Journal of Marine Systems 81 (2010) 34–43

Contents lists available at ScienceDirect

Journal of Marine Systems

journal homepage: www.elsevier.com/locate/jmarsys

Parameter optimisation techniques and the problem of underdetermination in marine biogeochemical models

Ben A. Ward [a,*], Marjorie A.M. Friedrichs [b], Thomas R. Anderson [a], Andreas Oschlies [c]

[a] National Oceanography Centre, Southampton, UK
[b] Virginia Institute of Marine Science, College of William & Mary, Gloucester Point, VA, USA
[c] IFM-GEOMAR, Leibniz Institute of Marine Sciences, Kiel, Germany

ARTICLE INFO

Article history:
Received 2 October 2008
Received in revised form 6 April 2009
Accepted 28 May 2009
Available online 28 December 2009

Keywords:
Genetic algorithm
Variational adjoint
Ecosystem model
Parameter optimisation
Underdetermination

ABSTRACT

Parameter values in marine biogeochemical models can strongly affect model performance, but can be hard to define accurately and precisely. When making quantitative comparisons between models it is helpful to objectively assign optimal parameter values, so it is the best model performance rather than the degree (or lack) of tuning which is assessed. The efficacy of two optimisation techniques, a variational adjoint (VA) and a micro genetic algorithm (μGA), was studied with respect to the calibration of two simple one-dimensional models for Arabian Sea data. Optimisations were randomly initialised a number of times, and given the level of uncertainty in the data, the two techniques performed equally well in terms of reducing model-data misfits. When ten parameters were optimised for either model, the Arabian Sea data were insufficient to constrain unique solutions; several parameters could be set anywhere across a wide range of values while providing a similarly good fit to the data. The significance of this underdetermination was assessed by evaluating model solutions against unassimilated equatorial Pacific data. When no prior information was used to assist the optimisation, the underdetermined solutions led to highly variable and often poor performance at the equatorial Pacific. Prior information was used to gain a more reliable solution in two ways: (1) by fixing all poorly-constrained parameters to their default prior values, optimising only parameters that were well-constrained by the data; or (2) by placing broad limits on the search to exclude unrealistic parameter values. Using the first approach the optimisation routines could constrain unique solutions and model performance in the equatorial Pacific was very consistent. The precise results were however sensitive to the uncertain *a priori* values of the fixed parameters. The second approach was less prescriptive, and consequently led to a more variable performance in the equatorial Pacific. It is argued that the first approach is unrealistically precise as it ignores any uncertainty in the unconstrained parameters, while solutions from the second approach may be unnecessarily broad. In conclusion, unconstrained parameter optimisation procedures should be assisted by stating all that is known *a priori* about the parameters, but no more.

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. The parameterisation of those functions is often set according to the results of laboratory based studies of plankton monocultures (Pahlow and Oschlies, 2009), and as such may not be representative of the diverse range of organisms present in oceanic ecosystems. An alternative approach has been to use optimisation techniques to adjust model parameters, with the aim of minimising the misfit between the model output and a set of *in situ* observations. This method has the advantage that the parameters are assigned values based on the behaviour of all the components in the observed ecosystem. Studies have repeatedly shown, however, that observations are currently insufficient to accurately constrain the number of parameters required by even the simplest ecosystem models (Matear, 1995; Fennel et al., 2001; Friedrichs et al., 2007). For example, three ecosystem models, with between 11 and 20 tunable parameters were optimised by (Matear, 1995) against *in situ* data at Station P in the North Pacific using a simulated annealing algorithm. Analysis of the error covariance matrix for the optimal parameters revealed that, as a consequence of partially-correlated parameters, no more than ten independent parameters could be constrained by the available data. Regardless of the model, the number of parameters that could be constrained was always found to be less than the total number of parameters.

If ecosystem models such as these are to be optimised, it is necessary to find a way of dealing with the many unconstrained parameters. It is also important to know that the optimised parameter values are sensible, and that a model is not getting the right answer for the wrong

* Corresponding author.
E-mail address: benw@mit.edu (B.A. Ward).

