Ocean Biogeochemical Optimisation in Earth System Models

Sophy Oliver

Oriel College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2021

Acknowledgements

This research was supported by the Natural Environmental Research Council (NE/L002612/1), the Oxford Doctoral Training Partnership in Environmental Research, and the Met Office. I am also grateful for the use of and support provided by the University of Oxford Advanced Research Computing (ARC) facility (`http://dx.doi.org/10.5281/zenodo.22558`).

My greatest thanks must go to my primary supervisor, Samar Khatiwala, for his consistent support, guidance and advice throughout this project. He never deemed any question of mine stupid (or he hid it well), allowing me to always seek help no matter the hour or how busy he was, for which I am very grateful. I would also like to thank my supervisor Coralia Cartis for her unfailing enthusiasm, encouragement and support, particularly in the beginning of this project when she tirelessly helped me to get my non-mathematical brain wrapped around numerical optimisation. Also, to my supervisors Adrian Martin and Andrew Yool, thank you for the insightful discussions, helpful advice and continuous feedback throughout this project.

I would also like to thank Iris Kriest, Simon Tett, Lindon Roberts, Anna Rufas-Blanco, Tom Anderson, Daniel Meyer, those at GEOMAR Helmholtz Centre for Ocean Research and those at UQLab for their help, support and advice at different stages of this project.

Finally, my thanks go to my family and friends, in particular to my mum Carolyn Oliver for forcing me to follow revision timetables through school, to my dad Paul Oliver for building my resilience and focus, and lastly to James Harvey, for always managing to make me laugh when I need it most.

Abstract

Global ocean biogeochemical models are important tools to understand the cycling of nutrients and carbon in the global ocean, and when embedded within Earth System Models allow us to investigate how the oceans interact in the past, present and future with the entire earth system. Biogeochemical parameters within such models can be systematically calibrated by minimising the misfit between modelled tracers and observations, yet due to their large computational expense this is rarely carried out. In this thesis, the aim is to first investigate efficient ways to tune global ocean biogeochemical parameters using optimisation algorithms, so to encourage more frequent objective model improvement. We then apply them to the global ocean biogeochemical model MEDUSA (Model of Ecosystem Dynamics, nutrient Utilisation, Sequestration and Acidification), the marine biogeochemical component of the U.K. Earth System Model (UKESM1) .

First, we carry out a parameter optimisation and sensitivity study on the global ocean biogeochemical model MOPS (Model of Oceanic Pelagic Stoichiometry). We investigate the performance of DFO-LS (Derivative-Free Optimisation using Least Squares), an optimisation algorithm which has been designed for computationally expensive models, and compare it to CMA-ES (Covariance Matrix Adaptation Evolution Strategy), another algorithm that was applied in a previous study in one of the first successful attempts at calibrating a global model. DFO-LS successfully calibrated MOPS in less than $1/30^{th}$ of the computational expense required by CMA-ES, making it feasible to move on to calibrating more computationally expensive models.

We then investigate the influence on parameter optimisation of shortening the length of the ocean biogeochemical spin up, which is required for the model to reach an equilibrated state. If this can be reduced from over 3000 years during optimisation, then the whole process will become less computationally expensive. Unfortunately we show for MOPS the spin up could only be reduced by one third, to a minimum length of 2000 years, due to the influence of certain biogeochemical parameters on the deeper ocean, where timescales are longer.

Lastly, parameter sensitivity and optimisation studies are carried out on the more complex model MEDUSA. We provide an assessment of how different biogeochemical parameters influence global distributions of modelled oceanic variables, and find the

most influential parameters to be those which control diatom and meso-zooplankton abundance, as these heavily influence the transport of organic matter to the deeper ocean. We also provide several optimised configurations of the model with a closer fit to observed nutrients and oxygen than the current default configuration, however with some unrealistic modelled primary production and plankton abundance.

Contents

1
Introduction

1.1 Predicting climate change

1.1.1 Atmospheric carbon dioxide

In 2021, the monthly mean atmospheric carbon dioxide (CO_2) measured at the Mauna Loa Observatory, Hawaii, peaked at 419 ppm (NOAA: Global Monitoring Laboratory, 2021). Historically, throughout the last 800 thousand years of glacials and interglacials, atmospheric CO_2 has fluctuated between approximately 180-300 ppm (Lüthi et al., 2008). However, since the industrial revolution, anthropogenic activities such as burning fossil fuels has driven the atmospheric CO_2 concentrations to beyond these bounds, and is still increasing at a rate much faster than seen before (NOAA: Global Monitoring Laboratory, 2021). CO_2 is a greenhouse gas, which allows incoming long-wave radiation from the sun into our atmosphere, but blocks the outgoing short-wave radiation from leaving our atmosphere, acting to warm our planet, therefore increasing atmospheric CO_2 induces global climate change (Shakun et al., 2012).

1.1.2 Earth System Model predictions

Global climate change has been observed since the 1960s, such as a warming of land and oceans, and a reduction in sea ice, and has been described in multiple reports by the Intergovernmental Panel on Climate Change (IPCC). These reports also include future climate change predictions under different greenhouse gas emission scenarios, written to aid policy makers around the world to help limit global warming.

These predictions are simulated by Earth System Models (ESMs) which try to represent the complex processes within all components of the earth system, and simulate how the system is likely to respond to rising atmospheric CO_2 concentrations. As many different ESMs contribute to IPPC reports, they are compared by the Coupled Model Intercomparison Project (CMIP), now in its 6th phase, which show the spread in predictions modelled by the various ESMs. For example, Figure 1.1 taken from Knutti and Sedláček (2013), shows a comparison of 30 ESMs predicting a global surface warming of approximately 3.5°C to 5°C by 2100 under a "worst case"

scenario (RCP 8.5), where influencing factors such as greenhouse gas emissions, land use change and population growth keep increasing with no regulatory controls.

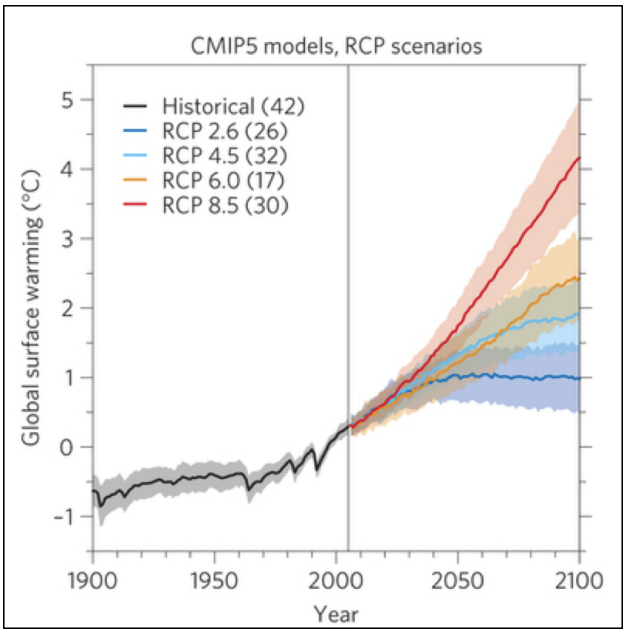

Figure 1.1: Global temperature change (mean and one standard deviation as shading) relative to 1986–2005 for the Representative Concentration Pathways (RCP) scenarios run by CMIP5, taken from Knutti and Sedláček (2013). The number of models is given in brackets. The scale of RCP scenarios goes from RCP 8.5 ("worst case scenario" of highest gas emissions) to RCP 2.6 (lowest gas emissions).

1.2 Ocean biogeochemistry

1.2.1 Ocean carbon pumps

The ocean is a great sink of CO_2 and has absorbed much of the CO_2 emitted by humans since the industrial revolution (see Figure 1.2 by the Australian Government Bureau of Meteorology (2016)). The global mean total inorganic carbon (DIC) is higher in deeper waters than in surface waters (Volk and Hoffert, 1985; Sarmiento and Bender, 1994). The suite of processes which maintain this vertical gradient in the global ocean can be grouped into three different "ocean carbon pumps"; (1) the solubility carbon pump, (2) the organic carbon pump, also known as the soft-tissue pump, and (3) the carbonate pump. The latter two pumps can collectively be known as the biological carbon pump (BCP), which acts to keep atmospheric CO_2

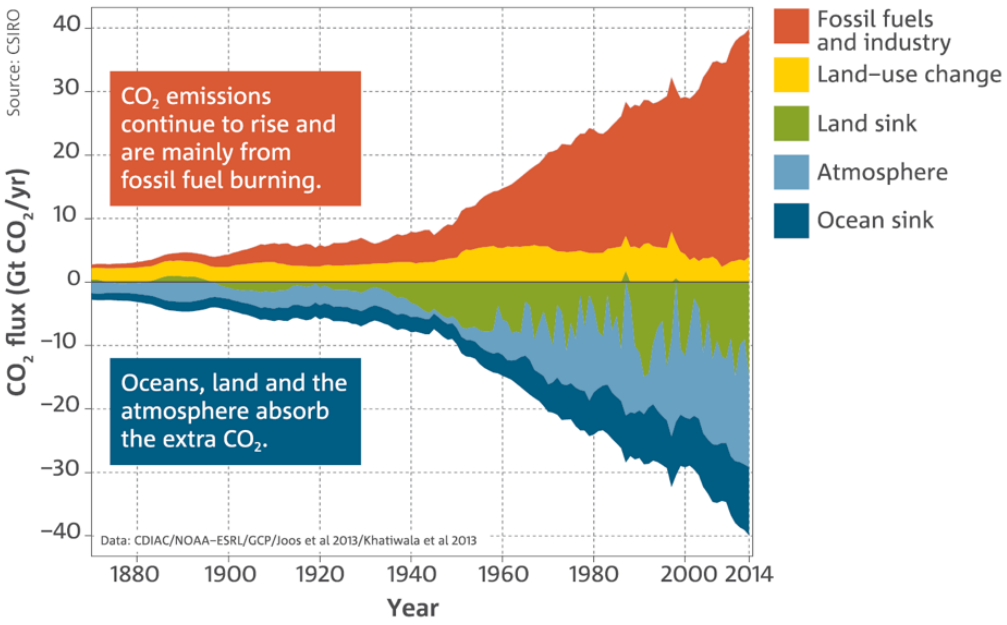

Figure 1.2: Annual fluxes, sources and sinks of CO_2, taken from Australian Government Bureau of Meteorology (2016). Data by Joos et al. (2013) and Khatiwala et al. (2013).

200 pmm lower than it would be otherwise (Parekh et al., 2006). See Figure 1.3, taken from Chisholm (2000), for a schematic of the ocean carbon pumps.

Solubility carbon pump Ocean drawdown of CO_2 is primarily due to the high solubility of CO_2 in seawater. Atmospheric CO_2 both dissolves into the ocean and out-gasses back into the atmosphere. CO_2 is more soluble in colder, more saline water, therefore there is a net flux of CO_2 into the ocean in areas of dense water formation, such as the northern North Atlantic and the Southern Ocean. This dense water, rich in DIC, then sinks to the deeper ocean and is transported to lower latitudes at depth, thereby this DIC content can be kept out of contact with the atmosphere for hundreds to thousands of years before resurfacing (Broecker and Peng, 1992). If this parcel of water resurfaces in a warmer location than where it was last in contact with the atmosphere, then the lower solubility of CO_2 in warmer water causes out-gassing of CO_2 from the ocean back into the atmosphere as it re-equilibrates. It has been suggested that the efficiency of the solubility carbon pump will be reduced due to global warming, as increased stratification will act to inhibit transport of DIC to the deeper ocean (Joos et al., 1999).

Organic carbon pump This pump involves complex interactions between nutrients, plankton and carbon (Turner, 2015). Dissolved carbon is taken up by phytoplankton during photosynthesis, and a percentage of this is then taken up by zooplankton, who feed on phytoplankton. The amount of DIC converted to organic carbon this way is dependent on phytoplankton growth, which in itself is controlled by the availability of light and nutrients. When the the plankton die or defecate, the organic matter sink as particles, called marine snow or detritus. As they sink, bacteria act to break down organic matter back into inorganic dissolved compounds via the process of remineralisation. Therefore, this particle sinking acts to bring organic carbon from the surface ocean into the ocean interior, and the depth at which it remineralises is influenced by the rate at which the particle sink through the water column. Often, the deeper particles remineralise, the longer the remineralised DIC is kept out of contact with the atmosphere. A small part of the sinking detritus may reach the sea floor and get buried. Many different factors play a role in determining the efficiency of the transfer of surface organic matter to the deeper ocean, such as temperature, plankton community structure, particle size, particle aggregation, mineral content, zooplankton vertical migration, and many more (Turner, 2015).

Carbonate pump As well as the organic carbon pump, various species of phytoplankton and zooplankton form inorganic calcium carbonate ($CaCO_3$) shells, such as coccolithophores (Müller, 2019). $CaCO_3$ is typically more dense than organic matter and may provide some protection from remineralisation, causing organic particles enclosed in a $CaCO_3$ shell to sink faster and be remineralised at greater depths (Armstrong et al., 2002; Klaas and Archer, 2002).

1.2.2 Biogeochemical distributions

Surface ocean Ocean biogeochemical distributions are not uniform across the vertical and horizontal space (Lalli and Parsons, 1997; Williams and Follows, 2011), for example see the global nitrate distribution in Figure 1.4. This is due to the

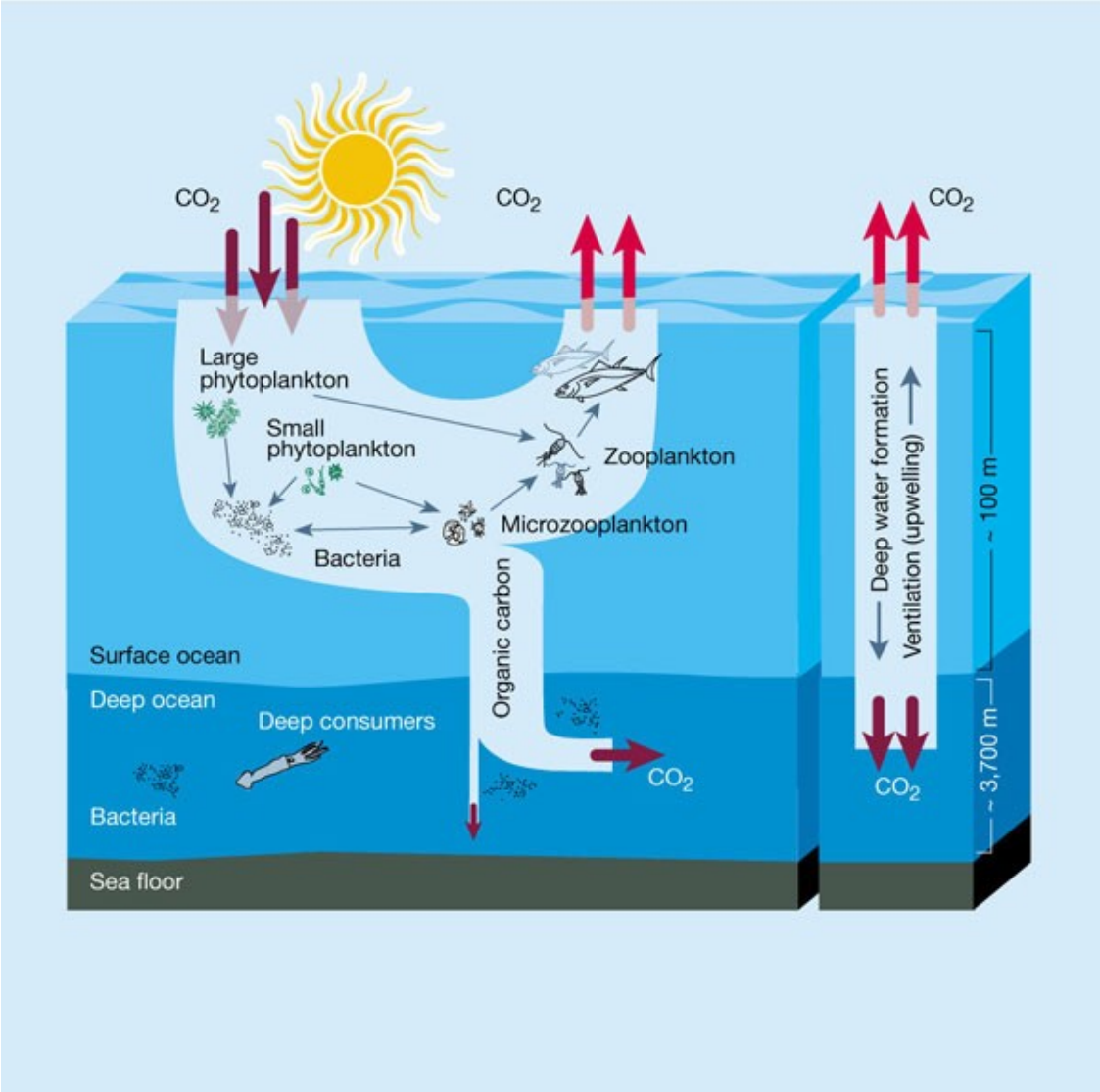

Figure 1.3: Schematic taken from Chisholm (2000), a modified version of the original graphic by Z. Johnson, of the biological (left) and solubility (right) carbon pumps.

spatial variability of factors such as light limitation, local upwelling and downwelling of water masses and riverine inputs. Phytoplankton are limited to the surface ocean due to light availability, therefore nutrients necessary for phytoplankton growth are depleted at the surface, particularly in the mid-ocean gyres where there is convergence of nutrient-depleted surface waters. There are exceptions where there are high concentrations of nutrients such as nitrate and phosphate at the surface, for example regions of upwelling nutrient-rich water, such as off the Peruvian coast, and "high nutrient, low chlorophyl" (HNLC) regions, such as the Southern Ocean,

where phytoplankton growth is limited by iron availability.

Deep ocean Nutrient and oxygen distributions at depth are also heavily determined by oceanic circulation. North Atlantic Deep Water (NADW) carries oxygenated, nutrient-poor water from the surface to the deep Atlantic, while Antarctic Bottom and Intermediate Water (AABW and AAIW) inject oxygenated, nutrient-rich water from the Southern Ocean into the South Atlantic (Chapter 12 of Williams and Follows, 2011). Deep water also becomes more nutrient-rich and depleted in oxygen as it travels along the "deep ocean conveyor" (Broecker, 1991), driven by the Thermohaline Circulation (THC). This occurs because remineralising sinking organic matter continues to add nutrients and strip oxygen from the deep water travelling from the Atlantic Ocean, through to the Indian Ocean, and finally to the North Pacific Ocean. Therefore, nutrients are generally most concentrated and oxygen most depleted in the deep North Pacific.

Oxygen Minimum Zones For oxygen, its global vertical profile is also influenced by sinking organic matter, the majority of which remineralises in the mesopelagic layer, thereby depleting the water of oxygen and creating the oxygen minimum layer. In regions of poor ventilation and high primary production, like the east equatorial Pacific, this oxygen minimum layer can be over 1km thick, called an "Oxygen Minimum Zone" (see Cabré et al., 2015).

1.3 Ocean biogeochemical models

1.3.1 NPZD models

To simulate the biogeochemical processes within the ocean, there are simple nutrient - phytoplankton - zooplankton - detritus (NPZD) models (e.g. Tian et al., 2015). They consist of these four pools, with several equations which control the growth/decay rates of each, and the processes which transfer matter between them (Lalli and Parsons, 1997; Williams and Follows, 2011), as shown in Figure 1.5. This is of course a simplified system, which reduces the effect of the rest of the ecosystem (e.g. higher

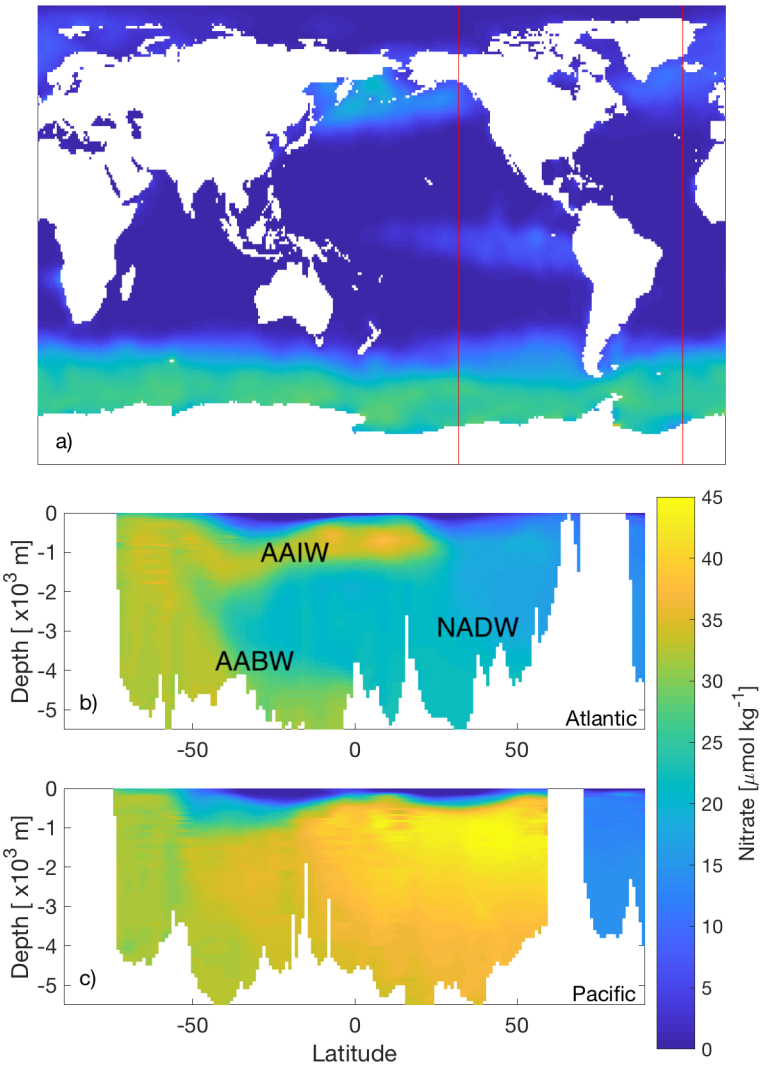

Figure 1.4: World Ocean Atlas interpolated objectively analysed mean concentration of nitrate in sea water [μmol kg^{-1}] (Garcia et al., 2018b), plotted for the (a) global surface, which also shows the locations of the longitudinal transects (red lines) for the nitrate data plotted at (b) 23°W and (c) 140°W. The location of the Antarctic intermediate water (AAIW), Antarctic bottom water (AABW) and North Atlantic deep water (NADW) masses have also been annotated.

predators, bacteria) to very simple parameterisations. These parameterisations are based on strong assumptions, such as zooplankton mortality is represented as a function of zooplankton abundance only, instead of the abundance of higher trophic predators, which aren't represented in the NPZD model (Edwards and Yool.A, 2000).

The nutrient is the core, or currency, of this NPDZ structure. Nitrogen and

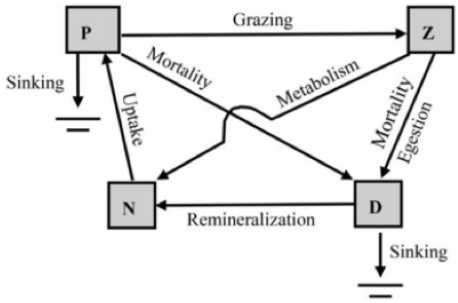

Figure 1.5: NPZD model structure and energy flow (Tian et al., 2015).

phosphate are both good tracers element to use for the NPZD currency, as they indicate the amount of biomass growth for every unit of nutrient uptake by the phytoplankton pool, according to the Redfield Ratio (Redfield, 1934; Gruber and Deutsch, 2014). This states the relative amounts of carbon to nitrogen to phosphorous is 106:16:1 in organic matter. Though a good approximation, it is important to be aware that the Redfield ratio is an average, and therefore masks considerable variability in plankton stoichiometry (Lomas et al., 2021).

1.3.2 Modelled processes

The nutrient (e.g. nitrogen) is taken up by phytoplankton, and is a product of the remineralisation process of detritus and zooplankton metabolism. Phytoplankton biomass is controlled by growth (dependent on light availability, temperature and nutrient limitation), grazing by zooplankton, mortality and sinking. Zooplankton biomass is controlled by growth and mortality rates. Detritus is non-living biomass falling through the water column, therefore is controlled by phytoplankton and zooplankton mortality rates, and zooplankton egestion. Matter leaves the detritus pool by either sinking to the sea floor, or by remineralisation back into dissolved inorganic nitrogen. More complex biogeochemical models build on this NPZD structure, for example by increasing the number of tracers modelled, which may include separating phytoplankton and sinking detrital pools into classes determined by size and particle sinking speed, respectively. Increased complexity allows more biogeochemical processes to be modelled, however they are often more computationally expensive. Increased complexity also brings in additional uncertainty associated

with the extra parameterisations, and it has been shown that increased ocean biogeochemical model complexity does not always ensure a better fit to observations (e.g. Kriest et al., 2010; Ward et al., 2013; Kwiatkowski et al., 2014; Kuhn and Fennel, 2019).

1.3.3 Global models

When ocean biogeochemical models are used to simulate the 3-dimensional global ocean (global ocean biogeochemical models), they need to be run in conjunction with a general circulation model so to include the transport of water with the oceanic currents, which will physically move chemical tracers around the ocean. They need to be spun up to an equilibrated state after which the biogeochemical tracer distributions show little change with time, which takes 3000-10,000 simulated years (Wunsch and Heimbach, 2008).

1.4 Biogeochemical parameterisations

Parameterisations are mathematical expressions which summarise complex processes, instead of modelling them explicitly. Parameters, or constant single values, regulate these parameterisations. These are required to represent the highly complex real world in a simplified and tractable model, and they allow models to account for problems such as grid resolution limitations and certain biological processes which aren't fully described by fundamental equations.

1.4.1 Plankton parameterisations

Processes such as phytoplankton nutrient uptake and growth, and zooplankton grazing can be described by a Michaelis-Menten curve:

$$v = \frac{V_{max}[S]}{K + [S]}, \tag{1.1}$$

where v is the modelled process, such as nutrient uptake or zooplankton grazing, V_{max} the maximum rate of change possible, $[S]$ the availability of the substance

required, such as nutrient concentration or abundance of prey, and K the half-saturation constant.

This creates a functional form as shown in Figure 1.6. As the availability of nutrients increase, the faster the uptake of nutrients by phytoplankton, however this slows at higher nutrient concentrations as the phytoplankton approach their maximum possible rate of uptake (Michaelis and Menten, 1913; Monod, 1949; Droop, 1974; Burmaster, 1979). The parameters in this modelled process are V_{max} and K. If the parameter K were increased, this acts to reduce the gradient of the initial slope of the function curve, causing phytoplankton to have a slower uptake of nutrients at lower concentrations than with a lower K value. If the parameter V_{max} were decreased, phytoplankton uptake nutrients slower at all nutrient concentrations than with a higher V_{max}.

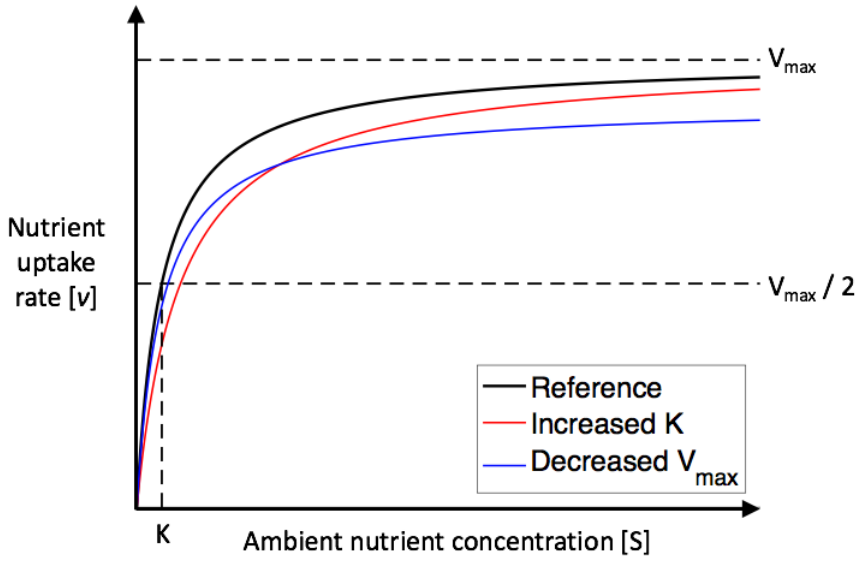

Figure 1.6: Nutrient uptake by phytoplankton as a function of the ambient nutrient concentration in the surrounding water, calculated as in Equation 1.1. At low concentrations uptake is limited by the availability of the nutrient, and at high concentrations uptake is limited by the maximum rate at which phytoplankton can transport nutrients into their cells. A reference function is plotted (black line), for which the corresponding V_{max} and K parameters are shown, and functions with an increased K (red line) and a decreased V_{max}, relative to the reference function, are also plotted.

1.4.2 Particle sinking parameterisations

Particle sinking Another example of ocean biogeochemical parameterisation is the sinking speed of detritus throughout the water column. Small particles often sink slower than larger ones, and remineralisation acts to reduce the size of the particles as they sink. Larger particles (e.g. dead diatoms and zooplankton, egested faecal pellets) may be slightly protected from remineralisation, therefore they may sink without reducing much in size. Therefore, when considering a single individual particle, its sinking may slow with depth as it becomes smaller. However, when considering multiple sinking particles with a distribution of sizes, their combined sinking speed may actually increase with depth because remineralisation removes the smallest particles from the size distribution, while the protected, fast-sinking particles remain (Ploug et al., 2008).

It is important to correctly model particle sinking and remineralisation in ocean biogeochemical models because this is the driver of the biological carbon pump, where sinking particulate organic matter (POC) draws down carbon from the upper ocean to the deep ocean (Turner, 2015).

Common parameterisations Various methods have been used to model sinking particles within ocean biogeochemical models (Kriest and Oschlies, 2008), such as a constant sinking speed with depth; linear increase in sinking speed with depth; particle size-dependent (Gregg et al., 2003) and temperature-dependent (Henson et al., 2011) sinking; according to a remineralisation profile described by a power law (Martin et al., 1987); and according to a ballasting model whereby minerals protect sinking particles from remineralisation (Armstrong et al., 2002). More complex models may use a combination of these, separating detrital pools and modelling their sinking rates differently. See the first column of Table A3 in Cabré et al. (2015) for a description of how different CMIP5 ocean biogeochemical models parameterise the POC flux profile.

Martin b parameter One of the parameters investigated in this thesis is the Martin **b** parameter within the "Martin Curve" power law (Martin et al., 1987) describing the sinking flux of organic matter, fit to observed flux data from 6 stations in the Pacific Ocean. The equation for the Martin Curve is

$$F(z) = F(z_{eu}) \left(\frac{z}{z_{eu}} \right)^{-\mathbf{b}}, \tag{1.2}$$

where $F(z)$ is the flux at depth z, z_{eu} the depth of the base of the euphotic zone within which 99% of sunlight is attenuated, $F(z_{eu})$ the flux leaving the euphotic zone, and **b** was originally estimated to be 0.858 by Martin et al. (1987).

This results in a function as shown in Figure 1.7. There is more sinking organic matter at the surface, which then decreases with depth due to remineralisation. If the remineralisation rate is assumed constant and uniform, when the parameter **b** is increased, more remineralisation occurs closer to the surface, therefore there is less flux at depth. Conversely, when the parameter **b** is decreased, more remineralisation occurs deeper in the water column. In the latter case, more organic matter is transported to the deeper ocean.

The Martin **b** exponent has far reaching consequences for ocean biogeochemistry and climate (e.g. Kwon et al., 2009). The value of 0.858 was derived by Martin et al. (1987) from a handful of POC flux profiles and has been revised by many more studies (e.g. Buesseler et al., 2020; Marsay et al., 2015). However, it does provide a simple global parameterisation to control sinking in ocean biogeochemical models, and is considered a tunable parameter during model calibration.

1.4.3 Calcium carbonate sinking

$CaCO_3$ influence on DIC and total alkalinity Calcium carbonate ($CaCO_3$) is created by certain phytoplankton to form their protective shells (see Section 1.2.1). Both the formation of $CaCO_3$ at the surface, and the dissolution of $CaCO_3$ at depth after sinking influences the amount of DIC, total alkalinity (TA), dissolved CO_2 and pH in the ambient water (see Figure 1.8, taken from (Zeebe and Wolf-Gladrow, 2001)). When one molecule of $CaCO_3$ is formed, this removes one atom of carbon

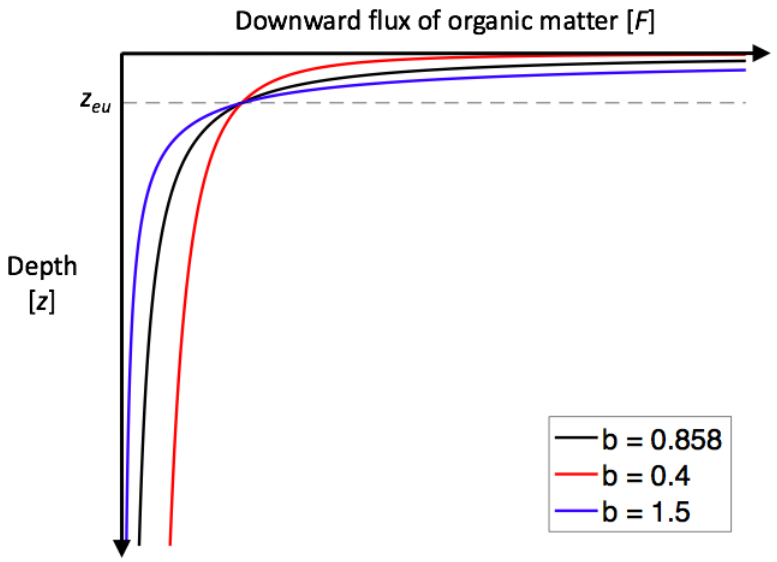

Figure 1.7: Flux of organic matter according to the Martin Curve power law (Martin et al., 1987) as in Equation 1.2, calculated with a **b** value of 0.858 (black line), 0.4 (red line) and 1.5 (blue line).

from the DIC pool, and simultaneously removes one carbonate ion (which has a negative charge of 2). As TA is a measure of the charges in the water, removing a carbonate ion from the dissolved pool decreases the TA of the water. This response to $CaCO_3$ formation acts to lower the pH and shifts the carbonate system of the surrounding water to higher levels of dissolved CO_2.

Calcite saturation The solubility of $CaCO_3$ increases with lower pH and temperature, and higher pressure, such as at higher latitudes and in the deeper ocean. The saturation state of $CaCO_3$ ($\Omega_{calcite}$) describes this global distribution, where values of $\Omega_{calcite}>1$ states the ambient water is supersaturated and promotes $CaCO_3$ formation, while $\Omega_{calcite}<1$ indicates the water is undersaturated and corrosive to $CaCO_3$.

$CaCO_3$:POC The relative fraction of $CaCO_3$ in fast sinking particulate organic carbon (POC) is thus a function of $\Omega_{calcite}$, and can be described by

$$fo(\Omega_{calcite}) = (\Omega_{calcite} - 1)\eta \cdot r0, \tag{1.3}$$

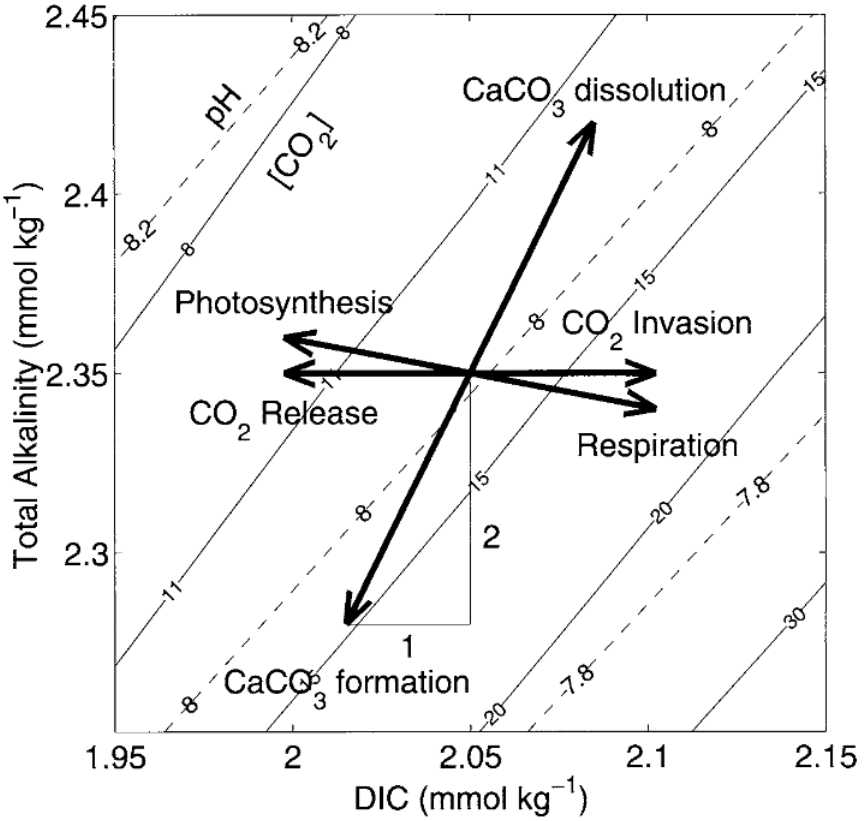

Figure 1.8: The effect of several processes on DIC and TA (Figure 1.1.3 by Zeebe and Wolf-Gladrow, 2001). Shown are levels of constant dissolved CO_2 in μmol kg^{-1} (solid lines) and pH (dashed lines), and the effect of a process on DIC (x-axis) and TA (y-axis) are shown as arrows.

where η is the thermodynamic calcification rate power, and $r0$ the $CaCO_3$:POC export rain ratio scalar, formulated by Ridgwell et al. (2007).

Again, this parameterisation has parameters which could be calibrated, such as the $CaCO_3$:POC export rain ratio scalar, $r0$. If this were to be increased, there would be a greater amount of $CaCO_3$ in sinking POC (see Figure 1.9). As described above (see Figure 1.8), this increased downward transport of $CaCO_3$ would increase the amount of dissolved CO_2 at the surface, and therefore may lead to an increase in atmospheric CO_2 concentration.

1.4.4 Parameter tuning

The biogeochemical parameters described above are only a few of many parameterisations used in ocean biogeochemical models. Each will have been given a value

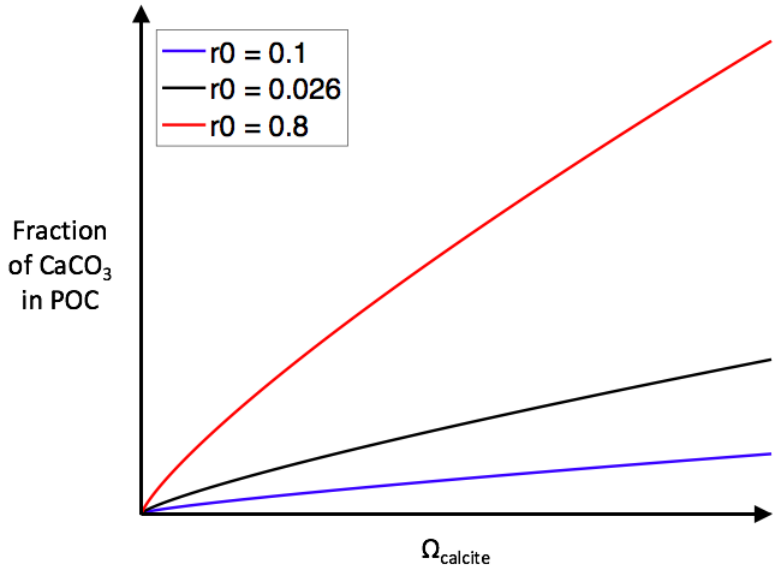

Figure 1.9: Fraction of $CaCO_3$ in sinking POC as a function of ambient $\Omega_{calcite}$, according to Equation 1.3, calculated with an $r0$ value of 0.1 (blue), 0.026 (black) as selected by Yool et al. (2013), and 0.8 (red).

constrained by theory or observations, and will have a range within which lie feasible values for that parameter. These parameters can be optimised by calibrating the model to observations, as changing the parameter values can improve how closely the model simulates the real ocean. This is not dissimilar to tuning a piano, whereby the ocean biogeochemical model is the piano, the biogeochemical parameters the piano keys, and oceanic observations the acoustic tuning fork.

1.5 Quantifying model performance

1.5.1 Model misfit

It is not feasible for global ocean biogeochemical models to be able to simulate absolutely everything within the complex system, therefore the output simulated by a model will always be different to reality. It is the magnitude of this difference which provides insight as to how well the model captures the real system. To measure this, one uses the model to simulate the recent past, then compares this to reliable observations for the same time period. This difference is called the model error, or misfit error. This misfit provides a relative model performance, which can

either be increased or decreased by changing the biogeochemical parameters within the model, corresponding to a worsened or improved model performance.

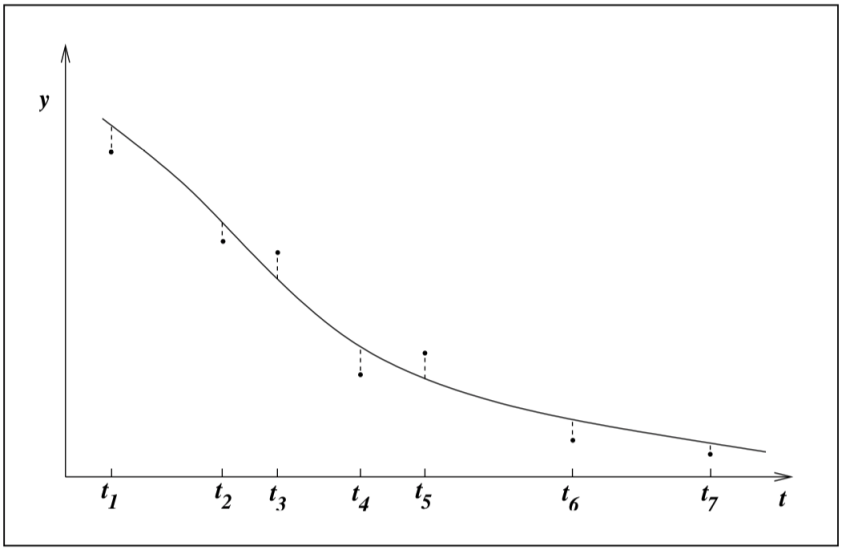

Figure 1.10: Taken from Figure 10.1 of Nocedal and Wright (2006), showing the model (solid line) and the observations (dots) of the concentration (y) of a substance in a patient's blood at time (t) after the patient takes a dose of that substance. The distances between the modelled and observed concentrations are also shown (dashed lines).

1.5.2 Model RMSE

The model misfit is often some version of a root mean squared error (RMSE), which indicates the overall distance between the measured observations and the corresponding modelled values. An example of these observations, modelled values and distances between the two are shown in Figure 1.10, which is taken from Figure 10.1 of Nocedal and Wright (2006).

To calculate the RMSE the differences between the model output and observations are squared so to remove negative values, averaged to provide a mean error, then square rooted to undo the previous squaring (see Chapter 10 of Nocedal and Wright (2006)). Some additional weighting may be required to account for issues such as different model grid point volume sizes. For example, the misfit equation may be

$$f = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(m_i - o_i)^2 w_i}, \tag{1.4}$$

where f is the misfit, N the number of model data points being compared to observations, m the model output, o the corresponding observations, and w the additional weighting given to that data point i.

1.5.3 Observations for calibration

Oceanic observations For the misfit to incorporate the model's success at simulating all aspects of the biogeochemical system, it needs to include the comparison of model outputs to observations of a range of oceanic variables, such as dissolved nutrients, oxygen and carbon concentrations, and plankton abundance. To quantify the model performance of a global ocean biogeochemical model the outputs need to be compared to global observations. Some oceanic variables can be derived from satellite data, such as primary production, however observations below the surface and of variables such as nutrient and oxygen concentrations need to be measured in situ via ship samples or float data, for example.

Available databases The need for global observations has led to international collaborations to gather observations into readily available databases, such as World Ocean Atlas (e.g. Garcia et al., 2018a) and GLODAP (Lauvset et al., 2016) for observations such as oceanic temperature, salinity, dissolved nutrients, oxygen and inorganic carbon, GEOTRACES (Schlitzer and Others, 2018) for dissolved macro and micro nutrients, and COPEPOD (O'Brien, 2014) for plankton observations.

Caveats Observations of some variables are more sparsely scattered than others, for example see Figure 1.11 of the location of dissolved inorganic nitrate observations, which will cause interpolation errors if the scattered observations are interpolated onto the model domain. Other caveats associated with observations include human and calibration error, under-sampling of regions difficult to access by ship and during winter months with harsher sea conditions, and extra parameterisations during the conversion from the measured variable (e.g. ocean colour viewed by satellites) to the target variable (e.g. depth-integrated phytoplankton carbon fixation).

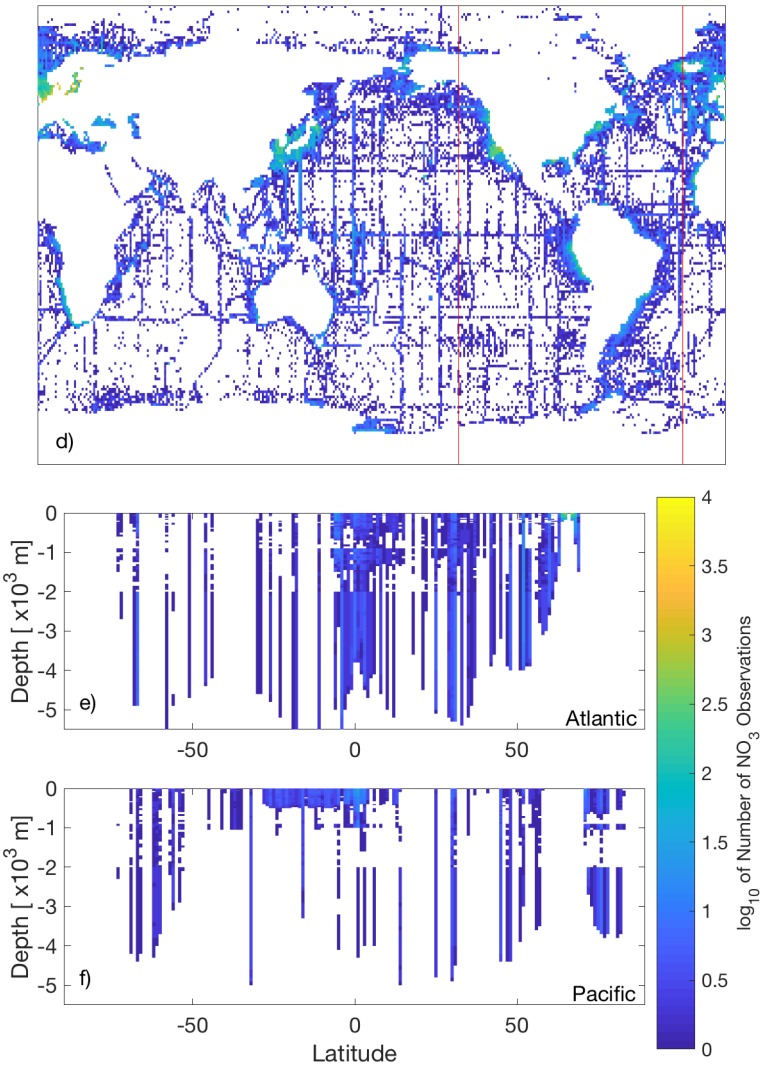

Figure 1.11: Number of World Ocean Atlas nitrate observations (Garcia et al., 2018b) plotted on a $\log_{10}$ colour scale (white oceanic areas show areas of no nitrate observations), plotted for the (a) global surface (0-2.5m depth), which also show the locations of the longitudinal transects (red lines) for the nitrate data plotted at (b) 23°W and (c) 140°W.

1.6 Parameter optimisation

1.6.1 n-dimensional misfit function

The misfit between the model output and real observations will vary as the parameters within the governing equations of the model are varied, as they will determine how well the model simulates reality. The minimisation of this misfit is a least squares problem, whereby the aim is to reduce the distances between the modelled function and the observations. The misfit will vary in n dimensions as n

number of parameters are varied within the model, resulting in an n-dimensional misfit function. As with most functions there will be a global minimum (see Figure 1.12), indicating the set of parameter values which result in the smallest misfit, and hence the best model performance for the chosen misfit function and observations used.

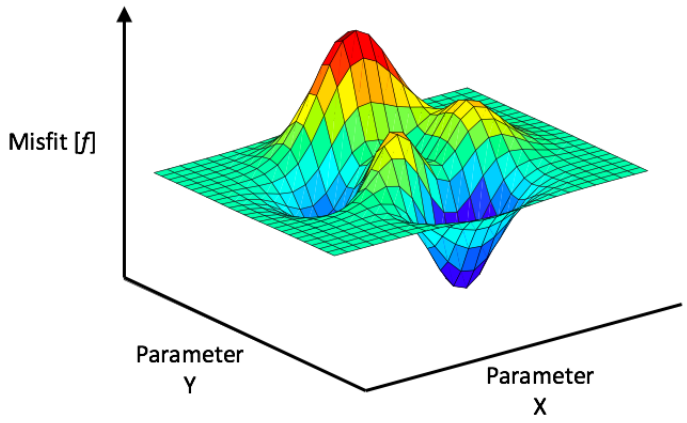

Figure 1.12: Example of the misfit changing with 2 varying model parameters. There are two minima, but only one is the global minimum. Note that this is a hypothetical example function used for explanatory purposes only.

1.6.2 Various tuning methods

To find the minima in the misfit function; for a computationally inexpensive model one could sample the n-dimensional parameter space on a very fine grid, evaluate the model at each of these sampled parameter configurations, then locate which of the samples provides the smallest misfit. However, for large parameter spaces and for computationally expensive models which take a long time to run for just one configuration of model parameters, this is not feasible. Instead, some manual tuning can be carried out, whereby a few parameter values are subjectively changed to gain some reduction in the misfit, or a more systematic method can be used by applying a numerical optimisation algorithm.

1.6.3 Optimisation algorithms

Optimisation algorithms work to locate the minimum of a function (Nocedal and Wright, 2006). They are usually iterative, as they often begin with an initial guess of the unknown optimal parameter values then generate a sequence of improved estimates (or "iterations") until they stop after a predefined number of iterations, or a misfit threshold is reached. The way in which an algorithm moves from one iteration to another is dependent upon the type of algorithm. Some will calculate the first and/or second order derivatives of the function at each iteration's location within the parameter space, so as to choose a direction in which to move down-slope of the function. These are called derivative-based methods, such as the Gauss-Newton line search method (Hartley, 1961). Others may not require true function derivative information at all, which are called derivative-free optimisation methods (Conn et al., 2009), such as those used and described later on in this thesis (see Section 2.3). Some will accumulate information from previous iterations, while others will only use the information at the current point (Nocedal and Wright, 2006).

1.6.4 Local and global solvers

It is desirable for an optimisation algorithm to be both accurate, whereby it identifies the solution close to the truth, and efficient, requiring minimal computational expense. However, increasing the performance in one of these usually results in a sacrifice in the other, as each additional function evaluation gives the optimiser more information about the function landscape within the parameter space, and another attempt to converge closer to the minima. Therefore, one has to find a trade-off between the two when the function is particularly computationally expensive to evaluate.

One aspect which will influence both the computational expense and accuracy of an optimisation algorithm is whether it is a local or a global solver. In the example misfit function shown in Figure 1.12 there are two minima; a smaller ("local") minimum for a lower Parameter Y value, and a larger ("global") minimum for a

higher Parameter Y value. Local optimisation algorithms will search the parameter space for local minima. These algorithms will not guarantee the solution to be the global minimum, and the solution may change depending on where in parameter space the algorithm starts its search. Conversely, global optimisation algorithms carry out a more global search of the parameter space to locate the global minimum. The location of the global minimum is preferred to a local one, however global optimisation algorithms usually require more evaluations of the function, therefore are more computationally expensive.

1.6.5 Multiple solutions

Locating the global optimum or solution is not always the most useful goal. Due to uncertainties in both the model and the observations there can be a range of parameter vectors which result in similarly low misfit values, and therefore these solutions are not significantly better or worse than each other. Instead of finding the global optimum, one can locate several good solutions to evaluate and investigate within the model (e.g. Gregoire et al., 2011). In Chapter 7 where we optimise a model to real oceanic observations, we locate and evaluate several "good" solutions. However, in Chapters 3 and 5 we calibrate to synthetic observations, whereby the model-observation misfit of zero can be achieved, therefore in these chapters we search for the global optimum. In Chapter 3, despite calibrating to synthetic observations, we will show that there are multiple parameter solutions which result in misfit values low enough to be within observational uncertainty had the observations be real and not synthetic, and therefore these solutions are not significantly different.

1.6.6 Derivative-free methods

Due to the large computational expense of ocean biogeochemical models, the efficiency of an optimisation algorithm is very important, as we want to evaluate (i.e., run to equilibrium) the biogeochemical model as few times as possible. An additional challenge these complex models pose is how unsuitable they are to

optimisation algorithms which require derivatives. It is difficult to estimate derivative matrices from ocean biogeochemical models because their internal equations often include unpredictable thresholds, and they are too computationally expensive for the optimisation algorithm to estimate using finite differences by increasing the parameter values by a fixed increment. Even if derivatives were calculated at each iteration of the optimisation they may not be useful, because using too small an increment to produce these derivatives of a non-linear, non-convex misfit function can produce incorrect function curvature information. Instead, derivative-free optimisation algorithms, of which there are several (Rios and Sahinidis, 2013), allow the optimisation of a problem without knowledge of the true function derivatives. As such methods are better suited to expensive ocean biogeochemical models, this thesis will focus on their application. The algorithms used in this thesis are described in Section 2.3.

1.6.7 Previous tuning studies

Atmospheric calibration A number of different approaches have been taken to tuning the parameters of different components of ESMs. Parameters of atmospheric models have been optimised by Severijns and Hazeleger (2005) using the Nelder-Mead simplex method, while Yang et al. (2013) applied the simulated stochastic approximation annealing method (Liang et al., 2014), Zou et al. (2014) using the Multiple Very Fast Simulated Annealing approach (Vakil-Baghmisheh and Navarbaf, 2008), and Tett et al. (2013, 2017) the derivative-free Gauss-Newton line-search variants, a local solver.

Biogeochemical calibration Previous ocean biogeochemical calibration studies have more frequently been carried out on computationally less expensive 0-dimensional (e.g. Kidston et al., 2013) and 1-dimensional models (e.g. Chen and Smith, 2018; Xiao and Friedrichs, 2014; Ward et al., 2010; Spitz et al., 1998), and regional models (e.g. Melbourne-Thomas et al., 2015; Zhao et al., 2005). For less-complex models, an optimisation approach requiring lots of model evaluations can

be applied to tune the parameters, such as the MATLAB function fminsearch (e.g. Kwon and Primeau, 2006; Devries et al., 2014; Holzer et al., 2014). However, for global ocean biogeochemical models the simulation times often need to be reduced before efficient optimisation procedures are applied, as explained in Kriest (2017). A tool to reduce the simulation time is the Transport Matrix Method (TMM; Khatiwala et al., 2005) which removes the need to run an ocean model alongside, by instead applying an "offline" circulation. Another method would be to find the equilibrium solution to the governing equations within the model, using either the inverse Jacobian as in Kwon and Primeau (2006), or a Newton-Krylov method as in Li and Primeau (2008); Khatiwala (2008); Piwonski and Slawig (2016). Once the simulation time has been sufficiently reduced an optimisation algorithm can then be applied, such as the derivative-free optimisation algorithm CMA-ES (e.g. Kriest et al., 2017, 2020; Sauerland et al., 2019; Niemeyer et al., 2019; Kriest, 2017).

Surrogate calibration Optimisation time may be further reduced by surrogate-based optimisation, where the computationally expensive biogeochemical model is replaced with an ensemble of one-dimensional models (e.g. Kuhn and Fennel, 2019) or a physics-based coarse model combined with the TMM (Priess et al., 2013a,b), allowing optimisation of the parameters within the computationally cheaper version.

1.6.8 Parameter sensitivity

Zhang et al. (2015) suggest a three step method to efficient parameter optimisation;

- parameter screening using a sensitivity study,
- determining a good starting point in parameter space,
- applying an optimisation algorithm.

Regarding the first of these, it is useful to carry out a sensitivity analysis of the biogeochemical parameters within an ocean biogeochemical model, as some parameters may have a greater influence on the misfit than others, thereby optimising this parameter may be more worthwhile than optimising a parameter the misfit

function is less sensitive to. Also, a parameter which has little influence on the misfit function is likely to be more difficult to successfully calibrate (Shen et al., 2016). A sensitivity study will also provide insight into the internal processes of the model, and hence marine biogeochemistry as a whole, and how varying certain parameters influence the model outputs.

1.6.9 Local and global parameter sensitivity

To determine the extent to which each model parameter influences the misfit function, one can calculate "local", or "global" sensitivities. Local sensitivities provide information of the sensitivity of the misfit function to each parameter at one, isolated location in parameter space. These can be calculated by perturbing each parameter one at a time by a small amount and calculating the derivative of the misfit function. This is relatively computationally inexpensive, as the misfit function need only be evaluated twice per parameter to calculate each derivative. However, as they only provide information at one point in parameter space, they can lead to misleading conclusions as to which parameters most heavily influence the misfit function over the entire n-dimensional parameter space. For example, in Figure 1.12, the results of a local sensitivity study on this example misfit function would vary significantly depending on the chosen location in the parameter space, with low sensitivities where the misfit landscape is flat, and high sensitivities where the landscape slopes up to a peak or down to a minimum.

Global sensitivities provide a lot more information, but are very costly to compute as they require the misfit function to be evaluated many times. Global sensitivities indicate how strongly each parameter influences the misfit function across the entire n-dimensional parameter space. One simple example of how to calculate the global sensitivity of one parameter would be to first calculate its local sensitivity at multiple locations in parameter space sampled on a fine n-dimensional grid, then take an average. Sampling the entire parameter space is very computationally expensive, therefore local sensitivities are more often calculated for global ocean biogeochemical models (e.g. Kriest et al., 2012), while global sensitivities are carried

out on less computationally expensive models (e.g. Prieur et al., 2019; Wang et al., 2018; Fennel et al., 2001).

1.7 Project aims

Global ocean biogeochemical models are a key tool in understanding the cycling of nutrients and carbon in the global ocean, and are an important component of the Earth System Models (ESMs) used to project future climate change. They are heavily parameterised, yet rarely systematically tuned due to their large computational expense. Efficient and robust optimisation methods are thus of considerable interest to the ocean biogeochemical and broader climate modeling community, and are therefore the main focus of this thesis.

Previous ocean biogeochemical calibration studies have been carried out on computationally less expensive 0-dimensional (e.g. Kidston et al., 2013) and 1-dimensional models (e.g. Chen and Smith, 2018; Xiao and Friedrichs, 2014; Ward et al., 2010; Spitz et al., 1998), regional models (e.g. Melbourne-Thomas et al., 2015; Zhao et al., 2005), or steady-state global models (e.g. Kwon and Primeau, 2006, 2008). However, with the aid of fast "offline" circulation schemes, such as the Transport Matrix Method (Khatiwala et al., 2005; Li and Primeau, 2008) which can be applied to time-dependent biogeochemical models, more recently, complex global ocean biogeochemical models have also begun to be systematically optimised to observations (e.g. Kriest et al., 2017, 2020; Sauerland et al., 2019; Niemeyer et al., 2019; Kriest, 2017).

One of the first successful attempts at systematic tuning of a global biogeochemical model was completed by Kriest et al. (2017) using the optimisation algorithm "Covariance Matrix Adaptation Evolution Strategy" (CMA-ES; Hansen, 2016). While the development and application of CMA-ES is an important step forward, it is still computationally too expensive for routine use. In Kriest et al. (2017), for example, the misfit function had to be evaluated at least 950 times to achieve a sufficiently low misfit to observations, which is prohibitively expensive for the more complex models run at resolutions typical of the current generation of ESMs.

Therefore, the first aim of this project was to explore the application of another, computationally less expensive algorithm, "Derivative Free Optimisation by Least Squares" (DFO-LS; Cartis et al., 2019), to the same problem set-up as in Kriest et al. (2017) on the global biogeochemical model MOPS (Kriest and Oschlies, 2015). This work is presented in Chapter 3.

Further aims were to improve the efficiency and success of global ocean biogeochemical optimisation, such as with the use of a parameter sensitivity study to determine which ocean biogeochemical parameter should be including during optimisation, and the reduction of the ocean biogeochemical spin up length during optimisation. These results are reported in Chapters 4-5.

Lastly, upon the availability of a TMM interface (Khatiwala et al., 2021) for the global ocean biogeochemical model of "intermediate" complexity called MEDUSA (Yool et al., 2013), the efficient optimisation methods investigated first on MOPS could then be applied to the more complex model. This will have wider relevance to the modelling community as MEDUSA is the chosen ocean biogeochemical component of the UK's contribution to CMIP6, the Earth System Model UKESM1 (Sellar et al., 2019). This latter work is presented in Chapters 6-7.

This project consists of interdisciplinary work, crossing the boundaries of both ocean biogeochemistry and numerical optimisation, where we apply very recent optimisation tools to topically relevant climate models.

2
Methods

2.1 Selected ocean biogeochemical models

2.1.1 Model of Oceanic Pelagic Stoichiometry (MOPS)

The global ocean biogeochemical model investigated in Chapters 3 - 5 of this thesis was the Model of Oceanic Pelagic Stoichiometry (MOPS), developed by Kriest and Oschlies (2013, 2015).

Development of MOPS MOPS originates from "NPZ-DOP" (Kriest et al., 2010), a five-component model with a phosphorus core, which simulates nutrients, detritus, dissolved organic matter (DOP), phytoplankton and zooplankton. A new feature was then added to NPZ-DOP by (Kriest and Oschlies, 2013) to include burial of detritus at the sea floor, with the loss of phosphorus then balanced by riverine input. This acted to reduce the sensitivity of oxygen to changes in the remineralisation profile, as allowing some particles to be buried in the sea bed means less organic matter remineralises in the water column to influence the oxygen distribution. NPZ-DOP was then further built upon to create MOPS-1.0 by (Kriest and Oschlies, 2015), with two main new features. One feature was the inclusion of oxidant-dependent remineralisation, whereas in NPZ-DOP remineralisation continued in the absence of oxygen. The second feature was the explicit representation of the nitrogen cycle, bringing this to a seven-component model. A schematic of the structure of MOPS-1.0 is provided in Figure 2.1, as in Kriest and Oschlies (2015). Some modifications were made to MOPS-1.0 to allow it to be used within an optimisation framework by Kriest et al. (2017), specifically to read in biogeochemical parameter values and compute the misfit to observations, creating MOPS-2.0. There are approximately 35 parameters within MOPS-2.0, regulating stoichiometry, nitrogen fixation, plankton, detritus, burial and oxic degradation.

NPZ-DOP studies Kriest et al. (2010) used NPZ-DOP model and several less complex versions to investigate the influence model complexity has on model performance, concluding that a higher complexity does not always lead to improved

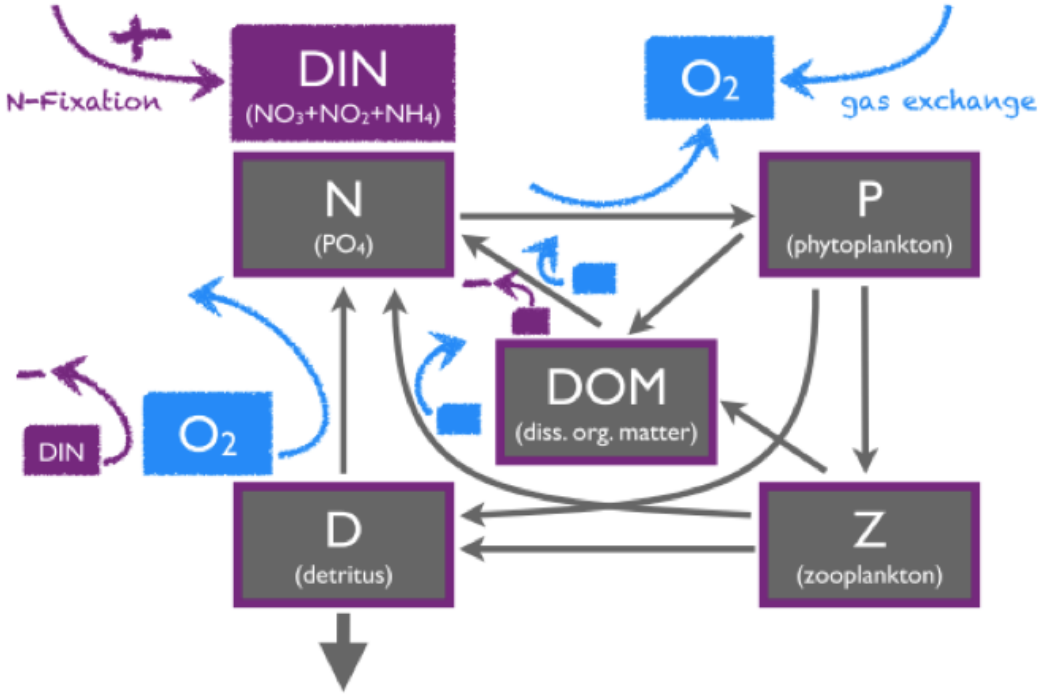

Figure 2.1: MOPS-1.0 model structure of the phosphorous core (in grey), oxygen fluxes (in blue) and nitrogen fluxes (in purple). Nitrogen is coupled to the phosphorous core, indicated by the purple borders. The purple plus and minus signs show fixed nitrogen loss and gain (Kriest and Oschlies, 2015).

performance, and instead parameter optimisation of less complex models can lead to a greater improvement in model performance. Then in a sensitivity study of NPZ-DOP by Kriest et al. (2012) they found the three most influential biogeochemical parameters on the misfit function between modelled and observed global phosphate and oxygen were the Martin **b** parameter, the maximum growth rate for phytoplankton, and the half-saturation constant for light limitation.

MOPS studies Kriest et al. (2017) optimised 6 biogeochemical parameters to observations of phosphate, nitrate and oxygen. The Martin **b** was robustly optimised from 0.858 to 1.1-1.2, however in optimisation experiments with wide parameter bounds the zooplankton and phytoplankton parameters were optimised to a solution where phytoplankton thrived while zooplankton were almost extinct, which was solved by narrowing the bounds. MOPS has since been used to investigate how to improve the modelled locations of oxygen minimum zones (Sauerland et al.,

2019; Niemeyer et al., 2019), and the influence the chosen general circulation has on biogeochemical parameter optimisation (Kriest et al., 2020). The latter study provided the optimal MOPS parameter configuration to be used within the Earth System Model FOCI, into which MOPS has recently been embedded (Matthes et al., 2020).

Why MOPS? MOPS was primarily chosen to be used in this thesis because it had already been adapted to be run with an "offline" general circulation (see Section 2.2.1). This offline circulation greatly reduces the computational expense of MOPS, making it suitable for an optimisation study whereby the model needs to be evaluated multiple times. MOPS was also chosen because it has been used in a previous optimisation study by Kriest et al. (2017) where they applied the optimisation algorithm CMA-ES (see Section 2.3.1), to which we wanted to compare the success and efficiency of another algorithm DFO-LS (see Section 2.3.2). MOPS was also a good model to use when testing our chosen methods because it is a global ocean biogeochemical model which simulates well the global distribution of biogeochemical tracers, yet is less complex and computationally expensive than many other global biogeochemical models, such as MEDUSA which will be described in the following section.

2.1.2 Model of Ecosystem Dynamics, nutrient Utilisation, Sequestration and Acidification (MEDUSA)

Chapters 6 and 7 of this thesis then move on to investigate a more complex global ocean biogeochemical model, namely the Model of Ecosystem Dynamics, nutrient Utilisation, Sequestration and Acidification (MEDUSA: Yool et al., 2011, 2013).

Development of MEDUSA MEDUSA was first introduced as MEDUSA-1.0 by (Yool et al., 2011). This is a plankton ecosystem model of "intermediate complexity", with a nitrogen core and able to include the silicon and iron biogeochemical cycles within the global ocean. They divided the plankton groups into relatively small and large size classes, and allow for both slow and fast sinking detritus. This model

was then built upon to include representation of both the organic and inorganic carbon processes (Yool et al., 2013), allowing it to fully capture the carbon cycle and its interaction with atmospheric carbon dioxide and ocean acidification.

Model structure The model structure of MEDUSA-2.0 is shown in Figure 2.2 (Yool et al., 2013). MEDUSA-2.0 resolves 15 state variables for the 3-dimensional ocean. Of these, 6 use nitrogen to describe the two classes of both phytoplankton and zooplankton, slow sinking detrital nitrogen and dissolved nitrate, and two describe the chlorophyll within both classes of phytoplankton. Of the rest, two use silicon to describe diatom phytoplankton and dissolved silicate, then there is

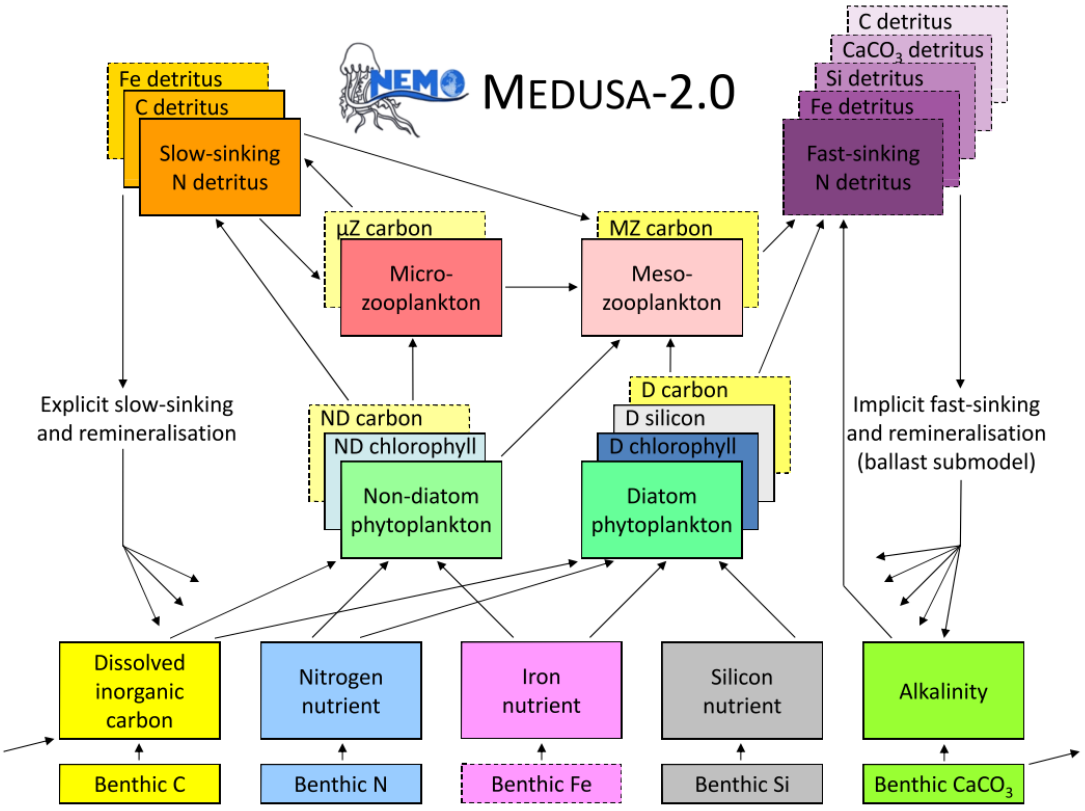

Figure 2.2: Model structure and interactions of MEDUSA-2.0, as seen in Yool et al. (2013). State variables are either modelled explicitly (boxes with solid borders) or implicitly (dashed borders). Overlapping boxes of non-diatom phytoplankton (ND), diatom phytoplankton (D), micro-zooplankton (μZ) and meso-zooplankton (MZ) indicate multiple modelled variables for these components. The benthic reservoirs at the bottom are fed by slow- and/or fast-sinking detritus, and release their respective dissolved matter into the ocean (with benthic $CaCO_3$ releasing both alkalinity and dissolved inorganic carbon). Note that dissolved oxygen interactions are not shown in this schematic.

one state variable each for iron, dissolved inorganic carbon (DIC), total alkalinity, dissolved oxygen and slow sinking detrital carbon. In addition to these, 4 state variables describe the 2-dimensional pools of nitrogen, carbon, silicon and calcium carbonate ($CaCO_3$) in the seafloor.

Tracer inventories MEDUSA-2.0 has fixed inventories of nitrogen, silicon and alkalinity, as there are no inputs or outputs into the system such as via river runoff or nitrogen fixation. However, this is not the case for iron, carbon and oxygen. Iron is removed via scavenging and input via dust deposition and dissolution from the benthos, while carbon and oxygen interact with the atmosphere via air-sea gas exchange.

Plankton interactions In MEDUSA, both size classes of phytoplankton take up nutrients and carbon from the surrounding seawater, with their growth limited by nutrient and light availability. The smaller size class of phytoplankton (non-diatoms) are then grazed on by both micro and meso-zooplankton, and the larger size class (diatoms) are grazed on by meso-zooplankton.

Slow-sinking detritus Non-diatom phytoplankton and micro-zooplankton losses, and a fraction of diatom and meso-zooplankton losses contribute to the slow sinking detrital pool. This is grazed on by both classes of zooplankton while sinking at a constant rate and remineralising at a temperature-dependant rate according to microbial respiration.

Fast-sinking detritus Diatom and meso-zooplankton losses contribute to the fast-sinking detrital pool, whose attenuation throughout the water column is modelled implicitly (whereby sinking and remineralisation is instantaneous) by one of three modes. The default mode is to determine the attenuation of the flux profile by a ballast model (Armstrong et al., 2002), whereby some fraction of the sinking material is protected from remineralisation by the ballasting minerals biogenic silica and calcium carbonate. The two other sinking modes allow the

attenuation of the fast sinking detritus according to either the Martin power law (Martin et al., 1987) or a variation of this power law where the parameter **b** in the exponent is a function of local surface temperature (Henson et al., 2012).

MEDUSA studies Alongside MEDUSA the physical model which provides the ocean general circulation and sea-ice interactions has been NEMO (Madec, 2008). Due to the large combined computational expense of this configuration, and a choice of over 70 parameters to optimise, only minor manual tuning of several parameters was carried out during the transition from MEDUSA-1.0 to MEDUSA-2.0 (Yool et al., 2013). MEDUSA has been used in many studies to provide insight into global ocean biogeochemistry. For example, NEMO-MEDUSA has been used to simulate how Arctic productivity may be influenced by retreating sea ice due to global warming (Lawrence et al., 2015; Yool et al., 2015), and to investigate the controls, pathways and efficiency of transporting organic matter and carbon from the upper ocean to the deeper ocean (e.g. Henson et al., 2015; Baker et al., 2017; Aldama-Campino et al., 2020). The outputs of NEMO-MEDUSA provide globally gridded ocean biogeochemical data which is useful to force other models, such as seafloor POC flux to force a benthic biomass model to understand how seafloor heterotrophs may respond to climate change (Yool et al., 2017), and various output data was used to force a carbon-isotope model to predict distributions of carbon isotope composition of phytoplankton (Magozzi et al., 2017).

Why MEDUSA? During this thesis we move on from MOPS to MEDUSA because it is a more complex model, which has not yet undergone systematic parameter tuning. MEDUSA-2.0 was also chosen by the Met Office to be the ocean biogeochemical component of the Earth System Model UKESM1 (Sellar et al., 2019; Yool et al., 2020, 2021), after a comparison of 6 possible candidates for the task (Kwiatkowski et al., 2014), therefore it may have wider relevance to the climate modelling community.

2.2 General circulation framework

To understand the transport of biological and chemical oceanic tracers such as nutrients (e.g. phosphate, nitrate) and carbon, knowledge of the ocean currents are necessary. This can be done by combining ocean biogeochemical models with ocean general circulation models (GCM), allowing the transport of oceanic tracers to be explicitly calculated according to the volume transport of water by the modelled ocean currents, such as NEMO-MEDUSA described in the previous section. However, ocean biogeochemical models require a long spin up time of several thousand years before reaching equilibrium. The cost of computing a GCM alongside this long spin up time is very large and time consuming, therefore an efficient yet accurate alternative has been sought (e.g. Khatiwala et al., 2005; Primeau, 2005).

