UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS

SCHOOL OF OCEAN & EARTH SCIENCE

Marine Ecosystem Model Analysis Using Data Assimilation

by

Ben Andrew Ward

A thesis submitted in partial fulfillment for the
degree of Doctor of Philosophy

October 2009

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS
SCHOOL OF OCEAN & EARTH SCIENCE

Doctor of Philosophy

MARINE ECOSYSTEM MODEL ANALYSIS USING DATA ASSIMILATION

by Ben Andrew Ward

Numerical modelling of the marine ecosystem requires the aggregation of diverse chemical and biological species into broad categories. To avoid large bias errors it is preferable to resolve as many explicit state variables and processes as possible. The cost of this increased complexity is greater uncertainty in model parameters and output. When comparing models, the importance of quantifying both bias error and the variability of unconstrained solutions was revealed as two marine ecosystem models were calibrated to data. Results demonstrated that all prior parameter information must include realistic error estimates if model uncertainty is to be quantified.

Five simple ecosystem models were calibrated to observations from two North Atlantic sites; the Bermuda Atlantic Time-series Study (BATS) and the North Atlantic Bloom Experiment (NABE). Model-data misfits were reduced by between 45 and 50%. The addition of model complexity (a parameterised microbial loop, a variable chlorophyll *a* to nitrogen ratio and dissolved organic nitrogen) led to larger improvements in model performance at BATS relative to NABE. Calibrated parameter values developed at NABE performed better than the default parameter values when applied at BATS. Solutions developed at BATS performed worse than the default values at NABE. The models lacked sufficient ecological complexity to function well at BATS. Errors in the model were masked by errors in the calibrated parameters and the models did not perform well with regard to independent data. The models were well suited to reproducing the NABE data, and the calibrated models performed relatively well at BATS.

The models were sensitive to the underlying physical forcing. Although the ecosystem models were originally calibrated within a poor representation of the physical environment at BATS, results from experiments using an improved physical model support the conclusion that the ecosystem models lacked the required complexity at that site.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Ben Andrew Ward, declare that this thesis entitled, "Marine ecosystem model analysis using data assimilation" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

Date: _____

Graduate School of the National Oceanography Centre, Southampton

This PhD dissertation by Ben A. Ward has been produced under the supervision of the following persons:

Supervisors:

- Thomas R. Anderson, Andreas Oschlies and Michael J.R. Fasham

Chair of Advisory Panel:

- John G. Shepherd

Acknowledgements

It seems appropriate to start by thanking the people who made this work possible when they generously provided me with their code, namely Andreas Oschlies and Marjy Friedrichs. Without their knowledge and experience, I would not have known where to begin. I would also like to thank Bablu Sinha, Andrew Yool and Markus Pahlow for providing me with physical forcing data, as well as Iris Kriest, Debbie Steinberg and Joe Cope who sent me their biogeochemical data.

I have been lucky enough to work with many insightful and experienced people both at the National Oceanography Centre and further afield. Tom Anderson provided valuable supervision and encouragement. Other members of the ecological modelling community who were generous enough to help me out with advice and support include John Hemmings, Adrian Martin, Katya Popova, Bablu Sinha and Andrew Yool. My dodgy statistical ideas have been frequently guided back onto something like the right path with a few wise words from Peter Challenor, Robin Hankin, John Shepherd and Markus Schartau. The typesetting of this thesis was made possible with the help of Sunil Patel at Highfield, who provided the template on which it is based. I would also like to say thank you to Joanna Waniek, who's enthusiasm and encouragement helped me to get started in oceanography.

I would also like to acknowledge the support given to me by Mike Fasham, who's advice always came with a great deal of realism and a good sense of humour.

Over the past three and a half years I have made many good friends in Southampton. Although I have probably alienated most of them with my chronic grumpiness, big shouts nonetheless go out to: Pete and Jude, My Sisters, Cyntoche, Nick Nack, Craig N.I.C.E., Wolfgang Morningstar, Gimli, Ice Man, Hao, Poooly, Liao, Sue-Ann, James R. Watson, Dazza, Bomber, Harry Potter, Matt P, Luke, Doug, Spoff, Pierre, Gerry, Merlin, Lev, Johnny Wilkinson from finance, Moritz, Tom B, Keith Herbert, J.A., Aurelie and Nico, Roz, Cyd, Shorty, Stinch, Ross, Isa, Francesca, Lorna, Beth, Zissou, The Lunch Bunch, Dunc (for the caps), NOCCC, MCC, Trojans, The Students, The Staff, Titanics and to Jane Mullet, for feeding me so well.

I was funded by a NERC MarQUEST Scholarship.

Abbreviations

(μ)**GA**	(micro) **G**enetic **A**lgorithm
BATS	**B**ermuda **A**tlantic **T**ime-series **S**tudy
Chl:N	**Chl**orophyll *a*:**N**itrogen
DCM	**D**eep **C**hlorophyll **M**aximum
DIN	**D**issolved **I**norganic **N**itrogen
DON	**D**issolved **O**rganic **N**itrogen
GCM	**G**eneral **C**irculation **M**odel
HNLC	**H**igh-**N**utrient, **L**ow-**C**hlorophyll
NABE	**N**orth **A**tlantic **B**loom **E**xperiment
NPZ	**N**utrients **P**hytoplankton **Z**ooplankton
NPZD	**N**utrients **P**hytoplankton **Z**ooplankton **D**etritus
MLD	**M**ixed-**L**ayer **D**epth
OCCAM	**O**cean **C**irculation and **C**limate **A**dvanced **M**odel
PAR	**P**hotosynthetically **A**vailable **R**adiation
pdf	**p**robabilty **d**ensity **f**unction
PFT	**P**lankton **F**unctional **T**ype
PON	**P**articulate **O**rganic **N**itrogen
POP	**P**arallel **O**cean **P**rogram
VA	**V**ariational **A**djoint

To Pete and Jude.

Chapter 1

Introduction

1.1 The marine ecosystem and export

The ocean has a large capacity to sequester CO_2 of anthropogenic origin. From 1990 to 1999, anthropogenic emission of CO_2 from burning of fossil fuels was estimated at 6.3 ± 0.4 Gt C yr^{-1}. The corresponding increase in atmospheric CO_2 was estimated at only 3.2 ± 0.1 Gt C yr^{-1} (IPCC, 2001). Much of the shortfall can be attributed to the effect of the ocean and terrestrial biosphere.

Oceanic uptake of CO_2 is driven by the solubility pump, which occurs as a consequence of the increased solubility of carbon in colder water. Because deep water is formed primarily in colder, high latitude regions of the ocean, the deep ocean interior contains more carbon than would otherwise be expected in the absence of these effects . Further to this abiotic pathway, the biological pump also drives the transport of carbon from surface waters to depth. Photosynthesis leads to the uptake of dissolved inorganic carbon (DIC) into dissolved and particulate organic forms, thus lowering the concentration of DIC in the well-lit surface ocean. The sinking of carbon rich particulate organic matter, as well as the preferential consumption near the surface by vertically migrating heterotrophic zooplankton (Honjo et al., 2008), result in a net downward flux of particulate organic carbon. As a result of these transports, and the fact that respiration of organic matter is not tied to the availability of light, the conversion of organic carbon back to DIC occurs on average deeper in the water column than photosynthesis (Oschlies, 2006). This decoupling of uptake and release results in a net downward transport of DIC, which is augmented in certain areas by the physical transport to depth of accumulated dissolved organic carbon (Najjar et al., 2007). Alongside these effects, the depletion of DIC in the surface waters establishes an atmosphere to ocean gradient of pCO_2 that enhances the effect of the solubility pump (Anderson and Totterdell, 2004).

While the physical and chemical processes of the solubility pump are reliably described by the Navier-Stokes and carbonate equations, and modelling it is simply a case of trying to find an approximation of unknown, but exact, solutions (Oschlies, 2006), the complex interactions of the food webs underlying the biological pump have (as yet) no such fundamental theoretical framework. This is because they belong to a extremely heterogeneous and dynamic system, where even the most fundamental (biological) components are subject to change. The marine ecosystem is so complex (i.e. consisting of many different and connected parts) that there must always be some aggregation of its components if the system is to be described at all.

To model the response of the carbon cycle to scenarios of heightened atmospheric CO_2 concentration, it is necessary to make decisions regarding the level of complexity that any such models incorporate (Anderson, 2005; Oschlies, 2006). Is it enough just to model

the simple and well understood solubility pump? or will changes to the system brought about by feedbacks between the environment and ecology affect the ability of the ocean to act as a sink for anthropogenic CO_2?

A number of marine biogeochemical models have been developed in an attempt to understand and predict the way the marine carbon cycle functions. The simplest of these ignore the role of biology altogether, treating carbon as an inorganic tracer (Bacastow and Maier-Reimer, 1990). Other, more sophisticated models have attempted to incorporate the biological pump, either through some very basic parameterisation (e.g. Najjar et al., 1992), or by explicitly including representations of the marine biota (e.g. Wroblewski et al., 1988; Fasham et al., 1990; Anderson and Pondaven, 2003; Aumont et al., 2003). A selection of different modelling approaches are briefly discussed in section 1.2, but it is first necessary to discuss the broad structure of the marine ecosystem with regard to the action of the biological pump.

Dissolved inorganic CO_2 is fixed into particulate organic form by marine autotrophs in the presence of light. Primary production also requires various nutrient elements, most notably nitrogen and phosphorous. The solubility of CO_2 in water means that carbon is not usually considered as a limiting requirement for photosynthesis, and so it is more sensible to track the fate of N and P as the resultant complex organic compounds flow through the various trophic levels of the ecosystem.

Broadly speaking, nutrient compounds may be rapidly recycled within the euphotic zone, or exported from the surface ocean via sinking, advection or turbulent mixing. Exported material is either lost to sedimentation or remineralised and later returned to the euphotic zone by physical processes. It is the export of organic compounds to the deep ocean that drives the biological pump. The relative partitioning of the two pathways is therefore critical. The proportion of nitrate that is recycled within the euphotic zone (i.e. the level of 'regenerated' production) is assumed to be dependent on two key factors, the size distribution of the autotrophic population, and the degree of coupling between the production and consumption of organic matter (Anderson and Totterdell, 2004).

The size spectrum of the plankton community affects the average surface-area to volume ratio such that smaller plankton species are consumed more rapidly (Kriest and Oschlies, 2007) with smaller phytoplankton also able to absorb nutrients at a higher rate in relation to their biomass. Smaller particulates also sink at a slower rate (Smayda, 1970). Under oligotrophic conditions, smaller phytoplankton species tend to have an advantage on account of these factors. The dominance of the smaller phytoplankton types under these conditions leads to reduced export as a consequence of the associated low sinking and high remineralisation rates. Furthermore, small phytoplankton are grazed by small

herbivores which proliferate more rapidly than larger groups. The zooplankton response is thus more closely coupled to primary production and organic compounds are more rapidly recycled to the inorganic form via this grazing pathway. Primary production in oligotrophic regions is thus dominated by high regenerated production and low export. In eutrophic regions, dominated by the larger phytoplankton, sinking rates are high and organic matter is recycled less efficiently, leading to a greater proportion of exported material. This framework suggests that the fate of organic matter is determined by the size structure of the plankton community, although physiological and morphological effects must also be considered. For example, the effects of "ballast" minerals, such as silicate, bicarbonate and lithogenic dust must also be considered on account of the fact that their high density increases the sinking rate of particulate organic material (Billet et al., 1983), which may also be protected from remineralisation by its association with these less reactive minerals (Armstrong et al., 2002). Physiological and morphological adaptations of marine plankton, such as motile flagella or calcium carbonate appendages may conversely serve to counteract or delay sinking (Smayda, 1970; Padisák et al., 2003)

An even more complicated picture emerges on consideration of the "microbial loop" (Azam et al., 1983; Ducklow, 1983). Dissolved organic matter - released from plankton by exudation, excretion, messy feeding and other processes - is consumed by bacteria and subsequently microzooplankton. The microbial loop can significantly increase the proportion of regenerated production, as organic matter is rapidly converted back to inorganic form, along a mediating pathway to the traditional trophic chain (Fasham et al., 1990).

The rate at which biogenic carbon is removed from the surface to depth is determined by the activity and relative efficiency of different pathways in the marine food web. Many factors including the size structure of the plankton community, diel migration of zooplankton and the composition of mineral ballasts associated with organic matter determine the rate of sinking. Association with less biologically and chemically reactive ballast materials, as well as the activity of zooplankton and bacterial populations will affect the rate at which organic carbon is remineralised in the mesopelagic zone (Anderson and Ryabchenko, in press). These factors are influenced by environmental conditions and as such the overall effect of the biological pump may be altered under different climate scenarios. Mechanisms by which these changes may occur have been described in the literature (Anderson and Totterdell, 2004; Le Quéré et al., 2005; Hood et al., 2006) and are many and varied. They will not be discussed in detail here but include the effect of iron fertilisation in high-nutrient, low-chlorophyll (HNLC) regions, the effect of increased stratification due to weakening of the thermohaline circulation and the effect of ocean acidification on the carbonate cycle. If the response of the marine ecosystem and the biological pump are to be correctly modelled, it is important

to capture these processes, either explicitly, or through some parameterisations. This process is addressed in the following section.

1.2 Marine biogeochemical models

As described above, the response of the biological pump to changed environmental forcing is the combined response of the many components of a highly complex system reacting in many different ways. Complexity is so high in fact that it has not yet been possible to apply any fundamental theoretical framework (Oschlies, 2006), and it has been necessary instead to make many simplifying assumptions. This has been achieved through a semi-empirical approach, where models are developed on the basis of theoretical considerations (i.e. our conceptual understanding of how the marine ecosystem functions, Anderson, submitted), but with much complexity simplified into empirically derived parameterisations (Baird, 1999). Many diverse components of the system are aggregated into relatively broad ecological classifications and models are thus reduced to a level of complexity that is both intellectually and computationally tractable.

Simplification along semi-empirical lines is not without it disadvantages, and models that are highly aggregated will have less ability to reproduce the more subtle interactions of the real system. For this reason, progress in marine ecosystem modelling has typically been from the very simple models to the more complex. In the following sections, three broad categories of marine biogeochemical models are examined, ranging from the simplest to some of the most complex.

1.2.1 Nutrient restoring models

The abundance of dissolved inorganic carbon in sea water is generally assumed to mean that increasing atmospheric CO_2 does not limit photosynthesis in any way, and so biological production is primarily controlled by the availability of light and the macronutrients, nitrogen and phosphorous. In a steady state, the physical supply of these nutrients to the well-lit ocean surface is balanced by biological consumption and export (Eppley and Peterson, 1979). If it is also assumed that any indirect effects of increasing atmospheric CO_2 (such as ocean acidification, or increased thermal stratification as a consequence of global warming) have no effect on biological activity, then the action of the biological pump can be very simply parameterised as the removal of nutrients that are supplied to the ocean surface, with carbon export assumed to occur alongside in a fixed stoichiometric ratio.

This theory is the basis of nutrient restoring models, where within a physical modelling framework the abiotic supply of nutrients is balanced by restoration of surface nutrients to observed levels (e.g. Bacastow and Maier-Reimer, 1990; Najjar et al., 1992). Biological production is usually passed to dissolved and particulate forms according to a specified fraction, with sinking and the subsequent remineralisation of particulate organic matter typically parameterised according to some remineralisation profile, such as that of Martin et al. (1987).

Nutrient restoring models represent an extreme example of this approach of simplification in model development. They typically have between two and four parameters, and in their simplicity they are able to focus on a very specific part of the ecosystem dynamics, namely the downward transport of organic matter. They are however unable to answer questions on other parts of the ecosystem structure or to represent interactions between those parts, simply because they are not resolved. Furthermore, because they are based on climatological observations, they are only suitable for diagnostic applications and are unable to model any deviations from the current ocean state.

To allow models to respond to changes in environmental forcing it has been necessary to explicitly resolve ecosystem structure. Models have been developed from the simplest nutrient, phytoplankton and zooplankton (NPZ) (Wroblewski et al., 1988), through to the more complex food web resolving plankton functional type (PFT) models (e.g. Le Quéré et al., 2005). Some aspects of these approaches are examined below.

1.2.2 NPZD type models

Despite their name, NPZD type models are not strictly limited to containing state variables for nutrients, phytoplankton, zooplankton and detritus. The term applies more generally to everything from NPZ through to more advanced formulations with additional state variables for detritus, bacteria and dissolved organic matter (e.g. Fasham et al., 1990). What links these models is that the state variables are very broadly based on trophic level. They typically contain between four and seven state variables, and this simplicity has facilitated their application within ocean general circulation models, where the physical transport of biogeochemical tracers tends to be the dominant computational expense (Oschlies and Garçon, 1999). The NPZD type models usually require between 10 and 30 parameters.

NPZD models require the aggregation of many diverse nutrient and plankton species into broad trophic categories (Le Quéré, 2006), with the fluxes between these groups described by empirically parameterised functions. As it is the relative abundance of different species within these aggregated groups that determines the cycling of nutrients,

and hence the parameter values, if the community composition changes then these values may cease to be applicable (Le Quéré et al., 2005; Friedrichs et al., 2007).

1.2.3 Plankton functional type models

While simple models have been shown to be capable of reflecting the bulk properties of ecosystems on both regional and global scales (Palmer and Totterdell, 2001; Spitz et al., 2001; Schartau and Oschlies, 2003b; Anderson, 2005) many important biogeochemical processes and climate feedbacks can only be resolved by more complex, food-web models (Le Quéré et al., 2005; Hood et al., 2006). Additionally, as composition of the broad ecological categories used to define NPZD models will be subject to change through both time and space, PFT models should be more generally applicable because they resolve relatively more fundamental ecological components (Friedrichs et al., 2007). These are the key motivations behind the development of PFT models, where the trophic categories of the NPZD models are replaced with more specific groups based on biogeochemical function.

The biologically mediated fluxes of elements in these models are resolved between groups of organisms defined by their biogeochemical role (e.g. nitrogen fixers, denitrifiers, silica producers, calcifiers and dimethylsulphide producers; Hood et al., 2006), rather than on trophic or phylogenetic grounds. Using an example of Hood et al. (2006), globally significant contributions to marine calcification are made by both phytoplankton (coccolithophorid species) and zooplankton (foraminifera and pteropods) as well as by the hard corals. By resolving these key processes it is possible for the models to capture important climate feedbacks (Le Quéré et al., 2005).

1.2.4 Model complexity

Simple models of the marine ecosystem are cheap to evaluate and their results can be relatively easily understood, but because they do not resolve certain marine biogeochemical processes, they will not be able to reproduce the full range of observations. More complex models on the other hand are computationally expensive, and include many more complex and non-linear interactions that can make their behaviour difficult to understand (Anderson, 2005), but because they explicitly resolve a more diverse number of processes they will be able to reproduce a wider range of observations and may be more generally applicable (Friedrichs et al., 2007).

Alongside these issues, the benefits of increased complexity need to be weighed against uncertainty in a model's parameters. The number of parameters in marine ecosystem

models may increase exponentially with the number of model components (Denman, 2003), and studies have consistently revealed that the parameters of even the simplest of NPZD models cannot be uniquely constrained by *in situ* observations (Matear, 1995; Schartau and Oschlies, 2003a; Friedrichs et al., 2006). As complexity increases, models will be better able to reproduce observations, resolving more of the individual processes that give rise to them, but the inclusion of extra complexity should also be treated with caution. Observations of the marine ecosystem are sparse, and adding extra model components while not in possession of the data required to constrain them will always lead to increased uncertainty (Anderson, 2005; Raick et al., 2006).

It is not sufficient to assess a model's performance in terms of its ability to fit the data alone, and any increase in this regard should only be considered alongside the increased uncertainty associated with any extra parameters. The addition of more complexity than is required to explain the data is advised against in Ockham's razor, which states that "it is vain to do with more what can be done with less", but given that no marine ecosystem model has yet provided a consistent and comprehensive fit to all available data, the idea framed by Karl Menger that "it is vain to do with less what requires more" seems equally sensible. To find the optimal level of complexity for a marine ecosystem model is to find the point where any additional ability to fit the data is outweighed by the extra uncertainty associated with the additional complexity. Data assimilation techniques allow the formal comparison of models with data, so that both these aspects may be assessed. In the following sections a methodology for evaluating model performance in terms of this balance is presented.

1.3 Observations

The semi-empirical methodology of marine ecosystem modelling means that observations are essential. Parameters may be constrained with data gained from laboratory experiments (e.g. Le Quéré et al., 2005; Pahlow and Oschlies, 2009), although care must be taken with this approach. Laboratory cultures often consist of one genetically and physiologically homogenous species, grown under ideal conditions (Ross and Geider, in preparation), which is in sharp contrast to the situation in the ocean, where plankton populations consist of individuals with high interspecific and intraspecific variability, growing in an ever changing environment.

An alternative source of data is provided by *in situ* observations, which have the advantage of measuring the behaviour of the system of interest, but are typically restricted to long-term time-series sites with little spatial perspective or short-term ship transects

with limited temporal perspective. Regardless of the context, most ship-based measurements also exhibit a strong seasonal bias, with autumn and winter measurements particularly difficult in high-latitude areas.

Observations are often subject to large random and systematic errors, and often do not correspond directly to model parameters, or even state variables. To allow comparison of models and data, they must be converted using often uncertain empirical conversion algorithms. For example, before modelled phytoplankton biomass can be compared to measurements of chlorophyll *a* fluorescence, these must first be converted to chlorophyll *a* biomass, and then subsequently to phytoplankton nitrogen content, using some fixed or variable chlorophyll to nitrogen ratio (e.g. Cloern et al., 1995). Similar concerns exist with regard to measurements of zooplankton (the net collection method only accounts for those individuals larger than 200 μm and biomass is often recorded as wet weight, Madin et al., 2001) and particulate organic nitrogen (settling of particulates within the collection vessel may lead to underestimating biomass by more than 25%, Gundersen et al., 2001). Measurements of biological rates are also subject to high errors, as they are difficult to study without perturbing the system in question (Oschlies, 2006), as typified by the incubation experiments used to measure primary production (e.g. BATS, 2008).

1.4 Data assimilation

The field of marine biogeochemical modelling includes a diverse range of models, but the approach to developing these models has, to a greater or lesser extent, been largely empirical and the one thing common to them all is a number of uncertain and adjustable parameters. The parameters of any one model, together with various auxiliary variables (such as initial conditions, physical forcing, *etc.*), comprise a control parameter vector $\mathbf{p}$ which uniquely determines the model output. This is the basis of "forward" modelling, where a model is run forward from a set of initial conditions in order to generate predictions about some subsequent period.

This approach has been applied with some success in the field of operational ecological modelling, for example in the prediction of harmful algal blooms (Schofield et al., 1999), although accurate assignment of the required parameter values can be difficult. This is especially true when parameters apply across broad functional groups accounting for many different biological or chemical species (Oschlies, 2006). For example, while the nutrient requirements for growth of a single phytoplankton species can be accurately measured in a controlled laboratory experiment, it is much harder (if not impossible) to assign a similar value to an aggregated group of species in the ocean, when the

composition of that group is not only unknown, but also highly variable in time and space.

While model parameters may be difficult, or even impossible to measure directly, ecosystem properties that often correspond to model state variables and fluxes have been the subject of extensive (although sparse) observation programs for a number of decades (Ducklow and Harris, 1993; Steinberg et al., 2001). Given that the forward modelling approach might crudely be described as the process of using poorly-known parameters to estimate well-known ecosystem properties, inverting this approach would seem sensible. This is the basis of variational data assimilation, or "inverse" modelling, where an optimal set of parameters is sought by fitting models to data. In contrast to the forward modelling approach, where the model output is derived on the basis of the model control variables $\mathbf{p}$, the inverse approach seeks the configuration of $\mathbf{p}$ that yields the best model performance. (The mathematical notation is introduced in appendix A.)

The objective optimisation of ecosystem model parameters usually requires the model performance to be quantified by a scalar error statistic, to be minimised by an automated optimisation routine. For the optimisation of n parameters, this process represents a search through an n-dimensional parameter space for the point (or region) corresponding to the best model fit. The coordinates of such a point (or region) thus specify the optimum parameters. A method for quantifying "goodness-of-fit" is given in section **??**, while techniques for locating the best parameter values in terms of this quantity are examined in section 1.5.1. In section 1.5.3 various applications of the inverse approach to biogeochemical modelling are described.

1.5 Quantifying model performance

Model performance can be quantified in a number of ways (Gregg et al., 2008), but most parameter optimisation studies make use of a "least-squares" cost function (e.g. Matear, 1995; Fasham et al., 1990; Spitz et al., 1998; Schartau and Oschlies, 2003a; Evans, 2003). For N observations $d_k, k = 1, 2, \ldots, N$ and model equivalents m_k such cost functions typically take the form

$$J(\mathbf{p}) = \sum_{k=1}^{K} \frac{(d_k - m_k)^2}{\sigma_k^2} \tag{1.1}$$

Where σ_k is an estimate of the (assumed Gaussian) error associated with the observations. The selection of such a metric can be justified with reference to Bayes' theorem,

which lays out the formal mathematical relationship that links the forward and inverse problems that were described in the preceding section.

Bayes' theorem is based on the premise outlined in equation 1.2. Our belief in proposition Y is quantified in the term prob(Y). Furthermore, our belief that proposition X is true, if it is assumed that Y is true, is quantified in the term prob($X|Y$). From this point, our belief that X and Y are both true (prob(X, Y)) has already been explicitly stated, because

$$\text{prob}(X, Y) = \text{prob}(X|Y) \times \text{prob}(Y) \tag{1.2}$$

Within the brackets, a comma denotes "and", while the vertical bar " | " means "given". If the X and Y terms in equation 1.2 are reversed, we obtain the expression

$$\text{prob}(Y, X) = \text{prob}(Y|X) \times \text{prob}(X) \tag{1.3}$$

As prob(X, Y) is clearly identical to prob(Y, X) (the probability of X and Y is the same as the probability of Y and X), the right hand side in equations 1.2 and 1.3 may be equated to each other.

$$\text{prob}(X|Y) \times \text{prob}(Y) = \text{prob}(Y|X) \times \text{prob}(X) \tag{1.4}$$

Dividing both sides by prob(Y), we obtain Bayes' theorem.

$$\text{prob}(X|Y) = \frac{\text{prob}(Y|X) \times \text{prob}(X)}{\text{prob}(Y)} \tag{1.5}$$

The importance of this equation with regard to parameter estimation is shown if we equate the propositions X and Y with *parameters* and *data* respectively.

$$\text{prob}(\textit{parameters}|\textit{data}) \propto \text{prob}(\textit{data}|\textit{parameters}) \times \text{prob}(\textit{parameters}) \tag{1.6}$$

Ignoring the denominator on the right-hand side of equation 1.5 (which is constant for any particular model), this expression relates the quantity of interest, namely the probability of a particular parameter set being true for the given data, to two terms that we have a much better chance of estimating. The second term on the right-hand side of equation 1.6 is the prior belief in a particular parameter set. If we are completely ignorant of the prior values of each parameter, this might simply be reflected in a uniform

prior, which is independent of the value of the parameters and thus assigned a constant value

$$\text{prob}(parameters) = \text{constant} \tag{1.7}$$

Bayes' theorem now simply relates the probability of the parameters given the data to the probability of the data given the parameters

$$\text{prob}(parameters|data) \propto \text{prob}(data|parameters) \tag{1.8}$$

or

$$\text{prob}(\mathbf{p}|\mathbf{d}) \propto \text{prob}(\mathbf{d}|\mathbf{p}) \tag{1.9}$$

The term on the right-hand side is labelled the joint probability, or "likelihood", of the the observations being reproduced by the model and parameters. For any parameter vector $\mathbf{p}$, the likelihood can be evaluated as the product of the individual probability for each observation. If the data are assumed to be independent, then

$$\text{prob}(\mathbf{d}|\mathbf{p}) = \prod_{k=1}^{N} \text{prob}(d_k|\mathbf{p}) \tag{1.10}$$

where d_k represents one of k observations $\mathbf{d}$. If it is also assumed that all observational errors are Gaussian, then the probability of each individual datum, d_k can be approximated as a function of the amount of observational noise and the distance between the modelled value m_k, and the observed data d_k.

$$\text{prob}(d_k|\mathbf{p}) = \frac{1}{\sigma_k\sqrt{2\pi}} \cdot \exp\left[-\frac{(d_k - m_k)^2}{2\sigma_k^2}\right] \tag{1.11}$$

so that

$$\text{prob}(\mathbf{d}|\mathbf{p}) \propto \exp\left(-\frac{1}{2} \cdot \sum_{k=1}^{N} \frac{(d_k - m_k)^2}{\sigma_k^2}\right) \tag{1.12}$$

Substituting into the simplified version of Bayes' theorem given in statement 1.9, and taking the log of both sides, the logarithm L of the posterior pdf is given by

$$L = \log_e \left(\text{prob}(\mathbf{p}|\mathbf{d})\right) = \text{constant} - \frac{1}{2} \cdot \sum_{k=1}^{N} \frac{(d_k - m_k)^2}{\sigma_k^2} \tag{1.13}$$

The second term on the right-hand side is directly proportional to the cost function outlined in equation 1.1, and as this is minimised the posterior probability of the model is maximised, thus defining the least-squares estimate of the optimal solution.

The cost function is usually complicated by the existence of different data types with different units and error statistics. Although the value of σ may be changed to reflect the uncertainties associated with different types of observations, in practice the values are often assigned in a very subjective manner, and the optimal parameters are often highly sensitive to this process. An alternative approach that has been applied in the field of ocean climate modelling is Pareto optimisation (Price et al., 2006). Using this technique misfits from inherently different observations (e.g. chlorophyll *a*, zooplankton, *etc.*) are held as incomparable and thus are not combined into a single metric. Instead, configurations of $\mathbf{p}$ are sought such that model performance with regard to each of the data types cannot be further improved without diminishing the fit in any other categories. The optimisation process seeks a range of such "Pareto optimal" solutions, which will vary in the relative errors from each category of misfits. These solutions, collectively known as a Pareto front, do not represent a definitive answer to the optimisation problem because they will likely include a diverse range of model behaviours. By describing possible trade-offs between the misfit for different observation types however, Pareto solutions might provide valuable information regarding how poor performance in one aspect of the model might be related to and possibly compensated for in the response of other aspects.

1.5.1 Systematic optimisation

In this study model-data misfit is only estimated using cost functions similar to the one outlined in equation 1.1. In such cases the desired solution is the parameter vector $\mathbf{p_{opt}}$ that minimises the value of $J(\mathbf{p})$. The search for $\mathbf{p_{opt}}$ with n free parameters can be thought of as a search through an n-dimensional parameter space for the point with the lowest misfit. In theory this could be done by mapping out the entire parameter space in terms of misfit cost, but for problems with anything more than a few parameters the space rapidly becomes much too large. For example, if each unknown parameter of a model was to be evaluated across a range of 10 discrete points, and the model took just one minute to run, the full evaluation of four unknown parameters would take approximately a week, while six would take close to two years. For a medium complexity

NPZD model with, say, 16 unknown parameters, the evaluation would take longer than the current age of the universe.

This "curse of dimensionality" means that for all but the simplest of models there is clearly a need for a more efficient approach. A wide range of choice is available among parameter optimisation algorithms, but although these are often fascinating for their ingenuity, the field is far too diverse to allow a full review here. The following paragraphs are thus restricted to introducing those techniques that have already been applied for the optimisation of marine biogeochemical models.

Deterministic techniques. The cost function landscape is determined by the response of $J(\mathbf{p})$ to changes in the control parameters $\mathbf{p}$. If the gradient of the cost function can be calculated (or at least estimated) with respect to those control parameters, then $\mathbf{p}$ can be adjusted accordingly to reduce the value of $J(\mathbf{p})$. This is the basis of a number of deterministic gradient descent algorithms, including Powell's conjugate gradient method (Press et al., 1992), which was applied by Fasham and Evans (1995) to optimise 28 parameters of the Fasham et al. (1990) model to observations from the North Atlantic Bloom Experiment (NABE, 47°N, 20°W). The algorithm, which has also been used by Dadou et al. (2004) and Hemmings et al. (2004), estimates the cost-function gradient through a finite-difference method where the model parameters are systematically adjusted around estimates of $\mathbf{p}$. Estimates derived in this way assume that the cost function is smooth on the scale of these perturbations, although this is not a strict requirement for convergence upon the correct solution (Oschlies, 2006).

An alternative, and more reliable way of calculating the cost-function gradient is provided by the adjoint method. Adjoint models may allow the calculation of the exact gradient of J (or any other function) with respect to $\mathbf{p}$. When coupled to a gradient descent algorithm (e.g. Gilbert and Lemaréchal, 1989) the parameters may be rapidly adjusted to minimise the value of J. This iterative search technique, known as the variational adjoint method, has frequently been applied in the field of marine ecosystem modelling (e.g. Lawson et al., 1995, 1996; McGillicuddy et al., 1998; Spitz et al., 1998, 2001; Schartau et al., 2001; Friedrichs, 2002; Faugeras et al., 2003, 2004; Kuroda and Kishi, 2004; Friedrichs et al., 2006, 2007). In contrast to finite-difference methods, adjoint based gradient estimates do not require the cost function to be locally smooth (although convergence will be faster if it is), but are based on an assumed linear functional response to the control parameters. It has been demonstrated in practice however, that this is not strictly necessary for a good solution (Gunson and Malanotte-Rizzoli, 1996; Friedrichs et al., 2007).

One benefit of the variational adjoint technique is that it automatically generates the diagonal elements of the error covariance matrix for $\mathbf{p_{opt}}$. This describes the distance that each parameter can be perturbed from its optimum value before incurring a significant increase in cost, as well as the degree to which pairs of parameters may be varied together, without causing a similar rise in J. These error estimates may also be derived through finite-difference approximations (Matear, 1995; Prunet et al., 1996), but regardless of the method such estimates should be treated with caution as local solutions are in no way guaranteed to apply globally (e.g. Prunet et al., 1996).

Once a purely gradient descent based method finds a configuration of $\mathbf{p}$ where the overall gradient is zero, it will stop. This will be the case when the overall solution (or global minimum) has been found and $\mathbf{p} = \mathbf{p_{opt}}$. However, unless the value of J increases monotonically away from $\mathbf{p_{opt}}$, then there will be other, non-optimal, configurations of $\mathbf{p}$ upon which a gradient descent method may also converge and stop. For example, Schartau et al. (2001) used a variational adjoint method to optimise the parameters of a simple NPZ model in terms of a misfit cost function. They found that the search method converged on a number of local minima which it was not able to escape from, despite the fact that other regions of the parameter space were shown to yield lower misfits. Gradient descent methods will only succeed in locating the global minimum if the search begins near enough to avoid all other local minima. For this reason an ensemble of searches may be initialised from many different points in the parameter space to increase the probability of finding $\mathbf{p_{opt}}$ (e.g. Schartau et al., 2001; Friedrichs, 2002; Hemmings et al., 2004). A larger ensemble will be more likely to locate the global minimum, although in practice the number of runs will be limited by computational resources.

Stochastic techniques. It is possible to efficiently search large parameter spaces even without knowledge of the cost-function gradient, and this can be very useful if following the gradient only ever leads to dead ends. A number of stochastic techniques have been developed where the parameter space is explored in a random way, but with the path of the search guided by probabilistic rules. One such method that has been favoured by ecosystem modellers is the simulated annealing algorithm (Kirkpatrick et al., 1983; Matear, 1995; Hurtt and Armstrong, 1996; Wallhead et al., 2006, 2009), which mimics the annealing process in metallurgy by which the slow cooling of a substance allows the formation of large crystals with a low energy state. The global minimum in this analogy is a perfect uniform crystal lattice which represents the lowest possible energy state for the substance.

Briefly, drawing from a probability distribution around a particular starting point $\mathbf{p}$, the algorithm evaluates point $\mathbf{p}'$. This new vector will always replace $\mathbf{p}$ if it leads to a reduction in cost, however if $J(\mathbf{p}') > J(\mathbf{p})$, then the new vector will only be accepted with a limited probability, determined by the cost difference between the two and a "temperature" parameter T. The probabilistic acceptance of parameter vectors with higher costs allows the simulated annealing algorithm to escape from local minima, but this also prevents convergence of a solution. By steadily reducing the value of T, as well as the width of the probability distribution used to generate $\mathbf{p}'$, the search eventually settles down and as T approaches zero it will converge on a solution. The algorithm is not guaranteed to find the global minimum, but if the cooling process is slow enough, it has been shown to provide good solutions (Matear, 1995).

Genetic algorithms (GAs) are another kind optimisation procedure to draw (loosely) on the emergence of optimality in nature. Unlike all of the algorithms described above, where candidate parameter vectors are evaluated in series, GAs are based on populations of model evaluations, with processes of selection, mutation and genetic crossover applied (according to configuration) so that a good solution may evolve. A modified GA, (the micro genetic algorithm - μGA) was used by Schartau and Oschlies (2003a) to optimise 13 parameters of a simple NPZD model simultaneously at three North Atlantic sites. The μGA was found to significantly improve model fit, but again, convergence on the global minimum could not be guaranteed.

Both simulated annealing and genetic algorithms can be computationally expensive, often requiring many tens of thousands of model evaluations. A GA has been shown in one study to converge faster than a simulated annealing algorithm (Athias et al., 2000), but the performance of both will be subject to configuration (Schartau and Oschlies, 2003a; Wallhead et al., 2009) and will also be problem dependent. The higher computational expense of stochastic search methods needs to be balanced against the ability of these algorithms to escape from local minima and to search through wider regions of the parameter space.

1.5.2 Solutions

Inverse modelling is the process of comparing a model of a system to data, with the ultimate goal of finding a unique set of parameter values that allow a perfect fit to the observations. To achieve this aim requires not only that the model is entirely consistent with the data, but also that there is enough information contained in those data to prevent any uncertainty in the optimum parameters.

The first requirement is unlikely to be satisfied, because even the most complex ecosystem models contain many simplifying assumptions (Anderson, 2005; Hood et al., 2006) and some residual misfits should always be expected. For example, the model studied by Fasham and Evans (1995) is so far one of the most complex marine ecosystem models to be subjected to full parameter optimisation, yet despite having so many degrees of freedom it was unable to provide a good fit to to the data available at NABE. By contrast however, a study published during the same year (Matear, 1995) found that data at Station P (50°N, 145°W) were inadequate to uniquely constrain more than 8 out of 14 parameters in a simple NPZ model. These results, and many since (Hurtt and Armstrong, 1996; Spitz et al., 2001; Schartau and Oschlies, 2003a,b; Friedrichs et al., 2006, 2007), suggest models that are at once too simple to explain the data, and yet too complex to be reliably constrained by them. In the language of inverse theory they are simultaneously overdetermined and underdetermined (see appendix A). When analysing such models it is important to take both these sources of error into account, because the benefits of additional model complexity should always be balanced against any associated rise in uncertainty. In the following paragraphs a number of different solutions to the inverse problem are briefly outlined with respect to quantifying the characteristics of this uncertainty.

Point solutions. Most optimisation procedures will readily produce a single optimal parameter vector $\mathbf{p_{opt}}$ as their solution. This might imply that $\mathbf{p_{opt}}$ is the only solution to the problem in hand, but depending on the number of free parameters and the amount of data available to constrain them there may be many configurations of $\mathbf{p_{opt}}$ that fit the data equally well. The parameter values within $\mathbf{p_{opt}}$ may be sensitive to noise in the observations and even the starting point of the search algorithm. Although they may be stated precisely, the actual values of $\mathbf{p_{opt}}$ will be drawn from a wider probability distribution. In such cases where there is a range of potential solutions from which $\mathbf{p_{opt}}$ is selected by chance, it will be misleading to present $\mathbf{p_{opt}}$ as the single best solution.

Uncertainty estimates. Instead of using a single parameter vector, the uncertainty in a solution can be more accurately quantified by describing the probability distribution from which it is drawn. For a linear model with normally distributed residuals, the sum-of-squares cost function landscape will be closely related to the probability distribution, with $\mathbf{p_{opt}}$ corresponding to the maximum likelihood estimate. The spread of this distribution will also be Gaussian, and in such cases it is not necessary to describe the whole distribution for each parameter, as it can be summarised by $\mathbf{p_{opt}} \pm \sigma$.

Where a parameter in $\mathbf{p_{opt}}$ is drawn from a narrow distribution, it is said to be well constrained by the data. Even small perturbations will cause a significant increase

in cost. Poorly constrained parameters, on the other hand, are drawn from a wider distribution and can be perturbed further without affecting J. This is the basis of the uncertainty estimates described in section 1.5.1, where the curvature of J at $\mathbf{p_{opt}}$ is used to estimate the distance over which increases in J become significant.

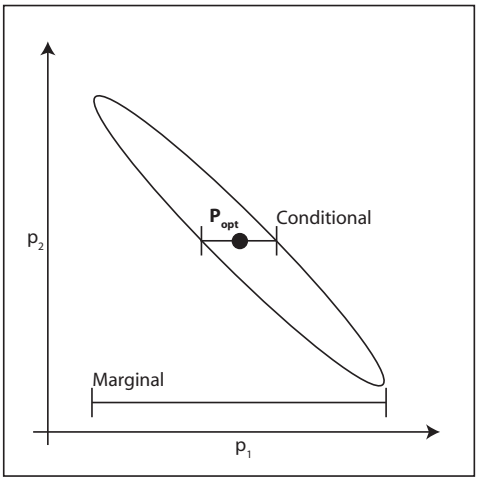

FIGURE 1.1: Conditional and marginal uncertainty estimates for the parameter p_1. The two parameters p_1 and p_2 are correlated with respect to $J(\mathbf{p})$, as illustrated by the elliptical solution region. The conditional solution assumes that p_2 is fixed at $\mathbf{p_{opt}}$. By integrating across p_2 the full range of possible solutions for p_1 (i.e. the marginal solution) can be found (modified from Sivia and Skilling, 2006)

Where uncertainty estimates are derived individually for each parameter, matters can be complicated by the existence of correlations among the parameters. For example, Schartau and Oschlies (2003a) found that while either of the two correlated parameters for sinking and remineralisation of detritus could not be perturbed far from their optimum value, when they were altered together in the correct ratio, they could be adjusted a lot further with no significant increase in cost. Conditional solutions describe the width of the probability distribution for each parameter as all other parameters are held constant. If correlations exist then this may represent an underestimation of the true level of uncertainty. By integrating across the full range of all the other parameters, marginal solutions account for the full level of uncertainty in each parameter. These concepts are illustrated in figure 1.1 and are discussed further in chapter 4.

1.5.3 Applications

As described above, the process of variational data assimilation is the search for parameter values that allow optimal model performance with regard to observations. In some cases these parameter estimates are all that is required from the process, as the

parameters may correspond to some interesting and otherwise unknown real world values. In many situations however, the parameters will be more abstract and will only be of interest when applied in a forward model to estimate something more tangible. Parameter optimisation may also be applied for purposes of model comparison, as models are analysed to see where they perform well and where they perform badly. Some examples of these approaches within the field of marine biogeochemical modelling are given below.

Parameter estimation. Parameter estimation is a fundamental part of all variational data assimilation studies, and in some cases it is the actual parameter estimates that are of scientific interest. This is especially true if parameters are hard to determine by other means. For example, phytoplankton growth parameters are required for the estimation of primary production from satellite data (Platt and Sathyendranath, 1999; Schartau and Oschlies, 2003a). Parameters such as the photosynthetic efficiency and the maximum growth rate cannot be measured directly *in situ*, so must either be estimated using laboratory experiments with cultured phytoplankton species (Le Quéré et al., 2005) or estimated through the variational assimilation of related observations (e.g. Schartau et al., 2001).

This latter approach provides estimates of the model parameters that allow the best fit to data, but it cannot be concluded that estimates derived in this way are always accurate. A number of different studies at the BATS alone have yielded estimates of α that vary over more than an order of magnitude (Fasham, 2000; Fennel et al., 2001; Schartau et al., 2001). This uncertainty can usually be attributed to an imperfect modelling framework. The process of minimising the cost function leads to model output that is most consistent with the data, but this may often be a case of getting the right answer for the wrong reasons. For example, if a model is forced by an inaccurate physical model, or contains unrealistic ecological equations, then the parameter values will seek to compensate for these errors (e.g. Schartau and Oschlies, 2003a). An example of this was demonstrated in a twin experiment conducted by Evans (1999), where a one-dimensional model was calibrated to fit synthetic data from a three-dimensional run of the same ecological model. Simply because horizontal effects such as advection were not accounted for, large errors were included in the calibrated parameters, particularly the photosynthetic efficiency parameter, α.

State and flux estimation. Biogeochemical observations of the ocean are sparse, and purely statistical interpolation can be unreliable for such complex and non-linear systems (Evans, 1999). A more sophisticated approach is to fit theoretical models to observations.

Once the model output has been brought in line with the available observations, gaps in the data can be filled in on the basis of the model theory. This approach goes beyond the purely statistical approach of methods such as krigging, but it should be noted that if the model theory is not correct then any interpolation will be prone to error.

This approach has been applied on both local and basin-wide scales. For example, Spitz et al. (2001) calibrated modified versions of the Fasham et al. (1990) model to observations at the Bermuda Atlantic Time-series Study (BATS) site. By adjusting approximately 50 model parameters, a good fit to the assimilated observations was achieved and subsequent application of the optimal parameters provided estimates of nitrogen fluxes in the system. While such estimates were dependent on the assumptions contained in the model structure, they provided an estimate of ecosystem fluxes that had not otherwise been measured. A similar approach was taken by Schartau and Oschlies (2003b), who sought a single parameter vector to minimise model-data misfit between a simple NPZD model and observations at three North Atlantic sites. By repeatedly calibrating the model parameters to noisy observations, error estimates were assigned for each optimal parameter, and these were used in the forward model to provide estimates of the standing stocks and nitrogen fluxes at each site. The solutions derived from the inverse approach were also applied in a three-dimensional model of the North Atlantic (Oschlies and Schartau, 2005), where the use of optimised parameters led to an improvement in ecological model accuracy not only at the three assimilated sites, but also at an independent site not included in the original optimisation process. The use of calibrated biogeochemical models for state and flux estimation can usually be considered as a diagnostic application, in that the models are applied to reproduce a set of observations, placing them in the context of the underlying model theory. The results present by Oschlies and Schartau (2005) go beyond this, examining the prognostic, or predictive ability of their model with regard to an independent set of data.

Model comparison. A wide range of ecological models have been applied in the study of marine environments, but until fairly recently (e.g. Friedrichs et al., 2006, 2007) no attempt has been made to compare different models in a quantitative and systematic way. One of the most basic requirements of a model is that it bears at least some resemblance to reality, and this quality can be estimated in terms of model-data misfit. The techniques described earlier in this chapter provide a way to objectively evaluate optimal model performance so that competing models can be compared in this regard. Simply comparing models in terms of minimum misfit is not enough however, when the extra degrees of freedom contained in more complex models are considered. For example, even an arbitrary high-order polynomial can in theory be perfectly fit to a set of observations if it is given enough adjustable coefficients, but such a model

would not have very much predictive skill if it were then applied in an attempt to reproduce independent data. This is because the polynomial would be largely fit to random fluctuations in the original observations, rather than to any consistent trend common to the two data sets (Friedrichs et al., 2006). Calibrated misfit by itself is not a sufficient criterion by which to compare models. It is more important to assess their ability to reproduce independent data that were not used in the assimilation process. If a model is consistently able to reproduce independent data it demonstrates that it is capable of explaining the underlying dynamics that give rise to the observed conditions, rather than just the observations themselves.

If models are best assessed against unassimilated data, then perhaps it would be better to evaluate uncalibrated models, assessing their predictive skill as they are applied with a set of *a priori* parameter values. Such an approach can be hard to achieve in practice however, because of the difficulties associated with assigning parameter values. For some parameter such as phytoplankton growth rates, for example, abundant data are available from laboratory based experiments, but as these are usually based on homogenous cultures of single species grown under uniformly optimal conditions they may not be entirely reliable when applied to modelling the natural environment, where many different species coexist with high genetic and phenotypic variability within a heterogeneous environment. Given these difficulties, and the fact that models are invariably subject to at least some degree of subjective parameter tuning in order to obtain coherent results, the application of parameter optimisation techniques can be applied to place these models on a level playing field, so that it the optimal model performance that is compared, rather than the degree (or lack) of tuning that has been applied.

The predictive skill of 12 ecosystem models was examined in a cross-validation experiment by Friedrichs et al. (2007), where models tuned at one site were subsequently evaluated against independent data at another. In an earlier paper (Friedrichs et al., 2006) the authors had noted that model predictive skill was strongly related to the number of parameters that were optimised. If too many parameters were tuned in an underdetermined model, the model could be made to fit the data well, but the parameters were unrealistic and the model performed very badly with regard to independent data. To overcome this problem, the authors chose to optimise only a limited subset of well-constrained parameters. Using this technique the optimisation process yielded well-defined and realistic parameter estimates that gave much better estimates of predictive skill in the cross-validation experiment. The results revealed that those models which included multiple plankton size-classes performed better than those which did not. The results suggest that size-classes are critical to increasing the predictive skill of ecosystem models. Aside from any ecological imperatives for increasing model complexity in this

way, it can be argued that adding (realistic) model complexity to models allows greater predictive skill on the grounds that a wider range of dynamic processes are included.

1.6 Objectives

In this thesis the aim is to use parameter optimisation techniques to examine a range of ecosystem models at several different time-series sites. A number of models of varying ecological complexity will be examined so that related issues of model bias and uncertainty may be investigated in this regard. In chapter 2, the modelling and optimisation framework of Friedrichs et al. (2007) will be used to compare a deterministic variational adjoint optimisation technique with a stochastic genetic algorithm. The two techniques will be examined not just in terms of their efficacy in returning a low value for the model-data misfit, but also in terms of their ability to quantify uncertainties in the optimal parameter values.

The results presented in chapter 2 will highlight the importance of quantifying the uncertainty in all model parameters, and in subsequent chapters a method of doing this is developed and applied. A new modelling and optimisation framework is introduced in chapter 3. Five simple NPZD based models are calibrated to time-series data from the North Atlantic using a micro genetic algorithm. In chapter 4 solutions from the optimisation process are analysed and a method for developing robust solutions that go some way towards quantifying parameter uncertainty is developed.

In chapter 5 the methods developed in chapters 3 and 4 are applied as the five ecosystem models are individually calibrated to biogeochemical observations from two time-series sites in the North Atlantic. As well as examining each model with regard to where it performs well, and where it does not, the models are also analysed with regard to where the optimal solutions cannot be constrained with a great deal of confidence. While the former process shows where additional model complexity is required in order to accurately reproduce observations, the latter highlights areas where model complexity is already more than can be constrained by the assimilated data. Following this analysis of the calibrated solutions, the optimised models are applied in chapter 6 in an effort to reproduce independent data. The calibrated models are essentially swapped between sites in a cross-validation experiment, in an attempt to gain estimates of each model's predictive skill. Examining the calibrated parameter sets as they are used to reproduce independent data provides a sterner test of the ecosystem models than simply calibrating to data. As highlighted by Friedrichs et al. (2006), a high-order polynomial with no underlying ecological theory could do a reasonable job of reproducing a given dataset, but would have little predictive skill. By applying the calibrated solutions with regard to

independent data it should be possible to highlight which models (if any) fit the data well at both sites because they account for dynamic processes that are consistent between sites, and conversely which models only get close to the data at either site because of the rigorous parameter calibration process that was applied.

The analyses presented in chapters 5 and 6 highlight the strengths and weaknesses of the five ecosystem models. These models are all however nested within an imperfect representation of the physical environment, so their performance will also be subject to unknown errors from this source. In chapter 7 the ecological models are calibrated within alternative representations of the physical environment. The results reveal that although a realistic representation of the physical environment is required for an accurate representation of the marine ecosystem, the conclusions drawn in previous chapters were not sensitive to the physical forcing scheme that was applied.

Finally, in chapter 8 the results and conclusions presented in previous chapters are summarised and their implications are discussed.

Chapter 2

Parameter optimisation and underdetermination

Preamble

This chapter is adapted from an article accepted for publication in a special issue of the Journal of Marine Systems.

Ward, B. A., Friedrichs, M. A. M., Anderson, T. R., Oschlies A., Parameter optimisation techniques and the problem of underdetermination in marine biogeochemical models. Journal of Marine Systems. *Resubmitted after corrections, 6^{th} April 2009.*

All of the work was carried out by Ben Ward, with the exception of the sensitivity analysis using the variational adjoint, which was carried out by Marjorie Friedrichs. The modelling framework and data were developed and compiled by Marjorie Friedrichs. The work was written up entirely by Ben Ward, although with significant intellectual input from Marjorie Friedrichs, Thomas Anderson and Andreas Oschlies. Comments from John Hemmings and two anonymous reviewers also helped to improve the text.

2.1 Introduction

Marine biogeochemical models typically rely on the aggregation of many diverse species into broad functional groups (Hood et al., 2006), with the fluxes of matter between model compartments described by empirical functions. As the output of a model will be determined by the values assigned in the parameterisation of these functions, performance can be significantly improved by the adjustment of these parameters. As a consequence of the many degrees of freedom present in even relatively simple models, this process is complicated, and the results may be highly variable, especially if only subjective measures of model performance (i.e. visual comparison with data) are used.

Data assimilation techniques allow model parameters to be objectively assigned through the minimisation of model-data misfit. By objectively and systematically calibrating model parameters, models can be compared in terms of their structure alone, rather than the degree (or lack) of tuning that was applied in their development. Although this is a useful approach, it was noted in the introduction that observations are currently insufficient to precisely constrain the number of parameters required by even the simplest models (Matear, 1995; Fennel et al., 2001; Friedrichs et al., 2007). If such models are to be compared using data assimilation, it is necessary to find a way of dealing with the many unconstrained parameters, either quantifying or reducing the degree of underdetermination in each case.

The latter can be achieved by including in the cost function prior information about the parameters, for instance by including extra terms that attach a high penalty to values that are thought to be unrealistic. For example, estimates of the probable distribution of each parameter can be included as a penalty term, so that any values that deviate too far from prior beliefs are assigned a large misfit cost (Fasham and Evans, 1995; Matear, 1995; Schartau et al., 2001; Evans, 2003). More simply, bounds can be defined that prevent each parameter from taking values outside of a certain range (Schartau and Oschlies, 2003a). A similar but related approach requires the optimisation of only a few well-constrained parameters, fixing the remainder to precise default values (i.e. restricting their allowed range to a single value). In a comparison of three ecosystem models (with 10-19 parameters), Friedrichs et al. (2006) used a variational adjoint method to demonstrate that the predictive ability of those models was a strong function of the number of parameters optimised. If too many unconstrained parameters were allowed to vary, the more complex models were unable to reproduce unassimilated data.

In this chapter, two optimisation routines, a variational adjoint (VA) method and a micro genetic algorithm (μGA) are compared. In the first instance, the ability of the two techniques to yield a good fit between the models and assimilated data is assessed.

The optimisation routine that returns the lowest misfit should ideally be preferred, but given most marine biogeochemical models are to some degree underdetermined, it is also necessary to examine the consequences of this uncertainty in the optimised parameter values. If the parameters are highly overfit to the assimilated data, it is likely that they will perform badly with regard to unassimilated data (Friedrichs et al., 2006), so in addition to looking at the minimum misfit costs achieved by the two techniques, the solutions are examined with regard to their ability to reproduce independent data.

The two optimisation techniques are compared in terms of their ability to calibrate a simple, single size class model and a slightly more complex, multiple size class model to Arabian Sea data, with solutions evaluated against independent, equatorial Pacific data, as an assessment of predictive skill. The techniques are applied to each model with first ten, and then three free parameters, as the following questions are addressed: What is the minimum misfit found by each technique? Do the two techniques yield the same consistent solutions? How useful are the solutions when they are used to model independent data? And how much information do different approaches yield with regard to assessing model skill?

2.2 Methods

The modelling and optimisation framework applied in this chapter was taken directly from Friedrichs et al. (2007), with the addition of the μGA optimisation technique. These aspects are summarised in sections 2.2.1 to 2.2.4. The numerical experiments performed here are described in section 2.2.5.

2.2.1 Biogeochemical models and physical forcing

Two models of different complexity were selected from the suite of biogeochemical models analysed by Friedrichs et al. (2007). The first model was chosen as the simplest of those with no size-class discrimination of plankton types; the second was the simplest of the models that did contain an explicit representation of different plankton size-classes. Both of these models were also examined in Friedrichs et al. (2006). A schematic diagram indicating the basic structure of both models is given in Figure 2.1. Table 2.1 lists both sets of model parameters and their default prior values.

The four-component (dissolved inorganic nitrogen, phytoplankton, zooplankton and detritus) ecosystem model, requiring ten ecological parameters, was developed specifically for the Arabian Sea. Regular entrainment of nutrients brought about by seasonal monsoon events make the use of a diatom-mesozooplankton based system appropriate at this

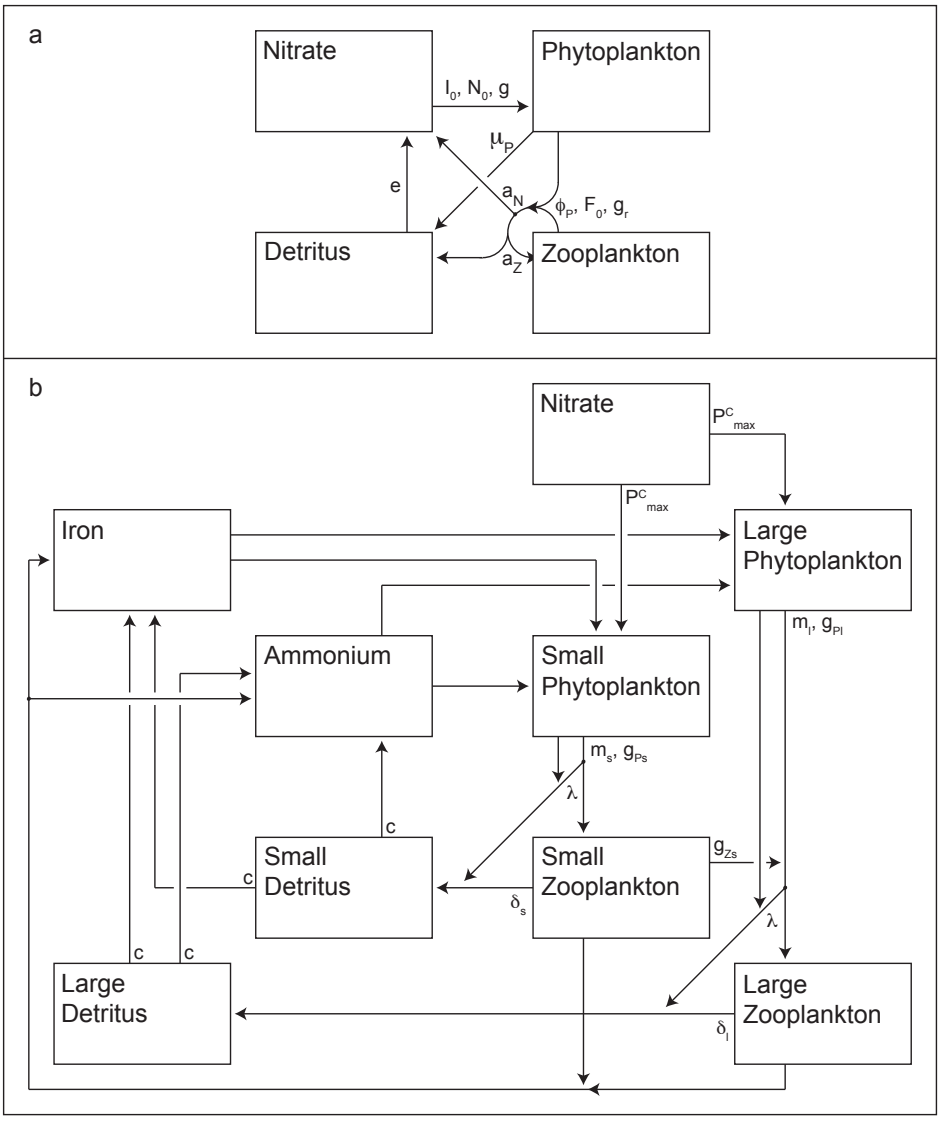

FIGURE 2.1: The four-component (a) and nine-component (b) biogeochemical models. State variables are shown as boxes, fluxes as arrows. The optimised parameters are shown next to the fluxes they describe.

site, with resolution of the microbial loop less important than at more oligotrophic sites. A full model description can be found in McCreary et al. (1996). The nine-component model, containing two size-classes of phytoplankton, zooplankton, and detritus, together with ammonium, nitrate, and iron was developed to simulate the high-nutrient-low-chlorophyll conditions observed in the equatorial Pacific (Christian et al., 2002) and requires 23 ecological parameters. The model structure incorporates both iron limitation of larger phytoplankton and the nanozooplankton-microzooplankton pathway that play an important role at that site. The implementation used here is identical to that described by Christian et al. (2002), except that the *a priori* grazing rate parameter for large phytoplankton was increased to 50 d^{-1} (Friedrichs et al., 2007).

