

CHRISTIAN-ALBRECHTS-UNIVERSITÄT ZU KIEL

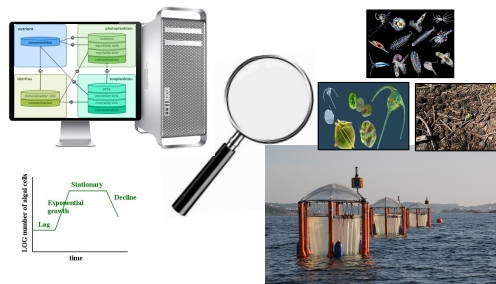


DOCTORAL THESIS

Model assessment and model-based data analyses of an ocean acidification mesocosm experiment

Author:
Shubham Krishna

Supervisor:
Prof. Andreas Oschlies
Dr. Habil. Markus Pahlow



*A thesis submitted in fulfillment of the requirements
for the degree of Dr.rer.nat.*

in the

*Faculty of Mathematics and Natural Sciences
at Christian-Albrechts-Universität zu Kiel*

*submitted by
Shubham Krishna*

December 5, 2017

First referee: Prof. Andreas Andreas Oschlies

Second referee: Dr. Habil. Markus Pahlow

Date of defense: 30/01/2018

Contents

Abstract	ix
Zusammenfassung	xi
1 Introduction	1
1.1 Stressors not directly related to OA	2
1.2 Ocean acidification as stressor	3
1.2.1 Ocean productivity	4
1.2.2 Calcification	4
1.2.3 Nitrogen fixation	5
1.2.4 Cellular stoichiometry of autotrophs	5
1.2.5 Zooplankton	5
1.2.6 Marine biodiversity, habitat and food web dynamics	6
1.3 Plankton ecosystem models	6
1.3.1 Monod-type models	7
1.3.2 Quota-type models	8
1.4 Making inference from experimental studies	9
1.4.1 Laboratory experiments	9
1.4.2 Mesocosm experiments	10
1.4.3 Analyses and interpretation of mesocosm data	10
1.5 Combining mesocosm data with models	12
1.5.1 Data assimilation	12
1.5.2 The cost function	13
1.5.3 Parameter optimisations	13
1.5.4 Uncertainty estimation	16
1.6 Thesis overview and contributions of authors	17
2 Metrics for the estimation of parameters of a plankton ecosystem model with mesocosm data	19
2.1 Introduction	19
2.1.1 Constraining parameter values of a plankton ecosystem model	19
2.1.2 Parameters of phytoplankton losses	21
2.2 Theory and methods	22
2.2.1 Data	22
2.2.2 The model	23
2.2.3 Data assimilation approach	28
2.3 Results	31
2.3.1 Parameter estimates and collinearities	32
2.3.2 Model-data comparisons	37
2.3.3 Cross-validation with unassimilated data	39
2.3.4 Simulating available data of zooplankton biomass	39
2.4 Discussion	41
2.4.1 Sensitivity of the data assimilation model to error statistics	41

2.4.2	Loss rates of phytoplankton biomass	42
2.5	Conclusion	44
3	A data–model synthesis to explain variability in calcification observed during a CO₂ perturbation mesocosm experiment.	47
4	Comparison of two carbon-nitrogen regulatory models calibrated with mesocosms data.	75
5	Summary & conclusive discussion	117
5.1	Effects of initial conditions on physio-ecological processes in mesocosm experiments	118
5.2	CO ₂ effects on cellular physiology of photoautotrophs and community level dynamics	118
5.3	Collinearity in model parameter estimates	119
5.4	Model biases	120
6	Outlook: What can be improved for similar future studies.	123
	Bibliography	125
	Acknowledgements	139
	Curriculum Vitae	141
	Declaration of Authorship	143

“There is no substitute for hard work.”

- Thomas A. Edison

Abstract

Ocean acidification (OA) has been dubbed as the “evil twin” of climate change. Studies suggest that OA has dramatic impacts on marine phytoplankton. Mesocosm facilities allow investigations on effects of changes in the carbonate chemistry of sea water on plankton communities in the vicinity of their natural habitats, e.g. Pelagic ecosystem CO₂ enrichment (PeECE) studies. Marine ecosystem models serve as an efficient tool to analyse and interpret mesocosm data, as they use mathematical equations to describe processes controlling dynamics of planktonic ecosystems.

The goal of this thesis is to investigate the effects of OA on phytoplankton growth dynamics by analysing data from an ocean acidification mesocosm experiment using different model approaches. To achieve this data assimilation (DA) methods are applied. These methods yield the optimised model solutions (with optimised parameter values) that maximize the likelihood probability of models explaining mesocosm data. In addition, DA methods estimate the ranges of uncertainty in optimised model parameter values.

In the first study (Chapter 2), the performance of different metrics (cost functions) that maximize the predictive capability of a plankton model are evaluated. Next, an optimality-based model is applied to investigate the large observed variability in calcification and total alkalinity during the PeECE-I experiment (Chapter 3). The model considers an explicit CO₂ dependency of calcification. Three model experiments are set up to simulate growth of bulk phytoplankton and coccolithophores in mesocosms with high, medium and low observed calcification rates.

Skills of two plankton models (OBM and CN-REcoM) that differ in their mechanistic description of nutrient uptake and algal growth are assessed against mesocosm data in the last study (Chapter 4) of this thesis. In contrast to the calcification study, the plankton models that are applied in Chapter 4 do not resolve any CO₂ effects on phytoplankton growth dynamics. The idea is to test whether this neglect of CO₂ dependencies is revealed in differences of model parameter estimates between different CO₂ treatments.

According to DA results, the cost function that is derived from a probabilistic approach and accounts for changes in correlations between observations performs better as metric for model calibration than other types of cost functions (e.g. Root mean squared errors). The model-based data analysis of the PeECE-I experiment suggests that the large variability that was observed in calcification could have been generated due to small differences in initial abundance of coccolithophores during initialisation (filling) of mesocosms. A pattern is seen in the estimates of two physiological parameters, the potential carbon fixation rate (V_0^C) and the subsistence quota (Q_{min}), between the CO₂ treatments for the OBM. It predicts high estimates of V_0^C and Q_{min} for phytoplankton in mesocosms treated with high CO₂ concentrations and vice versa for those in mesocosms with low CO₂. The OBM seems to suggest that OA may enhance carbon fixation rates in phytoplankton, but at the cost of

elevated metabolic stress. However, it is suggested to include mechanistic CO₂ dependencies of nutrient uptake and phytoplankton growth in the OBM for future studies on OA.

Zusammenfassung

Ozeanversauerung wurde bereits als der “böse Zwilling” des Klimawandels bezeichnet. Studien zeigen, dass Ozeanversauerung dramatische Auswirkungen auf marines Phytoplankton hat. Mesokosmen ermöglichen Untersuchungen des Effekts von Änderungen im Karbonatsystem des Meerwassers auf natürliche Planktongemeinschaften in ihrer natürlichen Umgebung, z.B. CO₂-Anreicherungsstudien im pelagischen Ökosystem (PeECE). Marine Ökosystemmodelle dienen als ein effizientes Werkzeug, um Mesokosmosdaten zu analysieren und zu interpretieren, da sie mit mathematischen Gleichungen die Prozesse beschreiben, die die Dynamik in Planktonökosystemen bestimmen.

Das Ziel dieser Dissertation ist die Untersuchung der Effekte von Ozeanversauerung auf die Planktonwachstumsdynamik, indem Daten eines Mesokosmosexperiments mit unterschiedlichen Modellansätzen analysiert werden. Um dies zu erreichen werden Datenassimilationsmethoden (DA) angewendet. Diese Methoden liefern die optimale Modelllösungen (mit optimisierten Parameterschätzungen), die die Likelihood der Modelle maximieren, die die Mesokosmosdaten erklären. Darüber hinaus schätzen die DA Methoden den Unsicherheitsbereich der optimierten Modellparameterwerte ab.

In der ersten Studie (Kapitel 2) wird die Performance von unterschiedlichen Metriken (Kostenfunktionen) bewertet, die die Voraussagefähigkeit des Planktonmodells maximieren. Anschließend wird ein optimalitätsbasiertes Modell verwendet, um die große beobachtete Variabilität der Kalzifizierung und Gesamtalkalinität während des PeECE-I-Experiments zu untersuchen (Kapitel 3). Das Modell berücksichtigt eine explizite CO₂-Abhängigkeit der Kalzifizierung. Drei Modellexperimente werden durchgeführt, um die Phytoplanktonwachstumsdynamik in Mesokosmen mit der beobachteten hohen, mittleren und geringen Kalzifizierungsrate zu simulieren.

Die Fähigkeit zweier Planktonmodelle (OBM und CN-REcoM), die sich in ihrer mechanischen Beschreibung der Nährstoffaufnahme und des Phytoplanktonwachstums unterscheiden, wird in der letzten Studie dieser Dissertation mittels Mesokosmosdaten bewertet (Kapitel 4). Im Gegensatz zu der Kalzifizierungsstudie lösen die Planktonmodelle, die in Kapitel 4 verwendet werden, nicht CO₂-Effekte auf die Phytoplanktonwachstumsdynamik auf. Die Idee ist, zu testen, ob diese Vernachlässigung der CO₂-Abhängigkeiten sich in Unterschieden der Parameterschätzungen für die unterschiedlichen CO₂-Behandlungen zeigen.

Die DA-Ergebnisse zeigen, dass die Kostenfunktion, die von einem probabilistischen Ansatz abgeleitet ist und die Änderungen der Korrelationen zwischen Beobachtungen berücksichtigt, besser als Metrik für Modellkalibration geeignet ist als andere Arten von Kostenfunktionen (z.B. RMSEs). Die modellbasierte Datenanalyse des PeECE-I-Experiments legt nahe, dass die große Variabilität, die in der Kalzifizierung beobachtet wurde, durch kleine Unterschiede der anfänglichen Konzentration von Coccolithophoriden während der Auffüllung der Mesokosmen verursacht worden sein könnte. Im OBM gibt es

eine Beziehung zwischen den CO_2 -Behandlungen und den Schätzwerten zweier physiologischer Parameter, der potentiellen Kohlenstofffixierungsrate (V_0^C) und der Subsistenzquote (Q_{min}). Es sagt hohe Schätzwerte für V_0^C und Q_{min} in Mesokosmen mit hoher CO_2 -Konzentration voraus und niedrige für Mesokosmen mit niedriger CO_2 -Konzentration. Das OBM deutet darauf hin, dass Ozeanversauerung die Kohlenstofffixierungsrate des Phytoplanktons erhöhen könnte, allerdings einhergehend mit erhöhtem metabolischen Stress. Für zukünftige Studien wird angeregt, auch die mechanistische CO_2 -Abhängigkeiten der Nährstoffaufnahme und des Phytoplanktonwachstums im OBM zu berücksichtigen.

List of Abbreviations

OA	Ocean acidification
BATS	Bermuda Atlantic Time-Series Study
PeECE	Pelagic Ecosystem CO ₂ Enrichment
OBM	Optimality-based Model
LHC	Light harvesting complex
RGR	Relative growth rate
CCM	Carbon concentrating mechanism
CN-REcoM	Carbon:Nitrogen Regulated Ecosystem Model
NPZD	Nitrogen Phytoplankton Zooplankton Detritus
C	Carbon
N	Nitrogen
DIN	Dissolved inorganic nitrogen
DIC	Dissolved inorganic carbon
DOC	Dissolved organic carbon
DON	Dissolved organic nitrogen
DOM	Dissolved organic matter
POC	Particulate organic carbon
POM	Particulate organic matter
TEP	Transparent exopolymer particles
dCCHO	Dissolved combined carbohydrates
PIC	Particulate inorganic carbon
HC	High calcification
MC	Medium calcification
LC	Low calcification
TA	Total alkalinity
DA	Data assimilation
IC	Initial conditions
RMSE	Root mean squared error
ML	Maximum likelihood
MAP	Maximum a posteriori
LM	Levenberg-Marquardt
GA	Genetic algorithm
MH	Metropolis-Hastings
SANN	Simulated annealing
BFGS	Broyden-Fletcher-Goldfarb-Shanno
MCMC	Markov Chain Monte Carlo
FME	Flexible Modelling Environment
ANOVA	Analysis of variance
SD	Standard deviation
RE	Relative error

To my loving dad (Late Mr. Ratan Srivastava)...

Chapter 1

Introduction

The overall goal of the thesis is to better understand phytoplankton growth dynamics in a changing climate. This is achieved by a combination of numerical models with observations from a mesocosm experiment on ocean acidification (OA). In the following, typical stressors in marine environments are introduced and their potential effects are explained (Sections 1.1 and 1.2). Subsequently, marine plankton ecosystem models are introduced and major modelling approaches of phytoplankton growth are described (1.3). In the end, data assimilation methods to calibrate models with mesocosm data are explained (Section 1.5).

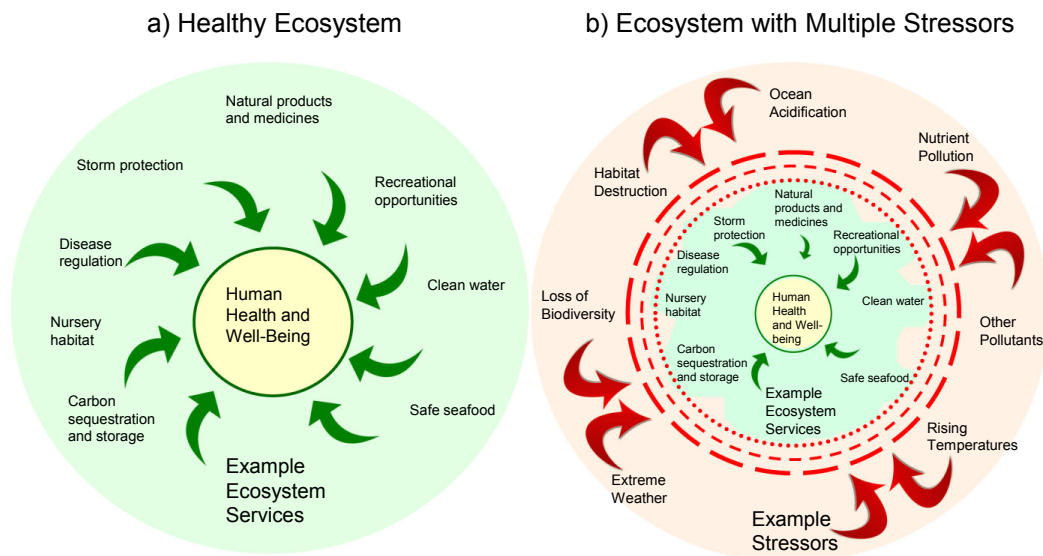


FIGURE 1.1: Comparison of a healthy marine ecosystem with the one that is stressed. Picture illustrates ecological and socio-economic ill effects of marine stressors. Taken from Sandifer and Sutton-Grier, 2014.

The Greek philosopher Heraclitus of Ephesus said, “nothing endures but change”. Indeed, this is true for our planet Earth. Since the beginning of the Anthropocene, Earth’s major ecosystems are dramatically changing. Anthropogenic pressures on the global ocean are believed to increase globally (Fig. 1.1). Nearly 66 % of the global ocean is impacted by abiotic stressors, e.g., abnormal salinities, hypoxia, ultraviolet radiation, pollutants, anomalous high temperatures, and OA (Halpern et al., 2015). Observed ecological responses of marine ecosystems to these multiple stressors are associated with a number of physiological, adaptive and behavioural responses of diverse species. Cumulative effects of environmental stressors on marine biota and habitats are

variable (additive, synergetic and antagonistic effects). In the following, major physiological and ecological impacts of marine stressors on aquatic organisms and their responses are highlighted.

1.1 Stressors not directly related to OA

Homeostases of cellular ionic composition is key to survival and growth in aquatic species. Changes in environmental salinity is believed to disrupt optimal ion regulation, and hence affecting cellular osmotic potential, protein phosphorylation, active transport across membranes and enzyme activity. In general, majority of the marine invertebrates cannot tolerate a wide fluctuation in salinity (stenohaline). Direct impact of changes in salinity at the cellular level triggers two subsequent responses: 1) cellular stress response (CSR), and 2) cellular homeostatic response (CHR). The role of CSR across the wide spectrum of marine species is to repair or compensate for the damage due to protein degradation caused by salinity stress. This is done by activating families of heat shock proteins. Following CSR, CHR prepares the cell to acclimatize to the new physical environment. It restores the ionic balance and cell volume by adjusting concentration of the free amino acid pool. However, pathways within the CSR and CHR are energy demanding and could have implications on cellular physiology and growth. In higher organisms, such as echinoderms, molluscs and crustaceans, impacts of salinity stress on larval development have been widely reported (Coward et al., 2009; Anger, 2003).

The term Hypoxia refers to a physical condition in aquatic environments when the oxygen concentration falls below $2 \text{ mgO}_2 \text{ L}^{-1}$ (Vaquer-Sunyer and Duarte, 2008). Hypoxia (low O_2) is a naturally occurring phenomenon in marine ecosystems, e.g., tide pools and oxygen minimum zones (OMZs). However, anthropogenic perturbations intensify the severity of hypoxia in these environments with subsequent ecological impacts that affect survival. Marine organisms have developed defence mechanisms to tackle environmental hypoxia, e.g., phytoplankton can migrate vertically to escape OMZs, zooplankton enhance activity of anaerobic enzymes (e.g. lactic dehydrogenase), and increased affinity for oxygen in circulatory fluid in crustaceans and fishes. But, exposure to severe hypoxic conditions cause significant reduction in growth in fishes and molluscs and also has negative effects on their reproductive health (Ekau et al., 2010).

Recent changes in ozone layer due to climate change have increase the irradiance of ultra violet radiations(UV-R) reaching earth's surface. Studies have shown that absorption of UV rays cause damage to nucleic acids (e.g. DNA) and proteins in marine organisms (Sinha and Häder, 2002; Van De Poll et al., 2001). When exposed to UV-R photosynthetic pigments in photoautotrophs can overproduce reactive oxygen species that can cause oxidative stress and can have detrimental effects on cellular macromolecules (Lesser, 2010).

Another stressors of concern is the impact due to pollutants. Anthropogenic and natural pollutants in marine ecosystems are increasing and can have harmful effects on organisms. Although species have some form of immune system to eliminate foreign substances from their body, higher concentrations of contaminants have adverse effects on physiology of marine biota. Severe immunotoxicological impacts of environmental pollutants (e.g., toxic metals, benzene, polycyclic aromatic hydrocarbons, and pesticides) have been demonstrated in

variety of marine organisms (Galloway and Depledge, 2001; Segner et al., 2012). Exposure to metal pollutants has a negative effect on the functioning of immune system in fishes. More recently, studies show that lower-trophic level marine organisms (e.g., zooplankton and invertebrates) are susceptible to microplastics (size less than 5 mm) as these pollutants can cause blockage in their digestive systems and may lead to reduced growth and reproductive output (Browne et al., 2008; Graham and Thompson, 2009). Bradshaw et al. (2015) observed changes in plankton community structure and a decline in biomass of *Macoma balthica* (a mollusc) in mesocosms enriched with a chemical pollutant.

The threat of increasing sea surface temperature due to global climate change has increased our interest to investigate impacts of thermal stress on marine organisms. Thermal tolerance limit in ectotherms and their physiological responses to changes in ocean temperature varies spatially. Species from temperate latitudes are adapted to variations in temperature and have broader ranges of thermal tolerances. Whereas, polar and tropical marine species have smaller phenotypic plasticity and seem vulnerable to thermal stress. In marine phytoplankton the rate of protein synthesis increases at high temperatures while number of phosphate-rich ribosomes decreases. Consequently, N:P ratio of algal cells increases in warmer oceans which in turn raises the demand for inorganic nitrogen, and hence affecting marine nitrogen cycle (Toseland et al., 2013).

1.2 Ocean acidification as stressor

The fifth assessment report of the Intergovernmental Panel on Climate Change states with 95 % confidence that global warming is the product of human activity. The report predicts an increase of 40 % in concentrations of anthropogenic carbon dioxide (CO₂) since pre-industrial times (18th century). During the last two decades, the oceans have absorbed about one-third of the CO₂ emitted due to human activities. This resulted in decline of surface ocean total pH by ≈ 0.1 units from 8.2 to 8.1. Furthermore, it is predicted that “if CO₂ emissions are not mitigated” then surface ocean pH could decline by 0.7 units by the year 2300 (Zeebe et al., 2008). Ocean acidification refers to a reduction in the pH of the global ocean over an extended period, on timescale of decades, in response to rising levels of atmospheric CO₂. At the equilibrium, the carbonate system of sea water depends on six variables: dissolved inorganic carbon (DIC), total alkalinity (TA), bicarbonate (HCO₃⁻¹), carbonate (CO₃⁻²) and protons (H⁺). The variables of the ocean carbonate system are quantified over the period of last three decades using state-of-the-art measurement techniques at three sub-tropical ocean-time series stations. DIC and TA have been observed since year 1989 at the station ALOHA of Hawai Ocean Time-Series (HOT). Same variables are measured at the Bermuda Atlantic Time-Series (BATS) station in North Atlantic since 1983, and at European Time-Series station (ESTOC) in Eastern Atlantic between year 1995 to 2004. Data collected at three ocean-time series stations reveal that the decline in surface ocean pH is approximately 0.0017 units every year. In addition, data also exhibit the mean reduction by 0.5 units in concentration of [CO₃⁻²] every year. Future trends suggest that as anthropogenic DIC increases, it will cause a reduction in the buffering capacity of the global ocean. Furthermore, predictions by sophisticated ocean models suggest that by 2100 the ocean’s buffering capacity would drop to 40 % of

what it was in pre-industrial period. Changes in the carbonate chemistry of the seawater can have wide range of effects on marine flora and fauna.

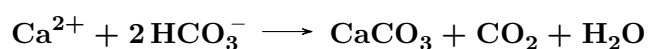
1.2.1 Ocean productivity

Marine primary producers encapsulate diverse phylogenetic groups, ranging from prokaryotes to angiosperms (Falkowski et al., 2004). These organisms differ widely in their light harvesting and carbon enrichment systems (Giordano et al., 2005). For photosynthesis, an algal cell requires sunlight, inorganic carbon, macronutrients (eg. Nitrogen and Phosphorus) and micronutrients (e.g. Iron). Amongst the species in DIC pool of the global ocean, HCO_3^{-1} and CO_3^{-2} ions are more abundant than CO_2 . However, an important substrate for the process of photosynthesis is CO_2 that is catalyzed by the enzyme ribulose-1,5-biphosphate carboxylase oxygenase (RubisCO). But, RubisCO has a low affinity for CO_2 . Therefore, many photoautotrophs employ carbon concentrating mechanisms (CCMs) that use energy to direct inorganic carbon to the proximity of RubisCO (Giordano et al., 2005). Species with effective CCMs are apparently less sensitive to increased CO_2 levels (Raven et al., 2011). Under high CO_2 concentrations many algae are able to downregulate CCM activity. The impact of this will be that less energy is required for the transport of inorganic carbon and more will be available for photosynthesis (Beardall and Giordano, 2002).

A variety of phytoplankton taxa, e.g., diatoms, coccolithophores, cyanobacteria, and dinoflagellates, show enhanced carbon fixation rates and primary production when exposed to elevated $p\text{CO}_2$ levels (Burkhardt and Riebesell, 1997; Barcelos e Ramos et al., 2010; Rost et al., 2006; Hutchins et al., 2009). Investigations of OA on the plankton community also show increase in net community production under high CO_2 levels (Egge et al., 2007). However, with respect to cell division rate (growth rates), responses of phytoplankton to OA are variable. For instance, some negative response was observed on growth rates of coccolithophores in a low pH environment by Barcelos e Ramos et al. (2010) and Iglesias-Rodriguez et al. (2008). But, Burkhardt and Riebesell (1997) and Gervais and Riebesell (2001) found positive effect on cell division rates of diatoms on decreasing the pH.

1.2.2 Calcification

Many marine organisms, including molluscs, echinoderms, corals and a variety of algae, produce exoskeletons, shells and other structures to protect themselves primarily from predation and from photodamage. Calcium carbonate (CaCO_3) is the most common building material used in the building of protective structures. The precipitation of CaCO_3 is supported by high pH and abundant supply of CO_3^{-2} ions, decreasing DIC and TA in ratio of 1:2 and increasing concentration of CO_2 (see the reaction below).



Thus, production of CaCO_3 in the ocean is always accompanied by a reduced uptake of atmospheric CO_2 (Ridgwell et al., 2007; Zondervan et al., 2002). This process serves as counter to Ocean's "biological pump" and termed as "CaCO₃ counter pump" (Gattuso and Hansson, 2011).

At taxonomic level, studies have shown differences in the sensitivity of calcification to changes in carbonate chemistry, e.g., between molluscs, corals, foraminifera, and coccolithophores (Lombard et al., 2009; Comeau et al., 2009; Gao et al., 2012; Engel et al., 2005). However, there is consensus that with decrease in the pH due to OA, energetic cost of calcification is expected to rise (Waldbusser et al., 2010; Monteiro et al., 2016).

Laboratory experiments suggest that unicellular calcifying algae (e.g. coccolithophores) are more vulnerable to OA than multicellular organisms showing a tendency towards reduced CaCO_3 precipitation rates under elevated CO_2 concentrations (Findlay et al., 2011; Meyer and Riebesell, 2015). However, it is not understood whether a decreased calcification rate will impact the competitive fitness of calcifying organisms relative to non-calcifying ones (Gattuso and Hansson, 2011).

1.2.3 Nitrogen fixation

A mechanistic understanding of effects of OA on diazotrophy is limited. Since iron (Fe) and phosphorus (P) are essential nutrients for nitrogen fixation, the impacts of OA on nitrogen fixers are expected to depend on Fe and P availability. In general, an enhanced response in carbon and nitrogen fixation is observed under elevated $p\text{CO}_2$ in majority of diazotrophs, e.g. *Trichodesmium* (Barcelos e Ramos et al., 2007).

1.2.4 Cellular stoichiometry of autotrophs

Only few studies have focussed on effects of CO_2 on algal C:N:P ratios, mainly because of long-standing notion that CO_2 is not limited in marine ecosystems due to rapid diffusion from the atmosphere (Schindler, 1977). In the light of multiple effects of OA on important metabolic processes, e.g., photosynthesis, calcification and nitrogen fixation, it is possible that cellular composition of photoautotrophs alter with strong changes in the carbonate chemistry of seawater. Studies (e.g., Burkhardt and Riebesell 1997; Iglesias-Rodriguez et al. 2008; Bellerby et al. 2008; Hutchins et al. 2009) showed a large variations in cellular C:N and N:P ratios under high CO_2 concentrations. For example, coccolithophores and cyanobacteria consistently showed positive response of OA on cellular C:N:P ratios, whereas cellular composition of dinoflagellates and diatoms remained unaffected under elevated $p\text{CO}_2$ (Leonardos and Geider, 2005; Iglesias-Rodriguez et al., 2008; Fu et al., 2007; Burkhardt et al., 1999). This likely depends on the actual growth conditions (e.g. light availability and nutrient supply).

1.2.5 Zooplankton

Little is known about OA effects on zooplankton relative to phytoplankton. Few studies have focussed on pteropods and foraminifera and they showed decreased calcification under high CO_2 conditions (Fabry et al., 2008; Ellis et al., 2009). Several studies have reported increase in cost of metabolic activities for zooplankton when exposed to high $p\text{CO}_2$ levels (Carotenuto et al., 2007; Edmunds, 2011; Lischka et al., 2011). Furthermore, OA could lead to mobility impairment in many microzooplankton (Caron and Hutchins, 2012). In contrast, no inferable CO_2 effects on zooplankton grazing were observed in a

mesocosms CO₂ perturbation study (Suffrian et al., 2008).

1.2.6 Marine biodiversity, habitat and food web dynamics

Organisms that lack mechanisms to cope up with internal acid-base disruptions are at the greater risk to OA. Mostly lower invertebrates, e.g., sponges, cnidarians, and ctenophores are vulnerable to changes in chemistry of seawater (Gattuso and Hansson, 2011). Amongst invertebrates, echinoderms appear to be least tolerable to low pH. This is evident by their absence from habitats with naturally high CO₂ levels, e.g., hydrothermal vents. Furthermore, it is speculated that many sponges and cnidarians may not be able to survive under OA levels projected for year 2100 under RCP6.0 scenario that predicts an increase in global mean temperature by $\approx 2.2^{\circ}\text{C}$. For calcifying taxa, the magnitude of susceptibility to OA depends on the type of carbonate mineral used for building skeletal structures. High-magnesium calcite organisms are believed to be at highest risk to elevated CO₂ levels. Many coral reef species are perceived to be vulnerable to decrease in pH of seawater. OA may slow down corals' growth by reducing calcification rates and reproduction. Furthermore acidification may also corrode the pre-existing coral skeletons. Since coral reefs are the most ecologically diverse and productive marine ecosystems, decline of such habitats would lead to loss of diverse flora and fauna.

OA is expected to disrupt trophic linkages of food webs, resulting in trophic cascades. An interesting mesocosm study on three-level food web dynamics reveals that under elevated CO₂ primary production and the biomass of top predator may increase, but secondary production would decline (Goldenberg et al., 2017). This is a significant finding as it suggests that OA may knock out some intermediate trophic levels of food web pyramids in marine ecosystems.

In summary, to predict synergistic impacts of OA on dynamics of marine ecosystems and their feedbacks to climate system is a challenging task, and therefore increasing experimental and modelling efforts are required.

1.3 Plankton ecosystem models

One of the key methodologies for predicting responses of aquatic ecosystems to environmental stressors, such as OA, warming and deoxygenation, is to develop mathematical models. Changes of ecosystems can be quantitatively assessed spatially and temporally from prognostic variables, e.g., primary production, grazing pressure, pH of the seawater and the saturation state of aragonite (Blackford, 2010). Furthermore, biogeochemical models can be applied to investigate how physiological-level responses of species to stressors in the marine environment may scale up to community and ecosystem levels. Models provide mathematical tools to describe, elucidate and quantify nutrient fluxes through marine food webs and can help to explain interactions with the atmosphere and ocean sediments. However, these models still contain some considerable degree of uncertainty, particularly since they depend on poorly known model parameters. Models are calibrated with diverse observational data sets. In practice, calibration of models boils down to fitting data by adjusting parameters. It is advisable to limit the number of adjustable parameters, as models

with many fitted parameters can lose their predictive power, which is a fundamental mathematical problem. How complex a model needs to be depends on the scientific question that is being addressed. Essentially, a model should explain data in its simplest form (Occam’s razor). As William of Occam famously said, “It is vain to do with more what can be done with fewer” and “A plurality is not to be posited without necessity”. Conversely, over-simplification in ecosystem models can lead to biases in parameter estimates. Thus, it is critical to find a balance between unwanted details and over-simplification in a marine ecosystem model (Ward et al., 2013).

Today a multitude of models with varying complexity exist to study marine food web dynamics. Simplest example of an aquatic food web model is a NPZD-type model that estimates fluxes of nutrients (N) between phytoplankton (P), zooplankton (Z) and bulk detritus (D) (Fasham et al., 1990). Phytoplankton are the basis of marine food chain as they are the source of food for higher trophic level organisms. Therefore, it is critical for models to represent phytoplankton growth dynamics appropriately when predicting impacts of stressors on marine ecosystem dynamics. In the following two typical approaches to modelling phytoplankton growth are described that differ in their assumptions for nutrient uptake.

1.3.1 Monod-type models

Monod-type models (MMs) relate growth of phytoplankton to external resources (Morel, 1987). MMs define growth rate (μ) of phytoplankton as:

$$\mu = \mu_{max} \cdot \left(\frac{N}{k_{\mu} + N} \right) \quad (1.1)$$

where μ_{max} is the maximum specific growth rate and N is the concentration of the limiting external nutrient. k_{μ} is the half-saturation constant for growth (the nutrient concentration at which growth is half of its maximum). The growth rate becomes zero at zero resource concentration and when N gets very large, μ approaches μ_{max} . Thus, concentration of nutrients in external medium mechanistically regulates the growth of photoautotrophs.

In the last few decades, several MMs for phytoplankton growth have been proposed (Fasham et al., 1990). Amongst them, many first-generation models (e.g. Dugdale 1967) are based on nitrogen and assume constant C:N ratios for estimating mass flux within the ecosystem. These models are known as Redfield–Monod-type models (Ayata et al., 2013). Some global marine ecosystem models (e.g., Faugeras et al. 2004; Aumont and Bopp 2006; Dutkiewicz et al. 2009) employ Monod kinetics for describing nutrient uptake and use the classical Redfield ratio to convert the nitrogen fluxes computed from the phytoplankton compartment to carbon fluxes.

A considerable drawback of the Monod equation describing nutrient limited growth is that values of k_{μ} have to be retrieved from measurements and they can change from one plankton species to another (Droop, 1983). Furthermore, studies have revealed that MMs have limitations when simulating plankton growth in unbalanced growth conditions (Flynn, 2003; Smith et al., 2009; Franks, 2009)

1.3.2 Quota-type models

A major advance in coupling phytoplankton growth with intracellular nutrient-to-carbon ratio (Q) came when it was noticed that μ was linearly related to $1/Q$, also known as the Droop equation (Droop, 1974). This gave birth to Quota-type plankton models (QMs) that define the growth rate of an algal cell as:

$$\mu = \mu_{max} \cdot \left(1 - \frac{Q_{min}}{Q}\right), \quad (1.2)$$

Where Q_{min} is the subsistence quota of a cell. The success of QMs in connecting growth rate to intracellular nitrogen, carbon, and/or phosphorus pool has motivated a variety of additional theoretical developments (e.g. Optimality-based models). These models have refined the Droop equation to consider the plasticity of cellular allocation of resources in different conditions, while investigating changes in C:N:P stoichiometry as adaptive responses to environmental forcing factors (e.g., Armstrong 2006; Pahlow 2005; Smith et al. 2011; Pahlow et al. 2013).

Shuter (1979) proposed one of the first optimality-based models for unicellular algal growth. Pahlow (2005) developed a carbon and nitrogen regulatory model that maximizes phytoplankton growth by optimally allocating resources among competing metabolic requirements for nutrient uptake, light-harvesting and growth. Pahlow and Oschlies (2009) extended the optimality-based formulations of Pahlow (2005) to include multi-nutrient (P and N) and light co-limitation of phytoplankton growth. In contrast to independent limitations by P and N as proposed by Flynn (2001), the model of Pahlow and Oschlies (2009)

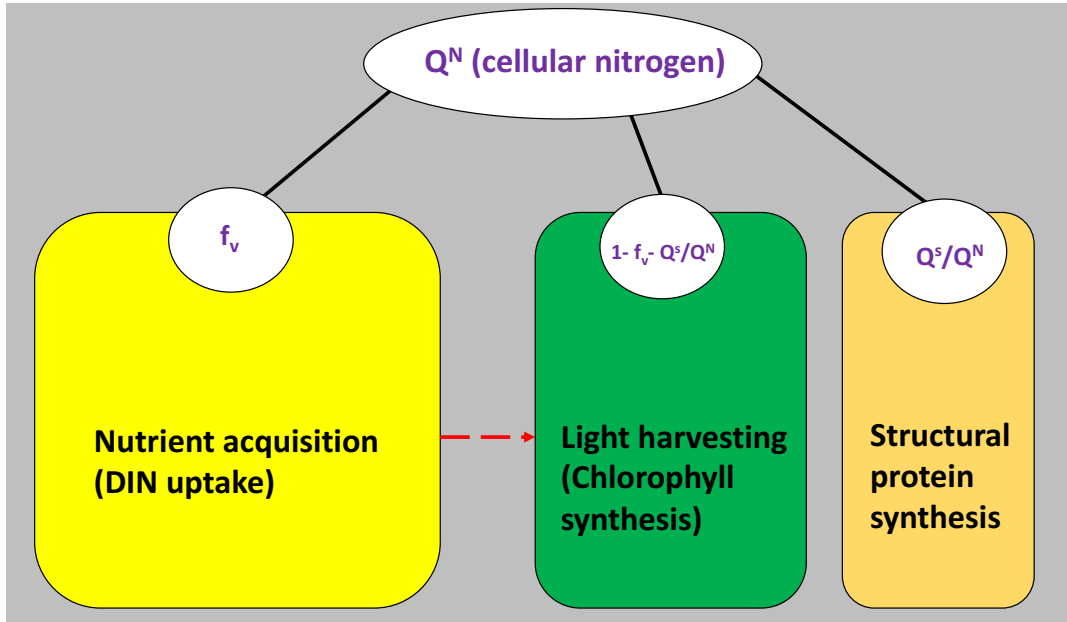


FIGURE 1.2: Schematic representation of allocation of cellular nitrogen (Q^N) for nutrient acquisition, light harvesting and protein synthesis functions of a cell. Fraction of Q^N assigned for DIN uptake is denoted by f_v . Whereas, Q^S/Q^N is the fraction allocated for synthesis of structural proteins. Finally, $1 - f_v - Q^S/Q^N$ accounts for the cellular N designated for light harvesting and carbon fixation. The red dotted arrow indicates that photosynthesis is limited by N exhaustion in the model. This figure is modified from Pahlow et al. (2013).

links C, Chl, N and P through a limitation chain in which P limits N acquisition and N limits photosynthesis, and photosynthesis limits growth. [Pahlow et al. \(2013\)](#) refined the chain model of [Pahlow and Oschlies \(2009\)](#) and introduced an orthogonal hierarchy for an optimal allocation of resources between cellular functions to maximimise the growth rate. At the top level, the cellular N is allocated between nutrient acquisition, light harvesting and structural protein synthesis functions of a cell (Fig 1.2). Subsequently, on the second level, the fraction allocated for nutrient uptake is again distributed between P- and N-acquisition compartments. At the lowest level, the resource for N acquisition is sub-divided to optimise energy demands for nitrogen fixation and DIN uptake in an algal cell. Predictions by the model of [Pahlow et al. \(2013\)](#) are consistent with major observed patterns in diazotrophy in the global ocean.

It is note that the optimality-based model (OBM) applied for the studies in this thesis do not resolve P limitation and N fixation, and hence accounts only for optimal allocation of resources devoted to nutrient acquisition, light harvesting and structural proteins in a cell.

1.4 Making inference from experimental studies

Experimental data are vital for the holistic understanding of effects of marine stressors on marine food web dynamics. They provide information about biological and physical processes over different temporal and spatial scales. Observations of biogeochemical variables are complex data types, coming from a variety of sources, e.g., laboratory experiments, mesocosm experiments, satellites, and in situ cruise observations. Some observational variables are correlated in space and time, e.g. measurements of dissolved inorganic nitrogen (DIN) and Chlorophyll (Chl) concentrations of phytoplankton. Field experiments, e.g. ship-borne measurements, are considered to provide most accurate representation of natural variables, but their availability are limited by poor coverage spatially and temporally ([Gregg et al., 2009](#)). Furthermore, these observations include strong seasonal signals, especially at high latitudes where harsh conditions limit accessibility. In addition, field observations are generally characterized by variability that originates from instrumental, environmental and sampling noise. Whereas, laboratory and mesocoms data are less noisy and easier to sample than collecting in situ measurements.

1.4.1 Laboratory experiments

Laboratory cultures often contain genetically identical species of phytoplankton that are grown under ideal conditions. Controlled, manipulative experiments in laboratories allow an investigation of impacts of stressors on individual species. They can help determining responses on cellular level to environmental perturbations (e.g., [Sunda et al. 2002](#); [Iglesias-Rodriguez et al. 2008](#); [Lohbeck et al. 2012](#)). Hence, laboratory experiments are an effective way of addressing questions like: Which species and communities are at risk? Will organisms adapt to a particular stressed environment? How evolution of an organism is effected by long term changes in physical environment? The major drawback of laboratory data is its representativeness for scaling up to the real world. In removing an organism from its real habitat it is difficult to obtain a natural and normal response to a stressor. Furthermore, these data sets do not provide any

information about changes on community level to variations in environmental conditions.

1.4.2 Mesocosm experiments

Mesocosms enclose a natural plankton community and expose it to environmental perturbations. Data from mesocosm experiments are considered to be more representative of natural situations. By keeping organisms in a more natural environment they may be less vulnerable to stressors than those in small laboratory experiments. Since mesocosms contain variety of biota, investigations on changes in food web dynamics in response to single or multiple stressors can easily be conducted. It has been demonstrated that mesocosm experiments can facilitate rigorous data-model comparisons and data assimilation (Vallino, 2000). The advantages of mesocosm experiments are: (1) They reduce system dimensions from space and time to only time or, if available, to 1D (vertically resolved), and hence greater effort can be placed on development and validation of the model; (2) Typically, mass is conserved in mesocosm experiments, which makes estimation of mass flux easier. (3) Environmental conditions in mesocosms can be better controlled, therefore these experiments prove to be an excellent method to investigate effects of changes in physical and chemical conditions on phytoplankton growth dynamics. Drawbacks of mesocosm experiments are variations and uncertainties in initial conditions that affect the plankton growth dynamics (Krishna and Schartau, 2017).

1.4.3 Analyses and interpretation of mesocosm data

The simplest qualitative method to interpret mesocosm data is graphical analysis. Time-series plots of observations provide useful information on the magnitude and timing of key biological processes, e.g., peak of chlorophyll bloom and exhaustion of nutrients, and also on the variability in the system derived (inferred from standard deviations in data). Line plots, bar charts, boxplots, contour maps and 2D images are some of the popular graphical methods to represent data sets of this kind. Quantitative data analyses include statistical methods, but also dynamical model approaches.

Some of the popular statistical analyses to extract information from mesocosm data are ANOVA (Engel et al., 2005; Egge et al., 2007; Jokiel et al., 2008; Kuffner et al., 2008), Linear regression (Short et al., 1995; Agawin et al., 2000; Findlay et al., 2011), p- and t-tests (Lancaster and Drenner, 1990; Engel et al., 2005; Tanaka et al., 2008) and F-statistics (Strecker et al., 2004; Roy et al., 2006). ANOVA test is conducted to examine the dependency of a particular variable on a certain factor. For example, Jokiel et al. (2008) performed an ANOVA test on their mesocosm data and found a significant negative impact of OA on calcifying coral reef organisms. Findlay et al. (2011) proposed a regression model to quantify the CO₂ effect on calcification by coccolithophores. Statistical significance test is done by computing t- and p-values to demonstrate whether the difference between two groups' means reflects the real difference in populations. A p value less than 0.05 indicates that there is a difference between the means and statistical significance does exists. Tanaka et al. (2008) measured higher concentration of particulate organic matter (POM) in mesocosms treated with high CO₂ concentration than the ones with low CO₂. Their

test statistic ($p < 0.05$) confirmed this result to be significant, indicating positive effect ocean acidification on POM. Engel et al. (2005) tested the influence of CO_2 on biological variables measured during a mesocosm experiment by computing p values. Their results show no effect of CO_2 on early picoplankton bloom ($p = 0.51$), but for calcifying algae (coccolithophores) they observed a significant CO_2 dependency ($p < 0.001$). The average coccosphere size was larger in mesocosms perturbed with low CO_2 concentrations. F-statistics are also employed to find if means of two populations are significantly different. However, F-value gives information on group of variables that are jointly significant. Strecker et al. (2004) applied F-statistics on their results and reported time-dependent negative impact of warming on total zooplankton biomass in their mesocosm experiment.

A different but complementary approach to statistical methods is application of dynamical models. Most models, describing phytoplankton acclimation and growth, have been calibrated with laboratory measurements (Geider et al., 1998; Flynn, 2001; Pahlow and Oschlies, 2009; Armstrong, 2006). On the one hand, Schartau et al. (2017) stressed that laboratory data are good for testing physiological mechanistic behaviour of the model, but inferences on parameter uncertainties remain difficult. On the other hand, ship-borne measurements are temporally and spatially sparse and provide limited information to constrain all the processes resolved in a marine ecosystem model. As mesocosm experiments overcome shortcomings of laboratory and ship-based measurements, configuration of plankton models with data from mesocosms has some advantages. Models are applied to interpret data collected from mesocosms, resolve nutrient and food web dynamics, and estimate mass fluxes in mesocosms (Baretta-Bekker et al., 1998; Vallino, 2000; Dueri et al., 2009; Joassin et al., 2011; Krishna and Schartau, 2017). Baretta-Bekker et al. (1998) calibrated their model with data from mesocosm experiments to study mixotrophy by bacterioplankton. Phosphorus budget estimated by their model suggests that bacteria competed with picoplankton for orthophosphate in nutrient depleted mesocosms. Baretta-Bekker et al. (1998) validated their food web model with mesocosm data and obtained good fits between observations and model results. However, they highlighted the problem of robustness in their model due to sensitivity of parameters on environmental conditions and species abundances. The modelling study of Joassin et al. (2011) classified three stages of coccolithophore (*Emiliana huxleyi*) bloom in mesocosms treated with present-day CO_2 concentrations. During the first phase, increase in biomass of *Emiliana huxleyi* by luxury consumption of nutrients is observed. The second phase is marked by phosphorus exhaustion and decoupling of carbon and nitrogen uptake takes place increasing C:N ratio of algal cells. The last phase is characterised by sharp decline in biomass of *Emiliana huxleyi* as a result of high phytoplankton mortality, which is in their study attributed to viral lysis. Krishna and Schartau (2017) configured an optimality-based model (Pahlow et al., 2013) with data from a mesocosm experiment on ocean acidification. Their results suggest that large variations in calcification seen between the mesocosms of the same CO_2 treatment are mainly due to small differences in initial concentrations of phytoplankton and photoacclimation states at the time of filling of mesocosms.

1.5 Combining mesocosm data with models

1.5.1 Data assimilation

Data assimilation (DA) is a method to combine the information in observations with dynamical models to predict state variables and to estimate model parameters (Dowd, 2007). Data assimilation can provide estimates of prognostic variables that combine strengths of observations and models while reducing their misfits and uncertainties in model results and observations (Fig. 1.3).

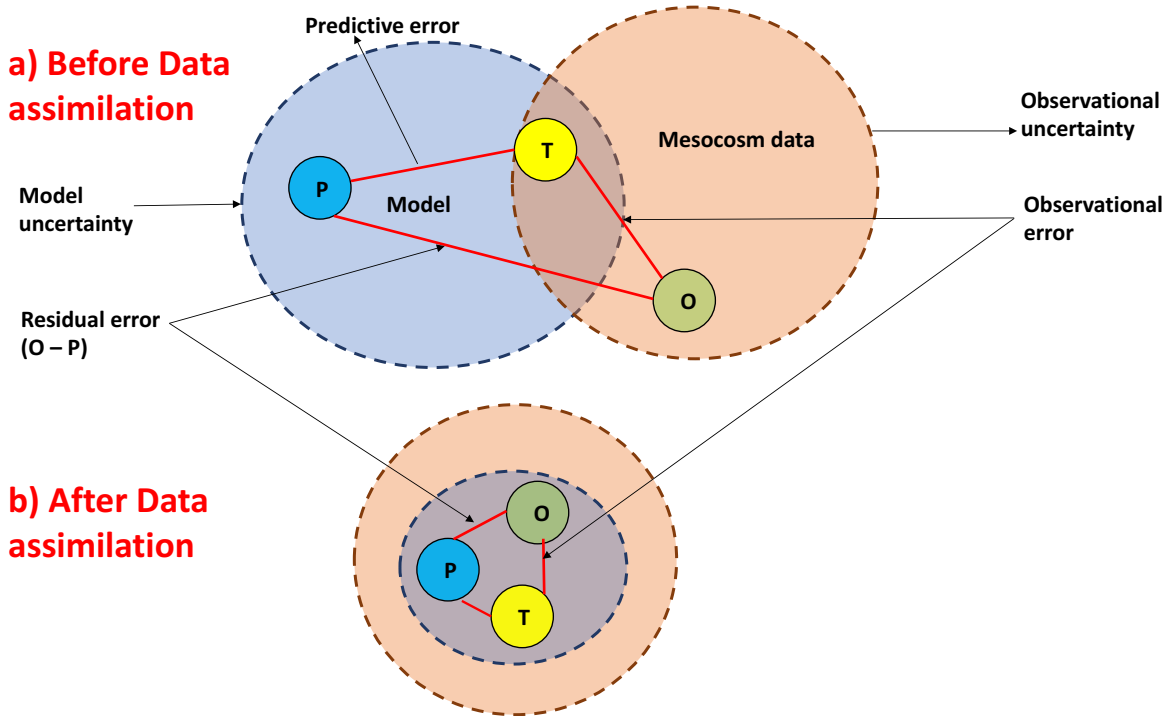


FIGURE 1.3: Schematic representations of model prediction (P), mesocosm observations (O), and true state (T). P and O are enclosed by their respective uncertainties. a) depicts situation before data assimilation when residual and prediction errors are large. b) shows the case after data assimilation when model has been calibrated by observations and model uncertainty has been significantly reduced. This figure has been modified from the one in Stow et al. (2009).

There are different types of data assimilation methods that have been applied. In general, we can distinguish two methodological classes of DA methods (Gregg et al., 2009). The first class typically assures that the models' solutions remain consistent with the dynamical equations imposed, and model parameter values are constant. In this case, only forcing data and model parameters are adjusted (optimised) to provide a model solution that gives a best representation of the data. The second class accounts for the fact that all models are imperfect and these data assimilation method introduce corrections to the models' solutions and sometimes also parameter values at locations and dates of observations (see e.g. Dowd et al. 2014). These “state” corrections not only depend on the misfit between model results and data but also include updated (and sophisticated) error information of the previous corrections. The corrections of model results occur in a temporal sequence at times of available data, and these methods are therefore referred to as sequential methods. Sequential methods are the basis for forecasting systems, e.g. in operational oceanography

or in meteorology (weather forecasts). If only the parameter values of a model are corrected sequentially, then a sequential method can provide temporally variable parameter values while model solutions are consistent with the model dynamics, i.e. assuring mass conservation.

1.5.2 The cost function

The cost function (J) quantifies the discrepancy between model outputs and corresponding observations (Schartau et al., 2017). There are different ways to evaluate model-data misfits, e.g., the root mean squared error (RMSE), the reliability index (RI) and the average error (AE) for univariate analyses (Stow et al., 2009), and weighted “least squares” for multivariate cases (Matear, 1995; Schartau and Oschlies, 2003).

For N number of observations, the residual vector (\vec{d}) at any point of time is given by the difference between vectors of observations (\vec{y}) and its model counterpart ($H(\vec{x})$):

$$\vec{d} = (\vec{y} - H(\vec{x})) \quad (1.3)$$

A non-probabilistic cost function is one that does not account for error information of observations (e.g. RMSE) and it is defined as a function of \vec{d} and N :

$$J = f(\vec{d}, N) \quad (1.4)$$

Whereas, a cost function based on a probabilistic approach (e.g. Negative Loglikelihood) considers errors in data (denoted by the vector $\vec{\epsilon}$) and described as a function of \vec{d} , N and $\vec{\epsilon}$:

$$J = f(\vec{d}, \vec{\epsilon}, N) \quad (1.5)$$

Furthermore, when a probabilistic based cost function is applied to estimate a set of model parameters (θ) then it is defined as:

$$J(\theta) = f(\vec{d}(\theta), \vec{\epsilon}, N) \quad (1.6)$$

1.5.3 Parameter optimisations

The optimal set of parameters (θ^*) is the one that yields the lowest value for the cost function (J) in the n_p dimensional parameter space. An optimum occurs when the gradient of the cost function becomes zero,

$$\left. \frac{\partial J(\vec{\theta})}{\partial \vec{\theta}} \right|_{\vec{\theta}=\vec{\theta}^*} = 0 \quad \left(\text{or } \nabla_{\vec{\theta}} J(\vec{\theta}) \Big|_{\vec{\theta}=\vec{\theta}^*} = 0 \right); \quad (1.7)$$

and the parameter Hessian matrix (H_θ) that contains second-order derivatives of $J(\vec{\theta}^*)$ is positive definite (Vallino, 2000):

$$H_\theta = \left. \frac{\partial^2 J(\vec{\theta})}{\partial \theta^2} \right|_{\vec{\theta}=\vec{\theta}^*} \quad (1.8)$$

The landscape of the cost function is determined by changes in $J(\vec{\theta})$ in response to variations in parameter values (θ). A simple example would be a convex

landscape where the minimum of the cost function lies at the bottom of a single valley. However, for a multi-dimensional parameter space the shape of the cost function is typically multi-modal or noisy (Fonseca and Fleming, 1995), indicating occurrence of many local minima, and it is difficult to determine the global minima of the cost function. There are numerous parameter optimisation algorithms, employing different methods to approximate the global minimum of the cost function. These algorithms can broadly be classified into two categories:

1.5.3.1 Gradient-based algorithms

Gradient-based algorithms make use of the gradient information to minimize the cost function (Gauthier et al., 1986; Snyman, 2005; Fennel et al., 2001). They compute the negative gradient of the cost function at every time step and descent in the direction of decreasing value of $J(\vec{\theta})$ until $\nabla_{\vec{\theta}} J(\vec{\theta})$ becomes zero, and the minimum is located. Gradient-based algorithms are well suited for convex error functions (cost functions). A simple, but common example of a gradient-based algorithm is one that uses the steepest-descent approach (Lawson et al., 1995). The disadvantage of the steepest-descent approach is that they have tendency to get trapped in a local minimum (e.g. when observations are very noisy). More advanced, the conjugate-gradient (CG) algorithm (Gilbert and Nocedal, 1992) finds an optimal search direction by combination of negative gradients of $J(\vec{\theta})$ at the current iteration step and the previous direction. Several studies, e.g., Matear (1995), Fennel et al. (2001), Lawson et al. (1995), have employed gradient-based optimisation techniques to estimate parameters for regional and global marine ecosystem models.

Another example of gradient-based algorithms are Newton and quasi-newton optimisation algorithms (also known as second order optimisation algorithms) that use the curvature (second-derivative) information stored in the Hessian matrix to approximate the global minimum. The Hessian matrix is constructed from gradient evaluations of the cost function. The optimal search direction is given by a vector describing the angle of the direction according to the inverse of H_{θ} . Newton and quasi-Newton algorithms are well-suited for finding the global optimum of convex non-linear cost functions. In theory, these algorithms fit a quadratic approximation to $J(\vec{\theta})$ at a given θ using both gradient and second derivative information. A Taylor expansion of $J(\vec{\theta})$ around the optimum (θ^*) is given as:

$$J(\vec{\theta}) = J(\vec{\theta}^*) + \frac{\partial J(\vec{\theta})}{\partial \vec{\theta}} \delta\theta + \frac{\partial^2 J(\vec{\theta})}{\partial \vec{\theta}^2} \delta\theta^2 + \mathcal{O}(\delta\theta^2) \quad (1.9)$$

Where $\delta\theta$ is a step size computed from the first and second derivatives of the cost function:

$$\delta\theta = -\frac{\nabla_{\vec{\theta}} J(\vec{\theta})}{\nabla_{\vec{\theta}}^2 J(\vec{\theta})} \quad (1.10)$$

The convergence criterion is to find the $\delta\theta$ that minimizes the above described quadratic approximation. Some of the popular Newton and quasi-Newton methods applied in inverse data assimilation approaches are: Levenberg-Marquardt (LM) algorithm (Ranganathan, 2004; Marquardt, 1963), Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (Broyden, 1970; Fletcher, 1970; Goldfarb,

1970; Shannon, 1970) and Davidon-Fletcher-Powell (DVP) algorithm (Leggett, 2013). Lawson et al. (1995) and Spitz et al. (1998) used Newton and quasi-Newton methods to fit their marine ecosystem models to observations collected from the Bermuda Atlantic Time-Series Study (BATS) site. Likewise, Fennel et al. (2001) applied a quasi-Newton method to optimize model parameters in their model sensitivity analysis. Rose et al. (2007) used a gradient-based optimization technique (Levenberg-Marquardt method) to calibrate their ocean food web model.

1.5.3.2 Stochastic algorithms

Gradient-based technique is usually considered to be not efficient for noisy cost functions (Todorov, 2005; Raupach et al., 2005), primarily because computations of the gradients and Hessian of the cost function are not exact. Many stochastic techniques for parameter optimization have been proposed that apply probabilistic techniques for approximating the global optimum. One stochastic technique that has been often used by ecosystem modellers is simulated annealing (SA) method (Bélisle, 1992). The name of this algorithm comes from annealing in metallurgy, a technique in which successive heating and slow cooling of a material increases size of its crystals with a low energy state. The SA method randomly selects a parameter set and evaluates the cost function value. It employs an acceptance probability function that decides whether to accept a particular parameter set or not. This process is iterated in the direction of decreasing cost function values until the global minimum is approximated. Although it is not guaranteed that the SA method will find the global optimum, it has been shown to provide good solutions (Matear, 1995).

Another class of stochastic search strategies are genetic algorithms (GAs). GAs are a heuristic optimisation technique inspired by Darwin's principle of natural evolution. A GA algorithm randomly selects a population of model evaluations on a fitness-based criteria (representing cost function values) and then carries out a process of recombination of parameter vectors to generate a successor population (new sets of parameter vectors). As this process is iterated a sequence of successive generations evolve and average fitness tends to increase (McCall, 2005). Increase in fitness of a population of parameter sets corresponds to lower cost function values. The recursive process of selection and recombination continues until a pre-determined convergence threshold is reached, and this way a GA algorithm finds a best solution to the problem. Schartau and Oschlies (2003) applied a Micro-genetic algorithm (μ GA), a small population size variant of GA, to optimise 13 parameters of a simple NPZD model at three locations in the North Atlantic. Their results show significant reduction in the model-data misfit.

More recently, Covariance Matrix Adaptation Evolution Strategy (CMA-ES) optimisation algorithms are becoming popular (Hansen et al., 2003). These algorithms make use of an approximated covariance matrix to generate a successor population with increased fitness. Kriest et al. (2017) calibrated a global 3D biogeochemical ocean model against global mean observations using a CMA-ES algorithm. They found that the optimised model produces a better fit to observed global biogeochemical fluxes.

1.5.4 Uncertainty estimation

Uncertainties in optimised model solutions can be estimated by finding all those parameter vectors that yield model-data misfit costs that are statistically indistinguishable from the global minimum of the misfit function (e.g. [Schartau et al. 2017](#)), and the distribution of these optimised parameters is called the posterior distribution or the posterior (e.g. [Harmon and Challenor 1997](#)). An inference on identifiability of a model parameter can be made by the spread of its posterior distribution (in the vicinity of the global optimum).