2.2.1 The Transport Matrix Method

The Transport Matrix Method (TMM) developed by Khatiwala et al. (2005) and described by Khatiwala (2007) allows efficient "offline" simulation of passive tracers by replacing the space- and time-discretized advection-diffusion equation by a sequence of sparse matrix-vector products:

$$\begin{aligned} &\mathbf{c}^{n+1} = \mathbf{A}_\mathrm{i}{}^{n}(\mathbf{A}_\mathrm{e}{}^{n}\mathbf{c}^{n} + \mathbf{q}^{n}) \\ &n = \text{time step index} \\ &\mathbf{c} = \text{tracer concentration} \\ &\mathbf{q} = \text{sources and sinks of the tracer} \\ &\mathbf{A}_\mathrm{i} = \text{Implicit TM} \\ &\mathbf{A}_\mathrm{e} = \text{Explicit-in-time TM} \end{aligned} \tag{2.1}$$

Here, the $\mathbf{A}_\mathrm{i}$ and $\mathbf{A}_\mathrm{e}$ represent the Transport Matrix (TM) at this point in time, which is a combination of the discretised advection-diffusion operator and any parameterised processes on the sub-grid scale. To ensure stability at relatively large time steps, GCMs explicitly calculate horizontal advection and diffusion, yet implicitly calculate mixing and advection in the vertical. Therefore the TM needs to be split into the two separate $\mathbf{A}_\mathrm{i}$ and $\mathbf{A}_\mathrm{e}$ components. These two sparse matrices

can be calculated for every time step by integrating forward a GCM with a suite of passive tracers, saving the two matrices, resetting the tracers to a particular known spatial pattern after each time step, and repeating.

Time-averaging For computational efficiency these can be time-averaged into an annually averaged $\mathbf{A}_\mathrm{i}$'s and $\mathbf{A}_\mathrm{e}$'s, or into 12 monthly averaged ones to include the circulation variability associated with the seasonal cycle. Using linear interpolation and these monthly averages the TM at any time instant can be calculated, without any further use of the GCM.

Efficient ocean biogeochemical modelling The TMM can be used to drive any ocean biogeochemical model, and is typically orders of magnitude faster than running the same model "online" alongside the GCM. This is very useful when multiple simulations of the ocean biogeochemical model are required, such as sensitivity analyses and parameter optimisation studies. Studies which have applied the TMM in this context include Kriest et al. (2017) and Sauerland et al. (2019) with the MOPS global ocean biogeochemical model.

2.2.2 General circulation models

Available transport matrices At the time of writing this thesis there were 6 different pre-extracted transport matrices (available for download at `http://kelvin.earth.ox.ac.uk/spk/Research/TMM/TransportMatrixConfigs/`), three of which are derived from the MITgcm general circulation model (Marshall et al., 1997) and three from the ocean component of version 2.9 of the University of Victoria Earth System Climate Model (UVic ESCM: Weaver et al., 2001; Eby et al., 2009). They each provide monthly mean transport matrices, and other physical required data such as oceanic temperature, salinity and wind speed. At the time of writing, work was undergoing to extract transport matrices from the NEMO circulation model (Madec, 2008), which would be particularly useful when considering the MEDUSA ocean biogeochemical model with which it has historically been coupled,

however they are not yet available. Therefore several of the available transport matrices were used in this thesis instead.

MITgcm ocean model The transport matrices used to provide the circulation for the MOPS ocean biogeochemical model in Chapters 3-5 of this thesis were extracted from the MITgcm ocean model (Marshall et al., 1997) configured with a horizontal resolution of 2.8° x 2.8° degrees and 15 vertical levels, with a closed boundary at 80°N (therefore the Arctic Ocean is not represented) (Dutkiewicz et al., 2005). These transport matrices were chosen for Chapters 3-5 so to remain consistent with a previous ocean biogeochemical parameter optimisation study (Kriest et al., 2017), for which Chapter 3 in particular is a comparison of optimisation methods.

UVic ocean model For Chapters 6-7 the predominantly used transport matrices were derived from the ocean model from version 2.9 of the University of Victoria Earth System Climate Model (Weaver et al., 2001; Eby et al., 2009), as extracted by Khatiwala et al. (2019). This has a resolution of 1.8° x 3.6° and 19 ocean depth layers, with an ocean time step of 8 hours.

ECCO ocean model Finally, the highest resolution transport matrices used in this thesis were extracted from the ECCO ocean general circulation model, which is derived from the MITgcm model (Marshall et al., 1997) and has been optimised to observations (Stammer et al., 2004). This has a resolution of 1° x 1° over ±80° latitude with 23 vertical levels, and an ocean time step of 12 hours. These circulations have been compared by Kriest et al. (2020) in a study showing how ocean biogeochemical parameter optimisation can depend on the chosen general circulation model. Due to its higher resolution the ECCO circulation is the most computationally expensive option of the three described transport matrices, therefore it is used sparingly in this thesis, despite a better overall agreement with observations.

2.3 Optimisation techniques

The calibration of ocean biogeochemical models involves minimising the misfit between the modelled tracers and oceanic observations (see Sections 1.5 and 1.6). To do this, optimisation algorithms can be applied. Described below are two such algorithms applied in this thesis. Both of the following algorithms can be applied to a "black box", whereby the optimiser only needs to communicate with the model parameters and misfit, while the inner workings of the model of interest (i.e. the biogeochemical model) do not need to be known, such as derivatives, initial conditions, model equations, etc. (see Figure 2.5 in Section 2.4).

2.3.1 CMA-ES

One of the optimisation algorithms used in Chapter 3 of this thesis is the Covariance Matrix Adaptation Evolution Strategy (CMA-ES; Hansen, 2016), which is a widely used stochastic, evolutionary, derivative-free algorithm used to solve the problem

$$\min_{x \in \mathbb{R}^n} \quad f(x),$$

where f is the function to minimise. By design, CMA-ES is an unconstrained solver, that is, parameters are not restricted to be within specified bounds. To ensure that parameters lie within reasonable bounds, a penalty term is added to the misfit when any parameter value goes outside of their specified range.

How CMA-ES works The process of CMA-ES is illustrated in Figure 2.3 (Kriest et al., 2017). During each iteration CMA-ES evaluates the misfit function multiple times with various parameter configurations and updates a covariance matrix. It returns only the parameter configurations which provide the best misfits to a multivariate normal distribution of parameters, then in the next iteration it randomly draws several more parameter configurations, and repeats. With each iteration, the population should be guided towards areas of the parameter landscape which provide lower expected misfit values, and therefore aim to converge on the parameter configurations which provide the best misfits.

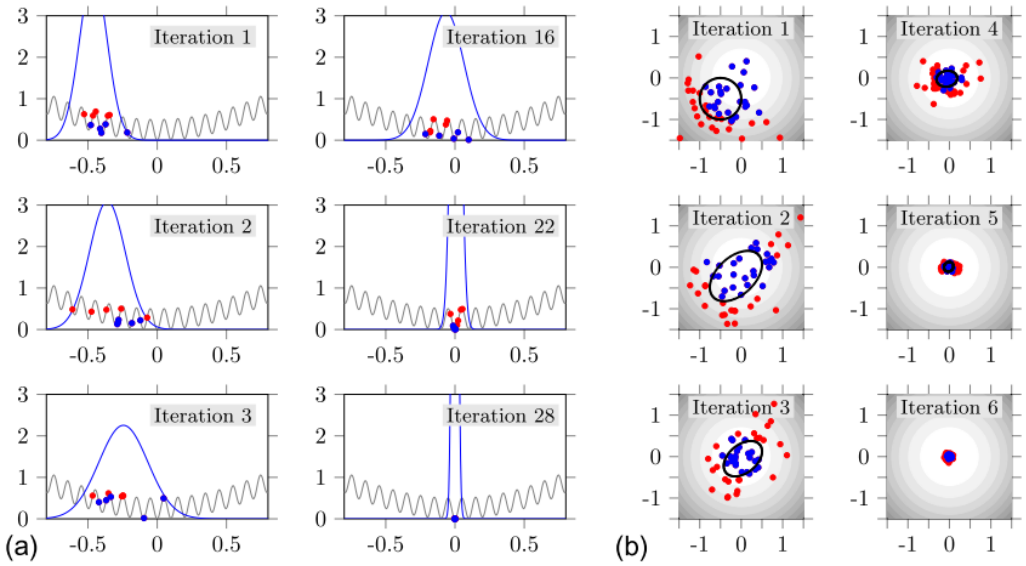

Figure 2.3: Figure by (Kriest et al., 2017) of the CMA-ES optimisation process applied to test functions (grey lines). The plots in (a) show how during each iteration a sample population (dots) was chosen from the normal distribution (blue curve) of parameter values on the x-axis, with their corresponding misfits on the y-axis. Half of the samples yielding the worst misfits (red dots) were removed from the next iteration's normal distribution. This is also shown in the (b) plots in two-dimensional parameter space, when applied to a simple sphere function. The black line indicates the updating distribution by its standard deviation ellipse, and again the samples are coloured dots.

As CMA-ES searches randomly the entire parameter space, it is therefore more likely to converge to a global optimum. In order to achieve this it requires many function evaluations, and therefore can be quite computationally expensive in practice. The CMA-ES code used in this thesis, which is based on the (μ/μ_{W},λ)-CMA-ES algorithm of Hansen (2016), was sourced from Kriest et al. (2017).

2.3.1.1 Mathematical Description

The following description follows Hansen (2016). The candidate biogeochemical parameter vectors are sampled from a multi-variate normal distribution:

$$\begin{aligned} &\mathcal{N}(\mathbf{m}, \mathbf{C}) \\ &\mathbf{m} = \text{Vector of mean parameter values} \\ &\mathbf{C} = \text{Positive definite matrix of covariances.} \end{aligned} \tag{2.2}$$

The distribution is sampled for every generation number k. Each generation can be considered one iteration of the algorithm, within which the distribution is sampled

λ times:

$$
\begin{aligned}
&{x_g}^{k+1} \sim \mathbf{m}^k + \sigma^k \mathcal{N}(\mathbf{0}, \mathbf{C}^k) \text{ for } g = 1, ..., \lambda \\
&{x_g}^{k+1} = \text{Parameter vector for generation } k{+}1 \\
&\mathbf{m}^k = \text{Mean of distribution at generation } k \\
&\sigma^k = \text{Overall standard deviation, or "step size", at generation } k \\
&\mathbf{C} = \text{Covariance matrix at generation } k \\
&\lambda = \text{Population size.}
\end{aligned}
\tag{2.3}
$$

After each generation, $\mathbf{m}^{k+1}$, $\mathbf{C}^{k+1}$ and σ^{k+1} need to be calculated before the equation can be repeated. This corresponds to determining the new mean of the search distribution, the updated covariance matrix and the new step size.

The new mean is created by updating the previous mean, according to a weighted average of the best μ (where $\mu = \lambda/2$) sampled points from the previous population (${x_{1...\lambda}}^{k+1}$), which are ordered and weighted according to their performance, as in Equation 2.4.

$$
\begin{aligned}
&\mathbf{m}^{k+1} = \mathbf{m}^k + \sum_{i=1}^{\mu} w_i (x_{i:\lambda}^{k+1} - \mathbf{m}^k) \\
&\sum_{i=1}^{\mu} w_i = 1, \; w_1 \geq w_2 \geq ... \geq w_\mu \geq 0 \\
&\mathbf{m}^{k+1} = \text{Updated mean of distribution} \\
&\mathbf{m}^k = \text{Mean of distribution at previous generation} \\
&\mu = \text{Number of selected good/positive points, e.g. } \lambda/2 \\
&w_{i=1...\mu} = \text{Positive weights, favouring the better selected points} \\
&x_{i:\lambda}^{k+1} = \text{The } i\text{-th best point from } x_1^{k+1} ... x_\lambda^{k+1}
\end{aligned}
\tag{2.4}
$$

Note that $i : \lambda$ corresponds to the i-th ranked point, ranked according to the misfit function to be minimised (with the lowest misfit being i_1).

To enable a fast search, we calculate $\mathbf{C}^{k+1}$ using the Rank-μ-Update as in Section 3.2 of Hansen (2016) which requires a relatively small population size, whereby the covariance matrix is calculated from information from previous generations, but is more heavily weighted to most recent generations. The step-size σ^{k+1}, which

controls the overall scale of the distribution, is calculated as in Section 4 of Hansen (2016).

2.3.1.2 Algorithm Description

Below is a description of the algorithm of $(\mu/\mu_{\mathrm{W}},\lambda)$-CMA-ES algorithm used in this thesis. A more extensive description is provided by Hansen (2016).

Algorithm 1 $(\mu/\mu_{\mathrm{W}},\lambda)$-CMA-ES

0: INPUT: Set initial parameters as in Table 1 of Kriest et al. (2017), population size $\lambda = 10$, $\mu = \lambda/2$, evolution paths, covariance matrix C=I, distribution mean, step size and maximum generation number.
1: **while** maximum generation is not reached and fitness distribution is not flat **do**
2: **Sample population of new probability distribution**
3: **for** $g = 0, 1, 2, ...\lambda$ **do**
4: Sample search point for this g
5: **end for**
6: **Update probability distribution**
7: Update the mean of the search distribution according to a weighted average of the best half of the previously sampled population
8: Update the overall standard deviation ("step size")
9: Update covariance matrix
10: **end while**

2.3.2 DFO-LS

The main optimisation algorithm used in this thesis is a Python implementation of the Derivative-Free Optimisation using Least Squares (DFO-LS) algorithm (Cartis et al., 2019; Roberts, 2019), which can be found here: `https://github.com/numericalalgorithmsgroup/dfols`. DFO-LS is an iterative method for minimising a function $f(\mathbf{x})$ with bounds on the variables $\mathbf{x}$. It can be applied to a black box general objective problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \quad f(\mathbf{x}) := ||\mathbf{r}(\mathbf{x})||^2, \tag{2.5}$$

where f has continuous derivatives, yet its evaluations can be stochastic and inexact, and where $\mathbf{r}$ is a vector of residuals measuring the misfit between the model and

observations. Instead of using information from just f, like other optimisation algorithms such as CMA-ES and BOBYQA (Powell, 2009), DFO-LS can take in more information from its problem structure via the misfit vector of terms $\mathbf{r}$.

Quadratic approximation The terms of $\mathbf{r}$ are used to calculate a local quadratic approximation of f using several initial evaluations of the true function. Then DFO-LS uses the derivatives of this quadratic approximation to locate a minimum. Hence, DFO-LS does not require derivatives of the true function, but of the approximation. This is extremely efficient when f is a computationally expensive climate model, as calculating derivatives of f takes a lot of time to compute, whereas the time taken to compute derivatives of a known quadratic function is negligible by comparison.

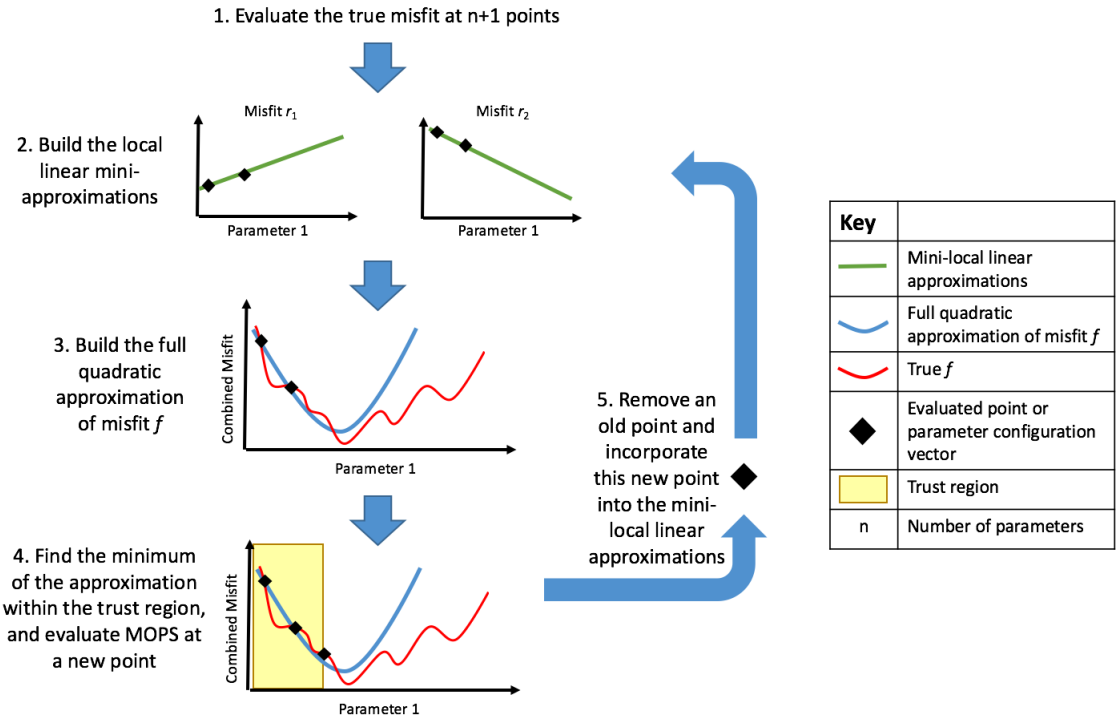

Figure 2.4: Schematic of DFO-LS optimising one parameter. 1) Two individual misfits r_i are evaluated at two locations in parameter space. 2) Two mini-local regressions are built (green lines) using these two r_i points (black diamonds). 3) Information from these are combined to build a quadratic approximation (blue line) to the true misfit function. 4) Within the trust-region (shaded in yellow) the minimum of the approximation is found, at which the true misfit function is evaluated. 5) If accepted, this new information is used to update the mini-local regressions. Steps 2-5 are then repeated until the specified termination criterion or maximum evaluations is reached.

Initiation The algorithm is illustrated in Fig. 2.4. DFO-LS must first be given a starting location within the parameter space from which to initialise. In the initial iteration of DFO-LS, the misfit function is evaluated at the user-specified starting location, and at n additional locations nearby, for a total of $n + 1$ function evaluations. In subsequent iterations only one function evaluation is needed. This is important to note because it means the computational expense of the initial sampling by DFO-LS increases only linearly with the number of parameters to be optimised.

How DFO-LS works Using the initial set of evaluated points, DFO-LS creates a quadratic approximation to the underlying true (unknown) misfit function (Cartis et al., 2019). DFO-LS then calculates the minimum of this function within a "trust region" centred around the starting point, with a default normalised radius of 0.1. The true misfit function is then evaluated at this location. If it is found to be worse than the misfit at the existing $n + 1$ points, it is rejected, the trust region is shrunk according to pre-defined trust region management DFO-LS settings, and the procedure is repeated. On the other hand, if it is found to be lower than the best of the $n+1$ points then it is accepted, and the point corresponding to the highest misfit amongst the previous $n + 1$ points is discarded. A new quadratic approximation is calculated or updated for this $n + 1$ set of points, and the procedure is repeated. Thus, at any iteration DFO-LS keeps track of $n + 1$ points in parameter space and the point with the lowest misfit is considered as that iteration's best set of parameters.

Termination The algorithm is terminated based on three specified criteria: 1) the maximum allowed number of function evaluations is exceeded, 2) the trust region radius is shrunk below a specified size, or 3) misfit reduction progress is identified as being too slow.

Improving geometry During the optimisation process the algorithm monitors the quality of the quadratic approximation, so as to prevent it from becoming degenerate and not representing the true cost function well across all dimensions of parameter space. This would occur for example if some of the sampled points fell on the same plane within parameter space, causing another plane to not be sufficiently sampled. If DFO-LS notices it becoming degenerate, it will undergo a geometry phase whereby it will shift one or more evaluated points to a location in the parameter space which improves the geometry, and re-evaluate the true function at this point.

Globalising DFO-LS DFO-LS is a "local" optimisation method (see Section 1.6.4). However, there is strong numerical evidence from the derivative-free optimisation algorithm Py-BOBYQA, upon which DFO-LS is based (Cartis et al., 2018), that it is able to find global minima. To increase the likelihood of finding the global minimum, DFO-LS can either be manually re-initiated from different starting locations in the parameter space, or automatically "restarted" once it determines that the reduction in the misfit is progressing too slowly. During a restart the trust region expands, allowing DFO-LS to search for points potentially outside the local minimum it may be trapped in, and move towards a lower minimum value elsewhere. This can be done by either a "hard" restart, whereby the (expensive) misfit function is re-evaluated at $n + 1$ new locations within the expanded trust region, or by a "soft" restart, whereby DFO-LS only "shifts" some of the current $n + 1$ points in parameter space to geometry-improving points (Cartis et al., 2019). The former is much more computationally expensive than a soft restart, therefore soft restarts are predominantly used in this thesis. To increase confidence that DFO-LS has found the global minimum, in this thesis DFO-LS is often initiated from multiple starting points.

Misfit vector Regarding the $\mathbf{r}$ misfit vector in Equation 2.5 given to DFO-LS, the length of the vector must be greater than or equal to 1, and there is no maximum vector length. However, to keep the system of linear equations to be

solved in Equation 2.10 from being underdetermined (with fewer observations than parameters to be determined), the length of this vector should be larger than the number of parameters being optimised. DFO-LS can solve an underdetermined problem if necessary, but it is better to have more terms of $\mathbf{r}$ than unknowns. Therefore, the preferred minimum length of $\mathbf{r}$ is $n + 1$. Despite there not being a maximum preferred length, there is usually a limit to how many terms of $\mathbf{r}$ that can be provided, as they should be independent of each other. Also, there are problems associated with being overdetermined, whereby there are too many observations for a simplistic model to ever manage to successfully capture.

Formulating the misfit One could provide a misfit for each grid point of the ocean biogeochemical model being optimised, providing an $\mathbf{r}$ vector length of many thousands. However, many of the terms of $\mathbf{r}$ which are physically close to each other in the model will respond very similarly to perturbations in the biogeochemical parameters being optimised, which will essentially result in a heavier weighting to this location of the ocean model. Therefore, the misfit terms of $\mathbf{r}$ should be provided in such a way that they respond to parameter perturbations independently of each other, such as by providing misfits for different observational types (e.g. nitrate, oxygen, phosphate, silicate, etc), for different spatial regions (e.g. North Atlantic, Southern Ocean, etc) and for different depths in the water column (e.g. 0-1000m, 2000-4000m, etc).

2.3.2.1 Mathematical Description

The mathematical description of the solver DFO-LS is provided in Cartis et al. (2019), and is summarised below.

The problem to solve is the nonlinear least-squares problem in Equation 2.5, which can be further written as

$$\min_{\mathbf{x}\in\mathbb{R}^n} \quad f(\mathbf{x}) := ||\mathbf{r}(\mathbf{x})||^2 = \sum_{i=1}^{m} r_i(\mathbf{x})^2, \tag{2.6}$$

where $\mathbf{r}(\mathbf{x}) := [r_1(\mathbf{x})...r_m(\mathbf{x})]^T$ is a function from $\mathbb{R}^n$ to $\mathbb{R}^m$ with unavailable, but well-defined, derivatives. In our case, n is the number of tunable parameters and m is the number of misfit values. Note that DFO-LS essentially squares the example misfit in Equation 1.4 so to make the problem more smooth.

Using a set of $n+1$ points

$$Y_k = \{\mathbf{x}_k, y_1, ..., y_{\mathrm{n+1}}\} \subset \mathbb{R}^n, \tag{2.7}$$

DFO-LS creates m mini-local models, one for each r_i. It builds a vector model $\mathbf{m}_k$

$$\mathbf{r}(\mathbf{x}_k + s) \approx \mathbf{m}_k(\mathbf{s}) := \mathbf{r}_k + J_k\mathbf{s}, \tag{2.8}$$

where J_k approximates the Jacobian of $\mathbf{r}$ at $\mathbf{x}_k$. It does this by solving the regression problem

$$\min_{\mathbf{r}_k, J_k} \quad \sum_{t=0}^{n} ||\mathbf{m}_k(\mathbf{y}_t - \mathbf{x}_k) - \mathbf{r}(\mathbf{y}_t)||^2. \tag{2.9}$$

This corresponds to finding the least square solutions to those mini-local models, which are overdetermined linear systems

$$\mathbf{W}_k \begin{bmatrix} r_{\mathrm{k,i}} \\ \mathbf{j}_{\mathrm{k,i}} \end{bmatrix} := \begin{bmatrix} 1 & (\mathbf{y}_0 - \mathbf{x}_k)^T \\ \vdots & \vdots \\ 1 & (\mathbf{y}_n - \mathbf{x}_k)^T \end{bmatrix} \begin{bmatrix} r_{\mathrm{k,i}} \\ \mathbf{j}_{\mathrm{k,i}} \end{bmatrix} = \begin{bmatrix} r_i(\mathbf{y}_0) \\ \vdots \\ r_i(\mathbf{y}_n) \end{bmatrix}, \tag{2.10}$$

where $i = 1, ... ,m$.

After the vector model $\mathbf{m}_k$ is built, the quadratic model m_k for the full function $f(\mathbf{x})$ is defined as

$$f(\mathbf{x}_k + \mathbf{s}) \approx m_k(\mathbf{s}) := ||\mathbf{m}_k(\mathbf{s})||^2 = ||\mathbf{r}_k||^2 + \mathbf{g}_k^T\mathbf{s} + \frac{1}{2}\mathbf{s}^T H_k\mathbf{s}, \tag{2.11}$$

where $\mathbf{g}_k := 2J_k^T\mathbf{r}_k$ (the approximation of the gradient) and $H_k := 2J_k^T J_k$ (the curvature information).

DFO-LS maintains a radius parameter $\Delta_k > 0$, within which m_k is expected to be a good local approximation for f in $B(\mathbf{x}_k, \Delta_k)$. This is the trust region.

After calculating a step $\mathbf{s}_k$ within the trust region by solving the trust region subproblem

$$\mathbf{s}_k \approx \underset{||\mathbf{s}|| \leq \Delta_k}{\arg\min}\, m_k(\mathbf{s}), \tag{2.12}$$

the algorithm evaluates $f(\mathbf{x}_k + \mathbf{s}_k)$. If

$$\mathbf{r}_k = \frac{f(\mathbf{x}_k) - f(\mathbf{x}_k + \mathbf{s}_k)}{m_k(0) - m_k(\mathbf{s}_k)}, \tag{2.13}$$

is sufficiently large (meaning the objective has been sufficiently decreased), then this new step is accepted and

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k, \tag{2.14}$$

and Δ_k is increased. Otherwise, if there is no sufficient decrease, the step is rejected and

$$\mathbf{x}_{k+1} = \mathbf{x}_k, \tag{2.15}$$

and Δ_k is decreased. Y_k is then updated to include $\mathbf{x}_{k+1}$ and points in Y_k may be moved to improve geometry (Cartis et al., 2019).

Each subsequent iteration includes constructing the quadratic model $m_k(\mathbf{s})$ and calculating and accepting/rejecting a new step $\mathbf{s}_k$. If all values within Y_k are within the specified noise level of $f(\mathbf{x}_k)$ or specified slow progress has been recognised, then either termination or a restart can be called.

For a soft restart the trust region radius is reset ($\Delta_k = \Delta_0$) and the best $\mathbf{x}_k$ is moved to a geometry-improving point (Cartis et al., 2019) within the larger trust region (shifting the trust region to this new point). The other points within Y_k closest to the old best $\mathbf{x}_k$ are also moved to geometry-improving points within the new trust region.

The iteration then continues from whichever of the new points has the lowest objective value.

2.3.2.2 Algorithm Description

Below is a brief description of the algorithm of DFO-LS. A more extensive description is provided in Cartis et al. (2019).

Algorithm 2 DFO-LS

0: INPUT: Number of parameters n, starting point $\mathbf{x}_0 \in \mathbb{R}^n$, minimum trust region radius (p_{end}), if hard or soft restarts are allowed, and maximum number of true misfit function evaluations.
1: Evaluate the true misfit function at $n+1$ points within the initial trust region to build the initial interpolation set $\{\mathbf{Y}_0\}$ (this can be done in parallel).
2: **for** $k = 0, 1, 2, ...$ **do**
3: **if** we have exceeded the maximum number of true misfit function evaluations **then**
4: terminate.
5: **end if**
6: Construct a quadratic approximation of the true misfit function (see Eq. 2.8-2.11).
7: Approximately solve the trust region subproblem (Eq. 2.12) to locate the minimum of the approximation within the trust region and get step $\mathbf{s}_k$ to this point.
8: Evaluate the true misfit function at $\mathbf{x}_k + \mathbf{s}_k$.
9: **if** the misfit is sufficiently decreased in the sense of Eq. 2.13 **then**
10: Accept Step:
11: Set $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k$.
12: **if** slow progress is observed **then**
13: call a hard or soft restart if allowed, or terminate.
14: **end if**
15: Form $\{\mathbf{Y}_{k+1}\}$ by replacing the worst point with the new accepted point to maintain a set of $n+1$ points.
16: **else**
17: Reject Step:
18: Set $\mathbf{x}_{k+1} = \mathbf{x}_k$ and shrink the trust region.
19: Check geometry.
20: **if** the trust region radius is smaller than p_{end} **then**
21: call a hard or soft restart if allowed, or terminate.
22: **end if**
23: Let $\{\mathbf{Y}_{k+1}\} = \{\mathbf{Y}_k\}$.
24: **end if**
25: **end for**

2.4 Optimisation framework

OptClim is a Python-based control system which allows any computational model to be optimised (or "tuned") by an optimisation algorithm. Neither the model nor the optimisation algorithm need to know much about the other as only the model's parameter values and resulting misfits need to be communicated, which the OptClim system reads, monitors and distributes. This way, the model, which is usually very computationally expensive and complex, can be treated as a "black box".

OptClim was created and written by Professor Simon Tett, University of Edinburgh, and used in several optimisation studies (e.g. Tett et al., 2013, 2017). The latest version of the software is available at `https://github.com/SimonTett/ModelOptimisation`. In 2018 this was modified to be used within this thesis, which is available at `https://doi.org/10.5281/zenodo.5517610`, and contains an extensive user guide.

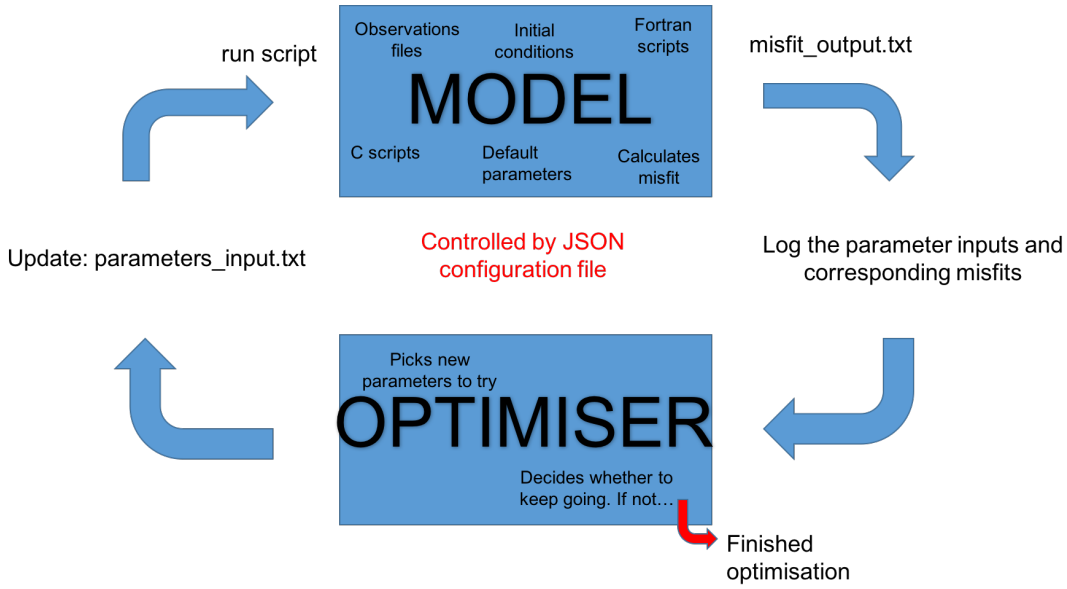

Figure 2.5: Schematic of OptClim's workflow.

2.4.1 Workflow

Essentially, OptClim allows a model to be run multiple times to iteratively improve the model parameters. It does this by using a specified optimisation algorithm to iteratively minimise the model misfit. As shown in the workflow diagram in Figure 2.5, the continuous loop of running a model then an optimisation algorithm is controlled by a JavaScript Object Notation (JSON) file in OptClim. The climate model in this example requires the input file "parameters_input.txt" and is run by the script "runscript". This model then runs, internally using any observational files, initial conditions files or scripts it needs to, and calculates the misfit error ("misfit_output.txt") to be minimised. The optimisation algorithm then reads in the previous history of parameter values and associated misfits, and determines the next parameter combination(s) to try which is likely to minimise the misfit. The model is then run again with this parameter combination(s), and the loop continues until the optimisation algorithm determines that the misfit has been sufficiently minimised.

2.4.2 OptClim Scripts

JSON File The JSON file is written in a way that is easy for machines to parse, as well as for us to read and write. Here the JSON file contains all the information provided by the user regarding their optimisation study, model parameters, location of reference files and scripts, optimiser to use, optimisation settings, etc.

runOptimise.py runOptimise.py is a Python script which begins the workflow loop shown in Figure 2.5. Inputs it requires include an output directory to save each iteration's climate model results to, and the JSON file specifying the OptClim settings. Below is a description of the code.

Code Description: runOptimise.py

INPUT: Output directory ("rootDir"), JSON file location, and if it is to RESTART.

1. Import packages.

```
2. Define functions.
3. Define the function for model creation using ModelSimulation.py and the
   class specific to the climate model.
4. Use StudyConfig.py to load in the config information from the JSON file.
5. Create MODELRUN to store all information on models previously run.
6. if RESTART is True do
7.     Delete all files and folders in the output directory.
8. end if
9. for every subfolder in rootDir do
10.    Read in that model run's information and store in MODELRUN.
11. end do
12. Read algorithm name from config.
13. if algorithm name == "algorithmX" do
14.    Retrieve user's settings specific to this optimisation algorithm.
15.    Run the optimiser.
16.    if optimiser states it's not yet at a solution do
17.       Append MODELRUN['modelsToMake'] with a new model run(s).
18.       Raise "LinAlgError": Go to line 22.
19.    end if
20. end if
21. Print and save solution status.
22. if LinAlgError do
23.    for every model in MODELRUN['modelsToMake'] do
24.       Create the next subfolder in rootDir for this model run.
25.       Submit the model run using Submit.py.
26.    end do
27. end if
28. Store optimisation history and updated configuration file.
29. Write optimisation history to the "final" JSON file.
```

2.4.3 Two job submission workflow options

There are two different versions of the runOptimise.py script, to allow for the two following options displayed in Figure 2.6. For more technical information on how to use OptClim, and how to incorporate new climate models and optimisation algorithms, please see the user guide provided here: `https://doi.org/10.5281/zenodo.5517610`.

Option 1: Individual Sequential Job Submissions If your model runtime is large, or maximum cluster wall clock time is short, then you will need to submit each iteration of the optimisation as an individual job. This is the default option

and you do this using the python script runOptimise.py.

Option 2: One Single Large Job Submission If your model runtime is small, or maximum cluster wall clock time is large, then you can avoid entering the submission queue before each optimisation iteration by submitting the entire optimisation as one large single job, thereby only waiting in the submission queue once. You do this using the python script runOptimise_oneJob.py, and start the optimisation via a bash script loop.

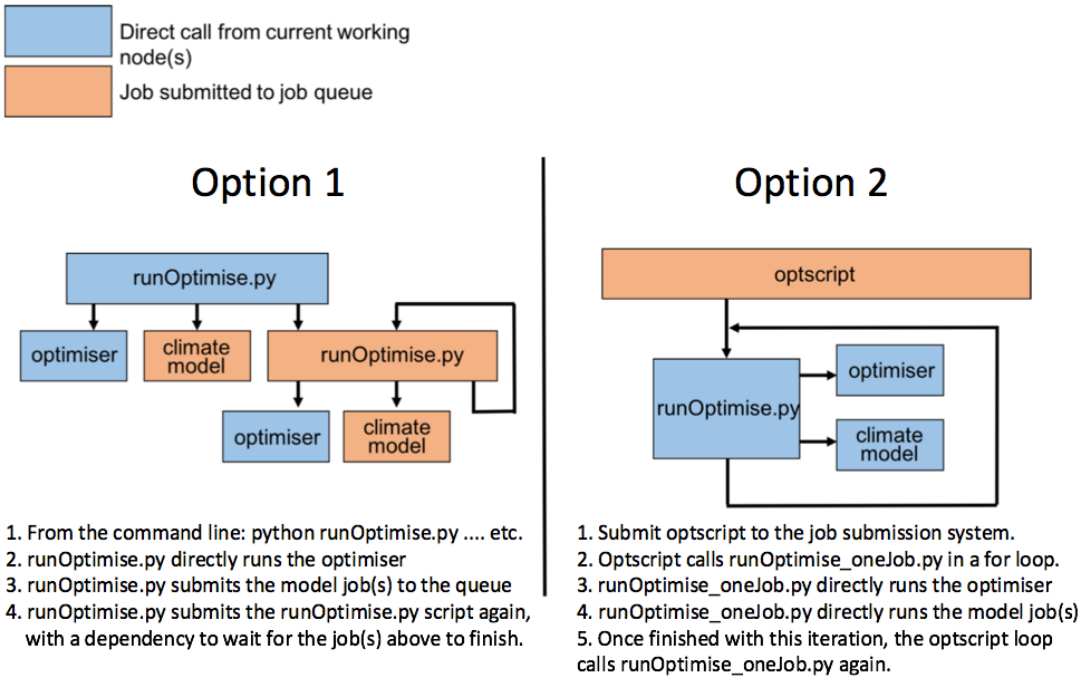

Figure 2.6: Schematic of two job submission workflow options.

2.4.4 Plotting Scripts

There are several plotting MATLAB scripts to visualise the progress of optimisation, described in Table 2.1.

Top Script — Function Called	**Purpose**
plotAll.m	Plot the progress of OptClim in this output directory, if using BOBYQA or DFO-LS.
— checkOptMatches.m	Check that the model run inputs/outputs match the text printed by BOBYQA / DFO-LS.

— plotParamTraject.m	Plot the parameter values throughout the optimisation by BOBYQA / DFO-LS.
— plotTracerMisfit.m	Plot the total and individual misfit values throughout the optimisation by BOBYQA / DFO-LS.
— plotParamPairMisfits.m	In one figure plot all parameter and total misfit values throughout the optimisation by BOBYQA / DFO-LS for each parameter pair.
— plotParamPairProgress.m	In one figure per iteration, plot the parameter and total misfit values during the optimisation by BOBYQA / DFO-LS for each parameter pair.
plotAllCmaes.m	Plot the progress of OptClim in this output directory, if using CMA-ES.
— plotParamTrajectCmaes.m	Plot the parameter values throughout the optimisation by CMA-ES.
— plotTracerMisfitCmaes.m	Plot the total and individual misfit values throughout the optimisation by CMA-ES.
— plotParamPairMisfitsCmaes.m	In one figure plot all parameter and total misfit values throughout the optimisation by CMA-ES for each parameter pair.
plotAllCompare.m	Compare the progress of OptClim by different optimisers in these output directories.
— plotCompare.m	Create 2 plots. (1) Plot on top of each other the total misfit values throughout the optimisation by different optimisers. (2) Plot on top of each other the parameter values throughout the optimisation by different optimisers.
— plotCompareMisfits.m	Compare the individual misfit values throughout the optimisation by different optimisers.
— plotParamRelations.m	Create one figure per parameter pair comparing the parameter tuning of one parameter pair by 2 different optimisers. This helps spot any non-independent parameters which may influence the tuning process.

Table 2.1: OptClimSO Plotting Scripts.

The OptClim package described here was used in the following chapters to control the optimisation of global ocean biogeochemical models.

3

A derivative-free optimisation method for global ocean biogeochemical models

The performance of global ocean biogeochemical models, and the Earth System Models in which they are embedded, can be improved by systematic calibration of the parameter values against observations. However, such tuning is seldom undertaken as these models are computationally very expensive. Here we investigate the performance of DFO-LS, a local, derivative-free optimisation algorithm which has been designed for computationally expensive models with irregular model-data misfit landscapes typical of biogeochemical models. We use DFO-LS to calibrate six parameters of a relatively complex global ocean biogeochemical model (MOPS) against synthetic dissolved oxygen, inorganic phosphate and inorganic nitrate "observations" from a reference run of the same model with a known parameter configuration. The performance of DFO-LS is compared with that of CMA-ES, another derivative-free algorithm that was applied in a previous study to the same model in one of the first successful attempts at calibrating a global model of this complexity. We find that DFO-LS successfully recovers 5 of the 6 parameters in approximately 40 evaluations of the misfit function (each one requiring a 3000 year run of MOPS to equilibrium), while CMA-ES needs over 1200 evaluations. Moreover, DFO-LS reached a "baseline" misfit, defined by observational noise, in just 11–14 evaluations, whereas CMA-ES required approximately 340 evaluations. We also find that the performance of DFO-LS is not significantly affected by observational sparsity, however fewer parameters were successfully optimised in the presence of observational uncertainty. The results presented here suggest that DFO-LS is sufficiently inexpensive and robust to apply to the calibration of complex, global ocean biogeochemical models. The work done in this chapter has been submitted to Geoscientific Model Development for review (Oliver et al., 2021). The structure of this chapter is as follows: first the methodology in Section 3.1, followed by the results in Section 3.2, discussion in Section 3.3, conclusions in Section 3.4, and finally the code availability in Section 3.5. Some separate investigative work is also provided in Appendix A.

3.1 Methodology

3.1.1 Ocean biogeochemical model

The chosen model to optimise is the Model of Oceanic Pelagic Stoichiometry (MOPS-2.0), described in Section 2.1.1. MOPS is a global ocean biogeochemical model, which simulates the cycling of 9 biogeochemical tracers, namely dissolved inorganic and organic phosphate, dissolved inorganic nitrate, dissolved oxygen, phytoplankton, zooplankton, detritus, dissolved inorganic carbon and alkalinity (Kriest and Oschlies, 2013, 2015). MOPS is coupled to the Transport Matrix Method (TMM; Khatiwala et al., 2005; Khatiwala, 2007, 2018), described in Section 2.2.1, which is an efficient numerical method for "offline" simulation of biogeochemical tracers. In this chapter we use monthly mean transport matrices and other physical forcing fields (including temperature, salinity, sea ice and winds) derived from a relatively coarse resolution ($2.8^\circ \times 2.8^\circ \times 15$ levels) configuration of MITgcm (Marshall et al., 1997), as described in Section 2.2.2.

3.1.2 Biogeochemical model parameters

The behaviour of MOPS is controlled by approximately 25 parameters, of which we have chosen 6 to consider for calibration, based on the previous optimisation study by Kriest et al. (2017). The detailed definitions and possible ranges of these parameters are described in that paper, and their descriptions are listed here in Table 3.1. Briefly, 4 of these parameters are mainly restricted to the epipelagic and mesopelagic zones of the ocean, as they involve phytoplankton and zooplankton. I_{C} and K_{PHY} are the phytoplankton half-saturation constants for light absorption and phosphate uptake, respectively. μ_{ZOO} is the zooplankton maximum grazing rate and k_{ZOO} the zooplankton quadratic mortality rate. The remaining two parameters influence the remineralisation and sinking of particulate organic matter (POM). $R_{\mathrm{O_2:P}}$ is the ratio of oxygen consumption to phosphate release during remineralisation when oxygen is available, and b^* is the exponent of the "Martin curve", a power law function that describes the attenuation of POM flux with depth (Martin et al., 1987).

$R_{O_2:P}$	Ratio of oxygen consumption to phosphate release
I_C	Phytoplankton half-saturation constants for light
K_{PHY}	Phytoplankton half-saturation constants for phosphate
μ_{ZOO}	Zooplankton maximum grazing rate
k_{ZOO}	Zooplankton quadratic mortality rate
b^*	Exponent of the Martin curve

Table 3.1: Selected MOPS biogeochemical parameters to optimise.

3.1.3 Biogeochemical observations

The "observations" used for calibration in this chapter are purely synthetic, created by a reference simulation of MOPS, hereafter referred to as MOPS-ref. MOPS-ref was run with the following parameter values: $R_{O_2:P}$ = 170 mmol O_2 : mmol P, I_C = 24 W m^{-2}, K_{PHY} = 0.03125 mmol P m^{-3}, μ_{ZOO} = 2 d^{-1}, k_{ZOO} = 3.2 (mmol P m^{-3})$^{-1}$ d^{-1}, and b^* = 0.858 (see Table 3.4). Using twin experiments, whereby the optimiser calibrates the model to synthetic observations, the success of the optimiser to recover these known parameter values can be determined.

3.1.4 Optimisation Algorithms

Using the optimisation framework described in Section 2.4 we iteratively evaluate the misfit between the model and observations, then use an optimisation algorithm to vary the model parameter inputs with the aim of finding a lower misfit. Here, every evaluation of the misfit requires running the biogeochemical model to equilibrium (3000 years), then calculating the misfit between the model outputs and real (or synthetic) observations of dissolved oxygen, phosphate and nitrate. The function of the misfit between MOPS and observations has previously been shown to contain multiple minima upon which an optimiser can converge (Kriest et al., 2017), therefore we need to confirm our chosen optimisation method is able to navigate multiple minima to locate the global one. To determine if an optimiser can find the global minimum within the misfit function landscape "twin" experiments are used, whereby the misfit is calculated between the model outputs and synthetic observations. The synthetic observations are created by the model with a known parameter

configuration, therefore the global minimum (and optimal parameter values) are known. We use twin experiments to compare the performance of the following two different optimisation algorithms.

3.1.4.1 CMA-ES

We first apply the Covariance Matrix Adaptation Evolution Strategy (CMA-ES; Hansen, 2016) optimisation algorithm, which is described in Section 2.3.1. CMA-ES is a widely used stochastic evolutionary algorithm, for use on a "black box" misfit function. CMA-ES carries out a global search of the parameter space, therefore it seeks to find the minimum over the parameter space. In order to achieve this it requires many function evaluations, and therefore can be quite computationally expensive in practice. The CMA-ES code used in this chapter, which is based on the $(\mu/\mu_{\mathrm{W}},\lambda)$-CMA-ES algorithm of Hansen (2016), was sourced from Kriest et al. (2017). As in the previous Kriest et al. study, we use a population size of 10, i.e., in each iteration of CMA-ES, the misfit function is evaluated 10 times.

3.1.4.2 DFO-LS

We then apply the derivative-free optimisation using least squares (DFO-LS) optimisation algorithm, which is described in Section 2.3.2. DFO-LS is an iterative algorithm for minimising a function $f(\mathbf{x})$ (Cartis et al., 2019), where $\mathbf{x} \in \mathbb{R}^n$ is the vector of parameters, each of which is constrained within specified bounds. DFO-LS can take into account the individual terms in the misfit function and mathematically incorporate their individual function structures into the problem to minimise.
DFO-LS must be given a starting location within the parameter space from which to initialise. In the initial iteration of DFO-LS, the misfit function is evaluated at the starting location and at n additional locations nearby, for a total of $n + 1$ function evaluations. In subsequent iterations only one function evaluation is needed.
Unlike CMA-ES, DFO-LS is more of a "local" optimisation method. However, there is strong numerical evidence from the derivative-free optimiser Py-BOBYQA, upon which DFO-LS is based (Cartis et al., 2018), that it is able to find global minima. This is not guaranteed in all cases, therefore to increase the likelihood of finding

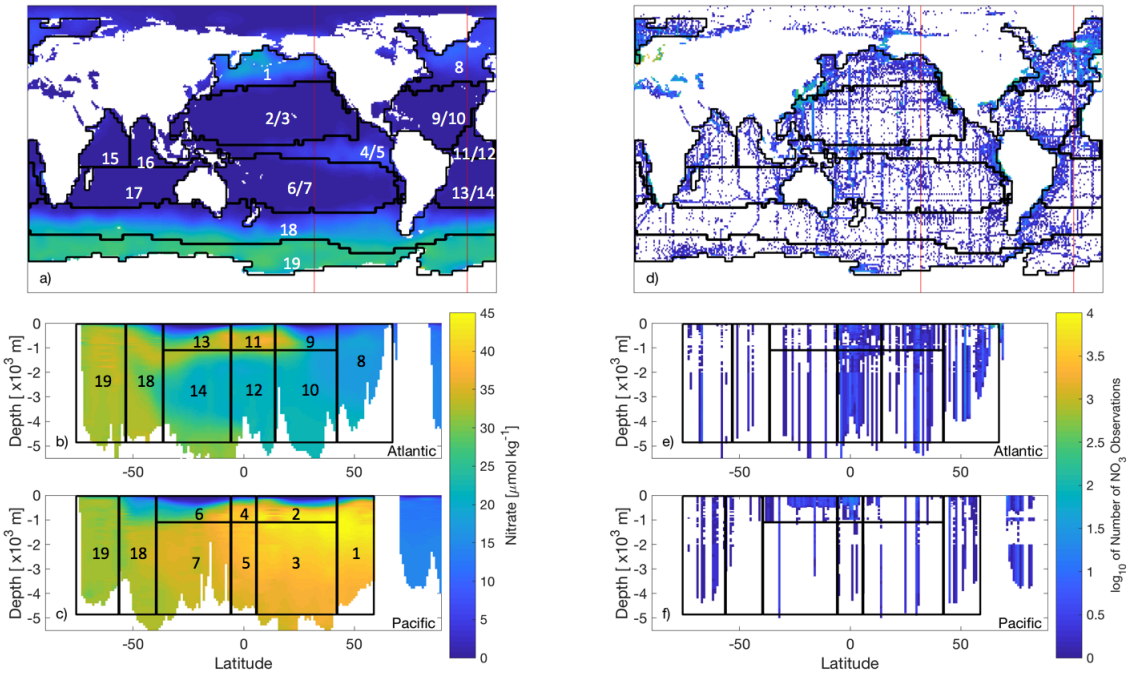

Figure 3.1: World Ocean Atlas nitrate data (Garcia et al., 2018a,b) of (left) interpolated objectively analysed mean concentration of nitrate in sea water [μmol kg^{-1}], and (right) number of true observations plotted on a $\log_{10}$ colour scale (white oceanic areas show areas of no nitrate observations). These have been plotted for the (a,d) global surface 0 m, which also show the locations of the longitudinal transects (red lines) for the nitrate data plotted at (b,e) 23°W and (c,f) 140°W. Overlain are the boundaries of 19 biomes (Henson et al., 2010; Weber et al., 2016) of similar biogeochemistry. Regions with 2 numbers have been further split by depth.

the global minimum, we initiate from multiple starting points and allow DFO-LS to apply soft restarts when there is slow progress. For more information see Section 2.3.2

3.1.5 Misfit functions

Every evaluation of our misfit requires running the biogeochemical model for 3000 years before calculating the misfit between the model outputs and synthetic observations. While both CMA-ES and DFO-LS minimise a single misfit function, DFO-LS can exploit the structure of the misfit function. Thus, if the misfit is defined as per Equation 2.6, we only provide CMA-ES with $f(\mathbf{x})$ whereas the individual $r_i(\mathbf{x})$ are supplied to DFO-LS. Here we define the r_i to take into account the spatial structure of the misfit by partitioning the ocean into previously established biome

regions of similar ocean biogeochemical properties (Henson et al., 2010; Weber et al., 2016), provided by Raffaele Bernardello (Barcelona Supercomputing Centre) via personal communication, several of which were further split by depth at 1000 m (see Fig. 3.1) for a total of 19 regions. For every region j, we further calculate a misfit for each of the 3 tracers q (phosphate, nitrate, oxygen) used in the optimisation. The objective $f(\mathbf{x})$ is thus composed of $19 \times 3 = 57$ terms of the form:

$$r^{\epsilon}_{qj}(\mathbf{x}) = \sqrt{\frac{V_j}{V_T}} \frac{\sqrt{\sum_{i=1}^{N_j} (m_{qi}(\mathbf{x}) - (o_{qi} + \epsilon_{qi}))^2 \frac{V_i}{V_j}}}{\sum_{i=1}^{N_j} o_{qi} \frac{V_i}{V_j}}, \tag{3.1}$$

where $m_{qi}(\mathbf{x})$ is the model solution with parameters $\mathbf{x}$ at grid point i for tracer q and o_{qi} the corresponding observation (the synthetic observations provided by a reference run of MOPS). N_j is the total number of model grid boxes in biome region j. The misfit is normalised by the volume-weighted mean tracer concentration for that region and weighted, first, by individual grid point volumes V_i relative to the volume V_j of region j and, second, by the region's total volume relative to the global ocean volume V_T. Real oceanic observations have a degree of uncertainty associated with them due to spatio-temporal oceanic processes, e.g., from small scale processes such as eddies. To account for this we add a noise term ϵ_{iq}, which is the added noise due to uncertainty associated with tracer q for every grid box i in the model. The global misfit $f^{\epsilon}_{global}(\mathbf{x})$ is then defined as

$$f^{\epsilon}_{global}(\mathbf{x}) = \sum_{q=1}^{3} \sum_{j=1}^{19} r^{\epsilon}_{qj}(\mathbf{x})^2. \tag{3.2}$$

The total misfit function f is broadly similar to Kriest et al. (2017), with the main difference being the incorporation of the 19 biome regions.

The non-noisy equivalent of these misfit terms are r_{qj} and f_{global} as in Equations 3.1 and 3.2 with $\epsilon = 0$. We also define "baseline" misfits r^{Base}_{qj} and f^{Base}_{global}, which are the misfits due to the noise alone in the special case where the model outputs equal the observations. f^{Base}_{global} equates to 7.04×10^{-4}, and gives an indication of the termination criteria when optimising the model to real noisy oceanic observations, as optimising below this threshold would serve no useful purpose.

To specify a realistic noise field we take the standard deviation variable provided in the World Ocean Atlas database (WOA18 Garcia et al., 2018a,b). Since our misfit is defined with respect to annual mean data we require an annual mean standard deviation without the variability of the seasonal cycle. To do so we take the numerical mean (weighted by number of observations) of the monthly standard deviations reported in the WOA18 dataset for the upper 800 m (phosphate and nitrate) or 1500 m (oxygen), and the annual standard deviation below those depths. These standard deviations fields were interpolated onto the model grid and then multiplied by three different Gaussian noise fields to create three separate noise (ϵ) realisations. The baseline misfit terms r_{qj}^{Base} and f_{global}^{Base} were calculated as an average over these realisations.

3.1.6 Optimisation experimental design and solver settings

In this chapter we seek to: 1) compare the performance of CMA-ES and DFO-LS on noise-free observations; 2) investigate DFO-LS's performance on noisy observations; and 3) investigate the impact of sparse observations on the ability of DFO-LS to recover the true parameters. In order to do so we carried out the following series of experiments (see Table 3.2 for the corresponding experiment labels):

Noise-free experiments In the noise-free experiments we attempted to recover all 6 parameters. For this a single run of CMA-ES was performed (labelled exp_c). For DFO-LS two experiments were carried out (exp_d1 and exp_d2), starting from two different locations in parameter space that were chosen to be relatively far from the target parameters. The parameter values for these and all other experiments are listed in Table 3.4.

Both CMA-ES and DFO-LS are controlled by various solver settings. For CMA-ES the main ones are the number of sequential generations and the population size. As per Kriest et al. (2017) we set these to 200 and 10, respectively. The solver settings used by DFO-LS are summarised in Table 3.3. Of the noise-free experiments exp_d1 and exp_d2, the former had DFO-LS settings regarding trust region management

twin experiments tuning to:		
noise-free observations	**noisy observations**	**sparse observations**
exp_c	expd_rng1	exp_d1_sparse
exp_d1	expd_rng2	exp_d2_sparse
exp_d2	expd_rng3	

Table 3.2: Names of each experiment tuning to noise-free, noisy and sparse twin observations. The characters within each experiment name, after "exp" follows the letter "c" or "d" depicting CMA-ES or DFO-LS. After this there is optionally a number to differentiate between different DFO-LS starting locations in parameter space, or text to identify the added noise scenario or whether the observations are sparse.

(**tr_radius**) which are more suitable for a noisy misfit function, while the latter for a smooth misfit function. As exp_d1 was slightly more successful, the trust region management settings were set to be more suitable for a noisy misfit function in all subsequent experiments.

DFO-LS experiments with observational uncertainty To understand optimisation performance in the presence of observational uncertainty, noise was added to the reference observations (see Section 3.1.5). Three such optimisation runs, each with a different noise realisation, were carried out with DFO-LS (expd_rng1, expd_rng2, expd_rng3) starting from the same location in parameter space, to minimise the noisy misfit function f^{ϵ}_{global}. The goal was to see if DFO-LS could recover all 6 of the MOPS-ref target parameter values.

DFO-LS experiments with sparse observations There are large areas of the ocean which have not been sampled adequately or at all (e.g., Fig. 3.1). While it is possible to fill in the gaps in the data using objective interpolation methods, this might not always work well in the presence of large gradients. In a last set of experiments we therefore compared how DFO-LS performs in the presence of data sparsity (exp_d1_sparse and exp_d2_sparse), by only using observations at model grid points for which the corresponding locations in WOA18 contain data, with its corresponding performance in the absence of data sparsity (exp_d1 and exp_d2).

DFO-LS Setting Name	Description	Group A*	exp_d2
maxfun	Maximum number of true misfit function evaluations	70	70
obj_fun_has_noise	Does the misfit function have stochastic noise?	False	False
rhobeg	Normalised radius of parameter trust region at start	0.1	0.1
rhoend	Normalised radius of parameter trust region for termination or restart	0.001	0.001
tr_radius.gamma_dec	Ratio to decrease trust region radius (Δ_k) in an unsuccessful iteration	0.98	0.5
tr_radius.alpha1	Ratio to decrease the lowest bound (ρ_k) for the trust region radius	0.9	0.1
tr_radius.alpha2	Ratio of ρ_k to decrease Δ_k by when ρ_k is reduced	0.95	0.5

Table 3.3: DFO-LS parameter settings for each optimisation experiment. * Group A = exp_d1, t6d_fullgrid, t6d_sparse, t6d_rng1, t6d_rng2, t6d_rng3.

3.2 Results

Tables 3.4 and 3.5 show the starting and optimised parameter values, and parameter recovery information for all twin optimisation experiments. In the following sections we plot both the global misfit and parameter values for every function evaluation throughout each optimisation experiment. During one CMA-ES iteration we evaluate the misfit function 10 times (the population size). Therefore for CMA-ES we plot the minimum (best) and maximum misfits, and we plot the parameter values corresponding to the best misfit, and the minimum and maximum parameter values of each population.

To both reiterate how DFO-LS works and fully explain the DFO-LS figures, we briefly describe the optimisation process in terms of expected misfit reduction and parameter trajectories. First, DFO-LS evaluates the misfit function $n+1$ times near to the chosen starting point, therefore in the first 7 evaluations we do not expect a significant misfit reduction, other than what may occur by chance. After these initial evaluations, DFO-LS attempts to minimise the misfit function and there will be both successful evaluations (the resulting misfit is lower than previously found in the optimisation), and unsuccessful ones (the misfit is not lower). There may also be restarts, directly after which unsuccessful evaluations are common as DFO-LS perturbs the parameter values to get out of a possible local minimum. Therefore on every DFO-LS figure we have plotted the misfit or parameter values for every evaluation (both successful and unsuccessful) as scattered points, and successful ones in a solid line.

Parameters	RO2:P	IC	KPHY	muZOO	kZOO	b*	Misfit
Upper Bound	200	48	0.5	4	10	1.8	
Lower Bound	150	4	0.0001	0.1	0	0.4	
Target	170	24	0.03125	2	3.2	0.858	0
Experiments							
exp_c Start	NA	NA	NA	NA	NA	NA	$4.231\text{x}10^{-2}$
Optimised	**170.003**	**24.001**	**0.031**	**2.000**	**3.200**	**0.858**	$\mathbf{2.909\text{x}10^{-10}}$
exp_d1	180.000	40.000	0.100	2.500	5.000	0.540	$5.248\text{x}10^{-2}$
	170.401	**24.026**	**0.051**	**2.062**	**3.448**	**0.860**	$\mathbf{4.143\text{x}10^{-6}}$
exp_d2	190.000	30.000	0.200	3.500	1.000	1.100	$2.715\text{x}10^{-1}$
	169.875	**23.663**	**0.153**	**2.013**	**3.211**	**0.859**	$\mathbf{6.747\text{x}10^{-6}}$
expd_rng1	180.000	40.000	0.100	2.500	5.000	0.540	$5.316\text{x}10^{-2}$
	170.812	**23.856**	**0.024**	**2.531**	**5.629**	**0.870**	$\mathbf{8.050\text{x}10^{-4}}$
expd_rng2	180.000	40.000	0.100	2.500	5.000	0.540	$5.316\text{x}10^{-2}$
	168.116	**25.321**	**0.007**	**2.086**	**4.634**	**0.852**	$\mathbf{8.717\text{x}10^{-4}}$
expd_rng3	180.000	40.000	0.100	2.500	5.000	0.540	$5.316\text{x}10^{-2}$
	169.234	**23.011**	**0.053**	**2.714**	**6.352**	**0.878**	$\mathbf{8.215\text{x}10^{-4}}$
exp_d1_sparse	180.000	40.000	0.100	2.500	5.000	0.540	$5.427\text{x}10^{-2}$
	169.816	**24.002**	**0.204**	**1.689**	**2.232**	**0.854**	$\mathbf{6.475\text{x}10^{-5}}$
exp_d2_sparse	190.000	30.000	0.200	3.500	1.000	1.100	$2.843\text{x}10^{-1}$
	170.022	**23.610**	**0.150**	**2.077**	**3.415**	**0.861**	$\mathbf{9.691\text{x}10^{-6}}$

Table 3.4: Optimised parameters for all twin experiments. Upper section shows parameter bounds and MOPS-ref target parameters to be recovered. Lower section shows each experiment's results. Columns 2-7: (1st row) starting parameter values and (2nd row) optimised parameters for $R_{O_2:P}$ [mmol O_2 : mmol P], I_C [W m^{-2}], K_{PHY} [mmol P m^{-3}], μ_{ZOO} [d^{-1}], k_{ZOO} [(mmol P $m^{-3})^{-1}$ d^{-1}], and b^*. Column 8: (1st row) the starting global misfit and (2nd row) the lowest global misfit. NA = not applicable for CMA-ES.

On every figure of global misfits the expected baseline misfit (f_{global}^{Base}, see Section 3.1.5) is also plotted, below which any misfit reduction is within observational noise levels. On every parameter trajectory plot the "recovery zone" is also shown. This indicates the range of parameter values within ±5% of the target value, normalised by the total range (upper bound minus lower bound) for that parameter. We consider a parameter to have been "recovered" by a certain number of evaluations, when all subsequent parameter values corresponding to successful evaluations remain within this recovery zone.

3.2.1 Noise-free experiments

CMA-ES experiment Figure 3.2 shows that during the optimisation exp_c by CMA-ES the global misfit decreases significantly from 10^{-1} to $\approx 10^{-4}$ within

Experiment	Number of evaluations required to recover parameter:						Maximum	Evaluation of	Evaluations to baseline
	RO2:P	IC	KPHY	muZOO	kZOO	b*	Evaluations	lowest misfit	misfit (7.0405e-04)
exp_c	420	370	1710	800	850	340	2000	1983	309
exp_d1	**16**	**25**	**12**	41	42	**13**	70	45	20
exp_d2	43	**9**	-	44	**29**	**26**	70	49	35
expd_rng1	**13**	**24**	**24**	-	-	**12**	70	29	-
expd_rng2	**38**	**39**	43	**11**	-	**12**	70	43	-
expd_rng3	**25**	**24**	57	-	-	**12**	70	57	-
exp_d1_subsel	**18**	46	-	-	-	**20**	70	62	21
exp_d2_subsel	**25**	**9**	-	**26**	**26**	**25**	70	64	29

Table 3.5: Number of evaluations required to recover each parameter for all twin experiments. Columns 2-7: number of misfit function evaluations required to successfully recover that parameter (- = never recovered). All evaluations required to recover a parameter which were fewer than 40 are typed in bold font. Column 8: the maximum number of evaluations completed. Column 9: the evaluation which provided the lowest or "best" global misfit. Column 10: the number of evaluations needed for the global misfit to be reduced below noise levels (- = never reached the baseline).

the first 500 function evaluations. Subsequently progress slows down as the misfit is reduced by only one more order of magnitude over the next 1000 evaluations. Progress then significantly improves, with the misfit decreasing from $\approx 10^{-5}$ to 10^{-9} within the final 500 evaluations. Note that the spikes in the maximum global misfit near the 1650^{th} evaluation was due to the added penalty factor when one of the parameter values in this population had a value just outside of its allowed range. While this experiment did not include noise, we note that CMA-ES required 309 evaluations to reach the baseline misfit, beyond which any misfit reduction would have been within observational noise levels.

Figure 3.3 shows how the 6 parameters were optimised towards the MOPS-ref target parameter values by CMA-ES. The targets were found relatively quickly within the initial 500 evaluations for the parameters $R_{\text{O}_2:\text{P}}$, I_{C} and b^*, corresponding to the initial fast misfit reduction previously shown in Fig. 3.2. The μ_{ZOO} and k_{ZOO} targets were found next after approximately 1000 evaluations, after which the optimiser began tuning K_{PHY} towards its target until it located after 1700 evaluations. If exp_c had been terminated once the observational noise level was reached after 309 evaluations, $R_{\text{O}_2:\text{P}}$, I_{C} and b^* would have been optimised to their MOPS-ref values relatively well, while K_{PHY}, μ_{ZOO} and k_{ZOO} would still be far from their target values.

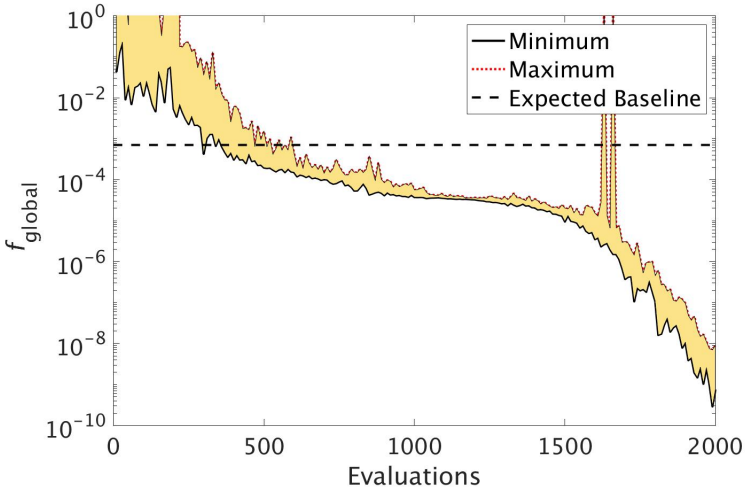

Figure 3.2: The reduction in global misfit of MOPS to the twin MOPS-ref observations by CMA-ES for the experiment exp_c. There were 10 MOPS evaluations within each CMA-ES iteration (population $\lambda = 10$), ran in parallel. Plotted is the baseline misfit (horizontal black dashed line), the minimum (black solid line) and maximum (red dotted line) misfit of each population, with the area between shaded yellow.

DFO-LS experiments To compare the performance of DFO-LS with CMA-ES we carried out two optimisation experiments with DFO-LS (exp_d1 and exp_d2) starting from two different locations in parameter space, with differing parameters controlling the DFO-LS trust-region shrinking speed. Optimisation exp_d1 had slower trust region shrinking parameters to allow it to better handle an irregular misfit function. Figure 3.4 shows the comparison between both experiments' reduction of the global misfit. In both cases there was rapid initial misfit decrease from near 10^{-1} to 10^{-3} within 30 model evaluations. Optimisation exp_d2 showed slightly slower misfit reduction, needing 35 evaluations to reach the baseline misfit, while exp_d1 only required 20 to reach the baseline, beyond which any misfit reduction is within observational noise levels. In both exp_d1 and exp_d2 DFO-LS managed to reduce the misfit to below 10^{-5} within 45-49 evaluations, then initiated restarts to reduce it further. As exp_d1 performed slightly better than exp_d2, especially between evaluations 15-45, exp_d1 trust region shrinking settings were used as defaults for all subsequent experiments.

Figure 3.5 shows how the 6 parameters were optimised towards the MOPS-ref target parameter values by exp_d1 and exp_d2. In both experiments within the first

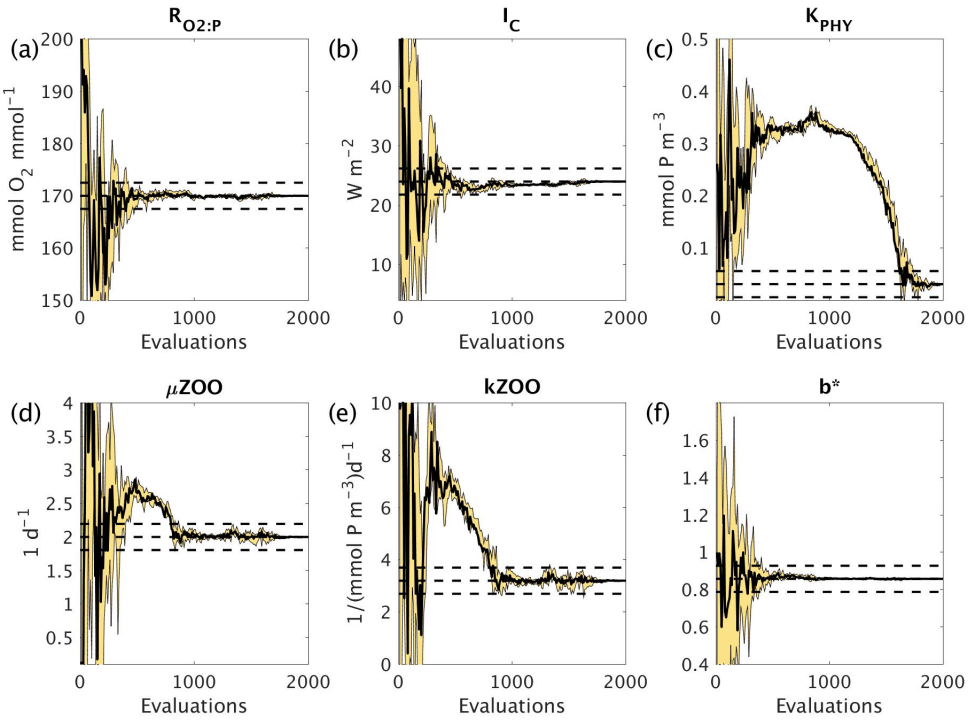

Figure 3.3: Parameter tuning by CMA-ES for the experiments exp_c for the parameters (a) $R_{\mathrm{O_2:P}}$, (b) I_{C}, (c) K_{PHY}, (d) μ_{ZOO}, (e) k_{ZOO} and (f) b^*. There were 10 MOPS evaluations within each CMA-ES iteration (population $\lambda = 10$), ran in parallel. Plotted are the parameter values associated with the minimum misfit of that population (thick black solid line), and the maximum and minimum of all parameter values within that population (thin black solid lines), with the area between shaded yellow. Also plotted are the MOPS-ref target parameter values and ±5% recovery zone (horizontal black dashed lines).

30 evaluations $R_{\mathrm{O_2:P}}$, I_{C}, k_{ZOO} and b^* were optimised to relatively close to their targets, and μ_{ZOO} within the first 45 evaluations. K_{PHY} was successfully optimised by exp_d1, however was not successfully optimised at all by exp_d2. If exp_d1 and exp_d2 had been terminated once reaching the noise baseline, after 20 and 35 evaluations respectively, $R_{\mathrm{O_2:P}}$, I_{C}, b^* and k_{ZOO} would have been optimised to their MOPS-ref values relatively well.

3.2.2 DFO-LS experiments with observational uncertainty

To assess the impact of observational uncertainty we carried out three experiments in which DFO-LS was initialised from the same starting location in parameter space, but with three different realisations of random noise added to the observations (see

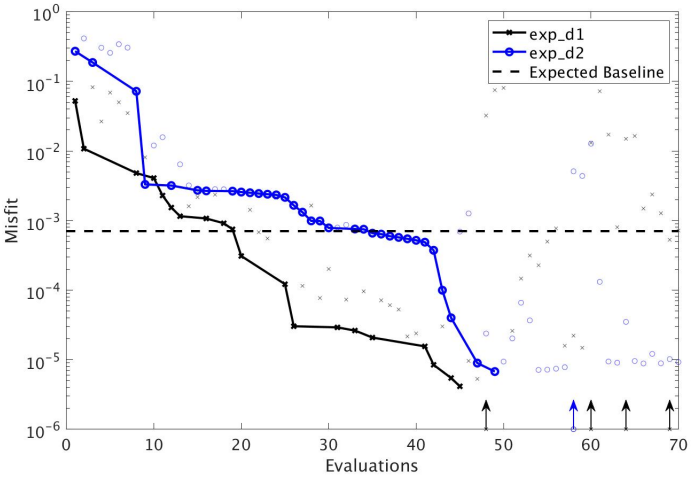

Figure 3.4: The reduction in global misfit of MOPS to the twin MOPS-ref observations by DFO-LS for the experiments exp_d1 (black line with crosses) and exp_d2 (blue line with circles). Also plotted is the baseline misfit (horizontal black dashed line). Vertical arrows indicate a soft restart, coloured and marked according to each experiment. Note that every MOPS evaluation has been plotted with a small marker, however only MOPS evaluations which resulted in a lower misfit than previously seen in each optimisation experiment has been plotted with a solid line.

Section 3.1.6). As seen in Fig. 3.6 DFO-LS managed to reduce the misfit to very close to the average baseline misfit within 30 evaluations with a reduction in misfit from $\sim 10^{-1}$ to $\sim 10^{-3}$. Closer to the end of the optimisation runs, restarts were initiated to encourage further misfit reduction, hence the large variations in misfit. Figure 3.7 shows that the initial misfit reduction corresponds to improved values for the parameters $R_{\mathrm{O_2:P}}$, I_{C}, K_{PHY} and b^*. DFO-LS seems to compensate for the noise by increasing the values for the parameters μ_{ZOO} and k_{ZOO}.

3.2.3 DFO-LS experiments with sparse observations

In a final set of experiments we examine whether DFO-LS is able to successfully optimise MOPS given a sparse set of observations (see Section 3.1.6). The experiments exp_d1_sparse and exp_d2_sparse were initialised from the same location in parameter space as exp_d1 and exp_d2, respectively, but the former were optimised using observations sub-sampled at grid points corresponding to locations in the un-interpolated WOA18 database. In these experiments no noise was added to the observations. Figure 3.8 shows that the two optimisations using full observations

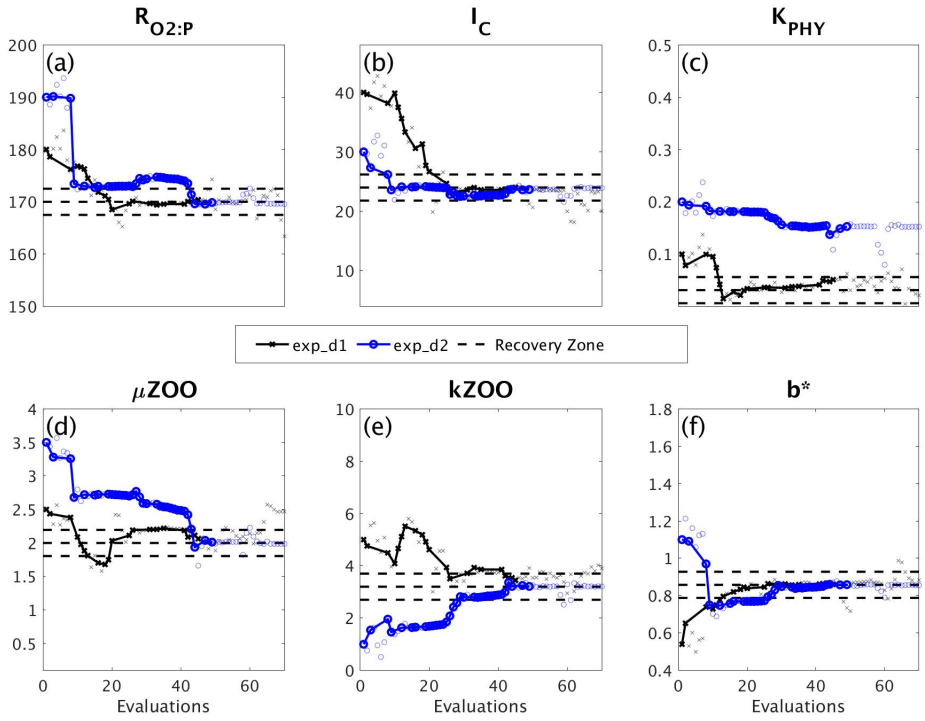

Figure 3.5: Parameter tuning by DFO-LS for the experiments exp_d1 (black line with crosses) and exp_d2 (blue line with circles) for the parameters (a) $R_{O_2:P}$, (b) I_C, (c) K_{PHY}, (d) μ_{ZOO}, (e) k_{ZOO} and (f) b^*. Also plotted are the MOPS-ref target parameter values and $\pm 5\%$ recovery zone (horizontal black dashed lines). Note that every MOPS evaluation has been plotted with a small marker, however only MOPS evaluations which resulted in a lower misfit than previously seen in each optimisation experiment has been plotted with a solid line.