TABLE 2.1: Tuned model parameters with prior values and minimum and maximum limits on the μGA search. The bold text indicates parameters that were varied in both the 10 and 3 parameter optimisations.

4-component model (McCreary et al., 1996)				
Parameter	Symbol	Default	Range	Units
$\frac{1}{2}$ saturation for Z grazing	F_0	1	0 - 5	mmol N m^{-3}
Light saturation constant	$\mathbf{I_0}$	40	5 - 80	Wm^{-2}
$\frac{1}{2}$ saturation for N uptake	N_0	1	0 - 10	mmol N m^{-3}
Grazing preference coefficient	$\phi_\mathbf{p}$	0.83	0 - 1	-
P growth rate parameter	$\mathbf{g}$	2.9e-5	1e-6 - 1e-4	d^{-1}
Z grazing rate parameter	g_r	4.6e-5	1e-5 - 1e-4	d^{-1}
P mortality	μ_p	1.2e-6	1e-8 - 1e-5	d^{-1}
Z assimilation coefficient	a_z	0.1	0 - 1	-
Z messy feeding to N	a_n	0.4	0 - 1	-
Detrital remineralisation	e	0.125	0.025 - 25	d^{-1}
9-component model (Christian et al., 2002)				
Parameter	Symbol	Default	Range	Units
Grazing rate parameter (Z_s on P_s)	g_{Ps}	50	5 - 60	d^{-1}
Grazing rate parameter (Z_l on P_l)	g_{Pl}	50	5 - 60	d^{-1}
Grazing rate parameter (Z_l on Z_s)	g_{Zs}	10	5 - 60	d^{-1}
Assimilation efficiency	λ	0.75	0.1 - 0.9	-
Z_s mortality rate	δ_s	0.05	0 - 0.2	d^{-1}
Z_l mortality rate	δ_l	0.2	0 - 2.0	d^{-1}
P_s mortality rate	m_s	0.05	0.05 - 1.2	d^{-1}
P_l mortality rate	m_l	0.2	0.05 - 1.2	d^{-1}
Detrital remineralisation rate	$\mathbf{c}$	0.35	0.1 - 0.5	d^{-1}
Max C specific growth rate	$\mathbf{P^C_{max}}$	1.0	0.1 - 2.0	d^{-1}

The ecosystem models are run in a one-dimensional framework and are forced by time-series profiles of temperature, vertical diffusivity and vertical velocity, together with mixed layer depth and surface values of photosynthetically available radiation. The physical framework is identical to that used in Friedrichs et al. (2007), and a full description can be found there.

2.2.2 Biogeochemical data

In situ observations of dissolved inorganic nitrogen, chlorophyll *a*, ^{14}C primary production and detrital flux from sediment traps at approximately 800∼880 m were available from the US Joint Global Ocean Flux Study (JGOFS) equatorial Pacific Process Study (February and November 1992) (Murray et al., 1995) and Arabian Sea Process Study (January to December 1995) (Smith et al., 1998). Data were restricted to those collected

at station S7 in the Arabian Sea (16.0°N, 62.0°W) and those collected within one degree of the equator during the equatorial Pacific (140°W) cruises. All observations were interpolated vertically onto the model grid, resulting in six dissolved inorganic nitrogen and five chlorophyll *a* and primary productivity profiles in the Arabian Sea, and 40 dissolved inorganic nitrogen and 27 chlorophyll *a* and primary productivity profiles for the equatorial Pacific. Sediment trap data for particulate nitrogen flux were available from the 800 m trap at 16.0°N, 61.5°W within the Arabian Sea, and from the 880 m trap in the equatorial Pacific (Honjo et al., 1995, 1999). Model detrital flux was extrapolated from the deepest model layer (150 m) to the level of the observations using the flux attenuation formula of Martin et al. (1987).

2.2.3 The cost function

The cost function J quantifies the misfit between observed values ($\hat{a}$) and modelled equivalents (a). It was evaluated at each site using a weighted sum of squares function,

$$J = \frac{1}{M} \sum_{m=1}^{M} W_m^2 \frac{1}{N_m} \sum_{j=1}^{N^m} (\hat{a} - a)_{nm}^2 \tag{2.1}$$

Individual misfits were summed over the number of different data types ($M = 4$; nitrate, chlorophyll a, primary productivity and export) and the number of observations (N_m). At each site, misfits to each category of data were weighted proportionally to the standard deviation of the observations, σ_m (Table 2.2), such that,

$$W_m = \frac{C_m}{\sigma_m} \tag{2.2}$$

The weighting factor C_m (set to 7 for productivity data and 3.5 for all other types) was included to bring the confidence assigned to each data type in line with subjective estimates of measurement uncertainty (Friedrichs et al., 2006). The higher value assigned to primary production data increases the associated weight of these observations, which would otherwise be very low as a consequence of their high variability. The number of observations for each data type at each site, N_m, was included in equation 2.1 so that more frequently observed data types would not dominate J. Models are not significantly different if the values of J differ by less than one.

TABLE 2.2: Standard deviations for Arabian Sea and equatorial Pacific data.

Data type	Arabian Sea	Equatorial Pacific	Units
Nitrate	2.48	1.87	mmol N m^{-3}
Chlorophyll *a*	0.22	0.09	mg chl m^{-3}
Primary Production	19.2	8.51	mmol C m^{-3} d^{-1}
Export	1.44	0.47	mmol C m^{-3} d^{-1}

2.2.4 Optimisation techniques

The VA technique applied by Friedrichs et al. (2007) is contrasted with another method, the μGA, which was taken from Carroll (1996) and was used by Schartau and Oschlies (2003a). While the VA is a deterministic, gradient descent technique that may be applied in an unrestricted search, the μGA is a stochastic technique that makes no use of the gradient of the cost function and is limited to searching only a predefined range of parameter values.

The variational adjoint (VA) technique

Adjoint models allow the calculation of the gradient of a model output function with respect to a (potentially large) set of model parameters. The VA technique seeks to minimise model-data misfit by efficiently adjusting model parameters based on the gradient information provided by the adjoint model. Such techniques have the additional benefit of allowing easy computation of the inverse of the Hessian matrix of second order partial derivatives of the cost function. This provides an estimate of the uncertainty and correlations among the optimal parameters.

The variational adjoint is an iterative process. The numerical model is run forward with an initial guess for the parameter values, and the cost function is evaluated. The adjoint model code - here automatically compiled using the Tangent linear and Adjoint Model Compiler (TAMC) (Giering and Kaminski, 1998) - is then run backwards in time, yielding the gradient of the cost function with respect to the model control parameters. The gradient information is passed to a limited memory quasi-Newton optimisation procedure (Gilbert and Lemaréchal, 1989), which calculates the optimal direction and step size as the parameters are adjusted towards the minimum of the cost function. The new parameter values are evaluated in the forward model, and the steps are repeated until a certain convergence criterion, based on the norm of the gradient of the cost function, is satisfied.

The calculation of the cost-function gradient by the adjoint is based on the assumption of a linear response to changes in the model parameters. While this is likely not the case

for biogeochemical models such as those applied here (Schartau and Oschlies, 2003a), the TAMC can nonetheless produce sensible and functional code for both the models presented here, as well as for number of others (Friedrichs et al., 2007).

The micro genetic algorithm (μGA)

The μGA is a stochastic optimisation technique analogous to evolution by natural selection. The algorithm begins with a set of randomly generated model parameter vectors (see figure 2.2). Each parameter vector is evaluated in the forward model and is assigned a misfit value as the model output is evaluated against observations. At the end of each generation the parameter vectors are duplicated and each set is randomly combined into pairs. Within each pair, individuals are compared in terms of model-data misfit, and the less fit parameter vector is discarded. The population size is maintained because the population was duplicated in the previous step. Each of the selected individuals is then encoded as a single string of binary digits and is assigned to a pair with another individual. Before each parameter vector is reproduced in the next generation of the algorithm, a process analogous to genetic crossover is applied. A single point along each pair of binary strings is selected at random, and all the digits occurring after this point are swapped between the individuals.

By selecting the fittest individuals, information describing the best parameters is passed into the next generation, whilst the crossover ensures that new points in the parameter space are evaluated. The μGA cycles through a predefined number of generations, and at the end of each generation, the fittest individual is passed directly to the next generation. This prevents the best solutions from being lost as the random processes of the algorithm are applied. Because the μGA does not use mutation (a feature of the traditional GA), members of the population tends to converge on the best parameter vector. In order to maintain the search across the entire prior parameter space, once the binary code describing the individuals contains less than 5% variability across the population, it is regenerated at random, with the best individual again conserved.

The μGA requires a number of its own parameters, such as the number of individuals within each generation. These parameters can be adjusted to improve the rate of convergence on a solution, but to avoid the possibly lengthy process of trial and error required to find the optimal configuration, the default values (Carroll, 1996) were applied. In all cases, the μGA population size was equal to the number of free parameters and all optimisations were run for 5,000 generations.

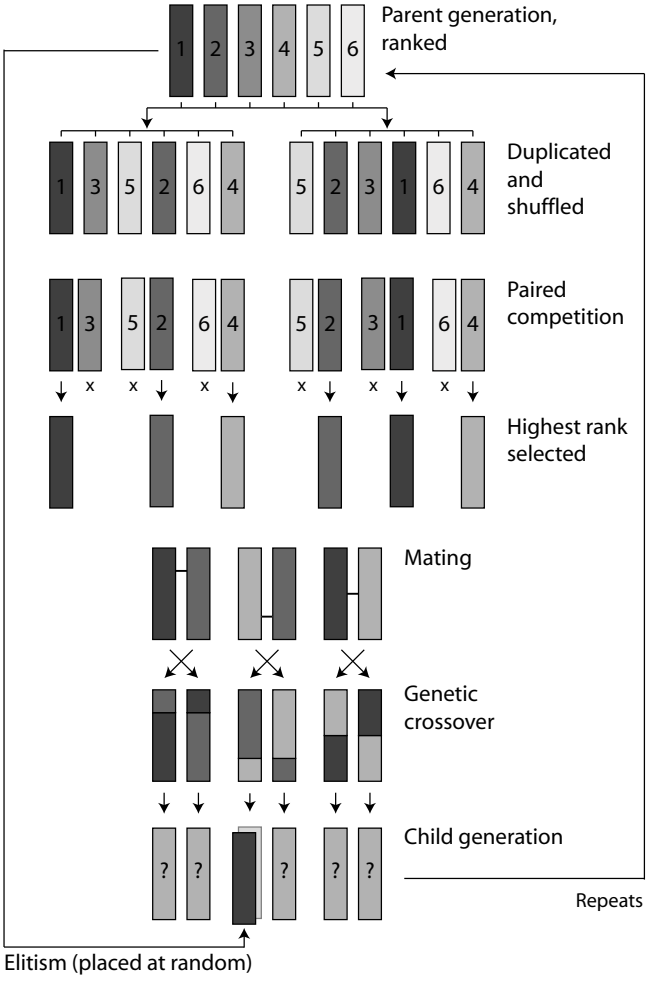

FIGURE 2.2: Schematic diagram of the micro genetic algorithm. Each rectangle corresponds to a single parameter vector to be evaluated in the forward model. The paired selection process means that the "fitter" parameter vectors are more likely to be selected for reproduction in the next generation.

Restricting the parameter space

Just as the values of three coordinates, x, y and z can be used to define any point in three-dimensional space, the values of n model parameters can be used to define a point in an n-dimensional parameter space. This analogy is useful for visualising the optimisation problem, which can be described as the search for the point (or region) of the parameter space associated with the minimum cost. The VA technique is free to evaluate an almost continuous range of parameter values across an infinite, or unbounded, parameter space (although if necessary, limits to the search may be applied, e.g. Schartau et al., 2001). The μGA by contrast, as a consequence of its stochastic approach to handling parameter

values, is restricted to searching a number of discrete points within a finite region of the model's parameter space.

The VA technique was applied here to search an unbounded parameter space because, in agreement with Friedrichs et al. (2006, 2007), preliminary experiments revealed that the application of a penalty term caused poorly constrained parameter values to become trapped by the sharp changes in gradient at the edges of the parameter space. The μGA by contrast does not become trapped as it makes no use of the cost function gradient. In fact it requires upper and lower limits to be set for each parameter because of the way a binary string is used to specify the discrete values of the parameters. Here, and in subsequent chapters, 6-bit strings were used, and so each parameter could only be assigned one of 64 discrete values in a predefined range. The upper and lower limits were set to exclude any unrealistic values for the parameters. For some parameters, such as the half-saturation constant for nutrient uptake (N_0) in the four-compartment model, the minimum values were set to zero, as any negative values would be nonsensical. Although some of the remaining minima and maxima were defined somewhat subjectively, they were set to conservatively broad values so that no realistic values were excluded from the solution. The limits of the μGA search space for both models are listed in Table 2.1.

The study by Friedrichs et al. (2007) found that the data available for assimilation were adequate to constrain only between two and four parameters for each model. It is likely that the application of an unbounded search with ten unconstrained parameters will lead to highly uncertain and perhaps unrealistic estimates of the optimal parameters. This can be prevented by placing limits on the parameter search space (e.g. Schartau et al., 2001) or by fixing the least constrained parameters to some sensible values and only optimising those parameters which are well-constrained by the data (Friedrichs et al., 2006, 2007).

A reduced subset of well-constrained parameters was identified for each model using the technique of Friedrichs et al. (2007). Each model was at first optimised for the full set of parameters, with parameter uncertainty identified using the inverse of the Hessian matrix. In each case the most unconstrained parameter was fixed to its prior default value and the optimisation process was repeated, sequentially fixing out all those parameters with uncertainty greater than 100% of their prior value. For each model, three well-constrained parameters were identified in this way, and these are listed in bold type in Table 2.1.

The two alternative approaches to incorporating prior information are applied here, optimising each model to Arabian Sea data with the unbounded VA and the bounded μGA, for at first ten, and then three, free parameters.

2.2.5 Numerical experiments

Each model was first optimised for ten free parameters. This was the total number of biological parameters in the four component model, and to maintain consistent degrees of freedom between models only ten out of the 23 biological parameters in the nine-component model were selected for optimisation. These ten parameters were identified in the sensitivity analysis of Friedrichs et al. (2007, unpublished result) as being the ten most well constrained by the Arabian Sea and equatorial Pacific data. The selected parameters relate to a broad range of the model pathways, from phytoplankton growth rates to detrital remineralisation and are listed in Table 2.1. The ten-parameter optimisations will include several poorly-constrained and partially-correlated parameters (Friedrichs et al., 2007), thus allowing some comparison of how well underdetermination is handled by the different optimisation approaches.

Both the VA and μGA techniques can be sensitive to the parameter values used to initialise the search (Friedrichs, 2002; Schartau and Oschlies, 2003a). Thus each model was optimised ten separate times for both the μGA and the VA, each time starting with a different set of randomly generated parameter vectors, all within the parameter ranges defined in Table 2.1. Each model was optimised to Arabian Sea data, with every solution subsequently applied in an attempt to reproduce unassimilated data from the equatorial Pacific. The misfit costs associated with the independent data were labelled the "predictive costs" of the solutions (Friedrichs et al., 2007), and they can be used to make a quantitative assessment of the predictive skill of optimised models (Gregg et al., 2008).

2.3 Results

The optimised model-data misfits for the assimilated Arabian Sea data, together with the associated predictive costs at equatorial Pacific are shown in Figure 2.3. As each optimisation technique was repeated ten times from different points in the parameter space, results are presented in groups of ten. The minimum misfit, J_{min}, achieved for each set of optimisations is taken as the best estimate of the global minimum yielded by a particular technique. Accounting for estimates of the observational error (Eqn. **??**), a misfit difference of less than one is not significant, and therefore for each set of optimisations, all calibrated solutions with a cost of $J < (J_{min} + 1)$ are equally valid. These solutions are subsequently referred to as "acceptable solutions" (Table 2.3), though as discussed in section 2.2.4, this does not necessarily imply that all the optimised parameter values are realistic. In Figure 2.3, the acceptable solutions are shown as filled

black bars. The unfilled bars represent the optimisations that did not converge on the (estimated) global minimum.

TABLE 2.3: Minimum optimised cost function values and associated predictive costs for each model/technique. The numbers in brackets gives the number of optimisations within each group that yielded a cost not significantly different from the minimum. The standard deviations of the mean predictive costs are also shown.

4-component model		Arabian Sea	Equatorial Pacific
Prior cost		35.7	82.9
		J_{min}	Mean predictive cost
10 parameters	VA	6.6 (6)	100.7 ± 11.5
	μGA	7.1 (10)	131.0 ± 44.1
3 parameters	VA	12.6 (10)	181.8 ± 0.0
	μGA	12.6 (10)	184.2 ± 23.0
9-component model		Arabian Sea	Equatorial Pacific
Prior cost		44.7	15.4
		J_{min}	Mean predictive cost
10 parameters	VA	6.8 (1)	48.0
	μGA	7.6 (10)	61.1 ± 28.6
3 parameters	VA	11.1 (10)	32.1 ± 0.2
	μGA	11.1 (10)	32.1 ± 3.2

The acceptable solutions from any one group of optimisations, while by definition very similar in terms of cost, often contained very different estimates of the optimal parameters. This variation is shown in Figure 2.4, where the acceptable solutions from the VA are shown in blue, while those from the μGA are shown in red. When applied to modelling the equatorial Pacific data, these solutions often resulted in highly variable predictive costs (e.g. Figure 2.3, panel a-ii). To allow a single metric estimating the predictive skill given by the acceptable solutions, the associated predictive costs were condensed into a single mean value (Table 2.3).

2.3.1 Four-component model

When comparing optimisations with the same number of free parameters using the four-component model, there were no significant differences in the minimum tuned costs (J_{min}) yielded by the two techniques. All of the acceptable solutions led to an improvement in terms of cost at Arabian Sea, when compared to the prior parameter values. The same solutions, however, performed worse than the priors when applied to the independent equatorial Pacific data (the misfit costs associated with the prior parameters are given in Table 2.3). Although these results held for both the ten-parameter and the

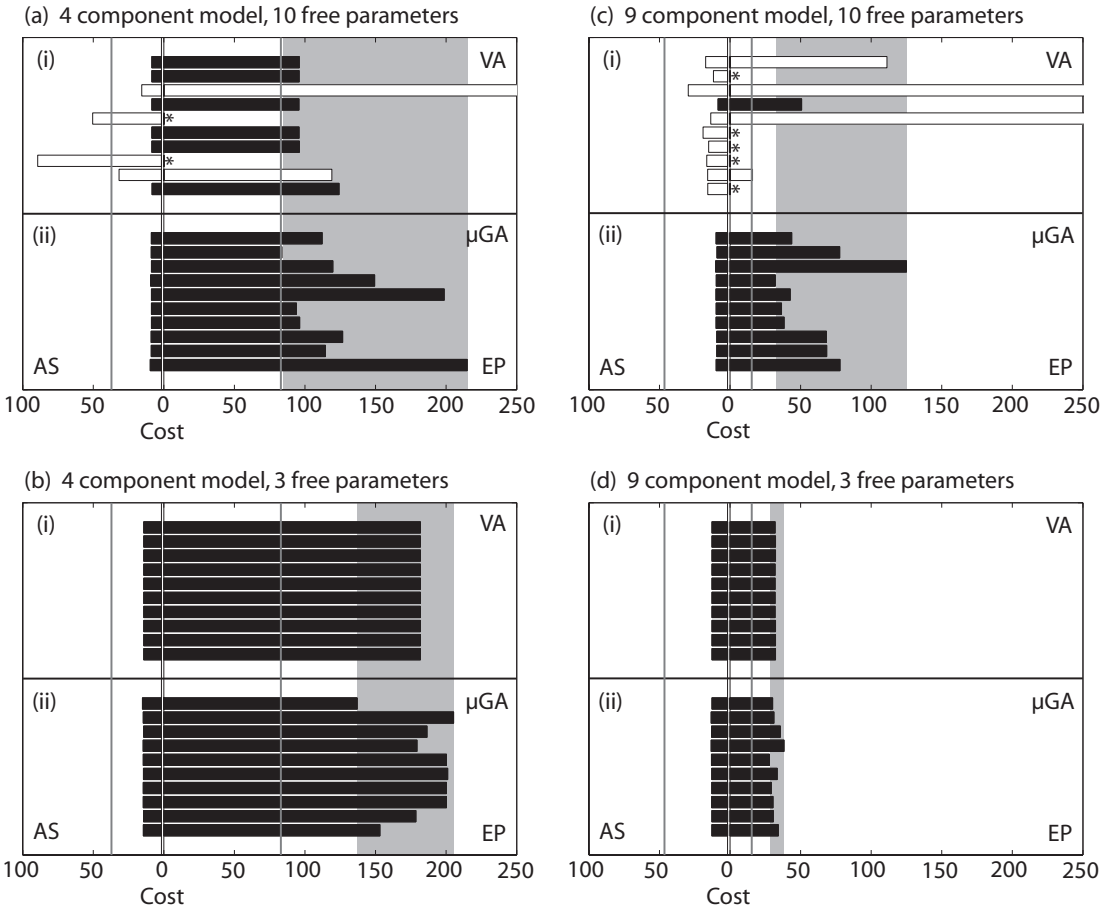

FIGURE 2.3: Cost function values for models calibrated to Arabian Sea data. The calibrated costs at Arabian Sea are shown on the left-hand side of the vertical axes, costs when the solutions were applied to equatorial Pacific are shown on the right. Unfilled bars represent optimisations that became trapped in local minima, filled bars denote where the results were statistically as good as the estimated global minimum from that technique. Asterisks denote solutions that led to the model crashing at equatorial Pacific. The grey boxes highlight the range of the predictive costs that are associated with acceptable solutions, and the dark grey lines show the costs at each site associated with the default prior parameters.

three-parameter optimisations, the minimum tuned costs when optimising ten parameters ($J_{min} = 6.6$, VA; $J_{min} = 7.1$, μGA) were between 44 and 48% lower than the cost when only three parameters were optimised ($J_{min} = 12.6$, both techniques).

Ten free parameters

When the four-component model was optimised with ten free parameters, the VA technique became trapped in local minima on four occasions. The costs for the six acceptable VA parameter solutions ($J_{min} = 6.6$) were not significantly different from those obtained with the μGA ($J_{min} = 7.1$). As expected, the VA solutions included unrealistic values

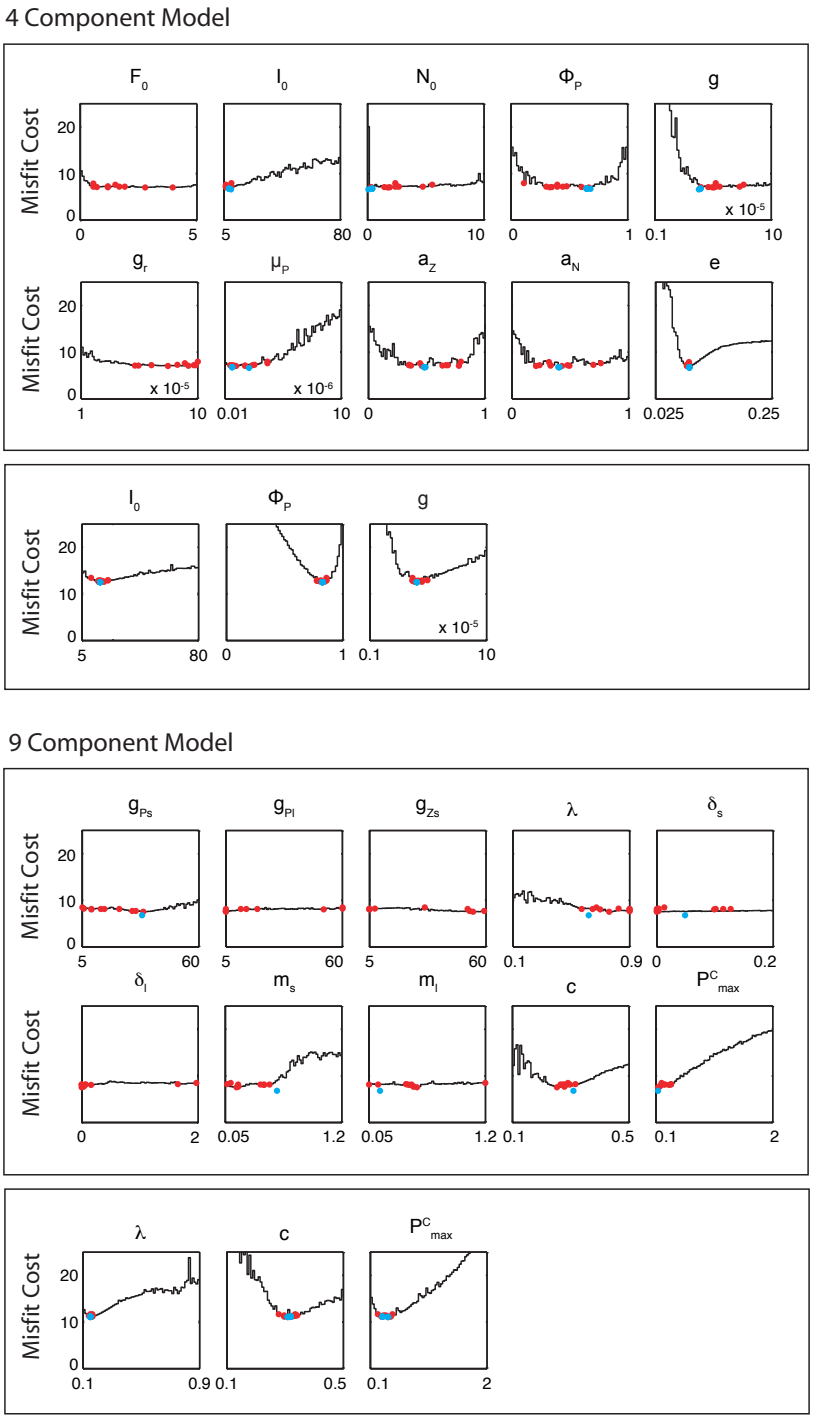

FIGURE 2.4: Collated output from the repeated optimisations. Any 10-dimensional parameter vector from the μGA can be represented by the abscissa of one point appearing in each of the 10 subplots. The shared ordinate of those points corresponds to the associated misfit value. Output from all runs was combined, with only the minimum misfit achieved for each discrete parameter value shown by the stepped black line. The red dots represent the estimates of acceptable optimal parameters, while the blue dots show the realistic and acceptable estimates of the optimal parameters from the adjoint technique. See Table 2.1 for symbols and units.

for the half-saturation for zooplankton grazing (F_0) and the light saturation constant (I_0), which were outside the range of credible values previously defined for the μGA. This is a consequence of using an unbounded search with too many poorly-constrained and partially-correlated parameters. In some cases the optimised values were physically or biologically meaningless: the first calibration of ten parameters to Arabian Sea data yielded an acceptable misfit cost of 6.79, but included a value for F_0 of 12,066 mmol N m^{-3}, with I_0 set to -8.35 Wm^{-2}. When the μGA was applied, the optimal parameters, although all realistic, were also highly variable within the defined range (Figure 2.4). All of the solutions converged to approximately the same cost of $J_{min} = 7.1$.

When the acceptable solutions were applied to model the equatorial Pacific, the VA and the μGA yielded mean predictive costs of 100.7 $\pm$ 11.5 and 131.0 $\pm$ 44.1, respectively, with the μGA showing substantially greater variability. The μGA solutions returned, on average, a worse fit to the independent data, even though these solutions did not contain unrealistic parameter values. Conversely, the very unrealistic VA solution described above yielded one of the lowest predictive costs at equatorial Pacific for the four-component model.

Three free parameters

When the number of free parameters was reduced to three, the VA returned a cost of 12.6 every time. The optimal parameters values were all realistic and showed very little variability (coefficient of variation, $C_v \leq 0.0000017\%$) (Figure 2.3, row 2), leading to very consistent costs at equatorial Pacific (181.8 $\pm$ 0.0). The μGA was equally invariant in terms of the calibrated Arabian Sea cost, but the optimal parameters were much more variable ($C_v \leq 18.4\%$), causing significant variability in cost at equatorial Pacific (184.2 $\pm$ 23.0). Although the three parameter optimisations were relatively well-constrained, the mean predictive costs at equatorial Pacific were larger than those from the more underdetermined optimisations with ten free parameters. This possibly counterintuitive result will be examined later in section 2.4.4.

2.3.2 Nine-component model

The nine-component model was optimised to Arabian Sea data as above, and the results are shown on the right hand side of Figure 2.3. In accordance with results from the four-component model, the minimum costs achieved when optimising ten parameters ($J_{min} =$ 6.8, VA; $J_{min} = 7.6$, μGA) were between 32 and 39% lower than when only 3 parameters were optimised ($J_{min} = 11.1$, both techniques). In every instance, with either ten or three free parameters, the optimisations successfully yielded lower calibrated costs at

Arabian Sea than the prior parameter values. All of the acceptable solutions performed worse than the prior parameters, in terms of cost at equatorial Pacific.

Ten free parameters

When ten parameters were optimised, the VA did not converge at all, with (at least) nine of the solutions trapped in local minima. Again, as expected, the best solution found by the unbounded VA (J_{min} = 6.8) contained four unrealistic parameter values (g_{P_s} = 0.332 d^{-1}, g_{P_l} = 3.3 d^{-1}, g_{Z_s} = -0.417 d^{-1} and m_l = -0.0314 d^{-1}). The μGA converged to a not significantly higher cost of 7.6 for all optimisations, but the solutions were very variable when applied to equatorial Pacific data with a mean predictive cost of 61.1 $\pm$ 28.6. The single acceptable VA solution had an predictive cost of 48.0. None of the techniques yielded solutions that outperformed the prior parameters at the equatorial Pacific.

Three free parameters

When only three key parameters of the nine-component model were optimised to Arabian Sea data, the VA converged every time to the same minimum cost (J_{min} = 11.07), with parameters showing little variability ($C_v \leq 4.9\%$). The μGA also converged every time to the same minimum cost, with only slightly larger parameter variability ($C_v \leq 7.1\%$). As in the four-component model, the associated costs at equatorial Pacific were more consistent than for the ten-parameter optimisations, at 32.08 $\pm$ 0.2 and 32.11 $\pm$ 3.19, for the VA and the μGA respectively. In contrast to the results from the four-component model, when only three parameters were optimised, the nine-component model yielded significantly lower mean predictive costs at equatorial Pacific than the solutions from the highly underdetermined ten-parameter optimisations.

2.3.3 Computational cost

Both techniques required multiple runs of the ecosystem models, and every 1,000 iterations took approximately 30 minutes on a Pentium 4 3.6 GHz processor with a 2 Mb cache. The VA was run for an indefinite number of iterations, with the algorithm terminating once the convergence criterion was satisfied. The number of iterations required for each run of the adjoint to converge on a global or local minimum point was between 37 and 2,649 for 10 parameters, and between 22 and 52 for 3 parameters.

The μGA was run each time for 5,000 generations, with a population size equal to the number of free parameters. This resulted in 50,000 iterations for the ten-parameter

optimisations, and 15,000 for the three-parameter optimisations. The number of μGA generations was chosen rather arbitrarily based on previous experiments with other models and data. In all but one case, the minimum cost at the 500^{th} generation was not significantly improved upon, even though the algorithm was run for an additional 4,500 generations. The longest any μGA search was able to run without any further improvement in the minimum cost was 4,964 generations (for the nine-component model with three free parameters), although in some cases improvements in the minimum misfit value occurred with less than 200 generations remaining.

2.4 Discussion

Models of marine biogeochemical systems are subject to error from a number of sources, including inadequate structure, parameter errors, physical forcing and initial conditions. For the purposes of comparing models it is possible to hold the last two sources of error constant, so that any differences in performance can be attributed to model structure and parameterisation. Furthermore, by using formal optimisation techniques, it is theoretically possible to objectively and fairly assign optimal model parameters on the basis of observations, so that models can be compared in terms of structure alone (Matear, 1995; Friedrichs et al., 2006, 2007). However, *in situ* observations of marine systems are sparse, and studies have consistently revealed that even the simplest marine ecosystem models are highly underdetermined (Matear, 1995; Hurtt and Armstrong, 1996; Fennel et al., 2001; Friedrichs et al., 2006). Parameter optimisation of non-linear systems is rarely a simple task, but even assuming the best fit can be located, the inherent underdetermination of these systems means that there will be considerable uncertainty associated with any optimal solution.

This study examined the use of two optimisation techniques, a variational adjoint (VA) technique and a micro genetic algorithm (μGA), to optimise two relatively simple (four- and nine-component) marine biogeochemical models to Arabian Sea data. The optimised solutions were subsequently evaluated against unassimilated equatorial Pacific data. The aim was to investigate the efficacy of different approaches in terms of reducing parameter error, handling underdetermination and quantifying uncertainty, rather than to compare the models themselves in terms of the residual structural errors and parameter uncertainty.

2.4.1 Model calibration

The VA and the μGA could not be distinguished with regard to their ability to lower the misfit cost in relation to assimilated data. In all experiments where Arabian Sea data were assimilated (for the four- and nine-component models, optimising both ten and three parameters) the minimum Arabian Sea costs produced by the μGA were not significantly different from the costs produced by the VA. When ten parameters were optimised, however, the VA frequently became trapped in local minima (four times out of ten with the four-component model and nine times out of ten for the nine-component model). This is attributable to the local search method of the VA, which uses the gradient of the cost function so that it always moves in the direction of a lower cost and can thus become trapped in local minima. For the same reason, however, the VA is able to descend very subtle gradients in the cost function. By contrast, the stochastic μGA, which does not utilise any gradient information, is not as proficient at descending fine gradients of the cost function, but is also not as prone to becoming trapped in local minima.

When ten parameters were optimised for either model, although similar costs were returned by both techniques, the fitted parameter values were often highly variable as the optimisations were repeated (e.g. Figure 2.4), demonstrating that the parameters were underdetermined by the data. Indeed, because the search was left unbounded, many of the solutions from the VA technique contained unrealistic and sometimes nonsensical parameter values. These values were not present in the μGA solutions, because that technique was restricted to searching only a finite and credible region of the parameter space.

By reducing the optimisation problem to just three well-constrained parameters (as defined by Friedrichs et al., 2007), although minimum costs were two to three times higher than for the ten-parameter optimisations, the problem of underdetermination was resolved. The VA did not become trapped in local minima and the solutions did not contain any unrealistic parameter values. The optimal parameter values yielded by the μGA were similarly well-constrained near the centre of the search space. Both techniques consistently returned the same minimum costs and the optimal parameters were also much more precisely defined than in the ten-parameter case. The VA returned almost identical values every time. The μGA did not converge quite as well as the VA in terms of optimal parameter values, but the solutions were equally valid in terms of minimum cost.

2.4.2 Uncertain estimates of model predictive skill

The importance of uncertainty in the optimal parameters becomes apparent when cross-validation experiments are applied, where the parameter solutions are used to generate simulations of a second, unassimilated data set. Take for example the optimisation of either model to Arabian Sea data using the μGA. With ten free parameters, the optimisations were poorly-constrained and there was much variability in the optimal parameters. Although this variation did not have a significant effect for the Arabian Sea (Figure 2.3a-ii and Figure 2.3b-ii), when the solutions were used to model the unassimilated equatorial Pacific data, the predictive costs varied by more than a factor of two.

As more parameters are optimised, the uncertainty associated with the parameters increases (Hastie et al., 2001). Each of the uncertain solutions can lead to very different estimates of the model predictive skill with regard to unassimilated data, and so if model predictive skill is to be assessed using cross-validation, it is either necessary to accurately estimate the full range of parameter uncertainty (section 2.4.3), or to fix unconstrained parameters to their default values such that all the remaining parameters are well constrained (Section 2.4.4). While the second approach has previously been effective in obtaining consistent solutions (Hemmings et al., 2003; Friedrichs et al., 2006, 2007), this technique does not necessarily take into account the uncertainty in the fixed parameter values.

2.4.3 Parameter uncertainty estimates

When optimising ten parameters using either the VA or μGA, the parameters were highly underdetermined, and their values could often be varied across a wide range without having a significant effect on the model-data misfit (Figure 2.4). Most of the ten parameters could not be precisely constrained by either the VA or the μGA (or for that matter by any other inverse technique). It is clear that in most cases, precise, scalar solutions for each parameter are unrealistically precise and will be inadequate to describe exactly what the data tell us.

When using inverse approaches to assess model performance it is important to take this uncertainty into account. Instead of looking at point solutions it will be more accurate to think of a solution region, across which the misfit costs are statistically indistinguishable from the absolute minimum. The dimensions of this region can be approximated by uncertainty estimates for the optimised parameters, and the VA and the μGA both offer ways of evaluating these.

The VA technique has the advantage of cheaply computing the full Hessian matrix, which can sometimes be inverted such that the diagonal elements provide uncertainty estimates for the parameters (Matear, 1995; Fennel et al., 2001; Friedrichs, 2001, 2002). These uncertainty estimates are defined as the amount by which each parameter can be individually perturbed from the optimum before introducing a significant increase in cost, and this distance is estimated by calculating the curvature of the cost function at the point of the minimum. Additionally, the off-diagonal elements of the inverted Hessian describe the correlations between pairs of parameters, where changes in one parameter are compensated for in terms of cost by changes in another parameter. The Hessian approach assumes that the curvature across the solution region is constant, but this is perhaps unlikely for a non-linear model, especially if the solution region is very large (Gunson and Malanotte-Rizzoli, 1996; Schartau et al., 2001). Sensitivity estimates are derived locally for what may initially be very poorly-constrained parameters, and it is possible that parameters that were insensitive at one point in the parameter space may become highly sensitive at another (Friedrichs et al., 2007). The Hessian may only be reliably inverted if it is well-conditioned, and analysis of the matrices derived here reveal that this was not the case when the models were optimised with ten free parameters.

The μGA provides no analytical estimate of the parameter uncertainties, but as a stochastic technique, it does explore a wide region of the parameter space. If all the thousands of parameter vectors evaluated by the μGA are collated together with their respective costs, all those with costs not significantly worse than the global minimum can be counted as acceptable solutions. If enough acceptable solutions are found, they will go some way to mapping out the solution region, and its extent across each parameter can be used to define the respective uncertainties. This method has an advantage over the Hessian based technique in that the uncertainty estimates are multidimensional and account for correlations between all parameters. Additionally it does not rely on the perhaps unrealistic assumption of a smooth cost function, as it makes no use of the cost gradient. Such estimates do require that the μGA has reliably located the absolute global minimum and is subsequently able to adequately map out the entire solution region. For high dimensional problems, the solution region will be convoluted and tiny relative to the searchable area (Tarantola, 2005), and so the computational expense required to reliably find and search it may be extremely high.

Although the VA and μGA are both able to provide estimates of parameter uncertainties, if too many unconstrained parameters are optimised, then no inverse method will be able to tell us very much about the parameters. More useful information may well be contained in good laboratory and field estimates of the parameter values, as long as they include accurate uncertainty estimates. If such sources of parameter information

were ignored in a model assessment, the results could easily overestimate the true level of model uncertainty.

2.4.4 Incorporating prior information

The previous section highlighted that inverse methods may not be very informative for highly underdetermined problems, and in such cases prior uncertainty estimates will do more to constrain the parameters. It would be sensible to use these prior estimates to specify the possible range of parameters that are otherwise poorly constrained by the data, but given that in many modelling studies no uncertainty estimates are given, it is perhaps not unreasonable to fix very underdetermined parameters to their precise prior values instead (Hemmings et al., 2003; Friedrichs et al., 2007).

The following section examines this approach as each model was optimised with at first ten, and then three free parameters. The analysis here is restricted to results from the μGA optimisations, so that the focus is on the effects of fixing unconstrained parameters, rather than on the differences between optimisation techniques. (The μGA was selected simply because it gave the most consistent results with ten free parameters, and it should be noted that the μGA and VA techniques produced almost identical results with three parameters; the only difference being that the VA parameter solutions were less variable.)

The mean predictive cost for the μGA solutions with the nine-component model decreased (from 61.1 $\pm$ 28.6 to 32.1 $\pm$ 3.2) when the number of free parameters optimised to Arabian Sea was reduced from ten to three. This is consistent with the findings of Friedrichs et al. (2007), where model predictive skill was seen to increase when only well-constrained parameters were optimised. The opposite pattern, however, was seen here in the four-component model, where the mean predictive cost went up (from 131.0 $\pm$ 44.1 to 181.8 $\pm$ 23.0) when only three, rather than ten, parameters were optimised. The grey boxes in Figure 2.3 represent the range of the predictive costs from all the acceptable solutions with ten and three free parameters. Although the general response of the mean predictive cost to removing unconstrained degrees of freedom was inconsistent, in both cases the solutions became much less variable when only three parameters were optimised.

The variability of the predictive costs was reduced when only three parameters were optimised because seven of the unconstrained parameters were fixed to precise values. In reality, those values may be quite uncertain, and could potentially influence the model misfit in relation to both the assimilated and unassimilated data. The pattern described in the previous paragraph is consistent with the fact that the four-component model

was developed for the Arabian Sea, while the nine-component model was developed for the equatorial Pacific (section 2.2.1). Bearing in mind that the default parameters were also assigned for those sites, it is perhaps not surprising that the performance of the nine-component model improved at the equatorial Pacific when more parameters were fixed to their prior values, while the performance of the Arabian Sea developed model was diminished.

This idea was examined with a cursory sensitivity analysis, where the values of the seven non-optimised parameters in the four component model (previously developed for the Arabian Sea (McCreary et al., 1996)) were replaced with values optimised to equatorial Pacific data. When the remaining parameters were optimised to Arabian Sea data, the mean predictive cost in the equatorial Pacific was reduced by 59%. Although this analysis was crude, it demonstrates the possibility that the unconstrained parameters can significantly affect the performance of the model with respect to unassimilated data. By fixing the unconstrained parameters to precise values, the uncertainty of the problem may be significantly underestimated, with the mean predictive cost perhaps strongly influenced by the values of the default parameters.

2.4.5 Implications

This chapter investigated two different approaches to using objective parameter optimisation for model evaluation and comparison, although the results are also relevant to applications where the goal is to develop a set of optimal parameters for forward modelling studies. The results presented here support the view that even simple marine biogeochemical models are underdetermined by the observations currently available at oceanic time-series sites, and thus no inverse techniques will be able to find uniquely determined solutions for all the parameters. This is not to say however, that we should always be satisfied just using off-the-shelf parameters. Although the solutions yielded by parameter optimisation techniques should only be used when they represent an improvement on our prior knowledge, that prior knowledge should not be assumed to be overly precise.

Estimation of parameter uncertainty is particularly important for the comparison of models of different complexity. Greater complexity allows models to resolve a more diverse range of biogeochemical dynamics, and this in turn may make them more generally applicable at different sites, but the additional parameters required to describe extra components are likely to add more unconstrained degrees of freedom. While such models can be more heavily tuned to provide a better fit to the data, and may resolve more explicit processes, the benefits of this additional ability to reproduce observations

should always be balanced against any increase in uncertainty associated with the extra parameters.

If unconstrained parameters are assigned precise prior values it is likely that the model uncertainty will be underestimated. On the other hand, if the only information used to constrain each model comes from a severely underdetermined optimisation experiment, then the solutions will exaggerate any uncertainty. Both of these scenarios were examined here. When only three well-constrained parameters were optimised, the results agree with previous work by Friedrichs et al. (2007) in that the more complex model showed much greater predictive skill than the simpler model. It can however be argued that this result does not account for uncertainty in the unconstrained parameters that were fixed to prior values. Some of this uncertainty was considered when 10 parameters were optimised within broad prior constraints, and this did indeed give more variable estimates of model predictive skill. After accounting for parameter uncertainty in this way however, the results still support the conclusion that the multiple size-class model had greater predictive skill, although further work would be needed to account for the effects of all 23 parameters.

Neither of the approaches presented in this chapter represent a perfect solution to dealing with uncertainty in parameter optimisation problems, but is is clearly important that this should somehow be quantified. The μGA approach is promising in this regard, and in subsequent chapters this method will be developed and applied using a different set of models and data. These are outlined in chapter 3, and a method of quantifying parameter uncertainty is developed in chapter 4.

Chapter 3

Methods

3.1 Introduction

Subsequent chapters in this thesis will focus on the performance of five closely related NPZD models at two North Atlantic sites, examining the influence of model physics on ecosystem model performance (chapter 7), as well as the ability of the models to reproduce (chapter 5) and predict (chapter 6) biogeochemical observations. The first of the two time-series sites, the Bermuda Atlantic Time-series Study (BATS) lies firmly within the oligotrophic subtropical gyre (31°40N, 64°10W), where the yearly cycle of phytoplankton growth is characterised by low biological productivity during the stratified summer months followed by a moderate bloom in response to winter mixing (Menzel and Ryther, 1961). The second location, the North Atlantic Bloom Experiment (NABE) site is located to the north-east of the subtropical gyre (47°N, 20°W), and is characterised by a strong spring bloom that occurs upon shoaling of the mixed layer (Ducklow and Harris, 1993). The two sites each show a contrasting response to changes in the mixed-layer depth, with phytoplankton growth triggered by the wintertime input of nutrients at the oligotrophic BATS, while the reduced light limitation upon shoaling of the mixed layer in spring is a more important factor at the eutrophic NABE (Follows and Dutkiewicz, 2002). The two sites thus provide a useful contrast with regard to model assessment.

The five models (introduced in section 3.5) are set up within a one-dimensional framework. The representation of vertical structure goes beyond the incorrect assumption that all biological matter and growth is entirely restricted to, and homogenous within, the ocean mixed layer, and so represents an improvement over zero-dimensional models. Increasing the vertical resolution of the biological model to include a number of different layers requires the mixing and advection of nutrients and particulate biomass. This represents a relatively large computational overhead, although the one-dimensional model is still cheap enough to evaluate many thousands of times in a few hours.

If the problem under investigation can be stated by asking how suitable are the five models in question for the modelling of the BATS and NABE ecosystems, or more ambitiously, what level of model complexity is required at these sites, then the data and techniques presented in this chapter represent the tools by which these questions may be addressed. How these tools are applied to find a solution is left for discussion in chapter 4. In this chapter, the physical forcing data are described first in section 3.2 and 3.3, with the biogeochemical observations at BATS and NABE subsequently described in section 3.4. The five ecosystem models that are used in an attempt to reproduce these data are introduced in section 3.5. In order to systematically optimise model parameters it is necessary to summarise the model-data misfit into a single value and the cost

function outlined in section 3.6 is used for this purpose. Finally, in section 2.2.4 the micro genetic algorithm optimisation routine is described.

3.2 Physical forcing

The ecosystem models require forcing data for surface irradiance, temperature, vertical diffusivity and optionally, vertical velocity. These were provided in the output of the Ocean Circulation and Climate Advanced Model (OCCAM) (Sinha and Yool, 2006) for BATS and NABE observation sites. For comparison, similar forcing data were also available at BATS from the Parallel Ocean Program (POP) model (Lu et al., 2007). The physical forcing resolves the daily cycle with a temporal resolution of one hour at BATS and two hours at NABE. The physical model output for those years where the biogeochemical models were evaluated is shown in figures 3.1.

3.2.1 Ocean Circulation and Climate Advanced Model (OCCAM)

The OCCAM model data were provided by Bablu Sinha and Andrew Yool at the National Oceanography Centre, Southampton. Data were derived from global coupled ocean biology-climate runs with 1° horizontal resolution. The 66 model levels vary in thickness from ∼5 m at the surface to ∼210 m at depth. Surface forcing was provided by NCEP/NCAR reanalysis data, with surface fluxes of heat, momentum and evaporation estimated from these data using empirical formulae. Surface salinity was nudged using a weak relaxation to Levitus et al. (1998) climatology. Mesoscale eddies were parameterised after Gent and McWilliams (1990). A K-profile parameterisation (KPP) mixing scheme (Large et al., 1994) was used to parameterise unresolved vertical mixing. A more complete description of the model physics is given in Sinha and Yool (2006).

3.2.2 Parallel Ocean Program (POP) general circulation model

The POP model data were provided for BATS by Markus Pahlow at the Leibniz Institute of Marine Sciences at the University of Kiel. The data were originally derived from an Atlantic run of version 2.0.1 of the Parallel Ocean Program (Lu et al., 2007). The horizontal resolution at BATS is 1° of longitude and 0.97° of latitude. The model grid contains 23 vertical levels, varying in thickness from ∼10 m at the surface up to ∼500 m at depth. As with OCCAM, surface fluxes were derived from NCEP/NCAR reanalysis data. Both temperature and salinity were restored to observed climatologies using a

"time-mean state nudging" (Lu et al., 2007). This is a more rigourous data assimilation scheme than was used in the OCCAM run, where only surface salinity was restored. Mesoscale eddies and vertical mixing were parameterised as in OCCAM.

The POP data were provided for only the surface 200 m at BATS, and on a different temporal and spatial grid to the OCCAM data. To ensure that any differences in the ecological model performance could be attributed to differences in the physical forcing itself, rather than to the use of a different gridding scheme, the POP data were linearly interpolated to correspond directly to the OCCAM data. Where no data were available (i.e. below 200 m), the physical fields were extrapolated as follows; temperature at each time-step was set to the mean value at 200 m; vertical velocity below 200 m was et to the level at 200 m; and vertical diffusivity was set to 10^{-4} m^2s^{-1}, after Friedrichs et al. (2007). The extrapolation of temperature and vertical velocity below 200 m will not affect the ecosystem model output because the only temperature-dependent biological process (phytoplankton growth) is not active below this depth, and advection due to vertical velocities was not applied for any runs using the POP physics.

It is possible that the linear interpolation of data could have a significant effect on the ecological model as a consequence of non-linear responses to the forcing. This may be particularly true for the vertical diffusivity field, which is highly variable in the raw dataset on short time-scales. This variability was suppressed as the POP data were interpolated onto the OCCAM model grid, but it should be noted that the averaged vertical diffusivity was always sufficient to homogenise the modelled nitrate field within the mixed layer during preliminary model runs.

3.3 Physical model assessment

The physical forcing data derived from the two GCMs are compared to *in situ* observations in figure 3.2. Temperature and mixed-layer depth data were downloaded from online databases for BATS (BATS, 2008) and NABE (Kleypas and Doney, 2001). The OCCAM model consistently underestimates the surface temperature at BATS by ∼1°C during the summer and by up to 3°C in the winter. The temperature at 200 m is ∼1°C colder than observations at the start of 1991, with the error increasing to ∼2°C by the end of 1995. The POP model, which was nudged to climatological temperature and salinity data (Lu et al., 2007), does a better job of reproducing observations, with no error in the surface or 200 m temperature fields discernible by eye.

Mixed-layer depth (MLD) was defined as the minimum depth at which the local temperature exceeds the surface value by more than 0.1°C. Although this statistic was not

explicitly required by any of the ecological models, which use vertical diffusivity to redistribute matter within the water column, it provides a useful diagnostic for evaluating the depth of the homogenous surface layer. At BATS, the maximum MLD in the POP model is never greater than 190 m, whereas the OCCAM model has strong mixing down to as much as 450 m in some years. The maximum MLD observed at BATS was approximately 300 m, although it was generally not seen to be greater than 200 m.

The limited data available at NABE suggest that the OCCAM model underestimates the temperature at both the surface and 200 m by between 2 and 3°C. Observations also indicate that the mixed layer in the OCCAM model shoals too early. The accurate timing of shoaling events is critical for the correct modelling of phytoplankton blooms (Friedrichs et al., 2006), so to allow a better match to data all the physical forcing data (solar radiation, temperature and vertical diffusivity) were held back by 10 days. The original dataset is shown in figure 3.2 with grey lines, the time-shifted dataset that is subsequently used in all experiments is shown with black lines.

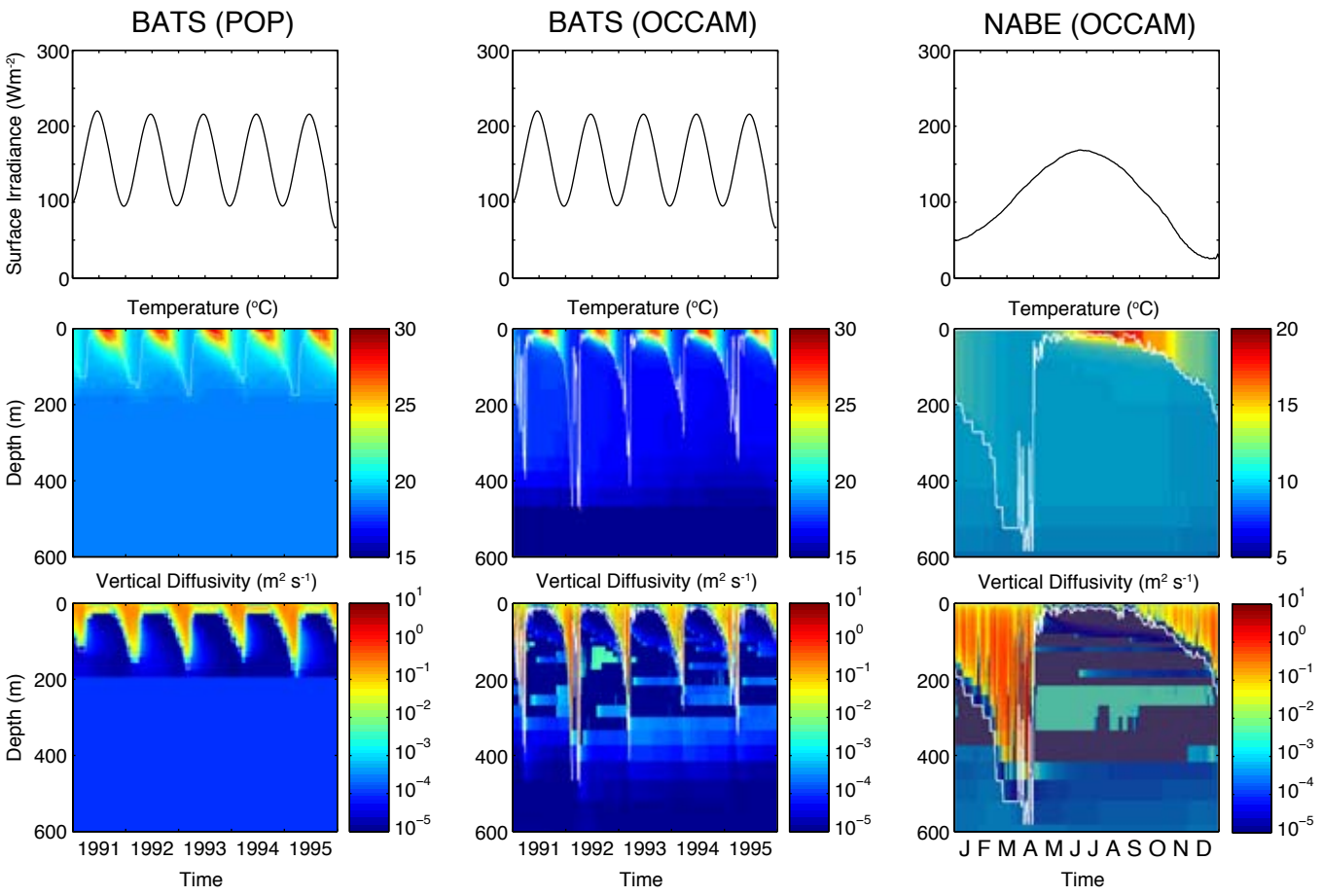

FIGURE 3.1: Physical forcing derived from the OCCAM and POP GCMs. Note that the POP model forcing has been resampled to fit the vertical and temporal resolution of the OCCAM forcing and the surface irradiance data is taken directly from the OCCAM model. Below 200 m the POP temperature and vertical diffusivity fields are filled in as described in the text. Estimated mixed layer depths are shown as white lines.