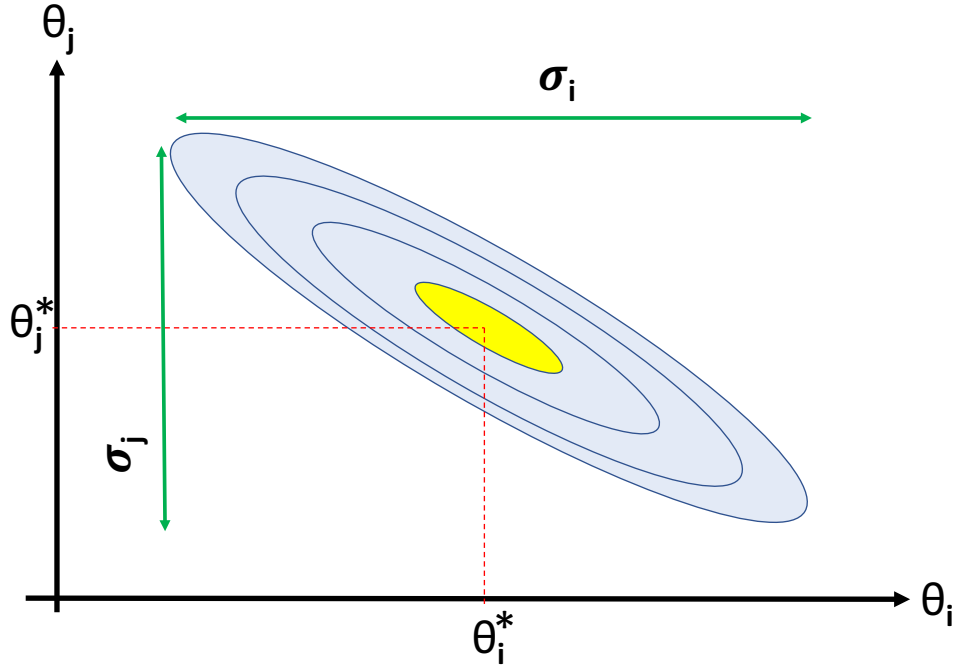


FIGURE 1.4: Graphical illustration of marginal errors (σ_i and σ_j) in posterior estimates of parameters θ_i and θ_j . Contours represent the cost function (or misfit) values while varying θ_i and θ_j . The inner most contour (depicted in yellow) encloses the optimum (minimum) approximated by optimisation algorithms. Whereas, θ_i^* and θ_j^* are the “best” optimised estimates for θ_i and θ_j . The idea of this figure has been taken from [Sivia \(1996\)](#).

The larger the spread of the posterior distribution, the higher the uncertainty associated with the parameter estimate. Uncertainties in parameter estimates are quantified by their marginal error-bars (e.g. σ_i and σ_j in Fig. 1.4). The innermost cost function contour (shaded in yellow) in Figure 1.4 represents the credible region of parameter uncertainties (that includes the optimum). All parameter combinations between θ_i and θ_j in this region are statistically indistinguishable.

Markov Chain Monte Carlo (MCMC) methods are considered as an efficient tool to determine the posterior distribution of parameter estimates ([Harmon and Challenor, 1997](#)). These algorithms employ Bayesian selective (acceptance/rejection) sampling schemes, e.g., Metropolis-Hastings (MH) algorithm ([Metropolis et al., 1953](#); [Hastings, 1970](#)) and Metropolis-Hastings-Delayed-rejection algorithm ([Haario et al., 2006](#)), to subsample the posterior and identify a credible region of random distribution of parameter values, which is interpreted as uncertainties in model solutions ([Dowd et al., 2014](#); [Schartau](#)

et al., 2017). Fiechter et al. (2013) have demonstrated the success of a MCMC method in separating the uncertainties in parameters estimates due to sampling issues and uncertain ecosystem dynamics for a regional marine ecosystem model. Smith and Yamanaka (2007) applied the MH algorithm for a maximum likelihood estimation while comparing two photoacclimation models against culture experiment data of Flynn et al. (1994).

1.6 Thesis overview and contributions of authors

The ultimate scientific objective of this thesis is to test and apply different model approaches to investigate effects of changes in CO₂ concentrations on phytoplankton growth dynamics. A major focus is put on understanding physiological acclimation (e.g. variable stoichiometry) of phytoplankton to changes in physical environment and some mean cellular state. To answer these overarching questions, a workbench in R (RDevelopment, 2012) has been set up that facilitates data-model syntheses. In this modelling framework it is possible to test different model versions against mesocosm experimental data. Models are calibrated with data from a mesocosm experiment on ocean acidification. Data assimilation is one of the key aspects of this thesis as a lot of emphasis is put on identification of metrics that maximise predictive capability of models. All analyses in this thesis were carried out using an optimality-based model (OBM) and Carbon-Nitrogen regulated ecosystem model (CN-REcoM) that includes parameterisations of Geider et al. (1998). The second chapter of the thesis addresses problems with respect to parameter identification in the OBM. Different metrics to quantify discrepancy between data and model results have been evaluated. Eventually, an inference on the most dominating sink pathway of phytoplankton biomass has been made. The best metric identified for the OBM from the chapter two is then applied to understand the cause of observed variability in calcification by coccolithophores between mesocosms treated with identical CO₂ concentrations (Chapter 3). The last chapter of the thesis provides the comprehensive overview of skill assessment of the OBM and CN-REcoM with mesocosm data. This study focusses on CO₂ effects on phytoplankton growth dynamics. Finally, thesis ends with discussion of main results obtained and providing some prospects for future research.

Chapter 2. *Metrics for estimation of loss parameters of plankton ecosystem model with data from an ocean acidification mesocosm experiment.* Shubham Krishna (SK) and Markus Schartau (MS). Original idea was jointly developed by MS and SK. MS provided insight into the experimental design. All model experiments and analyses were performed by SK. The chapter was prepared by SK with the help of comments by MS. The study focusses on the estimation of major loss parameters of phytoplankton biomass for the OBM set up. A comprehensive overview of performances of probabilistic and non-probabilistic based metrics to configure the model with mesocosm data is presented. The best model performance is obtained when it is calibrated with a metric that considers data covariances. According to the optimisation results, the exudation rate of photoautotrophs is the best constrained loss parameters by the data available. Model predicts phytoplankton exudation as some major sink pathway, as a large fraction of phytoplankton carbon is lost via production of labile dissolved organic carbon.

Chapter 3. *A data – model synthesis to explain variability in calcification observed during a CO₂ perturbation mesocosm experiment.* Shubham Krishna and Markus Schartau. The idea for this work and experimental design were developed by MS. All model experiments and data – model analyses were performed by SK. The chapter was prepared by SK which eventually got published after major revision by MS. In this study, data from an ocean acidification mesocosm experiment are reanalysed with the OBM to resolve observed differences in total alkalinity and particulate inorganic carbon (PIC) between mesocosms. The model suggests that small variations in initial abundance of coccolithophores and initial photoacclimation states generate differences in calcification that are larger than those induced by ocean acidification. Data assimilation results also reveal collinearities between initial conditions and physiological model parameters.

Chapter 4. *Comparison of two carbon-nitrogen regulatory models calibrated with mesocosms data.* Shubham Krishna, Markus Pahlow (MP) and Markus Schartau. The scientific question was initiated by MP and MS. The idea for experimental design was developed by MP and SK. All model simulations and analyses were carried out by SK. In this study, skills of OBM and CN-REcoM were assessed with major focus on potential CO₂ effects on plankton growth dynamics that can be interpreted from mesocosm data on ocean acidification. Here it is important to stress that no CO₂ effect was considered in the model approach, which is in contrast to the calcification study. The idea is to check whether this “neglect” becomes visible (or expressed) in differences between parameter estimates. The OBM reveals a systematic dependency of maximum photosynthesis rates and subsistence quota of phytoplankton on $p\text{CO}_2$ levels. Their estimates increase with the rise of CO₂ concentrations. According to the OBM, although carbon fixation rates of algal cells were stimulated in high CO₂ conditions, they face the challenge of high energetic costs for maintenance simultaneously. On the other hand, no such relationship was captured by CN-REcoM, mainly because it remained unconstrained by the data.

Please note, SK also performed some model simulations and optimisations for [Schartau et al. \(2017\)](#) which is not shown here. However, these optimisation results were used in studies described in this thesis.

Chapter 2

Metrics for the estimation of parameters of a plankton ecosystem model with mesocosm data

2.1 Introduction

The complex nature of plankton interaction makes it difficult to correctly interpret experimental data. This is particularly true for the outcomes of mesocosm experiments that typically comprise a natural plankton community, in contrast to e.g. batch experiments with plankton monocultures. To devise and apply dynamical plankton ecosystem models can be particularly helpful in this respect. A plankton ecosystem model resolves mass flux and these models are typically constructed to provide a coherent picture of the prevailing processes that are responsible for the observed changes, e.g. in nutrient and chlorophyll *a* concentrations. However, plankton ecosystem models are far from being perfect. These models include parameterisations whose parameter values need to be calibrated. This is typically achieved by comparing model results with observational data and by an evaluation of the data-model misfit. For the quantification of a data-model misfit some metric has to be defined, which expresses the deviation (or distance) between model solution and the data. Metrics are thus important to find those model solutions that can represent the observational data best.

2.1.1 Constraining parameter values of a plankton ecosystem model

The identification of parameter values for a plankton ecosystem model is a central aspect of model calibration (Schartau et al., 2017). The values that are assigned to the parameters eventually determine model solutions, as they control a variety of physio-ecological processes described in a dynamical plankton model, e.g., nutrient assimilation, photosynthesis, grazing, aggregation and mortality. Hence the choice of model parameter values strongly affects a model's performance. An optimal parameter estimate is the one that yields the minimum of data-model misfit that is quantified by the metric, a distance measure between data and the model results, also referred to as objective function or cost function. The values of many model parameters are usually unknown and difficult to determine as they may actually vary temporally and

spatially (Weir et al., 2013). For instance, optimised parameter values at one location might not be appropriate for the same model to reproduce observations at a different location (Friedrichs et al., 2006; Schartau and Oschlies, 2003). Likewise, parameter estimates that provide a single best model fit to data of some mesocosms of an experiment may not produce similarly good solutions to describe data of other mesocosms of the same experiment. This is why it is important to determine an ensemble of model solutions that represent uncertainty ranges in estimates of model parameters.

A well-constrained parameter is the one that shows small uncertainty in its estimates obtain from model calibrations with diverse data sets. The predictive capability of a model depends on how well its parameters can be constrained by the given data (Baretta et al., 1995; Spitz et al., 1998; Fulton et al., 2004; Schartau et al., 2017). Poorly known model parameters could produce estimates of physical and biological variables that are unacceptable. Models with a large number of unconstrained parameters have lower skills as they are likely to fit noise in data (Friedrichs et al., 2006). Thus, producing erroneous and unreliable model results. Model parameters may remain poorly identified due to several reasons. Often in situ data are sparse in space and time, and not powerful enough to calibrate a model and constrain all its parameters (Matear, 1995). For example, Friedrichs et al. (2007) assimilated primary productivity, chlorophyll and nitrate data in models of varying complexity, and found that models fitted data well but predicted different balances of primary productivity and grazing. Furthermore, they discussed if zooplankton data were available, models would have demonstrated greater skills. When observations are sparse then parameter estimation becomes sensitive to the “prior uncertainty”, and the prior knowledge about parameter values becomes vital. In such cases, a good practice is to gather best estimates of parameters from previous studies that employed similar parameterisations, and use them for prior information in a cost function (Schartau et al., 2017).

Uncertainty in model formulation and structure is another limitation with respect to the identifiability of parameters. Inappropriate physics, aggregation of diverse species in broad functional groups, and finite limitation of spatial and temporal resolution are some factors that contribute to structural errors or biases in marine biogeochemical models, which may affect parameter estimation (e.g. Hood et al. 2006; Schartau et al. 2017). On the one hand, overly simple models usually fail to capture key biological processes of a marine ecosystem, leading to some misrepresentation of the truth. In addition, parameter estimates of these models are not credible. On the other hand, models with high complexity tend to have a large number of unconstrained parameters, and hence may yield larger uncertainties in model results. Therefore, it is critical to find a right balance between model complexity and the degree to which parameters are constrainable (Matear, 1995; Ward et al., 2013; Friedrichs et al., 2007).

Usually, models describe dynamics of marine ecosystems by differential equations that require initial conditions (IC). Some model setups, like those used for simulations of mesocosms, are typically sensitive to initial conditions. Therefore large errors in ICs can cause problems in identification of parameters and model divergence from the true state (Vallino, 2000). Schartau et al. (2017) stressed that small absolute errors in IC are large relative errors that may affect model predictions, e.g. magnitude and timing of phytoplankton bloom. Furthermore, collinearity exists between values of IC and biological

parameters (Gibson and Spitz, 2011), and hence noise in IC of state variables could introduce uncertainties in parameter estimates.

Typically, inverse methods have been used for parameter estimation (Gregg et al., 2009). These methods minimise a cost function that quantifies the discrepancy between model output and data. Depending on the underlying statistical and probabilistic assumptions, cost functions can have different forms (Stow et al., 2009), e.g., root mean squared errors (RMSEs), or based on a maximum likelihood (ML) estimator, or a the maximum a posteriori (MAP) estimator. While computing RMSEs no assumptions about probability distributions of the variables of interest are made. Although RMSE is a simple quantitative metric, it is informative, and therefore several studies (e.g., Oschlies and Schartau 2005; Allen et al. 2007; Nerger and Gregg 2007; Doney et al. 2009) have evaluated the model performance by computing RMSEs. ML and MAP estimators are two well known probabilistic based methods to estimate model parameters (Weir et al., 2013). In ML estimation we seek to obtain a set of parameter values that maximises the *likelihood* probability (Lehmann and Casella, 2006), the probability of the data being described by a model and its parameters. In the Bayesian approach, we treat parameters as random variables and include some prior information about them as an additional probability. The conditional probability of the parameter values given the data (posterior) is then given by the product of a likelihood and a prior probability, and the mode of the posterior leads to a MAP estimate. Ward et al. (2010) have shown that a Bayesian approach to data assimilation (which includes prior information about parameters) generates more reliable model results than the ML estimation. From the optimisation perspective, the only difference between ML and MAP estimates is the cost function that is minimised. Lawson et al. (1996) applied a ML estimator in an adjoint method to estimate parameters of two ecological models calibrated with twin data. Likewise, Spitz et al. (2001) configured their ecosystem model with data from the Bermuda Atlantic Time Series (BATS) using a ML method. Matear (1995) computed the MAP estimates of the parameters for three different models of increasing complexity with data from the station Papa in the North Pacific. In this study he also determined uncertainty ranges of his parameter estimates, and collinearities were identified.

2.1.2 Parameters of phytoplankton losses

Several studies have focussed on the estimation and sensitivity analysis of loss parameters of phytoplankton biomass, while assimilating in situ observations (e.g., Hurtt and Armstrong 1996; Gunson et al. 1999; Moore et al. 2001; Losa et al. 2004; Schartau et al. 2017). Losa et al. (2004) calibrated their 0-D model with the data from a Bermuda station in the North Atlantic, and found that parameterisations for phytoplankton and zooplankton losses in the model make the major contribution to the model uncertainties. Ward et al. (2013) have demonstrated that processes corresponding to unconstrained parameters in the model can be removed with no significant reduction in the model's ability to fit the data. The parameters related to zooplankton (maximum grazing rate, the half saturation constant for ingestion and the zooplankton mortality) were poorly identified in the sensitivity experiment of Fennel et al. (2001) when they configured their model with only nitrate and phytoplankton data. Schartau et al. (2017) calibrated a model with data from a mesocosm experiment

and obtained large posterior errors in estimates of the parameter that determines the aggregation rate of phytoplankton. [Gibson and Spitz \(2011\)](#) reported collinearities between the phytoplankton loss parameters and IC in a regional model. They stressed that ignorance of parameter correlations in sensitivity experiments could generate spurious parameter sets, and hence vague model results.

In this study, we investigate how the identification of phytoplankton loss parameters of an optimality-based (OB) is affected when it is calibrated with data using different metrics. For our data assimilation (DA) approach we set up four sets of simulations that employ different cost functions to configure the OB model with data from a mesocosm experiment. We seek to disentangle phytoplankton losses due to the export of aggregated cells and the loss because of grazing in the experiment. We assess uncertainties and collinearities in the estimates of plankton sink parameters in the absence of explicit zooplankton observations like microzooplankton and mesozooplankton abundance or grazing rates. Based on a rigorous assessment of credible values of parameters and their uncertainty ranges obtained in each case we can identify the metric that yields small uncertainties in parameter estimates while showing credible model performance. Eventually, we make inference on the most dominant sink pathway of phytoplankton losses in the mesocosms.

2.2 Theory and methods

We apply an OB model based on phytoplankton growth parameterisations of [Pahlow et al. \(2013\)](#) that resolves variations in the uptake of carbon (C) and nitrogen (N) (with variable N:C). Another major benefit is the formulation of photoacclimation by the phytoplankton in the model. The idea of the study presented here is to evaluate differences in the estimation of the critical phytoplankton loss parameters between different cost function formulations. Two common cost functions are used that do not rely on explicit assumptions about observational errors. The other two cost functions do include error assumptions with respect to the data: i) one that imposes all data to be independent, and ii) another that considers correlations between the different types of observations. We apply the OB model to provide insight on dynamics of phytoplankton losses under temporally varying conditions (nutrients and light). We calibrate our model with data of six selected mesocosms. We designed four model experiments where we minimise four different cost functions to estimate the maximum grazing rate (g_{max}), aggregation rate of phytoplankton (ϕ) and exudation rate (γ) parameters of the model. Of particular interest are differences in the identified ranges of uncertainty of these estimates as well as the correlations between the parameter estimates (collinearities). This analysis is based on some preceding parameter optimisations (Chapter 3 of this thesis). It means that all other parameters have been optimised before, based on a ML estimator.

2.2.1 Data

We consider data from the first experiment of Pelagic Ecosystem CO₂ Enrichment studies (PeECE-I) ([Engel et al., 2005](#); [Delille et al., 2005](#)), a study

conducted in the large-scale mesocosm facilities at the Marine field station of University of Bergen (Raunefjorden) in year 2001. The objective of this study was to investigate effects of ocean acidification (OA) on plankton community dynamics. Nine mesocosms were filled with post-bloom, nutrient-depleted fjord water and grouped into three replicates (“Glacial”, “Present” and “Future”) that were subject to different dissolved inorganic carbon (DIC) concentrations mimicking past, present and 21st century levels of CO_2 in the atmosphere (Engel et al., 2005; Delille et al., 2005). All nine mesocosms were enriched with similar dissolved inorganic nitrogen (DIN) concentration, approximately 15 mmol m^{-3} . Daily physical measurements of temperature and salinity are available for the entire period (23 days) of the PeECE-I experiment. We interpolate these daily measurements to obtain hourly values and use them as environmental forcing for our model simulations. Hourly photosynthetically active radiation (PAR) data were derived from meteorological global irradiance measurements of by the Geophysical Institute of the University of Bergen, Norway (Skartveit et al., 2001).

Since we do not resolve CO_2 effects and calcification, we exclude total alkalinity (TA) and particulate inorganic carbon (PIC) data from our DA method. In principle, we could have used data from one treatment or replicate (three mesocosms) for our analysis. But we realised that data sets for a single CO_2 treatment (exposed to similar CO_2 levels) are insufficient to provide sufficient information about the observational error (i.e. standard deviation). Therefore, we decided to assimilate mean data of six mesocosms from “Glacial” and “Future” treatments in the model. Observational standard errors could then be computed for these mean data at dates of sampling.

2.2.2 The model

An OB model is applied for simulation. As noted before we do not resolve calcification by coccolithophores, therefore only a single model compartment of all phytoplankton is used. The model estimates carbon (C) and nitrogen (N) fluxes between phytoplankton (PhyC, PhyN, Chl), zooplankton (ZooC, ZooN), detritus (DetC and DetN), labile dissolved organic compounds (lDOC, lDON) and dissolved combined carbohydrates (dCCHO) compartments. dCCHO acts as a precursor for transparent exopolymer particles (TEP). Figure 2.1 depicts a schematic representation of the general structure of the model and various physio-ecological processes connecting different compartments.

We do not include any vertical export of particulate organic matter in our model (0-D model approach). Furthermore, we assume light gradients in mesocosms. The attenuation of the irradiance with depth is represented by an exponential function. The mean irradiance is thus derived by integrating this function over depth (divided by depth of the mesocosm, $d = 4.5 \text{ m}$).

Since the model employs optimality-based parameterisations of Pahlow et al. (2013) to describe phytoplankton growth, an optimal allocation is assumed for the resources between light harvesting and nutrient acquisition functions of a cell. The resource allocation depends on the cell quota (N :C ratio). In the following, source-minus-sink (sms) equations of mass flux of C and N for the temporal evolution of the respective prognostic variable are described.

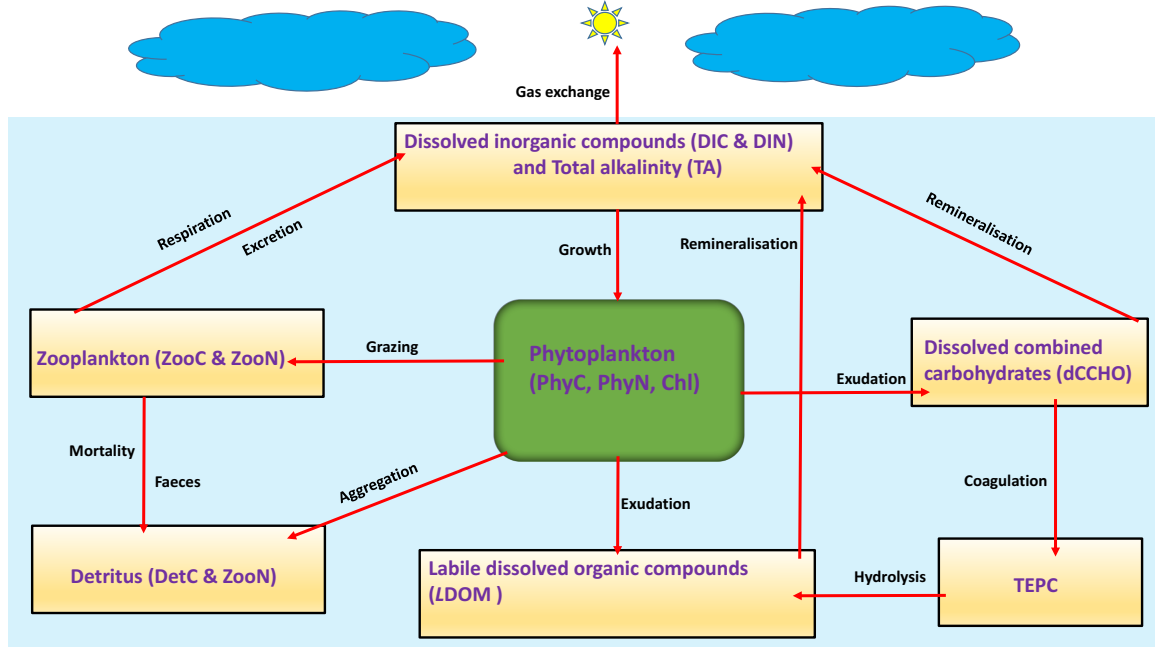


FIGURE 2.1: Schematic representation of different model compartments and processes connecting them.

2.2.2.1 Phytoplankton

The net growth rate of photoautotrophs depends on assimilated C and N, and losses due to aggregation, grazing by zooplankton and exudation of dissolved organic matter (DOM).

$$\frac{d}{dt}\text{PhyC} = (\mu - \gamma) \cdot \text{PhyC} - \frac{1}{Q^N} (A + G) \quad (2.1)$$

$$\frac{d}{dt}\text{PhyN} = V_C^N \cdot \text{PhyC} - \gamma \cdot \text{PhyN} - A - G \quad (2.2)$$

Where μ and V_C^N are the net growth and nitrogen uptake rates in units d^{-1} and $\text{mol N (mol C)}^{-1} \text{d}^{-1}$. γ is the exudation rate given in d^{-1} and Q^N is the cell quota in $\text{mol N (mol C)}^{-1}$. G and A are the nitrogen loss rates due to grazing by zooplankton and aggregation in $\text{mmol N m}^{-3} \text{d}^{-1}$. Parameterisations of aggregation and grazing are described in [Krishna and Schartau \(2017\)](#).

Chlorophyll synthesis: Chlorophyll synthesis of photoautotrophs depends on the optimised size of the chloroplast, the rate of change in Q^N , light conditions, A and G ([Pahlow et al., 2013](#)).

$$\frac{d}{dt}\text{Chl} = \left(\mu + \frac{\dot{\theta}}{\theta} \right) \cdot \text{Chl} - \theta^N (A + G) \quad (2.3)$$

Where θ is the chlorophyll *a*-to-carbon ratio in units $\text{g Chl (mol C)}^{-1}$.

2.2.2.2 Zooplankton

Any gain and loss in zooplankton biomass is attributed to Holling type III grazing on phytoplankton and quadratic mortality.

$$\frac{d}{dt}\text{ZooC} = \frac{G}{Q^N} - r_{\text{zoo}} - \frac{M_{\text{zoo}}}{Q_{\text{zoo}}} \quad (2.4)$$

$$\frac{d}{dt}\text{ZooN} = G - \gamma_{\text{zoo}}^N - M_{\text{zoo}} \quad (2.5)$$

Where r_{zoo} is the zooplankton respiration in unit $\text{mmol C m}^3 \text{ d}^{-1}$ and Q_{zoo} is the N:C ratio of zooplankton in $\text{mol N (mol C)}^{-1}$. γ_{zoo}^N and M_{zoo} is the nitrogen-specific zooplankton excretion and mortality in $\text{mmol N m}^3 \text{ d}^{-1}$.

2.2.2.3 Detritus

Phytoplankton aggregation and zooplankton mortality are sources for detrital C and N biomass, whereas hydrolysis of detritus is the primary sink.

$$\frac{d}{dt}\text{DetC} = \frac{A}{Q^N} + \frac{M_{\text{zoo}}}{Q_{\text{zoo}}} - \omega_{\text{det}} \cdot T_f \cdot \text{DetC} \quad (2.6)$$

$$\frac{d}{dt}\text{DetN} = A + M_{\text{zoo}} - \omega_{\text{det}} \cdot T_f \cdot \text{DetN} \quad (2.7)$$

With ω_{det} as the hydrolysis/degradation rate of detritus in d^{-1} and the term T_f is a function for the temperature dependence of metabolic rates.

2.2.2.4 Dissolved inorganic compounds

Dissolved inorganic carbon (DIC): The source for DIC comprises of respiration by Zooplankton, remineralisation of labile dissolved organic carbon (*l* DOC) compounds and sugars, and gas exchange with the atmosphere. The carbon assimilated for growth by photoautotrophs is the only sink for DIC.

$$\frac{d}{dt}\text{DIC} = -\mu \cdot \text{PhyC} + r_{\text{zoo}} + \rho \cdot T_f \cdot (\text{LDOC} + \text{dCCHO}) + F_{\text{DIC}} \quad (2.8)$$

Where ρ is the remineralisation rate of dissolved organic matter in d^{-1} and F_{DIC} is the flux due to air-sea gas exchange and has unit $\text{mmol C m}^{-3}\text{d}^{-1}$.

Dissolved inorganic nitrogen (DIN): Zooplankton excretion and remineralisation of labile dissolved organic nitrogen compounds are the sources for DIN pool. Nitrogen utilisation by photoautotrophs is the main sink of DIN.

$$\frac{d}{dt}\text{DIN} = -(V_C^N \cdot \text{PhyC}) + \gamma_{\text{zoo}}^N + \rho \cdot T_f \cdot \text{LDON} \quad (2.9)$$

2.2.2.5 Labile dissolved organic compounds

Sinks for *LDOM* are exudation of by phytoplankton, hydrolysis of detrital matter and hydrolysis of TEP, and the temperature-dependent remineralisation of *LDOC* constitute the sink for *LDOC*.

Labile dissolved organic carbon (*LDOC*):

$$\begin{aligned} \frac{d}{dt}LDOC = & \gamma \cdot (1 - f_{dCCHO}) \cdot \text{PhyC} + \omega_{det} \cdot T_f \cdot \text{DetC} \\ & + \omega_{gel} \cdot T_f \cdot \text{TEPC} - \rho \cdot T_f \cdot LDOC \end{aligned} \quad (2.10)$$

Where f_{dCCHO} is the fraction of exudates assigned to dCCHO and ω_{gel} is the hydrolysis/degradation rate of TEPC in d^{-1} .

Labile dissolved organic nitrogen (*LDON*):

$$\frac{d}{dt}LDON = \gamma \cdot \text{PhyN} + \omega_{det} \cdot T_f \cdot \text{DetN} - \rho \cdot T_f \cdot LDON \quad (2.11)$$

2.2.2.6 dCCHO

Formation of dCCHOs depend on carbon exudation by phytoplankton and the rate of formation of TEPCs.

$$\begin{aligned} \frac{d}{dt}dCCHO = & \gamma \cdot f_{dCCHO} \cdot \text{PhyC} - \phi_{dCCHO} \cdot dCCHO^2 \\ & - \phi_{TEPC} \cdot dCCHO \cdot \text{TEPC} - \rho \cdot T_f \cdot dCCHO \end{aligned} \quad (2.12)$$

2.2.2.7 Carbon content of transparent exopolymer particles (TEPC)

Sources for TEPC are coagulation of dCCHOs and aggregation of dCCHO with TEPC (Engel et al., 2004a) and the sink is the hydrolysis or degradation of TEPC.

$$\frac{d}{dt}TEPC = \phi_{dCCHO} \cdot dCCHO^2 + \phi_{TEPC} \cdot dCCHO \cdot \text{TEPC} - \omega_{gel} \cdot T_f \cdot \text{TEPC} \quad (2.13)$$

Detailed description of auxiliary variables and parameters introduced in above equations is listed in Table 2.1.

2.2.2.8 Initial conditions

The total initial biomass in our model is given by the parameter PON_0 . We assume 10% of PON_0 is assigned for initial N concentration of detritus. The remaining fraction (90%) of PON_0 is assigned to living biomass and distributed between zooplankton and phytoplankton. The parameter f_{zoo} determines the fraction of PON_0 that is assigned to initial zooplankton concentration. The remaining is allocated to initial phytoplankton.

$$PON_0 = \text{DetN}_0 + \text{ZooN}_0 + \text{PhyN}_0 \quad (2.14)$$

with the individual fractions:

$$\text{DetN}_0 = 0.1 \cdot PON_0 \quad (2.15)$$

$$\text{ZooN}_0 = f_{zoo} \cdot (PON_0 - \text{DetN}_0) \quad (2.16)$$

$$\text{PhyN}_0 = (1 - f_{zoo}) \cdot (PON_0 - \text{DetN}_0) \quad (2.17)$$

We compute initial C biomass of detritus, zooplankton and photoautotrophs from their respective N biomasses by applying a constant C:N ratio of 6.625. Initial conditions of DIC, TA, TEPC and dCCHO are taken from the data of the mesocosms, whereas for DOC and DON we assume small initial values close to zero (with C:N = 10) as explicit measurements of refractory and labile

Model parameters	Description	Value	Unit
1) PON_0	Initial concentration of particulate organic nitrogen	1.94	mmol N m ⁻³
2) f_{det}	fraction of PON_0 assigned to non-living detritus	0.51	-
3) f_{zoo}	fraction of living PON_0 assigned to zooplankton	0.77	-
4) V_0^N	photoautotrophic potential N fixation rate	4.1	mol N (mol C) ⁻¹ d ⁻¹
5) R_M^{Chl}	cost of chlorophyll maintenance	0.1	d ⁻¹
6) R_M	total respiration maintenance cost	0.05	d ⁻¹
7) ζ^{Chl}	cost of photosynthesis coefficient	0.4	mol C (g Chl a) ⁻¹
8) A_0	potential nutrient affinity	1	m ³ mol C ⁻¹ d ⁻¹
9) Q_{min}	subistence quota (minimum cellular N:C ratio)	0.027	mol N (mol C) ⁻¹
10) α	Photosynthetic efficiency of phytoplankton	3.3	mol C (g Chl a) ⁻¹ m ² W ⁻¹ d ⁻¹
11) ζ^N	cost of N uptake	0.7	mol C (mol N) ⁻¹
12) ω_{det}	hydrolysis rate of detritus	0.1	d ⁻¹
13) ω_{gel}	hydrolysis rate of TEPC	0.05	d ⁻¹
14) ρ	remineralsation rate of dissolved organic matter	0.1	d ⁻¹
15) ϕ_{dCCHO}	coagulation parameter of dCCHO	$7.48 \cdot 10^{-4}$	m ³ (mmol C) ⁻¹ d ⁻¹
16) ϕ_{TEPC}	coagulation parameter of dCCHO-TEPC	$2.56 \cdot 10^{-2}$	m ³ (mmol C) ⁻¹ d ⁻¹
17) R_{ref}	degradation rate constant at reference temperature	0.07	d ⁻¹
18) T_{ref}	reference temperature for A_E relation	293.15	K
19) A_E	slope of arrhenius relationship	4500	K
20) a_w	light attenuation due to water column	0.04	m ⁻¹
21) a_c	light attenuation due to chlorophyll a	0.05	(mg Chl a) ⁻¹ m ³
Auxiliary variables			
1) Q^N	N:C ratio of a cell		mol N (mol C) ⁻¹
2) V_C^N	C-specific N uptake rate of photoautotrophs		mol N (mol C) ⁻¹ d ⁻¹
5) V_C^C	C fixation rate of a cell		mol C (mol C) ⁻¹ d ⁻¹
6) μ	net growth rate of photoautotrophs		d ⁻¹
7) A	N-specific phytoplankton aggregation		mmol N (m) ⁻³ d ⁻¹
8) G	N-specific grazing		mmol N (mol C) ⁻¹
9) Q^{zoo}	N:C ratio of zooplankton		mmol C m ⁻³ d ⁻¹
10) r_{zoo}	zooplankton respiration		mmol N (m) ⁻³ d ⁻¹
11) M_{zoo}^N	N-specific zooplankton mortality		mmol N m ⁻³ d ⁻¹
12) γ_{zoo}^N	N-specific zooplankton excretion		mmol N m ⁻³ d ⁻¹
13) θ	Chl:C ratio of a cell		g Chl a (mol C) ⁻¹
14) θ	time derivative of θ		g Chl (mol C) ⁻¹ d ⁻¹
15) θ^N	Chl:N ratio of a cell		g Chl a (mol N) ⁻¹
16) S_{chl}	Chlorophyll synthesis rates of photoautotrophs		d ⁻¹
17) I	Irradiance		W m ⁻²
18) F_{DIC}	flux due to air-sea gas exchange		mmol C m ⁻³ d ⁻¹

TABLE 2.1: Fixed model parameters with their values as obtained from preceding optimisations and description of auxiliary variables.

DOC are not available ($\text{DON} = 0.005 \text{ mmolN m}^{-3}$ and $\text{DOC} = 0.05 \text{ mmolC m}^{-3}$).

2.2.3 Data assimilation approach

2.2.3.1 Control and fixed parameters

We want to determine estimates and uncertainty ranges of loss rates of phytoplankton biomass due to grazing, aggregation and exudation or leakage of organic matter during the PeECE-I experiment. Therefore, we have selected three control parameters (g_{max} , ϕ and γ) that affect the loss of phytoplankton biomass. Other parameter values were taken from a preceding optimisation (Chapter 4). These preceding estimates include IC parameters. The remaining biological and physical parameters were assigned values adapted from a series of previous studies (e.g., [Schartau et al. 2007](#); [Pahlow et al. 2013](#)) that employed parameterisations and definitions of state variables consistent to ours. The decision to optimise only three desired parameters and fix others is owing to the fact that we here focus on differences in uncertainty ranges and how these depend on the metrics employed. Table 2.1 lists the description of fixed model parameters along with their assigned values.

2.2.3.2 Cost functions

In our study we consider five different types of measurements ($N_y = 5$) from the PeECE-I experiment: (1) dissolved inorganic carbon (DIC, mmol m^{-3}), (2) dissolved inorganic nitrogen (DIN) (nitrate + nitrite, mmol m^{-3}), (3) chlorophyll *a* (Chl *a*, mg m^{-3}), (4) particulate organic nitrogen (PON, mmol m^{-3}), and (5) particulate organic carbon (POC, mmol m^{-3}). Concentrations of nitrate and nitrite are not explicitly resolved and we refer their sum as DIN. Observations are available on daily basis over a period of 23 days ($N_t = 23$). The observation vectors (\vec{y}_i) contains daily means of observations from nine mesocosms for five variables on a given day. Likewise, the corresponding vector for model counterparts ($H_i(\vec{x})$) consists of daily means of model states for the same variables. Please note, modelled PON (PON^{mod}) is the sum of simulated N concentrations of photoautotrophs (PhyN), zooplankton (ZooN) and detritus (DetN). Simulated POC (POC^{mod}) includes contributions from C biomasses of phytoplankton (PhyC), zooplankton (ZooC), detritus (DetC), and TEPC and dCCHO. The residual vector (\vec{d}_i) quantifies differences between observation vector and the model output vectors at a given day.

$$\vec{d}_i = \vec{y}_i - H_i(\vec{x}) = \underbrace{\begin{pmatrix} \text{DIC}_i \\ \text{DIN}_i \\ \text{Chl } a_i \\ \text{PON}_i \\ \text{POC}_i \end{pmatrix}}_{\text{data}} - \underbrace{\begin{pmatrix} \text{DIC}_i^{mod} \\ \text{DIN}_i^{mod} \\ \text{Chl}_i^{mod} \\ \text{PON}_i^{mod} \\ \text{POC}_i^{mod} \end{pmatrix}}_{\text{model results}} \quad (2.18)$$

As noted before, model-data misfits can be quantified in different ways. We set up four set of simulations or cases where we minimise cost functions that have different forms. We can broadly categorise our cost functions in two

categories: 1.) non-probabilistic cost functions, and 2) probabilistic cost functions. For the first and second model experiments we employ non-probabilistic approaches, whereas in the third and fourth cases probabilistic cost functions have been considered.

For the first sets of simulations our cost function (J_1) is defined by the summation of RMSEs.

$$J_1 = \sum_{i=1}^{N_t} \sum_{j=1}^{N_y} \sqrt{\frac{(\bar{y}_{i,j} - H_{i,j}(\vec{x}))^2}{n}} \quad (2.19)$$

Where n is the number of available observations on a given day and $\bar{y}_{i,j}$ is the spatial mean of observations between mesocosms for the respective measurement types on a day given by i . Every observational type may have its own unit and therefore, in principle, the N_y terms must not be added. However, addition of RMSEs of different units in equation 2.19 actually implies that a standard error of 1 (in the respective unit) is assumed. Dividing RMSEs by standard errors make them dimensionless and hence their summation becomes possible to compute the total model-data misfit.

We apply a relative RMSE (rRMSE) as the quantitative metric to evaluate our model in the second case. It differs from the RMSE in the normalisation technique. rRMSEs are normalised by the temporal mean of observations. Thus, our cost function for the second set of simulations is the summation of rRMSEs of different variables:

$$J_2 = \sum_{i=1}^{N_t} \sum_{j=1}^{N_y} \left(\sqrt{\frac{1}{n} \cdot (\bar{y}_{i,j} - H_{i,j}(\vec{x}))^2} \right) \cdot \frac{1}{\bar{y}_j} \quad (2.20)$$

Where \bar{y}_j is the temporal mean over the entire period of the experiment for the respective observational types.

For the third and fourth cases, we impose probabilistic assumptions on variables in our cost functions. In addition, we assume an additive Gaussian error model for our simulations:

$$\bar{y}_i = H_i(x_i) + \epsilon_i \quad (2.21)$$

Where H_i represents a measurement operator that accounts for interpolation and unit conversions, and $\epsilon \sim N(0, \sigma^2)$ is Gaussian distributed with mean zero and σ as the standard error. Notably, the standard errors are different from the standard deviations. The standard error accounts for the uncertainty of the sample mean, whereas the standard deviation is the measure of variability in data of a certain variable. Thus, σ depends on both the standard deviation (SD) and the sample size or the number of samples (n) at respective dates of observation, by the simple relation $\sigma = SD/\sqrt{n}$. As the sample size increases, the uncertainty of the sample mean decreases, but SD will remain the same. Our cost functions for the third and fourth model experiments are derived from the likelihood probability. As we disregard any prior information about

parameter values, the likelihood $p(\vec{y} | \Theta)$ is maximised by actually minimising its negative logarithm $J(\Theta)$ (Sivia and Skilling, 2006):

$$\hat{\Theta} = \operatorname{argmax} p(\vec{y} | \Theta) \rightarrow \operatorname{argmin} [-2 \ln(p(\vec{y} | \Theta))] \simeq \operatorname{argmin} J(\Theta) \quad (2.22)$$

The estimate of the parameter vector ($\hat{\Theta}$) corresponds with the maximum of the likelihood, or likewise with the minimum of its negative logarithm.

Next, we consider a cost function (J_3) in which all data are regarded as being independent. In this case, J_3 reduces to a Chi-squared cost function (Ward et al., 2010):

$$J_3 = \sum_{i=1}^{N_t} \sum_{j=1}^{N_y} \frac{(\vec{y}_{i,j} - H_{i,j}(\vec{x}))^2}{\sigma_{i,j}^2} \quad (2.23)$$

For the fourth case, we minimise a cost function (J_4) that accounts for correlations in the data, using a data covariance matrix (\mathbf{R}). Assuming Gaussian probabilities, for the multivariate case, J_4 can be written as:

$$J_4 = \sum_{i=1}^{N_t} (\vec{y}_i - H_i(\vec{x}))^T \mathbf{R}_i^{-1} (\vec{y}_i - H_i(\vec{x})) \quad (2.24)$$

On a given sampling day \mathbf{R}_i can be derived as:

$$\mathbf{R}_i = \mathbf{S}_i \cdot \mathbf{C}_{(\vec{y})} \cdot \mathbf{S}_i \quad (2.25)$$

With \mathbf{S}_i being diagonal matrices (containing σ_i as diagonal elements) and $\mathbf{C}_{(\vec{y})}$ being correlation matrices. Here, correlations between observations of biological variables change temporally. For example, observed DIN and Chl are strongly negatively correlated during the pre-bloom period but for the period after the algal bloom they are weakly correlated. Likewise, PON and POC observations are positively correlated before the bloom and negatively correlated after the bloom. As the correlations between measurements can change from the exponential growth period to the post-bloom period, we employ a novel approach in which we consider period-specific correlation matrices (\mathbf{C}_{pre} , \mathbf{C}_{bloom} and \mathbf{C}_{post}) in Equation (2.25) for the computation of the covariance matrices.

Matrix \mathbf{C}_{pre} contains correlations between observations of different measurement types during the pre-bloom period ($i = 1, 2, \dots, 10$):

$$\mathbf{C}_{(y)} = \mathbf{C}_{pre} = \begin{pmatrix} & \text{DIC} & \text{DIN} & \text{Chl } a & \text{PON} & \text{POC} \\ \text{DIC} & 1 & 0.66 & -0.53 & -0.72 & -0.78 \\ \text{DIN} & . & 1 & -0.53 & -0.75 & -0.48 \\ \text{Chl } a & . & . & 1 & 0.81 & 0.54 \\ \text{PON} & . & . & . & 1 & 0.63 \\ \text{POC} & . & . & . & . & 1 \end{pmatrix} \quad (2.26)$$

For the bloom period ($i = 10, 11, \dots, 14$) we assume measurements to be uncorrelated as we do not have any information about the exact time when correlations change. Thus, \mathbf{C}_{bloom} have off-diagonal elements of zero and diagonal

elements of one:

$$C_{(y)} = C_{bloom} = \begin{pmatrix} & \text{DIC} & \text{DIN} & \text{Chl } a & \text{PON} & \text{POC} \\ \text{DIC} & 1 & 0 & 0 & 0 & 0 \\ \text{DIN} & . & 1 & 0 & 0 & 0 \\ \text{Chl } a & . & . & 1 & 0 & 0 \\ \text{PON} & . & . & . & 1 & 0 \\ \text{POC} & . & . & . & . & 1 \end{pmatrix} \quad (2.27)$$

The third correlation matrix (C_{post}) comprises of correlations from the post-bloom period ($i = 15, 16, \dots, 23$):

$$C_{(y)} = C_{post} = \begin{pmatrix} & \text{DIC} & \text{DIN} & \text{Chl } a & \text{PON} & \text{POC} \\ \text{DIC} & 1 & 0.03 & -0.20 & \mathbf{0.14} & -0.50 \\ \text{DIN} & . & 1 & -0.29 & -0.27 & \mathbf{0.33} \\ \text{Chl } a & . & . & 1 & 0.62 & \mathbf{-0.32} \\ \text{PON} & . & . & . & 1 & \mathbf{-0.64} \\ \text{POC} & . & . & . & . & 1 \end{pmatrix} \quad (2.28)$$

2.2.3.3 Parameter optimisation and uncertainty estimation

For optimisations we applied “optim” function available in R package FME (Soetaert and Petzoldt, 2010). We follow a three steps optimisation procedure. In the first step, we applied *simulated annealing* algorithm (SANN), (Bélisle, 1992) to perform a stochastic search of the parameter space and approximate the global minima of our cost function $J(\Theta)$ without getting trapped in local minima. Next step is to refine the result of the SANN algorithm by using the *Broyden-Fletcher-Goldfarb-Shanno* (BFGS) algorithm (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shannon, 1970). The BFGS algorithm also approximates the Hessian matrix of the parameter estimates (Thacker, 1989; Tziperman et al., 1992). In the final step, we applied a Markov Chain Monte Carlo (MCMC) method (Harmon and Challenor, 1997) to estimate parameter uncertainties in the vicinity of the optimum also using the Hessian information provided by the BFGS algorithm. The MCMC method derives posterior uncertainty limits (credible regions) of the parameter estimates and resolves collinearities (correlations) between them. Furthermore, the MCMC method also provides the final “best” parameter vector ($\hat{\Theta}$) that generates the lowest cost function value (J_{min}).

2.3 Results

Model parameters that account for loss rates of phytoplankton biomass (g_{max} , ϕ and γ) were optimised in four model experiments. For each case, we obtain an ensemble of statistically indistinguishable model solutions and posterior estimates of loss parameters from the MCMC method. The optimisation results differ for all four cases. In general, the model appears to perform better when calibrated with cost functions that were derived from probabilistic considerations (e.g. based on a likelihood), especially J_4 . Our optimisation results suggest that exudation loss parameter (γ) is best constrained with the assimilated data. Furthermore, optimised estimates of γ are higher than the ones

of aggregation and grazing. Thus, according to our model, the exudation by phytoplankton appears to be an important pathway in the mesocosms.

2.3.1 Parameter estimates and collinearities

Our results clearly show differences in estimates of parameters and their respective uncertainties between the four model experiments. In general, we

Cases	RE_γ	RE_ϕ	$RE_{g_{max}}$
First case	0.21	0.82	2.61
Second case	1.13	2.14	2.71
Third case	0.06	0.22	1.12
Fourth case	0.05	0.36	0.25

TABLE 2.2: Relative errors ($RE = (\theta_{max} - \theta_{min})/\bar{\theta}$) for the three sink parameters (RE_γ , RE_ϕ , $RE_{g_{max}}$) in the respective cases according to MCMC results.

find the likelihood based cost functions to be a better metric for parameter identification (Table 2.2), as they yield smaller posterior errors in parameter estimates, quantified by relative errors (REs in Table 2.2). For the first and second cases, g_{max} and ϕ remain poorly constrained by the mesocosms data. This is evident from the large uncertainties in their posterior estimates (Figs. 2.2a and 2.2b). Furthermore, estimates of g_{max} , obtained from the MCMC method seem spurious as they tend to approach extremely low values when considering rRMSE and RMSE as cost functions. Interestingly, parameters are worst constrained for the second case (Fig. 2.2b, Table 2.2). This is surprising because we expected the rRMSE to behave as a better metric than the RMSE, because of the normalisation. Figures 2.3a and 2.3b show values of J_1 and J_2 (as contours) owing to variations in g_{max} and ϕ , while γ is fixed to its optimal estimate. The innermost ellipses of J_1 and J_2 around their respective optimums are not closed from the one end indicating insensitivity of cost functions to changes in values of g_{max} .

The mapping of the credible region of posterior uncertainties (green dots on Figs. 2.3a and 2.3b) on the 2-D parameter space illustrates that the algorithm fails to constrain g_{max} and ϕ . The performance of the MCMC method is slightly better in terms of convergence when J_1 is applied compared to J_2 . Overall, both J_1 and J_2 turn out to be ineffective metrics to constrain model parameters (especially g_{max} and ϕ). However, posterior errors in the estimates of γ are smaller for the first case than for the second. The aggregation parameter ϕ , which could not be estimated by applying rRMSE and RMSE cost functions, become better constrained by the assimilated data in third and fourth cases. There are nuanced differences in posterior estimates of ϕ between third and fourth model experiments. The model predicts low optimised estimate of g_{max} when data is assumed to be independent. Furthermore, uncertainties in g_{max} are large when J_3 is applied (Table 2.2, Figs. 2.5a and 2.5b). Overall, the credible regions of posterior uncertainties are smallest for the fourth case, if compared with the other cases. This suggests that the major loss parameters of phytoplankton biomass are best constrained if we employ J_4 as metric to calibrate the model.

The alignment of cost function contours exhibit negative collinearities between values of ϕ and g_{max} , which is more pronounced for the cases where

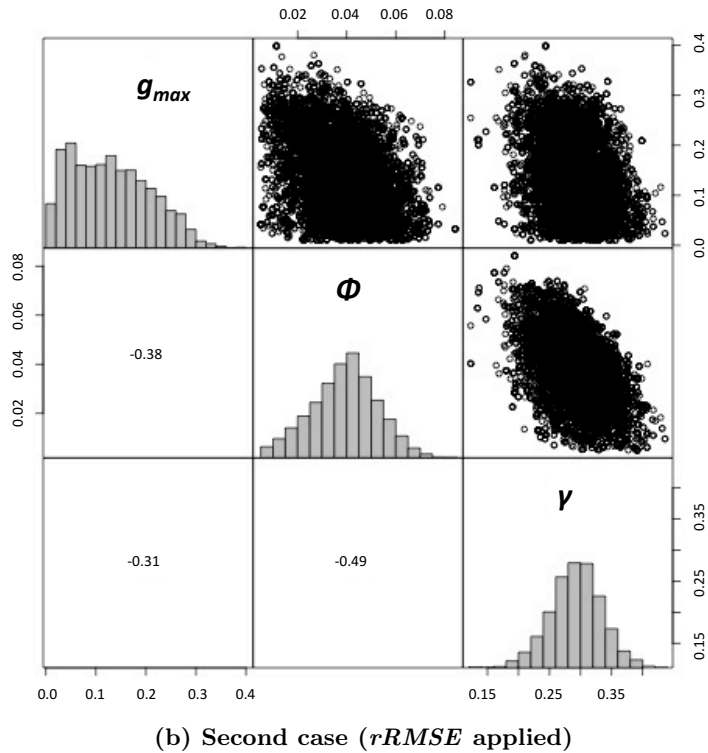
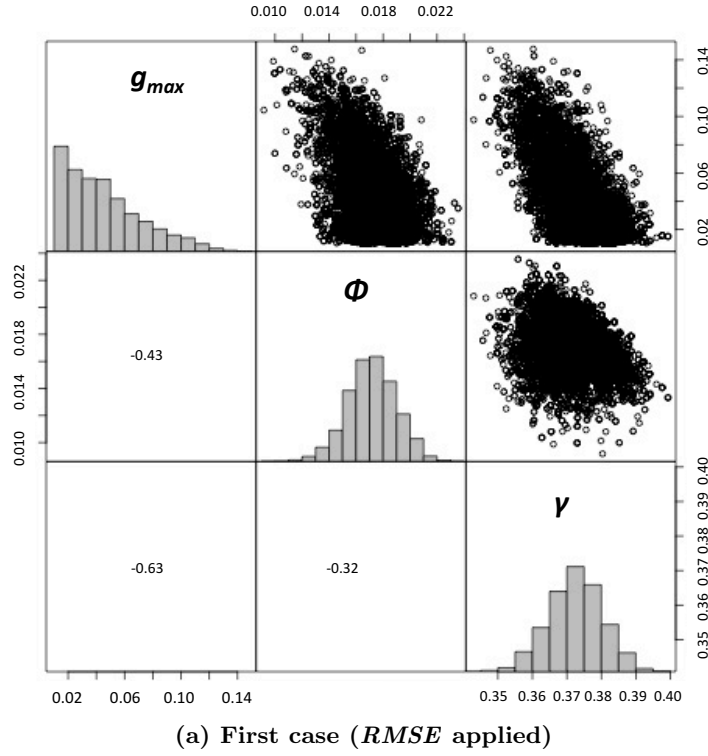
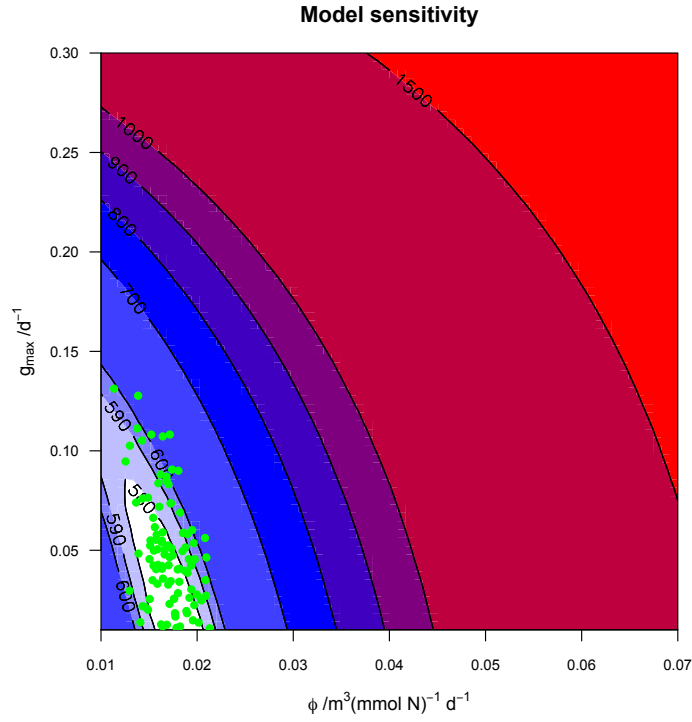
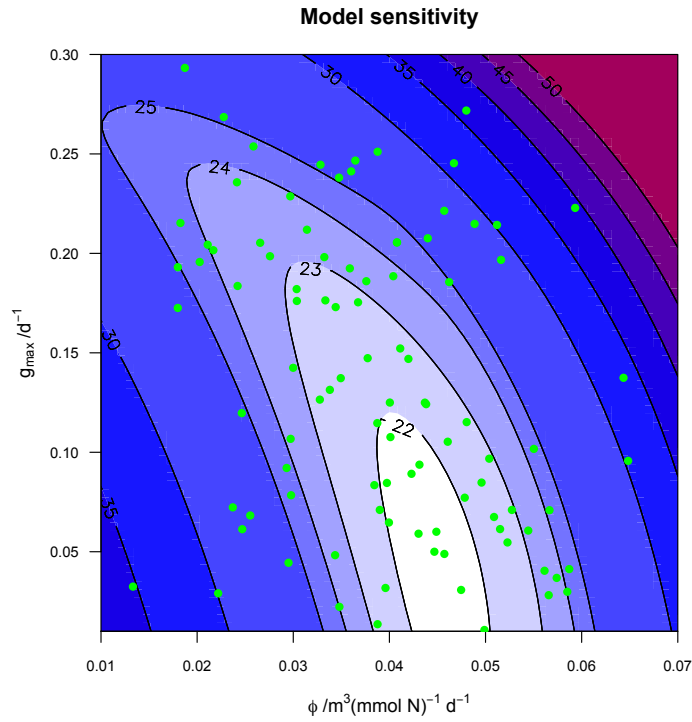


FIGURE 2.2: Posterior distribution of three control parameters (g_{max} , ϕ , γ in units d^{-1} , $m^3(\text{mmol N})^{-1} d^{-1}$, d^{-1}) according to Markov chain Monte Carlo (MCMC) results for the first and second cases (non-probabilistic approaches). Black clouds (markers) on the upper triangular matrix of each figure depict credible regions of posterior uncertainties for a combination of two parameters. Histograms on the diagonal represent distribution of posterior errors in parameter estimates. Numbers ($-1 >$ and < 1) on the lower triangular matrix are correlation coefficients among control parameters.

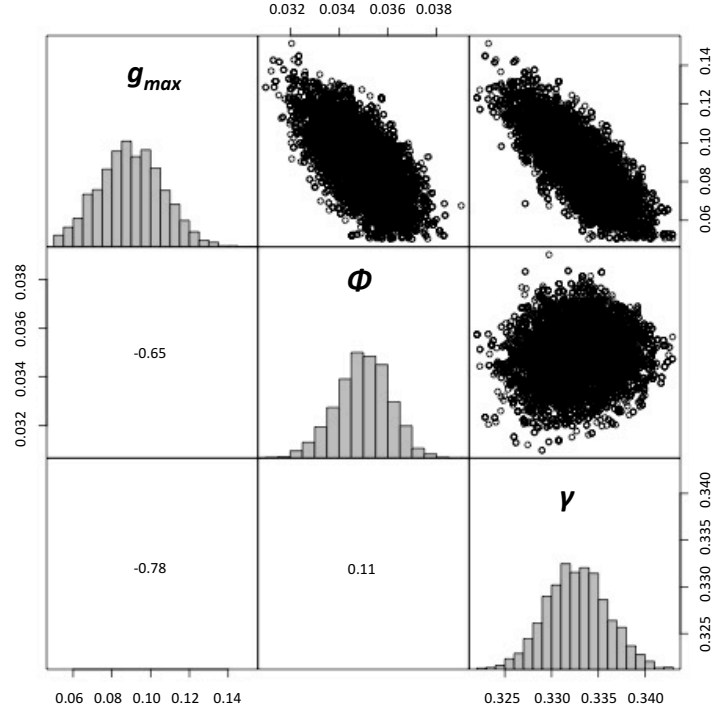


(a) First case ($RMSE$)

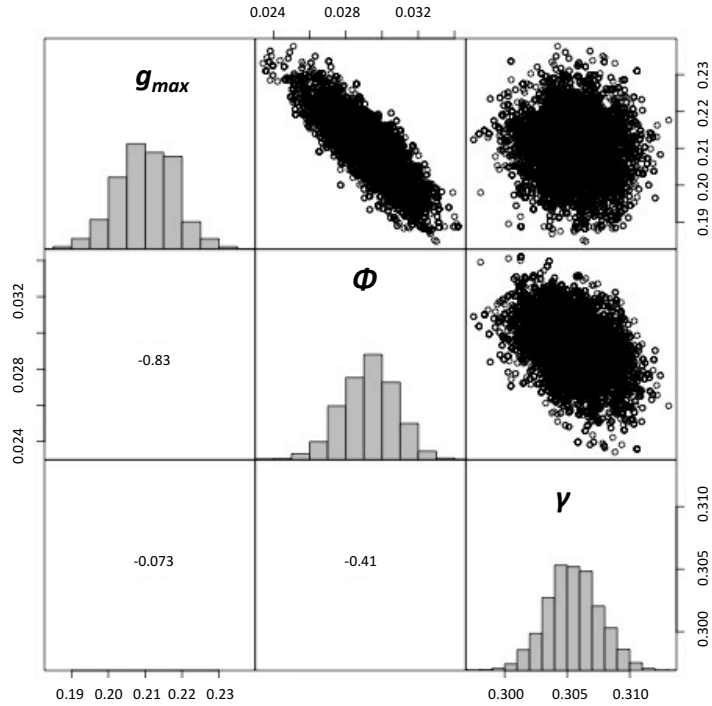


(b) Second case ($rRMSE$)

FIGURE 2.3: J_1 and J_2 contours when varying ϕ and g_{max} around the optimum estimate (J_{min}) at $\hat{\theta}$, while γ is fixed to its optimised point estimate. Green markers represent credible regions of parameter estimates obtained with the MCMC method.