(exp_d1 and exp_d2) converged to slightly lower misfits than when using sparse observations. Figure 3.9 shows that exp_d1 successfully recovered all 6 parameters within 42 evaluations, while exp_d1_sparse only successfully recovered $R_{O_2:P}$, I_C and b^* throughout the optimisation. The above results would indicate a poorer optimisation when using sparse observations, however, from the other starting point exp_d2 successfully recovered 5 parameters within 44 evaluations, while exp_d2_sparse recovered the same 5 after only 26 evaluations. This suggests that even with sparse observations it is possible to successfully optimise a global ocean biogeochemical model such as MOPS.

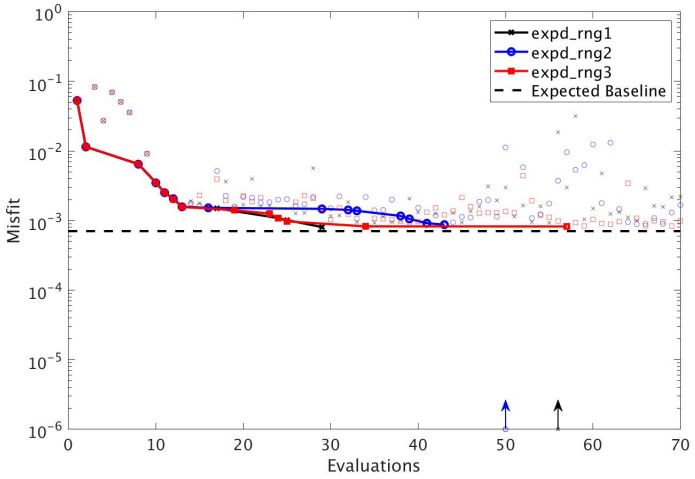

Figure 3.6: As in Fig. 3.4, but for experiments expd_rng1 (black line with crosses), expd_rng2 (blue line with circles), and expd_rng3 (red line with squares).

3.3 Discussion

3.3.1 CMA-ES vs DFO-LS optimisation performance

Our comparison of the two optimisation algorithms shows that DFO-LS could recover all 6 target parameter values within ∼40 evaluations of MOPS, while CMA-ES achieved the same goal within ∼1700 evaluations. By "recover" we mean optimised to within ±5% (normalised by the parameter range) of the target value. DFO-LS reduced the misfit to below the observational uncertainty threshold within 20-35 evaluations, while CMA-ES required 309 evaluations. DFO-LS is thus significantly more efficient for this particular problem and may, in general, be more practical for optimising more than a small handful of parameters. However we note that the multiple evaluations CMA-ES requires can be run in parallel. In contrast, DFO-LS, except for the initial $n + 1$ evaluations, runs sequentially.

CMA-ES is a single-objective optimiser, while DFO-LS can use information from multiple misfit values instead of just one. Therefore it can exploit more information to allow for a faster reduction in the misfit. Neither algorithm can completely guarantee a global optimum solution, although CMA-ES carries out a more global search than DFO-LS. There is significant evidence DFO-LS can find the global optimum (Cartis et al., 2018), but to increase confidence in the final solution it can

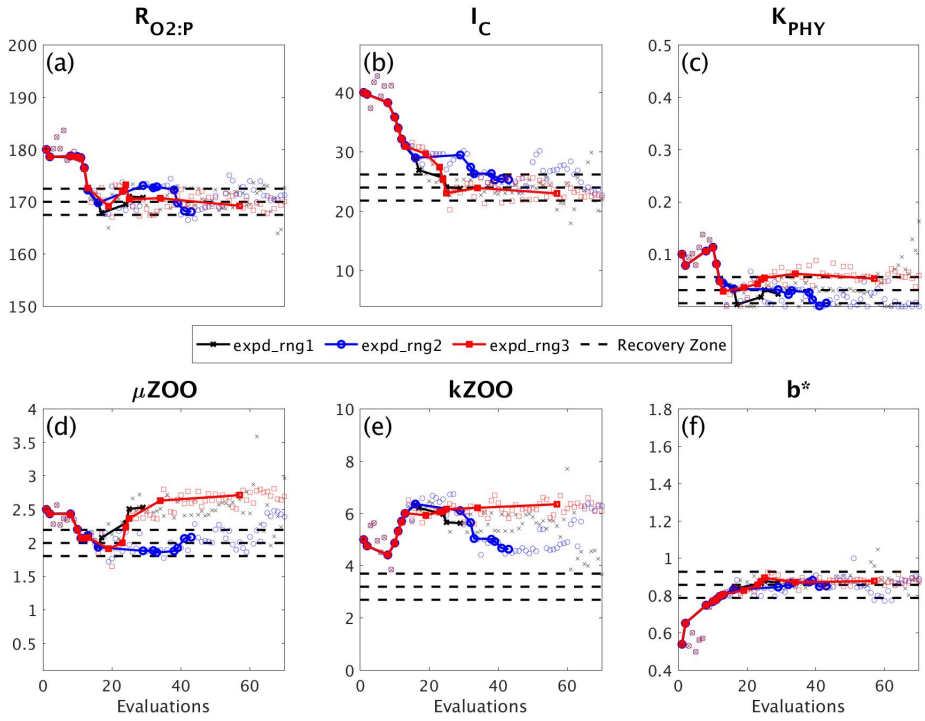

Figure 3.7: As in Fig. 3.5, but for the experiments expd_rng1 (black line with crosses), expd_rng2 (blue line with circles), and expd_rng3 (red line with squares).

be combined with a globalising method such as starting from different points in parameter space or using the DFO-LS restart functionality.

Both methods struggled with one of the parameters K_{PHY}, due to the misfit function's low sensitivity to this parameter (as found by perturbing the parameter values in each direction and computing the gradient). DFO-LS had not begun to tune this parameter for one of the experiments (exp_d2) before we terminated it at a maximum of 70 evaluations, although it did find K_{PHY} when initiated from a different starting point (exp_d1). CMA-ES also had difficulty in tuning K_{PHY} and only started optimising this parameter after all the other parameters were recovered at $\sim$1200 evaluations. The maximum number of DFO-LS evaluations was set to 70 as it is a sequential algorithm, therefore it was impractical to allow too many more evaluations. Had it been allowed to run longer the expectation is it would begin tuning K_{PHY} once the other 5 were sufficiently tuned, as was the case with CMA-ES.

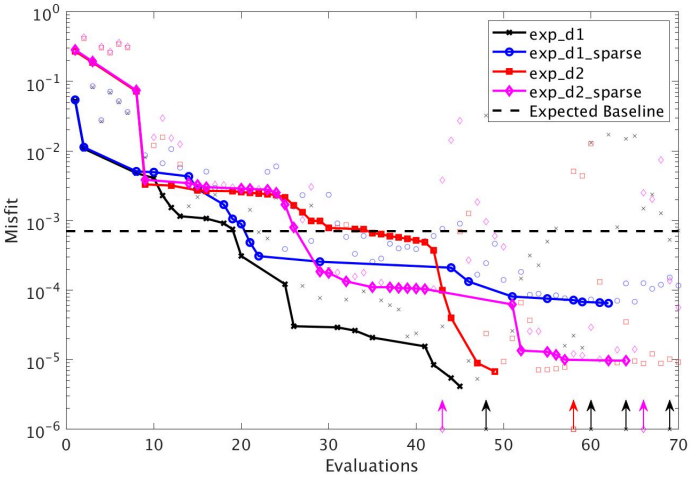

Figure 3.8: As in Fig. 3.4, but for experiments exp_d1 (black line with crosses), exp_d1_sparse (blue line with circles), exp_d2 (red line with squares) and exp_d2_sparse (magenta line with diamonds). Note that the baseline misfit (horizontal black dashed line) was calculated using the full grid noisy observations.

3.3.2 Calibrating to uncertain observations

Real oceanic observations come with associated uncertainty due to measurement error, temporal variations such as seasonal and diurnal cycles, and meso-scale variability due to factors such as eddies and the movement of fronts. Here we have studied how this uncertainty introduces a baseline to the misfit function, below which any optimisation of the biogeochemical model would be within the uncertainty level. We determined this baseline for the misfit function (or termination threshold) using the standard deviations of the observational data, however others have defined it as the global optimum of a surrogate formulation of the biogeochemical model (Sauerland et al., 2017).

In the present case and the chosen set of oceanic observations, the model was significantly optimised before reaching levels of observational uncertainty, particularly due to optimisation of the parameters which the model is most sensitive to, as was determined by perturbing each parameter while holding the others fixed and calculating the misfit. In this case the misfit function was most sensitive to b^* (the increase in particle sinking speed with depth) and I_C (the phytoplankton half saturation for light). The parameter b^* has a particularly strong non-local effect, as

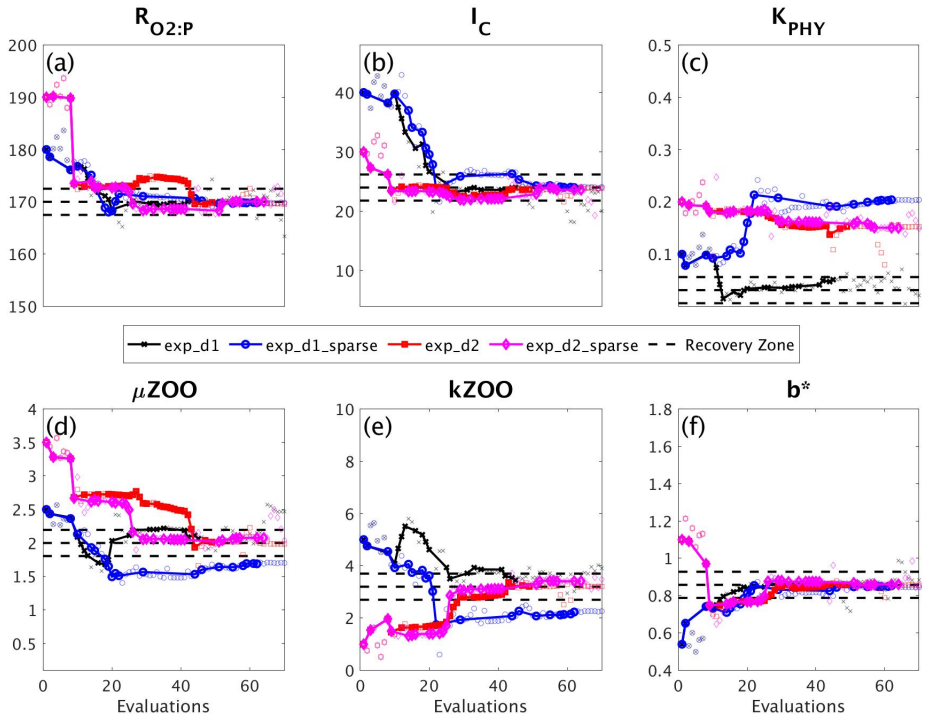

Figure 3.9: As in Fig. 3.5, but for the experiments exp_d1 (black line with crosses), exp_d1_sparse (blue line with circles), exp_d2 (red line with squares) and exp_d2_sparse (magenta line with diamonds).

it influences the flux of oxygen and nutrients to the deeper ocean, hence to ocean basins further along the "conveyor belt" (Kwon and Primeau, 2006). Somewhat surprisingly, a parameter the model is less sensitive to, $R_{O_2:P}$ (the ratio of oxygen consumption to phosphate release during remineralisation) was also well optimised before reaching the baseline. This parameter also has a non-local effect, as it influences the flux of oxygen and phosphate to the deeper ocean, therefore it has the potential to strongly influence the misfit function, but its low sensitivity in this study could be caused by narrow parameter bounds. In order to optimise less sensitive parameters before reaching noise levels one could introduce more metrics into the misfit calculation to help constrain these parameters, such as phytoplankton and zooplankton data or the location of oxygen minimum zones (Niemeyer et al., 2019).

3.3.3 Calibrating to sparse observations

We also investigated the ability of DFO-LS to optimise MOPS in the presence of sparsity in the observational data. The results shown here suggest that there is no significant difference in the performance of DFO-LS when tuning to data at every grid point versus a subset of grid points. Interpolating can introduce large errors, on the order of 20% (Garcia et al., 2018a,b), particularly in poorly sampled regions such as the Southern Ocean. However, our experiments suggest that it is possible to use un-interpolated observations, but it is important to start the optimiser from multiple locations in parameter space, or to generously allow restarts, in the presence of a complex misfit function with many local minima. These multiple runs clearly can be run in parallel.

3.4 Conclusions

This chapter compared the efficiency and performance of two derivative-free optimisation algorithms, CMA-ES and DFO-LS, applied to MOPS, a global ocean biogeochemical model with 7 prognostic tracers. The two methods were used to tune 6 of the key parameters that control the behaviour of MOPS. We found that DFO-LS has a significantly lower computational cost when compared to CMA-ES, between one and two orders of magnitude, which is important considering that global ocean biogeochemical models are computationally expensive, as they must be integrated for several thousand years to reach equilibrium. DFO-LS exploits more information when minimising the misfit function, therefore has more scope for reducing the misfit faster than CMA-ES. However, as DFO-LS is more of a local optimiser than CMA-ES, it should be paired with a globalising method such as starting from different initial points in parameter space, which can easily be run in parallel.

3.5 Code Availability

The base TMM and MOPS code used for the ocean biogeochemical simulations are available to download from `https://doi.org/10.5281/zenodo.1246300`. Transport matrices and forcing fields required to perform the simulations can be downloaded from `https://doi.org/10.5281/zenodo.5517238`. Modifications to the MOPS code for the specific experiments described in this chapter, along with model output and scripts to recreate the figures shown here, are available from `https://doi.org/10.5281/zenodo.5517626`. The OptClim optimisation framework used in this chapter to couple any climate model to any optimiser is available at `https://doi.org/10.5281/zenodo.5517610`. This includes the CMA-ES optimisation code taken from the Supplement of Kriest et al. (2017) and adapted to work with OptClim.

4

An inexpensive parameter sensitivity analysis of a global ocean biogeochemical model

Prior to parameter calibration, it is useful to have some understanding of how sensitive the misfit function is to perturbations in each of the parameters. Unfortunately, this was discovered by the author during the previous parameter optimisation chapter, particularly when the optimisation algorithms struggled to successfully calibrate the parameter K_{PHY} (phytoplankton half-saturation for phosphate) in the global ocean biogeochemical model MOPS. In hindsight, the sensitivity study completed in this current chapter would have been carried out prior to the previous chapter. However, to keep a consistent timeline, we display the MOPS parameter sensitivity study here.

Global ocean biogeochemical models are computationally expensive to run, therefore parameter sensitivities are often too costly to obtain. In an attempt to remedy this we use statistical techniques to perform a global parameter sensitivity study of 6 parameters within MOPS, derived from an inexpensive "surrogate" formulation of the true misfit function. We show that this prior study can be done inexpensively with the use of the Kriging method when trained with as few as 50 samples of the true misfit function. These 50 samples can be run in parallel, therefore reducing the sequential computational time required. We then go on to show that the expense of each individual sample can be reduced by shortening the spin-up time of the biogeochemical model to 2000 years, without significantly influencing the sensitivity results when compared to a longer 3000 year spin up to assumed equilibrium. This work will hopefully allow more extensive investigation of the misfit function and parameter sensitivities prior to an optimisation study of expensive models, such as the marine biogeochemical component within Earth System Models used to predict future climate change. The structure of this chapter is as follows: first the introduction in Section 4.1, followed by the methodology in Section 4.2, results in Section 4.3, discussion in Section 4.4, conclusions in Section 4.5, and finally the code availability in Section 4.6.

4.1 Introduction

Prior to an optimisation study it is useful to carry out a sensitivity analysis of the chosen parameters to be calibrated, as some parameters may have a smaller influence on the misfit than others, thereby likely making them more difficult to successfully calibrate (Shen et al., 2016). A sensitivity study will also provide insight into the internal processes of the model, and hence marine biogeochemistry as a whole, and how varying certain parameters influence the model outputs.

Sensitivities can broadly be categorised into two groups, local and global. Local sensitivities are relatively computationally inexpensive, yet provide the sensitivity of the misfit to only one parameter, at only one location within the n-dimensional parameter space. Local sensitivities are calculated by perturbing the parameters by only a small fraction from a chosen value while all other parameters remain constant, therefore the sensitivity to this parameter in the rest of the parameter space has been left unexplored. On the other hand, global sensitivities show the misfit's sensitivity to each parameter across the entire n-dimensional parameter space. However, they are very costly to compute and therefore rarely applied to ocean biogeochemical models (e.g. Prieur et al., 2019; Wang et al., 2018; Fennel et al., 2001).

One such global sensitivity method is the Sobol' index (Sobol', 1993; Sobol, 2001), which has been shown by Tang et al. (2006) to be one of the more robust global sensitivity methods by applying various methods to a watershed model, and has since been introduced to the biogeochemical modelling community by Prieur et al. (2019). In the latter the authors calculated Sobol' indices for parameters of a 1-dimensional biogeochemical model, requiring more than 10 thousand model evaluations. As their model took 50 seconds to run, this large number of parallel evaluations was feasible, however in our case with a global ocean biogeochemical model coupled with an "offline" circulation the run times have units of hours, making this method much less feasible.

In this chapter we strive to discover a computationally inexpensive yet accurate method to provide a global parameter sensitivity study. Here we use the Sobol'

indices approach, similar to Prieur et al. (2019), to present a global parameter sensitivity study of 6 chosen MOPS parameters. However, instead of evaluating the true expensive misfit function thousands of time, we formulate a computationally inexpensive "surrogate" model of the misfit function, then carry out the sensitivity analysis on the inexpensive surrogate. We train the surrogate with a range of sample sizes (50-110) to determine the fewest required to calculate accurate sensitivities. We then investigate the impact on the sensitivity results of shortening the biogeochemical model spin up time, to discover if the computational expense of the surrogate training data can be further reduced.

4.2 Methodology

4.2.1 Ocean biogeochemical model misfit function

The chosen global ocean biogeochemical model is the Model of Oceanic Pelagic Stoichiometry (MOPS-2.0: Kriest and Oschlies, 2013, 2015), coupled to an "offline" version of the MITgcm (2.8 degree) ocean model (Marshall et al., 1997) via the Transport Matrix Method (TMM; Khatiwala et al., 2005; Khatiwala, 2007). For more information see Section 2.2. One MOPS simulation initially starts with uniform distributions of biogeochemical tracers, then simulates the cycling of tracers for a specified spin up length (e.g. 3000 years to an assumed equilibrium state (Wunsch and Heimbach, 2008)), and finally outputs 3-dimensional fields of averaged final year global oceanic tracers.

For this sensitivity study 6 biogeochemical parameters were selected, the same as by Kriest et al. (2017) and as in Chapter 3, and are described in Table 4.1.

The misfit function describes how the model-observations misfit varies with parameter values, throughout the 6-dimensional parameter space. The observations used here are synthetic 3-dimensional tracer distributions, provided by a previous model run with a known parameter configuration, as in Chapter 3. As explained in Chapter 3, we combine regional oxygen, phosphate and nitrate misfits, weighted by volume and normalised by the mean of the observations, into a single global misfit (see Equations 3.1 and 3.2, where $\epsilon = 0$).

Abbrev.	Description	Units	Synthetic observations	Lower bound	Upper bound
$R_{O_2:P}$	Ratio of oxygen consumption to phosphate release	mmol O_2 : mmol P	170	150	200
I_C	Phytoplankton half-saturation for light	W m^{-2}	24	4	48
K_{PHY}	Phytoplankton half-saturation for phosphate	mmol P m^{-3}	0.03125	0.0001	0.5
μ_{ZOO}	Zooplankton maximum grazing rate	phytoplankton d^{-1}	2	0.1	4
k_{ZOO}	Zooplankton quadratic mortality rate	mmol P $(m^{-3})^{-1}$ d^{-1}	3.2	0	10
b^*	Exponent of the "Martin curve" (Martin et al., 1987)		0.858	0.4	1.8

Table 4.1: The 6 MOPS biogeochemical parameters chosen for the sensitivity study, including the parameter values used to create the synthetic observations, and the lower and upper bounds within which are feasible parameter values.

4.2.2 Kriging method

To be able to calculate expensive global parameter sensitivities, the misfit function needs to be evaluated at many different points (often thousands) within parameter space. A cheaper alternative would be to sample the true misfit function at several points, then interpolate the misfit between these sampled points to get an approximation of the misfit function. Evaluating the interpolated surface at any required points could then be used to calculate the global sensitivities, without having to evaluate the expensive true misfit function thousands of times.

Our interpolated version, or "surrogate" misfit function was created using the Kriging method, first developed by Krige (1951), and recently described by Martin and Simpson (2005). This method assumes the misfit function is a Gaussian process, and therefore can be fully described by a mean and covariance. The Kriging surrogate was trained on samples of the true misfit function, which were sampled according to the Latin Hypercube Sampling (LHS) strategy (McKay et al., 1979) with the MATLAB function **lhsdesign**. This is a stratified-random strategy, whereby the sampling space is divided into equal sections and then samples are chosen randomly from within each section. This is preferable over an expensive uniform sampling strategy (see Figure 4.1), as it requires fewer samples yet still ensures good coverage of the parameter space with no large unsampled areas. The Kriging model was trained using the MATLAB-based software package UQLab (Marelli and Sudret, 2014), which contains extensive Kriging documentation and coding examples (Lataniotis et al., 2019).

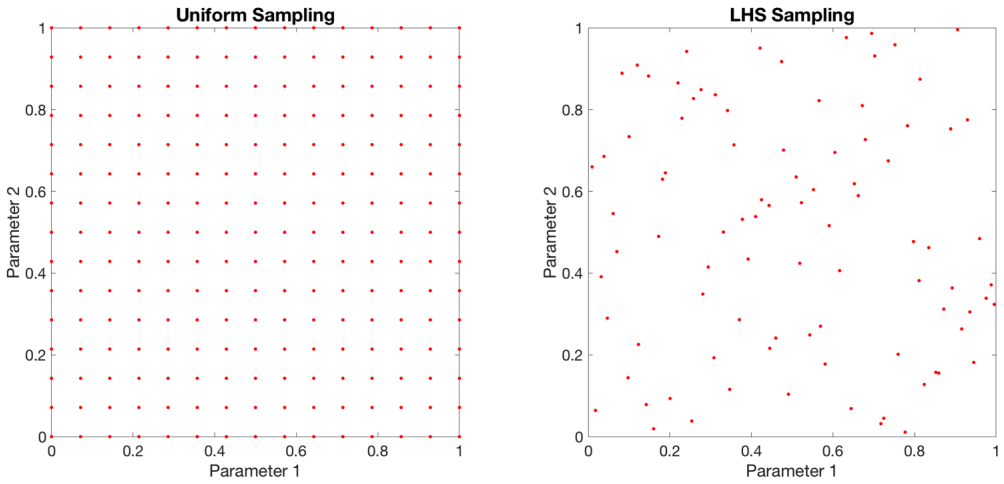

Figure 4.1: An example of sampling a 2-dimensional parameter space using (left) uniform sampling, and (right) LHS sampling.

The Kriging surrogate estimates the global misfit at an un-sampled yet requested location within parameter space ($M^K(\mathbf{x})$) as a superposition of both the trend and variance of the Gaussian process. This is described by Santner et al. (2003);

$$M^{\mathrm{K}}(\mathbf{x}) = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{h}(\mathbf{x}) + \sigma^2 Z(\mathbf{x}, w). \tag{4.1}$$

T is the transpose, causing the first term $\boldsymbol{\beta}^T\mathbf{h}(\mathbf{x})$ to be the summed component-wise multiplication of $\boldsymbol{\beta}$ and $\mathbf{h}(\mathbf{x})$. This is the polynomial function representing the mean or trend of the Gaussian process over parameter space in $\mathbf{x}$. After some investigation of the Kriging performance, we chose an Ordinary Kriging trend whereby the trend has a constant yet unknown value, instead of for example a Universal Kriging trend whereby the trend could be a polynomial. The second part $\sigma^2 Z(\mathbf{x}, w)$ is its variance σ^2 and zero-mean Gaussian process $Z(x, w)$, a function of the parameter values $\mathbf{x}$ and underlying probability space w.

As an indication of surrogate model error, we calculate leave-one-out cross validation errors and validation errors. The former involves creating $N - 1$ Kriging surrogates (where N is the number of training samples), each trained by the N sized sample set but with one left out, and comparing to the Kriging surrogate trained by the full N sized sample set, and the latter involves comparing the surrogate to the validation data set (see Section 1.7.1 and 1.7.2 respectively in Lataniotis et al. (2019)).

4.2.3 Parameter sensitivities

Local sensitivities for each of the 6 parameters were calculated at one location within parameter space, whereby only the parameter whose sensitivity was to be calculated was perturbed by 10% of its range while the other parameter values remained constant.

We then derived global sensitivities from the Kriging surrogate of the misfit function. We used the state-of-the-art variance-based Sobol' method (Sobol', 1993; Sobol, 2001), recently explained by Marelli et al. (2019). This relies on the ANOVA (ANalysis of VAriance) method whereby the variance of a function $Y = f(X)$ can be decomposed into fractions and attributed to a specific function variable X_i. A function can be decomposed as such:

$$Y = f_0 + \sum_{i=1}^{d} f_i(X_i) + \sum_{i<j}^{d} f_{ij}(X_i, X_j) + ... + f_{1,2,...,d}(X_1, X_2, ..., X_d), \text{ where} \quad (4.2)$$

f_0 = a constant, $E(Y)$,

$f_i(X_i) = E(Y|X_i) - f_0$,

$f_{ij}(X_i, X_j) = E(Y|X_i, X_j) - f_0 - f_i - f_j$.

Therefore the function's variance can also be decomposed;

$$\text{Var}(Y) = \sum_{i=1}^{d} V_i + \sum_{i<j}^{d} V_{ij} + ... + V_{1,2,...,d}, \text{ where} \quad (4.3)$$

$V_i = \text{Var}_{X_i}(E_{\mathbf{X}_{\sim i}}(Y|X_i))$,

$V_{ij} = \text{Var}_{X_{ij}}(E_{\mathbf{X}_{\sim ij}}(Y|X_i j)) - V_i - V_j$,

$\mathbf{x}_i = (X_1, ..., X_{i-1}, X_{i+1}, ..., X_d)$.

The First- and higher-order Sobol' indices $S_{i,j,...}$ represent the relative contribution to the function's total variance by a single variable X_i (First-order) or group of variables (e.g. Second-order X_{ij}, Third-order X_{ijk}, etc), averaged over variations in the excluded variables ($X_{\sim i}$, $X_{\sim ij}$, or $X_{\sim ijk}$, etc);

$$S_i = \frac{V_i}{\text{Var}(Y)} \text{ and } S_{i_1,...,i_s} = \frac{V_{i_1,...,i_s}}{\text{Var}(Y)}. \quad (4.4)$$

The Total Sobol' Index S_i^T is used to show the contribution of one variable X_i including all of the interactions with all other variables $X_{\sim i}$, and it is the sum of the first- and higher-order indices;

$$S_i^T = \sum_{i_1,\ldots,i_s \supset i} S_{i_1,\ldots,i_s}. \tag{4.5}$$

Essentially, in this chapter, Sobol' indices allow us to understand the relative influence of each input parameter on how well the model MOPS fits to some observations. Sobol' indices are able to distinguish between direct effects of these parameters individually, and effects due to interactions between multiple parameters. This aids one to pinpoint which parameters are most and least influential to the misfit function.

The Sobol' method has been shown to be one of the more robust methods to use (Tang et al., 2006), and has previously been applied to climate models (Zhang et al., 2015; Shi et al., 2019) and a biogeochemical model (Prieur et al., 2019). There are different ways in which you would sample parameter space to calculate Sobol' Indices, such as Monte Carlo (MC). For MC-based estimation the number of function evaluations scales with $M(n+2)$, where n is the number of parameters and M is preferably more than 1000. MC based estimation is robust but needs thousands of evaluations, hence the need for a cheap way of evaluating the misfit function, such as with the Kriging surrogate.

In UQLab (Marelli and Sudret, 2014) these Sobol' Indices were computed from the Kriging surrogate models by means of Monte Carlo-based estimation (Marelli et al., 2019), whereby the surrogate models were evaluated ∼8000 times.

4.2.4 Experimental Design

To sufficiently sample parameter space for surrogate training samples using the LHS method, Loeppky et al. (2009) found the number of samples should be at least ten times the number of parameters, yet Afzal et al. (2017) and Wang et al. (2014) suggest at least fifteen times. With 6 parameters, this would mean a training sample size of 60-90. To discover the smallest sample size necessary to still derive accurate

sensitivities from the surrogate, we create several Kriging surrogates using various training sample sizes. MOPS was spun up for 3000 years to assumed equilibrium and misfits were calculated for 4 different LHS sample set sizes, with 50 ("**LHS 50**"), 70 ("**LHS 70**"), 90 ("**LHS 90a**"), 110 ("**LHS 110**") samples. A 5th LHS set ("**LHS 90b**") was also used to create a surrogate, with 90 samples at different parameter locations to **LHS 90a**. The misfits from the 5 different LHS sample sets were then used as training data to create 5 different Kriging models respectively, each of which were a surrogate of the MOPS 6-dimensional misfit function. These 5 Kriging models were each validated against the 4 other LHS sample sets not used as its training data (shown in Section 4.3.1), then used to derive Sobol' indices. The relative importance of each parameter suggested by the global sensitivities of the Kriging surrogates were then compared to the relative importance of each parameter by the local sensitivities derived directly from the MOPS model (shown in Section 4.3.2).

After the above investigations to reduce the surrogate training sample size, we attempt to reduce the computational cost of the training samples further by shortening the spin up time of the biogeochemical model. A long model spin up is required to reach an equilibrated state due to the long time scales involved with deep ocean circulation (Wunsch and Heimbach, 2008; Primeau and Deleersnijder, 2009; Khatiwala et al., 2012). However, if the Sobol' indices derived from a Kriging surrogate trained by samples with a shortened spin up are similar to those by samples with a long, equilibrated spin up, then the computational cost of the training samples can be reduced without deteriorating the sensitivity results. Therefore, using an example LHS training sample size of 110, we created 6 more Kriging surrogates of the misfit using training samples with a spin up time of 100, 500, 1000, 1500, 2000, 2500 years (and we have the Kriging surrogate trained with 3000 year spin up samples from previously). Sobol' indices were then calculated from these surrogates to investigate the influence of a shortened spin up on the parameter sensitivities (shown in Section 4.3.2).

4.3 Results

4.3.1 Kriging surrogate validation

The performance of the Kriging surrogate models were validated against the true MOPS misfits at more than 300 validation points, provided by the LHS sets not used as training data for that particular Kriging model, and shown in Figure 4.2. The relationship between the true misfit (Y_{true}) and the surrogate misfit ($Y_{Kriging}$) is relatively good for all 5 surrogates, in the sense that even though there is some scatter around the 1:1 line, the overall correspondence is realised. The Kriging model trained by **LHS 110** predicted the true misfit most accurately, and least accurately when trained by **LHS 90a** and **LHS 90b**. Note that validation points with a low Y_{true} misfit can correspond to a $Y_{Kriging}$ misfit between approximately just below zero to 0.2, therefore the surrogates may not be useful for locating specific minima in the true misfit function but may provide insight to the overall shape of the misfit landscape and global parameter sensitivities.

4.3.2 Parameter sensitivities

The results of the local parameter perturbation study derived from the MOPS model are shown in Figure 4.3, where the greatest change in the true MOPS misfit was caused by perturbing b^* by 10% of its range. The next greatest change in the misfit was caused by I_C and k_{ZOO}, while the parameter with the lowest influence on the misfit at this local point in parameter space was K_{PHY}.

The global sensitivity results derived from the Kriging surrogates trained by various sample sizes are shown in Figure 4.4. The first-order Sobol' indices, which indicate the fraction of the surrogate misfit variance caused by each individual parameter on its own, without interactions with other parameters, showed the dominant cause of variance to be by b^*, followed by I_C. The second-order Sobol' indices, which indicate the fraction of the surrogate misfit variance caused by interactions between two parameters, showed the strongest interactions to be between b^* and I_C, and then between b^* and μ_{ZOO}. This then explained why the total Sobol'

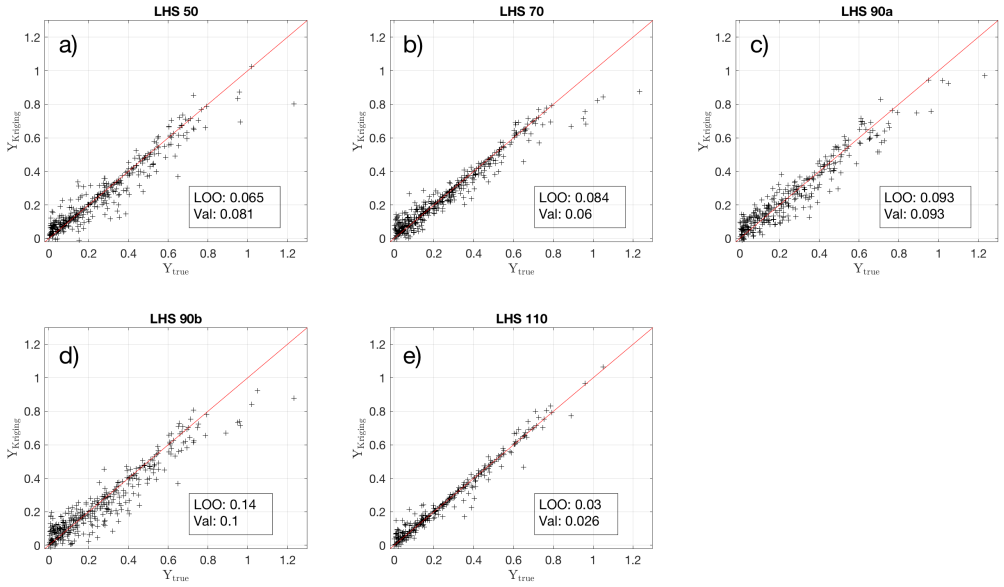

Figure 4.2: Validation of the Kriging surrogate models trained with sample sets (a) **LHS 50**, (b) **LHS 70**, (c), **LHS 90a**, (d), **LHS 90b** and (e) **LHS 110** against the true MOPS model. The misfits of both the true MOPS model (Y_{true}) and Kriging surrogates ($Y_{Kriging}$) were evaluated and plotted at more than 300 validation points in parameter space. The function $y = x$ has been plotted (red solid line) to show the ideal location for the points to plot, in the case where the Kriging correctly predicts the true misfit. Leave-one-out cross validation errors ("LOO") and validations errors ("Val") are also shown (see Section 4.2.2).

indices, which indicate the fraction of the surrogate misfit variance caused by each individual parameter allowing for all interactions with all other parameters, showed the dominant causes of misfit variance to be by b^*, I_C and μ_{ZOO} (in decreasing order). The parameter which causes the lowest variance in the misfit, and therefore the parameter which the surrogate misfit function had the least sensitivity to, is K_{PHY}. It was encouraging to note that varying the surrogate training sample locations and sizes did not significantly influence the Sobol' indices results, indicating how robust this method is, and indicating a smaller training sample size of as few as 50 is required for reliable sensitivities of the 6-dimensional misfit function.

When comparing to the local sensitivities calculated using the true MOPS misfit, the global sensitivities of the Kriging surrogates also showed the MOPS misfit was most sensitive to changes in b^* and I_C, and least sensitive to K_{PHY}. However, the local sensitivities of the true MOPS misfit indicated slightly higher and lower

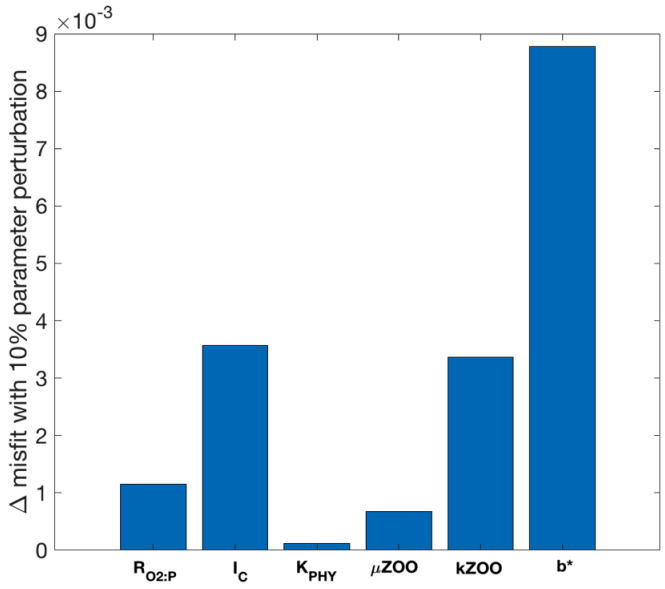

Figure 4.3: Bar graph of the misfit change due to individual parameter perturbations by +10% of their range.

sensitivities to the parameters k_{ZOO} and μ_{ZOO} respectively, than indicated by the global sensitivities derived from the surrogates. Therefore, if only local sensitivities had been calculated, this may have led us to place a greater importance to k_{ZOO} and less to μ_{ZOO}.

The global sensitivity results derived from the Kriging surrogates trained by the **LHS 110** samples with various biogeochemical model spin up lengths are shown in Figure 4.5. Sobol' indices derived from surrogates trained by samples with spin up lengths of 2000 years and longer showed similar results to those seen previously in Figure 4.4. However, below 2000 years, the shorter the spin up the greater the deterioration in sensitivities. A shortened spin up resulted in decreased second-order, increased first-order, and decreased total Sobol' indices. Summarily, a shortened spin up led to less misfit variance being attributed to parameter interactions but a greater importance given to individual parameters excluding parameter interactions, resulting in a slight net decrease in misfit variance attributed to parameters including all parameter interactions.

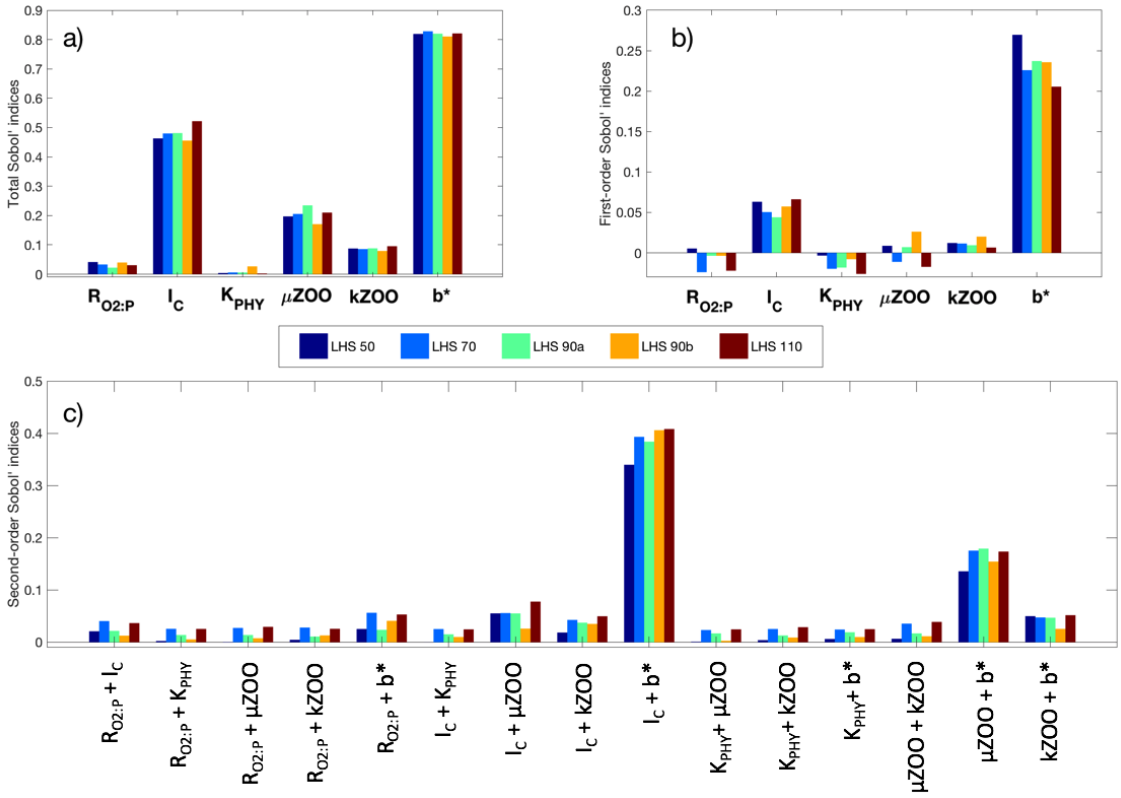

Figure 4.4: Sobol' indices derived from the 6 Kriging surrogate models trained by LHS sets **LHS 50**, **LHS 70**, **LHS 90a**, **LHS 90b**, and **LHS 110** (see figure legend for corresponding colours). (a) Total Sobol' indices, indicating the fraction of misfit variance caused by each parameter, including its interactions with all other parameters. (b) First-order Sobol' indices, indicating the fraction of misfit variance caused by each parameter, excluding its interactions with all other parameters. (c) Second-order Sobol' indices, indicating the fraction of misfit variance caused by parameter interactions between specified parameter pairs. Note that Sobol' indices are not constrained to be positive values.

4.4 Discussion

4.4.1 Performance of the Kriging surrogate model

The Kriging surrogate model was based on relatively inexpensive training data on a Latin Hypercube and allowed cheap and efficient exploration of the global misfit function within the 6-dimensional parameter space. We only emulate the 6-dimensional misfit function, which can be created using fewer true MOPS evaluations than an emulator of the three-dimensional ocean biogeochemical outputs, such as done by Kasim et al. (2020). We used 50-110 MOPS evaluations to create the surrogates, which can be generated relatively quickly as they can be run in

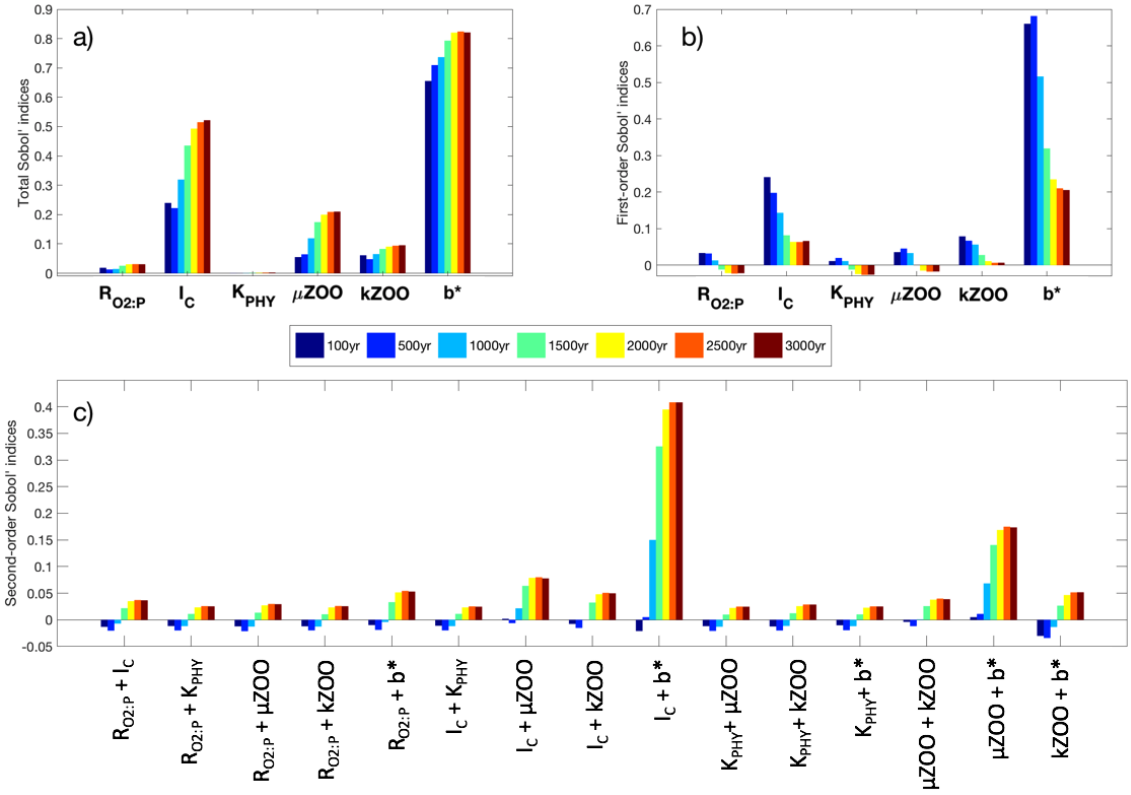

Figure 4.5: As in Figure 4.4, but for Sobol' indices derived from the 7 Kriging surrogate models trained by the **LHS 110** sample sets with spin up lengths of 100, 500, 1000, 1500, 2000, 2500 and 3000 years (see figure legend for corresponding colours).

parallel. All surrogates, whether trained by as few as 50 or as many as 110 training samples, provided a continuous function from which to gain reliable global parameter sensitivities across the whole parameter domain, instead of a partial scan of parameter space (Kriest et al., 2012). When comparing misfits calculated from the true MOPS model and the surrogate models, we found that they compared quite well, in the sense that low/high surrogate misfits corresponded to low/high true MOPS misfits respectively, however there was some scatter around a 1:1 relationship between them.

When comparing the surrogate-derived global parameter sensitivities with the MOPS-derived local sensitivities, again the surrogates correctly predicted the overall shape, with the correct general pattern of which parameters MOPS was most and least sensitive to, but with some discrepancies for the intermediate sensitive parameters. It is difficult to attribute differences due the possible inadequacies in

the surrogate models, or due to the difference in global and local sensitivities. This may be useful for an initial global sensitivity study, but your confidence in resulting sensitivities relies on the quality of the surrogate model. In our case we found all surrogates, even trained by as few as 50 training samples, were adequate for an initial global parameter sensitivity study.

4.4.2 Importance of a parameter sensitivity study

An initial sensitivity study of the parameters (such as done here, or for example by Kwon and Primeau (2006); Ji et al. (2015); Kriest et al. (2012); Kwon and Primeau (2008)) indicates which parameters have the most influence on the misfit function, and hence which should be included in any subsequent optimisation study. Conversely, it will also help decide which parameters an optimiser will have difficulty optimising, as they may have little to no influence on the misfit function. This can help reduce the dimensionality of the problem and therefore the expense of an optimisation study.

In the case of our misfit function, both the surrogate-derived global and MOPS-derived local sensitivities indicated b^* and I_{C} significantly influence the global misfit, therefore explaining why Kriest et al. (2017) and our work in Chapter 3 managed to optimise these parameters successfully and more quickly than other parameters. On the other hand, Kriest et al. (2017) and Chapter 3 had some difficulty when optimising K_{PHY}, which we have found here corresponded to a very low parameter sensitivity. This agrees with Kriest et al. (2012) who found their biogeochemical model misfit metric to be particularly sensitive to parameters influencing particle flux and sinking yet conversely less sensitive to plankton parameters. Note that K_{PHY} is not unimportant in an ocean biogeochemical sense, but in the sense that it had little effect on our defined misfit function which only included phosphate, nitrate and oxygen information and is dominated by the global deeper ocean. An initial parameter sensitivity study would allow this insensitive parameter to be screened out prior to optimisation, or lead to a modification of the misfit function

to better constrain this parameter, such as by weighting to the surface ocean or including phytoplankton data in the calibration observations (Zhao et al., 2005).

4.4.3 Can we reduce the ocean biogeochemical spin up time?

The long spin up of ocean biogeochemical models allow the global chemical distributions to reach equilibrium, after which there is very little change with time, which takes approximately 1500 years for the Atlantic Ocean and at least 3000 years for the Pacific (Wunsch and Heimbach, 2008; Primeau and Deleersnijder, 2009; Khatiwala et al., 2012). The surrogate models used for a sensitivity analysis in this chapter required a number of training samples of the MOPS model, which were relatively computationally expensive with a 3000 year spin up. We aimed to discover if the spin up of the training samples could be shortened, thereby reducing the computational expense of the surrogate. We found the spin up could only be shortened to 2000 years before the Sobol' sensitivity results deviated significantly from those of a 3000 year spin up, as spin ups shorter than this started to reduce the importance of interactions between parameters on the misfit variance.

4.5 Conclusions

This chapter carried out a computationally inexpensive global sensitivity analysis of 6 parameters within the global ocean biogeochemical model MOPS, with the use of the Kriging method to create a surrogate formulation of the model-observations misfit function throughout the 6-dimensional parameter space. This inexpensive surrogate of the misfit function allowed us to calculate Sobol' indices, which indicate the fraction of misfit variance caused by each parameter, both individually and allowing for parameter interactions, across the entire parameter space. Sobol' indices need thousands of function evaluations, which would be unfeasible without the inexpensive surrogate of the misfit function. We found the training sample size required by the surrogate to still result in reliable Sobol' indices was as few as 50, which of course can be run in parallel. We also found the initial expense of these 50 samples can be further reduced by shortening the ocean biogeochemical

model spin up from 3000 years to 2000 years, while still resulting in reliable Sobol' indices. Future work will include further investigation of this, and how shortening the spin up influences the misfit function, as shortening the spin up of ocean biogeochemical models during sensitivity and optimisation studies would greatly reduce their computational expense.

The work done here investigated a computationally cheap method of otherwise expensive global sensitivity analysis, which is especially useful when dealing with complex models such as global ocean biogeochemical models. It will allow for easier parameter screening prior to ocean biogeochemical optimisation studies, hence reduce the misfit function dimensionality and therefore computational expense of model calibration.

4.6 Code Availability

The base TMM and MOPS code used for the ocean biogeochemical simulations are available to download from `https://doi.org/10.5281/zenodo.1246300`. Transport matrices and forcing fields required to perform the simulations can be downloaded from `https://doi.org/10.5281/zenodo.5517238`. The MATLAB framework used to create the Kriging surrogates and calculate Sobol' indices were sourced from the Framework for Uncertainty Quantification (UQLab) which can be downloaded here: `https://www.uqlab.com/download`.

5

An investigation of a computational "shortcut" in ocean biogeochemical model optimisation

We investigate two ideas to reduce the computational expense of global ocean biogeochemical optimisation: Does optimisation of a biogeochemical model with a short spin-up provide the same optimised parameters as one with a full-length equilibrated spin-up? Can one provide more information to the optimiser in the form of monthly-averaged data instead of annually to ensure faster calibration? To test these, we used the derivative-free optimiser DFO-LS to calibrate up to 6 biogeochemical parameters within the global ocean biogeochemical model MOPS to synthetic oxygen, phosphate and nitrate observational data. Disappointingly, neither idea ensured successful, faster optimisation. Optimisation of a short model spin-up resulted in incorrect optimised parameters because the misfits after a short spin-up and a 3000 year spin-up do not always relate in a monotonic, predictable way. This is particularly true when optimising parameters which influence the flux of organic matter to the deeper ocean, where oceanic timescales are longer. This finding indicates caution should be used when utilising shortened spin ups, as is frequently the case in biogeochemical modelling. Also, providing the optimiser with more information in the form of monthly observational data did not ensure quicker successful optimisation, possibly because our chosen misfit function is heavily weighted to the deeper ocean where the seasonal cycle has less influence. This chapter has hopefully satisfied these lines of enquiry, allowing investigation of other methods to continue to strive for more efficient, systematic global ocean biogeochemical optimisation. The structure of this chapter is as follows: first the introduction in Section 5.1, followed by the methodology in Section 5.2, results in Section 5.3, discussion in Section 5.4, conclusions in Section 5.5, and finally the code availability in Section 5.6.

5.1 Introduction

The spin-up of an ocean biogeochemical model is to allow the global tracer distributions to reach a sufficient equilibrium state, after which there is very little change with time. During parameter optimisation, for a deviation in a parameter value to be fully felt throughout the entire global ocean biogeochemical model the

model must be spun up to full equilibrium, after which the quality of the model can be re-assessed. However, as this is very computationally expensive, the spin-up time of these models have been shortened and vary greatly in length in the many ESM predictions (Séférian et al., 2016). When the spin-up time is not long enough for the system to reach equilibrium then the outcomes are dependent on the initial tracer distributions. These initial distributions are commonly either (1) uniform distributions, (2) distributions from real present day observations (when available), or (3) distributions output by a previous, equilibrated model run.

What needs to be known is whether tuning the parameters of a non-equilibrated ocean biogeochemical state still provides correct optimal parameter values for that ocean biogeochemical model, as if it had been allowed to reach full equilibrium. If yes, then a computational “short cut" can be taken, as these models can be calibrated to observations more cheaply by optimising a shorter spin-up. However, if no, then full equilibrium is needed, and any optimisation of a shorter spin up needs to be considered carefully. With this possible computational short cut comes certain questions, such as does this short cut depend on the biogeochemical parameters being optimised? How does this depend on initial tracer distributions? And, can we tailor an ocean biogeochemical optimisation study to have the correct conditions to allow for this short cut to be utilised?

Another aspect often overlooked is systematic optimisation of the ocean biogeochemical seasonal cycle. Typically, ocean biogeochemical models are optimised using the annually-averaged tracer distributions of the final year of a specified spin-up time, which can be condensed into a single misfit value to be optimised (e.g. Kriest et al., 2017, 2020), or into multiple misfits to be optimised (e.g. Sauerland et al., 2019; Oliver et al., 2021). If the seasonal cycle were considered also, such as by optimising the monthly-averaged tracer distributions of the final year of a specified spin-up time, then more information is considered in the optimisation process. If this increased information were to enable faster optimisation, then it may be worth optimising the monthly-averaged model outputs instead of the annually-averaged

model output. The added bonus is that the optimisation process will improve the model's simulated seasonal cycle.

5.2 Methods

5.2.1 Ocean biogeochemical model

As in Chapters 3 and 4, here we use the Model of Oceanic Pelagic Stoichiometry (MOPS; Kriest and Oschlies, 2013, 2015) coupled to the "offline" MITgcm (2.8 degree) ocean model (Marshall et al., 1997) via the Transport Matrix Method (TMM; Khatiwala et al., 2005; Khatiwala, 2007, 2018). For more information see Sections 2.1.1 and 2.2.

MOPS simulations initially started with either uniform distributions of tracers or equilibrated tracer distributions output by a previous 3000 year model run. MOPS simulated the cycling of tracers for a specified spin-up time (either 3000 years to assumed equilibrium state, or shorter). MOPS outputs either annually-averaged or monthly-averaged 3-dimensional fields of global biogeochemical tracers at specified intervals throughout the spin-up, including the final year.

5.2.2 Biogeochemical parameters

In this chapter, as in Kriest et al. (2017) and Chapter 3, the parameters to be optimised are: (1) $R_{O_2:P}$, ratio of oxygen consumption to phosphate release [mmolO_2 :mmolP] during remineralisation when oxygen is available, (2) I_C, the phytoplankton half-saturation for light [Wm^{-2}], (3) K_{PHY}, the phytoplankton half-saturation for phosphate [mmolPm^{-3}], (4) μ_{ZOO}, the zooplankton maximum grazing rate [phytoplankton d^{-1}], (5) k_{ZOO}, the zooplankton quadratic mortality rate [(mmol Pm^{-3})$^{-1}$ d^{-1}], and (6) **b***, the exponent of the remineralisation Martin Curve (Martin et al., 1987) where the vertical flux of sinking organic matter is related to a constant raised to the power of -b*. Their bounds are stated in Table 5.2.

5.2.3 Misfit function

To indicate model performance during the optimisation process, we calculate the misfit between the model outputs and observations. This chapter does not move beyond "twin" experiments where the model is optimised using synthetic observations output by a previous model run with known parameter values. As in Chapter 3 the twin synthetic observations were output by the reference MOPS run "MOPS-Ref", initiated with uniform tracer distributions and spun up for 3000 years, with the reference parameter values stated in Table 5.2. The twin observations were global distributions of specified tracers, either annually-averaged or monthly-averaged for the final year of the 3000 year spin-up. The aim of the optimiser is to recover these reference parameters.

For each parameter configuration investigated by the optimiser, the misfits were calculated for three tracers (oxygen, phosphate and nitrate), 19 oceanic biome regions (a combination of Henson et al. (2010) and Weber et al. (2016), as in Chapter 3, either averaged over one year or over 12 separate months after a specified spin-up length. Therefore each calculation results in 57 (annual) or 684 (monthly) misfit values r_{qjt} to be given to the optimiser, shown in Equation 5.1. These are a root-mean square error RMSE between the biogeochemical model output m and twin observations o. They are weighted first by individual grid point volume within the region V_i as a fraction of the region's total volume V_j, then further weighted according to the region's volume V_j as a fraction of the sum of all regional volumes V_T. The square root of this is normalised by the twin observations, weighted the same as the previous term, so that the units of each tracer are normalised. When this was calculated for monthly averages instead of only annual averages, the calculation included the indexing term t to indicate the month of the year.

$$r_{qjt}(\mathbf{x}) = \frac{1}{(\sum_{i=1}^{N_j} o_{iqt}\frac{V_i}{V_j})\frac{V_j}{V_T}}\sqrt{(\sum_{i=1}^{N_j}(m_{iqt}(\mathbf{x}) - o_{iqt})^2\frac{V_i}{V_j})\frac{V_j}{V_T}}, \tag{5.1}$$

where $\mathbf{x}$ are the parameters, $V_T = \sum_{j=1}^{19} V_j$ and N_j is the total number of model grid boxes in biome region j. These individual misfits were given straight to the

optimiser, however for plotting clarity in the results section we commonly plot a global misfit, which for the annually-averaged misfits is

$$f_{annual}(\mathbf{x}) = \sum_{q=1}^{3} \sum_{j=1}^{19} r_{qj}(\mathbf{x})^2, \tag{5.2}$$

and for the monthly-averaged misfits is

$$f_{monthly}(\mathbf{x}) = L \frac{\sum_{q=1}^{3} \sum_{j=1}^{19} \sum_{t=1}^{12} r_{qjt}(\mathbf{x})^2}{M}, \tag{5.3}$$

where L is the number of annual individual misfits (in our case 57) and M is the number of monthly individual misfits (in our case 684), so that the annual and monthly global misfits are comparable.

A baseline misfit was also defined, where the term $(m(\mathbf{x}) - o)$ of the misfit equation is replaced by the standard deviation of the observations (Garcia et al., 2018a,b), below which any misfit reduction by an optimisation algorithm is within observational noise. In this chapter we determine an optimisation as successful if the misfit is reduced to below this baseline misfit.

5.2.4 Optimisation algorithm

Here we use the derivative-free optimisation using least squares (DFO-LS) algorithm (Cartis et al., 2019), described in Section 2.3.2. DFO-LS is a local solver, therefore to increase the likelihood of finding the global minimum, we enable "soft restarts" and initiate DFO-LS from different starting locations.

5.2.5 Equilibrated vs non-equilibrated misfits

To investigate how the relationships between assumed equilibrated misfits (calculated after a 3000 year spin-up) and non-equilibrated misfits (calculated after various shorter spin-ups) vary throughout the parameter space, MOPS was evaluated multiple times across the parameter space and the misfits calculated at every 100 years throughout a 3000 year spin-up. To inexpensively yet sufficiently sample the parameter space we use the stratified-random strategy Latin hypercube Sampling (LHS; McKay et al., 1979) with the MATLAB function **lhsdesign**. We sample

the space n x 15 times (where n is the number of parameters) resulting in MOPS evaluations at 90 different locations within the 6-dimensional parameter space.
This latin hypercube sampling was done for three different scenarios; LHS1, LHS2 and LHS3. In scenario LHS1 each MOPS evaluation was initiated from uniform tracer distributions; in LHS2 from the synthetic observations (output by MOPS-Ref); and in LHS3 from tracer distribution output by a poor previous model run, hereafter known as MOPS-Worst. MOPS-Worst is the LHS1 model run which resulted in the largest global misfit after 3000 years, and is used as a "worst case" model run to provide initial tracer distributions. The LHS1 model run which resulted in the lowest misfit after a 500 years short spin-up is hereafter known as MOPS-NonLinear (to be used later). For the parameter configurations for each of the named model runs, see Table 5.2.

5.2.6 Twin experiments

Table 5.1 provides the names, initial MOPS conditions, and misfit formulation choice for each of the twin optimisation experiments, and 5.2 shows the starting parameter values of each.
The investigation of optimising MOPS with various spin-up lengths was carried out on the annually-averaged misfit functions. DFO-LS was used to optimise 6 parameters within MOPS with uniform initial tracer distributions spun up for 500, 1000, 2000 and 3000 years ("**t6d_un_X**"). This was then done for MOPS initiated from MOPS-Ref equilibrated tracer distributions ("**t6d_eqR_X**", with an additional experiment with a 100 year spin up), and repeated for MOPS initiated from MOPS-Worst equilibrated tracer distributions ("**t6d_eqW_X**").
To compare the optimisation of annually-averaged and monthly-averaged misfit functions, MOPS was always initiated with uniform tracer distributions and spun up for 3000 years to equilibrium. DFO-LS was initiated from two different starting points within parameter space, and optimised 5 parameters within MOPS for the experiments minimising the annually-averaged ("**t5d_ann_1**" and "**t5d_ann_2**") and monthly-averaged ("**t5d_mon_1**" and "**t5d_mon_2**") misfit functions.

	Twin experiments where MOPS was initiated from:		
Spin-up length	uniform tracer distributions	MOPS-Ref equilibrated tracer distributions	MOPS-Worst equilibrated tracer distributions
100		**t6d_eqR_100**	
500	**t6d_un_500**	**t6d_eqR_500**	**t6d_eqW_500**
1000	**t6d_un_1000**	**t6d_eqR_1000**	**t6d_eqW_1000**
2000	**t6d_un_2000**	**t6d_eqR_2000**	**t6d_eqW_2000**
3000	**t6d_un_3000** **t5d_ann_1** **t5d_ann_2** **t5d_mon_1** **t5d_mon_2**	**t6d_eqR_3000**	**t6d_eqW_3000**

Table 5.1: Names of each twin optimisation experiment. The prefix naming structure for these experiments follow the rule "**tNd_**" (where **t** = twin, **N** = number of parameters to optimise, and **d** = DFO-LS). Then the following can be specified: either "**un**", "**eqR**" or "**eqW**" to indicate uniform, equilibrated MOPS-Ref or equilibrated MOPS-Worst initial tracer distributions respectively, the spin-up length in years, and either "**ann**" or "**mon**" to indicate annual or monthly-averaged misfits. A final number may be used to distinguish between different DFO-LS starting points within the parameter space.

5.3 Results

5.3.1 Equilibrated vs non-equilibrated misfits after shorter spin-ups

The results of the three LHS sets are shown in Figure 5.1. Not surprisingly, all of them showed the longer the spin-up, the stronger the linear relationship between the global misfits after a 3000 year spin-up and shortened spin-ups. When initiating MOPS from non-target observations (LHS1 and LHS3), the misfit relationships between a 3000 year spin-up and spin-ups shorter than 2000 years contained a non-linear "tail" below the y-intercept, where a low misfit after a short spin-up could correspond to a wide range of misfits after 3000 years. The non-linear tail was more prominent for shorter spin-ups, and had merged into the linear relationship by a 2000 year spin-up. However, when initiating MOPS from target observations (LHS2), there is no non-linear tail. Instead, there was some scatter above the linear relationship for shorter spin-ups, especially for a 100-year and 500-year spin-up, yet even with this scatter the lower misfits after a short spin-up generally corresponded

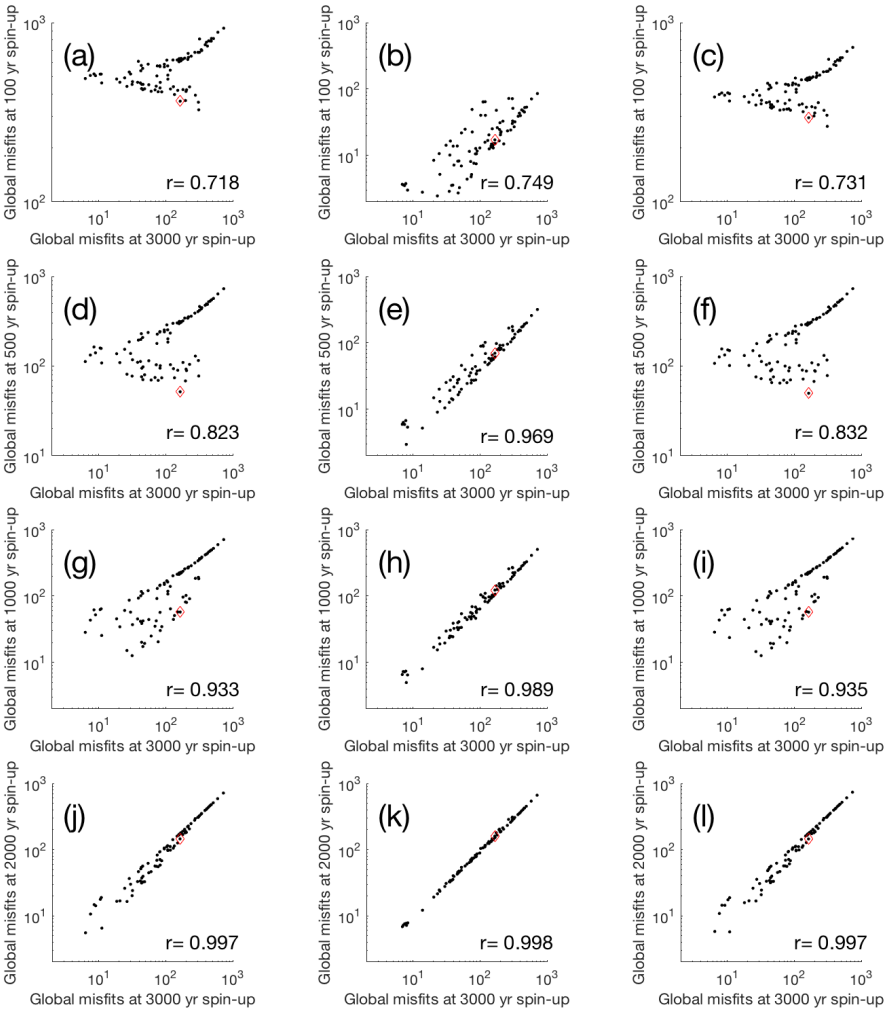

Figure 5.1: Global misfit values for each of the 90 MOPS latin hypercube sets (left subplot column) LHS1 initiated with uniform tracer distributions, (middle subplot column) LHS2 initiated with target twin observations output by MOPS-Ref, and (right subplot column) LHS3 initiated with equilibrated tracer distributions output by MOPS-Worst. These were calculated after a spin-up length of 3000 years on the x-axes, and (a-c) 100, (d-f) 500, (g-i) 1000 and (j-l) 2000 years on the y-axes. The correlation coefficients (r) are given for each scatter relationship. Note the log scale, and the y-axis limits change slightly for subplots a, c, d and f. The red diamonds in the three subplot columns highlight the LHS MOPS runs with the MOPS-NonLinear parameter values, used in Figure 5.8 subplots a, b and c respectively.

Parameters		RO2:P	IC	KPHY	muZOO	kZOO	b*	Evaluations (max/best)	Starting Misfit	Best Misfit	Evaluations to within noise (misfit < 1.076)
Upper Bound		**200**	**48**	**0.5**	**4**	**10**	**1.8**				
Lower Bound		**150**	**4**	**0.0001**	**0.1**	**0**	**0.4**				
Target MOPS-ref		**170**	**24**	**0.03125**	**2**	**3.2**	**0.858**				
MOPS-Worst		**183.590**	**39.630**	**0.474**	**2.204**	**0.445**	**1.659**				
MOPS-NonLinear		**198.775**	**18.901**	**0.055**	**2.925**	**3.155**	**0.543**				
MOPS-Linear		**155.010**	**36.282**	**0.433**	**1.425**	**3.353**	**1.395**				
Experiments: Various spin up lengths											
t6d_un_..	Start	197.000	18.901	0.055	2.925	3.155	0.543				
500	Optimised	200.000	30.730	0.000	1.771	2.246	0.400	80/76	52.160	45.110	Not reached
1000	Optimised	199.465	32.131	0.006	1.611	0.759	0.573	71/48	55.807	9.727	Not reached
2000	Optimised	177.169	25.404	0.076	2.243	2.702	0.796	80/77	141.008	0.579	23
3000	Optimised	169.927	24.064	0.003	1.998	3.249	0.859	72/61	159.209	0.002	13
t6d_eqR_..	Start	197.000	18.901	0.055	2.925	3.155	0.543				
100	Optimised	170.317	23.960	0.022	2.259	4.201	0.876	70/70	16.586	0.017	19
500	Optimised	169.302	23.808	0.183	1.990	3.549	0.871	70/56	67.937	0.058	16
1000	Optimised	168.075	23.096	0.182	2.144	4.190	0.885	70/48	119.736	0.095	18
2000	Optimised	167.991	23.707	0.000	2.045	3.897	0.882	70/44	155.746	0.085	18
3000	Optimised	167.801	23.715	0.000	2.025	3.762	0.883	70/48	161.226	0.091	14
t6d_eqW_..	Start	197.000	18.901	0.055	2.925	3.155	0.543				
500	Optimised	200.000	35.087	0.004	1.800	3.270	0.400	70/49	50.214	42.562	Not reached
1000	Optimised	199.488	28.887	0.000	2.442	1.790	0.622	70/55	55.410	9.639	Not reached
2000	Optimised	177.477	25.278	0.034	2.310	3.585	0.814	70/56	140.658	0.623	27
3000	Optimised	170.589	24.128	0.004	2.102	3.886	0.867	70/45	158.893	0.041	13
Experiments: Annual vs monthly observations											
t5d_ann_1	Start	180.000	40.000	0.03125*	2.500	5.000	0.540				
t5d_ann_1	Optimised	170.135	24.195	0.03125*	1.954	3.098	0.857	70/59	31.949	0.002	14
t5d_ann_2	Start	190.000	30.000	0.03125*	3.500	1.000	1.100				
t5d_ann_2	Optimised	169.967	24.022	0.03125*	2.140	3.670	0.858	70/43	187.824	0.008	12
t5d_mon_1	Start	180.000	40.000	0.03125*	2.500	5.000	0.540				
t5d_mon_1	Optimised	169.752	23.986	0.03125*	2.203	3.642	0.854	70/70	31.894^	0.022^	17
t5d_mon_2	Start	190.000	30.000	0.03125*	3.500	1.000	1.100				
t5d_mon_2	Optimised	169.909	23.975	0.03125*	2.022	3.264	0.858	70/22	188.630^	0.0007^	12

Table 5.2: Table of results for all twin experiments. Upper section states the parameter bounds, MOPS-Ref target parameters to be recovered, and parameters of several MOPS configurations. Lower section shows each experiment's results. Columns 2-7: either the starting or optimised parameter values, as specified in Column 1. Column 8: the maximum number of evaluations completed / the evaluation which provided the lowest or "best" global misfit. Column 9: the starting global misfit. Column 10: the lowest or "best" global misfit. Column 11: the number of evaluations needed for the global misfit to be reduced to below the baseline misfit of observational noise. * = fixed value during optimisation. ^ = normalised monthly misfit as in Equation 5.3.

to a lower misfit after a 3000-year spin-up. As expected, the misfits within LHS2 after a 100-year spin-up were much lower than for LHS1 and LHS3, as the latter were initiated with tracer distributions very different from the target tracer distributions.

Figures 5.2 to 5.4 further investigate the relationship between misfits after a 500-year and 3000-year spin-up for LHS1-3 respectively, by highlighting the non-linearities according to their ocean biogeochemical parameter values. The non-linear tail in the LHS1 and LHS3 misfit relationships were strongly associated with the parameters b* and I_C, with the former seeming to have more of an influence. Generally LHS

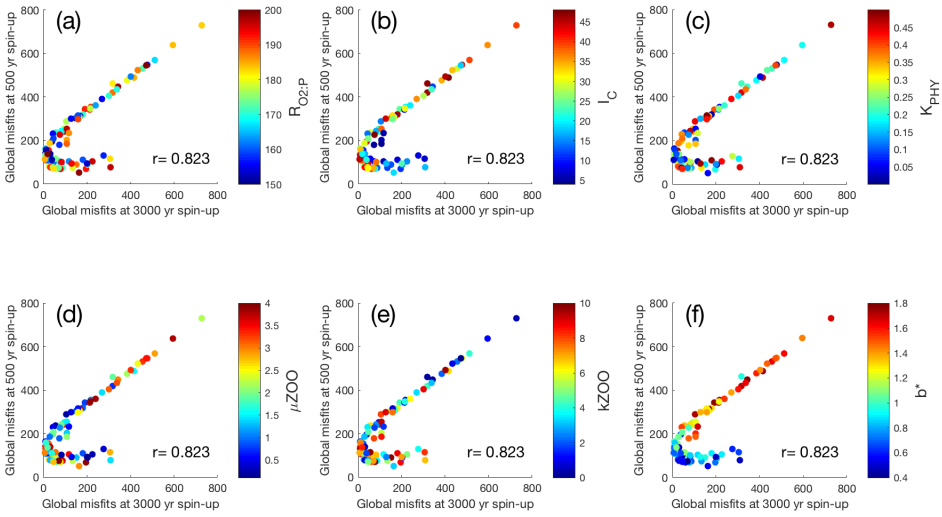

Figure 5.2: Global misfit values by the 90 MOPS latin hypercube points LHS1 initiated with uniform tracer distributions, calculated after a spin-up length of 3000 years on the x-axes and 500 years on the y-axes. The 90 misfits are coloured according to parameter values of (a) $R_{O_2:P}$, (b) I_C, (c) K_{PHY}, (d) μ_{ZOO}, (e) k_{ZOO} and (f) b*. The correlation coefficients (r) are given for each scatter relationship.

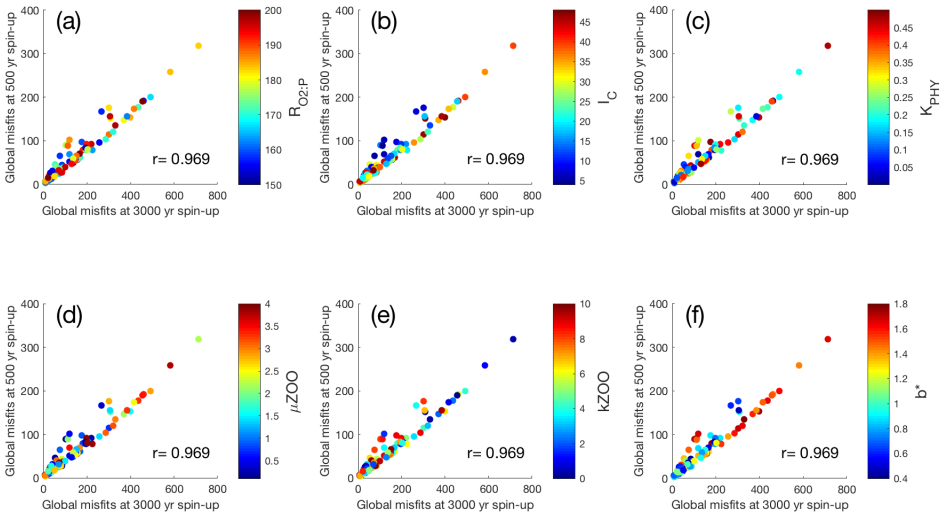

Figure 5.3: As in Figure 5.2, but for the 90 MOPS latin hypercube points LHS2 initiated with target twin observations output by MOPS-Ref.