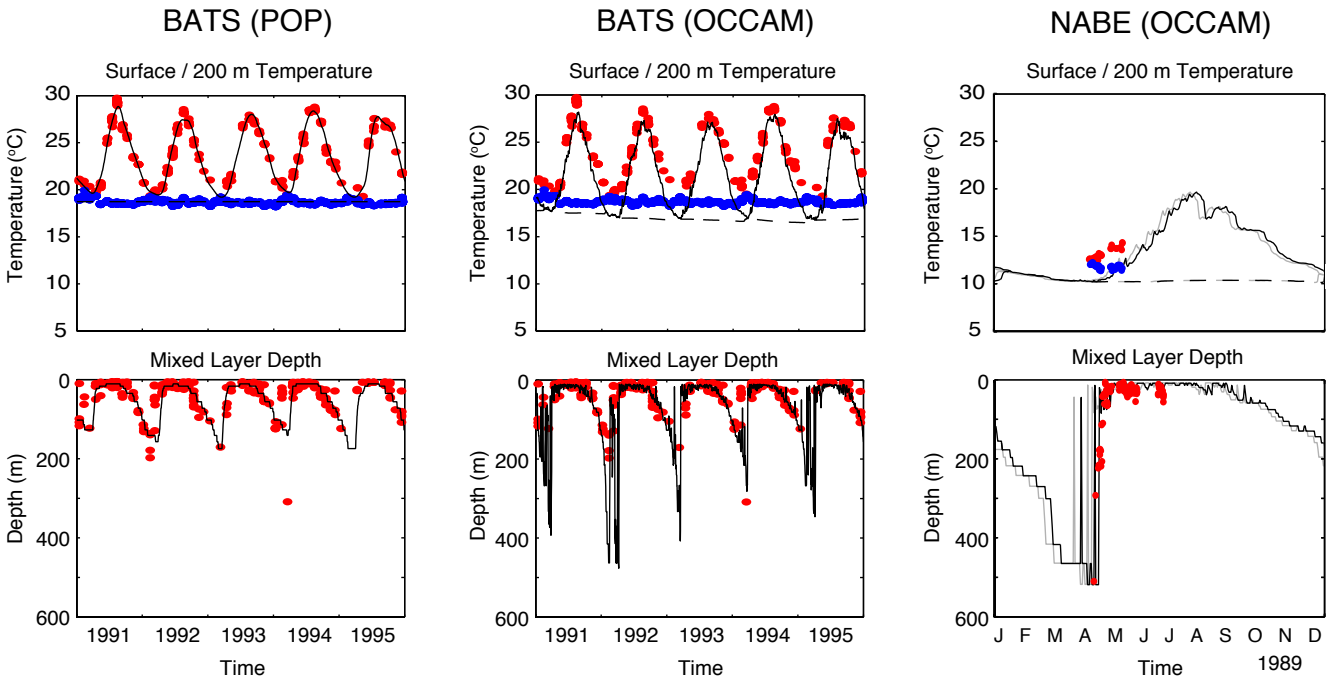

FIGURE 3.2: Upper row: Water temperature (°C) from observations and equivalent POP and OCCAM model output for the surface (red dots, solid line) and 200 m (blue dots, dashed line). Lower row: Observed and modelled mixed-layer depth (m) based on a $\Delta 0.1$°C temperature criterion. The physical (and biogeochemical) observations at NABE were moved forward by 10 days to allow a better fit. The original trajectories of the OCCAM temperature and mixed-layer depths at NABE are shown by grey lines.

3.4 Biogeochemical data

Biogeochemical data from BATS and NABE were provided by Iris Kriest at the Leibniz Institute of Marine Sciences at the University of Kiel. Data include dissolved inorganic nitrogen (DIN), particulate organic nitrogen (PON), chlorophyll *a* and primary production as carbon uptake (CUp). Zooplankton data from BATS were provided by Debbie Steinberg and Joe Cope at the Virginia Institute of Marine Science (Madin et al., 2001). BATS data (excluding zooplankton) were from bi-weekly to monthly surveys, for the years 1991 - 1995 (Michaels and Knap, 1996). Zooplankton data were collected on a monthly basis from April 1994 - December 1995 (Madin et al., 2001). NABE data were restricted to a short period coincident with the 1989 spring bloom and data were only available from 25$^{\text{th}}$ April to 8$^{\text{th}}$ May of that year. The mean and standard deviations of those observations entering the cost function (i.e. those in the surface 200 m) are shown in table 3.1.

Observed DIN corresponds to nitrate plus nitrite, except where nitrite data were not available and nitrate observations are used alone. Discrete bottle samples were measured using a reagent based colourimetric analysis technique (Knap et al., 1993). The limit of detection is 0.03 mmol N m^{-3} for nitrate plus nitrite and 0.05 mmol N m^{-3} for nitrate. Chlorophyll *a* measurements where taken from discrete bottle samples using a fluorometric technique. PON was measured as nitrogen oxides released from combusted samples of particulate matter. Primary production was estimated using a 24-hour radiocarbon technique (Knap et al., 1993), where water samples were spiked with ^{14}C and incubated *in situ* in light and dark bottles. The differential rate of ^{14}C uptake into particulate matter was used to estimate net primary production.

Mesozooplankton data were collected at BATS during day and night oblique towed casts of a 1 m^2, 200 μm mesh net across the surface layer (approximately 200 m). The resultant data consist of the integrated mesozooplankton biomass across the surface layer. The net collection technique measures only those zooplankton larger than the 200 μm mesh. The modelled zooplankton however must account for the effects of all sizes of zooplankton and so observations of mesozooplankton alone may be inappropriate for constraining this state variable. Observations of small (< 200 μm) zooplankton are sparse, especially during the study period, and none were available for assimilation at any site. Microzooplankton are likely to be most important at oligotrophic sites (Caron et al., 1995), and so observations of mesozooplankton at BATS (Z_m) were converted to estimates of the total zooplankton (Z) biomass after the data of Roman et al. (1995), with a relationship of $Z = 1.23Z_m + 0.097$ mmol [N] m^{-3} (figure 3.3). The y-intercept term of 0.097 mmol [N] m^{-3} is relatively large when compared to the mean mesozooplankton biomass of 0.018 mmol [N] m^{-3} and suggests a baseline concentration of microzooplankton.

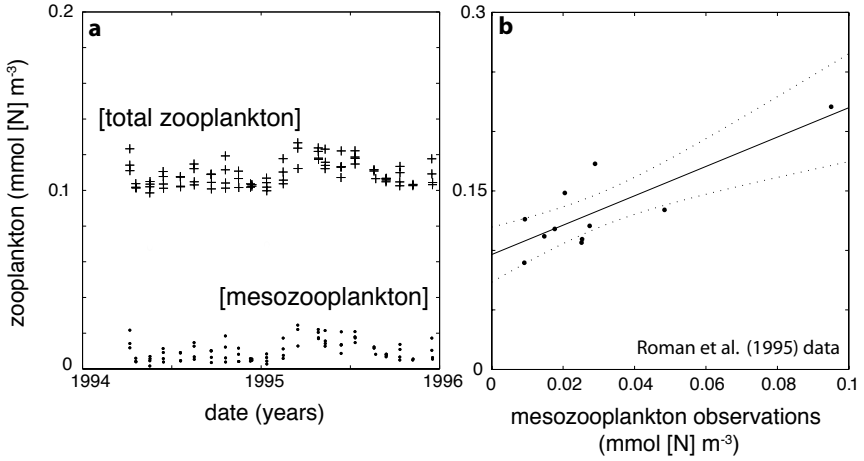

FIGURE 3.3: Conversion from observed mesozooplankton ($> 200\mu m$) to estimated total zooplankton at BATS. Observed mesozooplankton data are plotted against time as dots in panel **a**. They are converted to estimates of total zooplankton (crosses in panel **a**) using a linear regression of observed mesozooplankton and total zooplankton at Bermuda (Roman et al., 1995) (panel **b**).

TABLE 3.1: Summary statistics ($\hat{x} \pm \sigma^2$) for observations entering the cost function.

	BATS	NABE	Units
DIN	0.48±0.67	5.88±1.39	mmol N m^{-3}
Chlorophyll *a*	0.15±0.11	0.88±0.49	mg Chl m^{-3}
Zooplankton	0.11±0.007	n/a	mmol N m^{-3}
PON	0.31±0.14	1.22±0.67	mmol N m^{-3}
1° production	0.27±0.33	1.97±1.56	mmol C $m^{-3}d^{-1}$

3.5 Biogeochemical models

The five models selected for analysis are all of basic NPZD structure and are nested within a one-dimensional framework with same the 66 level vertical grid as the OCCAM model. These models lie firmly towards the lower end of the range of complexity currently employed in marine biogeochemical modelling. They were selected on the grounds that models with different components were also very closely related in their equations, thus making for easier comparison. The basic ecosystem model was taken from Oschlies and Garçon (1999) and was adapted to include a number of additional ecological pathways. The modelled state variables represent a highly simplified version of the marine ecosystem, within a one-dimensional vertical grid divided into 66 discrete and homogenous layers. The vertical distribution of state variables is modelled as a function of time, driven by time-varying fields of light, turbulent diffusion and (optionally) vertical

TABLE 3.2: Estimated standard deviations of observational errors, used as the weighting term σ in the cost function. These figures are taken from Schartau and Oschlies (2003a)

	BATS	NABE	Units
DIN	0.1	0.1	mmol N m^{-3}
Chlorophyll *a*	0.01	0.01	mg Chl m^{-3}
Zooplankton	0.01	n/a	mmol N m^{-3}
PON	0.0357	0.0357	mmol N m^{-3}
1° Production	0.3	1.0	mmol C $m^{-3}d^{-1}$

advection. The general equation for the time varying tracer concentration, Ψ_i, within each level (k) is given by

$$\frac{\partial \Psi_k}{\partial t} = -w_k \frac{\partial \Psi_k}{\partial z} + \frac{\partial}{\partial z}(K_\rho \frac{\partial \Psi_k}{\partial z}) + SMS(\Psi_k) \tag{3.1}$$

where K_ρ is the turbulent mixing coefficient and w_k is the vertical advection of tracers, which includes the sinking velocity of detritus, w_s. The final term on the right-hand side represents the source-minus-sink terms for the ecosystem model state variables. In all cases the biogeochemical model equations were stepped forward using a 4^{th}-order Runga-Kutta scheme with a time step of 15 minutes, regardless of the physical time step. A detailed description of the model equations is given in Appendix C.

The ecosystem model was initially set up with four state variables, namely DIN, phytoplankton, zooplankton and detritus. This model requires 16 biogeochemical parameters, and these are listed in table 3.3. Of the 16 parameters, only 12 were subjected to optimisation, with the remaining four fixed to well-established prior values (Schartau and Oschlies, 2003a).

The model was subsequently adapted to include a number of additional ecosystem pathways. Model two was developed after Oschlies (2001) with the inclusion of a "fast-recycling pathway" from phytoplankton to DIN. The microbial loop has been previously shown to be very important at oligotrophic sites such as BATS (Fasham et al., 1990), and the additional pathway represents an implicit parameterisation of this component of the ecosystem. The new linear pathway is parameterised using the existing phytoplankton loss rate, Φ_m^P. The loss of phytoplankton biomass to detritus is changed from linear to quadratic form, requiring an additional parameter, Φ_P^*.

The decoupling of chlorophyll *a* concentration from phytoplankton biomass allows accumulation of chlorophyll *a* in high-nutrient, low-light environments such as deep chlorophyll maxima (**?**), where the additional chlorophyll *a* biomass allows more efficient use of the dim light. Model three was thus adapted to include an empirical function for a variable chlorophyll *a* to nitrogen (Chl:N) ratio (Cloern et al., 1995). Photoacclimation was also included in the form of an adaptable phytoplankton growth efficiency (Anderson et al., 2007). The Cloern et al. (1995) function contains 4 parameters estimated from experimental data heavily biased towards coastal diatoms. Although it initially seemed sensible to optimise these parameters, preliminary experiments revealed that the data at both BATS and NABE were inadequate to do so, with variable and unrealistic solutions obtained. For this reason the parameter values of Cloern et al. (1995) were retained and model three had only 12 free parameters, as for model one. Photoacclimation was incorporated by making the initial slope of the photosynthesis irradiance curve α directly proportional to the Chl:N ratio, such that a doubling of Chl:N relative to the default fixed value of 1.59 mg Chl (mmol [N])$^{-1}$ led to a doubling of α relative to the parameterised value. The model equations are described in more detail in Appendix C.

Model four includes the fast-recycling pathway, the variable Chl:N ratio and photoacclimation, requiring a total of 13 tunable parameters.

Finally, in addition to the pathways included in model four, model five includes a pathway for recycling via semi-labile dissolved organic nitrogen (DON), loosely based on the models of Anderson and Williams (1998) and Dadou et al. (2001). This model contains 15 tunable parameters. Models one to four are said to be nested within model five, in that they are all a special cases of model five, albeit with certain pathways switched off. The use of nested models allows model output and performance to be more readily compared in terms of model structure and complexity. The five ecosystem models are shown schematically in figure 3.4.

Each model was initialised with a nitrate profile derived from the mean of the observations. Other state variables were initialised to very small values. Models were spun-up for at least one year prior to the assimilation of biogeochemical data, a time period that was found in preliminary experiments to bring the model state variables to the same order of magnitude as the observations.

3.6 The cost function

The cost function was used as a means of quantitatively comparing the model with observations. It is modified from Schartau and Oschlies (2003a), where it is defined as

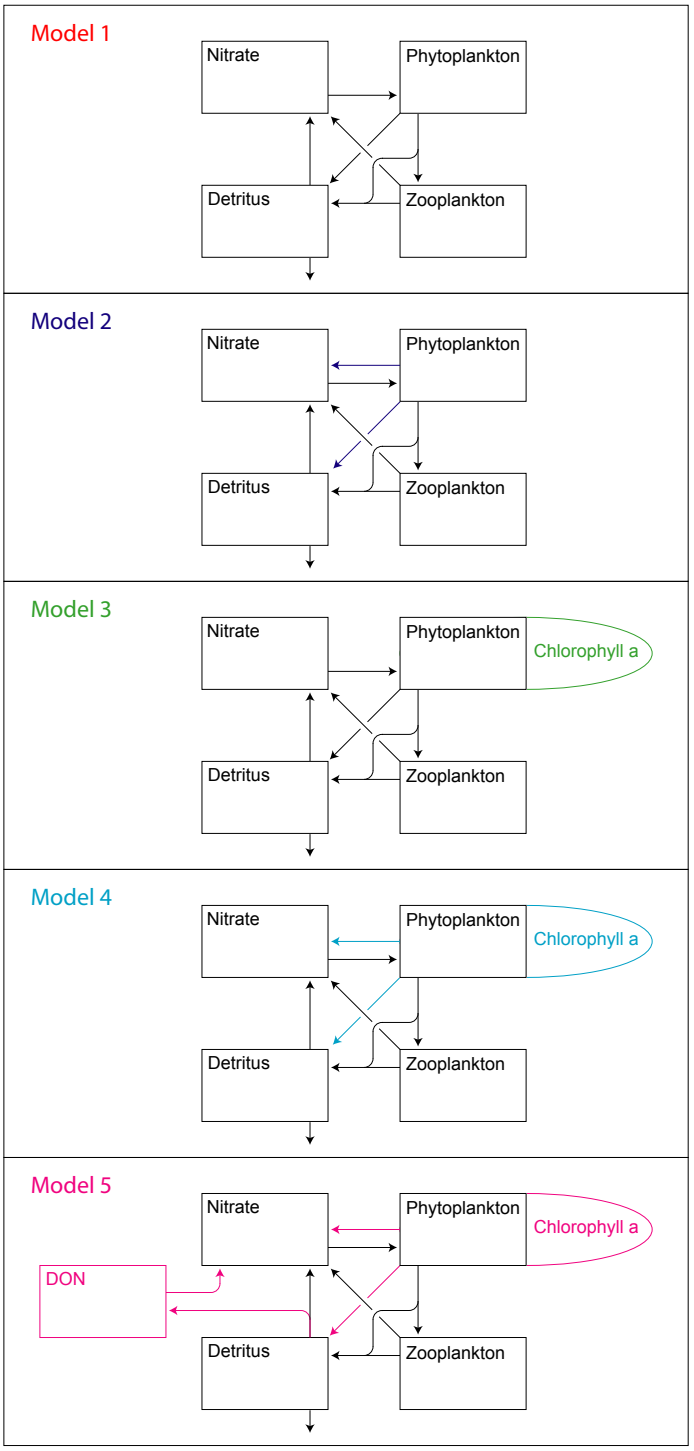

FIGURE 3.4: Nested NPZD type models

a weighted sum of squared errors between model results and observations. The misfit is dimensionless. The optimisation process seeks the smallest possible misfit value.

The cost function outlined by Schartau and Oschlies (2003a) used observations binned into monthly mean values, which were compared to climatological monthly mean values derived from the model output. This was done to avoid the situation where small temporal and spatial mismatches between the observations and an otherwise consistent model lead to significant increases in cost. Such "phase errors" may significantly affect calibrated parameter estimates, as compensatory parameter errors may be introduced to counteract any increase in cost. A disadvantage of the averaging approach, as highlighted by Schartau and Oschlies (2003a), is that the observations are sparse in both time and space, and so any mean values derived from these data are not necessarily representative. In this study each observation was compared with an equivalent value generated by the model. Model output was interpolated in time and space to match the observations. Where the observed values represent integrated measurements, the model output was also integrated. For example, modelled zooplankton were integrated across the surface ∼200 m before they were compared with observations. Likewise, primary production was measured across a 24 hour period using the ^{14}C technique, and so modelled primary production was integrated over 24 hours before entering the cost function. This approach was necessary because the diel cycle of solar irradiance was resolved in the physical forcing.

The approach of manipulating the model output to match the observations was selected as it is more accurate to interpolate the relatively dense model output than it is to process the sparse observations in a similar way. A disadvantage of this approach is that any phase errors in the underlying physics will be mapped directly into the cost function (see chapter 4). The observations were presented in a way that resolves interannual variability at BATS, and so the steady-state constraint of Schartau and Oschlies (2003a) was not required at that site. It was also omitted at NABE. The total misfit cost over the model run time (not including the spin-up year) is subsequently defined by

$$J = \frac{1}{M} \sum_{m=1}^{M} \frac{1}{N_m \sigma_m^2} \sum_{n=1}^{N_m} (\hat{a} - a)_{nm}^2 \tag{3.2}$$

Model output was compared to observations of the five different data types, m (DIN, Chlorophyll a, Zooplankton, PON and primary production). The squared differences between observations, $\hat{a}_{nm}$ and model equivalent a_{nm}, were summed and weighted according to the number of observations of each type, N_m, and the estimated standard deviation of the measurement error, σ_m^2 (see table 3.2). Only those observations within the surface 200 m were included in the cost function, with the total misfit given by J.

Schartau and Oschlies (2003a) applied a station scaling term to ensure a similar misfit values from each site, which would otherwise be skewed by the higher dynamic range of more eutrophic sites such as NABE. Such "simultaneous optimisation" is not considered in this study, and so the station scaling term is not applied. The form of equation 3.2 is roughly equivalent to that of equation 2.1 in chapter 2, although different weighting terms were applied there to remain consistent with Friedrichs et al. (2007).

3.7 The micro genetic algorithm (μGA)

The optimisation routine is described in section 2.2.4 of chapter 2. The upper and lower limits for the parameters of models one to four were based on the values of Schartau and Oschlies (2003a), although they were adapted so that each range of discrete values included the exact prior value for that parameter, while also making sure that every increment within the range could be expressed completely by using no more than three significant figures. The limits set on the DON parameters of model five were similarly defined with reference to Anderson and Williams (1998) and Dadou et al. (2001). The prior ranges for all parameters (given in table 3.3) were set to conservatively broad values so that no realistic values were excluded from the solutions.

3.8 Summary

The modelling and optimisation framework that was outlined in this chapter provides the means by which the performance of the five ecosystem models can be objectively quantified and optimised at BATS and NABE. Objective parameter optimisation is useful in model comparison studies, as model performance is highly dependent on what are often subjectively defined parameters (Friedrichs et al., 2007), but the results presented in chapter 2 confirm that it is not enough to simply compare models in terms of minimum cost. The high number of unconstrained parameters in almost all ecosystem models means that it is essential to account for uncertainty in some way, especially when comparing models of different complexity. In chapter 2 this was achieved by repeatedly initialising the parameter optimisation algorithms from different points in the parameter space, although it was suggested that more reliable methods exist (e.g. Schartau and Oschlies, 2003a).

In subsequent chapters of this thesis the aim is to qualitatively and quantitatively evaluate the performance of the five ecosystem models while also analysing ecosystem model behaviour at the two North Atlantic sites. The optimal performance of the ecosystem

models is relatively easily assessed using parameter optimisation, but is not clear that any such solution will be reliable or robust. In the following chapter the tools and techniques that were outlined in this chapter are examined with a view to developing robust solutions that goes some way to quantifying the uncertainty that is inherent in almost all biogeochemical models (e.g. Matear, 1995; Schartau and Oschlies, 2003a; Raick et al., 2006; Friedrichs et al., 2006). These solutions will be applied in the remaining chapters as the study goals are addressed.

TABLE 3.3: Model parameters with prior values and bounds placed on the μGA search.

Parameter	Symbol	Default	Range	Units
Fixed parameters common to all model implementations				
⋆ Growth coefficient	C_{ref}	1.066	-	-
⋆ Growth coefficient	c	1.0	-	$(^\circ\text{C})^{-1}$
⋆ PAR extinction coefficient	k_w	0.04	-	m^{-1}
⋆ Short-wave PAR fraction	f_{PAR}	0.43	-	-
Tunable parameters common to all model implementations				
⋆ P growth rate parameter	μ_m	0.6	0.2 - 1.46	d^{-1}
⋆ Slope of P:I curve	α	0.025	0.001 - 0.253	$\text{m}^2\text{W}^{-1}\text{d}^{-1}$
⋆ N uptake half-saturation	k_N	0.5	0.1 - 0.73	mmol N m^{-3}
⋆ P loss rate	Φ_m^P	0.01	0.0 - 0.63	d^{-1}
⋆ Maximum grazing rate	g	2.0	0.04 - 2.56	d^{-1}
⋆ Grazing encounter rate	ε	1.0	0.025 - 1.6	$\text{m}^6(\text{mmol N})^{-2}\text{d}^{-1}$
⋆ Assimilation efficiency	β	0.75	0.3 - 0.93	-
⋆ Z loss rate parameter	Φ_m^Z	0.01	0.0 - 0.063	d^{-1}
⋆ Z quadratic mortality	Φ_Z^*	0.205	0.01 - 0.955	$\text{m}^3(\text{mmol N})^{-1}\text{d}^{-1}$
⋆ D breakdown rate	γ_m	0.02	0.02 - 0.146	d^{-1}
⋆ D sinking velocity	w_s	6.0	2.0 - 128.0	m d^{-1}
⋆ Light attenuation by P	κ	0.03	0.01 - 0.073	$\text{m}^2(\text{mmol N})^{-1}$
Tunable parameters common to models 2, 4 and 5				
⋆ P quadratic mortality	Φ_P^*	0.045	0.0 - 0.945	$\text{m}^3(\text{mmol N})^{-1}\text{d}^{-1}$
Fixed parameters common to models 3, 4 and 5				
† Cloern parameter a	C_a	0.003	-	mg Chl $(\text{mmol N})^{-1}$
† Cloern parameter b	C_b	0.0155	-	mg Chl $(\text{mmol N})^{-1}$
† Cloern parameter c	C_c	0.05	-	$(^\circ\text{C})^{-1}$
† Cloern parameter d	C_d	0.059	-	$(\text{Wm}^{-2})^{-1}$
Tunable parameters in model 5				
∗ D to semi-labile DON	Φ_D	0.24	0.0 - 0.945	-
∗ Remineralistion of DON	γ_D	0.003	0.001 - 0.064	d^{-1}
Sources for default parameter values				
⋆ Schartau and Oschlies (2003a)				
† Cloern et al. (1995)				
∗ Anderson and Williams (1998); Dadou et al. (2001)				

Chapter 4

Defining a solution

4.1 Introduction

The results presented in chapter 2 agree with previously published work (e.g. Matear, 1995; Schartau and Oschlies, 2003a; Friedrichs et al., 2006) in suggesting that point estimates are inadequate to define the solutions to underdetermined parameter optimisation problems. This issue can be overcome by fixing any unconstrained parameters to precise prior values (Friedrichs et al., 2006), but if these parameters are not well defined then the values to which they are fixed may become an unresolved source of error.

Instead of removing unconstrained degrees of freedom, a number of studies have attempted to quantify the uncertainty associated with a full set of unknown parameters (Matear, 1995; Schartau and Oschlies, 2003a; Schartau et al., 2007). One way of achieving this is to compute the full Hessian matrix of second order partial derivatives. This can often be produced when a variational adjoint method is used (e.g. Lawson et al., 1995; Spitz et al., 1998; Friedrichs, 2001). The Hessian matrix can sometimes be inverted such that the diagonal elements provide uncertainty estimates for the parameters (Matear, 1995; Fennel et al., 2001; Friedrichs, 2001, 2002). These are defined as the amount by which each parameter can be individually perturbed from the optimum before introducing a significant increase in cost, and this distance is estimated by calculating the curvature of the cost function at the point of the minimum. Additionally, the off-diagonal elements of the inverted Hessian describe the correlations between pairs of parameters, where changes in one parameter are compensated for in terms of cost by changes in another. The Hessian approach assumes that the curvature across the solution region is constant, but this is perhaps unlikely for a non-linear model, especially if the solution region is very large (Gunson and Malanotte-Rizzoli, 1996; Schartau and Oschlies, 2003a). Sensitivity estimates are derived locally for what may initially be very poorly-constrained parameters, and it is possible that parameters that were insensitive at one point in the parameter space may become highly sensitive at another (Friedrichs et al., 2007). The Hessian may only be reliably inverted if it is well-conditioned, and again this seems unlikely to be the case for highly underdetermined models such as are presented here.

An alternative approach to defining a solution was taken in Schartau and Oschlies (2003a) and Schartau et al. (2007). The authors used a genetic algorithm to optimise a large set of parameters in each study (13 and 14) and attempted to quantify the uncertainty for each parameter by performing a number of different optimisations, each time with a different set of synthetic noise added to the original observation data. This additional noise caused variability among the optimal parameter estimates and this was recorded in both variance and bias estimates. The latter statistic was derived by calculating the distance between the optimum estimate for each parameter and the mean

estimate from all the optimisations. An advantage of this technique is that it does not rely on the perhaps unrealistic assumption of a smooth cost function, as it makes no use of the cost gradient, and the uncertainty estimates are not sensitive to the precise locations of what may be poorly defined optima. These statistics provide useful information on the bias and variance of the optimal parameters, but do not consider parameter correlations as the Hessian does. Furthermore the high computational expense of the genetic algorithm makes it very difficult to build up a statistically robust sample.

In this chapter the first model described in section 3.5 is calibrated and the optimal parameter estimates are examined in terms of their accuracy and uncertainty. In order to do this a twin experiment was developed where the real data outlined in section 3.4 are substituted for a set of synthetic data generated by the model itself. In each case model output generated using the default parameters given in table 3.3 was sampled at those times and depths where real data exist, thus generating synthetic data equivalent to the real data described in section 3.4. As with the data entering the cost function, the synthetic data were processed to reflect the sampling schemes employed for the real observations (i.e. integration of zooplankton data across the surface layers, etc.). Within such a framework the quality and quantity of information contained in the parameter estimates can be easily assessed because the "true" parameter values are known with certainty. From this starting point a method of quantifying the uncertainty in each parameter is developed and evaluated.

As stated in the introduction chapter, an important goal of the marine biogeochemical modelling community is prediction of the future magnitude of biologically driven carbon export to the deep ocean (Le Quéré et al., 2005). In light of this it would be helpful to be able to accurately model the rate of vertical particle fluxes in the ocean, especially given that field estimates have often yielded somewhat unreliable and inconsistent results (Buesseler et al., 2007). This uncertainty can make it difficult to incorporate particle trap data into the optimisation process (even though they may provide a very useful constraint on model performance, e.g. Friedrichs et al., 2006, 2007) and such data were not used in the cost function described in chapter 3. Given this omission, it would be useful to know how well particle flux is constrained by data that are used (DIN, chlorophyll *a*, zooplankton, PON and primary production). In order to address this question, inverse estimates of particle flux are compared to the precise known values from the twin experiment. In the remainder of this chapter, the term "export" is assumed to be synonymous with "particle flux at 300 m", despite the fact that removal of organic material from the surface 300 m is not guaranteed to be anything more than a transient process (e.g. Yool et al., 2007).

4.2 Numerical experiments

For the sake of simplicity, this chapter focuses only on the application of model one at BATS, as the aim is not to examine model performance, but rather to look at different types of solution to the inverse problem with regard to how well issues of underdetermination and observational error are handled. A twin experiment was set up using model one (section 3.5) and the default parameters given in table 3.3. The model output was sampled at those times and depths where real data exist (section 3.4), thus generating equivalent synthetic data. These were processed to reflect the sampling schemes employed for the real observations and are equivalent to f in equation 3.2. When the model is run with the true parameters the resultant misfit cost is zero. Subsequently, m easurement errors were approximated by simply assuming the error distributions given in table 3.2 and adding appropriately scaled Gaussian noise. In sections 4.3.1 and 4.3.2, the searchable parameter space is not as described in table 3.3, but is instead bounded in each dimension from zero to twice the true parameter value. The exception is the zooplankton assimilation efficiency β, which as a ratio is inherently restricted between 0 and 1.

4.3 Results and Discussion

4.3.1 Sensitivity analysis for individual parameters

Before undertaking the full optimisation procedure, the cost function sensitivity to individual parameters was analysed. Starting with the set of true parameter values (table 3.3), the misfit was evaluated upon perturbations of each individual parameter. The results of this process are shown in figure 4.1 (blue lines). While the misfit cost function is sensitive to changes in some parameters (e.g. growth rate parameter (μ_m), initial slope of the PI curve (α) and remineralisation rate (γ_m)), it can also be seen that other parameters (e.g. half-saturation constant for N uptake (k_N), grazing rate (g) and zooplankton loss rate (ϕ_m^Z)) have a much smaller effect on the misfit value across their given range.

Also shown in figure 4.1 is the sensitivity of the particle flux at 300 m (green lines). All perturbations that cause large errors in this flux (e.g. μ_m, α and γ_m) are also associated with large increases in the misfit cost. Conversely, those parameters that can be perturbed without strongly affecting the misfit cost (e.g. k_N, g and ϕ_m^Z) also seem to be fairly unimportant with regard determining export. These results suggest that the cost function (which utilises the available DIN, chlorophyll a, zooplankton, PON and

primary production observations) may be a useful tool for constraining particle flux. It is also apparent that while some parameter settings are capable of reducing export to zero (either by reducing photosynthesis or sinking to zero), no individual perturbations of the parameters are capable of raising the export to more than $\sim 6\%$ above the "true" value of 37.0 mg C m^{-2} day^{-1}. This can be related to the idea that the biological pump can only export as much biogenic matter as can be formed using new nitrogen supplied by physical processes. By killing off certain biological pathways it is possible to shut down export, but a strict upper limit is placed on this process by the supply of nutrients from depth.

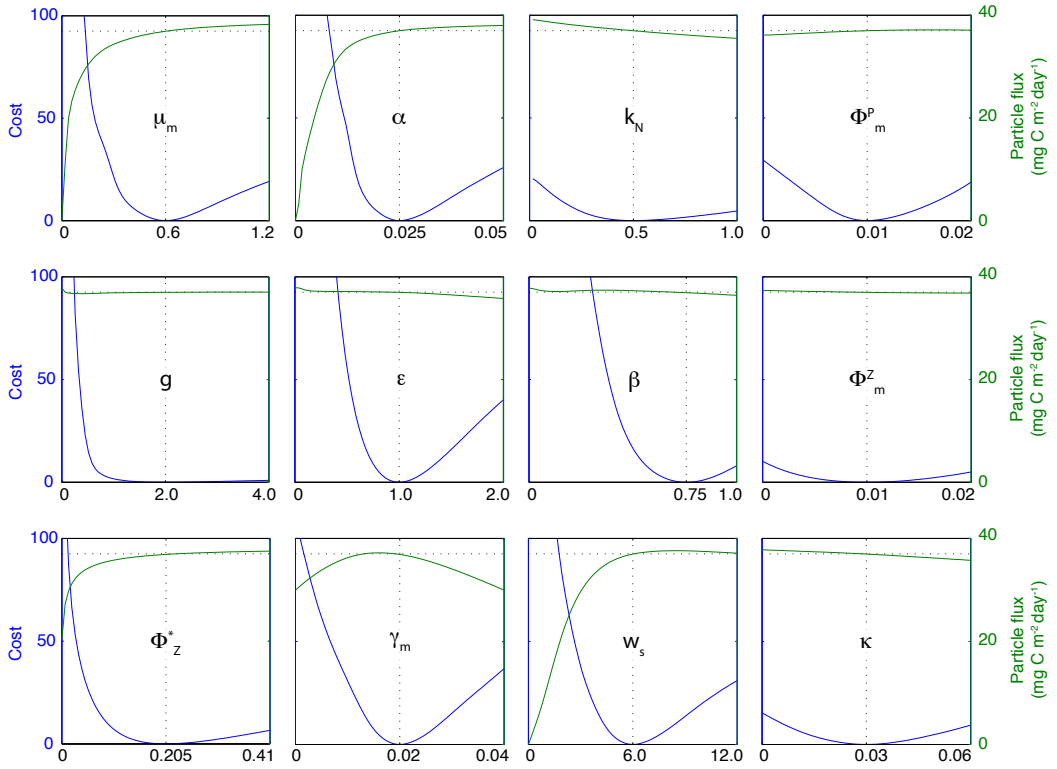

FIGURE 4.1: Misfit sensitivity to perturbations in individual parameters (blue lines). Also shown is the sensitivity of export production to the same perturbations (green lines). The subplot labels correspond to the abbreviations given in table 3.3. The true value for each parameter is shown by a vertical dotted line, the true level of export is shown by a similar horizontal line.

4.3.2 Parameter correlations

Parameter correlations are shown in figure 4.2, where the 66 possible pairings of the 12 free parameters are mapped in terms of the cost function (upper left) and the particle flux at 300 m (lower right). This was achieved by evaluating each pairing in a 31 by 31 grid across a planar search space, with all other parameters fixed at their default

values. The crosses show the location of the true parameters values (cost = 0, export = 37.0 mg C m^{-2} day^{-1}), while the contours outline the region of the parameter space where the cost is less than 1.1 (this threshold value is introduced in section 4.3.4). A number of parameters show signs of covariance with regard to this region of low cost. For example the detrital remineralisation rate (γ_m) and the detrital sinking velocity (w_s) show a strong linear correlation in this regard, indicating that their ratio may be more important than their individual values. This is not an unexpected result, as both the parameter values describe the removal of detritus, and it is clear how for example a slow sinking rate can compensate for a slow remineralisation rate because the detritus simply spends a longer time in the water column as remineralisation takes place.

Figures 4.1 and 4.2 serve to demonstrate how correlations can affect the optimisation problem. For a number of parameter pairings, the cost sensitivity is relatively high when the parameters are varied individually. That is, the value of any one parameter cannot be perturbed far from the true value without incurring a sharp increase in cost. When however these parameters are varied together and kept in the correct relationship, each value can be moved further away from the true value, before incurring the same cost. Parameter values that are individually well constrained can become poorly constrained when optimised simultaneously with other correlated parameters (Matear, 1995; Spitz et al., 1998; Schartau and Oschlies, 2003a).

The lower-right half of figure 4.2 shows the particle flux at 300 m mapped for each pairing of parameters. The regions of low cost, described above, are superimposed as contours. While the red areas indicate regions of the parameter space associated with excessive export, and the blue regions indicate too little, the white areas represent parameter combinations where the export is correct ($\pm$ 1%). In most cases, the regions of low cost correspond to these areas, indicating that, as in figure 4.1, the cost function offers a useful constraint on export. This is not always the case though, and returning to the γ_m and w_s parameters, figure 4.2 shows that when these parameters are combined in an approximately optimal ratio, but with artificially high values, the export is too high. At these high values of γ_m and w_s the remineralisation pathway is very active, with fast sinking and fast remineralisation. Artificially high estimates of export can occur with a low misfit cost because this process is not directly constrained by the data.

Other parameter combinations, such as the half-saturation for N uptake (k_N) and the phytoplankton loss rate (Φ_m^P), as well as the grazing encounter rate (ϵ) and the zooplankton quadratic mortality (Φ_Z^*), show similar regions of low cost with erroneous particle flux. This problem, which did not appear in the sensitivity analysis for individual parameters, is likely to get worse as more parameters are optimised and the parameter relationships become more complicated. For this reason, the constraints on

export production are examined further in section 4.3.5, as 12 parameters are optimised simultaneously.

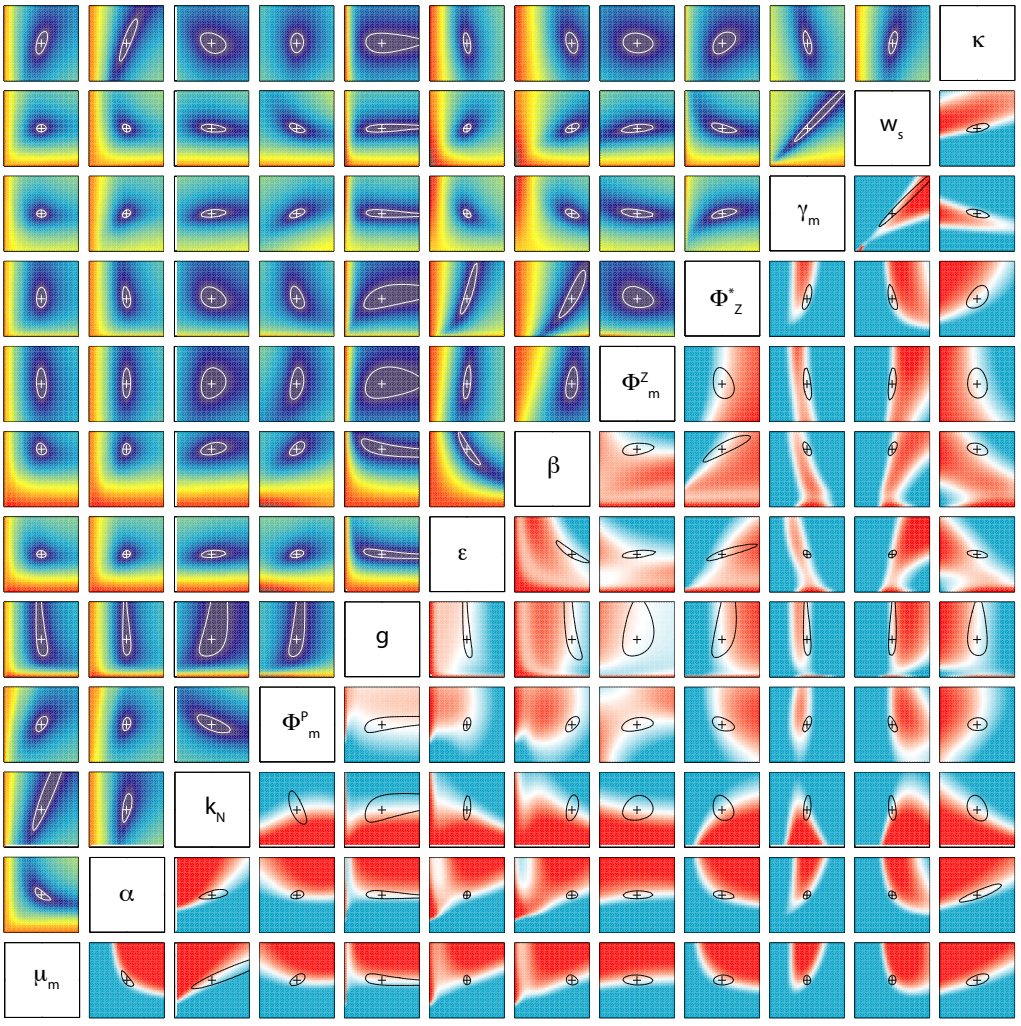

FIGURE 4.2: Misfit cost and export sensitivity to paired parameters. Each subplot represents a two-dimensional section through the parameter space, where all but the two parameters in question are set at truth. The top-left half of the figure maps the response of the cost function to variations in parameter pairs, with blue regions representing low cost. The label boxes correspond to the axes of each subplot; the key to the abbreviations used is given in table 3.3. The white crosses represent the location of the true parameters within each pairing. The black contours mark the region of the threshold cost function (section 4.3.4). Axis labels are not shown, but each corresponds to the parameter ranges described in section 4.2. The lower-right side of the figure shows the response of the particle flux at 300 metres, with the true parameters again shown with crosses and the regions of low cost superimposed as black contours. Red areas indicate excessive export, blue areas highlight regions of the parameter space where it is too low.

4.3.3 Preliminary optimisation

The μGA was initially run eight times for 5,000 generations, with a population size of 12. Each run made use of the same synthetic observations, but with a different set of Gaussian measurement noise added in each case. The output from these runs was combined so as to achieve a more thorough examination of the parameter space. This process yields many thousands of unique parameter vectors, each one associated with its own misfit value. In the same way that figure 4.1 shows the sensitivity of the cost function to perturbations of individual parameters, the μGA output can be used to estimate the sensitivity of the cost as many parameters are simultaneously adjusted.

Figure 4.3 shows each of the 12 parameters that were varied throughout the optimisation procedure, together with the associated response in the misfit value. Each subplot corresponds to an individual parameter, and a single parameter vector can thus be represented by the combined x-coordinates of a point duplicated within each subplot. The common y-coordinate of those points corresponds to the associated misfit value. In this way, the optimum parameter values found by the search are represented by the abscissae of the lowest points within each subplot. Figure 4.3 represents a composite of the eight optimisation runs and the eight red dots represent the optimal solutions yielded in each case.

For each discrete value of a parameter, highly variable values of the misfit cost may be returned, each dependent on the values of the 11 other free parameters. Only the minimum cost associated with each discrete parameter value is reproduced in figure 4.3, and so the plots describe the sensitivity of the cost to changes in each parameter, as the remaining free parameters are adjusted to optimal values. This sensitivity is defined by how rapidly the minimum misfit increases as each parameter is perturbed from its optimum value, and accounts for correlations between all the tuned parameters. For a well-posed problem in the twin experiment, with no correlated parameters, the optimum solution will be well defined by a set of smooth and symmetric parabolas centred on the true parameters (Schartau and Oschlies, 2003a).

The results shown in figures 4.1, 4.2 and 4.3 indicate that this is not a well posed problem. Even when considering the misfit sensitivity to perturbations in individual parameters, where correlations are of no concern, the shape of the cost function is flat across large areas of the parameter space.

The quality of the solutions yielded by the μGA may be assessed in part by asking, are they robust? or will different solutions arise with different realisations of the random observational errors? The red dots in figure 4.3 represent the best point estimates from each of the 8 runs of the μGA. Using the same synthetic data, but with different synthetic

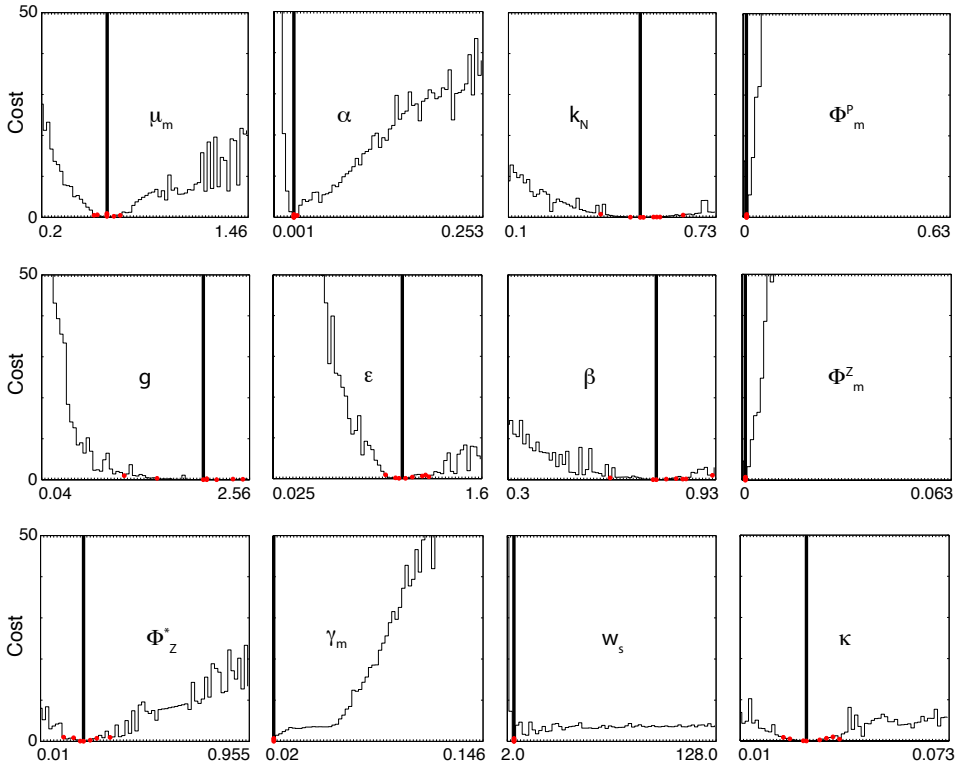

FIGURE 4.3: Output from 8 initial optimisations with the μGA, each based on an independent set of synthetic Gaussian measurement noise. Any 12-dimensional parameter vector can be represented by the abscissa of one point within each subplot. The shared ordinate of those points corresponds to the associated misfit value. Output from all runs was combined, with only the minimum misfit achieved for each discrete parameter value shown in the plot. The red dots represent the estimates of the global minimum from each of the 8 optimisations. See table 3.3 for symbols and units.

noise added, led to different estimates of the optimum parameters. This was because the large flat regions in the parameter space visualised in figures 4.1 and 4.2 allowed small changes in the observations to cause large differences in the optimum parameter estimates. In additional analysis, presented in chapter 2, it was found that simply using a different random seed to generate the first population in the search led to potentially large changes in the solution. Any confidence in a robust point estimate of the global minimum would require convergence on the same parameters and misfit value in at least some cases. It is clear is that point solutions to such a high-dimensional and ill-posed problem can never be considered robust. An approach to gaining robust solutions by fixing all poorly constrained parameters to their prior values was examined in chapter 2. It was found that the sensitivity of the misfit cost to these prior values made it is unacceptable to just ignore this source of uncertainty. In the remainder of this chapter a method of estimating parameter uncertainty is described and its efficacy and utility are examined using twin experiments. This method will be applied in subsequent chapters whenever it becomes necessary to quantify parameter uncertainty or its effects.

4.3.4 Quantifying uncertainty

Given estimates of the observational errors, how much of this uncertainty is carried through to the parameter estimates yielded by the μGA? To answer this question it is necessary to try and map the uncertainty in the observations through the μGA and into the solutions. The simplest and most reliable way to do this is to run the optimisation many times, each time with a different set of measurement errors, as in figure 4.3. This method was first outlined by Schartau and Oschlies (2003a) and subsequently Schartau et al. (2007), where new observation datasets were generated by adding noise to the real observations. The problem with this approach is that each optimisation routine takes many hours ($t \lesssim 24$ in this case) to complete. Schartau and Oschlies (2003a) carried out 5 optimisations in this way, and Schartau et al. (2007) managed 10, but it would take many hundreds or perhaps thousands to build up a statistically robust distribution.

Instead of mapping the effect of observational uncertainty through the genetic algorithm, which is too computationally expensive, a more efficient approach was achieved by taking advantage of the parameter sensitivity information hinted at in figure 4.3. Using a similar sensitivity plot Schartau and Oschlies (2003a) noted qualitatively that some parameters could be moved a long way from their "optimal" value without incurring a large increase in cost. This estimate of parameter uncertainty is quantified here by assessing the range over which each parameter can be varied without incurring a significant cost increase in relation to the true parameters.

Given that the true parameters (in the absence of measurement noise) result in a misfit cost of $J(\mathbf{m_t}) = 0$, how much of an increase is considered to be significant? This question can be answered by evaluating the true model output with respect to very many noisy observation sets. To do this the model only needs be run once, and the output can then be compared to very many sets of noisy observations. In this way the distribution of $J(\mathbf{m_t})$ can be calculated in only a few minutes. Figure 4.4 shows a histogram of the misfit distribution after 10,000 comparisons between the true model output and noisy synthetic observations. The cumulative distribution of the misfit is also shown. The lower panel is included as a demonstration of the robustness of the estimate, with the mean and standard deviation of the distribution both effectively constant after approximately 2,000 of the 10,000 evaluations. The distribution of the misfit due to measurement noise is always positive because of the use of a least squares cost function. Every individual model-data misfit is squared, and so contributes a positive value to the overall misfit term regardless of its sign. As the distribution is normal, approximately 97.7% of the cost function evaluations result in a cost that is less than or equal to the mean plus two standard deviations. For the true parameters this threshold is 1.165.

Any parameter vector that yields a lower cost than this can be said to be statistically indistinguishable from the true parameters, so counts as a "good" solution.

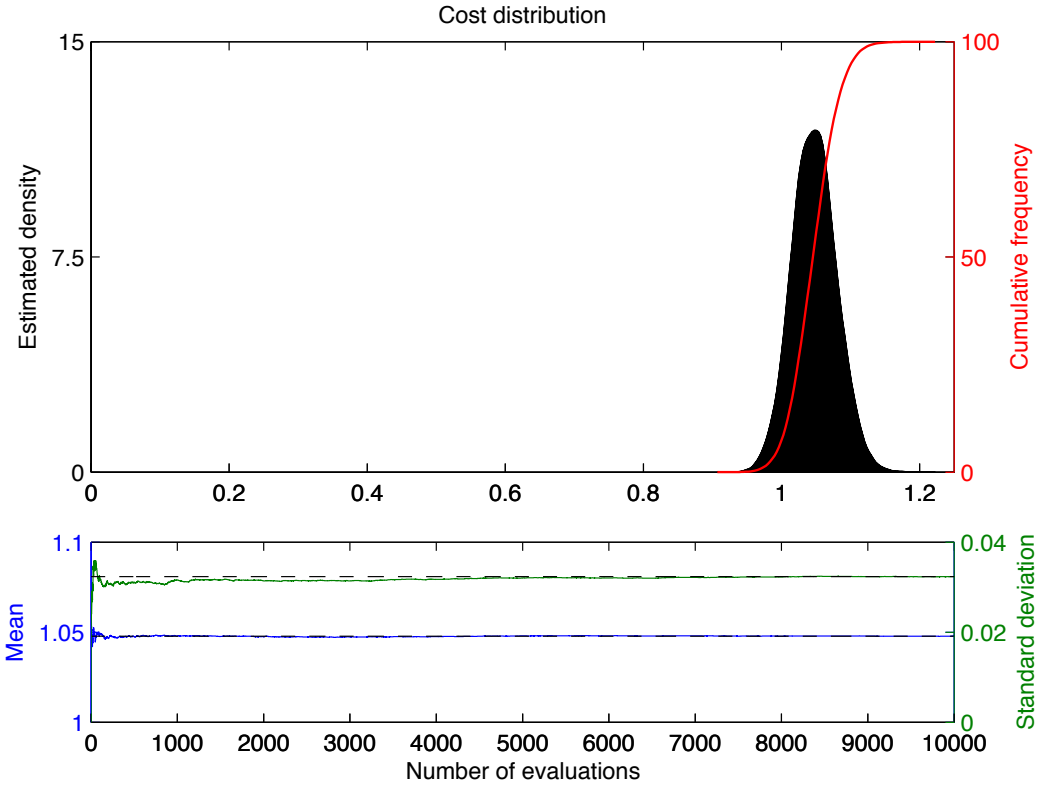

FIGURE 4.4: Misfit distribution attributable to Gaussian measurement noise. The lower panel shows the evolution of the mean and standard deviation throughout the 10,000 evaluations. These are stable from about 2,000 evaluations onwards, indicating a robust solution.

For a 12-dimensional problem, such as is posed here, there are far too many points in the parameter space to evaluate every one (where each parameter is resolved by 64 discrete increments there are 64^{12} possible combinations). Instead, it is possible to randomly sample the parameter space using Monte Carlo sampling. If enough randomly drawn points are evaluated in the forward model, the associated weights can be used to map out the solution region where the cost is lower than the threshold. A space-filling Latin hypercube sample was used to describe 5,000 points across the range of the parameter search space. These were evaluated in the forward model. Not one evaluation yielded a misfit less than the upper 95^{th} percentile of the true parameter misfit distribution. In fact the lowest misfit found by the Latin hypercube was 137.3; more than 4,000 standard deviations higher than the threshold. The region of the global minimum is clearly too small, and the prior parameter search space too large, for the efficient use of such a crude Monte Carlo technique. It would require the evaluation of far too many points in the search space to build up a statistically robust sample of good solutions.

Instead of blindly searching through such a huge parameter space it is possible to focus the search around the region of interest (i.e. the global minimum) by using output from the μGA. Although the search is not random, and thus may not be be used to evaluate the probability density function (pdf) for the parameter solutions, the scope of the solution region may be easily assessed. Using this technique the output from the eight runs of the μGA described in section 4.3.3 were used to generate the parameter error estimates shown by red dots in figure 4.5. Outside these ranges parameter values were excluded from the solution on the grounds that they were not found to yield credible output in the forward model.

These marginal solutions describe the range of each parameter where a low cost may be achieved. By assuming that the parameter error estimates independently define the solution region, any parameter correlations are ignored. The solution is thus defined as a solid hypercube whereas the true solution may represent a much more convoluted region within that space (see figure 1.1 in chapter 1). The significance of this is examined in the following section.

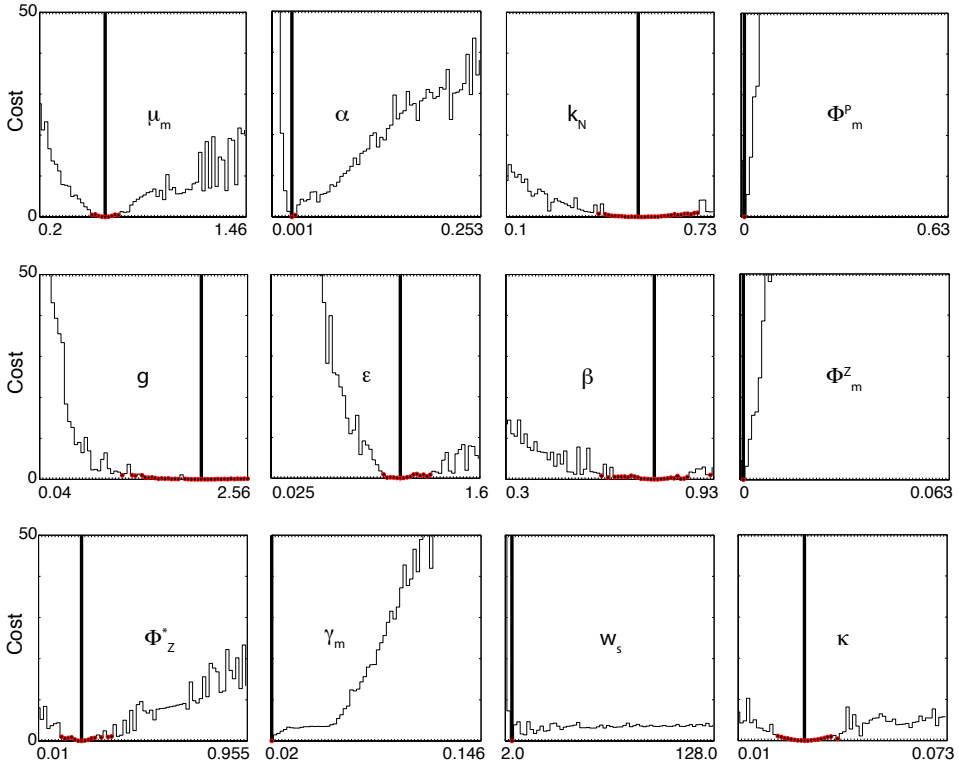

FIGURE 4.5: Solution to the preliminary optimisations given as error bounds. Red points indicate parameter values where the misfit cost was not significantly different from the cost associated with the true parameters. They were retained as potential solutions.

4.3.5 Evaluating the marginal solutions

The marginal solutions described in the previous section define the limits of a solution region. This region should represent a more informative solution to the inverse problem as it describes exactly what the data can say about the range of each parameter, rather than just providing an arbitrarily precise point solution. To investigate the degree to which these marginal estimates can constrain the forward model output the solution region was resampled with a Latin hypercube. The forward model was then evaluated at those points in terms of misfit and export.

Figure 4.6 shows the prior (upper row) and posterior (lower row) distributions for misfit cost and export. The prior distribution for the misfit cost was widely distributed (it is greatly truncated in figure 4.6, the mean is 351.9 and the maximum value is 3044.6). The 98$^{\text{th}}$ percentile of the cost for the true parameters (1.165) is shown in red. The export distribution is also poorly constrained by the prior parameter space and a large number of evaluations resulted in zero export (mostly model crashes). The maximum export flux seen in any of the model runs was 24% larger than the true estimate of 37.0 mg C m^{-2} day^{-1} (shown in red).

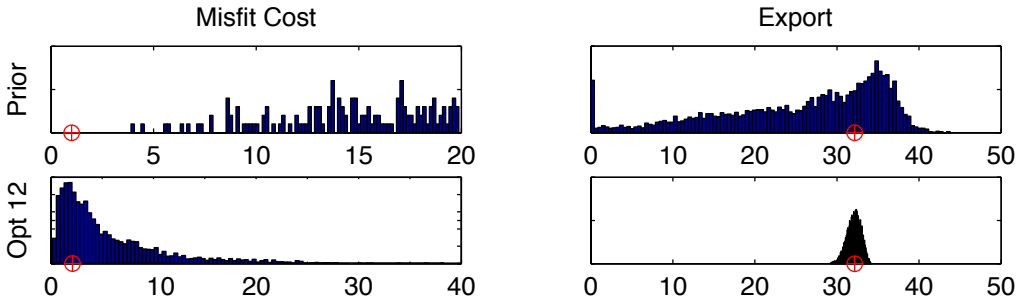

FIGURE 4.6: Probability distributions for misfit cost (left column) and export (right column). The upper row shows the distributions when parameters were drawn from the prior parameter space. The lower row represents the same distributions after optimisation of 12 free parameters.

The second row panels indicate the distributions given by the marginal solutions after the optimisation of 12 parameters. Variability in both misfit and export is greatly reduced, with the median value for export close to the value derived from the true parameters, $\mathbf{m_t}$. It is encouraging that the export production is much more narrowly defined after the optimisation of 12 parameters, despite the fact that many of the parameters are so poorly constrained. This result can be related back to figures 4.1 and 4.2, where it is shown that those parameters which have little effect on the cost function also do not have a large effect on export. In some cases in figure 4.2 it was shown that export could be erroneous even in areas of low cost, but it appears that even with 12 free

parameters this effect is not strong enough to cause very large errors in the estimated export. It is important to note however that these result can by no means be seen as proof that the inverse approach will work as well when real data are assimilated. In the twin experiment the modelled system was perfectly consistent with the synthetic data in terms of ecosystem structure and the underlying physics, and this will not be the case with real data.

Despite the fact that the parameter solution region was defined by a set of parameter vectors all with misfit less than 1.1, the resampled cost distribution is heavily skewed to the right. This is because, as mentioned previously, parameter correlations were not accounted for by the marginal solutions and resampling them will often yield non-optimal parameter combinations. Figure 4.7 shows the cost distribution for two pairs of parameters drawn from figure 4.2. One is highly correlated in terms of cost ($r^2 = 0.97$) while the other is much less so ($r^2 = 0.06$). In both cases the marginal solutions are shown by a red box around the true solution, shown in white. Even in panel *b*, where the correlation coefficient suggests the parameters are uncorrelated, the marginal solutions define a region considerably larger than the true solution. If the true solution in this case was a circle, then the marginal solution would be a square around that was 27% larger. If this relationship is scaled up for all the free parameters, a 12-dimensional hypercube is over 3,000 times larger than the hypersphere it surrounds. This ratio is made considerably worse when relations such as those shown in panel *a* are considered.

