\frac{1}{Chl}\left.\frac{\partial Chl}{\partial t}\right|_{bio}$ in $[\frac{1}{d}]$ at site S1 calculated by the *BFM* , normalised by its concentration. From experiment POMStep-Ref.

integration in BFMStep, the error must be caused by the splitting. This error dominates over all other errors intrinsic to the BFMStep version of the model, so that the choice of the integration scheme used for the global integration becomes irrelevant.

The independence of the BFMStep experiments at time steps $\Delta t_{glob} \leqslant \Delta t_{phys}$ with respect to time step size and integration schemes, the high splitting error of the BFMStep method and on the other side the better performance of the POMStep method suggests that it is actually the physical processes, that evolve faster than the biogeochemical ones.

This suspect is quantitatively supported by a comparison of the rates computed by the model. In figures 1.53 to 1.56 this comparison is shown for the key variables of the investigated ecosystem phosphate and chlorophyll-a. The highest rates overall, that determine the maximum time step, can be observed in the transport rates for chlorophyll-a connected to the deepening of the pycnocline (fig. 1.54): the high surface concentration of chlorophyll-a of the late summer (see e. g. fig. 1.15) is mixed downwards

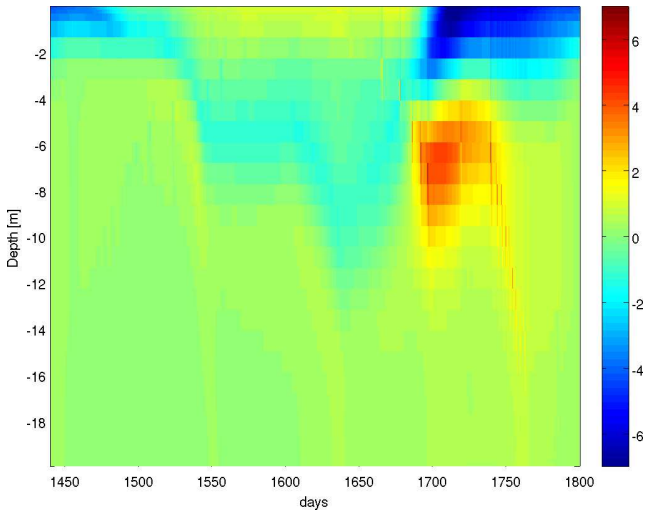


Figure 1.54: The chlorophyll-a rates $\frac{1}{Chl}\ \frac{\partial Chl}{\partial t}\Big|_{phys}$ in $[\frac{1}{d}]$ at site S1 calculated by the *POM* , normalised by its concentration. From experiment POMStep-Ref.

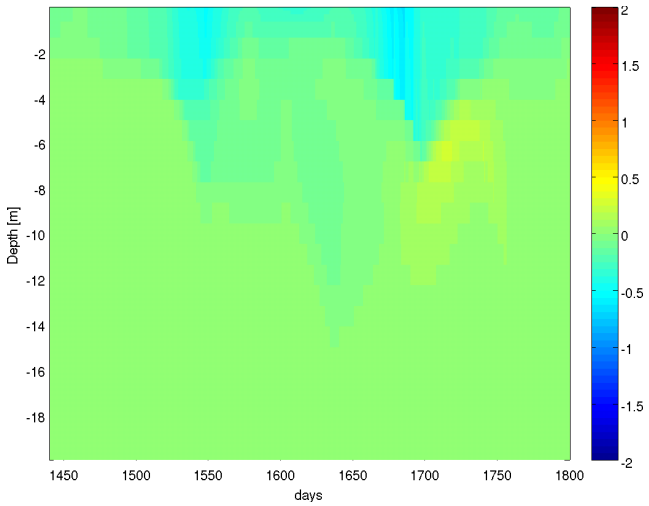


Figure 1.55: The phosphate rates $\frac{1}{PO_4}\ \frac{\partial PO_4}{\partial t}\Big|_{bio}$ in $[\frac{1}{d}]$ at site S1 calculated by the *BFM* , normalised by its concentration. From experiment POMStep-Ref.

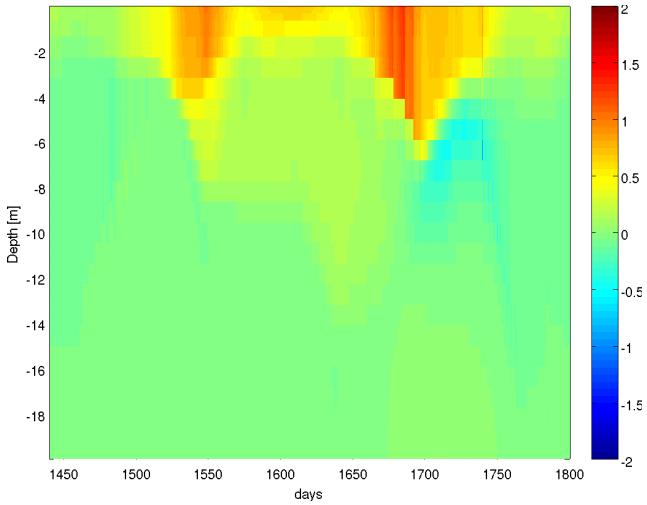


Figure 1.56: The phosphate rates $\frac{1}{PO_4}\ \frac{\partial PO_4(t_n)}{\partial t}\Big|_{phys}$ in $[\frac{1}{d}]$ at site S1 calculated by the *POM* , normalised by its concentration. From experiment POMStep-Ref.

into regions previously separated by the salinity gradient. The opposite holds, to lesser degree for PO_4, that was enclosed under the pycnocline and is now mixed into the surface layer, where in addition consumption decreases. The high values of superficial primary production during autumn and winter (fig. 1.53) are clearly exceeded by strong mixing that spreads the superficial chlorophyll overplus efficiently over the whole water column. Also for phosphate the dynamics are dominated by vertical mixing and the physical rate is significantly higher with respect to the biogeochemical one.

The fact, that this relation was observed in all three sites under rather different set-ups of physical and biogeochemical forcing leads to the conclusion, that it is in fact the physical scale, that imposes the time resolution of the numerical integration. Therefore an implementation, that integrates the transport equation directly under use of the biogeochemical rates is preferable. However, the general validity of this statement will have to be further investigated in different configurations, especially in

the context of three-dimensional models with advection and high-frequency forcing.

Finally, the possibility of further reducing the time resolution was highlighted. In this context an interpolated estimation of the biogeochemical rates based on the Adams-Bashforth approach was suggested (POMStep-2-inp experiments). Results showed a clear improvement over the classical piecewise-constant approach (POMStep-2-con). However, the gain in computational resources through the estimation of the biogeochemical processes on coarse intervals is mostly limited, as the higher effort in the computations is usually spent in the often implicit solvers of the physical model and the sub-model for the transport of the biogeochemical state variables.

Appendix A

Model Implementation

In this appendix the model results (POMStep-Ref) are validated at the seasonal scale (following the procedure proposed by Vichi et al., 2003) against the seasonal profiles of selected variables obtained from the observations contained in the ABCD dataset (Zavatarelli et al., 1998), for locations S1 and S3 and in the LBM (Laboratorio di Biologia Marina, Trieste) data set for the AA1 location. The choice of the seasonal time scale is motivated by the large temporal and spatial inhomogeneities of the available data pertinent to the implementation sites. This prevented to carry out the validation at a more resolved temporal scale. The data used for the S1 and S3 validation were extracted from ABCD by considering an area of about $0.4°x0.4°$ centred around the implementation sites. All the data of the LBM dataset pertinent to the Gulf of Trieste were considered valid for the AA1 validation.

The seasonal means of model state variables are in the following plotted against the means and the range of variability of all the available observations. Here it is shown a selection of model results focusing on temperature, (T), salinity (S) phosphate (PO_4) and chlorophyll-a (Chl). Figures are organised in rows corresponding to the three implementation sites (S1, S3 and AA1). Column represents the (calendar) season.

Hydrological properties

Fig. 1.57 shows the temperature seasonal profiles at at all sites and is characterised by well mixed conditions in winter and thermal stratification from spring to autumn. The observed variability is satisfactorily captured by the simulated profiles as the simulated profiles lies almost entirely within the observed variability. Although deviations (under and over estimations) from the observed mean can be noted, particularly in the shallower locations S1 and AA1, while the deeper S3 location show an overall better agreement.

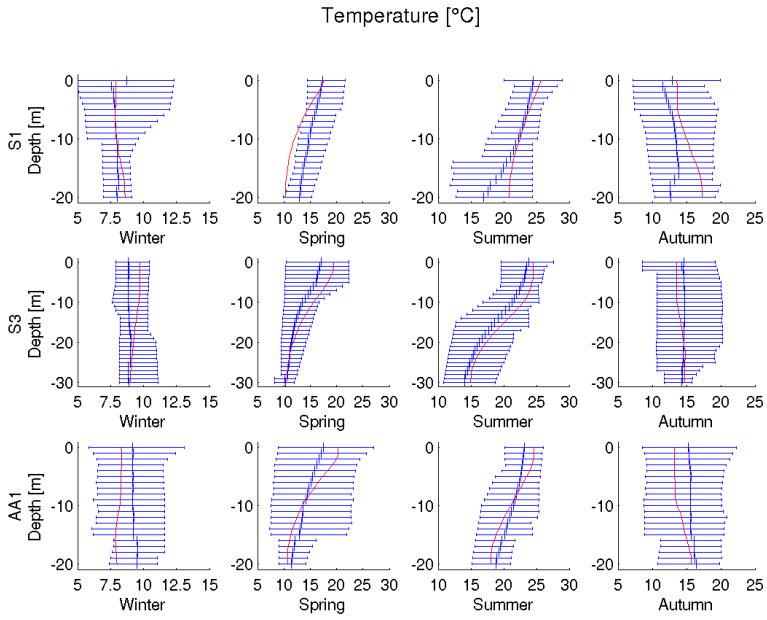


Figure 1.57: Comparison of seasonal mean temperature $T\,[^{\circ}C]$ of the model with climatological data of the four seasons from left to right. *Red:* model, *blue:* data with mean and standard deviation; *Top:* S1, *middle:* S3, *bottom:* AA1.

Fig. 1.58 shows the salinity seasonal profiles that highlight the main differences among the implementation sites. S1 shows the largest surface salinity variability due to the strong influence of the Po river runoff variability. This signal, despite the fast relaxation constant (see table 1.2), is only partially captured by the model as the simulated surface values do not exactly match the seasonal mean. Apart from that, the salinity profile show an overall good agreement with the observations. This applies even more to the other two locations over all seasons.

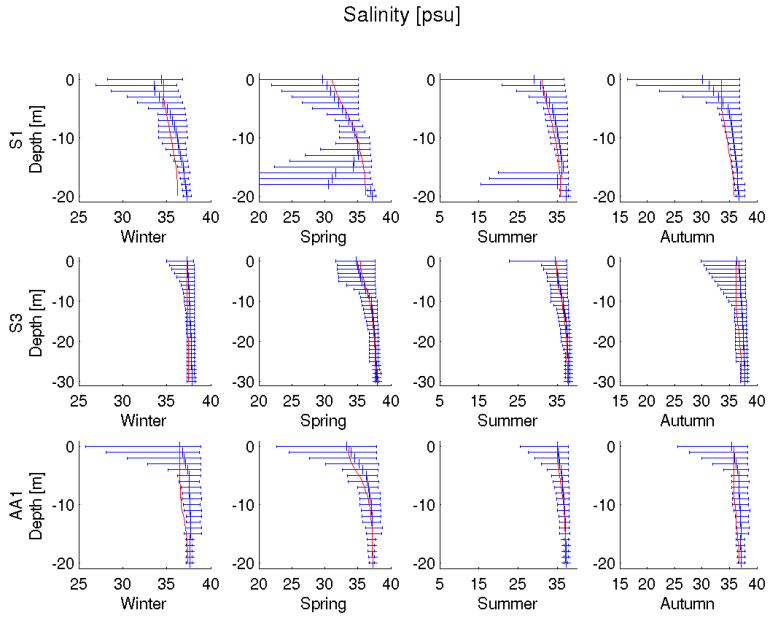


Figure 1.58: Comparison of seasonal mean salinity $S[psu]$ of the model with climatological data of the four seasons from left to right. *Red:* model, *blue:* data with mean and standard deviation; *Top:* S1, *middle:* S3, *bottom:* AA1.

Biogeochemical properties

The simulated Phosphate seasonal profiles are shown in Fig. 1.59.

At location S1, the model clearly overestimate the concentration all over the water column, even if the general vertical trend appears correct. The spring profile indicates a good agreement with data in the upper part of the water column, while below 10 m depth there is a clear model tendency to accumulate phosphate. Much better is the situation in summer as the simulated profile entirely lies within the observed variability, even if there is a slight tendency (contrary to the preceding season) to underestimate the concentration in the lower water column. Finally, the autumn profile appears as the more problematic, being characterised by a tendency to underestimate phosphate at surface and overestimate again in the lower water column. Much better is the model behaviour at location AA1 and S3, since the simulated profiles lies within the observed variability at all seasons, with a remarkable similarity to the average profile in summer and autumn.

The corresponding chlorophyll profiles are shown in Fig. 1.60: Also in this case location S1 is the one that shows the larger departure from the observations, particularly in winter, where the chlorophyll profile is clearly underestimated. As in Vichi et al. (2003) the reason for that can be traced to a not satisfactorily representation of the light distribution in the water column, as a consequence of the scanty information on the suspended sediment vertical distribution and variability. The underestimation persists also in spring, while the model recovers a satisfactory profile in summer and autumn. As for phosphate the model behaviour is much more satisfactory at locations AA1 and S3, where the simulated profiles always lies within the range of the observed variability, with the only exception of a slight underestimation at location AA1 in spring.

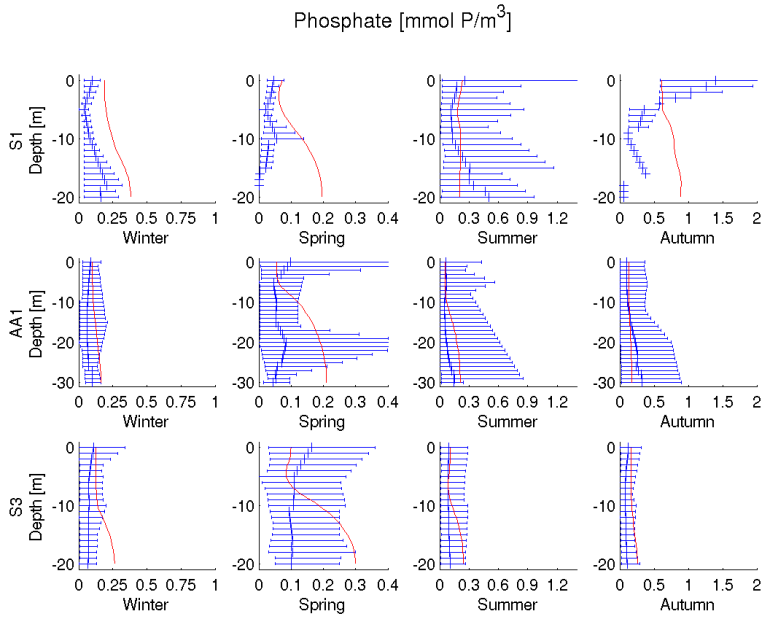
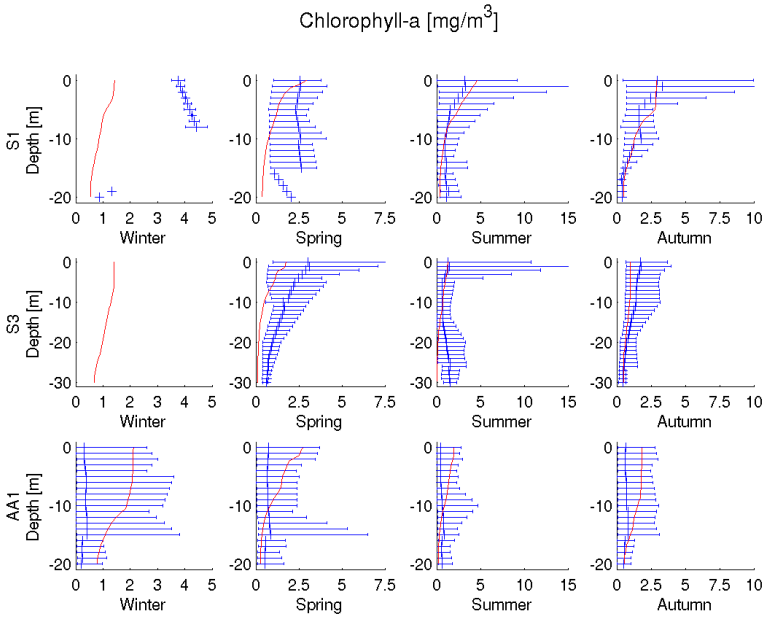


Figure 1.59: Comparison of seasonal mean phosphate PO_4 $\left[\frac{mmol}{m^3}\right]$ of the model with climatological data of the four seasons from left to right. *Red:* model, *blue:* data with mean and standard deviation; *Top:* S1, *middle:* S3, *bottom:* AA1.



\

Figure 1.60: Comparison of seasonal mean chlorophyll-a concentration $Chl\left[\frac{mg}{m^3}\right]$ of the model with climatological data of the four seasons from left to right. *Red:* model, *blue:* data with mean and standard deviation; *Top:* S1, *middle:* S3, *bottom:* AA1.

Appendix B

POMStep simulation: Results for selected biogeochemical state variables

An overview over the results is given completing the fields of chlorophyll and phosphate with the results over other state variables representing the remaining elements of the food web: bacteria, microzooplankton and particulate organic matter (all given in terms of carbon biomass), completed by oxygen.

The figures show, that the level of error for the POMStep scheme lies generally well below 1%. Only in the case of particulate organic carbon isolated peak values of 2% are reached caused by extremely steep gradients whose dynamics appears slightly shifted with respect to the reference field, causing these very localised error structures.

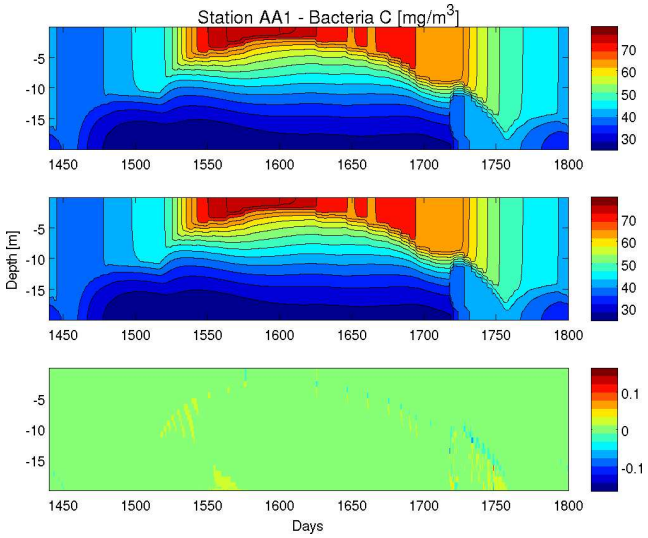


Figure 1.61: Bacteria carbon content $\left[\frac{mg\,C}{m^3}\right]$ in site AA1.
Top: POM-Step-1; *middle:* Reference Solution; *bottom:* Differences to reference solution in %

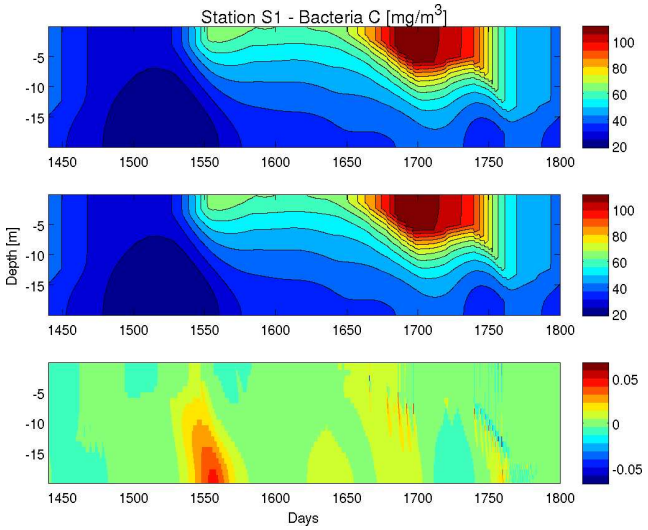


Figure 1.62: Bacteria carbon content $\left[\frac{mg\,C}{m^3}\right]$ in site S1.
Top: POM-Step-1; *middle:* Reference Solution; *bottom:* Differences to reference solution in %

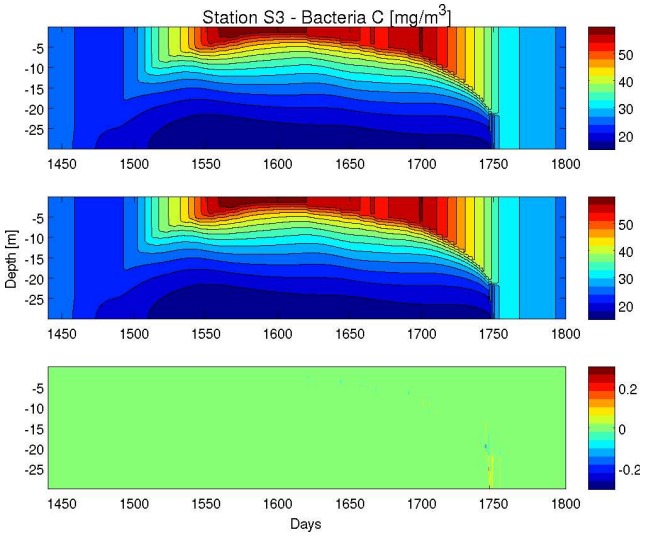


Figure 1.63: Bacteria carbon content $\left[\frac{mg\,C}{m^3}\right]$ in site S3.
Top: POM-Step-1; *middle:* Reference Solution; *bottom:* Differences to reference solution in %

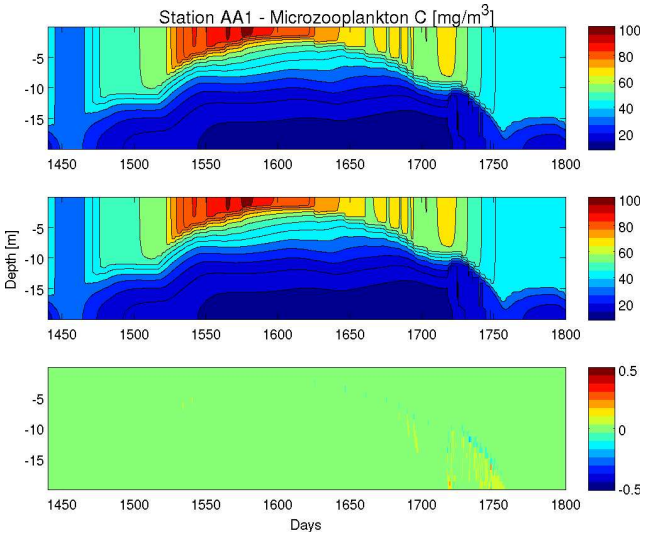


Figure 1.64: Microzooplankton carbon content $\left[\frac{mg\,C}{m^3}\right]$ in site AA1.
Top: POM-Step-1; *middle:* Reference Solution; *bottom:* Differences to reference solution in %

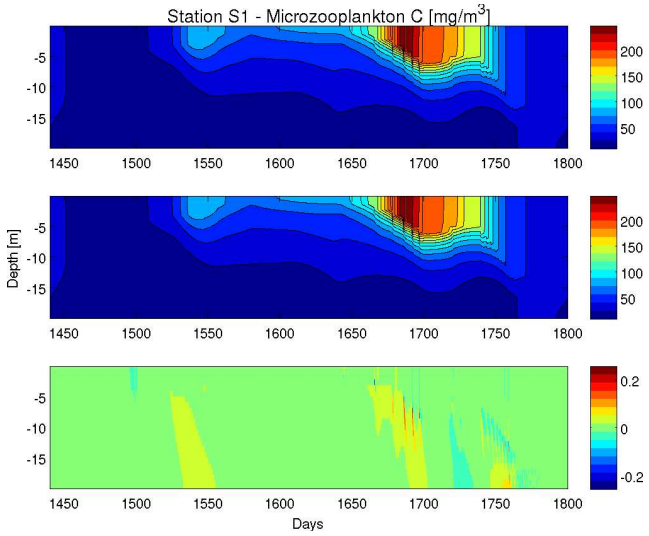


Figure 1.65: Microzooplankton carbon content $\left[\frac{mg\,C}{m^3}\right]$ in site S1. *Top:* POM-Step-1; *middle:* Reference Solution; *bottom:* Differences to reference solution in %

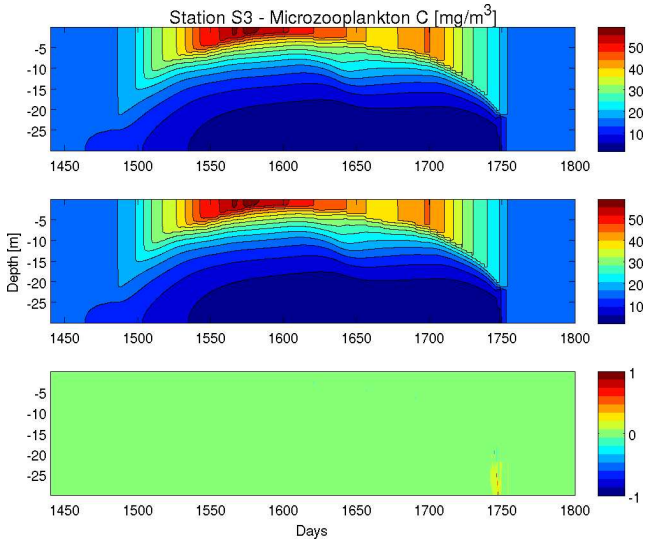


Figure 1.66: Microzooplankton carbon content $\left[\frac{mg\,C}{m^3}\right]$ in site S3. *Top:* POM-Step-1; *middle:* Reference Solution; *bottom:* Differences to reference solution in %

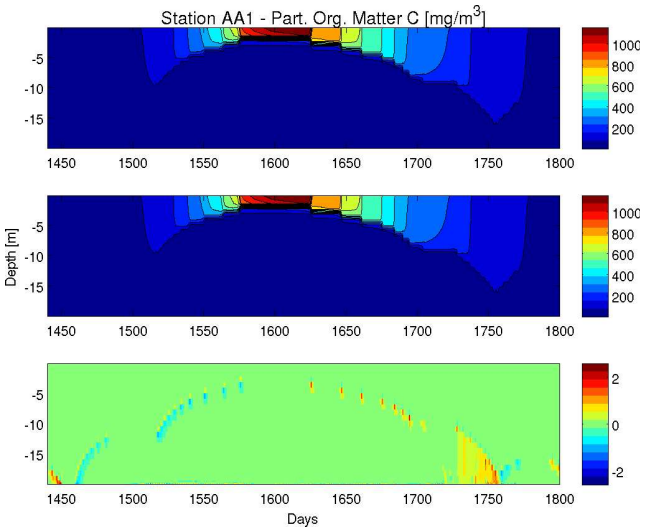


Figure 1.67: Particulate organic carbon $\left[\frac{mg\,C}{m^3}\right]$ in site AA1.
Top: POM-Step-1; *middle:* Reference Solution; *bottom:* Differences to reference solution in %

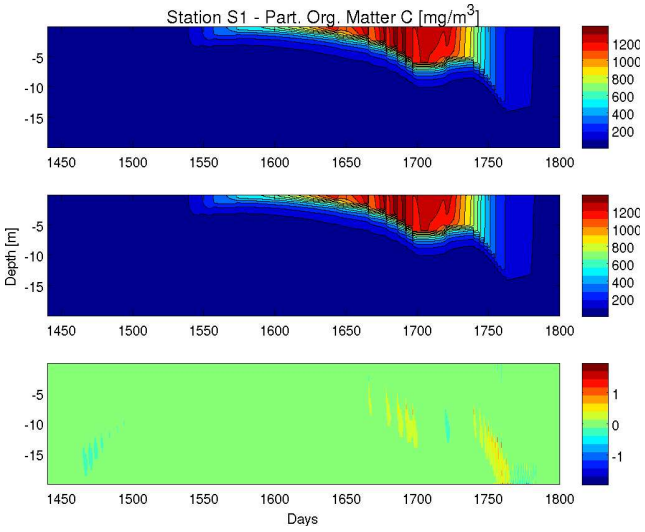


Figure 1.68: Particulate organic carbon $\left[\frac{mg\,C}{m^3}\right]$ in site S1.
Top: POM-Step-1; *middle:* Reference Solution; *bottom:* Differences to reference solution in %

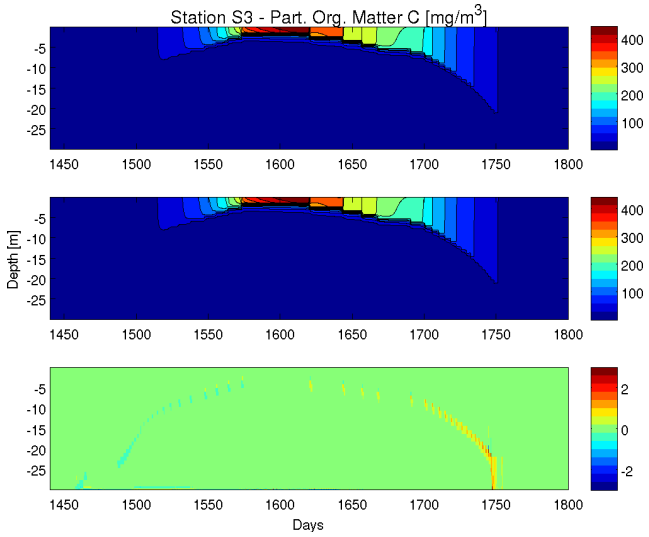


Figure 1.69: Particulate organic carbon $\left[\frac{mg\,C}{m^3}\right]$ in site S3.
Top: POM-Step-1; *middle:* Reference Solution; *bottom:* Differences to reference solution in %

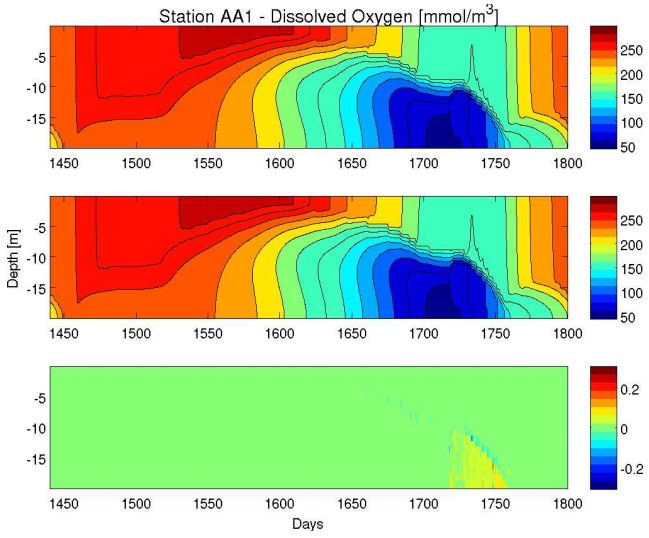


Figure 1.70: Dissolved oxygen $\left[\frac{mmol}{m^3}\right]$ in site AA1.
Top: POM-Step-1; *middle:* Reference Solution; *bottom:* Differences to reference solution in %

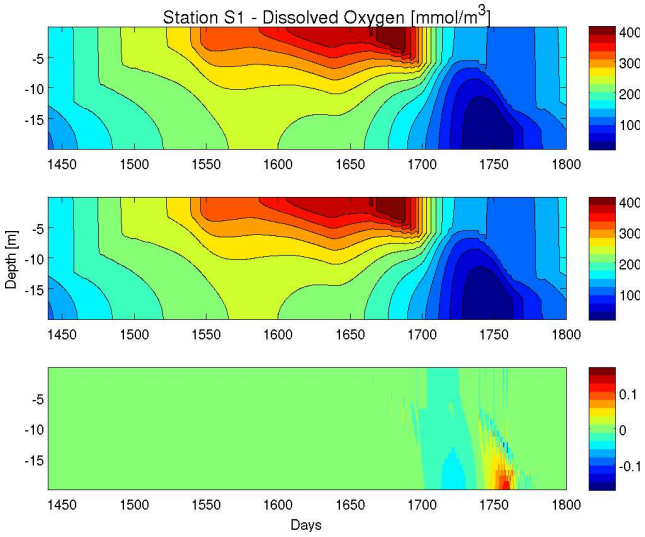


Figure 1.71: Dissolved oxygen $\left[\frac{mmol}{m^3}\right]$ in site S1.
Top: POM-Step-1; *middle:* Reference Solution; *bottom:* Differences to reference solution in %

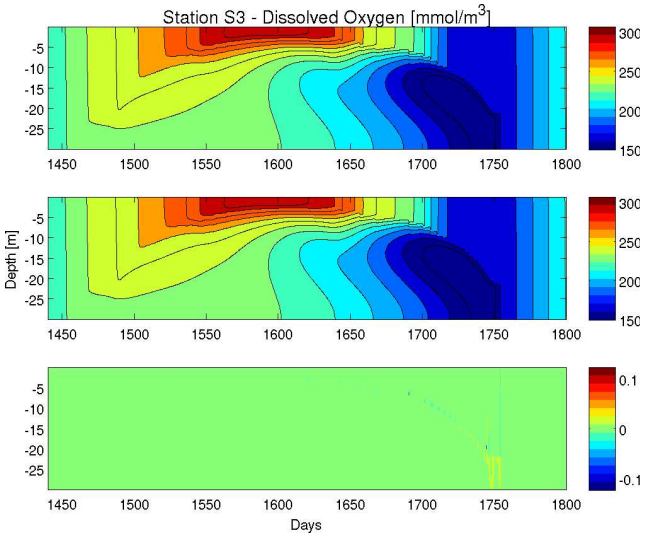


Figure 1.72: Dissolved oxygen $\left[\frac{mmol}{m^3}\right]$ in site S3.
Top: POM-Step-1; *middle:* Reference Solution; *bottom:* Differences to reference solution in %

2 A data assimilation scheme for dynamic marine ecosystem models

Introduction - The concept of data assimilation

A major concern of geophysical research and, in particular, of meteorology and oceanography is forecasting. The need for a detailed prediction of future natural events is driven by a variety of motivations that go from simple little problems of everyday's life over economical interests (e. g. harvest and fishery) to the severest concerns of life like natural hazards and climate change.

Weather forecasting is the most known and successful example of "operational science" applied to the prediction of the future state of a natural system (Navarra, 1996). In fact numerical weather forecasting constitutes now a mature field in the applied geophysical sciences. In particular the knowledge and the advances gained in this field constitute an extremely value asset for the development of an analogous prediction system applied to the ocean.

In extreme simplification a geophysical fluids forecasting system is constituted by two fundamental components: a deterministic numerical model and an observing system. Numerical models offer a description of the physical and ecological dynamics of the system of interest at the required time and space scales. In the case of physical models the limit of detail is now mainly dictated by the computational resources only. However, models are and will always be only a simplified representation of reality. They include a number of approximations, assumptions and uncertainties that cause the model to drift away from the real state of the system that it is designed to predict. Moreover, and this is principally the case with ecological models, with the increasing complexity of the models, also their calibration becomes more and more difficult, due to the increasing number of system parameters. In addition to that, even if a model could yield the correct solution of the system equations, it still relies on the accurate description of the exact boundary and initial conditions. Finally, one of the fundamental insights in physics of the last century was that nature is *not* deterministic, i. e. even if the current state of a natural system would

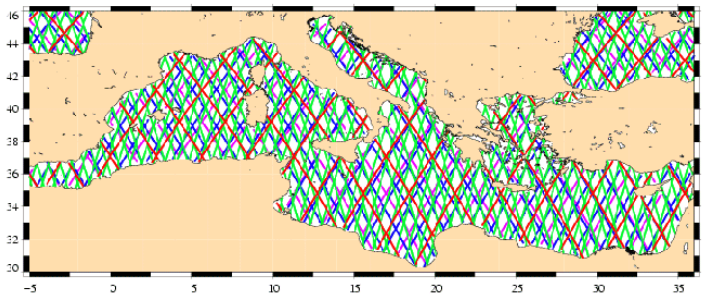


Figure 2.1: Altimetry data coverage from satellite tracks for the Mediterranean Sea as used in the MFSTEP project (http://www.bo.ingv.it/mfstep). *Green:* ENVISAT, *Red:* JASON-1, *Pink:* TOPEX/POSEIDON, *Blue:* Geosat Follow On.

be fully known, its future state is not uniquely defined. Therefore approaches are needed, that take probabilistic considerations into account, e. g. instead of determining a theoretical future state, give rather an estimation of the most likely realisation of future events, along with its probability and an estimation of the eventual error (Evensen, 2006).

The other cornerstone in forecasting is the observational system. Thanks to the achievements in research and technology, measurements have become quite efficient and more and more accurate. Considerable effort has been spent to extend and refine the networks of data acquisition and to collect the data (e. g. the TOGA observing system, see McPhaden (1998)). In particular the use of satellite missions (e. g. TOPEX/POSEIDON, ENVISAT, etc.) has largely increased the availability of essential observational data, such as sea-level anomaly, sea surface temperature or ocean colour (an example for the Mediterranean Sea is given in fig. 2.1). However, "in situ" data collection occurs forcefully at discrete locations in time and space, thus data availability remains restrictive, so that extrapolation strategies are needed to obtain informations on uncovered areas. The first systematic approaches of this kind go back to the middle of the 19th century when the first synoptic charts of the meteorological conditions of the atmosphere were accomplished. The method applied consists in

interpolating available observational data manually to isolines and vector fields of the involved physical quantities by sophisticated graphical techniques under the consideration of the connected physical laws. (This kind of data processing is often referred to as subjective analysis, as it requires direct human interaction and experience). The constructed charts were then used to project the contained information in future and provide short-time forecasts. (A classical description is given in Richardson, 1922)

The improvements in the observational capacity of the atmosphere yields a large amount of data, making subjective analysis no longer operable and called for (thanks also to computer developments) automatic data processing through the use of an algorithm based on mathematical interpolation techniques taking into account physical constraints. The first steps into this direction were attempted by Panofsky (1949), but major progress arose in the field with the concepts presented by Gilchrist and Cressman (1954), who introduced the concept of the regions of influence defining weight factors on the base of local distances to the observations and suggested the use of prior estimates as background information to merge with the observational data. These concepts were soon supplemented by statistical methods that considered the uncertainties of prior estimate and observations to find a proper balance for the best estimate of the analysis (Gandin, 1963), leading to the procedures now generally referred to as objective analysis.

With the automatic processing of data to project information on regular grids, the missing link between the two cornerstones (models and observing system) was provided. It is clear that the most accurate forecast possible, should take into account all the informations that are available and merge them into a forecasting system. With the tools of objective analysis, the information contained in observations can be projected on the discrete grid of a numerical model to provide a more realistic initial condition. (See Daley, 1991 for an extensive discussion on the subject of data analysis.) More than that, the observations can be merged dynamically in a model to correct the simulated state of the system, whenever data be-

comes available. Such a technique is called data assimilation system, that can be formally defined as the solution to the inverse problem of finding the best estimate of a system state given the dynamical equations and initial conditions of a system and a set of measurements (see also Bennett, 1992).