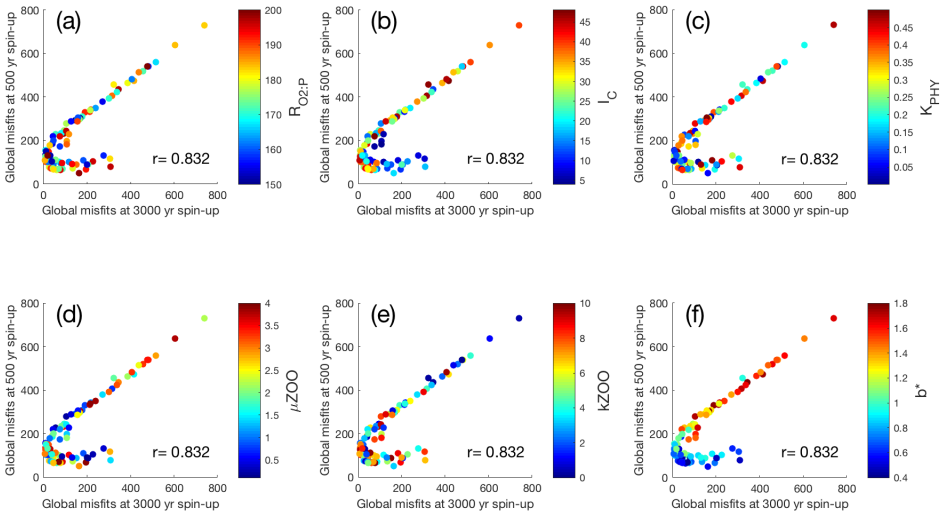

Figure 5.4: As in Figure 5.2, but for the 90 MOPS latin hypercube points LHS3 initiated with equilibrated tracer distributions output by MOPS-Worst.

samples with b* values less than approximately 0.9 fell on the non-linear tail (note that a value of 0.858 is widely used in ocean biogeochemical models, and currently used in MOPS), with the end of the tail furthest from the y-intercept containing LHS samples also with I_C values less than approximately 20 Wm^{-2}. The linear section of the misfit relationship corresponded to values of b* greater than 1, where the misfits increased with b*. For LHS2, where MOPS was initiated from target tracer distributions, there was a stronger linear relationship than LHS1 and LHS3, though again the misfits increased with increasing values of b*. There were two areas of scatter above the linear relationship, one corresponding to low values of both b* and I_C, and the other (closer to the y-intercept) to high values of b* and low values of I_C. The remaining 4 parameters did not seem to significantly influence the misfit relationships.

The 500-year to 3000-year spin-up misfit relationships were further investigated by separating the global misfits into the three tracer misfits, which were the sum of the squared regional individual misfits for each tracer phosphate, oxygen and nitrate. Figure 5.5 shows the tracer specific LHS results corresponding to values of b*, and Figure 5.6 to values of I_C. For all LHS1-3 the oxygen misfits contribute

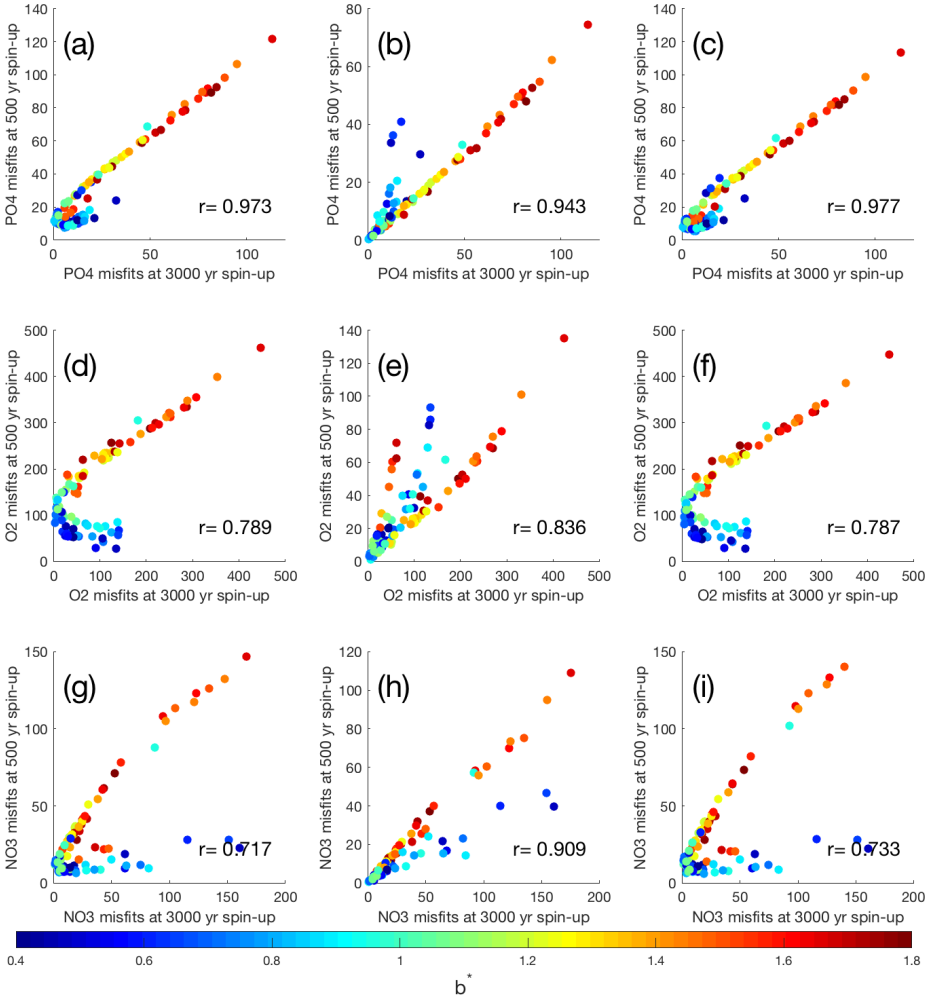

Figure 5.5: Individual tracer misfits for each of the 90 MOPS latin hypercube sets (left subplot column) LHS1 initiated with uniform tracer distributions, (middle subplot column) LHS2 initiated with target twin observations output by MOPS-Ref, and (right subplot column) LHS3 initiated with equilibrated tracer distributions output by MOPS-Worst. These were calculated after a spin-up length of 3000 years on the x-axes and 500 years on the y-axes, for (a-c) PO_4, (d-f) O_2 and (g-i) NO_3. The 90 misfits are all coloured according to b* parameter values. The correlation coefficients (r) are given for each scatter relationship.

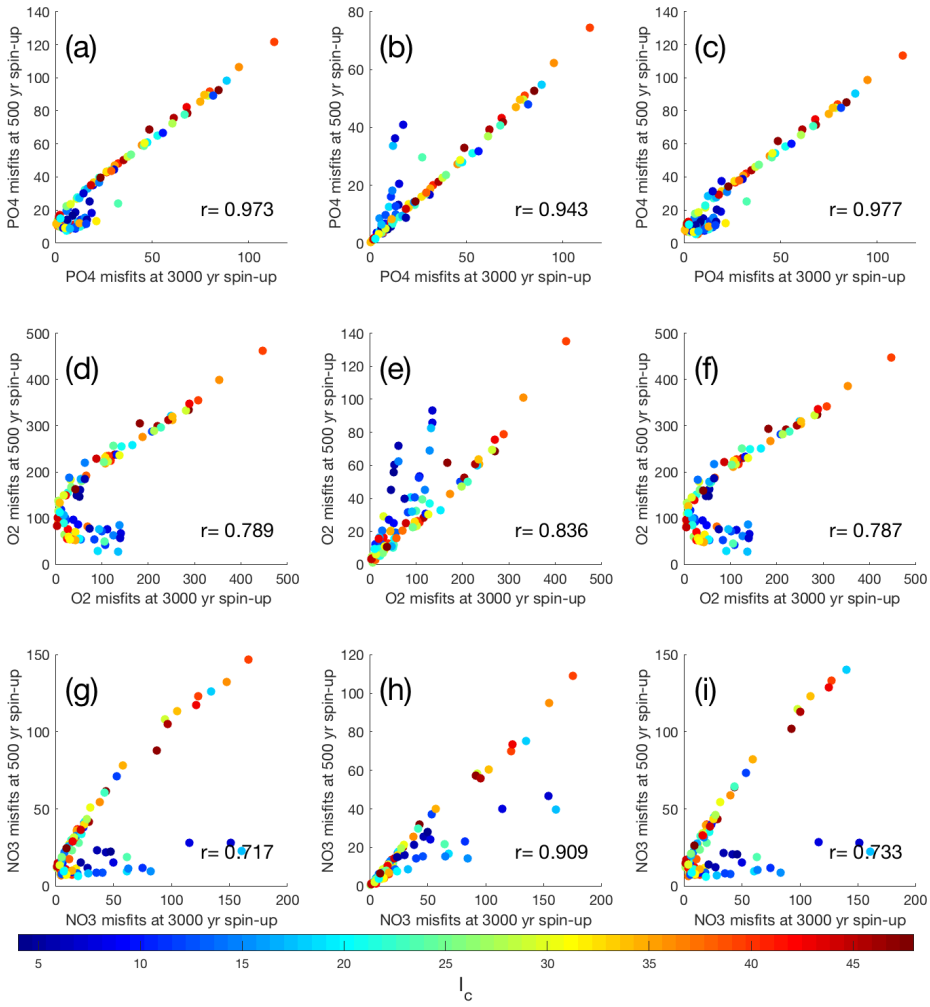

Figure 5.6: As in Figure 5.5, but coloured according to I_C parameter values.

the most to the global misfit, therefore the shape of the misfit relationships seen in Figure 5.1 are most similar to the oxygen misfit relationships. The phosphate misfit relationships for all LHS1-3 had the strongest linear relationship of the three tracers. The shape of the phosphate and oxygen LHS2 relationships greatly differed from the LHS1 and LHS3 relationships as most non-linear scatter fell above the linear fit for LHS2, and below for LHS1 and LHS3. However, all three showed similar nitrate relationships. The shape of the nitrate relationships differed from the non-linear tail previously seen, as instead of a straight linear line with a curved non-linear tail below the y-intercept, the nitrate relationship shape consisted of two separate linear branches, with the lower branch corresponding to low values of b* and I_C.

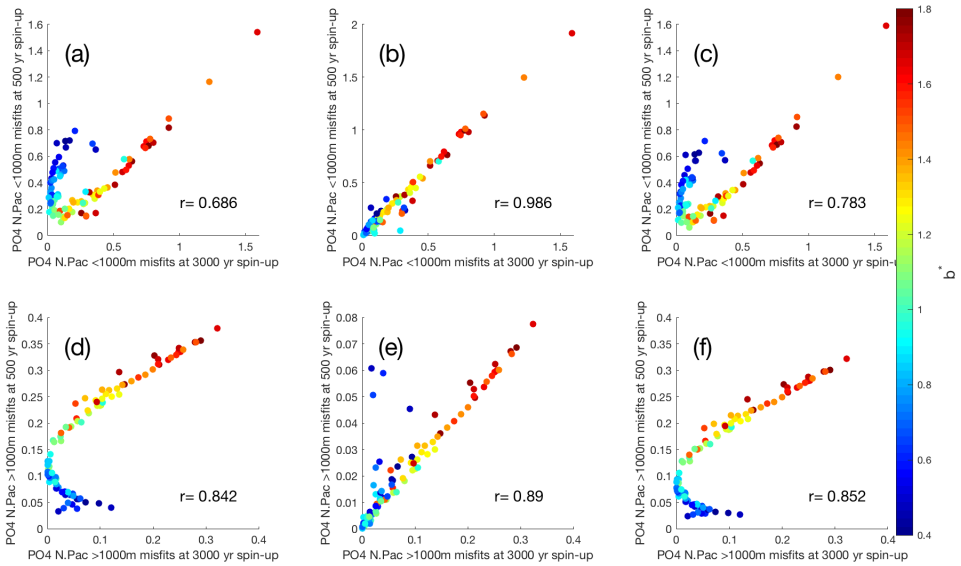

Figure 5.7: Regional phosphate misfit values for each of the 90 MOPS latin hypercube sets (left subplot column) LHS1 initiated with uniform tracer distributions, (middle subplot column) LHS2 initiated with target twin observations output by MOPS-Ref, and (right subplot column) LHS3 initiated with equilibrated tracer distributions output by MOPS-Worst. These were calculated after a spin-up length of 3000 years on the x-axes and 500 years on the y-axes, specifically of the PO_4 misfit within the (a-c) surface <1000m, and (d-f) deep >1000m North Pacific. The 90 misfits are coloured according to b* parameter values. The correlation coefficients (r) are given for each scatter relationship.

Figure 5.7 shows the last of the misfit relationships results, isolated for the North Pacific's individual phosphate misfits where there was a particularly strong opposing misfit relationship between the upper and lower water column. For the deep North Pacific (>1000m), the misfit relationships within LHS1 and LHS3 were similar to previously seen in their respective global misfit relationships, with a non-linear tail below the y-intercept. However, this was found to be very different for the shallower (<1000m) North Pacific, as for LHS1 and LHS3 the non-linear tail flipped to lie above the y-intercept instead. Again, most which did not fall on the linear fit corresponded to low values of b*.

5.3.2 Equilibrated vs non-equilibrated misfit timeseries

To better understand the non-linear tail seen in the LHS1 and LHS3 misfit relationships, time series of the misfits throughout a 3000-year spin-up of MOPS-

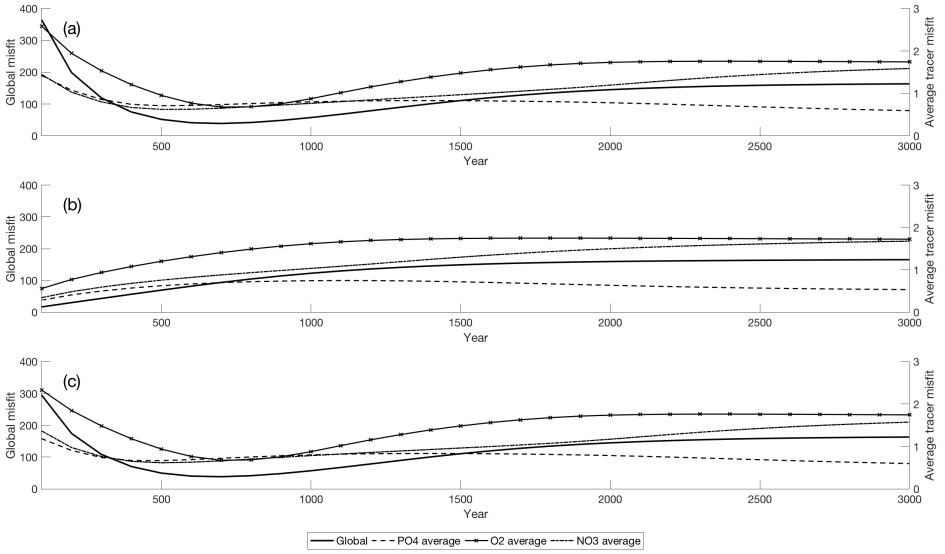

Figure 5.8: The misfit during a 3000 year spin-up for the MOPS run with MOPS-NonLinear parameter values, initiated from (a) uniform tracer distributions, (b) target twin observations output by MOPS-Ref, and (c) tracers output by MOPS-Worst. These MOPS runs are highlighted as a red diamond in the (a) left, (b) middle, and (c) right subplot column of Figure 5.1. Shown are the global misfit (black solid line) on the left y-axis, and the averaged individual misfits of PO_4 (black coarsely-dashed line), O_2 (black solid line with crosses) and NO_3 (black finely-dashed line) on the right y-axis.

NonLinear initiated from the 3 different initial tracer distributions have been shown in Figure 5.8. The time series were similar when initiated from uniform or equilibrated MOPS-Worst tracer distributions. For both, the large initial global misfit rapidly decreased with spin-up time until a minimum at 600-800 years, where it then gradually increased until a steady plateau after approximately 2000 years. The dominating oxygen misfit matched the global misfit time series well, while after a minimum at 700-800 years the nitrate misfit did not stop increasing for the remaining spin-up, and phosphate after a 500-year minimum rose slightly until 1500 years then gradually decreased for the remaining spin-up. However for LHS2 all misfits showed changes early on in the spin-up (within 1000 years) before remaining relatively uniform. In every case, nitrate had not reached an equilibrated plateau before the 3000 years.

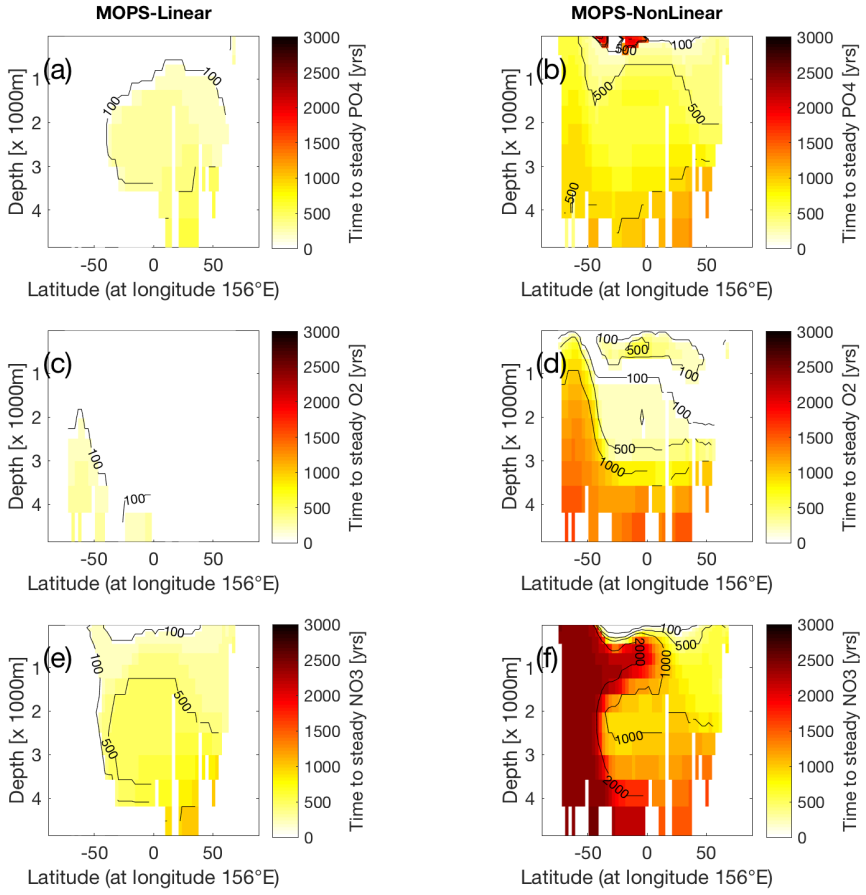

Figure 5.9: Length of spin-up needed to reach and remain within +-10% of the final 3000 year tracer values, indicated by both contour lines and colour shading. Plotted are the results along the Atlantic Ocean longitudinal transect (156°E) for MOPS initiated from equilibrated MOPS-Worst tracer distributions with the parameter configurations as in (left column) MOPS-Linear, and (right column) MOPS-NonLinear, for the tracers (a-b) phosphate, (c-d) oxygen and (e-f) nitrate.

5.3.3 Oceanic response time to certain parameter configurations

A second latin hypercube sample from LHS3 was chosen to further investigate the difference between a MOPS evaluation which lay on the linear section of the misfit relationships (MOPS-Linear), and one which lay on the non-linear tail (MOPS-NonLinear). The time taken for both MOPS spin-ups to reach and remain close to (specifically within +-10% of) the final 3000 year tracer average are shown for the Atlantic and Pacific Oceans in Figures 5.9 and 5.10 respectively.

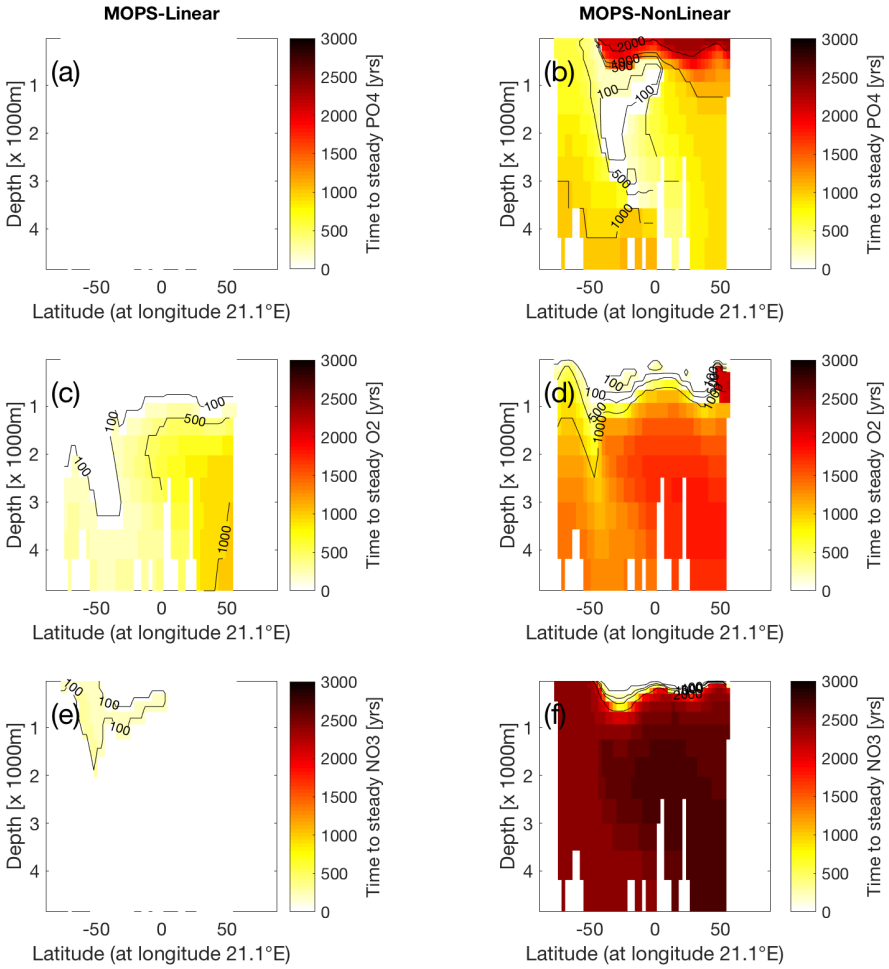

Figure 5.10: As in Figure 5.9 but for the Pacific Ocean longitudinal transect (21.1°E).

In both oceans, MOPS took less time to equilibrate all tracers when configured with MOPS-Linear biogeochemical parameters (relatively high I_C and b*) than with MOPS-NonLinear parameters (low b*). In both the Atlantic and Pacific transects, MOPS-Linear equilibrated all tracers within close to 500 years, with the exception of nitrate in the deep Equatorial Atlantic and oxygen in the deep North Pacific, which took up to 1000 years. MOPS-NonLinear was very different, with all tracers taking longer to equilibrate. In the surface South Atlantic and surface Pacific, phosphate took over 2000 years to equilibrate, while elsewhere it took approximately 500 years. Oxygen and nitrate showed the same spatial patterns for time to tracer equilibrium, though with nitrate generally taking longer. In the Atlantic, time to equilibrate

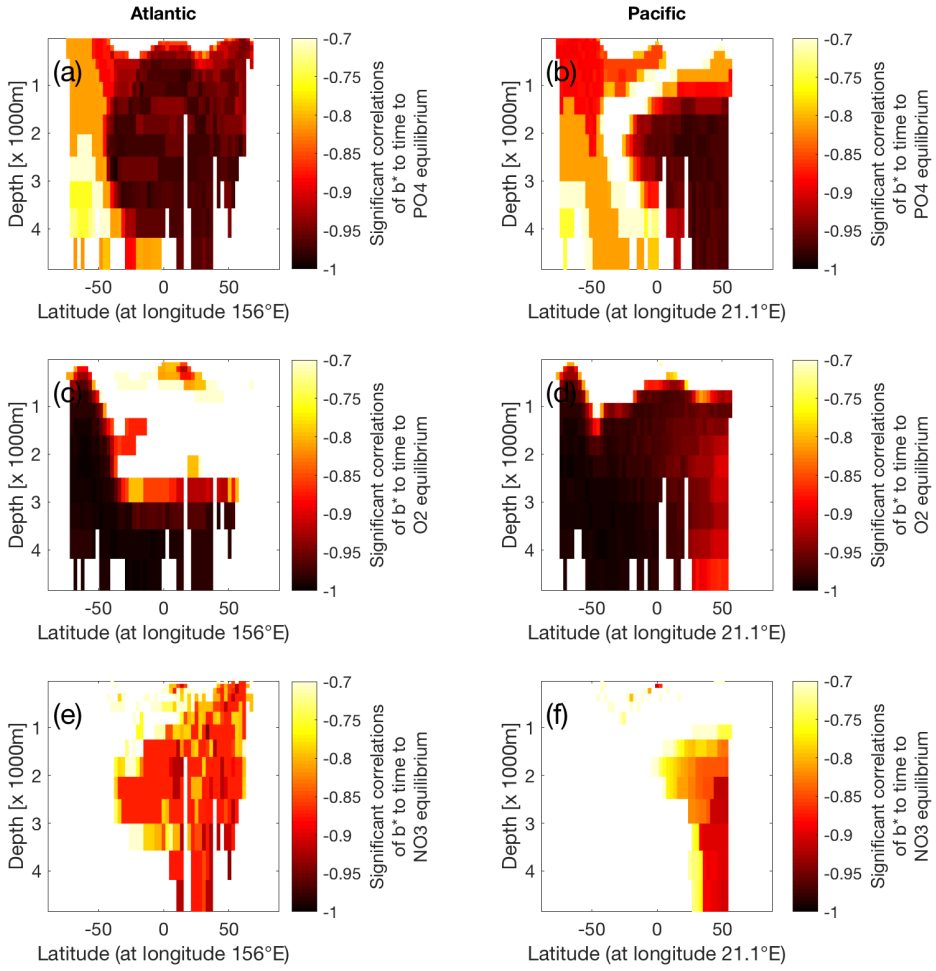

Figure 5.11: Significant correlation coefficients between b* parameter values and length of spin-up needed to reach and remain within +-10% of the final 3000 year tracer values. Plotted are the correlations along the (left column) Atlantic Ocean longitudinal transect at 156°E, and (right column) Pacific Ocean longitudinal transect at 21.1°E, for the tracers (a-b) phosphate, (c-d) oxygen and (e-f) nitrate. Note that MOPS was initiated from equilibrated MOPS-Worst tracer distributions.

both tracers increased with decreasing latitude and increasing depth, taking over 1000 and 2000 years respectively to equilibrate in the Antarctic Bottom Water for oxygen and nitrate respectively. Nitrate also took over 2000 years to equilibrate in the Antarctic Intermediate Water. In the Pacific, oxygen and nitrate equilibration time increased with depth and latitude toward the northern deep Pacific, reaching over 1000 and 2000 years respectively.

These previous plots show the influence of varying multiple biogeochemical values on

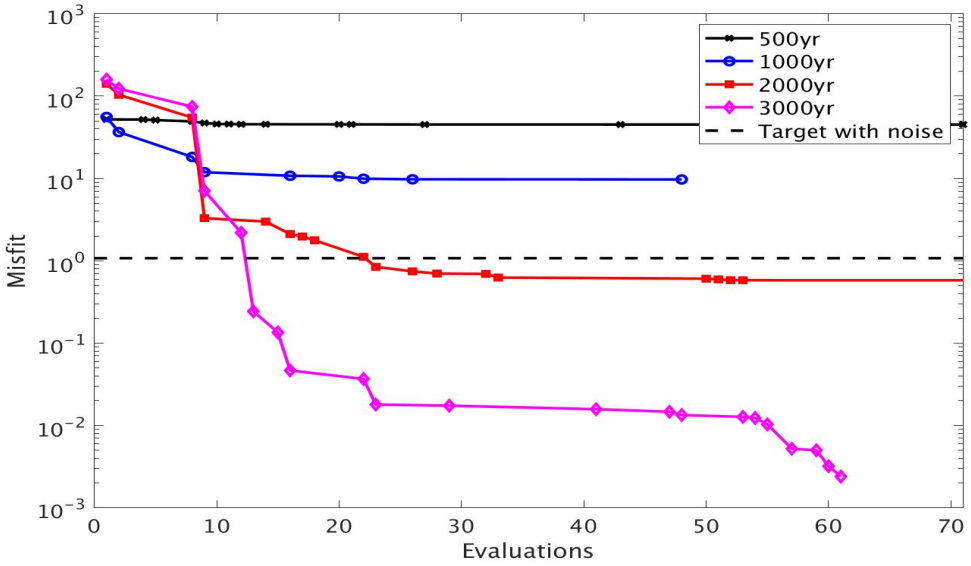

Figure 5.12: DFO-LS misfit reduction for the experiments with various MOPS spin-up lengths **t6d_un_500** (black line with crosses), **t6d_un_1000** (blue line with circles), **t6d_un_2000** (red line with squares) and **t6d_un_3000** (magenta line with diamonds). Target with noise (horizontal black dashed line) indicates the baseline misfit, below which any reduction in misfit is within observational noise. There were no restarts triggered in these experiments. Note that all experiments completed 70 MOPS evaluations, however for clarity only evaluations which decreased the misfit have been plotted.

equilibration time, particularly b* and I_C. However to understand the influence of b* in isolation, MOPS was run 10 more times with MOPS-Ref parameter values for all but b* which was sampled equally spaced between its bounds, and correlations were calculated between values of b* and time taken to reach equilibrium (see Figure 5.11). In both oceans, there were significant negative correlations between b* and time to equilibrium for all three tracers. Oxygen showed the most widespread and strongest correlations near -1, particularly in the Southern Ocean, deep Atlantic below 2500m and deep Pacific. Second was phosphate with strong correlations near -1 in the Equatorial and North Atlantic and deep North Pacific. Lastly was nitrate, with significant correlations near -0.85 in a similar pattern to phosphate, only with weaker correlations.

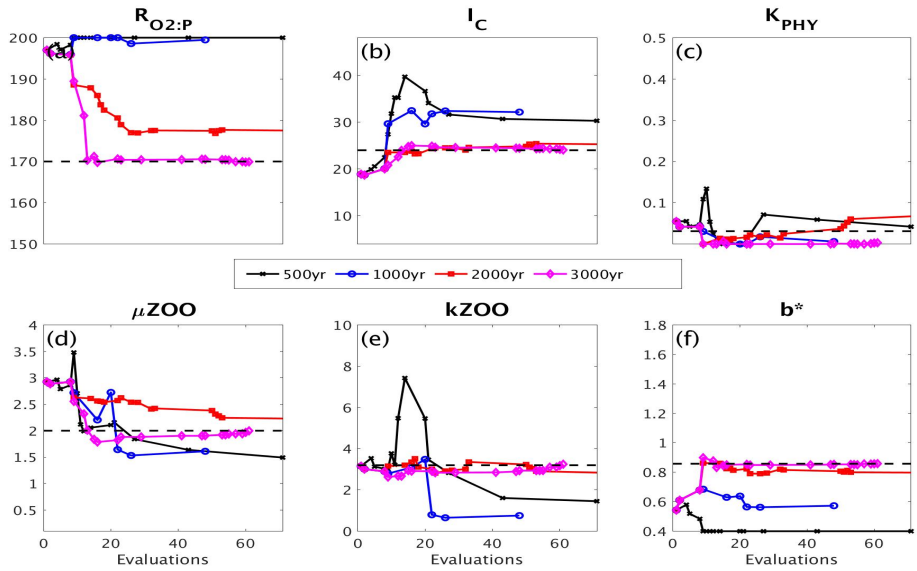

Figure 5.13: Parameter tuning by DFO-LS for the experiments where MOPS was initiated with uniform conditions: **t6d_un_500** (black line with crosses), **t6d_un_1000** (blue line with circles), **t6d_un_2000** (red line with squares) and **t6d_un_3000** (magenta line with diamonds), for the parameters (a) $R_{O_2:P}$, (b) I_C, (c) K_{PHY}, (d) μ_{ZOO}, (e) k_{ZOO} and (f) b*. Also shown are the MOPS-Ref target parameter values (horizontal black dashed line). Only evaluations which decreased the misfit have been plotted.

5.3.4 Equilibrated vs non-equilibrated optimisation of shorter spin-ups

DFO-LS attempted to optimise the 6 biogeochemical parameters of MOPS with various spin-up lengths, always calibrating towards the 3000-year MOPS-Ref twin observations. This was repeated for the three different initial tracer distributions. Figures 5.12 - 5.13 show the results of optimising MOPS initiated from uniform tracer distributions. The amount of global misfit reduction by DFO-LS shown in Figure 5.12 increased with MOPS spin-up length, therefore so did the optimiser's success. In experiment **t6d_un_500** DFO-LS could barely reduce the misfit of the 500 year spin-up within 70 evaluations, and only managed to reduce the 1000 year misfit to approximately 10 in **t6d_un_1000**. **t6d_un_2000** was significantly more successful, as the misfit was reduced to below the baseline misfit due to observational noise. The 2000 year misfit was rapidly reduced from above 100 to 3.29 within 9 MOPS evaluations, then more slowly to 0.62 by 33 evaluations.

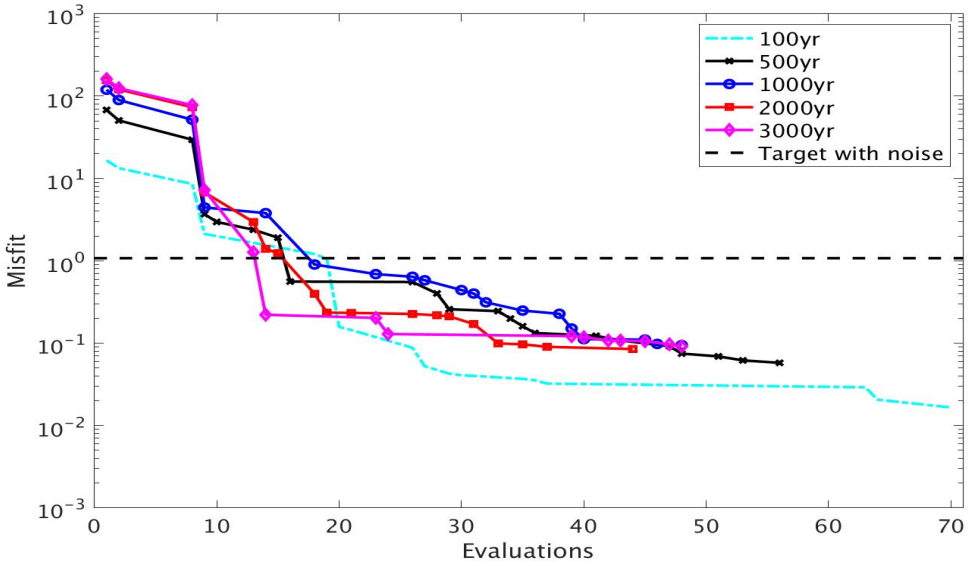

Figure 5.14: Misfit reduction as in Figure 5.12 but with MOPS initiated from equilibrated target tracer distributions output by MOPS-Ref, and including an extra experiment with a 100 year spin-up **t6d_eqR_100** (cyan dashed line).

t6d_un_3000 was the most successful with the largest misfit reduction, initially from above 100 to 0.047 after 16 evaluations, then slow progress to 0.01 after 55 evaluations before more rapid misfit decrease to 0.0024 after 61 evaluations.

Figure 5.13 shows how the 6 parameters were optimised towards the MOPS-Ref target parameter values during these 4 experiments. The success of **t6d_un_3000** previously seen in the misfit reduction is explained as all parameters but one (K_{PHY}) were recovered well, many within only 20 evaluations. DFO-LS managed to recover b*, I_C and k_{ZOO} in **t6d_un_2000** while in the shorter spin-up experiments failed to recover any, sometimes getting trapped on a parameter bound. How well the parameters were recovered by the shorter spin-up experiments generally reduced with spin-up length, with this statement being most to least true (in decreasing order) for b*, $R_{O_2:P}$, I_C, k_{ZOO} and μ_{ZOO}. K_{PHY} did not seem to be successfully optimised by any experiment (see Chapters 3 and 4).

Figures 5.14 - 5.15 show the results of optimising MOPS initiated from target equilibrated MOPS-Ref distributions. Unlike the results for uniform initial tracer distributions, DFO-LS greatly reduced the global misfit below the baseline misfit for

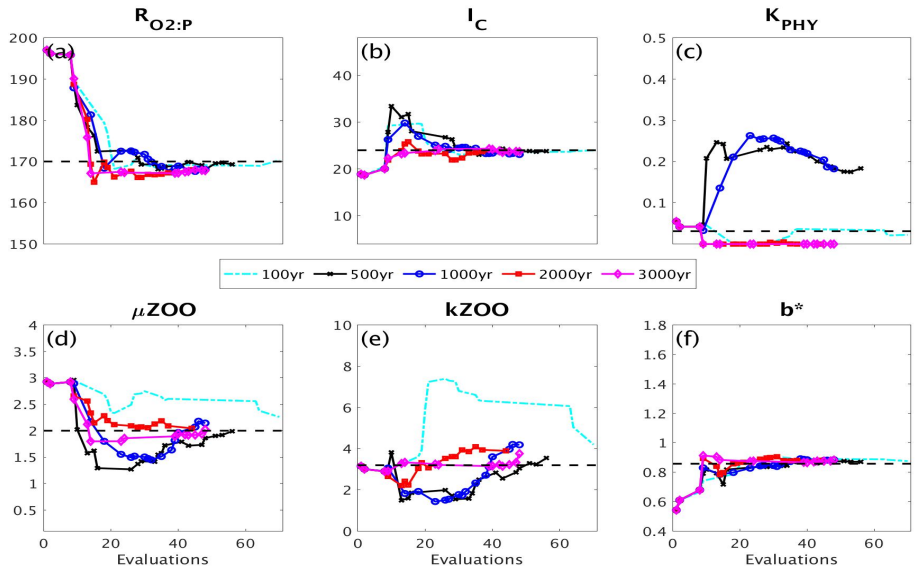

Figure 5.15: Parameter optimisation as in Figure 5.13 but with MOPS initiated from equilibrated target tracer distributions output by MOPS-Ref, and including an extra experiment with a 100 year spin-up **t6d_eqR_100** (cyan dashed line).

all experiments to below 0.1, with the shortest spin-up **t6d_eqR_100** performing best as it reduced the misfit below 0.1 after 26 evaluations and reached 0.017 after 70 evaluations. Figure 5.15 shows that when initiating MOPS from target tracer distributions DFO-LS recovered b*, $R_{O_2:P}$ and I_C quickly very well in all experiments, and μ_{ZOO} and k_{ZOO} more slowly by relatively well by all but the 100 year spin-up. For the two zooplankton parameters **t6d_eqR_100** was slow to optimise until approximately 63 evaluations when DFO-LS suddenly started to optimise them towards their targets, and would possibly have fully recovered them if DFO-LS were allowed to continue. **t6d_eqR_100** was the only one to remain close to the K_{PHY} target.

Figures 5.16 - 5.17 show the results of optimising MOPS initiated from equilibrated MOPS-Worst distributions, which look extremely similar to the optimisation results when MOPS was initiated from uniform tracer distributions in Figures 5.12 - 5.13.

5.3.5 Annual vs monthly optimisation

Figures 5.18 - 5.19 show the twin optimisation experiments where DFO-LS attempted to optimise 5 biogeochemical parameters using either annual or monthly misfits,

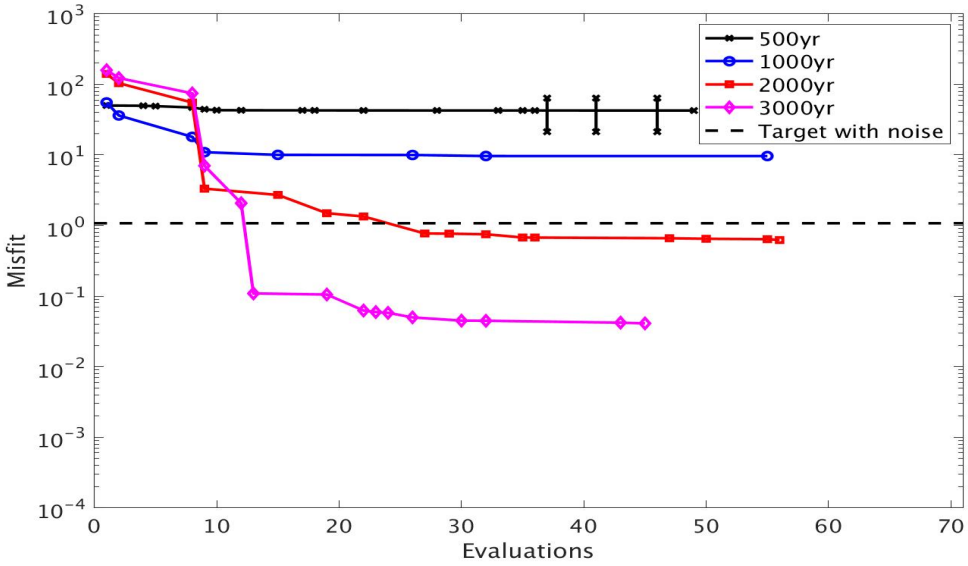

Figure 5.16: Misfit reduction as in Figure 5.12 but with MOPS initiated from equilibrated tracer distributions output by MOPS-Worst. Vertical lines indicate a soft restart.

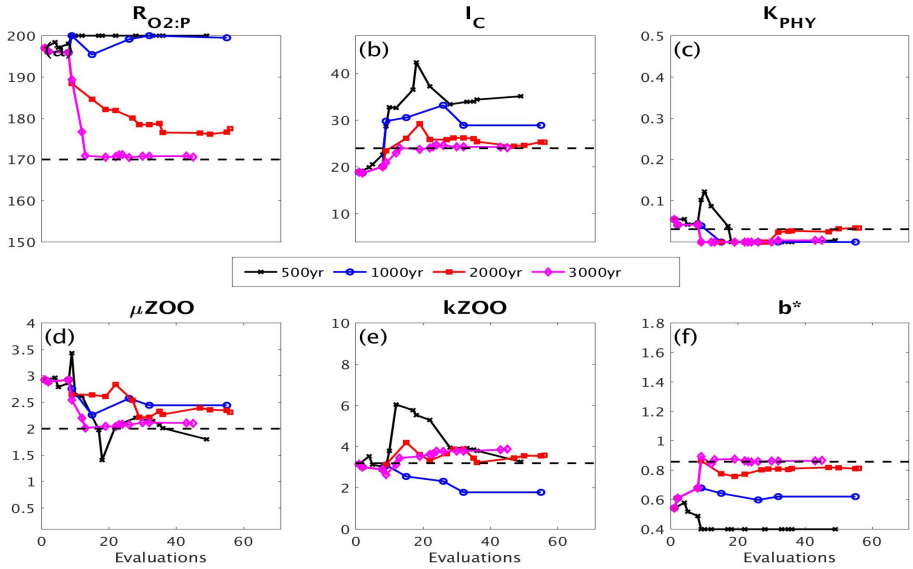

Figure 5.17: Parameter optimisation as in Figure 5.13 but with MOPS initiated from equilibrated tracer distributions output by MOPS-Worst.

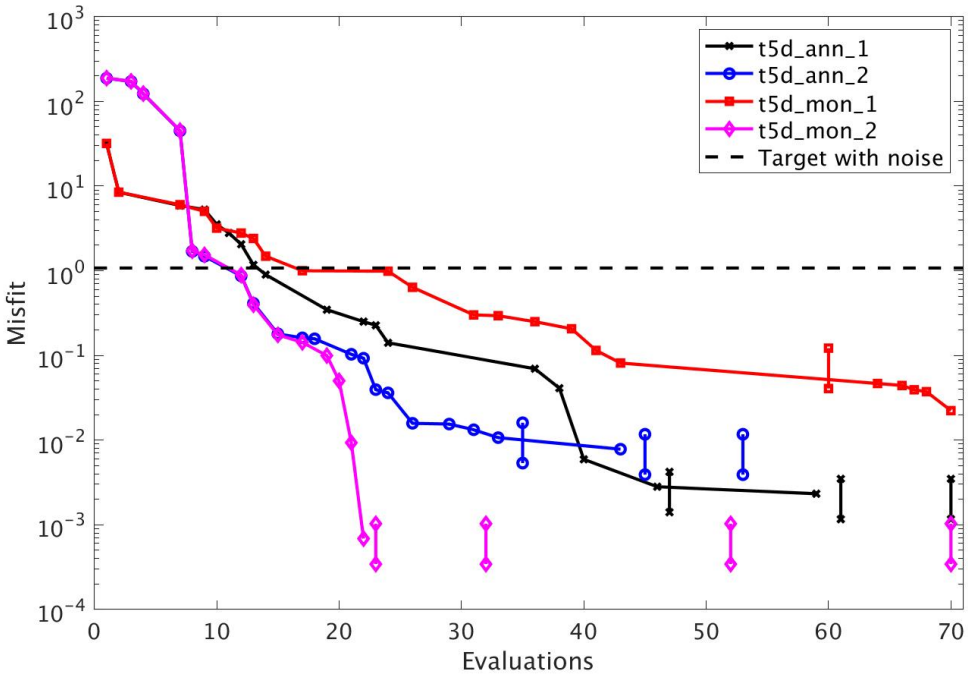

Figure 5.18: The reduction in annual and monthly global misfits by DFO-LS from two different starting points, for the experiments **t5d_ann_1** (black line with crosses), **t5d_ann_2** (blue line with circles), **t5d_mon_1** (red line with squares) and **t5d_mon_2** (magenta line with diamonds). Target with noise (horizontal black dashed line) indicates the baseline misfit, below which any reduction in misfit is within observational noise. Vertical lines indicate a soft restart. Only evaluations which decreased the misfit have been plotted.

starting from two different points in the parameter space (location 1 and 2). Note that for these MOPS was always initiated with uniform tracer distributions and spun up for 3000 years before the misfit was calculated. When starting from location 1, DFO-LS managed to significantly decrease both the annual and monthly misfits, reducing the annual misfit the most to 0.0023 after 59 evaluations, and the equivalent monthly misfit to 0.022 in 70 evaluations. However, when starting from location 2, DFO-LS was more successful with the monthly misfit, reducing it to 0.00069 within 22 evaluations, and the annual misfit to 0.0078 by 43 evaluations. Figure 5.19 shows all experiments recovered $R_{O_2:P}$, I_C and b* almost equally well but there were differences in how well the remaining two zooplankton-related parameters were recovered. The order of which experiment reached the lowest global misfit coincided with which recovered these zooplankton parameters the best, which was (from best

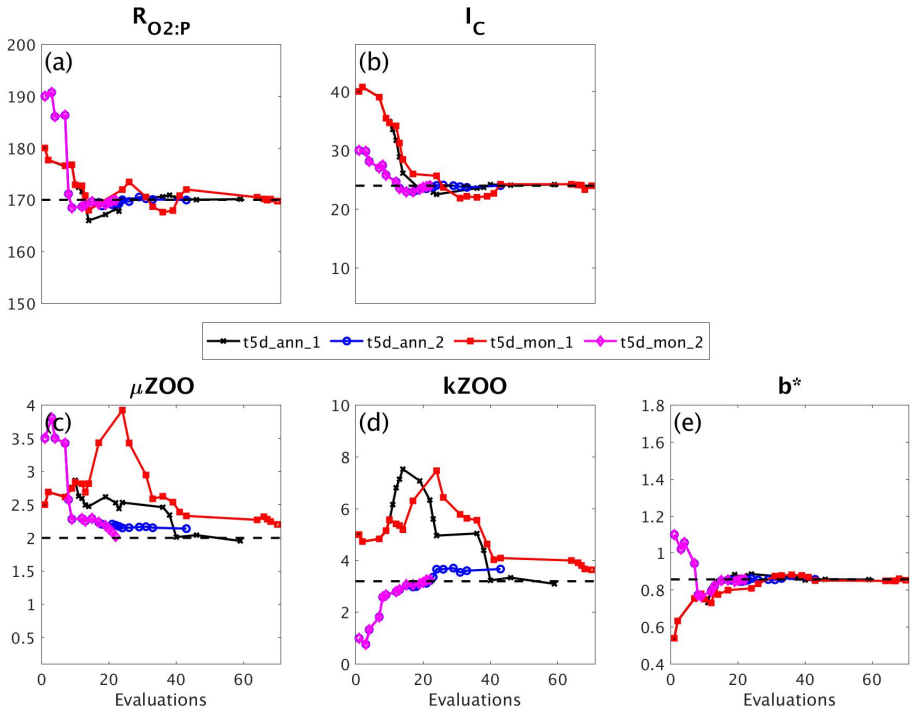

Figure 5.19: Parameter tuning by DFO-LS for the experiments **t5d_ann_1** (black line with crosses), **t5d_ann_2** (blue line with circles), **t5d_mon_1** (red line with squares) and **t5d_mon_2** (magenta line with diamonds) for the parameters (a) $R_{O_2:P}$, (b) I_C, (c) K_{PHY}, (d) μ_{ZOO}, (e) k_{ZOO} and (f) b*. Also shown are the MOPS-Ref target parameter values (horizontal black dashed line). Only evaluations which decreased the misfit have been plotted.

to worst): **t5d_mon_2**, **t5d_ann_1**, **t5d_ann_2**, and **t5d_mon_1**.

5.4 Discussion

5.4.1 Can we optimise shorter spin-ups?

When comparing the misfits after a 3000 year spin-up and misfits after shorter spin-ups, we found they do not always relate in a linear, predictable way, and depend on the biogeochemical parameters and model initial conditions. This unfortunately means that a global optimum in the misfit function after a short spin-up is not guaranteed to be at the same point of the parameter space as the global optimum after an equilibrated misfit, therefore optimisation of a short spin-up may not produce the correct optimal parameters. This is because a short spin-up does not give the model long enough for the ocean to feel the effects of running the

model with different parameter values, especially in the oceanic regions which are further along the "great ocean conveyor" (Broecker, 1991; Lumpkin and Speer, 2007). Perturbations take time to propagate further throughout the global ocean (Wunsch and Heimbach, 2008; Primeau and Deleersnijder, 2009; Siberlin and Wunsch, 2011; Yang and Zhu, 2011; Khatiwala et al., 2012), taking longest for the Pacific (Skinner and Shackleton, 2005; Shah et al., 2017). Therefore, the longer the spin-up to calculate the misfit, the more likely it is to correctly represent the equilibrated misfit, with a minimum of a 2000 year spin-up to allow for successful optimisation in our specific case.

However, there is no universal minimum spin-up length to ensure sufficient representation of the equilibrated misfit, because this is dependent on such factors as the chosen general circulation, misfit calculation, biogeochemical parameters, and initial tracer distributions. If the biogeochemical model is coupled with a GCM with faster high-latitude deep-water formation rates, any perturbations in the model configuration may propagate quicker through the global ocean (e.g. Broecker et al., 1999; Gnanadesikan et al., 2004; Primeau and Holzer, 2006; Marinov et al., 2008), and will require a shorter spin-up to reach a sufficiently equilibrated misfit. The chosen misfit function also strongly influences the spin-up necessary to reach an equilibrated misfit (Evans, 2003), because it may take into account tracers which are slower to equilibrate, such as nitrate. Additionally, it may only focus on the surface ocean which generally equilibrates quickly, whereas a misfit heavily weighted to the deeper ocean (as in our case) requires a longer spin-up to propagate the misfit to these depths (Wunsch and Heimbach, 2008; Khatiwala et al., 2012). Discussed in the following sections are the final two influential factors; the chosen biogeochemical parameters and initial tracer distributions.

5.4.2 How biogeochemical parameters influence time to equilibrium

An important finding of this chapter is that in our specific case of MOPS and 6 chosen biogeochemical parameters, the relationship between equilibrated misfits

and those calculated after a shortened spin-up are linear only for parts of parameter space. It was non-linear for the regions of parameter space with low values of the Martin exponent, b*, and sometimes low values of the half saturation constant for light, I_C, due to how these parameters influence how long the model needs to reach equilibrium. This may also have been true for the other parameters, but was possibly overshadowed by the influence of b* and I_C as the defined misfit function was most sensitive to these parameters.

The parameter b* influences the remineralisation depth of falling organic matter (Martin et al., 1987). A high b* causes the model to simulate slow sinking of organic matter, allowing remineralisation to occur predominantly in the surface ocean. Conversely, a low b* causes rapid sinking particles, allowing the sinking organic matter to be transported to the deep ocean before full remineralisation (Kwon and Primeau, 2006; Kwon et al., 2009). The latter case results in greater transport of organic matter and nutrients to the deep ocean conveyor belt, hence its influence is far-reaching (Kriest and Oschlies, 2013). During a low b* scenario any misfit signal which reaches the deep conveyor belt will take longer to reach equilibrium, compared with a high b* scenario where most remineralisation occurs within the surface and the misfit signal is not transported as far.

The value chosen for the parameter I_C directly relates to the amount of light necessary for the phytoplankton within the model to photosynthesise at their maximum rate. Therefore, a lower I_C results in phytoplankton which need less light to be more productive, therefore allowing them to be more productive at higher latitudes where there is less light. This could influence the flux of organic matter and nutrients to the deeper ocean by increasing the productivity of phytoplankton in regions of downwelling at high latitudes, hence this may be the cause as to why the misfit signal takes longer to equilibrate within the model. Also, the amount of nutrients leaving the large areas of upwelling in the Southern Ocean would be influenced by this I_C parameter, and hence influence the time required to propagate these nutrients globally.

5.4.3 The importance of initial conditions

The outcome of optimising the parameters within a short model spin-up is very different if the model is initialised with target tracer distributions, or non-target tracer distributions. Here the term "target" refers to the observational term within the misfit calculation, which is the data one decides to calibrate the model to. In a twin experiment this is the synthetic observations, and in a real experiment this is the real oceanic observations, provided by sources such as World Ocean Atlas (e.g. Garcia et al., 2018a,b). Non-target tracer distributions are anything else, such as either uniform distributions or equilibrated distributions output by a previous model run.

In this chapter, we found optimising MOPS initiated from target tracer distributions was successful at much shorter spin-ups than when initiated from non-target tracer distributions. This is a good result, as it indicates global ocean biogeochemical models can be optimised successfully when initiated from target tracer distributions. However, some further investigation uncovered the false sense of security this finding gave, which one will need to be aware of before utilising this short cut. After a short model spin-up (e.g. 100 years) the model has not had time to influence much of the initial ocean state and is restricted to the areas of the ocean which are fast-responding, such as the surface ocean and regions of downwelling (Wunsch and Heimbach, 2008). Therefore, the tuning process only tunes the surface ocean, and ignores the deeper ocean where the misfit signal does not reach within the short spin-up time.

In our twin experiments we found this restricted surface-tuning effect was enough to recover the reference optimal parameter configuration, only when the initial conditions were the target tracer distributions. For this case it did not matter that the misfit signal did not reach the deeper ocean during the short spin-up, because the deeper ocean did not need to be tuned as it was already at it's target state. Whereas in the case where MOPS was initialised from non-target tracer distributions, the misfit signal must reach the deeper ocean for it to be tuned as well as the surface, therefore a longer spin-up is needed. Therefore, when calibrating to

the real ocean instead of synthetic observations, potentially one could successfully optimise their global ocean biogeochemical model if the model is initiated from a good estimate of true biogeochemical distributions. This has practical implications for optimisation to real data when, as is frequently done, one might initialise from observed tracer distributions so as to get away with a shorter spin-up. However, the fact that only the surface ocean will been tuned during optimisation of a shortened spin up should be kept in mind.

5.4.4 Does optimising the seasonal cycle encourage faster calibration?

We tested to see whether it was possible to increase the speed of optimisation progress (hence also reduce computational cost and time) by including more data into the misfit function, by calibrating to monthly data instead of annual. Unfortunately, the results of 4 twin experiments showed that improved optimisation progress is not guaranteed by providing monthly data, therefore optimising the seasonal cycle of the biogeochemical model is not a reliable method to improve optimisation efficiency. This may have been due to the fact that the monthly data was heavily weighted to the deeper ocean, therefore the seasonal cycle was not strongly represented, and may have been too similar to the annual mean to have made much difference to the twin experiments using only annually-averaged data. A possibility for future study would be to optimise to the monthly data while weighting the misfit to the surface ocean (Evans, 2003), to ensure the seasonal cycle in the upper water column influences the misfit function, without ignoring the deeper ocean.

5.5 Conclusions

The main purpose of this chapter was to test methods to reduce the computational cost and time of ocean biogeochemical model optimisation. Due to the long model spin-up time required for the model to reach an equilibrated state after each parameter re-configuration (Wunsch and Heimbach, 2008; Khatiwala et al., 2012), ocean biogeochemical parameter optimisation is a notoriously laborious and

therefore often subjective step in model development (Hourdin et al., 2017). The methods tested here in an attempt to reduce the cost of this step were optimisation within the global ocean biogeochemical model MOPS (Kriest and Oschlies, 2015) with a shortened model spin-up, and optimisation using monthly observational data. It was found that successful optimisation of MOPS was not possible with a spin-up length of less than 2000 years, due to the slow misfit response throughout the deeper ocean. This was due to optimising parameters which caused the model to have increased vertical transport of organic matter, and hence increased remineralised nutrients to the deep global oceanic conveyor belt (Broecker, 1991; Kwon and Primeau, 2006; Kwon et al., 2009). This problem was masked when initiating the model with target tracer distributions, which are the observations used for calibration in the misfit function, because spin-up times shorter than 2000 years did not largely deviate the deeper ocean from it's target optimised state. Therefore, we advise caution to all who utilise shortened ocean biogeochemical model spin-ups (e.g. Priess et al., 2013b; Séférian et al., 2016). It was also found that increasing the information given to the optimisation algorithm in the form of monthly observational data instead of annual did not reliably increase the speed of successful optimisation. Sadly, neither tested method was a reliable or robust way to reduce the expense of global ocean biogeochemical optimisation. One possible conclusion is that oceanic spin-up times could be reduced to 2000 years instead of the full suggested 3000 years to equilibrium. However this must be carefully considered on a case-by-case basis as this 2000 year suggestion is dependent on both the general circulation and ocean biogeochemical models, as well as the chosen initial conditions, misfit function and parameters to be tuned. This chapter has hopefully satisfied these two lines of enquiry, allowing investigation of other methods to continue to strive for more efficient, systematic global ocean biogeochemical optimisation.

5.6 Code Availability

The base TMM and MOPS code used for the ocean biogeochemical simulations are available to download from `https://doi.org/10.5281/zenodo.1246300`. Trans-

port matrices and forcing fields required to perform the simulations can be downloaded from `https://doi.org/10.5281/zenodo.5517238`. The OptClim optimisation framework used in this study to couple any climate model to any optimiser is available at `https://doi.org/10.5281/zenodo.5517610`.

6

A parameter sensitivity study of the MEDUSA-2.0 global ocean biogeochemical model

Global ocean biogeochemical models are necessary for our understanding of global marine biogeochemistry, and how the oceans interact in the past, present and future with other parts of the earth system when embedded within Earth System Models. Investigations of how biogeochemical parameters within these models influence global distributions of important tracers such as dissolved nutrients, oxygen and carbon are required to develop a better understanding of the complex processes being modelled, and are useful prior to a parameter optimisation study. Here we individually perturb 25 biogeochemical parameters within the global ocean biogeochemical model MEDUSA version 2.0, including controls on plankton growth and loss, and particle sinking speed and composition. We provide an assessment of how these parameters influence 16 different simulated tracer fields, and we rank local parameter sensitivities ordered from most to least influential. Most influential were those which control plankton abundance, particularly diatoms and meso-zooplankton, as they heavily influence the transport of organic matter to the deeper ocean and hence the global distribution of dissolved nutrients. The work done here provides the first detailed investigation of the inner workings of MEDUSA, and will aid future parameter optimisation studies by highlighting which parameters are more influential and therefore more important to calibrate to real observations. The structure of this chapter is as follows: first the introduction in Section 6.1, followed by the methodology in Section 6.2, results and discussion in Section 6.3, conclusions in Section 6.4, and finally the code availability in Section 6.5. Work carried out to determine the preferred biogeochemical model step size to take is presented in Appendix B, and the preferred biogeochemical model spin up length to reach an equilibrated state in Appendix C.

6.1 Introduction

The global ocean biogeochemical model MEDUSA (Yool et al., 2011, 2013) has a nitrogen core, two size classes of phytoplankton and zooplankton, and two different sized sinking detrital pools. MEDUSA-2.0 was chosen by the Met Office to be the biogeochemical component of their Earth System Model UKESM1, after

a comparison of 6 possible candidates for the task (Kwiatkowski et al., 2014). MEDUSA is considered of "intermediate complexity" as it is neither the simplest nor the most complex of the existing global ocean biogeochemical models, and in the comparison study by Kwiatkowski et al. (2014) MEDUSA was the 4^{th} most complex of the 6 models.

Understanding how biogeochemical parameters influence tracer distributions throughout the global ocean provides insight on the complex processes being modelled, and indicates what may be the result of shifting biogeochemical processes during future climate change. The impact of several parameters within MEDUSA-1.0 on surface chlorophyll have been investigated before by Hemmings et al. (2015), however due to the large computational expense of MEDUSA, which is embedded within the NEMO ocean GCM (Madec, 2008), this was carried out on an emulator, built from several 1-D site-based simulators, and not on the full 3-D global model. Aside from some manual tuning of several parameters by Yool et al. (2013) during the transition from MEDUSA-1.0 to MEDUSA-2.0, the influence of parameters within the full 3-D model have not been investigated. However, MEDUSA has recently been coupled (Khatiwala et al., 2021) to the Transport Matrix Method (TMM) framework (Khatiwala et al., 2005; Khatiwala, 2007, 2018), described in Section 2.2.1. This not only allows fast "offline" integration of MEDUSA, but also makes it possible to couple it to different circulations derived from various other ocean GCMs. In particular, it is now feasible to run MEDUSA to equilibrium in a few hours, instead of several weeks to months, and hence makes it possible for the first time to perform a parameter sensitivity study.

Here we carry out a simple, computationally inexpensive parameter perturbation study, to indicate which parameters most influence different MEDUSA tracers. We perturb each parameter individually, keeping the other parameters constant, thereby provide local sensitivities only, whereby the sensitivities are calculated at one specific location in parameter space. Global sensitivities, which indicate parameter sensitivities throughout the entire parameter space, are much more computationally expensive to compute, and are therefore not attempted in this

chapter. For more information on previous parameter sensitivity studies within an ocean biogeochemical modelling context, see the introduction of Chapter 4 in this thesis.

In this chapter a varied selection of 25 biogeochemical parameters were chosen, including but not limited to controls on particle sinking speed, plankton growth and loss rates, and nutrient uptake half-saturation constants. Their influence on 16 different global tracer and variable distributions, such as dissolved nutrients, oxygen, carbon and plankton abundance, were described and interpreted, and ranked according to a defined misfit metric which incorporates 5 of the 16 oceanic variables.

6.2 Methodology

6.2.1 Ocean Biogeochemical Model

The chosen global ocean biogeochemical model to investigate is MEDUSA-2.0 (Yool et al., 2013), described in Section 2.1.2. We use the TMM interface (see Section 2.2.1) to MEDUSA recently developed by Khatiwala et al. (2021). As mentioned above, MEDUSA is embedded within the NEMO ocean GCM (Madec, 2008) and until now could only be used with that model. As a result, due to the large computational expense of NEMO-MEDUSA, MEDUSA was rarely spun-up for more than a few hundred years.

Here we use the transport matrices extracted from the ocean model in version 2.9 of the University of Victoria Earth System Climate Model (Weaver et al., 2001; Eby et al., 2009), which has a resolution of 1.8 x 3.6 degrees and 19 ocean depth layers. See Section 2.2.2 for more information. After some investigation we selected an ocean biogeochemical step size of 2 hours (4 per ocean model step: see Appendix B). The offline method allows us to efficiently spin up MEDUSA from uniform tracer distributions to assumed equilibrium after 3000 years (see Appendix C). The final year modelled tracer and variable distributions are averaged. All 15 of MEDUSA's prognostic tracers are used for analysis, as well as the diagnostic variable net primary production (NPP), as described in Table 6.1.

Abbreviation	Description	Units
OXY	Dissolved oxygen	mmol O2/m3
ALK	Alkalinity	meq/m3
SIL	Silicate	mmol Si/m3
DIN	Dissolved inorganic nitrogen	mmol N/m3
DET	Detrital nitrogen	mmol N/m3
FER	Dissolved iron	mmol Fe/m3
DTC	Detrital carbon	mmol C/m3
DIC	Dissolved inorganic carbon	mmol C/m3
CHN	Chl-a concentration in non-diatom phytoplankton	mg Chl/m3
CHD	Chl-a concentration in diatom phytoplankton	mg Chl/m3
PHN	Non-diatom phytoplankton	mmol N/m3
PHD	Diatom phytoplankton	mmol N/m3
PDS	Biogenic silicon in diatom phytoplankton	mmol Si/m3
ZMI	Micro zooplankton	mmol N/m3
ZME	Meso zooplankton	mmol N/m3
NPP	Non-diatom and diatom primary production	mmol N/m2/d

Table 6.1: The 16 annually-averaged oceanic variables used for analysis. MEDUSA models 15 tracers, however NPP was then calculated from diagnostic variables.

6.2.2 Misfit function

With any perturbation in model biogeochemical parameter values, there is to be expected a change in the model tracer distributions after a 3000 year spin up. To quantify this relative change we compare the model to observations. While there are adequate real observations of oceanic oxygen, surface primary production, and some nutrients, there are only very sparse observations of other MEDUSA model tracers for which to compare the model to, such as the zooplankton tracers. Therefore, during this perturbation sensitivity study we compare the model to synthetic observations, created by a MEDUSA simulation with default parameter values. The misfit between the model and the synthetic observations is defined below.

We calculate regional misfit values (Equation 6.1) for each of the 16 chosen tracers/variables (see Table 6.1), similar to in Chapter 3 where we calculated regional misfits for 3 tracers for each of their 19 defined oceanic biomes. One averaged misfit (Equation 6.2) is then calculated for each of the 16 tracers, which is the mean of the regional misfits for that specific tracer. Then, we calculate a

single combined misfit (Equation 6.3), which is the summed square of the regional misfits for DIN, OXY, ALK, SIL and NPP. We limit the combined misfit to these 5 variables because there are adequate real observations of the former 4 tracers, and then including NPP would encourage better constraints on the plankton parameters, therefore this would be a suitable misfit metric to be used during an optimisation study to calibrate MEDUSA to real observations.

The regional misfits are defined as

$$r_{qj}(\mathbf{x}) = \frac{\sqrt{\sum_{i=1}^{N_j}(m_{qi}(\mathbf{x}) - o_{qi})^2 \frac{V_i}{V_j^T}}}{\sum_{i=1}^{N_j} o_{qi} \frac{V_i}{V_j^T}} \sqrt{\frac{V_j^T}{V_T}} \sqrt{\frac{V_T}{V_{Global}}}, \tag{6.1}$$

where $m_{qi}(\mathbf{x})$ is the model solution with parameters $\mathbf{x}$ at grid point i for tracer q, and o_{qi} the corresponding synthetic observations. N_j is the total number of model grid boxes in biome region j. The misfit is normalised by the volume-weighted mean tracer concentration for that region and weighted, first, by individual grid point volumes V_i relative to the total volume V_j^T of biome region j and, second, by the region's total volume V_j^T relative to the total volume of all regions for this tracer V_T. As only surface regions are included for primary production, a final weighting by the V_T relative to the global ocean volume V_{Global} is applied. It is important to note that grid point volumes at depth can be over one order of magnitude larger than at the surface, therefore the misfit is weighted more heavily to the deeper ocean.

The averaged misfits are

$$\mathbf{r}_q(\mathbf{x}) = \frac{\sum_{i=1}^{J_q} r_{qj}(\mathbf{x})}{J_q}, \tag{6.2}$$

where J_q is the number of regions for that specific tracer. The combined misfit is then defined as

$$f(\mathbf{x}) = \sum_{q=1}^{5} \sum_{j=1}^{J_q} r_{qj}^{\epsilon}(\mathbf{x})^2, \tag{6.3}$$

where tracers 1 to 5 consist of DIN, OXY, ALK, SIL and NPP.

Description	Abbreviation	Lower bound	Upper bound	Reference value	Perturbed value	Combined misfit
Maximum meso-zooplankton grazing rate [d−1]	xgme	0.25	1.15	0.5	0.59	0.043753534
Diatom Fe nutrient uptake half-saturation constant [mmolFem−3]	xfld	0.000335	0.00134	0.00067	0.0007705	0.026528962
Diatom N nutrient uptake half-saturation constant [mmolNm−3]	xnld	0.1	6	0.75	1.34	0.020579814
Maximum growth rate for non-diatoms [d−1]	xvpn	0.25	2	0.8	0.975	0.019923755
Non-diatom chl-specific initial slope of P–I curve [gC(gchl)−1 (Wm−2)−1 d−1	xaln	0.3	30	15	17.97	0.006264509
Maximum growth rate for diatoms [d−1]	xvpd	0.25	2	0.75	0.925	0.005838385
Diatom chl-specific initial slope of P–I curve [gC(gchl)−1 (Wm−2)−1 d−1]	xald	0.3	30	11.25	14.22	0.005265045
Non-diatom Fe nutrient uptake half-saturation constant [mmolFem−3]	xfln	0.000165	0.00066	0.00033	0.0003795	0.005158644
Non-diatom N nutrient uptake half-saturation constant [mmolNm−3]	xnln	0.1	6	0.5	1.09	0.005120757
Meso-zooplankton maximum loss rate [d−1]	xmzme	0.05	0.8	0.2	0.275	0.004908510
Maximum micro-zooplankton grazing rate [d−1]	xgmi	0.1	4	2	2.39	0.004886489
O2 consumption by C remineralisation [molO2 (mol C)−1]	xthetarem	0.5	2.5	1.1226	1.3226	0.004854933
Micro-zooplankton maximum loss rate [d−1]	xmzmi	0.05	0.8	0.1	0.175	0.003535864
Fast-sinking fraction of diatom silicon grazing losses	xfdfrac3	0	1	0.8	0.9	0.001809975
Biogenic Si dissolution length scale [m]	xfastsi	1000	4000	2000	2300	0.001096607
Detritus gravitational sinking rate [m s-1]	vsed	1.45E-05	6.94E-05	3.47E-05	4.02E-05	0.000748186
Excess organic carbon dissolution length scale [m]	xfastc	94	376	188	216.2	0.000700417
CaCO3 : POC: export rain ratio scalar Ridgwell et al. (2007)	r0	0.02	0.082	0.026	0.0322	0.000511988
Detrital C remineralisation rate [d−1]	xmdc	0.00635	0.038	0.019	0.022165	0.000435672
Detrital N remineralisation rate [d−1]	xmd	0.0079	0.0474	0.0237	0.02765	0.000414066
Fast-sinking fraction of diatom nat. mort. losses	xfdfrac1	0	1	0.333	0.433	0.000381340
Fast-sinking fraction of mesozooplankton mort. losses	xfdfrac2	0	1	1	0.9*	0.000263152
O2 consumption by N remineralisation [molO2 (mol N)−1]	xthetanit	1	4	2	2.3	0.000250821
Si nutrient uptake half-saturation constant [mmolSim−3]	xsld	0.15	6	3	3.585	0.000011427
Calcium carbonate dissolution length scale [m]	xfastca	1750	7000	3500	4025	0.000000004

Table 6.2: MEDUSA parameter descriptions, abbreviations, bounds, reference and perturbed parameter values, and combined misfits (see Equation 6.3) for each of the 25 parameter perturbation MEDUSA simulations. The table has been sorted according to the combined misfits, with the parameter perturbation resulting in the largest combined misfit at the top.

6.2.3 Gradient-based local sensitivities

Prior to a parameter optimisation study aiming to minimise the misfit function, it is important to understand how sensitive the misfit function is to each of the parameters, as it may influence the success of an optimisation algorithm. A computationally expensive option would be to use a global sensitivity method which captures the global sensitivity of the misfit function throughout the entire parameter space, yet requires a large number of model runs. Alternatively, one could calculate local sensitivities which are inexpensive yet do not provide knowledge of the sensitivities throughout the entire parameter space. Here we have chosen the latter.

We chose 25 biogeochemical parameters (see Table 6.2) to investigate within MEDUSA, influencing important processes such as particle sinking speed, nutrient uptake, and phytoplankton and zooplankton growth and loss. See Section 1.4 and Yool et al. (2013) for further descriptions of these parameters. Table 6.2 shows

upper and lower bounds for each parameter, determined either by in-situ oceanic observations, or by doubling and halving the existing MEDUSA reference parameter values. All other parameter values and code switches were kept constant. We calculated gradient-based local sensitivities by evaluating MEDUSA 25 times, each with 24 parameters set to their reference values and the remaining parameter with their reference value perturbed by +10% of their range. This perturbation was always an increase, with the exception of the parameter xfdfrac2 whose reference value was already at the upper bound of the range, therefore here the perturbation was a decrease of 10%. The misfits to the synthetic observations were calculated after each perturbation run as in Section 6.2.2, with a slight difference for xdfrac2 in the misfit calculation where we switch the modelled and synthetic observations terms, so to consistently calculate the misfit according to a positive perturbation in each of the parameters.

This method of calculating sensitivities is strongly influenced by the chosen parameter bounds. If one parameter has wider bounds than another, then the perturbation will be larger, and therefore the resulting sensitivity may be exaggerated. One alternative to this method, independent of parameter bounds, would have been to perturb each parameter by a very small but equal amount, such as machine epsilon. However, as parameters often have different units, an equal perturbation in each parameter may not be comparable (such as a perturbation of 1 metre in one parameter, and 1 kilometre in another). For that reason, and the fact that this sensitivity study would go on to aid an optimisation study with the same specified parameter bounds as in this chapter, we chose to perturb each parameter according to their parameter bounds. Therefore, the quantitative results shown in this chapter are useful for the following chapter, to understand which parameters to optimise within MEDUSA, however a more qualitative approach should be taken when using this information to understand which parameters most influence modelled tracer distributions.

6.3 Results and Discussions

The results of the perturbation sensitivity study are in Table 6.3, showing the 16 average misfits between the synthetic observations and the model after each of the 25 parameter changes. The combined misfit is in the final column, which combines the misfits of DIN, OXY, ALK, SIL and NPP, and has been used to sort the perturbed parameters from the largest to smallest combined misfit. Also, to provide some perspective to the real ocean, the misfits to real observations were also calculated for DIN and SIL (Garcia et al., 2018b), OXY (Garcia et al., 2018a), ALK (Lauvset et al., 2016), and NPP (Behrenfeld and Falkowski, 1997) for the reference model run with default MEDUSA parameter values. These misfits to real observations are unsurprisingly higher than those to synthetic.

Distributions of the change in modelled tracers after a parameter perturbation has been shown in Figures 6.1 to 6.6 for the parameters which caused the greatest change in the 5 tracers included in the combined misfit. The maximum meso-zooplankton grazing rate (*xgme*) caused the largest change in silicate and primary production, and the $CaCO_3$: POC export rain ratio scalar (*xridg_r0*) the largest in nitrate and alkalinity. The oxygen consumption by carbon remineralisation (*xthetarem*) caused the largest change in oxygen, however as oxygen was the only tracer to be influenced by this parameter perturbation, we instead show the tracer change distributions caused by a perturbation in the second most influential parameter to oxygen, which was the detrital carbon remineralisation rate (*xmdc*). To see tracer change distributions due to all 25 parameter perturbations, and for all 16 tracers/variables, see Section 6.5 for where to view and download the supplementary figures for this chapter. The tracer responses to all 25 parameter perturbations are described and discussed below.