0924-7963/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jmarsys.2009.12.005

reasons (Anderson, 2005; Friedrichs et al., 2006). One technique is to provide additional information in the optimisation that attaches a high penalty to parameter values that are thought to be unrealistic. For example, estimates of the probable distribution of each parameter can be included as a penalty term in the model-data misfit function, 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, 2003).

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 the variational adjoint method to demonstrate that the predictive ability of those models was a strong function of the number of parameters optimised. A set of cross-validation experiments illustrated that if too many unconstrained parameters were allowed to vary, the more complex models were unable to reproduce any unassimilated data.

The use of formal parameter optimisation becomes particularly important for studies concerned with making quantitative comparisons between models (Friedrichs et al., 2006, 2007). It would not be fair, for example, to test a model parameterised for the English Channel against one that had been set up in the Sargasso Sea, if the goal was to reproduce biogeochemical observations from the North Atlantic subtropical gyre. Parameter optimisation techniques allow the fair and objective assignment of parameter values, so that any differences in model performance can be attributed to differences in model structure, rather than to the relevance of the parameter values (Friedrichs et al., 2007).

In this paper, 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. Methods

The modelling and optimisation framework was taken directly from Friedrichs et al. (2007), with the addition of the μGA optimisation technique. These aspects are summarised in Sections 2.1 to 2.4. The numerical experiments performed here are described in Section 2.5.

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 Fig. 1. Table 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 makes the use of a diatom-mesozooplankton based system appropriate at this 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* maximum grazing rate parameter for large phytoplankton was increased to 50 d^{-1} (Friedrichs et al., 2007).

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. 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).

2.3. The cost function

Following Friedrichs et al. (2007), 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,

$$J = \frac{1}{M}\sum_{m=1}^{M} W_m^2 \frac{1}{N_m}\sum_{j=1}^{N_m}(a-\hat{a})_{jm}^2. \quad (1)$$

Individual misfits were summed over the number of different data types ($M=4$; nitrate, chlorophyll a, primary productivity and export)