Models of marine ecosystem dynamics represent an attempt to predict the biogeochemical state of the ocean under the influence of the physical dynamics. The marine ecosystem is characterised by complex and strongly coupled interactions between a huge number of organisms and chemical constituents. In the design of a model, the trophic webs have to be idealised considerably by grouping the involved elements in generic classes or functional groups (see section1.2.2) and the food chain has to be cut off at adequate levels to close the system at the upper and lower ends of model resolution. Thus, the level of approximation is high and even if considerable effort has been spend in recent years on the design and calibration of deterministic models (Hofmann and Lascara, 1998) and reasonable success has been achieved in qualitative and quantitative modelling of the involved dynamical biogeochemical cycles, this is a research field still far from maturity. However, such dynamic descriptions of the (approximate) ecosystem are of valuable interest for a variety of reasons (e. g. impact of climate change on the marine ecosystem and vice versa, prediction of harmful algal blooms, mucilage events and anoxic crisis etc.), so that a number of recent works (e. g. Evensen, 2003; Berline et al., 2007; Raick et al., 2007) are concentrating on the application of data assimilation filters to marine ecosystem models to compensate for the large inaccuracies caused by the approximations and gain statistical information on the involved interactions leading to a more precise picture of the marine ecosystem dynamics.

In this chapter the implementation of different versions of the Singular Evolutive Extended Kalman Filter (*SEEK*, an evolution of the Extended Kalman Filter for sequential data assimilation) into the coupled physical biogeochemical model described in the previous chapter is proposed. The

first two sections give a broad overview of the formalism involved, in section 3 the *SEEK* Filter, as introduced by Pham et al. (1998), is presented along with two subversions that have been applied in marine ecosystem modelling. The fourth section gives the details of the implementation of the filter, including a description of the statistical background needed for the TWIN experiments carried out for validation, while section 5 gives the results of theses experiments. Section 6 gives the description of a new version of the filter based on the general formulation of Pham et al. (1998) and the results of the experiments performed with this new filter will be given in section 7. Section 8 finally offers the conclusions of this study.

2.1 Sequential data assimilation - the Kalman Filter

2.1.1 Optimal Interpolation

Objective analysis deals with the inverse problem to find the most probable realisation x_a of a system state given a measurement y^o and a first guess estimate, often also referred to as background field x^b. Both, the measurements and the background field are subject to errors (ϵ and p^b respectively), whose statistics are described by the error covariance $\boldsymbol{P}^b = \overline{p^b p^{b^{\mathrm{T}}}}$ for the background field and $\boldsymbol{R} = \overline{\epsilon\epsilon^{\mathrm{T}}}$ for the measurements. Consequently, also the analysis will give an error p_a with the covariance $P_a = \overline{p^a p^{aT}}$. The errors are given by the equations:

$$x^b = x^t + p^b \tag{2.1a}$$

$$y^o = \boldsymbol{H}x^t + \epsilon \tag{2.1b}$$

where x^t is the true state of the system and $\boldsymbol{H}$ the operator transforming a vector from model space to observation space. Then, for a linear case one can assume a solution of the form

$$x^a \;=\; x^t + p^a = \boldsymbol{A}x^b + \boldsymbol{K}y \; . \tag{2.2}$$

defining $\overline{p^a} = 0$.[1]

Inserting eq. 2.1a and 2.1b into 2.2 gives then

$$x^t + p^a = \boldsymbol{A}\left(x^t + p^b\right) + \boldsymbol{K}\left(\boldsymbol{H}x^t + \epsilon\right) \; . \tag{2.3}$$

Under the assumption, that the errors are unbiased ($\overline{p^b} = 0$ and $\bar{\epsilon} = 0$)

[1]Notation: italic letters refer to column vectors, bold italic letters refer to matrices or matrix-operators, calligraphic letters to gradient matrices. Lower Indices refer to time levels, upper indices to qualities such as *b*: background field, *a:* analysis, *t:* true state, *o:* observation, f: forecast.

and not correlated ($\overline{p^b \epsilon} = 0$), the expectation of this equation becomes

$$x^t = \boldsymbol{A} x^t + \boldsymbol{K}\boldsymbol{H} x^t \tag{2.4}$$

and therefore

$$\boldsymbol{A} = \boldsymbol{I} - \boldsymbol{K}\boldsymbol{H} \ . \tag{2.5}$$

where I is the identity matrix with

Thus, the analysis state is described by the equations

$$x^a = x^b + \boldsymbol{K}\left(y^o - \boldsymbol{H} x^b\right) \tag{2.6a}$$
$$p^a = p^b + \boldsymbol{K}\left(\epsilon - \boldsymbol{H} p^b\right) \tag{2.6b}$$

With the above assumptions on the errors, one obtains for the analysis covariance

$$\boldsymbol{P}^a = \boldsymbol{P}^b - \boldsymbol{K}\boldsymbol{H}\boldsymbol{P}^b - \boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}}\boldsymbol{K}^{\mathrm{T}} + \boldsymbol{K}\boldsymbol{R}\boldsymbol{K}^{\mathrm{T}} + \boldsymbol{K}\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}}\boldsymbol{K}^{\mathrm{T}} \ . \tag{2.7}$$

The optimal solution is then given by the minimum of the total variance of $\boldsymbol{P}^a$, i. e.

$$\mathfrak{d}_{\boldsymbol{K}} Tr\left(\boldsymbol{P}^a\right) = 0 = -2Tr\left(\boldsymbol{H}\boldsymbol{P}^b + \boldsymbol{R}\boldsymbol{K}^{\mathrm{T}} + \boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}}\boldsymbol{K}^{\mathrm{T}}\right) \ . \tag{2.8}$$

Thus, the optimal analysis state is given for

$$\boldsymbol{K} = \boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}}\left(\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H} + \boldsymbol{R}\right)^{-1} \tag{2.9}$$

as

$$x^a = x^b + \boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}}\left(\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H} + \boldsymbol{R}\right)^{-1}\left(y^o - \boldsymbol{H} x^b\right) \tag{2.10}$$

with an error covariance of

$$\boldsymbol{P}^a = \left(\boldsymbol{I} - \boldsymbol{K}\boldsymbol{H}\right)\boldsymbol{P}^b \ . \tag{2.11}$$

This analysis is usually referred to as linear optimal interpolation and it

gives the estimate of the system state that minimises the root mean square error (rms) with respect to the true state of the system for linear observation operators $\boldsymbol{H}$ under the assumption of unbiased and uncorrelated errors p^b and ϵ .

2.1.2 Sequential data assimilation

So far the problem has been posed not taking into account the time, as the method presented in the previous section computes the best estimate of a system state under certain assumptions, given the complete knowledge of the background and observation error covariances. The application of this method to assimilate measurements into a time dependent model could then be achieved in the following sequence:

- A model forecast step is performed providing the background information on system state and error covariances.
- Available measurements and the connected uncertainty are fed to the system.
- The best estimate and its error covariances are calculated according to equations (2.10) and (2.11).
- A new forecast step is launched starting from the obtained analysis state.

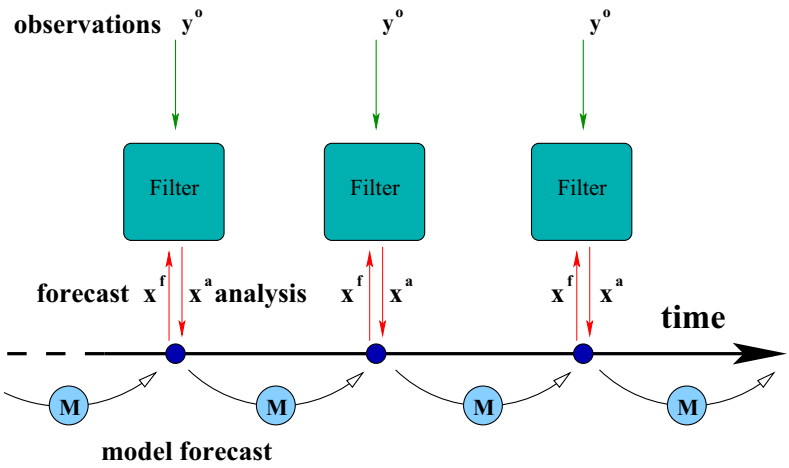


Figure 2.2: Sequential data assimilation cycles. M: model, x^{f}: forecast state, x^{a}: analysis state, y^o: observations.

This method is defined as sequential data assimilation and is characterised by the propagation in time of the error covariances. While for the measurement error some information might be available from calibration, statistics or instrumental documentation, this is not the case for the error connected to the background information, that is now better defined as forecast error, being a consequence of the model dynamics and of the previous analysis errors and is evolving in time.

2.1.3 The Kalman Filter

Kalman (1960) introduced an equation for the time evolution of the forecast error covariance in the case of linear model propagation:

The forecast state is given by the application of the model on the previous analysis step, propagating it from time t_{n-1} to time t_n:

$$x_n^f = \boldsymbol{M}_{n-1} x_{n-1}^a \ . \tag{2.12}$$

where $\boldsymbol{M}_{n-1}$ is the model operator on the time interval $[t_{n-1}; t_n]$. The pure

model error is then given by the equation

$$x_n^t = \boldsymbol{M}_{n-1} x_{n-1}^t + q_n \tag{2.13}$$

where q_n is the model error or dynamic noise.

With the limitation to linear model forecast operators, the two equations can be combined as

$$x_n^t - x_n^a = \boldsymbol{M}_{n-1} \left(x_{n-1}^t - x_{n-1}^a \right) + q_n \ . \tag{2.14}$$

The forecast error is then given by

$$\begin{aligned}
\boldsymbol{P}_n^f &= \overline{\boldsymbol{M}_{n-1} \left(x_{n-1}^t - x_{n-1}^a \right) \left(x_{n-1}^t - x_{n-1}^a \right)^{\mathrm{T}} \boldsymbol{M}_{n-1}^{\mathrm{T}}} + \overline{q_n {q_n}^{\mathrm{T}}} + \overline{\boldsymbol{M}_{n-1} \left(x_{n-1}^t - x_{n-1}^a \right) {q_n}^{\mathrm{T}}} \\
&\quad + \overline{q_n \left(x_{n-1}^t - x_{n-1}^a \right)^{\mathrm{T}} \boldsymbol{M}_{n.1}^{\mathrm{T}}} \\
&= \overline{\boldsymbol{M}_{n-1} p_{n-1}^a {p_{n-1}^a}^{\mathrm{T}} \boldsymbol{M}_{n-1}^{\mathrm{T}}} + \overline{q_n {q_n}^{\mathrm{T}}} + \overline{\boldsymbol{M}_{n-1} p_{n-1}^a {q_n}^{\mathrm{T}}} + + \overline{q_n {p_{n-1}^a}^{\mathrm{T}} \boldsymbol{M}_{n-1}^{\mathrm{T}}} \\
&= \boldsymbol{M}_{n-1} \boldsymbol{P}_{n-1}^a \boldsymbol{M}_{n-1}^{\mathrm{T}} + \boldsymbol{Q}_n \ ,
\end{aligned}$$

so that finally

$$\boldsymbol{P}_n^f = \boldsymbol{M}_{n-1} \boldsymbol{P}_{n-1}^a \boldsymbol{M}_{n-1}^{\mathrm{T}} + \boldsymbol{Q}_n \ , \tag{2.15}$$

assuming no correlation between the analysis error at the previous time step and the model error at the actual time step.

This, together with the equations of the optimal interpolation (eq. (2.10), (2.11)) provides a consistent set of dynamical equations for the model evolution and the connected error covariances, that define the *Kalman Filter*:

$$x_n^f = \boldsymbol{M}_{n-1} x_{n-1}^a \tag{2.16a}$$

$$\boldsymbol{P}_n^f = \boldsymbol{M}_{n-1} \boldsymbol{P}_{n-1}^a \boldsymbol{M}_{n-1}^{\mathrm{T}} + \boldsymbol{Q}_n \tag{2.16b}$$

$$\boldsymbol{P}_n^a = (\boldsymbol{I} - \boldsymbol{K}\boldsymbol{H}) \, \boldsymbol{P}^b \tag{2.16c}$$

$$x_n^a = x_n^f + \boldsymbol{P}_n^f \boldsymbol{H}^{\mathrm{T}} \left(\boldsymbol{H} \boldsymbol{P}_n^f \boldsymbol{H} + \boldsymbol{R}_n \right)^{-1} \left(y_n^o - \boldsymbol{H} x_n^f \right) \tag{2.16d}$$

The model is started from an appropriate initial condition for the system state and the analysis covariance. Whenever measurements are avail-

able, they can be readily assimilated following the given sequence.

However, the analysis state provided by this algorithm is optimal only under the following assumptions:

- unbiased errors p^f, ϵ, q.
- no correlation between $\left(p^f; \epsilon\right)$ and $\left(q_n; p^a_{n-1}\right)$.
- linearity of the observation operator $\boldsymbol{H}$.
- linearity of the model operator $\boldsymbol{M}$.
- the covariances $\boldsymbol{Q}_n$, $\boldsymbol{R}_n$ and $\boldsymbol{P}^a_0$ are known and positive definite.

2.1.3.1 The Extended Kalman Filter

The assumption of linearity in the above filter is a strong limitation for its application in particular in the context of atmospheric and oceanographic models that are characterised by non-linear dynamics. An improvement, that extends this limitation to the near linear case, is provided by the Extended Kalman Filter (Jazwinski, 1970; Gelb, 1974) applying the tangent linear hypothesis. In this approach, the forecast error given in the linear case by equation (2.15) is calculated using a Taylor expansion for the forecast step, neglecting all terms of second order or higher

$$\boldsymbol{M}_{n-1}x^t_{n-1} = \boldsymbol{M}_{n-1}x^a_{n-1} + \mathcal{M}_{n-1}\left(x^t_{n-1} - x^a_{n-1}\right) + \mathcal{O}\left(x^2\right) \tag{2.17}$$

where $\mathcal{M}_{n-1}$ is the Jacobian of the model operator

$$\mathcal{M}_{n-1} = \left.\frac{\partial \boldsymbol{M}\left(x\right)}{\partial x}\right|_{x=x^a_{n-1}}, \tag{2.18}$$

so that

$$x^t_n - x^a_n = \boldsymbol{M}_{n-1}x^t_{n-1} - \boldsymbol{M}_{n-1}x^a_{n-1} + q_n = \mathcal{M}_{n-1}\left(x^t_{n-1} - x^a_{n-1}\right) + q_n \tag{2.19}$$

which is analogous to eq. (2.14) with the model operator replaced by its Jacobian.

A similar reasoning can be applied to the linearisation of the observation operator (Gelb, 1974), so that the Extended Kalman Filter is given by the same equations as in (2.16a), (2.16b), (2.16c) where the model and observation operators are replaced by their gradients when using the tangent linear hypothesis. In summary, the Extended Kalman Filter is given by the equations:

$$x_n^a = \boldsymbol{M}_{n-1} x_{n-1}^a + \boldsymbol{K}_n \left(y_n^o - \boldsymbol{H}\boldsymbol{M}_{n-1} x_{n-1}^a \right) \tag{2.20a}$$

$$\boldsymbol{K}_n = \boldsymbol{P}_n^f \mathcal{H}^{\mathrm{T}} \left(\mathcal{H}\boldsymbol{P}_n^f \mathcal{H} + \boldsymbol{R}_n \right)^{-1} \tag{2.20b}$$

$$\boldsymbol{P}_n^f = \mathcal{M}_{n-1} \boldsymbol{P}_{n-1}^a \mathcal{M}_{n-1}^{\mathrm{T}} + \boldsymbol{Q}_n \tag{2.20c}$$

$$\boldsymbol{P}_n^a = \left(\boldsymbol{I} - \boldsymbol{K}\mathcal{H} \right) \boldsymbol{P}^f \ . \tag{2.20d}$$

where

x : is the state variable,

$\boldsymbol{M}_i$: the model operator from time level $i-1$ to level i

$\boldsymbol{H}$: is the observational operator,

$\boldsymbol{K}$: is the Kalman gain matrix,

$\boldsymbol{Q}_i$: the dynamic noise covariance matrix generated by model errors,

$\boldsymbol{R}_i$: the observational noise covariance matrix generated by measurement and sampling errors.

2.2 The Singular Evolutive Extended Kalman Filter - *SEEK*

However, two major obstacles remain, that hinder the full implementation of the Extended Kalman Filter in oceanographic or atmospheric model. (two rare examples of successful attempts are given by Evensen, 1992; Miller, 1994).

First, the needed covariances $\boldsymbol{Q}_n$, $\boldsymbol{R}_n$ and $\boldsymbol{P}_0^a$ are not easily specified. With regard to the measurement errors, correlations between the measurements and the evolution of the error in time can often be neglected, so that $\boldsymbol{R}$ might be given as a diagonal matrix. The initial forecast error, i. e. the error statistics of the given initial condition x_0^f, is difficult to estimate, but might be approximated from a properly prepared statistical background. On the contrary, little can be said a priori on the evolution of the dynamic noise matrix $\boldsymbol{Q}_n$.

The second difficulty lies in the dimension of the system. In complex models the state variable has often very large dimensions, so that the full representation of the various covariance matrices becomes not realisable in terms of computation time and computer memory.

Pham et al. (1998) have proposed the Singular Evolutive Extended Kalman Filter (*SEEK).* This filter avoids the above limitations by approximating the analysis covariance in a low rank reduction, that shall be briefly described here.

In the derivation of this filter an alternative, equivalent formulation of the Kalman Filter equations are applied: from equations (2.20b) and (2.20d) follows, that if the observation error covariance $\boldsymbol{R}$ is invertible, the gain matrix can be expressed as

$$\boldsymbol{K}_i = \boldsymbol{P}_n^a \mathcal{H}_n^T \boldsymbol{R}_n^{-1} \,, \tag{2.21}$$

so that the equations of the Extended Kalman Filter can be rewritten:

$$\boldsymbol{P}^a = \mathcal{M}_{n-1}\boldsymbol{P}^a_{n-1}\mathcal{M}_{n-1} + \boldsymbol{Q}_n - \left(\mathcal{M}_{n-1}\boldsymbol{P}^a_{n-1}\mathcal{M}^{\mathrm{T}}_{n-1} + \boldsymbol{Q}_n\right)\mathcal{H}^{\mathrm{T}}_n \tag{2.22a}$$
$$* \left[\mathcal{H}_n\left(\mathcal{M}_{n-1}\boldsymbol{P}^a_{n-1}\mathcal{M}^{\mathrm{T}}_{n-1} + \boldsymbol{Q}_n\right)\mathcal{H}^{\mathrm{T}}_n + \boldsymbol{R}_n\right]^{-1}\mathcal{H}_n\left(\mathcal{M}_{n-1}\boldsymbol{P}_{n-1}\mathcal{M}_{n-1} + \boldsymbol{Q}_n\right)$$
$$x^a_n = \boldsymbol{M}_{n-1}x^a_{n-1} + \boldsymbol{K}_n\left[y^o_n - \boldsymbol{H}_n\boldsymbol{M}_{n-1}x^a_{n-1}\right] \quad . \tag{2.22b}$$

The basic idea of the *SEEK*-Filter is the reduction of the rank of the analysis covariance matrix $\boldsymbol{P}_a$ by expressing it in terms of a few main directions of error propagation $\boldsymbol{L}$. The reduced rank analysis covariance matrix $\boldsymbol{U}_i$ is then given by

$$\boldsymbol{P}^a = \boldsymbol{L}\boldsymbol{U}\boldsymbol{L}^T \, . \tag{2.23}$$

This is motivated by the fact that in dynamic systems, that are described by an attractor (and the ocean and the atmosphere can be seen as such a system with their dynamics, dominated by external forcing and dissipation), only a small part of the phase space is passed by the state trajectories. Consequently, as correction directions chosen are those directions, where errors are less attenuated by the attractor itself. These directions lie in the linear sub space tangent to the attractor and thus are given by the eigenvectors of the tangent linear model $\mathcal{M}$. A limited number of the first eigenvectors will then be sufficient to reduce the error of the system significantly.

Moreover, it is shown in Pham et al. (1998), that if the error propagation directions $\boldsymbol{L}_n$ evolve according to

$$\boldsymbol{L}_n = \mathcal{M}_{n-1}\boldsymbol{L}_{n-1} \tag{2.24}$$

and the analysis error covariance matrix projected into the corresponding subspace evolves according to

$$\boldsymbol{U}_i^{-1} = \left[\boldsymbol{U}_{i-1} + \left(\boldsymbol{L}_i^T\boldsymbol{L}_i\right)^{-1}\boldsymbol{L}_i^T\boldsymbol{Q}_i\boldsymbol{L}_i\left(\boldsymbol{L}_i^T\boldsymbol{L}_i\right)^{-1}\right]^{-1} + \boldsymbol{L}_i^T\mathcal{H}_i^T\boldsymbol{R}_i^{-1}\mathcal{H}_i\boldsymbol{L}_i \, , \tag{2.25}$$

then the rank of the approximated analysis covariance matrix is maintained, and the subspace spun by the basis $\boldsymbol{L}$ converges to the subspace, that is associated to the first eigenvalues of $\mathcal{M}_{n-1}$„ and the filter is generally stable. (This holds under the given assumptions, i. e. most of all the linearity of $\boldsymbol{M}$. Note, however, that the stability is independent of the specification of the error covariances $\boldsymbol{R}$ and $\boldsymbol{Q}$ and the chosen initial conditions. The speed of convergence will largely depend on it though.)

An assimilation cycle is then given by the steps:

- Forecast: $x_n^f = \boldsymbol{M}_{n-1} x_{n-1}^a$
- Propagation of the correction directions $\boldsymbol{L}$
- Update of the reduced analysis error covariance $\boldsymbol{U}$
- Correction: $= x_n^f + \boldsymbol{K}_n \left[y_n^o - \boldsymbol{H}_n x_n^f \right]$

2.2.1 Versions of the *SEEK* Filter

The *SEEK* Filter still requires the specification of the dynamic noise covariance on each assimilation step. Pham et al. (1998) already proposed an alternative and more viable version of the filter that avoids the unknown dynamic noise covariance. From this simplified formulation another version evolved, that skips the propagation of the error correction directions. These two sub-version are described here, as they have been used for the numerical experiments.

2.2.1.1 Version with forgetting factor (SEEK-FF)

In this version the dynamic noise covariance in eq. (2.25) is specified as

$$\alpha \boldsymbol{U}_{i-1} = \left(\boldsymbol{L}_i^T \boldsymbol{L}_i \right)^{-1} \boldsymbol{L}_i^T \boldsymbol{Q}_i \boldsymbol{L}_i \left(\boldsymbol{L}_i^T \boldsymbol{L}_i \right)^{-1} \tag{2.26}$$

with $\alpha > 0$. This was proposed mainly for simplicity in the resulting formula to update $\boldsymbol{U}$. However, it was motivated as an amplification to

take into account the increasing error due to the linearisation (Pham et al., 1998).

The error propagation directions are, as before, updated at each assimilation step according to:

$$\boldsymbol{L}_i = \mathcal{M}_{i-1}\boldsymbol{L}_{i-1} \, . \tag{2.27}$$

The covariance matrix in the reduced space is then updated according to

$$\boldsymbol{U}_i^{-1} = \rho\boldsymbol{U}_{i-1}^{-1} + \boldsymbol{L}_i^T\mathcal{H}_i^T\boldsymbol{R}_i^{-1}\mathcal{H}_i\boldsymbol{L}_i \, , \tag{2.28}$$

where $\rho = \frac{1}{1+\alpha} \, \epsilon \, (0; 1]$ is called forgetting factor, as it can be interpreted as a down-weighting of previous errors in the iterative process.

2.2.1.2 Semi-static version (SEEK-0.5)

Here, the error propagation directions are kept constant in time assuming persistence of the error sub-space:

$$\boldsymbol{L}_i = \boldsymbol{L}_0 \tag{2.29}$$

and the covariance matrix is updated according to eq. 2.28:

$$\boldsymbol{U}_i^{-1} = \rho\boldsymbol{U}_{i-1}^{-1} + \boldsymbol{L}^T\mathcal{H}_i^T\boldsymbol{R}_i^{-1}\mathcal{H}_i\boldsymbol{L} \, . \tag{2.30}$$

This version was introduced by Brasseur et al. (1999) in the context of a circulation model merely as a test case for comparison with the more sophisticated filters. However, since then, it has widely used in dynamic ecosystem models thanks to its simplicity and the low computational effort (e. g. Hoteit et al., 2003, 2004; Magri et al., 2005; Raick et al., 2007).

2.3 Implementation

2.3.1 Model

The data assimilation filters were applied to the same 1D dynamic ecosystem model of section 1.2.

As coupling method the POMStep technique described in section 1.3.8 was applied. Computations were performed at a synchronous time step of $864s$.

The forcing functions for the perpetual year simulations used are described in 1.3.4.

The experiments were performed at sites AA1 and S3 of fig. 1.9, described in section 1.3.1 with the same model parametrisation given in tables 1.2 and 1.3.

2.3.2 Implemented filters

To validate the quality of the results obtained for the two described filters SEEK-FF and SEEK-0.5 with evolving statistics also a filter without any propagation was implemented. This scheme corresponds to an Optimal Interpolation as described in section 2.1.1, but is expressed in terms of the analysis covariance instead of the background covariance. This should be seen as a pure test aimed to estimate the benefit of the propagation techniques used. For better comparison the error covariance matrix was also reduced to the low rank representation:

$$\boldsymbol{L}_n = \boldsymbol{L}_0 \tag{2.31}$$

$$\boldsymbol{U}_n = \boldsymbol{U}_0 \tag{2.32}$$

$$\boldsymbol{P}_n^a = \boldsymbol{L}_0 \boldsymbol{U}_0 \boldsymbol{L}_0^{\mathrm{T}} \ . \tag{2.33}$$

Hence, the filter was implemented in the following versions:

- SEEK-FF (section 2.2.1.1)

- SEEK-0.5 (see section 2.2.1.2)
- OI-SEEK: Optimal Interpolation scheme in *SEEK* formulation.

In addition, a new subversion of the *SEEK* Filter (named SEEK-Q) was developed, implemented and tested in the frame of this work. Details of the filter characteristics and the results obtained are given in sections 2.6 and 2.7

2.3.3 TWIN experiment preparation

In this work the strategy of TWIN experiments (Daley, 1991) was chosen to validate the analysis results, that where applied. This approach is often used in data assimilation to calibrate an implementation before applications with real data, where the true state is not known. In these experiments a reference run is performed to provide both "pseudo measurements", and the "true state" of the system, that is then compared to a free model run (without assimilation) and to an assimilation run, both started from initial conditions adequately perturbed with respect to those used for the reference run.

In this study the model was initialised for winter conditions and spun up for 10 years. The reference run was then initialised from the final state of the spin-up run, while the free run was carried out in the same time frame and forcing, but with initial profiles from summer conditions in the biogeochemical state variables. The same initial conditions were applied then to the assimilation run.

2.3.4 Initialisation of the filter

While the *SEEK* Filter generally corrects for bad initialisation due to the convergence of the error propagation directions $\boldsymbol{L}$ (Pham et al., 1998), this is not valid for the version SEEK-0.5 as it is not propagating these directions. Also the speed of convergence is increased by a proper statistical preparation, so special care was taken in the preprocessing step.

It was chosen to obtain the initial analysis error covariance matrix from an EOF (empirical orthogonal functions) analysis of a series of ensemble runs of the model. A series of 50 Ensemble Runs was carried out with perturbed initial conditions $c_{pert}(z,0)$, that where obtained by applying vertically smooth random oscillations $r(z)$ (as in Evensen, 2003 for horizontal fields) with zero mean and unit variance to a 5% fraction of the unperturbed initial condition:

$$c_{pert}(z,0) = c(z,0) + 0.05r(z)c(z,0)\ . \tag{2.34}$$

Each ensemble member was run over a full season of $90d$ and the process was repeated for all four seasons storing the states in two day intervals, resulting in a data set of 9000 column vectors for each biogeochemical state variable. From the data the solution of the perpetual climatological annual cycle was subtracted and remaining biases removed. The data was then normalised as

$$\delta\phi_{n,k} = \frac{\Delta z}{h}\frac{\delta c_{n,k}(t)}{\overline{\sigma_n}} \tag{2.35}$$

with k indicating the vertical level, σ_n as the mean column variance of state n and $\frac{\Delta z}{h}$ being a geometric factor (Dobricic et al., 2005) applied, to avoid an over-stress of information contained in the surface and bottom layer (where vertical levels are very thin due to the logarithmic distribution of grid points, see 1.3.2) in the computations of the error correction directions. The data $\delta\phi_{n,k}$ is then fed into a singular value decomposition algorithm to obtain an appropriate sub-set of eigenvalues λ_l and the connected eigenvectors V_l of the data covariance matrix, used as the initial conditions for $\boldsymbol{L}_0$ and $\boldsymbol{U}_0$.

In fact, the maximum variability contained in an approximation of rank r of the covariance matrix $\boldsymbol{P}_0^a$, originally of rank m, is obtained by a projection of the matrix into the sub space spanned by its first r eigenvectors and the percentage of variability explained by this approximation is given

by

$$\frac{\sum_{l=1}^{r} \lambda_l}{Tr\left(\boldsymbol{P}_0^a\right)} = \frac{\sum_{l=1}^{r} \lambda_l}{\sum_{l=1}^{m} \lambda_l} \,. \tag{2.36}$$

Figures 2.3 and 2.4 give a graphical representation of the percentages of variability contained in the first 30 EOFs for the covariances resulting form the Ensemble Runs in site AA1 and Site S3. It can be seen, that in both cases more than 80% of the variability is explained by the first 15 EOFs. This was considered to give a sufficiently good approximation for the simulations and therefore used in all experiments (see also Brasseur and Verron, 2006).

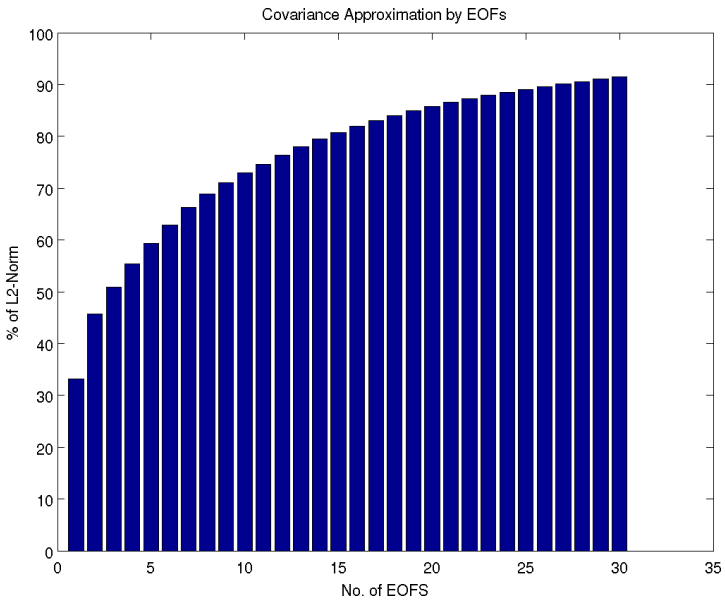


Figure 2.3: Variability [%] contained in EOFs (accumulative). Site AA1.

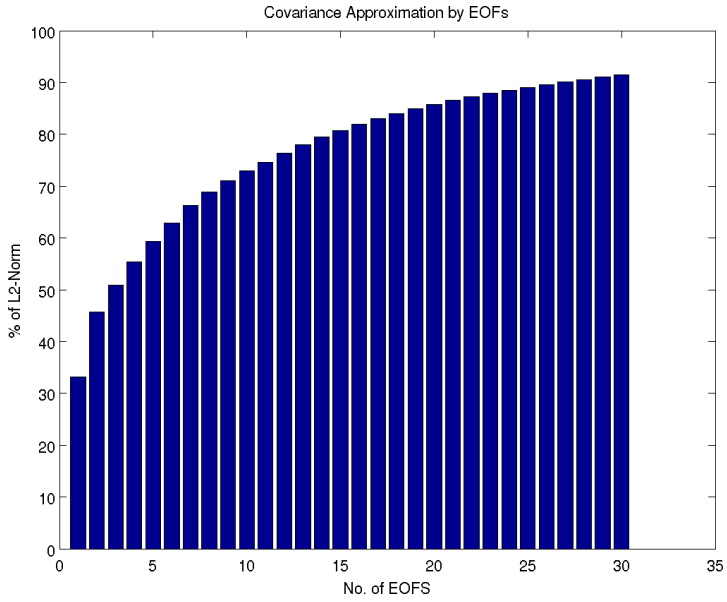


Figure 2.4: Variability [%] contained in EOFs (accumulative). Site S3.

OI-SEEK	Observations	Location	-	Δt_{ass}
O2-AA1	O2	AA1	-	2d
O2-S3	O2	S3	-	2d
SEEK-0.5	Observations	Location	Forgetting Factor ρ	Δt_{ass}
O1-AA1	O1	AA1	0.9	1d
O2-AA1	O2	AA1	0.85	2d
O2-S3	O2	S3	0.9	2d
SEEK-FF	Observations	Location	Forgetting Factor ρ	Δt_{ass}
O2-AA1	O2	AA1	0.6	2d
O2-S3	O2	S3	0.6	2d
SEEK-Q	Observations	Location	NRCoefficient β	Δt_{ass}
O2-AA1	O2	AA1	0.5	2d
O2-S3	O2	S3	0.5	2d

Table 2.1: Overview of experiments carried out with various filters. O1 and O2 refer to the observation vector as given in table 2.2; ρ is the forgetting factor as in eq. (2.28), β is the noise reduction coefficient as in eq. (2.37) and Δt_{ass} is the assimilation interval.

2.3.5 Experiment overview

In all the experiments performed the full BFM state vector was used, the observation covariance was assumed to be not time dependent and diagonal (assuming no correlation between the measurements) and the variance vector of the observations was taken proportional to the vector obtained by the application of the observation operator $\boldsymbol{H}$ on the diagonal elements of the ensemble error covariance. Two different sets of data (constituting the observational vector) were extracted from the reference run (see table 2.2):

- Observational vector 1 (O1): constituted by 1 nitrate and 2 oxygen observations relative to the depth given in table 2.2.

- Observational vector 2 (O2): constituted by oxygen and the full nutrient pool (Phosphate - N1; nitrate - N3, ammonium - N4, silicate

observation vector 1 (O1)																
y^o	1			2			3									
state	N3			O			O									
$d[m]$ at AA1	2			5			18									
observation vector 2 (O2)																
y^o	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
state	O			N1			N3			N4			N5			Chl
$d[m]$ at AA1	0	9.5	13.4	0	9.5	13.4	0	9.5	13.4	0	9.5	13.4	0	9.5	13.4	0
$d[m]$ at S3	0	14.5	29.3	0	14.5	29.3	0	14.5	29.3	0	14.5	29.3	0	14.5	29.3	0

Table 2.2: The two different observation vectors y^o applied in the experiments. O=oxygen, N1=Phosphate, N3=Nitrate, N4=Ammonium, N5=Silicate, Chl=Chlorophyll. d: depth level of observation

- N5) at the depth given in table 2.2. In addition also the surface chlorophyll concentration value was considered.

It follows, that the observational operator $\boldsymbol{H}$ is linear in both cases, as all observations directly correspond to elements of the model state vector or a simple sum of them (chlorophyll total concentration arises from the sum of the chlorophyll content of the three phytoplankton functional groups resolved by the model (see section 1.2.2 and table 1.1). The rationale behind these choices is given in the sections pertinent to each experiment.

A listing of the experiments performed is given in table 2.1. In defining the experiments it was attempted to reach three objectives

1. Test the implementation (section 2.4).
2. Compare the various versions of the filter to evaluate the role of the propagation schemes (section 2.5).
3. Elaborate, implement and validate a new subversion of *SEEK* (section 2.7).

The assimilation interval was 2 days in all experiments, with the only exception of experiment OI-AA1-SEEK0.5, where the assimilation interval

was chosen to be 1 day in order to allow comparison with the previous works based on *SEEK*.

Experiment results will be presented in two formats

- Hovmoeller diagrams (e. g. fig. 2.8), showing the time evolution of the water column to give an overview of the state variable as a whole. The field of the reference solution will be given at the bottom of each figure, the differences of the free run with respect to the reference solution, i. e. $x^{free} - x^t$, will be given on top, and in between the differences of the respective analysis states of the various filters $p^a = x^a - x^t$ are given. These plots are adequate to get a general overview of the filter fields, but are not accurate enough for a proper analysis.

- Time series (e.g. fig. 2.6), that give the evolution in time of the respective state variable at three distinct levels of the water column. The three levels are chosen in order to represent the different layers of the dynamical structures and are placed off the "measurement" points. In these plots the reference solution can be found in the left column, while in the right column errors of the free run and the various filters with respect to the true state are superimposed in one plot. The lines of the assimilation error plots are constructed connecting the analysis error $p^a_{n-1} = x^a_{n-1} - x^t_{n-1}$ to the forecast error $p^f_n = x^f_n - x^t_n$ which is then connected to the new analysis error p^a_n yielding a representation of the assimilation cycle, that allows the separate identification of the dynamics of the error induced by the model forecast and induced by the correction step of the filter, as illustrated in figure 2.5.

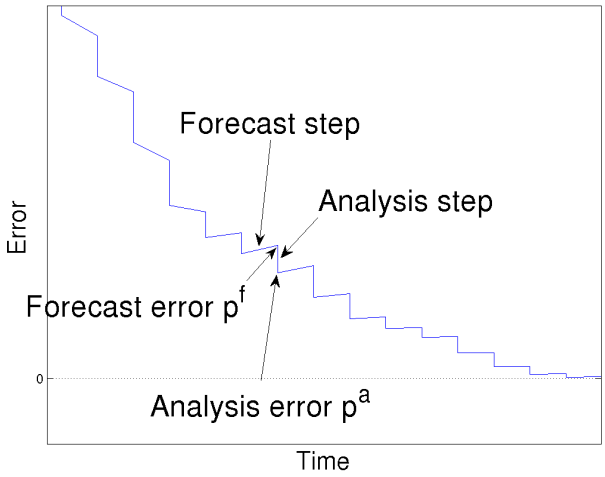


Figure 2.5: Forecast step and analysis step of an assimilation cycle in time series plots.

2.4 Implementation test

As first implementation test an experiment was carried out that adopts the set-up previously applied by Hoteit et al. (2003) for the implementation of the *SEEK* Filter in a 1D marine ecosystem dynamics model of the Cretan Sea. The observation vector O1 chosen (2.2) has the same characteristics of the observation vector used in that implementation and the filter used was SEEK-0.5. The experiment was set up in site AA1. (The parametrisation of the assimilation is given in table 2.1.)