	DIN	OXY	ALK	SIL	NPP	DET	DIC	DTC	FER	CHN	CHD	PHN	PHD	PDS	ZMI	ZME	Combined Misfit
Reference	2.76E-02	6.74E-02	2.74E-03	9.22E-02	2.18E-02												4.05E-01
xgme	1.66E-03	2.78E-03	2.80E-05	3.96E-02	3.72E-03	3.66E-01	1.84E-04	3.13E-01	1.56E-03	3.01E-02	1.49E-01	2.72E-02	1.41E-01	1.28E-01	9.63E-02	3.32E-02	4.38E-02
xfld	2.35E-03	3.49E-03	2.55E-05	3.12E-02	1.25E-03	1.96E-02	2.11E-04	1.90E-02	9.66E-04	1.64E-02	1.40E-01	1.19E-02	1.10E-01	9.67E-02	2.47E-02	1.40E-02	2.65E-02
xnld	1.66E-03	2.15E-03	1.97E-05	2.60E-02	1.27E-03	2.16E-02	1.39E-04	2.14E-02	7.34E-04	1.28E-02	1.01E-01	9.42E-03	9.04E-02	8.11E-02	2.03E-02	8.37E-03	2.06E-02
xvpn	8.01E-04	1.09E-03	2.53E-05	2.70E-02	1.54E-03	2.21E-02	8.60E-05	1.99E-02	8.36E-04	2.49E-02	6.84E-02	1.99E-02	6.82E-02	6.04E-02	3.17E-02	1.78E-02	1.99E-02
xaln	8.70E-04	1.34E-03	5.58E-05	1.48E-02	1.68E-03	4.88E-02	1.30E-04	4.22E-02	9.31E-04	3.20E-02	8.13E-02	3.02E-02	7.33E-02	6.67E-02	3.45E-02	2.46E-02	6.26E-03
xvpd	1.58E-03	2.92E-03	2.58E-05	1.52E-02	9.51E-04	1.39E-02	1.71E-04	1.24E-02	6.50E-04	1.19E-02	7.76E-02	1.19E-02	5.68E-02	5.86E-02	1.78E-02	1.09E-02	5.84E-03
xald	1.44E-03	2.27E-03	2.96E-05	1.43E-02	1.10E-03	2.03E-02	1.40E-04	1.89E-02	6.86E-04	1.61E-02	9.91E-02	1.60E-02	8.42E-02	6.35E-02	2.48E-02	1.04E-02	5.26E-03
xfln	7.30E-04	1.01E-03	4.21E-05	1.45E-02	1.26E-03	2.99E-02	1.06E-04	2.70E-02	7.92E-04	4.20E-02	7.58E-02	2.37E-02	7.15E-02	6.12E-02	3.26E-02	1.63E-02	5.16E-03
xnln	9.22E-04	1.40E-03	6.81E-05	1.39E-02	2.32E-03	5.46E-02	1.64E-04	5.14E-02	9.98E-04	4.87E-02	1.30E-01	2.92E-02	1.33E-01	7.22E-02	4.89E-02	1.82E-02	5.12E-03
xmzme	4.53E-04	8.76E-04	8.38E-05	1.42E-02	1.21E-03	3.85E-02	1.53E-04	2.97E-02	5.02E-04	2.28E-02	1.40E-01	1.64E-02	1.24E-01	5.92E-02	6.79E-02	9.46E-02	4.91E-03
xgmi	6.18E-04	9.28E-04	5.04E-05	1.40E-02	1.06E-03	3.17E-02	1.03E-04	2.65E-02	6.94E-04	3.55E-02	8.19E-02	4.62E-02	7.12E-02	5.43E-02	2.62E-02	1.33E-02	4.89E-03
xthetarem	0.00E+00	1.41E-02	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	4.84E-03
xmzmi	5.54E-04	6.76E-04	1.09E-04	1.13E-02	8.57E-04	5.76E-02	1.23E-04	3.90E-02	4.98E-04	3.81E-02	4.16E-02	3.18E-02	4.02E-02	3.50E-02	6.28E-02	1.12E-02	3.54E-03
xfdfrac3	3.35E-04	3.69E-04	8.48E-06	8.47E-03	1.28E-04	2.32E-03	2.55E-05	1.79E-03	1.62E-04	1.68E-03	4.50E-03	1.31E-03	3.63E-03	2.06E-02	3.08E-03	1.57E-03	1.81E-03
xfastsi	2.90E-04	3.35E-04	3.32E-06	6.85E-03	5.21E-05	7.65E-04	2.02E-05	7.61E-04	1.30E-04	5.34E-04	1.44E-03	4.78E-04	1.19E-03	8.75E-03	9.78E-04	6.80E-04	1.10E-03
vsed	7.81E-04	2.93E-03	2.50E-04	4.23E-03	1.31E-03	1.89E-01	4.38E-04	1.62E-01	6.91E-04	1.13E-02	2.16E-02	1.17E-02	2.16E-02	2.07E-02	1.65E-02	2.17E-02	7.47E-04
xfastc	1.90E-03	2.37E-03	1.52E-04	4.07E-03	1.12E-03	1.65E-02	1.77E-04	1.55E-02	1.63E-03	9.79E-03	2.52E-02	1.13E-02	2.13E-02	1.79E-02	1.27E-02	1.52E-02	7.00E-04
xridg_r0	2.46E-03	2.81E-03	6.26E-04	1.75E-03	6.75E-04	1.04E-02	5.89E-04	1.00E-02	7.35E-04	6.41E-03	8.79E-03	5.75E-03	7.61E-03	6.13E-03	6.24E-03	7.48E-03	5.12E-04
xmdc	1.09E-04	3.84E-03	6.28E-06	1.68E-03	3.20E-04	4.91E-03	4.50E-04	1.43E-01	1.16E-04	3.29E-03	9.40E-03	2.35E-03	8.65E-03	6.37E-03	5.76E-03	3.20E-03	4.34E-04
xmd	6.27E-04	2.10E-03	1.11E-04	3.13E-03	1.30E-03	1.48E-01	2.63E-04	2.57E-02	5.66E-04	8.88E-03	1.99E-02	7.82E-03	1.56E-02	1.47E-02	1.37E-02	1.28E-02	4.13E-04
xfdfrac1	3.22E-04	6.42E-04	1.75E-05	3.76E-03	3.04E-04	6.38E-03	4.47E-05	5.63E-03	1.31E-04	5.66E-03	1.37E-02	4.44E-03	1.04E-02	1.49E-02	9.99E-03	6.29E-03	3.81E-04
xfdfrac2	6.15E-04	7.12E-04	5.78E-05	2.92E-03	8.52E-04	2.15E-02	7.33E-05	1.65E-02	3.13E-04	5.23E-03	1.84E-02	4.90E-03	1.25E-02	1.22E-02	5.64E-03	2.26E-02	2.63E-04
xthetanit	0.00E+00	3.19E-03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	2.50E-04
xsld	2.54E-05	3.17E-05	8.49E-07	6.18E-04	1.44E-05	2.12E-04	2.37E-06	1.88E-04	1.20E-05	1.48E-04	1.30E-03	1.10E-04	7.71E-04	2.61E-03	2.24E-04	1.20E-04	1.14E-05
xfastca	6.13E-06	7.57E-06	3.09E-06	1.82E-06	9.77E-07	1.64E-05	2.33E-06	1.58E-05	3.54E-06	5.29E-06	1.04E-05	5.55E-06	9.57E-06	7.42E-06	7.66E-06	9.56E-06	3.94E-09

Table 6.3: Misfit results for the 25 perturbed parameter simulations. (Column 1) Perturbed parameter, (Columns 2:17) averaged misfits (see Equation 6.2) for each of the 16 tracers, and (Column 18) combined misfits (see Equation 6.3). As in Table 6.2, the table has been sorted according to the combined misfits, with the parameter perturbation resulting in the largest combined misfit at the top. Each column has been colour formatted to highlight the largest values in that specific column in red, and smallest in white. The top row shows averaged misfits for the reference model run with default parameters, calculated similar to Equation 6.2 but the misfits were calculated to real observations (Garcia et al., 2018a,b; Lauvset et al., 2016; Behrenfeld and Falkowski, 1997) instead of synthetic observations.

6.3.1 Perturbation sensitivities rankings

The parameters within Table 6.3 have been ordered according to the combined misfit (see Equation 6.3) caused by each parameter perturbation. The combined misfit (both to synthetic and real observations) seems to be dominated by deviations in dissolved silicate, despite the fact that all tracer deviations are normalised by the averaged synthetic observations, indicating that perturbation-induced variations in silicate are larger than for the other 4 tracers included in the combined misfit. Parameters which heavily influence silicate also influence phytoplankton and zooplankton tracers, causing the upper half of Table 6.3 to be highlighted red in the silicate, primary production, phytoplankton and zooplankton tracer columns. This is likely due to the fact that any parameter perturbation heavily influencing silicate has had an effect on diatom abundance, which, with time, can influence non-diatom abundance due to competition, hence also influencing the zooplankton in the trophic level above. The parameters in the lower half of Table 6.3 have less influence on silicate, but some have strong controls on dissolved and detrital nitrogen and carbon, oxygen, and alkalinity, so despite their lower ranking in the combined misfit their biogeochemical importance should not be overlooked.

6.3.2 Half-saturation constants

Perturbing the half-saturation constants within MEDUSA influences the growth rates of the two different phytoplankton types. An increase in a phytoplankton's half-saturation constant of a certain limiting nutrient means it reaches its maximum nutrient uptake rate (and hence growth rate) at a higher nutrient concentration, therefore the nutrient is not reduced to such a low level. The most limiting nutrients within MEDUSA are predominantly silicate, followed by iron then nitrate for diatoms, and predominantly iron then nitrogen for non-diatoms (Yool et al., 2013, 2021).

The most influential perturbed half-saturation constants within MEDUSA were the diatom half-saturation constant for first iron then nitrogen, followed by the

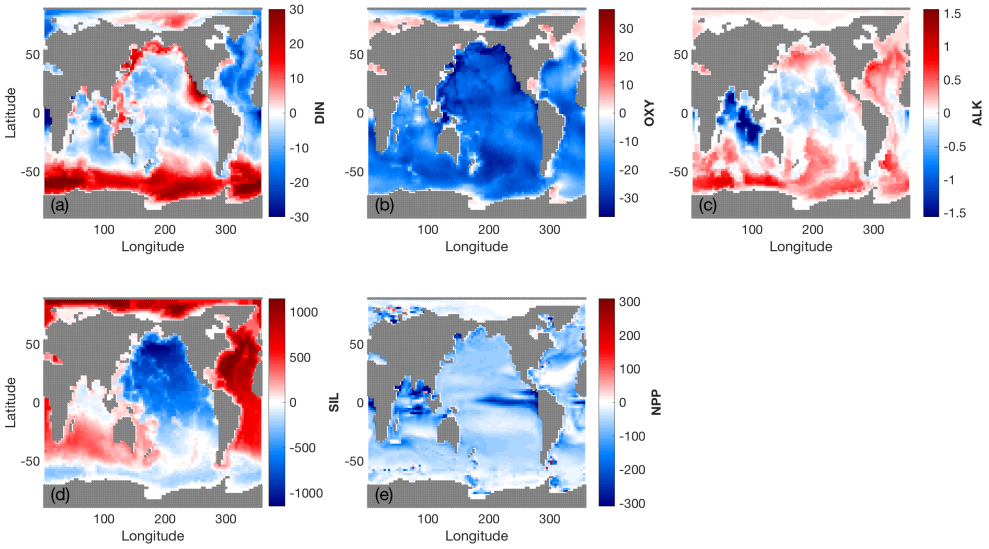

Figure 6.1: Change in column inventory of (a) DIN, (b) OXY, (c) ALK, (d) SIL, and (e) NPP after a +10% perturbation in *xgme*, the maximum meso-zooplankton grazing rate.

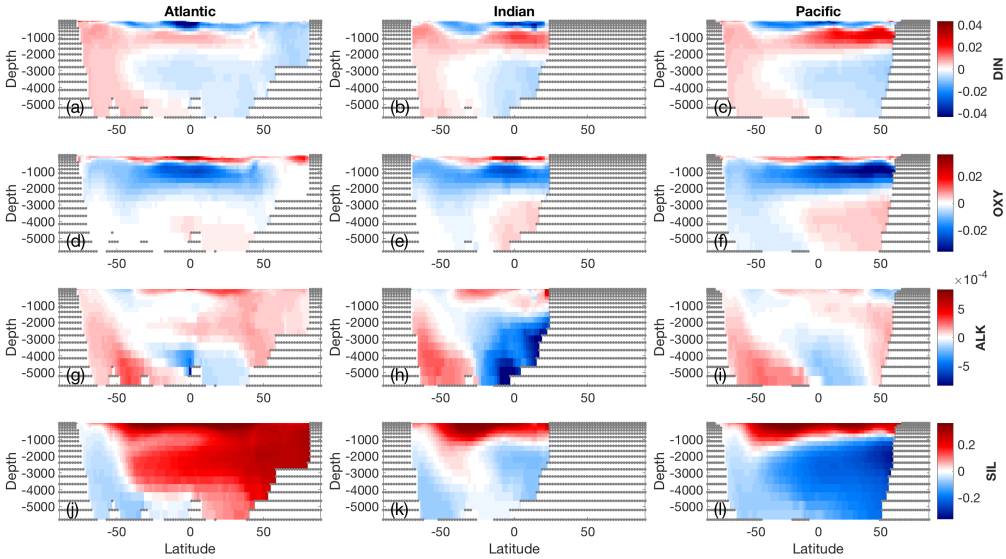

Figure 6.2: Change in volume-weighted zonally averaged (a-c) DIN, (d-f) OXY, (g-i) ALK, and (j-l) SIL after a +10% perturbation in *xgme*, the maximum meso-zooplankton grazing rate, for the (left) Atlantic, (middle) Indian, and (right) Pacific Oceans.

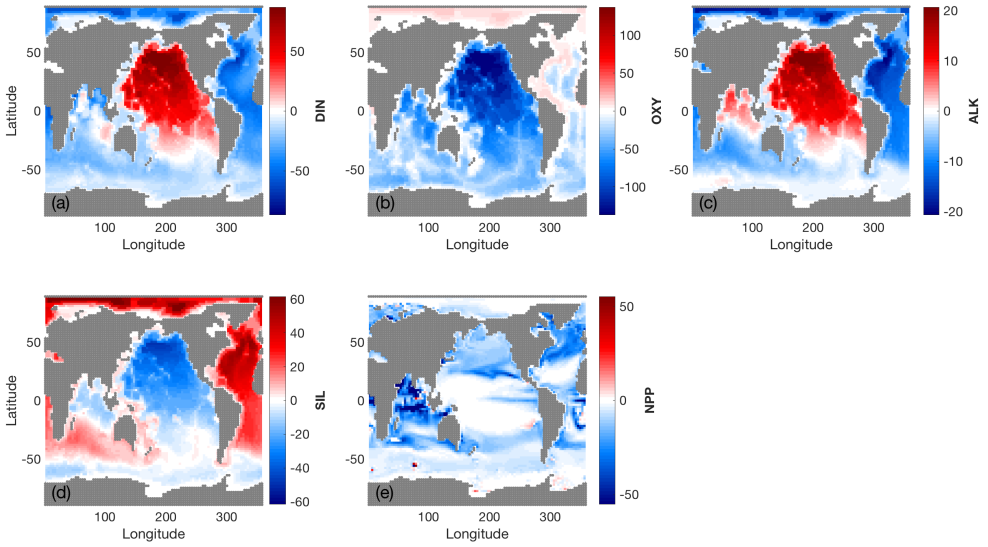

Figure 6.3: As in Figure 6.1, but for *xridg_r0*, the $CaCO_3$: POC: export rain ratio scalar.

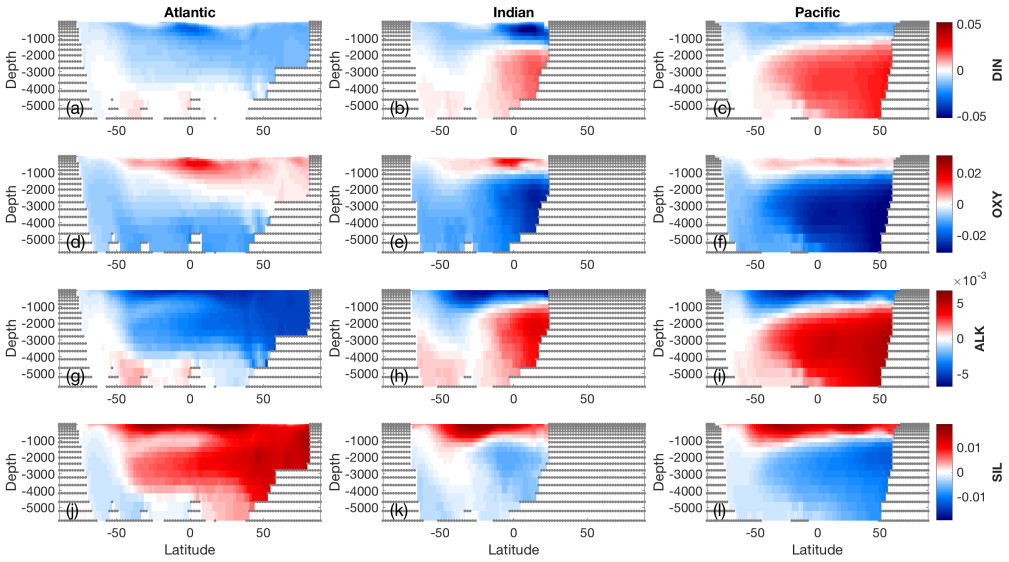

Figure 6.4: As in Figure 6.2, but for *xridg_r0*, the $CaCO_3$: POC: export rain ratio scalar.

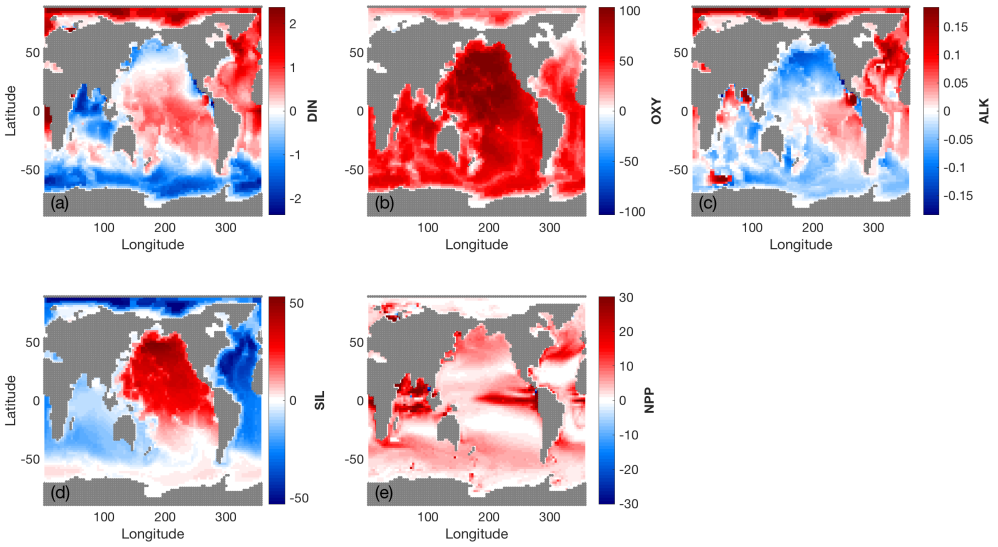

Figure 6.5: As in Figure 6.1, but for *xmdc*, the detrital C remineralisation rate.

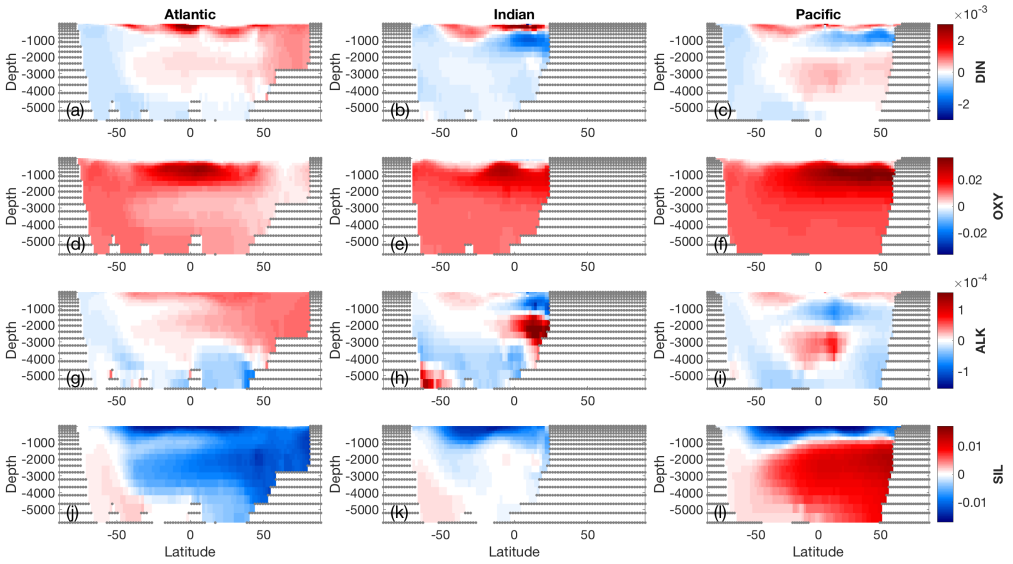

Figure 6.6: As in Figure 6.2, but for *xmdc*, the detrital C remineralisation rate.

non-diatom half-saturation constant for iron then nitrogen, with the diatom half-saturation constant for silicon being one of the least influential parameters. Therefore the controls on diatoms seem more important than on non-diatoms, and the controls on iron seem more important than on nitrogen, with silicon controls having little influence. This agrees well with the limiting nutrients for non-diatoms within MEDUSA, however as diatoms are dominantly limited by silicate, it is surprising the half-saturation for silicon had so little influence. It may be that silicate is so depleted at the surface, that a change in this half-saturation constant makes little difference.

The three perturbations in diatom nutrient uptake half saturation constants caused a decrease in surface diatoms, and therefore resulting in an increase in surface nutrients, and a decrease in dissolved nutrients such as nitrate and silicate being transported along the deep ocean conveyor belt to the deep north Pacific Ocean. The two perturbations in non-diatom nutrient uptake half saturation constants had a similar result, in that they caused a decrease in surface non-diatoms. However, in the dissolved nutrients tracer change distributions, we saw conflicting responses caused by both a decrease in non-diatoms and an increase in diatoms, as the latter could grow faster due to reduced competition. This may have caused the sub-surface decrease in nitrate after an increase in the non-diatom nitrogen uptake half saturation constant, as more diatoms would remove more nitrogen from the water.

6.3.3 Phytoplankton controls

The parameters controlling the maximum growth rate and initial slope of the P-I curve will have a similar influence on the biogeochemistry, as an increase in either will increase the abundance of that phytoplankton group, if zooplankton grazing remains constant. An increase in the maximum growth rate allows phytoplankton to grow faster when sufficient nutrients and light are available, and an increase in the initial slope of the P-I curve allows the phytoplankton to photosynthesise (and hence grow) more at lower irradiance levels.

The most influential perturbed phytoplankton parameters within MEDUSA were the non-diatom initial slope of the P-I curve then the non-diatom maximum growth rate, closely followed by the equivalent parameters for diatoms, in the same order. Therefore, unlike for the half-saturation constants, the controls on non-diatoms seem more important than on diatoms, and the controls on the initial slope of the P-I curve seem more important than on maximum growth rates. The controls on diatoms have more influence on nitrate, while the controls on non-diatoms have more influence on primary production, and the controls on growth rates have more influence on silicate.

An increase in the diatom parameters (both maximum growth rate and initial slope of P-I curve) resulted in an increase in surface diatoms, and therefore a decrease in surface nutrients and an increase at depth, as more diatoms stripped the surface of nutrients and transported them to depth via the ballasting sinking process where they were then remineralised. The overall influence on dissolved inorganic carbon was a global increase, particularly in the deeper ocean, again due to an increase in sinking carbon-rich organic matter. A perturbation in the non-diatom parameters had a similar influence on nitrate and dissolved inorganic carbon, as more non-diatoms in the surface resulting in less dissolved nutrients, however as non-diatoms are not included in the ballasting fast-sinking process, instead of transporting these to the deeper ocean like the diatoms these nutrients were remineralised shallower in the water column, resulting in an increase in nitrate and dissolved inorganic carbon in the sub-surface. Despite non-diatoms not requiring silicate it is still heavily influenced, with an increase in the surface global ocean and most of the Atlantic Ocean and a decrease in the deeper ocean. This may be due to the fact than an increase in non-diatoms caused a decrease in diatom abundance, possibly due to competition, leading to less silicate uptake by diatoms in the surface ocean, and therefore less silicate transported to the deeper ocean via the ballasting model.

6.3.4 Zooplankton controls

An increase in zooplankton maximum grazing rate means zooplankton can keep eating and growing to a higher threshold when sufficient prey is available. Also, as the zooplankton grazing half saturation constants remain the same during this increase in maximum grazing rate, the initial slope of grazing rate with increasing prey availability is steeper. This allows the zooplankton to graze faster when prey is less available. Micro-zooplankton prey on non-diatom phytoplankton and on particles of slow-sinking detritus, while meso-zooplankton prey on a larger source pool of non-diatoms, diatoms, micro-zooplankton and slow-sinking detritus. The zooplankton maximum loss rate is the maximum loss rate of zooplankton density-dependent losses (such as disease, cannibalism, predation by higher trophic levels) that can be reached with increasing zooplankton density, represented by a hyperbolic function. Therefore an increase in this parameter increases the maximum losses that occur as plankton concentrations rise. Again, as the zooplankton loss half saturation constants remain the same, the initial loss slope is steeper, therefore more loss occurs at lower zooplankton concentrations. Losses from micro-zooplankton contribute to the slow-sinking detrital pool, and losses from meso-zooplankton contribute to the fast-sinking detrital pool.

The most influential perturbed zooplankton parameters within MEDUSA were the meso-zooplankton maximum grazing rate (the most influential of the 25 parameters) then the meso-zooplankton maximum loss rate, closely followed by the equivalent parameters for micro-zooplankton, in the same order. Therefore, similar to the half-saturation constants, the controls on diatoms (grazed by meso-zooplankton) seem more important than on non-diatoms (grazed by micro-zooplankton), and the controls on the zooplankton maximum grazing rates seem more important than maximum loss rates. An increase in meso-zooplankton grazing would remove organic matter from the slow sinking detritus pool, and allow more to be transferred into the fast sinking detritus pool, as a fraction of meso-zooplankton losses contribute to fast sinking POC. However, this is not the case for micro-zooplankton, which graze on slow sinking detritus but then also return matter to the slow sinking pool

through micro-zooplankton losses, hence why the controls on meso-zooplankton grazing rates seem more important than for micro-zooplankton.

An increase in the micro-zooplankton maximum grazing rate resulted in an increase in micro-zooplankton and a decrease in non-diatoms, upon which the graze. This reduction in non-diatoms may have gone on to increase the abundance of surface diatoms, due to reduced competition, which stripped the surface of dissolve silicate and transported it towards the deep north Pacific. The influence on nitrate is more variable, possibly due to the conflicting response from the decrease in non-diatoms but increase in diatoms. With fewer non-diatoms, less nitrate is utilised at the surface, and less is being remineralised in the sub-surface, leading to an increase in surface nitrate and decrease in the sub-surface in some areas. A reduction in non-diatoms also led to an overall decrease in dissolved inorganic carbon, and detrital nitrogen and carbon. The micro-zooplankton maximum grazing rate perturbation also caused an increase in sub-surface meso-zooplankton, likely due to the increased supply of one of their food sources (micro-zooplankton). While an increase in micro-zooplankton maximum grazing rate caused fewer non-diatoms and more diatoms, an increase in meso-zooplankton maximum grazing rate caused fewer diatoms, and therefore more non-diatoms. Therefore the tracer responses to the two different perturbations oppose for most tracers. An increase in micro-zooplankton maximum loss rate has a similar effect as an increase in meso-zooplankton maximum growth rate, as both result in a more non-diatoms and fewer diatoms, while the perturbation in meso-zooplankton maximum loss rate influences the biogeochemistry similar to the micro-zooplankton maximum growth rate.

6.3.5 Particle sinking parameters

An increase in the detritus gravitational sinking rate causes faster sinking of the slow-sinking detritus pool, the majority of which remineralises in the upper ocean. This parameter is particularly interesting when considering optimising MEDUSA, as Yool et al. (2021) found the model generally remineralises too shallow within the water column. If the slow-sinking detritus pool were to sink faster, then there would

be less remineralisation occurring in the surface ocean, and more remineralisation somewhere below the surface, depending on how fast it is now sinking. When perturbed in MEDUSA, we saw this was the case for dissolved inorganic nitrogen and carbon, due to the downward shift in detrital nitrogen and carbon. The reduced nutrient supply in the surface caused a decrease in abundance of all plankton groups. Fewer phytoplankton resulted in reduced new $CaCO_3$ production and therefore increased alkalinity at the surface, as also seen by Kwon and Primeau (2008) regarding a similar biogeochemical parameter, the exponent in the power law for the depth profile of the remineralisation of particle organic matter. This parameter had little effect on dissolved silicate, as this nutrient is more influenced by the fast-sinking detrital pool as this contains diatoms.

There are 3 parameters within MEDUSA which control the fraction of diatom and meso-zooplankton process which contribute to the fast-sinking detrital pool. The one which has most influence on our combined misfit metric is the fast-sinking fraction of diatom silicon grazing losses (xfdfrac3), which controls the amount of opal from meso-zooplankton grazing on diatoms which goes into the fast-sinking pool. This most influences dissolved silicate in the model, which seems to dominate our combined misfit metric. An increase in this parameter caused more silicate to be removed from the surface ocean as more opal was transported deeper with the fast-sinking detrital pool before being remineralised at depth. The other two are the fast-sinking fraction of diatom and meso-zooplankton losses (xfdfrac1 and xfdfrac2 respectively). The former influences silicate similar to xfdfrac3, while the latter has more control on nitrate, dissolved inorganic carbon, detrital nitrogen and detrital carbon, as an increase in this parameter removed more detrital carbon and nitrogen from the surface, which then remineralised at depth.

The $CaCO_3$: POC export rain ratio scalar (Ridgwell et al., 2007) is the ratio of calcium carbonate to particulate organic carbon in sinking particulate organic carbon, and therefore controls the relative quantity of fast-sinking calcium carbonate, as a function of the ambient $CaCO_3$ saturation state. If this Ridgwell rain ratio coefficient is increased, then more $CaCO_3$ is present in the fast-sinking ballasting

detritus, thereby making the ballasting process more efficient at transporting organic material to the deeper ocean. This was seen in the modelled tracer distributions after a perturbation in this parameter, as nitrate and dissolved inorganic carbon were reduced in the surface ocean and increased at depth. More organic matter was remineralised in the deeper ocean, and as the remineralisation process utilises oxygen the amount of dissolved oxygen was decreased in the deeper ocean and increased at the surface. More $CaCO_3$ was removed from the surface ocean and transported to the deeper ocean, where it was then broken down to increase the amount of calcium and carbonate ions here, therefore alkalinity was strongly influenced by this parameter, similar to as seen by Kwon and Primeau (2008). This parameter would be important to consider when optimising MEDUSA, as Yool et al. (2021) found the model exports too much alkalinity from the surface.

6.3.6 Dissolution parameters

Dissolution parameters within MEDUSA include dissolution e-folding length scales, which describe how many metres a compound needs to sink through the water column before it is dissolved by a factor of approximately 2.7. The dissolution parameters which influenced the MEDUSA tracers the most were the biogenic silicon dissolution length scale, followed by the excess organic carbon dissolution length scale. The former mainly influenced silicate, as increasing this dissolution length scales allowed the biogenic silicon within sinking organic matter to travel deeper into the ocean before dissolution. The latter mainly influenced nitrate, dissolved inorganic carbon, alkalinity and dissolved iron, as an increase in the excess organic carbon dissolution length scale forces more organic carbon to dissolve deeper in the water column, and also releasing nutrients like nitrogen and iron along with it. Perturbing the calcium carbonate dissolution length scale within MEDUSA had very little influence on all of the 16 tracers. Kwon and Primeau (2008) also found this parameter to have little influence, when compared to parameters such as the $CaCO_3$ rain ratio and controls on vertical sinking.

Overall, a perturbation in both detrital nitrogen and carbon remineralisation rates caused a similar combined misfit deviation, with that for carbon having a slightly larger influence. An increase in the detrital carbon remineralisation rate particularly influenced oxygen, dissolved inorganic carbon and detrital carbon because more detrital carbon was remineralised at the surface, which caused there to be less dissolved inorganic carbon and more oxygen at depth. The final two parameter perturbations, the oxygen consumption by nitrogen and carbon remineralisation, only influenced the modelled oxygen distributions. An increase in either caused a decrease in global dissolved oxygen, particularly in non-downwelled deep water where oxygen hadn't been replenished by air-sea gas exchange.

6.4 Conclusion

We carried out a simple parameter perturbation study on 25 biogeochemical parameters within the global ocean biogeochemical model MEDUSA-2.0 (Yool et al., 2013). A systematic study of the parameters within the full 3-D model MEDUSA has not previously been done due to its large computation expense. However, by coupling it to the TMM (Khatiwala et al., 2021), this is now feasible. The aim of this chapter was to better understand the various biogeochemical controls within MEDUSA, and how certain parameters influence the biogeochemical processes modelled. Each parameter was individually perturbed by +10% of their feasible range and their influence on 16 model fields was assessed relative to the default simulation, including variables such as dissolved nutrients, oxygen and carbon, alkalinity, and plankton abundance. Parameters which had the strongest influence on MEDUSA were those which control plankton abundance, particularly diatoms and meso-zooplankton, as they heavily influence the transport of organic matter to the deeper ocean and hence the distribution of dissolved nutrients. The parameters were ordered according to a sensitivity ranking, which will be useful for future MEDUSA parameter optimisation studies, as it will help screen parameters which are not worth optimising and highlight which are the more important parameters to calibrate to observations.

6.5 Code Availability

The base TMM code are available to download from `https://doi.org/10.5281/zenodo.1246300` and the transport matrices used available to download from `http://kelvin.earth.ox.ac.uk/spk/Research/TMM/TransportMatrixConfigs/`. The TMM interface to MEDUSA-2.0 (Yool et al., 2013) will soon be submitted (Khatiwala et al., 2021). All supplementary figures can be viewed in the supplementary document "Ch6_SupplementaryFigs.pdf" supplied with this thesis. They include both a vertically-integrated and zonally-averaged figure of tracer change due to each of the 25 parameter perturbations, as seen in Figures 6.1 and 6.2, with tracer changes shown for all 16 tracers spread across 4 figures.

7

The calibration of the global ocean biogeochemical model MEDUSA-2.0 to observed oxygen, nutrients, alkalinity and primary production

The global ocean biogeochemical model MEDUSA has been selected as the marine biogeochemical component of the Earth System Model UKESM1, used to predict future anthropogenic climate change. Due to its large computational expense, MEDUSA has previously only undergone manual parameter tuning. However, by running the biogeochemical model with an "offline" general circulation using the Transport Matrix Method, it is now for the first time feasible to carry out systematic calibration using a numerical optimisation algorithm. Here we apply the optimisation algorithm DFO-LS, which we have previously shown can successfully optimise a global ocean biogeochemical model, to optimise 9 biogeochemical parameters within MEDUSA. We achieved mixed success, as despite significant improvements to modelled silicate, oxygen and alkalinity, some of the optimised configurations' net primary production (NPP) were almost half that of observations, and they modelled far too much opal production. This, coupled with the fact that the optimisation algorithm pushes many parameter values to the edge of their allowed parameter bounds, indicated that the parameters were tuned to compensate for some limitations, particularly in how MEDUSA models silicate. To improve future optimisation of MEDUSA it is recommended to have a greater weighting within the misfit function for some global metric of NPP, so as to not optimise to an unrealistically low non-diatom population. It was also found in this chapter that parameter optimisation had little influence on the ocean's response to future carbon emission scenarios, which is likely because the selected parameters for optimisation did not influence the solubility pump through which most of the ocean's carbon uptake occurs. The structure of this chapter is as follows: first the introduction in Section 7.1, followed by the methodology in Section 7.2, results in Section 7.3, discussion in Section 7.4, conclusions in Section 7.5, and finally the code availability in Section 7.6. Supplementary work carried out to compare the local parameter sensitivities calculated in Chapter 6 to sensitivities calculated at the optimised parameter configuration found in this chapter is presented in Appendix D, investigative twin optimisation experiments in Appendix E, and work to compare

MEDUSA-2.0 model outputs when coupled to various online and offline general circulation models in Appendix F.

7.1 Introduction

Global ocean biogeochemical models are necessary to understand the cycling of nutrients and carbon in the global ocean, and when embedded within Earth System Models (ESMs), they allow us to investigate how the ocean's carbon cycle interacts with other components of the earth system. With these tools one can quantitatively study the impact that drivers of climate change, such as increasing atmospheric concentrations of carbon dioxide, have on ocean biology and chemistry, and the resulting implications for the connecting components of the earth system. As the oceans have absorbed roughly a third of anthropogenic carbon dioxide from the atmosphere (Khatiwala et al., 2009; DeVries, 2014), the resulting impact is not trivial, and has been described in multiple reports of the Intergovernmental Panel on Climate Change (IPCC).

Global ocean biogeochemical models are heavily parameterised, therefore their model performance can be improved by calibrating the biogeochemical parameters to the real present-day ocean, by minimising the misfit between model outputs and oceanic observations. Optimisation algorithms work to locate minima within the misfit function throughout n-dimensional parameter space, where n is the number of parameters to be optimised. Preferably the location of the "global minimum" is sought, whereby the parameter configuration at this location in parameter space results in the very lowest possible misfit to observations. However, in high dimensionality problems when optimising over many parameters, and in complex models, the misfit function can be very difficult for an optimisation algorithm to navigate, with many areas within parameter space containing containing relatively low misfits called "local minima" for the optimisation algorithm to become trapped within. Global optimisation algorithms carry out a more global search of the parameter space, while local optimisation algorithms a more local search until they converge to a local minima. The former, though more likely to locate the global

minima, are often more computationally expensive, therefore the use of a local optimisation algorithm paired with a globalising method, such as initiating the optimisation algorithm from multiple points within parameter space, may be more efficient and just as successful (e.g. as in Chapter 3).

In this chapter we use the optimisation algorithm DFO-LS (Cartis et al., 2019) to optimise a selection of ocean biogeochemical parameters within the global ocean biogeochemical model of intermediate complexity, MEDUSA (Yool et al., 2011, 2013), which was selected to be the ocean biogeochemical component within the Earth System Model UKESM1 (Sellar et al., 2019; Kwiatkowski et al., 2014; Yool et al., 2021). Previously, an emulator built from several 1-dimensional simulators of MEDUSA-1.0 was calibrated to observations of chlorophyll (Hemmings et al., 2015), and since then minor manual tuning of several parameters was carried out during the transition from MEDUSA-1.0 to MEDUSA-2.0 (Yool et al., 2013). However, systematic calibration of the global biogeochemical model has not been carried out before now because of the computational expense of running MEDUSA within the NEMO ocean GCM (Madec, 2008) to which it is coupled. Here, we use the TMM interface to MEDUSA (Khatiwala et al., 2021) which now makes it feasible to carry out such an optimisation. Work is still underway to extract transport matrices from NEMO (Dr. Adrian Martin, pers. comm.). Therefore, we use monthly transport matrices derived from the UVic general circulation (Weaver et al., 2001; Eby et al., 2009), described in Section 2.2.2. Kriest et al. (2020) have shown there are issues with optimising global ocean biogeochemical models with different general circulations, as biogeochemical parameters may be over-tuned to compensate for failings within each circulation, therefore we cannot assume the resulting optimised parameter configuration(s) found from MEDUSA coupled to UVic will ensure a better fit to observations for MEDUSA coupled to NEMO. However, we use our findings here to comment on how MEDUSA might have been optimised had it been coupled to the NEMO circulation. Lastly, we also investigate how biogeochemical parameter optimisation within MEDUSA may impact its response to future CO_2 emissions.

7.2 Methodology

7.2.1 Biogeochemical model and parameters

In this chapter the global ocean biogeochemical model chosen to optimise was MEDUSA (Yool et al., 2013), which was used as in Chapter 6. After an initial parameter perturbation sensitivity study (see Chapter 6) and several short investigative twin optimisation experiments (see Appendix E), we chose 9 biogeochemical parameters (see Table 7.1) to optimise within MEDUSA. These influence processes such as particle sinking, nutrient uptake, phytoplankton growth and zooplankton grazing. As shown in Table 7.1, upper and lower bounds were determined for each parameter, within which lie reasonable parameter values. These were determined either by oceanic observations, or for those less constrained by observations, by doubling and halving the existing MEDUSA reference parameter values, to keep the parameter bounds sufficiently wide enough to allow the optimisation algorithm more freedom to determine the optimal values. All other parameter values and code switches were kept constant.

7.2.2 Oceanic observations

To calibrate the model we used annually-averaged, oceanic observations of:

- dissolved oxygen, provided by World Ocean Atlas (Garcia et al., 2018a),
- dissolved inorganic nitrogen (DIN), provided by World Ocean Atlas (Garcia et al., 2018b),
- dissolved inorganic silicon, provided by World Ocean Atlas (Garcia et al., 2018b),
- alkalinity, provided by GLODAP (Lauvset et al., 2016),
- net primary production (NPP), based on MODIS satellite-derived chlorophyll and the Vertically Generalized Production Model (Behrenfeld and Falkowski, 1997), obtained from the Ocean Productivity site of Oregon State University (`http://orca.science.oregonstate.edu/1080.by.2160.monthly.hdf.vgpm.m.chl.m.sst.php`).

Description	Units	Abbrev.	Lower bound	Upper bound	Reference config.	Opt1 config.	Opt2 config.	Opt3 config.	Comb4 config.
Detritus gravitational sinking rate	$m\ s^{-1}$	vsed	1.45×10^{-5}	6.94×10^{-5}	3.47×10^{-5}	3.20×10^{-5}	2.97×10^{-5}	3.47×10^{-5}	2.97×10^{-5}
Diatom chl-specific initial slope of P-I curve	$gC(gchl)^{-1}\ (Wm^{-2})^{-1}\ d^{-1}$	xald	7.5	30	11.25	26.72	30.0	11.25	30.0
Non-diatom chl-specific initial slope of P-I curve	$gC(gchl)^{-1}\ (Wm^{-2})^{-1}\ d^{-1}$	xaln	7.5	30	15	9.73	7.50	15	7.50
Excess organic carbon dissolution length scale	m	xfastc	94	376	188	190	376.0	188	376.0
Maximum meso-zooplankton grazing rate	d^{-1}	xgme	0.25	1	0.5	0.25	0.25	0.5	0.25
Maximum micro-zooplankton grazing rate	d^{-1}	xgmi	1	4	2	3.40	4.0	2	4.0
Detrital nitrogen remineralisation rate	d^{-1}	xmd	0.01	0.1	0.0237	0.01	0.01	0.0237	0.01
Diatom nitrogen nutrient uptake half-saturation	$mmol\ N\ m^{-3}$	xnld	0.375	1.5	0.75	0.62	0.375	0.75	0.375
$CaCO_3$:POC export rain ratio Ridgwell et al. (2007)		xridg_r0	0.02	0.082	0.026	0.026	0.026	0.051	0.051
Global misfit to observed DIN, OXY, SIL, ALK, NPP					0.4043	0.1506	0.1336	0.4165	0.1369
Global misfit to observed ALK					1.62×10^{-4}	1.72×10^{-4}	2.17×10^{-4}	0.76×10^{-4}	0.73×10^{-4}

Table 7.1: Selected MEDUSA parameter descriptions, units, abbreviations, lower and upper bounds, parameter values and global misfits for the MEDUSA configurations **Reference**, **Opt1**, **Opt2**, **Opt3** and **Comb4**. The first 8 parameters were tuned together, and calibrated to DIN, oxygen, silicate, alkalinity and NPP observations. The final parameter, the $CaCO_3$:POC export rain ratio scalar (Ridgwell et al., 2007), is the ratio of calcium carbonate ($CaCO_3$) to particulate organic carbon (POC) in fast-sinking detrital particles (see Section 1.4.3). This parameter was tuned in isolation, with the other 8 parameters fixed at their reference values, and calibrated to alkalinity observations only.

These are all 3-dimensional fields, except for NPP which is 2-dimensional. They were interpolated onto the model domain, and then used to calculate the misfit between model outputs and observations. Prior to calibrating to real observations, we calibrate to synthetic observations, which were simulated fields from a previous model run with a known ("**Reference**") parameter configuration (see Table 7.1).

7.2.3 Misfit function

The quality of the model is assessed by calculating the misfit between model outputs and real observations. After every evaluation of MEDUSA we calculate 89 individual misfit values, and one single global misfit value. The individual misfits (Equation 7.1) are calculated for each of the 5 observational variables, and each of the 19 biome regions (as displayed in Figure 3.1). As NPP is 2-dimensional, we cannot calculate a misfit for this variable for 6 of the deep water biome regions. The individual misfits are defined as

$$r_{qj}(\mathbf{x}) = \frac{\sqrt{\sum_{i=1}^{N_j}(m_{qi}(\mathbf{x}) - o_{qi})^2 \frac{V_i}{V_j^T}}}{\sum_{i=1}^{N_j} o_{qi} \frac{V_i}{V_j^T}} \sqrt{\frac{V_j^T}{V_T}} \sqrt{\frac{V_T}{V_{Global}}}, \tag{7.1}$$

where $m_{qi}(\mathbf{x})$ is the model output with parameters $\mathbf{x}$ at grid point i for tracer q, and o_{qi} the corresponding synthetic observations. N_j is the total number of model grid boxes in biome region j. The misfit is normalised by the volume-weighted mean tracer concentration for that region and weighted, first, by individual grid point volumes V_i relative to the total volume V_j^T of biome region j and, second, by the region's total volume V_j^T relative to the total volume of all regions for this tracer V_T. As only surface regions are included for NPP, a final weighting by the V_T relative to the global ocean volume V_{Global} is applied. It is important to note that grid point volumes at depth can be over one order of magnitude larger than at the surface, therefore the misfit is weighted more heavily to the deeper ocean.

The global misfit is then defined as

$$f(\mathbf{x}) = \sum_{q=1}^{5} \sum_{j=1}^{J_q} r_{qj}(\mathbf{x})^2, \tag{7.2}$$

where J_q is the number of regions for tracer q. We also define a global misfit specifically for just the tracer alkalinity, as in some experiments we go on to calibrate MEDUSA to alkalinity alone, and exclude the other 4 variables. This is defined as in Equation 7.2, however we only sum over the squared individual misfits r_{qj} where the tracer q = alkalinity. As this sums fewer terms, the global alkalinity misfit is always much lower than the global misfit.

We also define "expected baseline" misfits, which are the global misfits due to observational uncertainty alone. These were calculated by replacing the term $(m_{qi}(\mathbf{x}) - o_{qi})$ in Equation 7.1 with (ϵ_{qi}), where ϵ is the standard deviation of the observations. These baselines give an indication of a suitable termination threshold, below which any reduction of the misfit would be within observational uncertainty.

7.2.4 Optimisation algorithm DFO-LS

We apply the DFO-LS optimisation algorithm (Cartis et al., 2019; Roberts, 2019), described in Section 2.3.2. This was shown in Chapter 3 to perform well when optimising a global ocean biogeochemical model, albeit one of lower complexity than MEDUSA. To encourage DFO-LS to find the global minimum, we both initiate DFO-LS from multiple starting locations within parameter space, and turn on the functionality of "soft restarts".

7.2.5 Optimisation experiment design

In this chapter we carry out "twin" optimisation experiments calibrating to synthetic observations, and "real" optimisation experiments calibrating to real oceanic observations. In the former, the optimal parameter values are known, therefore the success of the optimisation algorithm can be determined.

With DFO-LS we performed two sets of optimisation experiments. Firstly, we optimised 8 of the chosen biogeochemical parameters to synthetic observations, followed by real observations. In these experiments, as the detrital nitrogen remineralisation rate (xmd) was altered by DFO-LS during optimisation, we

concurrently changed the detrital carbon remineralisation rate such that their ratio was fixed at 1:0.8 respectively.

These experiments highlighted how insensitive alkalinity was to these parameters, and hence there was little improvement to the alkalinity distributions in the optimised configurations of the model. Therefore, in the second set of optimisation experiments, we optimised one parameter in isolation, the scale factor (**xridg_r0**) for the latitude-dependent $CaCO_3$:POC export rain ratio relationship (Ridgwell et al., 2007), which has a significant influence on the global alkalinity distribution (see Section 1.4.3). We tuned this parameter to both synthetic, then real observations of alkalinity only, while the other 8 parameters remained fixed at their reference values. As this parameter was tuned in isolation to the other 8 parameters, it is deemed a separate optimisation experiment.

In both sets of experiments, we initiated DFO-LS from several different starting points within parameter space. The naming convention for the different optimisation experiments are as follows:

- "**t**" or "**r**", depicting whether this is a "twin" or "real" optimisation experiment,
- "**8**" or "**1**", which is the number of parameters being explicitly optimised, and
- "**_N**", where N is a number used to differentiate between starting locations in parameter space.

From these experiments we chose three optimised parameter configurations, two from the first set of optimisation experiments calibrating 8 parameters to all 5 observations, and one from the second set of optimisation experiments calibrating one parameter to alkalinity observations, henceforth called "**Opt1**", "**Opt2**" and "**Opt3**", respectively.

A fourth configuration was created by combining the optimised parameters from **Opt2**, and the one optimised parameter from **Opt3**, to create "**Comb4**". The 8 parameter values in **Opt2** were used here for **Comb4** because they resulted in the lowest misfit to real observations, found by DFO-LS. These were then combined

with the optimised **xridg__r0** parameter as in **Opt3** because DFO-LS found this to provide the lowest misfit to real alkalinity observations, when the other 8 parameters were set to their **Reference** values. This was done so as to include calibration to alkalinity, which was overshadowed in **Opt2** alone as DFO-LS preferentially calibrated to the other 4 observations.

See Table 7.1 for the parameter values for these configurations. The model outputs for each of these different optimised states were then compared to the **Reference** configuration and observations.

7.2.6 Prescribed atmospheric CO_2 emissions

We next investigate how the parameter optimisation carried out here affects MEDUSA's behaviour in response to CO_2 emissions. For each of the **Reference**, **Opt1-3** and **Comb4** configurations, MEDUSA was first spun up to equilibrium for 3000 years with a constant, pre-industrial atmospheric CO_2 mixing ratio of 278 ppm. Next, for each of these 5 configurations, we carried out 4 different transient runs from 1765 to 2100 with prescribed historical (up to 2005) and projected (from 2005) emissions for the Shared Socioeconomic Pathways (SSP) scenarios, SSP1-1.9, SSP2-4.5, SSP3-7.0 and SSP5-8.5 (Meinshausen et al., 2019). The numbers in the scenario labels correspond to the approximate anthropogenic radiative forcing (in W/m^2) in 2100. Thus, SSP-1.9, which includes negative emissions, is the most optimistic scenario, while SSP-8.5 is the 'business-as-usual' scenario in which emissions continue unabated. Figure 7.1 shows the CO_2 concentrations and corresponding diagnosed emissions under these scenarios. Note that these runs have fixed preindustrial transport matrices and other physical forcing.

For these runs, MEDUSA was coupled to a single box atmosphere whose CO_2 concentration prognostically varied in response to the specified emissions and air-sea CO_2 fluxes. As these simulations do not include an interactive land vegetation model, to ensure that the atmospheric CO_2 trajectory remains as close as possible to the historical and future SSP CO_2 concentrations, the emissions were generated by first forcing the **Reference** configuration with prescribed atmospheric CO_2

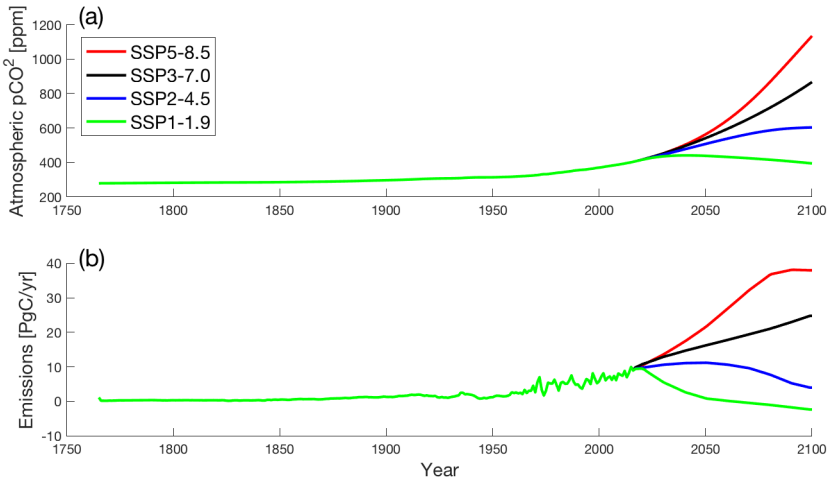

Figure 7.1: (a) Atmospheric CO_2 mixing ratio [ppm] (Meinshausen et al., 2019) and (b) diagnosed emissions [PgC/yr] from pre-industrial to the year 2100, for shared socio-economic pathways SSP5-8.5 (red), SSP3-7.0 (black), SSP2-4.5 (blue) and SSP1-1.9 (green).

concentrations for each of the historical+SSP scenarios. Emissions were then diagnosed as the sum of the rate of change of the atmospheric carbon inventory and simulated air-to-sea CO_2 flux.

7.3 Results

7.3.1 Optimisation experiments

Figure 7.2 shows the optimisation progress of the three twin experiments. Of the two which optimise 8 parameters to 5 tracers, both show significant misfit reduction. The misfit was reduced to below the expected baseline within approximately 30 evaluations in experiment **t8_1**, below which any further misfit reduction would be within observational uncertainty if we were calibrating to real data. During **t8_1**, two of the 8 parameters were successfully recovered within 80 evaluations, four were close to being recovered, and the optimisation of the remaining two were mostly unsuccessful, particularly **vsed** (the detritus gravitational sinking rate) which seemed to get trapped at the upper bound of the specified range. Despite a smaller misfit reduction, **t8_2** successfully recovered more parameters, recovering four of the 8 parameters, almost recovering three others, and only failing to tune

one parameter **xgmi** (the maximum micro-zooplankton grazing rate). This possibly highlights the importance of **xgmi**, as without its successful tuning **t8__2** was unable to reduce the misfit to within observational uncertainty. The experiment **t1__1**, which optimised only the parameter **xridg__r0** to synthetic alkalinity, required only 5 MEDUSA evaluations to both reach the expected baseline and successfully recover the parameter.

Figure 7.3 shows the optimisation progress of the 6 optimisation experiments to real data. All 4 experiments optimising 8 parameters to 5 tracers show significant misfit reduction, however there is a threshold below which none of the experiments manage to reduce the misfit by more than 64% of the **Reference** misfit. This is much higher than the observational uncertainty baseline of 5.6×10^{-4} for the 5 data fields in our misfit.

To achieve the misfit reduction, the optimisation experiments seemed to favour high values of **xald** (diatom chl-specific initial slope of P-I curve), **xfastc** (excess organic carbon dissolution length scale) and **xgmi** (maximum micro-zooplankton grazing rate), low values of **xgme** (maximum meso-zooplankton grazing rate), **xmd** (detrital nitrogen remineralisation rate) and **xnld** (diatom nitrogen uptake half-saturation), values of 3-4 ms^{-1} for **vsed** (detritus gravitational sinking rate), and various values for **xaln** (non-diatom chl-specific initial slope of P-I curve). During these optimisation experiments, increasing the initial slope of the P-I curve in diatoms and decreasing grazing by meso zooplankton favours diatom growth, while increasing grazing by micro zooplankton hinders non-diatom growth. Also, decreasing remineralisation rates and increasing dissolution length scales promotes vertical transport of organic matter.

The two experiments **t1__1** and **t1__2** optimising the parameter **xridg__r0** to observations of alkalinity only both reached a misfit of approximately 0.7×10^{-4}, again, above the observational uncertainty baseline for the global alkalinity misfit of 8.8×10^{-7}. Both experiments optimised **xridg__r0** to $\approx$0.05, which is slightly higher than the reference value currently set in MEDUSA-2.0, therefore suggesting

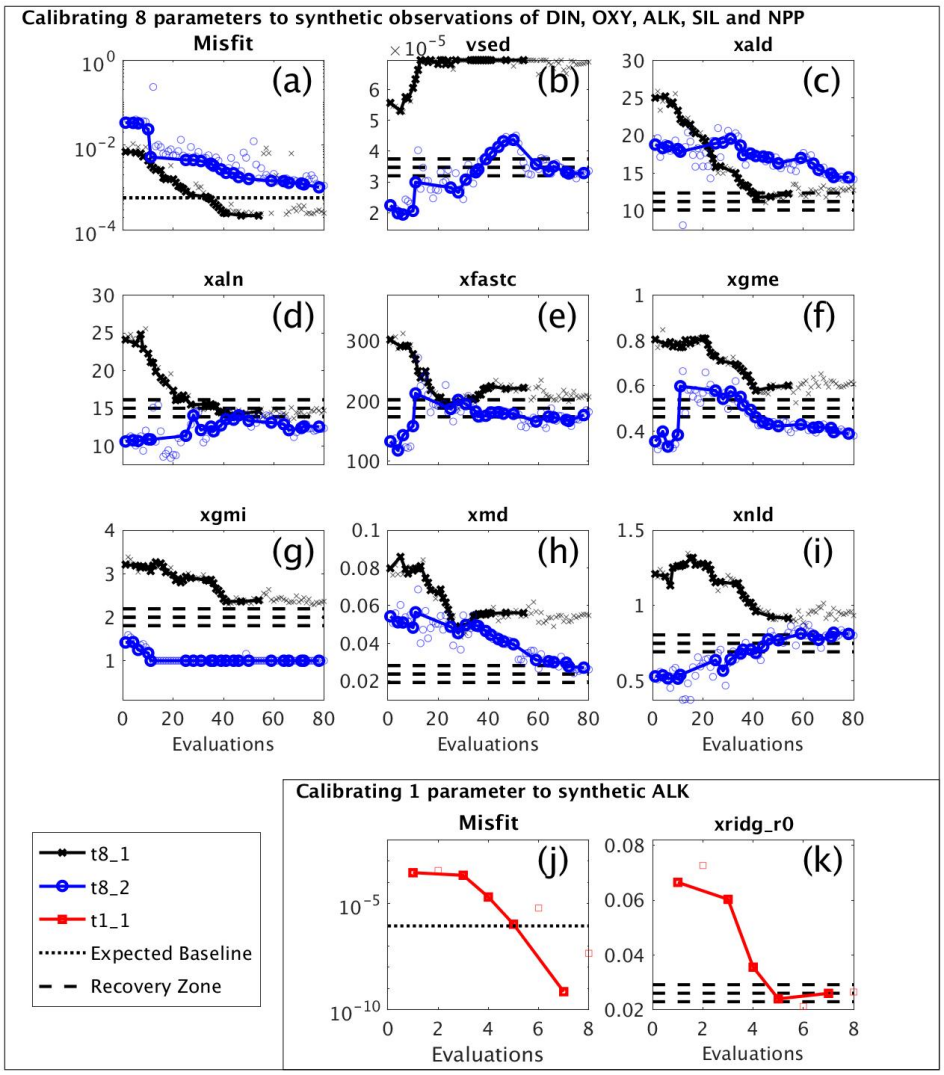

Figure 7.2: The misfit reduction and parameter calibration of the twin DFO-LS optimisation experiments. The top section shows the (a) global misfit reduction and (b-i) parameter optimisation progress of experiments **t8_1** (black line with crosses) and **t8_2** (blue line with circles), whereby DFO-LS calibrated 8 parameters to synthetic observations of DIN, oxygen, alkalinity, silicate and NPP. The bottom section shows (j) global alkalinity misfit reduction and (k) parameter optimisation progress of experiment **t1_1** (red line with squares), whereby DFO-LS calibrated 1 parameter to synthetic observations of alkalinity. Note that for each optimisation experiment every MEDUSA evaluation has been plotted with a small marker, however only MEDUSA evaluations which resulted in a lower misfit than previously seen in each optimisation experiment has been plotted with a solid line. Plotted on subplots (a) and (j) is the expected baseline misfit (horizontal black dotted line), which is the misfit due to observational uncertainty for (a) the 5 tracers and (j) alkalinity only, as described by Oliver et al. (2021). Plotted on subplots (b-i) and (k) is the parameter recovery zone (horizontal black dashed lines), which marks ±5% of the allowed parameter range either side of the reference parameter values, within which the target for that parameter is considered to be recovered, and therefore its optimisation is deemed successful.

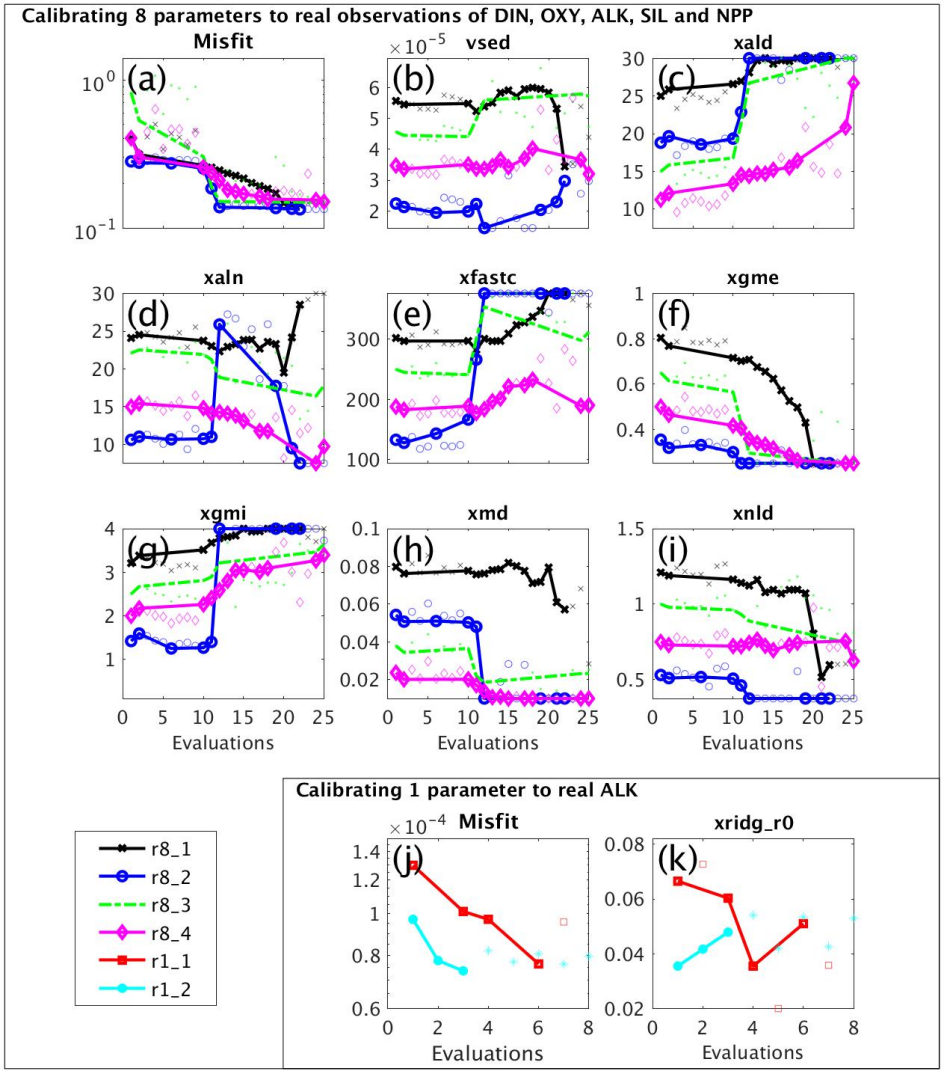

Figure 7.3: The misfit reduction and parameter calibration of all real DFO-LS optimisation experiments. The top section shows the (a) global misfit reduction and (b-i) parameter optimisation progress of experiments **r8__1** (black line with crosses), **r8__2** (blue line with circles), **r8__3** (green line with dots) and **r8__4** (magenta line with diamonds), whereby DFO-LS calibrated 8 parameters to real observations of DIN, oxygen, alkalinity, silicate and NPP. The bottom section shows (j) global alkalinity misfit reduction and (k) parameter optimisation progress of experiments **r1__1** (red line with squares) and **r1__2** (cyan line with filled circles), whereby DFO-LS calibrated 1 parameter to real observations of alkalinity. Note that for each optimisation experiment every MEDUSA evaluation has been plotted with a small marker, however only MEDUSA evaluations which resulted in a lower misfit than previously seen in each optimisation experiment has been plotted with a solid line.

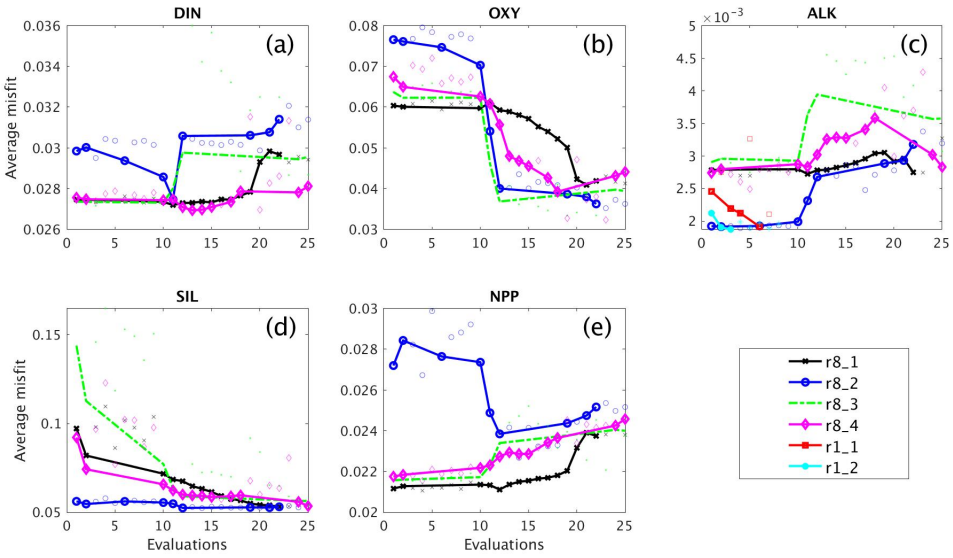

Figure 7.4: The averaged individual misfits for each of the tracers (a) DIN, (b) oxygen, (c) alkalinity, (d) silicate and (e) NPP, throughout the real DFO-LS optimisation experiments **r8_1** (black line with crosses), **r8_2** (blue line with circles), **r8_3** (green line with dots) and **r8_4** (magenta line with diamonds), and plotted just for alkalinity **r1_1** (red line with squares) and **r1_2** (cyan line with filled circles). Note that for each optimisation experiment every MEDUSA evaluation has been plotted with a small marker, however only MEDUSA evaluations which resulted in a lower misfit than previously seen in each optimisation experiment has been plotted with a solid line.

modelled alkalinity can be improved by strengthening export of $CaCO_3$ alongside POC export, for this chosen general circulation.

Figure 7.4 breaks down the global misfits into their averaged individual terms for each tracer during each optimisation experiment, so as to highlight which tracer distributions improve the most and cause the largest misfit reduction. In all optimisation experiments tuning 8 parameters, the misfit reduction was dominated by improvements to modelled oxygen and silica, while the model-data differences for DIN, alkalinity and NPP actually increased. This highlighted the need for optimisation experiments **t1_1** and **t1_2**, to isolate and improve modelled alkalinity, which is very sensitive to the parameter **xridg_r0**. However, similar experiments specifically to tune DIN and NPP would have been far too computationally expensive to carry out as they are influenced by many more biogeochemical parameters.

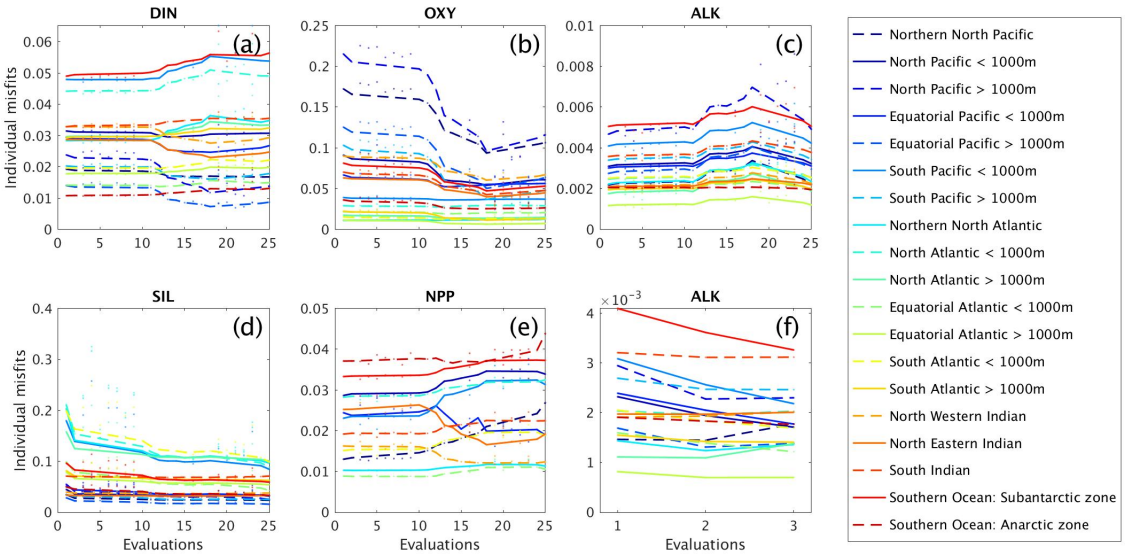

Figure 7.5: The individual misfits plotted for each of the 19 regions for each of the tracers (a) DIN, (b) oxygen, (c) alkalinity, (d) silicate and (e) NPP, throughout the real DFO-LS optimisation experiment **r8_4**, and (f) for alkalinity throughout **r1_2**.

Figure 7.5 further breaks down the global misfit by region for **r8_4** and **r1_2**, which show similar patterns of regional misfits as seen in the other optimisation experiments (not shown). The misfit reduction in oxygen in **r8_4** was dominated by improvements in the Pacific Ocean, particularly the deep North Pacific and northern North Pacific, and in silica by improvements in the entire North Atlantic, upper South Atlantic and upper South Pacific. The misfit reduction in alkalinity in **r1_2** was dominated by improvements in the subantarctic zone of the Southern Ocean, and the upper 1000 m of the entire Pacific Ocean.

7.3.2 Comparison of optimised configurations

The three resulting optimised MEDUSA configurations, **Opt1-3**, and the configuration **Comb4** (a combination of **Opt1** and **Opt3**), are compared with the **Reference** configuration and observations in Figure 7.6 and Table 7.2, and distributions of several diagnostic variables compared in Figures 7.7 and 7.8. Global observations were provided as in Section 7.2.2, and preindustrial dissolved inorganic carbon (DIC) provided by GLODAP (Lauvset et al., 2016). All 5 MEDUSA configurations produced global mean vertical profiles which:

- for DIN, fit well with observations, though with slightly lower concentrations between 600-3500m,
- for oxygen, modelled too low concentrations between 200-3000m,
- for silicate, modelled too high concentrations above 1500m and too low below 2000m,
- for alkalinity, modelled too much in the upper 2000m global ocean, and too little below 2000m,
- and for DIC, modelled too low concentrations below 1000m.

The configurations which included the 8 optimised parameters (**Opt1**, **Opt2** and **Comb4**) fit closest to observed vertical mean profiles of oxygen, silicate and DIC. We note that while oxygen and silica were two of the tracers used in the calibration, DIC was not. These three configurations modelled a lot less NPP than the **Reference** configuration, particularly in **Opt2** and **Comb4**, which was due to a massive reduction in non-diatom NPP throughout the global surface ocean (see Figure 7.7). This could have been caused by the very low value set for the non-diatom initial slope of the P-I curve (**xaln**), which meant the non-diatoms needed higher irradiance levels to grow well, and/or possibly caused by the high maximum micro-zooplankton grazing rate, which meant more non-diatoms were consumed. These three configurations also produced a lot more diatom biogenic opal than the **Reference** configuration, particularly in the northern North Pacific and Southern Ocean (see Figure 7.8), due to an increase in diatom abundance in these locations. This reduction in diatoms was possibly caused by a reduction in meso zooplankton (hence less grazing on diatoms) and an increase in micro-zooplankton (hence more grazing on non-diatoms, leading to less competition for diatoms) in the northern North Pacific and Southern Ocean (see Figure 7.7).

In the configurations which included the optimised parameter **xridg_r0** (**Opt3** and **Comb4**), the modelled vertical profiles of global mean alkalinity better fit observations, which again was to be expected as **xridg_r0** was calibrated to alkalinity observations alone. These configurations also produced and vertically transported more calcite throughout the global ocean (excluding the gyres) than in

Global Metrics	Downward Flux Particle Organic Carbon PgC/y	Dissolved Inorganic Carbon PgC	Downward Flux of calcite at 300m (observed) or 100m (modelled) PgC/y	Biogenic calcite Production PgC/y	Diatom biogenic opal production Pmol Si/y	Diatom and non-diatom NPP PgC/y
Observed	8.5-14.3 Xie et al. (2019)	36812 Lauvset et al. (2016)	0.912 Sulpis et al. (2021)	0.6-1.6 Ridgwell et al. (2009)	0.2-0.28 Nelson et al. (1995)	51.698* Behrenfeld et al. (1997)
Reference	8.71	36520	0.43	2.35	0.32	43.5
Opt1	8.55 (-2)	36710 (+0.5)	0.43 (-0.3)	2.48 (+6)	0.59 (+84)	34.6 (-21)
Opt2	7.61 (-13)	36780 (+0.7)	0.37 (-13)	2.14 (-9)	0.6 (+89)	27.5 (-37)
Opt3	8.27 (-5)	36180 (-0.9)	0.73 (+71)	4.02 (+71)	0.31 (-4)	41.2 (-5)
Comb4	7.26 (-17)	36440 (-0.2)	0.66 (+55)	3.78 (+61)	0.62 (+93)	26.0 (-40)

Table 7.2: A comparison of global metrics between those from observations or previous studies, and those calculated from the model outputs by the **Reference**, **Opt1**, **Opt2**, **Opt3** and **Comb4** MEDUSA configurations. Values in brackets show the directional percentage change from the **Reference** configuration. *NPP observations value was calculated from the observations used for calibration (see Section 7.2.2).

the **Reference** configuration (see Figure 7.8). This was due to the higher value of **xridg__r0** causing a greater ratio of $CaCO_3$ within POC export.

7.3.3 Response to CO_2 emissions

After the optimisation experiments, each of the **Reference** and 4 optimised MEDUSA configurations were then used to model future global ocean biogeochemistry under 4 different shared socio-economic pathways (see Section 7.2.6). Figures 7.9 and 7.10 show that for each emission scenario, there was negligible difference between the response by each MEDUSA configuration's modelled air-sea CO_2 flux, global DIC, surface ocean pH and $CaCO_3$ saturation state (Ω). For all scenarios, as expected CO_2 flux and DIC increased with emissions, while ocean pH and Ω calcite decreased.