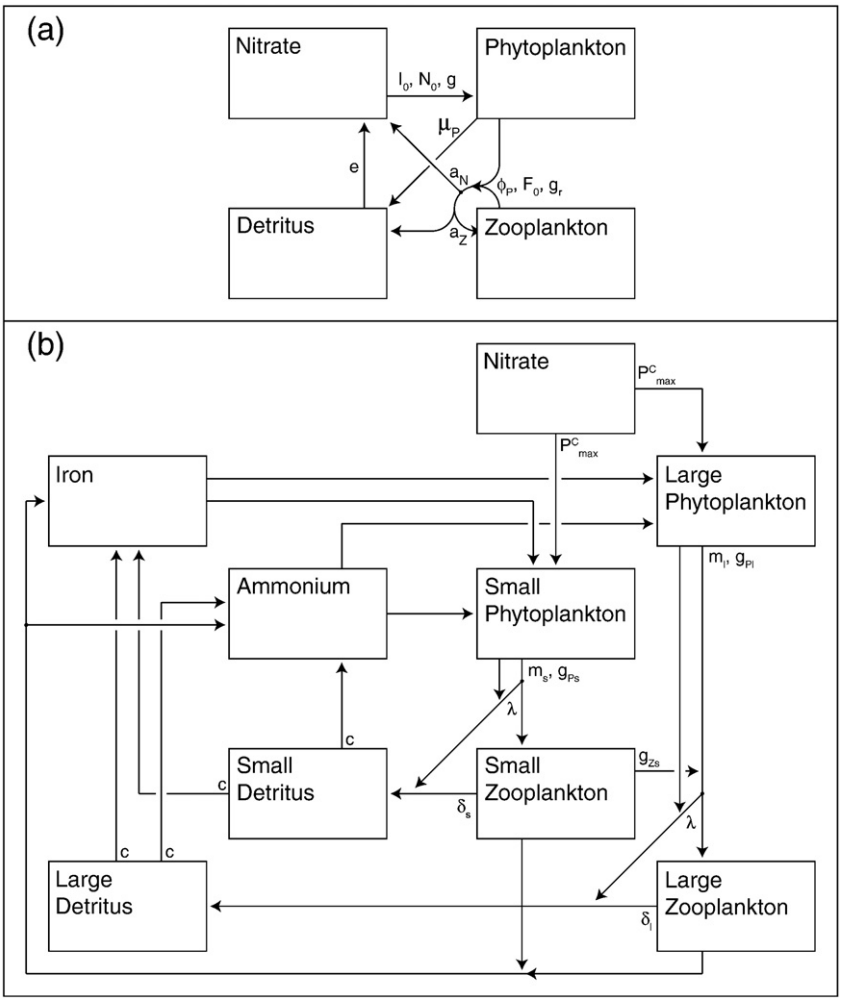

Fig. 1. The four-component (a) and nine-component (b) biogeochemical models. State variables are shown as boxes, fluxes as arrows. The optimised parameters are shown next to the fluxes they describe.

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), such that,

$$W_m = \frac{C_m}{\sigma_m}. \quad (2)$$

The weighting factor C_m (set to 7 for productivity data and 3.5 for all other types) was included to increase the confidence associated with observations of primary production, 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 Eq. (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.

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 (2003). 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.

2.4.1. 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 automatically computing 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

Table 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.

Parameter	Symbol	Default	Range	Units
4-component model McCreary et al. (1996)				
$\frac{1}{2}$ saturation for Z grazing	F_0	1	0 – 5	mmol N m^{-3}
Light saturation constant	$\boldsymbol{I_0}$	40	5 – 80	W m^{-2}
$\frac{1}{2}$ saturation for N uptake	N_0	1	0 – 10	mmol N m^{-3}
Grazing preference coefficient	$\boldsymbol{\phi_p}$	0.83	0 – 1	–
***P* growth rate parameter**	$\boldsymbol{g}$	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)				
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	$\boldsymbol{\lambda}$	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	$\boldsymbol{c}$	0.35	0.1 – 0.5	d^{-1}
Max C specific growth rate	$\boldsymbol{P^C_{max}}$	1.0	0.1 – 2.0	d^{-1}

such as those applied here (Schartau and Oschlies, 2003), 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).

2.4.2. 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. 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 randomly combined into pairs. Within each pair, individuals are compared in terms of model-data misfit, and the less fit parameter vector is discarded. This process of sorting and selecting the parameter vectors is repeated so that the population size is maintained. 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 a broad region of the 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 5000 generations.

Table 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}

2.4.3. 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 (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. In this study, 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 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, as is required for the μGA (Schartau and Oschlies, 2003), 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 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.

2.5. 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) 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 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, 2003). 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 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., 2009).

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 Fig. 2. 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 (Eq. 2), a misfit difference of less than one is not significant, and therefore for each set of optimisations, all solutions with a cost of $J<(J_{min}+1)$ are equally valid. These solutions are subsequently referred to as "acceptable solutions" (Table 3), though as discussed in Section 2.4.3, this does not necessarily imply that all the optimised parameter values are realistic. In Fig. 2 the acceptable solutions are shown as filled black bars. The unfilled bars represent the optimisations that did not converge on the (estimated) global minimum.

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 Fig. 3, 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. Fig. 2, 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 3).

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 3). Although these results held for both the ten-parameter and the 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).

3.1.1. 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 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 W m^{-2}. When the μGA was applied, the optimal parameters, although all realistic, were also highly variable within the defined range (Fig. 3). 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 ± 11.5 and 131.0 ± 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.

3.1.2. 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, $C_v\leq0.0000017\%$) (Fig. 2, row 2), leading to very consistent costs at equatorial Pacific (181.8 ± 0.0). The μGA was equally invariant in terms of the calibrated Arabian Sea cost, but the optimal parameters were much more variable ($C_v\leq18.4\%$), causing significant variability in cost at equatorial Pacific (184.2 ± 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 4.4.