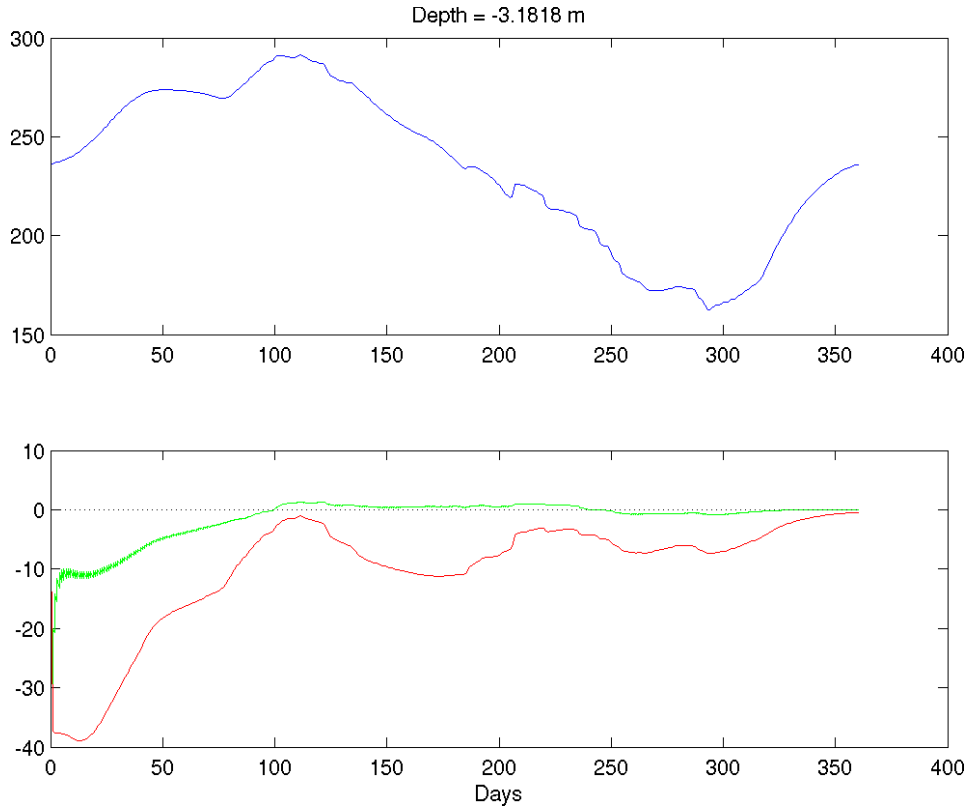


Figure 2.6: Dissolved oxygen $\left[\frac{mmol}{m^3}\right]$ at $-3.2m$, O1-AA1-SEEK-0.5 experiment. *Top:* reference run; *Bottom:* free model run and analysis errors $x - x^t$. *Red:* free model run; *Green:* SEEK-0.5

In figures 2.6 and 2.7 the time series of the two observed variables (oxy-

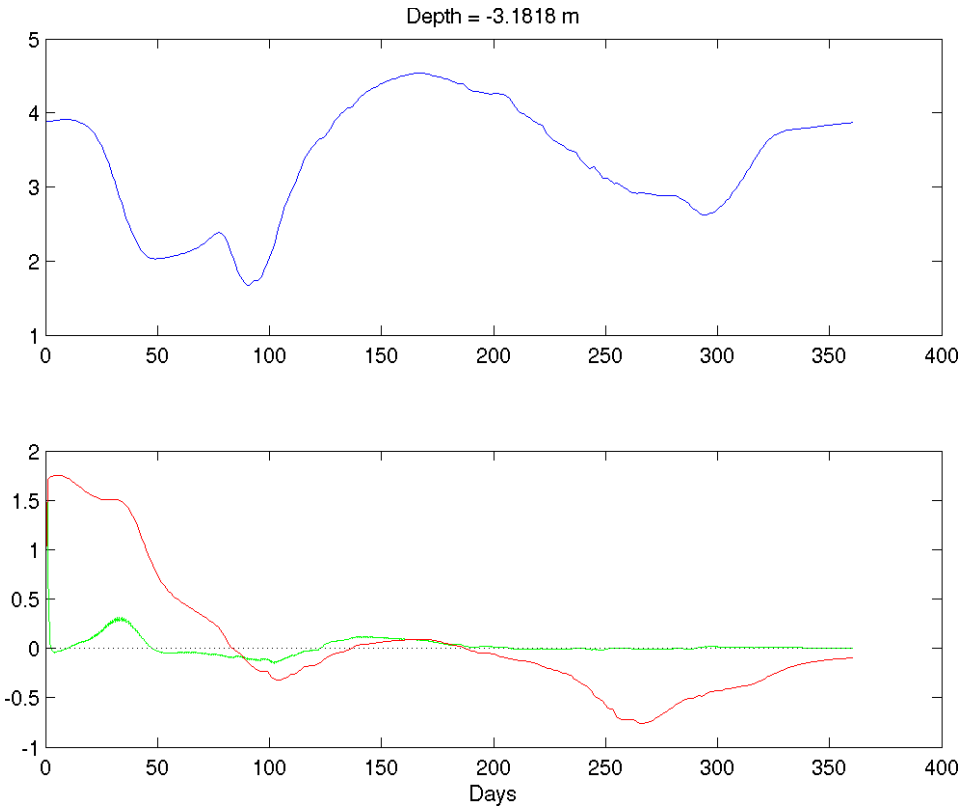


Figure 2.7: Nitrate $\left[\frac{mmol}{m^3}\right]$ at $-3.2m$, O1-AA1-SEEK-0.5 experiment. *Top:* reference run; *Bottom:* free model run and analysis errors $x-x^t$. *Red:* free model run; *Green:* SEEK-0.5

gen and nitrate) are shown at a depth of $3.2m$, a level that is strongly influenced by the spring bloom dynamics in this site. The information of the observation is well passed to the system, oxygen reaches the level of the reference solution within roughly $100d$, while nitrate is even quicker ($50d$). The free run on the contrary approaches in both cases the reference run just by the end of the simulation ($1y$).

Figures 2.8 and 2.9 show the water column annual cycle for 2 non observed, and key states of the ecosystem: chlorophyll and phosphate. A general better performance of the assimilation run (analysis) with respect to the free model run can be noted, though differences are rather small

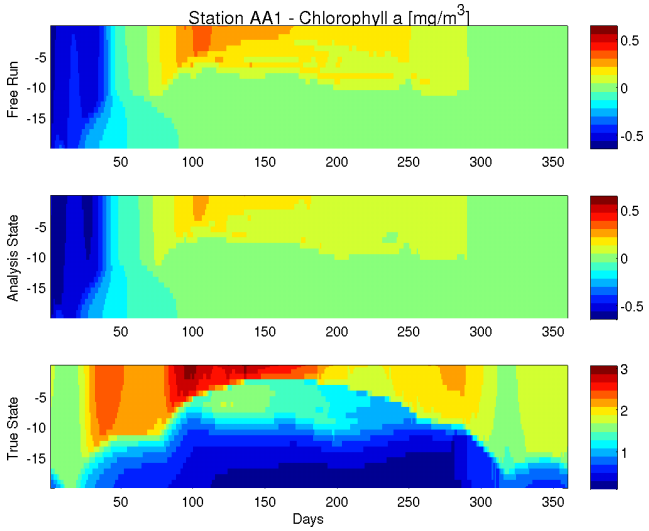


Figure 2.8: Chlorophyll-a content $\left[\frac{mg}{m^3}\right]$ for the O1-AA1-SEEK-0.5 experiment. *Bottom:* Reference Solution; *Middle:* Analysis error p^a; *Top:* free model run error $x^{free} - x^t$.

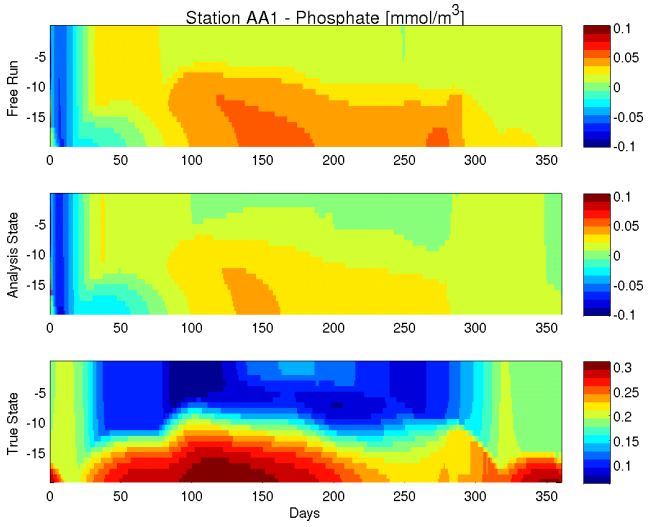


Figure 2.9: Phosphate $\left[\frac{mmol}{m^3}\right]$ field for the O1-AA1-SEEK-0.5 experiment. *Bottom:* Reference Solution; *Middle:* Analysis error p^a; *Top:* free model run error $x^{free} - x^t$.

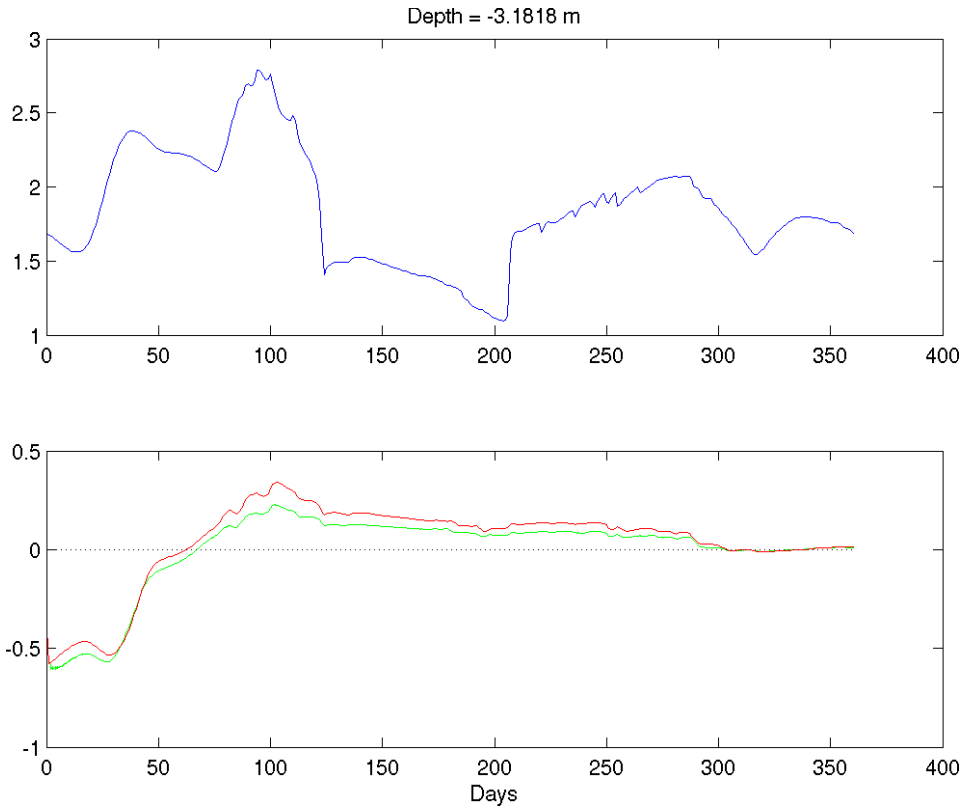


Figure 2.10: Chlorophyll-a content $\left[\frac{mg}{m^3}\right]$ at $-3.2m$, O1-AA1-SEEK-0.5 experiment. *Top:* reference run; *Bottom:* free model run and analysis errors $x - x^t$. *Red:* free model run; *Green:* SEEK-0.5

and in the first few assimilation steps, the analysis is even slightly worse than the free run. The time series for chlorophyll and phosphate at $3.2m$ depth (2.10 and 2.11 respectively) allow for a more detailed evaluation of the SEEK-0.5 performance.

The filter solution for chlorophyll (given in figure 2.10) is still closer to the reference, than the free run solution, but advantages are small and mainly a consequence of the model dynamics adapting to the corrected nitrate and oxygen levels, than a direct correction of chlorophyll by the filter. The reference solution is reached simultaneously by free run and filter solution by the end of the year. This situation repeats for the Phosphate

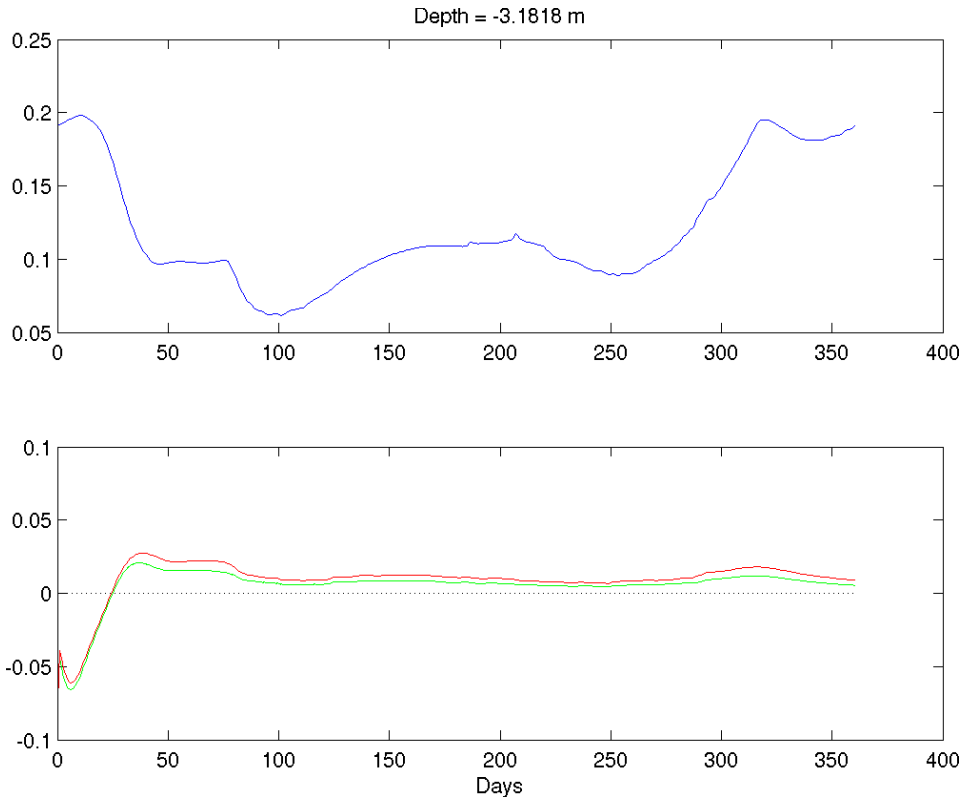


Figure 2.11: Phosphate $\left[\frac{mmol}{m^3}\right]$ at $-3.2m$, O1-AA1-SEEK-0.5 experiment. *Top:* reference run; *Bottom:* free model run and analysis errors $x - x^t$. *Red:* free model run; *Green:* SEEK-0.5

state variable (fig. 2.11), that even in the end of the simulation shows a small difference to the "true state".

Overall, the analysis quality is comparable to that obtained by (Hoteit et al., 2003). The correction of the non observed variables is probably working a little better in their case, the adaptation of the nitrate solution on the contrary is much more efficient in this case, as in the case of the Cretan Sea it did not reach the reference run even by the end of the year. Therefore, it might be concluded that the filter is implemented correctly and working properly though not performing efficiently.

As for the deficiencies in the non observed variables, a reason for the

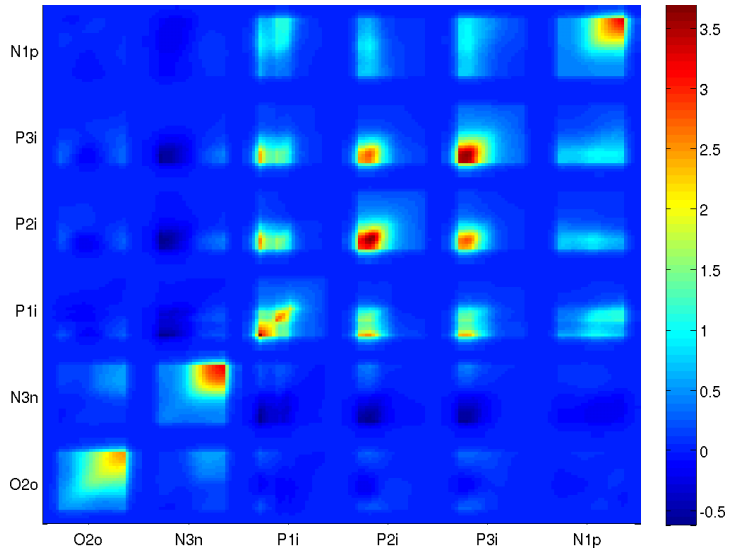


Figure 2.12: Section of the correlation matrix $\frac{\delta\phi * \delta\phi^{\mathrm{T}}}{N-1}$. State variables shown are: $O2o$ - Dissolved oxygen, $N3n$ - Nitrate, $P1i$ - Diatom chlorophyll-a content, $P2i$ - Nanoflagellates, $P3i$- Picophytoplankton, $N1p$- Phosphate. The matrix is organised by state variable blocks with the surface data for each state couple starting in the bottom left corner and depth increases to right and top. The data are normalised in a way that the mean of the main diagonal section taken for each state (the autocorrelations) is one, to guarantee comparable magnitudes between the various states, and maintaining variability distribution along the water column (see 2.3.4).

slow adaption rate might be, that few information of the observed states can be transferred on to the other states. A view on the error statistics performed to initialise the filter confirms few correlation especially of oxygen to the other state variables considered here. Figure 2.12 shows a section of the correlation matrix of the ensemble data (section 2.3.4) in site AA1. The variables shown in the section are those included in the observational vector (oxygen and nitrate) and the other two variables discussed in this section (chlorophyll in the three functional phytoplankton groups and phosphate). It can be seen, that oxygen is very weakly correlated to

the other state variables, so that few of its information can be used by the filter. The situation is merely better for the nitrate observations, in particular phosphate shows no correlation with the observed states at all, while on the contrary chlorophyll and phosphate, the two not observed states illustrated in this example, show reasonable correlations at the surface.

Thus, it can be concluded that the system is not well represented by the observational vector and this explains the weak performance of the filter. Therefore, it was decided to enlarge the range of pseudo-measurements on other state variables, i. e. increase the dimension of the observation vector.

2.5 Filter comparison

A closer look on the covariance maps of the ensemble statistics (Appendix A gives an overview of the covariance fields for the main functional groups) identifies phosphate as the key element in the model error dynamics in both sites as expected, given the overall character of phosphorus limitation of the northern Adriatic Sea ecosystem (Zavatarelli et al., 1998). Correlations to the surface values of all phytoplankton groups, microzooplankton and bacteria can be detected in both implementation sites over the whole water column. Also other nutrients show some correlation and anti-correlation patterns in the functional groups. Particularly strong is the surface correlation of the three phytoplankton groups to Microzooplankton. This is a direct consequence of the *BFM* trophic web since all the three phytoplankton functional groups are predated by microzooplankton.

Based on these considerations and the possible availability of "real" measurements it was decided to consider an observation vector (O2) that contains measurements of the four nutrient classes (Phosphate, Nitrate, Ammonium, Silicate) at three levels (surface, $-9.5m$, $-13.4m$ for site AA1; surface, $-14.5m$, $-29.3m$ for site S3).

Figures 2.13 and 2.14 show a comparison of the chlorophyll and phosphate field for the three filters OI-SEEK, SEEK-0.5 and SEEK-FF in the experiment session O2-AA1. The solution improves with the sophistication of the filter, especially the gap between the Optimal Interpolation scheme and the two SEEK Filters is significant. A detailed look on the time evolution at three distinct depth levels (fig. 2.15 and 2.16) shows, that in the lower level all the filters have some inertia before starting to work properly, in particular in the case of chlorophyll. During the initial 20 days they seem to merely follow the model dynamics. Beyond day 20 the fully propagating filter (SEEK-FF) separates from the free run and approaches the reference, while the other two filters don't show a significant improvement with respect to the free run at this level. For Phosphate the first days

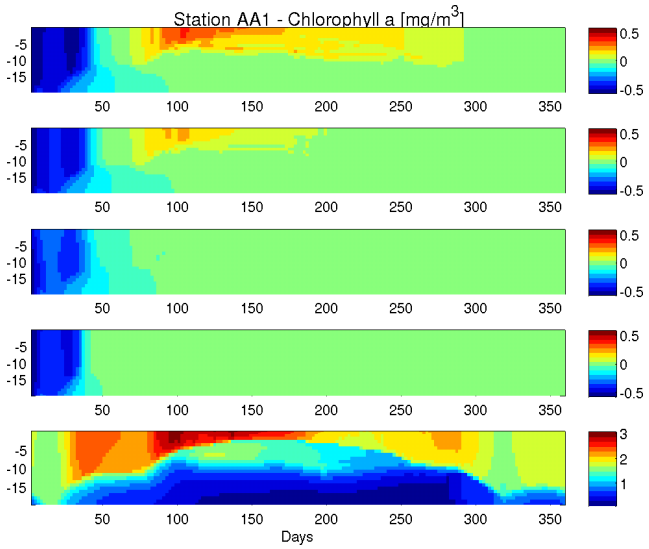


Figure 2.13: Chlorophyll-a content $\left[\frac{mg}{m^3}\right]$ for the O2-AA1 experiment. *From bottom to top:* Reference Solution, Analysis error p^a of SEEK-FF, SEEK-0.5 and OI-SEEK, free model run error $x^{free} - x^t$.

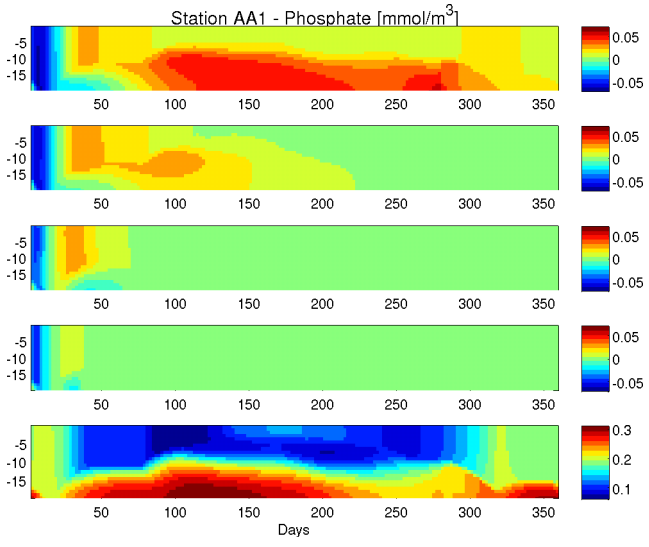


Figure 2.14: Phosphate $\left[\frac{mmol}{m^3}\right]$ field for the O2-AA1 experiment. *From bottom to top:* Reference Solution, Analysis error p^a of SEEK-FF, SEEK-0.5 and OI-SEEK, free model run error $x^{free} - x^t$.

of simulation are dominated by a strong oscillation in the model dynamics that is somewhat anticipated by the two SEEK Filters, after which they start to converge quickly towards the reference solution within 100 days, again with some advantage for the SEEK-FF version. The OI-SEEK on the contrary reaches the reference solution after 180 days. The three other nutrients (Nitrate, Ammonium and Silicate, fig. 2.17 , 2.18 and 2.19 respectively) and oxygen (fig. 2.20) show a quite similar behaviour, as they approach the "true state" of the reference run within hundred days, while the OI-SEEK solution essentially replicates the structure of the dynamics of the free run error in a strongly damped form.

For the other non observed variables, represented here by bacteria, particulate organic carbon and microzooplankton (2.21, 2.22 and 2.23 respectively), the three filters show rather different behaviours. While the OI-SEEK solution basically follows the free run with some improvement of the solution in the upper part of the water column, the SEEK-FF Filter starts very quickly to correct the field showing some overshooting (this is particularly visible in the bacteria solution in fig. 2.21), that is quickly recovered, generating an oscillating behaviour. The SEEK-0.5 Filter on the contrary converges slower but with a smoother fashion. The solution of Microzooplankton exhibits a particular difficulty of the filters with respect to the initial dynamics, as all three filter errors exceed the error of the free run. Again, the SEEK-FF Filter shows a clear advantage over the other two, recovering quickly from this strong initial oscillation towards the reference run, also here in an oscillating manner.

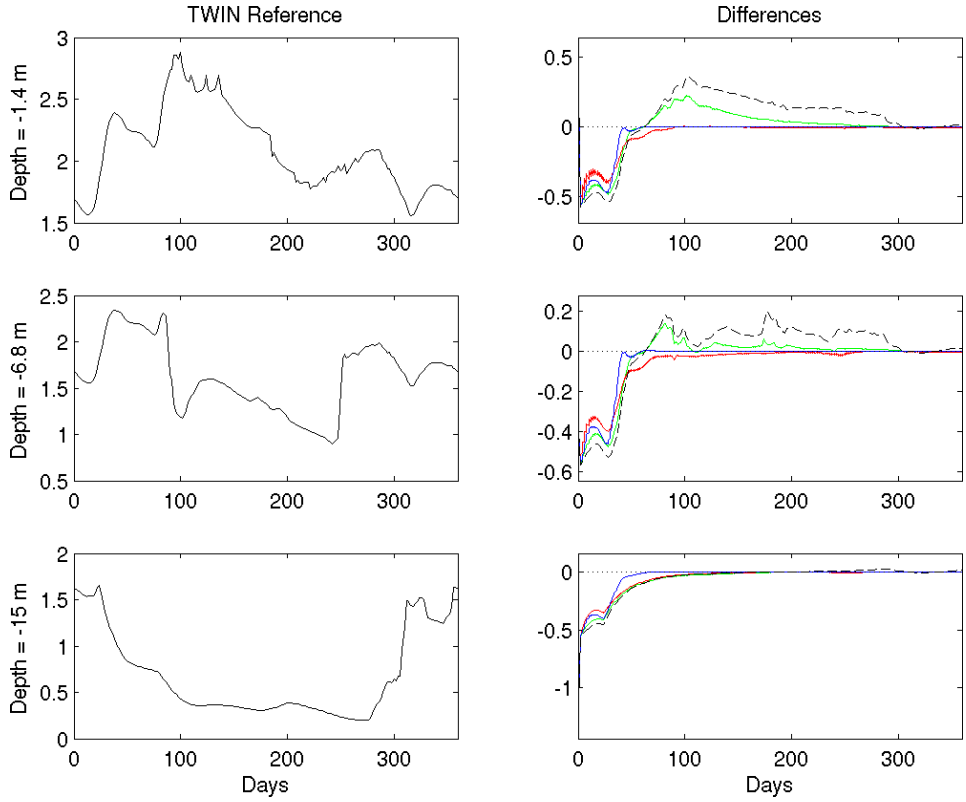


Figure 2.15: Chlorophyll-a content $\left[\frac{mg}{m^3}\right]$, O2-AA1 experiment at the depths of $1.4m$, $6.8m$ and $15m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* OI-SEEK; *Red line:* SEEK-0.5; *Blue line:* SEEK-FF.

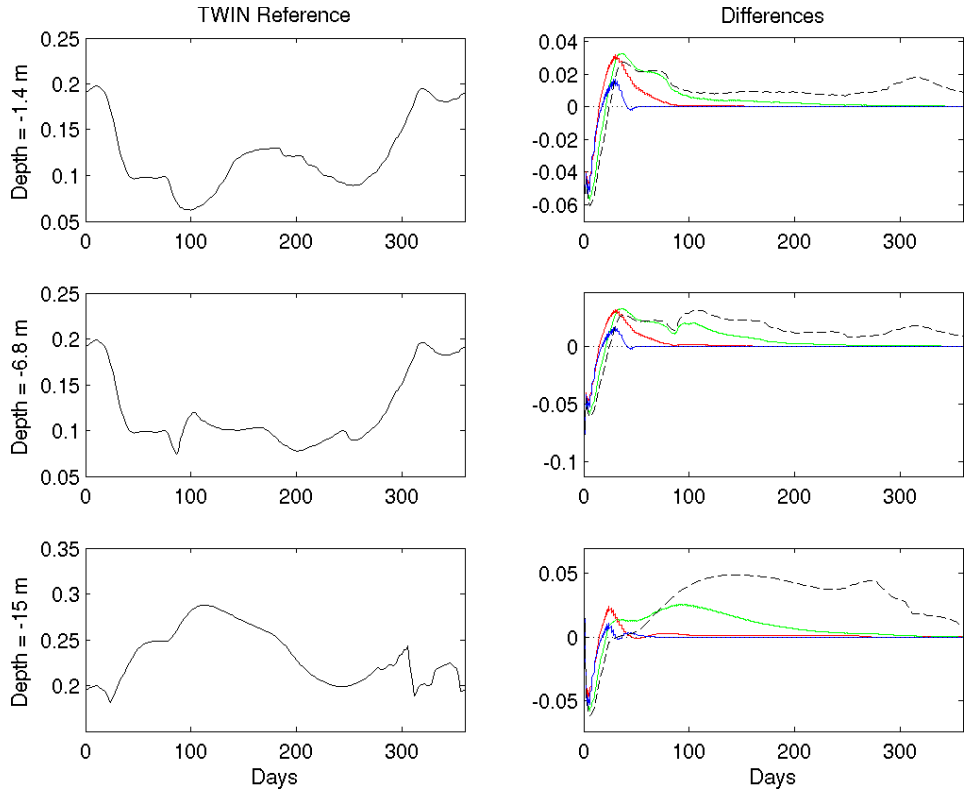


Figure 2.16: Phosphate $\left[\frac{mmol}{m^3}\right]$, O2-AA1 experiment at the depths of $1.4m$, $6.8m$ and $15m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* OI-SEEK; *Red line:* SEEK-0.5; *Blue line:* SEEK-FF.

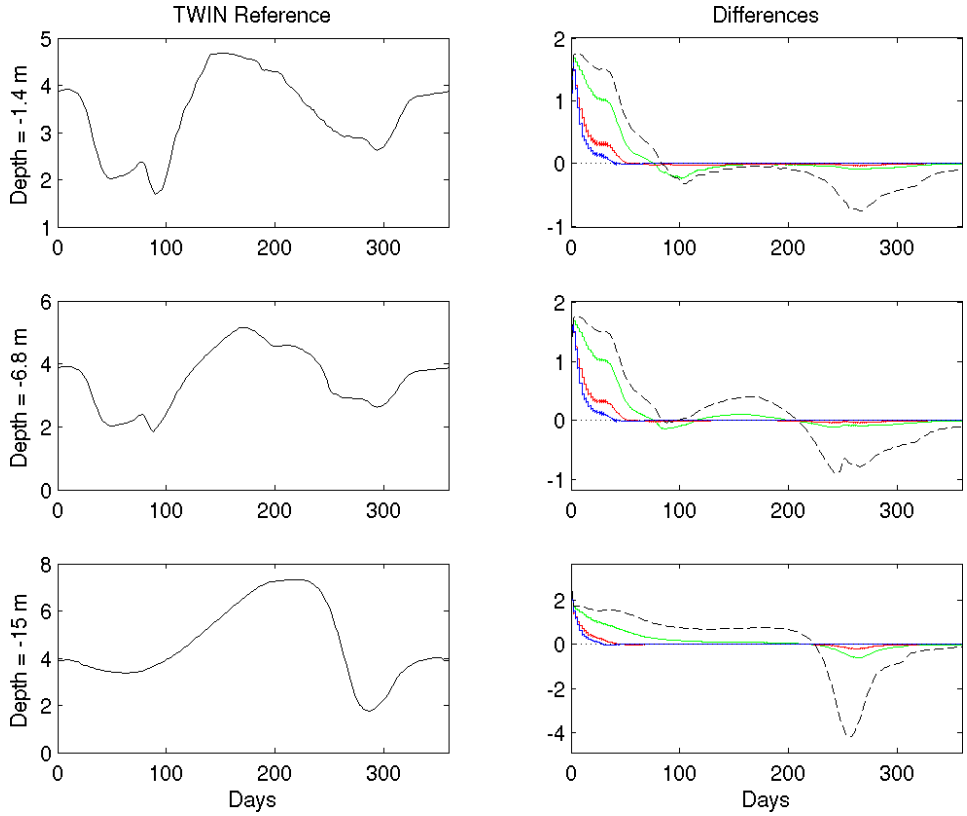


Figure 2.17: Nitrate $\left[\frac{mmol}{m^3}\right]$, O2-AA1 experiment at the depths of $1.4m$, $6.8m$ and $15m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* OI-SEEK; *Red line:* SEEK-0.5; *Blue line:* SEEK-FF.

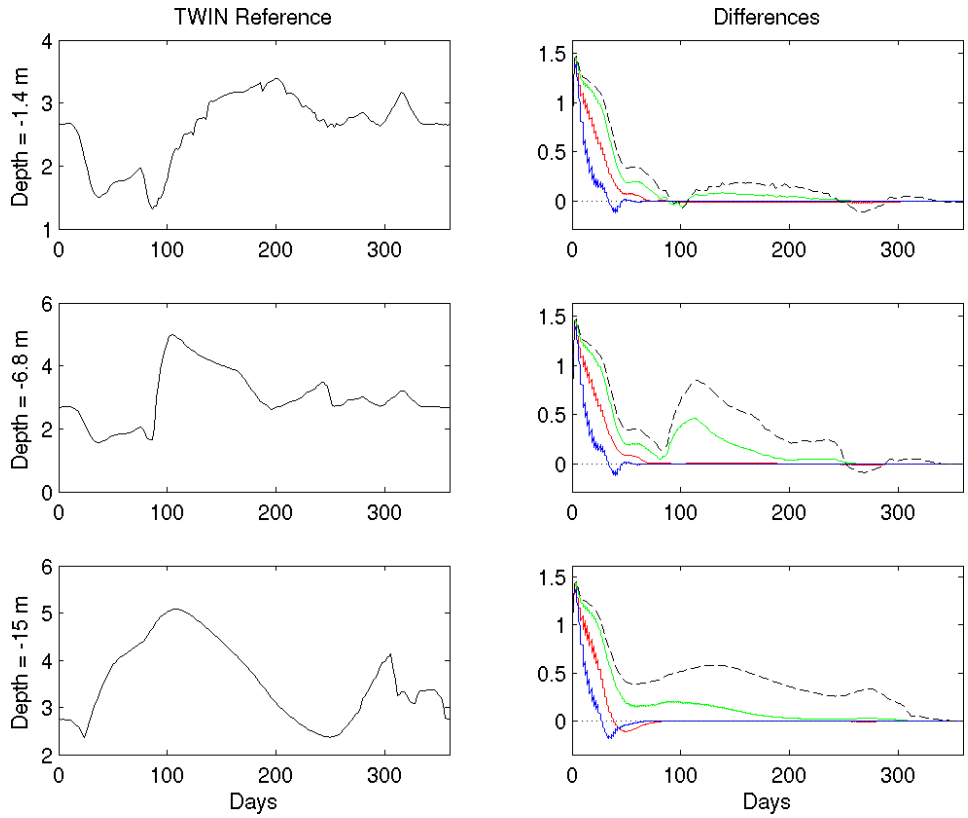


Figure 2.18: Ammonium $\left[\frac{mmol}{m^3}\right]$, O2-AA1 experiment at the depths of $1.4m$, $6.8m$ and $15m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* OI-SEEK; *Red line:* SEEK-0.5; *Blue line:* SEEK-FF.

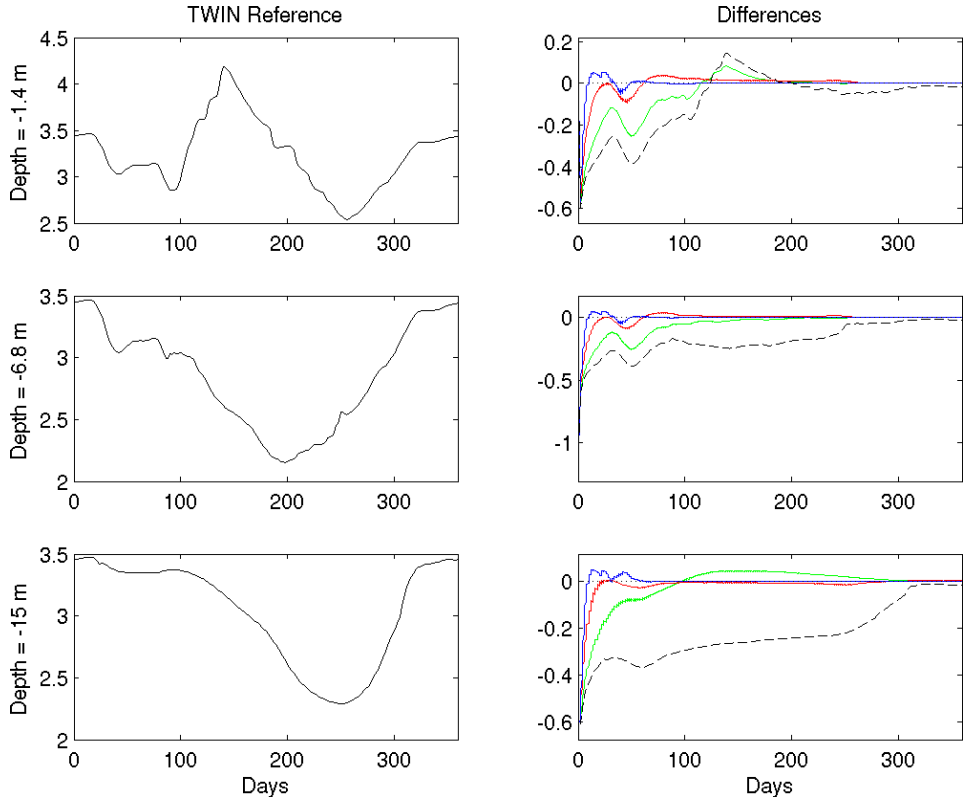


Figure 2.19: Silicate $\left[\frac{mmol}{m^3}\right]$, O2-AA1 experiment at the depths of $1.4m$, $6.8m$ and $15m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* OI-SEEK; *Red line:* SEEK-0.5; *Blue line:* SEEK-FF.

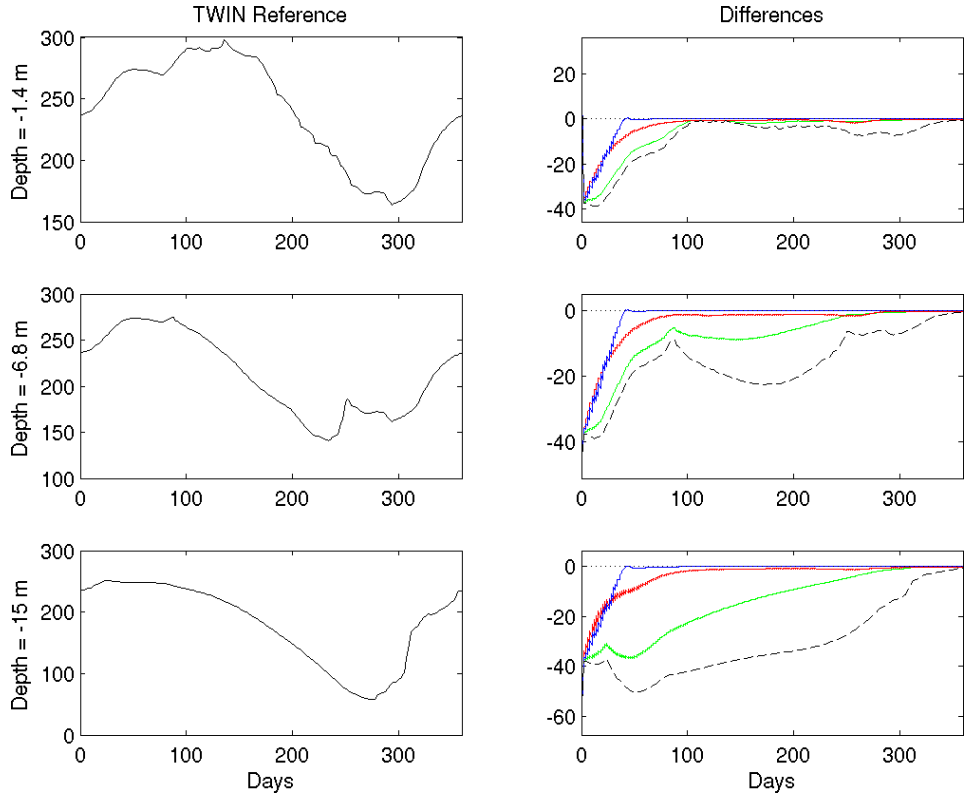


Figure 2.20: Dissolved oxygen $\left[\frac{mmol}{m^3}\right]$, O2-AA1 experiment at the depths of $1.4m$, $6.8m$ and $15m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* OI-SEEK; *Red line:* SEEK-0.5; *Blue line:* SEEK-FF.