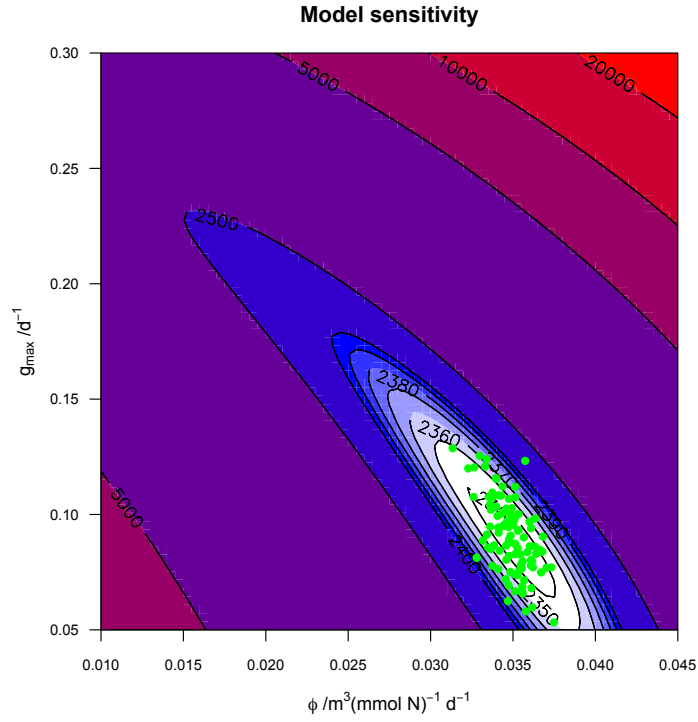


(a) Third case (without covariances)

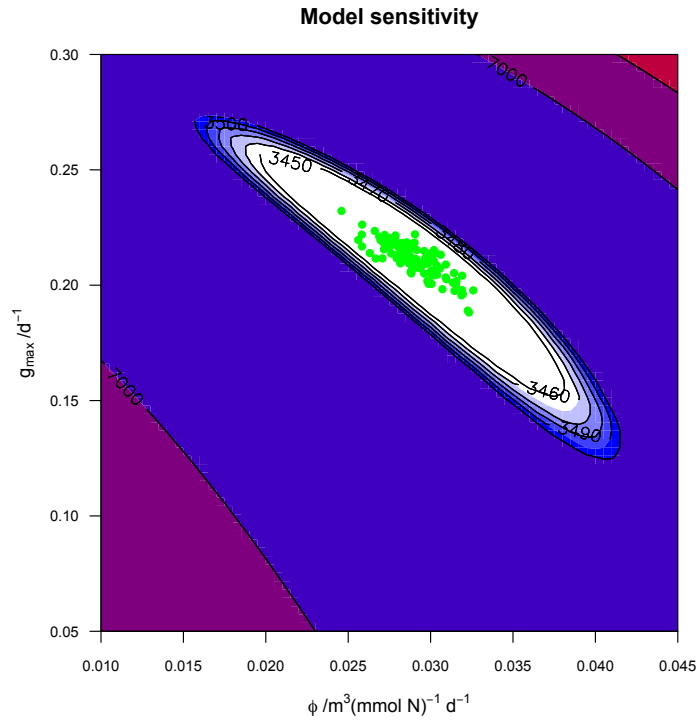


(b) Fourth case (with covariances)

FIGURE 2.4: Posterior distribution of three control parameters (g_{max} , ϕ , γ in units d^{-1} , $m^3(mmole\ N)^{-1}\ d^{-1}$, d^{-1}) according to Markov chain Monte Carlo (MCMC) results for the third and fourth cases (probabilistic approaches). Black clouds (markers) on the upper triangular matrix of each figure depict credible regions of posterior uncertainties for a combination of two parameters. Histograms on the diagonal represent distribution of posterior errors in parameter estimates. Numbers ($1 >$ and < 1) on the lower triangular matrix are correlation coefficients among control parameters.



(a) Third case (without covariances)



(b) Fourth case (with covariances)

FIGURE 2.5: J_3 and J_4 contours when varying ϕ and g_{max} around the optimum estimate (J_{min}) at $\hat{\theta}$, while γ is fixed to its optimised point estimate. Green markers represent credible regions of parameter estimates obtained with the MCMC method.

probabilistic based cost functions are applied (Fig. 2.5). Results from MCMC reveal similar collinearities between parameter combinations of ϕ and g_{max} (Fig. 2.4). These results exemplify the uncertainty in constraining major loss parameters for phytoplankton biomass in the presence of grazing, if no explicit information on grazing rates and zooplankton data are available. However, collinearities between parameter combinations that include γ are significantly reduced when the cost function with covariances is employed compared to other cases.

2.3.2 Model-data comparisons

Plots in Figure 2.6 depict time-series comparisons of data with model results (green trajectories) corresponding to the best parameter estimate ($\hat{\theta}$) in all four cases. For the fourth case, the model successfully reproduces all observations. The simulated DIC and PON is underestimated in the first, second and third cases. Model-data misfits in DIC and PON are largest for the case where RMSE is applied as the cost function. According to MCMC results, the optimised estimate for g_{max} is extremely low for the first case (Table 2.3), which means that zooplankton growth is apparently underestimated, and hence also heterotrophic respiration (one of the sources for DIC). Thus, the simulated PON and DIC concentrations are lower than the respective observations during the nutrient depletion phase (Figure 2.6).

Model parameters	First case	Second case	Third case	Fourth case
1) g_{max}	0.01	0.05	0.10	0.22
2) ϕ	0.02	0.04	0.03	0.03
2) γ	0.40	0.30	0.30	0.30

TABLE 2.3: Optimised point estimates of control parameters obtained from the MCMC method.

Joassin et al. (2011) used data of the same experiment, but only from the “present” treatment (Delille et al., 2005) to calibrate their model. They also underestimate observed PON in the post-bloom period. Furthermore, their model predicts low PON concentrations compared to the data also in the pre-bloom period, which is not the case for our model simulations.

For the first case, the “best” ensemble solution predicts much higher accumulation of POC than observed. Notably, the optimised estimate of γ is highest for the case where the RMSE is applied compared to other three cases (Table 2.3). This explains the overestimation of POC by the model, due to excessive TEPC production as a result of high exudation rates during the post-bloom period. For the other cases the model reproduces POC data well, especially when covariances are considered. Since TEPC contributes to POC, a good fit to observed accumulation of POC means that TEPC concentrations are also well resolved.

Our model predictions are consistent with chlorophyll observations when J_3 and J_4 are applied. For the third and fourth cases, the model matches observed chlorophyll concentrations in the post-bloom period, but this is because our model explicitly resolves the chlorophyll concentrations of detritus, and high estimates of ϕ (as obtained in the second, third and fourth cases) facilitate formation of detrital biomass. Simulated chlorophyll is underestimated for

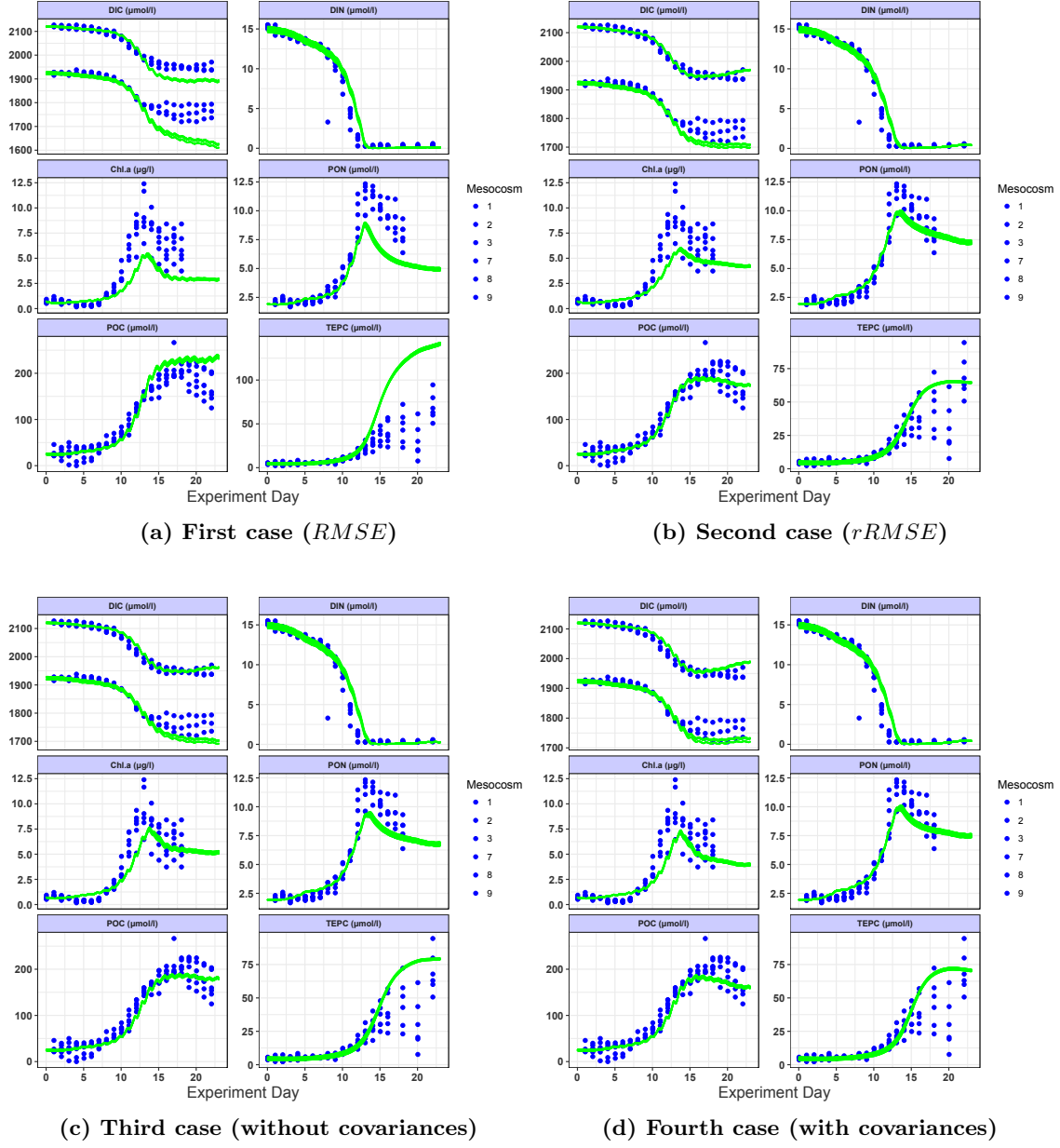


FIGURE 2.6: Comparison of observations (blue asterisks) with “best” optimised model results (green trajectories) corresponding to J_{min} at $\hat{\theta}$ obtained from the MCMC method for the four cases.

cases in which non-probabilistic cost functions (RMSE and $rRMSE$) are used for calibration. Nevertheless, our model predictions are in the same phase as observations. In contrast to Joassin et al. (2011), our model does not show early build up in phytoplankton chlorophyll in all cases.

A general model bias is seen in DIN uptake. In all four cases the model predicts some delayed draw down in the DIN. Since this bias is the characteristic of simulated DIN in optimised solutions of all four cases, it cannot be explained by differences in parameter estimates. This does not seem to be a severe model deficiency, it rather appears to be a mass balance problem in the data. The exponential draw down in DIN data and accumulation of observed

PON are not in phase. But, in the model mass balance is maintained. Thus, when the model is fitted against observed PON it introduces misfits in DIN, and vice versa.

2.3.3 Cross-validation with unassimilated data

Friedrichs et al. (2007) highlight the importance of validation of model solutions against independent, unassimilated data as it provides unbiased assessment of the model. To evaluate which metric maximises the robustness of the model we designed a cross-validation experiment. In this experiment we validated the “best” optimised solution obtained for the respective cases against data from those mesocosms that were part of the “Present” treatment. These data may be viewed as not being truly independent, because they are subject to the same experimental setup. But, these data were not used for optimisation. This way we compare our model solutions and assess model’s capability to reproduce “other” data. To quantify the robustness of the model solution we adapted a metric from Friedrichs et al. (2007), namely “Portability Index” (PI):

$$PI = J_{assim}/J_{unassim} \quad (2.29)$$

Where J_{assim} is the misfit cost with assimilated data (J_1 , J_2 , J_3 and J_4 in this study) and $J_{unassim}$ is the cross-validation cost (against the data that were not used for parameter optimisations). A model is increasingly portable or robust as its PI approaches value of 1.0.

Table 2.4 lists the PI values for the respective cases that are obtained when the optimised model solutions are cross-validated with independent data. We obtain higher PI values for the probabilistic cost functions (third and fourth cases) than the non-probabilistic ones (first and second cases). The highest PI value (0.99) is achieved for the fourth case, which is in further support of applying this kind of metric. This means the capability of the model to reproduce the unassimilated data is best when the covariances are applied. In contrast, the $rRMSE$ turns out to be the worst metric to configure the model and evaluate its robustness as for the second case we see a massive decrease in the PI value (by $\approx 95\%$) compared to the fourth case. A moderate decline in the PI index of the model is obtained when J_3 is considered than J_4 .

Cases	PI
First case	0.60
Second case	0.05
Third case	0.92
Fourth case	0.99

TABLE 2.4: PI denotes the portability index of the “best” optimised solution for the respective case.

2.3.4 Simulating available data of zooplankton biomass

Our results show that g_{max} remains unconstrained when the model is calibrated with RMSEs (first and second case) and independent data (third case). A possible reason for this could be unavailability of explicit zooplankton data. The

question then is to which extent some availability of zooplankton data would have helped in this respect. To seek for an answer, we designed a twin experiment where we calibrate our model with synthetic zooplankton data (generated by the model). We use the optimised solution corresponding to J_{min} at $\hat{\theta}$ obtained for the fourth case (without covariances) to generate temporal estimates for N and C concentrations of zooplankton. From these estimates, we sample the noon (mid day) values and treat them as model-generated zooplankton data. Subsequently, we evaluate our model with synthetic zooplankton data in addition to the data from the mesocosms. For this experiment, we assume data to be independent, hence similar to the third case. We employ the same data assimilation approach like in previous model experiments.

Results of the twin experiment show exactly what we expected, particularly in context of estimation of g_{max} . After assimilating synthetic zooplankton we obtain better constraints on g_{max} (Fig. 2.7).

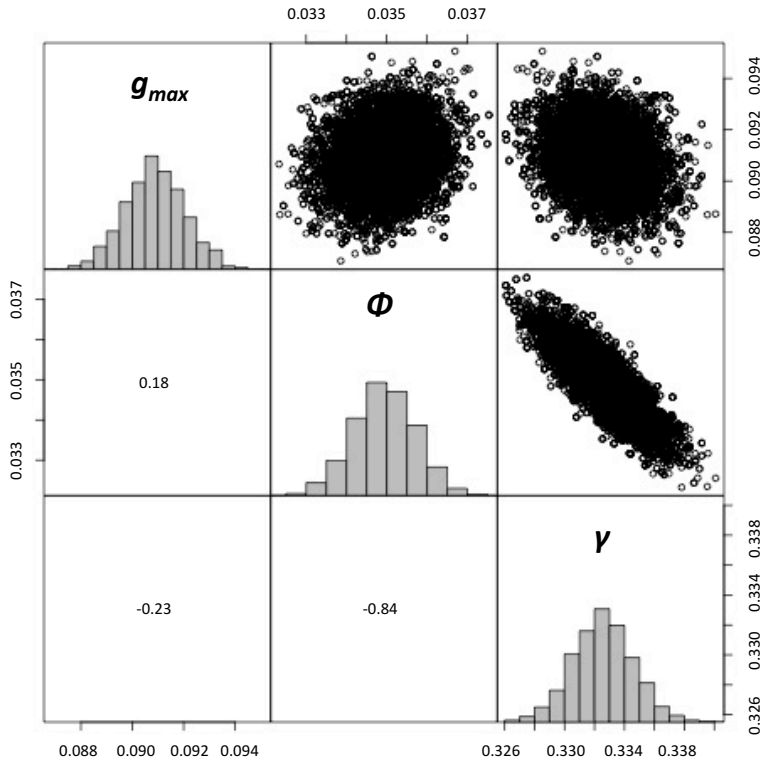


FIGURE 2.7: Posterior distribution of three control parameters (g_{max} , ϕ , γ) according to for the twin experiment with synthetic zooplankton data. Black clouds (markers) on the upper triangular matrix of each figure depict credible regions of posterior uncertainties for a combination of two parameters. Histograms on the diagonal represent distribution of posterior errors in parameter estimates. Numbers ($-1 >$ and < 1) on the lower triangular matrix are correlation coefficients among control parameters.

According to MCMC results, posterior errors in estimates of g_{max} have significantly reduced compared to the case in which we considered independent mesocosm data. The credible regions of posterior uncertainties in ϕ and γ are similar to ones derived for the fourth case. Thus, addition of zooplankton data facilitate estimation of g_{max} but have no effect on ϕ and γ . Another major improvement is in collinearities associated with g_{max} . Interestingly, when we

calibrate our model with data that include zooplankton N and C concentrations, we do not find strong correlations between γ and the other two sink parameters. This in contrast to results of the four model experiments described before. Thus, our twin experiment results exemplify that g_{max} can be well identified independent of other ϕ and γ if we assimilate data for zooplankton biomass.

MCMC algorithm finds similar estimates of $\hat{\theta}$ for the twin experiment and the case without covariances, and hence the model performance do not change significantly between these two cases in terms of reproducing of observations. However, we certainly find that parameters (especially g_{max}) are better constrained in the twin experiment.

2.4 Discussion

In this study we find optimised model results to be sensitive to the metric (cost function applied). The model seems to perform best when calibrated with cost function that accounts for covariances in data. Our results suggest that among all sink parameters only γ can be estimated comprehensively by the data. In contrast, g_{max} and ϕ parameters show strong collinearities in all four cases, suggesting they cannot be identified independently. We find large uncertainties in posterior estimates of g_{max} which is associated with the unavailability of zooplankton data. Fennel et al. (2001) reported a similar problem in constraining grazing parameters while calibrating their marine ecosystem model with DIN and Chl data of the Bermuda Atlantic Time-series study (BATS). They obtain large a posteriori errors for zooplankton maximum growth and mortality rates when zooplankton data were not available.

2.4.1 Sensitivity of the data assimilation model to error statistics

Stow et al. (2009) stressed that quantification of model skill provides useful insights on model selection and application. Our optimisation results show significant differences in parameter estimates and uncertainties. Essentially, the capability of our model to reproduce data depends on the information expressed in the cost function. Hence, it is important to comprehend how the information in the cost function affects model performance. According to our MCMC results, overall model performance is best seen for the fourth case when data covariances are considered. This inference can be made from some good agreement between model results and observations and the small magnitude of posterior errors in estimates of loss parameters (Figs. 2.6 and 2.4b).

Stow et al. (2009) highlighted that much of the art of constructing probabilistic based cost functions involves developing the covariance matrix that weights the contribution of every data points to the total cost in a time-series experiment. A cost function with covariances accounts for temporal changes that are correlated in observations (e.g. drawdown in DIN while PON increases). Our results show that even critical parameters, like those that determine phytoplankton loss rates are better constrained when the model is calibrated with J_4 compared to the other cases. This in line with Doney (1999),

who pointed out for assessing the time and space variability in pelagic ecosystem processes it is important to have good information on the error variance and cross correlations of biogeochemical variables. Dowd et al. (2014) stressed that by considering covariances in the cost function effects of non-normality in the observations and strong nonlinear dynamics are accounted for, and thus estimation of parameters is facilitated. Many studies have applied cost functions with covariances in their sequential and inverse data assimilations methods to estimate model parameters and state variables (e.g., Vallino 2000; Lawson et al. 1996; Natvik and Evensen 2003; Brasseur et al. 2005; Schartau et al. 2017; Krishna and Schartau 2017). Schartau et al. (2017) employed two probabilistic cost functions (with and without data covariances) to calibrate their model with mesocosm data to estimate sink parameters. They obtain smaller error margins in estimates of parameters when covariances are considered in the cost function. These results of Schartau et al. (2017) are consistent with ours (third and fourth cases). Fennel et al. (2001) applied a “Chi-squared” cost function to assimilate uncorrelated DIN and Chl data of the Bermuda Atlantic Time-series study (BATS) in a marine ecosystem model. Results of their sensitivity analysis show that many parameters are poorly constrained and they describe the magnitude of posterior errors in parameter estimates as “enormous”. Likewise, Ward et al. (2010) highlighted the problem of underestimation in 1-D marine biogeochemical model when validated it with the Arabian sea data using a cost function that disregarded covariances.

According to our results, RMSE and rRMSE appear to be less efficient metrics compared to probabilistic based cost functions in assessing model skill. Eknes and Evensen (2002) applied a normalised RMSE cost function to evaluate the performance of a 0-D model for uni- and multivariate cases. They obtain best result for the univariate case. Stow et al. (2009) pointed out that RMSEs are sensitive to phase errors in time and space, especially for multi-dimensional data sets. Therefore, it seems the RMSE is a plausible metric to calibrate models only when one observational type is considered. However, our model performs slightly better in terms of constraining parameters when calibrated with the RMSE than rRMSE. MCMC results show that parameter identification fails when rRMSE is applied, especially for g_{max} and ϕ . Large uncertainties in posterior estimates of g_{max} and ϕ suggest that rRMSE is not sensitive to variations in parameter values. This could happen because of the “smoothing” effect of the rRMSE. Since rRMSEs are typically normalised to smoothing constants (e.g. temporal means of observations), high weights cannot be assigned when large model-data discrepancy occurs. In other words, large errors cannot be penalised appropriately while considering the rRMSE, that may eventually result in biases. Losa et al. (2004) stressed that the weights for the cost function should be chosen carefully as they strongly affects the quality of the solution. Thus, while considering rRMSE we possibly impose some strong constraints, which is not the case with the RMSE.

2.4.2 Loss rates of phytoplankton biomass

In this study we focussed on only three parameters; all of them specifying loss rates of phytoplankton biomass. These losses are associated with different pathways of mass flux through the plankton ecosystem. We addressed the difficulty of estimating these parameters for the case of not having any

data available of dissolved organic matter, detritus and zooplankton concentrations. Also, information about rates of exudation, aggregation and grazing are missing. This situation is typical and it introduces uncertainties to the determination of mass flux. As discussed before, the choice of metric may become particularly relevant when trying to come up with small uncertainties in parameter estimates and model results. Figure 2.8 illustrates the range of the three pathways in association with posterior uncertainties of the parameter estimates as obtained from the MCMC method for the first and fourth cases.

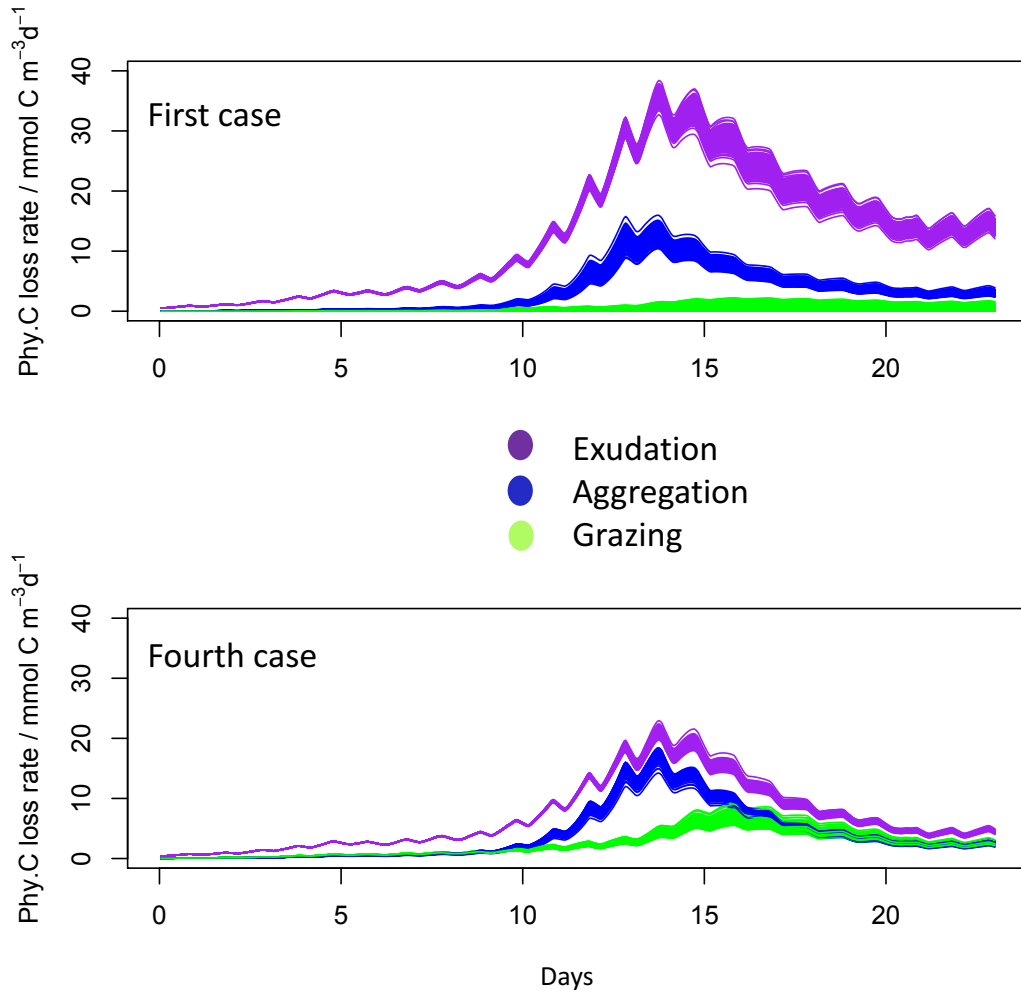


FIGURE 2.8: Optimised estimates of carbon losses from phytoplankton due to exudation, grazing and aggregation corresponding to ensemble solutions obtained from the MCMC method for the first case (when RMSE is applied) and the fourth case (when data covariances are considered in the cost function).

It is evident from the figure that loss of phytoplankton biomass due to exudation is greater than the one for grazing. According to our model, exudation seems to be an important pathway in this mesocosm experiment. The respective rate estimates seem reliable, because they also determine the amount of dCCHO being released and then transformed to TEPC. The comparison of the resulting TEPC concentrations correspond with those observed. It is consistent with the study of Engel et al. (2004b) in which they described algal exudates in form of dCCHOs as potential sink of organic carbon. Schartau

et al. (2007) highlighted that the release of dissolved organic matter (DOM) by phytoplankton followed by formation of TEPC is an alternative and fast pathway to produce POC that can be subject to carbon export. The TEP is likely involved in the process of aggregation and thus in the formation of particles described as detritus in our model approach. We did not resolve any dependency between the formation of detritus and TEP, but this can be of relevance because TEP may increase the apparent stickiness (Engel, 2000).

The reason why algal cells exude is discussed as being a release excess photosynthates that accumulate when carbon fixation exceeds incorporation into cellular materials (Björriksen, 1988). However, the question arises here is why our model predicts exudation as an important pathway for carbon export for the given setup. Schartau et al. (2007) highlighted two modes of carbon over-consumption. They associated first mode of DOC exudation to phytoplankton growth. The second mode is related to formation of TEPC that leads to rise in POC, while PON is constant. Apparently, data (DIN, DIC, Chl, POC and PON) we assimilated in cost functions for different cases are sufficient to constrain phytoplankton growth. Consequently, we are able to constrain values of γ as exudation is linearly coupled with the growth, especially during the bloom period. Since dCCHOs (exudates) are precursors of TEP, we are also able to reproduce TEPC concentrations in the mesocosms (see Fig. 2.6d). Joassin et al. (2011) used the data from the “present” treatment (Delille et al., 2005) of the PeECE-I experiment to calibrate their model, and they found a sharp increase in simulated TEPC concentrations between day 1 and 10. In contrast, our model do not show such behaviour. Although we do not assimilate TEPC data in our cost functions, we could still reproduce the corresponding observations well, especially for the case with covariances (Figure 2.6d). As we assimilated mean data of the “future” and “glacial” treatments in our cost functions, it may be possible they represent “high CO₂-like” conditions. Riebesell et al. (2007) have shown enhanced carbon fixation rates for photoautotrophs at high DIC concentrations. From an ecological perspective, our model suggests that phytoplankton may assimilate excess carbon when being exposed to elevated CO₂ levels, which they release in form of DOC and TEP. This assumption is in agreement with Engel et al. (2004a), they traced 63% fraction of over-consumed carbon in particulate polysaccharides (e.g. TEPC) and remaining 37% in DOC. Phytoplankton exudate serve as an energy source for the growth of pelagic bacteria (Larsson and Hagström, 1979). Hence, a positive feedback on microbial loop is likely to be associated with algal exudation. Results of Joassin et al. (2011) show that bacterial growth highly depends on the availability of DOM.

2.5 Conclusion

In an inverse data assimilation approach, it is critical to apply a quantitative metric that maximises the skill of a model for meaningful model simulations and the estimation of state variables. With our data assimilation approach we could test performances of four different cost functions that configure our model with mesocosm data.

From this study we learned that the choice of metric is critical when estimating parameter values. Our results show that uncertainty ranges can be larger for minimisations of RMSE in contrast to a probabilistic based cost

function that accounts for observational errors (i.e negative logarithm of likelihood). We found that parameters are best identified when data covariances are considered in the cost function (fourth case).

Correlations between parameter estimates can change depending on the type of metric applied. According to our MCMC results, the collinearity between γ and other loss parameters significantly reduces when the cost function with covariances is used to calibrate the model, and therefore exudation rates can be estimated independent of other parameters, which is not the case for the other metrics.

The optimised model results of carbon loss due to exudation are higher than the ones that correspond to aggregation and grazing. Our model results suggest that production of extracellular exudation is an important sink pathway of phytoplankton biomass in the mesocosm experiment.

The grazing parameter g_{max} remains unconstrained in most of the cases, which is associated to unavailability of explicit zooplankton data. To test this we designed a twin experiment in which model generated zooplankton data is assimilated. Our results show that some availability of data on zooplankton biomass already improves the estimation of g_{max} .

Chapter 3

A data–model synthesis to explain variability in calcification observed during a CO₂ perturbation mesocosm experiment.

Shubham Krishna and Markus Schartau

Published article: Krishna, S., and Schartau, M. (2017). A data–model synthesis to explain variability in calcification observed during a CO₂ perturbation mesocosm experiment. *Biogeosciences*, 14(7), 1857.

Biogeosciences, 14, 1857–1882, 2017
www.biogeosciences.net/14/1857/2017/
doi:10.5194/bg-14-1857-2017
© Author(s) 2017. CC Attribution 3.0 License.



A data–model synthesis to explain variability in calcification observed during a CO₂ perturbation mesocosm experiment

Shubham Krishna and Markus Schartau

GEOMAR Helmholtz Centre for Ocean Research Kiel, Düsternbrooker Weg 20, 24105 Kiel, Germany

Correspondence to: Shubham Krishna (skrishna@geomar.de) and Markus Schartau (mschartau@geomar.de)

Received: 22 September 2016 – Discussion started: 26 September 2016

Revised: 31 January 2017 – Accepted: 27 February 2017 – Published: 6 April 2017

Abstract. The effect of ocean acidification on growth and calcification of the marine algae *Emiliania huxleyi* was investigated in a series of mesocosm experiments where enclosed water volumes that comprised a natural plankton community were exposed to different carbon dioxide (CO₂) concentrations. Calcification rates observed during those experiments were found to be highly variable, even among replicate mesocosms that were subject to similar CO₂ perturbations. Here, data from an ocean acidification mesocosm experiment are reanalysed with an optimality-based dynamical plankton model. According to our model approach, cellular calcite formation is sensitive to variations in CO₂ at the organism level. We investigate the temporal changes and variability in observations, with a focus on resolving observed differences in total alkalinity and particulate inorganic carbon (PIC). We explore how much of the variability in the data can be explained by variations of the initial conditions and by the level of CO₂ perturbation. Nine mesocosms of one experiment were sorted into three groups of high, medium, and low calcification rates and analysed separately. The spread of the three optimised ensemble model solutions captures most of the observed variability. Our results show that small variations in initial abundance of coccolithophores and the prevailing physiological acclimation states generate differences in calcification that are larger than those induced by ocean acidification. Accordingly, large deviations between optimal mass flux estimates of carbon and of nitrogen are identified even between mesocosms that were subject to similar ocean acidification conditions. With our model-based data analysis we document how an ocean acidification response signal in calcification can be disentangled from the observed variability in PIC.

1 Introduction

Much knowledge about growth and mortality of phytoplankton has been inferred from experiments where environmental factors like light, temperature, and nutrient availability have been predominantly controlled, e.g. in laboratory experiments with batch cultures or with chemostats. Typically, these experiments are designed to determine a physiological response to variations of a single factor, e.g. explaining changes in photosynthetic rate when exposed to different light conditions (e.g. Platt et al., 1977; Marra and Heineemann, 1982; Lewis and Smith, 1983; Geider et al., 1985; Harrison and Platt, 1986; Harding et al., 1987). Many laboratory experiments are performed with monocultures, with the advantage that physiological responses may then become well expressed in measurements while variability between replicates or even between repeated experiments should remain low. In this context a series of laboratory studies with monocultures of calcifying coccolithophores were conducted to investigate responses in calcification to variations in carbonate chemistry, often with *Emiliania huxleyi*, (e.g. Zondervan et al., 2002; Iglesias-Rodriguez et al., 2008; Langer et al., 2009; Barcelos e Ramos et al., 2010). These studies were motivated by the expectation that the observed trend in ocean acidification (OA) will affect calcifying algae and that their physiology is likely sensitive to the seawater's calcite saturation state (Feely et al., 2004; Orr et al., 2005).

The repeated laboratory OA experiments showed ambiguous responses in calcification to variations in carbon dioxide (CO₂) concentrations and Findlay et al. (2011) pointed out that differences in laboratory methodology, but also details in experimental design, are likely the reason for the large observed variability in *E. huxleyi* responses to changes in carbonate chemistry. Similarly, Engel et al. (2014) stressed that

variations in the observed ratio between particulate inorganic carbon and particulate organic carbon (PIC : POC ratio) increase with the decrease of measured relative growth rates, depending on whether “low” growth conditions were balanced (as achieved with chemostats) or resulted from unresolved transient nutrient-limitation effects in batch cultures. This ongoing discussion is accompanied by the question of how representative the outcomes of monoculture laboratory experiments are, to allow for reliable future projections of OA effects on oceanic calcification rates of coccolithophores and on possible climate feedbacks.

If we seek to make inferences about future changes in calcification under oceanic conditions, experimental data are needed that consider more realistic environmental conditions with a natural phytoplankton community that may include calcifying algae like *Emiliania huxleyi*. A series of studies were conducted to investigate effects of OA on plankton dynamics. Among those were experiments with tanks or bags called mesocosms, with some enclosed water volume that typically comprised a natural plankton community. These mesocosms were typically perturbed and exposed to different CO₂ concentrations, e.g. Pelagic Ecosystem CO₂ Enrichment (PeECE) studies (Riebesell et al., 2008). Few studies focused on the impact of OA on growth of *E. huxleyi*. In contrast to monoculture laboratory experiments, CO₂ perturbation mesocosm experiments yield “net” community response signals that are anticipated to be more indicative for possible future changes in oceanic calcification of coccolithophores. Replicate mesocosms with similar initial nutrients, as well as initial dissolved inorganic carbon (DIC) concentrations typically show comparable temporal response patterns, i.e. an exponential growth phase until nutrients become depleted and a post-bloom period where chlorophyll *a* concentrations decline. However, replicate mesocosms that all included *E. huxleyi* exhibited large deviations in calcification responses, thereby altering carbonate chemistry. Such variability was well reflected in total alkalinity (TA) measurements of the PeECE-I experiment (Delille et al., 2005). Furthermore, during PeECE-I it happened that mesocosms with high and low calcification rates were revealed among replicates in all three CO₂ treatments. To find enhanced variability in calcification in mesocosm experiments is comprehensible and can be attributed to the likely mixture of superimposed responses of multiple plankton species even within replicates of similar CO₂ perturbation. Thus, small deviations in the initial relative mass distribution of photoautotrophs, zooplankton, and detritus between replicate mesocosms can translate into some pronounced variability in measurements even under similar environmental conditions (e.g. Eggers et al., 2014).

Here we investigate data and their variability of replicate mesocosms during the PeECE-I experiment. For this we take a modelling approach to simulate environmental conditions and the predominant dynamics of nine individual mesocosms as described in Engel et al. (2005) and in Delille et al. (2005). Joassin et al. (2011) presented a dynamical model to sim-

ulate the mass flux of carbon (C), nitrogen (N), and phosphorus (P) for the same PeECE-I experiment. Their model resolves growth and losses of *E. huxleyi* together with interdependencies between bacteria, viruses, detritus, and dissolved organic matter (DOM). The model of Joassin et al. (2011) also features the exudation and coagulation process of dissolved polysaccharides (here referred to as dissolved combined carbohydrates, dCCHO) to form transparent exopolymer particles (TEP). In the study of Joassin et al. (2011) some emphasis is put on the enhanced mortality of *E. huxleyi* due to viral lysis and on the variable stoichiometry (C : N ratio) of the particulate organic matter (POC : PON ratio). They did not attempt to resolve a dependency between calcification and CO₂ concentration and therefore restricted their simulations to one treatment with three replicate mesocosms that were exposed to present-day CO₂ concentrations.

The focus of our model approach is different in that we distinguish between two phytoplankton functional types, calcifying algae (e.g. *E. huxleyi*) and bulk non-calcifying algae, i.e. an unresolved combination of picoplankton, dinoflagellates, and diatoms. We assume a CO₂ sensitivity for the ratio of calcification to net carbon fixation (photosynthesis minus respiration), based on results from the meta-analysis of Findlay et al. (2011). In our data–model synthesis we concentrate on the initialisation (initial filling) of the mesocosms, with possible variations in the relative distribution of plankton and detritus resolved in our model. A data assimilation (DA) method is employed for the estimation of parameter values, which helps to disentangle and understand some of the differences and commonalities seen in observations, in particular in TA and PIC data, but also in measurements of dissolved inorganic nitrogen (DIN) and DIC, chlorophyll *a*, as well as in particulate organic nitrogen (PON) and particulate organic carbon (POC).

First we will briefly provide some background information about the experimental setup of PeECE-I, including irradiance, temperature, and salinity, as these environmental factors enter our model simulations. This will be followed by a description of the model equations that include components of the optimality-based approach to simulate algal growth, using parameterisations proposed by Pahlow et al. (2013). Thereafter, the data assimilation method for parameter estimation will be briefly explained. Specific details of the model and of the data assimilation method are given in the Appendix. Ensembles of three distinct model solutions will be presented together with their mass flux estimates of C and N. We will discuss the problem of identifying initial conditions in combination with important model parameters. We will also address the problem of resolving the variability observed in the accumulation of PIC and how this variability is related to the expression of the CO₂ effect introduced to the model.

2 Material and methods

For our analysis we consider the setup and data of the PeECE-I experiment, a study conducted at the Marine Biological Field Station (Raunefjorden, 60.3° N, 5.2° E) of the University of Bergen, Norway between 31 May and 25 June 2001 (Engel et al., 2005; Delille et al., 2005). The objective of this study was to investigate OA effects on marine calcifying algae (coccolithophores) captured in polyethylene bags of enclosed water volumes (mesocosms) and perturbed by different levels of CO₂ concentrations. A dynamical plankton ecosystem model is used for simulations of N and carbon C flux within each mesocosm. We apply a data assimilation method to identify best estimates of model parameter values together with initial conditions for model simulations.

2.1 Experimental data

Nine mesocosms of 2 m diameter and 11 m³ volume were filled with unfiltered, post-bloom, nutrient-depleted water from the fjord. After the filling of the mesocosms, nutrients were added so that all mesocosms had similar initial nutrient concentrations, approximately 15 mmol m⁻³ of nitrate together with nitrite and 0.5 mmol m⁻³ of phosphate. Like the nutrients, the initial TA in all nine bags was 2146 mmol m⁻³ approximately (or if normalised to unit mass $\approx 2200 \mu\text{mol kg}^{-1}$). The bags were covered with air-tight tents of tetra-fluoroethylene foil that allowed 95 % of photosynthetically active radiation (PAR) to pass through. The mesocosm bags were subject to three different levels of perturbation of partial pressure of CO₂: (a) mesocosms 1–3, referred to as M1, M2, and M3, were exposed to similarly high DIC levels (initial DIC = 2119, 2119, 2122 mmol m⁻³) with 700 ppmV of initial $p\text{CO}_2$; (b) M4, M5, and M6 started from DIC = 2048, 2056, 2040 mmol m⁻³ with a corresponding $p\text{CO}_2$ = 370 ppmV; and treatment (c) with initial DIC = 1919 mmol m⁻³, 1929 m⁻³, 1927 m⁻³ with 180 ppmV $p\text{CO}_2$ in mesocosms M7, M8, and M9. Thus, data from three replicate mesocosms are available for each of the three CO₂ treatments. For each mesocosm the partial pressure of atmospheric CO₂ above the surfaces was largely controlled by a continuous injection of gas with a treatment-specific, individually prescribed CO₂ content. Because there was an open space between surface of mesocosms and the tents, we assumed the $p\text{CO}_2$ in the air above the mesocosms' surfaces to be a mixture of 90 % of the perturbed $p\text{CO}_2$ inside a mesocosm and 10 % of the actual atmospheric $p\text{CO}_2$ (340 ppm) in all replicates.

Daily samples were collected and measured over a period of 23 days. For every mesocosm, temperature and salinity data were interpolated to hourly values for direct use as environmental input for model simulations (Fig. 1). Hourly photosynthetic available radiation (PAR) data were derived from meteorological global irradiance measurements of the Geophysical Institute at Bergen (Skartveit et al., 2001). Fig-

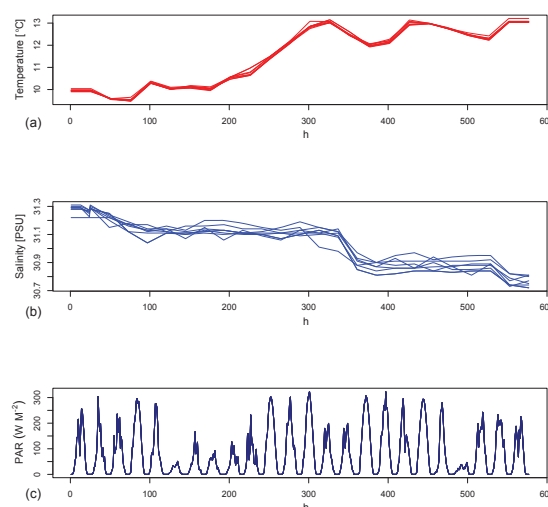


Figure 1. Forcing variables for all nine mesocosms: (a) shows temperature, linearly interpolated to hourly values between daily observations. (b) Displays hourly interpolated salinity values, and (c) reveals the irradiance data with hourly temporal variations resolved.

ure 1 shows that temperature increased by approximately 3 °C during the experiment and variations between the different mesocosms remained small. Small but noticeable differences exist between mesocosms with respect to salinity. In all mesocosms a gradual decrease in salinity was observed, from $S = 31.3$ to approximately $S = 30.8$. The PAR data exhibit variations on an hourly scale, due to changes in cloud cover.

2.2 Modelling approach

For model simulations we assume that all mesocosms are homogeneously mixed, as we neglect an explicit representation of vertical turbulent mixing (0-D-model approach). Furthermore, we assume no light gradient in mesocosms and use depth integrated hourly irradiance data to force the model. The applied model equations describe mass exchange rates of N and C between compartments of (1) dissolved inorganic nitrogen and carbon (DIN and DIC), (2) N and C biomass of coccolithophores and other phytoplankton (CoccoN and CoccoC, PhyN and PhyC), (3) zooplankton (ZooN and ZooC), (4) detritus (DetN and DetC), and (5) labile dissolved organic N and C (DON and DOC), Fig. 2. Due to the design of the PeECE-I experiment, our model includes some additional features. The first is that we consider an explicit representation of dissolved combined carbohydrates (dCCHO) that act as precursors for carbon content of transparent exopolymer particles (TEPC), similar to Schartau et al. (2007) and Joassin et al. (2011). Since our model resolves changes in TA along with DIC so that we can also derive pH val-

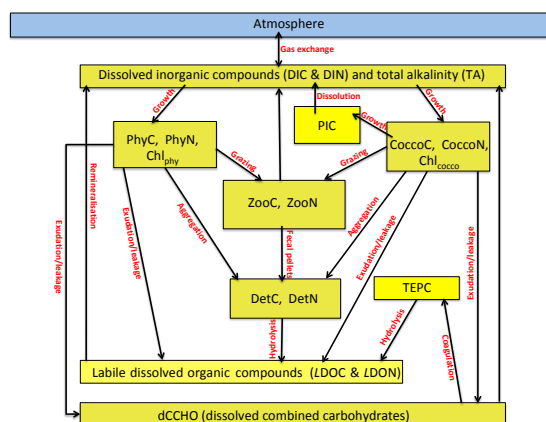


Figure 2. Schematic representation of the model: boxes characterise individual compartments that are represented by one or more model state variables. The arrows represent key biogeochemical processes (named in red) between compartments. One compartment includes dissolved inorganic carbon and nitrogen (DIC and DIN). This compartment also embeds total alkalinity (TA). Biomass and chlorophyll concentrations of photoautotrophs are resolved with respect to carbon and nitrogen explicitly (referred to as PhyC and CoccoC, PhyN and CoccoN, and Chl_{phy} and Chl_{cocco} respectively). Variations in carbon and nitrogen biomass are also resolved for zooplankton (ZooC and ZooN) and for detritus (DetC and DetN). Dissolved combined carbohydrates (dCCHO) are distinguished from other labile dissolved organic matter, described as LDOC and LDON. Only dCCHO are assumed to act as precursor for the formation of transparent exopolymer particles, whose carbon content is explicitly resolved (TEPC). One compartment represents the formation and dissolution of particulate inorganic carbon (PIC), affecting DIC as well as TA.

ues and the corresponding partial pressure of CO₂ ($p\text{CO}_2$). We resolve neither viral infections nor bacterial biomass explicitly, as done in Joassin et al. (2011). Microbial activity is implicitly considered by parameterisations of hydrolysis and remineralisation. Both processes are assumed to be temperature-dependent but are independent of changes in bacteria biomass. Instead, hydrolysis and remineralisation rates are calculated as being proportional to substrate availability only. Likewise, any effects by viral lysis remain unspecified and are an integral part of a single total mortality that is assigned to phytoplankton and coccolithophores. In the following, the general model equations of mass flux of C and N are described as sources and sinks, inducing changes in the mass concentration of the respective state variables.

2.2.1 Photoautotrophs

In our model we distinguish between calcifying and non-calcifying photoautotrophs, coccolithophores (Cocco), and other bulk phytoplankton (Phy). Respective net photoau-

trophic growth rates ($\mu_{\text{cocco/phy}}$) are described as rates of gross carbon fixation (V^C) minus some corresponding sum of respiration costs (r_C) due to the synthesis of chlorophyll *a*, nutrient assimilation, and maintenance: $\mu_{\text{cocco/phy}} = V^C - r_C$. The proportions of V^C and r_C are determined by optimal resource allocation while energetic trade-offs are imposed, as described in Pahlow et al. (2008). These physiological equations of optimal allocation have been shown to be well applicable for a series of different conditions (e.g. including diazotrophy) and scales (e.g. Smith et al., 2011; Pahlow et al., 2013; Arteaga et al., 2014; Fernández-Castro et al., 2016). Here we neglect diazotrophy as well as the effect of phosphorus availability on nitrogen uptake and thus on algal growth. From the data we could not infer any phosphorus limitation of growth prior to nitrogen depletion and we assume that cellular nitrogen (N) directly limits the net growth rate of photoautotrophs ($\mu_{\text{cocco/phy}}$). Nitrogen is generally necessary for synthesising enzymes. According to the model approach of Pahlow and Oeschies (2009), the major metabolic pathways within the algae are regulated by the resources allocated to produce these enzymes. Thus, key processes like photosynthesis, chlorophyll *a* synthesis, and net carbon fixation become affected by internal resource allocation. The model maximises the photoautotrophic growth rates by optimising the allocation of resources to nutrient acquisition sites and to the light-harvesting complex (LHC). The auxiliary variables mentioned above are described in Table A1 in Appendix A. The detailed equations are given in Appendix A2.

Biomass concentrations of photoautotrophs

The biomass build-up (net growth) of photoautotrophs depends on the amount of N and C assimilated by the algae minus losses because of aggregation, grazing by zooplankton, and exudation or leakage of organic matter. The sources minus sinks (sms) terms of the photoautotrophs' biomass are as follows:

$$\begin{aligned} \text{sms of photoautotroph biomass} &= \text{C and N uptake} \\ &- \text{exudation/leakage} - \text{aggregation} - \text{grazing}. \end{aligned}$$

The corresponding sms differential equations of C and N biomass for phytoplankton and coccolithophores are given in Appendix A2.

Chlorophyll *a* concentrations

The synthesis of chlorophyll *a* (Chl) is represented by an optimal trade-off between photosynthesis and respiratory costs in the chloroplast of a cell. The synthesis rate depends on the degree of light saturation (S_I), on the amount of net carbon fixed inside chloroplasts, and on the chlorophyll-to-carbon ratio (θ). Also, the chlorophyll synthesis rate is sensitive to changes in the cellular nitrogen-to-carbon ratio (N:C), Q^N . The descriptions of the above-introduced auxiliary variables are given in Table A1. Like for biomass, the parameterisa-

Table 1. Initial conditions and model parameters that are subject to optimisation.

Initial conditions & parameters for optimisation	Description	Unit
1. PON ₀	Initial concentration of particulate organic nitrogen	mmol N m ⁻³
2. f_{det}	Fraction of PON ₀ assigned to non-living detritus	–
3. f_{zoo}	Fraction of living PON ₀ assigned to zooplankton	–
4. f_{cocco}	Initial coccolithophore fraction of photoautotrophs	–
5. Q_0	Subsistence quota (minimum cellular N : C ratio)	mol mol ⁻¹
6. α_{cocco}	Photosynthetic efficiency of coccolithophores	mol C (g Chl <i>a</i>) ⁻¹ m ² W ⁻¹ d ⁻¹
7. α_{phy}	Photosynthetic efficiency of non-calcifying phytoplankton	mol C (g Chl <i>a</i>) ⁻¹ m ² W ⁻¹ d ⁻¹

tions for chlorophyll *a* are identical for the calcifying and non-calcifying phytoplankton in our model:

sms of chlorophyll *a* = synthesis of chlorophyll *a*

– aggregation – grazing.

The respective differential equations for chlorophyll *a* of non-calcifying phytoplankton (with subscripts phy) and coccolithophores (cocco) are listed in Appendix A2.

Formation of particulate inorganic carbon (PIC)

The process of calcification in our model depends on the amount of energy provided through photosynthesis and is simply expressed by a ratio of PIC formation per carbon fixed (f_{PIC} , Eq. A21). The differential equation of PIC describes a net accumulation rate (formation minus dissolution) and no explicit distinctions can be made with respect to how PIC becomes eventually distributed between algal biomass, detritus, or zooplankton:

sms of PIC = calcification by coccolithophores

– dissolution of coccoliths (calcite).

The differential equations for precipitation and dissolution of PIC are given in Appendix A4.

2.2.2 Zooplankton

The grazing losses of the photoautotrophs are resolved with an explicit representation of zooplankton biomass. With our grazing approach (Holling type III) no distinctions are made between micro- and meso-zooplankton or between different feeding types. Changes in zooplankton biomass are subject to a mortality (M_{zoo} ; e.g. losses to higher trophic levels). Other loss terms represent respiratory costs (r_{zoo}) as well as excretion (γ_{zoo}). Zooplankton restore C and N towards a constant N : C ratio ($Q_{\text{const}}^{\text{zoo}}$) of 0.19. The restoring time (τ) in our model is equal to 1 day. It mimics an increase in respiration (r_{zoo}) if the N : C ratio falls below $Q_{\text{const}}^{\text{zoo}}$ and an increase in excretion ($\gamma_{\text{zoo}}^{\text{N}}$) if N : C is above $Q_{\text{const}}^{\text{zoo}}$. Details of auxiliary variables related to the zooplankton compartment of

the model are given in Table A1. The buildup of zooplankton biomass depends on the total prey concentrations (phytoplankton and coccolithophores):

sms of zooplankton biomass = grazing on phytoplankton

+ grazing on coccolithophore

– respiration (or excretion) – mortality.

The differential equation for zooplankton biomass and grazing function are given in Appendix A5.

2.2.3 Detritus

Detritus comprises a variety of components with particles of different sizes and sinking rates (Fasham et al., 1990). The detritus resolved by our model simply combines dead plankton biomass and fecal pellets. Sources of detrital C and N mass are given in terms of phytoplankton aggregation and mortality of zooplankton. Aggregation is parameterised with quadratic loss terms of the photoautotrophs. These aggregation equations resolve interactions between two types of particles (small cells of photoautotrophs and large aggregates of detritus): (a) aggregation of cells of photoautotrophs and (b) aggregation of small photoautotrophs with larger detritus – see details in the Appendix A. The two-particle-type approach allows a trade-off between accuracy of estimated mass flux and the resolution of particle size (Ruiz et al., 2002). We assume that hydrolysis is temperature-dependent and that it is responsible for the degradation of detritus, acting as a source for (labile) LDON and LDOC. The equations of detrital C and N can thus be described as follows:

sms of detritus = aggregation of phytoplankton

+ aggregation of coccolithophore

+ zooplankton mortality – hydrolysis.

The respective differential equations of detrital C and N mass are given in the Appendix A7.

2.2.4 Dissolved inorganic compounds (DIN, DIC) and total alkalinity (TA)

Dissolved inorganic nitrogen (DIN)

The DIN pool represents the total concentration of nitrate, nitrite, and ammonium. Nitrogen utilisation by phytoplankton and coccolithophores is a sink of DIN, whereas heterotrophic excretion and remineralisation of LDON are the major sources:

$$\begin{aligned} \text{sms of DIN} = & -\text{N uptake by phytoplankton} \\ & -\text{N uptake by coccolithophores} \\ & +\text{excretion by zooplankton} + \text{remineralisation.} \end{aligned}$$

The sms differential equation for DIN is given in Appendix A8.

Dissolved inorganic carbon (DIC)

The DIC pool combines CO₂, bicarbonate, and carbonate. The primary sinks of DIC are net carbon fixation to support photoautotrophic growth ($\mu_{\text{cocco/phy}}$) and calcification of coccolithophores. We do not differentiate between the utilisation of CO₂ and bicarbonate for algal growth and calcification. Note that net carbon fixation ($\mu_{\text{cocco/phy}}$) in our model becomes slightly negative in the absence of light (dark respiration of the photoautotrophs). Total heterotrophic respiration acts as major DIC source and is expressed by zooplankton respiration and by the remineralisation of dissolved organic carbon (LDOC + dCCHO):

$$\begin{aligned} \text{sms of DIC} = & -\text{net C uptake by phytoplankton} \\ & -\text{net C uptake by coccolithophores} \\ & -\text{calcification} + \text{dissolution of PIC} \\ & +\text{zooplankton respiration} \\ & +\text{remineralisation} + \text{gas exchange.} \end{aligned}$$

The corresponding differential equation for DIC is listed in Appendix A8.

Total alkalinity (TA)

Temporal changes in TA in our model are due to the sinks and sources of DIN and DIP ($\Delta\text{DIP} = \frac{1}{16} \times \Delta\text{DIN}$), a process of precipitation and dissolution of calcite plates produced by the calcifying algae. We follow the nutrient-H⁺ compensation principle described in Wolf-Gladrow et al. (2007). In our model we are resolving the nitrogen flux of zooplankton excretion but we are eventually not resolving any associated net change in TA. This is because we cannot differentiate between the excretion of ammonium (NH₄⁺) and of nitrate (NO₃⁻) and nitrite (NO₂⁻). The excretion of 1 mole NH₄⁺ would increase TA by 1 mole, whereas the excretion of 1 mole NO₃⁻ or NO₂⁻ would decrease TA by 1 mole (Wolf-Gladrow et al., 2007). In other words, we indirectly impose

that half of the N excretion by zooplankton is NH₄⁺ and the other half is NO₃⁻ and NO₂⁻, which would introduce a net TA change of zero. Measured values of DIN, TA, and DIC on day one of the experiment were taken as initial conditions for respective mesocosms.

$$\begin{aligned} \text{sms of total alkalinity} = & \text{N and P uptake by phytoplankton} \\ & + \text{N and P uptake by coccolithophores} \\ & - \text{calcification by coccolithophores} + \text{dissolution of calcite} \\ & - \text{remineralisation of dissolved organic N and P.} \end{aligned}$$

The differential equation for TA is given in the Appendix A8.

2.2.5 Dissolved labile organic matter and transparent exopolymer particles

Dissolved organic matter (DOM) is produced by exudation of the photoautotrophs and by hydrolysis of detrital matter. The DOM is subject to remineralisation, being the source of DIN and DIC. The applied model distinguishes between dC-CHOs and a residual fraction of LDOC and LDON. This distinction is made because only dCCHOs are simulated to act as precursors for the formation of TEPs. In our model the DOM's primary source is freshly exuded and leaked organic matter from photoautotrophs. An additional source of DOM is due to degradation of detrital matter (hydrolysis and microbial exudation) in response to bacterial activity. The fraction of exudates that enter the dCCHO pool may vary between the exponential growth phase and during periods of nutrient limited growth, described as two modes of exudation in Schartau et al. (2007). We therefore introduced a parameterisation ($f_{\text{dCCHO}}^{\text{cocco/phy}}$, Eq. A39) that simulates such a shift in quality of the exudates, depending on the respective cell quota of the coccolithophores and of the other phytoplankton ($Q_{\text{cocco/phy}}^{\text{N}}$). Remineralisation and microbial respiration are respective sinks of LDOC and LDON. Table A1 lists all associated auxiliary variables. The equations for labile DOC and DON are described as follows (with details given in Appendix A9):

$$\begin{aligned} \text{sms of LDON} = & \text{exudation by photoautotrophs} \\ & + \text{hydrolysis/degradation of detritus} \\ & + \text{hydrolysis/degradation of gels} \\ & - \text{remineralisation/respiration of dissolved organic matter.} \end{aligned}$$

2.2.6 Dissolved combined carbohydrates (dCCHO)

By introducing dCCHO we account for an additional sink of DOC other than microbial degradation, which is the physical-chemical transformation of dissolved to particulate matter, here resolved as the coagulation of dCCHO to form transparent exopolymer particles (TEP) of carbon. This transformation is parameterised as an aggregation process, as proposed in Engel et al. (2004) and effectually applied in

Schartau et al. (2007) and in Joassin et al. (2011) (see details in Appendix A10):

sms of dCCHO = exudation – coagulation of dCCHO
 – aggregation of dCCHO with TEPC
 – remineralisation of dCCHO.