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 Fig. 2. 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

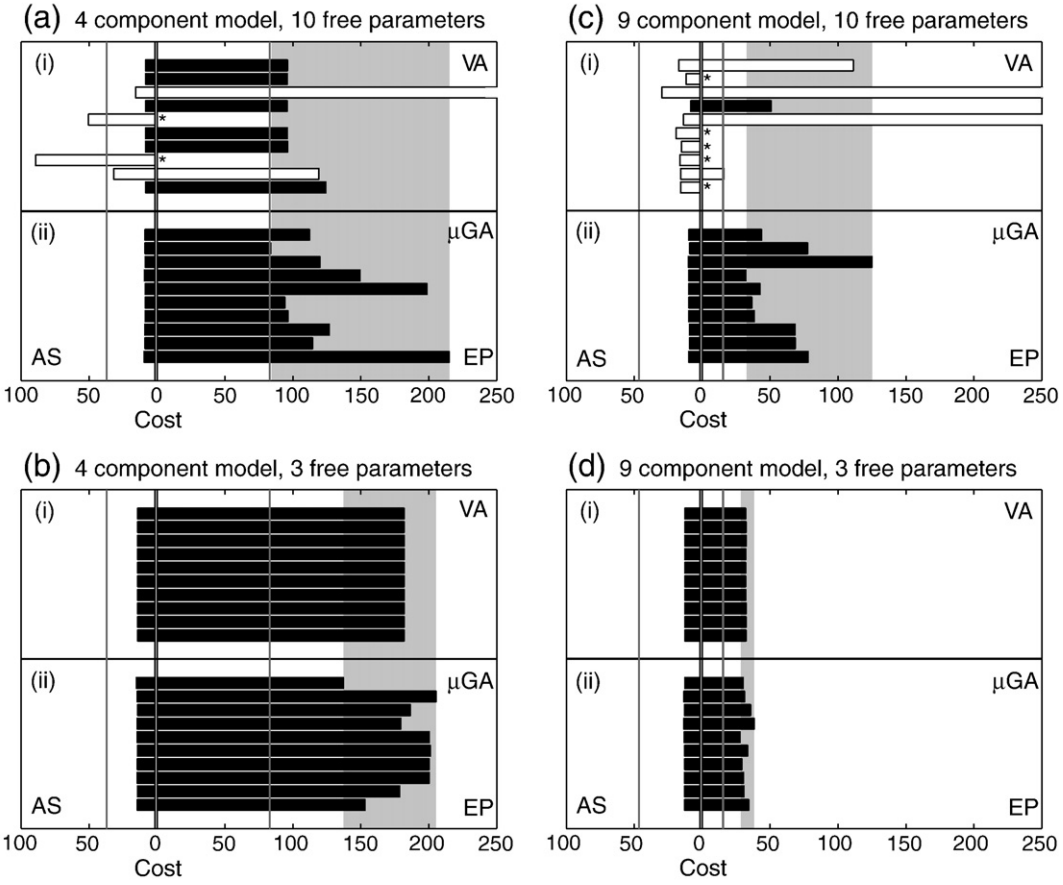

Fig. 2. Cost-function values for models calibrated to Arabian Sea data. The calibrated costs at Arabian Sea are shown on the left-hand side of the vertical axes, costs when the solutions were applied to equatorial Pacific are shown on the right. Unfilled bars represent optimisations that became trapped 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.

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.

3.2.1. 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 ± 28.6. The single acceptable VA solution had a predictive cost of 48.0. None of the techniques yielded solutions that outperformed the prior parameters at the equatorial Pacific.

Table 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.

		Arabian Sea	Equatorial Pacific
4-component model			
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			
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

3.2.2. 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 ± 0.2 and 32.11 ± 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.