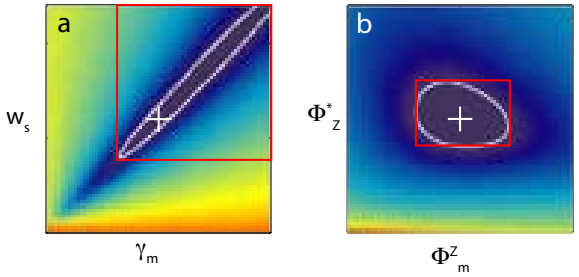

FIGURE 4.7: The marginal error estimates (red boxes) are defined by the range of the true solutions (white contours) across each parameter, but even where correlations are not strong, the marginal solutions cover more of the parameter space than is included in the true solutions.

For these reasons, the marginal parameter estimates are considered inadequate to define the solution to this inverse problem. The issue could be improved by fixing the most unconstrained and correlated parameters to their prior values (e.g. Friedrichs et al., 2006, 2007), but the analysis presented in chapter 2 revealed that if the prior values are not well known, then fixing to precise values falsely reduces the uncertainty of the problem

and provides an additional source of error. An alternative approach was adopted here, where all the good solution vectors yielded by the μGA (i.e. those where $J(\mathbf{m})$ was less than the threshold cost) were used to define the general solution. As the accepted parameter vectors all lie within the solution region (they are solutions themselves) they take full account of any correlations among the parameters. The distribution of these parameter vectors also provides useful information regarding any uncertainty in the parameters and the associated model output.

4.4 Conclusions

Results from the twin experiment suggest that although the inverse question was ill-posed, some useful information about the parameters could be gained from the synthetic observations. The misfit cost function penalises model-data discrepancies that are important in terms of export, and although small observational uncertainties were sometimes mapped through to large uncertainties in the parameter estimates, there was enough information in these solutions to provide a useful constraint on export.

A set of marginal solutions were used to define each parameter according to a threshold cost function value. The solution range for each parameter could then be randomly sampled to generate unbiased uncertainty estimates for the model output (Schartau and Oschlies, 2003a; Schartau et al., 2007). This approach however does not take any parameter correlation information into account, and the analysis presented above revealed that treating each parameter independently led to many non-optimal parameter combinations entering the solution. If these are subsequently used for state and flux estimation, the estimated uncertainty will almost certainly be to high. Likewise if such solutions are used to estimate a predictive cost function, as in Friedrichs et al. (2006, 2007), then the predictive costs will overestimate any uncertainty.

In order to overcome this problem it was necessary to preserve parameter correlation information in the solution, and this was achieved by defining the solution in terms of the optimal parameter vectors that were found by the μGA. By treating each solution vector as indivisible, when they were applied in the forward model a much tighter constraint could be maintained on any output.

Unfortunately this improvement in the handling of correlated parameters comes at a cost, in that the conditional solutions can no longer be sampled in the same unbiased way as the marginal solutions. Despite the fact that it is categorised as a stochastic technique in chapter 1, the μGA does not search the parameter space in an unbiased way. Instead it tends to focus around its best estimate of the optimal parameter vector. For

this reason the solutions developed in this chapter should be treated with caution. The conditional parameter estimates do not represent a comprehensive and unbiased solution to the inverse problem, but rather provide an estimate of the range of parameters and associated model outputs that allow a good fit to the data. A better approximation might have been achieved if an unbiased search method, such as a Metropolis-Hastings Markov chain method had been applied to search the parameter space (Andrieu et al., 2003). Although such an approach would represent an improvement on the μGA method outlined above, it was found in to be prohibitively difficult to apply in practice, and will not be considered further in this study.

Analysis of twin experiments is revealing, because the process provides a "ground truth" solution by which the efficacy of the inverse problem can be estimated. It is important to note however that this is only an estimate. The twin experiment initially represents a perfect model of the system. By using the true parameters and error free observations the model output exactly matches the observations and there is zero misfit. From this starting point it is relatively easy to define the threshold cost function because the true solution is known. However, when real data are applied in the inverse problem, the idea of a set of "true" parameters is no longer applicable. In subsequent chapters where real data are used this issue is addressed by comparing candidate parameter vectors with the best fit parameter vector ($\mathbf{m_{opt}}$) rather than with the true parameter vector ($\mathbf{m_t}$). In such cases, the threshold cost function is defined by repeatedly evaluating the optimal parameters against the real data, but with added synthetic noise. The threshold is taken as the optimum value, plus 2 standard deviations of the resampled costs.

Chapter 5

Calibrated model solutions

5.1 Introduction

Scientific models are often developed as a mathematical approximation of some more complex system or relationship. If such a model is assumed to represent a good approximation of reality, it can be used to analyse and generate predictions about the system in question. In the field of marine biogeochemistry, while models are developed on the basis of ecological theory (i.e. primary production by autotrophs, consumption by heterotrophs, and many other more sophisticated concepts), the high diversity of the real system means that many components must be aggregated and described by some empirical parameterisations. The behaviour of any such semi-empirical model is dependent on the values of its parameters, and thus if a model is to represent a good approximation of reality, it is necessary to find the most appropriate values for these parameters.

If a model can be said to represent a theory, then parameter optimisation is the search for a version of that theory that is closest to reality (or the data, at least). Through this process it is possible to investigate the implications, as well as any shortcomings, of the model or theory. Such an approach may be either diagnostic or prognostic. Examples of the former occur as models are used to analyse the behaviour of a particular system, filling in gaps in our knowledge about the present state of the marine ecosystem. This approach was taken by Spitz et al. (2001) and Schartau and Oschlies (2003b), who calibrated one-dimensional models to time-series data at various North Atlantic sites so that they could make estimates of unobserved fluxes and standing stocks of biomass. The prognostic approach, on the other hand, involves the use of models to make predictions, possibly about about some future climate scenario (e.g. Bopp et al., 2005), or the consequences of ocean iron fertilisation (e.g. Gnanadesikan et al., 2003; Dutkiewicz et al., 2006), for example.

While parameter optimisation techniques allow the objective assignation of model parameter values, there is a danger that calibration can go too far. If a model structure is unrealistic or inappropriate, any errors can potentially be corrected as parameter values are adjusted. This is especially true in the case of more complex models, which contain many more unconstrained parameters and hence more degrees of freedom. The process of parameter optimisation will not always solve the problems of a model, and may often hide them somewhere less visible instead. For this reason it is important to examine calibrated models, and their parameters, very closely. Solutions may also be validated against independent data to check their credibility. In this chapter the focus is on the calibration process, as the five ecosystem models are tuned to data from BATS and NABE (solutions will be assessed with regard to independent data in chapter 6). The optimised models will be assessed with regard to what they do well, and perhaps more

importantly, what they cannot do so well. By analysing the shortcomings of the calibrated models it will be possible to make suggestions about where improvements should be made.

The models that will be analysed in this chapter are all relatively simple, and omit many biogeochemical processes that are important in the real ecosystem. For this reason there will inevitably be shortcomings in their ability to fit the data. To increase this ability, the obvious solution is to include more (realistic) model complexity. If more specific dynamic processes are resolved this will allow the models to fit a wider range of data (Friedrichs et al., 2007), but any improvement in the dynamic range of models with increasing complexity must also be weighed against the cost of the greater uncertainty associated with extra parameters (Anderson, 2005; Raick et al., 2006). Throughout the calibration experiments that are described in this chapter, the amount of uncertainty in the optimal model solutions will be tracked using the technique outlined in chapter 4. Instead of examining just one optimal solution, all of the optimal solutions are included in the analysis. In this way, it will be possible to see not only how well the models fit the data, and where they are performing badly, but also where the models are particularly uncertain. By finding out where the models are inadequate, the results will highlight where additional model complexity might be needed. By also looking at where the models are particularly poorly constrained, the results will show where it will be necessary to collect more data before additional model complexity can be seriously considered.

5.2 Methods

The models, data and optimisation routine are described in chapter 3. In the experiments described in this chapter, models one to five were set up with the OCCAM physical forcing at BATS and NABE. The models were optimised to real data at each site, adjusting the free parameters listed in table 3.3. Initially the parameters for the variable Chl:N ratio used in models three to five were allowed to vary, but this was found to yield very unrealistic values for the Chl:N ratio. In all experiments presented here the parameters of this function were set to the default values given by Cloern et al. (1995). This decision is justified on the grounds of these unrealistic results, and the fact that the Cloern et al. (1995) function was empirically derived from a much more appropriate dataset than the one applied here (which includes no data for phytoplankton biomass).

Model solutions were defined as all those parameter vectors evaluated by the μGA which were not significantly worse than the optimum solution, or in other words, had a misfit cost less than or equal to the threshold cost (i.e. the minimum cost plus two standard deviations of the distribution produced by adding synthetic noise, as defined

in section 4.3.4). By evaluating each of these parameter solutions in the forward model, it was possible to find not only the state variables and fluxes that allowed an optimal fit to the data, but also the uncertainty associated with those properties.

Like all optimisation algorithms, the μGA cannot offer guaranteed convergence on the global minimum within a reasonable timescale. In order to increase confidence that the solutions returned in this chapter could not be improved upon, the optimisations were repeated until no further improvements were seen in terms of cost. Initially the μGA was run 8 times, each time using a different random seed to generate the first population. After these had run for 5,000 generations, the single lowest cost parameter vector from all eight runs was recorded and included in the first generation of four new runs of the μGA. Again, after 5,000 generations the optimisations were stopped. If the new runs had led to any significant improvement in the lowest misfit cost, the optimal parameter vector was used to initialise another four runs of the μGA. If no further improvements were seen, no further iterations were performed. This process is summarised in a flow chart in figure 5.1.

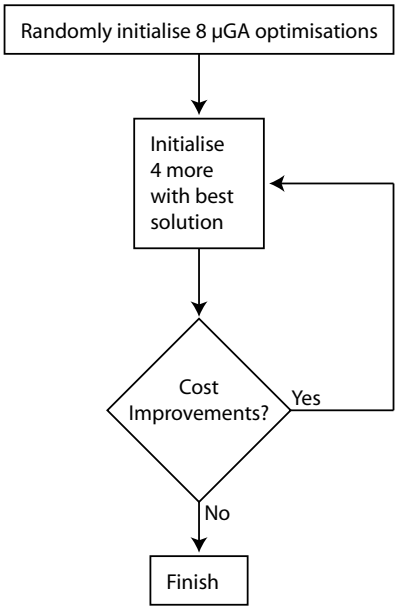

FIGURE 5.1: Flow chart showing iterations of the μGA.

5.3 Results

5.3.1 Calibrated model performance

Adjustment of the prior parameters using the μGA led to significant decreases in the cost function values for all five models at both sites. At BATS, the costs were reduced by between 46 and 56%, while at NABE the reductions ranged from 45 to 50%. Details of the individual prior and calibrated costs are given in table 5.1. These costs are broken down further in table 5.2, where the misfit contributions from each observation type are listed.

The introduction of additional complexity led to an improvement in the minimum calibrated costs at BATS, with models two to five achieving costs 9%, 9%, 21% and 31% lower than model one, respectively. At NABE this trend was less apparent, with model two showing an improvement of only 1%, while model three was actually 1% worse than model one (mainly as a consequence of increased chlorophyll *a* misfits, see table 5.2 and section 5.3.2). Model four showed a 1% improvement relative to model one, while model five was only 2% better in terms of cost.

TABLE 5.1: Misfit costs before and after optimisation at BATS and NABE. Optimisations used real observations and physical forcing derived from OCCAM. The threshold cost and the number of "acceptable" solutions are also shown.

BATS	Default	Optimised	Threshold	n solutions
Model 1	192.1	84.5	85.3	5532
Model 2	176.5	77.3	78.0	809
Model 3	142.2	76.9	77.7	2125
Model 4	137.6	66.6	67.3	994
Model 5	130.4	58.6	59.2	978

NABE	Default	Optimised	Threshold	n solutions
Model 1	833.5	425.8	429.2	2332
Model 2	760.0	421.6	424.9	820
Model 3	826.1	430.3	433.6	928
Model 4	832.7	420.4	423.7	914
Model 5	813.0	417.9	421.2	5127

TABLE 5.2: Misfit costs at BATS and NABE, broken down according to data type.

BATS	DIN	Chl a	Zoo	PON	CUp	Total
Model 1	6.6	29.5	4.2	5.9	38.3	84.5
Model 2	6.1	33.6	2.2	7.9	27.5	77.3
Model 3	6.5	28.6	4.7	9.2	28.0	76.9
Model 4	5.9	29.0	3.6	7.4	20.1	66.6
Model 5	6.0	23.1	0.3	5.1	24.0	58.6
Mean	6.2	28.8	3.0	7.1	27.7	72.9
Range	0.7	10.5	4.4	4.1	18.2	25.3

NABE	DIN	Chl a	Zoo	PON	CUp	Total
Model 1	27.3	304.5	-	63.6	30.5	425.8
Model 2	26.6	306.4	-	61.5	27.0	421.6
Model 3	29.7	317.9	-	64.7	17.9	430.3
Model 4	25.6	313.6	-	62.0	19.1	420.4
Model 5	25.2	313.3	-	61.9	17.5	417.9
Mean	26.9	311.1	-	62.7	22.4	423.3
Range	4.5	13.4	-	3.2	13.0	7.9

5.3.2 Model output and observations

Dissolved inorganic nitrogen (DIN)

Modelled mean DIN concentrations within the mixed layer are shown for both sites in figure 5.2. Each subplot shows the model output associated with the lowest misfit cost, together with an uncertainty estimate describing all solutions not significantly worse than that optimum.

Modelled DIN concentrations fit the data relatively well in comparison to other data types at both BATS and NABE, contributing on average 8.1% and 6.4% of the total costs. At BATS the models correctly reproduce the summer DIN concentrations of zero, but underestimate the peak winter concentrations. This discrepancy was improved by the introduction of the DON pathway in model five, but overall there were no significant differences in terms of DIN misfit cost between the five models.

There was also very little cost difference between the models at NABE, and the trajectories appear to be very similar. All five models underestimated the peak DIN concentrations seen in the mixed layer at the start of the observation period by up to 2 mmol N m^{-3}. Models one and two showed the greatest uncertainty in the rate at which DIN became depleted in the mixed layer, as indicated by the thickness of the shaded area

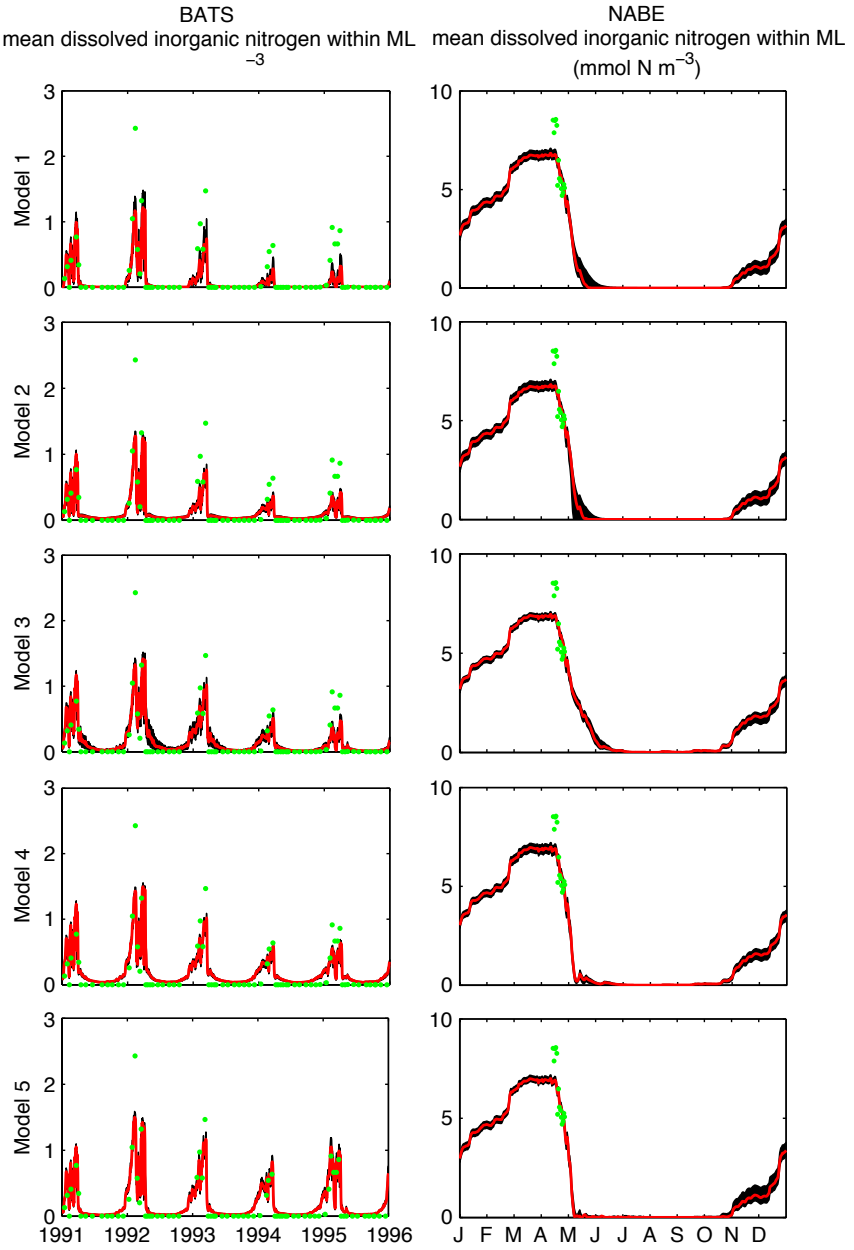

FIGURE 5.2: Dissolved inorganic nitrogen (DIN) concentrations at BATS and NABE, averaged over the mixed layer. Red lines represent the trajectory of the single optimal model solution, while the filled black areas cover the range of solutions not significantly different from that minimum (i.e. all the acceptable solutions, see text for details). Green dots represent observations, interpolated onto a vertical grid and similarly averaged over the modelled mixed layer. Note different scales on vertical axes between stations.

during May. The largest uncertainties in modelled DIN were found during the autumn period, when nutrients were entrained into the deepening mixed layer.

Figures 5.3 and 5.4 show profiles for observed and modelled DIN at BATS and NABE respectively, with the model-data residuals and uncertainty estimates for the optimal solutions also shown. Observations at BATS show that surface DIN concentrations rarely exceeded ∼1.5 mmol m^{-3} at any time during the study period, and that there was an intrusion of relatively nutrient rich water at the base of the euphotic zone towards the end of 1992.

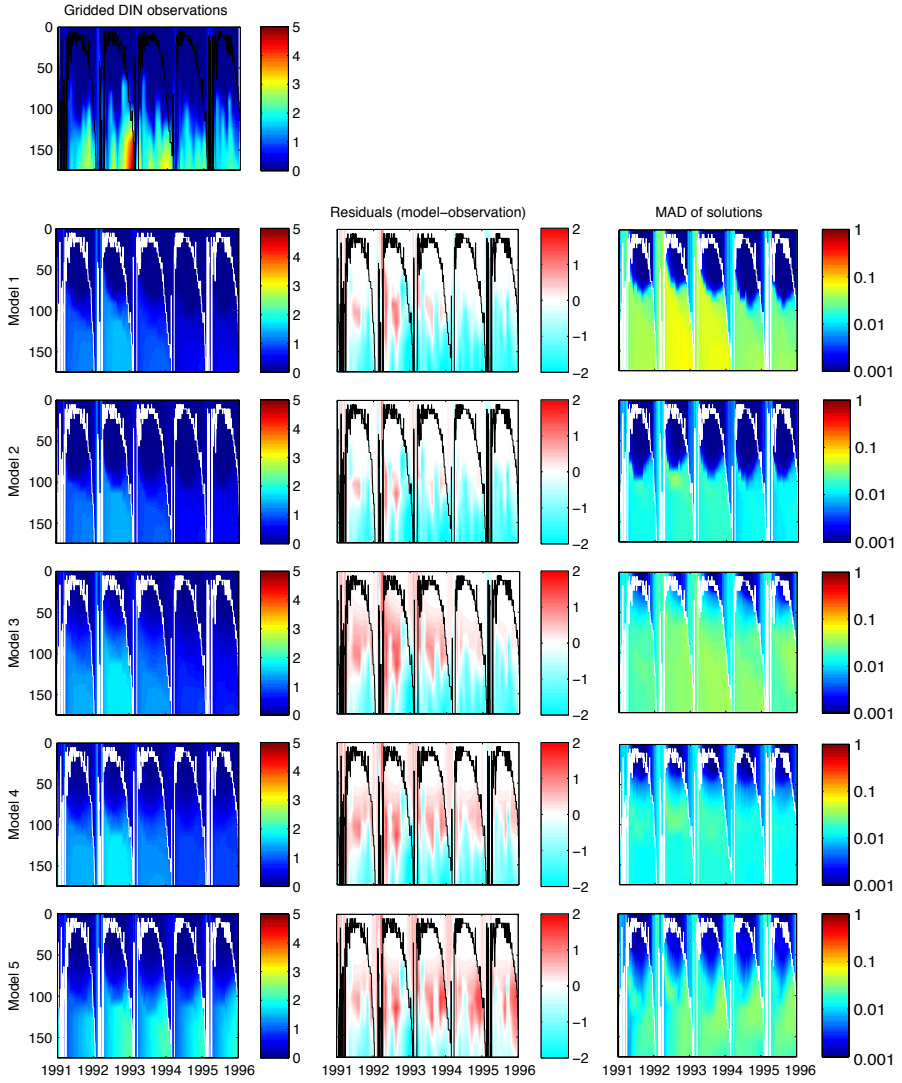

FIGURE 5.3: Dissolved inorganic nitrogen (DIN) at BATS (mmol N m^{-3}). The top panel shows observations taken between 1991 and 1995. The three columns below correspond to the optimised model output, model residuals and the median absolute deviation (MAD) of the optimal solutions, respectively. Complete rows represent output from each of models one to five. Modelled mixed-layer depths are shown by solid black or white lines.

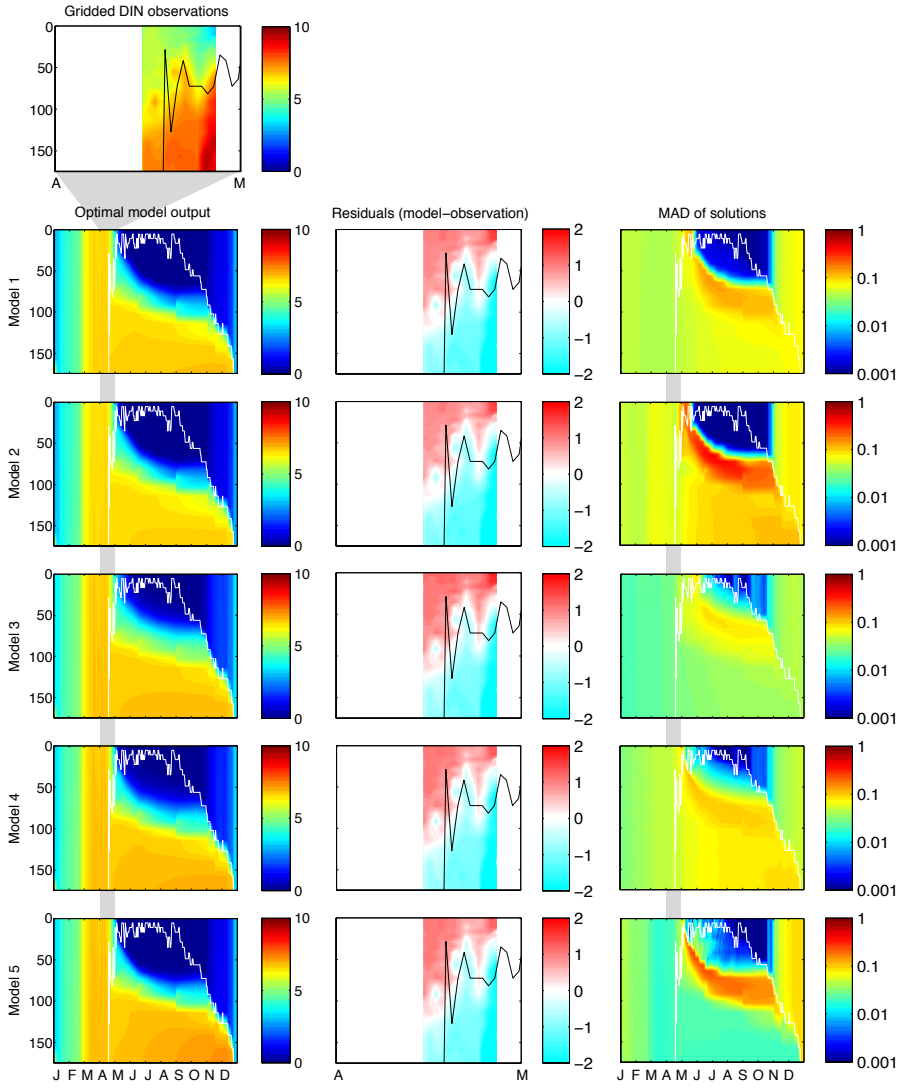

FIGURE 5.4: Dissolved inorganic nitrogen (DIN) at NABE (mmol N m^{-3}). The top panel shows observations taken during April 1989. The three columns below correspond to the optimised model output, model residuals and the median absolute deviation (MAD) of the optimal solutions, respectively. Complete rows represent output from each of models one to five. Modelled mixed-layer depths are shown by solid black or white lines.

Although all the models appear to have correctly reproduced the low nutrients in the surface waters, none picked up the high nutrient event. With the addition of a variable Chl:N ratio DIN concentrations at depth were seen to increase. The model-data residuals reveal that while modelled DIN was correctly depleted at the surface, levels were generally too high at intermediate depths and too low towards the base of the euphotic zone. The inclusion of a variable Chl:N ratio in models three to five generally increased the problem of overestimation at intermediate depths, but tended to improve the accuracy of deeper model estimates.

Uncertainty estimates for depth-resolved DIN at BATS are given in the third column of figure 5.3, and were calculated as the median absolute deviation (MAD) of the model solutions. This a robust statistic that is useful for describing the spread of non-Gaussian distributions. It is defined as the median of the absolute distance of all data from the median of those data.

All models exhibit very little uncertainty (i.e. $\sim$0.001 mmol m^{-3} $\ll$ DIN measurement uncertainty) in the nutrient depleted surface waters, while the higher variability in the nutrient rich deeper waters is still orders of magnitude smaller than observations. Surprisingly, the highest uncertainty is shown by the simplest of the five models, model one.

At NABE only observations from the month of April 1989 were included, and the x-axis is truncated to focus only on this month in some subplots of figure 5.4. DIN concentrations in the surface waters were $\sim$3-5 mmol m^{-3} during the observation period, rising sharply to $\sim$7-10 mmol m^{-3} across a shoaling nutricline. The model-data residuals reveal that all five models tend to overestimate DIN levels above the observed nutricline, while underestimating below it.

The modelled DIN profiles look very similar for all five models, with high nutrient levels throughout the euphotic zone until the shoaling of the mixed layer during April. From April through to the end of October, the surface layers were depleted of nutrients above a steadily deepening nutricline. As winter mixing began to set in from October, the euphotic zone was gradually homogenised, with DIN concentrations steadily increasing to their pre-bloom level.

The solution uncertainty estimates reveal that DIN was generally well-constrained at NABE, with the highest uncertainties seen around the nutricline in the months after the observation period. Model two shows the highest variability in this area, with the DIN uncertainty reaching up to 0.5 mmol N m^{-3}.

Chlorophyll *a*

Figure 5.5 shows modelled mean chlorophyll *a* and phytoplankton concentrations in the surface 200 m at BATS and NABE, together with observations of chlorophyll *a*. Chlorophyll *a* misfits were fairly high relative to other data types, comprising on average 37.2% of the total cost at BATS and 73.5% at NABE. Calibrated chlorophyll *a* misfits were much more variable than DIN misfits as the model structure was changed (table 5.2).

Chlorophyll *a* was tightly coupled to phytoplankton biomass in models one and two by the fixed Chl:N ratio, and the introduction of a variable Chl:N ratio in models three

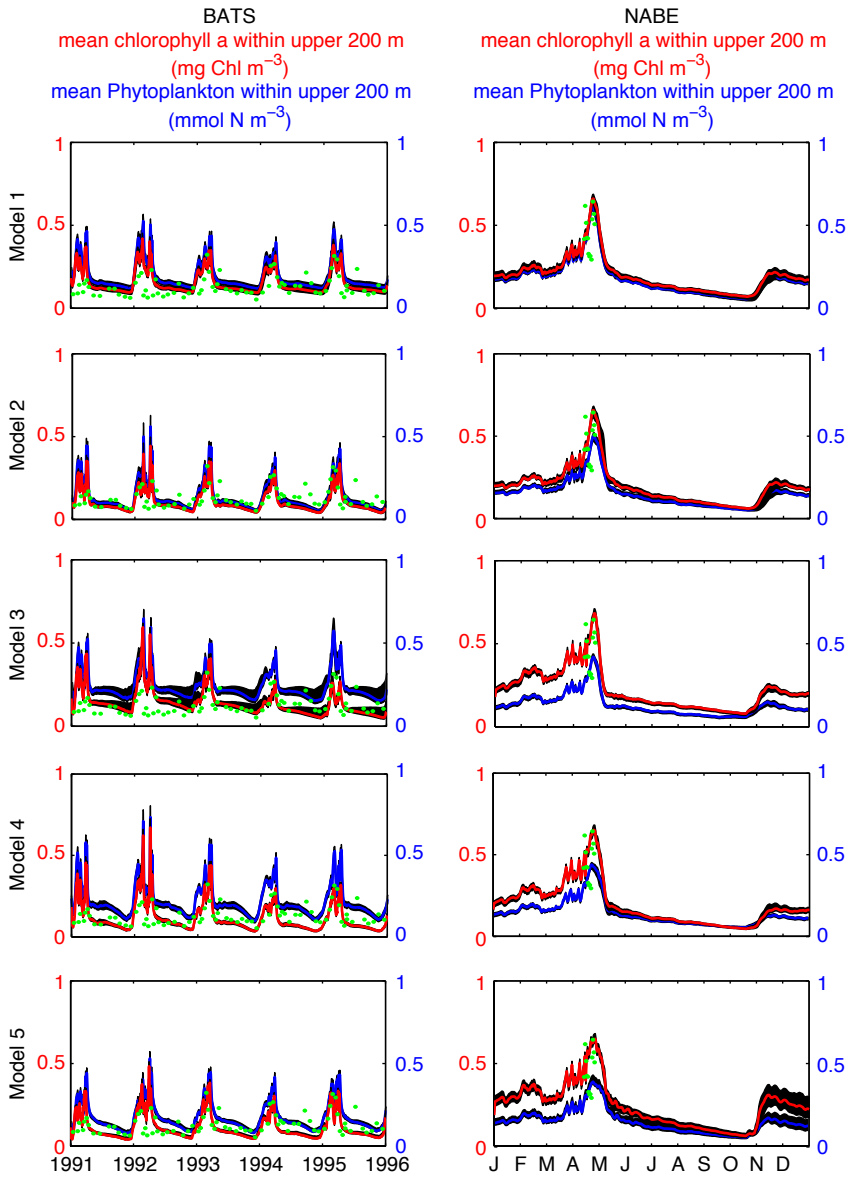

FIGURE 5.5: Chlorophyll *a* concentrations at BATS and NABE, averaged over the surface 200 m. Red lines represent the trajectory of the single optimal model solution, while the filled black areas cover the range of solutions not significantly different from that minimum (i.e. all the acceptable solutions, see text for details). Green dots represent observations, interpolated onto a vertical grid and similarly averaged over the modelled mixed layer.

to five relaxed this constraint. At BATS, although the average Chl:N ratio in the euphotic zone fell from 1.59 to between 0.97 and 1.16 g [chl a] (mol [N])$^{-1}$ in models three, four and five, the ratio increased to more than 2 g [chl a] (mol [N])$^{-1}$ below the nutricline. Phytoplankton biomass was significantly larger in these models, but chlorophyll *a* concentrations (which are directly constrained by observations) remained at a similar magnitude in all five cases. All the models overestimated the peak bloom chlorophyll *a* concentrations at BATS during 1991 and 1992. Modelled chlorophyll *a* misfits were lowest in models with a variable Chl:N ratio.

Despite the very high chlorophyll *a* misfits at NABE, modelled mean chlorophyll *a* biomass within the surface 200 m was closely comparable to observations, peaking at around 0.8 mg Chl m^{-3} for all five models. Modelled chlorophyll *a* profiles appear to be fairly consistent across all five models, although misfits were in fact made slightly larger with the addition of a variable Chl:N ratio, as the models were no longer able to capture the lowest chlorophyll a concentrations seen at the start of the bloom.

Depth-resolved profiles for modelled and observed chlorophyll *a* concentrations are shown in figures 5.6 and 5.7, together with model-data residuals and uncertainty estimates. Observations reveal that chlorophyll *a* concentrations in the euphotic zone were highly variable at BATS, but with a general pattern of higher chlorophyll *a* concentrations throughout the euphotic zone in the winter months, with biomass concentrated at intermediate depths (∼50-100 m) during the summer. The highest observed chlorophyll *a* concentrations occurred in this deep chlorophyll maximum (DCM).

Calibrated model chlorophyll *a* at BATS was quite variable between the five models, although all models produced peak chlorophyll *a* concentrations in the surface waters during the bloom period, rather than in the DCM. The plotted model-data residuals confirm that the high surface chlorophyll *a* concentrations were indeed erroneous. The models repeatedly underestimated chlorophyll *a* at intermediate depths, although this problem was ameliorated by the inclusion of a variable Chl:N ratio (models 3, 4 and 5). Figure 5.8 shows the mean vertical profiles of time-averaged chlorophyll *a* at BATS from each model, together with the observational mean. It appears that apart from a sharp peak at approximately 55 metres, all the models could correctly reproduce the observed (time-averaged) concentrations at intermediate depths (∼60-120 m). Surface concentrations were however overestimated if the variable Chl:N ratio was not included. Similarly, the fast-recycling pathway was required to simultaneously reproduce the concentrations seen below 120 m. It seems that both these model components played a role in reproducing the correct vertical structure for chlorophyll *a*. Model uncertainty, shown in the right-hand column of figure 5.6, was generally low for chlorophyll *a* at BATS, and corresponded closely to modelled biomass.

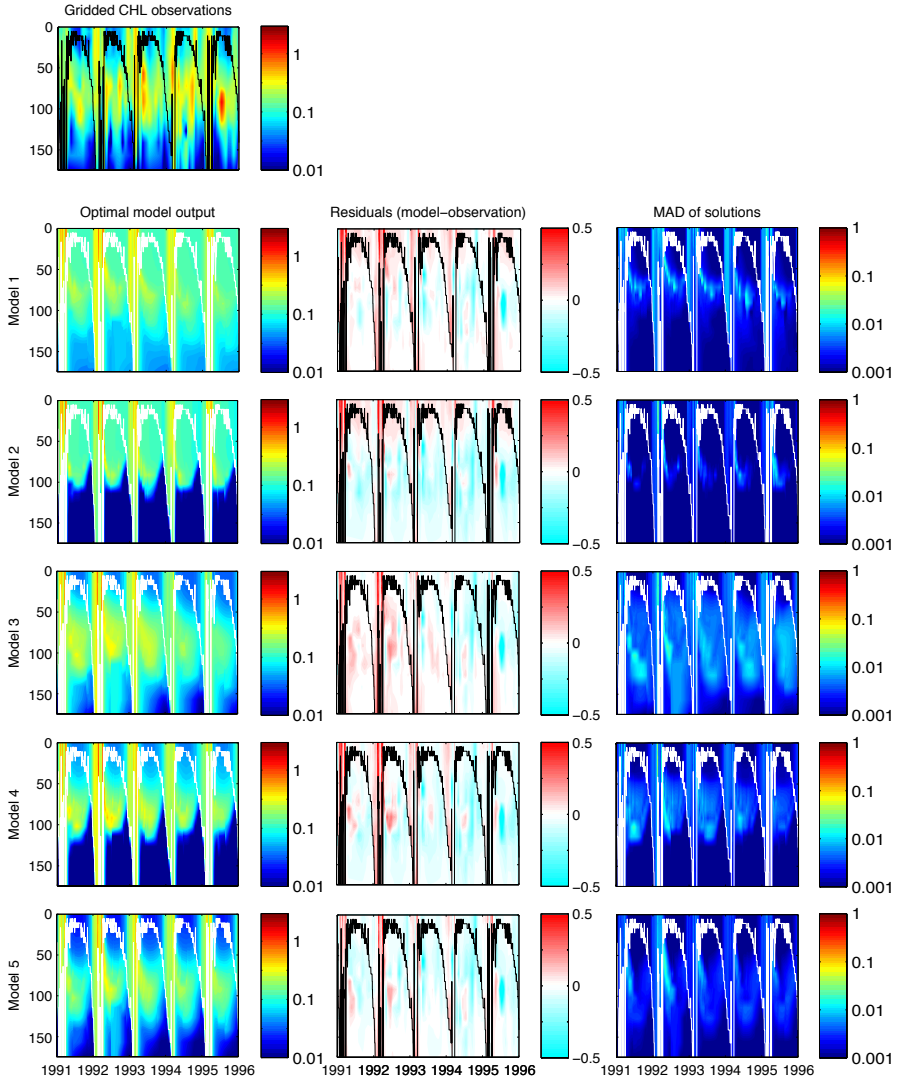

FIGURE 5.6: Chlorophyll *a* at BATS (mmol N m^{-3}). The top panel shows observations taken between 1991 and 1995. The three columns below correspond to the optimised model output, model residuals and the median absolute deviation (MAD) of the optimal solutions, respectively. Complete rows represent output from each of models one to five. Modelled mixed-layer depths are shown by solid black or white lines.

At NABE the observations collected during late April 1989 reveal peak chlorophyll *a* concentrations of ∼2 mg Chl m^{-3}, much higher than those seen at BATS. Model-data residuals show the chlorophyll *a* errors were patchy, although broadly speaking modelled concentrations within the mixed layer were too low before the shoaling of the mixed-layer depth, and too high after. Below the mixed layer the modelled chlorophyll *a* concentrations were always excessive. It appears that when the chlorophyll *a* biomass was integrated over 200 m the unrealistically high concentrations below the mixed layer could compensate for the very low concentrations within it (figure 5.5). Increasing model complexity did little to change the pattern of chlorophyll *a* misfits at NABE.

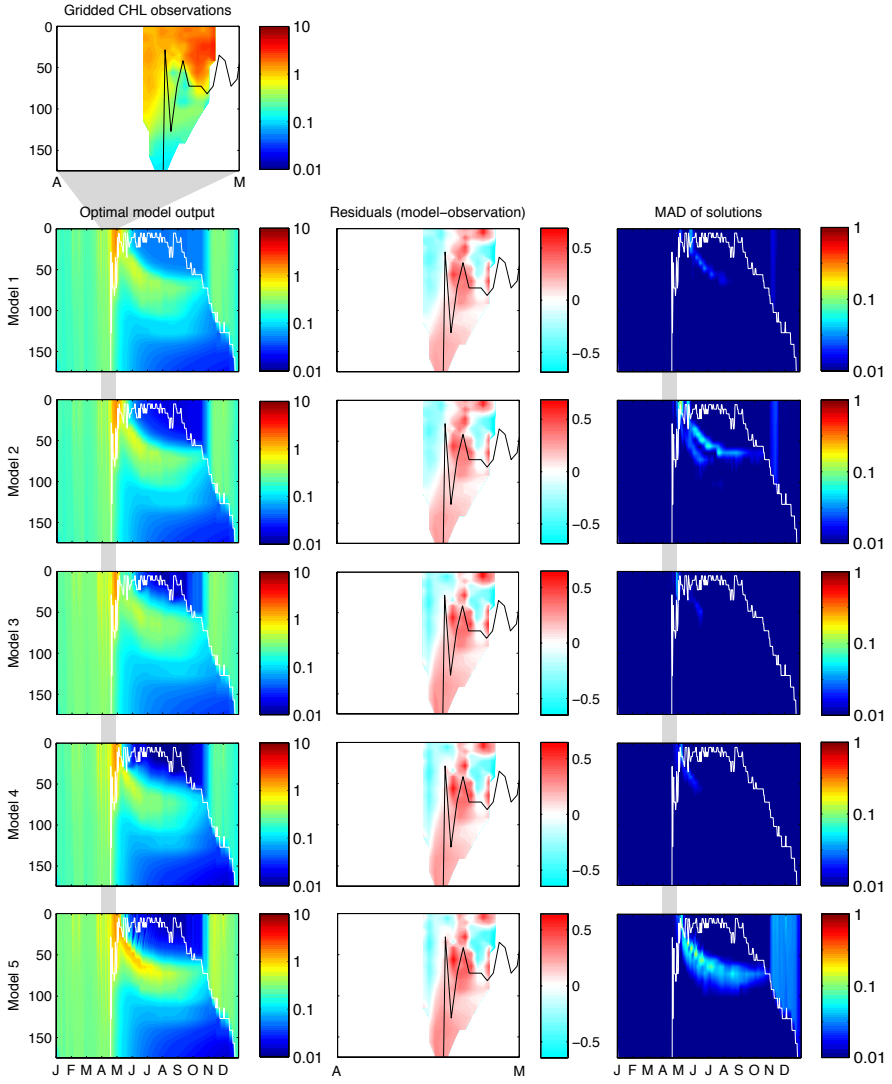

FIGURE 5.7: Chlorophyll *a* at NABE (mmol N m^{-3}). The top panel shows observations taken during April 1989. The three columns below correspond to the optimised model output, model residuals and the median absolute deviation (MAD) of the optimal solutions, respectively. Complete rows represent output from each of models one to five. Modelled mixed-layer depths are shown by solid black or white lines.

Over the broader yearly cycle, all models produced a spring bloom in the surface 50 to 100 m, with a clear DCM developing at the end of the bloom period and extending through the summer. The deepening of the mixed layer in the autumn was associated with a weak bloom in chlorophyll *a* biomass. Model uncertainty was generally greatest around the DCM and the autumn bloom, but it was not high relative to measurement uncertainty.

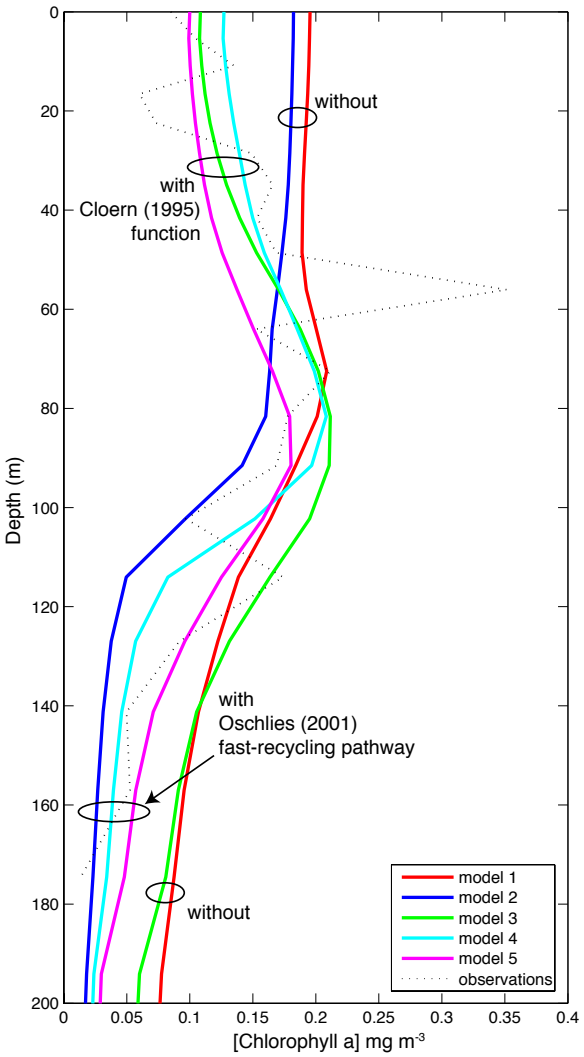

FIGURE 5.8: Vertical time-averaged chlorophyll *a* profiles at BATS.

Zooplankton

Zooplankton data were not incorporated at NABE, and were only available for years 1994 and 1995 at BATS. The data that were available were assigned a very low weight in the cost function (section 3.4) as it was unclear how well they correspond to modelled values. As a consequence the modelled zooplankton, shown in figure 5.9, have very low misfits, contributing on average only 4.1% of the total cost at BATS. This is despite the fact that concentrations in models one to four were much more variable than the assimilated observations. This excessive variability was most pronounced during years where no data were assimilated (1991 to 1993), and only model five produced output with a good visual match to the data.

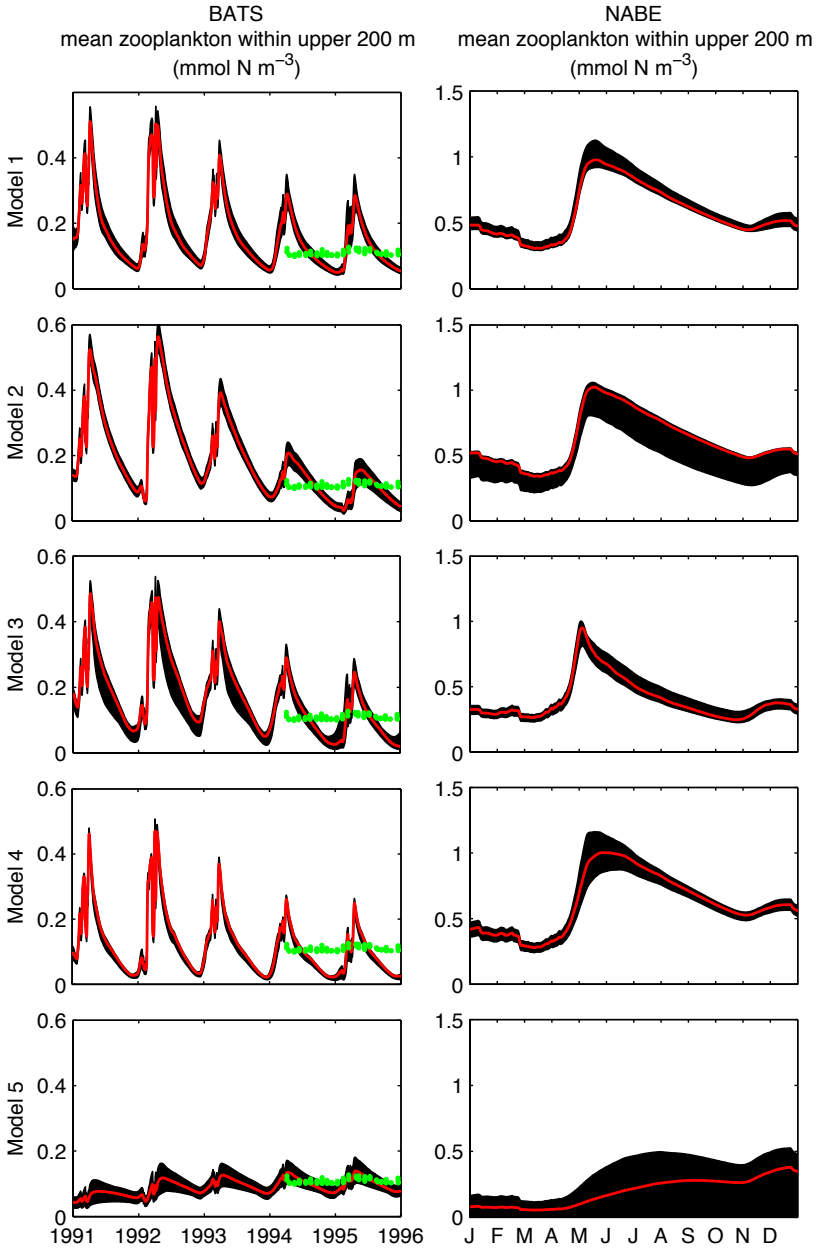

FIGURE 5.9: Zooplankton concentrations at BATS and NABE, averaged over the mixed layer. Red lines represent the trajectory of the single optimal model solution, while the filled black areas cover the range of solutions not significantly different from that minimum (i.e. all the acceptable solutions, see text for details). Green dots represent observations, interpolated onto a vertical grid and similarly averaged over the modelled mixed layer. Note different scales on vertical axes between stations.

The lack of zooplankton data appears to be an important omission at NABE, where optimised zooplankton concentrations were very high relative to observations used in previous modelling studies at NABE (peak values have been between 0.4 and 0.6 mmol N m^{-3}. e.g. Fasham and Evans, 1995; Waniek, 2003; Schartau and Oschlies, 2003b). Output was also highly uncertain, and this was particularly notable in model five, where the annual cycle of zooplankton biomass did not increase during the observed spring phytoplankton bloom and a number of the "optimal" solutions were associated with zero zooplankton biomass.

The optimal zooplankton trajectories did not always show a repeating annual cycle. Although such a condition was not included as a constraint in the cost function (as it was in Schartau and Oschlies, 2003a), and should not necessarily be expected at BATS where the physical forcing showed considerable interannual variability (figure 3.1), the performance of model five at NABE was poor to say the least. A longer model spin-up period and more carefully applied initial conditions may have led to more credible model output in this case.

Particulate organic nitrogen (PON)

All of the models correctly estimated the annual mean PON biomass at BATS, but tended to overestimate the variability. The mean and standard deviation for observations entering the cost function were 0.27±0.08, while the modelled equivalents have mean values of 0.27±0.14, 0.26±0.14, 0.29±0.13, 0.23±0.12 and 0.26±0.11 for models one to five respectively. On average PON contributed 9.1% of the model error at BATS and 14.8% at NABE.

In contrast to BATS, all the models underestimated the mean PON values at NABE while also underestimating the variability. The mean value for assimilated PON observations at NABE was 1.21±0.64, while the modelled equivalents were 0.72±0.21, 0.75±0.21, 0.54±0.20, 0.76±0.23 and 0.57±0.30 for models one to five respectively. The peak observed PON concentration within the mixed layer was ∼3.7 mmol N m^{-3}, while the maximum equivalent value returned by any of the models was only ∼1.5 mmol N m^{-3}.

Dissolved organic nitrogen (DON)

DON was only included as a state variable in model five. The time-evolution of the mean concentration within the mixed layer is shown for each site in figure 5.10. DON data were not included in the cost function, but although modelled DON was slightly

too high, it fitted the data at BATS reasonably well. As in the data, there was no strong seasonal pattern in the modelled DON concentrations, which tended to weakly accumulate during the winter bloom period before being drawn down over the summer.

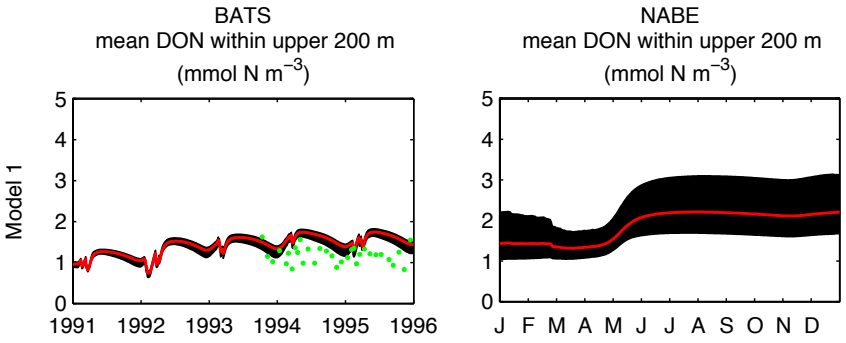

FIGURE 5.10: DON concentrations at BATS and NABE, averaged over the surface 200 m. Red lines represent the trajectory of the single optimal model solution, while the filled black areas cover the range of solutions not significantly different from that minimum (i.e. all the acceptable solutions, see text for details). Green dots represent observations, interpolated onto a vertical grid and similarly averaged over the modelled mixed layer.

There were no data available at NABE. The yearly cycle of modelled DON shows concentrations increased immediately after the spring bloom, but these were not drawn down during the summer. A repeating annual cycle was not achieved at NABE (see the previous comments on modelled zooplankton trajectories).

Primary production (CUp)

Primary production at BATS was the second largest source of model error after chlorophyll *a*, contributing on average 35.8% of the total error. Estimates of primary production were also very sensitive to model structure and this was the most variable source of misfit at BATS.

Modelled estimates of primary production within the surface 200 m are shown for both sites in figure 5.11. All models at BATS tended to underestimate the amount of primary production, especially during the summer. The addition of the fast-recycling pathway in model two improved the low summer production, but also resulted in erroneously high peak bloom production. The variable Chl:N ratio added in model three resulted in a lower misfit than models one and two, with the peak production much closer to observed values. The lowest misfit for primary production was achieved when both these components are included in model four. Although the inclusion of the DON pathway led

to a small improvement in the overall misfit, primary production misfits became slightly larger (table 5.2).

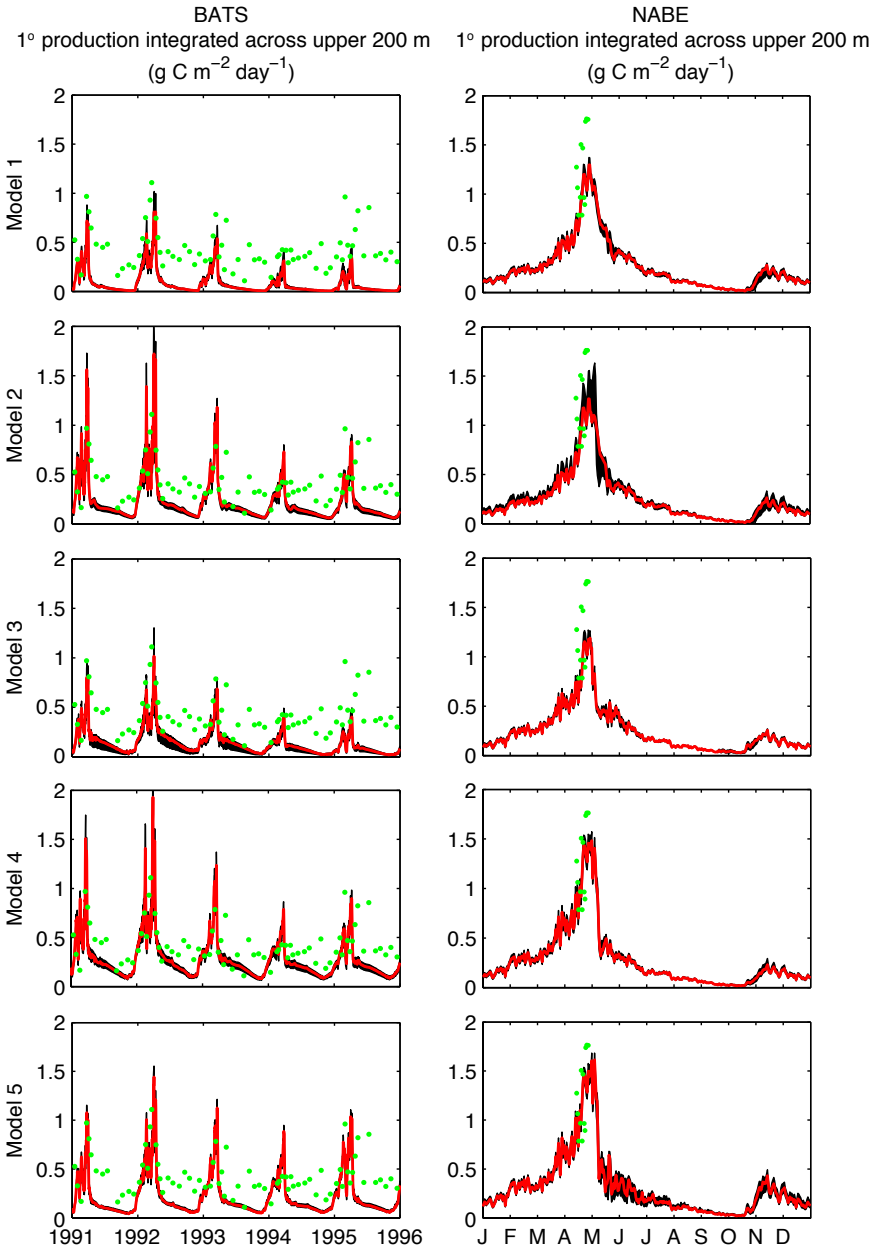

FIGURE 5.11: Primary production at BATS and NABE, averaged over the mixed layer. Red lines represent the trajectory of the single optimal model solution, while the filled black areas cover the range of solutions not significantly different from that minimum (i.e. all the acceptable solutions, see text for details). Green dots represent observations, interpolated onto a vertical grid and similarly averaged over the modelled mixed layer.

Primary production at NABE provided on average the smallest component of the total misfit, but was still significantly improved by changes to the model structure. None of the models were successful in capturing the peak observed values of primary production,

although this discrepancy was slightly reduced in models two, four and five (i.e. those with the fast-recycling pathway). The DON model was best in terms of misfit cost and peak spring bloom primary production.

Depth-resolved profiles for modelled and observed primary production are shown in figures 5.12 and 5.13, together with model-data residuals and uncertainty estimates. Observations at BATS reveal that primary production did not follow the same pattern observed for chlorophyll *a* in figure 5.6, with the fastest carbon uptake observed in the surface waters during the spring bloom. During the summer months primary production was roughly uniform down to approximately 100 m, with no strong increases associated with the DCM.

All five models showed a strong peak in primary production during the spring, but model one failed to capture the summertime primary production observed in the surface waters. The addition of the fast-recycling pathway in model two allowed for more production in the surface during this period, but also resulted in very slightly excessive primary production in the DCM. The addition of the variable Chl:N ratio in models three to five allowed for a more accurate distribution of production within the surface waters.

Model-data residuals at BATS reveal that production across the surface 100 m was too low throughout the year, except in the upper 50 m during the spring bloom, when there was too much phytoplankton growth. The general underestimation of primary production seems to have been reduced as model complexity was increased.

Primary production was only observed in the surface 50 m at NABE, with values of approximately $\sim$8 mmol N m^{-3} d^{-1} above 30 m dropping to nearer 0.1 mmol N m^{-3} d^{-1} below this depth. The modelled annual cycle was very similar between the five models, with a peak production occurring during April and May in the surface 50 m. The spring bloom was followed by a period where production is centred in the DCM between 50 and 100 m. The deepening of the mixed layer in October coincided with a slight increase in productivity. Model five produces unrealistic fluctuations in primary production during the summer. Model-data residuals were very closely correlated with observations, with production too low close to the surface and too high further down the water column. Model uncertainty was highest in the DCM, most notably in models two and five.