Transparent exopolymer particles (TEP)

The carbon content of TEP is explicitly resolved because it can be a significant constituent of POC measurements (Verdugo et al., 2004). This consideration is important for our data–model synthesis, in particular because it affects the stoichiometric C:N ratio of particulate organic matter. The sink terms of dCCHO, described before, are the only sources for TEPC in our model approach. The degradation of TEPC is parameterised similarly to the hydrolysis of detritus:

sms of TEPC = coagulation of dCCHO
 + aggregation of dCCHO with TEPC
 – degradation.

The corresponding differential equation for TEPC production is listed in the Appendix A10.

2.2.7 Model parameters and initial conditions

Out of 33 model parameters, 26 parameters are fixed and the remaining 7 parameters (4 initial condition parameters (f_{cocco} , f_{zoo} , f_{det} , PON_0) and 3 ecological parameters (α_{phy} , α_{cocco} , Q_0) enter the optimisation procedure. The decision on which parameters should become subject to optimisation is based on a series of preceding parameter optimisations and subsequent sensitivity analyses. A major objective is to reduce the number of parameters for optimisation to a meaningful minimum. This facilitates the identification of those parameter values that are of primary concern. Since we address differences in initial conditions in our study, we consider four parameters that determine these differences, and they need to become subject to optimisation. The additionally selected three growth parameters are amongst those to which the model solution is most sensitive. The model solutions are also highly sensitive to variations of the maximum potential nitrogen uptake rate (V_0^N). This parameter is excluded from optimisation, because it is not possible to obtain estimates of (V_0^N) that are independent of estimates of the photosynthetic efficiency. Therefore, a value is assigned to V_0^N that is typical and was used for simulations of other experiments (e.g. Pahlow et al., 2013), ensuring credible estimates of those parameters that are optimised in our study. The mesocosm experiment covers only a short post-bloom period and we found other parameters, like maximum grazing rates and the aggregation parameters, to be weakly constrained by the available data. Their consideration for optimisation would impede the identification of the other more

important parameters. Values assigned to those parameters that are excluded from optimisation are adapted from other studies (e.g. Pahlow et al., 2013; Schartau et al., 2007).

Initial condition values for some of the state variables in the model are computed by initial condition parameters, given in fractions. The initial biomass during the start of the experiments, specified by PON_0 , is distributed between living and non-living biomass, which is determined by the parameter of the initial detritus fraction (f_{det}). The living biomass is further distributed between photoautotrophs and zooplankton, specified by the initial zooplankton fraction parameter (f_{zoo}). Finally, the remaining relative distribution of photoautotrophic biomass is set by f_{cocco} . For example, a value of $f_{\text{cocco}} = 1$ would mean that all photoautotrophic biomass is associated with the presence of coccolithophores exclusively.

$$\text{PON}_0 = \text{DetN}_0 + \text{ZooN}_0 + \text{CoccoN}_0 + \text{PhyN}_0 \quad (1)$$

with the individual fractions:

$$\text{DetN}_0 = f_{\text{det}} \cdot \text{PON}_0 \quad (2)$$

$$\text{ZooN}_0 = f_{\text{zoo}} \cdot (\text{PON}_0 - \text{DetN}_0) \quad (3)$$

$$\text{CoccoN}_0 = f_{\text{cocco}} \cdot (\text{PON}_0 - \text{DetN}_0 - \text{ZooN}_0) \quad (4)$$

$$\text{PhyN}_0 = (1 - f_{\text{cocco}}) \cdot (\text{PON}_0 - \text{DetN}_0 - \text{ZooN}_0) \quad (5)$$

For initial zooplankton, coccolithophore, and phytoplankton biomass we apply a constant C:N ratio of 6.625. We consider a higher C:N ratio ($= 2 \times 6.625$) only for initial detritus. Since the mesocosms were filled with post-bloom, nutrient-depleted water masses, we assume that all dead particulate organic matter has a C:N ratio that is rather typical for such post-bloom conditions. Initial conditions of PIC, DIC, and TA are taken from the data for respective mesocosms, whereas we assume same small fixed values (e.g. $\text{DON} = 0.05$, $\text{DOC} = 102.5$, $\text{dCCHO} = 1.0$ and $\text{TEPC} = 3.5 \text{ mmol m}^{-3}$) as initial conditions for all mesocosms.

2.3 Design of data assimilation (DA) approach

A peculiarity of the PeECE-I experiment is that high and low changes in TA were found in all three CO₂ treatments, in response to differences in calcification (Delille et al., 2005). Because the three distinct patterns in calcification (Fig. 3) are attributable to all three treatments, a factor other than the CO₂ perturbations induced variations between the individual mesocosms. For all other observations no such clear pattern could be identified. We designed our data assimilation approach according to this finding and therefore investigate three possible situations (model solutions) that differ in their TA response: low, medium, and high calcification (referred to as LC, MC, and HC respectively). Thus, for each of these three (LC, MC, and HC) situations we find three mesocosms that were subject to three different CO₂ levels (initial 700, 370, and 180 ppmV). By adapting the same nomenclature as in Engel et al. (2005) and in Delille et al. (2005), we can

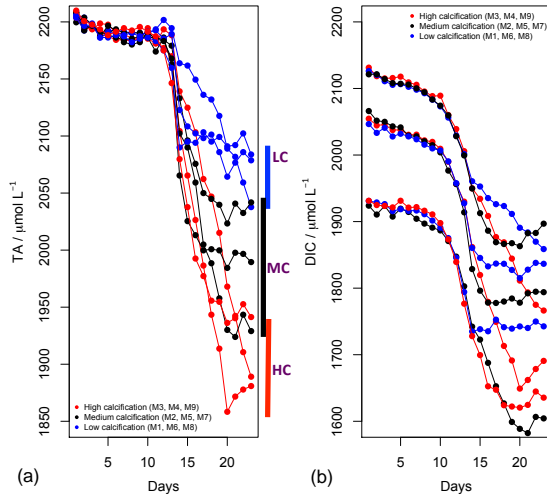


Figure 3. (a) shows three distinct calcification patterns, reflected in total alkalinity (TA) data. Those mesocosms that exhibit high TA values (a reduced drawdown during the bloom and post-bloom period) feature rates of low calcification (LC, in blue colour). Mesocosms with low TA values (a strong reduction of TA) reveal rates of high calcification (HC, marked red). Rates of medium calcification (MC) are assigned to the remaining mesocosms (with intermediate TA values, marked black). (b) shows the respective different CO₂ treatments in the same colours as for LC, MC, and HC. The figure shows that each calcification case (LC, MC, and HC) includes mesocosm of all three CO₂ treatments.

assign the mesocosms M1, M6, and M8 to those with low calcification rates (highest TA), M2, M5, and M7 to the ones with medium calcification, and finally M3, M4, and M9 to mesocosms with high calcification rates (lowest TA).

2.3.1 Definition of cost function (data–model misfit)

In our data assimilation approach we consider data from the three cases (LC, MC, and HC) separately, but we make identical statistical assumptions. The observation vector (y_i) contains daily means of three mesocosms of the following measurements:

1. dissolved inorganic carbon (DIC, mmol m⁻³),
2. dissolved inorganic nitrogen (DIN) (nitrate + nitrite, mmol m⁻³),
3. chlorophyll *a* (Chl *a*, mg m⁻³),
4. particulate organic nitrogen (PON, mmol m⁻³),
5. particulate organic carbon (POC, mmol m⁻³),
6. particulate inorganic carbon (PIC, mmol m⁻³),
7. total alkalinity (TA, mmol m⁻³).

Like the data vector y_i , the vector $H_i(x)$ represents mean values of three simulated mesocosms for each calcification case (LC, MC, and HC). It combines results of model states: C and N biomass concentrations of the photoautotrophs (PhyN & PhyC and CoccoN & CoccoC), of zooplankton (ZooN & ZooC), of detritus (DetN & DetC), and carbon concentration of transparent exopolymers particles (TEPC). The vector of differences (d_i) between observation (y_i) and model results $H_i(x)$ is given as follows.

$$d_i = y_i - H_i(x) = \underbrace{\begin{pmatrix} \text{DIC}_i \\ (\text{NO}_3^- + \text{NO}_2^-)_i \\ \text{Chl } a_i \\ \text{PON}_i \\ \text{POC}_i \\ \text{PIC}_i \\ \text{TA}_i \end{pmatrix}}_{\text{data}} - \underbrace{\begin{pmatrix} \text{DIC}_i \\ \text{DIN}_i \\ (\text{Chl}_{\text{phy}} + \text{Chl}_{\text{cocco}})_i \\ (\text{PhyN} + \text{CoccoN} + \text{ZooN} + \text{DetN})_i \\ (\text{PhyC} + \text{CoccoC} + \text{ZooC} + \text{DetC} + \text{TEPC})_i \\ \text{PIC}_i \\ \text{TA}_i \end{pmatrix}}_{\text{model results}} \quad (6)$$

For the cases LC, MC, and HC we calculated daily residual standard errors (σ_i) based on the measurements. Unlike other variables, the estimation of the standard errors for DIC is not straightforward because of the different CO₂ levels. For the derivation of the standard errors we considered the differences (offsets) of the mean initial DIC concentrations between the different CO₂ treatments. DIC concentrations of those mesocosms that were initially exposed to high-CO₂ (DIC) concentrations are “offset” – corrected so that their initial mean DIC matches the initial mean of the present-day DIC concentrations. Mesocosms of the low-CO₂ treatment were adjusted likewise. In this manner, all initial mean DIC concentrations have become identical, but changes and variations (between the mesocosms) with respect to these mean values remain. Thus, variances of the respective LC, MC, and HC mesocosms can be calculated after applying these (two) offset corrections to all DIC data of the high- and low-CO₂ treatments. Eventually, individual standard errors for the LC, MC, and HC mesocosms are derived for all sampling dates.

The time-varying covariance matrices \mathbf{R}_i are constructed with \mathbf{S}_i (with diagonal elements of standard errors, see Eq. B3 in Appendix B) together with some correlation matrices ($\mathbf{C}_{(y)}$). Correlations between measurements were computed based on data of all nine mesocosms. Two matrices $\mathbf{C}_{(y)}$ have been derived from data for two distinct periods: (1) the exponential growth phase and (2) the post-bloom pe-

riod.

$$\mathbf{R}_i = \mathbf{S}_i \cdot \mathbf{C}_{(y)} \cdot \mathbf{S}_i \quad (7)$$

Equation (7) is applied because correlations between observations can change from pre-bloom period to post-bloom period. For example, PON and DIC are strongly negatively correlated during the exponential growth phase but become weakly positively correlated during the post-bloom period, when both DIC and PON decrease. The correlation matrices, $\mathbf{C}_{(y)}$, for the two respective periods are also given in the Appendix B.

A maximum likelihood (ML) estimator is applied, meaning that no explicit prior information is considered for the estimation of parameter values. Eventually, we use three similar cost functions but with data (y) and covariances (\mathbf{R}) from the respective three mesocosms of each case. These daily data (y_i) are available for a period of $N_t = 23$ days, with subscript i indicating the day when measurements were made. The elements of the parameter vector of interest (Θ) are those parameters listed in Table 1, including the initial value of PON_0 and initial condition parameters that further specify how PON_0 is distributed between detritus, zooplankton, coccolithophores, and the remaining photoautotrophs. For a ML estimation of the parameters (including the initial conditions) we maximise the conditional probability of explaining the data, given our model, together with a set of values assigned to the parameters (to each element of Θ):

$$p(y|\Theta) = \text{constant} \cdot \exp\left[-\frac{1}{2} \sum_{i=1}^{N_t} d_i^T \mathbf{R}_i^{-1} d_i\right] \propto \exp\left[-\frac{1}{2} J(\Theta)\right]. \quad (8)$$

The ML estimate of parameter values can be found by actually identifying the minimum of the exponent of $p(y|\Theta)$ of Eq. (8), since the constant term is independent of Θ . We thus compute and minimise the following cost function $J(\Theta)$:

$$J(\Theta) = \sum_{i=1}^{N_t} (y_i - H_i(x))^T \mathbf{R}_i^{-1} (y_i - H_i(x)). \quad (9)$$

We not only wish to identify the minimum of $J(\Theta)$ that corresponds with one best estimate of parameter values ($\hat{\Theta}$) but also confine a credible region of parameter estimates. This credible region tells us how reliable the parameter estimates are (yielding lower and upper credibility limits) and resolves correlations (collinearities) between the parameters. The parameter optimisation procedure implemented in this study is described in detail in the Appendix B.

3 Results

3.1 Parameter estimates for specific mesocosms with low, medium, and high calcification

The same seven model parameters (Table 1) were optimised for all three calcification cases (LC, MC, and HC) independently, using data from respective mesocosms. With our data assimilation approach we can thus specify commonalities and differences between model solutions for mesocosms with LC, MC, and HC. Table 2 lists all ML estimates, which correspond with the best model solutions obtained with the Markov Chain Monte Carlo (MCMC) method. Collinearities are expressed by the correlation coefficients of two parameter combinations, which we have also calculated based on results of the MCMC method (Table 3).

Credible interval limits for each parameter were derived from nonparametric probability densities of the MCMC estimates. Figure 4 shows cumulative density function (CDF) for corresponding posterior probability distributions. The steeper the CDF increase is, the narrower the 95 % credible interval of the parameter estimate. According to the width of credible intervals we find uncertainty ranges of initial conditions parameters f_{det} , f_{zoo} , and PON_0 to be generally small for all three cases of calcification respectively. The initial condition parameters are best constrained for the solution of medium calcification (MC). The parameter f_{cocco} shows the largest uncertainty for the HC case. A large fraction ($\approx 90\%$) of initial biomass comprises of detrital matter in all three solutions. Table 4 shows mean concentration values of PON_0 , DetN_0 , ZooN_0 , CoccoN_0 , and PhyN_0 along with their uncertainties according to respective MCMC estimates. Initial zooplankton concentration is highest in HC solutions. Thus, more photoautotrophic biomass is lost due to grazing by zooplankton and less by aggregation in model solutions for HC, which is reflected by the negative correlation between initial condition parameters f_{zoo} and f_{det} . For those parameters that do not specify the initial conditions we hoped to find that all credible intervals overlap, which would have suggested insignificant differences between the estimates. A single set of values of these parameters could then be unambiguously used for simulations of all nine mesocosms, independently of how the values of the initial conditions turned out to be. This is not the case, as can be seen in Fig. 4 and in the correlation coefficients (Table 3). Estimates of the subsistence quota (Q_0) are lower for the mesocosms with high and medium calcification rates. Apparently, lower Q_0 and higher α_{cocco} values are required to build up high coccolithophores biomass in mesocosms with high calcification rates as initial coccolithophores concentration is low and grazing pressure is high.

During the post-bloom period, the mesocosms pooled in HC reveal TA changes that are consistently higher than in the LC mesocosms. In fact, these differences become well reflected in our parameter estimates. Thus, our optimised ensemble model solutions are providing the statistical evidence

Table 2. Maximum likelihood parameter estimates of three model solutions: low, medium, and high calcification (LC, MC, and HC).

Parameter	Description	LC	MC	HC	Units
PON ₀	Parameter of initial PON concentration	1.25	1.90	1.61	mmol N m ⁻³
<i>f</i> _{det}	Parameter of initial detritus fraction	0.89	0.89	0.89	–
<i>f</i> _{zoo}	Parameter of initial zoopl. fraction	0.72	0.63	0.88	–
<i>f</i> _{cocco}	Parameter of initial coccolithophore fraction	0.39	0.88	0.40	–
<i>Q</i> ₀	Subsistence N : C ratio	5.5 × 10 ⁻²	4.2 × 10 ⁻²	4.2 × 10 ⁻²	–
<i>α</i> _{cocco}	Photosynth. light absorpt. coeff. of coccolithoph.	1.40	0.50	1.66	mol C (g Chl <i>a</i>) ⁻¹ m ² W ⁻¹ d ⁻¹
<i>α</i> _{phy}	Photosynth. light absorpt. coeff. of non-calcifiers	1.73	3.10	1.71	mol C (g Chl <i>a</i>) ⁻¹ m ² W ⁻¹ d ⁻¹

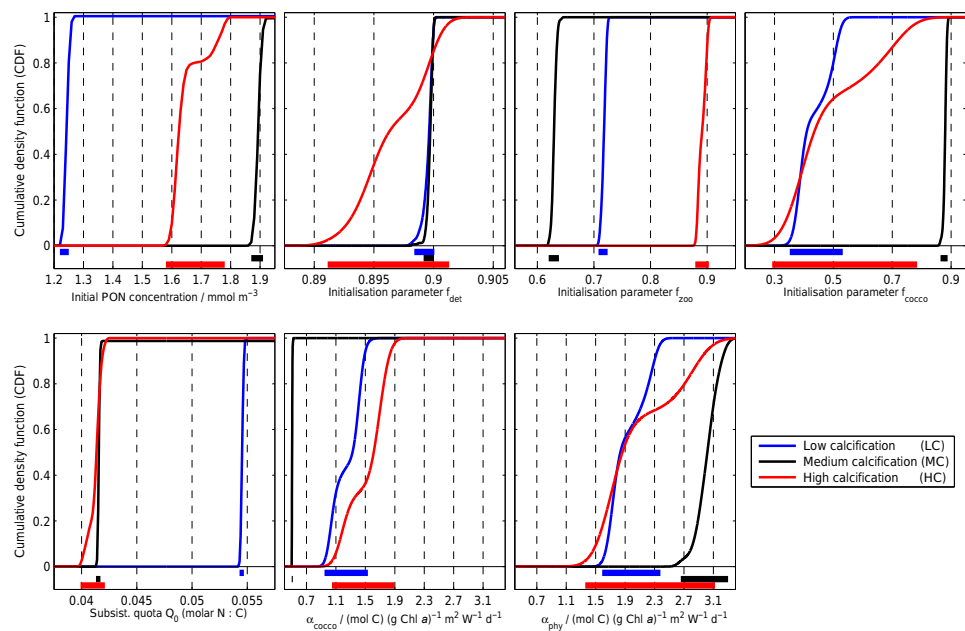


Figure 4. Probability distributions of the initial condition and physiological model parameters: the cumulative sum of non-parametric probability densities (CDF) were derived from the posteriors of the Markov Chain Monte Carlo (MCMC) approach. The bars on the bottom of each panel show respective 95 % credible (uncertainty) ranges of the parameter estimates.

that HC and LC are significantly different. With respect to the mesocosms assigned to the MC case we see in our parameter estimates and ensemble model solutions that they are rather close to conditions also met by the HC mesocosms. In this case the differences in parameter estimates (between MC and HC) are small, although we find significantly different estimates for α_{cocco} and for f_{zoo} between MC and HC (see Fig. 4). Thus, we may have one or two out of the three MC mesocosms that might have been better assigned to the HC case. However, this is reflected in our data assimilation results and we are primarily concerned with the upper and lower extremes in calcification, as resolved by the six mesocosms in the LC and HC cases.

3.2 Data–model comparison

The variational range of parameter estimates (Fig. 4) induce ensembles of model trajectories (model results) that are statistically indistinguishable (or equivalent). Based on these posterior ensemble parameter estimates of all three calcification solutions we find a general good agreement between model results and the data (Fig. 5).

The ensembles reflect uncertainty ranges in model solutions, which correspond nicely with most of the variability in observations. Almost the entire range of variability in TA is recovered with our three distinct solutions of calcification. The observed variability in POC is captured with the optimal ensemble model solutions. Only few maximum values

Table 3. Correlation coefficients of parameter estimates of low, medium, and high calcification model solutions (LC, MC, and HC). Correlation coefficients ≥ 0.6 are marked bold face.

	f_{det}	f_{zoo}	f_{cocco}	Q_0	α_{cocco}	α_{phy}
PON ₀	−0.03/0.03/−0.30	0.57/0.48/0.51	−0.10/0.29/ 0.66	0.05/−0.20/−0.34	0.11/0.03/−0.56	−0.10/0.19/ 0.60
f_{det}	1	−0.51/−0.33/− 0.92	0.13/0.01/−0.28	0.23/0.25/0.11	−0.15/−0.10/0.10	0.13/0.03/−0.40
f_{zoo}		1	−0.47/0.24 / 0.5	−0.11/−0.30/−0.16	0.50/0.52/−0.38	−0.42/0.22/ 0.63
f_{cocco}			1	0.10/−0.12/−0.25	− 0.99 /−0.15/− 0.95	0.99 / 0.93 / 0.93
Q_0				1	−0.10/−0.25/ 0.18	0.13/0.10/−0.26
α_{cocco}					1	− 0.97 /−0.18/ − 0.87
α_{phy}						1

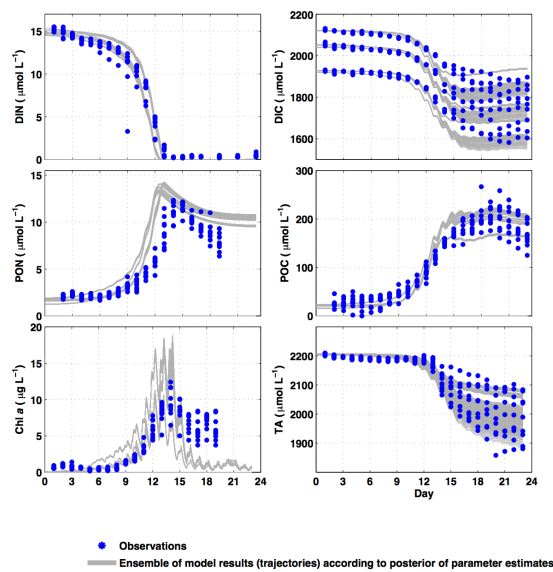


Figure 5. Full variational range of model outputs due to uncertainties in parameter estimates. Model ensembles of high, medium, and low calcification solutions compared with observations.

seen in POC data remain unresolved, likely because we have optimised parameters that hardly introduce changes in the solution of TEPC concentrations. The model solutions exhibit some faster increase in the accumulation of PON during the exponential growth phase, in spite of the fact that DIN data are well matched. Although this systematic model offset (bias) is pronounced, it does not correspond with any similar model bias in POC. Another general offset can be seen for simulated Chl *a* concentrations during the post-bloom period. Our model shows sharp draw down in Chl *a* in all three solutions (HC, MC, and LC) during the post-bloom period, whereas observed Chl *a* values are more variable.

Table 4. Mean initial values of PON (PON₀), detritus (DetN₀), zooplankton (ZooN₀), coccolithophores (CoccoN₀), and bulk phytoplankton (PhyN₀) according to posterior of the (initial condition) parameter estimates of three solutions: low, medium, and high calcification (LC, MC, and HC).

State variable name	LC/mmol N m ^{−3}	MC	HC
PON ₀	1.2 ± 0.01	1.9 ± 0.01	1.7 ± 0.1
DetN ₀	1.1 ± 4 × 10 ^{−4}	1.7 ± 1 × 10 ^{−3}	1.6 ± 0.01
ZooN ₀	0.1 ± 1 × 10 ^{−3}	0.1 ± 1 × 10 ^{−3}	0.2 ± 0.01
CoccoN ₀	0.02 ± 2 × 10 ^{−3}	0.06 ± 1 × 10 ^{−3}	0.01 ± 2 × 10 ^{−3}
PhyN ₀	0.02 ± 2 × 10 ^{−3}	0.01 ± 4 × 10 ^{−4}	0.01 ± 3 × 10 ^{−3}

3.2.1 Variations in calcification in response to growth conditions

According to our model approach we resolve changes in the rate of calcification relative to the carbon that is assimilated for growth of the coccolithophores. For the period of nutrient replenishment the values of the molar calcification-to-C-assimilation ratio ($\Delta\text{PIC} : \Delta\text{C} \approx 0.5$) are smaller than the values under nutrient-depleted growth conditions. All ensembles of model solutions (LC, MC, and HC) reveal a similar behaviour, with variations in $\Delta\text{PIC} : \Delta\text{C}$ greater than 0.5 (up to 2.2) for growth rates between 0 and 0.3 d^{−1}. These variations depend on the light-acclimation state (e.g. θ_{cocco}), fluctuations in irradiance, and cell quota ($Q_{\text{cocco}}^{\text{N}}$). The variations in $\Delta\text{PIC} : \Delta\text{C}$ during the nutrient-depleted period can be attributed to fluctuations in carbon assimilation due to production of TEPC.

3.2.2 Distinctions between model results of low and high calcification (LC and HC)

Optimised model results of LC yield the highest TA values of all mesocosms, being in accordance with the TA data. DIN concentrations are well resolved by the model, and variations of the ensemble DIN simulations are similarly low to in observations. The previously mentioned biases in PON and Chl *a* are most conspicuous in this LC ensemble of optimal model results. Variability in the POC data of the LC mesocosms is not captured by the model ensemble. But simula-

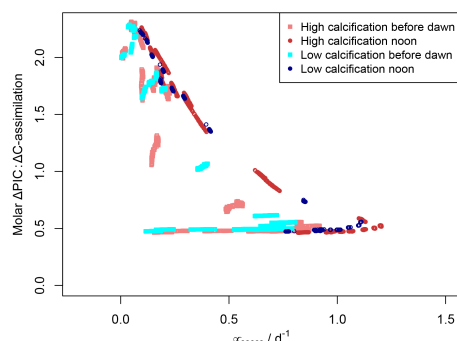


Figure 6. Molar calcification-to-C-fixation ratio compared to net growth rate of coccos (μ_{cocco}) in high and low calcification solutions.

tions results (solid lines in Fig. 7) match the POC mean of the three mesocosms. For PIC we also find a good agreement between model ensemble results and data. However, a noticeable potential bias exists for the PIC response in the high-CO₂ treatment (M1), where model results overestimate PIC data during the maximum bloom period and shortly after nutrient depletion. This overestimation is more pronounced in mesocosms with high CO₂ treatment. The LC ensemble successfully reproduces amplitude of Chl *a* peak seen in data; this is also the case in the solutions of HC mesocosms.

DIN is well resolved in the HC solutions (Fig. 8). Simulated Chl *a* also fits well to observations. HC solutions yield the largest variability in DIC, TA, and PIC amongst all optimised solutions, which we mainly attribute to the large uncertainty ranges of the model parameters f_{cocco} and α_{cocco} . The HC solutions show sharp drawdown in DIC during the bloom period compared to other solution (LC). This can be explained by an enhanced calcification activity due to high growth rates of coccolithophores in HC during the bloom period. Again, model overestimates observed PIC values (M3) under high-CO₂ conditions shortly after the maximum of bloom. PON is best reproduced in this HC case in comparison to LC. Although model HC solutions reproduce the entire variability in observed PIC, the corresponding best fits (to M3, M4, and M9) underestimate PIC data.

3.2.3 Integrated flux estimates of carbon and nitrogen (C and N budgets of mesocosms)

The ensemble model solutions for LC and HC constitute two extremes and we therefore concentrate on the C and N budgets of these two cases. Carbon and N flux estimates were computed as integrals over the entire 23-day period. Figure 9 shows mean C and N flux estimates and their standard errors of the LC solutions of the low- and high-CO₂ treatments. Figure 10 shows the corresponding flux estimates for the HC solution. We learn from these flux estimates that the simu-

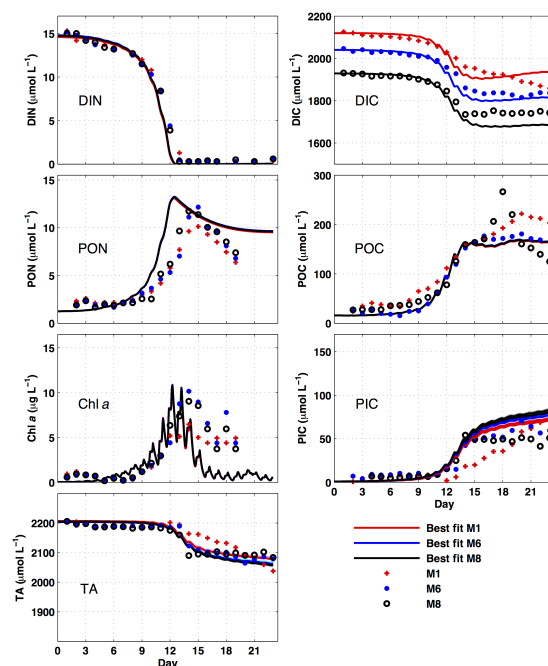


Figure 7. Low calcification solution. The coloured bands represent ensemble of model results according to the posterior and symbols show observations.

lated C and N mass flux estimates differ more between the mesocosms with different calcification rates than between the mesocosms exposed to different CO₂ levels. In both cases (LC and HC), most inorganic carbon and nitrogen (DIC and DIN) are utilised by non-calcifiers ($\approx 56\%$ in case of HC and $\approx 64\%$ in the LC solution), despite the differences between LC and HC. Generally, more carbon fixation (with C:N uptake ratio of $168:10 \approx 17$) occurs in the HC than in the LC mesocosms (C:N uptake ratio ≈ 13). Flux budgets show that non-calcifiers clearly dominate in mesocosms with low calcification rates, and in HC mesocosms coccolithophores and bulk phytoplankton biomasses are comparable (Figs. 9 and 10). Although grazing, in general, is high in HC mesocosms (Table 4), there is a trend of higher grazing pressure on bulk phytoplankton than on coccolithophores. This is shown by N flux estimates, where zooplankton gain nearly 57 % of their total biomass through grazing on non-calcifiers in HC and LC. According to our model solutions, the coccolithophores are always less vulnerable to grazing than the bulk phytoplankton. This model behaviour may not be fully conclusive, because we have no information about the actual grazing rates or about grazing preferences. A noticeable difference between high and low calcification model ensembles is in terms of mortality of zooplankton. Higher mortality is seen in HC solutions. Since the carbon fixation

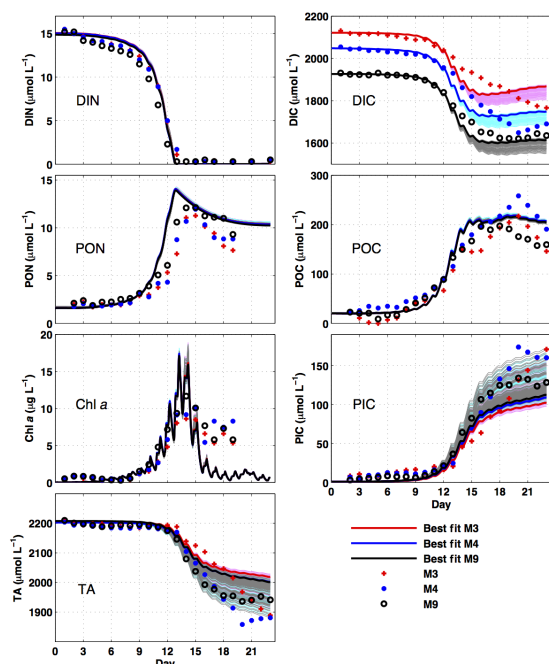


Figure 8. High calcification solution. The coloured bands represent ensemble of model results according to the posterior and symbols show observations.

in HC is high, exudation and leakage rates are also higher. Accordingly, TEPC production is enhanced in HC solutions. Unlike estimates of C flux, the N fluxes in HC and LC ensembles are similar, e.g. aggregation losses of phytoplankton and of coccolithophores are 3 ± 0.4 and 2 ± 0.4 mmol N m⁻³ in HC, and $3.4 \pm 2 \times 10^{-3}$ and $1.5 \pm 2 \times 10^{-3}$ mmol N m⁻³ in LC respectively. Similarly, flux estimates of all mesocosms show almost the same rates of DIN utilisation, excretion, exudation, and remineralisation.

4 Discussion

The data assimilation approach applied in this study was designed to resolve differences in TA and thus in calcification, while variations in other data (e.g. DIN, PON, and POC) should also be explained with our model. We distinguished between mesocosms with high, medium, and low calcification rates (HC, MC, and LC) and their respective data were used to come up with optimal estimates of initial conditions and of some important physiological model parameters. Ideally, we would have identified similar optimal values of the physiological parameters and would have obtained different estimates of the initial conditions for all three cases, HC, MC, and LC. However, our results reflect a more complex

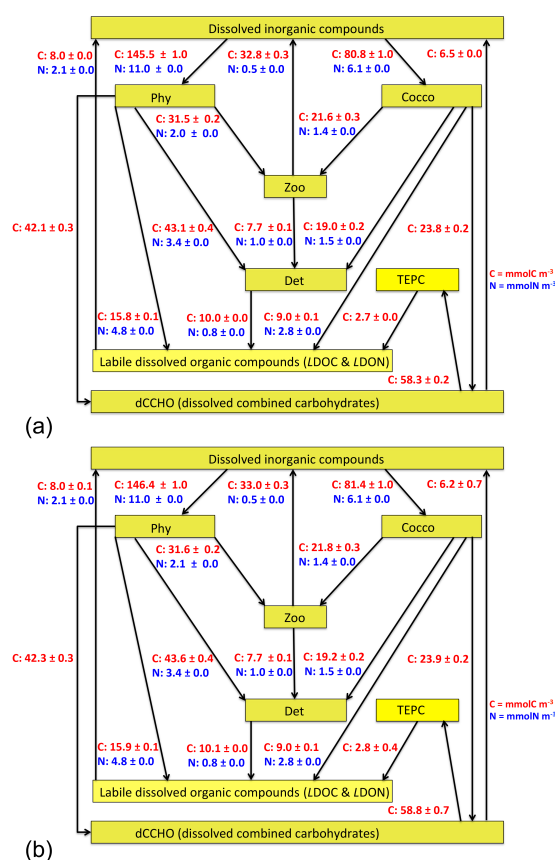


Figure 9. Carbon and nitrogen fluxes estimated by the model in mesocosms with low observed calcification but different CO₂ treatment, high (a) and low (b). All the arrows that point downwards show flux estimates from the respective compartment on the right hand side, whereas arrows pointing upwards show values on the left hand side.

picture and our optimised values for the initial conditions also depend on the best estimates for the model parameters. The initial conditions could not be constrained independently and model solutions of the HC case do not automatically imply a higher initial abundance of coccolithophores relative to the other, non-calcifying, phytoplankton. Likewise, the LC solution does not require a lower initial biomass of calcifying algae. Instead of differences in relative species abundance, the initial physiological conditioning, e.g. acclimation states of the algae, seems relevant as well, which is in the end reflected in the estimates of the physiological parameters Q_0 , α_{cocco} , and α_{phy} . An alternative data assimilation approach would be to optimise the physiological model parameters (Q_0 , α_{cocco} , and α_{phy}) together with the initial conditions (PON_0 , f_{det} , f_{zoo} , and f_{cocco}) for mesocosms of

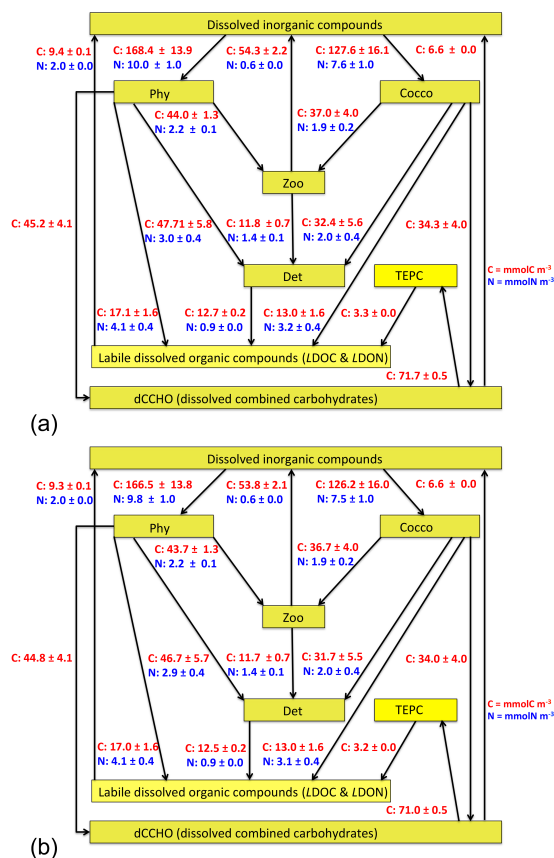


Figure 10. Carbon and nitrogen fluxes estimated by the model in mesocosms with high observed calcification but different CO₂ treatment, high (a) and low (b). All the arrows that point downwards show flux estimates from the respective compartment on the right hand side, whereas arrows pointing upwards show values on the left hand side.

one calcification case in a first step, e.g. the MC case (using data of mesocosms M2, M5, and M7). In a second step we could have fixed the optimised physiological model parameters Q_0 , α_{cocco} , and α_{phy} (as identified with data of, for example, the MC case) and would have then estimated only the initial condition parameters for the other mesocosms, e.g. low and high calcification (LC and HC). This alternative approach does work (not shown), but we learned that we may then put too much confidence into those estimates of Q_0 , α_{cocco} , and α_{phy} obtained first, e.g. estimates for the MC mesocosms. It can even obscure the fact that collinearities exist between some initial condition estimates and the other model parameters. Furthermore, with such an alternative approach we could end up with different estimates of the initial conditions, if we would have started with data of either the

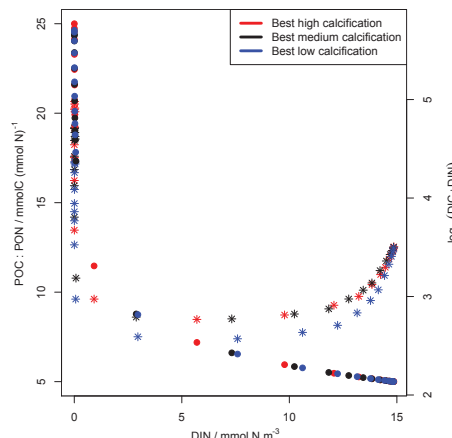


Figure 11. Ratios of [POC]:[PON] and [DIC]:[DIN] determined from daily sampled noon values of model results. Filled circles represent log₁₀([DIC]:[DIN]) ratios. Asterisk symbols represent [POC]:[PON] ratio over the duration of the experiment.

HC or LC mesocosms first instead. The design of our data-assimilation approach is more challenging but it is better suited to disclose major uncertainties and collinearities in estimating initial conditions together with model parameters of algal growth.

4.1 Uncertainty ranges in parameter estimates and variability in model solutions

Large variations can be seen in the data of PIC, reflecting the variability measured in TA. Since optimal ensembles of model solutions were derived for three distinct cases of calcification (LC, MC, and HC), we automatically capture most of the observed variability in PIC with our simulations. The spread of the ensemble solutions for TA and PIC is smaller in each of the three cases relative to the observed total range. This means that the respective uncertainties in our parameter estimates are small enough to obtain three distinctive ensembles of model solutions. However, as discussed before, it is not possible to identify optimal values of the initial condition parameter f_{cocco} independently from estimates of the other physiological model parameters. This situation is aggravating but not unusual (Schartau et al., 2016). For instance, in a sensitivity study with a regional marine ecosystem model, Gibson and Spitz (2011) stressed that collinearities exist between initial conditions and the values assigned to the biological parameters.

The posterior uncertainties in the estimates of the subsistence quota, (Q_0), are rather small, if compared with the uncertainty ranges of the other parameter estimates. Likewise, parameter estimates of the initial condition parameters PON_0 , f_{det} , and f_{zoo} are fairly confined. The variational

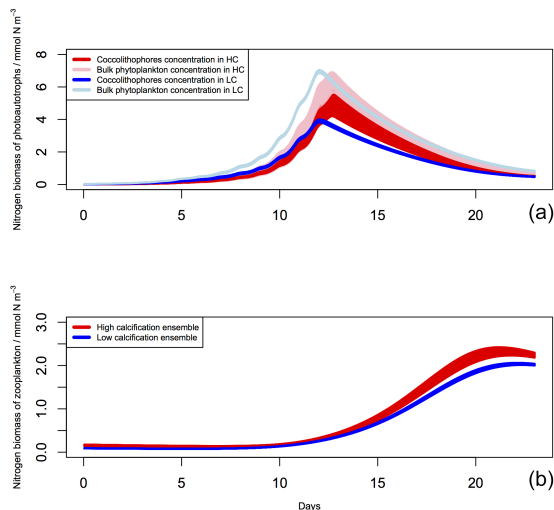


Figure 12. Simulated nitrogen biomass concentrations of photoautotrophs and zooplankton in high and low calcification solutions.

range that we see in our model solutions is mainly induced by uncertainties in estimates of the photosynthesis parameters α_{cocco} and α_{phy} and of f_{cocco} . The combination of these three parameters mainly determine the spread in model solutions with respect to the amount of C-fixation and also calcification. This also explains why the ensemble model solutions exhibit only small variations in DIN and PON concentrations and thus in our N-flux estimates.

Variability in POC is much more pronounced than in PON. All three model solutions show a steep increase in the POC : PON ratio as soon as algal growth becomes nutrient-limited (Fig. 11). The variability seen in the POC : PON ratio is thus mainly due to a temporal variation in Q^{N} (N : C ratio of both photoautotrophs) and thus of the algal growth conditions. The temporal variations in Q^{N} eventually disperse into zooplankton biomass and detritus, inducing elevations of their respective C : N ratios during the post-bloom period. Another contribution to the elevation of POC : PON ratios is also related to changes in POC because it constitutes concentrations of TEPC, which is explicitly resolved in our model.

Our results show an increase in molar $\Delta\text{PIC} : \Delta\text{C}$ -assimilation at low net growth rates (μ_{cocco}) under nutrient-limited conditions (Fig. 6) in both HC and LC cases. These variations are translated into some variability seen in the PIC : POC ratio. Variability in PIC : POC is discussed in Engel et al. (2014), where they collected and analysed data of diverse experiments and documented an increase (up to fourfold) in values of cellular PIC : POC at relative growth rate (RGR) $\approx 0.2 \text{ d}^{-1}$ and below in various CO₂ treatments. The reason for a sharp increase in the molar $\Delta\text{PIC} : \Delta\text{C}$ -assimilation ratio at low growth rates in our model is because

of a down regulation of LHC. Such model behaviour is in agreement with the interpretation of Barcelos e Ramos et al. (2012), who describe calcification as a process into which the coccolithophores can channel excess energy. In order to maximise (optimise) growth rate under nutrient-depleted and high-light conditions, the model allocates more resources and energy to support nutrient acquisition than to the LHC (indicated by low $f_{\text{cocco}}^{\text{LHC}}$ values). Since $\Delta\text{PIC} : \Delta\text{C}$ -assimilation is inversely related to $f_{\text{cocco}}^{\text{LHC}}$ in our model, an increase in calcification (relative to C-fixation) is obtained at low growth rates. The maximum of $\Delta\text{PIC} : \Delta\text{C}$ -assimilation ratio in our simulations are in accordance with those found in Barcelos e Ramos et al. (2010).

4.1.1 Differences between high and low calcification solutions (HC and LC)

The optimised model solutions for HC and LC reveal significant differences in the development of coccolithophore biomass. As discussed before, these differences are not solely attributable to differences in the relative proportions of initial biomass concentrations. In fact, the optimisations yielded estimates that suggest fairly similar initial coccolithophore biomass concentrations between all nine mesocosms. Eggers et al. (2014) stressed that variations in initial plankton composition can be responsible for large differences in the responses observed on community level, thereby masking any possible CO₂ effect on photosynthesis or calcification. Briefly, our results not only support the findings of Eggers et al. (2014), they provide additional insight to the problem of resolving a CO₂ response in the presence of variability in measurements. One added message compared to Eggers et al. (2014) is that our mass flux estimates are shown to differ more between the different calcification solutions than between the different CO₂ treatments. This situation exemplifies that simulation results (e.g. future model projections) may involve uncertainties in flux estimates that are larger than the CO₂ effect introduced to the model (e.g. by following Findlay et al., 2011). Another added message is that initial conditions may not be independently estimated from estimates of phytoplankton growth parameters, like α_{phy} and α_{cocco} . This is particularly relevant for model assessment and model analyses of mesocosm experiments. We stress that the original design of the experiment was meaningful, in particular with respect to the initial filling of the mesocosms in the PeECE-1 experiment. The retrospective separation of the CO₂ response signal from the system's variability was only possible because mesocosms with similar initial conditions were subject to different CO₂ concentrations. Such separation would be more difficult, in retrospect, if mesocosms with similar initial conditions would have been (by chance) exposed to similar CO₂ levels.

From a modelling perspective it is helpful to know about the initial individual mass contributions to PON_0 , including details in the initial composition of the plankton. But the level

of compositional detail remains unclear, since these variations in individual plankton composition will in the end always translate into some variational (uncertainty) range in, for example, the initial photo-acclimation state, since our model approach only distinguishes between calcifiers and all other, non-calcifying, phytoplankton. These considerations were disregarded when we designed this study and we originally thought of the importance of the relative mass distributions between the state variables resolved by our model, while imposing fixed initial stoichiometric ratios (C : N and Chl *a* : N). It seems plausible to allow for some variations of the initial stoichiometric ratios as well.

For now we are interested in the question: what induces the different model solutions for LC and HC, in spite of similar initial conditions in the concentrations of coccolithophores and phytoplankton? First of all, we have some differences between the relative proportions of initial detrital, zooplankton, and photoautotrophic biomass (e.g. DetN : ZooN : (PhyN + CoccoN) = 80 : 10 : 1 for HC and 28 : 3 : 1 for LC). The difference between these ratios point towards net photoautotrophic growth rates that are higher in the LC case than in the HC case, since losses due to grazing and aggregation must be lower in the LC case. However, the initial conditions in mesocosms of the LC case do not automatically yield model solutions of the highest photoautotrophic growth. Instead we find overall reduced growth rates but some pronounced differences in the relative proportions of biomass between the coccolithophores and the non-calcifying phytoplankton (Fig. 12). The reason for these differences lies primarily in the relative differences between the estimates of the physiological parameters, with estimates of α_{cocco} always being smaller than of α_{phy} . The photosynthetic efficiency of the coccolithophores remains clearly smaller (LC case) or can become similar (HC case) relative to the other, non-calcifying, phytoplankton. Major differences between the LC and HC solutions can thus be attributed to higher α_{cocco} values (median $\alpha_{\text{cocco}} = 1.7 \text{ mol C (g Chl } a)^{-1} \text{ m}^2 \text{ W}^{-1} \text{ d}^{-1}$) in HC posterior distribution compared to LC (median $\alpha_{\text{cocco}} = 1.4 \text{ mol C (g Chl } a)^{-1} \text{ m}^2 \text{ W}^{-1} \text{ d}^{-1}$). The estimates of α_{cocco} are negatively correlated with the estimates of f_{cocco} (Table 4) and we may therefore look on the combination of the two parameters. To do so we compare two extreme solutions, selected from the ensemble solutions of LC and HC respectively. One extreme solution yields the lowest calcification among all HC solutions, based on the parameter combination ($\alpha_{\text{cocco}} = 1.84 \text{ mol C (g Chl } a)^{-1} \text{ m}^2 \text{ W}^{-1} \text{ d}^{-1}$, $f_{\text{cocco}} = 0.34$). The other selected solution represents the highest calcification of all LC solutions, which corresponds with ($\alpha_{\text{cocco}} = 1.59 \text{ mol C (g Chl } a)^{-1} \text{ m}^2 \text{ W}^{-1} \text{ d}^{-1}$, $f_{\text{cocco}} = 0.35$). Thus, it is mainly the photosynthetic efficiency α_{cocco} to which the model solution is highly sensitive. Hence, a difference of $\approx 0.3 \text{ mol C (g Chl } a)^{-1} \text{ m}^2 \text{ W}^{-1} \text{ d}^{-1}$ can effectively determine the differences in our simulations with respect to rates of carbon fixation and calci-

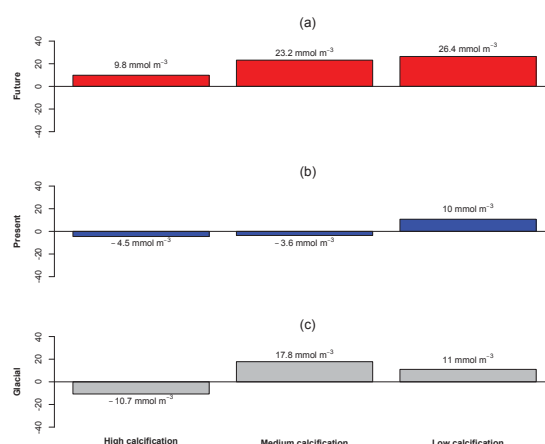


Figure 13. Bar plots depicting cumulative sum of PIC residual (model–data misfit) from day 13 to day 18 of the experiment for three replicates in mean solution of HC, MC, and LC ensembles. First row shows mesocosms with high CO₂ treatment (future), second row medium CO₂ treatment (present), and third row low CO₂ treatment (glacial).

fication. The build-up of comparable nitrogen biomass of coccolithophores and bulk phytoplankton in HC solutions are achieved with identical Q_0 values and only nuanced differences in values between α_{cocco} and α_{phy} . In contrast, bulk phytoplankton (non-calcifiers) out-compete coccolithophores during the bloom period in the LC solutions (Fig. 12).

Differences in photosynthetic efficiency estimates for the LC and HC cases could possibly be invoked for two reasons: (a) because of unresolved differences in initial photo-acclimation states (e.g. different light history during the filling period), since we assume identical initial Chl : N ($\theta_{\text{cocco}}^{\text{N}} = \theta_{\text{phy}}^{\text{N}}$) and N : C ($Q_{\text{cocco}} = Q_{\text{phy}}$) ratios for all nine mesocosms (and thus for LC, MC, and HC), or (b) because of unresolved varying conditions in irradiance. To impose identical surface PAR forcing on all nine mesocosms might not be appropriate, and the arrangement of neighbouring mesocosms may have caused some shading effects. From the available data and with our model approach it is not possible to resolve such varying conditions afterwards.

4.2 Model biases

Model biases disclose systematic deviations of simulation results from observations, which may point towards (i) erroneous model counterparts to observations (definition of $H(x)$ in Eq. 9) or (ii) deficiencies in model dynamics (errors in x). Some bias is related to the increase in PON concentration during the late phase of exponential growth (between days 10 and 12, Fig. 12). The noticeable bias (temporal offset) in simulated PON concentrations can be explained with an

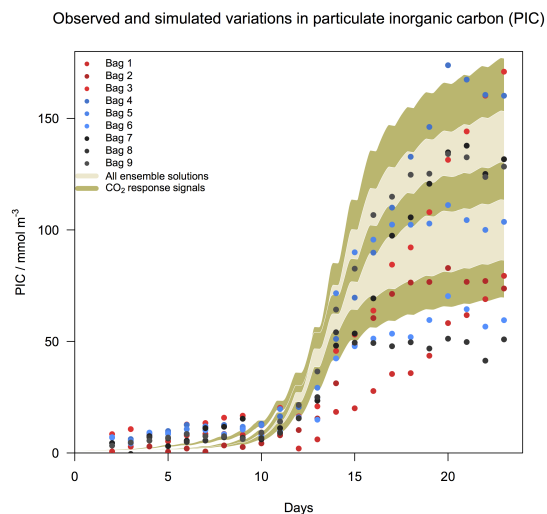


Figure 14. Full spread of model solutions according to credible range in parameter estimates, including ensemble solutions of high, medium, and low calcification (light brown shaded area). Symbols represent observations of all mesocosms. Khaki shaded bands show CO₂ effect in the model, for solutions with lowest, medium, and highest calcification rates.

apparent overestimation of initial coccolithophore biomass. The estimates of f_{cocco} turned out to be highest, if compared with the estimates for the low and high calcification (LC and HC) model solutions. Furthermore, the range of credible values for f_{cocco} is small (Fig. 4). Both estimates of f_{cocco} and of PON_0 lead to an initial biomass concentration of coccolithophores that is approximately three times higher than in the LC case and even six times the initial concentration of the model solutions for HC.

With our model we do not distinguish between growth of picoplankton and the other non-calcifying phytoplankton during the initial bloom phase. The initial abundance of picoplankton (mainly *Micromonas* spp.) and their decline was observed during the early pre-bloom period of the PeECE-I experiment (Engel et al., 2005). This explains why our simulated Chl *a* and PON concentrations are lower compared to observations between day 1 and day 4. Another discrepancy between simulated and observed Chl *a* exists during the post-bloom period. We assume that this bias is mainly because we do not account for detrital chlorophyll pigments (presumably of inactive or destroyed cells) in our model. Formation of detritus is associated with the aggregation of coccolithophores and of the other phytoplankton to form detritus (simulated as a transfer of algal biomass into detritus) in our model, and the fate of Chl *a* within the detritus compartment remains unresolved. Once N and C biomass of the photoautotrophs are transformed to detritus, an associated flux of Chl *a* is disregarded. An explicit consideration of the fate of Chl *a* would

likely improve model performance and some refinements in this respect are recommended for the future.

Results of our data–model synthesis also exhibit a small but distinctive bias in the calcification response to elevated CO₂ levels. The distinctions we made with respect to mesocosms of LC, MC, and HC helped us to identify such bias. This bias implies that the CO₂ effect on calcification, as introduced to our model, is slightly smaller than in the observations, which will be discussed in detail hereafter.

4.3 Disentangling CO₂ effect from the observed variability in PIC

We considered a simple CO₂ relationship that mimics only OA effects on calcification. It is a dependency that was adopted from the meta-analysis of Findlay et al. (2011). With this CO₂ dependence we can already capture differences in PIC formation. The CO₂ sensitivity that we introduced to our model is only effective with respect to the ratio of calcification versus C-fixation, thereby reducing the overall calcification rate under high-CO₂ conditions. This effect turned out to be small compared to the total variability seen in PIC data. According to our model setup we do not consider any potential changes in vulnerability to predation (or edibility) of the coccolithophores due to elevated CO₂. Likewise, any additional CO₂ effects, e.g. on the rate of aggregation, are not accounted for. Such effects remain unresolved, and therefore the comparison of our budget calculations yield only small differences between high and low CO₂ levels, in particular with respect to nitrogen flux estimates. Thus, differences in C and N budgets between the two extreme calcification cases, LC and HC, are more pronounced than between different levels of CO₂. To resolve consecutive ecological effects in response to a reduction of the relative calcification rate we would have needed explicit data, i.e. revealing differences in grazing and aggregation rates between the individual mesocosms. With the PON and POC data used in our data assimilation approach it is not possible to distinguish between different coccolithophore loss terms like grazing and aggregation, since detritus and zooplankton are both constituents of the same PON and POC measurements.

The advantage of resolving LC, MC, and HC solutions separately is that for each case we can compare data with model results of mesocosms individually, of low- (glacial), medium- (present), and high- (future) CO₂ treatments. In other words, for every LC, MC, and HC case we resolve three mesocosms, of which each was subject to different CO₂ levels. This way we have separated differences between LC, MC, and HC from variations induced by a CO₂ effect. Doing so reveals PIC formation to be systematically overestimated by the model for all mesocosms of the future treatment (Figs. 7 and 8, MC case not shown). In contrast to Delille et al. (2005), our results show an early onset of calcification in mesocosms of the high-CO₂ treatment between day 10 and day 15. It indicates that the CO₂ effect introduced to our

model is likely too weak. This becomes evident according to positive model–data residuals in PIC between day 13 and day 18 for those mesocosms with future treatment (Fig. 13). It is not evident for the glacial and present-day CO₂ treatments, where the corresponding residuals do not show a systematic positive offset.

Figure (14) shows the total variability seen in PIC data together with the full variational range of all ensemble model solutions. In addition, we depict those ranges in simulated PIC that are solely due to the CO₂ effect, based on the two extreme calcification solutions (lowest and highest simulated PIC) and the best model solution (according to the lowest cost function values) for the MC mesocosms. If we compare the simulated CO₂ response signal on calcification with the total variability in PIC (in Fig. 14), we find that the CO₂ effect remains small. This situation demonstrates the difficulty in isolating a distinctive CO₂ signal from the total variability seen in PIC observations. However, with our model-based analysis approach this CO₂ signal becomes detectable.

5 Conclusions

An analysis of data of a mesocosm experiment is often approached by first grouping individual mesocosms according to the level of perturbation (e.g. the level of DIC added). In some cases, such an apparently self-evident approach may not help to reveal some basic phenomenon in mesocosm experiments. For a meaningful data analysis the mesocosms need not be exclusively differentiated by the different levels of perturbation but may first be sorted by major differences between relevant response signals, as done with respect to the magnitude of calcification in our study (by differentiating between LC, MC, and HC). In mesocosm experiments these differences in responses are likely associated with variations in initial conditions.

With our data assimilation approach we could disentangle three distinctive ensembles of model solutions that represent mesocosms with high, medium, and low calcification rates. The results of our data–model synthesis show that the initial relative abundance of coccolithophores and the prevailing physiological acclimation states drive the bloom development and determine the amount of calcification in the mesocosms. Small variations of these two initial factors between the mesocosms can generate differences in calcification that are larger than the change in calcification induced by OA. In spite of this difficulty, a CO₂ response signal may still be identifiable, as long as mesocosms that reveal the strongest similarities (with respect to initial composition of plankton and their physiological state) are not used as replicates for similar CO₂ conditions (perturbations). Instead, mesocosms with similar initial conditions should be exposed to different levels of OA. Such favourable starting conditions were met in the mesocosm experiment described in Engel et al. (2005)

and Delille et al. (2005), as well as in the experiment of Eggers et al. (2014).

An alternative approach to setting up mesocosms is to gradually increase the level of perturbation for a series of mesocosms. This way a gradient of different perturbation levels is introduced. The advantage then is that mesocosms that have been collated according to, for example, the lowest and highest response signals (or likewise according to similarities in initial conditions) may then be separately analysed with respect to their responses to the individual levels of perturbation.

From this modelling study we infer that collinearities exist between estimates of initial conditions and physiological model parameters, in particular for the photosynthetic efficiencies α_{phy} and α_{cocco} and the initial fraction of coccolithophores determined by f_{cocco} . Therefore, it is not possible to identify initial concentration of photoautotrophs independently of parameters responsible for phytoplankton growth in HC, MC, and LC model solutions. This was only found because we optimised initial conditions together with physiological parameters for HC, MC, and LC mesocosms separately. By this separation we could better specify the CO₂ effect on PIC formation. In doing so, we could identify a systematic overestimation of calcification in our model and we conclude that our simulated CO₂ effect on PIC formation is even too weak.

Data availability. The results presented are made available by the authors. The model output data are centrally stored. Please send requests to skrishna@geomar.de or to mschartau@geomar.de.