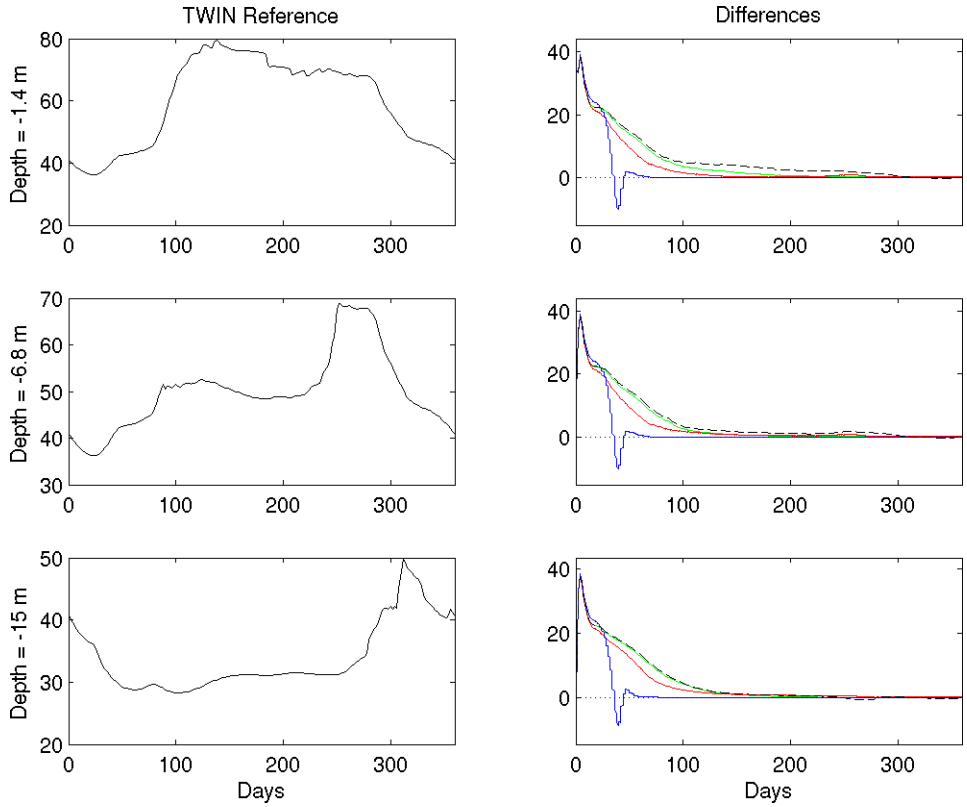


Figure 2.21: Bacteria carbon content $\left[\frac{mg\,C}{m^3}\right]$, O2-AA1 experiment at the depths of $1.4m$, $6.8m$ and $15m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* OI-SEEK; *Red line:* SEEK-0.5; *Blue line:* SEEK-FF.

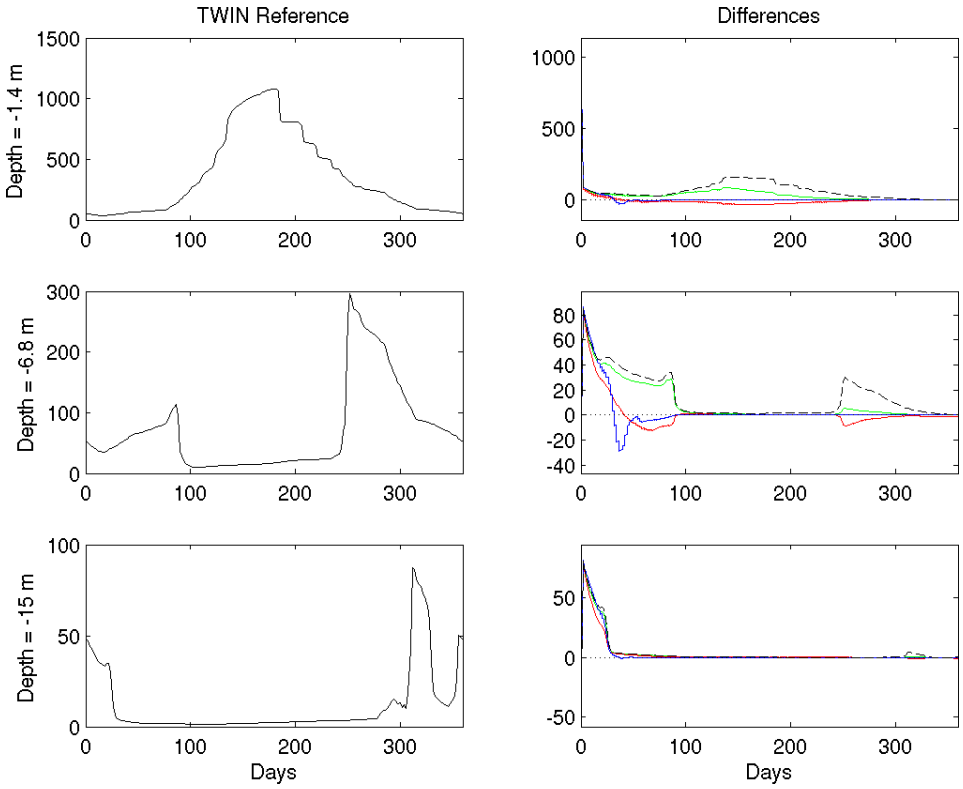


Figure 2.22: Particulate organic carbon $\left[\frac{mg\,C}{m^3}\right]$, O2-AA1 experiment at the depths of $1.4m$, $6.8m$ and $15m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* OI-SEEK; *Red line:* SEEK-0.5; *Blue line:* SEEK-FF.

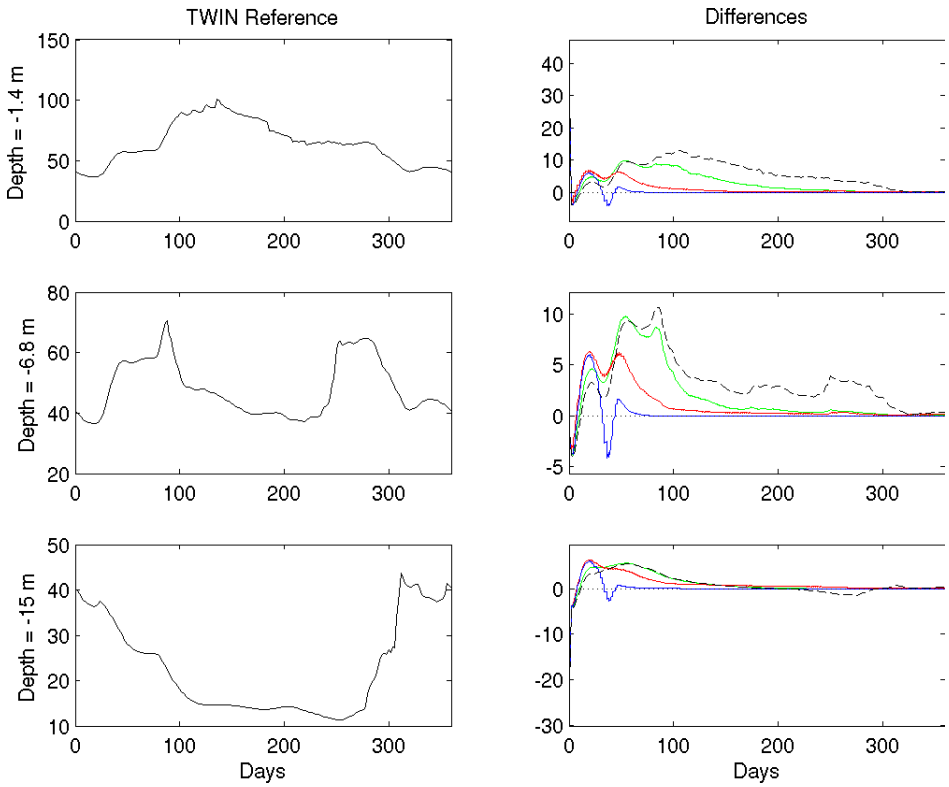


Figure 2.23: Microzooplankton carbon content $\left[\frac{mg\,C}{m^3}\right]$, O2-AA1 experiment at the depths of $1.4m$, $6.8m$ and $15m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* OI-SEEK; *Red line:* SEEK-0.5; *Blue line:* SEEK-FF.

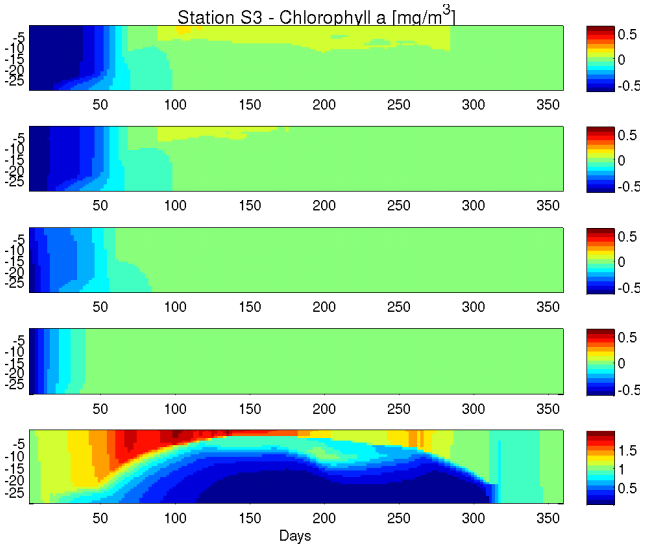


Figure 2.24: Chlorophyll-a content $\left[\frac{mg}{m^3}\right]$ for the O2-S3 experiment. *From bottom to top:* Reference Solution, Analysis error p^a of SEEK-FF, SEEK-0.5 and OI-SEEK, free model run error $x^{free} - x^t$.

For session O2-S3 for chlorophyll and phosphate respectively (see fig. 2.24 and 2.25), the quality of the analysis is gradually improving from the Optimal Interpolation filter over the semi-static *SEEK* to the fully propagating *SEEK* Filter. The time series for distinct depths reveal, that in this site SEEK-0.5 and SEEK-FF correct immediately the chlorophyll, all the nutrients and oxygen errors (fig. 2.26 to 2.31). The SEEK-FF Filter is able to reach the correct level of the reference run in this site in less then 50 days, while the SEEK-0.5 Filter is significantly slower (~100 days). The OI-SEEK Filter even needs 180-360 days to adapt to the reference run (with the only exception of chlorophyll (fig. 2.26), where even the free run recovers from the initial perturbation within 120-150 days). However, in the case of silicate the strong initial corrections result in an amplification of the initial model oscillation, an effect, that SEEK-FF and SEEK-0.5 recover fast, while the OI-SEEK needs some time, in particular in the lower part of the water column. The bacteria solution (fig. 2.32) is similarly well controlled by the two *SEEK* Filters, while the Optimal Interpolation

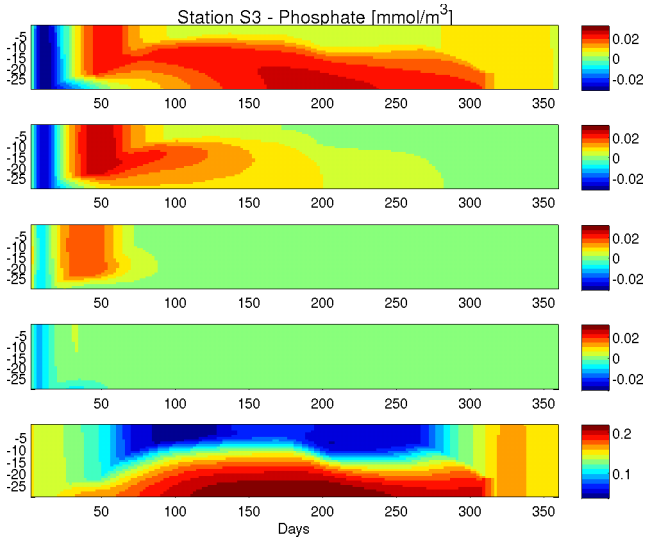


Figure 2.25: Phosphate $\left[\frac{mmol}{m^3}\right]$ field for the O2-AA1 experiment. *From bottom to top:* Reference Solution, Analysis error p^a of SEEK-FF, SEEK-0.5 and OI-SEEK, free model run error $x^{free} - x^t$.

scheme is not able to improve the free run.

For particulate organic matter (fig. 2.33) the behaviour is again quite different for the three assimilation schemes. SEEK-FF follows the free model run for the first 30 days, then suddenly corrections grow and the analysis state is deflected quickly towards the reference state. Then at around day 80, the concentration peaks in the middle and upper level of the plots of the reference solution, create strongly amplified errors in the analysis states. The SEEK-0.5 Filter on the contrary shows corrections with a rather linear trend from the beginning of the simulation, that however overpass the reference state and the filter is not able to recover until the mentioned peaks in the reference field vanish. The Optimal Interpolation scheme is also for this state variable essentially following the free model run without any significant improvement of the solution. For Microzooplankton all three filters show, similar to site AA1, some amplification of the initial oscillation in the model dynamics, that is faster or slower corrected with the same general tendencies, that were observed

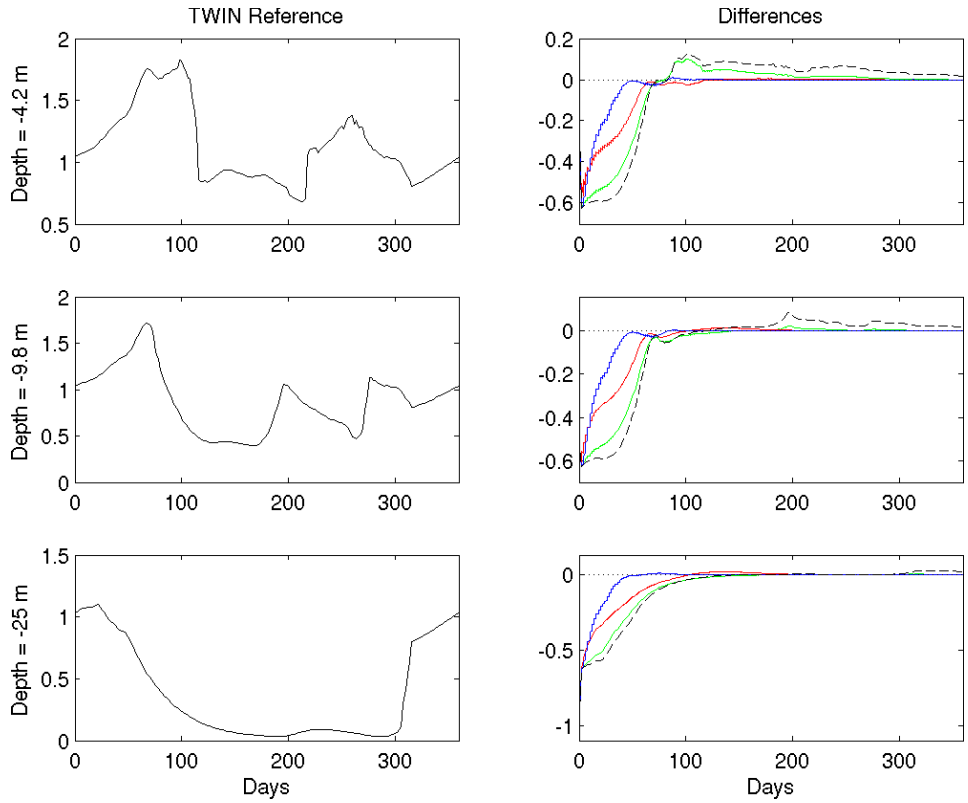


Figure 2.26: Chlorophyll content $\left[\frac{mg}{m^3}\right]$, O2-S3 experiment at the depths of $4.2m$, $9.8m$ and $25m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* OI-SEEK; *Red line:* SEEK-0.5; *Blue line:* SEEK-FF.

for the other states.

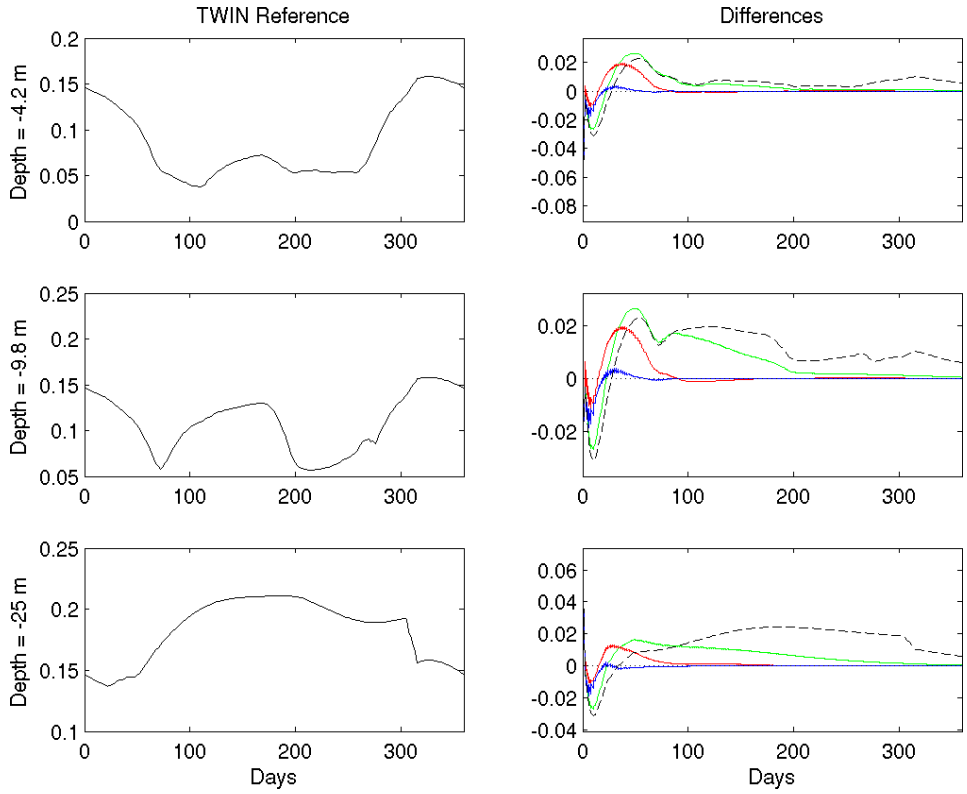


Figure 2.27: Phosphate $\left[\frac{mmol}{m^3}\right]$, O2-S3 experiment at the depths of $4.2m$, $9.8m$ and $25m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* OI-SEEK; *Red line:* SEEK-0.5; *Blue line:* SEEK-FF.

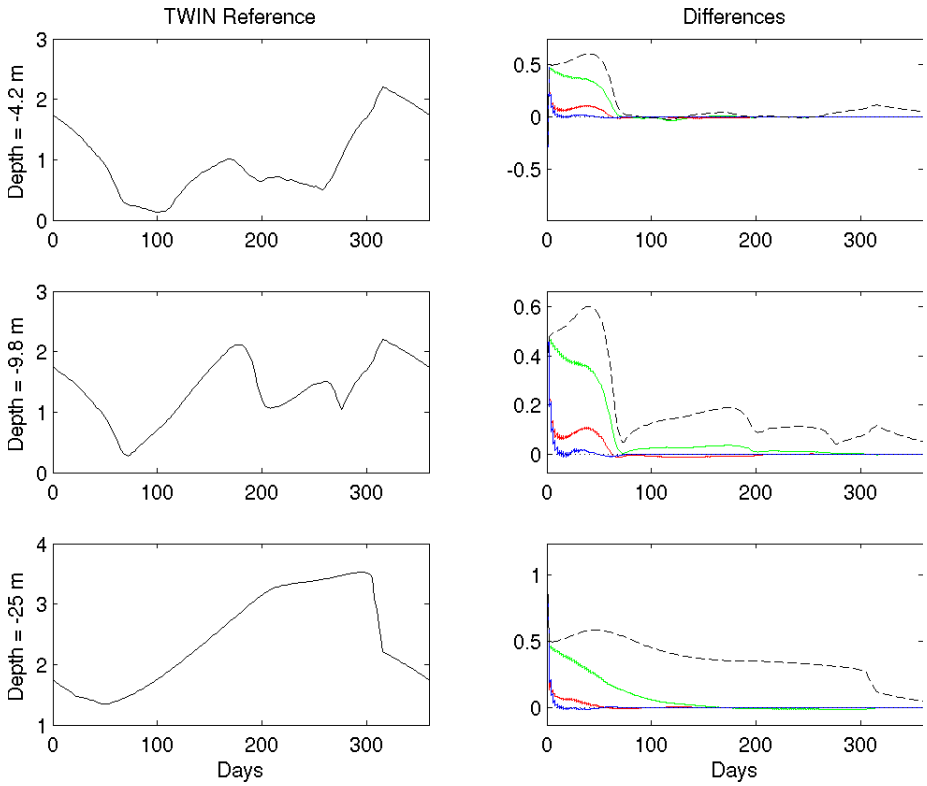


Figure 2.28: Ammonium $\left[\frac{mmol}{m^3}\right]$, O2-S3 experiment at the depths of $4.2m$, $9.8m$ and $25m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* OI-SEEK; *Red line:* SEEK-0.5; *Blue line:* SEEK-FF.

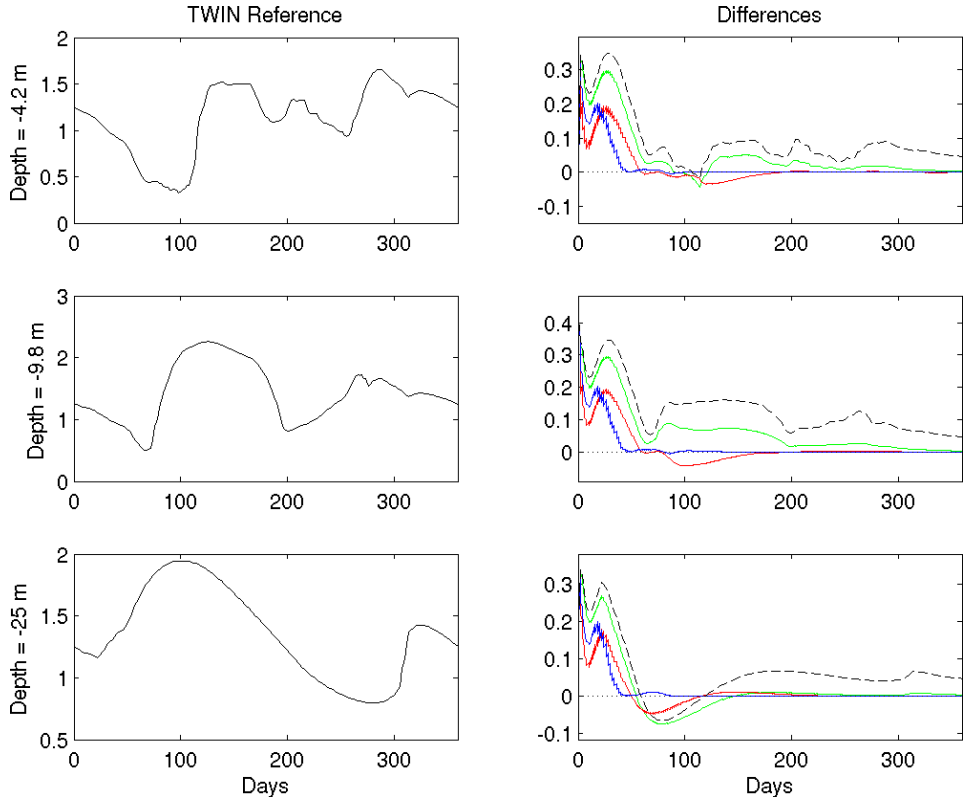


Figure 2.29: Nitrate $\left[\frac{mmol}{m^3}\right]$, O2-S3 experiment at the depths of $4.2m$, $9.8m$ and $25m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* OI-SEEK; *Red line:* SEEK-0.5; *Blue line:* SEEK-FF.

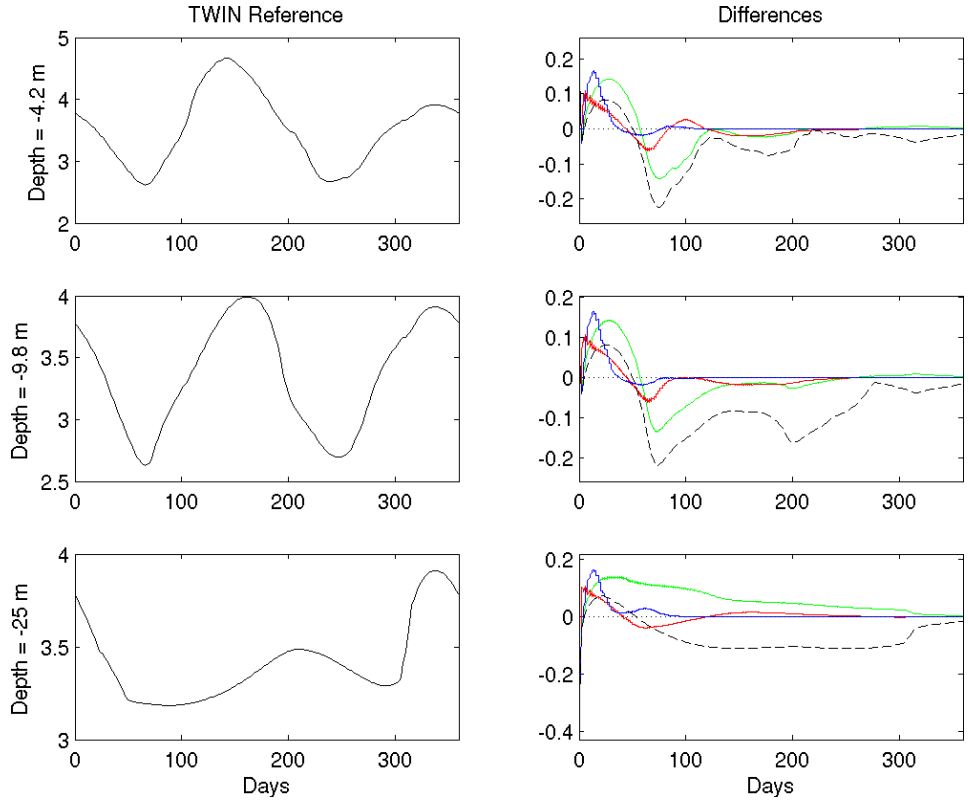


Figure 2.30: Silicate $\left[\frac{mmol}{m^3}\right]$, O2-S3 experiment at the depths of $4.2m$, $9.8m$ and $25m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* OI-SEEK; *Red line:* SEEK-0.5; *Blue line:* SEEK-FF.

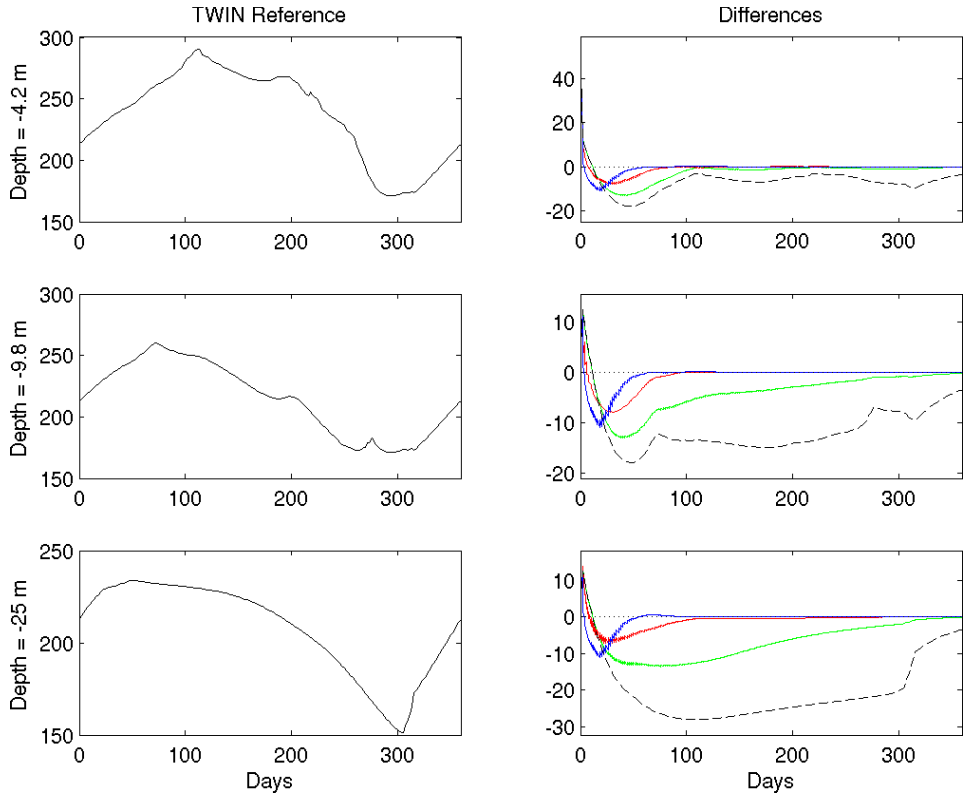


Figure 2.31: Dissolved oxygen $\left[\frac{mmol}{m^3}\right]$, O2-S3 experiment at the depths of $4.2m$, $9.8m$ and $25m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* OI-SEEK; *Red line:* SEEK-0.5; *Blue line:* SEEK-FF.

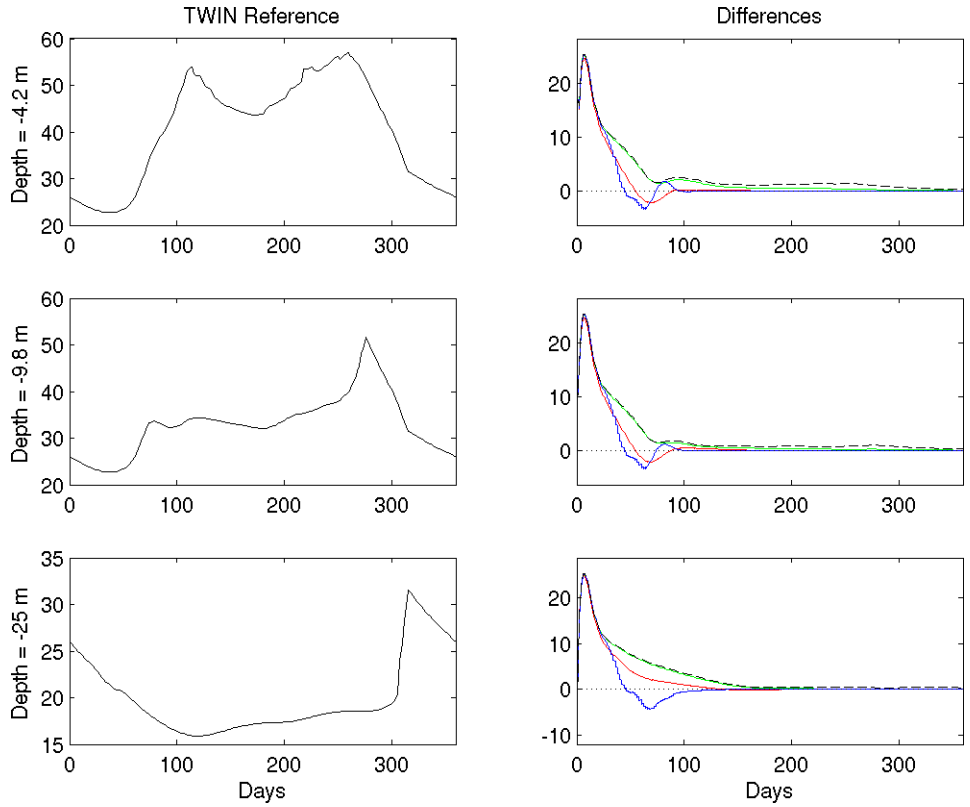


Figure 2.32: Bacteria carbon content $\left[\frac{mg\,C}{m^3}\right]$, O2-S3 experiment at the depths of $4.2m$, $9.8m$ and $25m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* OI-SEEK; *Red line:* SEEK-0.5; *Blue line:* SEEK-FF.

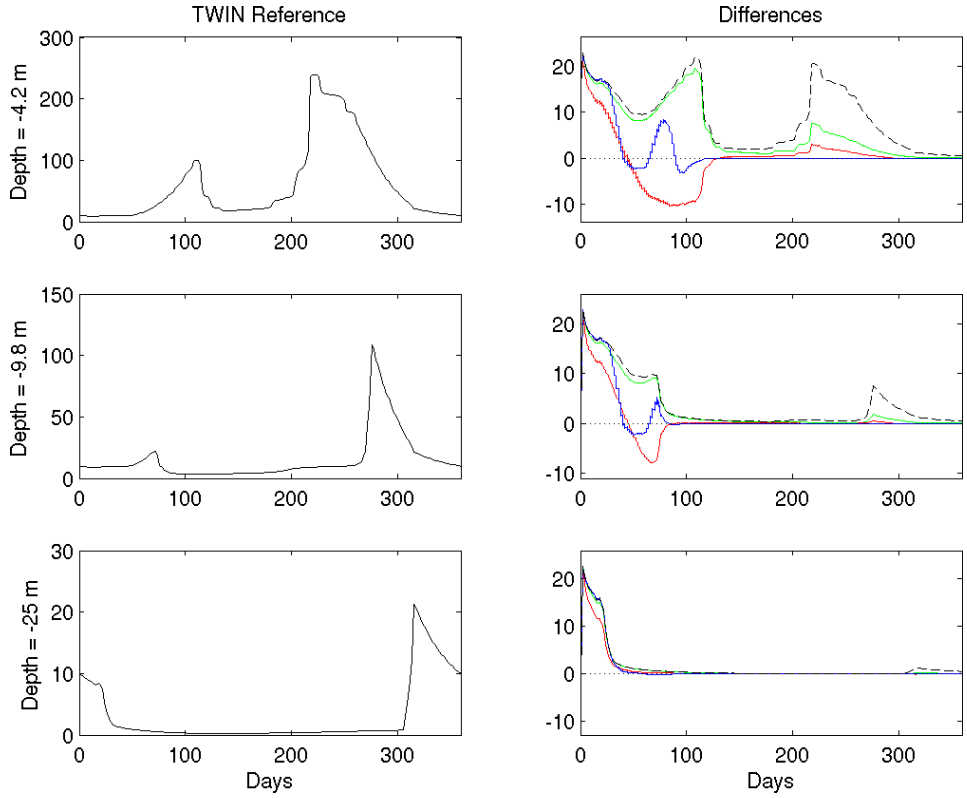


Figure 2.33: Particulate organic carbon $\left[\frac{mg\,C}{m^3}\right]$, O2-S3 experiment at the depths of $4.2m$, $9.8m$ and $25m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* OI-SEEK; *Red line:* SEEK-0.5; *Blue line:* SEEK-FF.

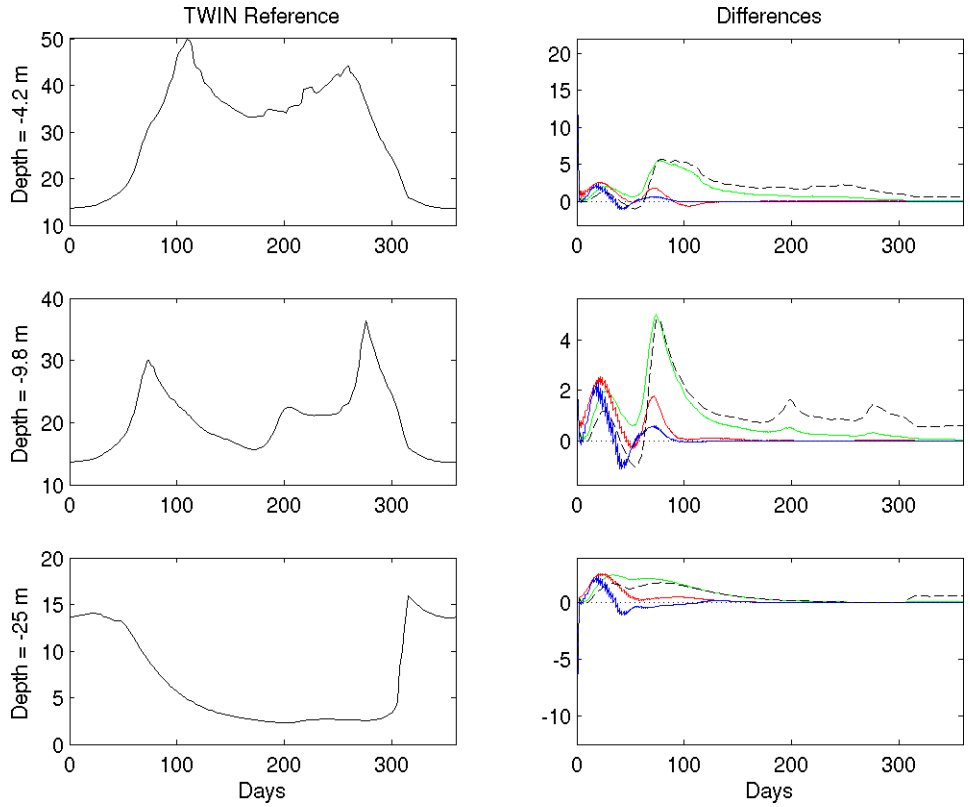


Figure 2.34: Microzooplankton carbon content $\left[\frac{mg\,C}{m^3}\right]$, O2-S3 experiment at the depths of $4.2m$, $9.8m$ and $25m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* OI-SEEK; *Red line:* SEEK-0.5; *Blue line:* SEEK-FF.

In summary both implemented *SEEK* filters showed clear and consistent improvements with respect to the free model run and with respect to the OI technique. A distinct gap in analysis quality is evident for both SEEK-FF and SEEK-0.5 with respect to the simple non evolving OI-SEEK scheme, that often simply damps the free run error maintaining its structure. SEEK-FF showed in general the best results, but demonstrated a tendency to a oscillatory behaviour due to overcorrection.

However, the main difference between the two schemes became evident in the calibration of the experiments. While SEEK-FF showed generally better results when adopting lower forgetting factors with respect to the forgetting factor used in SEEK-0.5, the latter exhibited serious stability problems with lower forgetting factor values. In these cases the analysis error tended to grow after initially well performing corrections, that eliminated the errors with respect to the reference run. An example is given in fig. 2.35, that shows the chlorophyll analysis obtained with SEEK-0.5 using a forgetting factor of 0.8, just below the 0.85 used in the simulation described. A small error that originates below the pycnocline after the onset of stratification, is amplified and spreads over the whole water column.