5.3.3 Optimal parameters and nitrogen fluxes

The μGA can provide a single "optimal" set of parameters that yields the lowest cost function value for each model, but given that the models were underdetermined and there

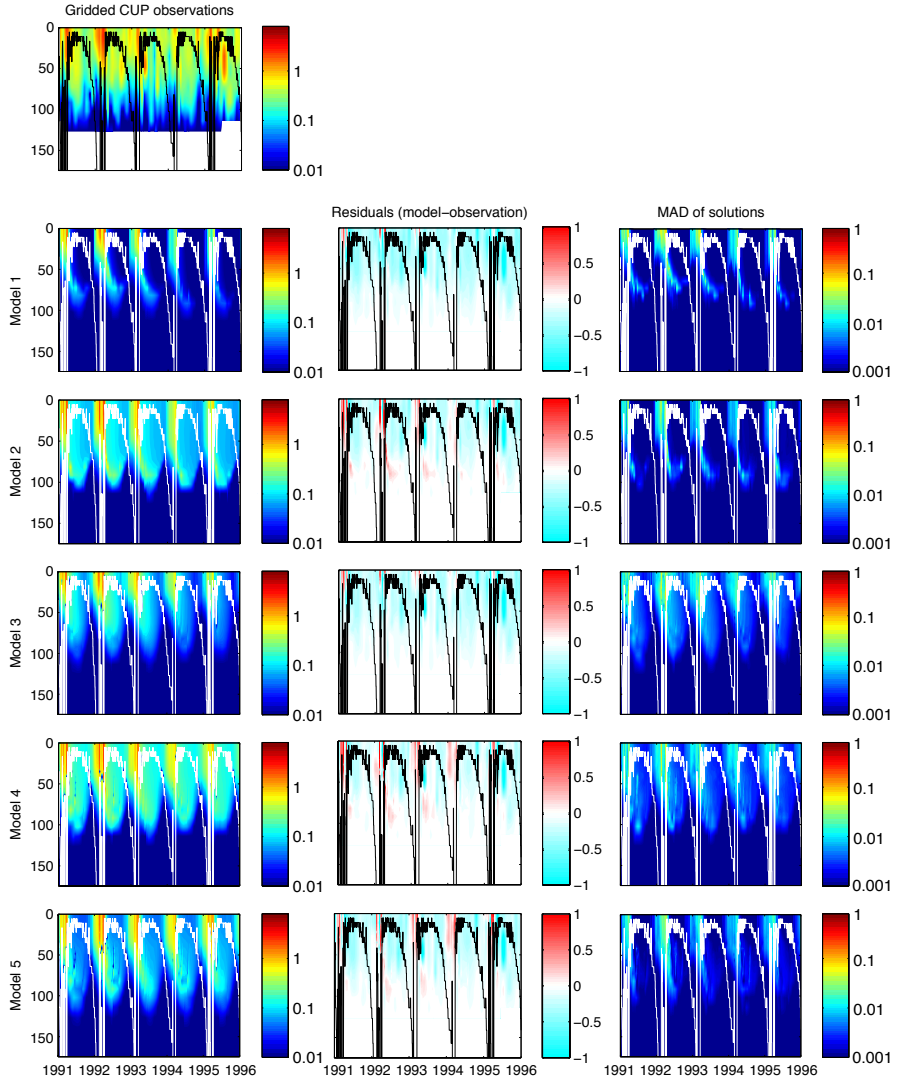

FIGURE 5.12: Primary production at BATS (mmol N m^{-3} d^{-1}). The top panel shows observations taken between 1991 and 1995. The three columns below correspond to the optimised model output, model residuals and the median absolute deviation (MAD) of the optimal solutions, respectively. Complete rows represent output from each of models one to five. Modelled mixed-layer depths are shown by solid black or white lines.

were a number of non-unique solutions that fit the data equally well, such a solution has to be considered as unrealistically precise. Instead, it will be more informative to state the range for each parameter over which statistically very similar misfit costs could be obtained.

Although the μGA is a computationally expensive algorithm, its stochastic nature does have the advantage of providing much useful information about the shape of the cost function (Schartau and Oschlies, 2003a), especially in the region of the global minimum. This information is shown for each model in figures 5.14 (BATS) and 5.15 (NABE).

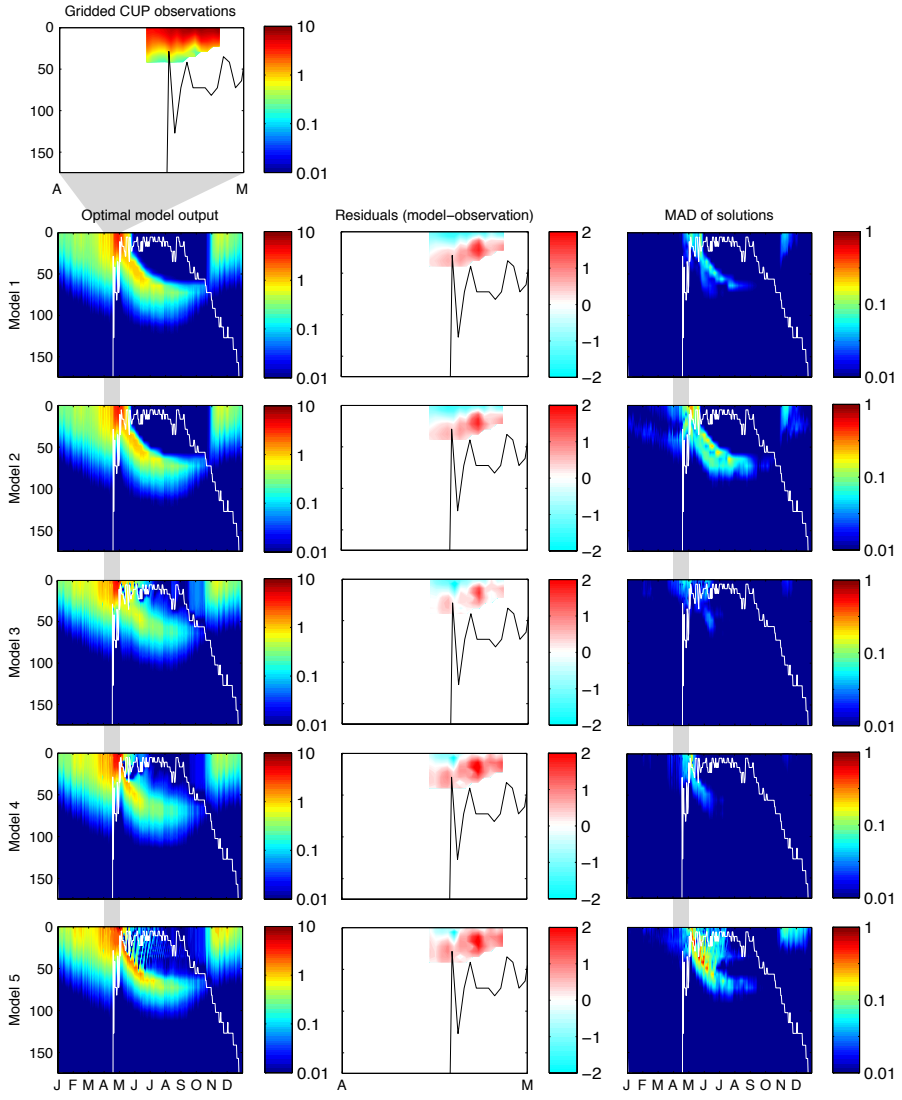

FIGURE 5.13: Primary production at NABE (mmol N m^{-3} d^{-1}). The top panel shows observations taken during April 1989. The three columns below correspond to the optimised model output, model residuals and the median absolute deviation (MAD) of the optimal solutions, respectively. Complete rows represent output from each of models one to five. Modelled mixed-layer depths are shown by solid black or white lines.

For each discrete value evaluated across the range of each parameter, the stepped line represents the minimum cost attained by the μGA, regardless of the values of the other parameters. Solutions are defined by the values for each parameter where a misfit cost equal to or lower than the threshold cost could be defined (see chapter 4). These solutions are marked as coloured dots in figures 5.14 and 5.15. It is important to note that the marginal solutions defined in figures 5.14 and 5.15 do not include all the information contained in the conditional solutions that were used to generate all the other figures used in this chapter (and the next). Any parameter correlations, which will often be

critical to gaining a good fit to the data, are ignored in the marginal solutions.

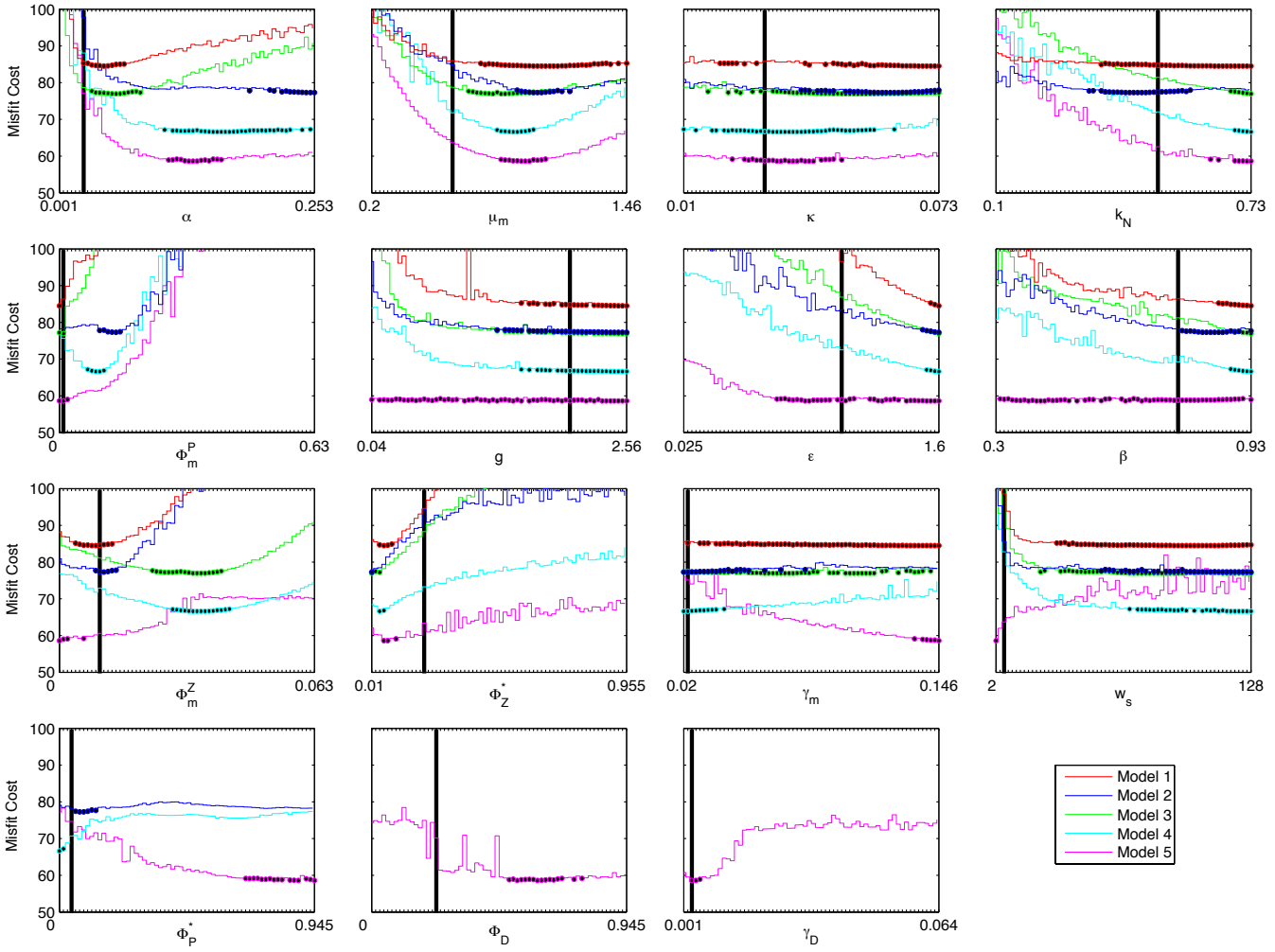

FIGURE 5.14: Misfit sensitivity to parameters and solution parameter values for the five models at BATS. Highlighted solutions (coloured dots) are defined as those parameter values returning a misfit cost not significantly worse than the minimum located cost. Each model is represented by a different colour, as described in the legend. The default parameter values are marked by thick vertical lines. For further details on the figure structure, see figure 4.5 and text in section 4.3.4.

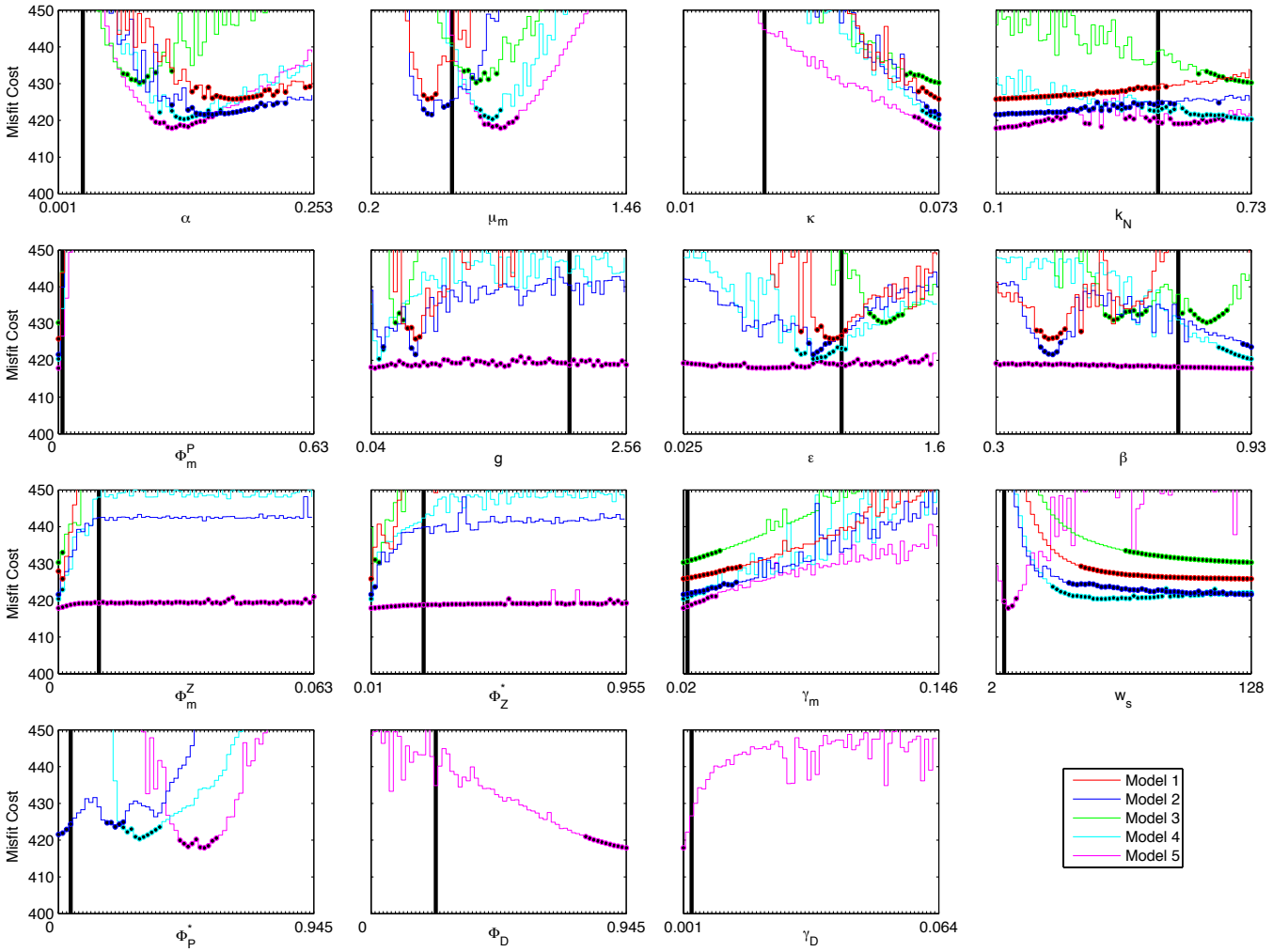

FIGURE 5.15: Misfit sensitivity to parameters and solution parameter values for the five models at NABE. See figure 5.14 for details.

The following section presents the nitrogen fluxes through each of the optimised models and the parameter solutions given in in figures 5.14 and 5.15 are referred to only in this context. Any further analysis of the optimal parameter values is undertaken in the discussion (section 5.4.1).

Nitrogen Fluxes at BATS

Estimates of the annual model nitrogen fluxes associated with the optimal solutions at BATS are shown in figure 5.16. An estimate of each flux is given as the median of all acceptable solutions, with variability quantified by the median absolute deviation (MAD). Black and grey arrows describe the minimum and maximum solution values for each flux. In the following paragraphs the fluxes will be described, with reference to the governing parameter values that were selected by the optimisation process.

Model one. Inorganic nitrogen was taken up by phytoplankton at a rate of 371±12 mmol N m^{-2} $year^{-1}$. This is very low when compared to observations and estimates from other modelling studies (e.g. 1279±359 mmol N m^{-2} $year^{-1}$ Schartau and Oschlies, 2003b). The half-saturation constant for dissolved inorganic nitrogen uptake (k_N) did not affect the cost function when it is set to values greater than 0.36 mmol N m^{-3}, indicating that nitrogen uptake at BATS was not saturated given conventional values of k_N (0.5 mmol N m^{-3}).

The maximum grazing rate for zooplankton, g, shows a similar sensitivity pattern to k_N, influencing the cost function only at relatively low values (< 1.52 d^{-1}). Z grazing did not saturate for typical values of g (2 d^{-1}) and the amount of grazing was thus determined by the prey encounter rate, ϵ, (1.55 - 1.66 m^{-6}(mmol N)$m^{-1}d^{-1}$) and concentrations of phytoplankton and zooplankton (both of which were constrained by observations). Phytoplankton loss was entirely through the grazing pathway, with linear phytoplankton loss to detritus (Φ_m^P) set to 0.0 d^{-1}.

A large proportion of grazing ($85\% \leqslant \beta \leqslant 93\%$) was assimilated into zooplankton biomass. This was then either returned to DIN by zooplankton mortality (optimally parameterised at a rate of approximately 0.0 - 0.01 d^{-1}) or lost to detritus at a rate of approximately 0.04 - 0.09 m^3(mmol N)$^{-1}d^{-1}$. Remineralisation and sinking of detritus were both relatively minor fluxes in this system.

Model two. A fast-recycling pathway was introduced in model two as an implicit representation of the microbial loop. The optimised model solutions yielded values of 667±12 mmol N m^{-2} $year^{-1}$ for this flux. The mean annual primary production was

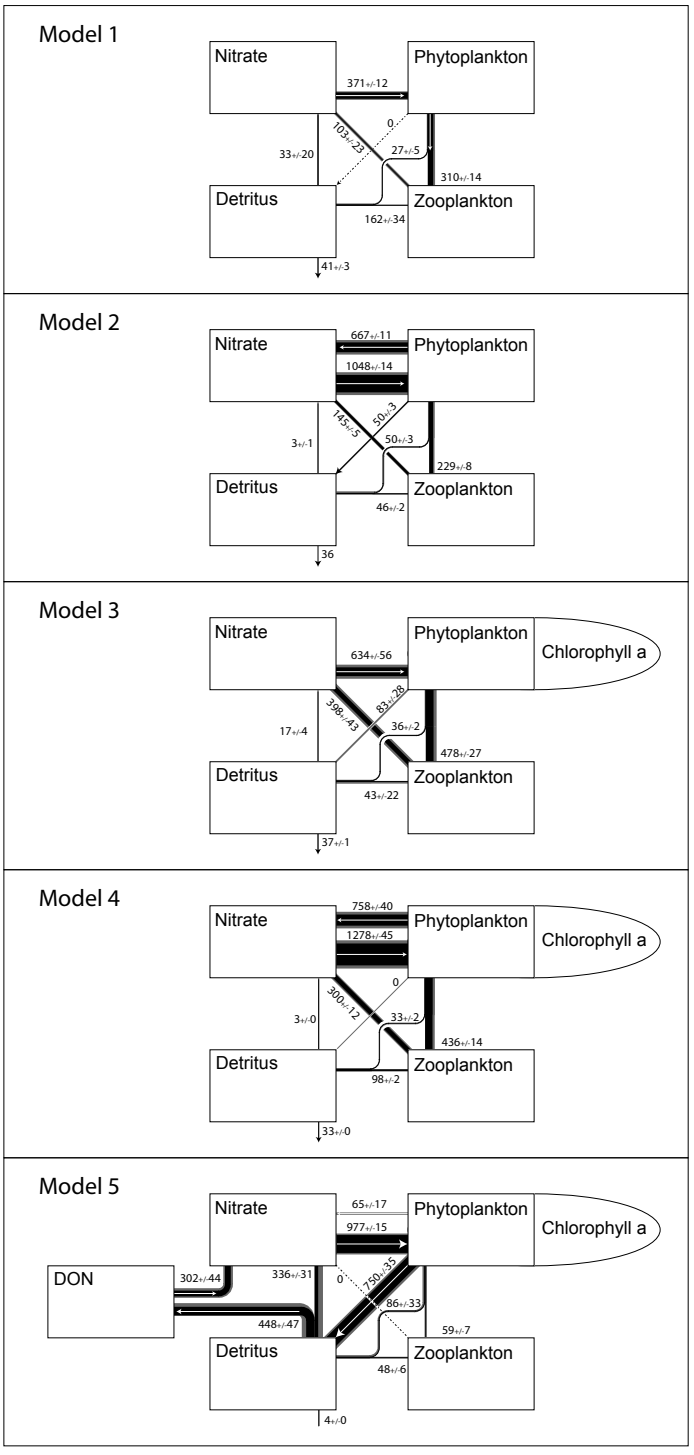

FIGURE 5.16: Optimal nitrogen fluxes at BATS. Optimised fluxes in mmol N m^{-2} $year^{-1}$ are shown for all the acceptable parameter solutions of each model. The widths of flux pathways are linearly scaled to indicate the magnitude of annual mean fluxes between state variables. Black flux pathways indicate the minimum flux associated with an acceptable solution, while grey lines indicate the maximum such flux. Figures next to the pathways correspond to the median and MAD of the solutions. Dotted lines indicate where the maximum optimised flux was zero.

increased over model one by approximately the same amount, to 1048±14 mmol N m^{-2} $year^{-1}$, bringing the estimate closer to observations and estimates from previous modelling studies. Optimal values for the parameter describing this flux, Φ_m^P, were set at 0.1 - 0.15 d^{-1}; higher than the prior value of 0.01 d^{-1}, and well-constrained near the centre of the prior range.

The linear term for non-grazing phytoplankton mortality was replaced in model two with a quadratic function. The flux from phytoplankton to detritus subsequently increased from zero in model one to 50±3 mmol N m^{-2} $year^{-1}$. Phytoplankton biomass was actively attenuated by three separate processes and grazing by zooplankton was not as strong as in model one. The optimal prey encounter rate, ϵ, was still high. The flux from zooplankton to DIN was larger than in model one, but losses to detritus were smaller. Sinking and remineralisation of detritus were both diminished from the already small fluxes seen in model one.

Model three. Model three did not include the fast-recycling pathway or quadratic phytoplankton loss of model two, but did incorporate a variable Chl:N ratio (Cloern et al., 1995). The decoupling of [chlorophyll *a*] from phytoplankton biomass meant that phytoplankton biomass was no longer linearly related to chlorophyll *a* observations, as shown in figure 5.5.

In relation to model one, primary production increased to 634±56 mmol N m^{-2} $year^{-1}$ with stronger recycling of nitrogen via grazing and exudation by zooplankton. The decoupling of phytoplankton from chlorophyll *a* allowed a significantly higher phytoplankton biomass to exist, with an associated increase in primary production, linear phytoplankton mortality and zooplankton grazing. Zooplankton excretion was also much larger than in model one. Zooplankton mortality and the sinking and remineralisation of detritus were all relatively weak, as in model one.

Model four. Here the fast-recycling pathway and quadratic phytoplankton mortality of model two were combined with the variable Chl:N ratio of model three. In this case primary production was estimated at 1278±40 mmol N m^{-2} $year^{-1}$, which was very similar to the value of yielded by 1279±359 mmol N m^{-2} $year^{-1}$ Schartau and Oschlies (2003b), who included a similar fast-recycling pathway and variable Chl:N ratio. As might be expected, the increased recycling of nitrogen, which was dealt with mostly by the flux from phytoplankton to DIN in model two and by the N-P-Z pathway in model three, was shared between these two pathways in model four. Once again, fluxes from phytoplankton and zooplankton to detritus were either non-existent, or weak, and sinking and remineralisation of detritus were also both relatively small fluxes.

Model five. The final model is the same as model four, but with the additional DON state variable, which was not constrained by any direct observations. Modelled primary production was optimally 977±15 mmol N m^{-2} $year^{-1}$. Phytoplankton loss was primarily straight into the detritus pool, with its governing quadratic phytoplankton loss parameter set to very high values of between 0.69 and the maximum value of 0.945 $m^3(mmol\ N)^{-1}d^{-1}$. Detrital nitrogen was either remineralised to DIN or passed to DON. Optimal values of the fraction of detrital breakdown passed to DON, Φ_D, were between 0.51 and 0.78, while the rate of hydrolysis of DON was set to relatively low values between 0.003 and 0.005 d^{-1}.

As well as altering the flux of nitrogen through the model, the addition of DON in model five dramatically reduced the sensitivity of a number of parameters, especially those related to the grazing of phytoplankton by zooplankton. Figure 5.14 shows that the parameters g, ϵ and β were all highly underdetermined in model five.

Nitrogen Fluxes at NABE

Estimates of the model nitrogen fluxes at NABE are shown in figure 5.17.

Model one. Primary production within the euphotic zone was estimated at 1176±12 mmol N m^{-2} $year^{-1}$, significantly lower than the estimate of 1891±273 mmol N m^{-2} $year^{-1}$ given by Schartau and Oschlies (2003b), where BATS, NABE and OWS-India data were assimilated simultaneously. There were no direct observational estimates of the annual primary production at NABE, as nearly all observations were collected during April and May 1989. Only data from that period were assimilated here, and results presented in figure 5.11 suggest that the peak modelled primary production slightly underestimated the observed values.

Unlike the results for BATS, the half-saturation for DIN uptake influences the cost function across the entire prior range, indicating that saturation of DIN uptake was an important factor at NABE. This is consistent with the stronger mixing and associated nutrient supply seen at this eutrophic site (Fennel et al., 2001). As at BATS, model one phytoplankton were grazed at a similar rate to primary production, although a much lower proportion was assimilated by zooplankton ($40\% \leqslant \beta \leqslant 51\%$). The majority of nitrogen was passed instead to detritus via messy feeding. Fluxes into the detrital compartment, which added up to approximately 930±5 mmol N m^{-2} $year^{-1}$ were balanced only in part by remineralisation within the euphotic zone and sinking out of this layer. The remainder was redistributed below the integration depth (200 m) by physical mixing.

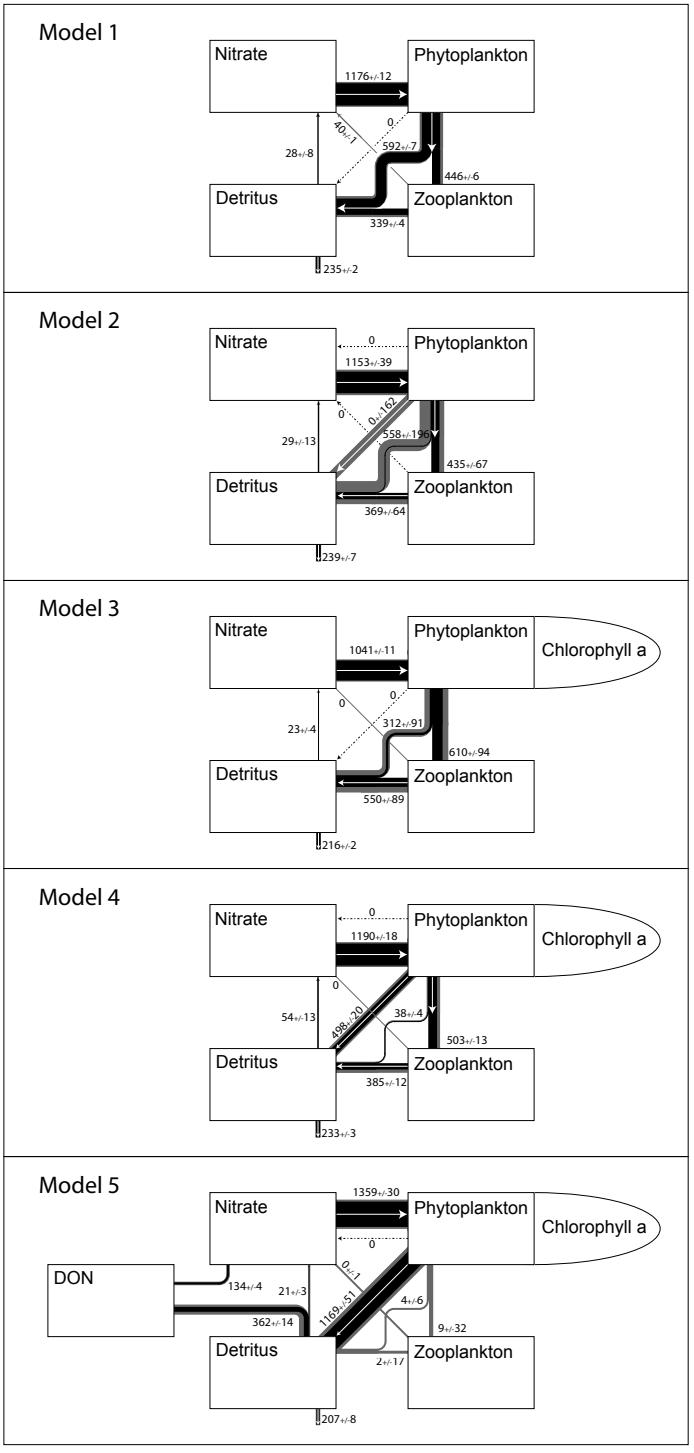

FIGURE 5.17: Optimal nitrogen fluxes at NABE. Optimised fluxes are shown for all the acceptable parameter solutions of each model. See description of figure 5.16 for details.

Model two. Fluxes in model two were generally similar to those in model one, although the introduction of quadratic phytoplankton mortality led to greater uncertainty surrounding phytoplankton loss. Annual primary production was similar to model one, at 1153±11 mmol N m^{-2} $year^{-1}$.

There were in fact two distinct configurations of model two that fitted the data at NABE equally well. The first was very similar to model one, with the new quadratic phytoplankton loss term and fast-recycling pathway both set to approximately zero. Phytoplankton losses in this mode were solely through the grazing pathway, at a yearly rate broadly equivalent to primary production. Alternatively however, the quadratic phytoplankton loss parameter Φ_P^* could adopt a slightly higher values of 0.165 - 0.24 $m^3(mmol\ N)^{-1}d^{-1}$, allowing a phytoplankton mortality rate of 0±162 mmol N m^{-2} $year^{-1}$ (estimates were distributed very irregularly between 0 and 495 mmol N m^{-2} $year^{-1}$). To balance this additional loss term the grazing pathway was reduced by an equivalent amount, mediated by a reduction of the maximum grazing rate g (from 0.4 - 0.52 d^{-1} to 0.16 - 0.48 d^{-1}), and by an increase in β (from 40 to 51% to 41 to 93%). Thus a greater proportion of the (reduced overall) grazing was assimilated by zooplankton.

Model three. Model three introduced a decoupling of chlorophyll a and phytoplankton biomass through the variable Chl:N ratio. Annual primary production was estimated at 1041±11 mmol N m^{-2} $year^{-1}$. This is slightly lower than the solutions for models one and two, but in general, the nitrogen fluxes were very similar. There was however a greater uncertainty associated with fluxes in and out of the zooplankton.

Model four. Model four added the fast-recycling pathway of model two to the variable Chl:N ratio of model three. The estimated primary production was 1190±18 mmol N m^{-2} $year^{-1}$. The fast-recycling pathway was strongly set to zero by the calibration process. Model four did not show the bi-modal solution seen in model two, and the optimal solutions all included a strong flux through the zooplankton compartment, with relatively little ($0.04 \leqslant (1-\beta) \leqslant 0.16$) passing through messy feeding.

Models one to four gave very similar estimates of primary production at NABE, with the the highest estimate (from model four) only 14% larger than the lowest (model three). This is in contrast to the results at BATS, where primary production from model four was 244% higher than the estimate given by model one.

Model five. As at BATS, the addition of DON to the model vastly changed the dynamics of the solution at NABE. Primary production was estimated at 1359±30 mmol N m^{-2} $year^{-1}$, an increase of an additional 14% relative to model four. The majority of

phytoplankton loss was through quadratic phytoplankton mortality, which was optimal between the relatively high values of 0.45 and 0.585 $m^3(mmol\ N)^{-1}d^{-1}$. The flux through the zooplankton compartment, unconstrained by any direct observations, was much smaller than in models one to four. In some of the optimal solutions there was zero flux through zooplankton. The weak flux through the grazing pathway is supported by observations that zooplankton are not a major factor controlling phytoplankton growth during the spring bloom at NABE (Dam et al., 1993), although the high rate of quadratic phytoplankton mortality seems unrealistic. This high rate may be due to the fact that DON production was channelled entirely through detritus, whereas direct fluxes from phytoplankton and zooplankton might also have been expected (Christian et al., 2002).

General comments. Wherever the fast-recycling pathway was included at NABE it was optimally set to zero. The flux from zooplankton to DIN could also be set optimally to zero for every model. Both of these pathways could have been completely removed from the models without adversely affecting the minimum misfits.

5.4 Discussion

The results given above reveal that model-data misfits consistently improve as model complexity is increased through a series of nested models. This is not surprising, given the additional degrees of freedom that are associated with the extra parameters (e.g. Anderson, 2005; Friedrichs et al., 2006), but while these extra terms generally lead to significant improvements in the calibrated cost at BATS, the effect at NABE is much less pronounced. Previously, Fennel et al. (2001) demonstrated that the sensitivity of the cost function to certain parameters was dependent on the dynamics of the site under investigation (or at least those captured by the available observations), and this fact will be used in section 5.4.2 to investigate the role and importance of the different model components at BATS and NABE. By analysing where different model pathways lead to improvements in model function, and where they do not, it should be possible to highlight those model components that are critical for an accurate representation of the two ecosystems, and those which are perhaps less important.

Following this analysis of model structure and ecosystem function, a first assessment of the cost of model complexity is made. At both sites, a large number of non-unique solutions were identified and evaluated. In section 5.4.3 the magnitude and distribution of these solution uncertainties are examined. Recommendations are made regarding observations that may help to constrain model output.

5.4.1 Parameter solutions

Although some parameters such as the phytoplankton growth terms α and μ_m are relatively well-constrained by the data, with sharp increases in cost on either side of the optimal values, not all parameters are reliably constrained by the data. Some parameters, such as the light attenuation by phytoplankton (κ) at BATS, can be adjusted across their entire range without seeming to cause any change in the misfit cost. In other cases, such as for the sinking rate of detritus, the data can only provide information about a lower (or upper) limit for a parameter. It appears that such parameters could be set anywhere above (or below) such limits with no appreciable increase in cost. In both these examples the prior bounds are required to prevent unrealistic parameter values from entering the solution, as the assimilated data were insufficient to do so.

Other parameters, such as the linear phytoplankton mortality rate (Φ_m^P) and the assimilation coefficient (β) are of more concern. Here it appears that a lower misfit cost might be achieved if the parameters were set to values outside of the prior limits. In these cases, the optimal parameter estimate is almost entirely defined by the prior information, although this is not necessarily a very serious issue in all cases. For example, a rate parameter could be easily (and sensibly) assigned a minimum prior value of zero, with optimality of this value over others simply indicating that the associated flux was not required to reproduce the assimilated data. In some cases however, particularly for the prior upper limits placed on parameters, the bounds cannot be assigned with such confidence. Where those bounds strongly influence the solutions (such as for the prey encounter rate ϵ at BATS, or the coefficient for light attenuation by phytoplankton κ at NABE), those solutions will be unavoidably sensitive to poorly defined assumptions.

Optimal estimates of the initial slope of the PI curve (α) together with the growth rate parameter (μ_m) are of particular interest as a consequence of their use in satellite primary production algorithms (e.g. Platt and Longhurst, 2000). Previously, inverse estimates have suggested higher than expected values for α (e.g. Fasham, 2000; Schartau et al., 2001; Schartau and Oschlies, 2003a) unless strong prior constraints were used to assist the optimisation process (e.g. Evans, 1999; Fennel et al., 2001). Schartau and Oschlies (2003a) noted that increased vertical stratification during the peak of the daily irradiance cycle, particularly during nutrient-replete and light-limited (i.e. winter/early spring) conditions could increase phytoplankton growth by raising the net amount of light available to the phytoplankton. This process was not resolved in their model and the authors suggested that the high estimate for photosynthetic efficiency α was in part compensating for the lower light levels experienced by the phytoplankton during the early spring and winter. Furthermore, because the increased value of α was constant

throughout the year, phytoplankton growth was more efficient at depth during the summer and had to be restricted by low values of the growth rate parameter μ_m.

In this study the modelling framework was similar to the one applied by Schartau and Oschlies (2003a), but the frequency of the physical forcing was increased from daily to hourly. For models one and three, where the fast-recycling pathway was not included, the optimal estimates of α are indeed lower and are much closer to prior estimates. These estimates should however only be considered in the knowledge that these optimised models strongly underestimated primary production. When the fast-recycling pathway was added, estimates of α were much larger, with the extra organic nitrogen produced by stronger primary production rapidly remineralised to DIN. The inadequate representation of primary production at BATS is discussed later in the text, but it appears that the high values of α were influenced by the need to maximise this flux. Given this model deficiency, it is not easy to assess the effects of resolving the daily cycle on estimates of α.

Models three to five used a depth-resolved Chl:N ratio that affected the overall phytoplankton growth rate through the chlorophyll specific photosynthetic efficiency $\alpha_{acclimated}$ (equation C.18 in appendix C). Surprisingly, the introduction of this function had little affect on the optimal values of α at BATS, where the fast-recycling pathway exerted a much stronger influence.

5.4.2 Model complexity and ecosystem function

BATS

Modelled estimates of primary production, although associated with low individual uncertainty, are highly variable between models. This is despite the fact that the models were all calibrated to identical observations of primary production. This emphasises the point that estimates of ecosystem fluxes are highly dependent on the underlying model structure (Oschlies, 2001) and that this issue is not overcome by the use of data assimilation (Schartau and Oschlies, 2003b; Spitz et al., 2001). All five models yield low estimates of primary production, relative to assimilated observations of inorganic carbon uptake, as shown in figures 5.11 and 5.12. The models also underestimate the observed yearly integrals of 110-144 g C m^{-2} $year^{-1}$ given by Lohrenz et al. (1992).

Model one, initially developed by Oschlies and Garçon (1999) to demonstrate the sensitivity of modelled primary production to physical forcing, contains no representation of the microbial loop. Given the relative importance of this pathway in the Sargasso Sea

(Fasham et al., 1990; Roman et al., 1995; Spitz et al., 2001), it is perhaps unsurprising that even the optimised model gives an estimate of only 30±1 g C m^{-2} $year^{-1}$.

The fast-recycling pathway of model two was introduced by Oschlies (2001) as an *ad hoc* method of increasing primary production, with the quadratic phytoplankton loss term introduced to prevent excessive recycling in eutrophic regions. The mechanistic analogy given for these two processes is that the fast-recycling pathway represents a parameterised microbial loop while quadratic phytoplankton mortality is related to cell aggregation and loss. Dependent on the value of Φ_m^P, the "microbial loop" should be able to drastically increase modelled primary production (Oschlies, 2001), and this is indeed the case in model two, where carbon uptake is estimated at 83±1 g C m^{-2} $year^{-1}$, an almost threefold increase relative to model one. Although this figure is much closer to observed values (Lohrenz et al., 1992) and corresponds to a 26% reduction in primary production misfit cost (table 5.2), the yearly integral is still approximately 24-43% too low and the model trajectories shown in figure 5.11 reveal that the model output consistently overestimates primary production during winter and spring, while underestimating throughout summer and autumn.

The variable Chl:N ratio applied in models three to five allows a degree of photoacclimation in darker, nutrient replete conditions (Cloern et al., 1995), with increased cellular chlorophyll *a* tied to an increase in the photosynthetic efficiency parameter α (Anderson et al., 2007). These acclimation processes allow some improvement relative to model one, with lower misfits for both chlorophyll *a* and primary production. In fact, total error and error from chlorophyll *a* are improved significantly more by the addition of the Chl:N ratio than by the addition of the fast-recycling pathway, despite the fact that the former change is not associated with any extra tunable parameters.

The results presented in figure 5.8 suggest that both the fast-recycling pathway and the variable Chl:N ratio are required to accurately model the vertical (time-averaged) chlorophyll *a* profile at BATS. While all five optimised models are capable of reproducing the biomass at intermediate depths, if the Cloern et al. (1995) function is omitted, surface concentrations are too high. Similarly, if the fast-recycling pathway is not included then fitting the chlorophyll *a* concentrations at intermediate depths leads to excessive chlorophyll *a* biomass below 120 m. In this respect the simple models are capable of fitting the (time-averaged) DCM, but only at the expense of poor performance elsewhere. Adding these two model components allows the model to find a good fit to more data simultaneously.

Addition of the DON pathway led to an overall decrease in misfit cost over model four, although primary production misfits became larger. Most of the improvements were in modelled chlorophyll *a* and zooplankton concentrations. Despite the fact that

DON dynamics were very crudely modelled, ignoring what are probably significant inputs directly from phytoplankton and zooplankton (Fasham et al., 1990; Anderson and Williams, 1998; Christian et al., 2002) and any variability in bacterial biomass, the modelled DON at BATS fit the data reasonably well. This is despite the fact that no DON data were assimilated at either site.

The DON pathway was introduced in model five to slow down the recycling of winter bloom biomass with the aim of increasing primary production during the summer. This was not achieved, although figure 5.16 does show that most recycled nitrogen passes through detritus and DON, rather than through the fast-recycling pathway. The rate parameter for hydrolysis of DON was optimally set to low values of between 0.003 and 0.005 d^{-1}. These very low values give a DON turnover time of between 200 and 333 days. Given that the aim of including the DON pathway was to provide a post-bloom source of DIN during the summer, a turnover rate on the order of a few weeks might reasonably might have been expected. It is possible that recycling via DON was too crudely parameterised to provide a useful supply of inorganic nutrients throughout the summer. Additionally, the one-dimensional framework employed here is probably too simple to account for all DON recycling, as it has been suggested that the horizontal supply of DON could be significant in the oligotrophic gyres, as a consequence of the slower rate of DON utilisation relative to DIN (Williams and Follows, 1998).

A number of processes that were not resolved in this study can be postulated as mechanisms by which primary production could be increased. For example the physical forcing applied here does not include vertical or horizontal advection, and so any physical supply of nutrients brought about by large-scale circulation and wind-driven advection are neglected. Within the subtropical gyre vertical Ekman pumping can be largely discounted as a source of new nutrient as it results in a net downwelling of water across the base of the mixed layer (additional analysis, not presented here, revealed that vertical oscillations in the water column did not affect the modelled biogeochemistry in any significant way). Aside from vertical motions, the horizontal component of the wind-driven circulation may well provide a significant flux of nutrient rich waters into the subtropical gyre (Williams and Follows, 1998). The contribution at BATS is however thought to be relatively small because most of the dissolved inorganic nitrogen supplied this way is consumed nearer to the gyre boundaries. As stated previously, the slower rate of DON breakdown could increase the amount of nitrogen biomass supplied by horizontal advection.

A study by (Popova et al., 2006) applied a six-component ecosystem model within a three-dimensional version of the OCCAM model. The modelled primary production at BATS is reproduced in figure 5.18 and bears resemblance to the output shown in

figure 5.11, particularly with regard to the low estimates during the summer. Although the physical model used by Popova et al. (2006) included the advection of nutrients by large scale wind and buoyancy driven circulation, these processes were not sufficient to raise summer and autumn primary production estimates to higher levels than were produced in the one-dimensional framework applied here (although this may well have been possible if the model had been rigorously calibrated to BATS data).

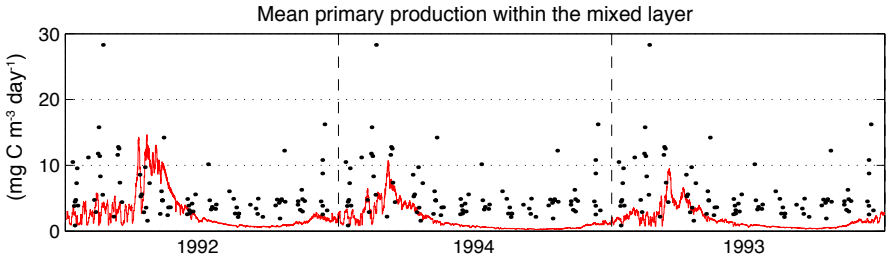

FIGURE 5.18: Primary production at BATS from a 3D coupled physical-ecosystem model, adapted from Popova et al. (2006). Reproduced with the permission of E. E. Popova.

.

The addition of nutrients to the surface waters by mesoscale activity has also been suggested as a mechanism by which extra new production can be fuelled within the subtropical gyres (Falkowski et al., 1991). Within a cyclonic eddy, divergent horizontal flow causes isopycnals to be deflected upwards, thus bringing deeper nutrient rich waters into the euphotic zone. If these waters are subsequently returned to depth following the rapid uptake of nutrients by phytoplankton, the result is a net upward flux of nutrients. This process was not accounted for in this study, nor by Popova et al. (2006), because the resolution of mesoscale features would require a three-dimensional physical model with possibly an order of magnitude improvement in the horizontal resolution (Oschlies, 2002). Eddy pumping has been estimated to increase new production by as much as 350 ± 150 mmol N m^{-2} $year^{-1}$ through increased entrainment of DIN (McGillicuddy and Robinson, 1997). This translates to 28 ± 12 g C m^{-2} $year^{-1}$ with the constant C:N ratio employed in this study. Although this additional nutrient flux would bring the best estimates of primary production into the observed range given by Lohrenz et al. (1992), it seems likely that these figures overestimate the efficiency of both the supply of nutrient from depth (Oschlies, 2002) and the utilisation of upwelled nutrients within the euphotic zone (Martin and Pondaven, 2003). Using a $^1/_9$ degree eddy resolving model, Oschlies (2002) estimated eddy induced nitrate supply to the North Atlantic oligotrophic gyre to be on the order of 50 mmol N m^{-2} $year^{-1}$, which is insufficient to rectify the shortfall presented here.

In spite of this, a simple NPZD model nested in an "eddy permitting" physical model (meridional resolution = $^1/_3^\circ$; zonal resolution $^2/_5^\circ$) did achieve better estimates of BATS summer primary production than are presented here (Schartau and Oschlies, 2003b). It is hard to say whether the extra production is due to the additional eddy-supplied nitrate, but even if it is, the uncertainty surrounding the efficiency of eddy pumping (Oschlies, 2002; Martin and Pondaven, 2003) casts doubt on whether this is a genuine effect. Assuming that primary production was otherwise too low, presented with this additional source of extra DIN, the cost function would always seek to use it as efficiently as possible, regardless of whether that was realistic.

An alternative mechanism for increased carbon uptake by phytoplankton, especially in oligotrophic waters, is through fixation of nitrogen by diazotrophs. Nitrogen fixing *Trichodesmium* species are known to favour warm, stratified, oligotrophic conditions (Capone et al., 1997), and it is believed that N_2-fixation activity is at its highest during the summer months, where modelled primary production in this study is particularly low. The omission of this seasonal increase in DIN supply may contribute in part to the low values for primary production seen in this study, but is not sufficient to fully account for them. Direct estimates of nitrogen fixation (as summarised by Hansell et al., 2004) typically range from 0.25-34 mmol N m^{-2} $year^{-1}$ ($\sim$0.02 - 2.7 g C m^{-2} $year^{-1}$), while the modelled primary production in this study is at best 25 g C m^{-2} $year^{-1}$ less than observed values.

The modelled shortfall is also consistent with the flawed assumption of a fixed molar C:N ratio of 6.625, particularly with regard to the conversion of modelled uptake of DIN to observed ^{14}C derived estimates (Schartau and Oschlies, 2003b). This fixed ratio was adopted on the basis of the long held paradigm that carbon, nitrogen and phosphorous are not only present in particulate organic matter and deep waters at a constant ratio of 106:16:1 (Redfield, 1934), but are also taken up, assimilated, exported and remineralised at this constant ratio.

While the Redfield ratio has generally been shown to apply over long time scales, particularly in the deep ocean (e.g. Copin-Montegut and Copin-Montegut, 1983; Anderson and Sarmiento, 1994), observations have shown that carbon uptake can be far in excess what might be expected through knowledge of nitrogen uptake and application of the Redfield ratio. This process of carbon overconsumption (Toggweiler, 1993) has been recorded in the north Atlantic with C:N ratios of over 25 (Kortzinger et al., 2001). Increased uptake of carbon over nitrogen can either lead to increased cellular C:N ratios (e.g. Goldman, 2000), or results in extracellular release of carbon rich photosynthate that cannot be incorporated into biomass in the absence of sufficient nitrogen (Fogg, 1983; Nagata, 2000). Neither of these processes are accounted for in the models presented here. The use of

the ^{14}C radiocarbon technique to measure primary production at BATS is defined as "the uptake of inorganic carbon into particulate matter" (BATS, 2008). If a significant amount of carbon rich phytoplankton exudate is transformed into particulates, as was demonstrated to be possible by Engel et al. (2002), the ^{14}C derived estimates of primary production may well be in excess of equivalent nitrogen based estimates. This process however, to the best of our knowledge, has not been shown to be significant at BATS.

Anderson and Pondaven (2003) assumed a C:N ratio for phytoplankton of 7.5 at BATS. As stated previously, a Redfield derived value of 6.625 was used in this study and adopting the Anderson and Pondaven (2003) value would lead to a 13% increase in estimates of carbon based primary production. Even higher values would be permissible during periods of nutrient stress if a variable ratio was used. The model of Geider et al. (1998) incorporates such a variable ratio through an independent cellular "quota" for carbon, nitrogen and chlorophyll *a*, and has been applied in numerous modelling studies (e.g. Moore et al., 2002; Lima and Doney, 2004; Schartau et al., 2007).

NABE

The obvious model deficiencies at NABE are consistent across all five models. Modelled DIN for example, shown in figure 5.2, fails to reproduce the peak concentrations observed at the start of the bloom period. It appears that this is associated with the unrealistically low concentrations of DIN seen below the mixed layer in figure 5.4. As the shortfall is mostly below the euphotic zone and the water column is vigourously mixed over winter, it seems likely that it is caused by either unrealistic physical forcing or a low initial DIN profile. The initial DIN profile was set to the mean observed nitrate profile for each site (section 3.5), which may not have been appropriate at NABE given the observational bias towards the spring. An observed nitrate profile from January 1994 was used to initialise the model of Waniek (2003), but the authors found a similar discrepancy in the peak concentrations at 47°N, 20°W. The initial nitrate profile was not subject to optimisation in either this study or Waniek (2003), but was in Fasham and Evans (1995), where the water column was simplified to a readily adjustable linear nitrate profile below a homogenous mixed layer. This study achieved a relatively good fit to observed nitrate concentrations within the mixed layer.

Another possible source of error is in the timing of the mixed layer shoaling. It was noted in section 3.3 that the observations were held back by 10 days to allow a better match between the modelled and observed physical properties. In spite of this step (or perhaps because of it) the chlorophyll *a* residuals shown in figure 5.7 reveal that modelled chlorophyll *a* in the surface 100 m or so quickly changes from being too low to

too high and then back again. Such an error distribution is consistent with a mistimed shoaling of the mixed layer.

Modelled chlorophyll *a* during the observation period is also consistently too high below the mixed layer. This error can perhaps be attributed to a number of temporary stratification events that occur before the permanent shoaling of the mixed layer in late April. There are no observations available to check if these events did or did not occur, but the importance of these short-term stratification events in determining the bloom characteristics was noted by Waniek (2003). It is possible that the high chlorophyll *a* below the modelled mixed layer is detrained biomass from an earlier bloom event. Indeed, the average chlorophyll *a* within the mixed layer presented in figure 5.5 shows a number of fluctuations just prior to the main bloom at NABE, which are consistent with such transient phytoplankton growth.

The suggestion that many of the model errors at NABE are associated with the mistiming of a number of stratification events is consistent with the fact that these errors were not reduced with the addition of model complexity, as they were at BATS. Friedrichs et al. (2006) found that while an ecosystem model that was embedded in a physical scheme with a consistently too shallow mixed layer could be made to fit the data by adjusting model parameters, no amount of parameter tuning could make up for errors in the timing of the mixed layer shoaling (although the optimal parameters may well be affected - see figure 7.1 and discussion in chapter 7).

Despite the consistent errors among the five models calibrated at NABE, inverse estimates of the model fluxes were highly dependent on the model structure. Uncertainty was particularly high in the phytoplankton sinks of grazing and mortality, which can be related to the choice between the linear and quadratic mortality terms applied in the different models.

A linear mortality term was used in models one and three, and in these cases the optimal output at BATS suggested that phytoplankton losses occurred only as a consequence of grazing by zooplankton. The introduction of a quadratic mortality term in models two, four and five, resulted in a similar misfit cost, but with significant phytoplankton losses also occurring through mortality as well as grazing. The equivalence of these two solutions in terms of misfit cost suggests that the data were inadequate to distinguish between mortality and grazing as mechanisms by which the phytoplankton population is controlled at NABE. The fact that the five models were all very similar in terms of cost while providing often quite different estimates of the nitrogen fluxes suggests that more data will be needed to select an appropriate model for NABE.

A number of model fluxes (figure 5.17) were optimally zero, which suggests that these processes might not be important at NABE. For example, the fast-recycling parameter Φ_m^P was set to zero wherever it was applied. This suggests that this process can be ignored at NABE, but it must be remembered that the solutions were based on a very limited dataset, with observations confined to a 2 week period in April 2009. For the remaining 50 weeks of the year the model was completely unconstrained and there was no indication (from the assimilated data at least) how the models perform during this period. This problem is highlighted by the fact that alternative solutions for the losses of phytoplankton cannot be distinguished by the cost function.

In this study, the inclusion of extra ecological complexity only serves to increase uncertainty at NABE, but as described above, this does not necessarily mean that the extra model components are either incorrect or even that they are not needed. This is emphasised by the fact that losses among the phytoplankton are so poorly constrained. It is clear that the model estimates of the data at NABE are far from perfect, and that certain processes such as phytoplankton aggregation may be important with regard to the recycling and export of nutrients, but until such processes can be constrained by the data, it is impossible to say which processes should be included, and in what form.

In addition to this caveat based on the poor temporal coverage at NABE, it should also be noted that only a very limited number of ecosystem model configurations were evaluated in this study. The extra model components that were assessed were primarily included with the goal of improving estimates of oligotrophic primary production during the summer stratified period at BATS (section 3). In light of this, it is not surprising that they were much less effective with regard to improving estimates of the eutrophic spring bloom at NABE. If extra model complexity had been added with the specific aim of improving the representation of the bloom dynamics, for example by including state variables for silicate and siliceous diatoms, or different phytoplankton and zooplankton size-classes, then larger improvements might have been achieved at NABE.

5.4.3 Non-unique solutions

The existence of non-unique solutions means that there are uncertainties in the model output, especially with regard to unconstrained properties such as zooplankton, or to periods during which no data were available. These uncertainties are examined below for both sites.

The analysed model solutions at BATS are associated with low uncertainty because observations there are available at frequent and regular intervals, providing good coverage of the entire seasonal cycle. Although none of the models could provide a completely

satisfactory fit to the data, there were at least no major temporal or spatial gaps in the dataset. As a consequence the model output could not deviate far from the optimal trajectory without incurring a significant increase in cost.

Model output was not so universally well-constrained at NABE, where the data were confined to a very limited period during the spring bloom. The significance of this lack of data is most notable in figures 5.4, 5.7 and 5.13. In each figure the third column represents uncertainty in the optimal solutions, and in some cases this can be several orders of magnitude larger than observational error. Much of this uncertainty is located in the DCM between May and October, or in the Autumn bloom in November. These results suggest that any further observations at the NABE site should be focussed on these periods, as such data would go furthest to increasing confidence in modelled estimates of the NABE ecosystem.

5.5 Concluding remarks

By adding model complexity at BATS, model data misfit was reduced by up to 31%, while at NABE the "best" model represented only a 2% improvement over the simplest. The results presented in this chapter suggest that the additional model complexity included in models one to four was not required for modelling of the spring bloom at NABE, but was essential for improved estimates of the biogeochemical dynamics at BATS. This is not to say however, that the fast-recycling pathway, the variable Chl:N ratio and the DON pathway are not important at NABE during the summer, autumn and winter where the model output was not constrained by observations. The high model uncertainty during these periods, discussed in the preceding section, emphasises the importance of investigating this matter further.

In addition to the poor temporal coverage at NABE, the models that were included in this study were somewhat biased towards improving estimates of the summer primary production at BATS. Given this experimental bias, it is not surprising that the extra model complexity seemed to have a greater effect at BATS than it did at NABE. It would clearly be premature to say that the complexity included in model one is sufficient for reproducing the NABE ecosystem, as residual errors were still very high and a number of potentially important model pathways were not investigated. Fasham and Evans (1995), for example, found that a more complex model with 7 state variables could not be tuned satisfactorily to NABE data, with their representation of zooplankton data (not even included in this study) singled out as being particularly poor. Another study of phytoplankton growth conditions in a northern region of the North Atlantic (Ward

and Waniek, 2007) found evidence of seasonal succession among three different phytoplankton functional types as phytoplankton growth became increasingly light during the autumn and winter. The maintenance of phytoplankton species within the mixed layer provides the 'seed population' for the spring bloom (Backhaus et al., 1999), and it has been shown that the species composition prior to the bloom affects its initial growth rate (Waniek, 2003). These factors may be important in correctly reproducing the magnitude of the North Atlantic spring bloom, a feat that was not achieved here with any of the simple NPZD models. Although there is still considerable room for improvement in the performance of all five models at NABE, a significant proportion of the errors can probably be attributed to the physical forcing, particularly the timing and frequency of the bloom and pre-bloom shoaling events (Waniek, 2003).