Appendix A: Supplementary model equations

A1 Arrhenius relation

The effect of temperature on the metabolic rates and biological activities of the vast majority of organisms is given by the Arrhenius relationship (Sibly et al., 2012).

$$T_f = \exp[-A_E \cdot (\frac{1}{T} - \frac{1}{T_{\text{ref}}})], \quad (\text{A1})$$

where T_{ref} is reference temperature, given in units Kelvin (K) and approximately equal to 293.15 K (Table A1).

A2 Photoautotrophs

The resource allocation depends on the cellular nitrogen-to-carbon (N:C) ratio, expressed by the cell quota (Q^N). Q^N is the cellular N biomass normalised to carbon and energy units. The availability of resources that can be allocated is estimated by the relative difference between Q^N and a subsistence quota (Q_0). Q_0 is the minimum N:C ratio required for a photoautotrophic cell to survive. As Q^N approaches Q_0 fewer resources can be allocated (e.g. to the light-harvesting complex, (LHC)) and algal growth becomes limited. Under balanced optimal conditions we can approximate $f_V \approx f_V^0$ for photoautotrophs. An optimal allocation of nutrients to specific cellular sites (or cell compartments) is thus determined by a trade-off between three fractions: (a) a fraction that is allocated to the nutrient acquisition complex (f_V), (b) a fraction attached to structural proteins (expressed as Q_s/Q^N), and (c) a remaining fraction ($1 - f_V - Q_s/Q$) that can be allocated to the LHC and thus promotes the synthesis of chlorophyll *a* (Pahlow et al., 2013). An optimal allocation factor (f_V^0) for nutrient uptake is derived by maximising net growth rate with respect to nutrient uptake and thus f_V (Eq. A3 in Appendix A). Under nutrient-depleted conditions, some higher growth rate of an algal cell can be maintained by increasing f_V^0 to the cost of resources that can be assigned to the light-harvesting complex (referred to as f_{LHC}^0 ; the optimal allocation factor for LHC). In consequence, the mobilisation of resources (N in this study) for nutrient acquisition (induced by an increase of f_V^0) reduces the rate of chlorophyll *a* synthesis; vice versa for light-limited conditions. Growth rate of a cell is optimised by investing more resources to the LHC of a cell, which enhances the rate of chlorophyll *a* synthesis. This is achieved for low values of f_V^0 .

In the model, the optimal allocation factor for the LHC in an algal cell is calculated from f_V^0 and Q_0 :

$$Q_s = \frac{Q_0}{2}, \quad (\text{A2})$$

$$f_{V_{\text{phy/cocco}}}^0 = \frac{Q_s}{Q_{\text{cocco/phy}}^N} - \zeta^N \cdot (Q_{\text{cocco/phy}}^N - Q_0), \quad (\text{A3})$$

$$f_{\text{LHC}_{\text{cocco/phy}}}^0 = 1 - \frac{Q_s}{Q_{\text{cocco/phy}}^N} - f_{V_{\text{cocco/phy}}}^0, \quad (\text{A4})$$

where ζ^N is the cost of N uptake in a photoautotrophic cell (mol mol^{-1}) and Q_s is the N quota attached with structural proteins ($\text{mol N (mol C)}^{-1}$). In our model, maximum N assimilation rate and maximum carbon fixation rates are numerically identical.

$$V_{\text{max}}^N = V_0^N \cdot T_f, \quad (\text{A5})$$

$$V_{\text{max}}^C = V_0^C \cdot T_f, \quad (\text{A6})$$

where V_{max}^N and V_{max}^C are maximum N assimilation and maximum carbon fixation rates ($\text{mol N (mol C)}^{-1} \text{d}^{-1}$ and $\text{mol C (mol C)}^{-1} \text{d}^{-1}$). Model parameters V_0^N and V_0^C are photoautotrophic potential N assimilation and C fixation rates ($\text{mol N (mol C)}^{-1} \text{d}^{-1}$ and $\text{mol C (mol C)}^{-1} \text{d}^{-1}$) (Table A1).

The total N uptake rate of photoautotrophs is calculated from the local N uptake rate (Pahlow et al., 2013). The latter is calculated from maximum N assimilation rate, potential nutrient affinity and DIN concentration.

$$\hat{V}^N = \left(\sqrt{\frac{1}{V_{\text{max}}^N}} + \sqrt{\frac{1}{A_0 \cdot (\text{DIN})}} \right)^{-2}, \quad (\text{A7})$$

$$V_{\text{phy/cocco}}^N = f_{V_{\text{cocco/phy}}}^0 \cdot \hat{V}^N, \quad (\text{A8})$$

where \hat{V}^N is the local N uptake of photoautotrophs ($\text{mol N (mol C)}^{-1} \text{d}^{-1}$). A_0 is potential nutrient affinity of respective algae ($\text{m}^3 (\text{mol C})^{-1} \text{d}^{-1}$) (Table A1).

The gross carbon fixation rate of calcifiers and non-calcifiers is calculated from day length, degree of light saturation, f_{LHC}^0 , and V_{max}^C :

$$V_{\text{cocco/phy}}^C = L_d \cdot f_{\text{LHC}_{\text{cocco/phy}}}^0 \cdot V_{\text{max}}^C \cdot S_I^{\text{cocco/phy}}, \quad (\text{A9})$$

where $V_{\text{cocco/phy}}^C$ is the gross carbon-fixation by photoautotrophs ($\text{mol C (mol C)}^{-1} \text{d}^{-1}$), L_d is the day length as a fraction of 24 h. For more details see Table A1. $S_I^{\text{phy/cocco}}$ is the degree of light saturation in photoautotrophs and calculated as follows:

$$S_I^{\text{cocco/phy}} = 1 - \exp\left(-\frac{\alpha \cdot \hat{\theta}_{\text{cocco/phy}} \cdot I}{V_0^C}\right), \quad (\text{A10})$$

where $\hat{\theta}_{\text{cocco/phy}}$ is the Chl:C ratio in the chloroplast of a cell (Pahlow and Oschlies, 2009; Pahlow et al., 2013) ($\text{mg Chl (mmol C)}^{-1}$).

The differential equations of C and N biomass for phytoplankton and coccolithophores are as follows.

$$\begin{aligned} \frac{d}{dt} \text{PhyC} &= (\mu_{\text{phy}} - \text{CN}_{\text{fact}} \cdot \gamma_N) \\ &\cdot \text{PhyC} - \frac{A_{\text{phy}}}{Q_{\text{phy}}^N} - \frac{G_{\text{phy}}}{Q_{\text{phy}}^N} \end{aligned} \quad (\text{A11})$$

$$\frac{d}{dt} \text{CoccoC} = (\mu_{\text{cocco}} - \text{CN}_{\text{fact}} \cdot \gamma_{\text{N}}) \cdot \text{CoccoC} - \frac{A_{\text{cocco}}}{Q_{\text{cocco}}^{\text{N}}} - \frac{G_{\text{cocco}}}{Q_{\text{cocco}}^{\text{N}}} \quad (\text{A12})$$

$$\frac{d}{dt} \text{PhyN} = V_{\text{phy}}^{\text{N}} \cdot \text{PhyC} - \gamma_{\text{N}} \cdot \text{PhyN} - A_{\text{phy}} - G_{\text{phy}} \quad (\text{A13})$$

$$\frac{d}{dt} \text{CoccoN} = V_{\text{cocco}}^{\text{N}} \cdot \text{CoccoC} - \gamma_{\text{N}} \cdot \text{CoccoN} - A_{\text{cocco}} - G_{\text{cocco}} \quad (\text{A14})$$

A description of auxiliary variables is given in Table A1. We stress that the parameterisations in Eqs. (A11) and (A12) are identical for both photoautotrophic groups (coccolithophores and non-calcifying algae), but some of the corresponding optimised parameter values may turn out to be different between the two.

The differential equations for chlorophyll *a* of non-calcifying phytoplankton (with subscripts phy) and coccolithophores (cocco) are as follows:

$$\frac{d}{dt} \text{Chl}_{\text{cocco/phy}} = \left(\mu_{\text{cocco/phy}} + \frac{\dot{\theta}_{\text{cocco/phy}}}{\theta_{\text{cocco/phy}}} \right) \cdot \text{Chl}_{\text{cocco/phy}} - A_{\text{cocco/phy}} \theta_{\text{cocco/phy}}^{\text{N}} - G_{\text{cocco/phy}} \theta_{\text{cocco/phy}}^{\text{N}}, \quad (\text{A15})$$

where $\theta_{\text{cocco/phy}}^{\text{N}}$ are the respective cellular Chl:N ratios (mg Chl (mmol N)⁻¹) (Table A1). The terms $\dot{\theta}_{\text{cocco/phy}}$ are the time derivatives of $\theta_{\text{cocco/phy}}$. The regulation of θ_{phy} and θ_{cocco} on the build-up and limitation of chlorophyll *a* is determined by optimality-based criteria.

The regulation term for chlorophyll *a* synthesis (S_{chl}) is given as follows:

$$S_{\text{chl}} = \frac{\dot{\theta}_{\text{cocco/phy}}}{\theta_{\text{cocco/phy}}} = \left(\frac{1}{\zeta^{\text{Chl}}} \cdot \frac{\partial A_{\text{cocco/phy}}}{\partial \hat{\theta}_{\text{cocco/phy}}} \right) + \dot{Q}_{\text{cocco/phy}}^{\text{N}} \cdot \frac{\hat{\theta}_{\text{cocco/phy}}}{\theta_{\text{cocco/phy}}} \cdot \left(\frac{2 \cdot Q_s}{Q_{\text{cocco/phy}}^{\text{N}} \cdot Q_{\text{cocco/phy}}^{\text{N}}} + \zeta^{\text{N}} \right), \quad (\text{A16})$$

$$\frac{\partial A_{\text{cocco/phy}}}{\partial \hat{\theta}_{\text{cocco/phy}}} = L_d \cdot V_{\text{max}}^{\text{C}} \cdot \left[\frac{\alpha_{\text{cocco/phy}} \cdot I}{V_{\text{max}}^{\text{C}}} \cdot (1 - S_I^{\text{cocco/phy}}) \cdot (1 - \zeta^{\text{Chl}} \cdot \hat{\theta}_{\text{cocco/phy}}) - S_I^{\text{cocco/phy}} \cdot \zeta^{\text{Chl}} \right] - R_{\text{M}}^{\text{Chl}} \cdot \zeta^{\text{Chl}}, \quad (\text{A17})$$

where *A* is an auxiliary variable that contains all light-dependent terms (Pahlow and Oschlies, 2009; Pahlow et al., 2013) (d⁻¹); ζ^{Chl} and ζ^{N} are costs of chlorophyll *a* synthesis and N assimilation (mol C (g Chl)⁻¹ and mol C (mol N)⁻¹) (Table A1). The derivative term ($\frac{\partial A}{\partial \theta}$) is given in units mol C (g Chl)⁻¹ d⁻¹.

A3 Respiration costs

Total respiration cost in a cell includes costs due to chlorophyll synthesis, nutrient acquisition, and cell maintenance.

$$r_{\text{phy/cocco}}^{\text{c}} = R_{\text{phy/cocco}}^{\text{Chl}} + \zeta^{\text{N}} \cdot V_{\text{phy/cocco}}^{\text{N}} + R_{\text{M}}, \quad (\text{A18})$$

where respiration cost due to synthesis of chlorophyll *a* is given as follows:

$$R_{\text{phy/cocco}}^{\text{Chl}} = (V_{\text{phy}}^{\text{C}} + f_{\text{phy/cocco}}^{\text{LHC}} \cdot R_{\text{M}}^{\text{Chl}}) \cdot \zeta^{\text{Chl}} \cdot \hat{\theta}_{\text{phy/cocco}}, \quad (\text{A19})$$

where R_{M} is maintenance respiration cost of a cell (d⁻¹). Detailed description of auxiliary variables is given in the Table A1.

A4 PIC formation and regulation of calcification

PIC formation can be written as a single differential equation:

$$\frac{d}{dt} \text{PIC} = (f_{\text{CO}_2} \cdot f_{\text{PIC}} \cdot \mu_{\text{cocco}}) \cdot \text{CoccoC} - \tau_{\text{dissol}} \cdot \text{PIC}, \quad (\text{A20})$$

where τ_{dissol} is the dissolution rate of PIC (d⁻¹). Parameterisation of the calcite-to-C_{organic} ratio is given by Eq. (A21), whereas the regression model of Findlay et al. (2011) to quantify effect of different CO₂ concentrations on PIC formation is represented by Eq. (A22).

$$f_{\text{PIC}} = \frac{1}{2} + \frac{s_{\text{PIC}}}{1 + \exp(s_{\text{PIC}} \cdot f_{\text{cocco}}^{\text{LHC}})}, \quad (\text{A21})$$

$$f_{\text{CO}_2} = -0.0097 \cdot \text{CO}_{2\text{aq}} + 0.9654, \quad (\text{A22})$$

with aqueous carbon dioxide CO_{2aq} concentrations normalised to water mass instead of volume (μmol kg⁻¹).

A reference rate of PIC formation under nutrient-replete and light-saturated conditions is prescribed as a molar ratio of $f_{\text{PIC}} = 0.5$ mol PIC formed per mol C assimilated into organic matter, Eq. (A21). The molar ratio (f_{PIC}) is assumed to increase when the fraction of resources allocated to the LHC of a cell ($f_{\text{cocco}}^{\text{LHC}}$) decreases. According to our model approach, the process of calcification can be interpreted as an additional pathway for dissipating excess energy (Barcelos e Ramos et al., 2012), as is the case under high-light conditions when chlorophyll *a* synthesis rates diminish (induced by a reduction of $f_{\text{cocco}}^{\text{LHC}}$). On the one hand, PIC formation becomes enhanced under high-light conditions, while fewer resources become allocated to LHC. On the other hand, calcification is reduced or ceases under conditions of low or no light. Under nutrient-depleted conditions, when more resources become allocated to nutrient uptake sites rather than to LHC, the rate of calcification per net carbon fixation also increases. For low (nutrient-limited) growth rates under saturated (or high) light conditions the parameterisation f_{PIC} can yield maxima in the calcite-to-C_{organic} ratio (of the calcifying algae) that may reach values of 2 and slightly above. The function f_{CO_2} in Eq. (A20) has no dimension and it simulates the effect of varying CO₂ concentrations on f_{PIC} .

A5 Zooplankton

The sms differential equations for zooplankton carbon and nitrogen biomass are as follows:

$$\frac{d}{dt}Z_{\text{ooC}} = \frac{G_{\text{phy}}}{Q_{\text{phy}}^{\text{N}}} + \frac{G_{\text{cocco}}}{Q_{\text{cocco}}^{\text{N}}} - r_{\text{zoo}} - \frac{M_{\text{zoo}}}{Q_{\text{zoo}}}, \quad (\text{A23})$$

$$\frac{d}{dt}Z_{\text{ooN}} = G_{\text{phy}} + G_{\text{cocco}} - \gamma_{\text{zoo}}^{\text{N}} - M_{\text{zoo}}. \quad (\text{A24})$$

Equations below represent Holling type 3 grazing dynamics.

$$G_{\text{phy}} = g_{\text{m}} \cdot \frac{(\text{PhyN}^2)}{\epsilon + (\text{PhyN}^2)} \cdot Z_{\text{ooN}}, \quad (\text{A25})$$

$$G_{\text{cocco}} = g_{\text{m}} \cdot \frac{(\text{CoccoN}^2)}{\epsilon + (\text{CoccoN}^2)} \cdot Z_{\text{ooN}}, \quad (\text{A26})$$

where g_{m} is the nitrogen-specific maximum grazing rate on photoautotrophs (d^{-1}) and ϵ is the half saturation constant for grazing ($(\text{mmol N})^2 \text{m}^{-6}$).

A6 Zooplankton respiration and excretion

Respiration is parameterized as a function of respiration maintenance rate coefficient, temperature-dependent metabolic rates, and carbon concentration of heterotroph.

$$r_{\text{zoo}} = R_{\text{basal}} \cdot T_f \cdot Z_{\text{ooC}} \quad (\text{A27})$$

Similarly, excretion is parameterised as a function of respiration maintenance rate to basal metabolism, temperature dependent metabolic rates and nitrogen concentration of heterotroph.

$$\gamma_{\text{zoo}} = R_{\text{basal}} \cdot T_f \cdot Z_{\text{ooN}} \quad (\text{A28})$$

A7 Detritus

The corresponding differential equations of detrital C and N mass are as follows.

$$\frac{d}{dt}\text{DetC} = \frac{A_{\text{phy}}}{Q_{\text{phy}}^{\text{N}}} + \frac{A_{\text{cocco}}}{Q_{\text{cocco}}^{\text{N}}} + \frac{M_{\text{zoo}}}{Q_{\text{zoo}}} - \omega_{\text{det}} \cdot T_f \cdot \text{DetC} \quad (\text{A29})$$

$$\frac{d}{dt}\text{DetN} = A_{\text{phy}} + A_{\text{cocco}} + M_{\text{zoo}} - \omega_{\text{det}} \cdot T_f \cdot \text{DetN} \quad (\text{A30})$$

Aggregation equations for bulk phytoplankton and coccolithophores are given below.

$$A_{\text{phy}} = \phi_{\text{agg}} \cdot \text{PhyN} \cdot \text{DetN} + \phi_{\text{agg}} \cdot \text{PhyN}^2 \quad (\text{A31})$$

$$A_{\text{cocco}} = \phi_{\text{agg}} \cdot \text{CoccoN} \cdot \text{DetN} + \phi_{\text{agg}} \cdot \text{CoccoN}^2 \quad (\text{A32})$$

A8 Dissolved inorganic compounds (DIN, DIC) and total alkalinity (TA)

The nitrogen uptake ($V_{\text{cocco/phy}}^{\text{N}}$) is carbon-specific and is therefore given as a rate of N utilisation per carbon ($\text{mol N} (\text{mol C})^{-1} \text{d}^{-1}$) (Pahlow and Oschlies, 2009):

$$\frac{d}{dt}\text{DIN} = - \left(V_{\text{phy}}^{\text{N}} \cdot \text{PhyC} + V_{\text{cocco}}^{\text{N}} \cdot \text{CoccoC} \right) + \gamma_{\text{zoo}}^{\text{N}} + \rho \cdot T_f \cdot \text{LDON}. \quad (\text{A33})$$

The sources of DIN are calculated from zooplankton excretion ($\gamma_{\text{zoo}}^{\text{N}}$) and the remineralisation of LDON.

The sms differential equation for DIC is given below.

$$\begin{aligned} \frac{d}{dt}\text{DIC} = & -\mu_{\text{phy}} \cdot \text{PhyC} - (1 + f_{\text{CO}_2} \cdot f_{\text{pic}}) \cdot \mu_{\text{cocco}} \\ & \cdot \text{CoccoC} + \tau_{\text{dissol}} \cdot \text{PIC} + r_{\text{zoo}} + \rho \cdot T_f \\ & \cdot (\text{LDOC} + \text{dCCHO}) + F_{\text{DIC}} \end{aligned} \quad (\text{A34})$$

Calculations of air-sea gas exchange (F_{DIC}) within mesocosms are based on original carbonate chemistry code provided by the Ocean Carbon-Cycle Model Intercomparison Project (Orr, 1999). The original code was refined to include an accelerated iteration scheme for pH and $p\text{CO}_2$ calculations (C. Völker, personal communication, 2007), as already applied in Schartau et al. (2007).

The differential equation listed below accounts for TA in the system.

$$\begin{aligned} \frac{d}{dt}\text{TA} = & (1 + 1/16) \cdot \left(\frac{V_{\text{phy}}^{\text{N}}}{Q_{\text{phy}}^{\text{N}}} \cdot \text{PhyN} + \frac{V_{\text{cocco}}^{\text{N}}}{Q_{\text{cocco}}^{\text{N}}} \cdot \text{CoccoN} \right) \\ & - 2 \cdot (f_{\text{CO}_2} \cdot f_{\text{PIC}} \cdot \mu_{\text{cocco}} \cdot \text{CoccoC} - \tau_{\text{dissol}} \cdot \text{PIC}) \\ & - (1 + 1/16) \cdot \rho \cdot T_f \cdot \text{LDON} \end{aligned} \quad (\text{A35})$$

Measured values of DIN, TA, and DIC on day one of the experiment were taken as initial conditions for respective mesocosms.

A9 Dissolved labile organic matter

The differential equations for dissolved organic matter are given below.

$$\begin{aligned} \frac{d}{dt}\text{LDOC} = & C_{\text{fact}} \cdot \gamma_{\text{N}} \cdot \left[(1 - f_{\text{dCCHO}}^{\text{phy}}) \cdot \text{PhyC} + (1 - f_{\text{dCCHO}}^{\text{cocco}}) \right. \\ & \cdot \text{CoccoC} \left. \right] + \omega_{\text{det}} \cdot T_f \cdot \text{DetC} \\ & + \omega_{\text{gel}} \cdot T_f \cdot \text{TEPC} - \rho \cdot T_f \cdot \text{LDOC} \end{aligned} \quad (\text{A36})$$

$$\begin{aligned} \frac{d}{dt}\text{LDON} = & \gamma_{\text{N}} \cdot (\text{PhyN} + \text{CoccoN}) + \omega_{\text{det}} \cdot T_f \\ & \cdot \text{DetN} - \rho \cdot T_f \cdot \text{LDON} \end{aligned} \quad (\text{A37})$$

A10 dCCHO and TEPC

The differential equation for dissolved combined carbohydrates (dCCHO) is given as follows.

$$\frac{d}{dt}\text{dCCHO} = C_{\text{fact}} \cdot \gamma_{\text{N}}$$

Table A1. Auxiliary model variables and model parameters.

Auxiliary variables & functions	Description	Unit
T_f	Arrhenius temperature dependency	–
f_V	resource fraction allocated for nutrient acquisition	–
f_V^0	optimal allocation value of f_V	–
f_{LHC}^0	optimal resource allocation to light-harvesting complex (LHC)	–
μ	net growth rates of respective photoautotrophs	d ⁻¹
Q_s	N quota attached with structural proteins	mol N (mol C) ⁻¹
\hat{V}^N	photoautotrophic local N uptake rate of rate	mol N (mol C) ⁻¹ d ⁻¹
V^C	photoautotrophic gross carbon fixation rates	mol C (mol C) ⁻¹ d ⁻¹
r^C	respiration rates	d ⁻¹
V_{max}^N	photoautotrophic maximum N assimilation rates	mol N (mol C) ⁻¹ d ⁻¹
V_{max}^C	photoautotrophic maximum C fixation rates	mol C (mol C) ⁻¹ d ⁻¹
V^N	carbon-specific nitrogen uptake rate	mol N (mol C) ⁻¹ d ⁻¹
Q^N	molar cellular nitrogen-to-carbon (N : C) ratio (cell quota)	mol N (mol C) ⁻¹
θ	chlorophyll- <i>a</i> -to-carbon (Chl : C) ratio of photoautotrophs	g Chl (mol C) ⁻¹
$\dot{\theta}$	time derivative of θ	g Chl (mol C) ⁻¹ d ⁻¹
θ^N	chlorophyll- <i>a</i> -to-nitrogen (Chl : N) ratio of photoautotrophs	g Chl (mol N) ⁻¹
S_I	degree of light saturation for photosynthesis	–
S_{chl}	regulation term for chlorophyll synthesis	mol C (mol N) ⁻¹
L_d	day length as a fraction of 24 h	–
I	Mean irradiance	W m ⁻² d ⁻¹
$\hat{\theta}$	photoautotrophic chloroplast Chl : C ratio	g Chl (mol C) ⁻¹
A	variable representing all light-dependent terms	d ⁻¹
G	nitrogen-specific rates of zooplankton grazing	mmol N m ⁻³ d ⁻¹
r_{zoo}	zooplankton respiration	mmol C m ⁻³ d ⁻¹
γ_{zoo}^N	zooplankton excretion of nitrogen	mmol N m ⁻³ d ⁻¹
M_{zoo}	nitrogen-specific zooplankton mortality	mmol N m ⁻³ d ⁻¹
A	nitrogen-specific rates of aggregation	mmol N m ⁻³ d ⁻¹
f_{PIC}	calcification relative to net carbon fixation	mol PIC (mol C) ⁻¹
F_{DIC}	flux due to air–sea gas exchange	mmol C m ⁻³ d ⁻¹
f_{CO2}	regression model of CO ₂ effect on calcification	–
f_{dCCHO}	fraction of exudates assigned to dCCHO	–
α_{dCCHO}	stickiness between dCCHO and dCCHO	–
β_{dCCHO}	C-specific collision rates between dCCHO and dCCHO	m ³ (mmol C) ⁻¹ d ⁻¹
α_{TEPC}	stickiness between dCCHO and TEPC	–
β_{TEPC}	C-specific collision rates between dCCHO and TEPC	m ³ (mmol C) ⁻¹ d ⁻¹
Model parameters (fixed)		Value
1. γ_N	photoautotrophic loss rate of organic nitrogen	0.1 d ⁻¹
2. $C N_{fact}$	enhancement factor of carbon exudation relative to γ_N	1.0 –
3. ρ	reminalisation rate of dissolved organic matter	0.05 d ⁻¹
4. ω_{det}	hydrolysis/degradation rate of detritus	0.02 d ⁻¹
5. ω_{gel}	hydrolysis/degradation rate of TEPC	0.01 d ⁻¹
6. τ_{dissol}	dissolution rate of particulate inorganic carbon	0.01 d ⁻¹
7. ϕ_{dCCHO}	coagulation parameter of dCCHO	7.48×10^{-4} m ³ (mmol C) ⁻¹ d ⁻¹
8. ϕ_{TEPC}	coagulation parameter of dCCHO-TEPC	2.56×10^{-2} m ³ (mmol C) ⁻¹ d ⁻¹
9. T_{ref}	reference temperature for A_E relation	293.15 K
10. A_E	slope of Arrhenius relationship	4500 K
11. a_w	light attenuation due to water column	0.04 m ⁻¹
12. a_c	light attenuation due to chlorophyll <i>a</i>	0.05 (mg Chl <i>a</i>) ⁻¹ m ³
13. R_M^{Chl}	cost of chlorophyll maintenance	0.1 d ⁻¹
14. R_M	total respiration maintenance cost	0.05 d ⁻¹
15. ζ^{Chl}	cost of photosynthesis coefficient	0.6 mol C (g Chl <i>a</i>) ⁻¹
16. ζ^N	cost of N uptake	0.7 mol C (mol N) ⁻¹
17. A_0	potential nutrient affinity	1 m ³ mol C ⁻¹ d ⁻¹
18. V_0^N	photoautotrophic potential N assimilation rate	4.0 mol C (mol N) ⁻¹
19. V_0^C	photoautotrophic potential C fixation rate	4.0 mol C (mol C) ⁻¹
20. γ_N	algal nitrogen loss rate	0.1 d ⁻¹
21. ϕ_{agg}	aggregation rate	0.01 m ³ (mmol N) ⁻¹ d ⁻¹
22. p_{dCCHO}	minimum DOC fraction allocated to dCCHO	0.2 –
23. g_m	nitrogen-specific maximum grazing rate	0.2 d ⁻¹
24. ϵ	prey capture rate normalised to maximum grazing rate	1 (mmol N) ² m ⁻⁶
25. M_{zoo}	mortality rate of zooplankton	0.05 d ⁻¹
26. R_{basal}	zooplankton basal respiration rate	0.05 d ⁻¹
27. sp_{PIC}	slope of ΔPIC formed per ΔC assimilated	5.0 mol PIC (mol C) ⁻¹

$$\begin{aligned} & \cdot \left[f_{\text{dCCHO}}^{\text{phy}} \cdot \text{PhyC} + f_{\text{dCCHO}}^{\text{cocco}} \cdot \text{CoccoC} \right] \\ & - \phi_{\text{dCCHO}} \cdot \text{dCCHO}^2 - \phi_{\text{TEP}} \cdot \text{dCCHO} \cdot \text{TEPC} \\ & - \rho \cdot T_f \cdot \text{dCCHO} \end{aligned} \quad (\text{A38})$$

Given below is the parameterisation to estimate the fraction of phytoplankton exudates that become available to be part of dCCHO during two distinct modes of carbon overconsumption described in Schartau et al. (2007).

$$f_{\text{dCCHO}}^{\text{cocco/phy}} = \left[1 + p_{\text{dCCHO}} \cdot \exp(1 - Q_s / Q_{\text{cocco/phy}}^N) \right]^{-1}, \quad (\text{A39})$$

where p_{dCCHO} is the fraction of DOC that enters the dCCHO pool. The coagulation parameter of dCCHO (ϕ_{dCCHO}) is derived from a product of α_{dCCHO} (stickiness between dCCHO and dCCHO) and β_{dCCHO} (C-specific collision rates between dCCHO and dCCHO). Likewise, the coagulation parameter of dCCHO-TEPC (ϕ_{TEP}) is computed from the product of α_{TEP} (stickiness between dCCHO and TEPC) and β_{TEP} (C-specific collision rates between dCCHO and TEPC); α_{dCCHO} and α_{TEP} have no units as they are probabilities, whereas β_{dCCHO} and β_{TEP} have units ($\text{m}^3 (\text{mmol C})^{-1} \text{d}^{-1}$). Values of α_{dCCHO} , α_{TEP} , β_{dCCHO} , and β_{TEP} are taken from (Schartau et al., 2007).

$$\begin{aligned} \phi_{\text{dCCHO}} &= \alpha_{\text{dCCHO}} \cdot \beta_{\text{dCCHO}} \\ \phi_{\text{dCCHO}} &= (0.87 \cdot 10^{-3}) \cdot 0.86 = 7.48 \cdot 10^{-4} \end{aligned} \quad (\text{A40})$$

$$\begin{aligned} \phi_{\text{TEP}} &= \alpha_{\text{TEP}} \cdot \beta_{\text{TEP}} \\ \phi_{\text{TEP}} &= 0.4 \cdot 0.064 = 2.56 \cdot 10^{-2} \end{aligned} \quad (\text{A41})$$

The differential equation for formation of TEPC is shown below.

$$\begin{aligned} \frac{d}{dt} \text{TEPC} &= \phi_{\text{dCCHO}} \cdot \text{dCCHO}^2 + \phi_{\text{TEP}} \cdot \text{dCCHO} \\ & \cdot \text{TEPC} - \omega_{\text{gel}} \cdot T_f \cdot \text{TEPC} \end{aligned} \quad (\text{A42})$$

Appendix B: Data assimilation

B1 Parameter optimisation procedure

The entire optimisation procedure of each (LC, MC, and HC) case is subject to five consecutive analysis steps:

1. adjustment of parameters while considering published typical values \rightarrow specify model solution that is in qualitative (visual) good agreement with observations of the medium calcification (MC) case.
2. application of simulated annealing algorithm (SANN) (see Bélisle (1992)), to effectively scan and minimise the seven-dimensional manifold $(\Theta, J(\Theta))$, while avoiding getting trapped into local minima of $J(\Theta) \rightarrow$ obtain global estimate of Θ .

3. local refinement of the parameter estimate, using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970) \rightarrow identify maximum likelihood estimate that corresponds with the global minimum $(\hat{\Theta}, J(\hat{\Theta}))$.
4. calculation of the inverse of second derivatives of $J(\Theta)$ with respect to every parameter ($\mathcal{H}_{jj} = \partial^2 J / \partial \Theta_j^2$ at $\hat{\Theta}$, which is a point-wise approximation of the diagonal elements of a Hessian matrix \mathcal{H}) \rightarrow derive marginal errors (standard errors, $\sqrt{\mathcal{H}_{jj}^{-1}}$) of the estimated parameter values.
5. application of a Markov Chain Monte Carlo (MCMC) method, using the marginal error information of item 4 above to confine credible range of optimal parameter values \rightarrow derive posterior confidence limits of parameter estimates and collinearities (correlations) between parameter estimates.

For steps 2, 3, and 5 the R package FME is applied, as coded and described by Soetaert and Petzoldt (2010). The plankton ecosystem model was coded and compiled as a shared library in FORTRAN so that we can apply a FORTRAN-R wrapper function. This wrapper allows us to take advantage of fast numerical Euler forward integrations of the model equations while, at the same time, we can benefit from the R platform and its freely available packages. The cost function $J(\Theta)$ is evaluated in R. The MCMC method employed here is based on the adaptive Metropolis–Hastings (AMH) algorithm (Haario et al., 2001), which is also available with the R package FME. The AMH algorithm generates a new parameter vector (Θ^*) by perturbing the original vector Θ , inferred from a “proposal” distribution (Metropolis et al., 1953). The standard deviation information required for generating the initial proposal (Gaussian) distribution in the AMH algorithm is derived from the diagonal elements of Hessian matrix. We approximated the diagonal elements of the Hessian with finite central differences, as described in, for example, Matear (1995), Kidston et al. (2011) and Kreuz and Schartau (2015). To do so we imposed an incremental step size of 1 % variation to the respective parameter values.

B2 Data correlation matrices

Correlations during pre-bloom (t_i ; $i = 1, \dots, 13$) between mesocosms with medium observed calcification in matrix form are given below.

$$C_{(y)} = \begin{pmatrix} & \text{DIC} & \text{DIN} & \text{Chl } a & \text{PON} & \text{POC} & \text{PIC} & \text{TA} \\ \text{DIC} & 1 & 0.57 & -0.95 & -0.77 & -0.95 & -0.89 & 0.88 \\ \text{DIN} & . & 1 & -0.56 & -0.52 & -0.53 & -0.58 & 0.53 \\ \text{Chl } a & . & . & 1 & 0.71 & 0.91 & 0.81 & -0.77 \\ \text{PON} & . & . & . & 1 & 0.87 & 0.77 & -0.65 \\ \text{POC} & . & . & . & . & 1 & 0.83 & -0.77 \\ \text{PIC} & . & . & . & . & . & 1 & -0.95 \\ \text{TA} & . & . & . & . & . & . & 1 \end{pmatrix} \quad (\text{B1})$$

1880

S. Krishna and M. Schartau: Modelling of CO₂ perturbation mesocosm experiment

Correlations during post-bloom period (t_i ; $i = 14, \dots, 22$) are as follows.

$$C_{(y)} = \begin{pmatrix} \text{DIC} & \text{DIC} & \text{DIN} & \text{Chl } a & \text{PON} & \text{POC} & \text{PIC} & \text{TA} \\ \text{DIC} & 1 & 0.22 & 0.27 & 0.29 & -0.83 & -0.93 & 0.94 \\ \text{DIN} & . & 1 & 0.3 & 0.31 & -0.23 & -0.22 & 0.24 \\ \text{Chl } a & . & . & 1 & 0.99 & 0.01 & -0.44 & 0.49 \\ \text{PON} & . & . & . & 1 & -0.02 & -0.45 & 0.50 \\ \text{POC} & . & . & . & . & 1 & 0.65 & -0.64 \\ \text{PIC} & . & . & . & . & . & 1 & -0.99 \\ \text{TA} & . & . & . & . & . & . & 1 \end{pmatrix} \quad (\text{B2})$$

Residual standard errors (σ_i) were calculated based on daily measurements between the mesocosms of similar observed calcification and can be written in matrix notation with off-diagonal elements being zero.

$$S_i = \begin{pmatrix} \sigma_i^{(\text{DIC})} & 0 & \dots & 0 \\ 0 & \sigma_i^{(\text{DIN})} & \dots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \dots & 0 & \sigma_i^{(\text{TA})} \end{pmatrix} \quad (\text{B3})$$

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The development of the modelling framework for mesocosm simulations and data assimilation was supported by the large integrated project Surface Ocean Processes in the Anthropocene (SOPRAN, 03F0662A), funded by the German Federal Ministry of Education and Research (BMBF). This study is a contribution to the BMBF-funded BIOACID (03F0728A) project. We gratefully acknowledge support by Markus Pahlow, who helped to refine equations in our model. We also like to acknowledge support given by Andreas Oschlies and by the GEOMAR data management team. We thank Sabine Mathesius for the compilation and inclusion of the forcing data into the mesocosm modelling setup. We thank Yonss Jose and Hadi Bordbar for helpful and constructive comments.

The article processing charges for this open-access publication were covered by a Research Centre of the Helmholtz Association.

Edited by: J. Middelburg

Reviewed by: two anonymous referees

References

- Arteaga, L., Pahlow, M., and Oschlies, A.: Global patterns of phytoplankton nutrient and light colimitation inferred from an optimality-based model. *Global Biogeochem. Cy.*, 28, 648–661, 2014.
- Barcelos e Ramos, J., Müller, M. N., and Riebesell, U.: Short-term response of the coccolithophore *Emiliania huxleyi* to an abrupt change in seawater carbon dioxide concentrations, *Biogeosciences*, 7, 177–186, doi:10.5194/bg-7-177-2010, 2010.
- Barcelos e Ramos, J., Schulz, K. G., Febiri, S., and Riebesell, U.: Photoacclimation to abrupt changes in light intensity by *Phaeodactylum tricornutum* and *Emiliania huxleyi*: the role of calcification, *Mar. Ecol.-Prog. Ser.*, 452, 11–26, 2012.
- Bélisle, C. J.: Convergence theorems for a class of simulated annealing algorithms on rd, *J. Appl. Probab.*, 885–895, 1992.
- Broyden, C. G.: The convergence of a class of double-rank minimization algorithms 1. general considerations, *IMA J. Appl. Math.*, 6, 76–90, 1970.
- Delille, B., Harlay, J., Zondervan, I., Jacquet, S., Chou, L., Wollast, R., Bellerby, R. G., Frankignoulle, M., Borges, A. V., Riebesell, U., and Gattuso, J. P.: Response of primary production and calcification to changes of $p\text{CO}_2$ during experimental blooms of the coccolithophorid *Emiliania huxleyi*, *Global Biogeochem. Cy.*, 19, GB2023, doi:10.1029/2004GB002318, 2005.
- Eggers, S. L., Lewandowska, A. M., Barcelos e Ramos, J., Blanco-Ameijeiras, S., Gallo, F., and Matthiessen, B.: Community composition has greater impact on the functioning of marine phytoplankton communities than ocean acidification, *Glob. Change Biol.*, 20, 713–723, 2014.
- Engel, A., Cisternas Novoa, C., Wurst, M., Endres, S., Tang, T., Schartau, M., and Lee, C.: No detectable effect of CO₂ on elemental stoichiometry of *emiliania huxleyi* in nutrient-limited, acidified continuous cultures. *Mar. Ecol.-Prog. Ser.*, 507, 15–30, 2014.
- Engel, A., Thoms, S., Riebesell, U., Rochelle-Newall, E., and Zondervan, I.: Polysaccharide aggregation as a potential sink of marine dissolved organic carbon, *Nature*, 428, 929–932, 2004.
- Engel, A., Zondervan, I., Aerts, K., Beaufort, L., Benthien, A., Chou, L., Delille, B., Gattuso, J. P., Harlay, J., Heemann, C., and Hoffmann, L.: Testing the direct effect of CO₂ concentration on a bloom of the coccolithophorid *Emiliania huxleyi* in mesocosm experiments, *Limnol. Oceanogr.*, 50, 493–507, 2005.
- Fasham, M., Ducklow, H., and McKelvie, S.: A nitrogen-based model of plankton dynamics in the oceanic mixed layer, *J. Mar. Res.*, 48, 591–639, 1990.
- Feely, R. A., Sabine, C. L., Lee, K., Berelson, W., Kleypas, J., Fabry, V. J., and Millero, F. J.: Impact of anthropogenic CO₂ on the CaCO₃ system in the oceans, *Science*, 305, 362–366, 2004.
- Fernández-Castro, B., Pahlow, M., Mouriño-Carballido, B., Marañón, E., and Oschlies, A.: Optimality-based *Trichodesmium* diazotrophy in the North Atlantic subtropical gyre, *J. Plankton Res.*, 38, 946–963, 2016.
- Findlay, H. S., Calosi, P., and Crawford, K.: Determinants of the PIC: POC response in the coccolithophore *Emiliania huxleyi* under future ocean acidification scenarios, *Limnol. Oceanogr.*, 56, 1168–1178, 2011.
- Fletcher, R.: A new approach to variable metric algorithms, *Comput. J.*, 13, 317–322, 1970.
- Geider, R., Osborne, B., and Raven, J.: Light dependence of growth and photosynthesis in *Phaeodactylum tricornutum* (*Bacillariophyceae*), *J. Phycol.*, 21, 609–619, 1985.
- Gibson, G. and Spitz, Y.: Impacts of biological parameterization, initial conditions, and environmental forcing on parameter sensitivity and uncertainty in a marine ecosystem model for the Bering sea, *J. Marine Syst.*, 88, 214–231, 2011.
- Goldfarb, D.: A family of variable-metric methods derived by variational means, *Math. Comput.*, 24, 23–26, 1970.
- Haario, H., Saksman, E., and Tamminen, J.: An adaptive metropolis algorithm, *Bernoulli*, 223–242, 2001.
- Harding Jr., L. W., Fisher Jr, T. R., and Tyler, M. A. Adaptive responses of photosynthesis in phytoplankton: specificity to time-scale of change in light, *Biol. Oceanogr.*, 4, 403–437, 1987.
- Harrison, W. and Platt, T.: Photosynthesis-irradiance relationships in polar and temperate phytoplankton populations, *Polar Biol.*, 5, 153–164, 1986.
- Iglesias-Rodriguez, M. D., Halloran, P. R., Rickaby, R. E., Hall, I. R., Colmenero-Hidalgo, E., Gittins, J. R., Green, D. R., Tyrrell, T., Gibbs, S. J., von Dassow, P., and Rehm, E.: Phytoplankton calcification in a high-CO₂ world, *Science*, 320, 336–340, 2008.
- Joassin, P., Delille, B., Soetaert, K., Harlay, J., Borges, A. V., Chou, L., Riebesell, U., Suykens, K., and Grégoire, M.: Carbon and nitrogen flows during a bloom of the coccolithophore *Emiliania huxleyi*: Modelling a mesocosm experiment, *J. Marine Syst.*, 85, 71–85, 2011.
- Kidston, M., Matear, R., and Baird, M. E.: Parameter optimisation of a marine ecosystem model at two contrasting stations in the sub-antarctic zone, *Deep-Sea Res. Pt. II*, 58, 2301–2315, 2011.
- Kreus, M. and Schartau, M.: Variations in the elemental ratio of organic matter in the central baltic sea: Part I—sensitivities of annual mass flux estimates to model parameter variations. *Cont. Shelf Res.*, 100, 46–63, 2015.

- Langer, G., Nehrke, G., Probert, I., Ly, J., and Ziveri, P.: Strain-specific responses of *Emiliania huxleyi* to changing seawater carbonate chemistry, *Biogeosciences*, 6, 2637–2646, doi:10.5194/bg-6-2637-2009, 2009.
- Lewis, M. R. and Smith, J. C.: A small volume, short-incubation-time method for measurement of photosynthesis as a function of incident irradiance, *Mar. Ecol.-Prog. Ser.*, 13, 99–102, 1983.
- Marra, J. and Heinemann, K.: Photosynthesis response by phytoplankton to sunlight variability, *Limnol. Oceanogr.*, 27, 1141–1153, 1982.
- Matear, R. J.: Parameter optimization and analysis of ecosystem models using simulated annealing: a case study at station p, *J. Mar. Res.*, 53, 571–607, 1995.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E.: Equation of state calculations by fast computing machines, *J. Chem. Phys.*, 21, 1087–1092, 1953.
- Orr, J. C.: Ocean Carbon-Cycle Model Intercomparison Project (OCMIP): Phase I (1995–1997), IGBP/GAIM Report Series, 7, 1999.
- Orr, J. C., Fabry, V. J., Aumont, O., Bopp, L., Doney, S. C., Feely, R. A., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., and Key, R. M.: Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms, *Nature*, 437, 681–686, 2005.
- Pahlow, M., Dietze, H., and Oschlies, A.: Optimality-based model of phytoplankton growth and diazotrophy, *Mar. Ecol.-Prog. Ser.*, 489, 1–16, 2013.
- Pahlow, M. and Oschlies, A.: Chain model of phytoplankton P, N and light colimitation, *Mar. Ecol.-Prog. Ser.*, 376, 1–16, 2009.
- Pahlow, M., Vézina, A. F., Casault, B., Maass, H., Malloch, L., Wright, D. G., and Lu, Y.: Adaptive model of plankton dynamics for the north atlantic, *Prog. Oceanogr.*, 76, 151–191, 2008.
- Platt, T., Denman, K., and Jossby, A.: Modelling the productivity of phytoplankton, *Sea*, 6, 807–856, 1977.
- Riebesell, U., Bellerby, R. G. J., Grossart, H.-P., and Thingstad, F.: Mesocosm CO₂ perturbation studies: from organism to community level, *Biogeosciences*, 5, 1157–1164, doi:10.5194/bg-5-1157-2008, 2008.
- Ruiz, J., Prieto, L., and Ortigón, F.: Diatom aggregate formation and fluxes: a modeling analysis under different size-resolution schemes and with empirically determined aggregation kernels, *Deep-Sea Res. Pt. I*, 49, 495–515, 2002.
- Schartau, M., Engel, A., Schröter, J., Thoms, S., Völker, C., and Wolf-Gladrow, D.: Modelling carbon overconsumption and the formation of extracellular particulate organic carbon, *Biogeosciences*, 4, 433–454, doi:10.5194/bg-4-433-2007, 2007.
- Schartau, M., Wallhead, P., Hemmings, J., Löptien, U., Kriest, I., Krishna, S., Ward, B. A., Slawig, T., and Oschlies, A.: Reviews and syntheses: Parameter identification in marine planktonic ecosystem modelling, *Biogeosciences*, 14, 1647–1701, doi:10.5194/bg-14-1647-2017, 2017.
- Shanno, D. F.: Conditioning of quasi-Newton methods for function minimization, *Math. Comput.*, 24, 647–656, 1970.
- Sibly, R. M., Brown, J. H., and Kodric-Brown, A.: *Metabolic ecology: a scaling approach*, John Wiley & Sons, 2012.
- Skartveit, A., Cleveland, F., and de Lange, T.: *Radiation Yearbook no. 37*, Meteorological Report Series, University of Bergen, Technical report, University of Bergen Geophysical Institute, Bergen, Norway, 2001.
- Smith, S. L., Pahlow, M., Merico, A., and Wirtz, K. W.: Optimality-based modeling of planktonic organisms, *Limnol. Oceanogr.*, 56, 2080–2094, 2011.
- Soetaert, K. and Petzoldt, T.: Inverse modelling, sensitivity and monte carlo analysis in r using package fme, *J. Stat. Softw.*, 33, 1–28, 2010.
- Verdugo, P., Alldredge, A. L., Azam, F., Kirchman, D. L., Passow, U., and Santschi, P. H.: The oceanic gel phase: a bridge in the DOM–POM continuum, *Mar. Chem.*, 92, 67–85, 2004.
- Wolf-Gladrow, D. A., Zeebe, R. E., Klaas, C., Körtzinger, A., and Dickson, A. G.: Total alkalinity: The explicit conservative expression and its application to biogeochemical processes, *Mar. Chem.*, 106, 287–300, 2007.
- Zondervan, I., Rost, B., and Riebesell, U.: Effect of CO₂ concentration on the PIC/POC ratio in the coccolithophore *Emiliania huxleyi* grown under light-limiting conditions and different daylengths, *J. Exp. Mar. Biol. Ecol.*, 272, 55–70, 2002.

Chapter 4

Comparison of two carbon-nitrogen regulatory models calibrated with mesocosms data.

Shubham Krishna, Markus Pahlow, Markus Schartau
Manuscript shall be submitted soon.

Comparison of two carbon-nitrogen regulatory models calibrated with mesocosms data

Shubham Krishna*, Markus Pahlow, Markus Schartau

GEOMAR Helmholtz Centre for Ocean Research Kiel, Düsternbrooker Weg 20, 24105 Kiel, Germany

Abstract

Marine phytoplankton can regulate their stoichiometric composition in response to variations in the availability of nutrients, light and the pH of seawater. Varying elemental composition of photoautotrophs affects several important ecological and biogeochemical processes, e.g., primary and export production, nutrient cycling, calcification, and grazing. Here we compare two plankton ecosystem models that consider regulatory mechanisms of cellular carbon and nitrogen driving the physiological acclimation of photoautotrophs. The Carbon:Nitrogen Regulated Ecosystem Model (CN-REcoM) and the Optimality-Based Model (OBM) differ in their representation of phytoplankton dynamics, i.e. nutrient acquisition, synthesis of chlorophyll *a*, and growth. All other model compartments (zooplankton, detritus, dissolved inorganic and organic matter) and processes (grazing, aggregation, remineralisation) are identical in CN-REcoM and OBM.

We assess the skills of the two models against data from an ocean acidification mesocosm experiment with three CO₂ treatments. Neither model accounts for any carbon dioxide (CO₂) effects explicitly. Instead, we assimilate data of the different CO₂ treatments separately. For the OBM, optimal parameter estimates of Q_{min} (subsistence N:C ratio) and V_C^0 (maximum potential photosynthesis rate of photoautotrophs) are higher for mesocosms exposed to high CO₂ compared to those with low CO₂ concentration. A possible physiological interpretation of higher estimates of Q_{min} and V_C^0 is that phytoplankton may experience environmental stress under more acidic conditions, and hence must invest more energy/resources for maintaining basic cellular functions. By contrast, a similar relationship is not observed for the CN-REcoM. Our data assimilation reveals that the parameters of the OBM are better constrained by the data than those of CN-REcoM. Furthermore, the OBM is more robust than CN-REcoM in reproducing data that were not used for parameter optimization.

Keywords: Models, Acclimation, mesocosms, ocean acidification, photoacclimation, optimality-based

1. Introduction

Several studies in the recent past have shown significant deviations in cellular C:N:P ratios of phytoplankton from the classical Redfield stoichiometry (C:N:P =

106:16:1) in the surface oceans (Burkhardt et al., 1999; Körtzinger et al., 2001; Geider and La Roche, 2002; Flynn, 2003; Martiny et al., 2013). Furthermore, it has been well documented that marine phytoplankton have the ability to change their stoichiometric composition, typically represented as ratios (e.g., C:N:P — carbon-to-nitrogen-to-phosphorus), in response to changes in nutrient and light availability, and the pH of seawater (Geider, 1987; Klausmeier et al., 2008; Flynn et al., 2001; Iglesias-Rodriguez et al., 2008). Such changes are subject to investigation as they affect food web dynamics at different trophic levels of aquatic ecosystems. As the elemental composition of photoautotrophs changes, several important ecological processes are affected, including primary production, nutrient cycling, calcification and cellular growth (Sterner, 1990; Urabe and Sterner, 1996; Cebrian and Lartigue, 2004; Orr et al., 2005). In addition, the content of chlorophyll *a* (Chl*a*) within photoautotrophic cells (Chl*a*:C and Chl*a*:N ratio) also varies because of complex influences by light, nutrient and temperature (Armstrong, 2006; Behrenfeld et al., 2002).

During the last two decades marine biogeochemical models have been applied to study the response of primary and export production to long-term changes in the physical environment of aquatic ecosystems over a wide range of temporal and spatial scales (Behrenfeld et al., 2006; Boyd and Doney, 2002; Quere et al., 2005; Harley et al., 2006; Smith et al., 2009; Ayata et al., 2013). These biogeochemical models employ various plankton-ecosystem components, ranging from simple nutrient, phytoplankton, zooplankton, detritus (NPZD) type models to highly complex (more than 20 compartments) ecosystem models (Friedrichs et al., 2007). However, most of these models are based on nitrogen, e.g. Fasham et al. (1990), while assuming constant C:N ratios for estimating mass flux within the ecosystem. In many of the first-generation biogeochemical models (Dugdale, 1967; Aksnes and Egge, 1991; Hurtt and Armstrong, 1999; Fennel et al., 2001; Klausmeier et al., 2004), phytoplankton growth is represented by a Monod model of nutrient concentration (Monod, 1949). Franks (2009) argued that there is no particular reason why an algal cell should behave like an enzyme and follow Michaelis-Menten kinetics for nutrient uptake.

Phytoplankton respond to changes in the physical environment and acclimate accordingly by varying their elemental composition (Goldman et al., 1979; Healey, 1985; Rhee et al., 1981; Burkhardt and Riebesell, 1997; Burkhardt et al., 1999). Models that do not account for photoacclimation (i.e., with a constant Chl:C) can not reproduce *in situ* data of nitrogen limited marine diatoms (Flynn et al., 2001). In addition, many model-data comparison studies have shown that including variable Chl:C and C:N ratios improves the ability of models to fit data (Hurtt and Armstrong, 1996, 1999; Spitz et al., 2001). Geider et al. (1998) presented a photoacclimation model that describes the changes in phytoplankton growth rates and their composition (C, N and chlorophyll) in response to changing physical and chemical environment. The predictions of the model are in good agreement with experimental observations from cultures under balanced and unbalanced growth. Flynn (2003) suggested that quota-type models with explicit inclusion of photoacclimation may work well for many oceanographic modelling scenarios. Armstrong (2006) proposed a dynamic model that includes an optimality-based relationship between Chl:C and N:C ratios, and inferred that the performance of the model was as good as the one of Geider et al. (1998). Saito et al. (2008) defined three categories of nutrient colimitation of primary productivity in aquatic ecosystems

on the basis of their mathematical formulations and visualizations. The chain model of [Pahlow and Oeschies \(2009\)](#) and [Pahlow et al. \(2013\)](#) assumes that phytoplankton growth in the surface ocean is colimited by N and P, in contrast to studies of [Flynn \(2001\)](#) and [Flynn \(2008\)](#) who assumed N and P limitations act independently.

It is difficult to compare and quantify model performance ([Evans, 2003](#); [Friedrichs et al., 2007](#)). Increasing the complexity of models can increase realism, but constraining large numbers of parameters is a major challenge, especially when observations are sparse ([Denman, 2003](#)). Devising and testing a plankton ecosystem model is challenged by two aspects: 1) data availability for comparison with model results (model assessment), and 2) identification of the best representative model solutions to explain these data (model calibration). Data assimilation is concerned with both aspects. Data available from mesocosm experiments can greatly facilitate assessment and calibration of model behaviour (e.g., [Vallino 2000](#); [Schartau et al. 2017](#)). In situ mesocosm perturbation studies provide an effective tool to study the impact of anthropogenic forcing on marine ecosystems ([Riebesell et al., 2008](#)). Environmental conditions (e.g., temperature, pH and salinity) can be manipulated in mesocosms to investigate marine ecosystem responses ([Sommer et al., 2007](#); [Kim et al., 2006](#)). A series of three Pelagic Ecosystem CO₂ enrichment Experiments (PeECE I–III) have been conducted between 2001 and 2005, ([Riebesell et al., 2008](#)). The PeECE experiments made use of large mesocosms that were exposed to different CO₂ concentrations to investigate effects of ocean acidification on marine plankton dynamics associated with increased stratification.

Modelling studies have focussed on the PeECE-I experiment ([Delille et al., 2005](#); [Engel et al., 2005](#)) and simulated observed biogeochemical processes, e.g., primary production, calcification, exudation of dissolved organic carbon (DOC), formation of transparent exopolymer particles (TEP) and viral lysis. In the study of [Joassin et al. \(2011\)](#) a mechanistic model was applied that resolves C, N, P fluxes for *E. Huxleyi*, with explicit representation of calcification and microbial (bacteria and viruses) dynamics. They stressed that the variability in the cellular C:N ratio of *E. Huxleyi* drives the dynamics of most of the processes (including calcification and DOC excretion) in the mesocosms, and attribute the collapse in algal abundance at the end of the bloom to viral lysis. In a recent study, [Krishna and Schartau \(2017\)](#) used an optimality-based model and they showed that large variability observed in calcification between the mesocosms during the PeECE-I experiment could be explained by the small variations in the initial abundance of coccolithophores and the prevailing physiological acclimation states. These small variations generate differences in calcification that are larger than those induced by ocean acidification.

In this study we compare differences in the performance of two carbon-nitrogen regulatory model approaches (Geider-based and Optimality-based) that differ in their representation of the physiology of the phytoplankton, such as the formulation of nutrient uptake (e.g., Droop and Monod kinetics), chlorophyll a synthesis, and growth. The focus of this study is to assess the skills of the models against data from the PeECE-I experiment. Although neither of the models resolve the effect of CO₂ on the plankton growth and carbon fixation, we employ them to investigate whether algal growth dynamics vary systematically in response to different CO₂ perturbations. We set up three sets of simulations representing different CO₂ treatments. For one set of simulations

we include data from mesocosms of high and normal CO₂ levels (future and present day treatments), referred to as "FUT-PRE". The second set consists of simulations of the mesocosms exposed to high and low CO₂ levels (future and glacial treatments), named "FUT-GLA". The third set represents simulations of low and present day CO₂ conditions (glacial CO₂ levels together with present day treatments), designated as "PRE-GLA". We apply the same data assimilation (DA) approach for all three sets. This way we obtain optimised parameter vectors for each CO₂ case. We compare the credibility of parameter values and their uncertainty ranges for both models, and assess how the parameter estimates vary between the CO₂ levels. Last but not the least, we cross-validate the "best" optimised solution for each ensemble case against data not used for parameter optimisation in order to test the general robustness of the model solutions.

2. Methods

We evaluate the performance of two marine ecosystem models, both being calibrated with data from the PeECE-I mesocosm experiment. We explore how the models represent plankton dynamics in the mesocosms under temporally varying conditions (nutrients and light), and their ability to reproduce observations. A data assimilation approach is applied for the identification of optimal model parameter values as well as for their respective ranges of uncertainty, including correlations (collinearities). The plankton ecosystem models (OBM and CN-REcoM) are applied to simulate the carbon and nitrogen fluxes in nine mesocosms.

2.1. Data

A series of three experiments was conducted in the large-scale mesocosm facilities at the marine field station of University of Bergen (Raunefjorden) between years 2001 and 2005 to investigate effects of ocean acidification (OA) on plankton community dynamics. These experiments are referred to as Pelagic Ecosystem CO₂ Enrichment (PeECE) studies (Riebesell and Tortell, 2011). PeECE-I is the first of the series. Nine mesocosms were grouped into three replicates each for three different levels of the initial partial pressure of CO₂ (Engel et al., 2005; Delille et al., 2005). Mesocosms 1–3 (M1, M2, M3) were subject to high dissolved inorganic carbon (DIC) levels (between 2119 and 2122 mmol m⁻³), corresponding to pCO₂ ≈ 710 ppmV. Mesocosms M7, M8 and M9 were initiated with low DIC levels, between 1919 and 1929 mmol m⁻³. The remaining three mesocosms (M4, M5, M6) were initialised with DIC concentrations between 2040 and 2056 mmol m⁻³, corresponding to present day pCO₂ level (≈ 410 ppmV). For our model simulations, we consider the air above the mesocosms' surface to be a mixture of 90 % air with the target pCO₂ and 10 % of air with ambient atmospheric pCO₂ (Krishna and Schartau, 2017). All nine mesocosms were enriched with similar initial concentrations of DIN (dissolved inorganic nitrogen), approximately 15 mmol m⁻³. Daily measurements of salinity and temperature are available for the entire period (23 days) of the PeECE-I experiment. We interpolate daily temperature and salinity values and use these as environmental forcing for our model simulations. We

also include hourly photosynthetically active radiation (PAR) data from meteorological global irradiance measurements of the Geophysical Institute of the University of Bergen, Norway (Skartveit et al., 2001).