In fact, the stability of the *SEEK* Filter is guaranteed by the propagation of the error correction directions through the tangent linear model, that converges (if the linear approximation holds) towards the directions for which errors are most amplified, while the iteration of the low rank representation of the analysis error $\boldsymbol{U}_n$ is a mere algebraic consequence to adapt to this propagation given by the formulation of the Kalman Filter (see the derivation of the general version of the filter in Pham et al., 1998). It is clear, that this stability characteristic is lost, if the correction directions are kept fixed and the behaviour of the filter is exclusively governed by the combination $\boldsymbol{L}_0^{\mathrm{T}}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{L}_0$, that is determined from the beginning and not propagated, but most of all reflects the observation error only. As a consequence the error covariance, given by equation 2.30, will approach

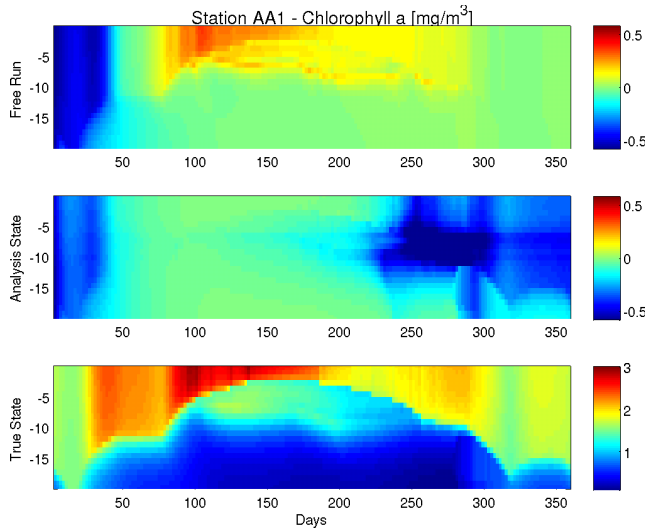


Figure 2.35: Instability of SEEK-0.5 solution in O2-AA1 experiment for forgetting factor $\rho = 0.8$. Chlorophyll-a content $\left[\frac{mg}{m^3}\right]$. *Bottom:* Reference Solution; *Middle:* Analysis error p^a; *Top:* Free model run error $x^{free} - x^t$.

an asymptotic limit

$$\begin{aligned}\lim_{n\to\infty} \boldsymbol{U}_n^{-1} &= \lim_{n\to\infty} \left[\rho^n \boldsymbol{U}_0^{-1} + \left(\rho^{n-1} + \rho^{n-2} + \ldots + \rho^0\right) \boldsymbol{L}_0^{\mathrm{T}} \boldsymbol{H}^{\mathrm{T}} \boldsymbol{R}^{-1} \boldsymbol{H} \boldsymbol{L}_0\right] \\ &= \frac{1}{1-\rho} \boldsymbol{L}_0^{\mathrm{T}} \boldsymbol{H}^{\mathrm{T}} \boldsymbol{R}^{-1} \boldsymbol{H} \boldsymbol{L}_0\end{aligned}$$

(as $0 < \rho \leqslant 1$). If, however the initial corrections are badly specified or the propagation of the model turns away from this directions, the amplitude of the error will not be properly controlled. The initially good performance of SEEK-0.5 also in the unstable cases indicates, that the experienced instabilities are not due to badly specified initial conditions, but rather due to isolated stiffnesses, that the filter was able to recover for higher forgetting factors, i. e. when it preserved enough memory of the background error structure from the initial condition.

As far as the vanishing effect of the initial condition is concerned, the problem of the exclusive dependency on the combination $\boldsymbol{L}^{\mathrm{T}} \boldsymbol{H}^{\mathrm{T}} \boldsymbol{R}^{-1} \boldsymbol{H} \boldsymbol{L}$

holds in principle also for SEEK-FF Filter, but, different from SEEK-0.5, with the propagation of $\boldsymbol{L}$ by the model, information on the model error is projected onto these directions.

However, in conditions of significant non-linearity, this propagation can not work properly and error amplifications might occur. As the filter is stable, the amplifications are reduced when the system returns to quasi-linear conditions and the filter recovers (this might be the case of the oscillations affecting the non-observed states). However, as the filter has down-weighted the older informations, it is relying mostly on the most recent and now inaccurate information. Therefore, it might be desirable, to reconsider the specification of the dynamic noise in eq. (2.25) to maintain some level of background information, that helps the filter to retain valuable information, that could go lost in the case of strongly non linear dynamics.

2.6 The *SEEK-Q* Variant

In order to achieve a more effective specification of the dynamic noise an alternative formulation of the Pham et al. (1998) filter is proposed. This new version can be considered intermediate between the general and the forgetting factor version.

Instead of expressing the dynamic noise covariance $\boldsymbol{Q}$ in terms of the previous analysis error covariance, it is specified explicitly by a formulation, that represents the generic structure of the model error in a statistical sense. With the preparation of the initial condition described in section 2.3 at hand, it seems natural to obtain this generic structure from the statistics of the 50 ensemble runs of the model, that defined the initial analysis covariance $\boldsymbol{P}_0^a$. Hence, one might assume, that

$$\boldsymbol{Q}_{ref} = \beta \boldsymbol{P}_0^a \tag{2.37}$$

with $0 < \beta \leqslant 1$ as noise reduction coefficient (NRC). This approach is similar to a formulation applied by Mitchell and Houtekamer (2000) and Houtekamer et al. (2005), who proposed in the context of an Ensemble Kalman Filter for the model error covariance $\boldsymbol{Q}$ to suppose a functional form similar to the forecast error covariance defined in a connected 3DVAR experiment, so they argued that in cases of low dynamic noise, it may be approximated as proportional.

The dynamic noise is then, coherently with the *SEEK* philosophy, formulated in low rank representation $\boldsymbol{S}$:

$$\boldsymbol{S} = \boldsymbol{L}_0^T \boldsymbol{Q}_{ref} \boldsymbol{L}_0 \tag{2.38}$$

$$\boldsymbol{Q}_{ref} = \beta \boldsymbol{L}_0 \boldsymbol{U}_0 \boldsymbol{L}_0^{\mathrm{T}} \ . \tag{2.39}$$

The idea behind this approach is to let the model maintain some background information on the essential error correlation structure in the analysis covariance to support the filter in cases, where the propagation in the *SEEK* formulation with forgetting factor, that relies entirely on the correct

functioning of the filter's propagation equations performs poorly, e. g. over short periods characterised by strong non-linearities.

In order to properly insert this modification into the propagation of the error correction directions and into the low rank representation of the analysis error covariance (eq. (2.24) and (2.25)), another modification with respect to eq. 2.27 and 2.28 is applied, i.e. the error propagation directions are renormalised each assimilation step in order to maintain their normality $\boldsymbol{L}^T\boldsymbol{L} = \boldsymbol{\delta}$. This can be easily obtained, e. g. by taking the normalisation factor $\boldsymbol{N}_i$ as the inverse Cholesky factor of $\boldsymbol{L}_{i-1}^T\mathcal{M}_{i-1}^T\mathcal{M}_{i-1}\boldsymbol{L}_{i-1}$, so that

$$\boldsymbol{L}_i = \mathcal{M}_{i-1}\boldsymbol{L}_{i-1}\boldsymbol{N}_i \,. \tag{2.40}$$

Then the covariance matrix has to be updated following the equation

$$\boldsymbol{U}_i^{-1} = \left[\boldsymbol{N}_i^{-1}\boldsymbol{U}_{i-1}\boldsymbol{N}_i^{-T} + \boldsymbol{L}_i^T\boldsymbol{L}_0\boldsymbol{S}\boldsymbol{L}_0^T\boldsymbol{L}_i\right]^{-1} + \boldsymbol{L}_i^T\mathcal{H}_i^T\boldsymbol{R}^{-1}\mathcal{H}_i\boldsymbol{L}_i \,.$$

and the new version of the *SEEK* Filter hereafter called SEEK-Q is given by:

$$\boldsymbol{S} = \boldsymbol{L}_0^T\boldsymbol{Q}_{ref}\boldsymbol{L}_0 = \beta\boldsymbol{L}_0^T\boldsymbol{P}^{f,ref}\boldsymbol{L}_0 \tag{2.41a}$$

$$\boldsymbol{L}_i = \mathcal{M}_{i-1}\boldsymbol{L}_{i-1}\boldsymbol{N}_i \tag{2.41b}$$

$$\boldsymbol{U}_i^{-1} = \left[\boldsymbol{N}_i^{-1}\boldsymbol{U}_{i-1}\boldsymbol{N}_i^{-T} + \boldsymbol{L}_i^T\boldsymbol{L}_0\boldsymbol{S}\boldsymbol{L}_0^T\boldsymbol{L}_i\right]^{-1} + \boldsymbol{L}_i^T\mathcal{H}_i^T\boldsymbol{R}^{-1}\mathcal{H}_i\boldsymbol{L}_i \,. \tag{2.41c}$$

The derivation of these equations is given in Appendix B.

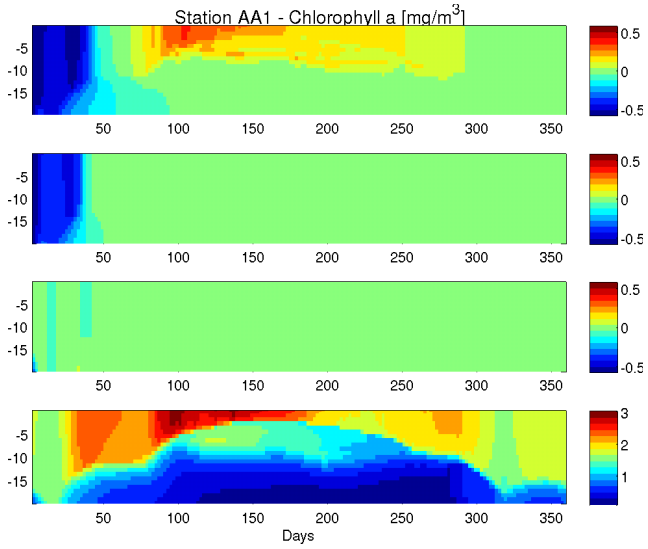


Figure 2.36: Chlorophyll-a content $\left[\frac{mg}{m^3}\right]$ for the O2-AA1 experiment. *From bottom to top:* Reference Solution, Analysis error p^a of SEEK-Q and SEEK-FF, free model run error $x^{free} - x^t$.

2.7 Comparing SEEK-Q with SEEK-FF

Experiments making use of observation vector O2 for locations AA1 and S3 were repeated using SEEK-Q and the results obtained were compared against the corresponding SEEK-FF experiments (O2-AA1-SEEK-Q and O2-S3-SEEK-Q), that is the more sophisticated of the implemented filters and performed the best analysis. This comparison showed a high efficiency of both analysis schemes over all state variables with differences often limited to the first 50 days of simulation, i. e. to the first 25 assimilation steps.

For chlorophyll and phosphate the analysis solutions relative to site AA1 are given in figures 2.36 and 2.37 respectively and demonstrate the efficiency of SEEK-Q, that corrects the analysis error virtually immediately. A closer look on the interaction between model dynamics and filter correction for the same state variables is given in figures 2.38 and 2.39 that reveals, how the forecast solution drifts away from the true state,

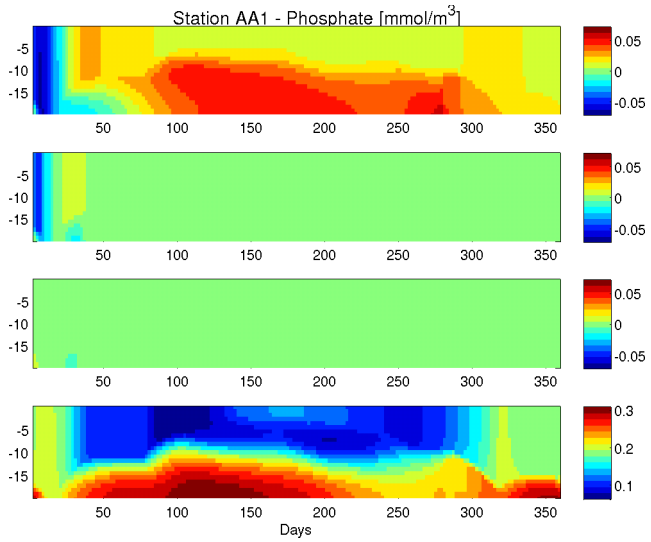


Figure 2.37: Phosphate $\left[\frac{mmol}{m^3}\right]$ field for the O2-AA1 experiment. *From bottom to top:* Reference Solution, Analysis error p^a of SEEK-Q and SEEK-FF, free model run error $x^{free} - x^t$.

but is continuously corrected by the analysis step. The SEEK-FF shows a slower convergence to the reference state. In the chlorophyll case the filter is rather late in adapting the solution to the measurements, as already observed in section 2.5. The solutions for the other nutrients and oxygen (fig. 2.40 to 2.43) repeat the behaviour observed for phosphate, with an immediate adaptation of the analysis for the SEEK-Q Filter, while the SEEK-FF Filter is more gradually adapting to the reference state, even if rather quickly. An extreme case can be observed for nitrate (fig. 2.40), where the reference state is reached with the first correction and never deflected anymore due to the quasi-equilibrium in the reference solution over winter.

More interesting is the comparison of the errors for bacteria (fig. 2.44). The sharp peak in the model dynamics caused by the strong perturbation of the initial condition is followed essentially by both, the SEEK-Q and the SEEK-FF solutions without improvement. Just after the peak, however, the SEEK-Q analysis manages to descend gradually towards the true

state, while the SEEK-FF Filter is for a short period unable to project the information of the observed states to the bacteria field and then, rather abruptly starts the oscillating correction dynamics already described (section 2.5). This behaviour is not observed in the SEEK-Q simulation. The errors for particulate organic matter show a similar structure with clearly better results achieved by SEEK-Q, even if at least in the middle section a smaller oscillation can be observed also in this case.

For microzooplankton (fig. 2.46) the situation in the first 30-40 days is rather critical. The initial oscillation of the free model run is amplified by both filters, in particular the new one. This strong effect might be induced by the initial chlorophyll concentrations, that are significantly lower for the free run and the assimilation runs. The lower phytoplankton concentration is corrected by the filters causing a corresponding positive correction of the microzooplankton due to the positive correlation of the two variables (predation), that amplifies the already growing error. However, both filters can recover rather quickly from this strong impact, again with an oscillatory behaviour in the SEEK-FF solution.

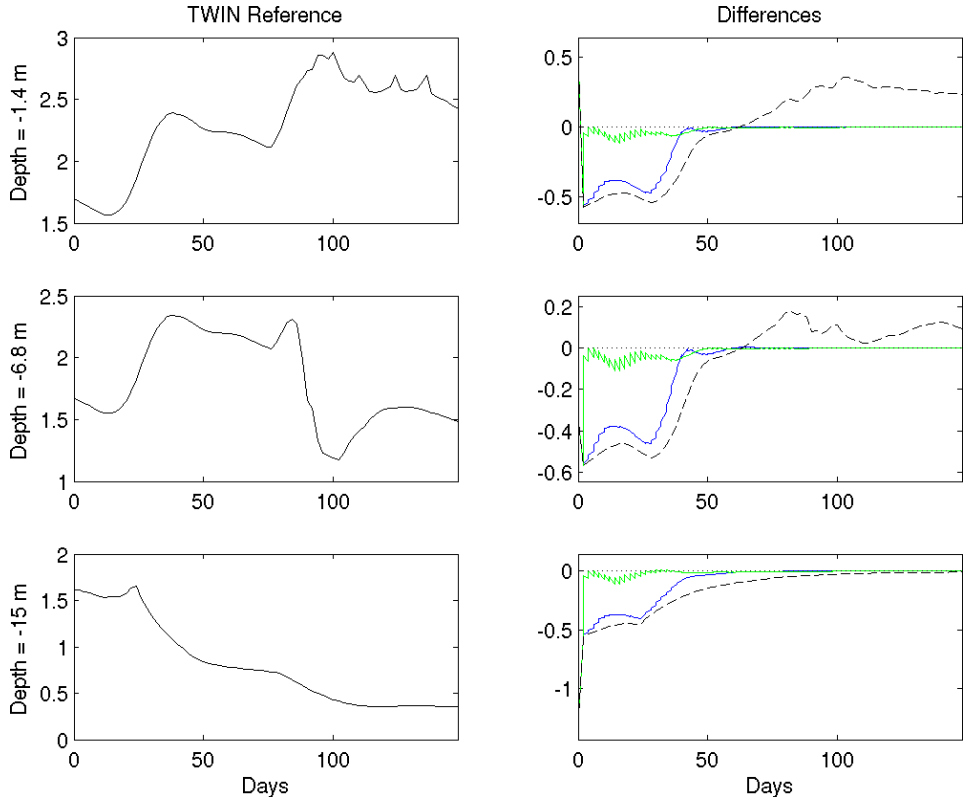


Figure 2.38: Chlorophyll-a content $\left[\frac{mg}{m^3}\right]$, O2-AA1 experiment at the depths of $1.4m$, $6.8m$ and $15m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF.

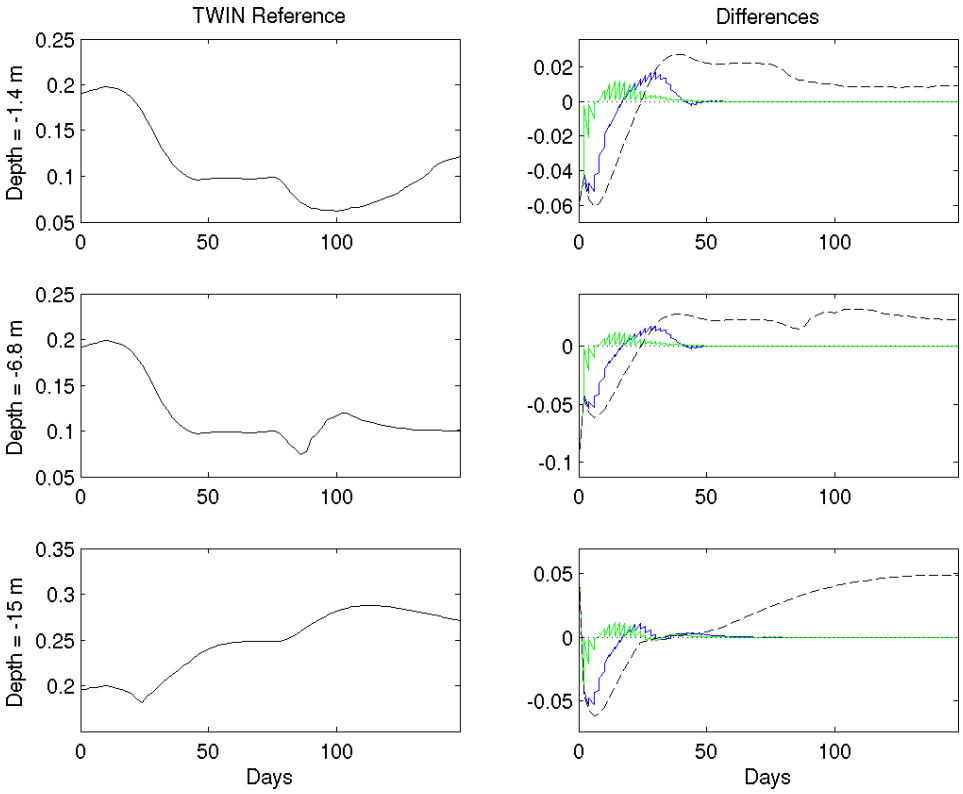


Figure 2.39: Phosphate $\left[\frac{mmol}{m^3}\right]$, O2-AA1 experiment at the depths of $1.4m$, $6.8m$ and $15m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF.

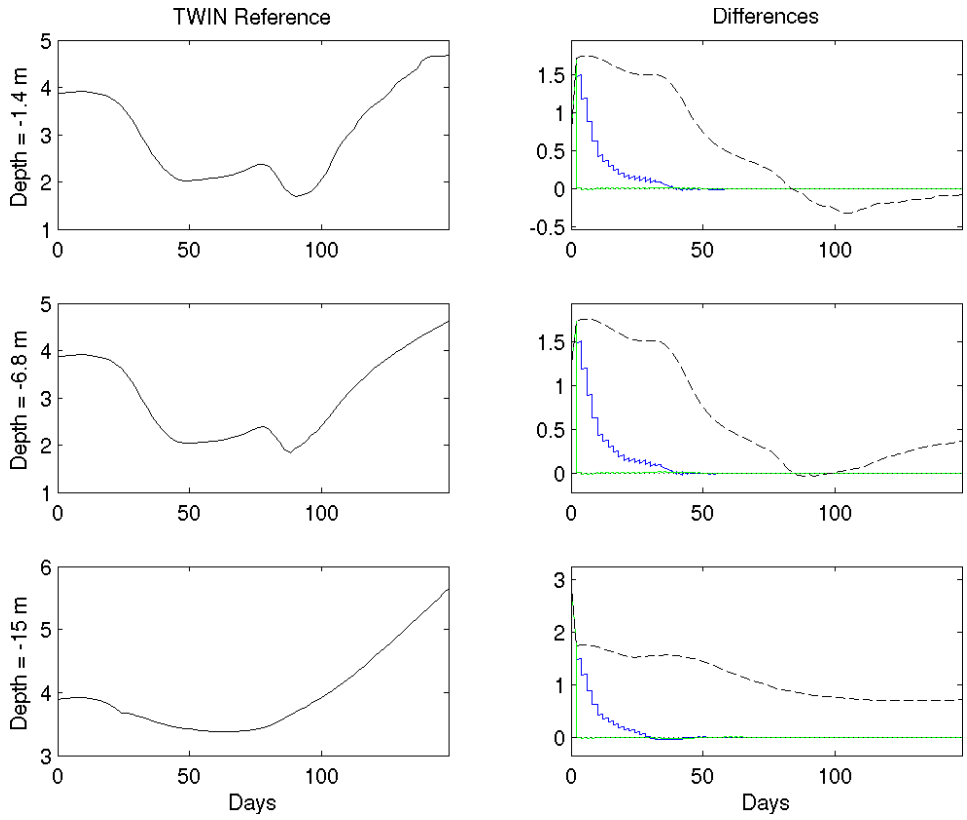


Figure 2.40: Nitrate $\left[\frac{mmol}{m^3}\right]$, O2-AA1 experiment at the depths of $1.4m$, $6.8m$ and $15m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF.

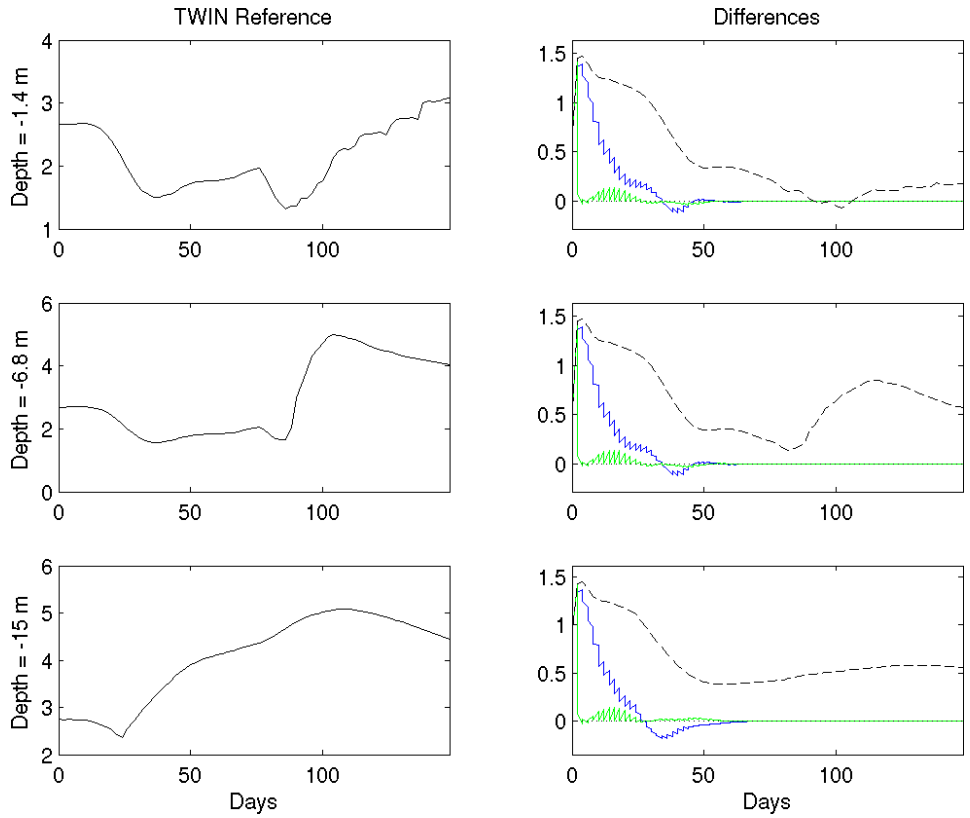


Figure 2.41: Ammonium $\left[\frac{mmol}{m^3}\right]$, O2-AA1 experiment at the depths of $1.4m$, $6.8m$ and $15m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF.

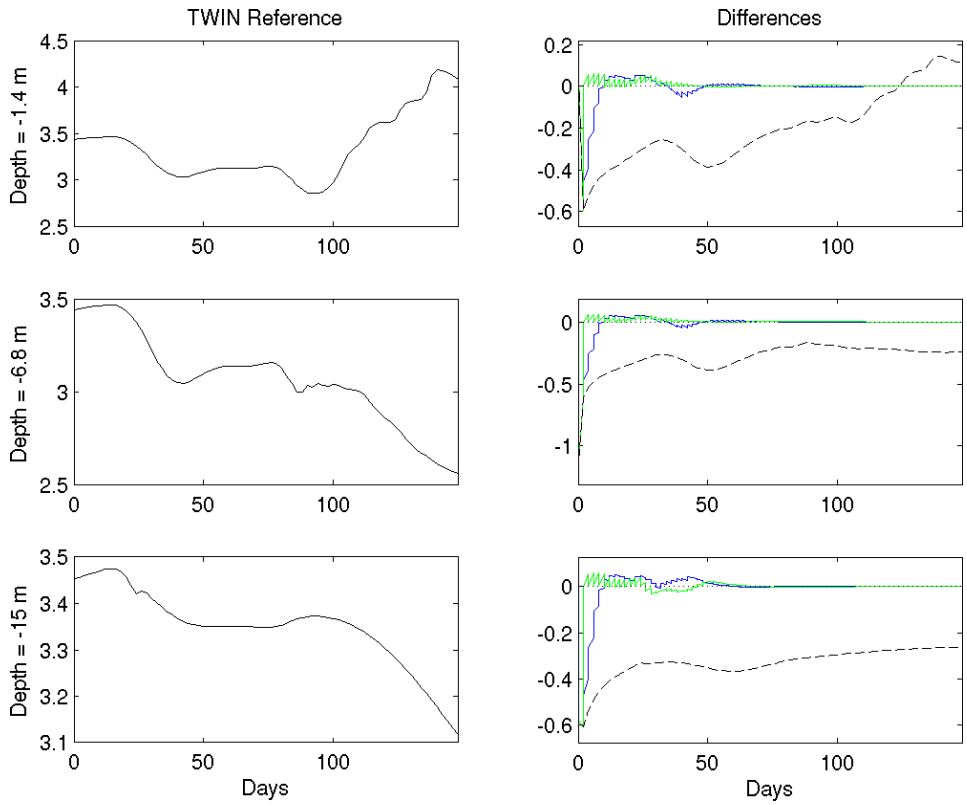


Figure 2.42: Silicate $\left[\frac{mmol}{m^3}\right]$, O2-AA1 experiment at the depths of $1.4m$, $6.8m$ and $15m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF.

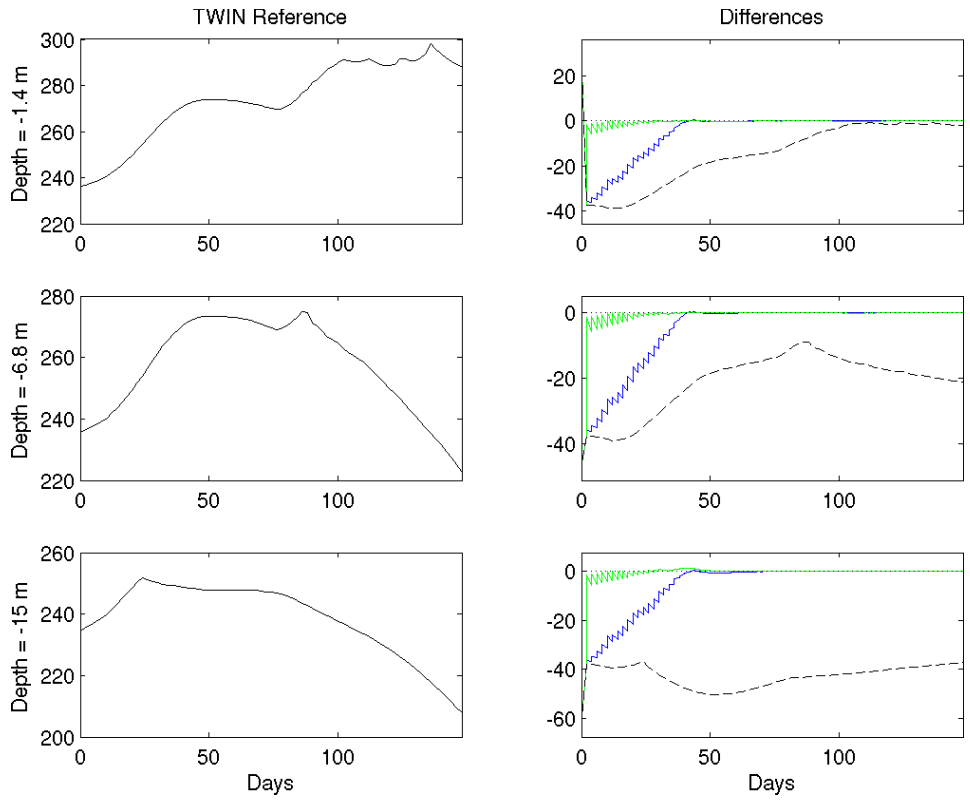


Figure 2.43: Dissolved oxygen $\left[\frac{mmol}{m^3}\right]$, O2-AA1 experiment at the depths of $1.4m$, $6.8m$ and $15m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF.

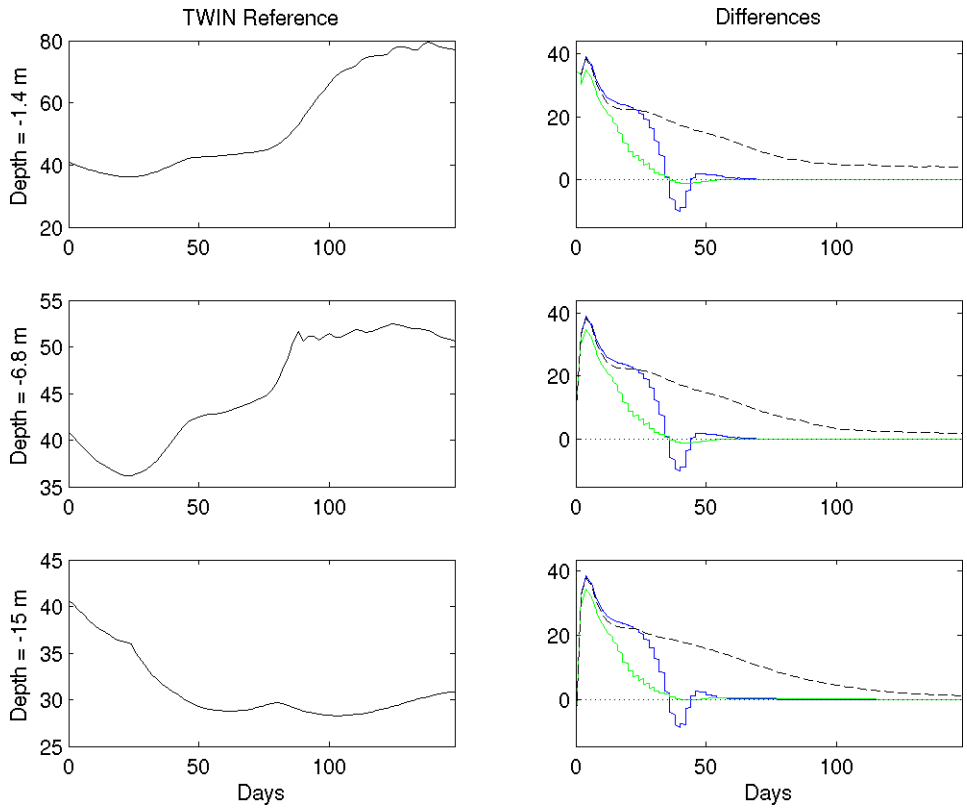


Figure 2.44: Bacteria carbon content $\left[\frac{mg\,C}{m^3}\right]$, O2-AA1 experiment at the depths of $1.4m$, $6.8m$ and $15m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF.

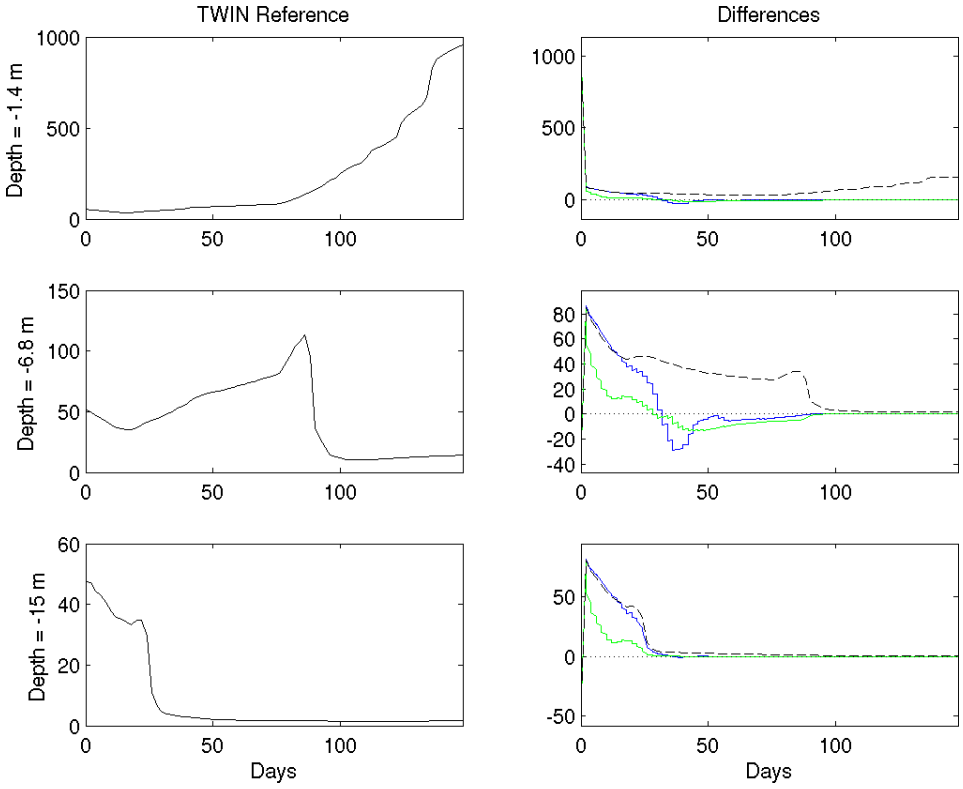


Figure 2.45: Particulate organic carbon $\left[\frac{mg\,C}{m^3}\right]$, O2-AA1 experiment at the depths of $1.4m$, $6.8m$ and $15m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF.

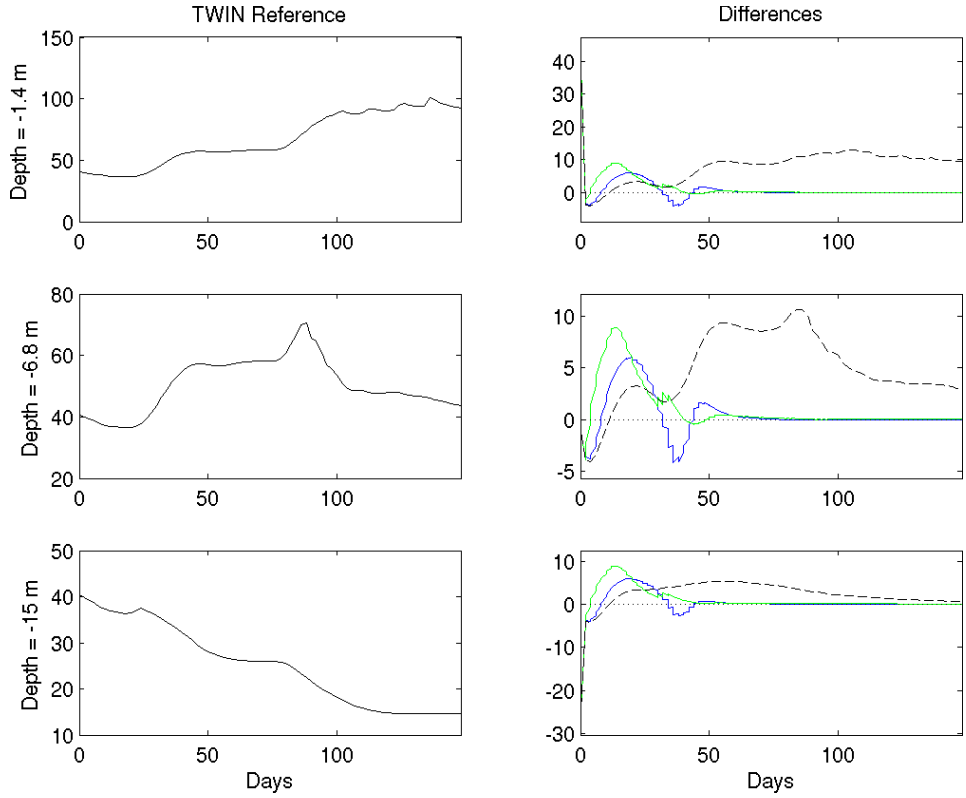


Figure 2.46: Microzooplankton carbon content $\left[\frac{mg\,C}{m^3}\right]$, O2-AA1 experiment at the depths of $1.4m$, $6.8m$ and $15m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF.

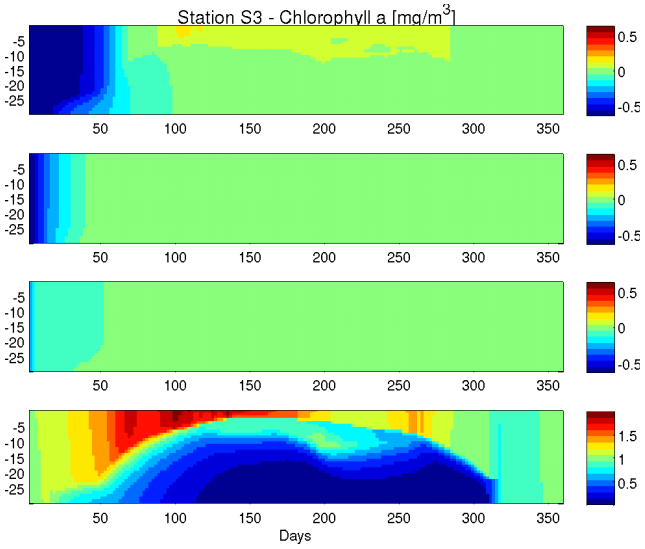


Figure 2.47: Chlorophyll-a content $\left[\frac{mg}{m^3}\right]$ for the O2-S3 experiment. *From bottom to top:* Reference Solution, Analysis error p^a of SEEK-Q and SEEK-FF, free model run error $x^{free} - x^t$.

The SEEK-Q effectiveness is confirmed by the results relative to site S3, as illustrated in figures 2.47 and 2.48 for the variables of chlorophyll and phosphate.

Again SEEK-Q is correcting most of the error of the initial perturbation in the first 2-3 assimilation steps, even if in this case a small error in the chlorophyll field persists for around 50 days. In this case also the SEEK-FF Filter starts correcting from the beginning. The correction is slower, but for chlorophyll it approaches the reference more precisely. Both schemes show little errors in the Phosphate field with the onset of stratification, that are quickly corrected though. Figure 2.49 shows, that the corrections in chlorophyll for the SEEK-Q scheme are not strong enough to counteract the drifting model dynamics immediately causing the small error seen also in figure 2.47.