However, there were differences in total NPP and downward flux of calcite. In all cases, NPP increased with emissions, as the increase in productivity caused by the rising temperatures were not remediated by increased stratification because the circulation in the transport matrix was fixed. The rate of increase in NPP varied depending on the chosen MEDUSA parameters. The configurations **Reference** and **Opt1** modelled the same NPP response, however **Opt2** with a particularly large excess organic carbon dissolution length scale (**xfastc**) modelled a slower increase, and **Opt3** with a larger $CaCO_3$:POC rain ratio (**xridg__r0**) modelled a faster increase. This link between increasing CO_2 emissions and NPP may be due to

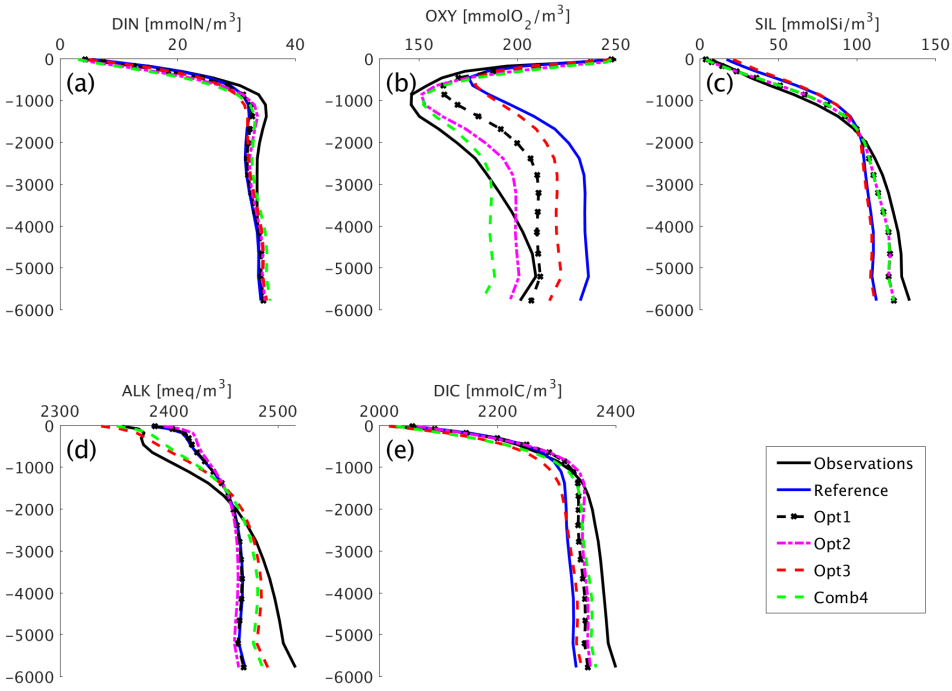

Figure 7.6: Vertical profiles of globally averaged (a) DIN [mmolN/m^3], (b) dissolved oxygen [mmolO_2/m^3], (c) silicate [mmolSi/m^3], (d) alkalinity [meq/m^3] and (e) DIC [mmolC/m^3], calculated from observations (black solid line), and the **Reference** (blue solid line), **Opt1** (black dashed line with crosses), **Opt2** (magenta dashed line with dots), **Opt3** (red dashed line) and **Comb4** (green dashed line) MEDUSA configurations. Observations were provided as in Section 7.2.2, and DIC provided by GLODAP (Lauvset et al., 2016).

a reduction in $CaCO_3$ production, and therefore a reduction in transport of organic matter to the deep ocean via the ballasting process, leaving more nutrients in the upper ocean to promote more NPP. As **Comb4** was a combination of the **Opt2** and **Opt3** parameters, their opposing responses seem to have cancelled out, causing NPP within **Comb4** to respond the same as **Reference** and **Opt1**. However, if a dynamic circulation had been used, the NPP response to warming would likely have been a decline, and there could have been bigger differences between simulations. The rate of decrease in downward flux of calcite with increasing emissions in each of the MEDUSA configurations seemed to be determined by the $CaCO_3$:POC rain ratio (**xridg_r0**), and configurations with a larger $CaCO_3$:POC rain ratio (**Opt3** and **Comb4**) predicted an extra decrease of 0.1 PgC/yr by the year 2100. As the

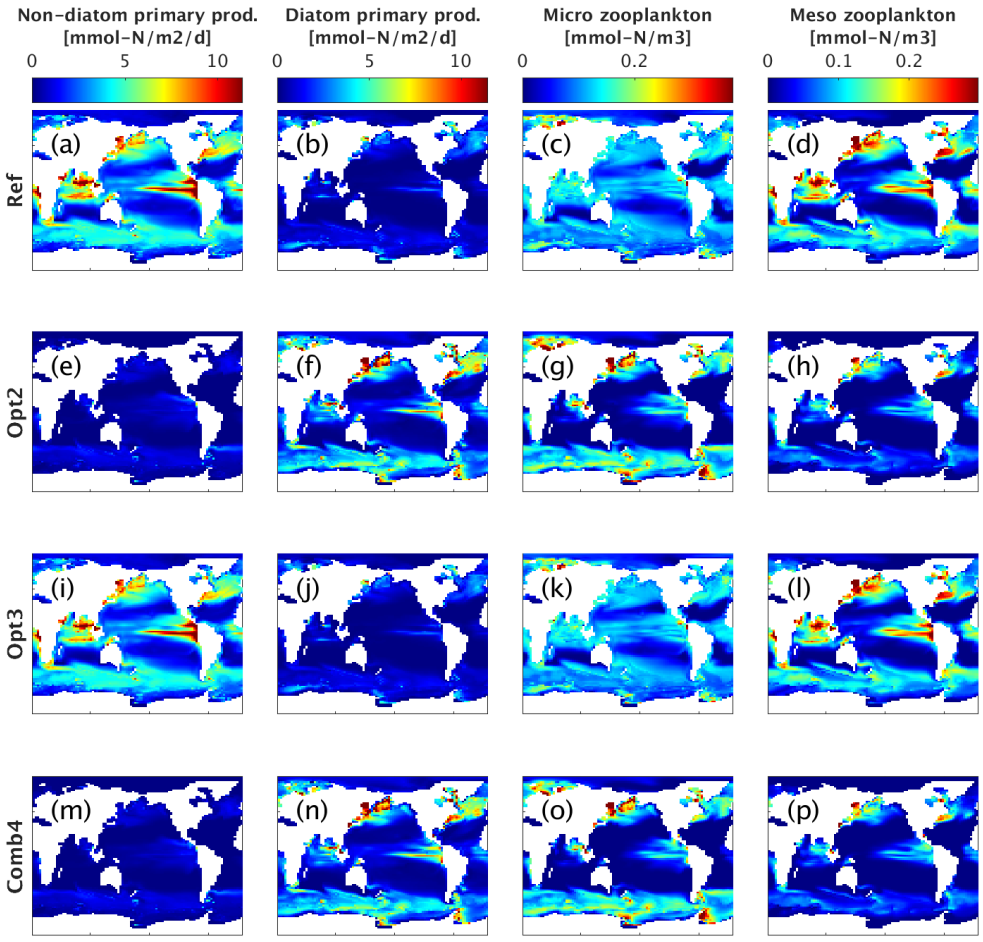

Figure 7.7: Global distributions of (far-left) non-diatom NPP [mmolN/m^2/d], (mid-left) diatom NPP [mmolN/m^2/d], (mid-right) micro zooplankton [mmolN/m^3] averaged in the surface 130m, and (far-right) meso zooplankton [mmolN/m^3] averaged in the surface 130m, for the (a-d) **Reference**, (e-h) **Opt2**, (i-l) **Opt3** and (m-p) **Comb4** MEDUSA configurations.

ocean becomes more corrosive to calcite with increasing carbon emissions, more calcite dissolves in the surface ocean and therefore the downward flux of calcite is reduced, and this effect was magnified when MEDUSA modelled a greater fraction of calcite within sinking matter.

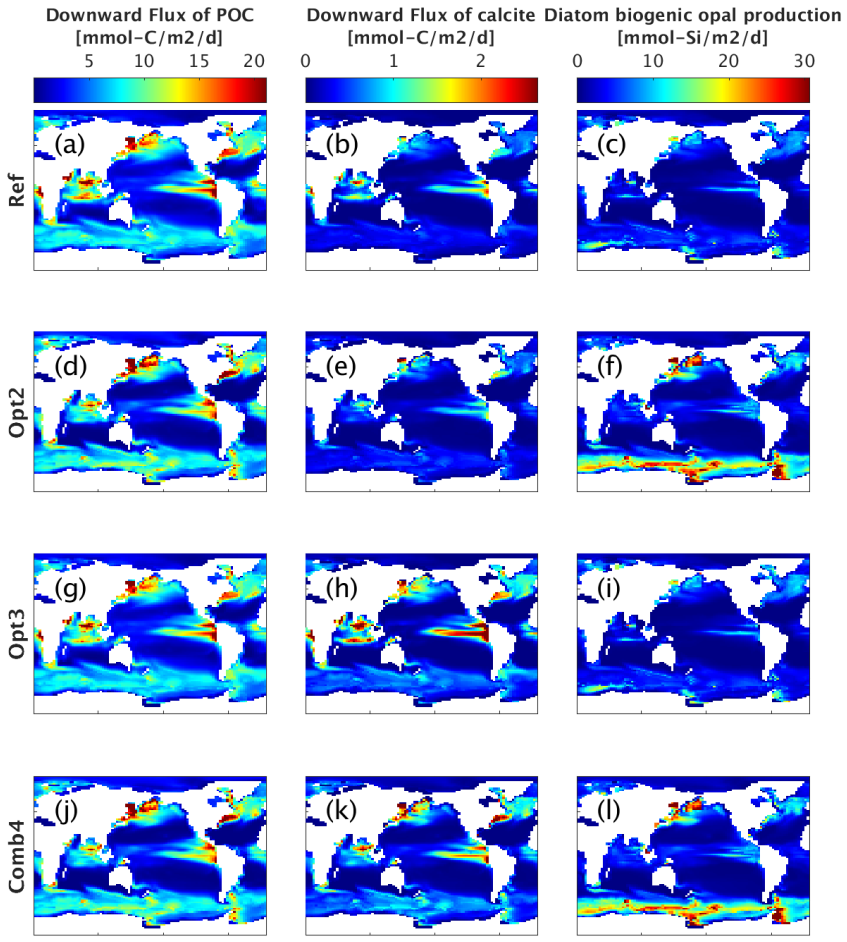

Figure 7.8: Global distributions of (left) downward flux of Particulate Organic Matter [mmolC/m^2/d], (middle) downward flux of calcite [mmolC/m^2/d], and (right) diatom biogenic opal production [mmolSi/m^2/d], for the (a-c) **Reference**, (d-f) **Opt2**, (g-i) **Opt3** and (j-l) **Comb4** MEDUSA configurations.

7.4 Discussion

7.4.1 Performance of the optimisation algorithm on the problem setup

Here we used the twin experiments to evaluate the success of the DFO-LS optimisation algorithm when calibrating multiple ocean biogeochemical parameters within MEDUSA. With similar twin experiments, in Chapter 3 we showed DFO-LS could reliably optimise 5 out of 6 biogeochemical parameters within the global ocean biogeochemical model MOPS, within approximately 40 evaluations of the

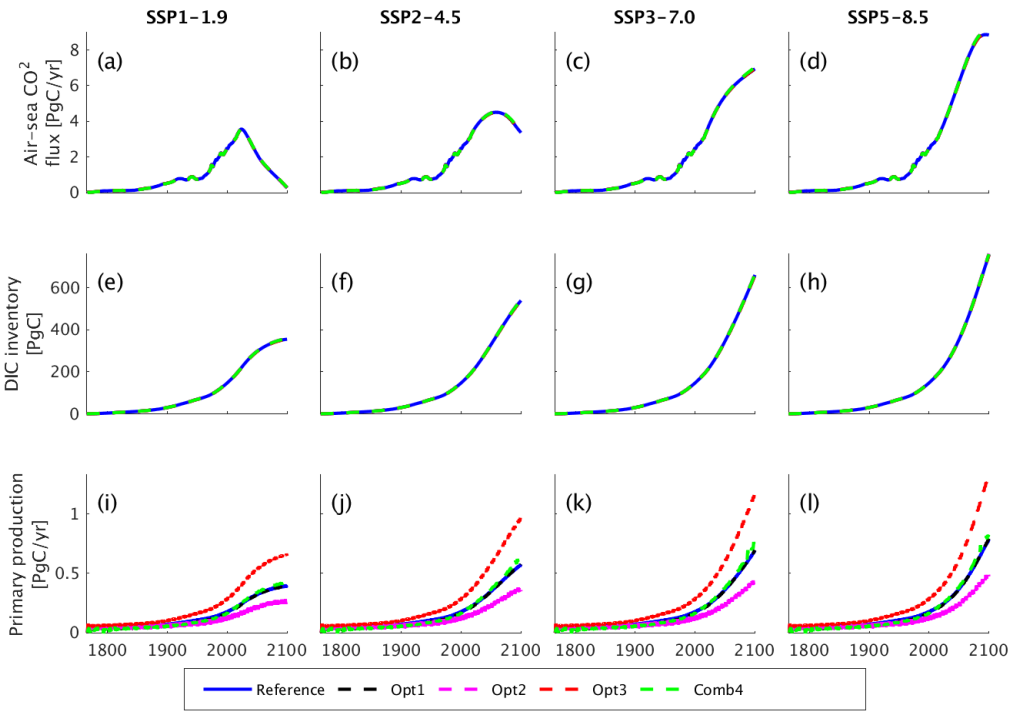

Figure 7.9: The change from modelled pre-industrial (a-d) air-sea CO_2 flux [PgC/yr], (e-h) DIC inventory [PgC], and (i-l) NPP [PgC/yr] from the year 1765 to 2100, modelled by the **Reference** (blue line), **Opt1** (black dashed line), **Opt2** (magenta dashed line), **Opt3** (red dashed line) and **Comb4** (green dashed line) MEDUSA configurations under the shared socio-economic pathways (far-left) SSP1-1.9, (mid-left) SSP2-4.5, (mid-right) SSP3-7.0, and (far-right) SSP5-8.5.

model. In this chapter we apply DFO-LS to a more complex biogeochemical model and attempt to optimise more biogeochemical parameters. This poses a greater challenge to DFO-LS than in Chapter 3.

We found that in both twin experiments optimising 8 parameters, the global misfit to synthetic observations was significantly reduced within 80 model evaluations (the maximum we allowed in our experiments), with one calibrating to within levels of real observational uncertainty. However, the target parameter values were not recovered as well as in Chapter 3, although the majority were calibrated to close to their target values. When the dimensionality of the problem was reduced to one and only alkalinity was used for calibration, the **xridg_r0** parameter was successfully optimised within only 5 evaluations of MEDUSA. This was reassuring to see that when the context was simpler, the optimisation algorithm works smoothly.

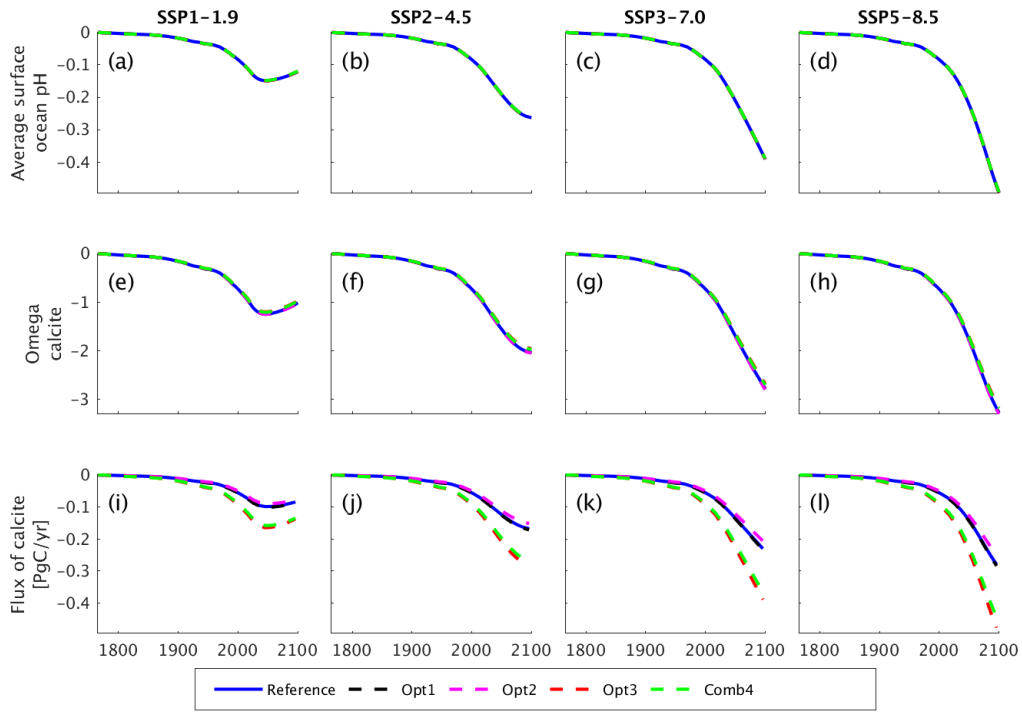

Figure 7.10: The change from modelled pre-industrial (a-d) average surface ocean pH, (e-h) omega calcite, and (i-l) downward flux of calcite [PgC/yr] from the year 1765 to 2100, modelled by the **Reference** (blue line), **Opt1** (black dashed line), **Opt2** (magenta dashed line), **Opt3** (red dashed line) and **Comb4** (green dashed line) MEDUSA configurations under the shared socio-economic pathways (far-left) SSP1-1.9, (mid-left) SSP2-4.5, (mid-right) SSP3-7.0, and (far-right) SSP5-8.5.

Conceivably, the more challenging 8 parameter optimisation problem would have benefited from more than the 80 model evaluations we allowed DFO-LS.

7.4.2 The quality of the optimised MEDUSA configurations

Each of the 4 optimisation configurations of MEDUSA had a significantly lower misfit to observations than did the **Reference** configuration, and on that metric could each be deemed a "better" version of the model. However, several issues were seen during the optimisation progress which highlighted problems with the resulting optimised configurations. Firstly, the optimisation experiments initiated from different starting locations in parameter space did not all converge to the same parameter configuration. This indicates that, instead of locating the global optimal parameter configuration, DFO-LS located several different local optima. Secondly,

these local optima all resulted in similarly low misfit values, indicative of a misfit topography with different areas of parameter space with similarly low misfits, as also seen by Kriest et al. (2017), for example. The same misfit reduction can thus be achieved by tuning towards different parameter values, making it more difficult to find the global optimum. Thirdly, DFO-LS consistently tuned three of the parameters to the edge of their specified limits. Therefore, when DFO-LS tries to push a parameter beyond these bounds this suggests it is not tuning that parameter to its correct optimal value, but to a value which will cause some limitation of the model to be compensated for. Lastly, when calibrating to all 5 observational variables only the misfits to oxygen and silicate observations were improved, while those to DIN, alkalinity and NPP actually worsened. As oxygen and silicate seemed to dominate the misfit, DFO-LS found parameter configurations which prioritised the improvement of these two variables, at the expense of the remaining variables. One way to improve a tracer which is less influential to the overall misfit (such as alkalinity) is to calibrate to that tracer alone. In summary, while DFO-LS succeeded in significantly reducing the misfit, the quality of the resulting "optimised" configurations were called into question.

Improvements to silica MEDUSA models too low concentrations of silica below 2000m. Diatoms create their hard shells out of silica, and via a ballasting process are transported to the deeper ocean. Hence an increase in diatoms would increase the modelled concentrations of silica at depth. To achieve this, DFO-LS tuned the biogeochemical parameters related to diatom growth and grazing by meso zooplankton, and essentially killed off most of the non-diatoms so as to reduce competition for the diatoms. This resulted in unrealistically low levels of non-diatom NPP, and too much opal production, all to improve the simulated distribution of silica.

Extinction of some plankton groups are not an uncommon consequence of ocean biogeochemical parameter optimisation (e.g. Kuhn and Fennel, 2019; Kriest et al., 2017; Schartau et al., 2001; Schartau and Oschlies, 2003). Schartau et al. (2001)

attempted to solve this problem by including zooplankton data to the observations used for calibration, with unsuccessful results for their zero-dimensional model. Kriest et al. (2017) narrowed their parameter bounds to better constrain the zooplankton grazing and mortality rates in their 3-dimensional model, which led their optimisation algorithm to converge to a more realistic optimised configuration. The limitation in modelled silica, which the optimisation algorithm tried to compensate for, is likely due to the parameter configuration and not due to the specific ocean circulation, as similar distributions of silicate have been modelled by MEDUSA coupled to other circulation models (see Figure F.1 in Appendix F).

Improvements to dissolved oxygen Another aspect that the optimised parameters had to compensate for was too high concentrations of oxygen in the global deep ocean. DFO-LS found that, of the parameters it was given to optimise, the misfit could be decreased by drastically reducing the remineralisation rate of detrital nitrogen (and also detrital carbon alongside it). This allowed more organic matter to sink to the deeper ocean before being remineralised, resulting in more overall oxygen consumption at depth and thus lower concentrations of dissolved oxygen, which is a better fit to observations. This discrepancy between modelled and observed oxygen concentrations at depth may be specific to the offline transport matrices of the UVic general circulation used in this chapter. When MEDUSA is coupled to the NEMO circulation it models an oxygen distribution which better fits observations (see Figure F.2 in Appendix F).

Improvements to alkalinity A final discrepancy that the optimised parameters had to compensate for was too high alkalinity in the surface ocean, and too low in the deep ocean. The alkalinity profile was improved by increasing the $CaCO_3$:POC rain ratio parameter, thereby increasing the relative quantity of fast-sinking $CaCO_3$ within sinking particulate organic matter. When $CaCO_3$ dissolves, calcium and carbonate ions are released, which increases alkalinity. Therefore, transporting more $CaCO_3$ from the surface to the deeper ocean acts to both decrease surface alkalinity and increase alkalinity at depth. As $CaCO_3$ acts as a ballast, this parameter also

influences dissolved oxygen and the transport of organic matter. Again, like for oxygen, this misfit to observations may be specific to the offline UVic general circulation (see Figure F.3 in Appendix F). When MEDUSA was coupled with the online NEMO circulation, the modelled alkalinity distribution was much closer to observations. In fact, Yool et al. (2021) found MEDUSA to model too low alkalinity in the surface when coupled to NEMO within UKESM1, therefore possibly this parameter would be optimised in the opposite direction to seen here.

7.4.3 The influence of ocean biogeochemical parameter optimisation on future predictions

When assessing the impact of biogeochemical parameter optimisation on MEDUSA's response to future carbon emission scenarios, we found it made little difference. This is unsurprising as most of the ocean's CO_2 uptake is via the solubility pump, therefore perturbing the biogeochemical parameters which mainly influence the biological pump, as we have done here, will have little influence on future oceanic CO_2 uptake. With this in mind, a change in ocean pH would influence the solubility pump, as carbon dioxide is less soluble in acidic water, however none of the optimised parameters influenced the future predictions of surface ocean pH, as this is controlled more by inorganic chemistry.

The main biogeochemical parameter which, when optimised, induced a change in the future predicted ocean behaviour was the $CaCO_3$:POC export rain ratio. This parameter was increased during optimisation, which increased the amount of $CaCO_3$ within sinking POC. This resulted in a greater increase in NPP by the year 2100, and a greater decrease in the downward flux of calcite by the year 2100.

Here we have shown MEDUSA's response to future carbon emission scenarios, but in reality the ocean is more complex, and therefore MEDUSA will not capture the full ocean response to carbon emissions. For example, if plankton characteristics are influenced by climate change, such as size and carbon to nutrient ratios, then this would impact carbon sequestration.

7.5 Conclusions

This chapter applied the derivative-free optimisation algorithm DFO-LS to MEDUSA (Yool et al., 2013), a global ocean biogeochemical model of intermediate complexity with 15 prognostic tracers, which was run with an "offline" general circulation provided by the UVic ocean model (Khatiwala et al., 2019). DFO-LS was used to tune 9 different biogeochemical parameters within various optimisation experiments, unfortunately with less reliable success than when the same optimisation algorithm was previously applied to a less complex ocean biogeochemical model in Chapter 3. The modelled distribution of silicate was improved by an increase in diatom abundance and hence opal transport to the deeper ocean, which was induced by favouring diatom growth and an unrealistic reduction in non-diatom abundance, resulting in a modelled global NPP almost half that of observations. Distributions of dissolved oxygen and alkalinity were also improved by certain parameter tuning, however this may be specific to the UVic circulation, and may not necessarily improve simulation of these tracers by MEDUSA when coupled to the NEMO circulation (Madec, 2008), as it is within the Earth System Model UKESM1 (Yool et al., 2021). It was also found that the optimisation of the selected ocean biogeochemical parameters had little influence on the ocean's response to future carbon emission scenarios, which is to be expected as it is the solubility pump through which most of the ocean's carbon uptake occurs, not the biological pump.

To better improve the optimisation of MEDUSA, without sacrificing the global non-diatom population, one could calibrate to more observational variables, such as zooplankton abundance. One could also weight the importance of each observational variable according to a covariance matrix, with relative weightings on the diagonal determined by the sensitivity of the misfit function to each observational variable (such as done in Tett et al. (2017)).

In particular, the observations of NPP used for calibration could be more heavily weighted. This would more forcibly constrain the reduction in non-diatom NPP caused by optimising MEDUSA to silicate observations. However, after some preliminary investigative optimisation experiments with this heavier weighting

(see Appendix E) we found DFO-LS struggled to reduce the misfit to observed distributions of NPP. This may have been due to the greater variability and sharp gradients in NPP, which are less pronounced in other observations such as dissolved nutrients and oxygen. Therefore, instead of attempting to reduce the misfit between each model grid point to equivalent observations, calibration to just the global total NPP may be more successful.

Future work could also involve applying this optimisation method to MEDUSA coupled to the NEMO circulation, once NEMO transport matrices have been extracted and made available.

7.6 Code Availability

The base TMM code are available to download from `https://doi.org/10.5281/zenodo.1246300` and the transport matrices used available to download from `http://kelvin.earth.ox.ac.uk/spk/Research/TMM/TransportMatrixConfigs/`. The TMM interface to MEDUSA-2.0 (Yool et al., 2013) will soon be submitted (Khatiwala et al., 2021). The OptClim optimisation framework used to couple any climate model to any optimisation algorithm is available at `https://doi.org/10.5281/zenodo.5517610`.

8
Conclusions

Earth System Models, and the various climate components contained within, are an important tool for understanding the complex natural world we live in. They allow us to predict future climate change, providing decision makers with necessary insight into how to limit global warming. The ocean biogeochemical component in particular is very important, as the oceans absorb a huge amount of carbon dioxide and transports much of it to the deep ocean where it can be kept out of contact with the atmosphere for thousands of years. The biological, physical and chemical processes which carry this out are complex, and global ocean biogeochemical models are often heavily parameterised. Due to the long model spin up time required for the model to reach an equilibrated state after each parameter re-configuration (Khatiwala et al., 2005; Wunsch and Heimbach, 2008), ocean biogeochemical parameter optimisation is a notoriously laborious and therefore often subjective step in model development (Hourdin et al., 2017). Efficient methods to improve these models are thus desirable. The overarching aim of this thesis is to first investigate such methods, and then apply them to the global ocean biogeochemical model MEDUSA (Yool et al., 2013), the chosen ocean biogeochemical component of the Earth System Model UKESM1 (Sellar et al., 2019). It is hoped that this thesis will provide a useful guide of what to (and what not to) do when investigating and improving parameterisations within such computationally expensive models, and serve as one more step toward systematic model calibration becoming common practice within the ocean biogeochemical modelling community.

This thesis starts off by investigating the global ocean biogeochemical model MOPS (Kriest and Oschlies, 2013). It is less computationally expensive than MEDUSA, yet provides a good representation of global oceanic tracer distributions. MOPS was therefore a good tool on which to test the efficiency of parameter optimisation methods before moving on to MEDUSA in the latter sections of the thesis. Also, the year before this PhD project began in 2018 MOPS, coupled to an "offline" general circulation model using the Transport Matrix Method (TMM; Khatiwala et al., 2005; Khatiwala, 2007, 2018), had been the first global ocean biogeochemical model to have several of its biogeochemical parameters systematically calibrated

to global observations (Kriest et al., 2017), therefore providing a "benchmark" to improve upon. Kriest et al. (2017) optimised 6 parameters within MOPS using the optimisation algorithm CMA-ES (Hansen, 2016). Chapter 3 of this thesis compared the efficiency and performance of CMA-ES with another derivative-free optimisation algorithm, DFO-LS, to see if one could optimise MOPS more efficiently. This was carried out by calibrating to synthetic observations only, whereby the global optimal parameter configuration was known and hence the success of the optimiser could easily be determined.

DFO-LS was found to have a significantly lower computational cost when compared with CMA-ES, between one and two orders of magnitude, which is important considering how computationally expensive these models are. DFO-LS exploits more information when minimising the misfit function, and therefore has more scope for reducing the misfit faster than CMA-ES. However, as DFO-LS is more of a local optimiser than CMA-ES, it has to be paired with a globalising method such as starting from different initial points in parameter space, but these can easily be run in parallel. The two optimisers CMA-ES and DFO-LS are very different and when comparing their performance one needs to consider the goal of the optimisation experiment. If one is looking for the precise absolute best parameter configuration at the global minimum of the misfit function, regardless of the computational cost and time, then the global evolutionary CMA-ES algorithm will in most cases perform better than DFO-LS. However, more realistically in the ocean biogeochemical (and wider climate) modelling community, a trade-off between misfit reduction and computational expense must be balanced. In that case the local deterministic optimiser DFO-LS is likely to be more appropriate than CMA-ES to find a significant decrease in the misfit function with a lower computational cost and time. This may not provide the globally optimal parameter configuration for the model, but will still lead to a significant improvement within the computational restraints associated with large parameterised computational models.

In Chapter 3 both optimisers had particular trouble when calibrating one specific biogeochemical parameter, which was determined to be caused by the misfit func-

tion's low sensitivity to this parameter. Therefore this highlighted the importance of carrying out a parameter sensitivity study prior to optimisation, to aid screening out insensitive parameters, and to enhance understanding of the influence each parameter has on the misfit function and biogeochemical model. This led to the work reported in Chapter 4, where a computationally inexpensive global Sobol' (Sobol', 1993; Sobol, 2001) sensitivity analysis of parameters within MOPS was carried out. The calculated Sobol' indices indicated the fraction of misfit variance caused by each parameter, both individually and allowing for parameter interactions, across the entire parameter space. They are computationally expensive to calculate, needing thousands of function evaluations. Therefore, they were calculated from a Kriging surrogate formulation (Krige, 1951) of the misfit function instead of on the full global ocean biogeochemical model. The training sample size required by the surrogate to still result in reliable Sobol' indices was as few as $n \times 8$ (where n = number of parameters), which of course could be run in parallel to reduce the sequential computational time required. The initial expense of these samples could be further reduced by shortening the ocean biogeochemical model spin up from 3000 years to 2000 years, while still resulting in reliable Sobol' indices. This method provides easier parameter screening prior to ocean biogeochemical optimisation studies, which will reduce the misfit function dimensionality and computational expense of model calibration. It also indicated the length of the ocean biogeochemical model spin up could be reduced and still produce correct sensitivity results, which was investigated further in Chapter 5.

The promising results of Chapter 3 suggested that it may be feasible to apply DFO-LS to more complex models. However, even with fast "offline" methods such as the TMM, it may not be practical to evaluate the misfit function for such models more than a few dozen times. Therefore, in Chapter 5 another method of reducing computational time during optimisation was investigated, by shortening the spin up length of the ocean biogeochemical model during optimisation. This was looked into briefly in Chapter 4 where it was found parameter sensitivity results became inaccurate when the spin up length was less than 2000 years. However, this may not

be the case for model calibration, and if one could optimise a less computationally expensive non-equilibrated model state then each iteration of the optimisation process would be quicker, significantly reducing the computational expense of the entire calibration.

This was investigated in Chapter 5 using MOPS, whereby the equilibrated model misfits were compared to non-equilibrated model misfits of various spin up lengths. If there was a simple, monotonic relationship between the two then it might be possible to perform optimisation without running the model to equilibrium. However this was not the case, as instead a non-monotonic relationship was found, causing optimisation of any spin up less than 2000 years to result in parameters which were not optimal for the equilibrated spin up. This was due to optimising parameters which caused the model to have increased vertical transport of organic matter, and hence increased remineralised nutrients to the deep global oceanic conveyor belt (Broecker, 1991; Kwon and Primeau, 2006; Kwon et al., 2009). A longer spin up was thus needed to transport the tracers correctly throughout the global ocean. This problem was masked when initiating the model with target tracer distributions, because spin up times shorter than 2000 years did not largely deviate the deeper ocean from it's target optimised state. Therefore, one should be cautious when utilising shortened ocean biogeochemical model spin ups (e.g. Priess et al., 2013b; Séférian et al., 2016). One might conclude from this that oceanic spin up times could be reduced to 2000 years instead of the more typical 3000 years (or longer). However, this must be carefully considered on a case-by-case basis as this 2000 year suggestion is dependent on the selected parameters to be tuned. Parameters influencing only the surface ocean require a shorter spin up to reach equilibrium than parameters that influence the flux of organic matter to the deeper ocean, where oceanic timescales are longer.

After first investigating efficient methods of ocean biogeochemical parameter optimisation the computationally less expensive model MOPS, this PhD project could then move on to apply some of these methods to the more complex model MEDUSA, which had yet to be systematically calibrated to observations. Due to the findings of

Chapter 3 whereby the success of an optimiser can be influenced by how sensitive the misfit function is to the parameters being tuned, it was decided that a sensitivity study should first be carried out on various biogeochemical parameters within MEDUSA (Chapter 6) prior to optimisation (Chapter 7). MEDUSA has historically always been coupled to the general circulation model NEMO (Madec, 2008), which vastly increases the computational expense of the biogeochemical model. Therefore here MEDUSA was coupled to an "offline" circulation via the TMM. Specifically, already available transport matrices from the University of Victoria Earth System Model (Weaver et al., 2001; Eby et al., 2009; Khatiwala et al., 2019) were used.

The aim of Chapter 6 was to inform the choice of which parameters to calibrate, and to better understand the various biogeochemical controls within MEDUSA. Each parameter was individually perturbed by +10% of their feasible range and their influence on 16 simulated fields was assessed relative to the default simulation, including variables such as dissolved nutrients, oxygen and carbon, alkalinity, and plankton abundance. Parameters which had the strongest influence on MEDUSA were those which control plankton abundance, particularly diatoms and meso-zooplankton, as they heavily influence the transport of organic matter to the deeper ocean and hence the global distribution of dissolved nutrients, at least in MEDUSA. The parameters were ordered according to a sensitivity ranking, which will be useful for future MEDUSA parameter optimisation studies, as it will help screen out parameters which may be difficult to optimise, and highlight which are the more important ones to calibrate.

Finally, in Chapter 7, DFO-LS was applied to optimise MEDUSA, calibrating several biogeochemical parameters previously investigated in Chapter 6. Unfortunately, DFO-LS had less reliable success than previously seen in Chapter 3, recovering fewer known optimal parameter values when calibrating to synthetic observations. When calibrating to real ocean observations, the modelled distribution of silicate was improved by an increase in diatom abundance and hence opal transport to the deeper ocean. This was induced by favouring diatom growth and an unrealistic reduction in non-diatom abundance, resulting in a modelled global NPP almost

half that of observations. Distributions of dissolved oxygen and alkalinity were also improved, however this may be specific to the UVic circulation, and therefore may not necessarily lead to similar improvements for other circulations.

This thesis describes several methods attempting to improve the efficiency (and therefore encourage more) of systematic parameter optimisation of global ocean biogeochemical models. Some were more successful than others. Suggested future work to expand on this thesis would be to better improve the optimisation of ocean biogeochemical models like MEDUSA, so as to not optimise certain tracers at the detriment of others, such as was seen with the global non-diatom population modelled by optimised configurations of MEDUSA in Chapter 7. To achieve this one can include more types of observations to calibrate the model to, for example, phytoplankton and zooplankton abundance data, with an additional global NPP constraint, which would better constrain the total zooplankton grazing and resulting reduction in non-diatoms (as discussed in Section 7.5).

To have more significance in the earth system modelling community, future work could involve applying the optimisation method used here to MEDUSA coupled to the NEMO circulation, once NEMO transport matrices have been extracted and made available, as it is the NEMO circulation which MEDUSA is coupled to within the Earth System Model UKESM1 (Sellar et al., 2019). The many lessons learned during this thesis allow the author to make several recommendations and suggestions for this future work. Firstly, compile all observational data that will best constrain the model parameters one might choose to optimise. For example, if these include phytoplankton and/or zooplankton parameters, then include observations of plankton abundance and primary production. If the observations have poor spatial coverage, instead of interpolating onto the model domain, keep as scattered data and interpolate the model onto the scattered observations when calculating the misfit. Secondly, formulate the misfit in a way that best suits the problem. If a lot of the parameters mainly influence the surface ocean, then do not volume-weight the misfit to the deeper ocean. Thirdly, determine the parameter bounds within which lie feasible parameter values, based on observations and theory. These bounds

will influence the misfit's sensitivity to this parameter, and hence how quickly an optimiser may calibrate it. Then carry out a sensitivity study to screen out less influential parameters, or to inform a modification of parameter bounds or a re-formulation of the misfit so to increase the misfit's sensitivity to certain parameters. If calibrating the surface ocean is the dominant goal, then initiate NEMO-MEDUSA with observed tracer distributions to allow a shortened biogeochemical model spin up. If one wishes to also calibrate the deeper ocean, one can still initiate from observed distributions and shorten the biogeochemical model spin up, however note that shorter than 2000 years will reduce how much the deeper ocean is being calibrated to the observations. Finally, use an optimisation algorithm requiring few evaluations of the model (e.g. DFO-LS) to find several "good" parameter configurations, then investigate these configurations further to check variables not used for calibration still look sensible when compared to observations. The best of these optimised configurations of NEMO-MEDUSA could then be used within UKESM1.

Bibliography

Afzal, A., Kim, K. Y., and Seo, J. W.: Effects of latin hypercube sampling on surrogate modeling and optimization, International Journal of Fluid Machinery and Systems, 10, 240–253, https://doi.org/10.5293/IJFMS.2017.10.3.240, 2017.

Aldama-Campino, A., Fransner, F., Ödalen, M., Groeskamp, S., Yool, A., Döös, K., and Nycander, J.: Meridional Ocean Carbon Transport, Global Biogeochemical Cycles, 34, https://doi.org/10.1029/2019GB006336, 2020.

Armstrong, R. A., Lee, C., Hedges, J. I., Honjo, S., and Wakeham, S. G.: A new, mechanistic model for organic carbon fluxes in the ocean based on the quantitative association of POC with ballast minerals, Deep-Sea Research Part II: Topical Studies in Oceanography, 49, 219–236, https://doi.org/10.1016/S0967-0645(01)00101-1, 2002.

Australian Government Bureau of Meteorology: Greenhouse gases, URL `http://www.bom.gov.au/state-of-the-climate/2016/greenhouse-gas-levels.shtml`, 2016.

Baker, C. A., Henson, S. A., Cavan, E. L., Giering, S. L., Yool, A., Gehlen, M., Belcher, A., Riley, J. S., Smith, H. E., and Sanders, R.: Slow-sinking particulate organic carbon in the Atlantic Ocean: Magnitude, flux, and potential controls, Global Biogeochemical Cycles, 31, 1051–1065, https://doi.org/10.1002/2017GB005638, 2017.

Behrenfeld, M. J. and Falkowski, P. G.: Photosynthetic rates derived from satellite-based chlorophyll concentration, Limnology and Oceanography, 42, 1–20, 1997.

Broecker, W. S.: The great ocean conveyor, Oceanography, pp. 79–89, https://doi.org/doi:10.5670/oceanog.1991.07, 1991.

Broecker, W. S. and Peng, T. H.: Interhemispheric transport of carbon dioxide by ocean circulation, Nature, 356, 587–589, https://doi.org/10.1038/356587a0, 1992.

Broecker, W. S., Sutherland, S., and Peng, T. H.: A possible 20th-century slowdown of southern ocean deep water formation, Science, 286, 1132–1135, https://doi.org/10.1126/science.286.5442.1132, 1999.

Buesseler, K. O., Boyd, P. W., Black, E. E., and Siegel, D. A.: Metrics that matter for assessing the ocean biological carbon pump, Proceedings of the National Academy of Sciences of the United States of America, 117, 9679–9687, https://doi.org/10.1073/pnas.1918114117, 2020.

Burmaster, D. E.: The unsteady continuous culture of phosphate-limited Monochrysis lutheri droop: Experimental and theoretical analysis, Journal of Experimental Marine Biology and Ecology, 39, 167–186, https://doi.org/10.1016/0022-0981(79)90012-1, 1979.

Cabré, A., Marinov, I., Bernardello, R., and Bianchi, D.: Oxygen minimum zones in the tropical Pacific across CMIP5 models: Mean state differences and climate change trends, Biogeosciences, 12, 5429–5454, https://doi.org/10.5194/bg-12-5429-2015, 2015.

Cartis, C., Roberts, L., and Sheridan-Methven, O.: Escaping local minima with derivative-free methods: a numerical investigation, arXiv:1812.11343v2, pp. 1–26, URL `http://arxiv.org/abs/1812.11343`, 2018.

Cartis, C., Fiala, J., Marteau, B., and Roberts, L.: Improving the flexibility and robustness of model-based derivative-free optimization solvers, ACM Transactions on Mathematical Software, 45, 1–35, https://doi.org/10.1145/3338517, 2019.

Chen, B. and Smith, S. L.: CITRATE 1.0 : Phytoplankton continuous trait-distribution model with one-dimensional physical transport applied to the Northwest Pacific, Geoscientific Model Development, 11, 467–495, https://doi.org/10.5194/gmd-11-467-2018, 2018.

Chisholm, S. W.: Stirring times in the Southern Ocean, Nature, 407, 685–687, https://doi.org/https://doi.org/10.1038/35037696, 2000.

Conn, A. R., Scheinberg, K., and Vicente, L. N.: Introduction to Derivative-Free Optimization, SIAM, https://doi.org/https://doi.org/10.1137/1.9780898718768, 2009.

DeVries, T.: The oceanic anthropogenic CO2 sink: Storage, air-sea fluxes, and transports over the industrial era, Global Biogeochemical Cycles, 28, pp631–647, https://doi.org/10.1002/2013GB004739, 2014.

Devries, T., Liang, J. H., and Deutsch, C.: A mechanistic particle flux model applied to the oceanic phosphorus cycle, Biogeosciences, 11, 5381–5398, https://doi.org/10.5194/bg-11-5381-2014, 2014.

Droop, M. R.: The nutrient status of algal cells in continuous culture, Journal of the Marine Biological Association of the United Kingdom, 54, 825–855, https://doi.org/10.1017/S002531540005760X, 1974.

Dutkiewicz, S., Follows, M. J., and Parekh, P.: Interactions of the iron and phosphorus cycles: A three-dimensional model study, Global Biogeochemical Cycles, 19, 1–22, https://doi.org/10.1029/2004GB002342, 2005.

Eby, M., Zickfeld, K., Montenegro, A., Archer, D., Meissner, K. J., and Weaver, A. J.: Lifetime of anthropogenic climate change: Millennial time scales of potential CO2 and surface temperature perturbations, Journal of Climate, 22, 2501–2511, https://doi.org/10.1175/2008JCLI2554.1, 2009.

Edwards, A. M. and Yool.A: The role of higher predation in plankton population models, Journal of Plankton Research, 22, 1085–1112, https://doi.org/10.1093/plankt/22.6.1085, 2000.

Evans, G. T.: Defining misfit between biogeochemical models and data sets, Journal of Marine Systems, 40-41, 49–54, https://doi.org/10.1016/S0924-7963(03)00012-5, 2003.

Fennel, K., Losch, M., Schroter, J., and Wenzel, M.: Testing a marine ecosystem model: sensitivity analysis and parameter optimization, Journal of Marine Systems, 28, 45–63, URL papers2://publication/uuid/90E701CC-93EE-430D-8932-0DBD847699BF, 2001.

Garcia, H., Weathers, K., Paver, C., Smolyar, I., Boyer, T., Locarnini, R., Zweng, M., Mishonov, A., Baranova, O., Seidov, D., and Reagan, J.: World Ocean Atlas 2018, Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Dissolved Oxygen Saturation, A. Mishonov Technical Editor. NOAA Atlas NESDIS 83, p. 38, 2018a.

Garcia, H., Weathers, K., Paver, C., Smolyar, I., Boyer, T., Locarnini, R., Zweng, M., Mishonov, A., Baranova, O., Seidov, D., and Reagan, J.: World Ocean Atlas 2018. Vol. 4: Dissolved Inorganic Nutrients (phosphate, nitrate and nitrate+nitrite, silicate), A. Mishonov Technical Editor, NOAA Atlas NESDIS 84, p. 35, 2018b.

Gnanadesikan, A., Dunne, J. P., Key, R. M., Matsumoto, K., Sarmiento, J. L., Slater, R. D., and Swathi, P. S.: Oceanic ventilation and biogeochemical cycling: Understanding the physical mechanisms that produce realistic distributions of tracers and productivity, Global Biogeochemical Cycles, 18, 1–17, https://doi.org/10.1029/2003GB002097, 2004.

Gregg, W. W., Ginoux, P., Schopf, P. S., and Casey, N. W.: Phytoplankton and iron: Validation of a global three-dimensional ocean biogeochemical model, Deep-Sea Research Part II: Topical Studies in Oceanography, 50, 3143–3169, https://doi.org/10.1016/j.dsr2.2003.07.013, 2003.

Gregoire, L. J., Valdes, P. J., Payne, A. J., and Kahana, R.: Optimal tuning of a GCM using modern and glacial constraints, Climate Dynamics, 37, 705–719, https://doi.org/10.1007/s00382-010-0934-8, 2011.

Gruber, N. and Deutsch, C. A.: Redfield's evolving legacy, Nature Geoscience, 7, 853–855, https://doi.org/10.1038/ngeo2308, URL http://dx.doi.org/10.1038/ngeo2308, 2014.

Hansen, N.: The CMA Evolution Strategy: A Tutorial, arXiv preprint arXiv:1604.00772, URL http://arxiv.org/abs/1604.00772, 2016.

Hartley, H. O.: The Modified Gauss-Newton Method for the Fitting of Non-Linear Regression Functions by Least Squares, Technometrics, 3, 269–280, 1961.

Hemmings, J. C. P., Challenor, P. G., and Yool, A.: Mechanistic site-based emulation of a global ocean biogeochemical model (MEDUSA 1.0) for parametric analysis and calibration: An application of the Marine Model Optimization Testbed (MarMOT 1.1), Geoscientific Model Development, 8, 697–731, https://doi.org/10.5194/gmd-8-697-2015, 2015.

Henson, S. A., Sarmiento, J. L., Dunne, J. P., Bopp, L., Lima, I., Doney, S. C., John, J., and Beaulieu, C.: Detection of anthropogenic climate change in satellite records of ocean chlorophyll and productivity, Biogeosciences, 7, 621–640, https://doi.org/10.5194/bg-7-621-2010, 2010.

Henson, S. A., Sanders, R., Madsen, E., Morris, P. J., Le Moigne, F., and Quartly, G. D.: A reduced estimate of the strength of the ocean's biological carbon pump, Geophysical Research Letters, 38, 10–14, https://doi.org/10.1029/2011GL046735, 2011.

Henson, S. A., Sanders, R., and Madsen, E.: Global patterns in efficiency of particulate organic carbon export and transfer to the deep ocean, Global Biogeochemical Cycles, 26, GB1028, https://doi.org/10.1029/2011GB004099, 2012.

Henson, S. A., Yool, A., and Sanders, R.: Global Biogeochemical Cycles carbon export : A model study, Global Biogeochemical Cycles, 29, 33–45, https://doi.org/10.1002/2014GB004965.Received, 2015.

Holzer, M., Primeau, F. W., DeVries, T., and Matear, R.: The Southern Ocean silicon trap: Data-constrained estimates of regenerated silicic acid, trapping efficiencies, and global transport paths, Journal of Geophysical Research: Oceans, 119, 313–331, https://doi.org/10.1002/2013JC009356, 2014.

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., Folini, D., Ji, D., Klocke, D., Qian, Y., Rauser, F., Rio, C., Tomassini, L., Watanabe, M., and Williamson, D.: The Art and Science of Climate Model Tuning, Bulletin of the American Meteorological Society, 98, 589–602, https://doi.org/10.1175/BAMS-D-15-00135.1, URL `http://journals.ametsoc.org/doi/10.1175/BAMS-D-15-00135.1`, 2017.

Ji, X., Liu, G., Gao, S., and Wang, H.: Parameter sensitivity study of the biogeochemical model in the China coastal seas, Acta Oceanologica Sinica, 34, 51–60, https://doi.org/10.1007/s13131-015-0762-0, 2015.

Joos, F., Plattner, G. K., and Stocker, T. F.: Global warming and marine carbon cycle feedbacks on future atmospheric CO2, Science, 284, 464–467, https://doi.org/10.1126/science.284.5413.464, 1999.

Joos, F., Roth, R., Fuglestvedt, J. S., Peters, G. P., Enting, I. G., Von Bloh, W., Brovkin, V., Burke, E. J., Eby, M., Edwards, N. R., Friedrich, T., Frölicher, T. L., Halloran, P. R., Holden, P. B., Jones, C., Kleinen, T., Mackenzie, F. T., Matsumoto, K., Meinshausen, M., Plattner, G. K., Reisinger, A., Segschneider, J., Shaffer, G., Steinacher, M., Strassmann, K., Tanaka, K., Timmermann, A., and Weaver, A. J.: Carbon dioxide and climate impulse response functions for the computation of greenhouse gas metrics: A multi-model analysis, Atmospheric Chemistry and Physics, 13, 2793–2825, https://doi.org/10.5194/acp-13-2793-2013, 2013.

Kasim, M. F., Watson-Parris, D., Deaconu, L., Oliver, S., Hatfield, P., Froula, D. H., Gregori, G., Jarvis, M., Khatiwala, S., Korenaga, J., Topp-Mugglestone, J., Viezzer, E., and Vinko, S. M.: Up to two billion times acceleration of scientific simulations with deep neural architecture search, arXiv:2001.08055, pp. 1–10, URL http://arxiv.org/abs/2001.08055, 2020.

Khatiwala, S.: A computational framework for simulation of biogeochemical tracers in the ocean, Global Biogeochemical Cycles, 21, 1–14, https://doi.org/10.1029/2007GB002923, 2007.

Khatiwala, S.: Fast spin up of Ocean biogeochemical models using matrix-free Newton-Krylov, Ocean Modelling, 23, 121–129, https://doi.org/10.1016/j.ocemod.2008.05.002, 2008.

Khatiwala, S.: Transport Matrix Method software for ocean biogeochemical simulations, https://doi.org/10.5281/ZENODO.1246300, URL https://zenodo.org/record/1246300, 2018.

Khatiwala, S., Visbeck, M., and Cane, M. A.: Accelerated simulation of passive tracers in ocean circulation models, Ocean Modelling, 9, 51–69, https://doi.org/10.1016/j.ocemod.2004.04.002, 2005.

Khatiwala, S., Primeau, F., and Hall, T.: Reconstruction of the history of anthropogenic CO2 concentrations in the ocean, Nature, 462, 346–349, https://doi.org/10.1038/nature08526, 2009.

Khatiwala, S., Primeau, F., and Holzer, M.: Ventilation of the deep ocean constrained with tracer observations and implications for radiocarbon estimates of ideal mean age, Earth and Planetary Science Letters, 325-326, 116–125, https://doi.org/10.1016/j.epsl.2012.01.038, 2012.

Khatiwala, S., Tanhua, T., Mikaloff Fletcher, S., Gerber, M., Doney, S. C., Graven, H. D., Gruber, N., McKinley, G. A., Murata, A., Ríos, A. F., and Sabine, C. L.:

Global ocean storage of anthropogenic carbon, Biogeosciences, 10, 2169–2191, https://doi.org/10.5194/bg-10-2169-2013, 2013.

Khatiwala, S., Schmittner, A., and Muglia, J.: Air-sea disequilibrium enhances ocean carbon storage during glacial periods, Science Advances, 5, 1–11, https://doi.org/10.1126/sciadv.aaw4981, 2019.

Khatiwala, S., Palmieri, J., Yool, A., Oliver, S., and Martin, A.: The Transport Matrix Method interface to the MEDUSA 2.0 global ocean biogeochemical model, To be submitted to Geoscientific Model Development, 2021.

Kidston, M., Matear, R., and Baird, M. E.: Phytoplankton growth in the Australian sector of the Southern Ocean, examined by optimising ecosystem model parameters, Journal of Marine Systems, 128, 123–137, https://doi.org/10.1016/j.jmarsys.2013.04.011, 2013.

Klaas, C. and Archer, D. E.: Association of sinking organic matter with various types of mineral ballast in the deep sea: Implications for the rain ratio, Global Biogeochemical Cycles, 16, 63–1–63–14, https://doi.org/10.1029/2001gb001765, 2002.

Knutti, R. and Sedláček, J.: Robustness and uncertainties in the new CMIP5 climate model projections, Nature Climate Change, 3, 369–373, https://doi.org/10.1038/nclimate1716, URL `http://dx.doi.org/10.1038/nclimate1716`, 2013.

Kriest, I.: Calibration of a simple and a complex model of global marine biogeochemistry, Biogeosciences Discussions, pp. 1–28, https://doi.org/10.5194/bg-2017-71, 2017.

Kriest, I. and Oschlies, A.: On the treatment of particulate organic matter sinking in large-scale models of marine biogeochemical cycles, Biogeosciences, 5, 55–72, https://doi.org/10.5194/bg-5-55-2008, 2008.

Kriest, I. and Oschlies, A.: Swept under the carpet: Organic matter burial decreases global ocean biogeochemical model sensitivity to remineralization length scale, Biogeosciences, 10, 8401–8422, https://doi.org/10.5194/bg-10-8401-2013, 2013.

Kriest, I. and Oschlies, A.: MOPS-1.0: Towards a model for the regulation of the global oceanic nitrogen budget by marine biogeochemical processes, Geoscientific Model Development, 8, 2929–2957, https://doi.org/10.5194/gmd-8-2929-2015, 2015.

Kriest, I., Khatiwala, S., and Oschlies, A.: Towards an assessment of simple global marine biogeochemical models of different complexity, Progress in Oceanography, 86, 337–360, https://doi.org/10.1016/j.pocean.2010.05.002, URL `http://dx.doi.org/10.1016/j.pocean.2010.05.002`, 2010.

Kriest, I., Oschlies, A., and Khatiwala, S.: Sensitivity analysis of simple global marine biogeochemical models, Global Biogeochemical Cycles, 26, 1–15, https://doi.org/10.1029/2011GB004072, 2012.

Kriest, I., Sauerland, V., Khatiwala, S., Srivastav, A., and Oschlies, A.: Calibrating a global three-dimensional biogeochemical ocean model (MOPS-1.0), Geoscientific Model Development, 10, 127–154, https://doi.org/10.5194/gmd-10-127-2017, 2017.

Kriest, I., Kähler, P., Koeve, W., Kvale, K., Sauerland, V., and Oschlies, A.: One size fits all ? - Calibrating an ocean biogeochemistry model for different circulations, Biogeosciences Discuss., https://doi.org/10.5194/bg-2020-9, 2020.

Krige, D.: A statistical approach to some mine valuation and allied problems on the Witwatersrand, Master's thesis, University of the Witwatersrand, South Africa, 1951.

Kuhn, A. M. and Fennel, K.: Evaluating ecosystem model complexity for the northwest North Atlantic through surrogate-based optimization, Ocean Modelling,

142, 101 437, https://doi.org/10.1016/j.ocemod.2019.101437, URL https://doi.org/10.1016/j.ocemod.2019.101437, 2019.

Kvale, K. F., Khatiwala, S., Dietze, H., Kriest, I., and Oschlies, A.: Evaluation of the transport matrix method for simulation of ocean biogeochemical tracers, Geoscientific Model Development, 10, 2425–2445, https://doi.org/10.5194/gmd-10-2425-2017, 2017.

Kwiatkowski, L., Yool, A., Allen, J. I., Anderson, T. R., Barciela, R., Buitenhuis, E. T., Butenschön, M., Enright, C., Halloran, P. R., Le Quéré, C., De Mora, L., Racault, M. F., Sinha, B., Totterdell, I. J., and Cox, P. M.: IMarNet: An ocean biogeochemistry model intercomparison project within a common physical ocean modelling framework, Biogeosciences, 11, 7291–7304, https://doi.org/10.5194/bg-11-7291-2014, 2014.

Kwon, E. Y. and Primeau, F.: Optimization and sensitivity study of a biogeochemistry ocean model using an implicit solver and in situ phosphate data, Global Biogeochemical Cycles, 20, 1–13, https://doi.org/10.1029/2005GB002631, 2006.

Kwon, E. Y. and Primeau, F.: Optimization and sensitivity of a global biogeochemistry ocean model using combined in situ DIC, alkalinity, and phosphate data, Journal of Geophysical Research: Oceans, 113, 1–23, https://doi.org/10.1029/2007JC004520, 2008.

Kwon, E. Y., Primeau, F., and Sarmiento, J. L.: The impact of remineralization depth on the air-sea carbon balance, Nature Geoscience, 2, 630–635, https://doi.org/10.1038/ngeo612, URL http://dx.doi.org/10.1038/ngeo612, 2009.

Lalli, C. M. and Parsons, T. R.: Biological oceanography, an introduction, Elsevier, second edn., 1997.

Lataniotis, C., Wicaksono, D., Marelli, S., and Sudret, B.: UQLab User Manual: Kriging (Gaussian Process Modeling) Report # UQLab-V1.3-105, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland, pp. 1–18, 2019.

Lauvset, S. K., Key, R. M., Olsen, A., van Heuven, S., Velo, A., Lin, X., Schirnick, C., Kozyr, A., Tanhua, T., Hoppema, M., Jutterström, S., Steinfeldt, R., Jeansson, E., Ishii, M., Pérez, F. F., Suzuki, T., and Watelet, S.: A new global interior ocean mapped climatology: the 1°x1° GLODAP version 2, Earth Syst. Sci. Data, 8, 325–340, https://doi.org/10.5194/essd-8-325-2016, 2016.

Lawrence, J., Popova, E., Yool, A., and Srokosz, M.: On the vertical phytoplankton response to an ice-free Arctic Ocean, J. Geophys. Res. Oceans, 120, 8571–8582, https://doi.org/10.1002/2015JC011180, 2015.

Li, X. and Primeau, F. W.: A fast Newton-Krylov solver for seasonally varying global ocean biogeochemistry models, Ocean Modelling, 23, 13–20, https://doi.org/10.1016/j.ocemod.2008.03.001, 2008.

Liang, F., Cheng, Y., and Lin, G.: Simulated stochastic approximation annealing for global optimization with a square-root cooling schedule, Journal of the American Statistical Association, 109, 847–863, https://doi.org/10.1080/01621459.2013.872993, 2014.

Loeppky, J. L., Sacks, J., and Welch, W. J.: Choosing the Sample Size of a Computer Experiment: A Practical Guide, Technometrics; Alexandria, 51, 366–376, 2009.

Lomas, M. W., Baer, S. E., Mouginot, C., Terpis, K. X., Lomas, D. A., Altabet, M. A., and Martiny, A. C.: Varying influence of phytoplankton biodiversity and stoichiometric plasticity on bulk particulate stoichiometry across ocean basins, Communications Earth & Environment, 2, 1–10, https://doi.org/10.1038/s43247-021-00212-9, 2021.

Lumpkin, R. and Speer, K.: Global ocean meridional overturning, Journal of Physical Oceanography, 37, 2550–2562, https://doi.org/10.1175/JPO3130.1, 2007.

Lüthi, D., Le Floch, M., Bereiter, B., Blunier, T., Barnola, J. M., Siegenthaler, U., Raynaud, D., Jouzel, J., Fischer, H., Kawamura, K., and Stocker, T. F.: High-resolution carbon dioxide concentration record 650,000-800,000 years before present, Nature, 453, 379–382, https://doi.org/10.1038/nature06949, 2008.

Madec, G.: NEMO Reference Manual, Ocean Dynamic Component: NEMO-OPA, Note du Pole de modélisation, Technical Report 27, Note du Pole de Modélisation, Institut Pierre Simmon Laplace, Paris, France, No. 27, ISSN 1288–1619, 2008.

Magozzi, S., Yool, A., Vander Zanden, H. B., Wunder, M. B., and Trueman, C. N.: Using ocean models to predict spatial and temporal variation in marine carbon isotopes, Ecosphere, 8, https://doi.org/10.1002/ecs2.1763, 2017.

Marelli, S. and Sudret, B.: UQLab: A framework for uncertainty quantification in Matlab, Proc. 2nd Int. Conf. on Vulnerability, Risk Analysis and Management (ICVRAM2014), Liverpool, United Kingdom, pp. 2554–2563, 2014.

Marelli, S., Lamas, C., Konakli, K., Mylonas, C., Wiederkehr, P., and Sudret, B.: UQLab User Manual: Sensitivity Analysis, , Report # UQLab-V1.3-106, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland, 2019.

Marinov, I., Gnanadesikan, A., Sarmiento, J. L., Toggweiler, J. R., Follows, M., and Mignone, B. K.: Impact of oceanic circulation on biological carbon storage in the ocean and atmospheric pCO2, Global Biogeochemical Cycles, 22, 1–15, https://doi.org/10.1029/2007GB002958, 2008.

Marsay, C. M., Sanders, R. J., Henson, S. A., Pabortsava, K., Achterberg, E. P., and Lampitt, R. S.: Attenuation of sinking particulate organic carbon flux through the mesopelagic ocean, Proceedings of the National Academy of Sciences of the United States of America, 112, 1089–1094, https://doi.org/10.1073/pnas.1415311112, 2015.

Marshall, J., Adcroft, A., Hill, C., Perelman, L., and Heisey, C.: A finite-volume, incompressible navier stokes model for studies of the ocean on parallel

computers, Journal of Geophysical Research, 102, 5753–5766, https://doi.org/10.1029/96JC02775, 1997.

Martin, J. D. and Simpson, T. W.: Use of kriging models to approximate deterministic computer models, AIAA Journal, 43, 853–863, https://doi.org/10.2514/1.8650, 2005.

Martin, J. H., Knauer, G. A., Karl, D. M., and Broenkow, W. W.: VERTEX: carbon cycling in the northeast Pacific, Deep Sea Research Part A, Oceanographic Research Papers, 34, 267–285, https://doi.org/10.1016/0198-0149(87)90086-0, 1987.

Matthes, K., Biastoch, A., Wahl, S., Harlaß, J., Martin, T., Brücher, T., Drews, A., Ehlert, D., Getzlaff, K., Krüger, F., Rath, W., Scheinert, M., Schwarzkopf, F., Bayr, T., Schmidt, H., and Park, W.: The Flexible Ocean and Climate Infrastructure Version 1 (FOCI1): Mean State and Variability, Geoscientific Model Development Discussions, 1, 1–53, https://doi.org/10.5194/gmd-2019-306, 2020.

McKay, M. D., Beckman, R. J., and Conover, W. J.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics, 21, 239–245, https://doi.org/10.1080/00401706.2000.10485979, 1979.

Meinshausen, M., Nicholls, Z., Lewis, J., Gidden, M., Vogel, E., Freund, M., Beyerle, U., Gessner, C., Nauels, A., Bauer, N., Canadell, J., Daniel, J., John, A., Krummel, P., Luderer, G., Meinshausen, N., Montzka, S., Rayner, P., Reimann, S., Smith, S., van den Berg, M., Velders, G., Vollmer, M., and Wang, H. J.: The SSP greenhouse gas concentrations and their extensions to 2500, Geoscientific Model Development Discussions, pp. 1–77, https://doi.org/10.5194/gmd-2019-222, 2019.

Melbourne-Thomas, J., Wotherspoon, S., Corney, S., Molina-Balari, E., Marini, O., and Constable, A.: Optimal control and system limitation in a Southern Ocean

ecosystem model, Deep-Sea Research Part II: Topical Studies in Oceanography, 114, 64–73, https://doi.org/10.1016/j.dsr2.2013.02.017, 2015.

Michaelis, L. and Menten, M.: Die Kinetik der Invertinwirkung, Biochemische Zeitschrift, 49, 333–369, 1913.

Monod, J.: The growth of bacterial cultures, Annual Reviews of Microbiology, 3, 371–394, 1949.

Müller, M. N.: On the genesis and function of coccolithophore calcification, Frontiers in Marine Science, 6, 1–5, https://doi.org/10.3389/fmars.2019.00049, 2019.

Nelson, D. M., Tréguer, P., Brzezinski, M. A., Leynaert, A., and Quéguiner, B.: Production and dissolution of biogenic silica in the ocean: Revised global estimates, comparison with regional data and relationship to biogenic sedimentation, Global Biogeochemical Cycles, 9, 359–372, https://doi.org/10.1029/95GB01070, 1995.

Niemeyer, D., Kriest, I., and Oschlies, A.: The effect of marine aggregate parameterisations on global biogeochemical model performance, Biogeosciences Discussions, pp. 1–29, https://doi.org/10.5194/bg-2019-122, 2019.

NOAA: Global Monitoring Laboratory: Monthly Average Mauna Loa CO2, URL `https://gml.noaa.gov/ccgg/trends/`, 2021.

Nocedal, J. and Wright, S. J.: Numerical Optimization, Springer, second edn., 2006.

O'Brien, T.: COPEPOD: The Global Plankton Database. An overview of the 2014 database contents, processing methods, and access interface., U.S. Dep. Commerce, NOAA Tech. Memo. NMFS-F/ST-37, p. 29p, 2014.

Oliver, S. E., Cartis, C., Kriest, I., Tett, S., and Khatiwala, S.: A derivative-free optimisation method for global ocean biogeochemical models, Geoscientific Model Development, In review, https://doi.org/https://doi.org/10.5194/gmd-2021-175, 2021.

Parekh, P., Dutkiewicz, S., Follows, M. J., and Ito, T.: Atmospheric carbon dioxide in a less duty world, Geophysical Research Letters, 33, 2–5, https://doi.org/10.1029/2005GL025098, 2006.

Piwonski, J. and Slawig, T.: Metos3D: The Marine Ecosystem Toolkit for Optimization and Simulation in 3-D - Part 1: Simulation Package v0.3.2, Geoscientific Model Development, 9, 3729–3750, https://doi.org/10.5194/gmd-9-3729-2016, 2016.

Ploug, H., Iversen, M. H., Koski, M., and Buitenhuis, E. T.: Production, oxygen respiration rates, and sinking velocity of copepod fecal pellets: Direct measurements of ballasting by opal and calcite, Limnology and Oceanography, 53, 469–476, https://doi.org/10.4319/lo.2008.53.2.0469, 2008.

Powell, M.: The BOBYQA algorithm for bound constrained optimization without derivatives, NA Report NA2009/06, p. 39, https://doi.org/10.1.1.443.7693, URL http://www6.cityu.edu.hk/rcms/publications/preprint26.pdf, 2009.

Priess, M., Koziel, S., and Slawigs, T.: Marine ecosystem model calibration through enhanced surrogate-based optimization, Advances in Intelligent Systems and Computing, 197, 193–208, URL http://dx.doi.org/10.1016/j.jocs.2013.04.001, 2013a.

Priess, M., Piwonski, J., Koziel, S., Oschlies, A., and Slawig, T.: Accelerated parameter identification in a 3D marine biogeochemical model using surrogate-based optimization, Ocean Modelling, 68, 22–36, https://doi.org/10.1016/j.ocemod.2013.04.003, URL http://dx.doi.org/10.1016/j.ocemod.2013.04.003, 2013b.

Prieur, C., Viry, L., Blayo, E., and Brankart, J. M.: A global sensitivity analysis approach for marine biogeochemical modeling, Ocean Modelling, 139, 101 402, https://doi.org/10.1016/j.ocemod.2019.101402, URL https://doi.org/10.1016/j.ocemod.2019.101402, 2019.

Primeau, F.: Characterizing transport between the surface mixed layer and the ocean interior with a forward and adjoint global ocean transport model, Journal of Physical Oceanography, 35, 545–564, https://doi.org/10.1175/JPO2699.1, 2005.

Primeau, F. and Deleersnijder, E.: On the time to tracer equilibrium in the global ocean, Ocean Science, 5, 13–28, https://doi.org/10.5194/os-5-13-2009, 2009.

Primeau, F. W. and Holzer, M.: The Ocean's Memory of the Atmosphere : Residence-Time and Ventilation-Rate Distributions of Water Masses, J. Phys. Oceanogr, pp. 1439–1456, https://doi.org/doi:10.1175/JPO2919.1, 2006.

Redfield, A. C.: in James Johnston Memorial Volume (ed. Daniel, R. J.), Univ. Press of Liverpool, pp. 176–192, 1934.

Ridgwell, A., Zondervan, I., Hargreaves, J. C., Bijma, J., and Lenton, T. M.: Assessing the potential long-term increase of oceanic fossil fuel CO 2 uptake due to CO2-calcification feedback, Biogeosciences, 4, 481–492, https://doi.org/10.5194/bg-4-481-2007, 2007.

Ridgwell, A., Schmidt, D. N., Turley, C., Brownlee, C., Maldonado, M. T., Tortell, P., and Young, J. R.: From laboratory manipulations to Earth system models: Scaling calcification impacts of ocean acidification, Biogeosciences, 6, 2611–2623, https://doi.org/10.5194/bg-6-2611-2009, 2009.

Rios, L. M. and Sahinidis, N. V.: Derivative-free optimization: A review of algorithms and comparison of software implementations, Journal of Global Optimization, 56, 1247–1293, https://doi.org/10.1007/s10898-012-9951-y, 2013.

Roberts, L.: Derivative-Free Algorithms for Nonlinear Optimisation Problems, Thesis, URL `https://ora.ox.ac.uk/objects/uuid:ec76e895-6eee-491a-88ed-b4ed10fa6003`, 2019.

Santner, T., Williams, B., and Notz, W.: The design and analysis ofcomputer experiments, Springer series in Statistics. Springer, 2003.

Sarmiento, J. L. and Bender, M.: Carbon biogeochemistry and climate change, Photosynthesis Research, 39, 209–234, https://doi.org/10.1007/BF00014585, 1994.

Sauerland, V., Löptien, U., Leonhard, C., Oschlies, A., and Srivastav, A.: Error assessment of biogeochemical models by lower bound methods, Geosci. Model Dev. Discuss., pp. 1–22, https://doi.org/10.5194/gmd-11-1181-2018, 2017.

Sauerland, V., Kriest, I., Oschlies, A., and Srivastav, A.: Multiobjective Calibration of a Global Biogeochemical Ocean Model Against Nutrients , Oxygen , and Oxygen Minimum Zones, Journal of Advances in Modeling Earth Systems, 11, https://doi.org/10.1029/2018MS001510, 2019.

Schartau, M. and Oschlies, A.: 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, 2003.

Schartau, M., Oschlies, A., and Willebrand, J.: Parameter estimates of a zero-dimensional ecosystem model applying the adjoint method, Deep-Sea Research Part II: Topical Studies in Oceanography, 48, 1769–1800, https://doi.org/10.1016/S0967-0645(00)00161-2, 2001.

Schlitzer, R. and Others: The GEOTRACES Intermediate Data Product 2017, Chemical Geology, 493, 210–223, https://doi.org/10.1016/j.chemgeo.2018.05.040, URL `https://doi.org/10.1016/j.chemgeo.2018.05.040`, 2018.

Séférian, R., Gehlen, M., Bopp, L., Resplandy, L., Orr, J. C., Marti, O., Dunne, J. P., Christian, J. R., Doney, S. C., Ilyina, T., Lindsay, K., Halloran, P. R., Heinze, C., Segschneider, J., Tjiputra, J., Aumont, O., and Romanou, A.: Inconsistent strategies to spin up models in CMIP5: Implications for ocean biogeochemical model performance assessment, Geoscientific Model Development, 9, 1827–1851, https://doi.org/10.5194/gmd-9-1827-2016, 2016.

Sellar, A. A., Jones, C. G., Mulcahy, J. P., Tang, Y., Yool, A., Wiltshire, A., O'Connor, F. M., Stringer, M., Hill, R., Palmieri, J., Woodward, S., de Mora, L.,

Kuhlbrodt, T., Rumbold, S. T., Kelley, D. I., Ellis, R., Johnson, C. E., Walton, J., Abraham, N. L., Andrews, M. B., Andrews, T., Archibald, A. T., Berthou, S., Burke, E., Blockley, E., Carslaw, K., Dalvi, M., Edwards, J., Folberth, G. A., Gedney, N., Griffiths, P. T., Harper, A. B., Hendry, M. A., Hewitt, A. J., Johnson, B., Jones, A., Jones, C. D., Keeble, J., Liddicoat, S., Morgenstern, O., Parker, R. J., Predoi, V., Robertson, E., Siahaan, A., Smith, R. S., Swaminathan, R., Woodhouse, M. T., Zeng, G., and Zerroukat, M.: UKESM1: Description and Evaluation of the U.K. Earth System Model, Journal of Advances in Modeling Earth Systems, 11, 4513–4558, https://doi.org/10.1029/2019MS001739, 2019.

Severijns, C. A. and Hazeleger, W.: Optimizing parameters in an atmospheric general circulation model, Journal of Climate, 18, 3527–3535, https://doi.org/10.1175/JCLI3430.1, 2005.

Shah, S. H. A. M., Primeau, F. W., Deleersnijder, E., and Heemink, A. W.: Tracing the ventilation pathways of the deep North Pacific Ocean using Lagrangian particles and eulerian tracers, Journal of Physical Oceanography, 47, 1261–1280, https://doi.org/10.1175/JPO-D-16-0098.1, 2017.

Shakun, J. D., Clark, P. U., He, F., Marcott, S. A., Mix, A. C., Liu, Z., Otto-Bliesner, B., Schmittner, A., and Bard, E.: Global warming preceded by increasing carbon dioxide concentrations during the last deglaciation, Nature, 484, 49–54, https://doi.org/10.1038/nature10915, 2012.

Shen, C., Shi, H., Liu, Y., Li, F., and Ding, D.: Discussion of skill improvement in marine ecosystem dynamic models based on parameter optimization and skill assessment, Chinese Journal of Oceanology and Limnology, 34, 683–696, https://doi.org/10.1007/s00343-016-5068-3, 2016.

Shi, Y., Gong, W., Duan, Q., Charles, J., Xiao, C., and Wang, H.: How parameter specification of an Earth system model of intermediate complexity influences its climate simulations, Progress in Earth and Planetary Science, 6, https://doi.org/10.1186/s40645-019-0294-x, 2019.

Siberlin, C. and Wunsch, C.: Oceanic tracer and proxy time scales revisited, Climate of the Past, 7, 27–39, https://doi.org/10.5194/cp-7-27-2011, 2011.

Skinner, L. C. and Shackleton, N. J.: An Atlantic lead over Pacific deep-water change across Termination I: Implications for the application of the marine isotope stage stratigraphy, Quaternary Science Reviews, 24, 571–580, https://doi.org/10.1016/j.quascirev.2004.11.008, 2005.

Sobol', I. M.: Sensitivity estimates for nonlinear mathematical models, Mathematical and Computer Modelling 1, pp. 407–414, 1993.

Sobol, I. M.: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Mathematics and Computers in Simulation, 55, 271–280, https://doi.org/10.1016/S0378-4754(00)00270-6, 2001.

Spitz, Y. H., Moisan, J. R., Abbott, M. R., and Richman, J. G.: Data assimilation and a pelagic ecosystem model: Parameterization using time series observations, Journal of Marine Systems, 16, 51–68, https://doi.org/10.1016/S0924-7963(97)00099-7, 1998.

Stammer, D., Ueyoshi, K., Köhl, A., Large, W. G., Josey, S. A., and Wunsch, C.: Estimating air-sea fluxes of heat, freshwater, and momentum through global ocean data assimilation, Journal of Geophysical Research C: Oceans, 109, 1–16, https://doi.org/10.1029/2003JC002082, 2004.

Sulpis, O., Jeansson, E., Dinauer, A., Lauvset, S. K., and Middelburg, J. J.: Calcium carbonate dissolution patterns in the ocean, Nature Geoscience, 14, 423–428, https://doi.org/10.1038/s41561-021-00743-y, URL `http://dx.doi.org/10.1038/s41561-021-00743-y`, 2021.

Tang, Y., Reed, P., Wagener, T., and Van Werkhoven, K.: Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation, Hydrology and Earth System Sciences, 11, 793–817, https://doi.org/10.5194/hess-11-793-2007, 2006.

Tett, S. F., Rowlands, D. J., Mineter, M. J., and Cartis, C.: Can top-of-atmosphere radiation measurements constrain climate predictions? Part I: Tuning, Journal of Climate, 26, 9367–9383, https://doi.org/10.1175/JCLI-D-12-00596.1, 2013.

Tett, S. F. B., Yamazaki, K., Mineter, M. J., Cartis, C., and Eizenberg, N.: Calibrating Climate Models Using Inverse Methods: Case studies with HadAM3, HadAM3P and HadCM3, Geoscientific Model Development Discussions, pp. 1–36, https://doi.org/10.5194/gmd-2016-305, URL https://www.geosci-model-dev-discuss.net/gmd-2016-305/, 2017.

Tian, R., Chen, C., Qi, J., Ji, R., Beardsley, R. C., and Davis, C.: Model study of nutrient and phytoplankton dynamics in the Gulf of Maine: patterns and drivers for seasonal and interannual variability, ICES Journal ofMarine Science, 72, 388–402, https://doi.org/doi:10.1093/icesjms/fsu090Original, 2015.

Turner, J. T.: Zooplankton fecal pellets, marine snow, phytodetritus and the ocean's biological pump, Progress in Oceanography, 130, 205–248, https://doi.org/10.1016/j.pocean.2014.08.005, URL http://dx.doi.org/10.1016/j.pocean.2014.08.005, 2015.

Vakil-Baghmisheh, M. T. and Navarbaf, A.: A modified very fast simulated annealing algorithm, 2008 International Symposium on Telecommunications, IST 2008, https://doi.org/10.1109/ISTEL.2008.4651272, 2008.

Volk, T. and Hoffert, M. I.: Ocean carbon pumps: analysis of relative strengths and efficiencies in ocean-driven atmospheric CO2 changes. In E. T. Sundquist, & W. S. Broecker (Eds.), The carbon cycle and atmospheric CO 2 : natural variations Archean to present., Chapman conference papers, 1985. American Geophysical Union; Geophysical Monograph, 32, 99–110, 1985.

Wang, C., Duan, Q., Gong, W., Ye, A., Di, Z., and Miao, C.: An evaluation of adaptive surrogate modeling based optimization with two benchmark problems, Environmental Modelling and Software, 60, 167–179, https://doi.org/10.1016/

j.envsoft.2014.05.026, URL http://dx.doi.org/10.1016/j.envsoft.2014.05.026, 2014.