3.3. Computational cost

Both techniques required multiple runs of the ecosystem models, and every 1000 iterations took approximately 30 min on a Pentium 4 3.6 GHz

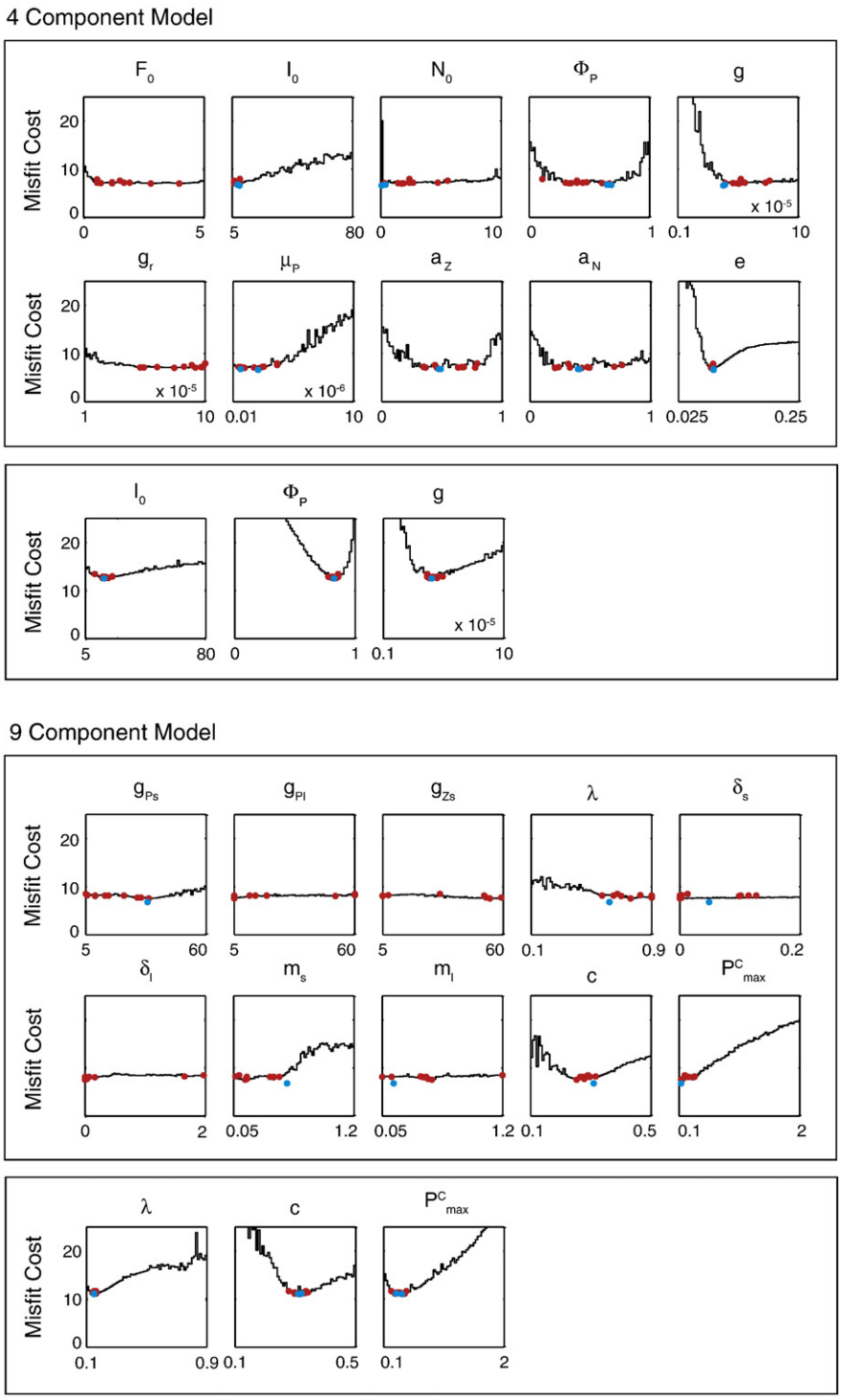

Fig. 3. Collated output from the repeated optimisations. Any 10-dimensional parameter vector from the μGA can be represented by the abscissa of one point appearing in each of the 10 subplots. The shared ordinate of those points corresponds to the associated misfit value. Output from all runs was combined, 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 1 for symbols and units.