The situation for the nutrients phosphate , nitrate and ammonium time series (fig. 2.50,2.51 and 2.52respectively) essentially confirms the AA1 results. A major difference is evident in the silicate errors (fig. 2.53), where

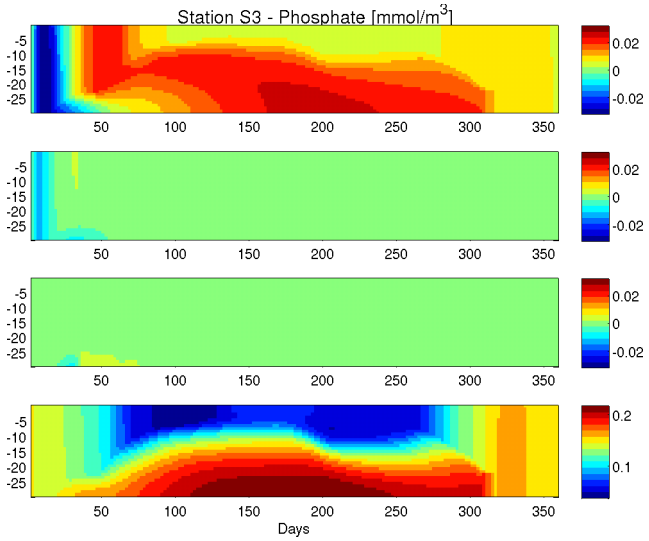


Figure 2.48: Phosphate $\left[\frac{mmol}{m^3}\right]$ field for the O2-S3 experiment. *From bottom to top:* Reference Solution, Analysis error p^a of SEEK-Q and SEEK-FF, free model run error $x^{free} - x^t$.

the SEEK-Q Filter again is able to correct against the strong model drift, while the SEEK-FF Filter is initially even amplifying the model signal correcting in the wrong direction. This causes an error that clearly exceeds even the initial error of the free model run. The oxygen error (fig. 2.54) is similar to the nutrients errors. It is very well controlled for the SEEK-Q solution, while the SEEK-FF Filter needs about 20 days to establish proper corrections. The time series of the bacteria error (fig. 2.55) show the behaviour already observed in the AA1 site with the new filter approaching the reference solution faster and avoiding the oscillatory behaviour of the SEEK-FF field.

However, both filters don't behave satisfactorily on particulate organic matter (fig. 2.56). While the SEEK-Q Filter starts rather correctly, it quickly develops overshooting, that brings the solution away from the true state and the correct level is only recovered by the model dynamics, when the innovations of the observations approach zero. The SEEK-FF Filter on the contrary starts correcting with a significant delay after around 30 days,

then approaches the reference solution well, when a too strong correction turns the solution away again until the filter manages to recover proper error propagation. For both filters the error is not correctly correlated to the observations and both filter reach the level of the reference run only after the errors in the other fields have disappeared.

For Microzooplankton (fig. 2.57), again, the contrasting strong initial perturbations causes problems in the initial assimilation in particular for the SEEK-Q Filter in the middle layer, as was observed for site AA1.

In summary, the new version of the SEEK Filter introduced in section 2.6 performed well in the present twin experiments. In most cases it considerably improved the SEEK-FF analysis, avoiding the oscillatory structures appearing occasionally when the true solution is approached. Problems were met in the initial correction of microzooplankton in both sides, that however, seem to be enforced by the heavy impact of strong contrasting initial perturbation due to the imposition of the initial summer profiles in winter conditions (see section 2.3.4 on the initialisation). More concerning is the relatively poor performance of both filters for particulate organic matter, where both filters don't seem to establish well correlated corrections. This will deserve a closer inspection of the information propagation along the trophic web.

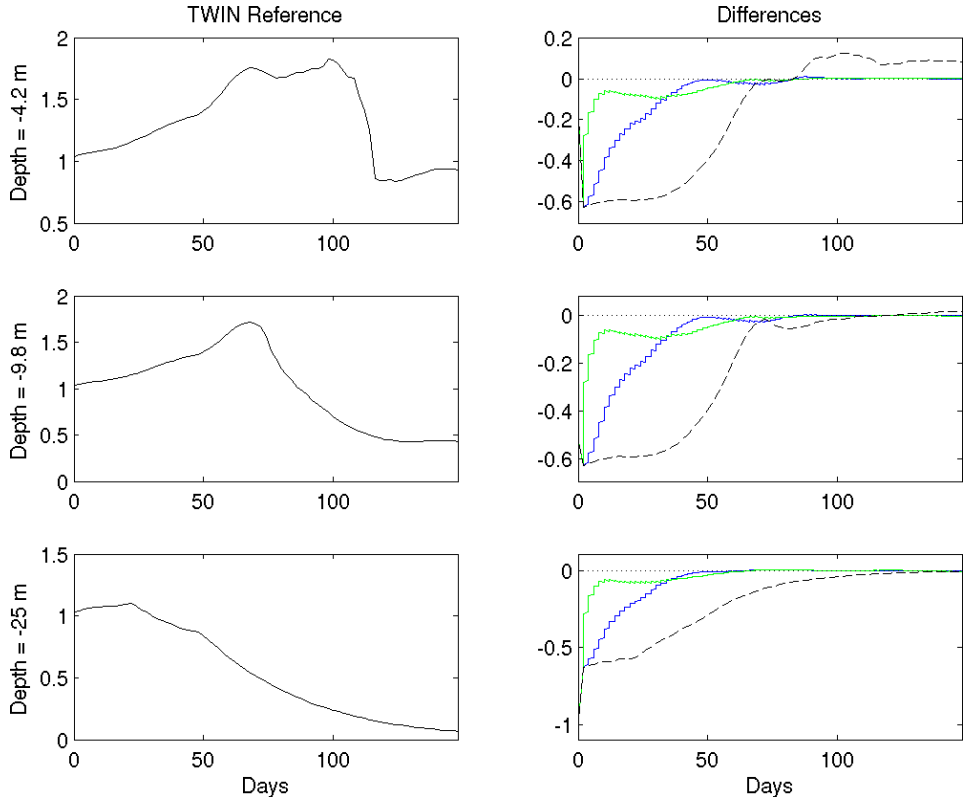


Figure 2.49: Chlorophyll-a content $\left[\frac{mg}{m^3}\right]$, O2-S3 experiment at the depths of $4.2m$, $9.8m$ and $25m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF.

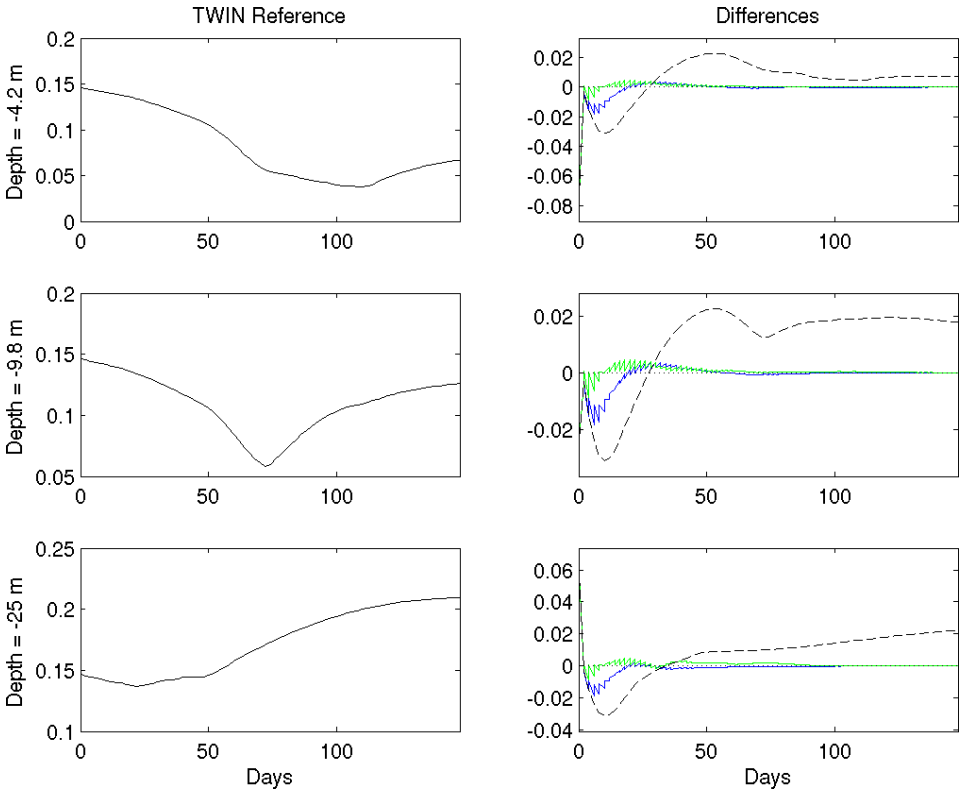


Figure 2.50: Phosphate $\left[\frac{mmol}{m^3}\right]$, O2-S3 experiment at the depths of $4.2m$, $9.8m$ and $25m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF.

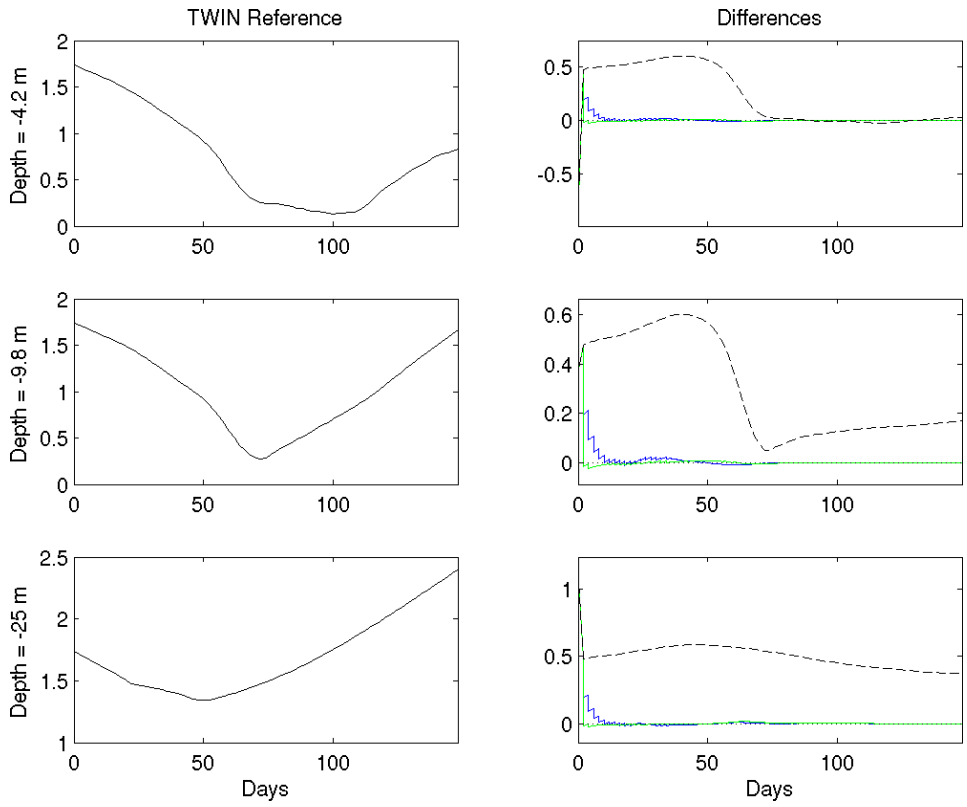


Figure 2.51: Nitrate $\left[\frac{mmol}{m^3}\right]$, O2-S3 experiment at the depths of $4.2m$, $9.8m$ and $25m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF.

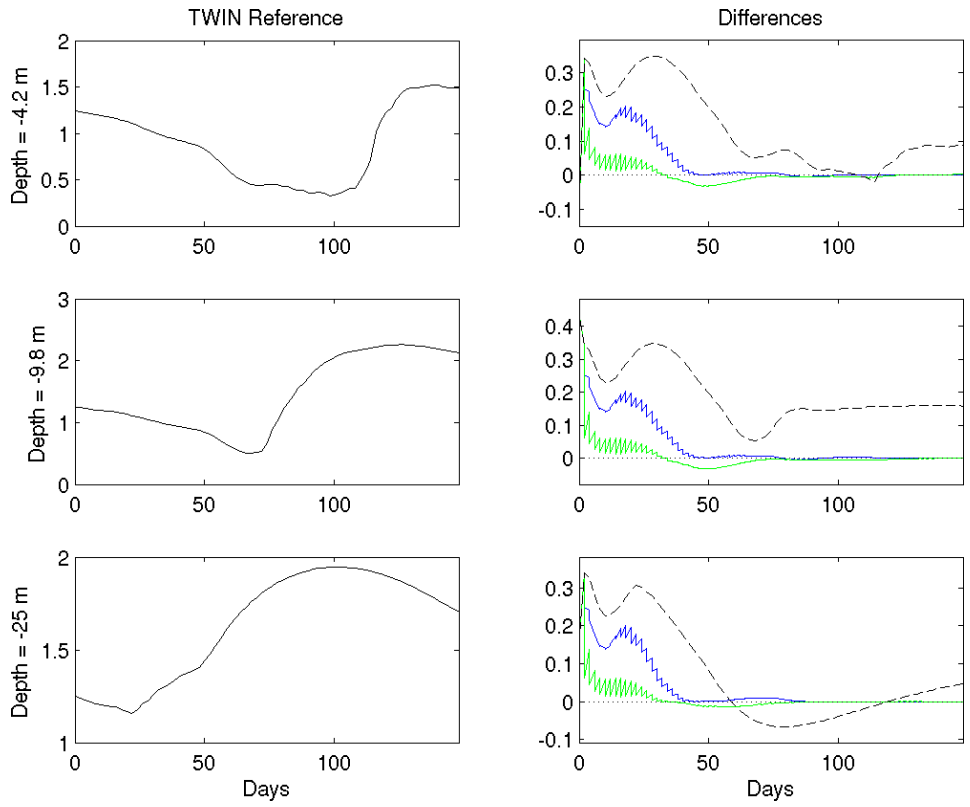


Figure 2.52: Ammonium $\left[\frac{mmol}{m^3}\right]$, O2-S3 experiment at the depths of $4.2m$, $9.8m$ and $25m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF.

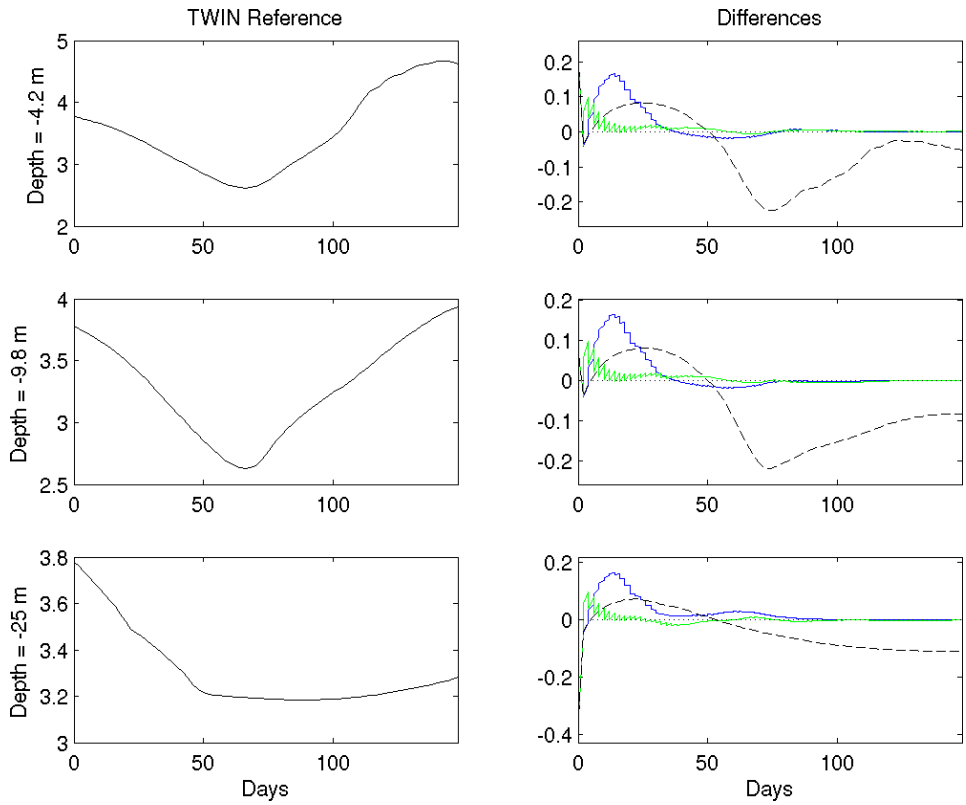


Figure 2.53: Silicate $\left[\frac{mmol}{m^3}\right]$, O2-S3 experiment at the depths of $4.2m$, $9.8m$ and $25m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF.

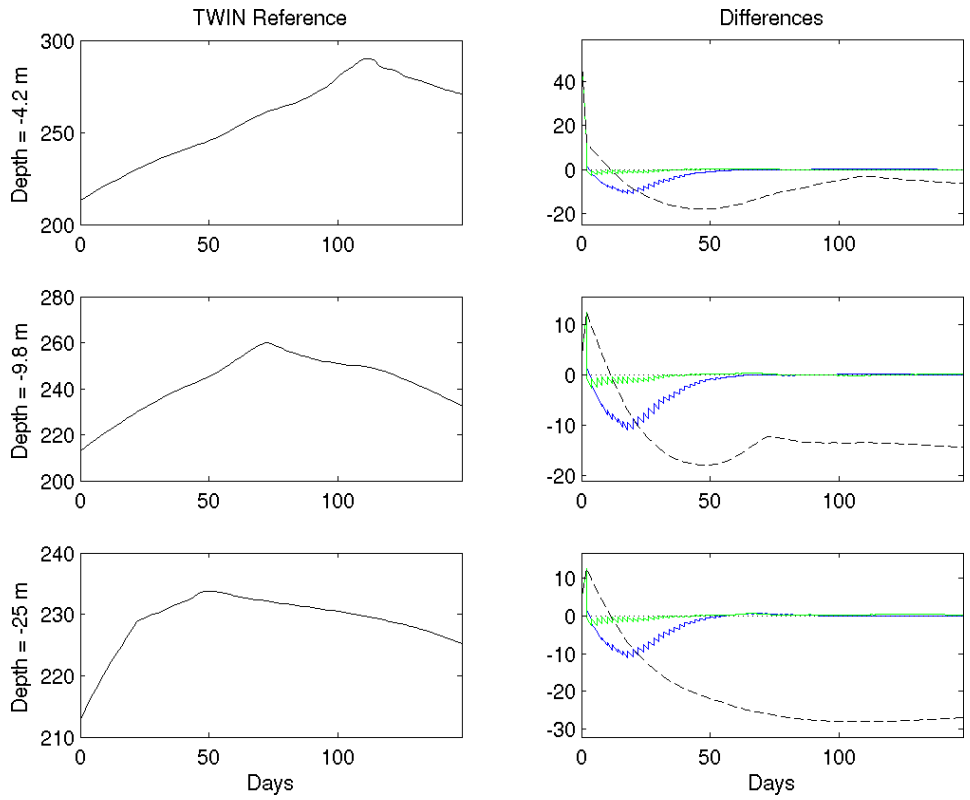


Figure 2.54: Dissolved oxygen $\left[\frac{mmol}{m^3}\right]$, O2-S3 experiment at the depths of $4.2m$, $9.8m$ and $25m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF.

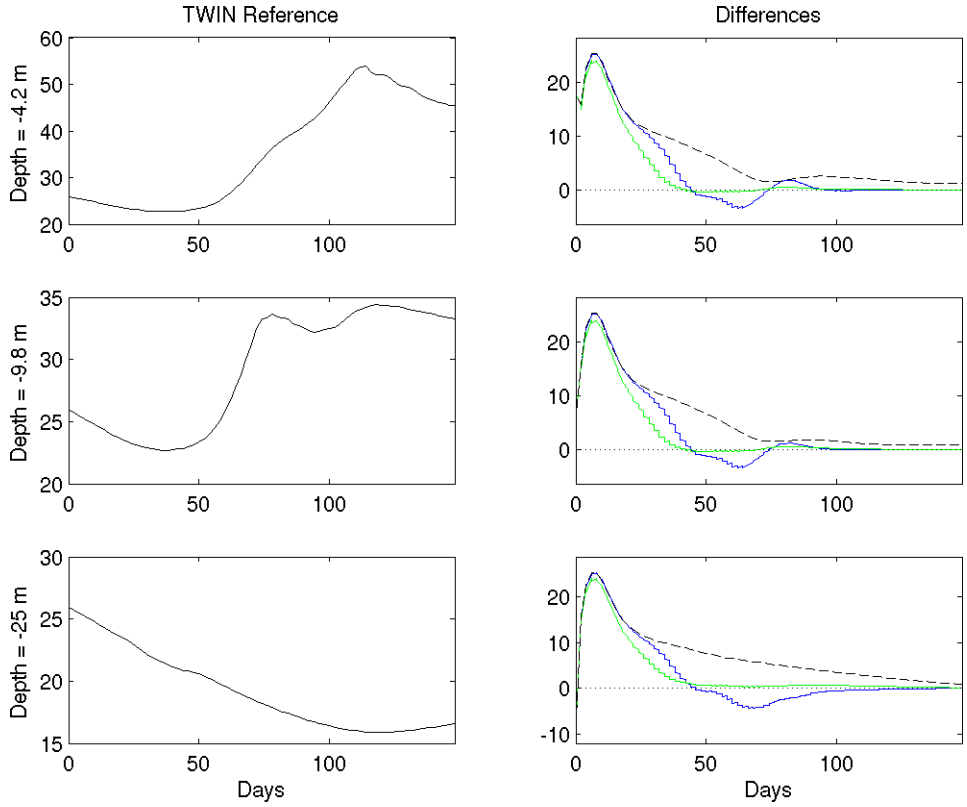


Figure 2.55: Bacteria carbon content $\left[\frac{mg\,C}{m^3}\right]$, O2-S3 experiment at the depths of $4.2m$, $9.8m$ and $25m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF.

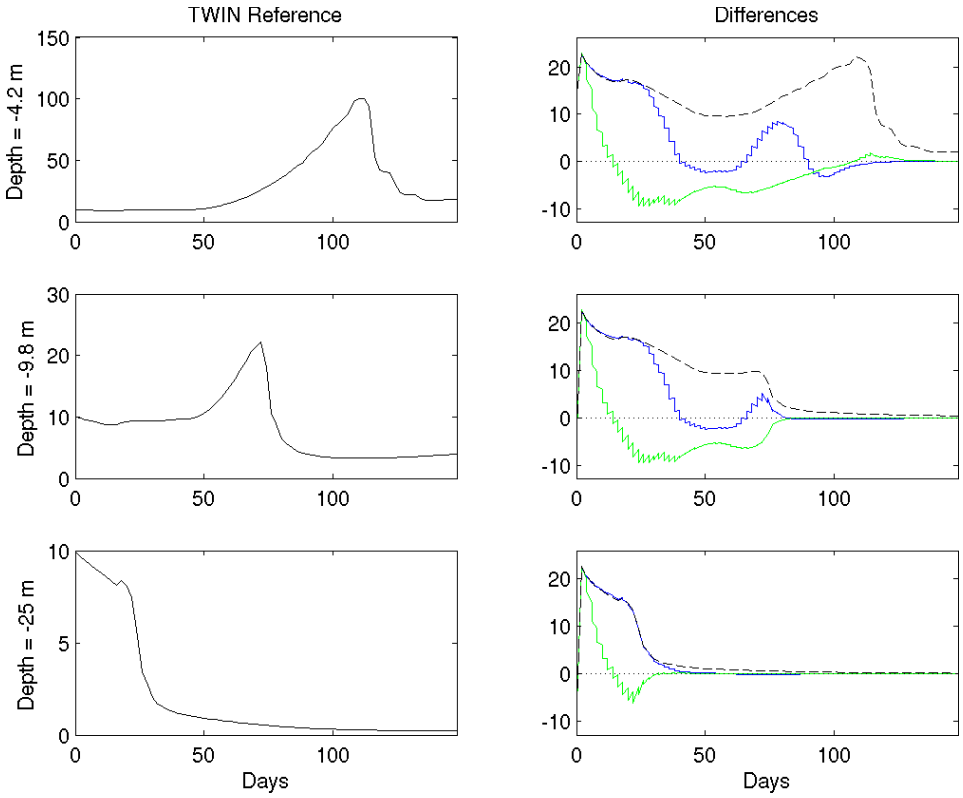


Figure 2.56: Particulate organic carbon $\left[\frac{mg\,C}{m^3}\right]$, O2-S3 experiment at the depths of $4.2m$, $9.8m$ and $25m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF.

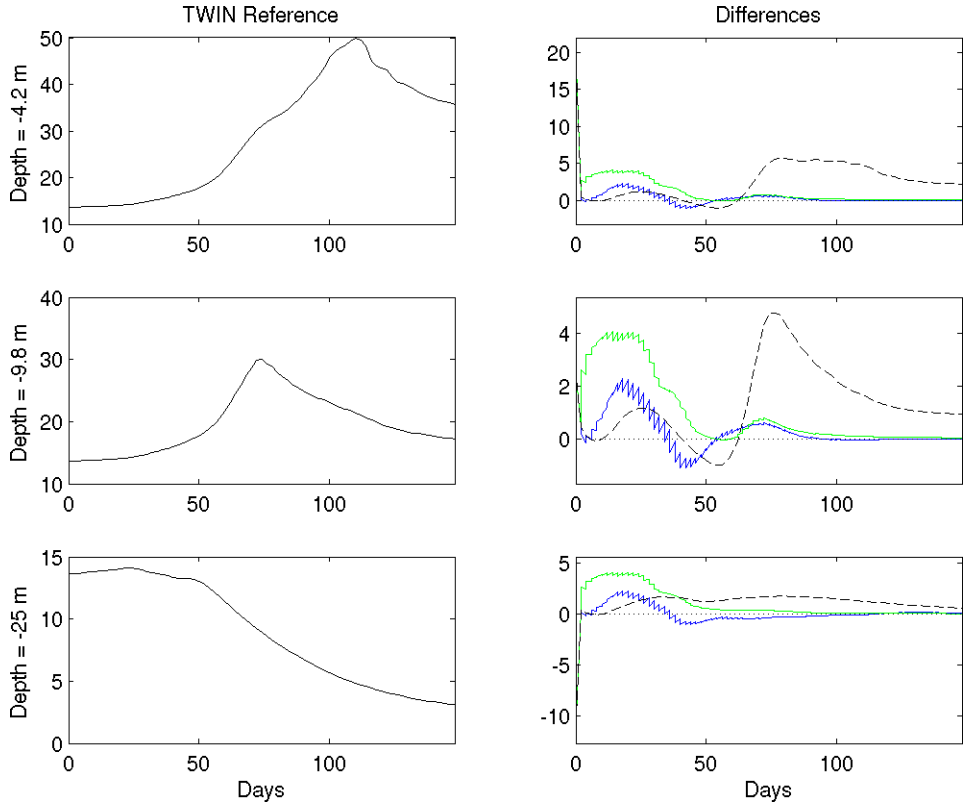


Figure 2.57: Microzooplankton carbon content $\left[\frac{mg\,C}{m^3}\right]$, O2-S3 experiment at the depths of $4.2m$, $9.8m$ and $25m$ (from top to bottom). *Left:* reference run; *Right:* free model run and analysis errors $x - x^t$. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF.

2.8 Summary and conclusions

This chapter presented a comparison of two sub-versions of the Singular Evolutive Extended Kalman Filter supplemented with the proposal of a new formulation. The SEEK-FF version was proposed by Pham et al. (1998). It was designed for practical implementations due to its simple representation of the model error covariance $\boldsymbol{Q}$, that is expressed as proportional to the analysis error covariance (see section 2.2.1.1). The SEEK-0.5, that is frequently used in marine ecosystem dynamics, applies a further simplification with respect to the SEEK-FF proposed by Brasseur et al. (1999), that assumes the persistence of the error-sub-space keeping the error propagation directions constant throughout the simulation, updating only the reduced rank covariance matrix (see section 2.2.1.2).

The new filter version, proposed here, uses a more explicit, but still rather simple formulation for the dynamic noise covariance $\boldsymbol{Q}$ providing a generic structure of the model error to support the filter propagation (see section 2.6). This was motivated by the not entirely satisfactory performance of SEEK-0.5 and SEEK-FF at calibration and simulation. Problems turned out to be connected to the propagation of the error covariance, that relies completely on the correct representation of the error correction directions by the filter. This however holds only in the near-linear limit. The new version, on the contrary, specifies the dynamic noise covariance in a way, that a level of background information is maintained in the propagation of the error covariance and therefore is less sensitive to local effects of non-linearity.

The performance of these filters was validated in a series of TWIN experiments comparing the errors of the filters against the free model run without assimilation and against a simple optimal interpolation scheme, that did not consider the evolution in time of the error statistics.

All filters showed for all state variables in both sites much better performance, than the free model run and the interpolation filter underlining the benefit of taking into account the evolution of the error statistics. The

low rank approximation was confirmed as a reasonable reduction in system dimension, as even the optimal interpolation filter performed considerably better than the free model run on all state variables. The more sophisticated, but also more costly filter SEEK-FF generally showed advantages over the semistatic SEEK-0.5 in the quality of the analysis, however, the major drawback of the SEEK-0.5 was identified in its tendency to instabilities when the forgetting factor was lowered towards the value that was used for the SEEK-FF implementation.

The absence of these stability problems in SEEK-FF (that only exhibited limited oscillatory behaviour for few state variables) confirms the statement of Pham et al. (1998) of the general stability of the filter, that holds in the limit of the linear approximation in which case the error correction directions are propagated properly.

The comparison of the newly proposed version SEEK-Q against the SEEK-FF showed a further improvement of the performance of the *SEEK* filter. Errors with respect to the SEEK-FF are significantly reduced in most of the state variables and the oscillatory behaviour of the SEEK-FF is greatly removed, making it a promising alternative to the already established versions. (An overview over the rms errors of the O2-AA1 and O2-S£ experiments is given in Appendix C.)

Overall, the results for the performed TWIN experiments were very promising. The errors appear strongly reduced with respect to the errors of the free model run in particular in the state variables essential to the primary production (nutrients and phytoplankton). Some problems were observed in the initial error fields of microzooplankton and particulate organic matter, that can be attributed to the rather strong impact of the initial perturbation and were controlled rather quickly by the filter (microzooplankton).

However, the general validity of these results will have to be confirmed in other configurations, to begin with the assimilation of real observational data rather than the pseudo-observations in a TWIN framework.

A drawback in the implementation of the two filters SEEK-FF and SEEK-

Q is given by the connected computational effort and the demand in computational resources. While the cost for the SEEK-FF is roughly determined by the cost of a model run times the number of EOFs used plus the model forecast itself, the SEEK-Q requires even more time due to the additional matrix inversions and normalisations. These operations are performed in low rank reduction though, so that the additional cost is limited with respect to the model propagation of the complete set of EOFS. Also the demand in memory is highly increased for these filters as each EOF is of the size of the model state and the SEEK-Q variant even requires two sets. These issues might become significant, especially in the case of a three-dimensional implementation with the connected preprocessing statistics, so that an alternative way to access to additional computer power such as GRID computing might be favourable.

Appendix A

Ensemble error correlations

In this appendix a selection of sections of the correlation matrix is shown, obtained from the analysis performed on the ensemble runs described in section 2.3.4.

An overview over the major biogeochemical components of the model is given to illustrate the (initial) flow of information in the filter. The sections include the state variables of the observation vector O2 (see table 2.2) and an additional state representing the various elements of the trophic web.

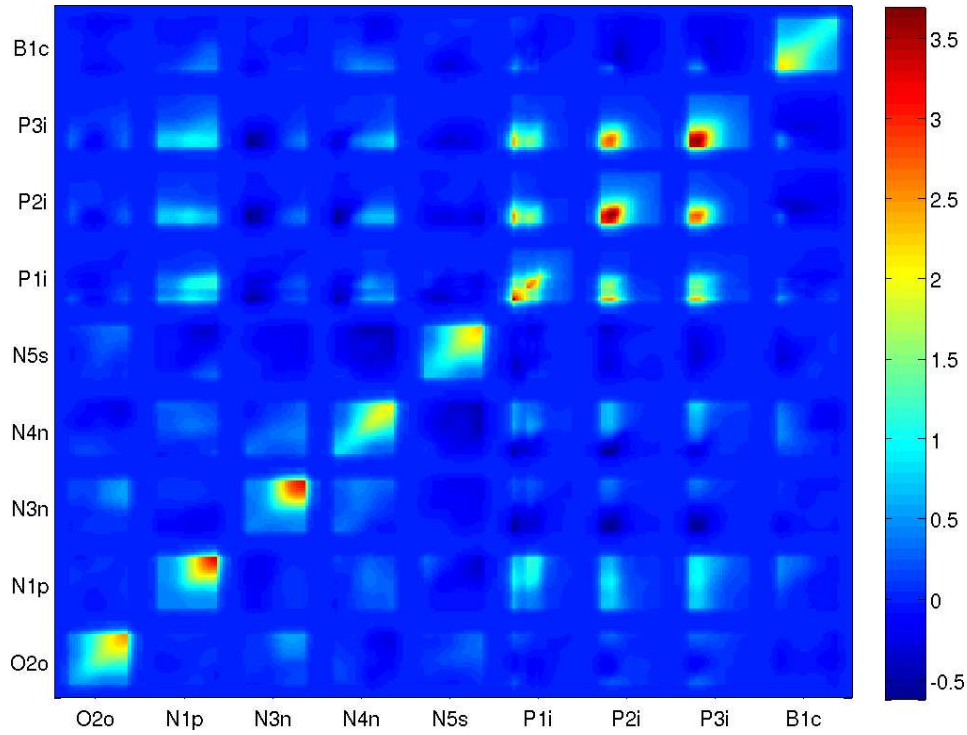


Figure 2.58: Correlation of the variables used in the observations O2 and bacteria carbon content. Section of the correlation matrix $\frac{\delta\phi * \delta\phi^{\mathrm{T}}}{N-1}$ in site AA1. State variables shown are: $O2o$ - Dissolved Oxygen, $N1p$- Phosphate, $N3n$ - Nitrate, Ammonium - $N4n$, Silicate - $N5s$, $P1i$ - Diatom chlorophyll-a content, $P2i$ - Nanoflagellates, $P3i$ - Picophytoplankton, $B1c$ - Bacteria carbon content. The matrix is organised by state variable blocks with the surface data for each state couple starting in the bottom left corner and depth increases to right and top. The data are normalised in a way that the mean of the main diagonal section taken for each state (the auto-correlations) is one, to guarantee comparable magnitudes between the various states, and maintaining variability distribution along the water column (see 2.3.4).

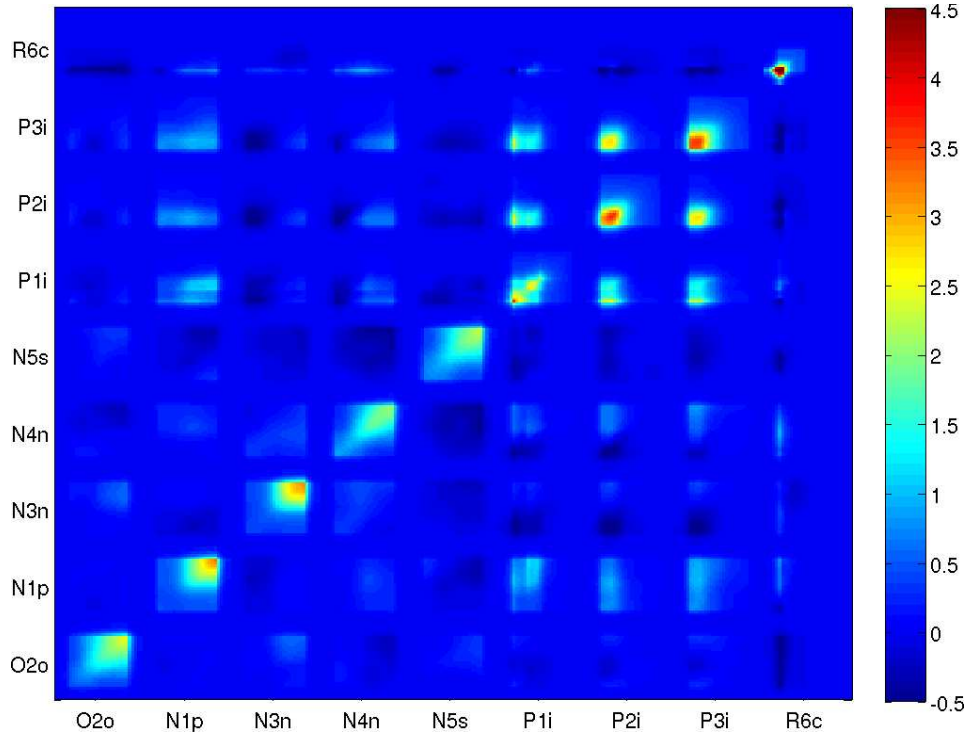


Figure 2.59: Correlation of the variables used in the observations O2 and particulate organic carbon. As fig. 2.59, state variables shown are: $O2o$ - Dissolved oxygen, $N1p$- Phosphate, $N3n$ - Nitrate, Ammonium - $N4n$, Silicate - $N5s$, $P1i$ - Diatom chlorophyll-a content, $P2i$ - Nanoflagellates, $P3i$ - Picophytoplankton, $R6c$ - Particulate organic carbon. Site AA1.

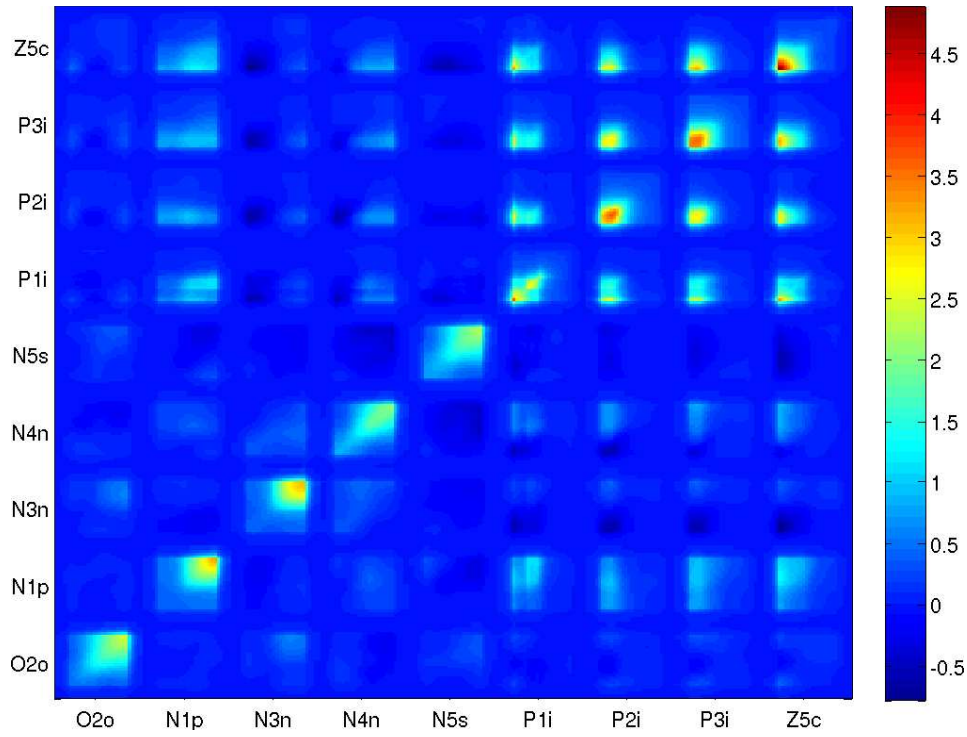


Figure 2.60: Correlation of the variables used in the observations O2 and microzooplankton carbon content. As fig. 2.59, state variables shown are: $O2o$ - Dissolved oxygen, $N1p$- Phosphate, $N3n$ - Nitrate, Ammonium - $N4n$, Silicate - $N5s$, $P1i$ - Diatom chlorophyll-a content, $P2i$ - Nanoflagellates, $P3i$- Picophytoplankton, $Z5c$ - Microzooplankton carbon content. Site AA1.