In this study we do not attempt to explicitly resolve coccolithophores and calcification, as done in Krishna and Schartau (2017). Only a single compartment of all phytoplankton is used instead, which reduces the number of free parameters and allows for a better comparison of the two model versions of interest. However, calcification affects the amount of dissolved inorganic carbon (DIC) utilized by the algae in the mesocosms. Therefore, in this study we employ “corrected” DIC data, which is the sum of observed DIC and PIC (particulate inorganic carbon). In this manner, we attempt to correct for the effect calcification on the DIC data.

2.2. Candidate models

We apply two carbon and nitrogen regulatory marine ecosystem models, namely, Carbon:Nitrogen Regulated Ecosystem Model (CN-REcoM) and Optimality-based model (OBM) to simulate the PeECE-I experiment. We assume the mesocosms to be homogeneously mixed and do not include any vertical export of particulate organic matter in either model. Furthermore, we consider light gradients in mesocosms. Attenuation of light in water column is estimated by Lambert-Beer’s law and depends on depth (d), the light absorption coefficients of seawater (a_w) and chlorophyll a (a_c), and the chlorophyll concentration (Chl). The mean irradiance (I_{avg}) is thus derived by integrating this function over depth.

$$\kappa = a_w + (a_c \cdot \text{Chl}) \quad (1)$$

$$I_{avg} = \frac{0.95 \cdot I}{\kappa \cdot d} \cdot [1 - \exp(\kappa \cdot d)] \quad (2)$$

Where κ is the attenuation coefficient for light at a given depth. It is to note that tops of mesocosms were covered with tetrafluoroethylene films in the PeECE-I experiment and allowed only 95 % transmission for photosynthetically active radiation (Delille et al., 2005). Therefore, we assume a fraction of incident irradiance, which is equal to $0.95 \cdot I$, that enters into the mesocosms. In both models effects of phosphorus on photoautotrophic growth are neglected and we do not simulate diazotrophy here.

2.3. General structure of models

Both models describe carbon and nitrogen fluxes between: (1) dissolved inorganic compounds (DIC and DIN), (2) Phytoplankton (PhyC, PhyN), (3) zooplankton (ZooC, ZooN), (4) detritus (DetC, DetN), (5) labile dissolved organic compounds (lDON, lDOC) and (6) dissolved combined carbohydrates (dCCHO). dCCHO acts as a precursor for transparent exopolymer particles (TEP). Figure 1 shows a schematic representation of the general structure of both models and processes connecting different model compartments. OBM and CN-REcoM only differ in the mechanistic representation

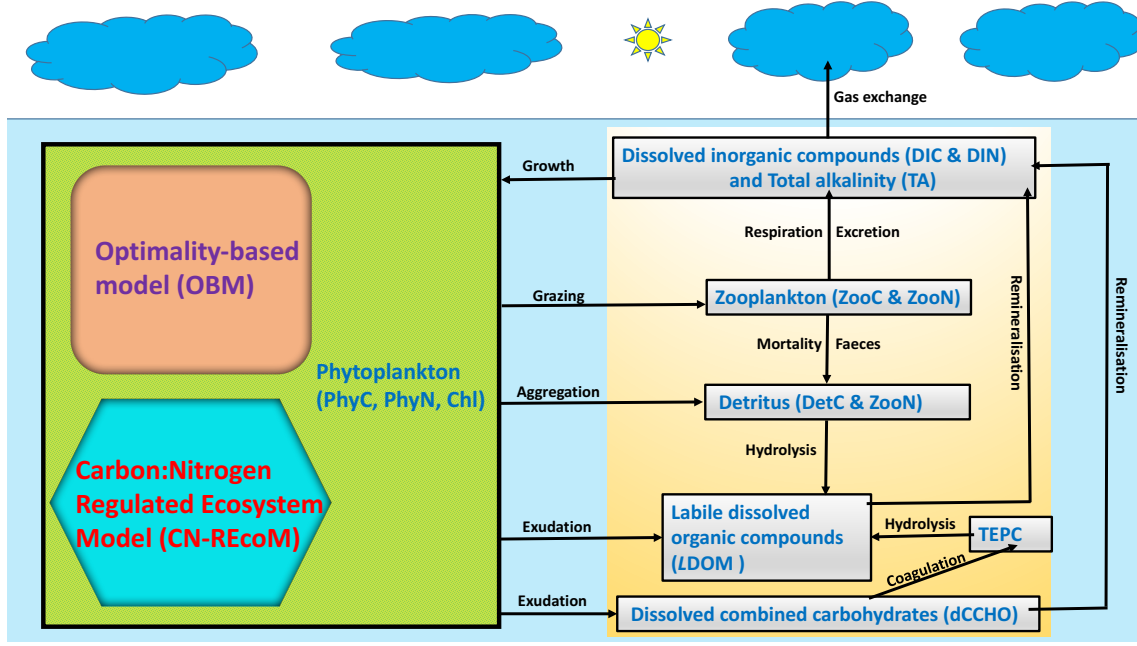


Figure 1: Schematic representation of model compartments connected by different physio-ecological processes. It is note that only phytoplankton compartment is different for the OBM and CN-REcoM, and rest are identical.

of phytoplankton growth dynamics (including chlorophyll synthesis). Formulations of all other model compartments (Zooplankton, Detritus, LDOM, dCCHO and TEP) are identical for both models (Fig. 1).

Phytoplankton: Algal growth is attributed to carbon (C) and nitrogen (N) assimilated by photoautotrophs minus losses due to aggregation, grazing by zooplankton and exudation of organic matter. The general ordinary differential equations (ODEs) for phytoplankton biomass build up in both models are given as:

$$\frac{d}{dt} \text{PhyC} = (\mu - \gamma) \cdot \text{PhyC} - \frac{1}{Q^N} (A + G) \quad (3)$$

$$\frac{d}{dt} \text{PhyN} = V_C^N \cdot \text{PhyC} - \gamma \cdot \text{PhyN} - A - G \quad (4)$$

Where μ and V_C^N are the net growth and nitrogen uptake rates in units d^{-1} and $\text{mol N} (\text{mol C})^{-1} \text{d}^{-1}$. γ is the exudation rate given in d^{-1} and Q^N is the cell quota in $\text{mol N} (\text{mol C})^{-1}$. Nitrogen loss rates due to grazing (G) by zooplankton and aggregation of phytoplankton (A) are given in $\text{mmol N m}^{-3} \text{d}^{-1}$. Parameterisations of A and G are described in [Krishna and Schartau \(2017\)](#). It is to note that the OBM and CN-REcoM employ different mechanistic formulations for the estimation of μ and V_C^N .

For both models, chlorophyll a content in phytoplankton is regulated by its synthesis rate (S_{chl}), A and G . Although parameterisations of A and G are identical for the OBM and CN-REcoM, the ones for S_{chl} are different. The ODE for Chl synthesis is given as:

$$\frac{d}{dt} \text{Chl} = S_{chl} \cdot \text{Chl} - \theta^N (A + G) \quad (5)$$

Details on auxiliary variables and parameters introduced in the above equations are given in table 1.

Model parameters	Description	Unit
1) PON_0	Initial concentration of particulate organic nitrogen	mmol N m ⁻³
2) f_{det}	fraction of PON_0 assigned to non-living detritus	-
3) f_{zoo}	fraction of living PON_0 assigned to zooplankton	-
4) Q_{min}	subsistence quota (minimum cellular N:C ratio)	mol N (mol C) ⁻¹
5) α	Photosynthetic efficiency of phytoplankton	mol C (g Chl a) ⁻¹ m ² W ⁻¹ d ⁻¹
6) γ	exudation rate of photoautotrophs	d ⁻¹
5) ζ^N	cost of N uptake	mol C (mol N) ⁻¹
6) V_0^C	maximum potential photosynthesis rate of photoautotrophs	mol C (mol C) ⁻¹
7) R_{ref}	degradation rate constant at reference temperature	d ⁻¹
Auxiliary Variables		
1) Q^N	N:C ratio of a cell	mol N (mol C) ⁻¹
2) V_{max}^N	maximum N uptake rates of photoautotrophs	mol N (mol C) ⁻¹ d ⁻¹
3) V_C^N	C specific N uptake rate of photoautotrophs	mol N (mol C) ⁻¹ d ⁻¹
4) V_{max}^C	maximum C fixation rate of photoautotrophs	mol C (mol C) ⁻¹ d ⁻¹
5) V^C	C fixation rate of a cell	mol C (mol C) ⁻¹ d ⁻¹
6) μ	net growth rate of photoautotrophs	d ⁻¹
7) A	N-specific phytoplankton aggregation	mmol N (m) ⁻³ d ⁻¹
8) G	N-specific grazing	mmol N (m) ⁻³ d ⁻¹
9) θ	Chl:C ratio of a cell	g Chl a (mol C) ⁻¹
10) θ^N	Chl:N ratio of a cell	g Chl a (mol N) ⁻¹
11) S_{chl}	Chlorophyll synthesis rates of photoautotrophs	d ⁻¹
12) I	Irradiance	W m ⁻²
13) T_{ref}	reference temperature for A_E relation	K

Table 1: Common parameters and auxiliary variables for the CN-REcoM and OBM.

Zooplankton: In both models, zooplankton grazing is described by a Holling type III function (Morozov, 2010) on photoautotrophs and loss in zooplankton biomass (carbon and nitrogen) is accounted for by mortality. Additional details on zooplankton compartment and ODEs for zooplankton biomass are given in Appendix A.1.

Detritus: Sources of detrital C and N biomass are represented in terms of phytoplankton aggregation and mortality of zooplankton in both models. Respective ODEs for detrital biomass can be found in Appendix A.2. Furthermore, the OBM and CN-REcoM also resolve detrital chlorophyll (see Appendix A.2).

Dissolved inorganic compounds: Respiration by Zooplankton, remineralisation of labile dissolved organic carbon (l DOC) compounds and sugars, and gas exchange with the atmosphere are sources for dissolved inorganic carbon (DIC). The only sink of DIC is the carbon assimilated by photoautotrophs. The sources of dissolved inorganic nitrogen (DIN) are zooplankton excretion and remineralisation of labile dissolved organic nitrogen (l DON) compounds. Nitrogen utilisation by an algal community is the primary sink of DIN in both models. The differential equation for DIN is given as: Differential equation on DIC and DIC are given in Appendix A.3 and Appendix A.4.

Dissolved labile organic matter (LDOM): Freshly exuded and leaked organic matter from phytoplankton, hydrolysis of detritus and transparent exopolymer particles (TEP) contribute to formation of L DOC and L DON. The conforming ODEs are de-

scribed in Appendix A.6.

Dissolved combined carbohydrates (dCCHO): dCCHO acts as an additional sink of dissolved organic compounds (DOC). dCCHOs coagulate to form TEP, which is a physical-chemical transformation. The ODEs for formation of dCCHO and carbon content of TEP(TEPC) are explained in Appendix A.6.

2.4. Differences between the OBM and CN-REcoM

The OBM and the CN-REcoM describe the regulation of C:N and Chl:C ratios of algal cells in response to changes in external nutrient, light and temperature conditions. CN-REcoM is largely based on the assumptions of Geider et al. (1998), asserting that maximum photosynthetic rate is achieved by an algal cell when Q^N reaches Q_{max}^N . By contrast, for the OBM μ is maximised by implementing an optimal resource allocation strategy between nutrient acquisition and light-harvesting functions in a cell (Pahlow et al., 2013).

Nutrient limitation: Nutrient assimilation (nitrogen in this study) in CN-REcoM is a Michaelis-Menten function of maximum carbon-specific nitrogen uptake rate (V_{max}^N) and DIN. Notably, formulation of V_{max}^N is modified from that of Geider et al. (1998), e.g. in Schartau et al. (2007):

$$V_{max}^N = V_{fact} \cdot V_0^C \cdot T_f \cdot \frac{Q_{max}}{1 + \exp(-100 \cdot (Q_{max} - Q^N))} \quad (6)$$

With V_0^C being the maximum potential photosynthesis rate in d^{-1} and V_{fact} is a factor (<1) that relates the maximum potential nitrogen to V_0^C . The term T_f is a function for the temperature dependence of metabolic rates as used in Geider et al. (1998), and Q_{max} is the maximum of cellular N:C ratio in unit $\text{mol N} (\text{mol C})^{-1}$. The exponential term introduced in the denominator of Eq. (6) prevents the singularity when V_{max}^N reaches 0 as Q^N increases towards Q_{max} . This makes CN-REcoM more robust and amenable to 1D and 2D modelling approaches where an explicit representation of vertical turbulent mixing is required. V_C^N is finally calculated with the Monod equation:

$$V_C^N = V_{max}^N \cdot \frac{\text{DIN}}{\text{DIN} + k_{\text{DIN}}} \quad (7)$$

In contrast, nitrogen assimilation in the OBM is described by the optimality-based kinetics as proposed in Pahlow et al. (2013). V_C^N is given as a function of optimal allocation factor for nutrient acquisition (f_v^o), DIN and T_f :

$$V_C^N = f_v^o \cdot \left(\sqrt{\frac{1}{(T_f \cdot V_0^N)}} + \sqrt{\frac{1}{A_0 \cdot (\text{DIN})}} \right)^{-2} \quad (8)$$

where V_0^N is the maximum potential nitrogen uptake rate and has unit $\text{mol N} (\text{mol C})^{-1} \text{d}^{-1}$ and A_0 is the nutrient affinity in $\text{m}^3 (\text{mol C})^{-1} \text{d}^{-1}$. Notably, Smith et al.

Parameters	Description	Unit
1) V_{fact}	conversion factor of maximum potential N uptake rate relative to V_0^C	-
2) θ_{max}	maximum Chl:C ratio of a cell	g Chl (mol C) ⁻¹
3) Q_{max}^N	maximum N:C ratio of a cell	mol N (mol C) ⁻¹
4) K_{DIN}	half-saturation constant for nitrate uptake	mmol N m ⁻³

Table 2: CN-REcoM specific parameters.

(2016) suggested A_0 to be a better metric of competitive ability of phototrophs for nutrients than the half-saturation constant of Michaelis-Menten kinetics. Details about the auxiliary variables and parameters related to nitrogen limitation for both models are given in Tables 1, 2 and 3.

Algal growth: The growth rate of an algal cell in the CN-REcoM is estimated by the difference between the carbon fixation rate (V^C) and losses owing to basal respiration and costs of biosynthesis. Furthermore, we assume that the maintenance metabolic rate constants describing basal respiration, remineralisation of algal N, degradation of algal Chlorophyll (Chl) are identical, and denoted as a single parameter R_{ref} in units of d⁻¹.

$$\mu = V^C - (\zeta_N \cdot V_C^N) - R_{ref} \quad (9)$$

Where ζ^N is the cost of N uptake in a photoautotrophic cell in units of mol C (mol N)⁻¹. V^C is affected by the light saturated maximum photosynthesis rate (V_{max}^C), nutrient limitation, light limitation and temperature (Geider et al., 1998).

$$V^C = V_{max}^C \cdot \left[1 - \exp\left(\frac{-\alpha \cdot \theta \cdot I_{avg}}{V_{max}^C}\right) \right] \quad (10)$$

where I_{avg} is the mean irradiance in Wm⁻². θ is the chlorophyll *a*-to-carbon ratio in units g Chl (mol C)⁻¹, and α is the photosynthetic efficiency of photoautotrophs in mol C (g Chl a)⁻¹ m² W⁻¹ d⁻¹.

Parameters	Description	Unit
1) V_0^N	photoautotrophic potential N fixation rate	mol N (mol C) ⁻¹ d ⁻¹
2) R_M^{chl}	cost of chlorophyll maintenance	d ⁻¹
3) R_M	total respiration maintenance cost	d ⁻¹
4) ζ^{chl}	cost of photosynthesis coefficient	mol C (g Chl a) ⁻¹
5) A_0	potential nutrient affinity	m ³ mol C ⁻¹ d ⁻¹
Auxiliary variables		
1) Q_s	Cell quota attached with structural proteins	mol N (mol C) ⁻¹
2) f_v^o	Optimised fraction of N:C allocated for nutrient acquisition	-
3) f_{LH}^o	Optimised fraction of N:C allocated for light harvesting	-
4) $\dot{\theta}$	time derivative of theta	g Chl a (mol C) ⁻¹
5) S_I	degree of light saturation	-

Table 3: OBM specific parameters and auxiliary variables.

In the OBM, μ is estimated by the difference between V^C and the sum of respiration costs (R) due to Chlorophyll a synthesis, nitrogen acquisition and basal respiration. In contrast to the CN-REcoM, the OBM resolves explicitly respiration costs due to light harvesting. V^C of photoautotrophs is calculated from day length (L_d) as a fraction of 24 hour, the degree of light saturation (S_I), optimal allocation factor for light harvesting f_{LH}^o , V_0^C and T_f (Pahlow et al., 2013).

$$V^C = L_d \cdot f_{LH}^o \cdot T_f \cdot V_0^C \cdot S_I \quad (11)$$

It is to note that for the OBM V_0^C and V_0^N are numerically identical. S_I is the (dimensionless) degree of light saturation.

The description of the auxiliary variables and parameters associated with phytoplankton growth for both models are given in Tables 1, 2 and 3.

Chlorophyll synthesis: Pigment synthesis in the CN-REcoM depends on light and V_C^N . Synthesis of chlorophyll ceases as chlorophyll a -to-nitrogen ratio of a cell, θ^N , approaches its maximum θ_{max}^N , which is a model parameter. Under high light conditions, S_{chl} is down-regulated and becomes inversely proportional to the light dependent term ($\alpha \cdot \theta \cdot I_{avg}$).

$$S_{chl} = V_C^N \cdot \frac{1}{2} \cdot (\theta_{max}^N - \theta^N) \cdot \frac{V^C}{\alpha \cdot \theta \cdot I} \quad (12)$$

With θ^N in g Chl (mol N)⁻¹.

Conversely, chlorophyll synthesis in the OBM is driven by a trade-off between carbon fixation and respiratory costs of photosynthesis within the chloroplast, and does not depend on θ_{max}^N . S_{chl} depends on the (optimised) size of the chloroplast (f_{LH}^o), the rate of change in Q^N and light conditions (Pahlow et al., 2013).

$$S_{chl} = \left(\mu + \frac{\dot{\theta}}{\theta} \right) \quad (13)$$

where the term $\left(\frac{\dot{\theta}}{\theta} \right)$ is the optimal regulation term for chlorophyll a synthesis.

Auxiliary variables and model parameters describing chlorophyll synthesis are listed in Tables 1, 2 and 3.

2.4.1. Control parameters

We have selected 7 model parameters of the OBM and 8 parameters of CN-REcoM for the optimisations. 6 parameters are common for the OBM and CN-REcoM. These parameters are:

1. PON_0 ,
2. f_{det} ,
3. f_{zoo} ,
4. Q_{min} ,
5. α ,

6. γ .

Increasing the number of parameters for optimisation could affect the estimation of model parameters (Denman, 2003; Friedrichs et al., 2007). Therefore, we have kept the number of model parameters to a meaningful minimum. Krishna and Schartau (2017) show that collinearities exist between estimates of initial conditions and ecological model parameters, in particular algal growth parameters, and it is not possible to determine either of them independently. Hence, we have decided to include initial conditions as parameters (PON_0 , f_{det} and f_{zoo}) along with the growth parameters in our optimisations. Although the definition of the parameter α slightly differs between OBM and CN-REcoM, it remains a major source parameter for the carbon (C) biomass build up in both models. The units ($\text{mol C (g Chl } a)^{-1} \text{ m}^2 \text{ W}^{-1} \text{ d}^{-1}$) of α are same in OBM and CN-REcoM. The subsistence cell quota (Q_{\min}) turns out to be a critical physiological parameter in estimating C biomass (Krishna and Schartau, 2017). We assume identical organic carbon and nitrogen exudation rates (γ) for photoautotrophs in both models. γ is a critical parameter in constraining TEPC production and hence POC. In CN-REcoM the maximum nutrient uptake rate is derived from V_0^C and V_{fact} , whereas in the OBM, the potential maximum rates of nutrient uptake (V_0^N) and photosynthesis (V_0^C) are numerically identical.

PeECE-I has a short post-bloom period (Krishna and Schartau, 2017), hence parameters like maximum grazing rates and the aggregation rates are poorly constrained by the available data (Schartau et al., 2017). Studies on skill assessment of models (e.g., Friedrichs et al. 2006, 2007; Ward et al. 2010) show that poorly constrained parameters can strongly affect model performance. Therefore, we decided to use the same values for grazing and aggregation-rate parameters as in Krishna and Schartau (2017).

2.4.2. Initial conditions

In OBM and CN-REcoM, the initial-conditions of some state variables are calculated by initial condition parameters. Total initial biomass is given by the parameter (PON_0) and distributed between living and non-living biomass. The parameter f_{det} determines the fraction of PON_0 that is assigned to initial detritus. The remaining fraction, living biomass, is distributed between photoautotrophs and zooplankton as specified by the initial zooplankton fraction parameter (f_{zoo}).

$$PON_0 = \text{Det}N_0 + \text{Zoo}N_0 + \text{Phy}N_0 \quad (14)$$

with the individual fractions:

$$\text{Det}N_0 = f_{det} \cdot PON_0 \quad (15)$$

$$\text{Zoo}N_0 = f_{zoo} \cdot (PON_0 - \text{Det}N_0) \quad (16)$$

$$\text{Phy}N_0 = (1 - f_{zoo}) \cdot (PON_0 - \text{Det}N_0) \quad (17)$$

We compute initial C biomass of zooplankton and photoautotrophs from their respective N biomasses by applying a constant C:N ratio of 6.625. For the C biomass of detritus we assume a higher C:N ratio ($= 2 \times 6.625$). We made this assumption because mesocosms were filled with late/post-bloom, nutrient depleted water, and hence all dead particulate organic matter should have a C:N ratio that is typical for such post-bloom conditions (Krishna and Schartau, 2017). We assume an initial Chl:N ratio of

one. Therefore, the estimates are identical to initial N biomass of photoautotrophs in both models. Initial conditions of DIC, TA, TEPC and dCCHO are taken from the data of the mesocosms, whereas for DOC and DON we assumed small initial values close to zero (with C:N = 10) as explicit measurements of refractory and labile DOC are not available (DON = 0.005 mmolN m⁻³ and DOC = 0.05 mmolC m⁻³).

2.5. Data assimilation

Although both modelling approaches (OBM and CN-REcoM) in this study do not resolve the effects of changes in carbonate chemistry of seawater on algal growth dynamics, we desire to assess the skills of the models in this respect. Thus, we would like to simulate the three different CO₂ treatments using the OBM and the CN-REcoM, and investigate if there are any differences in the parameter estimates or the data-model misfit function that suggest possible CO₂ effects on phytoplankton growth dynamics. Hence, we set up three simulation cases representing different CO₂ concentrations. For the first simulation set (FUT-PRE) we assimilated means of the data of mesocosms treated with high (FUT) and medium (PRE) CO₂ concentrations (high CO₂ scenario). For the FUT-GLA simulations means of the data of the FUT and GLA mesocosms were assimilated (medium CO₂ case). For PRE-GLA, we considered means of the data of the PRE and GLA mesocosms. Our decision to assimilate data from two treatments in each parameter optimisation is owing to the fact that the data sets for each individual treatment (3 mesocosms) are insufficient for constraining our model parameters. Thus, we included data from 6 mesocosms in each set (Table 4).

Simulated CO ₂ cases	Description	Assimilated mesocosms
1) "FUT-PRE"	High CO ₂ scenario	M1, M2, M3, M4, M5, M6
2) "FUT-GLA"	Today's CO ₂ scenario	M1, M2, M3, M7, M8, M9
3) "PRE-GLA"	Low CO ₂ scenario	M4, M5, M6, M7, M8, M9

Table 4: Grouping of mesocosms in the respective CO₂ cases.

2.5.1. Data-model misfit function (Cost function)

As described above, we consider the data for three cases (FUT-PRE, FUT-GLA, PRE-GLA). However, we impose the same statistical assumptions in all three cases. The observations vector (\vec{y}_i) contains daily means of observations from six mesocosms for five variables, namely: (1) dissolved inorganic carbon (DIC, mmol m⁻³), (2) dissolved inorganic nitrogen (DIN) (nitrate + nitrite, mmol m⁻³), (3) chlorophyll *a* (Chl *a*, mg m⁻³), (4) particulate organic nitrogen (PON, mmol m⁻³), (5) particular organic carbon (POC, mmol m⁻³).

Likewise, the model counterpart, the vector $H_i(\vec{x})$, consists of daily means of model states for the same variables. PON and POC are not state variables as they represent sums of biomasses of living organisms and the detritus (Vallino, 2000). Modelled PON (PON^{mod}) is the sum of N concentrations of photoautotrophs (PhyN), zooplankton (ZooN) and detritus (DetN) for both models.

$$\text{PON}^{\text{mod}} = \text{PhyN} + \text{ZooN} + \text{DetN} \quad (18)$$

Similarly, modelled POC (POC^{mod}) includes contributions from C biomasses of phytoplankton (PhyC), zooplankton (ZooC), detritus (DetC), and TEPC and dCCHO.

$$\text{POC}^{mod} = \text{PhyC} + \text{ZooC} + \text{DetC} + \text{TEPC} + \text{dCCHO} \quad (19)$$

The residual vector (\vec{d}_i) represents the differences between observation vector and the model output vectors.

$$\vec{d}_i = \vec{y}_i - H_i(\vec{x}) = \underbrace{\begin{pmatrix} \text{DIC}_i \\ \text{DIN}_i \\ \text{Chl } a_i \\ \text{PON}_i \\ \text{POC}_i \end{pmatrix}}_{\text{data}} - \underbrace{\begin{pmatrix} \text{DIC}_i^{mod} \\ \text{DIN}_i^{mod} \\ \text{Chl}_i^{mod} \\ \text{PON}_i^{mod} \\ \text{POC}_i^{mod} \end{pmatrix}}_{\text{model results}} \quad (20)$$

We assume the same observational error model for all three cases. We compute daily residual standard errors (σ_i) from the available measurements on a given day for each case. The estimation of residual standard errors for DIC is not as straightforward as for the other variables (Krishna and Schartau, 2017). Since there are three different CO_2 treatments, we first computed the differences (offsets) of the mean initial DIC concentrations for the FUT and GLA treatments from the mean initial DIC concentrations of the PRE treatment. Then, for those mesocosms that were treated with high CO_2 concentrations, the computed “offset” was subtracted from the DIC measurements. Likewise, DIC values for the mesocosms of the PRE treatment were adjusted (the “offset” value was added). Thus, initial mean DIC values for all three treatments become identical, but the variability with respect to the means of the observations between the treatments still persists.

We defined our cost function as the maximum likelihood (ML) estimator to find out the parameter vector ($\hat{\Theta}$) that maximises the conditional probability of explaining the data for a given model together with a set of values assigned to the parameter vector (Θ). Since we do not consider explicit prior information for the estimation of model parameters, the posterior probability, $p(\vec{y} | \Theta)$ becomes proportional to the negative logarithm of the maximum likelihood ($J(\Theta)$, or the cost function). For all three simulation cases we use a similar cost function but different data (\vec{y}_i) and covariances (\mathbf{R}_i) from the corresponding six mesocosms available for the period of $N_t = 23$ days. The subscript i indicates on which day measurements were taken.

$$p(\vec{y} | \Theta) \propto \exp\left[-\frac{1}{2}J(\Theta)\right] = \text{constant} \cdot \exp\left[-\frac{1}{2} \sum_{i=1}^{N_t} \vec{d}_i^T \mathbf{R}_i^{-1} \vec{d}_i\right] \quad (21)$$

Since the constant term in Eq. (21) is independent of Θ , the cost function we eventually

minimise and compute is:

$$J(\Theta) = \sum_{i=1}^{N_t} (\vec{y}_i - H_i(\vec{x}))^T \mathbf{R}_i^{-1} (\vec{y}_i - H_i(\vec{x})) \quad (22)$$

The time-varying covariance matrices \mathbf{R}_i are constructed with diagonal matrices \mathbf{S}_i (containing σ_i as diagonal elements at time t_i) and correlation matrices ($\mathbf{C}_{(\vec{y})}$). Correlations between measurements were computed based on data from six mesocosms for each case.

$$\mathbf{R}_i = \mathbf{S}_i \cdot \mathbf{C}_{(\vec{y})} \cdot \mathbf{S}_i \quad (23)$$

The covariance matrices (\mathbf{R}_i) are constructed for three distinct periods:

1. Exponential growth phase (Day 1 to 10) ,
2. Bloom period (Day 11 to 14) ,
3. Post-bloom period (Day 15 to 23) .

Since the correlations between observations can change from the exponential growth period to the post-bloom period (Schartau et al., 2017; Krishna and Schartau, 2017), we employ period-specific covariance matrices in our data assimilation approach. However, for the bloom period we assume measurements to be uncorrelated. We dedicate this period as a transition period because we do not have any information about the precise time when correlations change. Thus, for the bloom period covariance matrices have off-diagonal elements of zero and contain only variances as the diagonal elements.

2.5.2. Parameter optimisation and uncertainty estimation

We applied intrinsic functions available in the R package FME (Soetaert and Petzoldt, 2010) to optimise model parameters and estimate uncertainties associated with them in all three cases (FUT-PRE, FUT-GLA, PRE-GLA). The first step of the optimisation procedure is to come up with a parameter set corresponding to the model solution (reference solution) that is qualitatively in good agreement with observations of the FUT-GLA case. The second step is the application of *simulated annealing* algorithm (SANN), see (Bélisle, 1992), to perform a stochastic search of the parameter space and find out the global minima of $J(\Theta)$ without getting trapped in local minima. In the next step the result of the SANN algorithm is refined by using the *Broyden-Fletcher-Goldfarb-Shanno* (BFGS) algorithm (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970) — a gradient-based search algorithm that gives the best local parameter estimate ($\hat{\Theta}$). The fourth step is to approximate the Hessian matrix (\mathcal{H}) at $\hat{\Theta}$ using the “Hessian” function available in the “numDERIV package” in R. Further, the covariance matrix ($U_{\hat{\Theta}}$) is computed from \mathcal{H} (Schartau et al., 2017) as:

$$U_{\hat{\Theta}} = 2 \cdot \mathcal{H}^{-1} \quad (24)$$

Lastly, the marginal error or standard error information (u_l) around $\hat{\Theta}$ is derived from the covariance matrix, $U_{\hat{\Theta}}$ (Schartau et al., 2017).

$$u_l = t_{df}^{1-\frac{\alpha}{2}} \sqrt{U_{\Theta_{ll}}} \quad (25)$$

where $t_{df}^{1-\frac{\alpha}{2}}$ is the two-tailed student's t distributions for a given degree of freedom(df) and the quantile of the parametric probability density distribution (α), see Marsili-Libelli et al. (2003).

The fifth and the final step of the optimisation procedure is the application of a Markov Chain Monte Carlo (MCMC) method (Harmon and Challenor, 1997). The MCMC method employed in this study is based on the Metropolis-Hastings (MH) algorithm (Metropolis et al., 1953; Hastings, 1970). The MH algorithm uses the marginal error information around $\hat{\Theta}$ and samples the complete probability distribution of the parameter space to generate an ensemble of optimised model solutions. Application of a MCMC method serves three main purposes: first, it estimates the credible range of optimised parameter values (uncertainties); second, it derives collinearities (correlations) between the optimised parameter estimates. And third, it also provides the best “optimised” solution, corresponding to the lowest cost function value.

Both plankton ecosystem models were coded and compiled as a shared library in FORTRAN. This is done so that we can apply a FORTRAN-R wrapper function. The wrapper allows us to take advantage of fast numerical Euler forward integrations of the model equations in FORTRAN, and at the same time we can benefit from the open-source R platform.

3. Results

3.1. Optimised parameter estimates and associated uncertainties

Seven parameters from the OBM and eight from the CN-REcoM were optimised separately in the three CO₂ treatment cases (FUT-GLA, FUT-PRE and PRE-GLA). With our data assimilation approach we obtained an ensemble of optimised model solutions for each CO₂ case, including information on parameter uncertainties. We attribute the differences in optimised parameter values between the three independent optimisations to potential CO₂ effects. Table 5 shows the optimised parameter estimates for both models in all the three cases. Our results clearly indicate differences in optimised estimates of the respective control parameters and their associated uncertainties between the models. Figure 2 shows uncertainties in estimates of initial concentrations of PON, zooplankton and detritus for the FUT-PRE, FUT-GLA and PRE-GLA mesocosms. For the OBM, estimates of initial concentrations of PON, zooplankton and detritus are similar between the CO₂ treatments. However, we find differences in optimised values of initial concentrations between the cases for the CN-REcoM. The initial concentrations of detritus and zooplankton seem to be underestimated by CN-REcoM in the mesocosms treated with medium (present day) CO₂ levels (Fig. 2). Low grazing pressure in the FUT-GLA mesocosms promotes a faster algal biomass build up, which is corroborated by the strong negative correlation between the parameters α and f_{zoo} for the CN-REcoM (Table A.3).

We expected to identify systematic differences in parameter estimates (primarily growth parameters) between the three CO₂ treatments for both models, which would clearly point towards CO₂-effects on the algal growth dynamics that are not explicitly resolved by these two models. We observe a distinct pattern in the estimates of parameters Q_{min} and V_0^N (numerically identical to V_0^C) only for the OBM. The estimates of Q_{min} and V_0^N are highest for the mesocosms exposed to high CO₂ concentrations compared to those with low CO₂ (Fig. 3). By contrast, a similar relationship is not observed for the CN-REcoM, as no inferable differences in estimates of Q_{min} and V_0^C can be seen between the treatments.

For the CN-REcoM, parameters (mainly α and Q_{min}) seem to be poorly constrained by data, which is evident from the large uncertainties in the optimised estimates of α and unrealistically low values of Q_{min} for photoautotrophs (Fig. 3). The estimates of Q_{min} are as low as 0.01 for the CN-REcoM, which are lower than the minimum in N:C ratios seen for the extreme case of species with very high fat content, e.g. *Nannochloropsis oculata* (Flynn et al., 1993; Hoffmann et al., 2010). Estimates of α are too high for the CN-REcoM in all three cases (Fig. 3). Apparently, CN-REcoM minimises our cost function by unrealistic overestimation of α and underestimation of Q_{min} . However, this brute-force minimisation increases the chance of fitting noise to data (Friedrichs et al., 2007). On the contrary, estimates of α and Q_{min} are well constrained for the OBM by the data in all three cases (Fig. 3). Moreover, the optimised values are realistic and comparable to the ones found in laboratory and field experiments.

Parameters	Description	FUT-PRE	FUT-GLA	PRE-GLA	Unit
OBM					
1) PON_0	Initial concentration of particulate organic nitrogen	1.80	1.94	1.90	mmol N m ⁻³
2) f_{det}	fraction of PON_0 assigned to non-living detritus	0.37	0.52	0.40	-
3) f_{zoo}	fraction of living PON_0 assigned to zooplankton	0.94	0.77	0.77	-
4) α	Photosynthetic efficiency of phytoplankton	1.40	3.33	1.65	mol C (g Chla) ⁻¹ m ² W ⁻¹ d ⁻¹
5) γ	Algal loss rate of organic carbon and nitrogen	0.23	0.33	0.27	d ⁻¹
6) Q_0/Q_{min}	subsistence quota (minimum cellular N:C ratio)	0.04	0.03	0.02	mol N (mol C) ⁻¹
7) V_0^C	photoautotrophic maximum potential photosynthesis rate	3.27	3.26	3.31	d ⁻¹
CN-REcoM					
1) PON_0		1.86	1.80	2.01	mmol N m ⁻³
2) f_{det}		0.38	0.10	0.47	-
3) f_{zoo}		0.98	0.92	0.933	-
4) α		5.67	7.72	7.70	mol C (g Chla) ⁻¹ m ² W ⁻¹ d ⁻¹
5) γ		0.21	0.23	0.24	d ⁻¹
6) Q_0/Q_{min}		0.02	0.01	0.01	mol N (mol C) ⁻¹
7) V_0^C	photoautotrophic maximum potential photosynthesis rate	3.27	3.26	3.31	d ⁻¹
8) V_{fact}	conversion factor of maximum potential N uptake rate relative to V_0^C	0.31	0.22	0.24	-

Table 5: “Best” parameter estimates obtained from the MCMC method for both the models.

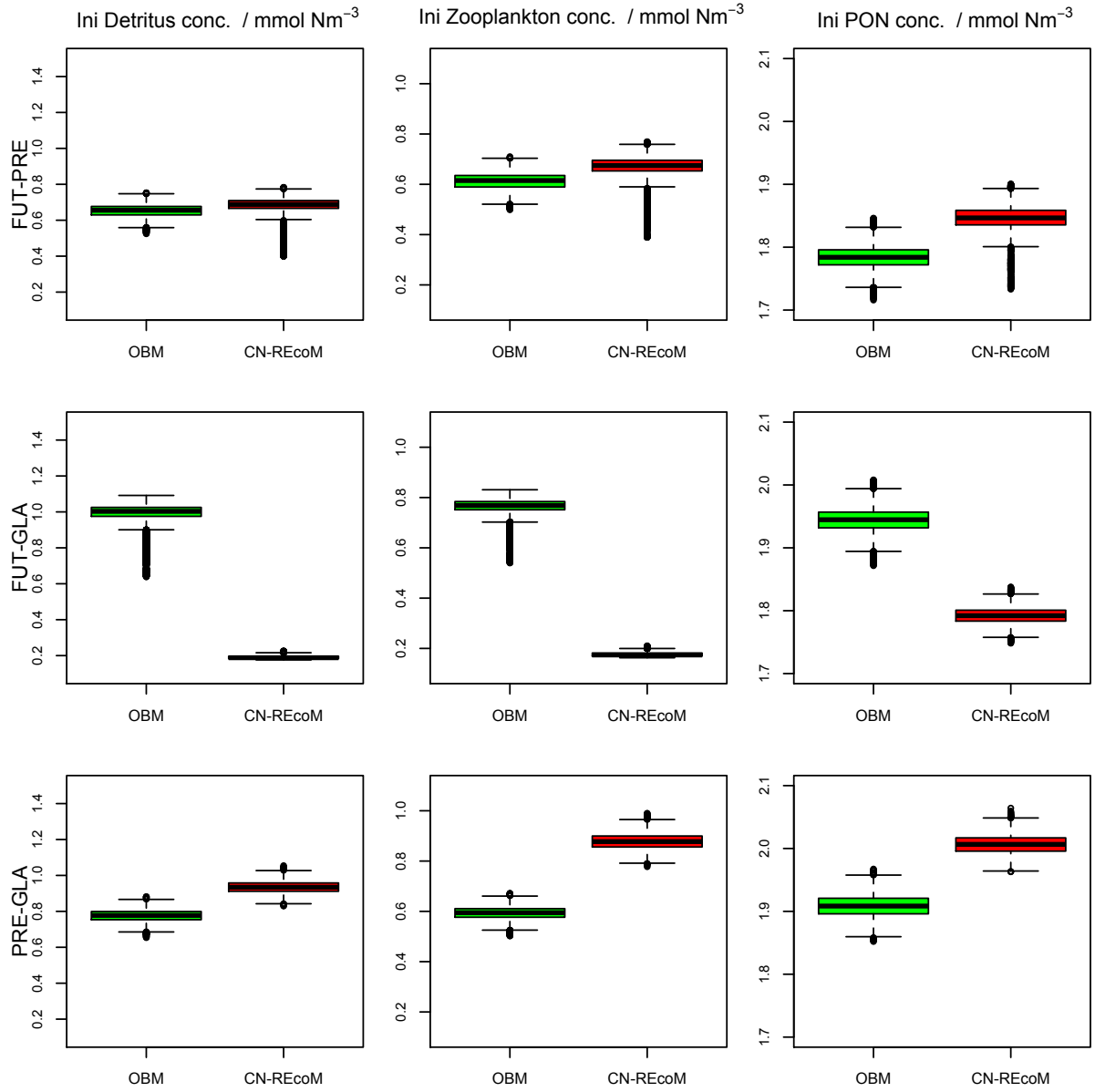


Figure 2: Boxplot representation of the posterior distribution of initial condition parameters as obtained from the MCMC method for the CN-REcoM and OBM in three CO₂ cases.

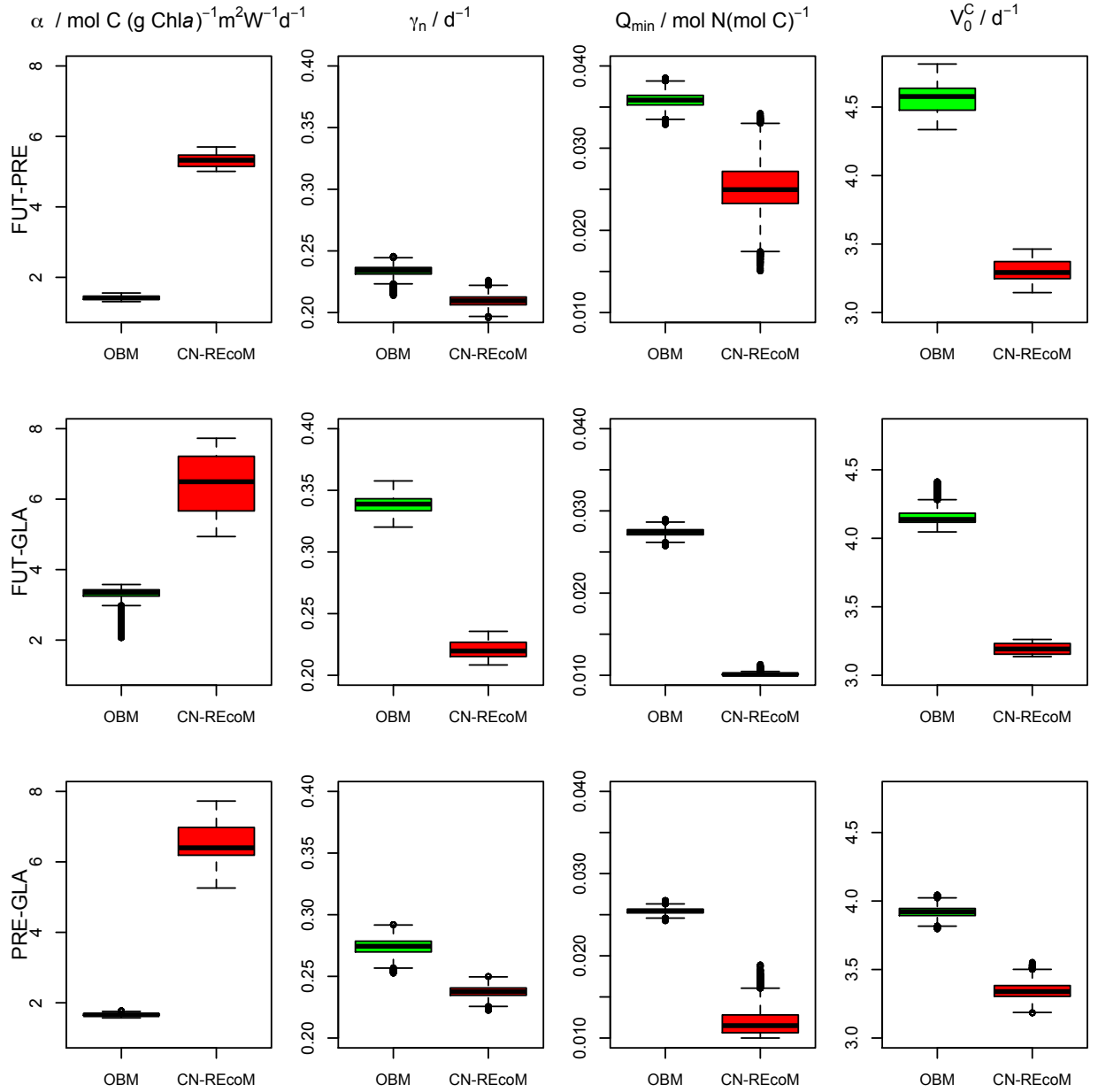


Figure 3: Boxplot representation of posterior distribution of ecological parameters as obtained from the MCMC method for the CN-REcoM and OBM in three CO₂ cases.

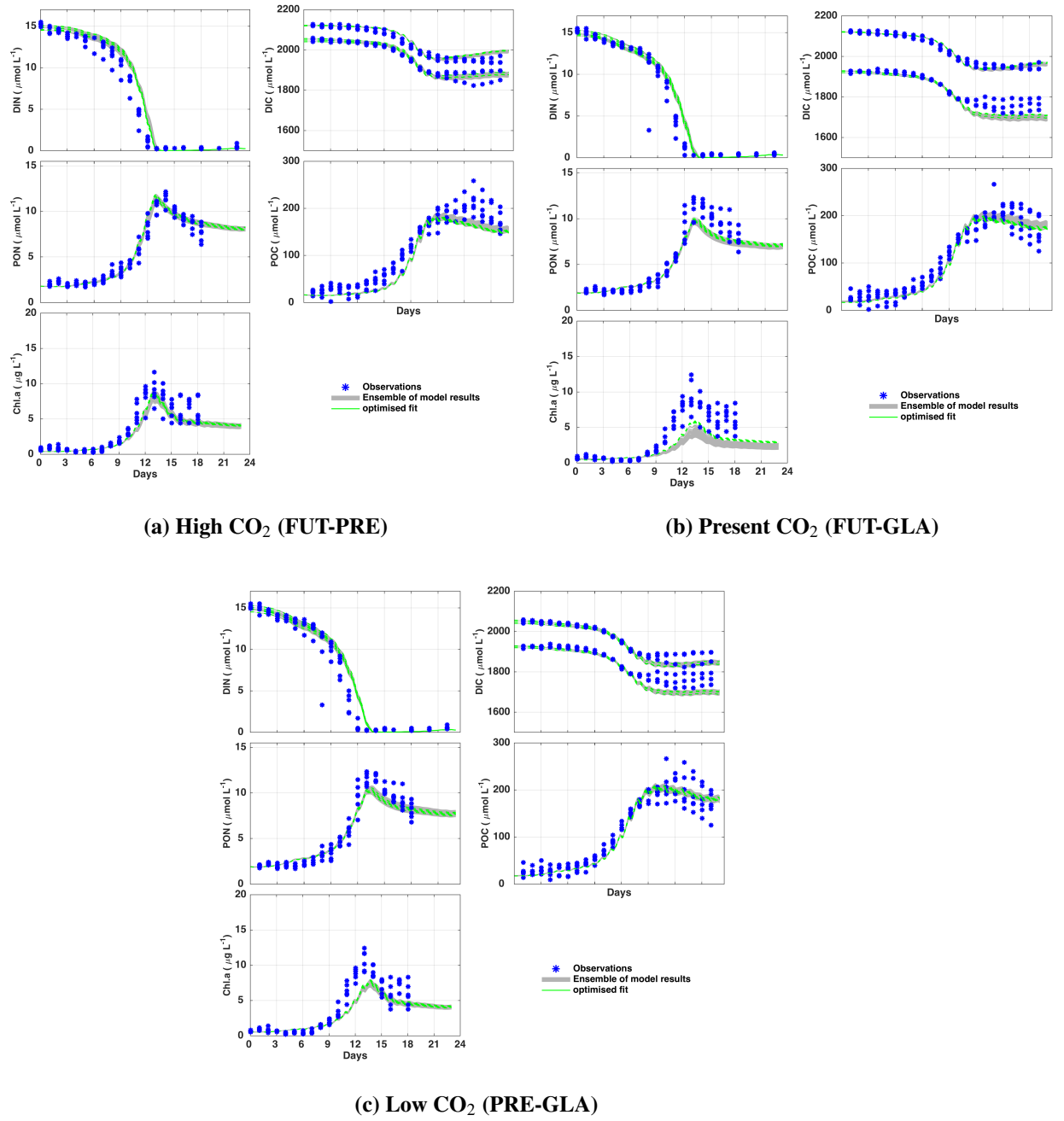


Figure 4: OBM ensemble plot for mesocosms with CO₂ treatment, high (a), present (b) and low (c). Symbols represent observations, grey trajectories ensemble solutions (100 members randomly subsampled) and dashed green line represents the "best" optimised solution corresponding to the lowest cost function value.

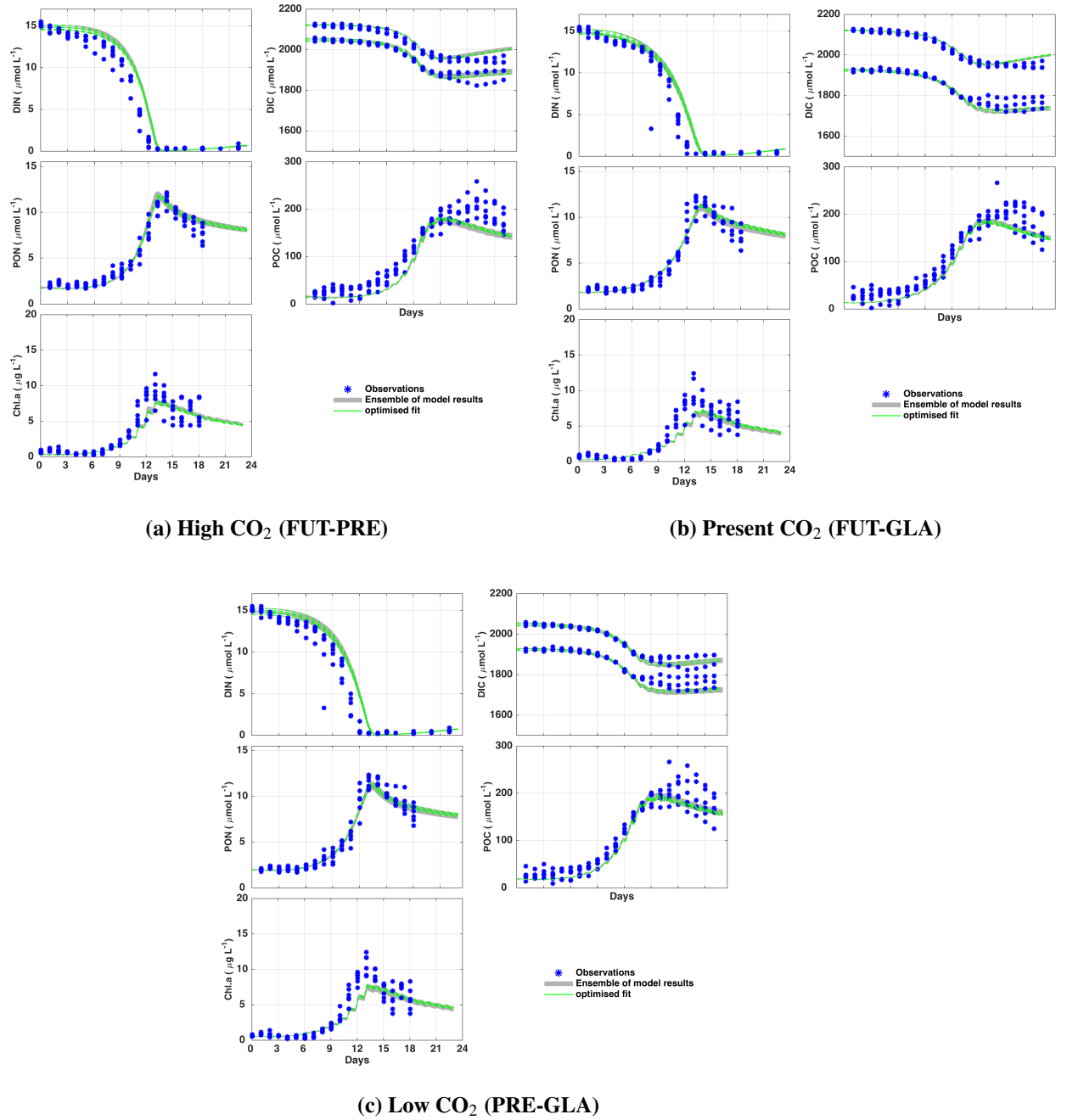


Figure 5: CN-REcoM ensemble plot for mesocosms with CO₂ treatment, high (a), present (b) and low (c). Symbols represent observations, grey trajectories ensemble solutions (100 members randomly subsampled) and dashed green line represents the "best" optimised solution corresponding to the lowest cost function value.

3.2. Data-Model intercomparison

Uncertainties in credible values of parameter estimates can generate model solutions that are statistically indistinguishable. For each CO₂ perturbation case we obtain an ensemble of model solutions from the MCMC method. Figures 4a, 4b and 4c show ensemble solutions for the FUT-PRE, FUT-GLA and PRE-GLA mesocosms compared to data for the OBM. Likewise, Figures 5a, 5b and 5c depict model-data comparisons for the CN-REcoM. In general, the OBM appears to perform better than the CN-REcoM in reproducing data. This inference is quantitatively supported by low cost function values obtained for the solutions of the OBM compared to those for the CN-REcoM. Large variations in cost function values highlight the parameter identification problem for the CN-REcoM.

Models	PRE-GLA	FUT-GLA	FUT-PRE
OBM	$J = 1966 \pm 4.0$	$J = 3143 \pm 20.4$	$J = 3099 \pm 5.0$
CN-REcoM	$J = 2185 \pm 20.0$	$J = 5693 \pm 129.4$	$J = 2982 \pm 10.5$

Table 6: Mean cost function values and standard deviations based corresponding to the posterior distribution of parameter estimates obtained from the MCMC method for the three CO₂ cases.

Both models reproduce DIC, PON and POC observations well in all three treatments (Figs. 4a, 4b, 4c, 5a, 5b and 5c). Optimised solutions of OBM and CN-REcoM show a systematic offset (delay) in the draw down of DIN which is more pronounced between day 9 and day 13. The bias persists in all three CO₂ cases despite differences in parameter estimates. It seems, data assimilation method compensate for phase-lagged PON and DIN data by fitting simulated PON to corresponding observations on the cost of misfits in DIN. By doing so it generates overall low data-model misfit for the entire period of the experiment in all three cases.

The accumulation of POC during the post bloom period is slower for the optimised solutions of CN-REcoM compared to OBM. Since dCCHOs are the precursors of TEPC, low estimates of γ for the CN-REcoM (Fig. 3) may result in the lower TEPC production, and hence low POC concentrations. The OBM underestimates the observed Chl *a* in the FUT-GLA case. This underestimation can be attributed to high optimised values of α for phytoplankton in the FUT-GLA mesocosms (Fig. 3). The higher the estimate of α , the lower the chlorophyll synthesis rate and higher the carbon fixation rate in an algal cell.

3.2.1. Differences between the High-CO₂ and low-CO₂ model solutions (FUT-PRE and PRE-GLA)

Both models reveal a CO₂-dependency of subsistence quota (Q_{min}) of phytoplankton in mesocosms. According to MCMC results, posterior estimates of Q_{min} increase with the CO₂ concentration (Fig. 3). A possible physiological interpretation of this could be that phytoplankton in high-CO₂ mesocosms experience environmental stress, and hence invest more resources/energy in maintenance of cells to sustain life, e.g., synthesis of enzymes and thickening of cell wall. Consequently, less resources are

available to support primary production and cellular growth. On the contrary, in the low CO₂ mesocosms phytoplankton are relatively less stressed. We find a pattern in estimates of V_0^C between CO₂ cases for the OBM. Potential cellular carbon fixation rates are higher for phytoplankton in the FUT-PRE mesocosms than the PRE-GLA ones. This is consistent with the results of [Riebesell et al. \(2007\)](#) and [Engel et al. \(2013\)](#), who observed increased photosynthetic rates of phytoplankton in response to elevated $p\text{CO}_2$. However, we do not find any relationship between CO₂ concentrations and V_0^C for the CN-REcoM.

Although the OBM predicts high estimates of V_0^C for the FUT-PRE mesocosms, the accumulation of POC is slow (Fig. 4a). Observations regarding effects of elevated CO₂ on marine phytoplankton have been ambiguous. On the one hand, studies have shown elevated carbon fixation and growth rates in variety of phytoplankton taxa in acidic environments ([Riebesell et al., 2007](#); [Hare et al., 2007](#); [Riebesell et al., 2007](#); [Engel et al., 2013](#)). On the other hand, primary production is often negatively impacted by increasing CO₂ concentrations ([Sciandra et al., 2003](#); [Langer et al., 2006](#)). Thus, the slow build up of biomass as predicted by the OBM for the high-CO₂ mesocosms is not unrealistic. Furthermore, from the POC data it is difficult to derive an inference on relationship between CO₂ treatment and POC build up, as the latter becomes highly variable under nutrient depleted conditions. Production of TEPC is considered one of the reasons for the variability in POC accumulation ([Engel et al., 2014](#)).

In general, models exhibit some delay in DIN uptake, but the draw down is slightly faster for the FUT-PRE mesocosms, especially between day 9 and 13. This may be because the maximum nitrogen uptake rates (V_0^N) of phytoplankton are higher for the high-CO₂ mesocosms than the other two treatments. For the CN-REcoM, the product of V_0^C and V_{fact} yields the estimate of V_0^N , whereas for the OBM V_0^C and V_0^N are numerically identical. Table 5 shows that both models find high V_0^N estimates for the FUT-PRE mesocosms. [Bowes \(1993\)](#) and [Beardall et al. \(2009\)](#) also reported enhanced nitrogen uptake rates for phytoplankton exposed to elevated CO₂ levels.

3.3. Validation and cross-validation of optimised model solutions

The cost function is a useful metric to quantify model-data misfit. However, obtaining low model-data misfits is not only the criterion to assess the performance of a model. Generating low cost function values does not necessarily indicate that a data-assimilation model is robust, until it is validated against data that were not assimilated or not used for parameter optimisations.

We cross-validated the “best” solution (optimised solution that yields the lowest cost function value) from the respective CO₂ ensemble solutions against the independent data. For instance, the cost function value is assessed for the optimised solution of the FUT-PRE mesocosms against data from GLA (glacial) mesocosms. Likewise, the solution for the respective models that yielded the lowest cost function value for the PRE-GLA mesocosms is cross-validated against the FUT (future) mesocosms. The cost function value for the optimised FUT-GLA solution is evaluated with data from the PRE (present) mesocosms. In this way we are able to assess the robustness of our optimised model solutions.

Figure 6 shows the cost function values corresponding to the optimised solutions for the three CO₂ treatments. In addition, the figure depicts the model-data misfits

that are obtained when optimised solutions for FUT-PRE, FUT-GLA and PRE-GLA mesocosms are validated with independent data.