At BATS the addition of model complexity consistently led to improvements in the model performance, but the representation of primary production (and export see figure 7.4 in chapter 7) was unsatisfactory throughout. Drawing on these results, and the knowledge that biogeochemical cycling at BATS seems to be dependent on the complex interaction of a number of ecological pathways and processes including the microbial loop (Fasham et al., 1990), DON recycling, nitrogen fixation and non-Redfieldian stoichiometry (Anderson and Pondaven, 2003), it seems that the ecological models applied here are too simple to accurately reproduce observations. Additionally, there are a number of physical processes not resolved in this one-dimensional study that may have a significant effect on model performance. These include the supply of both inorganic and organic nutrients by horizontal advection (Williams and Follows, 1998), and inorganic nutrients from depth by mesoscale activity (McGillicuddy and Robinson, 1997), although the significance of these processes is not yet clear.

The results presented in this chapter suggest that the levels of both physical and ecological complexity that were applied in this study were too low to accurately represent the BATS ecosystem. After parameter optimisation it is hard to distinguish between physical and ecological error, especially if only one physical framework is applied. If the modelling framework applied here was indeed too simple, then the process of parameter optimisation may well have been compensating for any inadequacies, with the model getting the right answer for the wrong reasons. The limits placed on the parameter search space should prevent unrealistic values from entering the solutions (see chapter 2), but is still possible that the optimisation process selected a number of *ad hoc* solutions that make no real ecological sense. For this reason it is important to assess the solutions in validation experiments, where the model solutions are examined with respect to independent data. This can be done by simply swapping parameter solutions between time-series sites (e.g. Friedrichs et al., 2007), in a process known as cross-validation. If the parameter values are sensible, and the modelled ecological components are consistent

between the sites, then the model will perform well with respect to independent data. If on the other hand the model solutions contain a number of *ad hoc* parameter values that are made up simply to compensate for some model inadequacy, then the model will perform badly under cross-validation. This concept is the basis of the next chapter, as BATS solutions are applied at NABE, and *vice versa*.

Chapter 6

Cross-validation

6.1 Introduction

An ideal model of the marine ecosystem will be able to accurately reproduce observations taken at any site, at any point in time. If such a model was calibrated at one site, the optimal parameters should be equally valid at another. In reality, models represent a simplification of the observed system, and calibrated parameters will be representative of the particular assimilated observations and any associated errors. A different site will have a different set of data and a new set of observational errors. Further to this, the ecosystem may function in a very different way than at the calibration site. For example results from the previous chapter suggest (in line with previous studies) that the primary production at BATS is strongly driven by the rapid recycling of nutrients, while at NABE the supply of new inorganic nitrogen from depth is more important.

By examining the performance of each calibrated model with regard to independent data from different sites, the ability of each model to cope with the diverse response of the ecosystem as a whole can be assessed. As stated above, the ideal model (with ideal data) will be able to model both sites perfectly, but back in the real world there will be a cost associated with simplifying a complex system. The cross-validation approach tests the generality of the models, which will be low for empirical models where many diverse aspects of the ecosystem are condensed into a few invariant parameterisations. More complex models that are based on mechanistic formulations should be more robust through time and space, as each model component has to account for a smaller number of diverse processes. By increasing model complexity to include processes that are less variable in time and space, models should prove to be more portable between different sites. There is however an obvious cost associated with increased complexity, which is the larger amount of data required to constrain the extra model parameters.

The optimal parameter vectors defined in chapter 5 show considerable variability without being associated with any significant changes in the misfit cost (e.g. figures 5.14 and 5.15). This insensitivity of the cost function occurs because, firstly, uncertainty about errors contained within the observations makes it difficult to distinguish between parameter vectors with similar costs (e.g. chapter 2), and secondly, (even assuming perfect observations) the available information may be insufficient to constrain a unique solution (e.g. chapter 4). This is either because changes in model parameters lead to changes in the model output that are not resolved by the data, or because the changes in the parameters lead to no changes in the model output.

In this chapter the parameter solutions developed in chapter 5 are applied in a cross-validation experiment. By using the entire range of parameters that led to an acceptable solution at the calibration site, an uncertainty estimate can be included in the results at

the cross-validation site. In this way the distribution of the costs at the validation site can be used to assess the model performance, accounting not only for the benefits associated with resolving extra complexity, but also the costs associated with the addition of perhaps unconstrained model parameters.

6.2 Methods

In chapter 5 the μGA was applied for each of the five models to find many "optimal" parameter vectors for BATS and NABE. In this chapter those solutions are applied in a cross-validation experiment, where the parameters calibrated for BATS data are applied at NABE, and *vice versa*. For the sake of clarity the model-data misfits yielded by the cross-validation experiment are labelled the "predictive costs", after Friedrichs et al. (2006). The minimum costs found by the μGA in chapter 5 are referred to as "calibrated costs", while the costs associated with the default parameters are labelled the "default costs".

6.3 Results

6.3.1 Predictive costs

The calibration process yielded many alternative parameter vectors that could not be distinguished in terms of the model misfit to the assimilated data. When these parameter vectors were applied in an effort to reproduce unassimilated data however, the costs were highly variable. The wide distributions of these predictive costs, shown for each site in figure 6.1, can be attributed to changes in the parameter values that, while not significantly affecting the misfit cost at the calibration site, lead to significant changes in cost when applied to independent data. The variability of the predictive costs will be discussed in further detail later in the text, but for the sake of a quick model comparison the predictive costs for each model are summarised by their median values. These are presented in table 6.1. The misfit costs associated with the default parameters and the optimised parameters are also shown.

The calibrated costs should represent the minimum cost that can be achieved by each model for a given dataset. When the NABE solutions were applied at BATS, the median predictive costs were between 38 and 117% higher than this baseline. When the BATS solutions were applied at NABE, the predictive costs were between 61 and 322% higher than the calibrated costs at the same site.

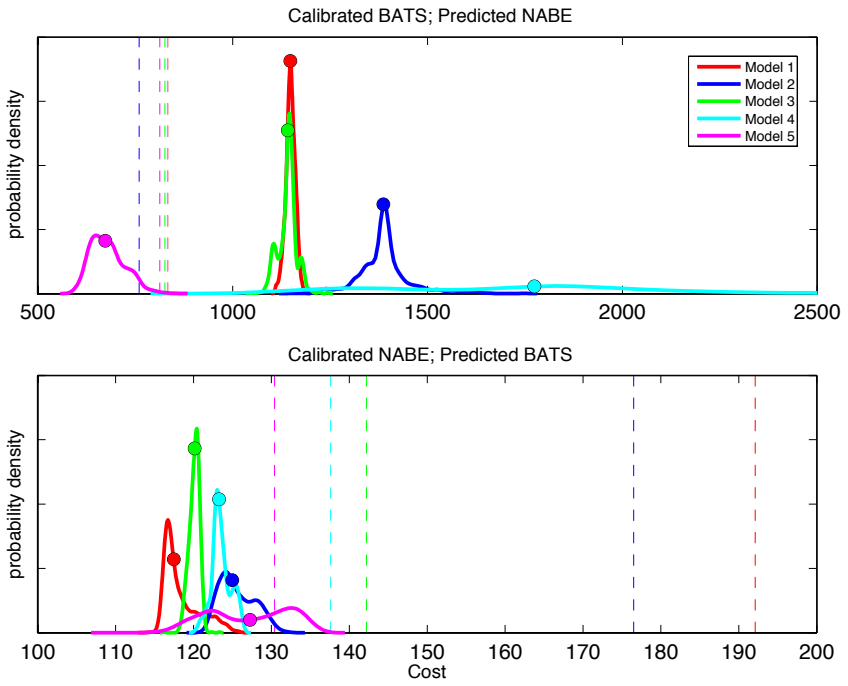

FIGURE 6.1: Distribution of the predictive costs for each of the five ecosystem models, at BATS and NABE. The median of each distribution is marked with a circle. The cost associated with the default parameters at the validation site are shown as dashed lines.

The parameters calibrated to NABE data provide a better fit to the data at BATS than the default parameters. The median predictive costs for models one to five were between 2 and 39% lower than the corresponding default costs. The parameters calibrated to BATS data generally performed worse than the default parameters when applied at NABE, with the exception of model five. The median predictive costs for models one to four were between 38 and 113% higher than the respective default costs at NABE. The median predictive cost for model five was 17% lower than the default cost.

The predictive costs are broken down according to contributions from different data types in table 6.2. When the NABE solutions were applied at BATS, the predictive costs were dominated by chlorophyll *a* and primary production misfits, as they were for the calibrated solutions (table 5.2). At NABE the predictive costs based on the BATS parameters were dominated by chlorophyll *a* misfits, as they were for the calibrated solutions. The model performance was worse for all data types when compared to the calibrated solutions.

There was no consistent relationship between model complexity and either the magnitude or the variability of the predictive costs at either site. Predictive costs for model four were highly variable at NABE, with a range of 4757. Given this very high uncertainty, it is perhaps surprising that model five, which is essentially the same as model four but

TABLE 6.1: Predictive costs at BATS and NABE. Costs from the default and optimised parameter vectors were included for comparison.

BATS	Default	Calibrated	Predictive
Model 1	192.1	84.5	116.8
Model 2	176.5	77.3	125.1
Model 3	142.2	76.9	120.2
Model 4	137.6	66.6	123.3
Model 5	130.4	58.6	127.4
NABE	Default	Calibrated	Predictive
Model 1	833.5	425.8	1148.9
Model 2	760.0	421.6	1388.1
Model 3	826.1	430.3	1143.8
Model 4	832.7	420.4	1775.9
Model 5	813.0	417.9	673.4

TABLE 6.2: Predictive costs at BATS and NABE, broken down according to data type.

BATS	DIN	Chl a	Zoo	PON	CUp	Total
Model 1	13.6	46.9	0.2	5.0	51.0	116.8
Model 2	13.6	52.6	0.5	5.2	53.1	125.1
Model 3	10.1	51.9	0.6	6.1	50.9	120.2
Model 4	10.6	56.2	0.3	5.3	50.1	123.3
Model 5	11.9	52.0	6.3	9.8	47.5	127.4
Mean	12.0	51.9	1.6	6.3	50.5	122.6
Range	3.5	9.3	6.1	4.8	5.6	10.6

NABE	DIN	Chl a	Zoo	PON	CUp	Total
Model 1	128.8	900.1	-	76.1	43.9	1148.9
Model 2	74.2	1068.6	-	75.0	170.1	1388.1
Model 3	130.6	880.1	-	76.6	56.5	1143.8
Model 4	60.3	1495.7	-	68.9	151.0	1775.9
Model 5	87.5	473.8	-	71.7	40.4	673.4
Mean	96.3	963.7	-	73.7	92.4	1226.0
Range	70.3	1021.9	-	8.0	129.7	1102.5

with two additional unknown parameters, had the lowest predictive costs at BATS, with the range of costs reduced by a factor of more than 18. Model five had the most variable predictive costs at NABE.

The predictive costs were much more variable between the models at NABE than at BATS, in absolute and relative terms. Normalising to the mean of the predictive costs at each site reveals that the range of predictive costs represent 89.9% of the mean at

NABE, and only 8.6% at BATS. In other words, the parameters developed at NABE perform much more consistently in the cross-validation experiment than the parameters developed at BATS. This reflects the fact that the calibrated costs at NABE were (relatively) more consistent than at BATS (table 6.1).

6.3.2 Predictive model output

Dissolved inorganic nitrogen (DIN)

Modelled mean DIN concentrations produced in the cross-validation experiment are shown in figure 6.2. The plots show the model output associated with the median predictive cost (red), as well as the range of solutions yielded by the cross-validated parameters (black).

At BATS the cross-validated DIN trajectories were considerably lower than the calibrated solutions shown in figure 5.2. The peak DIN concentrations in the mixed layer were strongly underestimated, and the predictive DIN costs were on average 94% higher than in the calibrated solutions. There was relatively little variability between the five models in terms if the predictive DIN costs, with the range of these figures corresponding to 29% of the mean.

The predictive DIN trajectories at NABE were slightly higher than the calibrated solutions shown in figure 5.2, with the peak DIN observations (underestimated in the calibration experiments) matched by all five models. Regardless of this improvement, the predictive DIN costs at NABE were on average 258% higher than the calibrated solutions at the same site. This was a much larger relative deterioration in model performance than was seen at BATS. The relative variability of the predictive DIN costs was also much higher at NABE than at BATS, with the range across the five models equivalent to 73% of the mean. This variability was confirmed through visual inspection of figure 6.2, where the NABE trajectories show a number of differences. Most notably, model three shows very high variability in the summer and autumn DIN concentrations, while model four shows considerable variability during the observation period. Although predictive outputs from model three were clearly highly variable for most of the year, model outputs corresponding to the observations were much less so. As a result the predictive costs given in table 6.1 were relatively consistent. By contrast, the output from model four was much less variable for most of the year, but the high variability seen during the observation period meant that the associated predictive costs were highly variable.

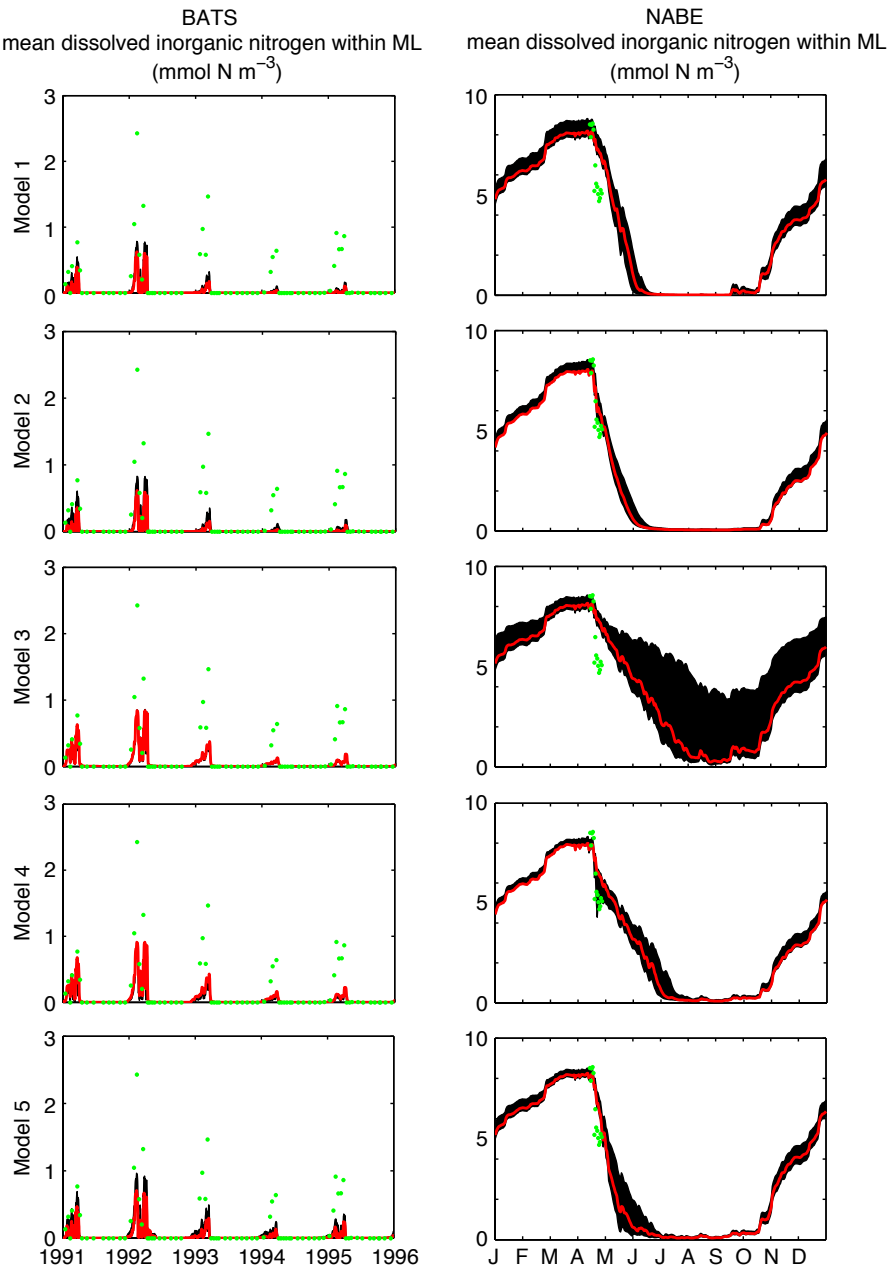

FIGURE 6.2: Dissolved inorganic nitrogen (DIN) concentrations at BATS and NABE, averaged over the mixed layer. Red lines represent the trajectory of one solution associated with the median cost, while the filled black areas cover the range of all solutions (see chapter 5). Green dots represent observations, interpolated onto a vertical grid and similarly averaged over the modelled mixed layer. Note different scales on vertical axes between stations.

Chlorophyll *a*

Cross-validated chlorophyll *a* profiles are shown in figure 6.3. The chlorophyll *a* concentrations at BATS reflect the diminishing levels of DIN within the mixed layer that are shown in figure 6.2. By 1994 and 1995, models two to four underestimate the levels of chlorophyll *a* throughout the entire year. There are some differences apparent between the different models, with model one showing the most realistic chlorophyll *a* trajectories. This was reflected in the cross-validated chlorophyll *a* misfits, which were lowest at BATS for model one.

At NABE the predictive costs were much more variable than at BATS. The range across all five models was 106% of the mean at NABE, but only 18% at BATS. Visual inspection also suggests that modelled chlorophyll *a* trajectories were more variable between the five models at NABE, with the most of the variability associated with the period from the autumn bloom to just before the spring bloom. As with DIN, model four shows considerable uncertainty during the bloom period, when all of the observations were collected.

The predictive costs at NABE deteriorate more relative to the calibrated costs than they do at BATS. Predictive chlorophyll *a* misfits at BATS were on average 80% higher than the calibrated costs. At NABE this figure was 310%.

Zooplankton

Mean zooplankton concentrations from the cross-validation experiment are shown in figure 6.4. Zooplankton data were only available at BATS, and at this site the misfits were very low as a consequence of the low weight applied in the cost function. Surprisingly, in terms of model-data misfits, the parameters developed at NABE led to a better representation of the observed zooplankton at BATS than parameters that were actually calibrated to those data. This fact should however be placed in the context of the multivariate optimisation process, where misfits to some data may be allowed to increase, as long as the total misfit is not made larger.

Zooplankton biomass in the cross-validation experiment was more variable at BATS during the years where no data were available, although clearly this has nothing to do with those data, as the solutions were derived at NABE. The drop in variability can however be related to the diminished DIN and chlorophyll *a* concentrations seen during these years in the previous two figures. At NABE the zooplankton concentrations appear to be more variable between the different models than at BATS.

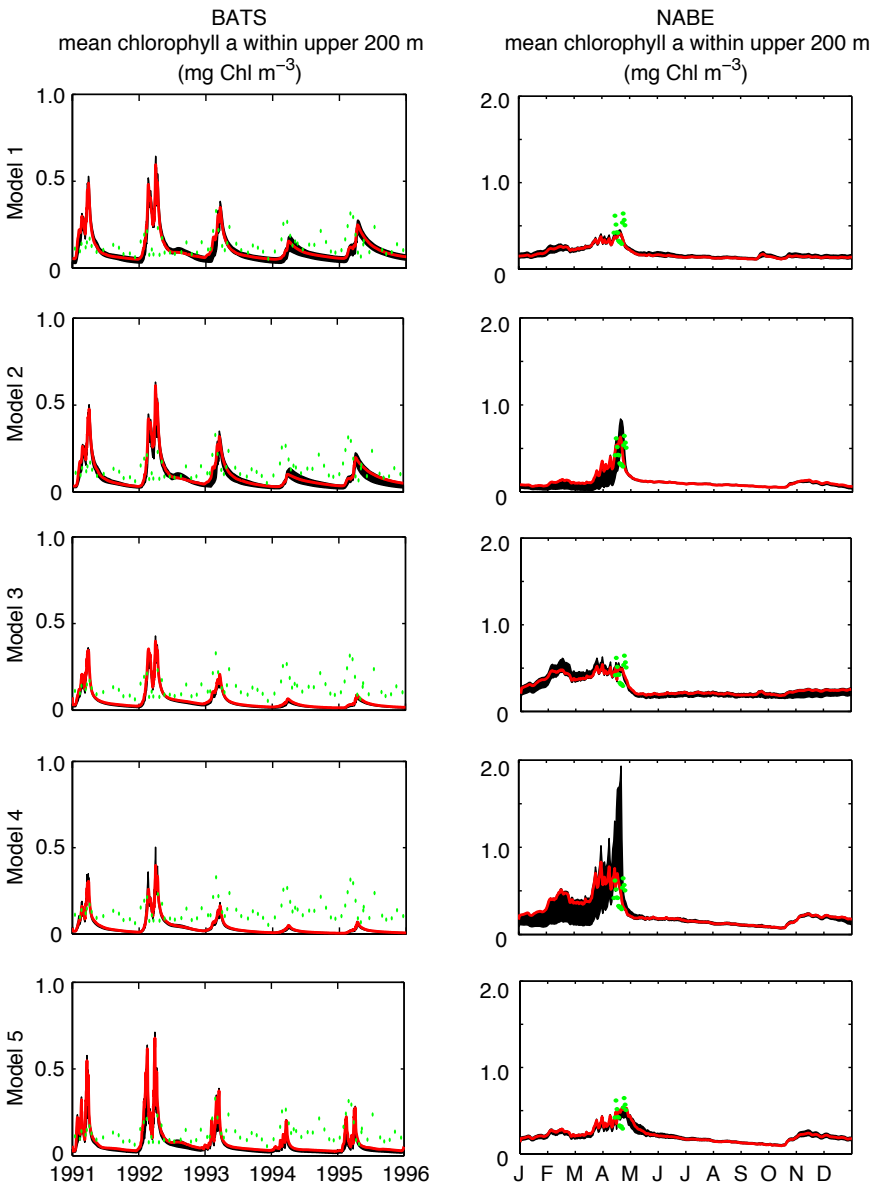

FIGURE 6.3: Chlorophyll *a* concentrations at BATS and NABE, averaged over the euphotic zone (200 m). Red lines represent the trajectory of one solution associated with the median cost, while the filled black areas cover the range of all solutions (see chapter 5). Green dots represent observations, interpolated onto a vertical grid and similarly averaged over the modelled mixed layer. Note different scales on vertical axes between stations.

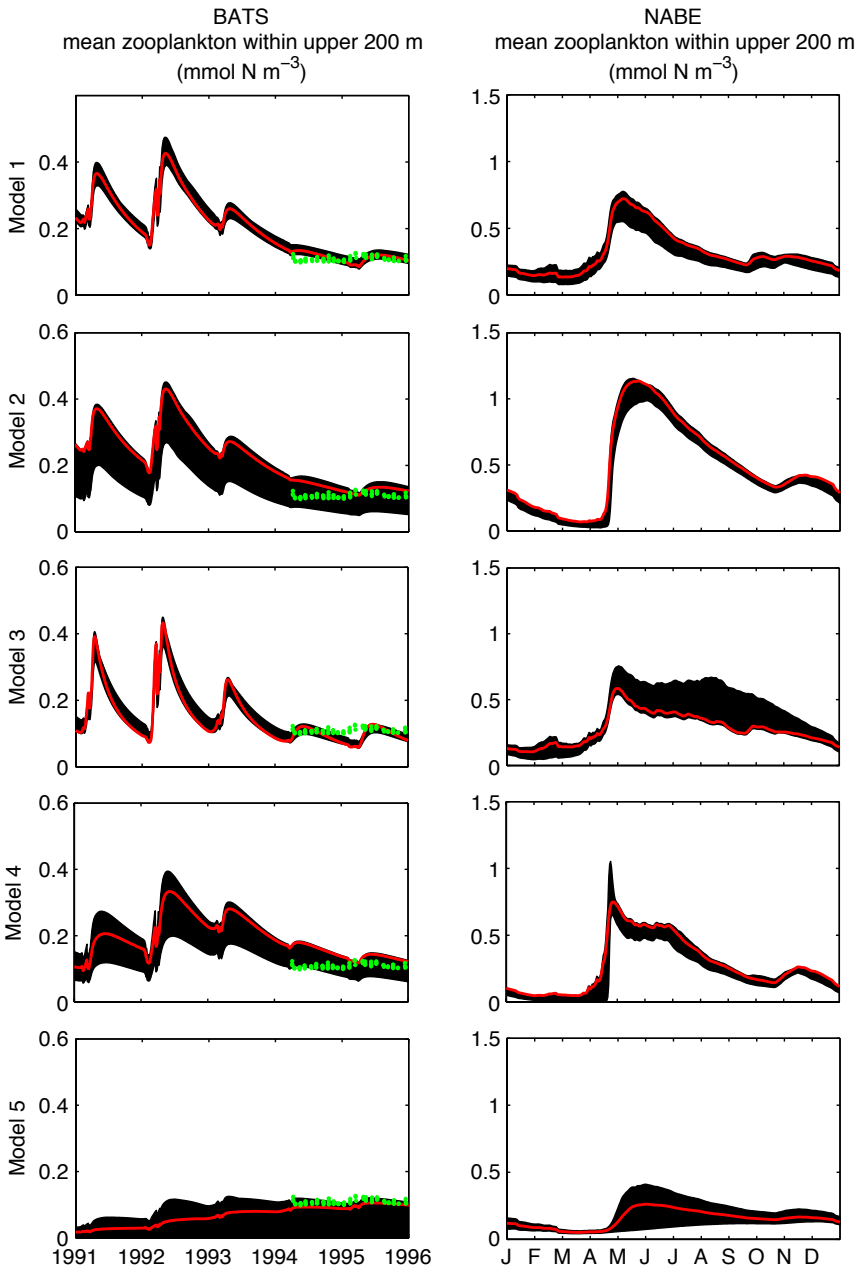

FIGURE 6.4: Zooplankton concentrations at BATS and NABE, averaged over the euphotic zone (200 m). Red lines represent the trajectory of one solution associated with the median cost, while the filled black areas cover the range of all solutions (see chapter 5). Green dots represent observations, interpolated onto a vertical grid and similarly averaged over the modelled mixed layer. Note different scales on vertical axes between stations.

Dissolved organic nitrogen (DON)

Modelled DON concentrations from the cross-validation experiment are shown in figure 6.5. At BATS the cross-validated DON profile was nearer to observations than the calibrated solution, although it was also associated with greater uncertainty. At NABE the cross-validated solutions were less uncertain than the calibrated solutions. There was also a stronger seasonal cycle, with DON concentrations diminishing over the winter and increasing in the post-bloom period.

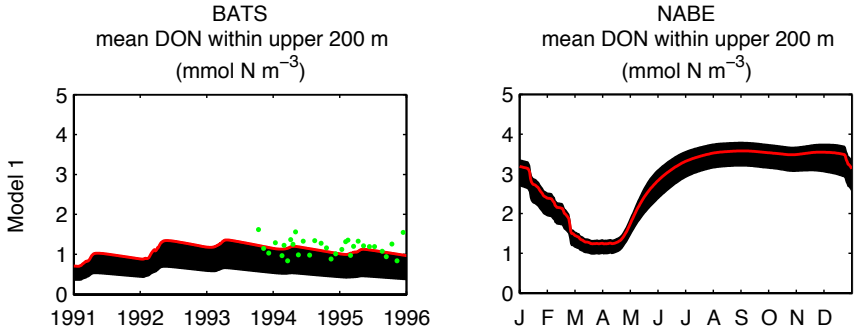

FIGURE 6.5: Dissolved organic nitrogen (DON) concentrations at BATS and NABE, averaged over the euphotic zone (200 m). Red lines represent the trajectory of one solution associated with the median cost, while the filled black areas cover the range of all solutions (see chapter 5). Green dots represent observations, interpolated onto a vertical grid and similarly averaged over the modelled mixed layer. Note different scales on vertical axes between stations.

Primary production

Figure 6.6 shows primary production within the surface 200 m from the cross-validation experiments at BATS and NABE. The cross-validated models at BATS underestimate summer primary production more strongly than the calibrated models, and there was no noticeable improvement as complexity was increased. This lack of improvement at BATS was reflected in the low variability of the cross-validated costs at BATS; 11% of the mean across the five models. At NABE the relative variability was much higher, at 140%. On average, the cross-validated estimates of primary production produce misfits 82% higher than the calibrated solutions at BATS, and 313% higher at NABE.

As in the calibration experiments, the cross-validated mean primary production at NABE was lowest in models without the fast-recycling pathway (models one and three). When this pathway was included and calibrated at BATS, the resultant primary production at NABE was too high. The very high uncertainty seen in the model four for the DIN and chlorophyll *a* profiles was also present in the primary production output.

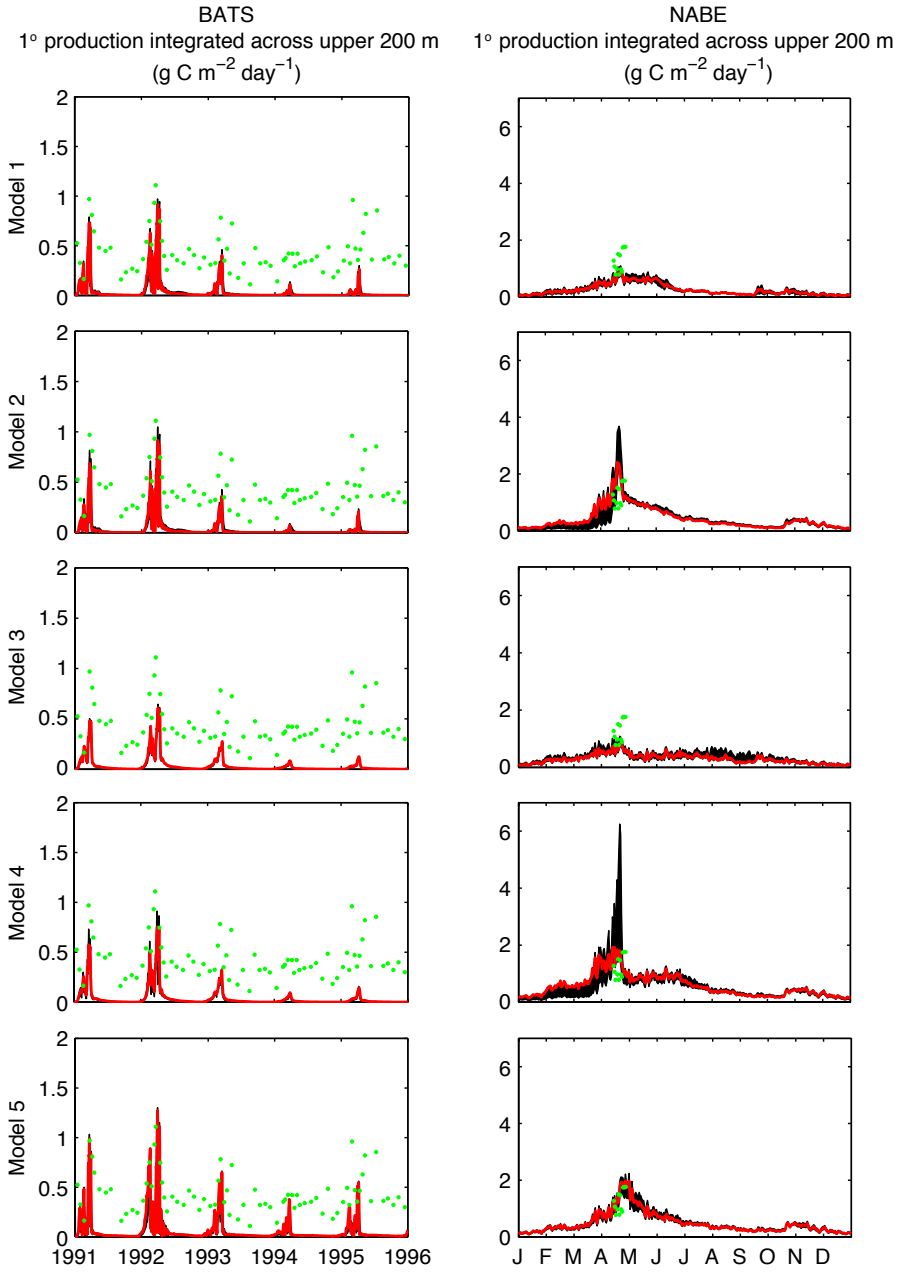

FIGURE 6.6: Primary production concentrations at BATS and NABE, integrated across the euphotic zone (200 m). Red lines represent the trajectory of one solution associated with the median cost, while the filled black areas cover the range of all solutions (see chapter 5). Green dots represent observations, interpolated onto a vertical grid and similarly averaged over the modelled mixed layer. Note different scales on vertical axes between stations.

6.3.3 Predictive nitrogen fluxes

Modelled nitrogen fluxes from the cross-validation experiments are shown in figures 6.7 and 6.8. Visual inspection suggests that the cross-validated fluxes at BATS were generally much weaker than the calibrated solutions given in figure 5.16. At NABE the cross-validated fluxes for model one were roughly comparable to the calibrated solutions shown in figure 5.17, but get disproportionately large in the more complex models.

Cross-validated estimates of annual primary production at BATS were reduced to between 13 and 45% of the calibrated solutions. At NABE they correspond to between 88 and 183% of the calibrated values. Estimates were much more consistent between models at BATS than they were at NABE, although the uncertainties surrounding the phytoplankton losses through mortality and grazing (described in the previous chapter) are reproduced in the cross-validation at BATS. At BATS the range of the cross-validated estimates corresponds to 57% of the mean value. At NABE this figure increases to 85%. In absolute terms, the range of cross-validated estimates of annual primary production was an order of magnitude higher at NABE than at BATS.

The high variability of primary production estimates at NABE, contrasted with the consistent solutions at BATS, mirrors the pattern seen in the calibrated solutions given in chapter 5. Figures 5.16 and 5.17 show the calibrated solutions were variable at BATS and more consistent at NABE. When the models were calibrated to NABE data, the solutions did not change a great deal as model complexity was added. At BATS however, the initially very low estimates of primary production consistently improved as new model components were added (up to model four). When the optimal parameter solutions were swapped between sites, this pattern was reversed. The NABE solutions applied at BATS lead to relatively consistent (but low) estimates of primary production across the five models. When the BATS solutions were applied at NABE, the modelled primary production was much more variable, becoming larger as model complexity increases (up to model four).

If it is assumed that the calibrated solutions with the lowest primary production misfits represent the best estimates of annual primary production, the best model estimates were 1278±45 mmol N m^{-3} $year^{-1}$ at BATS, and 1359±30 mmol N m^{-3} $year^{-1}$ at NABE. The cross-validated estimates at BATS were much lower than the best estimate, between 143 and 241 mmol N m^{-3} $year^{-1}$. At NABE, the cross-validated estimate from model one was a relatively low 1176±12 mmol N m^{-3} $year^{-1}$, but in the more complex models the estimates were all significantly larger than the best estimate, between 1501±121 and 2351±74 mmol N m^{-3} $year^{-1}$.

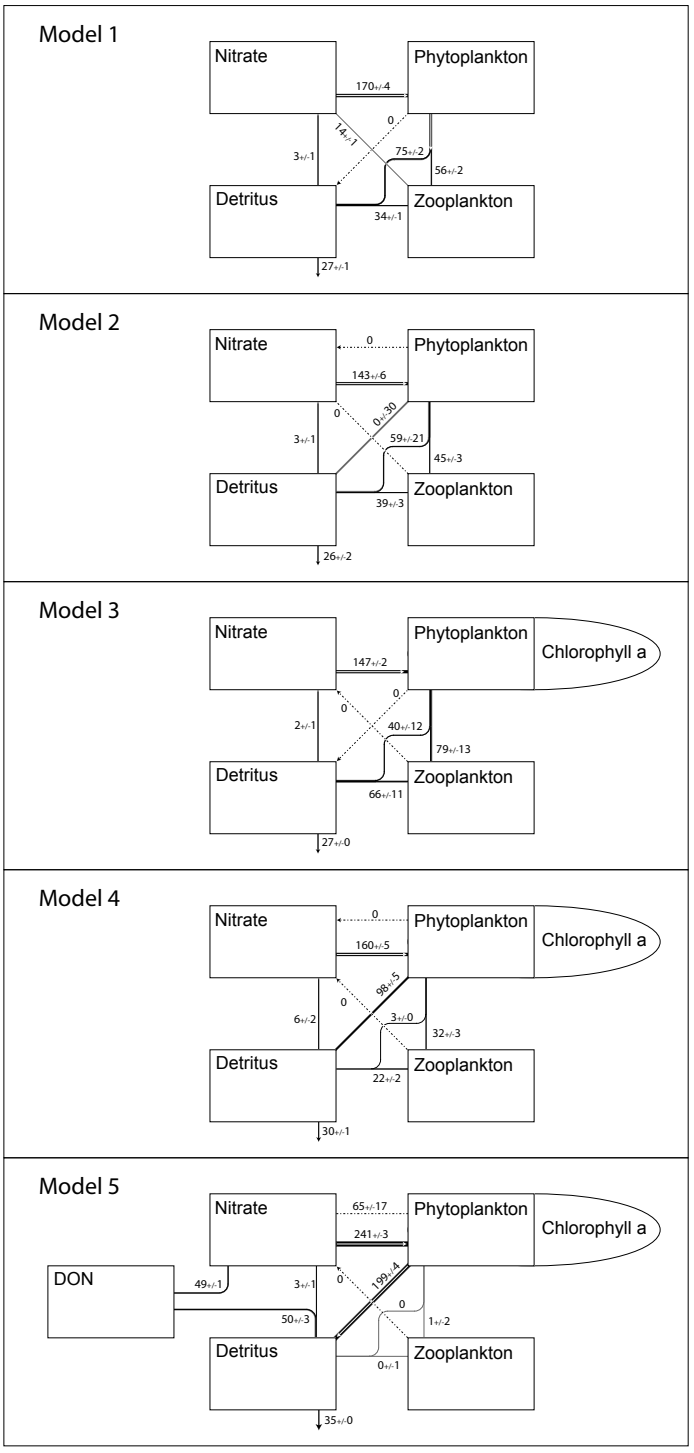

FIGURE 6.7: Predictive nitrogen fluxes at BATS, based on the NABE calibrated parameters. Fluxes were shown for all the acceptable parameter solutions of each model. The widths of flux pathways were linearly scaled to indicate the magnitude of annual mean fluxes between state variables. Black flux pathways indicate the minimum flux associated with an acceptable solution, while grey lines indicate the maximum such flux. Figures next to the pathways correspond to these median and and MAD of the solutions. Dotted lines indicate where the maximum optimised flux was zero.

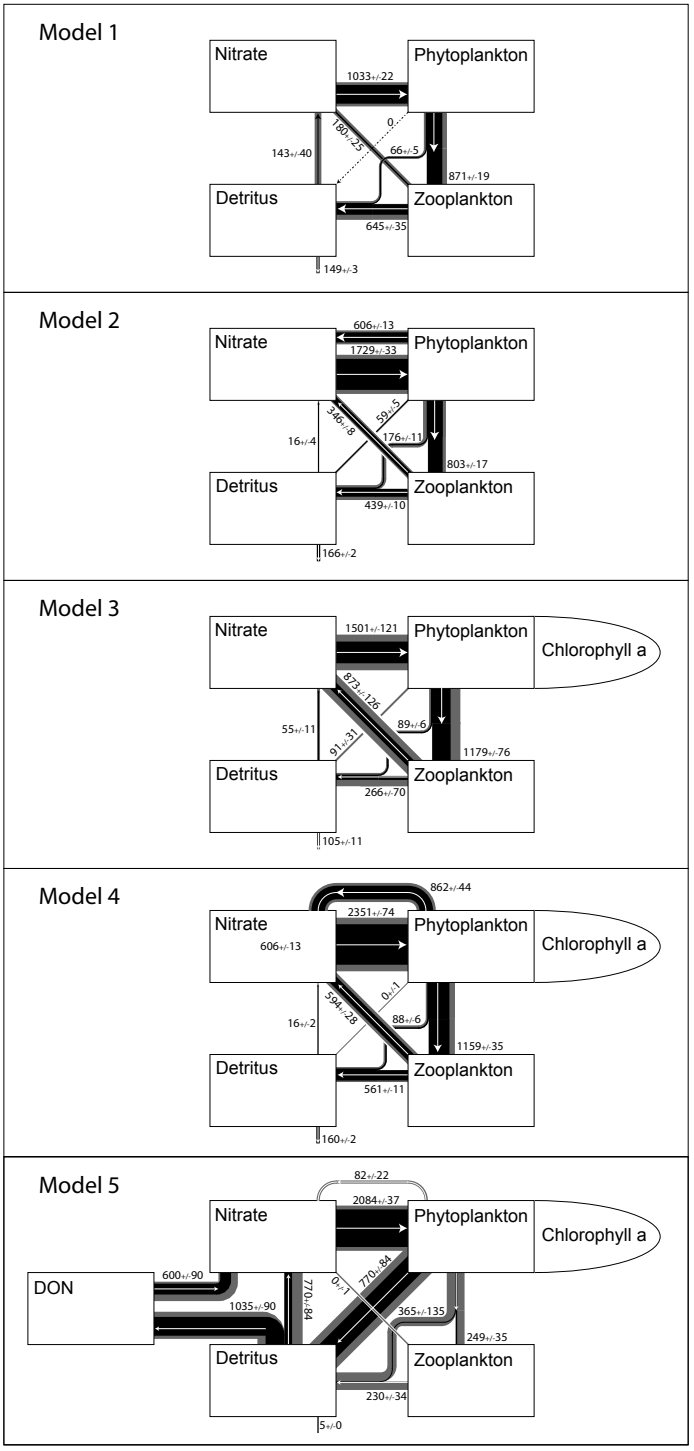

FIGURE 6.8: Predictive nitrogen fluxes at NABE, based on the BATS calibrated parameters. See description of figure 5.16 for details.

When the models were calibrated, adding complexity at BATS allowed the model solutions to produce larger, more realistic estimates of the primary production. When these solutions were applied at NABE, the more complex models again yielded higher estimates of primary production, but in this case they were probably too high. When the models were calibrated to NABE data, the estimated primary production was already relatively high in model one, and adding complexity did little to increase this value. When these solutions were applied at BATS, the estimated primary production remained fairly constant between the models, and severely underestimated the best estimate in all five cases.

6.4 Discussion

In this chapter the calibrated parameter solutions were examined in a cross-validation experiment. By applying the optimal parameters developed at one site in an effort to recreate observations at the other, the generality of the optimised models could be assessed. In other words, to what degree were the solutions outlined in chapter 5 universal to both the oligotrophic and eutrophic sites? The models will be general if they were based on equations that can account for the processes at both sites, using the same parameter values. If on the other hand the parameter values were highly specific to each site, and this will occur if the values were assigned to compensate for any model inadequacy (physical or ecological), then the solutions will not be generally applicable at different sites.

When the calibrated solutions were applied in the cross-validation, the NABE derived solutions performed well in relation to the default parameters, but the BATS derived solutions did not (figure 6.1). This result is consistent with the suggestion made in chapter 5 that the model framework was inadequate at BATS. If the optimal parameters serve to compensate for model inadequacies, rather than to describe processes that are universal in the open ocean, it is not surprising that they should perform very badly when applied at NABE.

6.4.1 NABE solutions applied at BATS

In its simplest form, inverse theory allows certain aspects of the marine ecosystem that are perhaps difficult to observe, and these include ecological parameters, to be estimated by fitting a model of the system to data. Estimates derived in this way are subject to error from a number of sources, including observational error, errors in the physical forcing, and model inadequacy. Optimal estimates of the model parameters will be partly

set to compensate for these errors, yet despite this, the solutions developed at NABE performed better than the default parameters when they were reapplied at BATS. Any compensatory errors carried from the optimisation framework through to the parameter solutions were not large enough to overwhelm any general improvements in the model parameters brought about by optimisation.

Out of the five models the simplest, model one, performed the best when cross-validated at BATS. In the limited context of this experiment, it was the most general of the model solutions, and adding model complexity just served to reduce this generality and increase uncertainty. The other model without the fast-recycling pathway, model three, also performed well in the cross-validation. The equivalent models that included fast recycling, models two and four, performed slightly worse. Of all the models calibrated at NABE the most complex, model five performed the worst in the cross-validation, and was subject to the highest predictive cost uncertainty at BATS (figure 6.1).

None of the solutions developed at NABE were so overfit that they could not outperform the default parameters at BATS, and that should be interpreted as evidence that the models were well suited to reproducing the NABE observations. It is important to bear in mind however, that when the optimised parameters were compared to the default parameters, it was not a very stern test of the actual model performance at BATS, as the models were compared only with themselves. It is however revealing that the calibrated parameters did not perform worse than the default parameters, because if it is accepted that calibrating the parameters of a flawed model gives the right answer (in terms of fit) for the wrong reasons (in terms of parameter values), the fact that the calibration process did not introduce large errors into the parameters suggests that the models calibrated at NABE were getting close to the right answer for the right reasons.

Regardless of this, it was shown in chapter 5 that none of the models could accurately reproduce observations at BATS, and that fact remains here. The problem of underestimated summer primary production seen in the calibrated solutions was even worse in the cross-validated solutions, and winter DIN concentrations were also too low. When the models were calibrated to BATS data, there was an imperative in the cost function to minimise these errors, but clearly there was not when NABE data were assimilated.

6.4.2 BATS solutions applied at NABE

The solutions defined at BATS did less well in the cross-validation experiment, with predictive costs from models one to four far in excess of the misfits yielded by the default parameters (table 6.1). Only in model five did the BATS calibrated solutions perform better than the default parameters at NABE. If the conclusions drawn in chapter 5

were correct, and the modelling framework applied here was inadequate for reproducing observations at BATS, then there was a good chance that many of the parameter values assigned in the calibration experiments will be in some ways compensating for that inadequacy. If the optimal parameters were assigned in order to mask errors in the models, rather than on the basis of some true ecological process, then this would explain the high predictive costs seen when the solutions were applied at NABE. Using the same analogy as was outlined in the previous section, the models were brought closer to the right (cost) answer, but for the wrong (parameter) reasons. When the calibrated parameters were applied against independent data at NABE, the flaws they contained were exposed.

The fast-recycling pathway included in models two, four and five can be described as *ad hoc*, as it was applied specifically for the purpose of increasing primary production in the oligotrophic gyre (Oschlies, 2001). Although a more realistic representation would include fluxes through DON, bacteria and ammonium state variables (Fasham et al., 1990), the microbial loop was instead condensed into a single linear function, controlled by the phytoplankton linear mortality parameter (Φ_m^P). The calibrated fluxes shown in figure 5.16 reveal that the fast-recycling pathway was most active at BATS in models two and four (nitrogen in model five was recycled primarily through an unrealistically high quadratic mortality of phytoplankton to detritus). In these two models at BATS the optimal values of Φ_m^P were relatively high in comparison to the default value of 0.01 d^{-1}. The high values of Φ_m^P, in conjunction with relatively high estimates of the initial slope of the PI curve α, allowed more primary production as nitrogen was rapidly recycled between DIN and phytoplankton.

When the BATS solutions were applied at NABE (where Φ_m^P was optimally zero - figure 5.15), the high values of Φ_m^P caused a strong (and most likely unrealistic) flux from phytoplankton back to DIN, and primary production was too high. This error was attributed to the fact that the microbial loop was strong at BATS, but not at NABE. By using such a simple parameterisation of the microbial loop, when the parameters were optimised at BATS, the fast recycling was hard-wired into the models. When these solutions were applied at NABE, the microbial loop was too strong, and models two and four performed the worst of all the five models. To compound this error, the value of Φ_m^P may well have been compensating for the lack of eddy-pumping, lateral advection, nitrogen fixation and carbon overconsumption in the model framework (see chapter 5). When this parameter was applied at NABE, where these processes were not important (especially not during the observed bloom period), it was not surprising that primary production was overestimated.

Adding the variable Chl:N ratio in model three does very little to change model predictive skill at NABE relative to model one. This was understandable given that there were no extra free parameters assigned with this function. When the same function was assigned in model four however, there was a massive increase in the predictive cost uncertainty relative to model two. This behaviour was perhaps harder to explain, but it suggests that the predictive costs and their uncertainty cannot be easily predicted as a function of the number of free parameters. They will rather be dependent on the finer details of the model structure and the non-linear way it responds to environmental forcing. Regardless of this however, the results presented in figures 6.2, 6.3 and 6.6 suggest the raw cost function values should be treated with caution at NABE. The high predictive cost uncertainty seen at NABE in model four was attributable to a very isolated patch of variability that happens to coincide with the NABE observations. By contrast, although uncertainty in model three was also very high, it was restricted to the period after the spring bloom where there were no observations. Whatever the reason for these different uncertainty distributions (they were somehow attributable to the fast-recycling pathway that was applied in model four but not in model three), it was clear that the data at NABE were not adequate to capture the true range of model performance at that site. The costs and benefits of the variable Chl:N ratio are thus very hard to disentangle.

The addition of the DON pathway in model five led to the best cross-validated model performance at BATS. It was the only one of the five models for which the calibrated solutions developed at BATS performed better than the default parameters at NABE. This could be interpreted as evidence that model five was the only model with sufficient complexity to produce sensible and general solutions when calibrated at BATS. In terms of the calibrated solutions, the fluxes in model five look fairly similar at BATS and NABE (figures 5.16 and 5.17). In both cases there was a strong primary production flux, with most of the phytoplankton biomass passing directly to detritus through the quadratic mortality function. Fluxes through the zooplankton compartment were relatively weak. Detrital biomass was strongly remineralised at both sites, with a significant proportion passing through the slower DON pathway. When the calibrated BATS solutions were applied at NABE, the pattern remained broadly the same, albeit with much greater uncertainty. While the simplified production of DON by detrital breakdown should be treated with caution, as it is in reality a much more complex process (Anderson and Williams, 1998; Christian et al., 2002), it appears that model five was more appropriate than any of the other models for reproducing observations at BATS, and the solutions derived there were not unreasonable when applied to NABE.

The optimised fluxes for model five were similar at both sites, and calibrating one does not compromise the performance at the other, at least in comparison to the default parameters. If this is framed in the same context as the discussion in section 6.4.1,

then it should be concluded that model five represents a considerable improvement over the other models at BATS. It should be remembered however, that the cross-validation experiment applied here was limited, especially as the data collected at NABE cover such a short period of time. Just as a limited data set will not be able to exclude certain solutions in a calibration exercise, the same dataset would similarly not be able to penalise those solutions in a cross-validation. If the data do not exist, it is impossible to falsify related model output. It is also helpful to remember the poor performance of all five models in chapter 5, where the summer primary production could not be reproduced even after calibration. Although model five performed best in terms of the cross-validation, it was still a long way short of reproducing the complex ecological processes observed at BATS.

6.5 Calibration and cross-validation at oligotrophic and eutrophic sites

The asymmetry in predictive costs as the models were cross-validated at BATS and NABE may also be interpreted in terms of the amount of information contained in the assimilated data. Fennel et al. (2001) noted that the weak nutrient supply at BATS, and the consequent low surface DIN concentrations and phytoplankton biomass, meant that models calibrated to fit data there were rarely saturated in terms of nutrient uptake or grazing. This lack of a saturating response resulted in parameters such as the maximum growth rate and maximum grazing rate being particularly poorly constrained. This was exactly the situation seen in the previous chapter, where the growth rate parameter μ and the maximum grazing rate g were mostly well constrained at NABE (figure 5.15), but poorly constrained at BATS (figure 5.14). This concept is illustrated in figure 6.9, which shows the optimal estimates of the phytoplankton growth and zooplankton grazing responses given for model one.

The light-saturated growth rate for all models is given by

$$\mu = \mu_m C_{\text{ref}}{}^{cT} \frac{N}{k_N + N} \tag{6.1}$$

while the zooplankton grazing response is given by

$$G = g \frac{\epsilon P^2}{g + \epsilon P^2} \tag{6.2}$$

Given that C_{ref} and c are constants (table 3.3), and assuming a constant temperature of T=20°C, the upper panel in figure 6.9 shows the optimal phytoplankton light-saturated growth response to nutrients, dependent on the covarying optimal values of μ_m and k_N from model one. The lower panel shows the zooplankton grazing response to model phytoplankton biomass, as a function of the optimal values of g and ϵ. For each of the two sites, BATS and NABE, the coloured lines represent the observed range of surface DIN and phytoplankton biomass (excluding the upper 1%). The figure demonstrates that at BATS, the saturated part of the curves are very poorly constrained, but this does not significantly affect the misfit cost because the very low standing stocks at this site mean that the upper parts of these curves very rarely affect the model performance. At NABE, surface DIN concentrations and phytoplankton biomass are much higher, and hence the saturated part of the curves are much better constrained.

The differences in the estimated responses at the two sites becomes important as the optimal parameters are switched between sites in the cross-validation experiment. Within the range of DIN and phytoplankton seen at BATS, there is very little difference between the growth and grazing responses, and hence the NABE solutions perform reasonably well at BATS. The situation is very different when the BATS solutions are applied at NABE. At the higher DIN concentrations and phytoplankton biomasses seen at this eutrophic site, the curves are very different, leading to the observed very high misfits.

This issue is highlighted by the contrasting fluxes shown in figures 6.7 and 6.8. As the NABE solutions are applied at BATS, the weak nutrient supply at BATS means that errors in the parameters are superimposed upon generally weak fluxes and low standing stocks, which may be within the range of observational errors. By contrast, as the solutions are applied at NABE, the errors in the optimal parameters may be magnified by the strong fluxes and high standing stocks at this site, leading to very high misfit costs.

6.6 Conclusions

The model results here support the tentative conclusions that were drawn in the previous chapter, namely that the five models seem to be better suited for modelling the ecosystem at NABE, than the one at BATS. These conclusions were arrived at because the models calibrated at NABE seemed to explain the data in a similar and consistent way, with additional (possibly superfluous) complexity making little difference. This was not the case at BATS, where although the additional complexity led to large improvements, the solutions were less consistent in their explanation of the data.

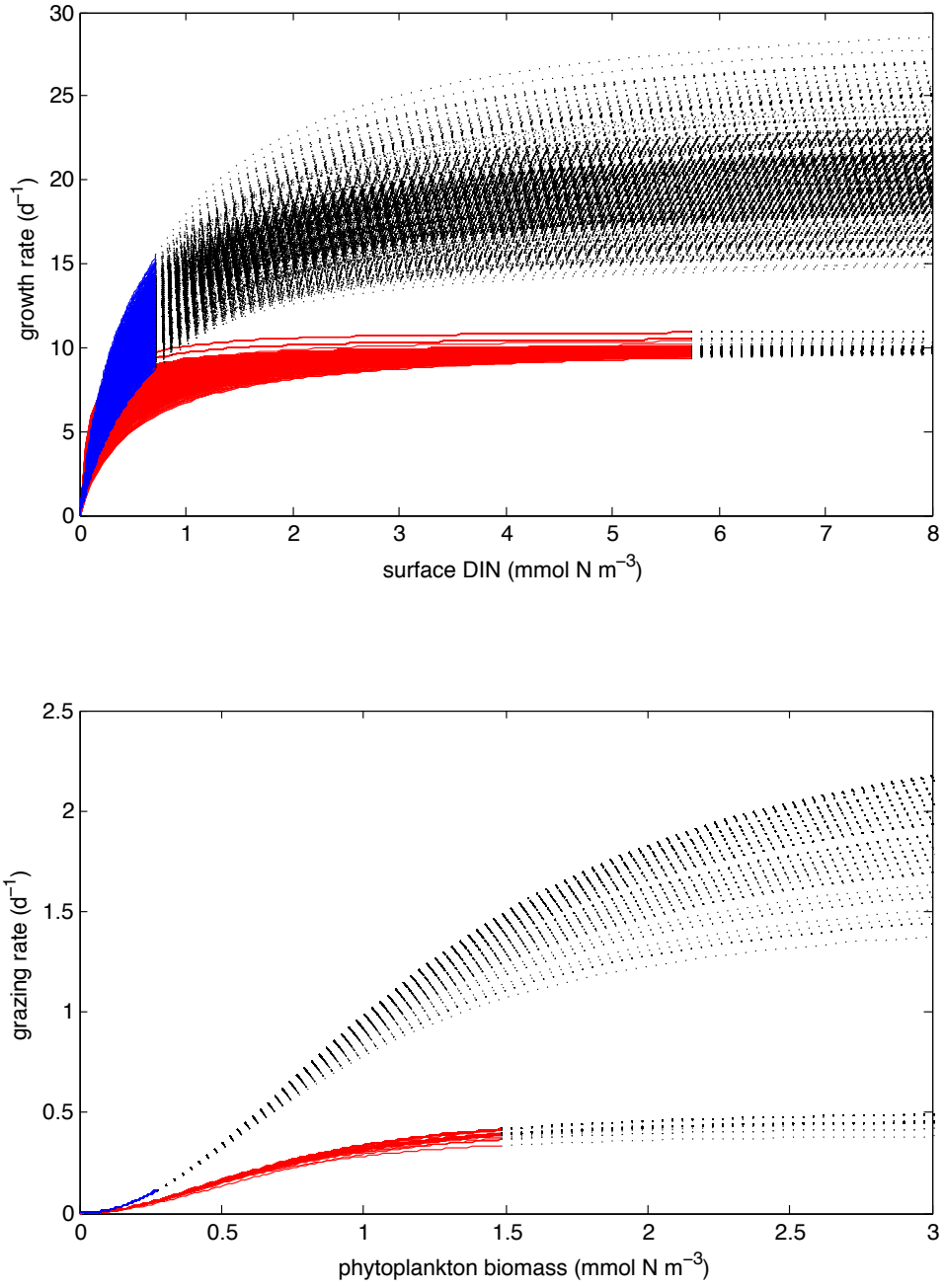

FIGURE 6.9: Growth and grazing response solutions for model 1. The blue and red lines represent the different optimal estimates of the light saturated growth and zooplankton grazing responses at BATS and NABE respectively. The lines are coloured for concentrations up to the 99$^{\text{th}}$ percentile of observed values for surface DIN and phytoplankton (chlorophyll/1.59) at each site.

These conclusions were supported in the results from this chapter, where the BATS calibrated parameters were of very low quality when it came to reproducing the data at NABE. It was suggested that the reason for this was the mapping of errors in the physical and ecological models into the optimised model parameters. Because the cost function cannot distinguish between errors caused by the model physics and errors caused by the ecological models, it was very hard to determine which of these was primarily responsible for the poor performance of the BATS calibrated parameters in the cross-validation experiment.

Regardless of their source, errors in the modelling framework were transformed into *ad hoc* errors in the model parameters, and the resultant solutions were of very poor quality. In order to successfully model the ecosystem at BATS, it will be necessary to make improvements in both the model physics and the ecological models. The process of parameter optimisation can be helpful in this respect. Through careful analysis of the model shortcomings, pre- and post-calibration, it is possible to highlight areas of the model where improvements are required (Schartau and Oschlies, 2003b; Oschlies and Schartau, 2005). In this study, the results suggest that improvements are needed at BATS with regard to the levels of new and regenerated nutrient supply during the summer months (e.g. Fasham et al., 1990; Spitz et al., 2001; McGillicuddy and Robinson, 1997; Williams and Follows, 1998). Further analysis will however be required to ascertain the roots causes of these model deficiencies. The relative contributions of physical error are examined further in the following chapter.

The problem outlined in the preceding section throws some doubt on the arguments made previously in this chapter, although it does not refute them. It also explains the much higher costs (calibrated and predictive) seen at the NABE site relative to BATS. Such a problem might have been overcome by weighting the model-data misfits according to the magnitude of the individual observations, rather than the errors, but this was not addressed in experiments here. The application of such a weighting term would also invalidate the Bayesian justification for the use of a least-squares cost function (section 1.5).

As described in the introduction, the use of cross-validation experiments has the potential to highlight where models are overfit to data, because unrealistic parameter estimates may be penalised when evaluated against independent data. In reality, such experiments are made very difficult as a consequence of the issues described here and in chapter 2. An alternative to the cross-validation approach may be to evaluate error and uncertainty separately, examining models in terms of their ability to fit available data, and their uncertainty with regard to their response to environmental change. In the context of the models evaluated here, the best approach may be to evaluate each model's ability

to fit to all the available data at a number of different sites simultaneously (although this could potentially cause problems where data were collected during different years, or even decades, e.g. Schartau and Oschlies, 2003a). Any attempt to fit a model to data from a diverse number of sites would provide a strong test of a model's generality, and it is likely that model performance would be relatively poor. Increasing model complexity would undoubtedly decrease the model-data misfits in such an experiment, but if these solutions, with their associated uncertainties, were then used to make predictions, it is likely that estimates from more complex models would be too uncertain to be of any practical use. In this way models could be compared in terms of their error with regard to reproducing current observations, and their uncertainty in terms of predicting future scenarios (e.g. Litchman et al., 2006).

Chapter 7

Physical errors and ecology

7.1 Introduction

As phytoplankton are circulated by the advective and diffusive motions of water they are exposed to varying levels of light and nutrients, the two primary factors controlling their growth. In this way the composition of the marine ecosystem is strongly influenced by the physical environment. For the same reasons, the behaviour of a model of the marine ecosystem will also be determined by whatever approximation of the physical environment it is nested within. If the ecological performance of a model is to be assessed, it is therefore important to understand the influence of any errors in the physical forcing (Doney, 1999; Popova et al., 2006).

The models applied in this study are directly and indirectly affected by a number of physical variables, specifically solar radiation, water temperature, vertical mixing and mixed-layer depth. Biogeochemical model performance has been shown to be affected by systematic physical model errors, such as a poor approximation of the mixed-layer depths (e.g. Friedrichs et al., 2006; Popova et al., 2006) or inadequate representation of double-diffusive mixing (Glessmer et al., 2008). Inconsistencies between the real environment (in which the data are collected) and the synthetic environment (in which a model is run) will also lead to model-data misfits, even if the ecological model offered a perfect representation of the ecosystem (e.g. Evans, 1999; Schartau and Oschlies, 2003a; Wallhead et al., 2006). This is of special interest in parameter optimisation studies, where the calibration process will usually seek to minimise such misfits through changes in the ecological parameters, regardless of whether the errors were ecological or physical in origin. For this reason, if a model is nested within a flawed representation of the physical environment, it is likely that errors in the physical forcing will be compensated for by errors in the ecological parameters.

The model results presented in chapters 4, 5 and 6 are based on physical forcing derived from the OCCAM model. When the model mixed-layer depths were compared to data in figure 3.2, it was apparent that the OCCAM environment was too cold and too deeply mixed at BATS. The results presented in chapters 5 and 6 revealed that none of the five models nested within the OCCAM physical environment could correctly reproduce the data at BATS, although the techniques that were applied did not allow for any clear distinction to be made between intrinsic errors in the ecological model structure and extrinsic errors that were caused by the underlying physical forcing. Before it can be concluded that extra ecological model complexity is required at BATS, it will be necessary to carry out a more detailed investigation into the importance of the physical errors that have been highlighted in the OCCAM model.

In this chapter the aim is to compare the performance of model one as it is run within the OCCAM environment, with its performance within the POP environment. Through this comparison it will be possible to examine how errors in the physical forcing relate to errors in the optimised model output and parameters, as well as to the model's predictive skill.

7.2 Temporal (and spatial) errors

Before moving on to look at the effects of large systematic errors in the physical forcing schemes, the potential importance of small temporal and spatial mismatches between the observed and modelled environment was examined. This was done in a twin experiment sensitivity analysis, with synthetic data generated using the OCCAM model at BATS. The individual sensitivity for each parameter of model one was assessed as the observations were shifted backwards and forwards in time by between 1 and 14 whole days relative to the model output. Through this technique it was possible to examine how perturbations of individual parameters from their true value served to compensate for temporal mismatches between the model and observations. The results are presented in figure 7.1, which shows the sensitivity of the misfit cost across the range of each parameter as the observation dataset was shifted backwards and forwards in time. Within each panel, the thick red line shows the optimal parameter value for each temporal shift in the data (as all other parameters were held at their default values).

When the data were not shifted in time, application of the true parameters led to a misfit cost of zero and any perturbation of the parameters invariably led to increases in the model-data misfit. In other words, the default, or true parameters were also the optimal parameters. As temporal errors were introduced to the data the misfit cost for the default parameters ceased to be zero, and a lower misfit could often be achieved by adjusting an individual parameter away from its true value. For many of the 12 parameters examined in figure 7.1, the optimal value for each parameter varied across almost the entire prior range (which themselves were very broad), and responded to the temporal errors in a very irregular way. For example, the optimal value for the maximum phytoplankton growth rate parameter μ_m decreased steadily as the observations were moved backward in time, but as the observations were moved progressively forwards, the optimal value fluctuates between high and moderate values. Another illustration is given by the optimal value for the maximum grazing rate g, which decreased steadily as the observations were held back in time, up until they were delayed by 10 days, at which point the maximum value suddenly became optimal.

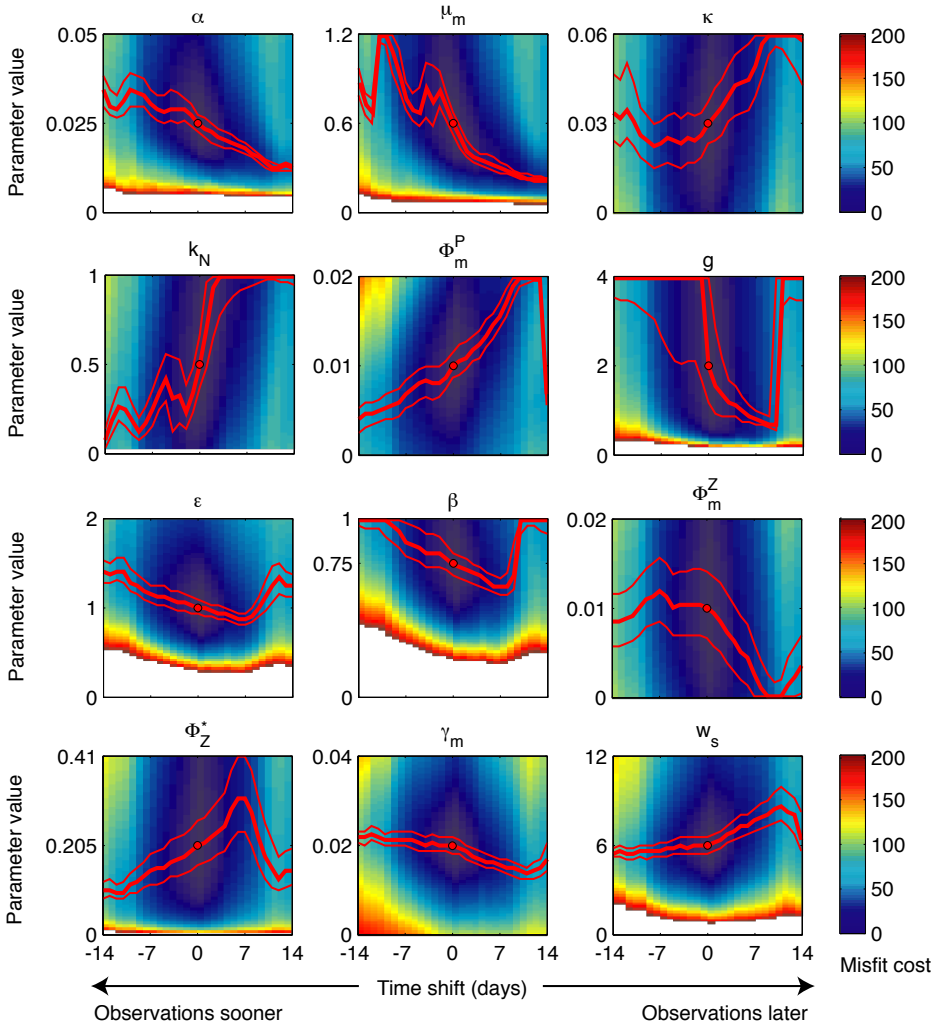

FIGURE 7.1: Misfit cost (J) as a function of parameter values and phase error in synthetic data. Each plot corresponds to one parameter of model 1 (table 3.3). Model-data misfit was examined across the prior range of each parameter (y-axis), as temporal errors were added to the data (x-axis). Thick red lines represent the optimum parameter values as phase errors were introduced, with thin red lines denoting the distance along the y-axis incurring a cost increase of unity. The black circles denote where the true parameters were used with no temporal shift in the data ($J = 0$). Regions with a cost of $J > 200$ are left blank. See table 3.3 for parameter symbols.

This cursory sensitivity analysis illustrates that even if the model framework is perfectly consistent with the observations in terms of ecosystem structure and the physical environment, small temporal and spatial mismatches, or "phase errors" between the physical model and the observed environment can lead to large errors in the calibrated parameters. Such mismatches will be unavoidable wherever the degree of between sample variability is more than can be accurately tracked (Wallhead et al., 2006). This is undoubtedly the case within the marine environment, where transient meteorological and oceanographic features (i.e. storms and eddies) can lead to rapid changes in the biogeochemical properties at a particular location.

Given that much of the variability in the marine environment cannot be constrained, ignoring its effects will result in optimal parameter estimates that contain unknown and potentially large errors. To overcome this problem in the optimisation of a simple ecological model, Schartau and Oschlies (2003a) chose to compare model output with observations in the form of monthly averages. By adopting this approach the possibly spurious temporal (and hence spatial) precision in the relationship between the observations and model output was reduced. The quantities entering the cost function were essentially integrated across each month, and so the potential for erroneously including (or ignoring) some important event was reduced. The main disadvantage of this approach (apart from the fact that less information is included in the cost function, thus making the optimisation problem more underdetermined) is that when sparse measurements are incorporated into monthly mean values they will probably not be representative of each period. This can introduce a new (and potentially important) source of error into the optimisation process, and it was for this reason (ignoring the previously mentioned phase errors) that model output in this study was interpolated to match the observations, rather than the other way around.

An alternative approach to handling these errors was taken by Wallhead et al. (2006), who allowed for temporal errors with the inclusion of time-lags around the data, such that each observation could be moved back and forth through time in a search for the nearest best-fitting modelled value. This method effectively loosens the constraint in time between between the modelled and observed environments, such that mismatches are effectively realigned. An *a priori* constraint on the maximum size of the time-lags was included such that the cost function was not given excessive freedom to smooth out important ecological misfits that are critical to finding the correct parameters. This technique would certainly have led to more accurate estimates of the optimal parameters in the twin experiment outlined in figure 7.1, although it has not yet been applied for the assimilation of seasonal data with unknown temporal errors.

7.3 Alternative physical forcing schemes

The two GCMs used to generate the physical forcing data were introduced in section 3.2, and were compared to observations in section 3.3. The main difference between the two models at BATS is that while mixed-layer depths are apparently realistic in the POP model, they are strongly overestimated in the OCCAM model. The water column described by the POP model data may however be unrealistically stable as a consequence of the interpolation to match the OCCAM temporal and spatial grid.

Using physical forcing derived from both GCMs, the first of the five models outlined in section 3 was evaluated before and after calibration to real data at BATS. The pre- and post-optimisation performance of the ecological model should reveal both its sensitivity to the physical forcing field, as well as the degree to which changes in the physical environment can be compensated for by changes in the model parameters. Model one was initially run with the default parameters given in table 3.3. The twelve most uncertain model parameters were then calibrated using the μGA. The model-data misfits associated with each type of observation are shown in table 7.1.

TABLE 7.1: Cost breakdown for models applied at BATS with alternative physical forcing schemes.

Default	DIN	Chl a	Zoo	PON	CUp	Total
OCCAM	6.0	130.8	5.0	12.5	37.8	192.1
POP	9.3	86.4	7.6	10.1	58.1	171.5
Calibrated	DIN	Chl a	Zoo	PON	CUP	Total
OCCAM	6.6	29.5	4.2	5.9	38.3	84.5
POP	6.3	25.4	4.3	5.3	50.7	90.0

7.3.1 Default model output

Model misfits were lower with the default parameters within the POP environment ($J = 171.5$) than within the OCCAM environment ($J = 192.1$). Most of the cost difference could be attributed to the chlorophyll a misfits, where the model performed 34% better when it was forced by the POP physics. Primary production misfits associated with the default parameters were also large, and in this respect the ecological model actually performed 35% better within the OCCAM physical environment. This improvement was masked in the total cost by the larger changes in the chlorophyll a misfits.

The default model trajectories are shown as black lines in figure 7.2. One of the most noticeable differences in model behaviour within the two physical environments is that

while mean DIN concentrations within the winter mixed layer were strongly underestimated when the POP model forcing was applied. The more accurate representation of DIN when the model was run within the OCCAM environment (table 7.1) occurred despite (or more likely because of) the fact that the depth of winter mixing - and hence the physical supply of nutrients - seems to be strongly overestimated by OCCAM. The POP GCM appears to provide much better estimates of the winter mixed-layer depth, but the biogeochemical model provided a poor representation of observed DIN concentrations when it was run within this environment.

These contrasting results can be followed through to the modelled chlorophyll *a* concentrations and zooplankton biomass. When the model was forced by the deeply-mixed physical environment provided by the OCCAM model, the chlorophyll *a* concentrations within the surface 200 m were too high, especially during winter and spring, although the phytoplankton biomass was almost certainly limited to some extent by the highly variable population of zooplankton that seems to be tightly coupled to chlorophyll *a* biomass in figure 7.2. When the model was run within the POP environment, the chlorophyll *a* concentrations were more realistic (although excessively high winter concentrations were produced in 1992, 1993 and 1995). The figures shown in table 7.1 reveal that while the OCCAM forcing led to a chlorophyll *a* misfit of 130.8, the POP forcing was associated with a much lower (although far from desirable) value of 86.4. Although the model produced a more realistic phytoplankton population when it was run within the POP environment, this was grazed by an unrealistically small and unstable zooplankton population (figure 7.2). The modelled zooplankton trajectories in both scenarios provide a very poor visual match to observations, but the associated misfits contribute very little to the cost function as a consequence of the low confidence weighting assigned to them in chapter 3.

The contrasting responses described above are best understood in terms of primary production, which although too low in both physical environments is both higher and more accurate with the OCCAM forcing. Within this environment deep winter mixing led to a larger supply of nutrients, thus allowing slightly stronger primary production in the euphotic zone. Chlorophyll *a* biomass subsequently became unrealistically large during the winter and spring, but was prevented from rising even higher by a tightly-coupled population of zooplankton grazers. Winter mixing was by contrast much weaker within the POP environment and, as a consequence of the low nutrient availability, there was even less primary production. Chlorophyll *a* biomass could only be maintained at realistic levels because the zooplankton population was very small and grazing pressure was relatively low.

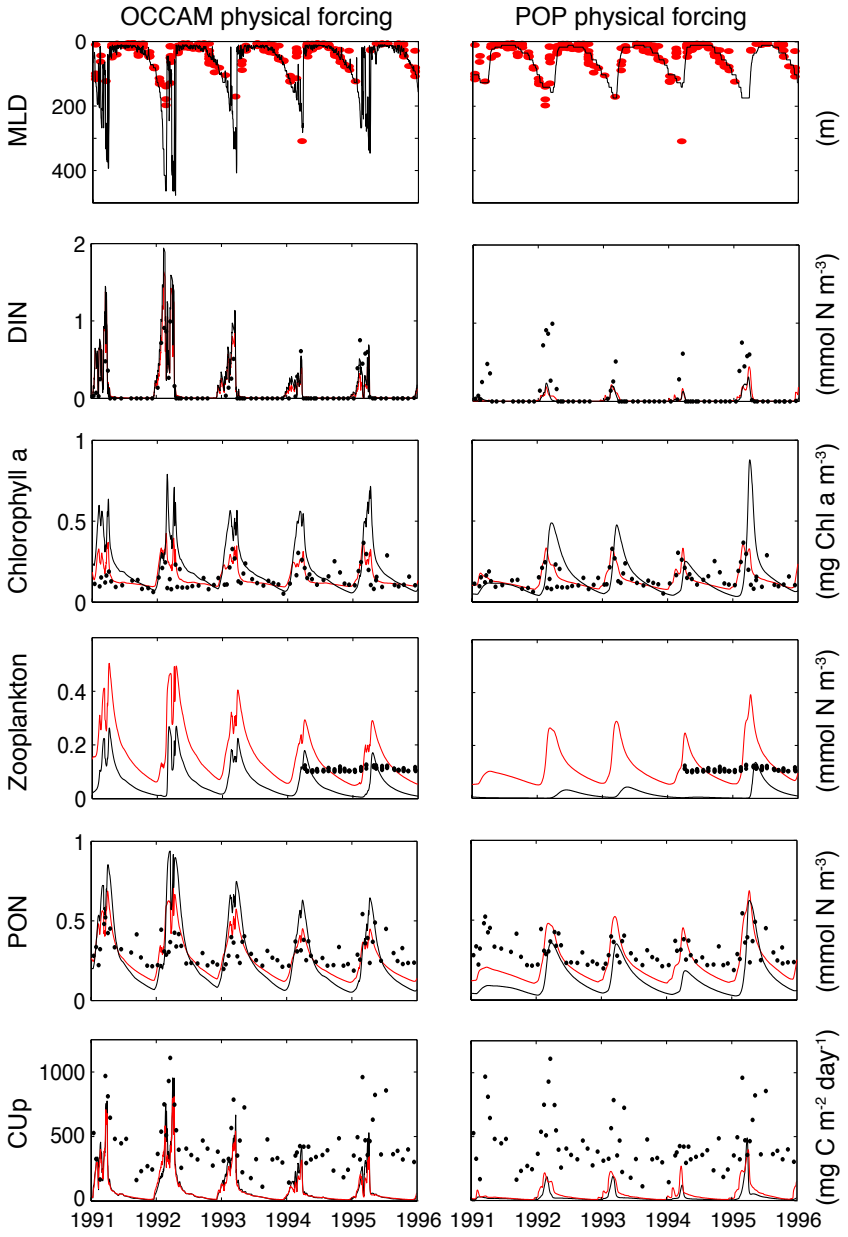

FIGURE 7.2: Default (black) and calibrated (red) model output at BATS using physical forcing data derived from the OCCAM and POP GCMs. Observations of the mixed-layer depth are shown as red dots. Biogeochemical observations are shown as black dots. For DIN the plotted values represent the mean concentration within the modelled mixed layer, which explains the difference between the plotted values for the OCCAM and POP models.

Although there were some differences in the chlorophyll *a* biomass within the different environments, the zooplankton populations seem to have been affected more (the low costs can be attributed to the low weightings in the cost function). As more nutrients were supplied to the system, the phytoplankton biomass remained relatively constant, but zooplankton biomass increased as more energy and biomass were passed through the system. This latter phenomenon is reflected in the higher estimates of primary production associated with the OCCAM model forcing.

7.3.2 Calibrated model output

By calibrating the model parameters to the BATS data, overall model-data misfits were significantly reduced (OCCAM: 192.1→84.5 and POP: 171.5→90.0, see table 7.1). The model performance was also less variable between the two physical frameworks. The largest improvements in the ecosystem model performance within both physical environments came in the chlorophyll *a* misfits, which themselves were very large prior to calibration. As with the default parameters, the OCCAM forced model had a lower primary production misfit but performed worse in terms of chlorophyll *a*. These differences were much smaller than they were before calibration.

The trajectories for the optimised model outputs are shown as red lines in figure 7.2. The calibration process led to similar dynamic adjustments within both physical frameworks. Neither primary production nor mean DIN within the mixed layer were dramatically improved by parameter optimisation (although very small improvements were seen within the POP framework). It is seems that in this very simple model these variables were more strongly limited by the physical environment than by the values of the parameters (although the lack of any improvement in modelled DIN may also be related to the low weighting that was assigned in the cost function).

The modelled chlorophyll *a* concentrations within both environments were significantly improved by the parameter optimisation process, with the misfit falling from 130.8 to 29.5 in the OCCAM framework and from 86.4 to 25.4 within the POP environment. The chlorophyll *a* trajectories plotted in figure 7.2 confirm that the model output was indeed much closer to observations.

In contrast to the pattern seen for chlorophyll *a*, the calibration process led to large increases in the zooplankton population within both environments. It appears that the model made significant improvements in the chlorophyll *a* field by maximising grazing pressure from zooplankton. The large and highly variable population of zooplankton (which were not tightly constrained as a consequence of their low weighting) meant that phytoplankton growth could be kept in check while primary production was either held

steady (OCCAM), or very moderately increased (POP). The optimal parameter values that were selected to facilitate these changes are discussed in the following paragraphs.

7.3.3 Optimal parameter values

The optimal parameter values yielded as model one was calibrated within the two physical environments are shown in figure 7.3. Instead of specifying an artificially precise point solution for each parameter, a broader range of values that were associated with costs not significantly worse than the optimal value are given (see chapter 4 for details).

For most parameters the range of near optimal or "acceptable" solutions are broadly similar within the two physical environments. For example, the photosynthetic efficiency parameter α was set to fairly high values relative to the default in both cases. This is not surprising given that large values of this parameter can assist the development of deep chlorophyll maxima, which have been shown to be a feature at BATS (e.g. Steinberg et al., 2001). The value of the maximum growth rate parameter μ_m presents a slightly different example however, because while it is only optimal at relatively high values within the OCCAM environment, it is completely unconstrained when the POP physics are used. This is most probably due to the relatively weak mixing in the POP model forcing, which does not supply enough DIN to allow saturated growth, even if the value of μ_m is extremely low (e.g. Fennel et al., 2001).

The three parameters responsible for zooplankton grazing (g, ϵ and β) are all optimal at relatively high values (the maximum grazing rate parameter g is fairly unconstrained and is optimal at high or moderate values). The parameters that determine the rate of zooplankton mortality (Φ_m^Z and Φ_Z^*) are by contrast optimal at the low end of the search space. This combination of high grazing and zooplankton growth efficiency with low zooplankton mortality indicates how very strong top-down control was selected by the optimisation process as a means to keep the chlorophyll a biomass in check while maintaining primary production. The selection of this pathway appears to be a realistic choice given the relative simplicity of the NPZD model and the conventionally held view that most of the phytoplankton growth at BATS is consumed and recycled through bacteria and microzooplankton before it can be exported to depth (Fasham et al., 1990; Spitz et al., 2001).

As mentioned previously, most of the parameters in model one were optimal across a similar range of values when they were optimised within the OCCAM and POP physical frameworks. The obvious exception to this is the value of the detrital sinking rate w_s, which could be optimally set to any value above 16 m d^{-1} within the OCCAM environment, or to any value below 16 m d^{-1} within the POP environment. These contrasting

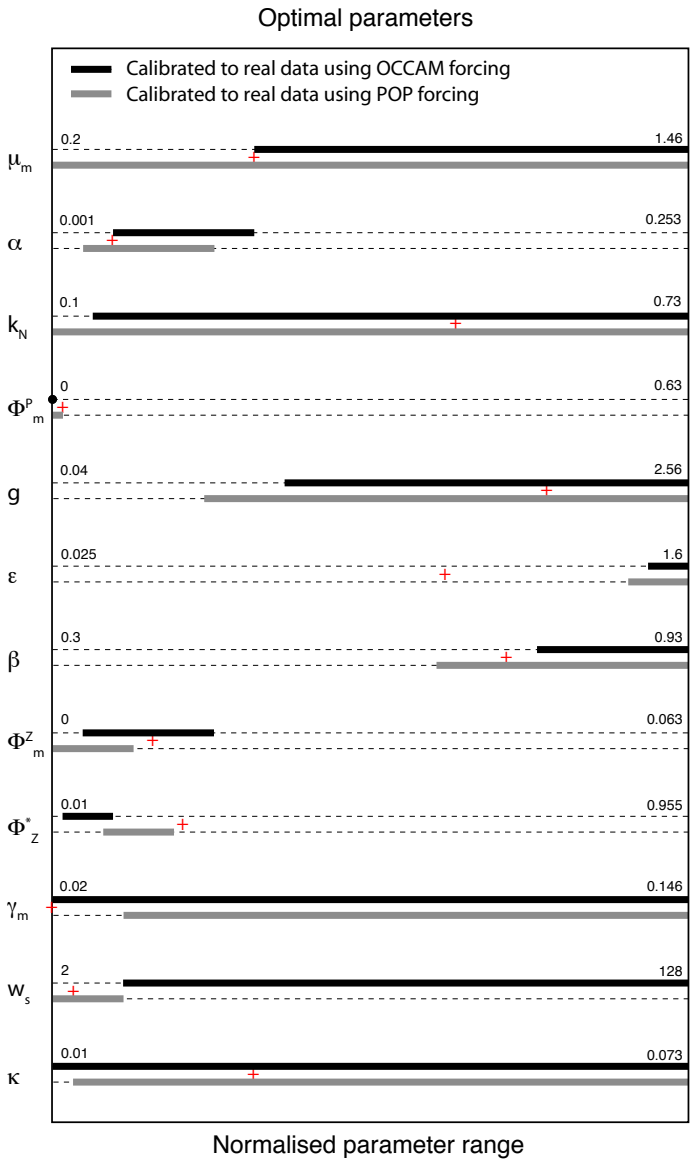

FIGURE 7.3: Optimal parameter values for model one as it was calibrated at BATS with physical forcing derived from the OCCAM and POP GCMs. Black bars represent the range of acceptable parameters as the model was calibrated with OCCAM forcing, grey bars show the same results for the POP forcing. Prior parameter values are shown by red crosses.

solutions can be related to the amount of physical mixing within each GCM. With the default parameters, strong winter mixing in OCCAM led to a large influx of nutrients that accumulated in dissolved inorganic form, as well as in excessive phytoplankton and PON biomass. Physical mixing in the POP environment was not nearly as strong, and the lower levels of DIN that were brought into the euphotic zone were almost immediately passed to phytoplankton biomass, leaving the mixed layer largely depleted of DIN. The two contrasting solutions for w_s occurred as the optimisation process attempted to balance the budget of nitrogen within the mixed layer. When the OCCAM forcing was applied w_s could be set to higher values, allowing the model to quickly remove excess detrital nitrogen before most of it could be remineralised. The much weaker mixing in the POP environment meant that less nitrogen was supplied to the euphotic zone and the low sinking rate allowed detrital matter to be more efficiently recycled to DIN. These adaptations are reflected in the fact that the peak levels of PON within the euphotic zone decreased after calibration in the OCCAM environment, while there was a general increase in PON concentrations in the POP forced model.

The results presented in figure 7.4 show the modelled sediment flux at 300 m, before and after calibration of the OCCAM and POP forced models. In comparison to the differences between the two physical environments, the calibration process had very little effect on the amount of exported material. Apart from the fact that sediment flux data were not incorporated into the cost function, this suggests that the rate of flux was determined primarily by the physical supply of nutrients to the mixed layer (Eppley and Peterson, 1979). This process was apparently realistically modelled within the POP environment, where good estimates of the mixed-layer depth led to a reasonable approximation of the sediment flux at 300 metres. When the model was run within the OCCAM environment, the physical supply of nutrients to the euphotic zone was excessive, and this was balanced by very high estimates for the detrital flux across 300 metres.

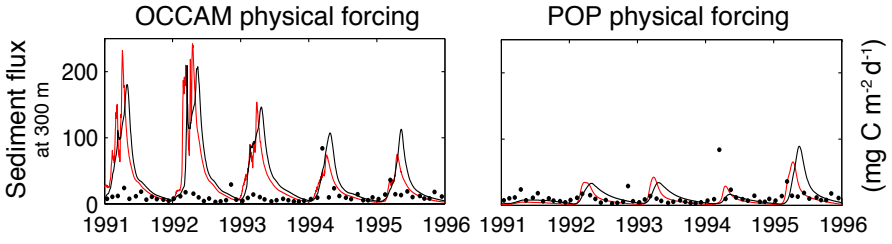

FIGURE 7.4: Sediment flux at 300 m in the default (black) and calibrated (red) models run with alternative physical forcing schemes.

7.3.4 Primary production

Even with the excessive physical supply of nutrients in the OCCAM environment, model one was incapable of correctly reproducing the observed levels of primary production at BATS, a problem that was particularly acute during the summer months. When the model was run within the improved physical environment of the POP model, primary production was too low all year round (figure 7.2). These results suggest two things. Firstly, model one is inadequate to reproduce the observed levels of primary production at BATS, regardless of which of the two physical models it is nested in. Secondly, the more reasonable estimates of winter primary production seen within the OCCAM environment could only occur because because winter nutrient supply was artificially high.

As model complexity was increased in chapter 5, estimates of summertime primary production did not improve as much as might have been expected. In light of the results presented in this chapter, this might be explained by the very high detrital sinking rates that were selected to remove the excess nitrogen mixed into the euphotic zone during the winter. With such high sinking rates it may be very difficult for the models to regenerate enough DIN to fuel primary production during summer. This hypothesis is supported by the results presented in figure 7.5, where the optimised trajectories for primary production are presented as all five models were calibrated within the two physical environments. On the left-hand-side of the plot the OCCAM based results that were presented in chapter 5 are recapitulated. As model complexity increases there are slight improvements in the modelled summer primary production, but even the best estimates are towards the low end of the observed values. The results as the models were optimised within the POP environment are presented on the right-hand-side of the plot. As described above, the simplest models strongly underestimate primary production throughout the year, but as complexity increases, both winter and summer primary production estimates improve. The best estimates come from those models (four and five) which include both the fast-recycling pathway and the variable chlorophyll a to nitrogen ratio, with the DON pathway in model five apparently making little difference. The misfit costs for the five models run within the POP environment are shown in table 7.2, with the equivalent OCCAM results included for comparison.

It appears that the excessive winter mixing within the OCCAM model is indeed responsible for the low summertime primary production that was seen within that environment. By compensating for the extra DIN that is supplied to the mixed-layer, the models adopt an unrealistic value for the detrital sinking rate, and are unable to retain nitrogen within the mixed layer. The model parameters have clearly been altered to compensate for an

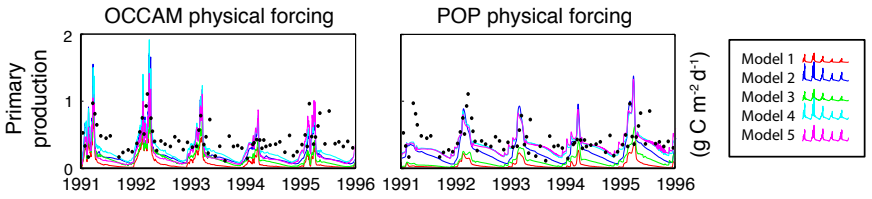

FIGURE 7.5: Optimised primary production from all five models after calibration at BATS with alternative physical forcing schemes.

TABLE 7.2: Optimised misfit costs for all five models at BATS, using alternative physical forcing schemes

OCCAM	DIN	Chl a	Zoo	PON	CUP	Total
Model 1	6.6	29.5	4.2	5.9	38.3	84.5
Model 2	6.1	33.6	2.2	7.9	27.5	77.3
Model 3	6.5	28.6	4.7	9.2	28.0	76.9
Model 4	5.9	29.0	3.6	7.4	20.1	66.6
Model 5	6.0	23.1	0.3	5.1	24.0	58.6
Mean	6.2	28.8	3.0	7.1	27.7	72.9
Range	0.7	10.5	4.4	4.1	18.2	25.3

POP	DIN	Chl a	Zoo	PON	CUP	Total
Model 1	6.3	25.4	4.3	5.3	50.7	92.0
Model 2	5.5	30.2	4.3	7.1	29.9	77.0
Model 3	5.9	22.8	3.9	6.0	40.2	78.8
Model 4	5.9	22.6	3.7	5.3	21.3	58.7
Model 5	5.6	21.9	2.2	5.3	22.9	57.9
Mean	5.8	24.6	3.7	5.8	33.0	72.9
Range	0.7	8.4	2.1	1.8	29.5	34.1

unrealistic artefact, and this may play a significant role in the predictive ability of the models that are calibrated at BATS within the OCCAM physical environment.

To investigate this possibility, the solutions developed at BATS with the POP forcing were applied within the NABE environment (still using the OCCAM forcing). The predictive costs, shown in table 7.3, are in most cases actually slightly worse than when the model was calibrated within the OCCAM environment. Although these results can not be considered as entirely reliable (the μGA optimisations were not iterated to increase confidence in finding the global minimum, as it was in chapter 5), the results suggest that the low predictive costs of the models at NABE were not just a consequence of the physical errors that were described in this chapter.

TABLE 7.3: Cross-validated costs at NABE as the models were optimised at BATS within alternative physical forcing schemes.

NABE	OCCAM	POP
Model 1	1149±10	1323±73
Model 2	1388±43	2722±653
Model 3	1144±17	1148±58
Model 4	1776±355	1265±99
Model 5	673±35	1212±118

The results presented here support the view that it is very important to get the model physics right before indulging in any ecological model development (Popova et al., 2006). This is especially true if formal parameter optimisation techniques are applied, because the indiscriminate nature of a cost function that condenses all errors into a single value means that problems can manifest in unexpected ways. In the previous chapters the models were calibrated at BATS within an environment that supported an excessive supply of nutrients to the mixed layer, yet the results continually indicated that primary production was too low. The fact that such an obvious error in the physical environment was allowed to remain highlights the danger of naively interpreting calibrated model results as "optimal" solutions. Despite these problems however, many of the results presented in chapters 5 and 6 are supported rather than refuted by the results presented in this chapter. The poor performance of the cross-validated solutions at NABE was attributed mainly to models lacking sufficient ecological complexity to describe the BATS ecosystem in a way that was consistent at NABE. Physical errors were also cited as a potential source of error as the parameters were calibrated at BATS, but the results presented here suggest that the excessive mixed-layer depths in OCCAM were not the main source of error. Other physical errors that were not resolved in either the POP or OCCAM forcing, such as eddy-pumping or horizontal advection, may yet be shown to be important.

7.4 Summary and conclusions

Results from a twin experiment demonstrated that small temporal and spatial errors, in a modelling framework that was otherwise perfectly consistent with the data (in terms of physics and ecology), could lead to large errors in the optimal parameters. Without further analysis it is not possible to say whether the parameter sensitivity to these phase errors would increase or decrease as many parameters were optimised simultaneously, but they will certainly affect any optimal solutions.

These small mismatches between the modelled and observed environment can be related to the fact that the largest contributor to overall misfit throughout this chapter was chlorophyll *a*. This was as a consequence of the high confidence in the chlorophyll *a* data, which gave rise to the high weights assigned in the cost function (table 3.1). Chlorophyll *a* can be measured to a high degree of accuracy relative to its observed range in the ocean (with a detection limit of ∼0.01 mg Chl m^{-3}, e.g. Honjo et al., 1989; BATS, 2008), but is highly spatially variable even at very fine scales (Gower et al., 1980). Given that it is not currently possible to accurately track this variability, if it is not properly accounted for in some other way (typically by loosening the constraints placed on individual misfits, e.g. Schartau and Oschlies, 2003a; Wallhead et al., 2006) it will lead to significant mismatches between the modelled chlorophyll *a* and data in an otherwise consistent model.

The high chlorophyll *a* misfits tended to dominate the cost function, and thus the influence of other types of observation was relatively small. The most striking example of this is given by the zooplankton. Although the model output clearly matches the synthetic zooplankton observations very badly, zooplankton consistently return the lowest misfit costs. While it is true that zooplankton observations are fairly unreliable, suffering from a number of random and systematic errors (e.g. undersampling of microzooplankton and swimmers, unreliable conversion of wet/dry weights to nitrogen biomass, etc.) it is not unreasonable to expect better performance than was achieved here. The unrealistic behaviour of the zooplankton may also have served to mask other model deficiencies, as artificially high grazing was often selected to compensate for excessive chlorophyll *a* biomass during the winter bloom.

To evaluate the significance of the unrealistically deep mixed layer seen in the OCCAM model, a simple one-dimensional ecosystem model was run within two different representations of the physical environment at BATS. The results show that the uncalibrated model behaviour was highly sensitive to the underlying physics. The assimilation of biogeochemical data allowed the differences between the two model runs to be reduced, but disparity in the physical supply of nutrients had a strong effect on the modelled sediment fluxes. Although the two physical environments were quite different, and calibration led to very different estimates of the detrital sinking rate, other changes post calibration were broadly similar, with strong grazing selected as a means to control phytoplankton growth.

Whichever physical model was used, the simple NPZD model was unable to correctly reproduce the seasonal cycle at BATS. Primary production was too low regardless of the modelled physical environment, suggesting that a poor representation of the ecological dynamics was key to the underestimated primary production in model one. In preceding

chapters the models were run at BATS within a flawed representation of the physical environment. Excessive winter mixing led to the inclusion of compensatory errors in the optimal parameters, such as artificially high values for the detrital sinking rate w_s. The strong correction of total nitrogen within the winter mixed layer prevented the model from making significant improvements in summer primary production as complexity was increased. When the physical environment was corrected with the application of the POP model, additional ecological complexity led to the expected improvements in summer primary production.

Just as the marine ecosystem behaves differently under different physical conditions, if an ecological model is forced by the wrong physical data it is likely that it will respond in the wrong way (Doney, 1999; Popova et al., 2006). Ecological models are thus subject to intrinsic errors in their representation of the ecosystem, as well as extrinsic errors in the physical environment. The process of parameter optimisation may serve to correct for both types of error, which can make the results difficult to interpret. The more confidence that can be placed in the physical forcing, the greater the confidence with which the ecological model performance, and errors, can be diagnosed.

The results presented in this chapter have shown that solving the same problem within different physical environments can lead to very different solutions, and for this reason it is important that the physical environment in which the ecosystem model is nested be as realistic as possible (Popova et al., 2006). This can be done either by using data assimilated physical models (e.g. Lu et al., 2007), or by forcing models with real *in situ* data Friedrichs et al. (e.g. 2006). At the same time however, the results presented in this chapter also support the conclusion that the NPZD model was quite inadequate for modelling the BATS ecosystem. The strong underestimation of observed primary production, as well as the very poor performance in the cross-validation experiment, occurred regardless of the physical forcing that was applied, supporting the conclusion that additional model complexity is required to reproduce the key processes at BATS.

Chapter 8

Summary and discussion

8.1 Summary

8.1.1 Parameter optimisation and underdetermination

Chapter 2 presented a comparison between two optimisation techniques, a variational adjoint method and a micro genetic algorithm, as two ecosystem models were calibrated to data. While the results suggested that both methods were equally effective in terms of reducing model-data misfits, the level of underdetermination in both models meant that estimates of the optimal parameter values included a high degree of uncertainty.

When ten parameters from each model were calibrated in the absence of any prior information, solutions were highly inconsistent and often contained unrealistic and occasionally nonsensical values. In the search for a robust and informative solution, two approaches to handling parameter uncertainty were examined. Unrealistic solutions could be prevented either by restricting each parameter to a predefined range of values, or by fixing unconstrained parameters to precise values, based on prior beliefs.

Using the latter approach, as outlined by Friedrichs et al. (2007), the few parameters that were optimised were relatively well-constrained by the data, and precise solutions were consistently returned. When solutions defined in this way were used to assess model predictive skill, estimates were more precise than when a larger number of unconstrained parameters were optimised. Further analysis showed that these precise estimates of predictive skill were however sensitive to the uncertain values of the fixed parameters. Fixing what were essentially poorly defined parameters to precise values led to the uncertainty in the model predictive skill being underestimated. Estimates derived in this way were thus strongly affected by factors that were not considered in the analysis. The less prescriptive approach of excluding unrealistic solutions by placing broad limits on each parameter led to the optimisations producing inconsistent solutions. Although this may initially seem to be a disadvantage, such estimates were all equally valid in terms of misfit cost, so actually presented a more informative solution to the problem.

The two approaches analysed in chapter 2 highlighted that without the incorporation of prior knowledge into unconstrained parameter optimisation problems, it is almost impossible to gain a sensible and useful solution. By incorporating prior information that is too strong however, the solutions become highly dependent on what is in fact poorly defined information. It was concluded that in order to make an accurate model assessment, in terms of residual misfit and uncertainty, it is important to incorporate prior information that is as precise as possible while including realistic uncertainty estimates.

8.1.2 Solutions to the inverse problem

A modelling framework was outlined in chapter 3, and was investigated in chapter 4 with regard to quantifying any uncertainty in the optimal parameters. Prior information was incorporated into the problem through broad limits placed on each parameter. Although more precise solutions may have been obtained by carefully defining more precise priors, broad limits were selected to minimise the chances of falsely excluding a valid solution.

As in chapter two, the optimisation of many poorly constrained parameters led to inconsistent solutions. Parameter sensitivity analyses revealed that this was attributable to a combination of partially correlated parameters and uninformative data. Uncertainty in the optimal parameters was quantified by estimating the range of each parameter over which the μGA found "acceptable" solutions, with a misfit cost not statistically worse than the global minimum cost.

Marginal solutions were defined independently for each parameter, and thus took no account of parameter correlations. When these were used to evaluate the forward model output, it was found that uncertainty was significantly overestimated, with the majority of solutions performing significantly worse than the optimum. To get around this problem, only those individual parameter vectors that yielded acceptable solutions were used to define the solution. These parameter vectors could be evaluated in the forward model, so that the range of optimal solutions could be assessed in terms of modelled standing stocks and fluxes, as well as parameter values.

8.1.3 State estimation at BATS and NABE

Chapter 5 described the optimisation of five closely related NPZD models to data at BATS and NABE. The calibrated models were analysed not only with respect to their ability (or lack thereof) to reproduce the assimilated data, but also with regard to any uncertainty in the optimal model output, which was estimated using the techniques developed in chapter 4.

The simplest of the five models underestimated annual primary production at BATS by approximately 73-79%. The addition of a highly parameterised microbial loop and a variable chlorophyll *a* to nitrogen ratio allowed significant improvements in modelled primary production. The best estimates of this flux were however still considerably lower than observed values, particularly during the summer months. Results demonstrated that the inclusion of both the variable Chl:N ratio and the implicit microbial loop were required for the model to correctly reproduce the time-averaged vertical chlorophyll *a* profile. It was suggested that the consistent shortfall in primary production was attributable to

missing ecological components (e.g. nitrogen fixation, variable stoichiometry), although the inclusion of unresolved physical processes (e.g. eddy-pumping, lateral advection) may also have led to better estimates. The high variability of solutions yielded by the different models at BATS was cited as evidence that at least some of the models were functioning in an unrealistic way.

Model performance was much more consistent at NABE than it was at BATS, in terms of both visualised output and misfit cost. It was suggested that all the models at NABE contain sufficient complexity to explain the dynamics underlying the data at that site (although large residual errors were unavoidable), and additional components that were evaluated were essentially redundant with regard to the data.

8.1.4 Model uncertainty

The regular and frequent collection of data at BATS allowed the optimal model output to be fairly well constrained at that site, despite the fact that many parameters were quite poorly determined. This suggests that those parameters had very little effect on any aspects of the model output, and so will probably remain unconstrained, even as additional data are incorporated. The model output at NABE was by contrast quite uncertain at some points in time and space, most notably around the summertime DCM and the autumn bloom. Here the uncertain estimates of the model parameters had a notable effect on model output (away from the data) and so parameter estimates would be improved if observations were incorporated during those periods.

8.1.5 Model predictive skill

The conclusions drawn with regard to the calibrated model performance at BATS and NABE were supported by the findings of a cross-validation experiment that was presented in chapter 6. As they were calibrated, the five models struggled to match the observed levels of primary production at BATS. Whether this was due to an inadequate representation of the physics or biology at that site (it is more likely both), the optimised parameters were selected to maximise this flux. Because this was done in an unrealistic, *ad hoc* way, when the calibrated parameters were applied to reproduce NABE data, model performance was very poor. The adaptations that were incorporated to maximise primary production at BATS were essentially compensating for model errors, and led to excessive fluxes of nitrogen at NABE.

When the NABE calibrated solutions were applied at BATS the resultant misfit costs were lower than the default values at that site, suggesting that strong compensatory

errors were not incorporated into the optimal parameters as the models were calibrated. It is argued that this was because the models were representative of the ecosystem dynamics in operation at NABE, and so unrealistic parameter adjustments were not required (the models however still performed very badly at BATS, for the same reasons as were described in section 8.1.3).

8.1.6 Physical forcing and ecological error

The results presented in chapters 5 and 6 were based on models that were run within a flawed representation of the physical environment at BATS. The OCCAM representation of the mixed layer was too deep during the winter at this site, and it is possible that any results and conclusions were compromised by this error. To examine this possibility, the models were calibrated again at BATS using physical forcing derived from the POP model, in which the mixed-layer depths were reproduced with greater accuracy.

The model results were indeed highly sensitive to the model physics at BATS, although the discrepancy was reduced as the parameters were calibrated to data. Model one was unable to reproduce the observed levels of primary production at BATS, regardless of which physical environment it was nested in. Strong grazing by zooplankton was selected in both cases as a means by which to maximise primary production. Modelled estimates of the detrital flux at 300 metres were highly sensitive to the physical forcing, and estimates were not reconciled through the calibration process.

Differences in the amount of winter mixing within the two physical forcing schemes led to very different estimates for the optimal value of the detrital sinking rate, with this parameter set to relatively high values within the OCCAM environment. It was suggested that these high values were selected as a means to remove excess nitrogen from the euphotic zone, but this also meant that too much nitrogen was lost from the mixed layer during the summer. As model complexity was increased within the euphotic zone, estimates of summer primary production did not increase as much as might have been expected, given the additional model components. When the POP model was used, the parameter optimisation process did not need to find a way to remove excess nitrogen, and additional ecological complexity led to much better estimates of summer primary production.

Despite improvements in the modelled mixed-layer depths, the parameter solutions developed at BATS within the POP environment still performed very badly as they were applied at NABE. This result supports the conclusions drawn in the previous two chapters, namely that the NPZD models could reproduce the data at BATS, but only by making a number of *ad hoc* and ecologically unfounded adaptations in the parameter

values. These compensatory errors were heavily penalised as they were applied to reproducing independent data at NABE.

8.2 Discussion

8.2.1 Multivariate optimisation

Model-data misfits were defined in this study using cost functions that condensed errors associated with many different types of observations into single scalar values. The weight of contribution from each category of error was determined by the size of the misfits in relation to some weighting terms. These were required so that errors were quantified in the same units, and were defined in terms of the confidence that the modelled values correspond closely to the observed "real world" values.

In the absence of any detailed study focussing on how these weighting terms should be defined, their inclusion is often fairly subjective (Evans, 2003). In this study almost no attention was paid to the magnitude of such terms, as they were taken directly from previous studies (Friedrichs et al., 2007; Schartau and Oschlies, 2003a). In chapter 2 this was not such a problem, as the models and data were also taken directly from the study of Friedrichs et al. (2007). The Schartau and Oschlies (2003a) weighting terms were by contrast applied for the comparison of a new set of models and data. The results presented in chapters 5, 6 and 7 demonstrated that extra care in the assignation of cost function weightings may have been beneficial. The cost function was generally dominated by chlorophyll *a* misfits, although other observation types were also very poorly represented by the models in some cases. The most striking example of this was the modelled zooplankton concentrations, which were almost always much more variable than the data. This was not just a problem because the zooplankton fits were so poor, but also because the unrealistic zooplankton behaviour may have served to compensate for errors in other aspects of the model.

Although more care should have been taken in the assignation of the misfit weighting terms, the process of doing so will always be somewhat inexact and will invariably include a degree of subjectivity in both the form and the magnitude of the weighting terms. Solutions based upon cost functions defined in this way will always be dependent on the balance of weighting terms. An alternative approach that has been applied in the field of climate modelling is based upon the use of multi-objective methods that search for a range of "non-dominated" or "Pareto optimal" solutions (Price et al., 2006, 2009). As was described in chapter 1, such an approach treats misfit contributions from different observation types as incomparable, and solutions are selected on the grounds that

individual costs attributable to any one type of observation cannot be reduced without increasing the costs associated with another type of observation. A range of solutions defined in this way are collectively known as a Pareto front, and offer a number of alternative solutions that allow the researcher to visualise how shortcomings in one aspect of the model might be compensated in another. If this approach had been employed in this study, it may have helped to reveal, for example, how a poor representation of the zooplankton might allow for a better fit in other modelled variables, and would perhaps have highlighted areas for further model development.

8.2.2 Physical forcing

The primary aim of this thesis was to examine a range of ecosystem models of different complexity. By using consistent physical forcing it was possible to highlight what differences in the model output were attributable to differences in the ecological models, but it was much harder to ascertain the cause of errors that were common to all the ecological models. When using parameter optimisation techniques it may also be difficult to establish if the model reflects the response of the real ecosystem to the physical environment, or whether it fits the data only because the calibrated parameters are able to compensate for errors in the ecosystem model, the physical forcing, or both.

The results presented in chapter 7 highlighted that the ecosystem models were very sensitive to the modelled physical environment and that interpretation of the results was made more difficult when the model was calibrated within the flawed OCCAM model. The results also suggested that the conclusions regarding the poor performance of the ecological models at BATS were valid regardless of which of the two physical forcing schemes was applied. As is apparent in the literature (e.g. Anderson and Pondaven, 2003), the basic NPZD structure employed here was inadequate to accurately reproduce the key processes at BATS. It is clear (at least for the models presented here) that there is some way to go with regard to improving both the representation of the physical environment, and the ecosystem. Where biogeochemical models are coupled to ocean GCMs care must be taken to ensure the representation of the physical environment is made as accurate as possible, with particular attention paid to the correct representation of mixed-layer depths (e.g. Popova et al., 2006). Despite the demonstrated sensitivity of the ecosystem models to the physical environments however, the fact that all five models showed very poor predictive skill when calibrated within both the POP and OCCAM environments suggests that the inclusion of additional ecological model complexity need not wait for model physics to be perfected.

8.2.3 Ecosystem model complexity

As complexity increases models resolve more dynamical processes, but also contain more parameters and hence more degrees of freedom. Any increased ability reproduce the observed system must be weighed against extra model uncertainty (Anderson, 2005; Raick et al., 2006). For the purposes of model analysis and comparison it is thus desirable not only to assess the degree to which models can fit certain observations, but also to quantify any uncertainty in model components that are not directly constrained by the data.

In the model comparison study of Friedrichs et al. (2007) the authors calibrated 12 different ecosystem models of differing complexity to data from the Arabian Sea and equatorial Pacific. The results suggested that those models which included explicit representation of multiple plankton size-classes had greater predictive skill when the calibrated models were used to reproduce independent data. The main reason cited for this improvement was that models with greater complexity resolved more key processes, and so were better able to reproduce a broad range of data. This difference can also be thought of in terms of the processes that are involved. More complex models resolve more explicit ecological functions. If these processes are less aggregated, and more fundamental, then they are less likely to change between sites. For this reason, more complex models may be shown to be more general, and thus more portable between sites.

The results presented in chapter 2 broadly support the findings of Friedrichs et al. (2007). Of the two models that were examined, the more complex of the two showed greater predictive skill, regardless of how this was assessed. The results also revealed however that previous work has tended to underestimate model uncertainty by adopting unrealistically precise values for unconstrained parameters. When additional parameter uncertainty was considered, the results still supported the conclusion that the more complex model had the better predictive skill. The current trend towards developing models of ever higher complexity is justified on the grounds that without incorporating many of the key ecological components that drive the biological pump it will be impossible to resolve feedbacks between such processes within a changing environment (Le Quéré et al., 2005; Sinha et al., in preparation). Although entirely valid, this imperative for the inclusion of additional complexity must be tempered by the fact that our knowledge of how the marine ecosystem functions is limited.

The increasing underdetermination of more complex models means that as models include more unknown parameters, they will be increasingly be able to fit more biogeochemical observations, possibly for the wrong reasons (Anderson, 2005). This is again a

valid argument, and as has been demonstrated, it is one that must be considered in the comparison of models of different complexity. Again however, such arguments must be counterbalanced by consideration of the broad simplification required in the formulation of NPZD type models. Such models are in some ways more empirical than the more complex PFT models, which resolve a larger number of explicit ecological functions. Given that previous analyses (e.g. Matear, 1995; Hurtt and Armstrong, 1996; Schartau and Oschlies, 2003a; Friedrichs et al., 2006) have consistently revealed that the observations are currently inadequate to constrain even these simple models, the possibility that they too are fitting the data for the wrong reasons must be considered.

This prospect was demonstrated in this study when the NPZD models were fitted to data at BATS. Although relatively simple, the calibrated models could be brought reasonably close to observations at that site (particularly when the physical forcing was corrected). The results however revealed that the calibrated parameters performed very badly when they were used in an attempt to reproduce independent data at NABE. It was argued that this result was attributable to the incorporation of *ad hoc* parameter adjustments that were attributable to model deficiencies rather than some genuine ecological processes. The most striking example of this was the inclusion of the implicit microbial loop that was parameterised in a simple linear relation to phytoplankton biomass. The optimisation process was able to raise modelled estimates of primary production by setting the parameter describing this flux to relatively high values, but this adaptation was not generally applicable at both sites, and led to excessive fluxes at NABE. When the models were calibrated to data at NABE, the model structure was broadly similar to the prevailing ecological theory at that site (i.e. light limited phytoplankton growth and rapid detrital sinking), and so less compensatory errors were incorporated into the parameters.

Simpler models can be made to fit data for the wrong reasons, and while the cost of unconstrained complexity is increased uncertainty, it appears that the problem of leaving key dynamic processes unresolved is increased bias error. The search for optimal model complexity will continue to be the search for the point where any improvements in model fit (both calibrated and predictive) are outweighed by the increased uncertainty of more complex models. When assessing this trade-off it will be necessary to find a balance between simple models that provide relatively precise, but potentially heavily biased solutions, and more complex models which although potentially quite uncertain, may be able to get closer to the true (i.e. observed) situation.

The models presented in the latter chapters of this thesis appear to be of sufficient complexity to explain the ecosystem behaviour at NABE (although large residuals were unavoidable, probably as a consequence of mistimed short-term stratification events).

The models were however short of the degree of complexity required to reproduce the important ecological processes that occur at BATS. At this site it is necessary to include more model components, and several areas of potential improvement were cited in chapter 5, including methods to make model diagnostics more consistent with observations. The results presented here are not sufficient to highlight which model components are required, but the techniques presented do offer a means of answering this question. As has been stated numerous times throughout the text, the addition of model complexity may lead to increased uncertainty in both diagnostic and prognostic model estimates, but given that the calibrated solutions at BATS were relatively well defined, and that there is a wealth of ecological and biogeochemical data available at BATS, it seems likely that complexity can be increased some way before increased uncertainty becomes a problem at that site.

The results presented at NABE suggest that model complexity was sufficient to capture the spring bloom at that site, with the models behaving in a credible fashion, as reflected by their performance in the cross-validation experiment. The relatively high uncertainty associated with the summer and autumn behaviour of the model would suggest that additional observations will be required if the models are to be used to estimate annually integrated values for fluxes such as primary production and export.

8.2.4 A global model?

One-dimensional studies allow for a detailed analysis of model performance. The models can be run many times, so detailed sensitivity and uncertainty analyses may be performed. At the same time, comprehensive, depth-resolved data sets may be used to examine the veracity of model outputs. The localised nature of these models means however that they are of little use when it comes to estimating the importance of the biological pump on a basin or global scale. To achieve such goals models must be coupled to three-dimensional ocean GCMs, in much the same way as they were nested in a one-dimensional environment here.

The results presented in this study would suggest that none of the models are suitable for application in a global model, if the goal is to estimate primary production. Although the models could reproduce observed fluxes at NABE and BATS (once the physical forcing had been corrected), they could not achieve consistent results at both sites using the same parameter values. If production in the BATS ecosystem was correctly reproduced it was probably for the wrong reasons, and the same flux was overestimated at NABE. When the NABE ecosystem was modelled correctly, primary production was grossly underestimated at BATS. A similar result was obtained by Schartau and Oschlies

(2003a), who calibrated an NPZD model (closely related to the second of the five models applied here) to data at three North Atlantic sites, including BATS and NABE. Their estimate of primary production at BATS was also close to observed values, but primary production at NABE was considerably higher than the best estimate achieved in this study. Although no data were available to constrain annual primary production at this latter site, it seems that the cost of fitting these simple models to data at BATS is high primary production at NABE. If the same biogeochemical model is to be used to model both these sites it will be necessary to increase model complexity to include a number of important processes. Many of these may be redundant at NABE, but it has been shown that if a consistent solution is to be obtained it is essential that they are included at BATS.

Appendix A

Inverse theory

Inverse problems are approached with the goal of finding the optimum estimate of the parameters (or more generally, the model control variables, which may also include additional factors such as initial conditions or time-step size). This estimate yields the minimum discrepancy between the model output and the observations. This section will focus on the different forms of the inverse problem, and what their solutions can reveal about a model of a system.

Explicit linear form of the inverse problem

Working from a set of N measurements, and assuming a model perfectly consistent with the real system, the data can be represented by a vector $\mathbf{d}$ of length N. The model parameters may also be represented this way, by a vector $\mathbf{p}$ of length M. This is shown in equation A.1 below, with T denoting a transpose vector:

$$\begin{aligned} &\text{data:} && \mathbf{d} = [d_1, d_2, d_3, ..., d_N]^T \\ &\text{model parameters:} && \mathbf{p} = [p_1, p_2, p_3, ..., p_M]^T \end{aligned} \tag{A.1}$$

The model describes the relationship between $\mathbf{d}$ and $\mathbf{p}$, and takes the form of one or more functions that the data and parameters are expected to be consistent with. The full set of model equations can be summarised in the vector equation $\mathbf{f}(\mathbf{d}, \mathbf{p})$.

In the explicit linear form of the inverse problem, the data and parameter vectors are both linear. The equation vector is represented by the relationship between $\mathbf{d}$ and $\mathbf{p}$, which is described by the data kernel $\mathbf{G}$.

$$\mathbf{f}(\mathbf{d}, \mathbf{p}) = 0 = \mathbf{d} - \mathbf{G}\mathbf{p} \tag{A.2}$$

or

$$\mathbf{G}\mathbf{p} = \mathbf{d} \tag{A.3}$$

The data kernel $\mathbf{G}$ takes the form of an $M \times N$ matrix which when multiplied out in equation A.3 represents the explicit model equations. The data kernel may contain constants as well as linear and non-linear expressions of auxiliary variables.

Auxiliary variables are not integral parts of the model equations but rather describe the geometry of the system. An example of an auxiliary variable varying non-linearly in this way is given by light-attenuation with depth in a water column. Local light intensity may be an integral parameter of a model, but it can remain as a linear member of the parameter vector by multiplying it by an exponential function of depth within the data kernel.

Time may also be considered as an auxiliary variable such that time dependent problems can be described by this form. The explicit linear form is central to the study of inverse problems, which often either take this form, or may be solved through assuming linearity as an approximation (Menke, 1989).

Over-determined and under-determined problems

Looking at the explicit linear form, the inverse problem is described by equation A.3. The best case solution to such a problem, is where $\mathbf{G}\mathbf{p} = \mathbf{d}$ contains just enough information to constrain a unique set of parameters that lead to a model perfectly consistent with the data. Unfortunately, as a result of sparse and noisy data, in combination with incomplete models, in practice such a solution is rarely, if at all, possible.

Over-determined problems are inverse problems where there is too much information in $\mathbf{G}\mathbf{p} = \mathbf{d}$ to produce an exact solution with zero misfit. The simplest example of this is fitting a straight line, $y = mx + c$, to a group of non-colinear points (Menke, 1989). In such a case, there is more information in the data than can be explained by the model, and the only possible solution to the inverse problem is a "best fit" to the data.

Under-determined problems represent the converse situation where there is not enough information in $\mathbf{G}\mathbf{p} = \mathbf{d}$ and there are a large (in fact infinite) number of different parameter sets that yield zero misfit with the data. The model generates information

that is not matched in the data. Because such output has no effect on the model-data misfit, the parameters governing that output are unconstrained and non-unique solutions exist. The complimentary example to the over-determined problem is trying to fit a high order polynomial the same set of non-colinear points. Where the straight line of the previous example is over-determined, and cannot fit every point, the high order polynomial is under-determined, as there are an infinite number of curves that fit the data equally well. The straight line fails to capture all the information in the data (of which much could be random noise), but it is of greater utility than the under-determined (or over-fit) polynomial, because it succeeds in capturing the general trend.

In a complex inverse problem, it is possible that there are parts of the problem which are over-determined, and parts which are under-determined. This situation is labelled a mixed-determined problem, and is characterised by non-unique optimum solutions that do not fully capture the variability in the data. The types of problems described by these three categories are known as ill-posed questions.

The key to getting the best solution to ill posed questions is to try and minimise the uncertainty associated with each case. Where the model is seriously over-determined, it may be sensible to add further complexity, so that important processes and feedbacks may be resolved. It is important however not to add complexity where it is not constrained in the available data, as this would make the problem under-determined and allow fitting to noise.

Where a model is under-determined, it makes sense to either try and find more data, or where this is not possible, to try and reduce the complexity of the model, so that it is not over-fit. It is also possible to add *a priori* information, not contained in $\mathbf{Gp} = \mathbf{d}$, that helps to reduce the uncertainty by eliminating some of the non-unique solutions.

Appendix B

Model equations (Chapter 2)

Both models were embedded in an identical physical forcing field, as described in chapter 2 and Friedrichs et al. (2007). Vertical advection and sinking of biological tracers were evaluated using a third-order direct space-time upwind based scheme (Hundsdorfer and Trompert, 1994) and a Sweby flux limiter (Sweby, 1984). These were adapted to estimate one-dimensional (vertical) advection. Conservation of mass was maintained by assuming horizontal gradients of zero in the biogeochemical tracers in the Arabian Sea. A scaling analysis of the equatorial Pacific (Friedrichs and Hofmann, 2001) revealed that horizontal advective divergence of nitrate is an important process in that region, and so output from a three-dimensional biogeochemical model (Murtugudde et al., 1996) was used to provide an additional source/sink when the model was run for the equatorial Pacific.

Light attenuation at depth was calculated in the same way for both models, with photosynthetically available radiation (PAR) attenuated through depth by water and chlorophyll molecules according to the following equation

$$k_I = k_w + k_{Chl} \cdot Chl$$

where $k_w = 0.05$ m^{-1} is the attenuation coefficient for water, $k_{Chl} = 0.1$ m^2 mg Chl is the light attenuation by chlorophyll, and Chl is the chlorophyll concentration at depth z, given in mg Chl m^{-3}. Phytoplankton to chlorophyll ratios were model dependent.

Four-component model

The four component (NPZD: dissolved inorganic nitrogen, phytoplankton, zooplankton and detritus) model was taken from (McCreary et al., 2001). Within each model level, j, the biological source-minus-sink terms are given by

$$\text{sms(DIN)} = a_n g_r \mathcal{F} Z + eD - g\overline{\mathcal{I}}\mathcal{N}P \tag{B.1}$$

$$\text{sms(P)} = g\overline{\mathcal{I}}\mathcal{N}P - g_r \mathcal{P} Z - \mu_p P \tag{B.2}$$

$$\text{sms(Z)} = a_z g_r \mathcal{F} Z - g_r \mathcal{Z} Z - \mu_z Z \tag{B.3}$$

$$\text{sms(D)} = (1 - a_z - a_n) g_r \mathcal{F} Z + \mu_p P + \mu_z Z - eD - \frac{w_s}{h_j}(D_{j-1} - D_j) \tag{B.4}$$

Within the equations listed above, $\overline{\mathcal{I}}$ describes the average light-limited phytoplankton growth rate within model level j, as a function of the photosynthetically available radiation (PAR).

The average light level within each model layer, calculated using a fixed phytoplankton nitrogen to chlorophyll ratio of 0.381 mg Chl (mmol N)$^{-1}$, was used to evaluate the degree of light limitation within each layer. The light-limitation term at depth z takes the form

$$\mathcal{I} = \frac{I_z}{\sqrt{I_z^2 + I_o^2}}$$

Nutrient limited growth is described by the function,

$$\mathcal{N} = \frac{N}{N + N_o}$$

where N is the ambient DIN concentration, and N_o is the half-saturation concentration for DIN uptake.

Zooplankton graze on both the phytoplankton and themselves, according to a Michaelis-Menten function that relates the total food supply (F), to a specific grazing rate ($g_r \mathcal{F}$) where

$$\mathcal{F} = \frac{F}{F + F_o}$$

and F_o is the half-saturation food concentration. The food supply is a function of the concentrations of P and Z, as well as the zooplankton preference (ϕ_p) for grazing on phytoplankton rather than on other zooplankton.

$$F = \phi_p P + (1 - \phi_p) Z$$

Individual losses to grazing from the phytoplankton and zooplankton compartments are thus respectively described by $g_r \mathcal{P}$ and $g_r \mathcal{Z}$, where

$$\mathcal{P} = \frac{\phi_p P}{F + F_o}$$

and

$$\mathcal{Z} = \frac{(1 - \phi_p) Z}{F + F_o}$$

The sinking rate for detritus is w_s.

Nine-component model

The nine-component ecosystem model, originally described by Leonard et al. (1999), was used in the modified form given by Christian et al. (2002). The model includes two state variables each (small and large) for phytoplankton, zooplankton and detritus, in addition to state variables for ammonium, nitrate and iron. Phytoplankton specific growth is calculated as a temperature dependent maximum growth rate (Eppley, 1972), multiplied by light and nutrient limitation terms. The source-minus-sink terms for changes in the two phytoplankton populations through time are

$$\mathrm{sms(P_s)} = \mu_{max}^{T} \cdot I_{lim} \cdot N_{lim} \cdot P_s - g_{P_s} \Lambda (1 - e^{-\Lambda P_s}) P_s Z_s - m_s P_s \tag{B.5}$$

and

$$\mathrm{sms(P_l)} = \mu_{max}^{T} \cdot I_{lim} \cdot N_{lim} \cdot P_l - g_{P_l} \Lambda (1 - e^{-\Lambda P_l}) P_l Z_l - m_l P_l \tag{B.6}$$

The maximum growth rate possible at temperature T is given by

$$\mu_{max}^{T} = \mu_0 e^{k_T T}$$

where μ_0 is the maximum specific phytoplankton growth rate at 0°C, and k^T is a constant. Light-limitation is calculated as

$$I_{lim} = 1 - e^{-\alpha I / P_{max}}$$

where α and P_{max} are empirically derived parameters for the initial slope of the P-I curve and the maximum photosynthetic rate, given as functions of time and depth (Cullen et al., 1992). Phytoplankton chlorophyll content was estimated using the Redfield ratio and the equations of Geider et al. (1996). The optimal chlorophyll-to-carbon ratio is given by

$$\theta = \frac{\theta_{max}}{1 + \theta_{max} \alpha I / 2 P_{max}^{C}}$$

where θ_{max} is the maximum chlorophyll-to-carbon ratio, P_{max}^{C} is the maximum carbon specific growth rate. The modelled values were relaxed towards θ with an acclimation "half-life" of approximately three days (Christian et al., 2002).

Nutrient limitation is calculated separately for the small and large phytoplankton as the minimum of nitrogen and iron limitation terms, with nitrogen limitation dependent on the sum of nitrogen available as nitrate and ammonium.

$$N_{lim} = \min(\mathrm{NO}_3' + \mathrm{NH}_4', \mathrm{Fe}') \tag{B.7}$$

The contribution of the preferred ammonium form is given by

$$\mathrm{NH}_4' = \frac{\mathrm{NH}_4}{k_{\mathrm{NH}_4} + \mathrm{NH}_4}$$

where NH_4 is the ambient concentration of ammonium, and k_{NH_4} is the half-saturation concentration of ammonium uptake. Wherever ammonium concentrations are sufficient to allow maximum growth (i.e. $\mathrm{NH}_4' = 1$), nitrate uptake is completely inhibited. NO_3 uptake is accordingly described by,

$$\mathrm{NO_3'} = \frac{\mathrm{NO_3}}{k_{\mathrm{NO_3}} + \mathrm{NO_3}}(1 - \mathrm{NH_4'})$$

with nitrate uptake increasing as ammonium becomes scarce. Iron limitation is given by the following expression,

$$\mathrm{Fe'} = \frac{\mathrm{Fe}}{k_{\mathrm{Fe}} + \mathrm{Fe}}$$

In the two preceding equations, the NO_3 and Fe terms describe ambient nutrient concentrations, with $k_{\mathrm{NO_3}}$ and k_{Fe} giving the half-saturation concentrations for uptake.

Small zooplankton graze on small phytoplankton, while large zooplankton graze on both large phytoplankton and the small zooplankton. The grazing response is described by a modified non-satiating Ivlev function (Ivlev, 1955).

$$\mathrm{sms(Z_s)} = \lambda g_{P_s} \Lambda (1 - e^{-\Lambda P_s}) P_s Z_s - g_{Z_s} \Lambda (1 - e^{-\Lambda Z_s}) Z_s Z_l - (r_s + \delta_s) Z_s \tag{B.8}$$

and

$$\mathrm{sms(Z_l)} = \lambda \Lambda (g_{P_l} (1 - e^{-\Lambda P_l}) P_l + g_{Z_s} (1 - e^{-\Lambda Z_s}) Z_s) Z_l - (r_l + \delta_l) Z_l \tag{B.9}$$

The larger grazer population has no explicit predator, so the "miscellaneous mortality parameter" δ_l is set to a relatively high value to provide closure (Steele and Henderson, 1992).

The equations for the formation of detritus and its sinking across the base of each model layer are

$$\mathrm{sms(D_s)} = (1 - \lambda) g_{P_s} \Lambda (1 - e^{-\Lambda P_s}) P_s Z_s + \delta_s Z_s + m_s P_s - (c_s + w_s) D_s \tag{B.10}$$

and

$$\mathrm{sms(D_{l1})} = (1 - \lambda) \Lambda (g_{P_l} (1 - e^{-\Lambda P_l}) P_l + g_{Z_s} (1 - e^{-\Lambda Z_s}) Z_s) Z_l + \delta_l Z_l - (c_l + w_l) D_{l1} \tag{B.11}$$

Ammonium and nitrate uptake are calculated in the following equations according to the ratios of nutrient limitation given in equation B.7

$$\begin{aligned}\mathrm{sms(NO_3)} = &- I_{lim}\Big(\mu^T_{max}\cdot N_{lim}\cdot\frac{\mathrm{NO_3'}}{\mathrm{NO_3'}+\mathrm{NH_4'}}\cdot P\Big)_{small}\\ &- I_{lim}\Big(\mu^T_{max}\cdot N_{lim}\cdot\frac{\mathrm{NO_3'}}{\mathrm{NO_3'}+\mathrm{NH_4'}}\cdot P\Big)_{large}\\ &+ NH_4\cdot k_{NO_3}\cdot(z>120m)\end{aligned} \tag{B.12}$$

$$\begin{aligned}\mathrm{sms(NH_4)} = &- \Big(\mu^T_{max}\cdot I_{lim}\cdot N_{lim}\cdot\frac{\mathrm{NH_4'}}{\mathrm{NO_3'}+\mathrm{NH_4'}}\cdot P\Big)_{small}\\ &- \Big(\mu^T_{max}\cdot I_{lim}\cdot N_{lim}\cdot\frac{\mathrm{NH_4'}}{\mathrm{NO_3'}+\mathrm{NH_4'}}\cdot P\Big)_{large}\\ &+ c(D_s+D_l)+r(Z_s+Z_l)\\ &+ F_{in}\\ &- NH_4\cdot k_{NO_3}\cdot(z>120m)\end{aligned} \tag{B.13}$$

Sinking particulate nitrogen is returned to inorganic form through the remineralisation term F_{in} (equation B.16). Below 120 m, ammonium is returned to nitrate through nitrification, hence the Boolean switch term $(z > 120m)$.

Iron cycling was handled by assigning a fixed Fe:N ratio to phytoplankton, zooplankton and detritus, with the exception of large phytoplankton, which were assigned a variable ratio. In order to maintain conservation of mass, excess iron released as this group of plankton were grazed was passed directly to the dissolved pool. An additional state variable, D_{l2}, was adopted to allow sedimentation of iron in aggregates of large phytoplankton.

$$\mathrm{sms(D_{l2})} = m_l P_l - (c_l + w_l)D_{l2} \tag{B.14}$$

The iron equation is

$$\begin{aligned}\mathrm{sms(Fe)} = &- (\mu_s P_s R + \mu_l P_l R_N) + R(r_s Z_s + r_1 Z_1 + c_s D_s + c_l D_{l1}) + R_{P_l}(c_l D_{l2})\\ &+ g_{P_l}\Lambda(1-e^{-\Lambda P_l})Z_l P_l(R_N - R)\end{aligned} \tag{B.15}$$

where μ_s $(= (\mu^T_{max}\cdot I_{lim}\cdot N_{lim})_s)$ and μ_l $(= (\mu^T_{max}\cdot I_{lim}\cdot N_{lim})_l)$ are the realised growth rates for small and large phytoplankton. The parameters R_{P_l} and R are the Fe:N

ratios for, respectively, large phytoplankton, and all other organic state variable (i.e. phytoplankton and zooplankton biomass, and detritus).

Particulate sinking out of each layer was attenuated according to the empirical formula of Martin et al. (1987). Remineralisation within each layer (F_{in}) was subsequently parameterised as the instantaneous redistribution of sinking particulates to inorganic form. This was calculated as the difference between the flux into the top of each layer, $F(z_t)$, and the flux out of the bottom $F(z_b)$.

$$F_{in} = F(z_t) - F(z_b) = F(z_0)\left(\frac{z_t^b - z_b^b}{z_0^b}\right)$$

where $F(z_0)$ is the magnitude of the flux at its original depth z_0, and z_t and z_b are the depths at the top and bottom of the layer in question. The parameter b is the attenuation coefficient.

For a particular level j, this supply of nutrients is given by the sum of remineralised sinking matter from all overlying layers.

$$F_{in}(j) = \sum_{i=1}^{j-1}(w_s D_s(i) + w_l(D_{l1}(i) + D_{l2}(i)))\left(\frac{z_{j-1}^b - z_j^b}{z_i^b}\right) \tag{B.16}$$

Primary production (PP) was calculated as

$$PP = (\mu_s P_s + \mu_l P_l)RR$$

where RR is the Redfield ratio, C:N = 6.625.

TABLE B.1: All *a priori* parameters for the nine-component model (Christian et al., 2002). Where applicable, the two columns give the parameter values for "small" and "large" size fractions respectively, except for the molar iron-to-nitrogen ratio, where value 2 applies to netplankton only. The initial slope of the P-I curve and the maximum photosynthetic rate are given as empirical functions of time and depth, after Cullen et al. (1992)

Parameter	Symbol	Value 1	Value 2	Units
Max. phytoplankton growth rate at 0°C	μ_0	0.59		d^{-1}
T-dependence of phytoplankton growth	k_T	0.0633		°C
Initial slope of P-I curve	α	$f(t,z)$		$m^2W^{-1}d^{-1}$
Maximum photosynthetic rate	P_{max}	$f(t,z)$		d^{-1}
Half-saturation for nitrate uptake	K_{NO_3}	0.25	0.3	μM
Half-saturation for ammonium uptake	K_{NH_4}	0.05		μM
Half-saturation for iron uptake	K_{Fe}	10.0	120.0	pM
Molar iron-to-nitrogen ratio	R	1.98×10^{-5}	3.3×10^{-5}	-
Phytoplankton grazing rate parameter	g_P	50.0	50.0	d^{-1}
Zooplankton grazing rate parameter	g_Z	10.0		d^{-1}
Ivlev coefficient for grazing rate	Λ	1.0		$(\text{mmol N m}^{-3})^{-1}$
Assimilation efficiency	λ	0.75		-
Zooplankton specific respiration rate	r	0.1		d^{-1}
Zooplankton mortality rate	δ	0.05	0.2	d^{-1}
Phytoplankton mortality rate	m	0.05	0.2	d^{-1}
Detrital remineralisation rate	c	0.03		d^{-1}
Specific detrital sedimentation rate	w	0.02	0.07	d^{-1}
Rate coefficient for nitrification	k_{NO_3}	0.04		d^{-1}
Particle flux attenuation	b	-0.858		-
Growth rate for chl.-to-carbon ratio	P^C_{max}	1.0		d^{-1}
Maximum chl.-to-carbon ratio	θ_{max}	0.0278		-
Rate coefficient for photoacclimation	k_{acc}	0.24		d^{-1}
Solubility limit for dissolved iron	S_{Fe}	0.6		nM

Appendix C

Model equations (Chapters 3 - 7)

Model 1

The model equations are taken from Oschlies and Garçon (1999). The evolution of the state variables (N, P, Z and D) is determined by vertical mixing and a source-minus-sink (sms) term describing biological activity. Vertical mixing is provided by the prognostic output of an offline general circulation model (either OCCAM or POP, see chapter 3).

The source-minus-sink terms are given by

$$\mathrm{sms(P)} = \overline{J}(z,t,\mathrm{N})\mathrm{P} - G(\mathrm{P})\mathrm{Z} - \phi_m^P \mathrm{P} \tag{C.1}$$

$$\mathrm{sms(Z)} = \beta G(\mathrm{P})\mathrm{Z} - \phi_m^Z \mathrm{Z} - \phi_Z^* \mathrm{Z}^2 \tag{C.2}$$

$$\mathrm{sms(D)} = (1-\beta)G(\mathrm{P})\mathrm{Z} + \phi_m^P \mathrm{P} + \phi_Z^* \mathrm{Z}^2 - \gamma_m \mathrm{D} - w_s \frac{\partial \mathrm{D}}{\partial z} \tag{C.3}$$

$$\mathrm{sms(N)} = \gamma_m \mathrm{D} + \phi_m^Z \mathrm{Z} - \overline{J}(z,t,\mathrm{N})\mathrm{P} \tag{C.4}$$

where $\overline{J}$ is the daily averaged phytoplankton growth rate and G is the grazing function. The remaining parameters are defined in table 3.3.

The phytoplankton growth rate is calculated as the minimum of the light- and nutrient-limited growth,

$$\overline{J}(z,t,\mathrm{N}) = \min\left(\overline{J}(z,t), J_{\max}\frac{\mathrm{N}}{k_N+\mathrm{N}}\right) \tag{C.5}$$

where $\overline{J}(z,t)$ is the light-limited growth rate, and $J_{\max}$ is the light-saturated growth.

The instantaneous light-limited growth rate $J(z,t)$ is averaged in time over $\tau_{24\mathrm{h}} = 1$ day, and vertically across each grid box, yielding

$$\overline{J}(z,t) = \frac{1}{\tau_{24\mathrm{h}}}\int_0^{24\mathrm{h}} \frac{1}{z_k - z_{k+1}}\int_{z_{k+1}}^{z_k} J(z,t)\,dz\,dt \tag{C.6}$$

where

$$J(z,t) = \frac{V_p \alpha I(z,t)}{\left[V_p^2 + (\alpha I(z,t))^2\right]^{1/2}} \tag{C.7}$$

$$I(z,t) = I(t)_{z=0}\, e^{\left(k_w \tilde{z} - \int_{\tilde{z}}^0 \kappa P\,dz\right)} \tag{C.8}$$

$$I(t)_{z=0} = \mathrm{f_{PAR}}\,\tau(t)\,2\,\frac{\tau_{24\mathrm{h}}}{\tau_{\mathrm{sun}}}\,I_{\mathrm{ECMWF}}(t) \tag{C.9}$$

$$J_{\max} = V_p = \mu_m C_{\mathrm{ref}}{}^{cT} \tag{C.10}$$

Here $\tilde{z} = z/\sqrt{1-(\cos\theta/1.33)^2}$ is the effective vertical depth of incident light at angle θ at noon. A triangular function $\tau(t)$ describes the evolution of the day, increasing linearly from 0 to 1 between daybreak and noon, then decreasingly linearly to zero at nightfall, over the variable day length τ_{sun}. The integral (C.6) is solved analytically, as described by Evans and Parslow (1985). The parameter T is the temperature in $^\circ C$, and I_{ECMWF} is the monthly mean shortwave radiation, provided by the European Centre for Medium-Range Weather Forecasts (ECMWF).

The zooplankton grazing function uses a Holling type III function.

$$G(\mathrm{P}) = \frac{g\epsilon \mathrm{P}^2}{g + \epsilon \mathrm{P}^2} \tag{C.11}$$

Model 2

Model 2 includes a fast recycling pathway from P to N, as an implicit representation of the microbial loop (Oschlies, 2001). The linear flux is described using the existing parameter Φ_m^P. The loss of phytoplankton to detritus was re-parameterised, replacing the existing linear term, $\Phi_m^P \mathrm{P}$, with the quadratic term, $\Phi_P^* \mathrm{P}^2$. Changes to the core model equations (C.1 to C.4) are highlighted in red.

$$\mathrm{sms(P)} = \overline{J}(z,t,\mathrm{N})\mathrm{P} - G(\mathrm{P})\mathrm{Z} \; - \; \phi_P^* \mathrm{P}^2 - \phi_m^P \mathrm{P} \tag{C.12}$$

$$\mathrm{sms(Z)} = \beta G(\mathrm{P})\mathrm{Z} - \phi_m^Z \mathrm{Z} - \phi_Z^* \mathrm{Z}^2 \tag{C.13}$$

$$\mathrm{sms(D)} = (1-\beta)G(\mathrm{P})\mathrm{Z} \; + \; \phi_P^* \mathrm{P}^2 \; + \; \phi_Z^* \mathrm{Z}^2 - \gamma_m \mathrm{D} - w_s \frac{\delta \mathrm{D}}{\delta z} \tag{C.14}$$

$$\mathrm{sms(N)} = \gamma_m \mathrm{D} \; + \; \phi_m^P \mathrm{P} \; + \; \phi_m^Z \mathrm{Z} - \overline{J}(z,t,\mathrm{N})\mathrm{P} \tag{C.15}$$

Model 3

Model 3 is identical to model 1, but incorporates a variable Chl *a* to nitrogen biomass ratio following Cloern et al. (1995), with the units converted from mg Chl *a* (mg C)$^{-1}$ to mg Chl *a* (mmol N)$^{-1}$ using Redfieldian stoichiometry (Redfield, 1934). The relationship is defined using the following function,

$$\mathrm{Chl:N} = 79.5 \left(\mathrm{C_a} + \mathrm{C_b} \exp(\mathrm{C_c T}) \; \exp(-\mathrm{C_d} \times 0.4\mathrm{I}) \; \mu' \right) \tag{C.16}$$

where

$$\mu' = \frac{\mathrm{N}}{k_N + \mathrm{N}} \tag{C.17}$$

defines the degree of nutrient limitation. Photoacclimation was included by allowing the initial slope of the photosynthesis:irradiance curve α is to vary in direct proportion to the Chl:N ratio, normalised to the fixed Chl:N ratio of 1.59 mg Chl *a* (mmol N)$^{-1}$, such that

$$\alpha_{acclimated} = \alpha \text{Chl} : \text{N} \frac{1}{1.59} \tag{C.18}$$

with $\alpha_{acclimated}$ replacing α in equation C.7 (following Anderson et al., 2007).

Model 4

Model 4 incorporates all the changes made in models 2 and 3, and thus includes the fast recycling pathway, the variable Chl:[N] ratio and photoacclimation.

Model 5

Model 5 builds on model 4, incorporating a new state variable for semi-labile DON. This is supplied by a constant fraction (ϕ_D) of the existing detrital remineralisation flux. DON is returned to the nutrient pool at a linear biomass specific rate, parameterised by γ_D, the hydrolysis of semi-labile DON.

$$\text{sms(P)} = \overline{J}(z,t,\text{N})\text{P} - G(\text{P})\text{Z} - \phi_P^*\text{P} - \phi_m^P\text{P} \tag{C.19}$$

$$\text{sms(Z)} = \beta G(\text{P})\text{Z} - \phi_m^Z\text{Z} - \phi_Z^*\text{Z}^2 \tag{C.20}$$

$$\text{sms(D)} = (1-\beta)G(\text{P})\text{Z} + \phi_P^*\text{P} + \phi_Z^*\text{Z}^2 - \gamma_m\text{D} - w_s\frac{\delta\text{D}}{\delta z} \tag{C.21}$$

$$\text{sms(DON)} = \phi_D\gamma_m\text{D} - \gamma_D DON \tag{C.22}$$

$$\text{sms(N)} = \gamma_D DON + (1-\phi_D)\gamma_m\text{D} + \phi_m^P\text{P} + \phi_m^Z\text{Z} - \overline{J}(z,t,\text{N})\text{P} \tag{C.23}$$

Bibliography

Anderson, L. A., Sarmiento, J. L., 1994. Redfield ratios of remineralization determined by nutrient analysis. Global Biogeochemical Cycles 8, 65–80.

Anderson, T. R., 2005. Plankton functional type modelling: Running before we can walk? Journal of Plankton Research 27 (11), 1073–1081.

Anderson, T. R., submitted. Progress in marine ecosystem modelling and the "unreasonable effectiveness of mathematics". Journal of Marine Systems.

Anderson, T. R., Pondaven, P., 2003. Non-Redfield carbon and nitrogen cycling in the Sargasso Sea: pelagic imbalances and export flux. Deep-Sea Research I 50, 573–591.

Anderson, T. R., Ryabchenko, V. A., in press. Geophysical monograph series. Ch. Carbon cycling in the mesopelagic zone of the central Arabian Sea: Results from a simple model.

Anderson, T. R., Ryabchenko, V. A., Fasham, M. J. R., Gorchakov, V. A., 2007. Denitrification in the Arabian Sea: A 3D ecosystem modelling study. Deep-Sea Research I 54, 2082–2119.

Anderson, T. R., Totterdell, I. J., 2004. The Ocean Carbon Cycle and Climate. NATO Science Series, Ch. Modelling the response of the biological pump to climate change, pp. 65–96.

Anderson, T. R., Williams, P. J. l. B., 1998. Modelling the seasonal cycle of dissolved organic carbon at Station e_1 in the English Channel. Estuarine, Coastal and Shelf Science 46, 93–109.

Andrieu, C., De Freitas, N., Doucet, A., Jordan, M. I., 2003. An introduction to mcmc for machine learning. Machine Learning 50, 5–43.

Armstrong, R. A., Lee, C., Hedges, J. I., Honjo, S., Wakeham, S. G., 2002. A new, mechanistic model for organic carbon fluxes in the ocean based on the quantitative association of POC with ballast materials. Deep-Sea Research II 49, 219–236.

Athias, V., Mazzega, P., Jeandel, C., 2000. Selecting a global optimization method to estimate the oceanic particle cycling rate constants. Journal of Marine Research 58, 675–707.

Aumont, O., Maier-Reimer, E., Blain, S., Pondaven, P., 2003. An ecosystem model of the global ocean including Fe, Si, P co-limitations. Global Biogeochemical Cycles 17 (2,1060), doi:10.1029/2001GB001745.

Azam, F., Fenchel, T., Field, J. G., Gray, J. S., Meyer-Reil, L. A., Thingstad, F., 1983. The ecological role of water-column microbes in the sea. Marine Ecology Progress Series 10, 257–263.

Bacastow, R., Maier-Reimer, E., 1990. Ocean-circulation model of the carbon cycle. Climate Dynamics 4, 95–125.

Backhaus, J. ., Wehde, H., Hegset, E. N., Kämpf, J., 1999. 'phyto-convection': the role of oceanic convection in primary production. Marine Ecology Progress Series 189, 77–92.

Baird, M. E., 1999. Towards a verified mechanistic model of plankton population dynamics. Ph.D. thesis, Department of Biological Sciences, University of Warwick.

BATS, 2008. Bermuda Atlantic Time-series Study.
URL `http://bats.bios.edu/`

Billet, D. S. M., Lampitt, R. S., Rice, A. L., Mantoura, R. F. C., 1983. Seasonal sedimentation of phytoplankton to the deep sea benthos. Nature 302 (520-522).

Bopp, L., Aumont, O., Cadule, P., Alvain, S., Gehlen, M., 2005. Response of diatoms to global warming and potential implications: A global model study. Geophysical Research Letters 32 (L19606).

Buesseler, K. O., Lamborg, C. H., Boyd, P. W., Trull, P. J. L. T. W., Bidigare, R. R., Bishop, J. K. B., Casciotti, K. L., Dehairs, F., Elskens, M., Honda, M., Karl, D. M., Siegel, D. A., Silver, M. W., Steinberg, D. K., valdes, J., Mooy, B. V., Wilson, S., 2007. Revisiting carbon flux through the ocean's twilight zone. Science 316, 567–570.

Capone, D., Zehr, J. P., Paerl, H. W., Bergman, B., Capenter, E. J., 1997. *Trichodesmium*, a globally significant marine cyanobacterium. Science 276, 1221–1229.

Caron, D. A., Dam, H. G., Kremer, P., Lessard, E. J., Madin, L. P., Malone, T. C., Napp, J. M., Peele, E. R., Roman, M. R., Youngbluth, M. J., 1995. The contribution of microorganisms to particulate carbon and nitrogen in surface waters of the Sargasso Sea near Bermuda. Deep-Sea Research I 42 (6), 143–972.

Carroll, D. L., 1996. Chemical laser modelling with genetic algorithms. AIAA Journal 34 (2), 338–346.

Christian, J. R., Verschell, M. A., Murtugudde, R., Busalacchi, A. J., McClain, C. R., 2002. Biogeochemical modelling of the tropical Pacific Ocean. I: Seasonal and interannual variability. Deep-Sea Research II 49, 509–543.

Cloern, J. E., Grenz, C., Vidergar-Lucas, L., 1995. An empirical model of the phytoplankton chrlorphyll:carbon ratio - the conversion factor between productivity and growth rate. Limnology and Oceanography 40 (7), 1313–1321.

Copin-Montegut, C., Copin-Montegut, G., 1983. Stoichiometry of carbon, nitrogen, and phosphorous in marine particulate matter. Deep-Sea Research 30, 31–46.

Cullen, J. J., Lewis, M. R., Davis, C. O., Barber, R. T., 1992. Photosynthetic characteristics and estimated growth rates indicate grazing is the proximate control of primary production in the equatorial pacific. Journal of Geophysical Research 97, 639–654.

Dadou, I., Evans, G. T., Garçon, V., 2004. Using JGOFS in situ and ocean color data to compare biogeochemical models and estimate their parameters in the subtropical North Atlantic Oceantetxitin situ and ocean color data to compare biogeochemical models and estimate their parameters in the subtropical North Atlantic Ocean. Journal of Marine Research 62, 565–594.

Dadou, I., Lamy, F., Rabouille, C., Ruiz-Pino, D., Andersen, V., Bianchi, M., Garçon, V., 2001. An integrated biological pump model from the euphotic zone to the sediment: a 1-D application in the Northeast tropical Atlantic. Deep-Sea Research II 48, 2345–2381.

Dam, H. G., Miller, C. A., Jonasdottir, S. H., 1993. The trophic role of mesozooplankton at 47°N, 20°W during the North Atlantic Bloom Experiment. Deep-Sea Research II 40 (1/2), 197–212.

Denman, K. L., 2003. Modelling planktonic ecosystems: parameterizing complexity. Progress in Oceanography 57, 429–452.

Doney, S. C., 1999. Major challenges confronting marine biogeochemical modelling. Global Biogeochemical Cycles 13 (3), 705–714.

Ducklow, H. W., 1983. Production and fate of bacteria in the oceans. BioScience 33, 494–501.

Ducklow, H. W., Harris, R. P., 1993. Introduction to the JGOFS North Atlantic bloom experiment. Deep-Sea Research II 40, 1–8.

Dutkiewicz, S., Follows, M. J., Heimbach, P., Marshall, J., 2006. Controls on ocean productivity and air-sea carbon flux: An adjoint model sensitivity study. Geophysical Research Letters 33 (L02603), doi:10.1029/2005GL024987.

Engel, A., Goldthwait, S., Passow, U., Alldredge, A., 2002. Temporal decoupling of carbon and nitrogen dynamics in a mesocosm diatom bloom. Limnology and Oceanography 47 (3), 753–761.

Eppley, R. W., 1972. Temperature and phytoplankton growth in the sea. Fisheries Bulletin (US) 70, 1063–1085.

Eppley, R. W., Peterson, B. J., 1979. Particulate organic matter flux and planktonic new production in the deep ocean. Nature 282, 677–680.

Evans, G. T., 1999. The role of local models and data sets in the Joint Global Ocean Flux Study. Deep-Sea Research I 46, 1369–1389.

Evans, G. T., 2003. Defining misfit between biogeochemical models and data sets. Journal of Marine Systems 40-41, 49–54.

Evans, G. T., Parslow, J. S., 1985. A model of annual plankton cycles. Biological Oceanography 3 (3), 327–347.

Falkowski, P. G., Ziemann, D., Kolber, Z., Bienfang, P. K., 1991. Role of eddy pumping in enhancing primary production in the ocean. Nature 352 (55-58).

Fasham, M. J. R., 2000. The changing ocean carbon cycle. Cambridge University Press, Ch. Advances in ecosystem modelling within JGOFS, pp. 417–446.

Fasham, M. J. R., Ducklow, H. W., McKelvie, S. M., 1990. A nitrogen-based model of plankton dynamics in the oceanic mixed layer. Journal of Marine Research 48, 591–639.

Fasham, M. J. R., Evans, G. T., 1995. The use of optimization techniques to model marine ecosystem dynamics at the JGOFS station at 47N 20W. Philosophical Transactions: Biological Sciences 348 (1324), 203–209.

Faugeras, B., Bernard, O., Sciandra, A., Lévy, M., 2004. A mechanistic modelling and data assimilation approach to estimate the carbon/chlorophyll and carbon/nitrogen ratios in a coupled hydrodynamical-biological model. Nonlinear Processes in Geophysics 11, 515–533.

Faugeras, B., Lévy, M., Mémery, L., Verron, J., Blum, J., Charpentier, I., 2003. Can biogeochemical fluxes be recovered from nitrate and chlorophyll data? a case study assimilating data in the Northwestern Mederterranean Sea at the JGOFS-DYFAMED station. Journal of Marine Systems 40-41, 99–125.

Fennel, K., Losch, M., Schroter, J., Wenzel, M., 2001. Testing a marine ecosystem model: Sensitivity analysis and parameter optimization. Journal of Marine Systems 28, 45–63.

Fogg, G. E., 1983. The ecological significance of extracellular products of phytoplankton photosynthesis. Botanica Marina 26, 3–14.

Follows, M. J., Dutkiewicz, S., 2002. Meteorological modulation of the North Atlantic spring bloom. Deep-Sea Research II 49, 321–344.

Friedrichs, M. A. M., 2001. A data assimilative marine ecosystem model of the central equatorial Pacific: Numerical twin experiments. Journal of Marine Research 59, 859–894.

Friedrichs, M. A. M., 2002. Assimilation of JGOFS EqPac and SeaWiFS data into a marine ecosystem model of the central equatorial Pacific Ocean. Deep-Sea Research II 49, 289–319.

Friedrichs, M. A. M., Dusenberry, J. A., Anderson, L. A., Armstrong, R., Chai, F., Christian, J. R., Doney, S. C., Dunne, J., Fujii, M., Hood, R. R., McGillicuddy, D., Moore, J. K., Schartau, M., Spitz, Y. H., Wiggert, J. D., 2007. Assessment of skill and portability in regional marine biogeochemical models: the role of multiple planktonic groups. Journal of Geophysical Research 112 (C08001), doi:10.1029/2006JC003852.

Friedrichs, M. A. M., Hofmann, E. E., 2001. Physical control of biological processes in the central equatorial pacific ocean. Deep-Sea Research I 48, 1023–1069.

Friedrichs, M. A. M., Hood, R. R., Wiggert, J. D., 2006. Ecosystem model complexity versus physical forcing: Quantification of their relative impact with assimilated Arabian Sea data. Deep-Sea Research II 53, 576–600.

Geider, R. J., MacIntyre, H. L., Kana, T. M., 1996. A dynamic model of photoadaptation in phytoplankton. Limnology and Oceanography 41, 1–15.

Geider, R. J., MacIntyre, H. L., Kana, T. M., 1998. A dynamic regulatory model of phytoacclimation to light, nutrients and temperature. Limnology and Oceanography 43 (4), 679–694.

Gent, P., McWilliams, J., 1990. Isopycnal mixing in ocean circulation models. Journal of Physical Oceanography 20, 150–155.

Giering, R., Kaminski, T., 1998. Recipes for adjoint code construction. ACM Transactions on Mathematical Software 24 (4), 437–474.

Gilbert, J. C., Lemaréchal, C., 1989. Some numerical experiments with variable-storage quasi-Newton algorithms. Progress in Mathematics 45, 405–435.

Glessmer, M. S., Oschlies, A., Yool, A., 2008. Simulated impact of double-diffusive mixing on physical and biogeochemical upper ocean properties. Journal of Geophysical Research 113 (C08029).

Gnanadesikan, A., Sarmiento, J. L., Slater, R. D., 2003. Effects of patchy ocean fertilization on atmospheric carbon dioxide and biological production. Global Biogeochemical Cycles 17 (2), 1050.

Goldman, J. C., 2000. Primary production in the sea. Ch. Physiological processes, nutrient availability, and the concept of relative growth rate in marine phytoplankton ecology, pp. 179–197.

Gower, J. F. R., Denman, K. L., Hoyler, R. J., 1980. Phytoplankton patchiness indicates the fluctuation spectrum of mesoscale oceanic structure. Nature 288, 157–159.

Gregg, W. W., Friedrichs, M. A. M., Robinson, A. R., Rose, K. A., Schlitzer, R., Thompson, K. R., Doney, S., 2008. Skill assessment in ocean biological data assimilation. Journal of Marine Systems 76 (1-2), 16–33.

Gundersen, K., Orcutt, K. M., Purdie, D. A., Michaels, A. F., Knapp, A. H., 2001. Particulate organic carbon mass distribution at the Bermuda Atlantic Time-series Study (BATS) site. Deep-Sea Research II 48, 1697–1718.

Gunson, J. R., Malanotte-Rizzoli, P., 1996. Assimilation studies of open-ocean flows. 2. Error measures with strongly nonlinear dynamics. Journal of Geophysical Research 101 (C12), 28473–28488.

Hansell, D. A., Bates, N. R., Olson, D. B., 2004. Excess nitrate and nitrogen fixation in the North Atlantic Ocean. Marine Chemistry 84, 243–265.

Hastie, T., Tibshirani, R., Friedman, J., 2001. The elements of statistical learning; data mining, inference, and prediction. Springer.

Hemmings, J. C. P., Srokosz, M. A., Challenor, P., Fasham, M. J. R., 2003. Assimilating satellite ocean-colour observations into oceanic ecosystem models. Philosophical Transactions of the Royal Society of London A 361, 33–39.

Hemmings, J. C. P., Srokosz, M. A., Challenor, P., Fasham, M. J. R., 2004. Split-domain calibration of an ecosystem model using satellite ocean colour data. Journal of Marine Systems 50, 141–179.

Honjo, S., Dymond, J., Collier, R., Manganini, S. J., 1995. Export production of particles to the interior of the equatorial Pacific Ocean during the 1992 EqPac experiment. Deep-Sea Research II 42 (831-870).

Honjo, S., Dymond, J., Prell, W., Ittekkot, V., 1999. Monsoon-controlled export fluxes to the interior of the Arabian Sea. Deep-Sea Research 46 (1859-1902).

Honjo, S., Manganini, S. J., Krishfield, R., 1989. Cruise report: JGOFS Leg 1 International study of the North Atlantic bloom. Tech. Rep. WHOI-89-22: 1-30, WHOI.

Honjo, S., Manganini, S. J., Krishfield, R. A., Francois, R., 2008. Particulate organic carbon fluxes to the ocean interior and factors controlling the biological pump: A synthesis of global sediment trap programs since 1983. Progress in Oceanography 76 (217-285).

Hood, R. R., Laws, E. A., Armstrong, R. A., Bates, N. R., Brown, C. W., Carlson, C. A., Chai, F., Doney, S. C., Falkowski, P. G., Feely, R. A., Friedrichs, M. A. M., Landry, M. R., Moore, J. K., Nelson, D. M., Richardson, T. L., Salihoglu, B., Schartau, M., Toole, D. A., Wiggert, J. D., 2006. Pelagic functional group modelling: Progress, challenges and prospects. Deep-Sea Research II 53, 459–512.

Hundsdorfer, W., Trompert, R. A., 1994. Method of lines and direct discretization:a comparison for linear advection. Applied Numerical Mathematics 16 (6), 469–490.

Hurtt, G. C., Armstrong, R. A., 1996. A pelagic ecosystem model calibrated with BATS data. Deep-Sea Research II 43 (2-3), 653–683.

IPCC, 2001. Climate Change 2001: The Scientific Basis. Intergovernmental Panel on Climate Change Report 2001. Cambridge University Press, Ch. The Carbon Cycle and Atmospheric Carbon Dioxide, pp. 183 – 237.

Ivlev, V. S., 1955. Experimental Ecology of the Feeding of Fishes. Pischepromizdat, Moscow.

Kirkpatrick, S., Gelatt, Jr., C. D., Vecchi, M. P., 1983. Optimization by simulated annealing. Science 220 (4598), 671–680.

Kleypas, J. A., Doney, S. C., 2001. Nutrients, chlorophyll, primary production and related biogeochemical properties in the ocean mixed layer. A compilation of data collected at nine JGOFS sites.
URL `http://dss.ucar.edu/datasets/ds259.0/`

Knap, A. H., Michaels, A. F., Dow, R. L., Johnson, R. J., Gundersen, K., Sorensen, J. C., Close, A. R., Howse, F., Hammer, M., Bates, N., Doyle, A., Waterhouse, T., 1993. BATS methods manual. US JGOFS Planning Office, Woods Hole, MA.

Kortzinger, A., Koeve, W., Kahler, P., Mintrop, L., 2001. C:N ratios in the mixed layer during the productive season in the northeast atlantic ocean. Deep-Sea Research I 48 (661-688).

Kriest, I., Oschlies, A., 2007. Modelling the effect of cell-size-dependent nutrient uptake and exudation on phytoplankton size spectra. Deep-Sea Research I 54 (9), 1593–1618.

Kuroda, H., Kishi, M. J., 2004. A data assimilation technique applied to estimate parameters for the NEMURO marine ecosystem model. Ecological Modelling 172, 69–85.

Large, W. G., McWilliams, J. C., Doney, S. C., 1994. Oceanic vertical mixing: a review and a model with nonlocal boundary layer parameterization. Reviews of Geophysics 32, 363–403.

Lawson, L. M., Hofmann, E. E., Spitz, Y. H., 1996. Time series sampling and data assimilation in a simple marine ecosystem model. Deep-Sea Research II 43 (2-3), 625–651.

Lawson, L. M., Spitz, Y. H., Hofmann, E. E., Long, R. B., 1995. A data assimilation technique applied to a predator-prey model. Bulletin of Mathematical Biology 57 (4), 593–617.

Le Quéré, C., 2006. Reply to Horizons article 'Plankton functional type modelling: running before we can walk?' Anderson (2005): I. Abrupt changes in marine ecosystems? Journal of Plankton Research 28 (9), 871–872.

Le Quéré, C., Harrison, S. P., Prentice, I. C., Buitenhuis, E. T., Aumont, O., Bopp, L., Claustre, H., Cotrim da Cunha, L., Geider, R., Giraud, X., Klaas, C., Kohfeld, K. E., Legendre, L., Manizza, M., Platt, T., Rivkin, R. B., Sathyendranath, S., Uitz, J., Watson, A. J., Wolf-Gladrow, D., 2005. Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. Global Change Biology 11, 2016–2040.

Leonard, C. L., McClain, C. R., Murtugudde, R., Hoffman, E. E., Harding, L. W., 1999. An iron-based ecosystem model of the central equatorial Pacific. Journal of Geophysical Research 104, 1325–1341.

Levitus, S., Burgett, R., Boyer, T. P., 1998. World ocean atlas, 1998, Volume 4, Salinity. NESDIS 3.

Lima, I., Doney, S. C., 2004. A three-dimensional, multi-nutrient and size-structured ecosystem model for the North Atantic. Global Biogeochemical Cycles 18 (3), GB3019.

Litchman, E., Klausmeier, C. A., Schofield, J. R. M. O. M., Falkowski, P. G., 2006. Multi-nutrient, multi-group model of present and future oceanic phytoplankton communities. Biogeosciences 3, 585–606.

Lohrenz, S., Knauer, G., Asper, V., Tuel, M., Michaels, A., Knap, A., 1992. Seasonal variability in primary production and particle flux in the northwestern Sargasso Sea: US JGOFS Bermuda Atlantic Time-series Study. Deep-Sea Research I 39, 1373–1391.

Lu, Y., Wright, D. G., Yashayaev, I., 2007. Modelling hydrographic changes in the labrador sea over the past five decades. Progress in Oceanography 73, 406–426.

Madin, L. P., Horgan, E. F., Steinberg, D. K., 2001. Zooplankton at the Bermuda Atlantic Time-series Study (BATS) station: diel, seasonal and interannual variation in biomass, 1994-1998. Deep-Sea Research II 48, 2063–2082.

Martin, A., Pondaven, P., 2003. On estimates for the vertical nitrate flux due to eddy pumping. Journal of Geophysical Research 108 (C11), 3359.

Martin, J. H., Knauer, G. A., Karl, D. M., Broenkow, W. W., 1987. Vertex: carbon cycling in the northeast Pacific. Deep-Sea Research 34, 267–285.

Matear, R. J., 1995. Parameter optimization and analysis of ecosystem models using simulated annealing: A case study at Station P. Journal of Marine Research 53, 571–607.

McCreary, J., Kohler, K., Hood, R., Olson, D., 1996. A four-comopnent ecosystem model of biological activity in the Arabian Sea. Progress in Oceanography 37, 193–240.

McCreary, J. P., Kohler, K. E., Hood, R. R., Smith, S. L., Kindle, J., Fischer, A. S., Weller, R. A., 2001. Influences of diurnal and intraseasonal forcing on mixed-layer and biological variability in the central Arabian Sea. Journal of Geophysical Research 106 (C4), 7139–7155.

McGillicuddy, D. J., Lynch, D. R., Moore, A. M., Gentleman, W. C., Davis, C. S., Meise, C. J., 1998. An adjoint data assimilation approach to diagnosis of physical and biological controls of *Pseudocalanus spp.* in the Gulf of Maine-Georges Bank region. Fisheries Oceanography 7 (3/4), 205–218.

McGillicuddy, D. J., Robinson, A. R., 1997. Eddy induced nutrient supply and new production in the Sargasso Sea. Deep-Sea Research I 44, 1427–1449.

Menke, W., 1989. Geophysical Data Analysis: Discrete Inverse Theory. Vol. 45 of International Geophysics Series. Academic Press Limited, 24-28 Oval Road, London NW1 7DX.

Menzel, D. W., Ryther, J. H., 1961. Annual variations in primary production of the Sargasso Sea of Bermuda. Deep-Sea Research 7, 282–288.

Michaels, A. F., Knap, A. H., 1996. Overview of the U.S. JGOFS Bermuda Atlantic Time-series Study and Hydrostation S program. Deep-Sea Research II 43 (2-3), 157–198.

Moore, J. K., Doney, S. C., Kleypas, J. A., Glover, D. M., Fung, I. Y., 2002. An intermediate complexity marine ecosystem model for the global domain. Deep-Sea Research II 49, 403–462.

Murray, J., Johnson, E., Garside, C., 1995. A US JGOFS process study in the equatorial Pacific (EqPac): Introduction. Deep-Sea Research II 45 (2-3), 275–293.

Murtugudde, R., Seager, R., Busalacchi, A. J., 1996. Simulation of the tropical oceans with an ocean GCM coupled to an atmospheric mixed-layer model. Journal of Climate 9, 1795–1815.

Nagata, T., 2000. Microbial Ecology of the Oceans. Wiley-Liss, Inc, Ch. Production mechanisms of dissolved organic matter, pp. 121–151.

Najjar, R. G., Jin, X., Louanchi, F., Aumont, O., Caldeira, K., Doney, S. C., Dutay, J. C., Follows, M. J., Gruber, N., Joos, F., Lindsay, K., Maier-Reimer, E., Matear, R. J., Matsumoto, K., Monfray, P., Mouchet, A., Orr, J. C., Plattner, G. K., Sarmiento, J. L., Schlitzer, R., Slater, R. D., Weirig, M. F., Yamanaka, Y., Yool, A., 2007. Impact of circulation on export production, dissolved organic matter, and dissolved oxygen in the ocean: Results from Phase II of the Ocean Carbon-cycle Model Intercomparison Project (OCMIP-2. Global Biogeochemical Cycles 21 (GB3007), doi:10.1029/2006GB002857.

Najjar, R. G., Sarmiento, J. L., Toggweiler, J. R., 1992. Downward transport and fate of organic matter in the ocean: Simulations with a general circulation model. Global Biogeochemical Cycles 6 (1), 45–76.

Oschlies, A., 2001. Model-derived estimates of new production: New results point to lower values. Deep-Sea Research II 48, 2173–2197.

Oschlies, A., 2002. Can eddies make ocean deserts bloom? Global Biogeochemical Cycles 16 (4), 1106.

Oschlies, A., 2006. Ocean Weather Forecasting. Springer, Ch. On the use of data assimilation in biogeochemical modelling, pp. 525–557.

Oschlies, A., Garçon, V., 1999. An eddy-permitting coupled physical-biological model of the North Atlantic. 1. Sensitivity to advection numerics and mixed layer physics. Global Biogeochemical Cycles 13 (1), 135–160.

Oschlies, A., Schartau, M., 2005. Basin-scale performance of a locally optimized marine ecosystem model. Journal of Marine Research 63, 335–358.

Padisák, J., Soróczki-Pintér, E., Rezner, Z., 2003. Sinking properties of phytoplankton shapes and the relation of form resistance to morphological diversity of plankton - An experimental study. Hydrobiologia 500, 243–257.

Pahlow, M., Oschlies, A., 2009. Chain model of phytoplankton p, n and light colimitation. Marine Ecology Progress Series 376, 69–83.

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

Platt, T., Sathyendranath, S., 1999. Spatial structure of pelagic ecosystem processes in the global ocean. Ecosystems 2, 384–394.

Platt, T. S. S., Longhurst, A., 2000. The changing ocean carbon cycle. Cambridge University Press, Ch. Remote sensing of primary production in the ocean: Promise and fulfilment, pp. 384–394.

Popova, E. E., Coward, A. C., Nurser, G. A., de. Cuevas, B., Fasham, M. J. R., Anderson, T. R., 2006. Mechanisms controlling primary and new production in a global ecosystem model - Part I: Validation of the biological simulation. Ocean Science 2, 249–266.

Press, W. H., Flannery, B. P., Teukolsky, S. A., Vetterling, W. T., 1992. Numerical recipes in Fortran 77: The Art of Scientific Programming, 2nd Edition. Cambridge University Press, Cambridge.

Price, A. R., Myerscough, R. J., Voutchkov, I. I., Marsh, R., Cox, S. J., 2009. Multi-objective optimization of GENIE Earth system models. Philosophical Transactions of the Royal Society A 367 (1898), 2623–2633.

Price, A. R., Voutchkov, I. I., Pound, G. E., Edwards, N. R., Lenton, T. M., Cox, S. J., 2006. Multiobjective tuning of grid-enabled earth system models using a non-dominated sorting genetic algorithm (NSGA-ii). In: Proceedings of the Second IEEE International Conference on e-Science and Grid Computing.

Prunet, P., Minster, J.-F., Ruiz-Pino, D., Dadou, I., 1996. Assimilation of surface data in a one-dimensional physical-biogeochemical model of the surface ocean 1. Method and preliminary results. Global Biogeochemical Cycles 10 (1), 111–138.

Raick, C., Soetaert, K., Grégoire, M., 2006. Model complexity and performance: How far can we simplify? Progress in Oceanography 70, 27–57.

Redfield, A. C., 1934. On the proportions of organic derivatives in sea water and their relation to the composition of plankton. Vol. James Johnstone Memorial Volume, Liverpool. pp. 176–192.

Roman, M. R., Caron, D. A., Kremer, P., Lessard, E. J., Madin, L. P., Malone, T. C., Napp, J. M., Peele, E. R., Youngbluth, M. J., 1995. Spatial and temporal changes in the partitioning of organic carbon in the plankton community of the Sargasso Sea off Bermuda. Deep-Sea Research I 42 (6), 973–922.

Ross, O. N., Geider, R. J., in preparation. Eulerian versus lagrangian modelling of phytoplankton photosynthesis and photoacclimation: new answers to an old question.

Schartau, M., Engel, A., Thoms, S., Volker, C., Wolf-Gladrow, D., 2007. Modelling carbon overconsumption and the formation of extracellular particulate organic carbon. Biogeosciences 4, 433–454.

Schartau, M., Oschlies, A., 2003a. Simultaneous data-based optimization of a 1D-ecosystem model at three locations in the North Atlantic: Part I - Method and parameter estimates. Journal of Marine Research 61, 765–793.

Schartau, M., Oschlies, A., 2003b. Simultaneous data-based optimization of a 1D-ecosystem model at three locations in the North Atlantic: Part II - Standing stocks and nitrogen fluxes. Journal of Marine Research 61, 795–821.

Schartau, M., Oschlies, A., Willebrand, J., 2001. Parameter estimates of a zero-dimensional ecosystem model applying the adjoint method. Deep-Sea Research II 48, 1769–1800.

Schofield, O., Grzymski, J., Bisset, W. P., Kirkpatrick, G. J., Millie, D. F., Moline, M., Roesler, C. S., 1999. Optical monitoring and forecasting systems for harmful algal blooms: Possibility or pipe dream? Phycology 35, 1477–1496.

Sinha, B., Anderson, T. R., Buitenhuis, E. T., Le Quéré, C., in preparation. Comparison of the emergent behavior of a complex ecosystem model in two ocean general circulation models.

Sinha, B., Yool, A., 2006. Extension of the occam 1° general circulation model to include the biogeochemical cycles of carbon and oxygen, Part I: Technical description. Research and Consultancy Report No. 5, National Oceanography Centre, Southampton.

Sivia, D. S., Skilling, J., 2006. Data analysis: A Bayesian tutorial. Oxford University Press.

Smayda, T. J., 1970. The suspension and sinking of phytoplankton in the sea. Oceanography and Marine Biology: An Annual Review 8 (353-414).

Smith, S. L., Codispoti, L. A., Morrison, J. M., Barber, R. T., 1998. The 1994-1996 Arabian Sea Expedition: An integrated, interdisciplinary investigation of the response of the northwestern Indian Ocean to monsoonal forcing. Deep-Sea Research II 45, 1905–1915.

Spitz, Y. H., Moisan, J. R., Abbot, M. R., 2001. Configuring an ecosystem model using data from the Bermuda Atlantic Time Series (BATS). Deep-Sea Research II 48, 1733–1768.

Spitz, Y. H., Moisan, J. R., Abbot, M. R., Richman, J. G., 1998. Data assimilation and a pelagic ecosystem model: Parameterization using time series observations. Journal of Marine Systems 16, 51–68.

Steele, J. H., Henderson, E. W., 1992. The role of predation on plankton models. Journal of Plankton Research 14, 157–172.

Steinberg, D. K., Carlson, C. A., Bates, N. R., Johnson, R. J., Michaels, A. F., Knap, A. H., 2001. Overview of the U.S. JGOFS Bermuda Atlantic Time-series Study (BATS): a decade-scale look at ocean biology and biogeochemistry. Deep-Sea Research II 48, 1405–1447.

Sweby, P. K., 1984. High resolution schemes using flux limiters for hyperbolic conservation laws. SIAM Journal of Numerical Analysis 5 (995-1011).

Tarantola, A., 2005. Inverse Problem Theory and Methods for Model Parameter Estimation. Society for Industrial and Applied Mathematics.

Toggweiler, J. R., 1993. Carbon overconsumption. Nature 363, 210–211.

Wallhead, P. J., Martin, A. P., Srokosz, M. A., Fasham, M. J. R., 2006. Accounting for unresolved spatial variability in marine ecosystems using time lags. Journal of Marine Research 64 (6), 881–914.

Wallhead, P. J., Martin, A. P., Srokosz, M. A., Franks, P. J. S., 2009. Skill assessment via cross-validation and Monte Carlo simulation: An application to Georges Bank plnakton models. Journal of Marine Systems 76, 134–150.

Waniek, J. J., 2003. The role of physical forcing in initiation of spring blooms in the northeast Atlantic. Journal of Marine Systems 39, 57–82.

Ward, B. A., Waniek, J. J., 2007. Phytoplankton growth conditions during autumn and winter in the Irminger Sea, North Atlantic. Marine Ecology Progress Series 334, 47–61.

Williams, R. G., Follows, M. J., 1998. The ekman transfer of nutrients and maintenance of new production over the North Atlantic. Deep-Sea Research I 45, 461–489.

Wroblewski, J. S., Sarmiento, J. L., Flierl, G. R., 1988. An ocean basin scale model of plankton dynamics in the North Atlantic. 1. Solutions for the climatological oceanographic conditions in May. Global Biogeochemical Cycles 2 (3), 199–218.

Yool, A., Martin, A. P., Fernandéz, C., Clark, D. R., 2007. The significance of nitrification for oceanic new production. Nature 447, 999–1002.