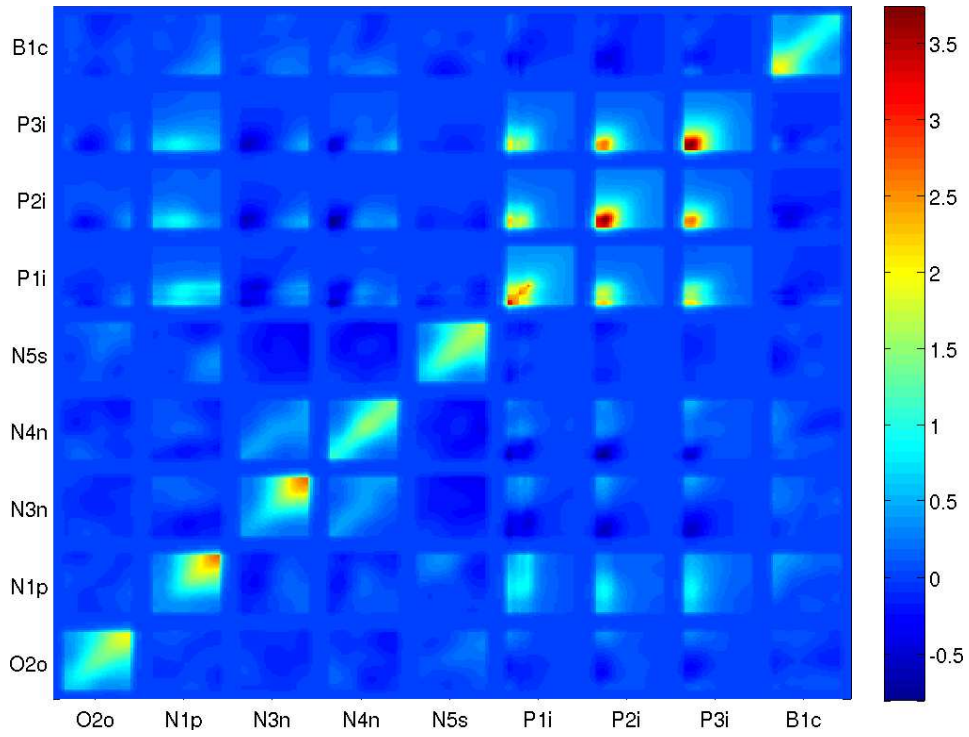


Figure 2.61: Correlation of the variables used in the observations O2 and bacteria carbon content. As in fig. 2.58 for site S3, state Variables shown are: $O2o$ - Dissolved oxygen, $N1p$- Phosphate, $N3n$ - Nitrate, Ammonium - $N4n$, Silicate - $N5s$, $P1i$ - Diatom chlorophyll-a content, $P2i$ - Nanoflagellates, $P3i$ - Picophytoplankton, $B1c$ - Bacteria carbon content.

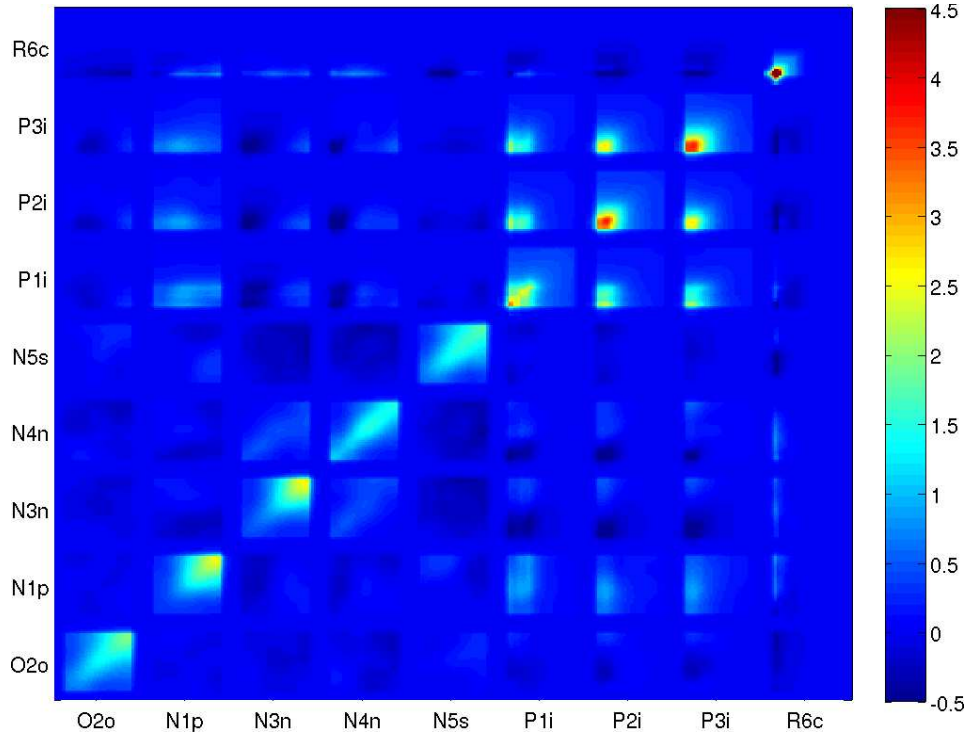


Figure 2.62: Correlation of the variables used in the observations O2 and particulate organic carbon. As in fig. 2.58 for site S3, state Variables shown are: $O2o$ - Dissolved oxygen, $N1p$- Phosphate, $N3n$ - Nitrate, Ammonium - $N4n$, Silicate - $N5s$, $P1i$ - Diatom chlorophyll-a content, $P2i$ - Nanoflagellates, $P3i$ - Picophytoplankton, $R6c$ - Particulate organic carbon.

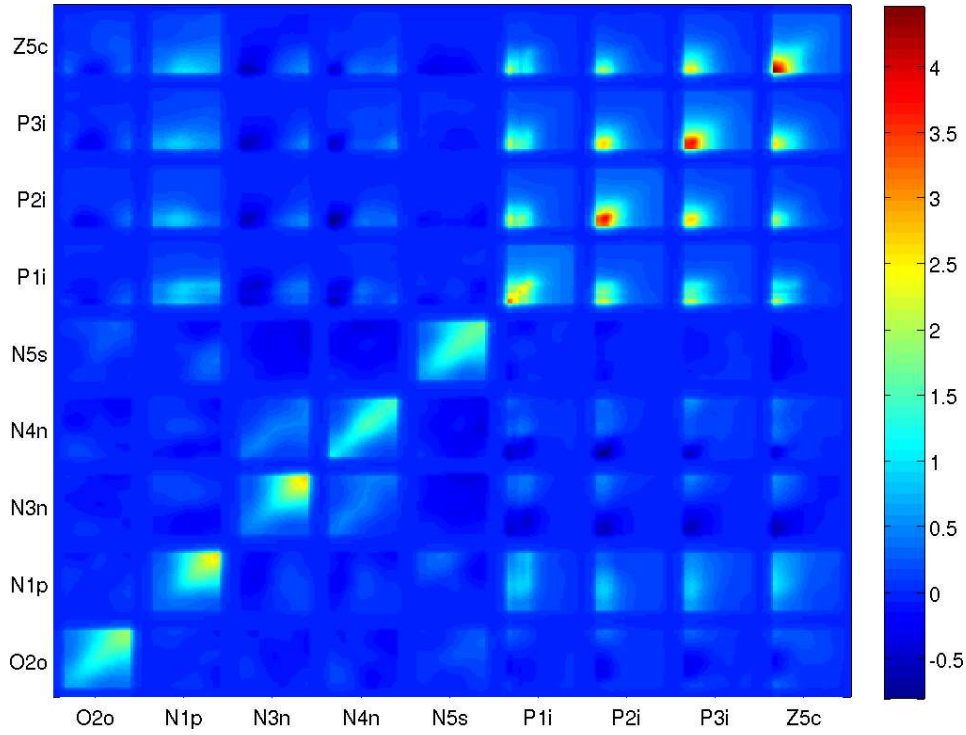


Figure 2.63: Correlation of the variables used in the observations O2 and microzooplankton carbon content. As in fig. 2.58 for site S3, state Variables shown are: $O2o$ - Dissolved oxygen, $N1p$- Phosphate, $N3n$ - Nitrate, Ammonium - $N4n$, Silicate - $N5s$, $P1i$ - Diatom chlorophyll-a content, $P2i$ - Nanoflagellates, $P3i$ - Picophytoplankton, $Z5c$ - Microzooplankton carbon content.

Appendix B

Derivation of the propagation equations for the SEEK-Q version

Following Pham et al. (1998) in the description of the general version of the *SEEK* Filter, the full analysis error covariance can be approximated expressing it in terms of a subset of eigenvectors $\boldsymbol{V}_i$ belonging to the first highest eigenvalues of $\mathcal{M}_i$:

$$\boldsymbol{P}_i^a = \boldsymbol{V}_i \boldsymbol{\Pi}_i \boldsymbol{V}_i^T \,. \tag{2.42}$$

He further stated that, for a slowly evolving model gradient $\mathcal{M}$ the iterative sequence

$$\boldsymbol{V}_i = \mathcal{M}\boldsymbol{V}\boldsymbol{\Pi}_i^{-1},\ \boldsymbol{\Lambda}_i = \boldsymbol{W}_{i-1}\mathcal{M}\boldsymbol{V}_{i-1},\ \boldsymbol{W}_{i-1} = \left(\boldsymbol{V}_{i-1}^T\boldsymbol{V}_{i-1}\right)^{-1}\boldsymbol{V}_{i-1}^T \tag{2.43}$$

converges starting from some $\boldsymbol{V}_0$ with appropriate rank towards the first eigenvectors of $\mathcal{M}$ and $\boldsymbol{\Lambda}$ against a matrix with the corresponding eigenvalues, such that $\mathcal{M}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{\Lambda}$. Thus, for a slowly evolving $\mathcal{M}_i$ this sequence can be used to propagate $\boldsymbol{V}$. [2] Supposing this, the covariance matrix expressed in terms of the eigenvectors evolves then according to

$$\boldsymbol{\Pi}_i^{-1} = \left(\boldsymbol{\Lambda}_i\boldsymbol{\Pi}_{i-1}\boldsymbol{\Lambda}_i + \boldsymbol{W}_i\boldsymbol{Q}_{i-1}\boldsymbol{W}_i^T\right)^{-1} + \boldsymbol{V}_i^T\mathcal{H}_i^T\boldsymbol{R}_i^{-1}\mathcal{H}_i\boldsymbol{V}_i \,. \tag{2.44}$$

Now, one can introduce the substitutions:

$$\boldsymbol{L}_i = \boldsymbol{V}_i\boldsymbol{\Lambda}_i\cdots\boldsymbol{\Lambda_1}\boldsymbol{N}_1\cdots\boldsymbol{N}_i \quad \boldsymbol{L}_0 = \boldsymbol{V}_0 \tag{2.45}$$

$$\boldsymbol{U}_i^{-1} = \boldsymbol{N}_i^T\cdots\boldsymbol{N}_1^T\boldsymbol{\Lambda}_1^T\cdots\boldsymbol{\Lambda}_i^T\boldsymbol{\Pi}_i^{-1}\boldsymbol{\Lambda}_i\cdots\boldsymbol{\Lambda}_1\boldsymbol{N}_1\cdots\boldsymbol{N}_i \quad \boldsymbol{U}_0 = \boldsymbol{\Pi}_0 \,. \tag{2.46}$$

where $\boldsymbol{N}$ are the invertible normalisation factors, as in section 2.6.

[2] Note, that this does not represent a further restriction to the system, as the extended Kalman Filter already involves the tangent linear approximation of the model operator $\mathcal{M}$.

As $\mathcal{M}_i \boldsymbol{V}_{i-1} = \boldsymbol{V}_i \boldsymbol{\Lambda}_i$(from eq. (2.43)), $\boldsymbol{L}_i$ propagates as

$$\boldsymbol{L}_i = \mathcal{M}_{i-1} \boldsymbol{L}_{i-1} \boldsymbol{N}_i \; . \tag{2.47}$$

Substituting eq.(2.44) into eq. (2.46), one obtains then

$$\boldsymbol{U}_i^{-1} = \left[\boldsymbol{N}_i^{-1} \boldsymbol{U}_{i-1} \boldsymbol{N}_i^{-T} + \boldsymbol{L}_i^T \boldsymbol{L}_0 \boldsymbol{S} \boldsymbol{L}_0^T \boldsymbol{L}_i\right]^{-1} + \boldsymbol{L}_i^T \mathcal{H}_i^T \boldsymbol{R}^{-1} \mathcal{H}_i \boldsymbol{L}_i \; .$$

as propagation of $\boldsymbol{U}_i$.

Here it was used, that by multiplication with $\boldsymbol{N}_i^T \cdots \boldsymbol{N}_1^T \boldsymbol{\Lambda}_1^T \cdots \boldsymbol{\Lambda}_i^T$ from the left and $\boldsymbol{\Lambda}_i \cdots \boldsymbol{\Lambda}_1 \boldsymbol{N}_1 \cdots \boldsymbol{N}_i$ from the right the terms of eq. (2.44) change as:

$$\begin{aligned}
\boldsymbol{\Lambda}_i^T \boldsymbol{\Pi}_{i-1} \boldsymbol{\Lambda}_i &\rightarrow \boldsymbol{N}_i^{-1} \boldsymbol{N}_{i-1}^{-1} \cdots \boldsymbol{N}_1^{-1} \boldsymbol{\Lambda}_1^{-1} \cdots \boldsymbol{\Lambda}_{i-1}^{-1} \boldsymbol{\Pi}_{i-1} \boldsymbol{\Lambda}_{i-1}^{-T} \cdots \boldsymbol{\Lambda}_1^{-T} \boldsymbol{N}_1^{-T} \cdots \boldsymbol{N}_{i-1}^{-T} \boldsymbol{N}_i^{-T} \; , \\
\left(\boldsymbol{V}_i^T \boldsymbol{V}_i\right)^{-1} \boldsymbol{V}_i^T &\rightarrow \boldsymbol{N}_i^{-1} \cdots \boldsymbol{N}_1^{-1} \boldsymbol{\Lambda}_1^{-1} \cdots \boldsymbol{\Lambda}_i^{-1} \left(\boldsymbol{V}_i^T \boldsymbol{V}_i\right)^{-1} \boldsymbol{V}_i^T \\
&= \left(\boldsymbol{V}_i^T \boldsymbol{V}_i \boldsymbol{\Lambda}_i \cdots \boldsymbol{\Lambda}_1 \boldsymbol{N}_1 \cdots \boldsymbol{N}_i\right) \boldsymbol{V}_i^T = \left(\boldsymbol{V}_i^T \boldsymbol{L}_i\right)^{-1} \boldsymbol{V}_i^T \\
&= \left(\boldsymbol{\Lambda}_{i-1}^{-T} \cdots \boldsymbol{\Lambda}_1^{-T} \boldsymbol{N}_1^{-T} \cdots \boldsymbol{N}_{i-1}^{-T} \boldsymbol{L}_i^T \boldsymbol{L}_i\right) \boldsymbol{V}_i^T \\
&= \left(\boldsymbol{L}_i^T \boldsymbol{L}_i\right)^{-1} \boldsymbol{L}_i^T = \boldsymbol{L}_i^T \\
\boldsymbol{W}_i \boldsymbol{Q} \boldsymbol{W}_i^T &\rightarrow \boldsymbol{L}_i^T \boldsymbol{Q} \boldsymbol{L}_i = \boldsymbol{L}_i^T \boldsymbol{L}_{ref} \boldsymbol{S} \boldsymbol{L}_{ref}^T \boldsymbol{L}_i
\end{aligned}$$

In summary, the new version of the *SEEK* Filter is given by the eq.:

$$\begin{aligned}
\boldsymbol{S} &= \boldsymbol{L}_0^T \boldsymbol{Q}_{ref} \boldsymbol{L}_0 = \beta \boldsymbol{L}_0^T \boldsymbol{P}^{f,ref} \boldsymbol{L}_0 \\
\boldsymbol{L}_i &= \mathcal{M}_{i-1} \boldsymbol{L}_{i-1} \boldsymbol{N}_i \\
\boldsymbol{U}_i^{-1} &= \left[\boldsymbol{N}_i^{-1} \boldsymbol{U}_{i-1} \boldsymbol{N}_i^{-T} + \boldsymbol{L}_i^T \boldsymbol{L}_0 \boldsymbol{S} \boldsymbol{L}_0^T \boldsymbol{L}_i\right]^{-1} + \boldsymbol{L}_i^T \mathcal{H}_i^T \boldsymbol{R}^{-1} \mathcal{H}_i \boldsymbol{L}_i \; .
\end{aligned}$$

Appendix C

Root mean square errors (rms) of the various filters

An overview over the time series of the mean analysis errors of the main variables of the *BFM* trophic web is given. The errors $\overline{\eta}$ for the three *SEEK* filters SEEK-Q, SEEK-FF and SEEK-0.5 are given as root mean square errors (rms), normalised by the reference state x^t

$$\eta = \sqrt{\left(\frac{x^a - x^t}{x^t}\right)^2}$$

and averaged over the water column

$$\overline{\eta} = \frac{1}{H}\int_0^{-H} \eta \mathrm{d}z \ ,$$

where H is the depth of the water column and x^a the computed analysis state.

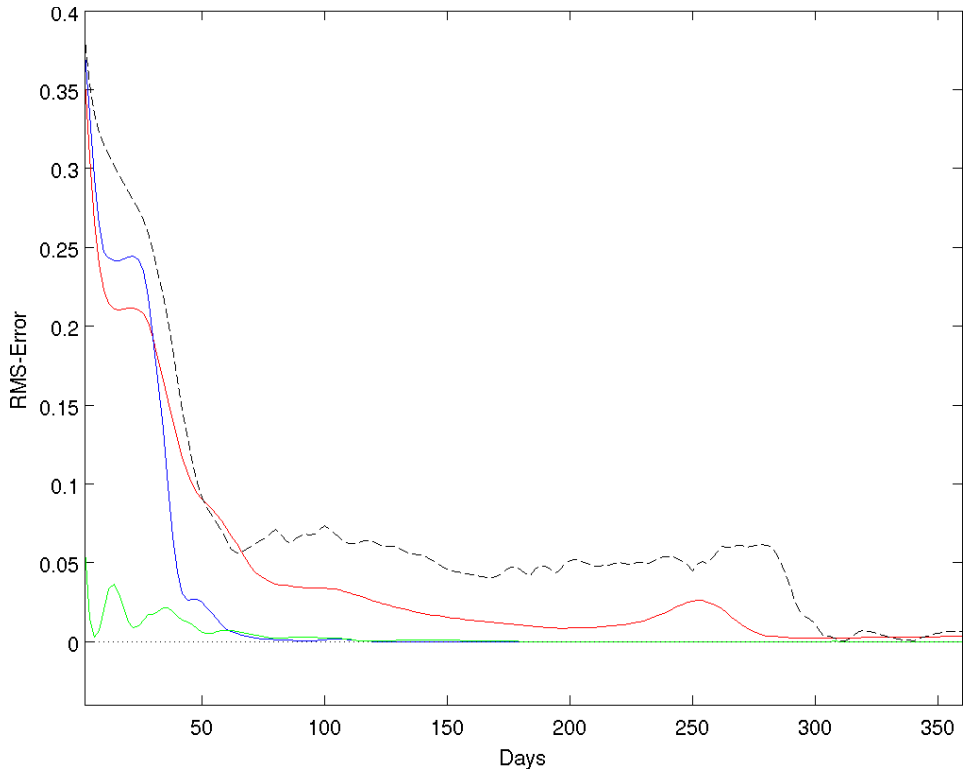


Figure 2.64: Column mean rms-error $\overline{\eta}$ for chlorophyll-a content in site AA1. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF; *Red line:* SEEK-0.5.

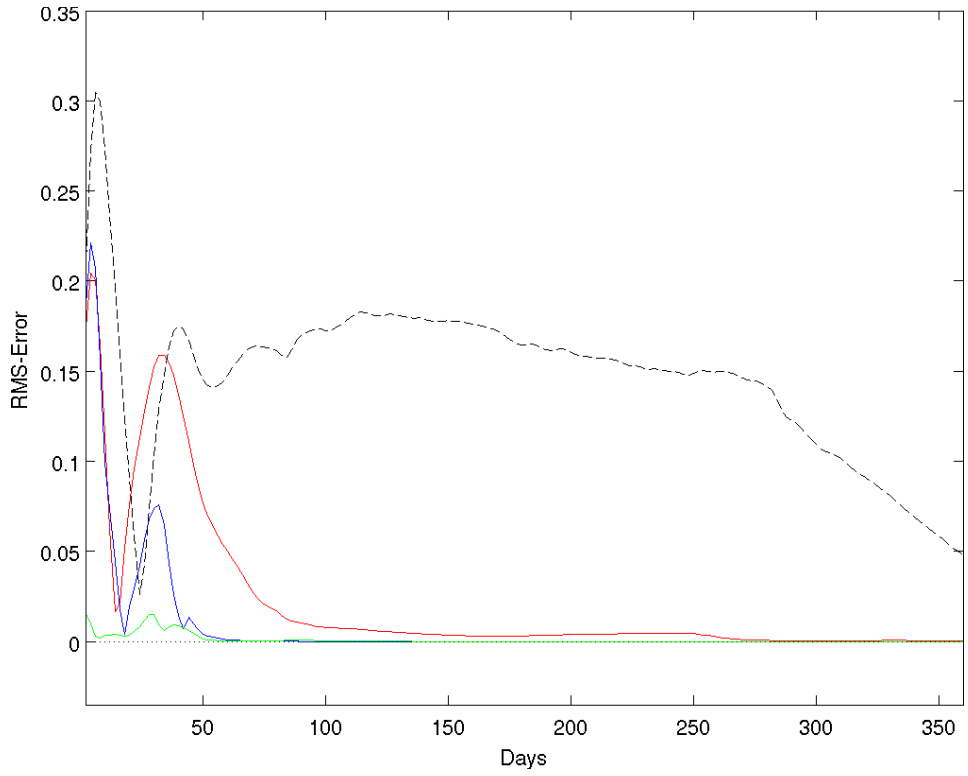


Figure 2.65: Column mean rms-error $\overline{\eta}$ for phosphate in site AA1. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF; *Red line:* SEEK-0.5.

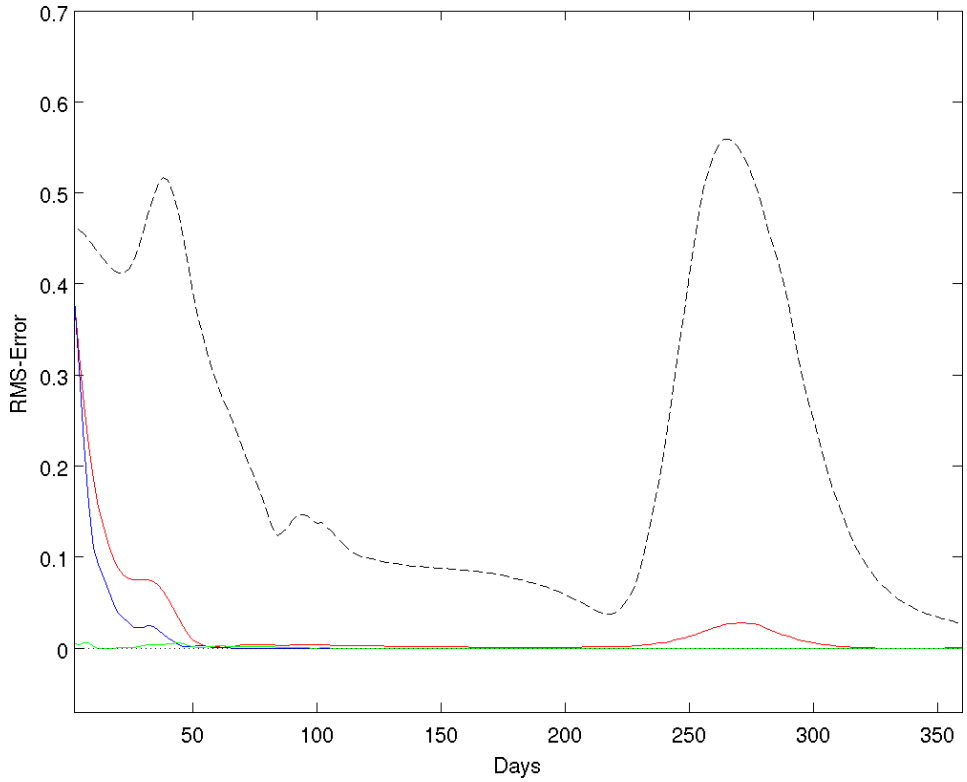


Figure 2.66: Column mean rms-error $\overline{\eta}$ for nitrate in site AA1. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF; *Red line:* SEEK-0.5.

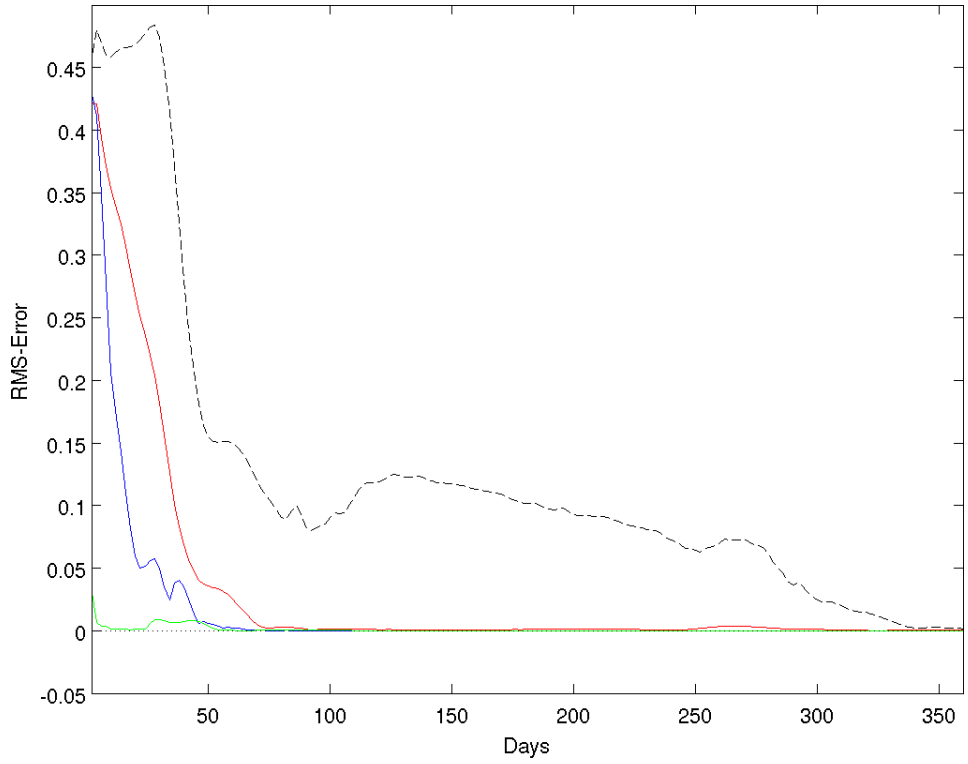


Figure 2.67: Column mean rms-error $\overline{\eta}$ for ammonium in site AA1. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF; *Red line:* SEEK-0.5.

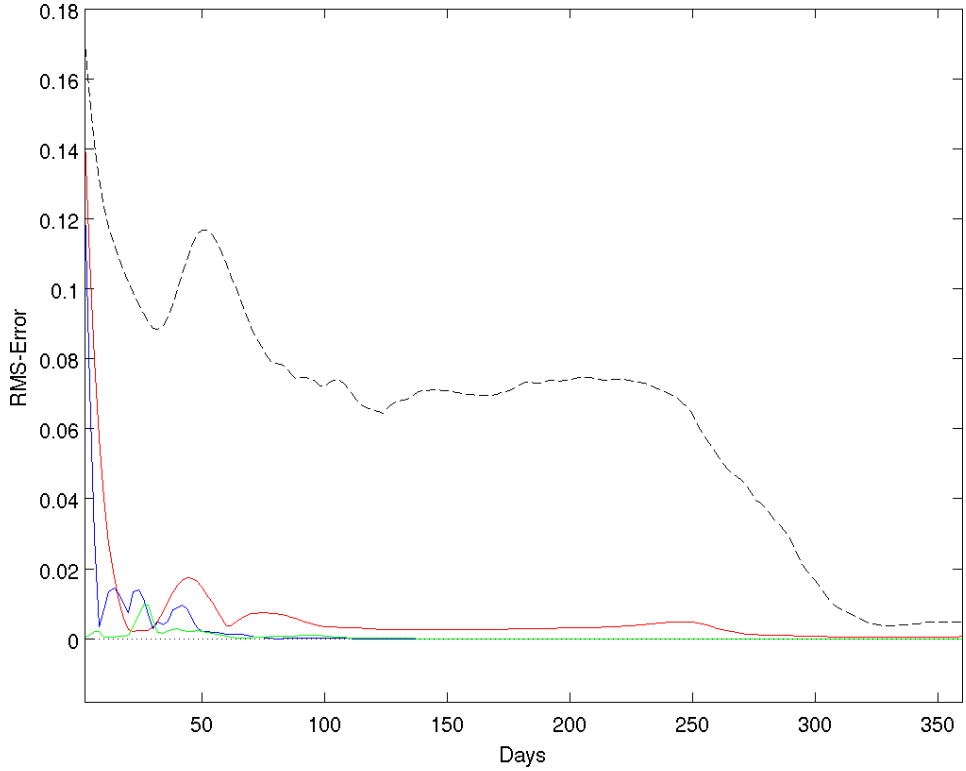


Figure 2.68: Column mean rms-error $\overline{\eta}$ for silicate in site AA1. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF; *Red line:* SEEK-0.5.

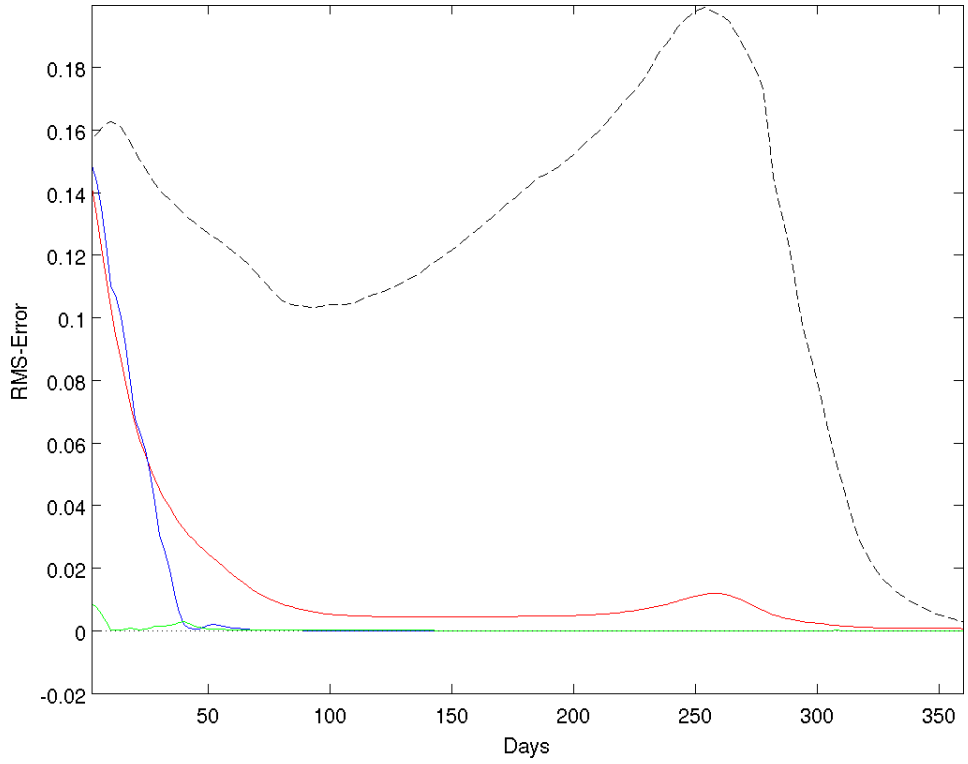


Figure 2.69: Column mean rms-error $\overline{\eta}$ for dissolved oxygen in site AA1. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF; *Red line:* SEEK-0.5.

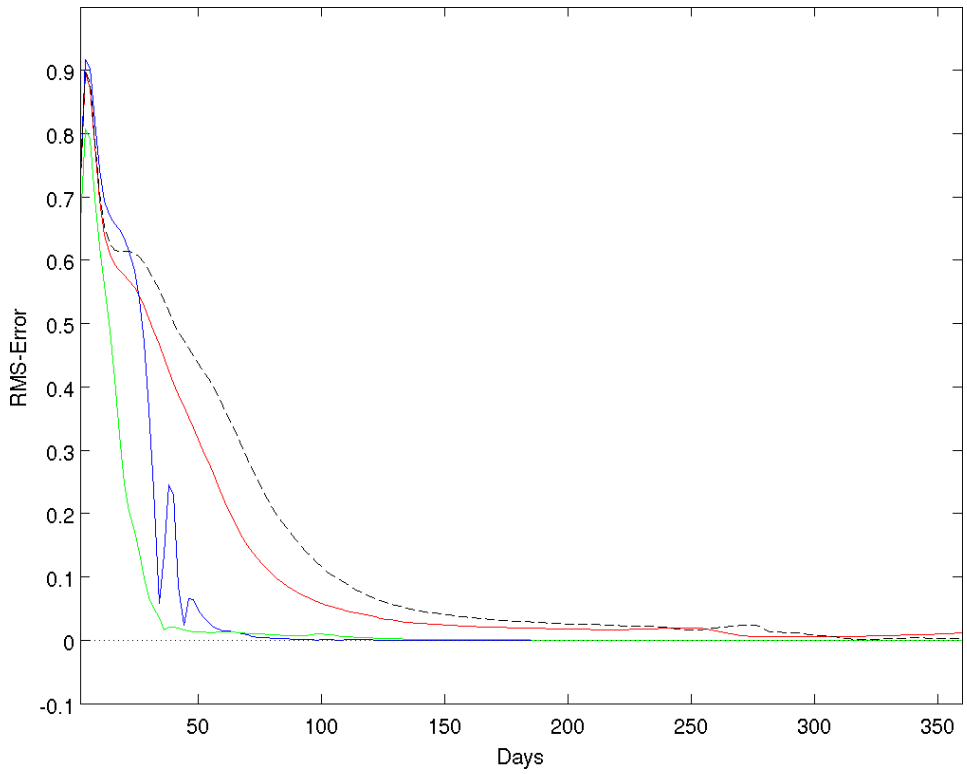


Figure 2.70: Column mean rms-error $\overline{\eta}$ for bacteria carbon content in site AA1. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF; *Red line:* SEEK-0.5.

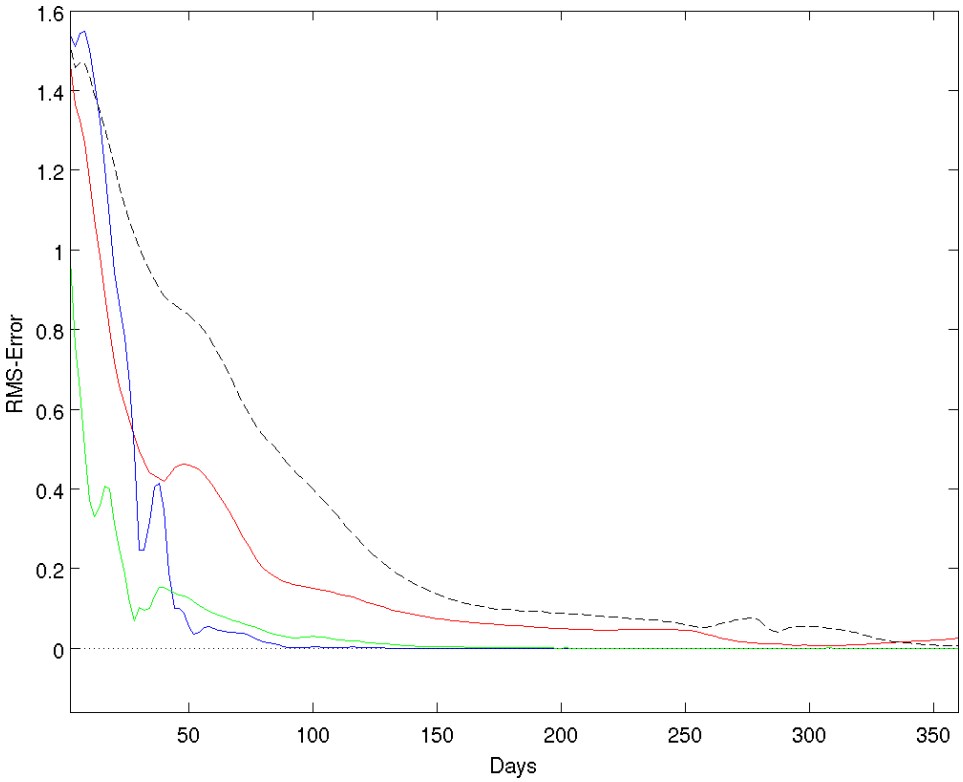


Figure 2.71: Column mean rms-error $\overline{\eta}$ for particulate organic carbon in site AA1. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF; *Red line:* SEEK-0.5.

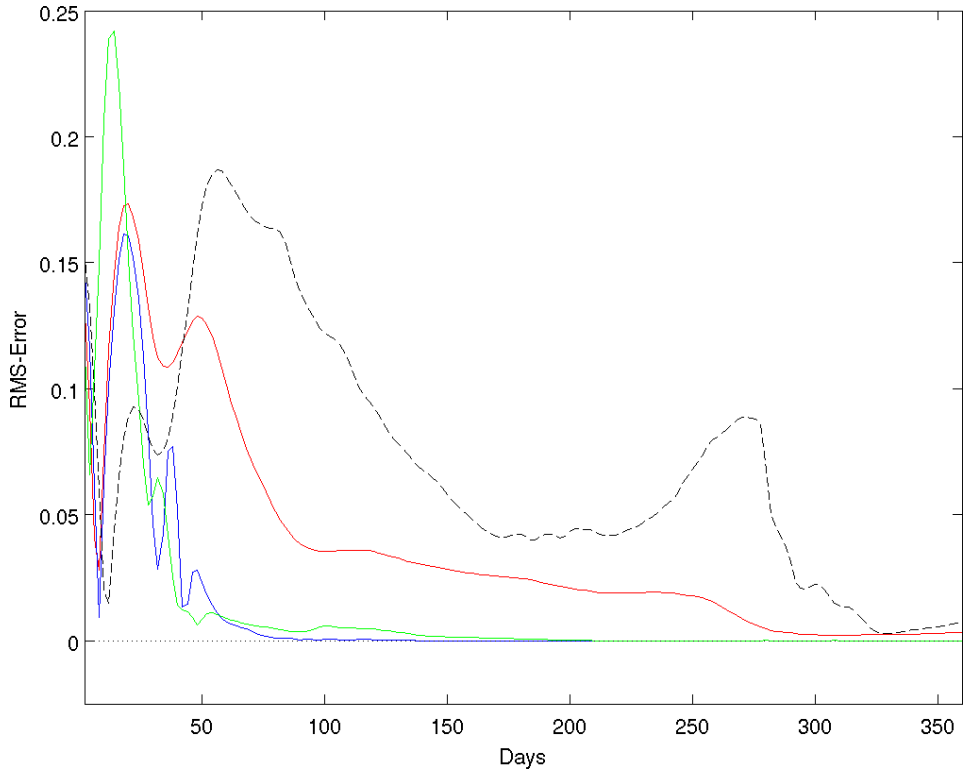


Figure 2.72: Column mean rms-error $\overline{\eta}$ for microzooplankton carbon content in site AA1. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF; *Red line:* SEEK-0.5.