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 2649 for 10 parameters, and between 22 and 52 for 3 parameters.

The μGA was run each time for 5000 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 500th generation was not significantly improved upon, even though the algorithm was run for an additional 4500 generations. The longest any μGA search was able to run without any further improvement in the minimum cost was 4964

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.

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.

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. Fig. 3), 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 (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.

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 (Fig. 2a-ii and b-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 4.3), or to fix unconstrained parameters to their default values such that all the remaining parameters are well constrained (Section 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.

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 (Fig. 3). 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, point solutions are unrealistically precise and will be inadequate to describe exactly what the data tell us about the parameters.

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. A useful review can be found in Schartau and Oschlies (2003).

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.

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 ± 28.6 to 32.1 ± 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 ± 44.1 to 181.8 ± 23.0) when only three, rather than ten, parameters were optimised. The grey boxes in Fig. 2 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 2.1). 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.

When ten parameters were optimised in each case, the optimal parameters were much more variable and sensible parameters could not be returned without the incorporation of some prior information. When this was done by fixing unconstrained parameters to precise prior values, the problem became well constrained, but it has been shown that estimates of model predictive skill are sensitive to the uncertain prior values of those parameters. When prior information was introduced through estimates of the credible range of each parameter, estimates of model predictive skill were highly uncertain as a consequence of variability in the unconstrained parameters. In some cases the prior estimates were the only constraint on the optimised parameters.

4.5. Implications

This study focusses on the use of 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 currently underdetermined by observations at oceanic time-series sites, and thus no inverse technique 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 study represent a perfect solution to dealing with uncertainty in parameter optimisation problems. Only by combining good *a priori* and *a posteriori* knowledge of the parameters and their uncertainties will it be possible to make a comprehensive comparison between models. Details of how this might be achieved are beyond the scope of this paper, but one possibility is the use of more sophisticated prior information than either the point estimates or the uniform distributions applied here. Such a method would favour the adoption of a Bayesian approach, where data assimilation can be used within a formal mathematical framework to seek improvements in the prior error estimates of all the uncertain parameter values.

Acknowledgements

The authors wish to thank Markus Schartau and two anonymous reviewers for comments on the manuscript. Ben Ward was funded by the Natural Environment Research Council through a MarQUEST PhD Studentship. He and Tom Anderson wish to acknowledge this support. Marjorie Friedrichs was supported by the National Aeronautics and Space Administration Ocean Biology and Biogeochemistry Program. This paper is Contribution No. 3012 of the Virginia Institute of Marine

Science, The College of William & Mary. The authors wish to thank John Hemmings for helpful comments on the manuscript.

References

Anderson, T.R., 2005. Plankton functional type modelling: running before we can walk? Journal of Plankton Research 27 (11), 1073–1081.
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.
Evans, G.T., 2003. Defining misfit between biogeochemical models and data sets. Journal of Marine Systems 40–41, 49–54.
Fasham, M.J.R., Evans, G.T., 1995. The use of optimization techniques to model marine ecosystem dynamics at the JGOFS station at 47 N 20 W. Philosophical Transactions: Biological Sciences 348 (1324), 203–209.
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.
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., 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.
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.
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.
Gregg, W., Friedrichs, M.A.M., Robinson, A.R., Rose, K., Schlitzer, R., Thompson, K.R., 2009. Skill assessment in ocean biological data assimilation. Journal of Marine Systems 76. doi:10.1016/j.marsys.2008.05.006.
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. Royal Society of London 361, 33–39.
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).
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.
Hurtt, G.C., Armstrong, R.A., 1996. A pelagic ecosystem model calibrated with BATS data. Deep-Sea Research II 43 (2–3), 653–683.
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.
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.
Pahlow, M., Oschlies, A., 2009. Chain model of phytoplankton p, n and light colimitation. Marine Ecology Progress Series 376, 69–83.
Schartau, M., Oschlies, A., 2003. 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., Willebrand, J., 2001. Parameter estimates of a zero-dimensional ecosystem model applying the adjoint method. Deep-Sea Research II 48, 1769–1800.
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.