Wang, S., Flipo, N., and Romary, T.: Time-dependent global sensitivity analysis of the C-RIVE biogeochemical model in contrasted hydrological and trophic contexts, Water Research, 144, 341–355, https://doi.org/10.1016/j.watres.2018.07.033, URL https://doi.org/10.1016/j.watres.2018.07.033, 2018.

Ward, B. A., Friedrichs, M. A. M., Anderson, T. R., and Oschlies, A.: Parameter optimisation techniques and the problem of underdetermination in marine biogeochemical models, Journal of Marine Systems, 81, 34–43, https://doi.org/10.1016/j.jmarsys.2009.12.005, 2010.

Ward, B. A., Schartau, M., Oschlies, A., Martin, A. P., Follows, M. J., and Anderson, T. R.: When is a biogeochemical model too complex? Objective model reduction and selection for North Atlantic time-series sites, Progress in Oceanography, 116, 49–65, https://doi.org/10.1016/j.pocean.2013.06.002, URL http://dx.doi.org/10.1016/j.pocean.2013.06.002, 2013.

Weaver, A. J., Eby, M., Wiebe, E. C., Ewen, T. L., Fanning, A. F., MacFadyen, A., Matthews, H. D., Meissner, K. J., Saenko, O., Schmittner, A., Yoshimori, M., Bitz, C. M., Holland, M. M., Duffy, P. B., and Wang, H.: The UVic earth system climate model: Model description, climatology, and applications to past, present and future climates, Atmosphere - Ocean, 39, 361–428, https://doi.org/10.1080/07055900.2001.9649686, URL https://doi.org/10.1080/07055900.2001.9649686, 2001.

Weber, T., Cram, J. A., Leung, S. W., DeVries, T., and Deutsch, C.: Deep ocean nutrients imply large latitudinal variation in particle transfer efficiency, Proceedings of the National Academy of Sciences, 113, 8606–8611, https://doi.org/10.1073/pnas.1604414113, 2016.

Williams, R. G. and Follows, M. J.: Ocean dynamics and the carbon cycle: Principles and mechanisms, Cambridge University Press, 2011.

Wunsch, C. and Heimbach, P.: How long to oceanic tracer and proxy equilibrium?, Quaternary Science Reviews, 27, 637–651, https://doi.org/10.1016/j.quascirev.2008.01.006, 2008.

Xiao, Y. and Friedrichs, M. A. M.: The assimilation of satellite-derived data into a one-dimensional lower trophic level marine ecosystemmodel, Journal of Geophysical Research: Oceans, pp. 2691– 2712, https://doi.org/10.1002/2013JC009433, 2014.

Xie, F., Tao, Z., Zhou, X., Lv, T., and Wang, J.: Spatial and temporal variations of particulate organic carbon sinking flux in global ocean from 2003 to 2018, Remote Sensing, 11, https://doi.org/10.3390/rs11242941, 2019.

Yang, B., Qian, Y., Lin, G., Leung, L. R., Rasch, P. J., Zhang, G. J., Mcfarlane, S. A., Zhao, C., Zhang, Y., Wang, H., Wang, M., and Liu, X.: Uncertainty quantification and parameter tuning in the cam5 zhang-mcfarlane convection scheme and impact of improved convection on the global circulation and climate, Journal of Geophysical Research Atmospheres, 118, 395–415, https://doi.org/10.1029/2012JD018213, 2013.

Yang, H. and Zhu, J.: Equilibrium thermal response timescale of global oceans, Geophysical Research Letters, 38, 1–5, https://doi.org/10.1029/2011GL048076, 2011.

Yool, A., Popova, E. E., and Anderson, T. R.: MEDUSA-1.0: A new intermediate complexity plankton ecosystem model for the global domain, Geoscientific Model Development, 4, 381–417, https://doi.org/10.5194/gmd-4-381-2011, 2011.

Yool, A., Popova, E. E., and Anderson, T. R.: MEDUSA-2.0: An intermediate complexity biogeochemical model of the marine carbon cycle for climate change and ocean acidification studies, Geoscientific Model Development, 6, 1767–1811, https://doi.org/10.5194/gmd-6-1767-2013, 2013.

Yool, A., Popova, E. E., and Coward, A. C.: Future change in ocean productivity: Is the Arctic the new Atlantic, Journal of Geophysical Research: Oceans, 120, 7771–7790, https://doi.org/10.1002/2015JC011167, 2015.

Yool, A., Martin, A. P., Anderson, T. R., Bett, B. J., Jones, D. O., and Ruhl, H. A.: Big in the benthos: Future change of seafloor community biomass in a global, body size-resolved model, Global Change Biology, 23, 3554–3566, https://doi.org/10.1111/gcb.13680, 2017.

Yool, A., Palmiéri, J., Jones, C. G., Sellar, A. A., de Mora, L., Kuhlbrodt, T., Popova, E. E., Mulcahy, J. P., Wiltshire, A., Rumbold, S. T., Stringer, M., Hill, R. S., Tang, Y., Walton, J., Blaker, A., Nurser, A. J., Coward, A. C., Hirschi, J., Woodward, S., Kelley, D. I., Ellis, R., and Rumbold-Jones, S.: Spin-up of UK Earth System Model 1 (UKESM1) for CMIP6, Journal of Advances in Modeling Earth Systems, 12, 1–36, https://doi.org/10.1029/2019MS001933, 2020.

Yool, A., Palmiéri, J., Jones, C., de Mora, L., Kuhlbrodt, T., Popova, E., Nurser, A. J. G., Hirschi, J., Blaker, A., Coward, A., Blockley, E., and Sellar, A.: Evaluating the physical and biogeochemical state of the global ocean component of UKESM1 in CMIP6 Historical simulations, Geoscientific Model Development, 14, 3437–3472, https://doi.org/https://doi.org/10.5194/gmd-14-3437-2021, 2021.

Zeebe, R. E. and Wolf-Gladrow, D.: CO2 in seawater: Equilibrium, kinetics, isotopes, Volume 65, 2001.

Zhang, T., Li, L., Lin, Y., Xue, W., Xie, F., Xu, H., and Huang, X.: An automatic and effective parameter optimization method for model tuning, Geoscientific Model Development, 8, 3579–3591, https://doi.org/10.5194/gmd-8-3579-2015, 2015.

Zhao, L., Wei, H., Xu, Y., and Feng, S.: An adjoint data assimilation approach for estimating parameters in a three-dimensional ecosystem model, Ecological Modelling, 186, 235–250, https://doi.org/10.1016/j.ecolmodel.2005.01.017, 2005.

Zou, L., Qian, Y., Zhou, T., and Yang, B.: Parameter tuning and calibration of RegCM3 with MIT-emanuel cumulus parameterization scheme over CORDEX East Asia domain, Journal of Climate, 27, 7687–7701, https://doi.org/10.1175/JCLI-D-14-00229.1, 2014.

Appendices

A

Investigative work to accompany Chapter 3

The work done here is an extension to that done in Chapter 3, describing avenues of research which were explored during that study to improve the quality or reduce computational expense of ocean biogeochemical model parameter calibration techniques. The work below shows (1) the precision to which the optimisation algorithm DFO-LS can get to a function minimum if allowed the computational freedom, and (2) whether the success of DFO-LS can be improved by combining with an initial sampling method.

A.1 DFO-LS initiated from close to the TWIN global minimum

A.1.1 Introduction

In Chapter 3 we investigated the success of the optimisation algorithm DFO-LS (Cartis et al., 2019) when optimising 6 parameter values within the ocean biogeochemical model MOPS (Kriest and Oschlies, 2015) towards a previous model configuration output with a known global optimum parameter configuration. In these optimisation experiments we started DFO-LS in the parameter space far away from the known target optimum values and allowed DFO-LS 70 evaluations to converge to a minimum. By the end of these 70 evaluations DFO-LS converged relatively close to 5/6 of the known optimal parameter values, and it is assumed that with more evaluations DFO-LS would only converge even closer to the optimal parameter configuration. However we did not provide evidence for this in Chapter 3, therefore here we show that DFO-LS can converge much closer to a minimum when given the chance.

A.1.2 Methodology

As in Chapter 3 the optimisation algorithm DFO-LS (Cartis et al., 2019) was applied to the ocean biogeochemical model MOPS (Kriest and Oschlies, 2015) coupled to the MITgcm (2.8 degree) ocean model (Marshall et al., 1997) via the Transport Matrix Method (Khatiwala et al., 2005; Khatiwala, 2007). The MOPS model was run for 3000 years with a known parameter configuration (MOPS-ref) containing the target

Parameters		RO2:P	IC	KPHY	muZOO	kZOO	b*	Evaluations (max/best)	Misfit (start/best)
Upper Bound		**200**	**48**	**0.5**	**4**	**10**	**1.8**		
Lower Bound		**150**	**4**	**0.0001**	**0.1**	**0**	**0.4**		
Target		**170**	**24**	**0.03125**	**2**	**3.2**	**0.858**		**0**
Experiment									
t6dC	Start	170.10000	24.01000	0.03225	2.01000	3.21000	0.85900	80	2.213E-03
	Optimised	170.00600	23.99821	0.03281	2.00075	3.20119	0.85798	80	2.557E-06

Table A.1: Table of parameter bounds, target parameter values, and optimisation results. Upper section shows parameter bounds and MOPS-ref target parameters to be recovered. Lower section shows the experiment results. Columns 2-7: (1st row) starting parameter values and (2nd row) optimised parameters. Column 8: (1st row) the maximum number of evaluations completed and (2nd row) the evaluation which provided the lowest or "best" global misfit. Column 9: (1st row) the starting global misfit and (2nd row) the lowest global misfit. Ocean biogeochemical parameter descriptions can be found in Chapter 3.

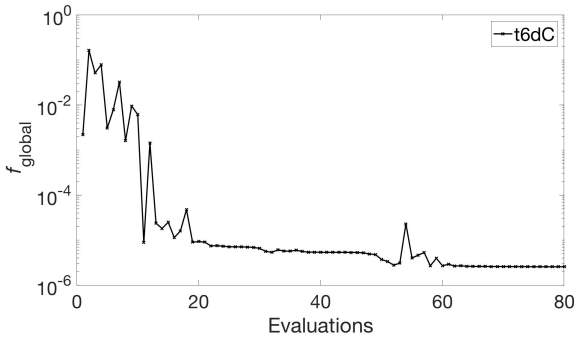

Figure A.1: The reduction in global misfit of MOPS to the MOPS-ref synthetic observations by DFO-LS for the optimisation experiment **t6dC** starting from very close to the target parameters. Note that every MOPS evaluation during the optimisation has been plotted.

parameter values (see Table A.1) to create synthetic observation of annual global oxygen, phosphate and nitrate. DFO-LS starts in one location of the 6-dimensional parameter space and tunes the 6 parameter values to minimise the misfit error (f_{global}) between the model output and the synthetic observations to recover the known optimal parameter values. Here in optimisation experiment **t6dC** DFO-LS was started very close to the known optimal parameter values in the parameter space, with the aim to show that DFO-LS can converge very closely to the optimum.

A.1.3 Results

Table A.1 shows the optimised results of the experiment **t6dC**, and Figure A.1 illustrates the misfit reduction throughout the optimisation progress. As DFO-LS

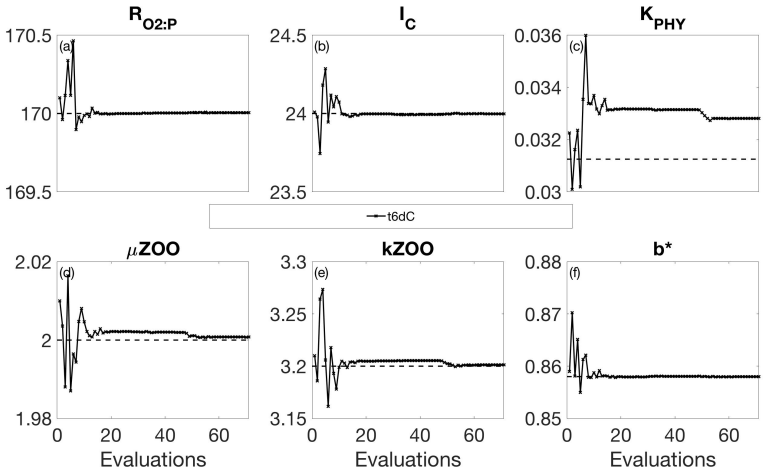

Figure A.2: Parameter tuning by DFO-LS for the experiment **t6dC** for the parameters (a) $R_{O_2:P}$, (b) I_C, (c) K_{PHY}, (d) μ_{ZOO}, (e) k_{ZOO} and (f) **b***. Also plotted are the MOPS-ref target parameter values (horizontal black dashed line). Note that every MOPS evaluation during the optimisation has been plotted.

was initiated very close to the target parameters in the parameter space, the starting misfit was very low at 2.2x10^{-3}. After an initial search of the parameter space DFO-LS rapidly reduced the misfit to 9.3x10^{-6} within 20 evaluations of MOPS. After this DFO-LS gradually continued to decrease the misfit to 2.6x10^{-6} after 80 evaluations.

Figure A.2 shows how DFO-LS calibrated the parameters during the optimisation. By 20 evaluations DFO-LS had converged to 6 parameter values, 3 of which are very close to the targets ($R_{O_2:P}$, I_C, **b***), 2 relatively close (μ_{ZOO}, k_{ZOO}) and 1 not close (K_{PHY}). By 80 evaluations μ_{ZOO} and k_{ZOO} were tuned to very close to the targets, while K_{PHY} was still far off.

A.1.4 Discussion

When initiating DFO-LS from very close to the global optimum, the optimiser managed to keep converging closer and closer to the optimal values of 5 out of the 6 parameters. This indicates that DFO-LS would have continued to decrease the misfit in the optimisation experiments done in Chapter 3 if it had been allowed to continue past 70 MOPS evaluations. However, as found in Chapter 3, the parameter K_{PHY} is difficult to optimise successfully because the misfit function is least sensitive

to this parameter. Due to the misfit function's low sensitivity to this parameter, one can argue this parameter is not important to the misfit function. Therefore, it may not be necessary to find its exact "true" value, as instead all values of K_{PHY} with the same low misfit are equally valid. However, if the formulation of the misfit were changed to better constrain this parameter, such as by increasing the weighting of surface observations (where this parameter has most influence), then the sensitivity to this parameter may increase and therefore be deemed important to optimise.

A.1.5 Conclusion

The work done here is an extension to that done in Chapter 3, to help show that DFO-LS can continue to converge closer and closer to the optimal parameter values within a global ocean biogeochemical model, if given the computational time to do so. This is useful information to have in the theoretical case when optimising towards synthetic observations where a misfit value of zero is possible, as DFO-LS can continue to converge closer to zero. In an applied case when optimising towards real oceanic observations, the ocean biogeochemical model will never be able to replicate the real ocean exactly and DFO-LS will not be able to reduce the misfit below a certain misfit threshold above zero. However the work done here indicates that DFO-LS would continue to converge closer and closer to this threshold. The user should determine when to terminate the optimiser by taking into account computational expense and time, observational noise and parameter precision, or allow DFO-LS to continue until misfit reduction stagnates.

A.2 Initiating DFO-LS from "good" starting locations

A.2.1 Introduction

In Chapter 4 we used the Latin Hypercube Sampling (LHS) strategy (McKay et al., 1979) to sample an ocean biogeochemical model's misfit function to synthetic observations within a 6-dimensional parameter space. We sampled the model MOPS (Kriest and Oschlies, 2015) 90 times with this sampling method, twice, and here

Region Number	Location
1	Northern North Pacific
2	Upper North Pacific
3	Lower North Pacific
4	Upper Equatorial Pacific
5	Lower Equatorial Pacific
6	Upper South Pacific
7	Lower South Pacific
8	Northern North Atlantic
9	Upper North Atlantic
10	Lower North Atlantic
11	Upper Equatorial Atlantic
12	Lower Equatorial Atlantic
13	Upper South Atlantic
14	Lower South Atlantic
15	North Western Indian
16	North Eastern Indian
17	South Indian
18	Southern Ocean: Subantarctic zone
19	Southern Ocean: Antarctic zone

Table A.2: Biomes.

we use the four parameter configurations resulting in the lowest global misfits from both of these sample sets "LHS1" and "LHS2" to better inform the starting locations for the optimiser DFO-LS. For every MOPS sample the global misfit value can be broken down into regional misfit values, and divided into specific oceanic tracers (oxygen, phosphate and nitrate). Here we show first show the regional misfit distribution for the LHS1 samples, then carry out twin optimisation experiments initiated from "good" starting points in parameter space determined by the samples yielding the lowest misfit values.

A.2.2 Methodology

In Chapters 3 and 4 we used the ocean biogeochemical model MOPS (Kriest and Oschlies, 2015) coupled to the MITgcm (2.8 degree) ocean model (Marshall et al.,

1997) via the Transport Matrix Method (Khatiwala et al., 2005; Khatiwala, 2007). In Chapter 3 we optimise using DFO-LS, initiated from random starting points in parameter space. Then in Chapter 4 we ran MOPS 90 times (twice) with different parameter configurations determined by the randomised LHS strategy, after which the model outputs of annual global oxygen, phosphate and nitrate were compared to the outputs of a previous model configuration. This was used to calculate the misfit between oxygen, phosphate and nitrate for 19 difference oceanic biome regions, which could then be condensed into a single global misfit value. See Section 3.1.5 for the misfit equations. Table A.2 describes the 19 biome regions.

In Chapter 3 we struggled to optimise the parameter K_{PHY}, due to the misfit function's low sensitivity to this parameter, as seen in Chapter 4. Here, in an attempt to better optimise the parameter K_{PHY}, we decided to better inform DFO-LS by finding a suitable starting point to initialise DFO-LS from. To do this the parameter space was sampled using the inexpensive strategy Latin hypercube Sampling (McKay et al., 1979) with the MATLAB function **lhsdesign**, as in Chapter 4. MOPS was evaluated and the individual and global misfits were calculated at 90 different locations in the parameter space according to the latin hypercube created using the random number seed 999 (LHS1), and then repeated for 90 different locations according to a second latin hypercube created using the random number seed 111 (LHS2). Four of the LHS1 sampled points resulting in the lowest global misfits were used to start DFO-LS from 4 "good" starting points (where 1 = location with the lowest misfit, and 4 = 4th lowest misfit), and these experiments are henceforth called "**t6d_LHS1_good1**", "**t6d_LHS1_good2**", "**t6d_LHS1_good3**", and "**t6d_LHS1_good4**". To determine the reliability and reproducibility of this method it was repeated with the sample set LHS2 and these experiments are henceforth called "**t6d_LHS2_good1**", etc.

A.2.3 Results

Figure A.3 breaks the results down into individual misfits for each of the 19 biome regions and 3 oceanic tracers. The largest misfits occur in the Pacific and North

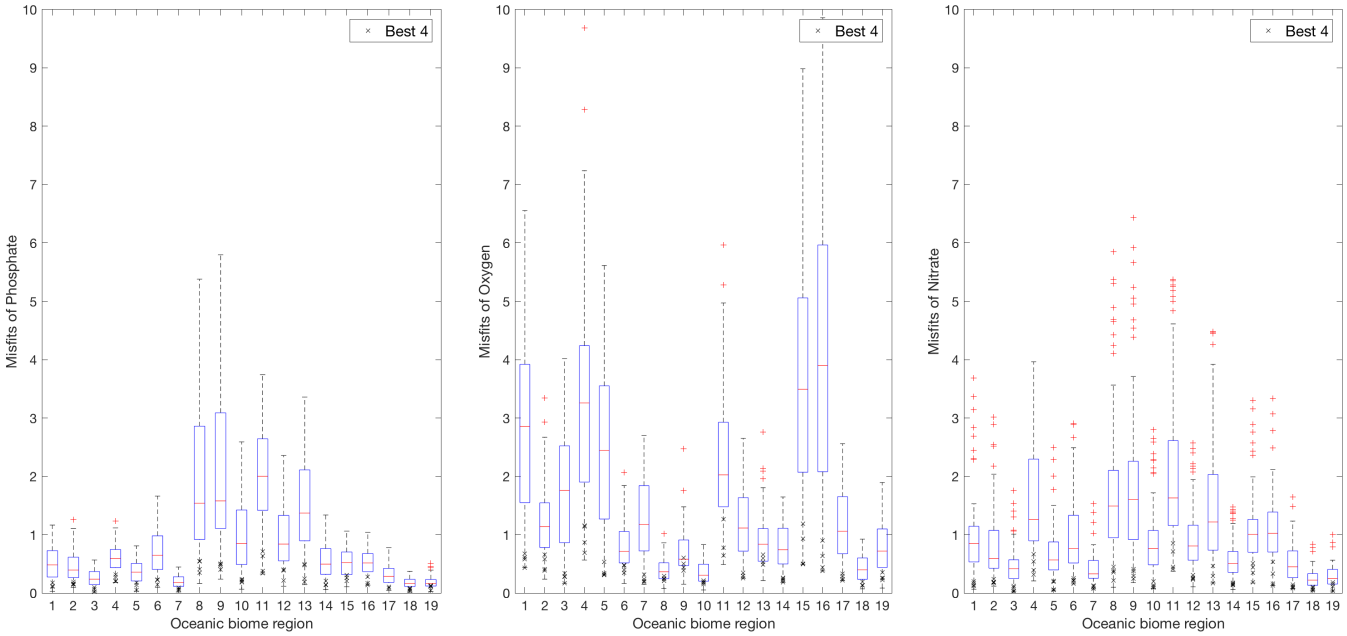

Figure A.3: Box-and-whisker plots of the (a) phosphate, (b) oxygen and (c) nitrate regional misfits for the 90 LHS1 samples. The red horizontal lines within each box indicate the median, the bottom and top edges show the 25th and 75th percentiles respectively, the whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted as red '+' symbols. The four samples with the lowest global misfits are shown by black crosses. Region numbers correspond to those in Table A.2.

Indian oceans for oxygen, and in the upper Atlantic for both phosphate and nitrate. All four LHS1 samples with the lowest global misfits only consist of low individual misfits.

Figures A.4 and A.5 show the results of the DFO-LS experiments initiated from "good" starting points. Regarding the difficult K_{PHY} parameter, of the 8 experiments only two (**t6d_LHS1_good1** and **t6d_LHS1_good3**) were most successful in "recovering" this parameter, most likely because they were initiated very close to the target. Therefore starting from "good" initial locations do not reliably encourage the recovery of this parameter.

A.2.4 Conclusions

Unsurprising, the four LHS1 samples with the lowest global misfits contain exclusively low misfit values for each oceanic tracer and biome region. Oxygen

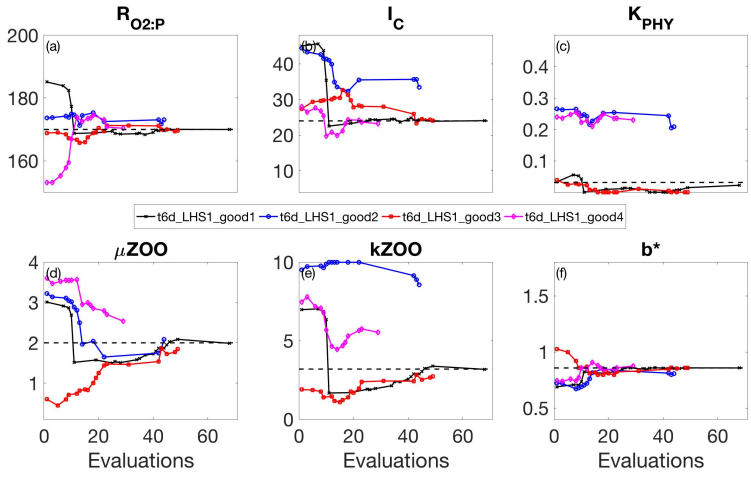

Figure A.4: Parameter tuning by DFO-LS for the experiments **t6d_LHS1_good1** (black line with crosses), **t6d_LHS1_good2** (blue line with circles), **t6d_LHS1_good3** (red line with squares) and **t6d_LHS1_good4** (magenta line with diamonds) for the parameters (a) $R_{O_2:P}$, (b) I_C, (c) K_{PHY}, (d) μ_{ZOO}, (e) k_{ZOO} and (f) **b***. Also plotted are the MOPS-ref target parameter values (horizontal black dashed line). Only evaluations which decreased the misfit have been plotted.

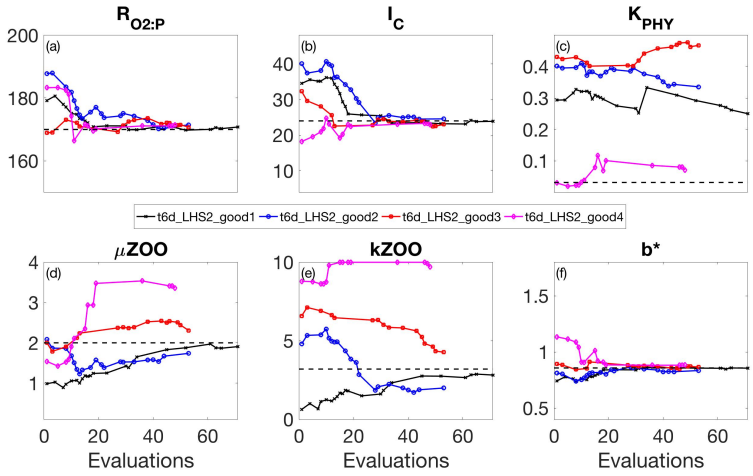

Figure A.5: Parameter tuning by DFO-LS for the experiments **t6d_LHS2_good1** (black line with crosses), **t6d_LHS2_good2** (blue line with circles), **t6d_LHS2_good3** (red line with squares) and **t6d_LHS2_good4** (magenta line with diamonds) for the parameters (a) $R_{O_2:P}$, (b) I_C, (c) K_{PHY}, (d) μ_{ZOO}, (e) k_{ZOO} and (f) **b***. Also plotted are the MOPS-ref target parameter values (horizontal black dashed line). Only evaluations which decreased the misfit have been plotted.

shows the largest range of misfits therefore is more influenced by the varying parameter values, particularly in the Pacific and Indian oceans. Modelled oxygen is dominantly controlled by surface air-sea gas exchange, however it is also influenced by the overturning circulation and decomposition of organic matter. Decomposing

organic matter depletes oxygen from the water, therefore by varying biogeochemical parameters such as particle sinking speed (Martin et al., 1987) one determines the amount of organic matter reaching the deep ocean, and hence the amount of oxygen depletion further along the ocean conveyor belt (Broecker, 1991; Lumpkin and Speer, 2007). This may be why the oxygen misfits vary so much in the Indian and Pacific oceans towards the end of the conveyor belt (Kwon and Primeau, 2006; Kwon et al., 2009). As phosphate and nitrate are both nutrients they behave relatively similar, therefore both show slightly larger changes in misfits in the upper (surface 1000m) Atlantic Ocean. Again this could be due to varying the parameter controlling particle sinking speed, as faster sinking of nutrient-rich organic particles strip the surface ocean of phosphate and nitrate (Kriest and Oschlies, 2013).

When starting from 2 random points in the parameter space (see Chapter 3), DFO-LS managed to recover the target parameters as well as 6/8 of the optimisation experiments shown here, therefore initiating from "good" starting locations does not reliably improve the success of DFO-LS. Therefore this method of combining initial sampling with DFO-LS did not force DFO-LS to find the global optimum, just a local one, which DFO-LS can do on it's own already without the expense of initial sampling.

B

How big an ocean biogeochemical step can we take?

B.1 Introduction

Ocean biogeochemical modellers frequently need to find a realistic balance between model accuracy and computational expense, such as when considering the size of both the ocean circulation and biogeochemical model time step. Ocean biogeochemical models coupled with "online" general circulation models (GCMs) are very computationally expensive, and can take weeks to run for a few hundred simulated years. Therefore an investigation of the size of the biogeochemical model time step is often too costly, and an estimate is used whereby the time step is safely small enough for the model to remain stable. An example of an unstable biogeochemical time step would be if it were longer than the time taken for sinking organic matter to pass through one single depth layer of the model.

Ocean biogeochemical models coupled with a "offline" GCM using the Transport Matrix Method (TMM: Khatiwala et al., 2005; Khatiwala, 2007) are much less expensive, allowing this investigation of the biogeochemical step to be carried out. The TMM is a numerical scheme for efficient simulation of passive oceanic chemical and biological tracers, which removes the need to run an expensive "online" GCM alongside the biogeochemical model. This method is compatible with any GCM as TMM matrices can be generated from any circulation model, and most biogeochemical models can be modified to be successfully coupled to the TMM. Kvale et al. (2017) described the process of extracting TMM matrices from version 2.9 of the University of Victoria Earth System Climate Model (Weaver et al., 2001; Eby et al., 2009), evaluated the performance of the "offline" biogeochemical model within UVic ESCM compared to the "online" version, and concluded the TMM is a reliable tool for its purpose. The reduction in computational expense, plus the functionality within the TMM to vary the number of biogeochemical steps per ocean step, allows us to investigate the preferred biogeochemical step size. They found an offline ocean circulation and biogeochemical time step of 8 hours to be most desirable in their model configuration, when considering both stability and accuracy.

Here we do something similar, but with the ocean biogeochemical model MEDUSA (Yool et al., 2013), coupled with the UVic circulation using the TMM. In this configuration of MEDUSA and UVic the ocean time step is 8 hours, and the biogeochemical step was 8 per ocean step (therefore 1 hour). We investigate the quality of the model outputs while decreasing the biogeochemical time step per ocean step from 8 to 1, so to decrease computational expense of the model, while maintaining model output quality. This work has been carried out prior to an optimisation study, as reducing the computational expense of this model set up will allow for faster and less expensive optimisation. The planned optimisation is to calibrate 15 of MEDUSA's parameters using global (exluding the Arctic Ocean) observations of oxygen, nitrate, silicate, alkalinity and net primary production, therefore the model tracers related to these will be particularly inspected when reducing the biogeochemical step size.

B.2 Methodology

The global ocean biogeochemical model MEDUSA (Yool et al., 2011, 2013) has been coupled to the general circulation model UVic (Weaver et al., 2001; Eby et al., 2009) using the TMM (Khatiwala et al., 2005; Khatiwala, 2007). Using an ocean step size of 8 hours and a biogeochemical step size of 1 hour, MEDUSA was initiated from uniform tracer distributions and spun up for 500 years. Usually we would spin up to assumed equilibrium after at least 3000 years, however we only use the model outputs to investigate the relative influence of varying the biogeochemical step size, therefore 500 years is sufficient. The 15 modelled tracer distributions (see Table B.1) for the final year of the spin up were then averaged and output by the model. This is henceforth called the "reference run", to which all runs with shortened biogeochemical steps were compared to. Six additional model model runs were carried out, each identical to the reference run except for the biogeochemical step size. In these six runs the number of biogeochemical steps per ocean step was decreased linearly from 6 to 1, which is equivalent to exponentially increasing the biogeochemical step size from 1.33 hours to 8 hours, considering the ocean

Model Tracer Outputs			
Abbreviation	**Description**	**Units**	Largest acceptable biogeochemical step size [hrs]
OXY	Dissolved oxygen	mmolO2/m3	8.00
ALK	Alkalinity	meq/m3	8.00
SIL	Silicate	mmolSi/m3	4.00
DIN	Dissolved inorganic nitrogen	mmol-N/m3	8.00
DET	Detrital nitrogen	mmol-N/m3	4.00
FER	Dissolved iron	mmolFe/m3	8.00
DTC	Detrital carbon	mmol-C/m3	4.00
DIC	Dissolved inorganic carbon	mmol-C/m3	8.00
CHN	Chl-a concentration in non-diatom phytoplankton	mg Chl/m3	2.67
CHD	Chl-a concentration in diatom phytoplankton	mg Chl/m3	2.00
PHN	Non-diatom phytoplankton	mmol-N/m3	4.00
PHD	Diatom phytoplankton	mmol-N/m3	2.00
PDS	Biogenic silicon in diatom phytoplankton	mmolSi/m3	2.67
ZMI	Micro zooplankton	mmol-N/m3	2.00
ZME	Meso zooplankton	mmol-N/m3	2.00

Table B.1: The 15 tracers modelled by MEDUSA.

step size is 8 hours. The output tracer distributions from these six runs were then compared to the reference run by calculating the percentage increase and/or decrease from the reference run for each of the 15 tracers, with the intention of finding the largest possible biogeochemical time step while maintaining an acceptably small difference from the reference run. See Table B.2 for the biogeochemical step sizes and computational time required for each of these runs.

The percentage change $\mathbf{P}$ from the reference run **ref** to the reduced biogeochemical step size run **mod**, for each of the 15 tracers t and 6 biogeochemical step sizes n were defined as

$$\mathbf{P}_{tn} = 100\frac{\mathbf{ref}_t - \mathbf{mod}_{tn}}{\mathbf{ref}_t}. \tag{B.1}$$

We determine the percentage change in a tracer distribution to be "acceptable" if at least 95% of the global tracer distribution has a percentage change of less than 1% from the reference run.

B.3 Results

See Figures B.1 to B.15 for maps of the percentage change in each of the 15 tracers due to increasing the biogeochemical step size. The results of these are more condensed into the final column of Table B.1, where the largest acceptable biogeochemical time step has been determined for each of the 15 tracers, whereby a step is "acceptable" if at least 95% of the global tracer distribution has a percentage change of less than 1% from the reference run.

MEDUSA runs		
Biogeochemical step size [hrs]	**Number of biogeochemical steps per ocean step**	**Computational wallclock time [hrs]**
1 (Reference run)	8	5.7
1.33	6	3.0
1.60	5	1.7
2.00	4	2.3
2.67	3	1.2
4.00	2	1.6
8.00	1	1.2

Table B.2: The (column 1) biogeochemical step size, (column 2) number of biogeochemical steps per 8 hour ocean step, and (column 3) computational wall clock time required to spin up to 500 years, for the reference run and each of the 6 shortened biogeochemical step runs. These were run on 16 nodes, 16 processors per node, on the ARC supercomputing cluster at Oxford University. Note that on ARC some nodes seemingly randomly sometimes perform faster than others, therefore any noted change in computational time may be due to that.

The tracers which change the least with an increase in biogeochemical step size are OXY, ALK, DIN, FER, and DIC, which all have a largest acceptable biogeochemical step size of 8 hours. Next with a largest acceptable biogeochemical step size of 4 hours are the tracers SIL, DET, DTC and PHN. The tracers CHN and PDS are significantly perturbed when increasing the biogeochemical step size, as they have a shorter largest acceptable biogeochemical step size of 2.67 hours, and the most perturbed tracers with a largest acceptable biogeochemical step size of only 2 hours are the tracers CHD, PHD, ZMI and ZME.

Figures B.1 to B.15 show how the model outputs increasingly differed from the reference run as the biogeochemical step size was increased, commonly in the Arctic Ocean, and the Weddell and Ross seas in the Southern Ocean. There are also three more areas where certain tracers are significantly perturbed; in the Indian Ocean for DTC, DET, DIN, and SIL, in the deep South Atlantic for CHD and ZMI, and in the deep Equatorial Pacific for CHN and ZME.

Table B.2 details the computational time required for each run, with the reference run with a 1 hour biogeochemical step taking 5.7 hours, and the run with a 2 hour biogeochemical step taking 2.3 hours. The computational expense is more than halved by making this change to the biogeochemical step size.

B.4 Conclusion

In order to find a less expensive model set up to optimise we intend to increase the biogeochemical model step size, while maintaining model outputs acceptably similar to those created by a smaller biogeochemical step size. We found that increasing the biogeochemical step size from 1 hour (the reference run) to 2 hours reduces the computational time to run a 500 year spin up by more than half, while keeping the difference in tracer outputs acceptably small for all of the 15 model tracers. This allows us to continue with an optimisation study of a less expensive model setup, while remaining confident that optimising the setup with a larger biogeochemical step size would reveal similar results to that of optimising the setup with the original smaller biogeochemical step size.

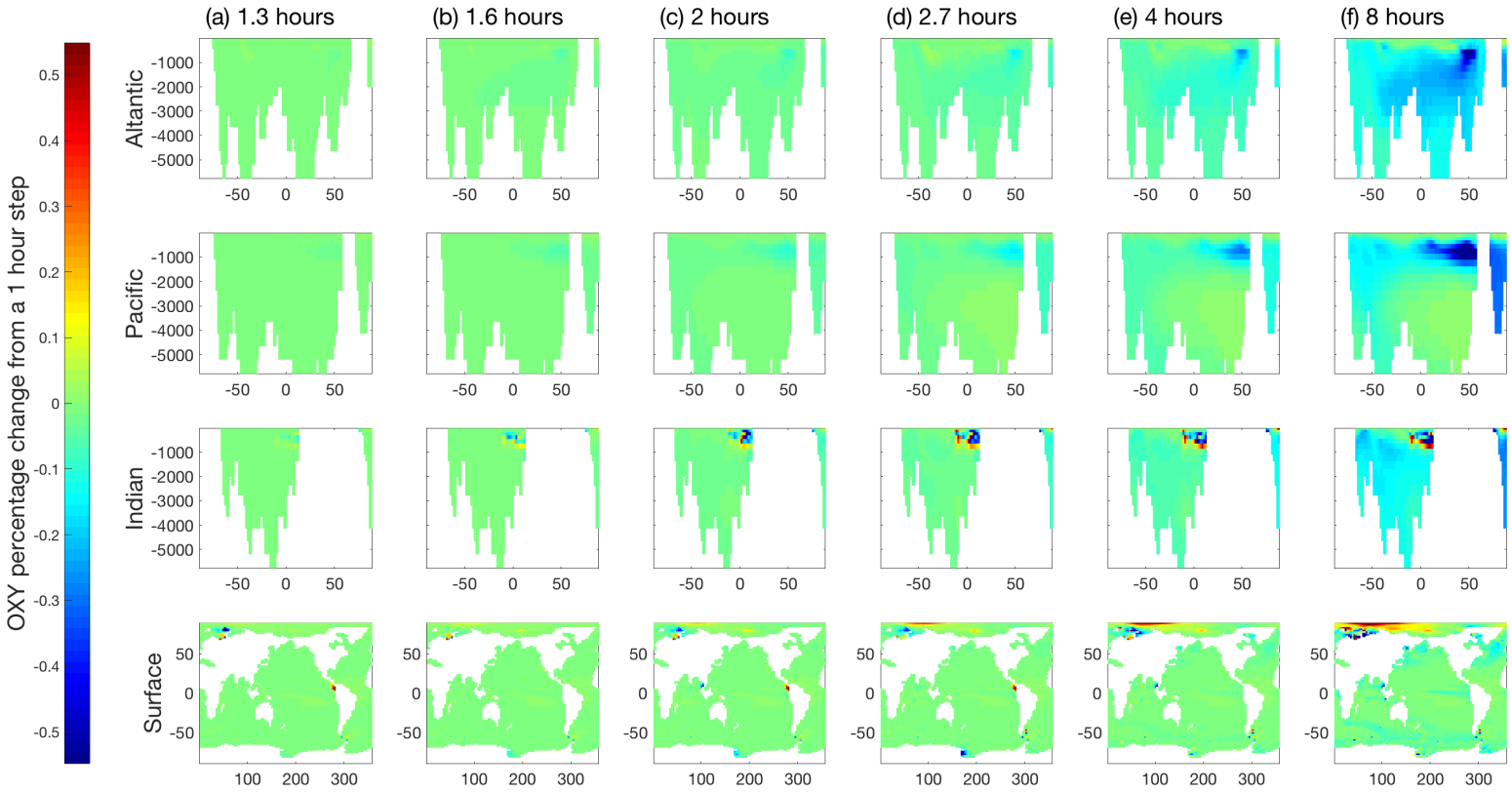

Figure B.1: Percentage increase and decrease in OXY after changing the biogeochemical step from 1 hour to (a) 1.3, (b) 1.6, (c) 2, (d) 2.7, (e) 4 and (f) 8 hours, for the (top) Atlantic, (2nd row) Pacific, (3rd row) Indian and (bottom) world surface ocean.

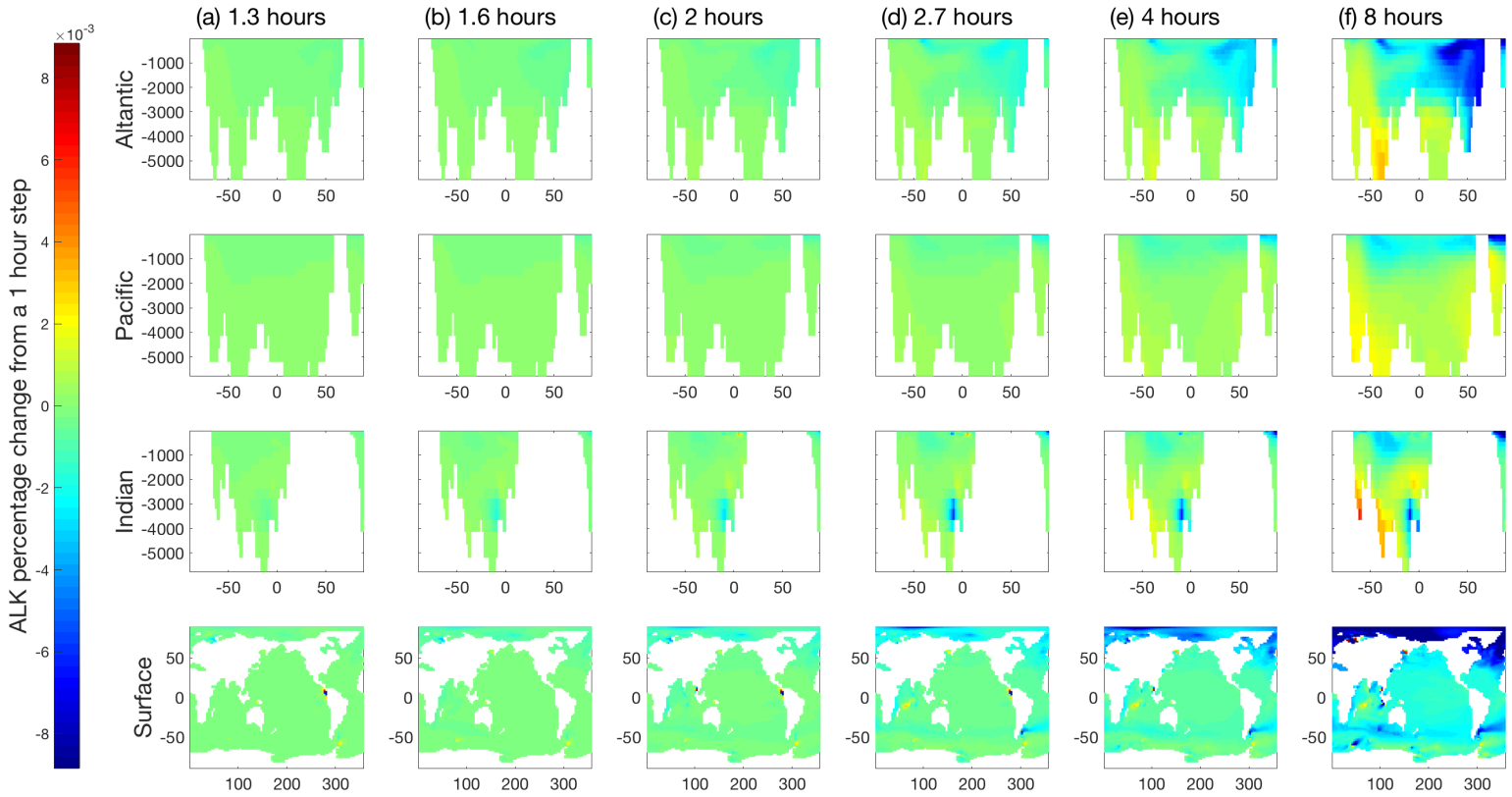

Figure B.2: Same as Figure B.1 but for ALK. Near-zero values are still mapped as green, however note the change in scale limits.

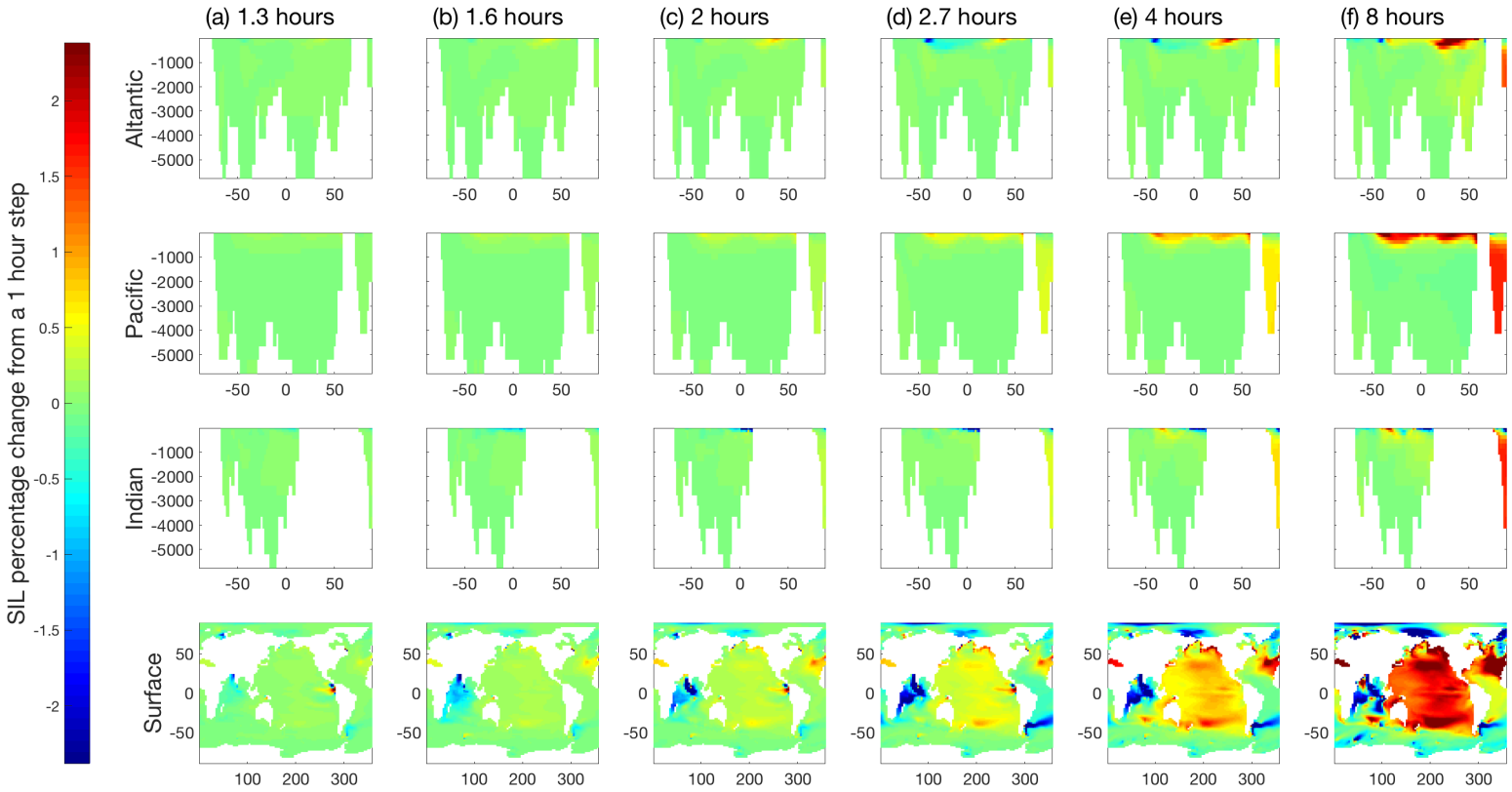

Figure B.3: Same as Figure B.1 but for SIL. Near-zero values are still mapped as green, however note the change in scale limits.

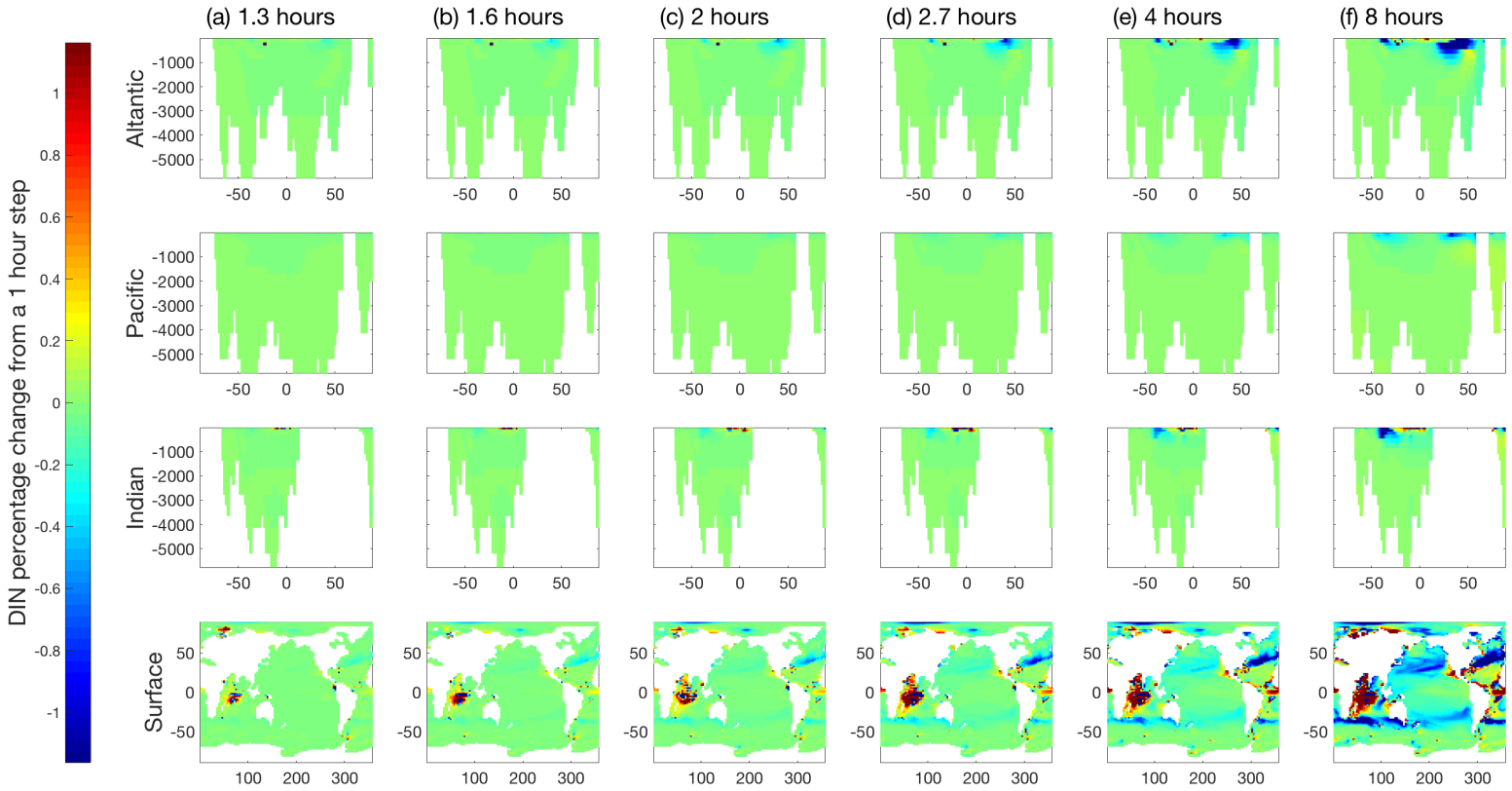

Figure B.4: Same as Figure B.1 but for DIN. Near-zero values are still mapped as green, however note the change in scale limits.

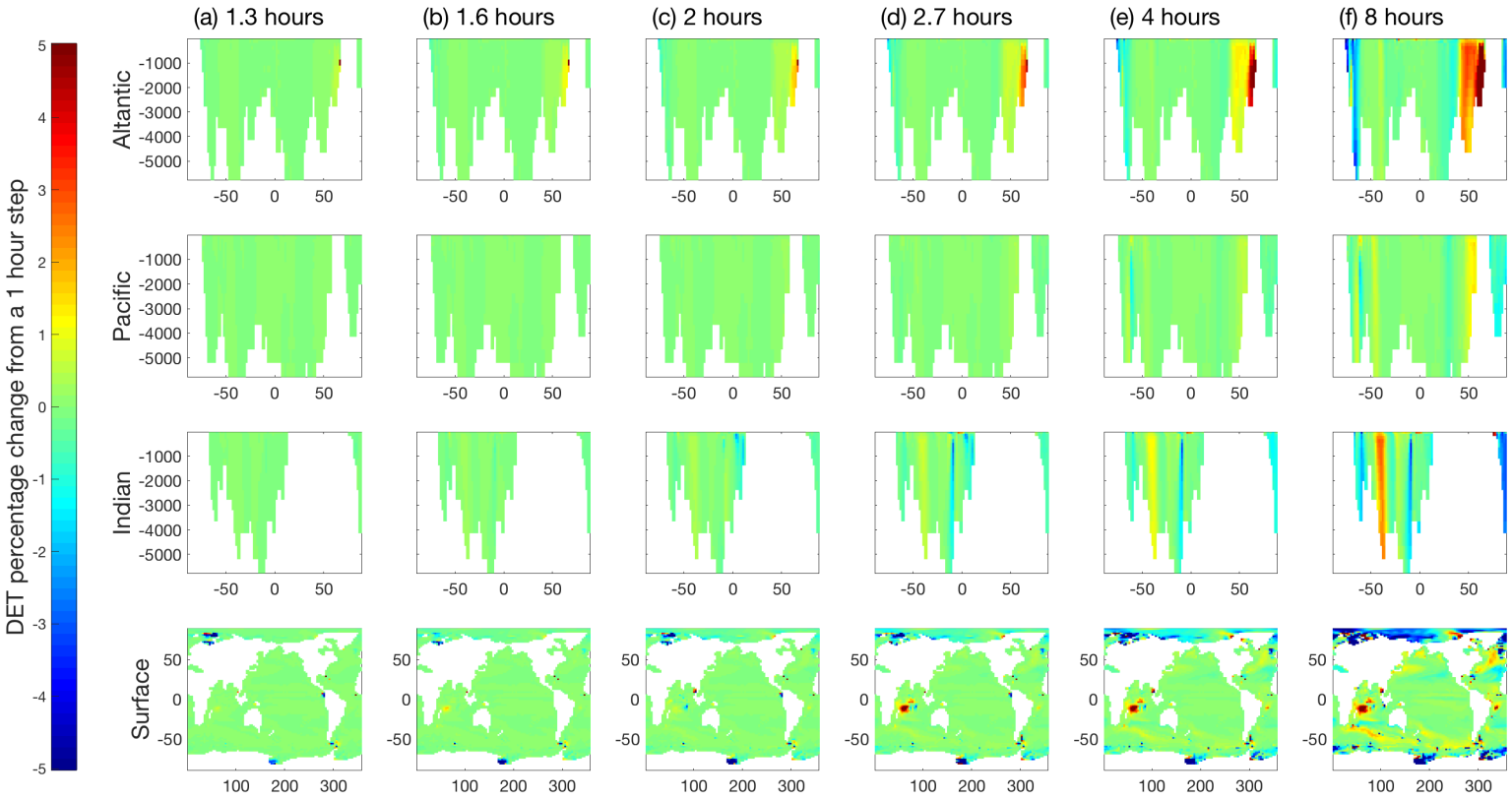

Figure B.5: Same as Figure B.1 but for DET. Near-zero values are still mapped as green, however note the change in scale limits.

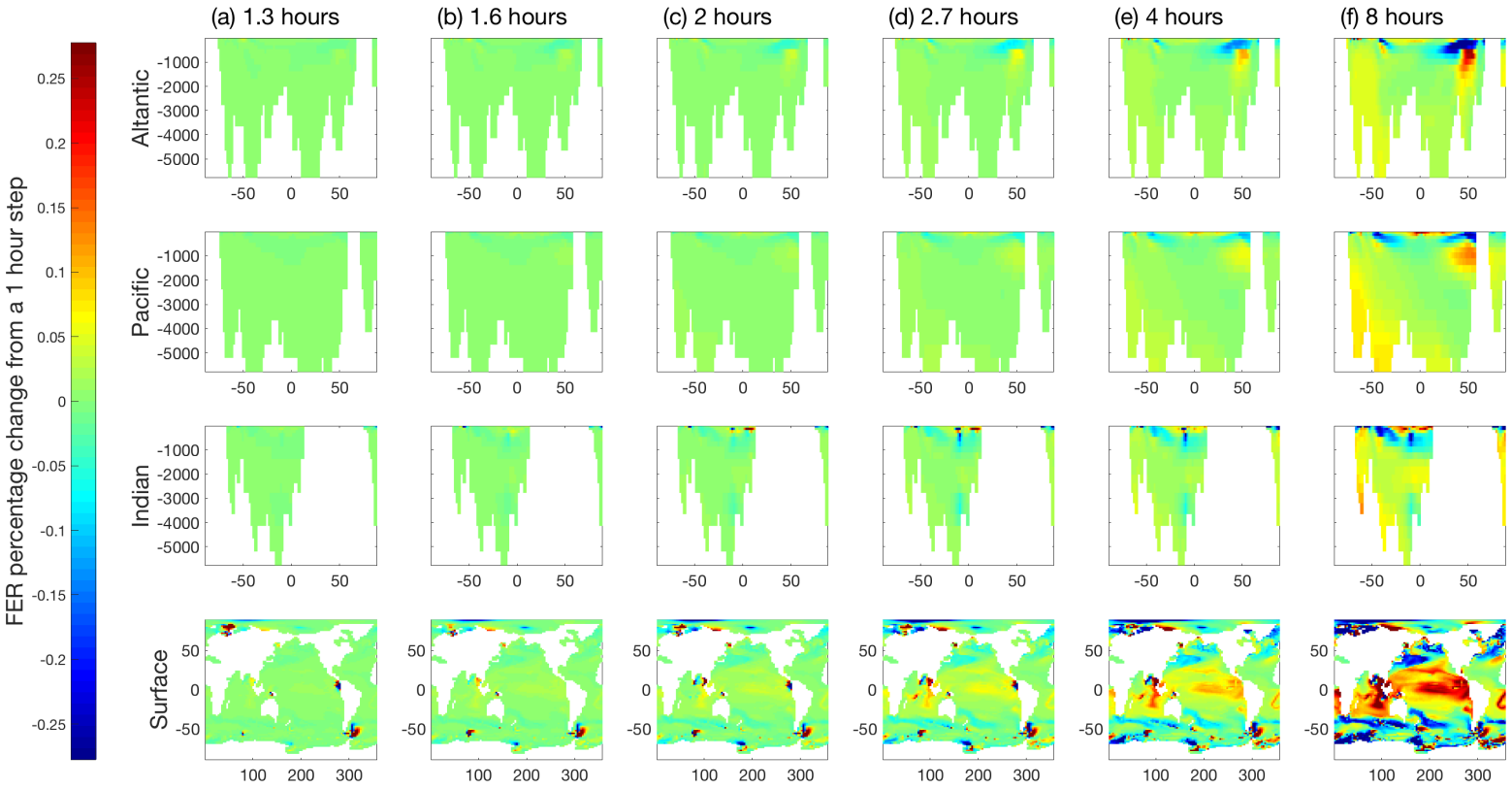

Figure B.6: Same as Figure B.1 but for FER. Near-zero values are still mapped as green, however note the change in scale limits.

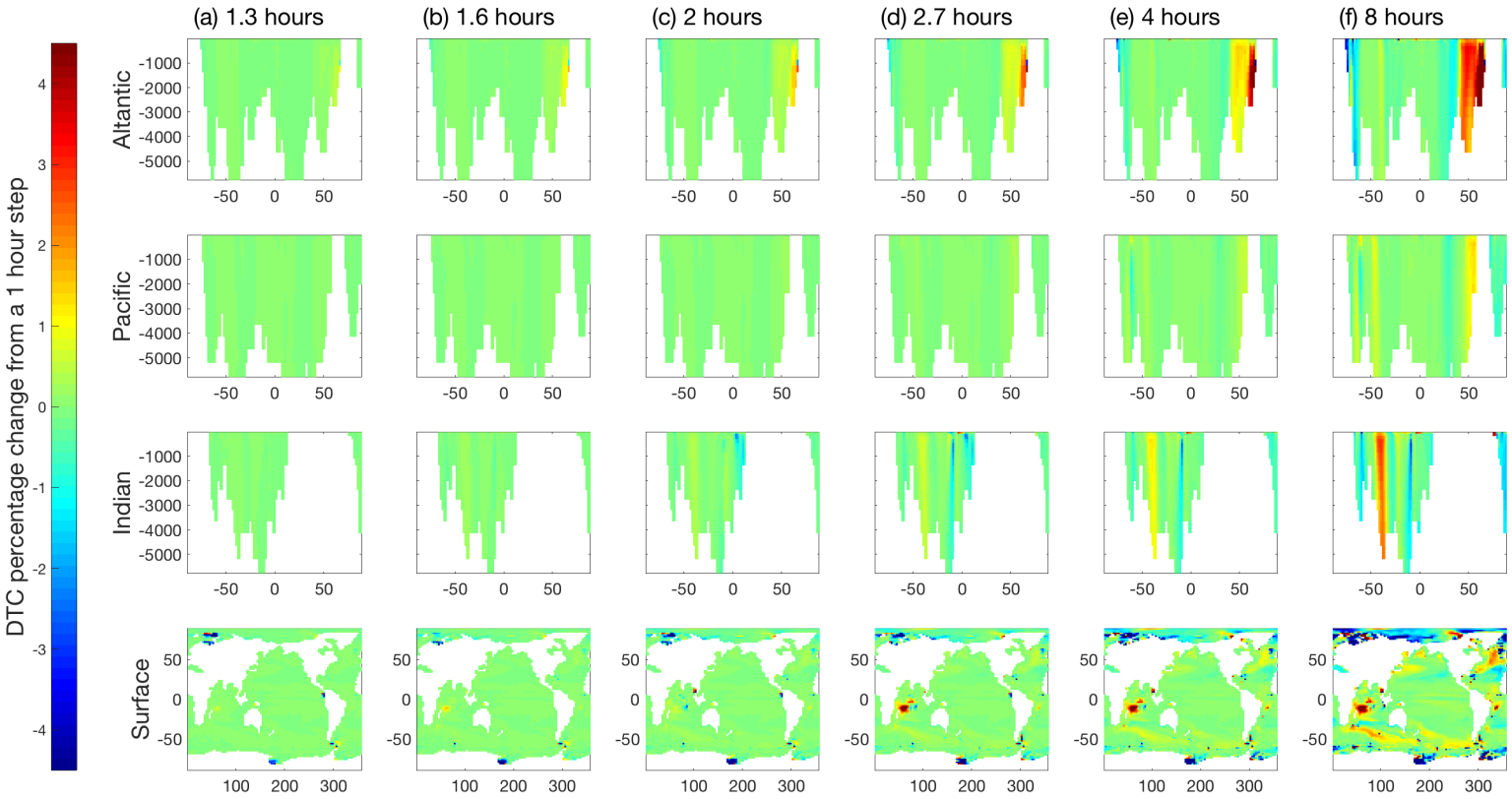

Figure B.7: Same as Figure B.1 but for DTC. Near-zero values are still mapped as green, however note the change in scale limits.

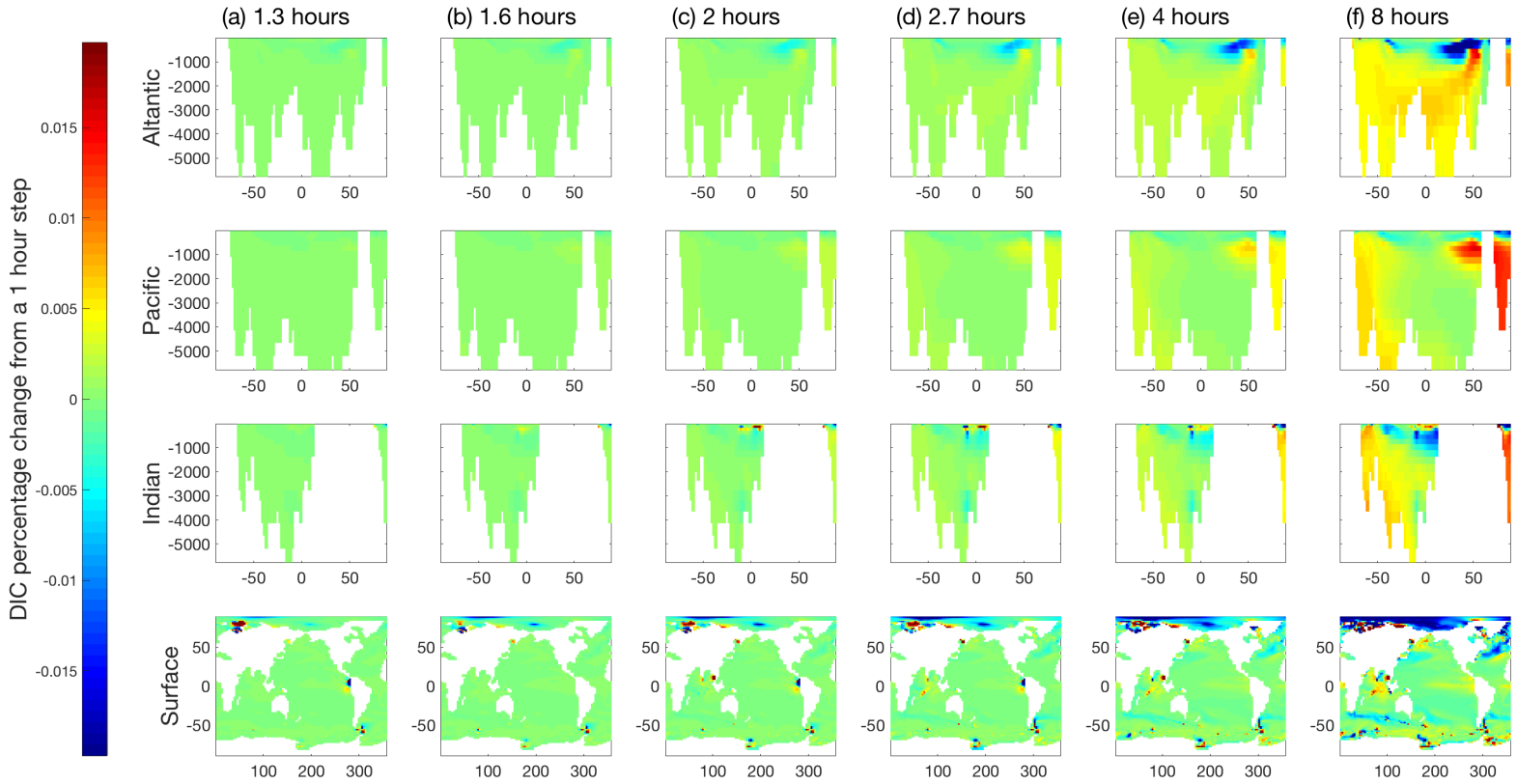

Figure B.8: Same as Figure B.1 but for DIC. Near-zero values are still mapped as green, however note the change in scale limits.

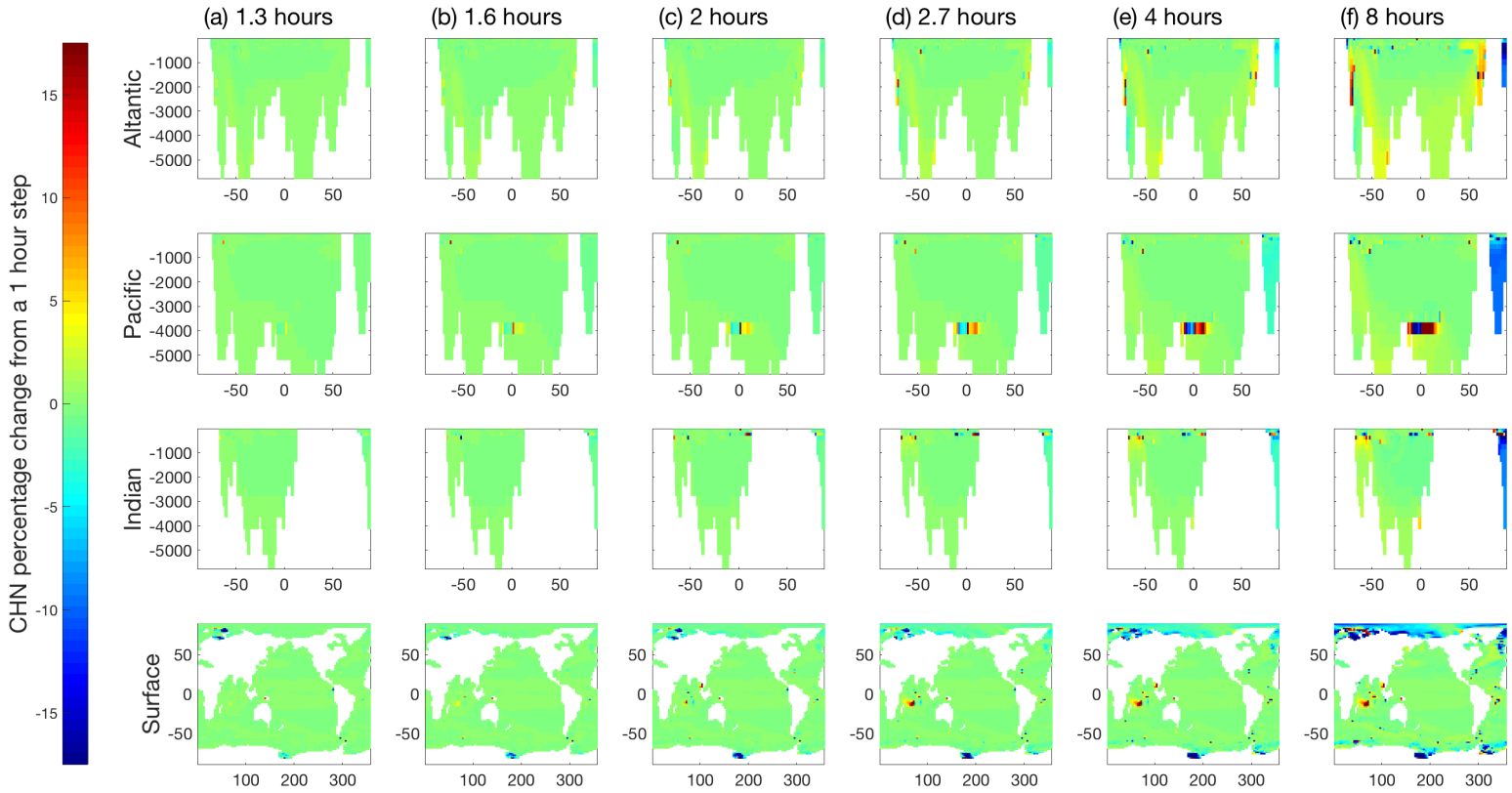

Figure B.9: Same as Figure B.1 but for CHN. Near-zero values are still mapped as green, however note the change in scale limits.

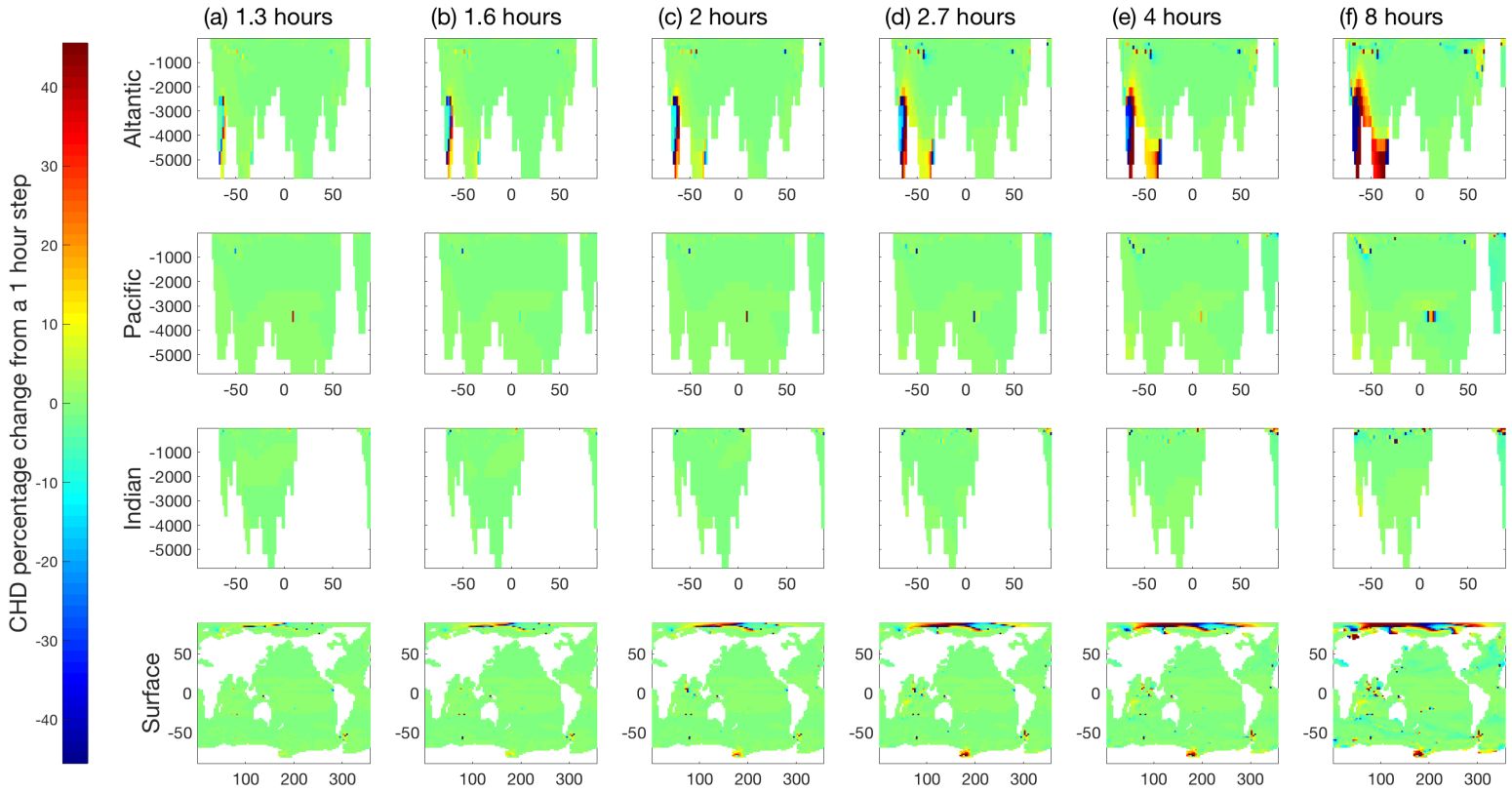

Figure B.10: Same as Figure B.1 but for CHD. Near-zero values are still mapped as green, however note the change in scale limits.