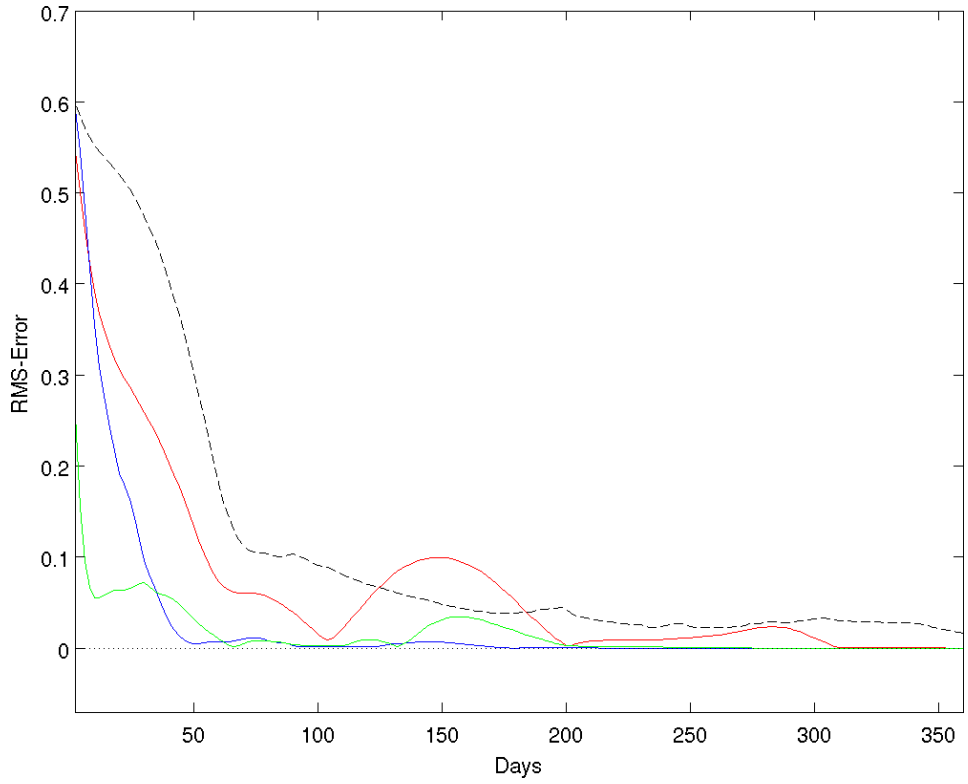


Figure 2.73: Column mean rms-error $\overline{\eta}$ for chlorophyll-a content in site S3. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF; *Red line:* SEEK-0.5.

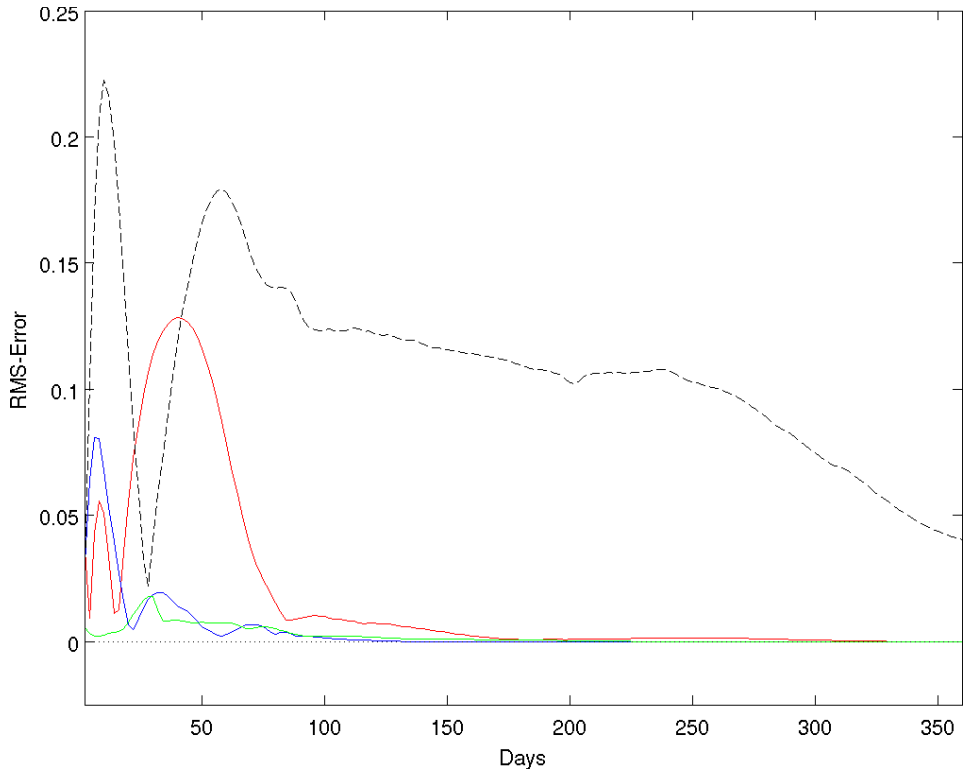


Figure 2.74: Column mean rms-error $\overline{\eta}$ for phosphate in site S3. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF; *Red line:* SEEK-0.5.

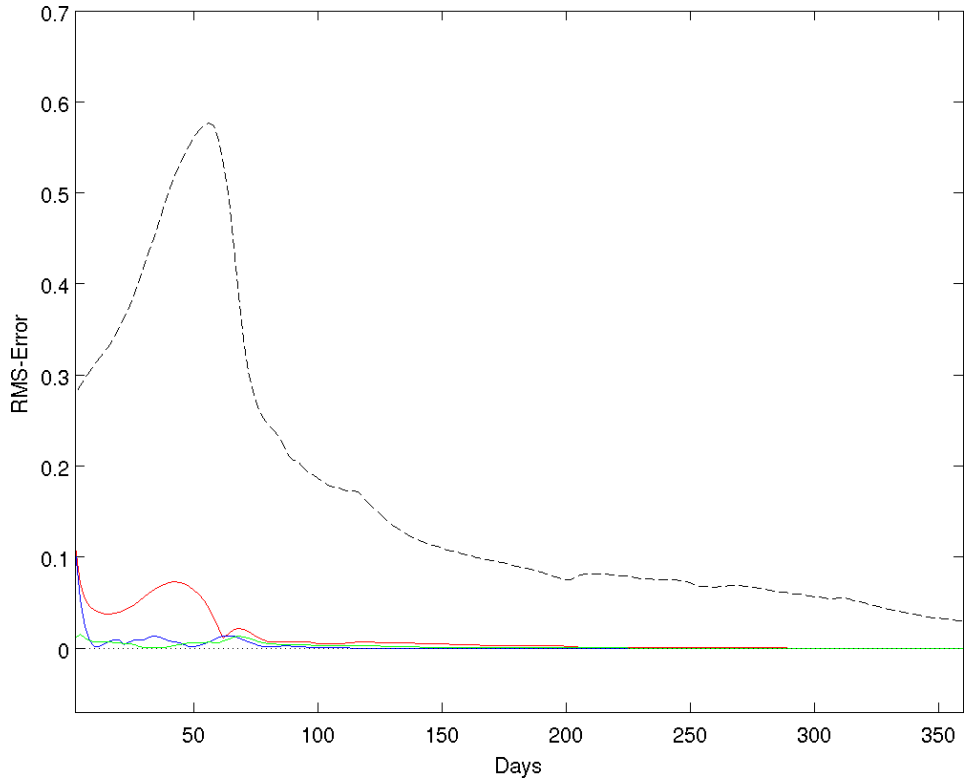


Figure 2.75: Column mean rms-error $\overline{\eta}$ for nitrate in site S3. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF; *Red line:* SEEK-0.5.

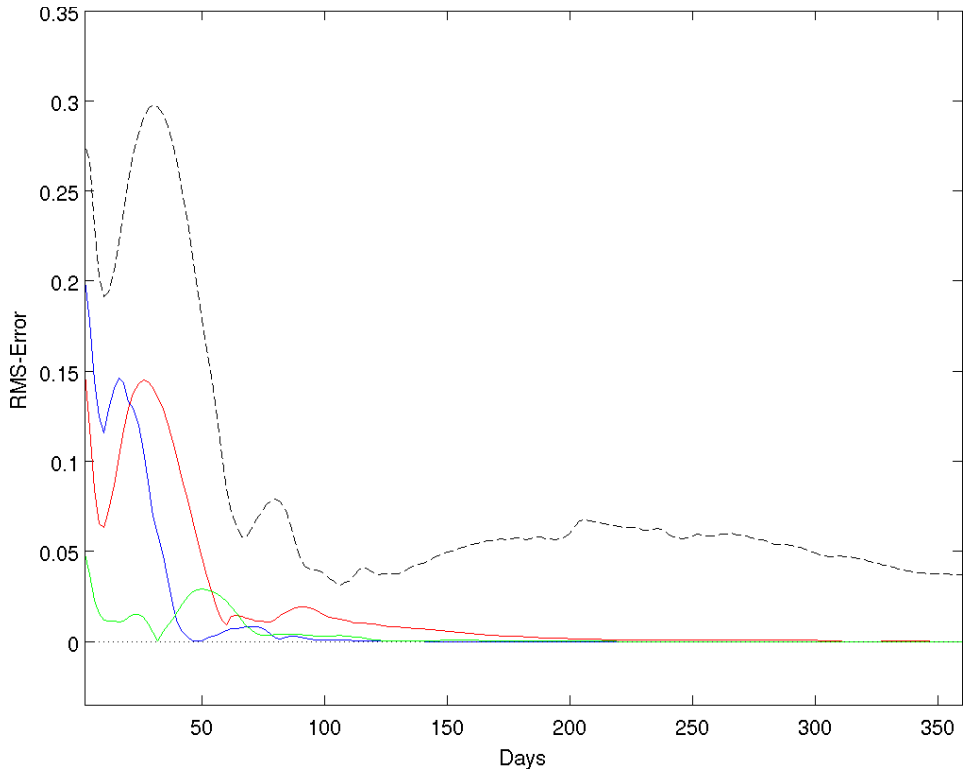


Figure 2.76: Column mean rms-error $\overline{\eta}$ for ammonium in site S3. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF; *Red line:* SEEK-0.5.

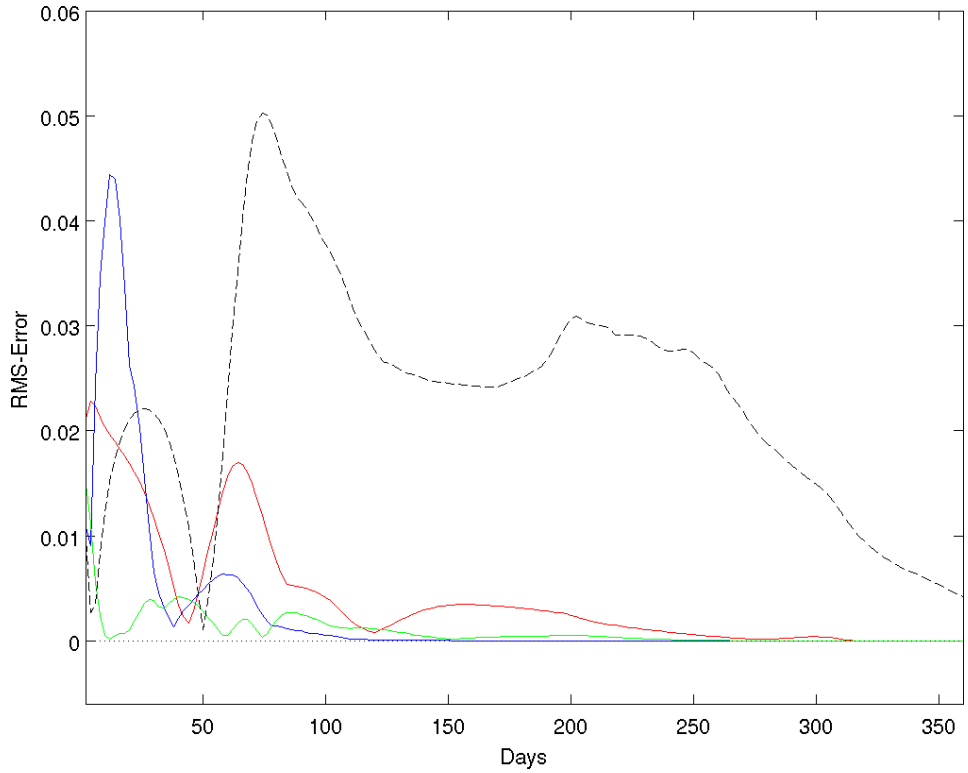


Figure 2.77: Column mean rms-error $\overline{\eta}$ for silicate in site S3. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF; *Red line:* SEEK-0.5.

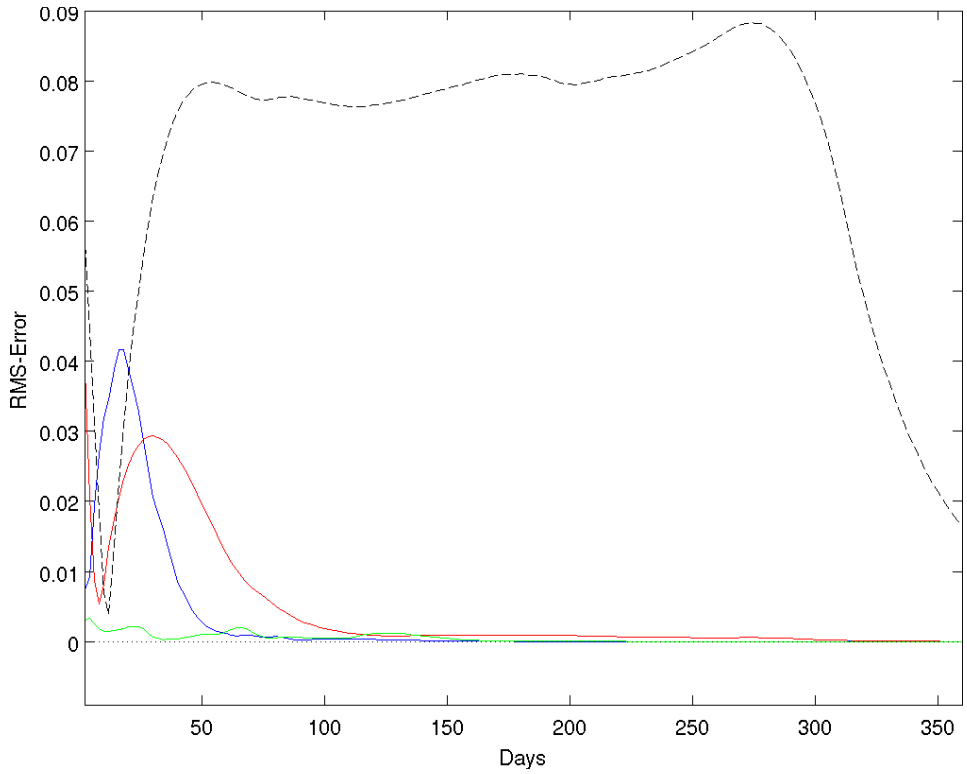


Figure 2.78: Column mean rms-error $\overline{\eta}$ for dissolved oxygen in site S3. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF; *Red line:* SEEK-0.5.

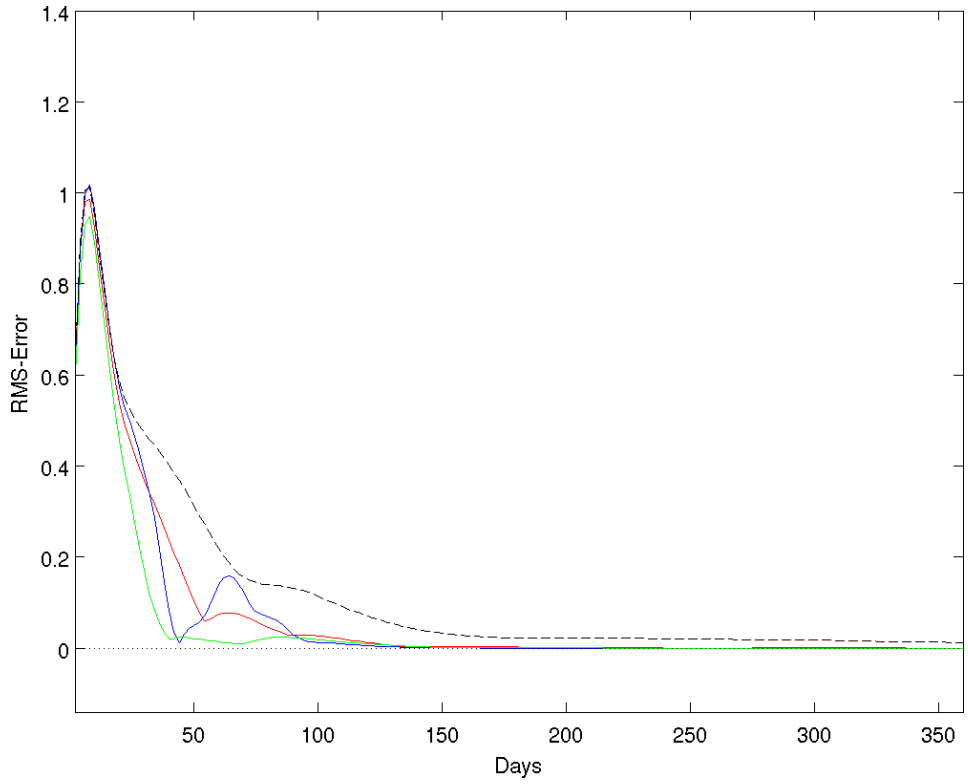


Figure 2.79: Column mean rms-error $\overline{\eta}$ for bacteria carbon content in site S3. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF; *Red line:* SEEK-0.5.

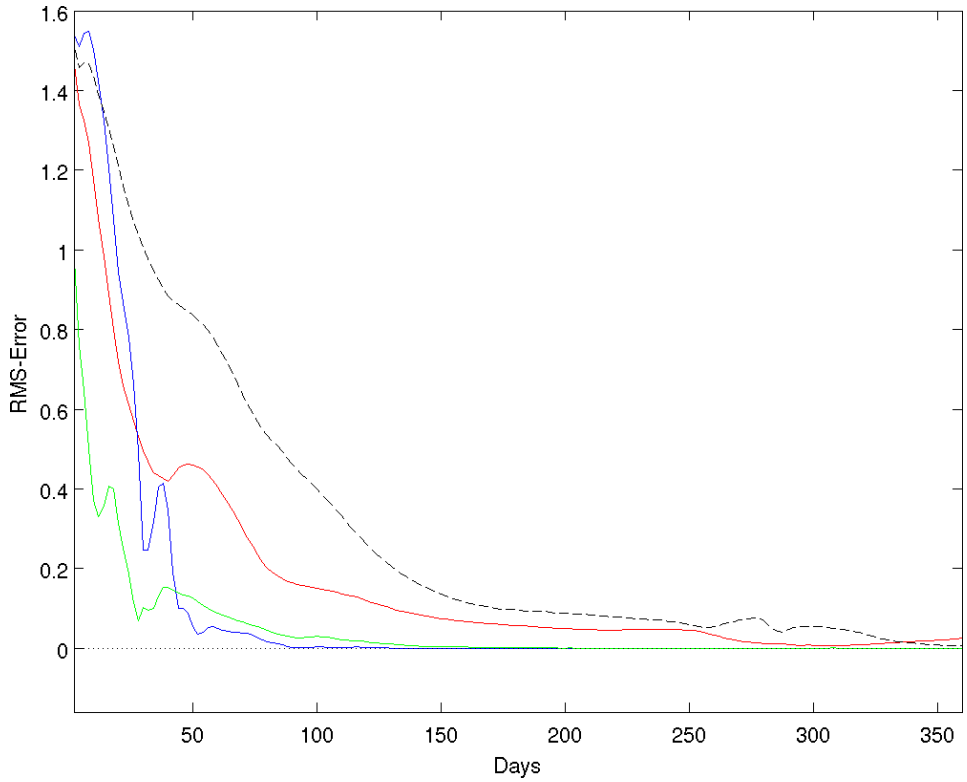


Figure 2.80: Column mean rms-error $\overline{\eta}$ for particulate organic carbon in site S3. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF; *Red line:* SEEK-0.5.

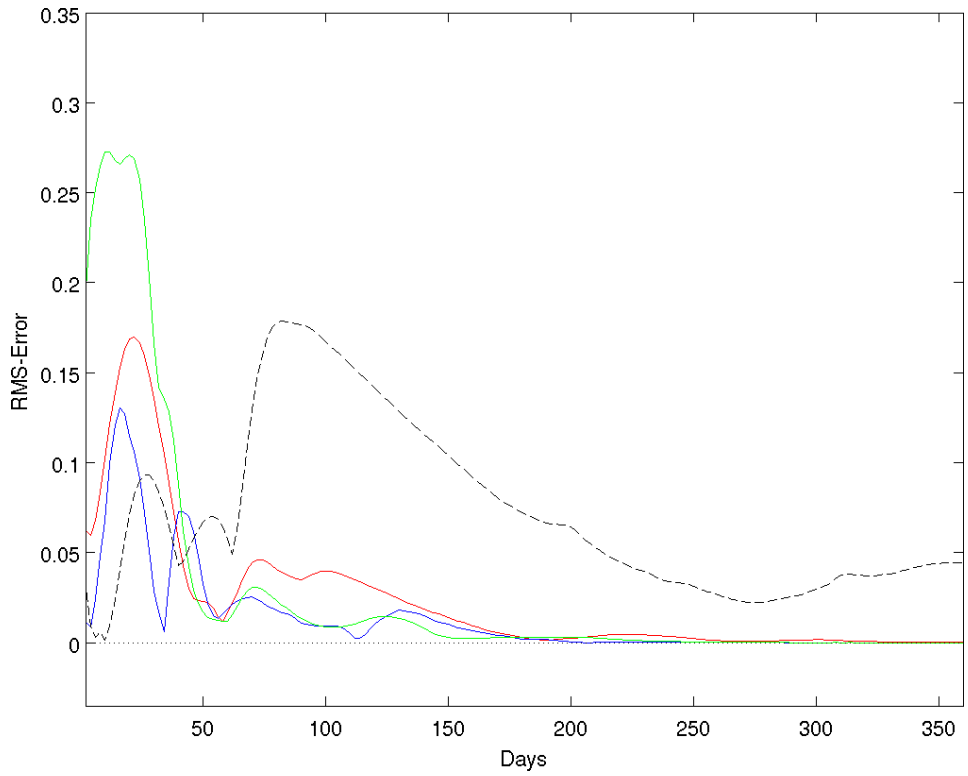


Figure 2.81: Column mean rms-error $\overline{\eta}$ for microzooplankton carbon content in site S3. *Black dashed line:* free model run; *Green line:* SEEK-Q; *Blue line:* SEEK-FF; *Red line:* SEEK-0.5.

Bibliography

Artegiani, A., Bregant, D., Paschini, E., Pinardi, N., Raicich, F., and Russo, A. (1997a). The Adriatic Sea general circulation. Part I. Air-sea interactions and water mass structure. *J. Phys. Oceanogr.*, 27:1492–1514.

Artegiani, A., Bregant, D., Paschini, E., Pinardi, N., Raicich, F., and Russo, A. (1997b). The Adriatic Sea general circulation. Part II. Baroclinic circulation structure. *J. Phys. Oceanogr.*, 27:1515–1532.

Asselin, R. (1972). Frequency filters for time integration. *Mon. Weather Rev.*, 100:487–490.

Baretta, J. W., Ebenhöh, W., and Ruardij, R. (1995). The European Regional Seas Ecosystem Model, a complex marine ecosystem model. *Netherlands Journal of Sea Research*, 33:233–246.

Baretta-Bekker, J. G., Baretta, J. W., Hansen, A. S., and Reimann B. (1998). An improved model of carbon and nutrient dynamcis in the microbial foodweb in marine enclosures. *Aquatic Microbial Ecology*, 14:91–108.

Baretta-Bekker, J. G., Baretta, J. W., and Rasmussen, E. K. (1995). The microbial food web in the European Regional Seas Ecosystem Model. *Netherlands Journal of Sea Research*, 33:363–379.

Bennett, A. F. (1992). *Inverse Methods in Physical Oceanography*. Cambridge University Press.

Berline, L., Brankart, J.-M., Brasseur, P., Ourmières, Y., and Verron, J. (2007). Improving the physics of a coupled physical-biogeochemical

model of the North Atlantic through data assimilation: Impact on the ecosystem. *Journal of Marine Systems*, 64(1-4):153–172.

Blom, J. G. and Verwer, J. G. (2000). A comparison of integration methods for atmospheric transport-chemistry problems. *J. Comput. Appl. Math.*, 126(1-2):381–396.

Blumberg, A. and Mellor, G. L. (1987). A description of a three-dimensional coastal ocean circulation model. Technical Report 1-16, American Geophysical Union. N. S. Heaps (Ed.).

Brasseur, P., Ballabrera-Poy, J., and Verron, J. (1999). Assimilation of altimetric data in the mid-latitude oceans using the Singluar Evolutive Extended Kalman filter with an eddy-resolving, primitive equation model. *Journal of Marine Systems*, 22:269–294.

Brasseur, P. and Verron, J. (2006). The SEEK filter method for data assimilation in oceanography: a synthesis. *Ocean Dynamics*, 56:650–661.

Burchard, H., Deleersnijder, E., and Meister, A. (2003). A high-order conservative patankar-type discretisation for stiff systems of production-destruction equations. *Appl. Numer. Math.*, 47(1):1–30.

Carmichael, G. R., Sandu, A., Potra, F., Damian-Iordache, V., and Damian-Iordache, M. (1996). The current state and future directions in air quality modeling. In Sydow, A., Carmichael, G. R., and Korn, G., editors, *Modelling and Simulation of Complex Environmental Problems*. Springer-Verlag.

Cataletto, B., Feoli, E., Fonda Umani, S., and Cheng-Yong, S. (1995). Eleven years of time-series analysis on the net-zooplankton community in the Gulf of Trieste. *ICES J. Mar. Sc.*, 52:669–687.

Cozzi, S., Catalano, G., and Lipizer, M. (1999). Importanza e distribuzione spazio-temporale delle frazioni organiche disciolte dell'azoto e del fosforo nel bacino del nord Adriatico. *Atti della Associazione Italiana di Oceanologia e Limnologia*, 13(1):21–32.

Daley, R. (1991). *Atmospheric Data Analysis*. Camebridge University Press.

Dobricic, S., Pinardi, N., Adani, M., Bonazzi, A., Fratianni, C., and Tonani, M. (2005). Mediterranean Forecasting System: an improved assimilation scheme for sea-level anomaly and its validation. *Q. J. R. Meteorol. Soc*, 131:3627–3642.

Droop, M. R. (1973). Some thoughts on nutrient limitation in algae. *Journal of Phycolgy*, 9:264–272.

Evensen, G. (1992). Using the extended Klaman filter with a multilayer quasi-geostrophic ocean model. *J. Geophys. Res.*, 97(17):905–17,924.

Evensen, G. (2003). The Ensemble Kalman Filter: theoritical formulation and practical implementation. *Ocean Dynamics*, 53:343–367.

Evensen, G. (2006). *Data Assimilation*. Springer.

Fonda Umani, S., Sun, C., Feoli, E., Cataletto, B., Cabrini, M., and Milani, L. (1995). Is it possible to identify any plankton succession in the Gulf of Trieste (Northern Adriatic Sea)? In Eleftheriou, A., Ansell, A., and Smith, C., editors, *Biology and Ecology of Shallow Coastal Waters*, pages 59–65. 28th European Marine Biology Symposium.

Gandin, L. (1963). Objective analysis of meteorological fields. *Gidromet*.

Geider, R. J., MacIntyre, H. L., and Kana, T. M. (1997). A dynamic model of phytoplankton growth and acclimation: responses of the balanced growth rate and chlorophyll a:carbon ratio to light, nutrient limitation and temperature. *Mar. Ecol. Progr. Ser.*, 148:187–200.

Gelb, A. (1974). *Applied Optimal Estimation*. MIT Press.

Giani, M., Boldrin, A., Matteucci, G., Frascari, F., Gismondi, M., and Rabitti, S. (2001). Downward fluxes of particulate carbon, nitrogen and phosphorus in the north-western Adriatic Sea. *Sc. Tot. Env.*, 266:125–134.

Giani, M., Gismondi, M., Savelli, F., Boldrin, A., and Rabitti, S. (2000). Carbonio organico, azoto e fosforo particellati nell'Adriatico settentrionale: distribuzione verticale e variabilità temporale. *Atti della Associazione Italiana di Oceanologia e Limnologia*, 13(2):55–65.

Gibson, J., Kallberg, P., Uppala, S., Nomura, A., and Serrano, E. (1997). Era description. Technical report, ECMWF.

Gilchrist, B. and Cressman, G. (1954). An experiment in objective analysis. *Tellus*, 6:309–318.

GLOBEC (1993). Sampling and Observational Systems: Report of the First Meeting of an International GLOBEC Working Group. GLOBEC Report No. 3. Technical report, SCOR.

Graedel, T. E. and Crutzen, P. J. (1995). *Atmosphere, Climate and Change*. Scinetific American Library.

Haitvogel, D. B., Blanton, J., Kindle, J. C., and Lynch, D. R. (2000). Coastal Ocean Modeling: Processes and Real-Time Systems. *Oceanography*, 13(1):35–46.

Hofmann, E. and Lascara, C. M. (1998). Overview of interdisciplinary modeling for marine ecosystems. In Brink, K. H. and Robinson, A. R., editors, *The Sea*, volume 10, chapter 19, pages 507–540. John Wiley & Sons.

Hoteit, I., Triantafyllou, G., Petihakis, G., and Allen, J. I. (2003). A singular evolutive extended Kalman filter to assimilate real in situ data in a 1-D marine ecosystem model. *Annales Geophysicae*, 21:389–397.

Hoteit, I., Tryantafyllou, G., and Petihakis, G. (2004). Towards a data assimilation system for the Cretan Sea ecosystem using a simplified Kalman filter. *Journal of Marine Systems*, 65(1-4):84–104.

Houtekamer, P., Mitchell, H., Pellerin, G., Buehner, M., Charron, M., Spacek, L., and Hansen, B. (2005). Atmospheric Data Assimilation with

an Ensemble Kalman Filter: Results with Real Observations. *Monthly Weather Review*, 133:604–620.

Jacobson, M. Z. (2005). *Fundamentals of Atmospheric Modeling*. Camebridge University Press, 2nd edition.

Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory*. Academic Press.

Johannessen, J. A., Le Trayon, P.-Y., Robinson, I., Nittis, K., Bell, M. J., Pinardi, N., and Bahurel, P. (2006). Marine environment and security for the European area - Toward operational oceanography. *Bulletin of the American Meteorological Society*, 87(8).

Kalman, R. E. (1960). A new approach to linear filter and prediction problems. *J. Bas. Eng.*, 82:35–45.

Legendre, L. and Rassoulzadegan, F. (1995). Plankton and nutrient dynamics in marine waters. *Ophelia*, 41:153–172.

Maggiore, A., Zavatarelli, M., Angelucci, M., and Pinardi, N. (1998). Surface heat and water fluxes in the Adriatic Sea: Seasonal and interannual variability. *Phys. Chem. Earth*, 23:561–567.

Magri, S., Brasseur, P., and Lacroix, G. (2005). Assimilation de donnèes dans un modèle d'ecosystème marin de la mer Ligure. *C. R. Geoscience*, 337:1065–1074.

Malej, A., Mozetic, P., Malacic, V., Terzic, S., and Ahel, M. (1995). Phytoplankton responses to freshwater inputs in a small semi-enclosed gulf (Gulf of Trieste, Adriatic Sea). *Mar. Ecol. Prog. Ser.*, 120:111–121.

Malone, T. C. and Cole, M. (2000). Toward a Global Scale Coastal Ocean Observing System. *Oceanography*, 13(1):7–11.

Mann, K. H. and Lazier, J. R. N. (1991). *Dynamics of marine ecosystems*. Blackwell Science, 1st edition.

McPhaden, M. J. (1998). The Tropical Ocean-Global Atmosphere observing system: A decade of progress. *Journal of Geophysical Research*, 103(C7):14169–14240.

McRae, G. J., Goodwin, W. R., and Seinfeld, J. H. (1982). Numerical solution of the atmospheric diffusion equation for chemically reacting flows. *J. Comp. Phys*, 45:1–42.

Mellor, G. L. (2004). *Users Guide For a Three-Dimensional, Primitive Equation, Numerical Ocean Model*.

Mellor, G. L. and Yamada, T. (1982). Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys. and Space Phys.*, 20(4):851–875.

Miller, C. B. (2006). *Biological Oceanography*. Blackwell Science, 1st edition.

Miller, R. N. (1994). Perspective on advanced data assimilation in strongly nonlinear systems. In Brasseur, P. P. and Nihoul, J. C. J., editors, *Data Assimilation: Tools for Modelling the Ocean in a Global Change perspective*, volume I of *NATO ASI*, pages 195–216. Springer.

Mitchell, H. and Houtekamer, P. (2000). An Adaptive Ensemble Kalman Filter. *Monthly Weather Review*, 128:416–433.

Mozetic, P., Fonda Umani, S., Cataletto, B., and Malej, A. (1998). Seasonal and inter-annual plankton variability in the Gulf of Trieste in the Northern Adriatic. *ICES J. Mar. Sc*, 55:711–722.

Navarra, A. (1996). *Le previsioni del tempo*. il Saggiatore.

Nihoul, J. C. J. and Djenidi, S. (1998). Coupled physical, chemical and biological models. In Brink, K. H. and Robinson, A. R., editors, *The Sea*, volume 10, chapter 18, pages 483–506. John Wiley & Sons.

Nyholm, N. (1977). Kinetics of nitrogen-limited algal growth. *Progress in Water Technology*, 8:347–358.

Oddo, P., Pinardi, N., and Zavatarelli, M. (2005). A numerical study of the interannual variability of the Adriatic Sea (2000-2002). *Sci. Total. Environ.*, 353:39–56.

Panofsky, H. (1949). Objective weather-map analysis. *J. Appl. Meteor.*, 6:386–92.

Pettine, M., Patrolecco, L., Carusso, M., and Crescenzio, S. (1998). Transport of carbon and nitrogen to the northern Adriatic Sea. *Est. Coast. Shelf Sci*, 46:127–142.

Pham, D., Verron, J., and Roubaud, M. (1998). A singular evolutive extended Kalman filter for data assimilation in oceanography. *Journal of Marine Systems*, 16:323–340.

Pinardi, N., Allen, I., Demirov, E., De Mey, P., Korres, G., Lascaratos, A., Le Traon, P.-Y., Maillard, C., Manzella, G., and Tziavos, C. (2005). The Mediterranean ocean forecasting system: first phase of implementation (1998-2001). *Annales Geophysicae*, 21:3–20.

Polimene, L., Pinardi, N., Zavatarelli, M., Allen, J. I., Giani, M., and Vichi, M. (2007). A numerical simulation study of dissolved organic carbon accumulation in the northern Adriatic Sea. *Journal Of Geophysical Research*, 112.

Polimene, L., Pinardi, N., Zavatarelli, M., and Colella, S. (2006). The Adriatic Sea ecosystem seasonal cycle: Validation of a three-dimensional numerical model. *Journal of Geophysical Research*, 111.

Raick, C., Alvera-Azcarate, A., Barth, A., Brankart, J. M., Soetaert, K., and Grégoire, M. (2007). Application of a SEEK filter to a 1D biogeochemical model of the Ligurian Sea: Twin experimetns and real in-situ data assimilation. *Journal of Marine Systems*, 65(1-4):561–583.

Richardson, L. F. (1922). *Weather prediction by numerical process*. Camebridge University Press.

Riley, G., Stommel, H., and Bumpus, D. F. (1949). Quantitative ecology of the plankton of the Western-North Atlantic. *Bull. Bingham Oceanogr. Coll.*, 12:1–169.

Sportisse, B. (2000). An analysis of operator splitting techniques in the stiff case. *J. Comput. Phys.*, 161(1):140–168.

Strang, G. (1968). On the construction and comparison of difference schemes. *SIAM J. Numer. Anal.*, 5:506–517.

Sun, P. (1996). A pseudo-non-time-splitting method in air quality modeling. *J. Comput. Phys.*, 127(1):152–157.

Supic, N. and Orlic, M. (1999). Seasonal and interannual variability of the northern Adriatic surface fluxes. *J. Mar. Sys.*, 20:205–229.

Sverdrup, H. U. (1953). On conditions for the vernal blooming of phytoplankton. *J. Cons. Perm. Int. Exp. Mer.*, 18:287–195.

Verwer, J., Hundsdorfer, W., and Blom, J. (2001). Numerical time integration for air pollution models.

Vichi, M., Masina, S., and Navarra, A. (2007a). A generalized model of pelagic biogeochemistry for the global ocean ecosystem. Part II: Numerical simulations. *Journal of Marine Systems*, 64(1-4):110–134.

Vichi, M., Masina, S., Patara, L., Baretta, J. W., Ruardij, P., Pinardi, N., Zavatarelli, M., Polimene, L., and Butenschön, M. (2006). The Biogeochemical Flux Model (BFM). Technical report, http://www.bo.ingv.it/bfm/.

Vichi, M., Oddo, P., Zavatarelli, M., Coluccelli, A., Coppini, G., Celio, M., Fonda Umani, S., and Pinardi, N. (2003). Calibration and validation of a one-dimensional complex marine biogeochemical flux model in different areas of the northern Adriatic shelf. *Annales Geophysicae*, 21:413–436.

Vichi, M., Pinardi, N., and Masina, S. (2007b). A generalized model of pelagic biogeochemistry for the global ocean ecosystem. Part I: Theory. *Journal of Marine Systems*, 64(1-4):89–109.

Vichi, M., Pinardi, N., Zavatarelli, M., Matteucci, G., Marcaccio, M., Bergamini, M., and Frascari, F. (1998). One-dimensional ecosystem model tests in the Po Prodelta Area (Northern Adriatic Sea). *Env. Mod. Soft.*, 13:471–481.

Walstad, L. J. and McGillicuddy, D. J. (2000). Data Assimilation for Coastal Observing Systems. *Oceanography*, 13(1):47–53.

Wollast, R. (1999). Evaluation and comparison of the global carbon cycle in the coastal zone and in the open ocean. In Brink, K. H. and Robinson, A. R., editors, *The Sea*, volume 10, pages 213–252. John Wiley & Sons.

Zavatarelli, M. and Pinardi, N. (2003). The Adriatic Sea modelling system: a nested approach. *Annales Geophysicae*, 21:345–364.

Zavatarelli, M., Pinardi, N., Kourafalou, V., and Maggiore, A. (2002). Diagnostic and prognostic model studies of the Adriatic Sea circulation: the seasonal variability. *J. Geophys. Res.*, 107:4/1–4/20.

Zavatarelli, M., Raicich, F., Bregant, D., Russo, A., and Artegiani, A. (1998). Climatological biogeochemical characteristics of the Adriatic Sea. *Journal of Marine Systems*, 18:227–263.