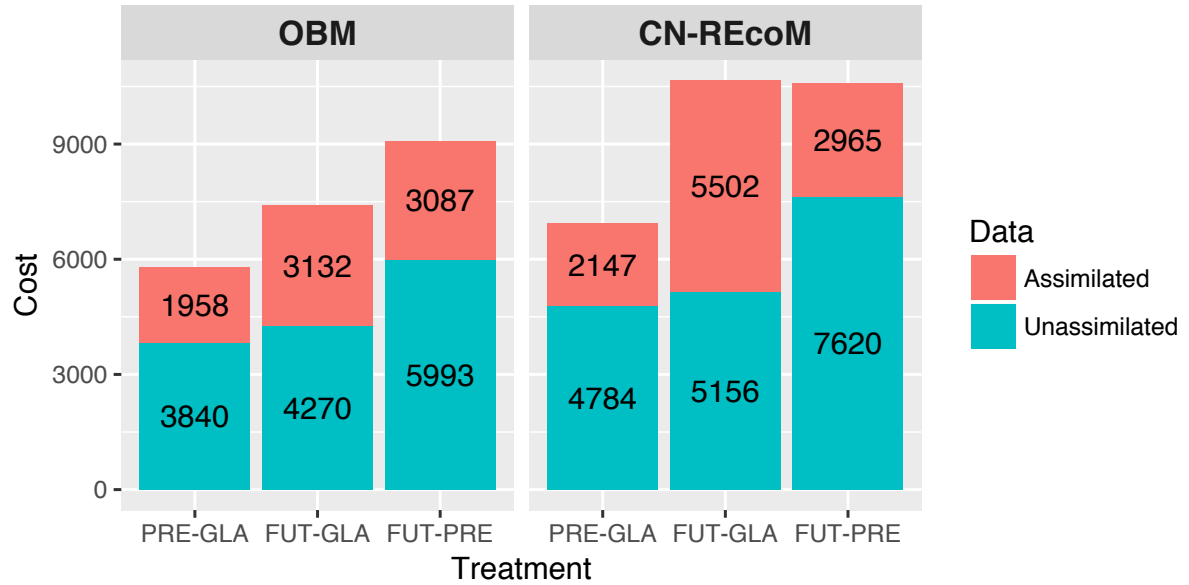


Figure 6: Stacked bar plot representation of the cost function values corresponding to validation of the “best” optimised solutions obtained in the three CO₂ case against assimilated and unassimilated data for the OBM and CN-REcoM.

In general, the OBM yields lower misfits than the CN-REcoM when configured with assimilated and unassimilated data. Overall cost owing to the optimised solution for the low CO₂ mesocosms for the OBM is $\approx 10\%$ lower than for the CN-REcoM. Likewise, the optimised FUT-GLA solution for the CN-REcoM produces $\approx 57\%$ higher model-data misfit compared to the OBM (Fig. 6). However, the solutions for the high-CO₂ mesocosms for the OBM and CN-REcoM show similar cost function values. The cross-validation results also give an edge to the OBM over the CN-REcoM (Fig. 6). For the CN-REcoM, the cost corresponding to the cross-validation of the PRE-GLA solution with the data from the FUT-PRE mesocosms is $\approx 24\%$ higher than the one of the OBM. Similarly, the optimised FUT-GLA solution for the CN-REcoM induces $\approx 20\%$ higher misfit relative to the OBM when compared with unassimilated data. The largest discrepancy in the cross-validation results is seen for the high-CO₂ case (FUT-PRE) between the models. The OBM yields $\approx 28\%$ lower misfit value than the CN-REcoM. To summarise, results of our cross-validation experiments suggest that the OBM is a more robust model than the CN-REcoM in reproducing the unassimilated, independent data.

3.4. Cellular ratios

Optimised estimates of V_0^C for the OBM are higher than those of CN-REcoM in all three cases (Fig. 3). Figure 7 shows variations in cellular photosynthetic rates owing to changes in Chl:C ratios. The OBM yields higher carbon fixation rates than the CN-REcoM, however Chl:C ratios remain low (especially in the post bloom period). It seems that, as the OBM accounts for optimal allocation of resources between light

harvesting and nutrient acquisition functions of a cell, it can maintain some high carbon fixation rates even when Chl:C ratios are low. This behaviour of the OBM is apparent in the FUT-GLA case. On the other hand, Figure 7 reveals a strong constraint in the formulation of the photosynthesis for the CN-REcoM: which is a cell can achieve high values of carbon fixation rates only when Q^N approaches Q_{max}^N and θ approaches θ_{max} . [Armstrong \(2006\)](#) pointed out a similar problem for the model of [Geider et al. \(1998\)](#).

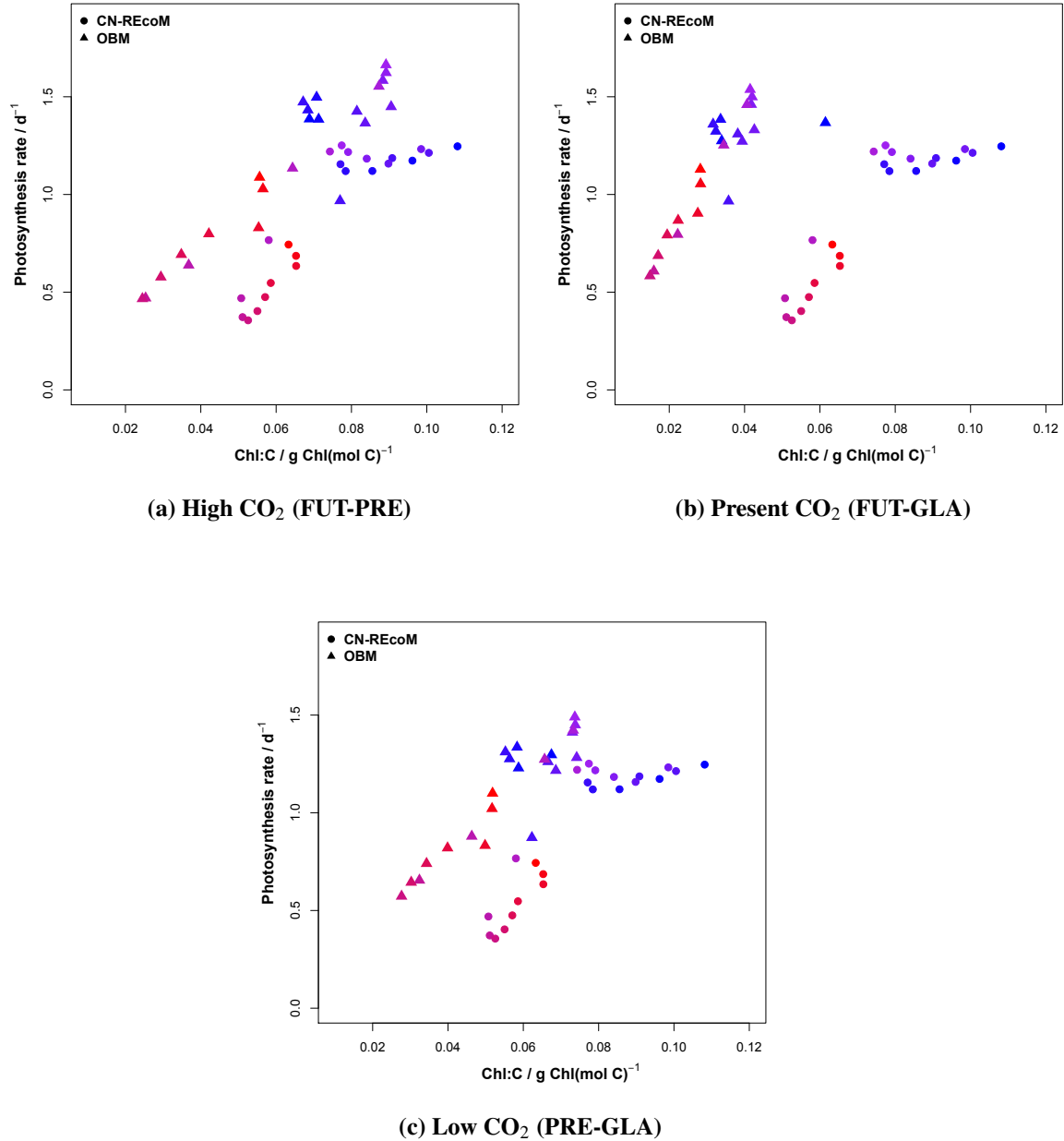


Figure 7: Daily mean values for cellular Chl:C vs carbon fixation rates estimated by models for mesocosms with CO_2 treatment, high (a), present (b) and low (c). The colour palette (blue to red) depicts daily variations in these ratios over the period of the experiment, starting from the day 1 (blue) to the day 23 (red).

For the FUT-GLA case, the OBM predicts high estimates of α , which means algal

cells are able to achieve high photosynthesis rates at low chlorophyll concentrations. Simulated phytoplankton exudation rates by the OBM are also higher for the FUT-GLA mesocosms compared to the PRE-GLA and FUT-PRE ones, which applies to TEPC production and POC concentrations as well. Apparently, the cost function finds high values for photosynthetic efficiencies and exudation rates to produce a good ensemble fit to observed accumulation of POC in the FUT-GLA mesocosms (Fig. 4b). But, by doing so it generates misfits in Chlorophyll as Chl:C ratios are underestimated. However, this trade-off in the cost function works well for the FUT-GLA solutions of the OBM as it reduces the overall discrepancy between data and model integrated over the time period of the experiment. More details on this are discussed below.

4. Discussion

A novel data assimilation approach is applied in this study to evaluate skills of two plankton models, CN-REcoM and OBM, with data from mesocosms treated with different CO₂ concentrations. We want to assess whether neglect of CO₂ dependency in models is expressed in differences between parameter estimates. Ideally, we would have obtained the similar optimised parameter values and associated uncertainty ranges for both models in the respective CO₂ case. However, our results reflect a more complex picture. Uncertainties in initial condition parameters are larger for the OBM compared to the CN-REcoM. But physiological parameters are better constrained for the OBM in all three CO₂ cases. Furthermore, a systematic pattern is seen in the estimates of Q_{min} and V_C^0 for solutions of the OBM between the cases. Owing to the fact that we assimilated data of three different CO₂ treatments in our models, it may be possible to attribute variations in the estimates of Q_{min} and V_C^0 to potential CO₂ effects. However, interestingly these effects are captured only by the OBM and not by the CN-REcoM.

4.1. Potential CO₂ effects on phytoplankton growth dynamics

The most striking result of our study is that estimates of Q_{min} and V_C^0 show a dependency on CO₂ concentrations for the OBM. According to MCMC results, the highest optimised values of V_C^0 and Q_{min} are obtained for the FUT-PRE mesocosms and lowest for the PRE-GLA case. Although we do not explicitly resolve CO₂ effects in the OBM, these results are consistent with findings of previous studies concerning impacts of OA on marine phytoplankton (e.g., Burkhardt et al. 1999; Riebesell et al. 2007; Gao et al. 2012). Burkhardt et al. (1999) and Gervais and Riebesell (2001) have observed an increase in rates of carbon fixation in diatom culture in response to elevated CO₂ levels. Likewise, Shi et al. (2009) and Barcelos e Ramos et al. (2010) have found stimulating effects of ocean acidification on coccolithophores. In addition, Riebesell et al. (2007) noted that primary production increased by $\approx 40\%$ in mesocosms treated with high CO₂. A major source of energy expenditure in phytoplankton is the active transport of inorganic carbon across the membranes through the CO₂ concentrating mechanisms (CCMs) (Hopkinson et al., 2011). It is believed that in an acidic environment many phytoplankton species can effectively down-regulate CCMs and use the energy saved for carbon fixation and growth (Raven and Johnston, 1991; Gao et al., 2012). This may explain high estimates of V_C^0 for phytoplankton in high CO₂ mesocosms, as predicted by the OBM.

Interestingly, OBM also predicts high estimates of Q_{min} for phytoplankton in FUT-PRE mesocosms and lowest for the FUT-GLA case. It appears that in the high- CO_2 mesocosms photoautotrophs may invest a higher fraction of cellular resources in cellular maintenance to sustain life than those in other mesocosms. The effect of high maintenance costs is apparently antagonistic to the positive effect of enhanced carbon fixation rates on algal growth in the high- CO_2 mesocosms. However, this situation is not unrealistic. Gao et al. (2012) investigated responses of marine primary producers to marine stressors (including OA). They found that diatoms exposed to high CO_2 conditions show enhanced growth rates but also increased maintenance costs of pH homeostasis. Schlüter et al. (2016) stressed that energetic costs of maintenance in phytoplankton are expected to rise in response of OA. Likewise for coccolithophores, with decrease in pH and concentration of CO_3^{2-} ions the total energetic cost of calcification (including maintenance costs) is believed to increase (Monteiro et al., 2016).

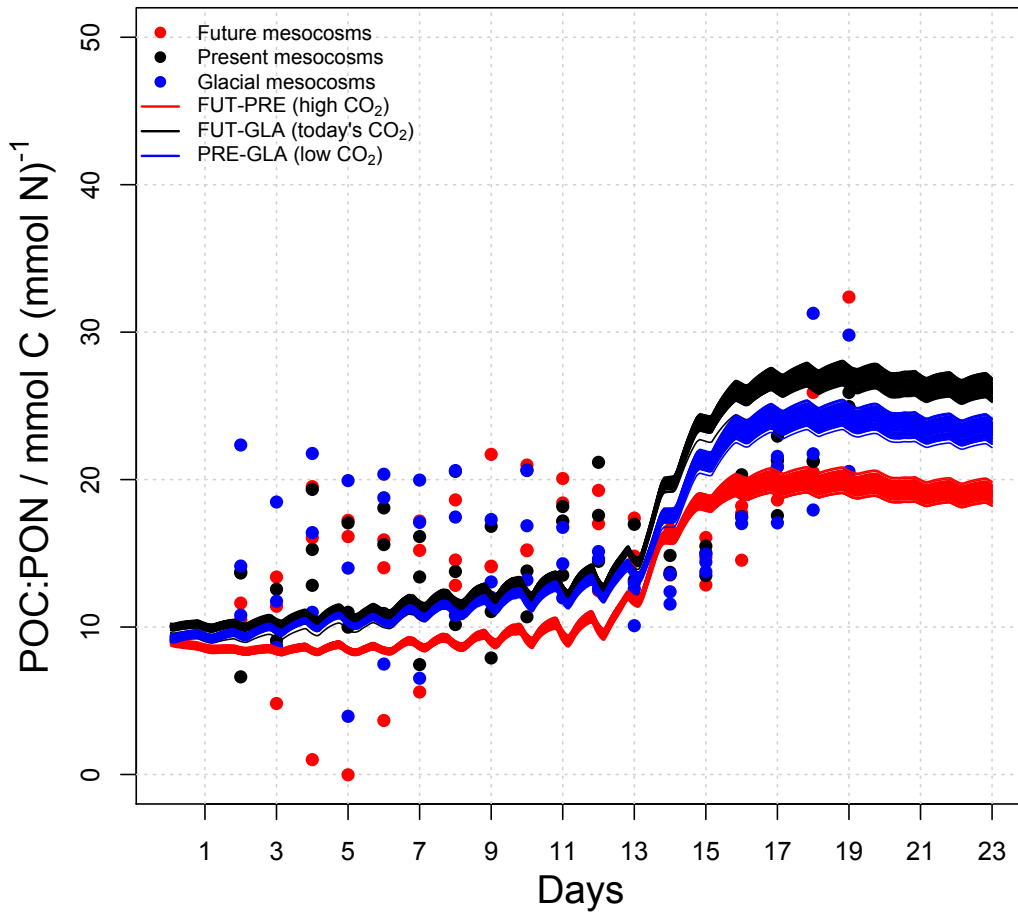


Figure 8: Comparison of simulated POC:PON ratios by the OBM with corresponding observations for the three CO_2 treatments.

As noted above, the OBM finds high carbon fixation rates for phytoplankton in

the FUT-PRE mesocosms. However, the predicted POC:PON ratios are lowest for the mesocosms from the high-CO₂ treatment, whereas the observed POC:PON ratios are highest (see Fig. 8). This discrepancy is the result of the more severe underestimation of POC in the FUT-PRE than in the other treatments. A possible explanation may be that phosphorus limitation is not included in the OBM and hence production of PCH-HOs and subsequently TEPC as cells enter P exhaustion phase is not captured by the model. Furthermore, studies have shown elevated TEPC production in phytoplankton when exposed to high CO₂ concentration (Engel et al., 2004; Riebesell et al., 2007). Since the OBM does not resolve any CO₂ effect on exudation and TEPC formation, this could be another reason for underestimation of POC.

On the one hand, the OBM gives us some insight into CO₂ effects on phytoplankton growth dynamics when evaluated with mesocosm data on OA. On the other hand, for CN-REcoM we do not see any pattern in estimates of Q_{min} and V_C^0 between the CO₂ treatments. Although optimised values of Q_{min} increase with the CO₂ concentrations, the estimates obtained for FUT-GLA and PRE-GLA are unrealistically low and thus spurious.

4.2. Plastic vs. rigid model behaviours

A rigorous model skill assessment should be based on several criteria. While the ability to reproduce observations is an important skill, it is equally important to assess parameter uncertainties and robustness of the model, as well as the plausibility of the parameter estimates obtained during calibration. For example, the CN-REcoM fits most of the data quite well, but the estimates of its parameters (e.g. Q_{min} and α) are unrealistic with large posterior errors. Friedrichs et al. (2007) stressed that models with a large number of unconstrained parameters may have lower predictive ability. Such models exhibit a high degree of plasticity (or flexibility) in their parameter estimates which may result from structural deficiencies in these models (Ward et al., 2010; Schartau et al., 2017). For instance, Q_{min} and α are poorly constrained for the CN-REcoM, but the same parameters are well constrained for the OBM with the identical data. In this case, inappropriate parameterization of processes in the CN-REcoM seems to be the reason for poor parameter estimates.

Posterior estimates of α and Q_{min} for CN-REcoM seem unrealistic as these values fall outside of their typical ranges inferred from previous modelling studies that employed similar parameterisations. CN-REcoM predicts much higher estimates of α than the OBM in all three CO₂ cases. For such high values of α , a model should ideally yield low Chl:C ratios and high primary production rates for algal cells (Lewis and Smith, 1983). However this is not the case, as we obtain low carbon fixation rates and high Chl:C ratios for the CN-REcoM (Fig. 7). Thus, the behaviour of CN-REcoM is not coherent with its parameter estimates. It appears that CN-REcoM contains a structural uncertainty in its parameterisations that include Q_{min} and α because the same parameters are well constrained for the OBM by the identical data.

Modelling studies before have highlighted structural deficiencies in the model of Geider et al. (1998) for phytoplankton acclimation (Flynn et al., 2001; Armstrong, 2006; Smith and Yamanaka, 2007). Flynn et al. (2001) calibrated four plankton models with culture data of Anning et al. (2000) and found significant shortcomings in the parameterization for N assimilation in the Geider model. Likewise, Armstrong (2006)

pointed out that some assumptions of the Geider model can not reproduce the data of [Laws and Bannister \(1980\)](#) and hence may yield unreliable estimates of Chl:C and N:C ratios. [Smith and Yamanaka \(2007\)](#) validated the Geider-based (G) model and the optimality plankton growth (P) model of [Pahlow \(2005\)](#) with the data of [Flynn et al. \(1994\)](#) and reported an unexpected behaviour of the G model in predicting Chl:C ratios. The CN-REcoM is also based on assumptions of [Geider et al. \(1998\)](#), and hence we cannot rule out deficiencies in the model formulations.

Optimised values of α and Q_{min} predicted by the OBM are comparable to estimates obtained in previous studies for the same parameters. In this regard, the OBM seems to display a rigid behaviour as it imposes strong constraints on estimates of its parameters during the calibration with data. However, this may make the OBM more skilled than CN-REcoM because it yields solutions for the CO₂ cases by finding parameter estimates that are well in the range of their uncertainties as inferred from preceding studies.

4.3. Model biases

Biases in multivariate model outputs can result from either structural deficiencies in the model or from trade-offs in the metric we employ to calibrate the model. However, since data have associated uncertainties as well, it is not trivial to analyse and assess model skill based on model-data discrepancies ([Allen et al., 2007a](#)).

Our data assimilation results reveal a few biases in both models. Some of these biases persist in all three CO₂ cases irrespective of models, whereas others are specific to a particular treatment. In the following, we disentangle biases due to uncertainty in model parameterisations, mass balance problem in the data and trade-offs in the cost function. A common bias that can be identified in solutions of the CN-REcoM and OBM is the underestimation of observed chlorophyll in the pre-bloom period between days 1-4. During the initial phase of the PeECE-I experiment some growth of picoplankton was observed ([Engel et al., 2005](#)). However, this initial growth is not captured by models as they neither resolve phytoplankton size spectrum nor functional groups. Thus, structural over-simplification of the phytoplankton compartment in the CN-REcoM and OBM could be the reason for both models to underestimate chlorophyll concentrations at the beginning of the experiment. A similar bias was highlighted in [Krishna and Schartau \(2017\)](#).

Both models predicts slower uptake of DIN in all three cases. As highlighted above, exponential drawdown in observed DIN and build-up of PON data do not occur in the same phase. In other words, the peak in observed PON is seen 1-2 days after the exhaustion of DIN. This may be occur due to: 1) some early loss of nitrogen via DON, 2) early grazing on phytoplankton and 3) plankton mortality due to bacterial infection. However, in both models the concentration of PON peaks on the same day when DIN is fully depleted (Figs. 4a, 4b, 4c, 5a, 5b and 5c), suggesting mass balance is maintained. This induces a discrepancy between observed DIN and PON and their model counterparts, especially during the exponential growth phase. Thus, when models are fitted to PON data, misfits in DIN are introduced, and vice versa.

In the FUT-GLA mesocosms, the OBM simulates lower chlorophyll concentrations than observed. This underestimation of chlorophyll concentration can be explained by low rates of chlorophyll synthesis and high exudation rates for phytoplankton in the

FUT-GLA mesocosms. The OBM finds higher estimates of α and γ for the FUT-GLA case than the other two. High values of α are typically found for low-light adapted phytoplankton as they are capable of utilising low light intensities to achieve high carbon fixation rates at low chlorophyll concentrations (Falkowski, 1980). Thus, there could be a possibility of self-shading effects in the FUT-PRE mesocosms according to the OBM. However, the same inference cannot be derived from observations. The OBM also predicts high values for γ which means higher exudation of DOC and PCHHOs, and also enhanced TEPC production that eventually becomes part of POC. The OBM reproduces POC observations quite well for the FUT-GLA case. We computed percentage model bias, Pbias (Allen et al., 2007b) for the OBM which is a measure of systematic under- or overestimation of observations. Interestingly, we found highest Pbias value for chlorophyll (62%) and lowest for POC (0.4%). Therefore, it is likely that the OBM predicts high estimates of α and γ to yield a good fit to observed POC and other variables except chlorophyll, and hence lowers the overall model-data misfit for the FUT-GLA case.

5. Conclusion

Schartau et al. (2017) highlight that representative model results are those that can explain data. The attraction of our data assimilation (DA) method is that it generates a set of model solutions that explains observations and gives insight into model parameter estimates and their associated uncertainty ranges. We applied our novel DA approach, which accounts for changes in correlations between observed measurement types in the cost function during pre-bloom, bloom and post-bloom phases of the experiment, to calibrate two plankton models, the optimality-based model (OBM) and the Carbon:Nitrogen Regulated Ecosystem Model (CN-REcoM).

Parameters are better constrained for the OBM than the CN-REcoM in three CO₂ treatments. Effects of CO₂ are expressed as differences in parameter estimates between CO₂ treatments. A systematic pattern in estimates of the maximum-rate parameter (V_C^0) and the subsistence quota (Q_{min}) between the three CO₂ treatments is revealed for the OBM. The highest estimates of V_C^0 and Q_{min} are obtained for phytoplankton in the mesocosms that were treated with high CO₂ concentrations and lowest for those in mesocosms with low CO₂ levels. The OBM seems to suggest that ocean acidification may stimulate carbon fixation rates in phytoplankton, but at the cost of increased metabolic stress. However, a firm conclusion can be made only after considering an explicit, mechanistic CO₂ dependency in the OBM. Therefore, for future similar studies we suggest to include parameterisations in the OBM that resolve CO₂ effects on nutrient uptake and algal growth.

By contrast, we do not see a clear relationship between parameter estimates and CO₂ concentrations for the CN-REcoM, suggestive of potential CO₂ effects. In addition, most of the parameters are poorly constrained for the CN-REcoM and some of their estimates seem spurious. Our cross-validation results show that OBM solutions are more robust than those of CN-REcoM as they yield lower misfits when compared with independent data. Although our results show that the OBM underestimates POC:PON ratios compared to those observed, in our assessment of two plankton models we find that the OBM is more skilled than CN-REcoM.

Appendix A. Identical model compartments

Appendix A.1. Zooplankton

For both models, the grazing response of the zooplankton community is according to Holling type III (Morozov, 2010), with no distinction between micro- and mesozooplankton. Any loss in zooplankton biomass (carbon and nitrogen) is accounted for by mortality. Zooplankton homeostasis in C and N stoichiometry is approximately maintained by restoring a constant N:C ratio (Q_{zoo}^{const}) of 0.19, with a restoring time (τ) of 24 hours. We assume that respiration and excretion by zooplankton form the means for restoring towards Q_{zoo}^{const} . The buildup and decline of zooplankton biomass depend on prey concentration (grazing) and losses due to mortality, respiration and excretion. In contrast to Joassin et al. (2011), the OBM and the CN-REcoM do not explicitly resolve viral or bacterial biomass. In both models the microbial activity is implicitly considered to be part of remineralisation and hydrolysis. Thus, the rates of remineralisation and hydrolysis depend on substrate concentration, and are assumed to be independent of bacteria biomass. The differential equations (ODEs) for zooplankton C and N concentrations are given as:

$$\frac{d}{dt}ZooC = \frac{G}{Q^N} - r_{zoo} - \frac{M_{zoo}}{Q_{zoo}} \quad (A.1)$$

$$\frac{d}{dt}ZooN = G - \gamma_{zoo}^N - M_{zoo} \quad (A.2)$$

Where r_{zoo} is the zooplankton respiration in unit $\text{mmol C m}^3 \text{ d}^{-1}$ and Q_{zoo} is the N:C ratio of zooplankton in $\text{mol N (mol C)}^{-1}$. γ_{zoo}^N and M_{zoo} is the nitrogen-specific zooplankton excretion and mortality in $\text{mmol N m}^3 \text{ d}^{-1}$ (see Table A.1).

Appendix A.2. Detritus

Detritus comprises all non-living particulate organic matter (Fasham et al., 1990). Aggregation loss of photoautotrophs is a quadratic function of nitrogen biomass. The aggregation equation resolves interactions between two types of particles: (a) aggregation of cells of photoautotrophs and (b) aggregation of small algal cells with large detritus. We assume that temperature- and substrate-dependent hydrolysis is responsible for the degradation of detritus. Detritus disintegration acts as a source for labile DOC (l DOC) and DON (l DON). Respective ODEs of detrital C and N mass are given as:

$$\frac{d}{dt}DetC = \frac{A}{Q^N} + \frac{M_{zoo}}{Q_{zoo}} - \omega_{det} \cdot T_f \cdot DetC \quad (A.3)$$

$$\frac{d}{dt}DetN = A + M_{zoo} - \omega_{det} \cdot T_f \cdot DetN \quad (A.4)$$

With ω_{det} as the hydrolysis/degradation rate of detritus in d^{-1} (Table A.1).

Krishna and Schartau (2017) showed that a discrepancy occurs between simulated and observed Chl *a* during the post-bloom period when detrital chlorophyll is not taken into account. Hence, we explicitly resolve the detrital chlorophyll. As for N and C, the source for detrital chlorophyll pigment is phytoplankton aggregation.

$$\frac{d}{dt}\text{DetChl} = A \cdot \theta - R_{ref} \cdot \text{Chl} \quad (\text{A.5})$$

Appendix A.3. Dissolved inorganic carbon (DIC)

Since the OBM allows for respiration by algae at night, the net carbon fixation may become negative in the absence of light. In both models no differentiation have been made between the utilisation of carbonate, bicarbonate ions and aqueous CO₂. The source-minus-sink (sms) equation for DIC is:

$$\frac{d}{dt}\text{DIC} = -\mu \cdot \text{PhyC} + r_{zoo} + \rho \cdot T_f \cdot (\text{LDOC} + \text{dCCHO}) + F_{\text{DIC}} \quad (\text{A.6})$$

Where ρ is the remineralisation rate of dissolved organic matter in d⁻¹ (Table A.1) and F_{DIC} is the flux due to air-sea gas exchange and has unit mmol C m⁻³d⁻¹.

Appendix A.4. Dissolved inorganic nitrogen (DIN)

The DIN pool in the OBM and CN-REcoM is represented by the total concentration of nitrate (NO₃⁻¹), nitrite (NO₂⁻¹) and ammonium (NH₄⁻¹). However, the contributions of individual constituents are not resolved. The differential equation for DIN uptake is:

$$\frac{d}{dt}\text{DIN} = -(V_C^N \cdot \text{PhyC}) + \gamma_{zoo}^N + \rho \cdot T_f \cdot \text{LDON} \quad (\text{A.7})$$

Appendix A.5. Total alkalinity (TA)

Since we do not resolve calcification, the only source of TA in both models is DIN and DIP (dissolved inorganic phosphorus) uptake by phytoplankton. Since models do not explicitly resolve sources and sinks of DIP, we assume a fixed stoichiometry ($\Delta\text{DIP} = \frac{1}{6} \cdot \Delta\text{DIN}$) to derive the contributions of changes in DIP from those in DIN. Additionally, we assume that the one half of the N excretion by zooplankton constitutes NH₄⁻¹ and the other half NO₃⁻¹ and NO₂⁻¹. Thus, the net TA change is zero (Krishna and

Schartau, 2017). Temperature-driven remineralisation of LDON and LDOP is the only process that decreases total alkalinity of the system in our study. The ODE for TA changes is:

$$\frac{d}{dt}TA = (1 + 1/16) \cdot \frac{V_C^N}{Q^N} \cdot \text{PhyN} - (1 + 1/16) \cdot \rho \cdot T_f \cdot \text{LDON} \quad (\text{A.8})$$

Appendix A.6. Dissolved organic matter (DOM) and TEP

DOM comprise DOC and DON, and produced by exudation of carbon and nitrogen by phytoplankton and hydrolysis of detrital matter. In both the models, sinks of DOC consist of residual labile dissolved organic carbon (LDOC) and dCCHO. The temperature-dependent remineralisation process acts as a sink for LDOC and LDON.

$$\begin{aligned} \frac{d}{dt}LDOC = & \gamma \cdot (1 - f_{\text{dCCHO}}) \cdot \text{PhyC} + \omega_{\text{det}} \cdot T_f \cdot \text{DetC} \\ & + \omega_{\text{gel}} \cdot T_f \cdot \text{TEPC} - \rho \cdot T_f \cdot LDOC \end{aligned} \quad (\text{A.9})$$

$$\frac{d}{dt}LDON = \gamma \cdot \text{PhyN} + \omega_{\text{det}} \cdot T_f \cdot \text{DetN} - \rho \cdot T_f \cdot LDON \quad (\text{A.10})$$

Where f_{dCCHO} is the fraction of exudates assigned to dCCHO (Krishna and Schartau, 2017). ω_{gel} is the hydrolysis/degradation rate of TEPC in d^{-1} . We estimate a temporally varying fraction of DOC exudates (f_{dCCHO}) that enters the dCCHO pool from Q^N :

$$f_{\text{dCCHO}} = \left[1 + p_{\text{dCCHO}} \cdot \exp(1 - Q_s/Q^N) \right]^{-1} \quad (\text{A.11})$$

With Q_s being the N quota attached with structural proteins in $\text{mol C (mol N)}^{-1}$ and p_{dCCHO} is the fraction of DOC allocated to dCCHO (see Table A.1). It is to note that dCCHO act as precursors for the formation of TEPs. In other words TEP production is the sink for dCCHO pool. The corresponding differential equation for dCCHO is:

$$\begin{aligned} \frac{d}{dt}dCCHO = & \gamma \cdot f_{\text{dCCHO}} \cdot \text{PhyC} - \phi_{\text{dCCHO}} \cdot dCCHO^2 - \phi_{\text{TEP}} \cdot dCCHO \cdot \text{TEPC} \\ & - \rho \cdot T_f \cdot dCCHO \end{aligned} \quad (\text{A.12})$$

Where ϕ_{dCCHO} and ϕ_{TEPC} are the coagulation parameters (Krishna and Schartau, 2017).

TEPC is explicitly resolved in models. The fluctuations in TEPC production can induce variations in the C:N ratio of particulate organic carbon (POC), especially during nutrient-limited periods (Engel et al., 2004; Schartau et al., 2007). Hence, TEPC is

a critical constituent of POC measurements (Verdugo et al., 2004). The formation of TEPC is described by coagulation of dCCHOs and aggregation of dCCHO with TEPC (Engel et al., 2004; Schartau et al., 2007). The mathematical equation describing TEPC production is following:

$$\frac{d}{dt} \text{TEPC} = \phi_{\text{dCCHO}} \cdot \text{dCCHO}^2 + \phi_{\text{TEPC}} \cdot \text{dCCHO} \cdot \text{TEPC} - \omega_{\text{gel}} \cdot T_f \cdot \text{TEPC} \quad (\text{A.13})$$

Parameters	Description	Unit
1) A_E	slope of arrhenius relationship	K
2) a_w	light attenuation due to water column	m^{-1}
3) a_c	light attenuation due to chlorophyll a	$(\text{mg Chla})^{-1} \text{m}^3$
4) ρ	rem mineralisation rate of dissolved organic matter	d^{-1}
5) ω_{det}	hydrolysis/degradation rate of detritus	d^{-1}
6) ω_{gel}	hydrolysis/degradation rate of TEPC	d^{-1}
7) ϕ_{dCCHO}	coagulation parameter of dCCHO	$\text{m}^3 (\text{mmol C})^{-1} \text{d}^{-1}$
8) ϕ_{TEPC}	coagulation parameter of dCCHO-TEPC	$\text{m}^3 (\text{mmol C})^{-1} \text{d}^{-1}$
9) ϕ_{agg}	aggregation rate	$\text{m}^3 (\text{mmol N})^{-1} \text{d}^{-1}$
10) p_{dCCHO}	minimum DOC fraction allocated to dCCHO	
11) g_m	nitrogen specific maximum grazing rate	d^{-1}
12) ϵ	prey capture rate normalised to maximum grazing rate	$(\text{mmol N})^2 \text{m}^{-6}$
13) $Q_{\text{const}}^{\text{zoo}}$	Constant N:C ratio in zooplankton	$\text{mol N} (\text{mol C})^{-1}$
14) M_{zoo}	mortality rate of zooplankton	d^{-1}
15) R_{basal}	zooplankton basal respiration rate	d^{-1}

Table A.1: Additional common parameters for CN-REcoM and OBM

	f_{det}	f_{zoo}	α	γ	Q_0/Q_{min}	V_0^N
OBM						
PON_0	0.43 / 0.54 / 0.37	0.38 / -0.41 / 0.17	-0.19 / 0.52 / 0.12	0.15 / 0.38 / 0.33	0.28 / -0.27 / -0.04	0.43 / -0.16 / 0.36
f_{det}	.	0.0 / -0.88 / -0.24	0.32 / 0.92 / 0.35	-0.1 / -0.13 / -0.17	0.30 / -0.15 / 0.34	0.10 / -0.74 / -0.17
f_{zoo}	.	.	-0.75 / -0.86 / -0.38	-0.33 / 0.19 / 0.27	0.79 / 0.41 / 0.17	0.96 / 0.89 / 0.74
α	.	.	.	0.41 / 0.10 / 0.42	-0.53 / -0.34 / -0.35	-0.70 / -0.70 / -0.01
γ	-0.46 / -0.33 / -0.40	-0.19 / 0.56 / 0.81
Q_0/Q_{min}	0.84 / 0.32 / -0.01

Table A.2: Correlation coefficients between the optimised parameter values of the OBM as estimated by the MCMC method for the three CO₂ cases (FUT-PRE / FUT-GLA / PRE-GLA). Correlation coefficients ≥ 0.6 are marked bold face.

	f_{det}	f_{zoo}	α	γ	Q_0/Q_{min}	V_0^C	V_{fact}
CN-REcoM							
PON_0	0.64 / 0.21 / 0.53	0.40 / -0.31 / 0.10	0.47 / 0.27 / 0.20	-0.36 / 0.29 / 0.14	-0.36 / -0.01 / -0.01	-0.32 / 0.30 / -0.10	0.48 / -0.28 / 0.21
f_{det}	.	0.40 / 0.14 / -0.04	0.38 / 0.04 / 0.29	-0.1 / 0.00 / 0.01	-0.05 / -0.01 / 0.17	-0.33 / 0.03 / 0.01	0.52 / 0.04 / 0.15
f_{zoo}	.	.	0.10 / -0.80 / -0.40	0.18 / -0.71 / -0.04	-0.04 / -0.10 / 0.31	0.02 / -0.81 / 0.38	0.41 / 0.91 / 0.02
α	.	.	.	0.26 / 0.91 / 0.39	-0.54 / 0.04 / -0.20	-0.67 / 0.98 / -0.22	0.65 / -0.90 / 0.12
γ	-0.58 / -0.02 / -0.22	-0.10 / 0.90 / 0.34	0.18 / -0.71 / -0.21
Q_0/Q_{min}	0.73 / 0.05 / 0.65	-0.61 / -0.06 / -0.50
V_0^C	-0.90 / -0.92 / -0.87

Table A.3: Correlation coefficients between the optimised parameter values of CN-REcoM as estimated by the MCMC method for the three CO₂ cases (FUT-PRE / FUT-GLA / PRE-GLA). Correlation coefficients ≥ 0.6 are marked bold face.

References

- Aksnes, D. and Egge, J. (1991). A theoretical model for nutrient uptake in phytoplankton. *Marine ecology progress series*. Oldendorf, 70(1):65–72.
- Allen, J., Somerfield, P., and Gilbert, F. (2007a). Quantifying uncertainty in high-resolution coupled hydrodynamic-ecosystem models. *Journal of Marine Systems*, 64(1):3–14.
- Allen, J. I., Holt, J. T., Blackford, J., and Proctor, R. (2007b). Error quantification of a high-resolution coupled hydrodynamic-ecosystem coastal-ocean model: Part 2. chlorophyll-a, nutrients and SPM. *Journal of Marine Systems*, 68(3):381–404.
- Anning, T., MacIntyre, H. L., Pratt, S. M., Sammes, P. J., Gibb, S., and Geider, R. J. (2000). Photoacclimation in the marine diatom *Skeletonema costatum*. *Limnology and Oceanography*, 45(8):1807–1817.
- Armstrong, R. A. (2006). Optimality-based modeling of nitrogen allocation and photoacclimation in photosynthesis. *Deep Sea Research Part II: Topical Studies in Oceanography*, 53(5):513–531.
- Ayata, S.-D., Lévy, M., Aumont, O., Sciandra, A., Sainte-Marie, J., Tagliabue, A., and Bernard, O. (2013). Phytoplankton growth formulation in marine ecosystem models: should we take into account photo-acclimation and variable stoichiometry in oligotrophic areas? *Journal of Marine Systems*, 125:29–40.
- Barcelos e Ramos, J., Müller, M., and Riebesell, U. (2010). Short-term response of the coccolithophore *Emiliania huxleyi* to an abrupt change in seawater carbon dioxide concentrations. *Biogeosciences*, 7:177.
- Beardall, J., Stojkovic, S., and Larsen, S. (2009). Living in a high CO₂ world: impacts of global climate change on marine phytoplankton. *Plant Ecology & Diversity*, 2(2):191–205.
- Behrenfeld, M. J., Marañón, E., Siegel, D. A., and Hooker, S. B. (2002). Photoacclimation and nutrient-based model of light-saturated photosynthesis for quantifying oceanic primary production. *Marine Ecology Progress Series*, 228:103–117.
- Behrenfeld, M. J., O'Malley, R. T., Siegel, D. A., McClain, C. R., Sarmiento, J. L., Feldman, G. C., Milligan, A. J., Falkowski, P. G., Letelier, R. M., and Boss, E. S. (2006). Climate-driven trends in contemporary ocean productivity. *Nature*, 444(7120):752–755.
- Bélisle, C. J. (1992). Convergence theorems for a class of simulated annealing algorithms on rd. *Journal of Applied Probability*, pages 885–895.
- Bowes, G. (1993). Facing the inevitable: plants and increasing atmospheric CO₂. *Annual review of plant biology*, 44(1):309–332.
- Boyd, P. W. and Doney, S. C. (2002). Modelling regional responses by marine pelagic ecosystems to global climate change. *Geophysical Research Letters*, 29(16).

- Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms 1. general considerations. *IMA Journal of Applied Mathematics*, 6(1):76–90.
- Burkhardt, S. and Riebesell, U. (1997). CO₂ availability affects elemental composition (c: N: P) of the marine diatom *Skeletonema costatum*. *Marine Ecology Progress Series*, pages 67–76.
- Burkhardt, S., Zondervan, I., and Riebesell, U. (1999). Effect of CO₂ concentration on c: N: P ratio in marine phytoplankton: A species comparison. *Limnology and Oceanography*, 44(3):683–690.
- Cebrian, J. and Lartigue, J. (2004). Patterns of herbivory and decomposition in aquatic and terrestrial ecosystems. *Ecological Monographs*, 74(2):237–259.
- Delille, B., Harlay, J., Zondervan, I., Jacquet, S., Chou, L., Wollast, R., Bellerby, R. G., Frankignoulle, M., Borges, A. V., Riebesell, U., et al. (2005). Response of primary production and calcification to changes of pCO₂ during experimental blooms of the coccolithophorid *Emiliana huxleyi*. *Global Biogeochemical Cycles*, 19(2).
- Denman, K. L. (2003). Modelling planktonic ecosystems: parameterizing complexity. *Progress in Oceanography*, 57(3):429–452.
- Dugdale, R. (1967). Nutrient limitation in the sea: dynamics, identification and significance. *Limnol. Oceanogr*, 12(4):685–695.
- Engel, A., Borchard, C., Piontek, J., Schulz, K. G., Riebesell, U., and Bellerby, R. (2013). CO₂ increases 14C-primary production in an arctic plankton community. *Biogeosciences (BG)*, 10(3):1291–1308.
- Engel, A., Cisternas Novoa, C., Wurst, M., Endres, S., Tang, T., Schartau, M., and Lee, C. (2014). No detectable effect of CO₂ on elemental stoichiometry of *Emiliana huxleyi* in nutrient-limited, acclimated continuous cultures. *Marine Ecology Progress Series*, 507:15–30.
- Engel, A., Delille, B., Jacquet, S., Riebesell, U., Rochelle-Newall, E., Terbrüggen, A., and Zondervan, I. (2004). Transparent exopolymer particles and dissolved organic carbon production by *Emiliana huxleyi* exposed to different CO₂ concentrations: a mesocosm experiment. *Aquatic Microbial Ecology*, 34(1):93–104.
- Engel, A., Zondervan, I., Aerts, K., Beaufort, L., Benthien, A., Chou, L., Delille, B., Gattuso, J.-P., Harlay, J., Heemann, C., et al. (2005). Testing the direct effect of CO₂ concentration on a bloom of the coccolithophorid *Emiliana huxleyi* in mesocosm experiments. *Limnology and Oceanography*, 50(2):493–507.
- Evans, G. T. (2003). Defining misfit between biogeochemical models and data sets. *Journal of marine systems*, 40:49–54.
- Falkowski, P. G. (1980). Light-shade adaptation in marine phytoplankton. In *Primary productivity in the sea*, pages 99–119. Springer.

- Fasham, M., Ducklow, H., and McKelvie, S. (1990). A nitrogen-based model of plankton dynamics in the oceanic mixed layer. *Journal of Marine Research*, 48(3):591–639.
- Fennel, K., Losch, M., Schröter, J., and Wenzel, M. (2001). Testing a marine ecosystem model: sensitivity analysis and parameter optimization. *Journal of Marine Systems*, 28(1):45–63.
- Fletcher, R. (1970). A new approach to variable metric algorithms. *The Computer Journal*, 13(3):317–322.
- Flynn, K., Davidson, K., and Leftley, J. (1993). Carbon-nitrogen relations during batch growth of *Nannochloropsis oculata* (eustigmatophyceae) under alternating light and dark. *Journal of Applied Phycology*, 5(4):465–475.
- Flynn, K., Davidson, K., and Leftley, J. (1994). Carbon-nitrogen relations at whole-cell and free-amino-acid levels during batch growth of *Isochrysis galbana* (prymnesiophyceae) under conditions of alternating light and dark. *Marine Biology*, 118(2):229–237.
- Flynn, K. J. (2001). A mechanistic model for describing dynamic multi-nutrient, light, temperature interactions in phytoplankton. *Journal of Plankton Research*, 23(9):977–997.
- Flynn, K. J. (2003). Do we need complex mechanistic photoacclimation models for phytoplankton? *Limnology and Oceanography*, 48(6):2243–2249.
- Flynn, K. J. (2008). The importance of the form of the quota curve and control of non-limiting nutrient transport in phytoplankton models. *Journal of Plankton Research*, 30(4):423–438.
- Flynn, K. J., Marshall, H., and Geider, R. J. (2001). A comparison of two n-irradiance interaction models of phytoplankton growth. *Limnology and Oceanography*, 46(7):1794–1802.
- Franks, P. J. (2009). Planktonic ecosystem models: perplexing parameterizations and a failure to fail. *Journal of Plankton Research*, page fbp069.
- Friedrichs, M. A., Dusenberry, J. A., Anderson, L. A., Armstrong, R. A., Chai, F., Christian, J. R., Doney, S. C., Dunne, J., Fujii, M., Hood, R., et al. (2007). Assessment of skill and portability in regional marine biogeochemical models: Role of multiple planktonic groups. *Journal of Geophysical Research: Oceans*, 112(C8).
- Friedrichs, M. A., Hood, R. R., and Wiggert, J. D. (2006). Ecosystem model complexity versus physical forcing: Quantification of their relative impact with assimilated Arabian Sea data. *Deep Sea Research Part II: Topical Studies in Oceanography*, 53(5):576–600.
- Gao, K., Helbling, E. W., Häder, D.-P., and Hutchins, D. A. (2012). Responses of marine primary producers to interactions between ocean acidification, solar radiation, and warming. *Marine Ecology Progress Series*, 470:167–189.

- Geider, R. and La Roche, J. (2002). Redfield revisited: variability of C:N:P in marine microalgae and its biochemical basis. *European Journal of Phycology*, 37(1):1–17.
- Geider, R. J. (1987). Light and temperature dependence of the carbon to chlorophyll a ratio in microalgae and cyanobacteria: implications for physiology and growth of phytoplankton. *New Phytologist*, 106(1):1–34.
- Geider, R. J., MacIntyre, H. L., and Kana, T. M. (1998). A dynamic regulatory model of phytoplanktonic acclimation to light, nutrients, and temperature. *Limnology and Oceanography*, 43(4):679–694.
- Gervais, F. and Riebesell, U. (2001). Effect of phosphorus limitation on elemental composition and stable carbon isotope fractionation in a marine diatom growing under different CO₂ concentrations. *Limnology and Oceanography*, 46(3):497–504.
- Goldfarb, D. (1970). A family of variable-metric methods derived by variational means. *Mathematics of Computation*, 24(109):23–26.
- Goldman, J. C., McCarthy, J. J., and Peavey, D. G. (1979). Growth rate influence on the chemical composition of phytoplankton in oceanic waters. *Nature*, 279(5710):210–215.
- Hare, C. E., Leblanc, K., DiTullio, G. R., Kudela, R. M., Zhang, Y., Lee, P. A., Riesenman, S., and Hutchins, D. A. (2007). Consequences of increased temperature and CO₂ for phytoplankton community structure in the bering sea. *Marine Ecology Progress Series*, 352:9–16.
- Harley, C. D., Randall Hughes, A., Hultgren, K. M., Miner, B. G., Sorte, C. J., Thornber, C. S., Rodriguez, L. F., Tomanek, L., and Williams, S. L. (2006). The impacts of climate change in coastal marine systems. *Ecology letters*, 9(2):228–241.
- Harmon, R. and Challenor, P. (1997). A Markov chain Monte Carlo method for estimation and assimilation into models. *Ecological Modelling*, 101(1):41–59.
- Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109.
- Healey, F. P. (1985). Interacting effects of light and nutrient limitation on the growth rate of *Synechococcus linearis* (cyanophyceae). *Journal of phycology*, 21(1):134–146.
- Hoffmann, M., Marxen, K., Schulz, R., and Vanselow, K. H. (2010). Tfa and epa productivities of *Nannochloropsis salina* influenced by temperature and nitrate stimuli in turbidostatic controlled experiments. *Marine drugs*, 8(9):2526–2545.
- Hopkinson, B. M., Dupont, C. L., Allen, A. E., and Morel, F. M. (2011). Efficiency of the CO₂-concentrating mechanism of diatoms. *Proceedings of the National Academy of Sciences*, 108(10):3830–3837.

- Hurtt, G. C. and Armstrong, R. A. (1996). A pelagic ecosystem model calibrated with bats data. *Deep Sea Research Part II: Topical Studies in Oceanography*, 43(2-3):653–683.
- Hurtt, G. C. and Armstrong, R. A. (1999). A pelagic ecosystem model calibrated with bats and owsa data. *Deep Sea Research Part I: Oceanographic Research Papers*, 46(1):27–61.
- Iglesias-Rodriguez, M. D., Halloran, P. R., Rickaby, R. E., Hall, I. R., Colmenero-Hidalgo, E., Gittins, J. R., Green, D. R., Tyrrell, T., Gibbs, S. J., von Dassow, P., et al. (2008). Phytoplankton calcification in a high-CO₂ world. *science*, 320(5874):336–340.
- Joassin, P., Delille, B., Soetaert, K., Harlay, J., Borges, A. V., Chou, L., Riebesell, U., Suykens, K., and Grégoire, M. (2011). Carbon and nitrogen flows during a bloom of the coccolithophore *Emiliania huxleyi*: Modelling a mesocosm experiment. *Journal of Marine Systems*, 85(3):71–85.
- Kim, J.-M., Lee, K., Shin, K., Kang, J.-H., Lee, H.-W., Kim, M., Jang, P.-G., and Jang, M.-C. (2006). The effect of seawater CO₂ concentration on growth of a natural phytoplankton assemblage in a controlled mesocosm experiment. *Limnology and oceanography*, 51(4):1629–1636.
- Klausmeier, C. A., Litchman, E., Daufresne, T., and Levin, S. (2008). Phytoplankton stoichiometry. *Ecological Research*, 23(3):479–485.
- Klausmeier, C. A., Litchman, E., and Levin, S. A. (2004). Phytoplankton growth and stoichiometry under multiple nutrient limitation. *Limnology and Oceanography*, 49(4):1463–1470.
- Körtzinger, A., Koeve, W., Kähler, P., and Mintrop, L. (2001). C:N ratios in the mixed layer during the productive season in the northeast atlantic ocean. *Deep Sea Research Part I: Oceanographic Research Papers*, 48(3):661–688.
- Krishna, S. and Schartau, M. (2017). A data-model synthesis to explain variability in calcification observed during a CO₂ perturbation mesocosm experiment. *Biogeo-sciences (BG)*, 14(6):1857–1882.
- Langer, G., Geisen, M., Baumann, K.-H., Kläs, J., Riebesell, U., Thoms, S., and Young, J. R. (2006). Species-specific responses of calcifying algae to changing seawater carbonate chemistry. *Geochemistry, Geophysics, Geosystems*, 7(9).
- Laws, E. A. and Bannister, T. (1980). Nutrient-and light-limited growth of *Thalassiosira fluviatilis* in continuous culture, with implications for phytoplankton growth in the ocean. *Limnology and Oceanography*, 25(3):457–473.
- Lewis, M. R. and Smith, J. C. (1983). A small volume, short-incubation-time method for measurement of photosynthesis as a function of incident irradiance. *Marine Ecology Progress Series*, pages 99–102.

- Marsili-Libelli, S., Guerrizio, S., and Checchi, N. (2003). Confidence regions of estimated parameters for ecological systems. *Ecological Modelling*, 165(2):127–146.
- Martiny, A. C., Pham, C. T., Primeau, F. W., Vrugt, J. A., Moore, J. K., Levin, S. A., and Lomas, M. W. (2013). Strong latitudinal patterns in the elemental ratios of marine plankton and organic matter. *Nature Geoscience*, 6(4):279–283.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092.
- Monod, J. (1949). The growth of bacterial cultures. *Annual Reviews in Microbiology*, 3(1):371–394.
- Monteiro, F. M., Bach, L. T., Brownlee, C., Bown, P., Rickaby, R. E., Poulton, A. J., Tyrrell, T., Beaufort, L., Dutkiewicz, S., Gibbs, S., et al. (2016). Why marine phytoplankton calcify. *Science Advances*, 2(7):e1501822.
- Morozov, A. Y. (2010). Emergence of holling type III zooplankton functional response: bringing together field evidence and mathematical modelling. *Journal of theoretical biology*, 265(1):45–54.
- Orr, J. C., Fabry, V. J., Aumont, O., Bopp, L., Doney, S. C., Feely, R. A., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., et al. (2005). Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms. *Nature*, 437(7059):681–686.
- Pahlow, M. (2005). Linking chlorophyll–nutrient dynamics to the redfield N:C ratio with a model of optimal phytoplankton growth. *Marine Ecology Progress Series*, 287:33–43.
- Pahlow, M., Dietze, H., and Oschlies, A. (2013). Optimality-based model of phytoplankton growth and diazotrophy. *Marine Ecology Progress Series*, 489:1–16.
- Pahlow, M. and Oschlies, A. (2009). Chain model of phytoplankton P, N and light colimitation. *Marine Ecology Progress Series*, 376:69–83.
- Quere, C. L., Harrison, S. P., Colin Prentice, I., Buitenhuis, E. T., Aumont, O., Bopp, L., Claustre, H., Cotrim Da Cunha, L., Geider, R., Giraud, X., et al. (2005). Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biology*, 11(11):2016–2040.
- Raven, J. A. and Johnston, A. M. (1991). Mechanisms of inorganic-carbon acquisition in marine phytoplankton and their implications for the use of other resources. *Limnology and Oceanography*, 36(8):1701–1714.
- Rhee, G., Gotham, I. J., et al. (1981). The effect of environmental factors on phytoplankton growth: Temperature and the interactions of temperature with nutrient limitation. *Limnology and Oceanography*, 26(4):635–648.

- Riebesell, U., Bellerby, R., Grossart, H.-P., and Thingstad, F. (2008). Mesocosm CO₂ perturbation studies: from organism to community level. *Biogeosciences (BG)*, 5(4):1157–1164.
- Riebesell, U., Schulz, K. G., Bellerby, R., Botros, M., Fritsche, P., Meyerhöfer, M., Neill, C., Nondal, G., Oschlies, A., Wohlers, J., et al. (2007). Enhanced biological carbon consumption in a high CO₂ ocean. *Nature*, 450(7169):545–548.
- Riebesell, U. and Tortell, P. D. (2011). Effects of ocean acidification on pelagic organisms and ecosystems. *Ocean acidification. Oxford University Press, Oxford*, pages 99–121.
- Saito, M. A., Goepfert, T. J., and Ritt, J. T. (2008). Some thoughts on the concept of colimitation: Three definitions and the importance of bioavailability. *Limnology and Oceanography*, 53(1):276–290.
- Schartau, M., Engel, A., Schröter, J., Thoms, S., Völker, C., and Wolf-Gladrow, D. (2007). Modelling carbon overconsumption and the formation of extracellular particulate organic carbon. *Biogeosciences*, 4(4):433–454.
- Schartau, M., Wallhead, P., Hemmings, J., Löptien, U., Kriest, I., Krishna, S., Ward, B. A., Slawig, T., and Oschlies, A. (2017). Reviews and syntheses: Parameter identification in marine planktonic ecosystem modelling. *Biogeosciences (BG)*, 14(6):1647–1701.
- Schlüter, L., Lohbeck, K. T., Gröger, J. P., Riebesell, U., and Reusch, T. B. (2016). Long-term dynamics of adaptive evolution in a globally important phytoplankton species to ocean acidification. *Science advances*, 2(7):e1501660.
- Sciandra, A., Harlay, J., Lefèvre, D., Lemée, R., Rimmelin, P., Denis, M., and Gattuso, J.-P. (2003). Response of coccolithophorid *Emiliana huxleyi* to elevated partial pressure of CO₂ under nitrogen limitation. *Marine Ecology Progress Series*, 261:111–122.
- Shanno, D. F. (1970). Conditioning of quasi-Newton methods for function minimization. *Mathematics of Computation*, 24(111):647–656.
- Shi, D., Xu, Y., and Morel, F. (2009). Effects of the pH/pCO₂ control method on medium chemistry and phytoplankton growth. *Biogeosciences*, 6(7):1199–1207.
- Skartveit, A., Cleveland, F., and de Lange, T. (2001). Radiation Yearbook no. 37, Meteorological Report Series, University of Bergen. Technical report, University of Bergen Geophysical Institute, Bergen, Norway.
- Smith, S. L., Pahlow, M., Merico, A., Acevedo-Trejos, E., Sasai, Y., Yoshikawa, C., Sasaoka, K., Fujiki, T., Matsumoto, K., and Honda, M. C. (2016). Flexible phytoplankton functional type (flexpft) model: size-scaling of traits and optimal growth. *Journal of Plankton Research*, 38(4):977–992.

- Smith, S. L. and Yamanaka, Y. (2007). Quantitative comparison of photoacclimation models for marine phytoplankton. *ecological modelling*, 201(3):547–552.
- Smith, S. L., Yamanaka, Y., Pahlow, M., and Oschlies, A. (2009). Optimal uptake kinetics: physiological acclimation explains the pattern of nitrate uptake by phytoplankton in the ocean. *Marine Ecology Progress Series*, 384:1–12.
- Soetaert, K. and Petzoldt, T. (2010). Inverse modelling, sensitivity and Monte Carlo analysis in R using package FME. *Journal of Statistical Software*, 33.
- Sommer, U., Aberle, N., Engel, A., Hansen, T., Lengfellner, K., Sandow, M., Wohlers, J., Zöllner, E., and Riebesell, U. (2007). An indoor mesocosm system to study the effect of climate change on the late winter and spring succession of baltic sea phyto- and zooplankton. *Oecologia*, 150(4):655–667.
- Spitz, Y., Moisan, J., and Abbott, M. (2001). Configuring an ecosystem model using data from the Bermuda Atlantic time series (BATS). *Deep Sea Research Part II: Topical Studies in Oceanography*, 48(8):1733–1768.
- Sterner, R. W. (1990). The ratio of nitrogen to phosphorus resupplied by herbivores: zooplankton and the algal competitive arena. *The American Naturalist*, 136(2):209–229.
- Urabe, J. and Sterner, R. W. (1996). Regulation of herbivore growth by the balance of light and nutrients. *Proceedings of the National Academy of Sciences*, 93(16):8465–8469.
- Vallino, J. (2000). Improving marine ecosystem models: use of data assimilation and mesocosm experiments. *Journal of Marine Research*, 58(1):117–164.
- Verdugo, P., Alldredge, A. L., Azam, F., Kirchman, D. L., Passow, U., and Santschi, P. H. (2004). The oceanic gel phase: a bridge in the DOM–POM continuum. *Marine Chemistry*, 92(1):67–85.
- Ward, B. A., Friedrichs, M. A., Anderson, T. R., and Oschlies, A. (2010). Parameter optimisation techniques and the problem of underdetermination in marine biogeochemical models. *Journal of Marine Systems*, 81(1):34–43.