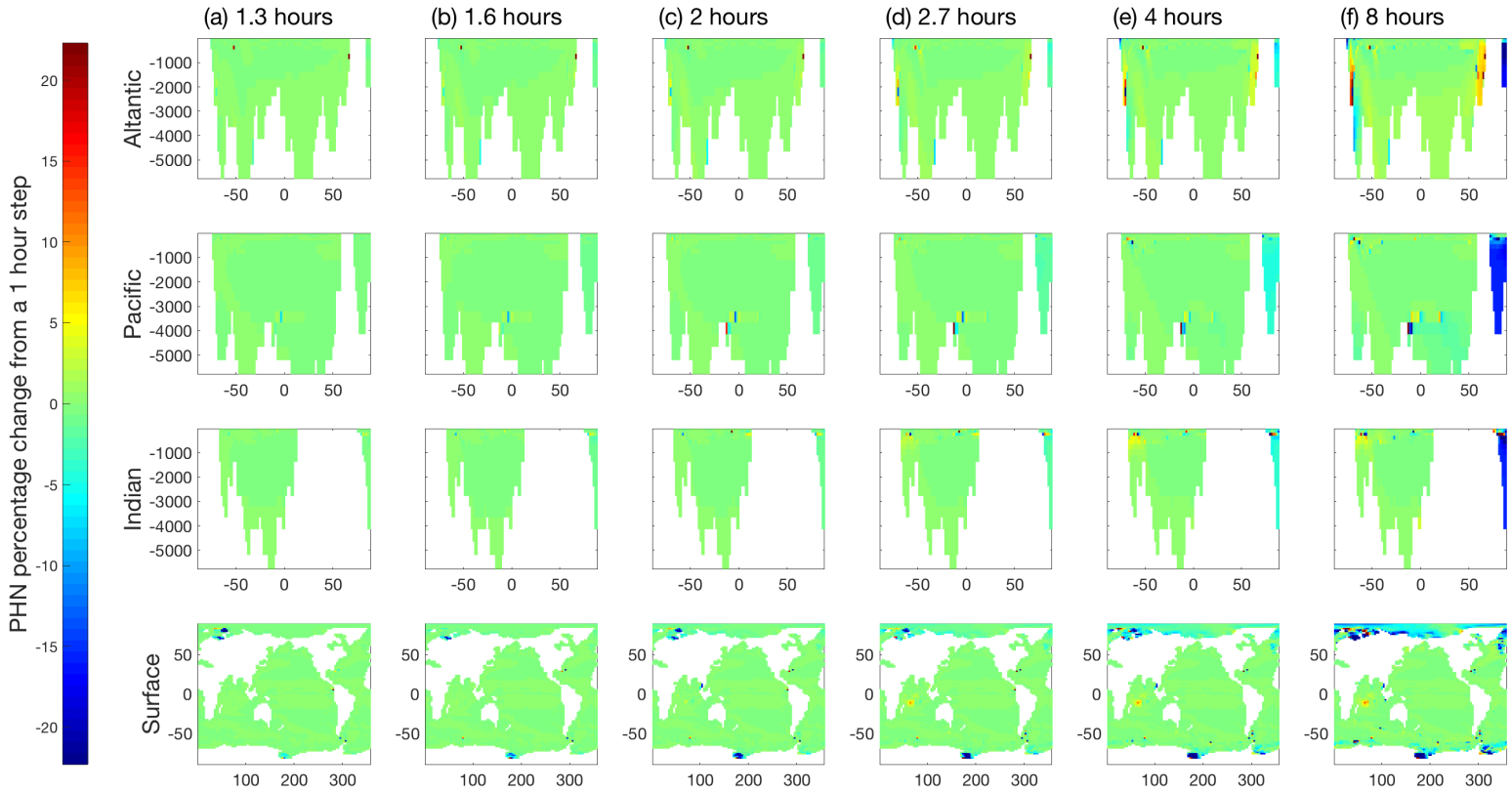

Figure B.11: Same as Figure B.1 but for PHN. Near-zero values are still mapped as green, however note the change in scale limits.

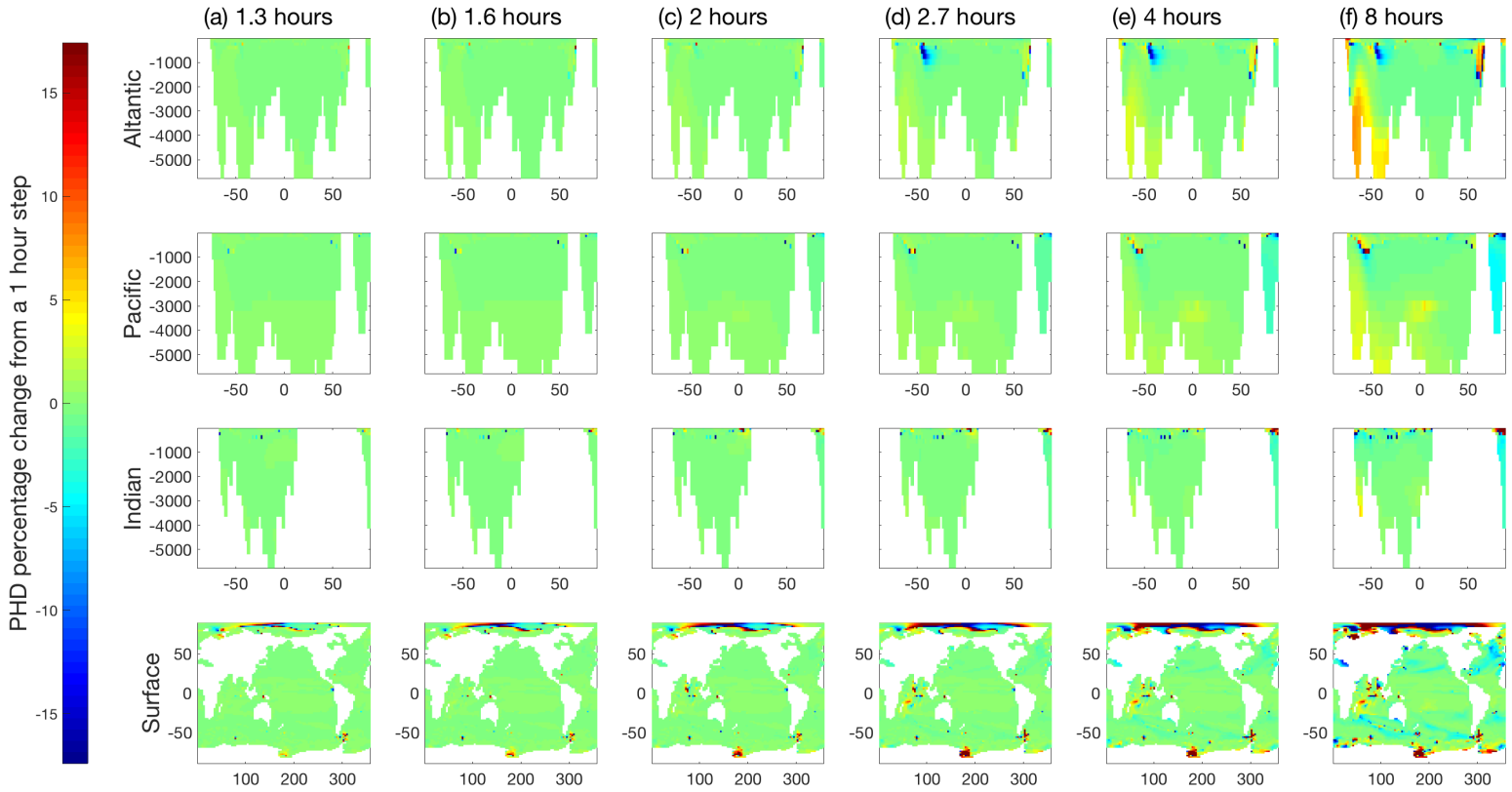

Figure B.12: Same as Figure B.1 but for PHD. Near-zero values are still mapped as green, however note the change in scale limits.

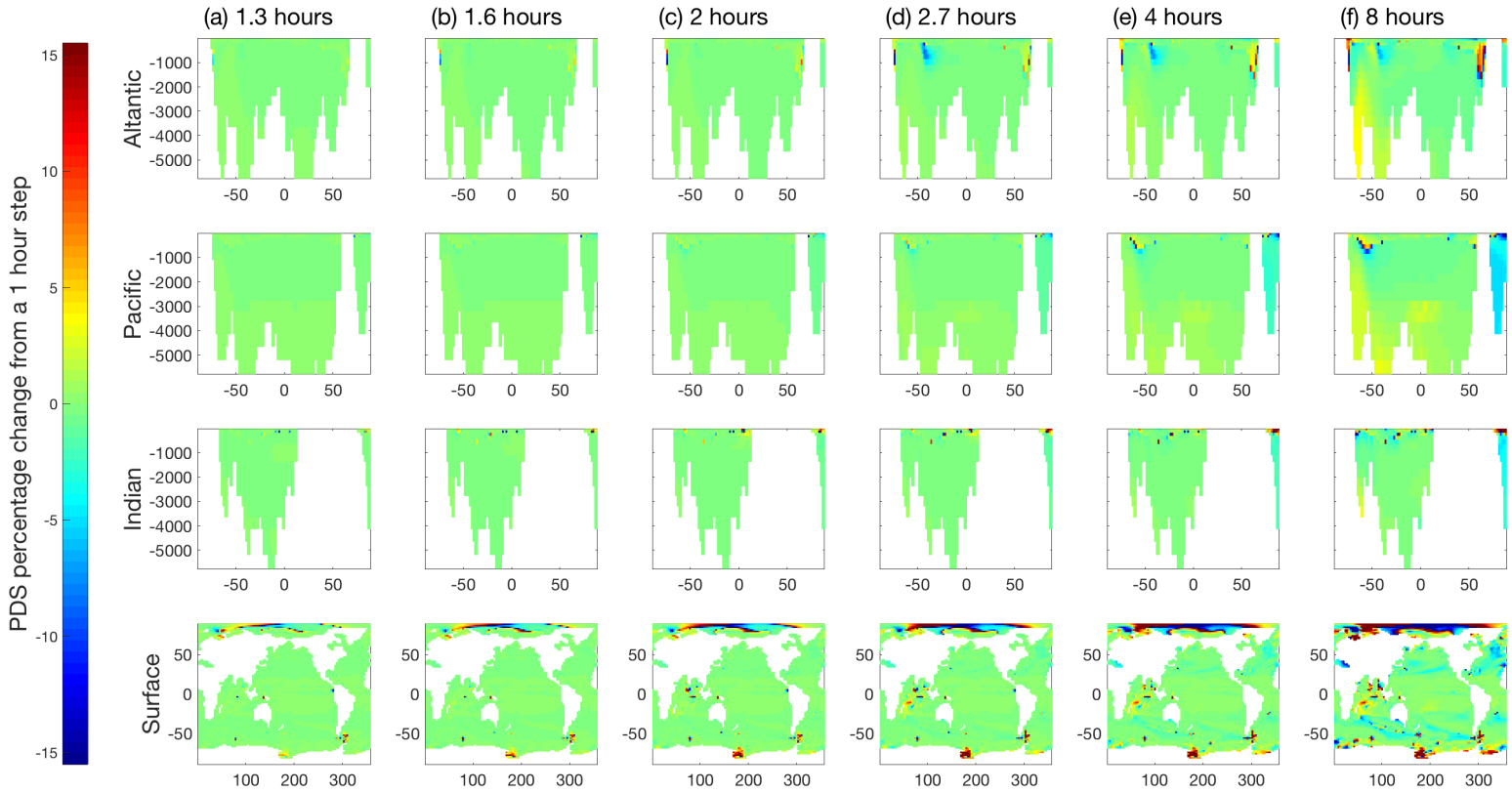

Figure B.13: Same as Figure B.1 but for PDS. Near-zero values are still mapped as green, however note the change in scale limits.

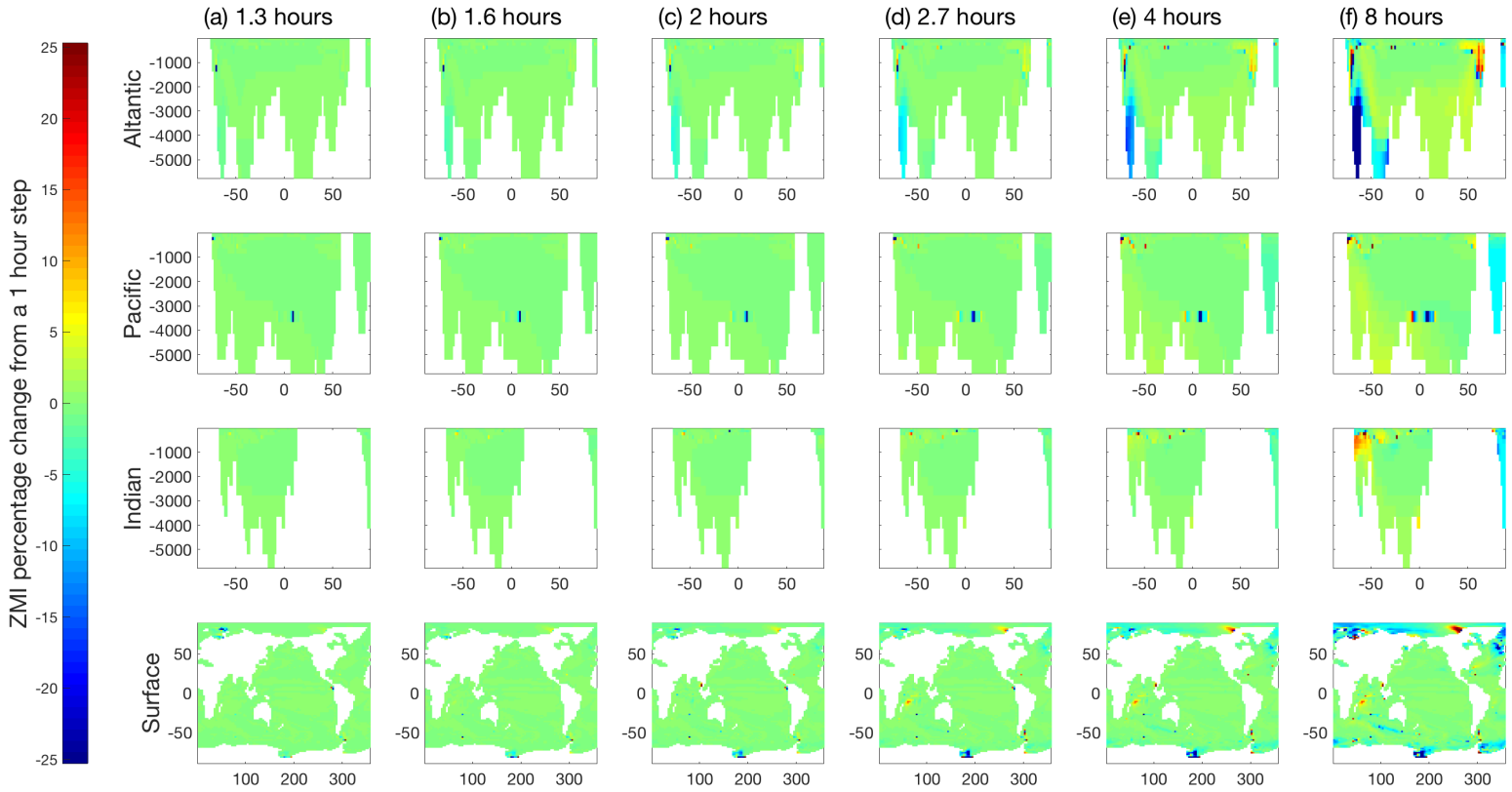

Figure B.14: Same as Figure B.1 but for ZMI. Near-zero values are still mapped as green, however note the change in scale limits.

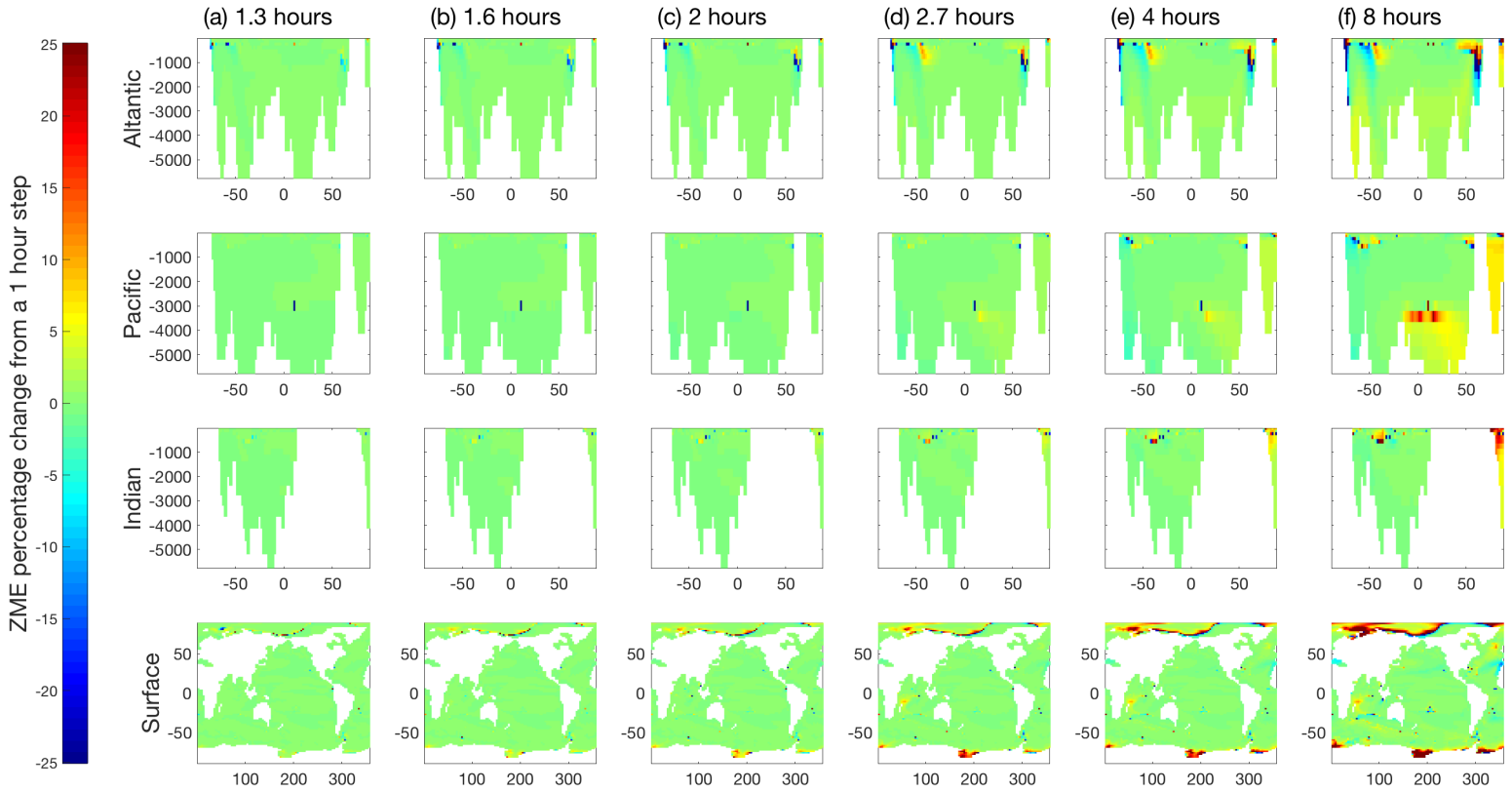

Figure B.15: Same as Figure B.1 but for ZME. Near-zero values are still mapped as green, however note the change in scale limits.

C

MEDUSA-2.0 spin up time required to reach an equilibrated state

C.1 Introduction

Ocean biogeochemical models need to be “spun-up" to allow the model to reach a sufficiently equilibrated state, which can take from 3000 to 10,000 years (Wunsch and Heimbach, 2008), with the Atlantic taking the shortest time to equilibrate, and the deep North Pacific the longest (Wunsch and Heimbach, 2008; Primeau and Deleersnijder, 2009).

After investigation of the behaviour of tracers through the spin up of the global ocean biogeochemical model MOPS (Kriest and Oschlies, 2015) using transport matrices (Khatiwala et al., 2005; Khatiwala, 2007) from the MITgcm2.8 general circulation (Marshall et al., 1997), Kriest et al. (2017) determined a 3000 year spin up was sufficient. Hence in Chapters 3 - 5 of this thesis we also used a 3000 year spin up to equilbrate MOPS.

In Chapters 6 and 7 we move on to the more complex global ocean biogeochemical model MEDUSA-2.0 (Yool et al., 2013), using transport matrices from the UVic general circulation (Weaver et al., 2001; Eby et al., 2009). Therefore we had to re-assess the spin up length required to reach equilibrium, which we do here by tracking the model misfit to observations and tracer behaviour through a 6000 year spin up. To keep computational expense as low as possible a short spin up length is desirable, yet it must still be long enough to allow the model to approach equilibrium.

C.2 Conclusion

It was found that a 3000 year spin up is sufficient for this model set up with MEDUSA-2.0, as all misfits (Figure C.1) and all regionally-averaged tracers (Figure C.2) have equilibrated by 3000 years, with oxygen and silicate in the north Pacific taking the longest. Therefore a 3000 year spin up is henceforth used in this thesis for MEDUSA-2.0.

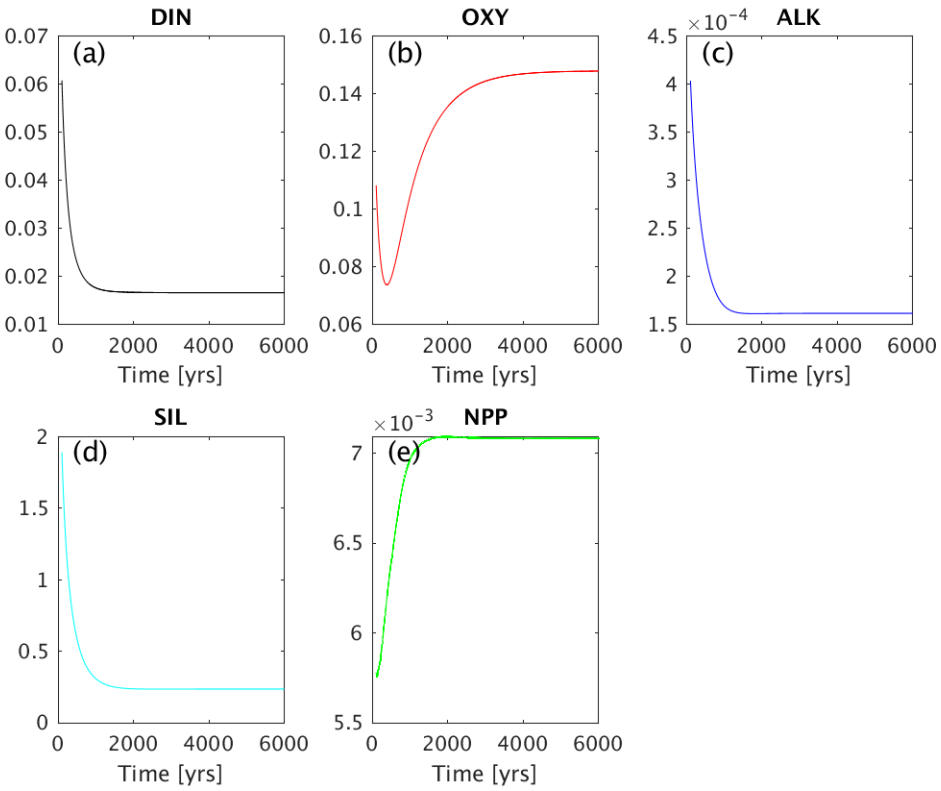

Figure C.1: Average global misfit for each of the tracers (a) dissolved inorganic nitrate, (b) oxygen, (c) alkalinity, (d) silicate, and (e) primary production.

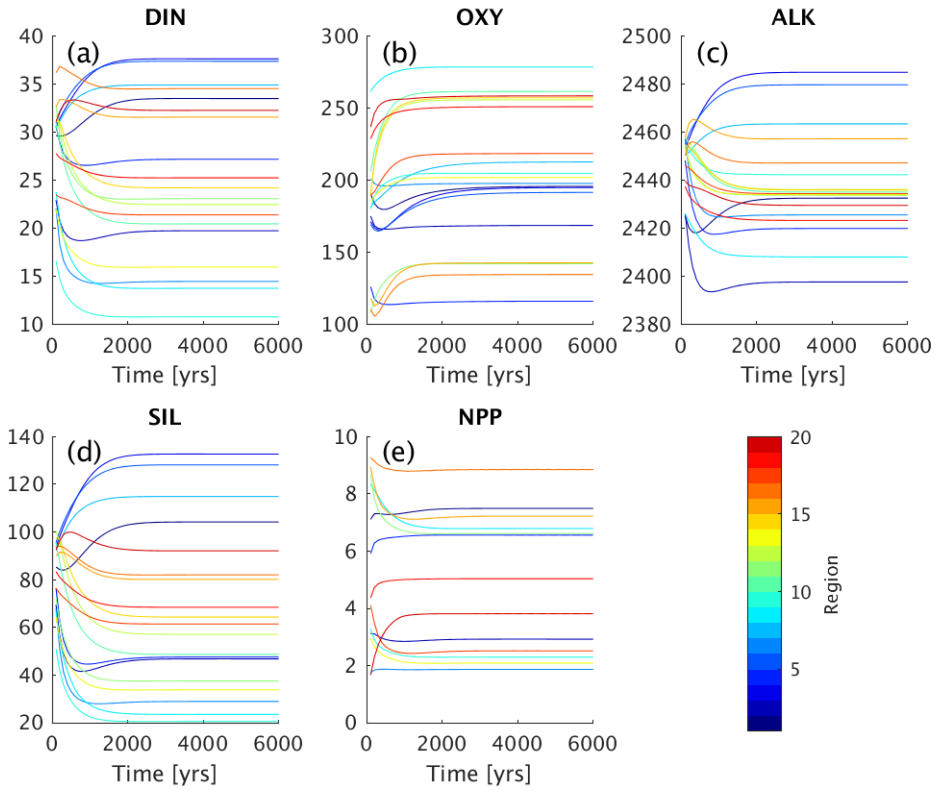

Figure C.2: Averaged (a) dissolved inorganic nitrate, (b) oxygen, (c) alkalinity, (d) silicate, and (e) primary production, for each of the 19 regions (see Table C.1).

Region number	Region description
1	Northern North Pacific
2	North Pacific < 1000m
3	North Pacific > 1000m
4	Equatorial Pacific < 1000m
5	Equatorial Pacific > 1000m
6	South Pacific < 1000m
7	South Pacific > 1000m
8	Northern North Atlantic
9	North Atlantic < 1000m
10	North Atlantic > 1000m
11	Equatorial Atlantic < 1000m
12	Equatorial Atlantic > 1000m
13	South Atlantic < 1000m
14	South Atlantic > 1000m
15	North Western Indian
16	North Eastern Indian
17	South Indian
18	Southern Ocean: Subantarctic zone
19	Southern Ocean: Antarctic zone

Table C.1: Specified regions, or oceanic biomes.

D

A comparison of MEDUSA-2.0 local parameter sensitivities at two locations in parameter space

D.1 Introduction

Here we re-calculate the parameter perturbation MEDUSA-2.0 sensitivities at the new, optimised location **Comb4** from Chapter 7, and compare to the parameter sensitivities previously calculated in Chapter 6 at the default **Reference** location. The purpose of this was to determine how the local parameter sensitivities change after optimisation. As the purpose of optimisation is to minimise the misfit to observations, this **Comb4** should now be in a minima within the misfit function, and therefore the sensitivity of the misfit function to perturbations in parameters at this new location is expected to be smaller than at the **Reference** parameter location. The methodology here follows Section 6.2.

D.2 Results

Table D.1 shows a summary of the parameter values for the respective perturbation runs. Tables D.2 show the sensitivity tables for both locations in parameter space, **reference** and **Comb4**, which is the top and bottom table respectively. These show the averaged misfits change from the initial configuration to the configuration where 1 individual parameter is perturbed, for each of the 16 tracers/variables, and in the final column the combined misfit which includes DIN, OXY, ALK, SIL and NPP.

The sensitivities change between the two locations in parameter space. Most obviously, oxygen now dominates the misfit changes at location **Comb4**, when before for the **Reference** configuration it was silicate.

D.3 Discussion

During the optimisation in Chapter 7, the misfit to silicate was reduced significantly from the **Reference** configuration to the **Comb4** configuration. Therefore, the large dominance of silicate on the parameter sensitivities seen at parameter location **Reference** has now been reduced at location **Comb4**, allowing oxygen to now be more dominant. The sensitivities overall have slightly reduced after optimisation to

Description	Abbreviation	Lower bound	Upper bound	Reference	Opt4	Perturbed ref	Perturbed opt	Combined misfit ref	Combined misfit opt
Maximum meso-zooplankton grazing rate [d−1]	xgme	0.25	1.15	0.5	0.25	0.59	0.3400	4.38E-02	7.97E-04
Diatom Fe nutrient uptake half-saturation constant [mmolFem−3]	xfld	0.000335	0.00134	0.00067	0.00067	0.0007705	0.0008	2.65E-02	1.80E-03
Diatom N nutrient uptake half-saturation constant [mmolNm−3]	xnld	0.1	6	0.75	0.375	1.34	0.9650	2.06E-02	3.06E-03
Maximum growth rate for non-diatoms [d−1]	xvpn	0.25	2	0.8	0.8	0.975	0.9750	1.99E-02	1.45E-05
Non-diatom chl-specific initial slope of P–I curve [gC(gchl)−1 (Wm−2)−1 d−1]	xaln	0.3	30	15	7.5	17.97	10.47	6.26E-03	1.55E-04
Maximum growth rate for diatoms [d−1]	xvpd	0.25	2	0.75	0.75	0.925	0.9250	5.84E-03	1.82E-03
Diatom chl-specific initial slope of P–I curve [gC(gchl)−1 (Wm−2)−1 d−1]	xald	0.3	30	11.25	30	14.22	27.03	5.26E-03	2.59E-04
Non-diatom Fe nutrient uptake half-saturation constant [mmolFem−3]	xfln	0.000165	0.00066	0.00033	0.00033	0.0003795	0.0004	5.16E-03	6.13E-05
Non-diatom N nutrient uptake half-saturation constant [mmolNm−3]	xnln	0.1	6	0.5	0.5	1.09	1.0900	5.12E-03	2.10E-04
Meso-zooplankton maximum loss rate [d−1]	xmzme	0.05	0.8	0.2	0.2	0.275	0.2750	4.91E-03	8.59E-05
Maximum micro-zooplankton grazing rate [d−1]	xgmi	0.1	4	2	4	2.39	3.6100	4.89E-03	1.22E-05
O2 consumption by C remineralisation [molO2 (mol C)−1]	xthetarem	0.5	2.5	1.1226	1.1226	1.3226	1.3226	4.84E-03	1.58E-02
Micro-zooplankton maximum loss rate [d−1]	xmzmi	0.05	0.8	0.1	0.1	0.175	0.1750	3.54E-03	2.30E-05
Fast-sinking fraction of diatom silicon grazing losses	xfdfrac3	0	1	0.8	0.8	0.9	0.9000	1.81E-03	3.48E-05
Biogenic Si dissolution length scale [m]	xfastsi	1000	4000	2000	2000	2300	2300	1.10E-03	5.59E-04
Detritus gravitational sinking rate [m s-1]	vsed	1.45E-05	6.94E-05	3.47E-05	2.97E-05	4.02E-05	3.52E-05	7.47E-04	5.07E-04
Excess organic carbon dissolution length scale [m]	xfastc	94	376	188	376	216.2	347.80	7.00E-04	2.22E-04
CaCO3 : POC: export rain ratio scalar Ridgwell et al. (2007)	r0	0.02	0.082	0.026	0.051	0.0322	0.0572	5.12E-04	3.84E-04
Detrital C remineralisation rate [d−1]	xmdc	0.00635	0.038	0.019	0.008	0.022165	0.0112	4.34E-04	2.29E-03
Detrital N remineralisation rate [d−1]	xmd	0.0079	0.0474	0.0237	0.01	0.02765	0.0140	4.13E-04	1.15E-03
Fast-sinking fraction of diatom nat. mort. losses	xfdfrac1	0	1	0.333	0.333	0.433	0.4330	3.81E-04	1.25E-03
Fast-sinking fraction of mesozooplankton mort. losses	xfdfrac2	0	1	1	1	0.9*	0.9000	2.63E-04	2.43E-05
O2 consumption by N remineralisation [molO2 (mol N)−1]	xthetanit	1	4	2	2	2.3	2.3000	2.50E-04	8.14E-04
Si nutrient uptake half-saturation constant [mmolSim−3]	xsld	0.15	6	3	3	3.585	3.5850	1.14E-05	1.64E-05
Calcium carbonate dissolution length scale [m]	xfastca	1750	7000	3500	3500	4025	4025	3.94E-09	8.13E-07

Table D.1: MEDUSA parameter descriptions, abbreviations, bounds, initial **Reference** and **Comb4** parameter values, perturbed **Reference** and **Comb4** parameter values, and combined misfits (see Equation 6.3) for each of the 25 parameter perturbation MEDUSA simulations from both the **Reference** and **Comb4** parameter locations. Each misfit column has been colour formatted to highlight the largest values in that specific column in red, and smallest in blue. Columns 5-6 are also colour formatted, to show which of the **Reference** and **Comb4** configurations have the larger (red) or smaller (blue) parameter values.

	DIN	OXY	ALK	SIL	NPP	DET	DIC	DTC	FER	CHN	CHD	PHN	PHD	PDS	ZMI	ZME	Combined Misfit
Reference	2.76E-02	6.74E-02	2.74E-03	9.22E-02	2.18E-02												4.05E-01
xgme	1.66E-03	2.78E-03	2.80E-05	3.96E-02	3.72E-03	3.66E-01	1.84E-04	3.13E-01	1.56E-03	3.01E-02	1.49E-01	2.72E-02	1.41E-01	1.28E-01	9.63E-02	3.32E-02	4.38E-02
xfld	2.35E-03	3.49E-03	2.55E-05	3.12E-02	1.25E-03	1.96E-02	2.11E-04	1.90E-02	9.66E-04	1.64E-02	1.40E-01	1.19E-02	1.10E-01	9.67E-02	2.47E-02	1.40E-02	2.65E-02
xnld	1.66E-03	2.15E-03	1.97E-05	2.60E-02	1.27E-03	2.16E-02	1.39E-04	2.14E-02	7.34E-04	1.28E-02	1.01E-01	9.42E-03	9.04E-02	8.11E-02	2.03E-02	8.37E-03	2.06E-02
xvpn	8.01E-04	1.09E-03	2.53E-05	2.70E-02	1.54E-03	2.21E-02	8.60E-05	1.99E-02	8.36E-04	2.49E-02	6.84E-02	1.99E-02	6.82E-02	6.04E-02	3.17E-02	1.78E-02	1.99E-02
xaln	8.70E-04	1.34E-03	5.58E-05	1.48E-02	1.68E-03	4.88E-02	1.30E-04	4.22E-02	9.31E-04	3.20E-02	8.13E-02	3.02E-02	7.33E-02	6.67E-02	3.45E-02	2.46E-02	6.26E-03
xvpd	1.58E-03	2.92E-03	2.58E-05	1.52E-02	9.51E-04	1.39E-02	1.71E-04	1.24E-02	6.50E-04	1.19E-02	7.76E-02	1.19E-02	5.68E-02	5.86E-02	1.78E-02	1.09E-02	5.84E-03
xald	1.44E-03	2.27E-03	2.96E-05	1.43E-02	1.10E-03	2.03E-02	1.40E-04	1.89E-02	6.86E-04	1.61E-02	9.91E-02	1.60E-02	8.42E-02	6.35E-02	2.48E-02	1.04E-02	5.26E-03
xfln	7.30E-04	1.01E-03	4.21E-05	1.45E-02	1.26E-03	2.99E-02	1.06E-04	2.70E-02	7.92E-04	4.20E-02	7.58E-02	2.37E-02	7.15E-02	6.12E-02	3.26E-02	1.63E-02	5.16E-03
xnln	9.22E-04	1.40E-03	6.81E-05	1.39E-02	2.32E-03	5.46E-02	1.64E-04	5.14E-02	9.98E-04	4.87E-02	1.30E-01	2.92E-02	1.33E-01	7.22E-02	4.89E-02	1.82E-02	5.12E-03
xmzme	4.53E-04	8.76E-04	8.38E-05	1.42E-02	1.21E-03	3.85E-02	1.53E-04	2.97E-02	5.02E-04	2.28E-02	1.40E-01	1.24E-01	5.92E-02	6.79E-02	9.46E-02		4.91E-03
xgmi	6.18E-04	9.28E-04	5.04E-05	1.40E-02	1.06E-03	3.17E-02	1.03E-04	2.65E-02	6.94E-04	3.55E-02	8.19E-02	4.62E-02	7.12E-02	5.43E-02	2.62E-02	1.33E-02	4.89E-03
xthetarem	0.00E+00	1.41E-02	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	4.84E-03
xmzmi	5.54E-04	6.76E-04	1.09E-04	1.13E-02	8.57E-04	5.76E-02	1.23E-04	3.90E-02	4.98E-04	3.81E-02	4.16E-02	3.18E-02	4.02E-02	3.50E-02	6.28E-02	1.12E-02	3.54E-03
xfdfrac3	3.35E-04	3.69E-04	8.48E-06	8.47E-03	1.28E-04	2.32E-03	2.55E-05	1.79E-03	1.62E-04	1.68E-03	4.50E-03	1.31E-03	3.63E-03	2.06E-02	3.08E-03	1.57E-03	1.81E-03
xfastsi	2.90E-04	3.35E-04	3.32E-06	6.85E-03	5.21E-05	7.65E-04	2.02E-05	7.61E-04	1.30E-04	5.34E-04	1.44E-03	4.78E-04	1.19E-03	8.75E-03	9.78E-04	6.80E-04	1.10E-03
vsed	7.81E-04	2.93E-03	2.50E-04	4.23E-03	1.31E-03	1.89E-01	4.38E-04	1.62E-01	6.91E-04	1.13E-02	2.16E-02	1.17E-02	2.16E-02	2.07E-02	1.65E-02	2.17E-02	7.47E-04
xfastc	1.90E-03	2.37E-03	1.52E-04	4.07E-03	1.12E-03	1.65E-02	1.77E-04	1.55E-02	1.63E-03	9.79E-03	2.52E-02	1.13E-02	2.13E-02	1.79E-02	1.27E-02	1.52E-02	7.00E-04
r0	2.46E-03	2.81E-03	6.26E-04	1.75E-03	6.75E-04	1.04E-02	5.89E-04	1.00E-02	7.35E-04	6.41E-03	8.79E-03	5.75E-03	7.61E-03	6.13E-03	6.24E-03	7.48E-03	5.12E-04
xmdc	1.09E-04	3.84E-03	6.28E-06	1.68E-03	3.20E-04	4.91E-03	4.50E-04	1.43E-01	1.16E-04	3.29E-03	9.40E-03	2.35E-03	8.65E-03	6.37E-03	5.76E-03	3.20E-03	4.34E-04
xmd	6.27E-04	2.10E-03	1.11E-04	3.13E-03	1.30E-03	1.48E-01	2.63E-04	2.57E-02	5.66E-04	8.88E-03	1.99E-02	7.82E-03	1.56E-02	1.47E-02	1.37E-02	1.28E-02	4.13E-04
xfdfrac1	3.22E-04	6.42E-04	1.75E-05	3.76E-03	3.04E-04	6.38E-03	4.47E-05	5.63E-03	1.31E-04	5.66E-03	1.37E-02	4.44E-03	1.04E-02	1.49E-02	9.99E-03	6.29E-03	3.81E-04
xfdfrac2	6.15E-04	7.12E-04	5.78E-05	2.92E-03	8.52E-04	2.15E-02	7.33E-05	1.65E-02	3.13E-04	5.23E-03	1.84E-02	4.90E-03	1.25E-02	1.22E-02	5.64E-03	2.26E-02	2.63E-04
xthetanit	0.00E+00	3.19E-03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	2.50E-04
xsld	2.54E-05	3.17E-05	8.49E-07	6.18E-04	1.44E-05	2.12E-04	2.37E-06	1.88E-04	1.20E-05	1.48E-04	1.30E-03	1.10E-04	7.71E-04	2.61E-03	2.24E-04	1.20E-04	1.14E-05
xfastca	6.13E-06	7.57E-06	3.09E-06	1.82E-06	9.77E-07	1.64E-05	2.33E-06	1.58E-05	3.54E-06	5.29E-06	1.04E-05	5.55E-06	9.57E-06	7.42E-06	7.66E-06	9.56E-06	3.94E-09

	DIN	OXY	ALK	SIL	NPP	DET	DIC	DTC	FER	CHN	CHD	PHN	PHD	PDS	ZMI	ZME	Combined Misfit
Opt4	3.51E-02	3.50E-02	1.85E-03	5.34E-02	2.54E-02												1.37E-01
xthetarem	0.00E+00	2.26E-02	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.58E-02
xnld	3.53E-03	6.12E-03	2.47E-04	2.05E-03	8.39E-03	1.18E-01	3.49E-04	1.08E-01	2.04E-03	5.06E-01	7.13E-02	4.60E-01	4.09E-02	3.63E-02	1.67E-01	4.20E-02	3.06E-03
xmdc	1.53E-04	8.89E-03	7.90E-05	6.58E-05	4.48E-04	1.25E-02	7.97E-04	1.99E-01	8.68E-05	2.91E-02	5.69E-03	2.77E-02	5.23E-03	5.40E-03	1.92E-02	1.13E-02	2.29E-03
xvpd	2.39E-03	5.81E-03	6.89E-05	3.56E-03	2.59E-03	3.11E-02	2.38E-04	2.75E-02	1.16E-03	7.58E-02	4.39E-02	7.12E-02	2.76E-02	3.03E-02	5.89E-02	3.53E-02	1.82E-03
xfld	2.58E-03	5.62E-03	6.37E-05	3.34E-03	3.95E-03	5.21E-02	2.43E-04	4.55E-02	1.33E-03	1.40E-01	5.69E-02	1.13E-01	3.03E-02	3.24E-02	7.01E-02	4.26E-02	1.80E-03
xfdfrac1	2.41E-03	4.73E-03	2.94E-04	2.04E-03	3.96E-03	8.29E-02	3.63E-04	7.55E-02	1.10E-03	4.60E-02	4.06E-02	4.65E-02	4.10E-02	4.99E-02	1.00E-01	6.96E-02	1.25E-03
xmd	1.35E-03	5.79E-03	2.24E-04	4.38E-04	3.46E-03	2.05E-01	5.82E-04	5.12E-02	1.01E-03	7.13E-02	1.89E-02	6.22E-02	1.74E-02	1.62E-02	4.40E-02	3.21E-02	1.15E-03
xthetanit	0.00E+00	5.14E-03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	8.14E-04
xgme	1.33E-03	3.79E-03	1.03E-04	3.98E-04	3.49E-03	5.19E-02	2.53E-04	3.66E-02	9.10E-04	1.44E-01	8.08E-02	1.55E-01	1.10E-01	1.01E-01	1.33E-01	1.84E-01	7.97E-04
xfastsi	1.91E-04	3.02E-04	7.26E-06	4.99E-03	4.08E-04	7.43E-03	1.58E-05	6.30E-03	1.42E-04	2.25E-02	6.18E-03	1.98E-02	5.01E-03	1.84E-02	1.23E-02	9.25E-03	5.59E-04
vsed	1.16E-03	3.39E-03	4.08E-04	1.04E-03	2.46E-03	1.40E-01	4.52E-04	1.14E-01	8.11E-04	3.85E-02	2.02E-02	3.55E-02	1.93E-02	1.66E-02	5.85E-02	3.64E-02	5.07E-04
r0	1.68E-03	2.68E-03	5.12E-04	1.38E-04	9.15E-04	1.31E-02	4.76E-04	1.18E-02	6.53E-04	2.15E-02	6.99E-03	2.13E-02	6.04E-03	5.09E-03	1.48E-02	9.18E-03	3.84E-04
xald	8.55E-04	1.99E-03	2.74E-05	1.17E-03	2.05E-03	2.91E-02	8.83E-05	2.51E-02	5.99E-04	6.21E-02	2.73E-02	4.87E-02	1.77E-02	1.91E-02	2.83E-02	2.32E-02	2.59E-04
xfastc	1.34E-03	2.15E-03	1.39E-04	3.08E-04	1.22E-03	1.36E-02	1.23E-04	1.27E-02	8.14E-04	2.72E-02	8.99E-03	2.23E-02	8.32E-03	6.52E-03	2.13E-02	1.45E-02	2.22E-04
xnln	6.20E-04	1.16E-03	4.81E-05	5.72E-04	2.57E-03	3.46E-02	7.30E-05	3.07E-02	5.29E-04	1.72E-01	1.80E-02	1.68E-01	1.73E-02	1.79E-02	7.68E-02	2.26E-02	2.10E-04
xaln	3.79E-04	6.10E-04	5.57E-05	4.06E-04	2.63E-03	4.36E-02	6.28E-05	3.73E-02	3.74E-04	2.60E-01	1.98E-02	2.14E-01	1.50E-02	1.59E-02	6.74E-02	1.75E-02	1.55E-04
xmzme	5.93E-04	1.03E-03	3.62E-05	4.43E-04	1.50E-03	2.08E-02	7.22E-05	1.75E-02	3.73E-04	7.80E-02	2.97E-02	7.26E-02	3.22E-02	3.56E-02	4.41E-02	1.05E-01	8.59E-05
xfln	3.39E-04	6.26E-04	3.25E-05	5.07E-04	1.31E-03	2.04E-02	4.48E-05	1.76E-02	2.66E-04	1.20E-01	1.38E-02	1.01E-01	1.17E-02	1.36E-02	3.97E-02	1.57E-02	6.13E-05
xfdfrac3	2.68E-04	2.46E-04	8.27E-06	9.75E-04	6.20E-04	9.42E-03	1.69E-05	8.45E-03	1.90E-04	3.04E-02	5.76E-03	2.65E-02	5.41E-03	1.61E-02	1.83E-02	9.50E-03	3.48E-05
xfdfrac2	4.20E-04	4.83E-04	1.58E-05	2.47E-04	8.61E-04	1.20E-02	3.18E-05	1.03E-02	2.29E-04	2.36E-02	6.69E-03	1.88E-02	7.36E-03	5.48E-03	2.22E-02	1.49E-02	2.43E-05
xmzmi	3.36E-04	4.61E-04	1.34E-04	4.70E-04	7.15E-04	3.42E-02	1.39E-04	2.77E-02	2.83E-04	3.96E-02	1.19E-02	4.49E-02	8.81E-03	8.97E-03	1.01E-01	1.62E-02	2.30E-05
xsld	1.06E-04	1.30E-04	3.64E-06	4.55E-04	6.18E-04	8.73E-03	7.81E-06	7.59E-03	1.31E-04	1.98E-02	8.24E-03	1.77E-02	6.05E-03	1.07E-02	1.28E-02	9.69E-03	1.64E-05
xvpn	9.76E-05	1.63E-04	5.65E-06	2.77E-04	7.08E-04	1.29E-02	1.01E-05	1.06E-02	9.70E-05	5.59E-02	8.59E-03	5.03E-02	8.19E-03	7.51E-03	2.50E-02	9.91E-03	1.45E-05
xgmi	1.59E-04	2.67E-04	6.06E-05	4.18E-04	4.53E-04	1.70E-02	7.00E-05	1.55E-02	1.26E-04	3.04E-02	5.98E-03	3.32E-02	5.65E-03	5.89E-03	1.52E-02	8.56E-03	1.22E-05
xfastca	2.50E-05	4.57E-05	1.12E-05	1.07E-05	1.22E-04	1.81E-03	8.79E-06	1.50E-03	1.80E-05	3.96E-03	1.80E-03	3.57E-03	1.82E-03	1.90E-03	2.92E-03	5.66E-03	8.13E-07

Table D.2: Misfit results for the 25 perturbed parameter simulations at locations (top) **Reference** and (bottom) **Comb4** in parameter space. (Column 1) Perturbed parameter, (Columns 2:17) averaged misfits (see Equation 6.2) for each of the 16 tracers, and (Column 18) combined misfits (see Equation 6.3). Each column has been colour formatted to highlight the largest values in that specific column in red, and smallest in blue.

Comb4 because the **Reference** default parameters were unlikely to have been in or near a minimum of the misfit function, therefore were more likely to lie on a steep gradient than the optimised **Comb4** parameters. As **Comb4** lies in a minimum, the shape of the misfit function is likely to be more like a shallow bowl around this point, with a smaller gradient around this location, resulting in smaller perturbation sensitivities.

E

Investigative MEDUSA-2.0 twin optimisations

E.1 Introduction

Here shown are the investigative twin optimisation experiments carried out prior to the experiments done in Chapter 7. These experiments were similar to those in Chapter 7, but DFO-LS was used to optimise 15 parameters instead of 8, and the weighting within the misfit formulation varied slightly (see figure captions for this information). For parameter descriptions see Table 6.2. For parameter bounds within which the optimisation of each parameter was constrained, see the y-axes limits of Figures E.1 to E.4, as they vary slightly from the bounds stated in Table 6.2.

E.2 Results

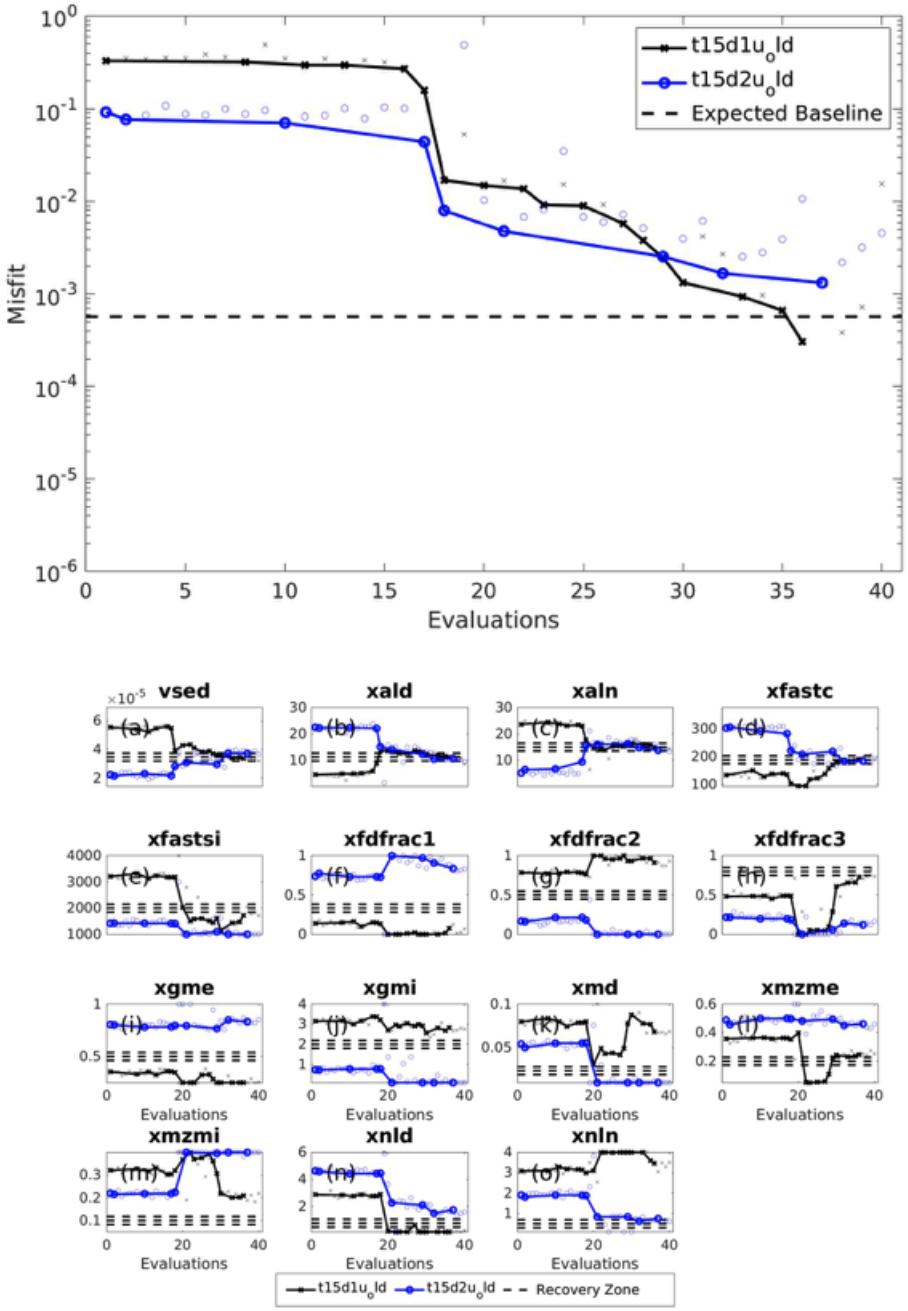

Figure E.1: The misfit reduction and parameter calibration of the twin DFO-LS optimisation experiments to optimise 15 parameters within MEDUSA-2.0. The top plot shows the global misfit reduction with the expected baseline misfit (horizontal black dotted line), and lower subplots (a-o) the parameter optimisation progress of experiments initiated from 2 different locations in parameter space. DFO-LS calibrated to synthetic observations of dissolved inorganic nitrogen, oxygen, alkalinity, silicate and primary production, with primary production weighted equally to the other 4 tracers, despite only having surface observations. Plotted on subplots (a-o) is the parameter recovery zone (horizontal black dashed lines), which marks +-5% of the allowed parameter range either side of the reference parameter values, within which the target for that parameter is considered to be recovered.

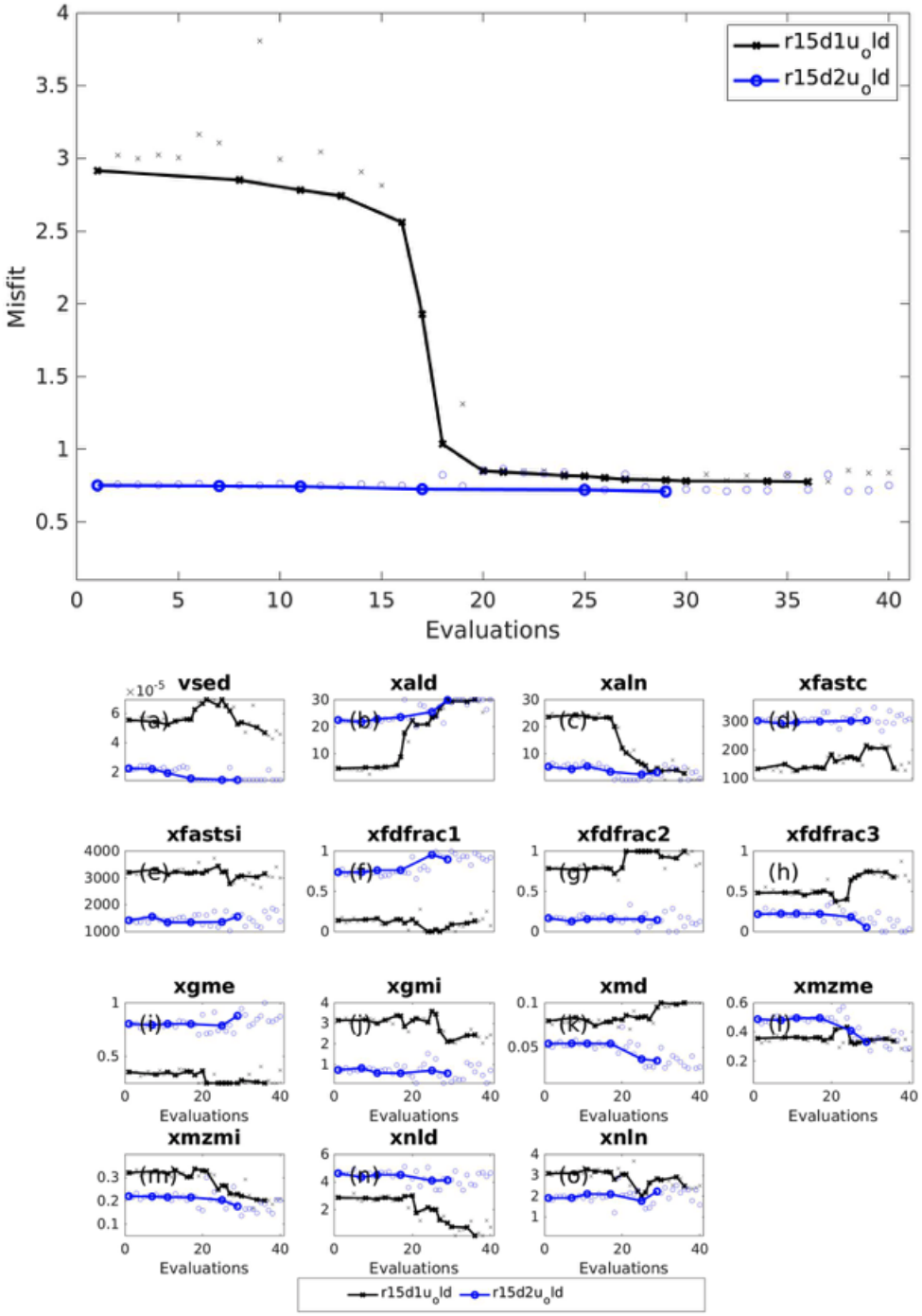

Figure E.2: The misfit reduction and parameter calibration of the real DFO-LS optimisation experiments to optimise 15 parameters within MEDUSA-2.0. The top plot shows the global misfit reduction, and lower subplots (a-o) the parameter optimisation progress of experiments initiated from 2 different locations in parameter space. DFO-LS calibrated to real observations of dissolved inorganic nitrogen, oxygen, alkalinity, silicate and primary production, with primary production weighted equally to the other 4 tracers, despite only having surface observations.

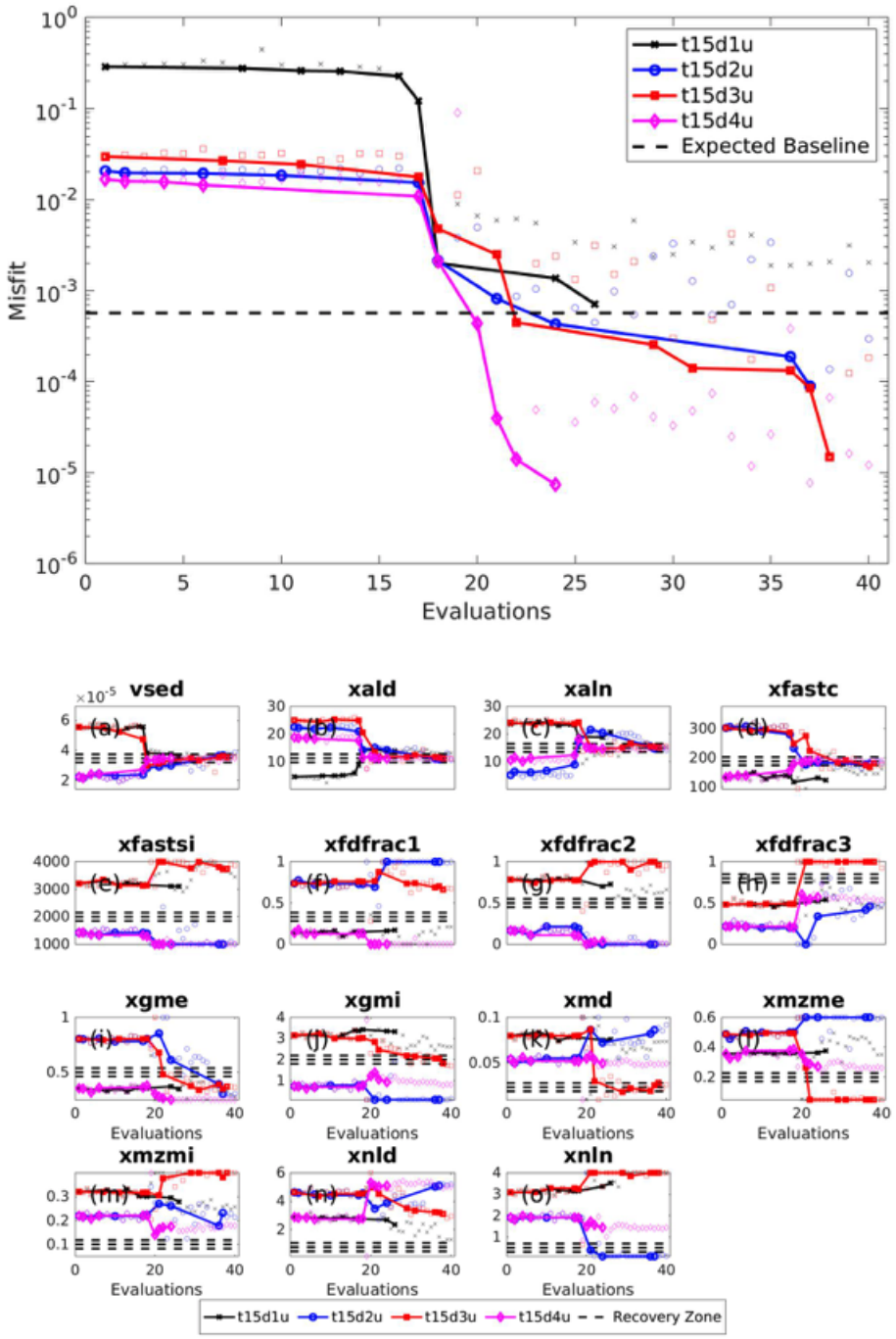

Figure E.3: The misfit reduction and parameter calibration of the twin DFO-LS optimisation experiments to optimise 15 parameters within MEDUSA-2.0. The top plot shows the global misfit reduction with the expected baseline misfit (horizontal black dotted line), and lower subplots (a-o) the parameter optimisation progress of experiments initiated from 4 different locations in parameter space. DFO-LS calibrated to synthetic observations of dissolved inorganic nitrogen, oxygen, alkalinity, silicate and primary production, with primary production weighted less that the other 4 tracers as there are only surface observations of NPP. Plotted on subplots (a-o) is the parameter recovery zone (horizontal black dashed lines), which marks +-5% of the allowed parameter range either side of the reference parameter values, within which the target for that parameter is considered to be recovered.

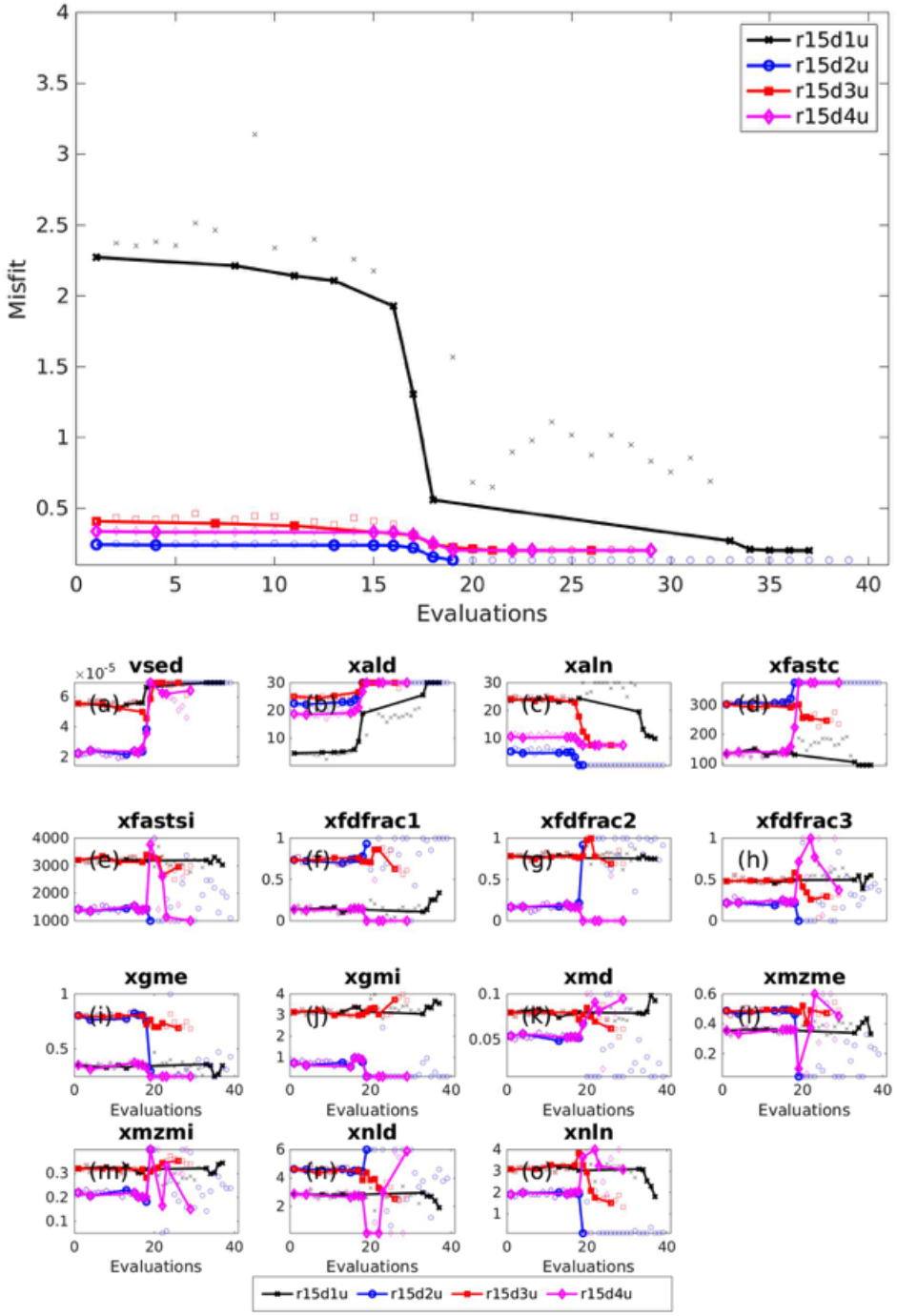

Figure E.4: The misfit reduction and parameter calibration of the real DFO-LS optimisation experiments to optimise 15 parameters within MEDUSA-2.0. The top plot shows the global misfit reduction, and lower subplots (a-o) the parameter optimisation progress of experiments initiated from 4 different locations in parameter space. DFO-LS calibrated to real observations of dissolved inorganic nitrogen, oxygen, alkalinity, silicate and primary production, with primary production weighted less than the other 4 tracers as there are only surface observations of NPP

F

A comparison of MEDUSA-2.0 model outputs when coupled to various online and offline general circulation models

F.1 Introduction

Here we show comparisons of "online" MEDUSA-NEMO and "offline" MEDUSA-TMM outputs of nitrate, silicate, oxygen and alkalinity, which are discussed in Chapter 7.

F.2 Methods

MEDUSA-2.0 (Yool et al., 2013) was coupled to two different "offline" general circulation models using the Transport Matrix Method (Khatiwala et al., 2005; Khatiwala, 2007, 2018):

- UVIC, from the ocean model from version 2.9 of the University of Victoria Earth System Climate Model(Weaver et al., 2001; Eby et al., 2009), and
- ECCO, derived from the MITgcm model (Marshall et al., 1997) and optimised to observations (Stammer et al., 2004).

See Section 2.2.1 for more information on the available transport matrices. These two "offline" model setups were compared to the much more computationally expensive model setup of MEDUSA-2.0 coupled to the "online" NEMO circulation (Madec, 2008), and also compared to observations, shown in Figures F.1-F.4.

F.3 Results

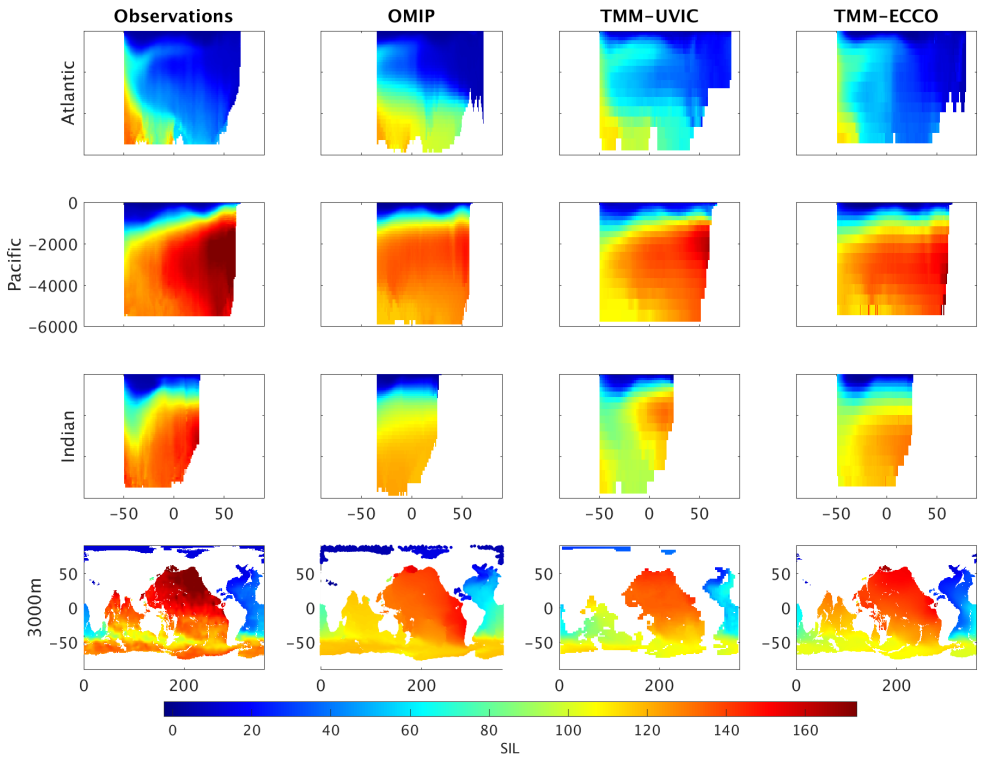

Figure F.1: Distributions of silicate [mmolSi/m^3] by (from left to right) observations (Garcia et al., 2018b), the MEDUSA-NEMO configuration, the MEDUSA-TMM-UVIC configuration, and the MEDUSA-TMM-ECCO configuration, plotted (from top to bottom) for the zonally-averaged Atlantic, Pacific and Indian oceans, and the global ocean at 3000m depth. Note all use the **Reference** parameter values.

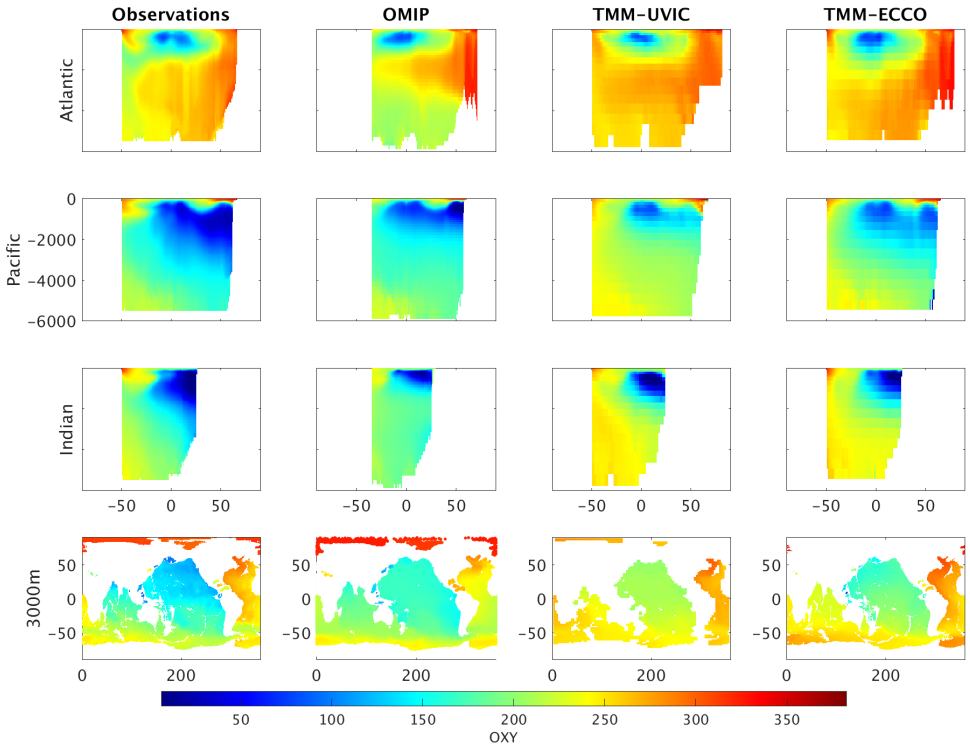

Figure F.2: Distributions of oxygen [mmolO_2/m^3] by (from left to right) observations (Garcia et al., 2018a), the MEDUSA-NEMO configuration, the MEDUSA-TMM-UVIC configuration, and the MEDUSA-TMM-ECCO configuration, plotted (from top to bottom) for the zonally-averaged Atlantic, Pacific and Indian oceans, and the global ocean at 3000m depth. Note all use the **Reference** parameter values.

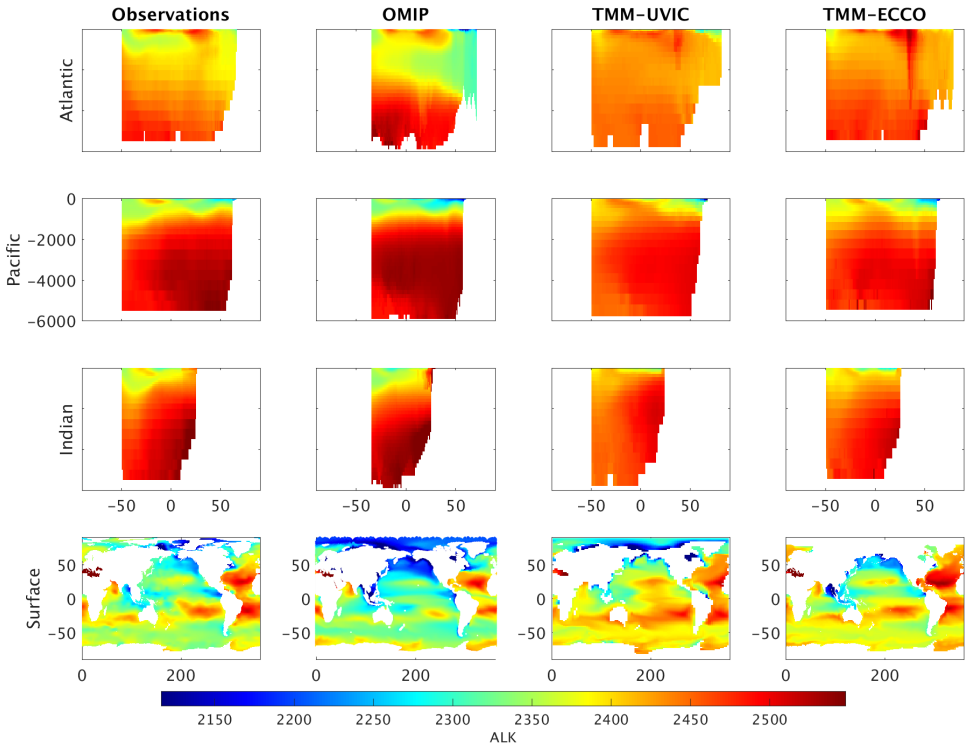

Figure F.3: Distributions of alkalinity [meq/m^3] by (from left to right) observations (Lauvset et al., 2016), the MEDUSA-NEMO configuration, the MEDUSA-TMM-UVIC configuration, and the MEDUSA-TMM-ECCO configuration, plotted (from top to bottom) for the zonally-averaged Atlantic, Pacific and Indian oceans, and the global surface ocean. Note all use the **Reference** parameter values.

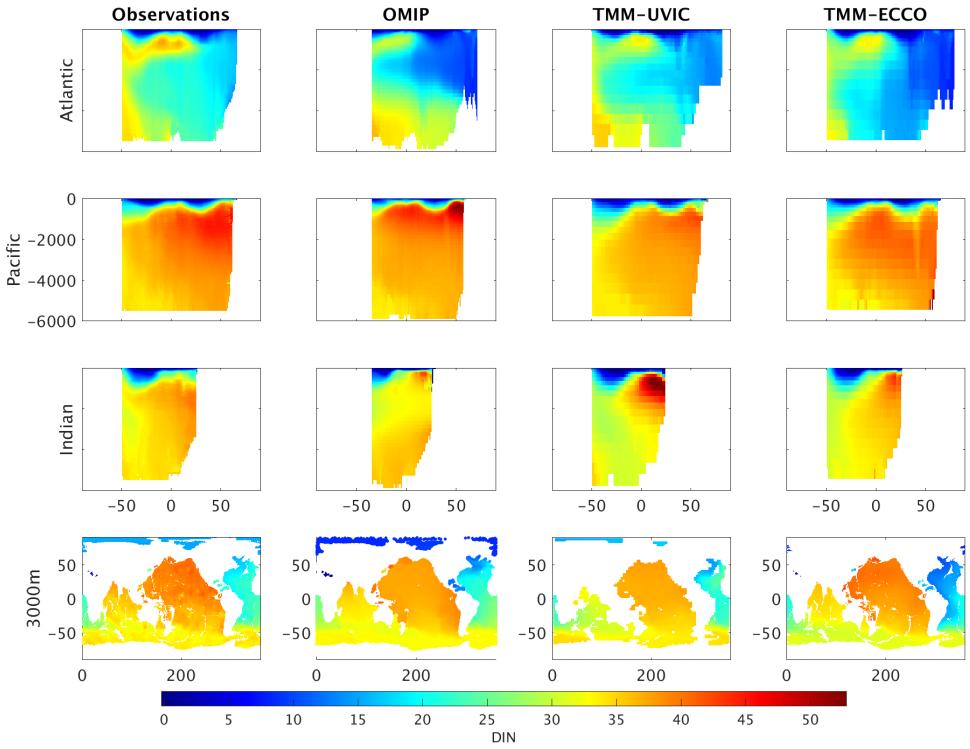

Figure F.4: Distributions of nitrate [mmolN/m^3] by (from left to right) observations (Garcia et al., 2018b), the MEDUSA-NEMO configuration, the MEDUSA-TMM-UVIC configuration, and the MEDUSA-TMM-ECCO configuration, plotted (from top to bottom) for the zonally-averaged Atlantic, Pacific and Indian oceans, and the global ocean at 3000m. Note all use the **Reference** parameter values.