Chapter 5

Summary & conclusive discussion

This thesis has addressed two important aspects of oceanography that are related to: 1) the response dynamics of marine planktonic ecosystems to OA, and 2) estimation of model parameters with a data assimilation (DA) approach. On the one hand, the thesis provides some valuable insights into physiological and community level responses of marine phytoplankton to OA. On the other hand, it highlights novel DA approaches that facilitate parameter identification in marine ecosystem models and maximise their predictive capabilities.

Model-based data analyses of an OA experiment were performed, in which observations from a mesocosm experiment were interpreted with different dynamical model approaches. The plankton models were calibrated with these data and were applied to investigate CO₂ effects on phytoplankton growth dynamics in the mesocosms. The analyses were somewhat different from classical statistical analyses of the data. With the definition of the model equations we become explicit about the underlying dynamics of the system. Two model approaches (and three model versions) have been applied in this thesis. The first model approach (an optimality-based model) resolves OA effects on calcification by coccolithophores (see Chapter 2). With this model approach the variability observed in TA and calcification during the PeECE-I experiment is investigated. In the next study the performance of two plankton models is assessed against data from the same mesocosm experiment. These models differ with respect to their description of nutrient uptake and phytoplankton growth. Intentionally, no OA effect was considered in both models in this particular analysis. The idea was to test whether this neglect of any CO₂ dependency is expressed in estimates of the physiological parameters when fitted with data of different CO₂ treatments.

All mesocosms have been assumed to be homogeneously mixed. Therefore the applied models do not resolve vertical transport of organic matter. However, these 0-D model approaches describe all important physio-ecological processes of a marine ecosystem and allow the estimation of mass flux through a planktonic ecosystem.

Data assimilation (DA) methods were used for a synthesis of mesocosm observations and model results, as well as for the optimisation of model parameter values, along with their uncertainties. The cost functions applied in these methods account for changes in correlations between observations during different phases of phytoplankton growth (e.g., pre-bloom, bloom and post-bloom periods). This approach has not been applied in any data assimilation studies so far. DA methods helped to successfully reproduce most of the observed data and give some valuable insights into physiological and ecological aspects of phytoplankton dynamics in the mesocosms (e.g. cellular level responses to elevated CO₂ levels and variability in calcification and accumulation of POC at

the community level). In the following synthesis, key results of this thesis are presented.

5.1 Effects of initial conditions on physio-ecological processes in mesocosm experiments

In mesocosm experiments initial conditions (ICs) play a critical role as these determine the food web dynamics in the mesocosms and also affect the estimates of mass fluxes that are exchanged at different trophic levels (Vallino, 2000). Likewise, marine ecosystem models that simulate mass flux at local oceanic sites are also sensitive of initial conditions and physical forcings (e.g. Gibson and Spitz 2011).

In Chapter 3 an optimality-based model has been applied to find the source of variability observed in calcification and TA between mesocosms (especially those from the same CO₂ treatment). To answer this question a data assimilation approach has been designed in which mesocosms are sorted into three groups depending on rates of calcification, and data from these mesocosms are assimilated in three separate sets (cases) of model simulations (see Chapter 3). These three groups of distinct calcification rates happened to be exposed to different CO₂ conditions. This way we could separate the initial condition problem from a possible CO₂ effect on calcification. Data assimilation method yields an ensemble of model solutions for respective cases. The spread of these solutions captures the large part of observed variability in data (particularly PIC and TA). With small variations in initial biomass of coccolithophores the model is able to reproduce most of the variability in PIC and TA data (see Figs. 5 and 14 in Chapter 3). Thus, according to the model, the large variability that was observed in calcification and TA in the PeECE-I could have been generated due to small differences in initial abundance of coccolithophores and respective photo-acclimation states during initialisation (filling) of mesocosms.

Eggers et al. (2014) stressed the relevance of initial plankton community composition in driving dynamics of planktonic ecosystems. They found that initial abundance of phytoplankton had a greater impact than OA on the standing stock biomass at the end of the experiment. Results obtained in Chapter 3 reveal similar finding. It has been shown that the variability generated due to uncertainties in ICs is much larger than the simulated CO₂ effect (Fig. 14 in Chapter 3). Eggers et al. (2014) also stressed that initial ratio between algal species (e.g., cyanobacteria, diatoms, and dinoflagellates) is the key to the outcome of competitive dynamics. This corroborates the model results obtained in Chapter 3. According to the model, in mesocosms with high calcification rates coccolithophores outcompete bulk phytoplankton as their initial abundance is higher, and vice versa for low calcification mesocosms.

5.2 CO₂ effects on cellular physiology of photoautotrophs and community level dynamics

To discriminate a CO₂ response signal from all the other factors that controlled/affected the calcification rates in the PeECE-I experiment, a CO₂ dependency of calcification has been adopted from Findlay et al. (2011) and

applied to an optimality-based model (Chapter 3). The model suggests that the effect of CO₂ on calcification (accumulation of PIC) and TA is small. To obtain a weak OA response signal on calcification is not unrealistic. There are studies (e.g., [Buitenhuis et al. 1999](#); [Langer et al. 2006](#); [Rickaby et al. 2010](#); [Engel et al. 2014](#)) that show no conclusive evidence of stimulating effects of OA on calcification.

In contrast to the calcification study, no CO₂ dependency is considered in model approaches (OBM and CN-REcoM) applied in Chapter 4. Neither of these models explicitly resolve CO₂ effects on algal growth dynamics. They are employed to investigate whether there are any systematic variations in estimates of model parameters between mesocosms of different CO₂ treatments that can be related to potential OA effects. This is achieved with a data assimilation approach in which three sets of simulations have been set up for both models, mimicking three different CO₂ treatments. Results obtained for the OBM show high estimates of parameters V_0^C (potential maximum photosynthesis rate) and Q_{min} (subsistence quota) for phytoplankton in mesocosms with high CO₂ and vice versa for those in low CO₂ mesocosms. However, no such relationship is seen in parameter estimates of the CN-REcoM.

As noted above, the pattern in the estimates of V_0^C and Q_{min} between CO₂ treatments (in the neglect of CO₂ dependency) is obtained only for the OBM. A possible reason for this could be that the OBM accounts for the optimal allocation of resources between nutrient uptake and light harvesting functions of a cell, which is not the case for the CN-REcoM.

Results obtained in Chapter 4 support the OBM to be more “skilful” than the CN-REcoM. Owing to this fact, differences in parameter estimates of the OBM can be attributed to possible CO₂ effects. From an ecological perspective, the OBM seems to suggest that OA may stimulate carbon fixation rates in phytoplankton, but at the expense of elevated metabolic costs. This is supported by findings of previous studies that investigated OA effects on physiology of marine phytoplankton ([Zondervan et al., 2002](#); [Riebesell et al., 2007](#); [Gao et al., 2012](#); [Monteiro et al., 2016](#)). [Gao et al. \(2012\)](#) studied effects of OA along with other stressors on growth dynamics of phytoplankton. Their results show enhanced primary production rates and increased maintenance costs of pH homeostasis in algae treated with high CO₂. [Monteiro et al. \(2016\)](#) highlighted that energy demand for calcification in coccolithophores, that includes transport, metabolic, structure and maintenance costs, is believed to increase in acidic conditions.

5.3 Collinearity in model parameter estimates

The identification of collinearities in parameter estimates inform about those parameters that cannot be estimated independent of others ([Matear, 1995](#); [Schartau et al., 2017](#)). [Gibson and Spitz \(2011\)](#) stressed that it is of great interest to explore the sensitivity of an ecosystem model dynamics to uncertainty in biological parameters, initial conditions and environmental forcing and collinearities between them. In this thesis, collinearities between model parameters and initial conditions were clearly identified, as well as between different physiological parameters for different model approaches.

Results of Chapter 2 show that parameter collinearities are sensitive to the type of metric that is used to calibrate the model. In Chapter 2 the performance of four cost functions has been evaluated while configuring a plankton model with mesocosm data. Correlations between optimised values of loss parameters (mainly the exudation parameter γ) are significantly reduced when the cost function based on probabilistic assumption of data covariances is applied (Figs. 2.2 and 2.4 in Chapter 2). Similar results are discussed in [Schartau et al. \(2017\)](#), who calibrated a plankton model with mesocosm data to estimate algal loss parameters. They get better information on parameter estimates and obtain reduced collinearities when cost function with covariances was used.

Data assimilation results in Chapter 3 reveal strong negative collinearities between estimates of f_{cocco} and estimates of α and α_{cocco} . This indicates that initial abundance of coccolithophores in mesocosms can not be estimated independently of physiological model parameters (e.g. α and α_{cocco}). As these parameter dependencies are resolved by the data assimilation approach in Chapter 3, it became possible to understand the contribution of two initial factors (initial abundance and photo-acclimation states of coccolithophores) in generating a large variability in calcification. [Gibson and Spitz \(2011\)](#) pointed out a similar sensitivity between estimates of initial phytoplankton and nitrate concentrations and values of annual primary productivity in their regional marine ecosystem model.

In Chapter 4 collinearities between two physiological parameters, maximum potential photosynthesis rate of photoautotrophs (V_0^C) and subsistence quota (Q_{min}), have been revealed for two different model setups (OBM and CN-REcoM). Both models find strong positive correlation between V_0^C and Q_{min} only for phytoplankton in the mesocosms treated with high CO_2 concentrations (Table 6 in Chapter 4). As this dependency between V_0^C and Q_{min} is only obtained for high CO_2 mesocosms, it may be reasonable to speculate that OA can invoke antagonistic physiological responses on growth of marine algae.

5.4 Model biases

Few model biases are identified in this thesis. All model approaches described tend to underestimate phytoplankton biomass (PON) and chlorophyll during the start of the experiments (see Chapters 3 and 4). Initial picoplankton bloom was observed during the pre-bloom period of the PeECE-I experiment ([Engel et al., 2005](#)). Models introduced in this thesis do not resolve size spectrum of phytoplankton. As no size-based distinction has been, models may fail to capture early picoplankton bloom which could lead to underestimation of PON and Chl compared to observations between day 1 and day 4. [Joassin et al. \(2011\)](#) proposed a dynamical plankton model to simulate mesocosms of “Present” treatment from the PeECE-I experiment. Their model also underestimates Chl and PON at the start of the experiment, may be for the same reason as noted before.

Another bias has been observed in the simulated calcification response to elevated CO_2 level (see Chapter 3). The optimality-based model systematically overestimates calcification by coccolithophores in mesocosms treated with high CO_2 concentrations. This is evident from positive values of PIC residuals (model minus data) for mesocosms treated with high CO_2 (Figure 13 in Chapter 3). Although the model yields low cellular PIC:C-assimilation ratio

for the “Future” treatment mesocosms, it simulates a weak OA response on calcification at the plankton community level.

A common bias is identified for the OBM and CN-REcoM in Chapter 4. Both models predict some slower uptake of DIN compared to observations. Graphical visualization of observations from the PeECE-I experiment show that increase in PON do not match up with drawdown in DIN. This is why we assumed observations to be uncorrelated between day 11 and day 14 in the cost function applied in Chapter 4. However, for models peak in PON is obtained on the same day when DIN gets fully exhausted. This leads to a discrepancy between data and model results. To compensate for this, it seems the cost function tolerates misfits in DIN to generate good fit to PON data (see Figs. 4 and 5 in Chapter 4). And by doing so it yields lower overall model-data misfit for the entire period of the experiment.

Chapter 6

Outlook: What can be improved for similar future studies.

It has been shown that small uncertainties in initial conditions can lead to a large variability in response of dynamics of a planktonic ecosystem (e.g. in calcification, see Chapter 3). Therefore, for future mesocosm studies on OA we should put more efforts to measure initial conditions, e.g. data about initial plankton composition and measurements of initial nitrogen-to-carbon and chlorophyll-to-carbon ratios. For the experimental set up of future OA experiments, it is suggested to sort mesocosms into groups with similar initial conditions (e.g. phytoplankton abundances) first, and then perturbed these groups with different levels of CO₂ concentrations. This way we can resolve the effects of initial conditions and we may increase the chance of detecting a robust OA response signal, as noise-to-signal ratio will be reduced.

The model parameter describing the maximum grazing rate can remain poorly constrained in absence of explicit data about zooplankton and detritus (as shown in Chapter 2). As this zooplankton parameter is not identified, no firm model-based inferences on zooplankton dynamics and their top-down control on phytoplankton can be made. Therefore, more emphasis should be put on measurements of zooplankton abundance, composition and grazing rates in future mesocosm experiments, as these data may facilitate identification of zooplankton related model parameters.

Although plankton models (OBM and CN-REcoM) do not explicitly resolve effects of CO₂ on growth dynamics of phytoplankton, they (mainly OBM) suggest a dependency of physiological parameters (related to phytoplankton growth) on CO₂ (see Chapter 4). Therefore, for the future modelling studies with the OBM addressing similar scientific questions it might be beneficial to include parameterisations for CO₂ effects on nutrient uptake and phytoplankton growth in the model. In Chapter 4, the simulated POC:PON ratios are underestimated by the OBM compared to observations. This may be resolved by including phosphorus (P) limitation in the model for the future studies. Introduction of P-limitation may improve the solutions of carbon exudation as it induces delayed reduction in nitrogen uptake, and hence continuous uptake of carbon can take place until N:C ratio reaches its minimum (Q_{min}).

Different aspects of data assimilation, e.g. parameter optimisation and quantification of uncertainties in parameter estimates, have been extensively addressed here. A probabilistic based cost function, that accounts for temporal covariances between measurement types, is shown to perform much better as metric for model calibration than other cost functions (e.g. RMSEs). Data assimilation results presented in this thesis show that cost functions (probabilistic approaches) with covariances facilitate parameter identification and

uncertainty estimation as well as help to reduce parameter collinearities (Chapters 2, 3 and 4). Thus, for subsequent data-model syntheses of multivariate nature, this thesis recommends application cost functions that are derived from probabilistic approaches and account for temporal and/or spatial correlations between observations.

Bibliography

- Agawin, N. S., Duarte, C. M., and Agustí, S. (2000). Nutrient and temperature control of the contribution of picoplankton to phytoplankton biomass and production. *Limnology and Oceanography*, 45(3):591–600.
- Allen, J., Somerfield, P., and Gilbert, F. (2007). Quantifying uncertainty in high-resolution coupled hydrodynamic-ecosystem models. *Journal of Marine Systems*, 64(1):3–14.
- Anger, K. (2003). Salinity as a key parameter in the larval biology of decapod crustaceans. *Invertebrate reproduction & development*, 43(1):29–45.
- Armstrong, R. A. (2006). Optimality-based modeling of nitrogen allocation and photoacclimation in photosynthesis. *Deep Sea Research Part II: Topical Studies in Oceanography*, 53(5):513–531.
- Aumont, O. and Bopp, L. (2006). Globalizing results from ocean in situ iron fertilization studies. *Global Biogeochemical Cycles*, 20(2).
- Ayata, S.-D., Lévy, M., Aumont, O., Sciandra, A., Sainte-Marie, J., Tagliabue, A., and Bernard, O. (2013). Phytoplankton growth formulation in marine ecosystem models: should we take into account photo-acclimation and variable stoichiometry in oligotrophic areas? *Journal of Marine Systems*, 125:29–40.
- Barcelos e Ramos, J., Biswas, H., Schulz, K. G., LaRoche, J., and Riebesell, U. (2007). Effect of rising atmospheric carbon dioxide on the marine nitrogen fixer *Trichodesmium*. *Global biogeochemical cycles*, 21(2).
- Barcelos e Ramos, J., Müller, M., and Riebesell, U. (2010). Short-term response of the coccolithophore *Emiliana huxleyi* to an abrupt change in seawater carbon dioxide concentrations. *Biogeosciences*, 7:177.
- Baretta, J., Ebenhöf, W., and Ruurdij, P. (1995). The european regional seas ecosystem model, a complex marine ecosystem model. *Netherlands Journal of Sea Research*, 33(3-4):233–246.
- Baretta-Bekker, J., Baretta, J., Hansen, A., and Riemann, B. (1998). An improved model of carbon and nutrient dynamics in the microbial food web in marine enclosures. *Aquatic microbial ecology*, 14(1):91–108.
- Beardall, J. and Giordano, M. (2002). Ecological implications of microalgal and cyanobacterial CO₂ concentrating mechanisms, and their regulation. *Functional Plant Biology*, 29(3):335–347.
- Bélisle, C. J. (1992). Convergence theorems for a class of simulated annealing algorithms on Rd. *Journal of Applied Probability*, pages 885–895.

- Bellerby, R., Schulz, K., Riebesell, U., Neill, C., Nondal, G., Johannessen, T., and Brown, K. (2008). Marine ecosystem community carbon and nutrient uptake stoichiometry under varying ocean acidification during the pecee III experiment. *Biogeosciences (BG)*, 5(6):1517–1527.
- Bjørrisen, P. K. (1988). Phytoplankton exudation of organic matter: Why do healthy cells do it? *Limnology and oceanography*, 33(1):151–154.
- Blackford, J. C. (2010). Predicting the impacts of ocean acidification: Challenges from an ecosystem perspective. *Journal of Marine Systems*, 81(1):12–18.
- Bradshaw, C., Näslund, J., Hansen, J., Kozłowsky-Suzuki, B., Sundström, B., and Gustafsson, K. (2015). Hexabromocyclododecane affects benthic-pelagic coupling in an experimental ecosystem. *Environmental Pollution*, 206:306–314.
- Brasseur, P., Bahurel, P., Bertino, L., Birol, F., Brankart, J.-M., Ferry, N., Losa, S., Rémy, E., Schröter, J., Skachko, S., et al. (2005). Data assimilation for marine monitoring and prediction: the mercator operational assimilation systems and the mersea developments. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3561–3582.
- Browne, M. A., Dissanayake, A., Galloway, T. S., Lowe, D. M., and Thompson, R. C. (2008). Ingested microscopic plastic translocates to the circulatory system of the mussel, *Mytilus edulis* (L.). *Environmental science & technology*, 42(13):5026–5031.
- Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms 1. general considerations. *IMA Journal of Applied Mathematics*, 6(1):76–90.
- Buitenhuis, E. T., De Baar, H. J., and Veldhuis, M. J. (1999). Photosynthesis and calcification by *Emiliana Huxleyi* (prymnesiophyceae) as a function of inorganic carbon species. *Journal of Phycology*, 35(5):949–959.
- Burkhardt, S. and Riebesell, U. (1997). CO₂ availability affects elemental composition (C:N:P) of the marine diatom *Skeletonema costatum*. *Marine Ecology Progress Series*, pages 67–76.
- Burkhardt, S., Zondervan, I., and Riebesell, U. (1999). Effect of CO₂ concentration on C:N:P ratio in marine phytoplankton: A species comparison. *Limnology and Oceanography*, 44(3):683–690.
- Caron, D. A. and Hutchins, D. A. (2012). The effects of changing climate on microzooplankton grazing and community structure: drivers, predictions and knowledge gaps. *Journal of Plankton Research*, 35(2):235–252.
- Carotenuto, Y., Putzeys, S., Simonelli, P., Paulino, A., Meyerhöfer, M., Sufrian, K., Antia, A., and Nejstgaard, J. (2007). Copepod feeding and reproduction in relation to phytoplankton development during the PeECE III mesocosm experiment. *Biogeosciences Discussions*, 4(5):3913–3936.
- Comeau, S., Gorsky, G., Jeffree, R., Teyssié, J.-L., and Gattuso, J.-P. (2009). Impact of ocean acidification on a key arctic pelagic mollusc (*Limacina helicina*). *Biogeosciences*, 6(9):1877–1882.

- Cowart, D. A., Ulrich, P. N., Miller, D. C., and Marsh, A. G. (2009). Salinity sensitivity of early embryos of the antarctic sea urchin, *Sterechinus neumayeri*. *Polar biology*, 32(3):435–441.
- Delille, B., Harlay, J., Zondervan, I., Jacquet, S., Chou, L., Wollast, R., Bellerby, R. G., Frankignoulle, M., Borges, A. V., Riebesell, U., et al. (2005). Response of primary production and calcification to changes of $p\text{CO}_2$ during experimental blooms of the coccolithophorid *Emiliana huxleyi*. *Global Biogeochemical Cycles*, 19(2).
- Doney, S. C. (1999). Major challenges confronting marine biogeochemical modeling. *Global Biogeochemical Cycles*, 13(3):705–714.
- Doney, S. C., Lima, I., Moore, J. K., Lindsay, K., Behrenfeld, M. J., Westberry, T. K., Mahowald, N., Glover, D. M., and Takahashi, T. (2009). Skill metrics for confronting global upper ocean ecosystem-biogeochemistry models against field and remote sensing data. *Journal of Marine Systems*, 76(1):95–112.
- Dowd, M. (2007). Bayesian statistical data assimilation for ecosystem models using markov chain monte carlo. *Journal of Marine Systems*, 68(3):439–456.
- Dowd, M., Jones, E., and Parslow, J. (2014). A statistical overview and perspectives on data assimilation for marine biogeochemical models. *Environmetrics*, 25(4):203–213.
- Droop, M. (1974). The nutrient status of algal cells in continuous culture. *Journal of the Marine Biological Association of the United Kingdom*, 54(4):825–855.
- Droop, M. (1983). 25 years of algal growth kinetics a personal view. *Botanica marina*, 26(3):99–112.
- Dueri, S., Dahllöf, I., Hjorth, M., Marinov, D., and Zaldívar, J. (2009). Modeling the combined effect of nutrients and pyrene on the plankton population: Validation using mesocosm experiment data and scenario analysis. *Ecological Modelling*, 220(17):2060–2067.
- Dugdale, R. (1967). Nutrient limitation in the sea: dynamics, identification and significance. *Limnol. Oceanogr*, 12(4):685–695.
- Dutkiewicz, S., Follows, M. J., and Bragg, J. G. (2009). Modeling the coupling of ocean ecology and biogeochemistry. *Global Biogeochemical Cycles*, 23(4).
- Edmunds, P. J. (2011). Zooplanktivory ameliorates the effects of ocean acidification on the reef coral *Porites* spp. *Limnology and oceanography*, 56(6):2402–2410.
- EGGE, J., Thingstad, T., Engel, A., Bellerby, R., and Riebesell, U. (2007). Primary production during nutrient-induced blooms at elevated CO_2 concentrations. *Biogeosciences Discussions*, 4(6):4385–4410.
- Eggers, S. L., Lewandowska, A. M., Barcelos e Ramos, J., Blanco-Ameijeiras, S., Gallo, F., and Matthiessen, B. (2014). Community composition has greater impact on the functioning of marine phytoplankton communities than ocean acidification. *Global Change Biology*, 20(3):713–723.

- Ekau, W., Auel, H., Pörtner, H.-O., and Gilbert, D. (2010). Impacts of hypoxia on the structure and processes in pelagic communities (zooplankton, macro-invertebrates and fish). *Biogeosciences*, 7(5):1669–1699.
- Eknes, M. and Evensen, G. (2002). An ensemble kalman filter with a 1-D marine ecosystem model. *Journal of Marine Systems*, 36(1):75–100.
- Ellis, R. P., Bersey, J., Rundle, S. D., Hall-Spencer, J. M., and Spicer, J. I. (2009). Subtle but significant effects of CO₂ acidified seawater on embryos of the intertidal snail, *Littorina obtusata*. *Aquatic Biology*, 5.
- Engel, A. (2000). The role of transparent exopolymer particles (tep) in the increase in apparent particle stickiness (α) during the decline of a diatom bloom. *Journal of Plankton Research*, 22(3):485–497.
- Engel, A., Cisternas Novoa, C., Wurst, M., Endres, S., Tang, T., Schartau, M., and Lee, C. (2014). No detectable effect of CO₂ on elemental stoichiometry of *Emiliania huxleyi* in nutrient-limited, acclimated continuous cultures. *Marine Ecology Progress Series*, 507:15–30.
- Engel, A., Delille, B., Jacquet, S., Riebesell, U., Rochelle-Newall, E., Terbrüggen, A., and Zondervan, I. (2004a). Transparent exopolymer particles and dissolved organic carbon production by *Emiliania huxleyi* exposed to different CO₂ concentrations: a mesocosm experiment. *Aquatic Microbial Ecology*, 34(1):93–104.
- Engel, A., Thoms, S., Riebesell, U., Rochelle-Newall, E., and Zondervan, I. (2004b). Polysaccharide aggregation as a potential sink of marine dissolved organic carbon. *Nature*, 428(6986):929–932.
- Engel, A., Zondervan, I., Aerts, K., Beaufort, L., Benthien, A., Chou, L., Delille, B., Gattuso, J.-P., Harlay, J., Heemann, C., et al. (2005). Testing the direct effect of CO₂ concentration on a bloom of the coccolithophorid *Emiliania huxleyi* in mesocosm experiments. *Limnology and Oceanography*, 50(2):493–507.
- Fabry, V. J., Seibel, B. A., Feely, R. A., and Orr, J. C. (2008). Impacts of ocean acidification on marine fauna and ecosystem processes. *ICES Journal of Marine Science*, 65(3):414–432.
- Falkowski, P. G., Katz, M. E., Knoll, A. H., Quigg, A., Raven, J. A., Schofield, O., and Taylor, F. (2004). The evolution of modern eukaryotic phytoplankton. *science*, 305(5682):354–360.
- Fasham, M., Ducklow, H., and McKelvie, S. (1990). A nitrogen-based model of plankton dynamics in the oceanic mixed layer. *Journal of Marine Research*, 48(3):591–639.
- Faugeras, B., Bernard, O., Sciandra, A., and Lévy, M. (2004). A mechanistic modelling and data assimilation approach to estimate the carbon/chlorophyll and carbon/nitrogen ratios in a coupled hydrodynamical-biological model. *Nonlinear Processes in Geophysics*, 11(4):515–533.
- Fennel, K., Losch, M., Schröter, J., and Wenzel, M. (2001). Testing a marine ecosystem model: sensitivity analysis and parameter optimization. *Journal of Marine Systems*, 28(1):45–63.

- Fiechter, J., Herbei, R., Leeds, W., Brown, J., Milliff, R., Wikle, C., Moore, A., and Powell, T. (2013). A bayesian parameter estimation method applied to a marine ecosystem model for the coastal Gulf of Alaska. *Ecological modelling*, 258:122–133.
- Findlay, H. S., Calosi, P., and Crawford, K. (2011). Determinants of the PIC: POC response in the coccolithophore *Emiliana huxleyi* under future ocean acidification scenarios. *Limnology and Oceanography*, 56(3):1168–1178.
- Fletcher, R. (1970). A new approach to variable metric algorithms. *The Computer Journal*, 13(3):317–322.
- Flynn, K., Davidson, K., and Leftley, J. (1994). Carbon-nitrogen relations at whole-cell and free-amino-acid levels during batch growth of *Isochrysis galbana* (prymnesiophyceae) under conditions of alternating light and dark. *Marine Biology*, 118(2):229–237.
- Flynn, K. J. (2001). A mechanistic model for describing dynamic multi-nutrient, light, temperature interactions in phytoplankton. *Journal of Plankton Research*, 23(9):977–997.
- Flynn, K. J. (2003). Modelling multi-nutrient interactions in phytoplankton; balancing simplicity and realism. *Progress in Oceanography*, 56(2):249–279.
- Fonseca, C. M. and Fleming, P. J. (1995). An overview of evolutionary algorithms in multiobjective optimization. *Evolutionary computation*, 3(1):1–16.
- Franks, P. J. (2009). Planktonic ecosystem models: perplexing parameterizations and a failure to fail. *Journal of Plankton Research*, 31(11):1299–1306.
- Friedrichs, M. A., Dusenberry, J. A., Anderson, L. A., Armstrong, R. A., Chai, F., Christian, J. R., Doney, S. C., Dunne, J., Fujii, M., Hood, R., et al. (2007). Assessment of skill and portability in regional marine biogeochemical models: Role of multiple planktonic groups. *Journal of Geophysical Research: Oceans*, 112(C8).
- Friedrichs, M. A., Hood, R. R., and Wiggert, J. D. (2006). Ecosystem model complexity versus physical forcing: Quantification of their relative impact with assimilated Arabian Sea data. *Deep Sea Research Part II: Topical Studies in Oceanography*, 53(5):576–600.
- Fu, F.-X., Warner, M. E., Zhang, Y., Feng, Y., and Hutchins, D. A. (2007). Effects of increased temperature and CO₂ on photosynthesis, growth, and elemental ratios in marine *Synechococcus* and *Prochlorococcus*(cyanobacteria). *Journal of Phycology*, 43(3):485–496.
- Fulton, E. A., Parslow, J. S., Smith, A. D., and Johnson, C. R. (2004). Biogeochemical marine ecosystem models II: the effect of physiological detail on model performance. *Ecological modelling*, 173(4):371–406.
- Galloway, T. S. and Depledge, M. H. (2001). Immunotoxicity in invertebrates: measurement and ecotoxicological relevance. *Ecotoxicology*, 10(1):5–23.
- Gao, K., Helbling, E. W., Häder, D.-P., and Hutchins, D. A. (2012). Responses of marine primary producers to interactions between ocean acidification, solar radiation, and warming. *Marine Ecology Progress Series*, 470:167–189.

- Gattuso, J.-P. and Hansson, L. (2011). *Ocean acidification*. OUP Oxford.
- Gauthier, O., Virieux, J., and Tarantola, A. (1986). Two-dimensional nonlinear inversion of seismic waveforms: Numerical results. *Geophysics*, 51(7):1387–1403.
- Geider, R. J., MacIntyre, H. L., and Kana, T. M. (1998). A dynamic regulatory model of phytoplanktonic acclimation to light, nutrients, and temperature. *Limnology and Oceanography*, 43(4):679–694.
- Gervais, F. and Riebesell, U. (2001). Effect of phosphorus limitation on elemental composition and stable carbon isotope fractionation in a marine diatom growing under different CO₂ concentrations. *Limnology and Oceanography*, 46(3):497–504.
- Gibson, G. and Spitz, Y. (2011). Impacts of biological parameterization, initial conditions, and environmental forcing on parameter sensitivity and uncertainty in a marine ecosystem model for the Bering Sea. *Journal of Marine Systems*, 88(2):214–231.
- Gilbert, J. C. and Nocedal, J. (1992). Global convergence properties of conjugate gradient methods for optimization. *SIAM Journal on optimization*, 2(1):21–42.
- Giordano, M., Beardall, J., and Raven, J. A. (2005). CO₂ concentrating mechanisms in algae: mechanisms, environmental modulation, and evolution. *Annu. Rev. Plant Biol.*, 56:99–131.
- Goldenberg, S. U., Nagelkerken, I., Ferreira, C. M., Ullah, H., and Connell, S. D. (2017). Boosted food web productivity through ocean acidification collapses under warming. *Global Change Biology*.
- Goldfarb, D. (1970). A family of variable-metric methods derived by variational means. *Mathematics of Computation*, 24(109):23–26.
- Graham, E. R. and Thompson, J. T. (2009). Deposit-and suspension-feeding sea cucumbers (Echinodermata) ingest plastic fragments. *Journal of Experimental Marine Biology and Ecology*, 368(1):22–29.
- Gregg, W. W., Friedrichs, M. A., Robinson, A. R., Rose, K. A., Schlitzer, R., Thompson, K. R., and Doney, S. C. (2009). Skill assessment in ocean biological data assimilation. *Journal of Marine Systems*, 76(1):16–33.
- Gunson, J., Oschlies, A., and Garçon, V. (1999). Sensitivity of ecosystem parameters to simulated satellite ocean color data using a coupled physical-biological model of the North Atlantic. *Journal of Marine Research*, 57(4):613–639.
- Haario, H., Laine, M., Mira, A., and Saksman, E. (2006). DRAM: Efficient adaptive MCMC. *Statistics and Computing*, 16(4):339–354.
- Halpern, B. S., Frazier, M., Potapenko, J., Casey, K. S., Koenig, K., Longo, C., Lowndes, J. S., Rockwood, R. C., Selig, E. R., Selkoe, K. A., et al. (2015). Spatial and temporal changes in cumulative human impacts on the world’s ocean. *Nature communications*, 6.

- Hansen, N., Müller, S. D., and Koumoutsakos, P. (2003). Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). *Evolutionary computation*, 11(1):1–18.
- Harmon, R. and Challenor, P. (1997). A Markov chain Monte Carlo method for estimation and assimilation into models. *Ecological Modelling*, 101(1):41–59.
- Hastings, W. K. (1970). Monte carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.
- Hood, R. R., Laws, E. A., Armstrong, R. A., Bates, N. R., Brown, C. W., Carlson, C. A., Chai, F., Doney, S. C., Falkowski, P. G., Feely, R. A., et al. (2006). Pelagic functional group modeling: Progress, challenges and prospects. *Deep Sea Research Part II: Topical Studies in Oceanography*, 53(5):459–512.
- Hurtt, G. C. and Armstrong, R. A. (1996). A pelagic ecosystem model calibrated with bats data. *Deep Sea Research Part II: Topical Studies in Oceanography*, 43(2-3):653–683.
- Hutchins, D. A., Mulholland, M. R., and Fu, F. (2009). Nutrient cycles and marine microbes in a CO₂-enriched ocean. *Oceanography*, 22(4):128–145.
- Iglesias-Rodriguez, M. D., Halloran, P. R., Rickaby, R. E., Hall, I. R., Colmenero-Hidalgo, E., Gittins, J. R., Green, D. R., Tyrrell, T., Gibbs, S. J., von Dassow, P., et al. (2008). Phytoplankton calcification in a high-CO₂ world. *Science*, 320(5874):336–340.
- Joassin, P., Delille, B., Soetaert, K., Harlay, J., Borges, A. V., Chou, L., Riebesell, U., Suykens, K., and Grégoire, M. (2011). Carbon and nitrogen flows during a bloom of the coccolithophore *Emiliana huxleyi*: Modelling a mesocosm experiment. *Journal of Marine Systems*, 85(3):71–85.
- Jokiel, P., Rodgers, K., Kuffner, I., Andersson, A., Cox, E., and Mackenzie, F. (2008). Ocean acidification and calcifying reef organisms: a mesocosm investigation. *Coral Reefs*, 27(3):473–483.
- Kriest, I., Sauerland, V., Khatiwala, S., Srivastav, A., and Oschlies, A. (2017). Calibrating a global three-dimensional biogeochemical ocean model (MOPS-1.0). *Geoscientific Model Development*, 10(1):127.
- Krishna, S. and Schartau, M. (2017). A data–model synthesis to explain variability in calcification observed during a CO₂ perturbation mesocosm experiment. *Biogeosciences*, 14(7):1857.
- Kuffner, I. B., Andersson, A. J., Jokiel, P. L., Ku’ulei, S. R., and Mackenzie, F. T. (2008). Decreased abundance of crustose coralline algae due to ocean acidification. *Nature Geoscience*, 1(2):114.
- Lancaster, H. F. and Drenner, R. W. (1990). Experimental mesocosm study of the separate and interaction effects of phosphorus and mosquitofish (*Gambusia affinis*) on plankton community structure. *Canadian Journal of Fisheries and Aquatic Sciences*, 47(3):471–479.
- Langer, G., Geisen, M., Baumann, K.-H., Kläs, J., Riebesell, U., Thoms, S., and Young, J. R. (2006). Species-specific responses of calcifying algae to changing seawater carbonate chemistry. *Geochemistry, Geophysics, Geosystems*, 7(9).

- Larsson, U. and Hagström, A. (1979). Phytoplankton exudate release as an energy source for the growth of pelagic bacteria. *Marine biology*, 52(3):199–206.
- Lawson, L. M., Hofmann, E. E., and Spitz, Y. H. (1996). Time series sampling and data assimilation in a simple marine ecosystem model. *Deep Sea Research Part II: Topical Studies in Oceanography*, 43(2-3):625–651.
- Lawson, L. M., Spitz, Y. H., Hofmann, E. E., and Long, R. B. (1995). A data assimilation technique applied to a predator-prey model. *Bulletin of Mathematical Biology*, 57(4):593–617.
- Leggett, D. J. (2013). *Computational methods for the determination of formation constants*. Springer Science & Business Media.
- Lehmann, E. L. and Casella, G. (2006). *Theory of point estimation*. Springer Science & Business Media.
- Leonardos, N. and Geider, R. J. (2005). Elevated atmospheric carbon dioxide increases organic carbon fixation by *Emiliana huxleyi* (haptophyta), under nutrient-limited high-light conditions. *Journal of Phycology*, 41(6):1196–1203.
- Lesser, M. P. (2010). Depth-dependent effects of ultraviolet radiation on survivorship, oxidative stress and DNA damage in sea urchin (*Strongylocentrotus droebachiensis*) embryos from the Gulf of Maine. *Photochemistry and photobiology*, 86(2):382–388.
- Lischka, S., Büdenbender, J., Boxhammer, T., and Riebesell, U. (2011). Impact of ocean acidification and elevated temperatures on early juveniles of the polar shelled pteropod *Limacina helicina*: mortality, shell degradation, and shell growth. *Biogeosciences*, 8(4):919.
- Lohbeck, K. T., Riebesell, U., and Reusch, T. B. (2012). Adaptive evolution of a key phytoplankton species to ocean acidification. *Nature Geoscience*, 5(5):346.
- Lombard, F., da Rocha, R. E., Bijma, J., and Gattuso, J.-P. (2009). Effect of carbonate ion concentration and irradiance on calcification in foraminifera. *Biogeosciences Discussions*, 6:8589–8608.
- Losa, S. N., Kivman, G. A., and Ryabchenko, V. A. (2004). Weak constraint parameter estimation for a simple ocean ecosystem model: what can we learn about the model and data? *Journal of marine systems*, 45(1):1–20.
- Marquardt, D. W. (1963). An algorithm for least-squares estimation of non-linear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2):431–441.
- Matear, R. J. (1995). Parameter optimization and analysis of ecosystem models using simulated annealing: A case study at station P. *Journal of Marine Research*, 53(4):571–607.
- McCall, J. (2005). Genetic algorithms for modelling and optimisation. *Journal of Computational and Applied Mathematics*, 184(1):205–222.

- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092.
- Meyer, J. and Riebesell, U. (2015). Reviews and syntheses: Responses of coccolithophores to ocean acidification: a meta-analysis. *Biogeosciences*, 12(6):1671.
- Monteiro, F. M., Bach, L. T., Brownlee, C., Bown, P., Rickaby, R. E., Poulton, A. J., Tyrrell, T., Beaufort, L., Dutkiewicz, S., Gibbs, S., et al. (2016). Why marine phytoplankton calcify. *Science Advances*, 2(7):e1501822.
- Moore, J. K., Doney, S. C., Kleypas, J. A., Glover, D. M., and Fung, I. Y. (2001). An intermediate complexity marine ecosystem model for the global domain. *Deep Sea Research Part II: Topical Studies in Oceanography*, 49(1):403–462.
- Morel, F. M. (1987). Kinetics of nutrient uptake and growth in phytoplankton. *Journal of Phycology*, 23(1):137–150.
- Natvik, L.-J. and Evensen, G. (2003). Assimilation of ocean colour data into a biochemical model of the North Atlantic: Part 1. data assimilation experiments. *Journal of Marine Systems*, 40:127–153.
- Nerger, L. and Gregg, W. W. (2007). Assimilation of seawifs data into a global ocean-biogeochemical model using a local seik filter. *Journal of Marine Systems*, 68(1):237–254.
- Oschlies, A. and Schartau, M. (2005). Basin-scale performance of a locally optimized marine ecosystem model. *Journal of Marine Research*, 63(2):335–358.
- Pahlow, M. (2005). Linking chlorophyll–nutrient dynamics to the redfield N:C ratio with a model of optimal phytoplankton growth. *Marine Ecology Progress Series*, 287:33–43.
- Pahlow, M., Dietze, H., and Oschlies, A. (2013). Optimality-based model of phytoplankton growth and diazotrophy. *Marine Ecology Progress Series*, 489:1–16.
- Pahlow, M. and Oschlies, A. (2009). Chain model of phytoplankton P, N and light colimitation. *Marine Ecology Progress Series*, 376:1–16.
- Ranganathan, A. (2004). The Levenberg-Marquardt Algorithm. *Tutorial on LM algorithm*, 11(1):101–110.
- Raupach, M., Rayner, P., Barrett, D., DeFries, R., Heimann, M., Ojima, D., Quegan, S., and Schmulius, C. (2005). Model–data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications. *Global Change Biology*, 11(3):378–397.
- Raven, J. A., Giordano, M., Beardall, J., and Maberly, S. C. (2011). Algal and aquatic plant carbon concentrating mechanisms in relation to environmental change. *Photosynthesis Research*, 109(1):281–296.

- RDevelopment, C. (2012). Team 2009: R: A language and environment for statistical computing. Vienna, Austria. Internet: <http://www.R-project.org>.
- Rickaby, R. E., Young, J., et al. (2010). Perturbing phytoplankton: response and isotopic fractionation with changing carbonate chemistry in two coccolithophore species. *Climate of the Past*, 6(6):771.
- Ridgwell, A., Zondervan, I., Hargreaves, J., Bijma, J., and Lenton, T. (2007). Assessing the potential long-term increase of oceanic fossil fuel CO₂ uptake due to CO₂-calcification feedback. *Biogeosciences*, 4(4):481–492.
- Riebesell, U., Schulz, K. G., Bellerby, R., Botros, M., Fritsche, P., Meyerhöfer, M., Neill, C., Nondal, G., Oschlies, A., Wohlers, J., et al. (2007). Enhanced biological carbon consumption in a high CO₂ ocean. *Nature*, 450(7169):545.
- Rose, K. A., Megrey, B. A., Werner, F. E., and Ware, D. M. (2007). Calibration of the NEMURO nutrient–phytoplankton–zooplankton food web model to a coastal ecosystem: Evaluation of an automated calibration approach. *Ecological modelling*, 202(1):38–51.
- Rost, B., Richter, K.-U., Riebesell, U., and Hansen, P. J. (2006). Inorganic carbon acquisition in red tide dinoflagellates. *Plant, Cell & Environment*, 29(5):810–822.
- Roy, S., Mohovic, B., Giancesella, S. M., Schloss, I., Ferrario, M., and Demers, S. (2006). Effects of enhanced UV-B on pigment-based phytoplankton biomass and composition of mesocosm-enclosed natural marine communities from three latitudes. *Photochemistry and Photobiology*, 82(4):909–922.
- Schartau, M., Engel, A., Schröter, J., Thoms, S., Völker, C., and Wolf-Gladrow, D. (2007). Modelling carbon overconsumption and the formation of extracellular particulate organic carbon. *Biogeosciences*, 4(4):433–454.
- Schartau, M. and Oschlies, A. (2003). Simultaneous data-based optimization of a 1D-ecosystem model at three locations in the North Atlantic: Part I—method and parameter estimates. *Journal of Marine Research*, 61(6):765–793.
- Schartau, M., Wallhead, P., Hemmings, J., Löptien, U., Kriest, I., Krishna, S., Ward, B. A., Slawig, T., and Oschlies, A. (2017). Reviews and syntheses: Parameter identification in marine planktonic ecosystem modelling. *Biogeosciences Discussions*, pages 1–79.
- Schindler, D. (1977). Evolution of phosphorus limitation in lakes. *Science*, 195(4275):260–262.
- Segner, H., Wenger, M., Möller, A. M., Köllner, B., and Casanova-Nakayama, A. (2012). Immunotoxic effects of environmental toxicants in fish—how to assess them? *Environmental science and pollution research*, 19(7):2465–2476.
- Shannon, D. F. (1970). Conditioning of quasi-Newton methods for function minimization. *Mathematics of Computation*, 24(111):647–656.

- Short, F. T., Burdick, D. M., and Kaldy, J. E. (1995). Mesocosm experiments quantify the effects of eutrophication on eelgrass, *Zostera marina*. *Limnology and oceanography*, 40(4):740–749.
- Shuter, B. (1979). A model of physiological adaptation in unicellular algae. *Journal of theoretical biology*, 78(4):519–552.
- Sinha, R. P. and Häder, D.-P. (2002). UV-induced DNA damage and repair: a review. *Photochemical & Photobiological Sciences*, 1(4):225–236.
- Sivia, D. and Skilling, J. (2006). *Data analysis: a Bayesian tutorial*. OUP Oxford.
- Sivia, D. S. (1996). *Data Analysis: A Bayesian Tutorial*. Oxford Univ. Press.
- Skartveit, A., Cleveland, F., and de Lange, T. (2001). Radiation Yearbook no. 37, Meteorological Report Series, University of Bergen. Technical report, University of Bergen Geophysical Institute, Bergen, Norway.
- Smith, S. L., Pahlow, M., Merico, A., and Wirtz, K. W. (2011). Optimality-based modeling of planktonic organisms. *Limnology and Oceanography*, 56(6):2080–2094.
- Smith, S. L. and Yamanaka, Y. (2007). Quantitative comparison of photoacclimation models for marine phytoplankton. *ecological modelling*, 201(3):547–552.
- Smith, S. L., Yamanaka, Y., Pahlow, M., and Oschlies, A. (2009). Optimal uptake kinetics: physiological acclimation explains the pattern of nitrate uptake by phytoplankton in the ocean. *Marine Ecology Progress Series*, 384:1–12.
- Snyman, J. (2005). *Practical mathematical optimization: An introduction to basic optimization theory and classical and new gradient-based algorithms*, volume 97. Springer Science & Business Media.
- Soetaert, K. and Petzoldt, T. (2010). Inverse modelling, sensitivity and Monte Carlo analysis in R using package FME. *Journal of Statistical Software*, 33.
- Spitz, Y., Moisan, J., and Abbott, M. (2001). Configuring an ecosystem model using data from the Bermuda Atlantic time series (BATS). *Deep Sea Research Part II: Topical Studies in Oceanography*, 48(8):1733–1768.
- Spitz, Y., Moisan, J., Abbott, M., and Richman, J. (1998). Data assimilation and a pelagic ecosystem model: parameterization using time series observations. *Journal of Marine Systems*, 16(1):51–68.
- Stow, C. A., Jolliff, J., McGillicuddy, D. J., Doney, S. C., Allen, J. I., Friedrichs, M. A., Rose, K. A., and Wallhead, P. (2009). Skill assessment for coupled biological/physical models of marine systems. *Journal of Marine Systems*, 76(1):4–15.
- Strecker, A. L., Cobb, T. P., and Vinebrooke, R. D. (2004). Effects of experimental greenhouse warming on phytoplankton and zooplankton communities in fishless alpine ponds. *Limnology and Oceanography*, 49(4):1182–1190.

- Suffrian, K., Simonelli, P., Nejstgaard, J., Putzeys, S., Carotenuto, Y., and Antia, A. (2008). Microzooplankton grazing and phytoplankton growth in marine mesocosms with increased CO₂ levels. *Biogeosciences (BG)*, 5:1145–1156.
- Sunda, W., Kieber, D., Kiene, R., and Huntsman, S. (2002). An antioxidant function for DMSP and DMS in marine algae. *Nature*, 418(6895):317.
- Tanaka, T., Thingstad, T., Lovdal, T., Grossart, H.-P., Larsen, A., Allgaier, M., Meyerhöfer, M., Schulz, K., Wohlers, J., Zöllner, E., et al. (2008). Availability of phosphate for phytoplankton and bacteria and of labile organic carbon for bacteria at different pCO₂ levels in a mesocosm study. *Biogeosciences (BG)*, 5:669–678.
- Thacker, W. C. (1989). The role of the Hessian matrix in fitting models to measurements. *Journal of Geophysical Research: Oceans*, 94(C5):6177–6196.
- Todorov, E. (2005). Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system. *Neural computation*, 17(5):1084–1108.
- Toseland, A., Daines, S. J., Clark, J. R., Kirkham, A., Strauss, J., Uhlig, C., Lenton, T. M., Valentin, K., Pearson, G. A., Moulton, V., et al. (2013). The impact of temperature on marine phytoplankton resource allocation and metabolism. *Nature Climate Change*, 3(11):979.
- Tziperman, E., Thacker, W. C., Long, R. B., Hwang, S.-M., and Rintoul, S. R. (1992). Oceanic data analysis using a general circulation model. part II: A North Atlantic model. *Journal of Physical Oceanography*, 22(12):1458–1485.
- Vallino, J. (2000). Improving marine ecosystem models: use of data assimilation and mesocosm experiments. *Journal of Marine Research*, 58(1):117–164.
- Van De Poll, W. H., Eggert, A., Buma, A. G., and Breeman, A. M. (2001). Effects of UV-B-induced DNA damage and photoinhibition on growth of temperate marine red macrophytes: habitat-related differences in uv-b tolerance. *Journal of Phycology*, 37(1):30–38.
- Vaquer-Sunyer, R. and Duarte, C. M. (2008). Thresholds of hypoxia for marine biodiversity. *Proceedings of the National Academy of Sciences*, 105(40):15452–15457.
- Waldbusser, G. G., Bergschneider, H., and Green, M. A. (2010). Size-dependent pH effect on calcification in post-larval hard clam *Mercenaria* spp. *Marine Ecology Progress Series*, 417:171–182.
- Ward, B. A., Friedrichs, M. A., Anderson, T. R., and Oschlies, A. (2010). Parameter optimisation techniques and the problem of underdetermination in marine biogeochemical models. *Journal of Marine Systems*, 81(1):34–43.
- Ward, B. A., Schartau, M., Oschlies, A., Martin, A. P., Follows, M. J., and Anderson, T. R. (2013). When is a biogeochemical model too complex? objective model reduction and selection for North Atlantic time-series sites. *Progress in Oceanography*, 116:49–65.

- Weir, B., Miller, R. N., and Spitz, Y. H. (2013). Implicit estimation of ecological model parameters. *Bulletin of mathematical biology*, 75(2):223–257.
- Zeebe, R. E., Zachos, J. C., Caldeira, K., and Tyrrell, T. (2008). Carbon emissions and acidification. *Science*, 321(5885):51.
- Zondervan, I., Rost, B., and Riebesell, U. (2002). Effect of CO₂ concentration on the PIC/POC ratio in the coccolithophore *Emiliana huxleyi* grown under light-limiting conditions and different daylengths. *Journal of Experimental Marine Biology and Ecology*, 272(1):55–70.

Acknowledgements

I take this opportunity to express my deepest sense of gratitude to my employer Prof. Andreas Oschlies for giving me an opportunity to work in this wonderful research group (Biogeochemical modelling), and to pursue my dreams and make them reality. I am thankful to Andreas for his immense support throughout the period of my doctoral research. Words are not enough to thank my graduate supervisor Dr. Markus Schartau. He is a “father-figure” in my doctoral career. It is because of his constant guidance and support I could get this ahead in research. For this I would like to thank him from the bottom of my heart.

Although not being my official PhD supervisor on paper, Dr. Markus Pahlow has always been a fantastic mentor to me. I got vital help from Markus during my research, and his work ethics and poise are inspiration to we all young scientists. Here, I have an opportunity to thank Markus for his generosity. Further, I would like to thank Prof. Hermann Bange for giving me an opportunity to work in SOPRAN project, which was a stepping-stone in my research career. Towards the end of my PhD, I was involved in BIOACID project (Phase III), for which I thank Prof. Ulf Riebesell. I am also thankful to my dear colleague Ulrike Löptien for reviewing the introduction of my thesis. In general, I gained a lot in terms of scientific knowledge from my colleagues in Biogeochemical Modelling group (especially during our weekly group meetings). Therefore, I would like to thank you all as in one way or the other you all contributed in my doctoral research. It is my duty to thank the ISOS team (particularly Dr. Avan Antia) for supporting (partially) my travels to attend conferences and summer schools during the course of my PhD programme and also for providing some very useful courses. I also would like to acknowledge here the constant assistance I got from Mrs. Monika Peschke (our secretary) while dealing with bureaucratic and administrative issues. Thank you Monika for your help throughout my employment in GEOMAR.

Life is empty without friends. I consider this thesis as a platform to thank all my friends. Firstly, friends from GEOMAR (Hadi, Evan, Daniel, Sabine, Eduardo, Fabian, Bei, Alex, Lionel and many more). Coffee breaks with Hadi certainly helped to boost my productivity at work and at times also helped me to come out from mental blocks. So cheers to our coffee breaks, Hadi. My good friend and a wonderful office mate Sabine helped me with german translation of my thesis abstract, for which I am thankful to her. I am grateful to my friends outside GEOMAR (Sumanth, Apurva, Ashish bhai, Shanthi, Mahasweta, Rathi, Gaurav, Rajesh, Tarunjot, Shubhranshu, Mathias). Special thanks to my american friend Daniel for reviewing my thesis introduction. Friends, you all kept me motivated and were always there whenever I needed you. I wish you all love, success and prosperity in your lives.

My parents are mine greatest hero. It is because of their sacrifice, support and love I could reach to this level in my life. My mother (Mrs. Priyambada Srivastava) is a personification of compassion and hard work. My father (Mr.

Ratan Srivastava) has shown me how to dream big and taught me how to be self confident in life while sticking to the path of honesty. He unfortunately passed away in year 2013, but his teachings and blessings are always with me. I love you Papa. I also want to thank two special girls in my life, Liliana (my love) and Swarnim (my sister) for their unconditional love.

Last but not the least thanks to almighty for good health and wisdom.

Shubham Krishna

Kirchhofallee 37,
Kiel 24114, Germany
phone: (+49) 431-6004270
cell: (+49) 151-23051371
skrishna@geomar.de

PERSONAL

Date of Birth: May 5, 1986
Place of Birth: Allahabad, India
Citizenship: Indian
Visa Status: Permanent residence permit in Germany/Schengen Area

EDUCATION

GEOMAR (Helmholtz Center for Ocean Research), Kiel, Germany

PhD in Biogeochemical Modelling
May 2013 – present (Expected defense in January 2018)
Thesis: Model assessment and model-based data analyses of an ocean acidification mesocosm experiment.
Supervisors: Prof. Andreas Oschlies, Dr. Markus Schartau

LMU (Ludwig Maximilians University), Munich, Germany

MSc in Geophysics
October 2010 - March 2013
Thesis: Rayleigh numbers and earth convection modeling.
Supervisor: Prof. Hans-Peter Bunge

D.Y Patil University, Pune, India

Bachelor of Technology (B.Tech) in Bioinformatics
July 2005 - September 2009
Thesis: Phylogenetic analysis and 3D structure prediction of NS-5 protein in Mosquito-borne flaviviruses using bioinformatics tools.

RESEARCH INTERESTS

- Optimality-based modelling of phytoplankton.
- Bayesian statistics.
- Data assimilation.
- Methods of parameter optimization and uncertainty estimation.
- Nutrients limitation of primary production in the global ocean.
- Skill assessment of marine ecosystem models.
- Phytoplankton functional types (PFTs).
- Field experiments on marine stressors

ACADEMIC & RESEARCH and PROFESSIONAL EXPERIENCE

2013 – present	PhD Candidate, GEOMAR <ul style="list-style-type: none"> • Establishing a workbench in R for data-model synthesis study. • Applying different model approaches to investigate effects of abiotic stressors (e.g. pH/CO₂) on phytoplankton dynamics during diverse mesocosm experiments. • Resolving phytoplankton functional types in a marine ecosystem model. • Explaining variability seen in calcification and total alkalinity in an ocean acidification mesocosm experiment using an optimality-based model approach. • A Bayesian approach to data assimilation (inverse method). • Model parameter optimization (stochastic and gradient-based approaches). • Estimation of model parameter uncertainties using Markov Chain Monte Carlo (MCMC) methods.
2012 – 2013	Robotic programmer in a biotechnology project at Europe's one of the leading space company, Kayser-Threde GmbH (OHB System AG now), Munich, Germany

Shubham Krishna

- 2010 – 2013 Master's student, Department of Geophysics, LMU
- A modelling approach to solve and visualize earth's mantle convection problem using Navier-Stokes equations and investigate relationship between Nusselt number and Rayleigh number under varying viscosities and thermal conductivities.
 - Solving an inverse problem to investigate the use of geomagnetic signatures for navigation by animals in the Pacific Ocean.

MEMBERSHIPS

- 2014 – 2016 Member of Helmholtz Juniors (HeJu), the representation of PhD students at the Helmholtz association (Head office) in Berlin, Germany.
- 2014 – 2015 Speaker of HeJu communication group.
- 2014 – 2016 PhD representative at GEOMAR.

PUBLICATIONS/MANUSCRIPTS

Articles in scientific journals

- Krishna, S. and Schartau, M.: *A data-model synthesis to explain variability in calcification observed during a CO₂ perturbation mesocosm experiment*, Biogeosciences, 14, 1857-1882, 2017, doi:10.5194/bg-14-1857-2017.
- Schartau, M., Wallhead, P., Hemmings, J., Löptien, U., Kriest, I., Krishna, S., Ward, B. A., Slawig, T., and Oschlies, A.: *Reviews and syntheses: Parameter identification in marine planktonic ecosystem modelling*, Biogeosciences, 14, 1647-1701, 2017, doi:10.5194/bg-14-1647-2017.

Articles in preparation

- Krishna, S., Pahlow, M., and Schartau, M.: *Comparison of two carbon-nitrogen regulatory models calibrated with mesocosms data.*
- Krishna, S. and Schartau, M.: *Metrics for the estimation of parameters of a plankton ecosystem model with mesocosm data.*

CONFERENCES/SUMMER SCHOOLS ATTENDED

- IMBER ClimEco4 Summer School, August 2014, Shanghai, China.
- Data-Assimilation in Biogeochemical cycles autumn school, September 2014, Trieste, Italy.
- MEOPAR/ GOV marine ecosystem analysis and prediction workshop and MEAP-TT Meeting, June 2015, Halifax, Canada.
- 'Future Ocean' symposium, September 2015, New York, USA.
- Trait-based Approaches to Ocean Life international workshop, October 2015, Waterville valley resort, New Hampshire, USA.

Software skills

- Numerical computing: R, Fortran, MATLAB, C++, C, Java
- Operating Systems: Unix (Mac), Linux, Windows
- General: LaTeX, Microsoft office

Languages

Hindi (Mother tongue), English (Fluent), German (Level: B2), Spanish (Level: A1)

References

- Prof. Andreas Oschlies, Head of research unit Biogeochemical Modelling, GEOMAR (Primary supervisor). E-mail: aoschlies@geomar.de
- Dr. Markus Schartau, research unit Biogeochemical Modelling, GEOMAR (thesis supervisor). E-mail: mschartau@geomar.de.
- Dr. Markus Pahlow, research unit Biogeochemical Modelling, GEOMAR. E-mail: mpahlow@geomar.de.

Shubham Krishna

Declaration of Authorship

I, Shubham Krishna, declare that this thesis titled, “Model assessment and model-based data analyses of an ocean acidification mesocosm experiment” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly by me while registered as a PhD student at the Christian-Albrechts University, Kiel.
- And it has been prepared following the “Rules of Good Scientific Practice” of the German Research Foundation (DFG).
- Where I have used data of other studies, those are referenced.
- Where I have quoted from the work of others, the source is always given.
- I have acknowledged all main sources of help.
- I have made clear in author contributions what all work was done by me and what was done by others.

Signed:

Date:
