Lake ecosystems in a changing climate : strategy for coupled hydrodynamic and biogeochemical modelling

Francesco Piccioni

THÈSE DE DOCTORAT

de l'École des Ponts ParisTech

Lake ecosystems in a changing climate: strategy for coupled hydrodynamic and biogeochemical modelling

École doctorale, Science Ingénierie et Environnement

Spécialité du doctorat : Sciences et Techniques de l'Environnement

Thèse préparée au sein du Laboratoire Eau Environnement et Systèmes Urbains de l'Ecole des Ponts ParisTech et du Laboratoire MISTEA de INRAe Montpellier

Thèse soutenue le 14 septembre 2021, par :

Francesco PICCIONI

Composition du jury :

Marieke Anna FRASSL Researcher, Bundesanstalt für Gewässerkunde (DE)	*Rapporteur*
Tom SHATWELL Researcher, Helmholtz-Zentrum für Umweltforschung (DE)	*Rapporteur*
Damien BOUFFARD Directeur de recherche, EAWAG (CH)	*Examinateur*
Pierre-Alain DANIS Chargé de recherche, Office Français de la Biodiversité	*Examinateur*
Magali JODEAU Ingénieur-Chercheur, EDF R&D, LHSV	*Examinatrice*
Delphine LATOUR Maitre de conférence, Université Clermont-Auvergne	*Examinatrice*
Céline CASENAVE Chargée de recherche, UMR MISTEA – INRAE	Co-*Directrice de thèse*
Brigitte VINÇON-LEITE Directrice de recherche, LEESU – ENPC	*Directrice de thèse*

Abstract

The ecological state of freshwater ecosystems worldwide has deteriorated along the past decades. Anthropogenic pressures have altered their physical and biogeochemical dynamics, acting both within their watershed and on the climatic conditions. Eutrophication and climate change contributed to the increase of algal blooms, and in particular of toxic cyanobacteria blooms, which currently constitute one of the main concern in the management of water resources.

With the advance of urbanization, an increasing number of lakes are located in metropolitan areas. The high loads of nutrients and pollutants coming from the watershed often lead urban lakes to eutrophic conditions and cyanobacteria blooms, which cause bathing bans and restrictions for aquatic sports. Responsive surveys and long-term climate change impact studies are essential for the management of such sites, but rarely addressed.

In this respect, modelling tools are of central importance to better understand the functioning of aquatic ecosystems, the factors triggering harmful algal blooms, and to support the management of water resources. However, aquatic ecological models are often complex and highly parametrized, and their implementation and calibration are challenging. Automated strategies for parameter calibration are available but are rarely applied. Furthermore, data from traditional periodical limnological survey do not allow to test the models on dynamics quicker than the span between two successive campaigns, and to thoroughly assess the uncertainty of their outcomes.

In this context, this PhD thesis focuses on the use of deterministic models to reproduce the thermal dynamics and phytoplankton dynamics, notably cyanobacteria, in a small and shallow urban lake on different time-scales. To do so, two coupled hydrodynamic and biogeochemical three-dimensional (3D) models are implemented and analysed. The models used here are the FLOW and BLOOM modules from the Delft3D modelling suite, and the biogeochemical library Aquatic EcoDynamics coupled with the hydrodynamic model TELEMAC3D. The models are applied on Lake Champs-sur-Marne, an urban lake located in the East of Paris that suffers from strong cyanobacterial blooms and for which an extensive data set is available.

This work aims to address in detail three strategic elements in lake ecosystem modelling:

(i) The impact of climate change on the thermal regime of small and shallow lakes, and its relation to cyanobacterial growth. This is assessed through long-term 3D hydrodynamic simulations that allowed to hindcast the evolution of the study site during the past six decades.

(ii) The applicability and the benefits of automated calibration for complex bio-

geochemical models. This is done through an innovative methodology for parameter estimation: Approximate Bayesian Computation (ABC), tested here for the first time on a complex, highly-parametrized model.

(iii) The coupling and the feedbacks between hydrodynamic and biogeochemical models focusing on different time scales, and the importance of an extensive data set, that includes continuous high-frequency observations.

The results show that the impact of climate change on small and shallow lakes can be severe, with consequences on the stratification dynamics and that thermal conditions increasingly favourable for cyanobacterial growth have established over time in the study site. This suggests that cyanobacteria dominance could become a widespread issue in the near future, if such trends are confirmed. Furthermore, this work proves that automated calibration strategies, and ABC in particular, can be profitably applied to complex physically-based biogeochemical models in order to improve their results over the period chosen for calibration. Eventually, this work also highlights the importance of an extensive data set to set-up a coupled 3D hydrodynamic / biogeochemical model, and analyse and exploit its results over different time scales.

Key words: *Lake, Hydrodynamics, Ecological Modelling, Phytoplankton, Cyanobacteria, Climate change, Automated calibration, Approximate Bayesian Computation*

Résumé

L'état écologique des écosystèmes d'eau douce s'est détérioré au cours des dernières décennies. Les pressions anthropiques ont modifié leurs dynamiques physiques et biogéochimiques en agissant à la fois au sein de leur bassin versant et sur les conditions climatiques. L'eutrophisation et le changement climatique ont contribué à l'augmentation des proliférations phytoplanctoniques, notamment des cyanobactéries, qui constituent aujourd'hui une préoccupation majeure pour la gestion des ressources en eau.

Avec l'avancée de l'urbanisation, un nombre croissant de lacs se trouve dans des zones métropolitaines, où les apports élevés de nutriments et de polluants en provenance du bassin versant peuvent favoriser les proliférations de cyanobactéries. Des systèmes de mesure en temps réel peuvent faciliter la surveillance à court terme de ces milieux. Pour une gestion durable à long terme, des études d'impact du changement climatique sont indispensables.

Les modèles déterministes sont essentiels pour améliorer la compréhension du fonctionnement des écosystèmes aquatiques, des facteurs de contrôle des efflorescences phytoplanctoniques et pour optimiser la gestion des ressources en eau. Cependant, les modèles écologiques aquatiques sont souvent complexes et hautement paramétrés. Leur mise en œuvre et leur calage sont des tâches complexes, rarement automatisées. De plus, les données issues des suivis limnologiques traditionnels, basés sur des campagnes de mesure périodiques, ne permettent pas de tester les modèles sur des dynamiques rapides, ni d'évaluer de manière approfondie l'incertitude de leurs résultats.

Dans ce contexte, cette thèse porte sur l'utilisation de modèles déterministes pour reproduire le fonctionnement thermique et la dynamique du phytoplancton, notamment des cyanobactéries, d'un lac urbain peu profond. Pour ce faire, deux modèles couplés hydrodynamique et biogéochimique tridimensionnels (3D) sont mis en œuvre et leurs résultats analysés : les modules FLOW et BLOOM de la suite Delft3D, et la librairie biogéochimique Aquatic EcoDynamics couplée au modèle hydrodynamique TELEMAC3D. Le site d'étude est le lac de Champs-sur-Marne, un lac urbain situé à l'Est de Paris caractérisé par de fortes proliférations de cyanobactéries et pour lequel un grand jeu de données est disponible.

Ce travail vise à analyser trois éléments stratégiques pour la modélisation des écosystèmes lacustres:

(i) L'impact du changement climatique sur le régime thermique des lacs et ses conséquences sur le développement des cyanobactéries, au travers de simulations hydrodynamiques 3D et la reconstitution de l'évolution thermique du site d'étude au cours des six dernières décennies.

(ii) Les bénéfices du calage automatique pour des modèles biogéochimiques complexes, par l'application d'une méthodologie innovante pour l'estimation des paramètres: Approximate Bayesian Computation (ABC).

(iii) Les effets de différentes approches de modélisation et de couplage entre modèles hydrodynamiques et biogéochimiques, et les avantages des observations en continu en haute fréquence pour les évaluer sur différentes échelles temporelles.

Les résultats de ce travail montrent que l'impact du changement climatique sur les lacs urbains sont importants, avec de fortes conséquences sur la dynamique de stratification et des conditions thermiques de plus en plus favorables aux cyanobactéries. Si de telles tendances se confirment, le problème posé par la dominance des cyanobactéries, pourrait encore s'aggraver dans un proche avenir. En outre, ce travail illustre l'intérêt de méthodes de calage automatique, l'ABC en particulier, pour améliorer les résultats de modèles biogéochimiques déterministes complexes sur la période choisie pour le calage. Enfin, ce travail met également en évidence l'importance d'un jeu de données étendu pour mettre en oeuvre un modèle couplé hydrodynamique/biogéochimique 3D et analyser ses résultats à différentes échelles de temps et d'espace.

Key words: *Lac, Hydrodynamique, Modélisation écologique, Phytoplancton, Cyanobactéries, Changement climatique, Calage automatique, Approximate Bayesian Computation*

Acknowledgements

I would like, first of all, to express my deepest gratitude to my thesis Directors Brigitte Vinçon-Leite and Céline Casenave: your close and valuable supervision, your continued guidance and your scientific advice were indispensable for this work to be accomplished.

I would also like to thank the members of my thesis committee, Antoine Rousseau, Olivier Boutron and Alexandrine Pannard for their insightful advice, which helped strengthen the robustness of this work. A special thanks goes to Minh-Hoang Le, for his crucial scientific support.

I also wish to thank Tom Shatwell and Marieke A. Frassl for their thorough and stimulating review of this dissertation. It has been a privilege to discuss my work with you. Thank you also to Magali Jodeau, Delphine Latour, Pierre-Alain Danis and Damien Bouffard for their role as examiners.

My appreciation also goes to École des Ponts ParisTech and to the ANR ANSWER Project for funding my doctoral scholarship. I would further like to thank École des Ponts ParisTech for granting the academic and working environment, crucial for scientific research, that allowed this work to be realized.

I also wish to thank the *Base de loisirs de Champs-sur-Marne* (CD93) for their logistic support in the field campaigns.

To the teams at LEESU and MISTEA, that both welcomed me with warmth and solid scientific knowledge, also goes a good chunk of my gratitude. Thank you Bruno for your scientific advice and for being a constant test for my French comprehension skills. Thank you Mohamed, Philippe and Nicolas for your vital and continued help with the field campaigns, and thank you Annick for always being so helpful. Daniel, sharing the office with you has been a pleasure and a honour.

To all my fellow doctoral and post-doctoral colleagues, thank you all for your support, for the *apéros* at *Le Descartes* and, of course, for the countless coffee breaks. Najwa, thank you for your support during all of our conferences and for constantly setting an example of commitment and integrity. Natalie, thank you for always seeming so relaxed, showing me that a PhD can be finished with calm and ease. To the Brazilian crew – Guilherme, Natalia and Alexsandro: these years would not have been the same without you. Giulia and Riccardo, your support as PhD researchers was much apprechiated, no matter how distant. Nicola, thank you greatly for showing up when it was mostly needed and least expected.

Eventually, Domitille, my *bridge over troubled waters*, without you none of this would have been possible.

Paris, 2021

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Freshwater only constitutes 2.5% of the total water volume contained in the hydrosphere. A large fraction, around 68.7%, of the global freshwater volume is stocked in glaciers and ice caps, or is ground water (roughly 30%, Shiklomanov & Rodda 2004). Lakes account for around 0.26% of the global freshwater volume (or 0.007% of the total water volume, Shiklomanov & Rodda 2004), but constitute the main stock of readily available surface freshwater on Earth. Freshwater sustains all life on land. Lakes are therefore ecosystems of utmost importance, which goes beyond their natural and aesthetic significance. They are an essential source of biodiversity (Balian et al. 2008, Sayer et al. 2009, Hassall 2014) and are vital for human society, as they serve drinking water purposes, hydropower production, industrial process cooling and, depending on their size, transport.

Despite their importance, inland water bodies and, more broadly, estuarine/-coastal environments have been widely deteriorating along the last decades, mainly due to climate change and eutrophication. Eutrophication can be a natural phenomenon: the geologically-slow ageing process of lakes, during which the water body becomes nutrient-enriched, more biologically productive, and slowly fills in to become a pond or marsh (Rast & Thornton 1996, Anderson et al. 2002). However, eutrophication processes in freshwater bodies have been greatly accelerated in the last decades. This is due to various anthropogenic stressors that modified lakes environmental status and natural biogeochemical cycles of nutrients within a watershed (Le Moal et al. 2019).

One of the main symptoms of eutrophication are harmful algal blooms (HABs). The Encyclopedia of Aquatic Ecotoxycology defines algal blooms as "a rapid increase or accumulation of free-floating or attached eukaryotic algae or cyanobacteria in a surface waterbody", while harmful algal blooms as "an algal/cyanobacterial bloom that has harmful socioeconomic or ecological effects" (Watson & Molot 2013).

The main concern around HABs in inland waters is related to the phylum of cyanobacteria, often predominant in this type of events and capable to produce toxins that concern both animals and humans. In terms of their toxicological target, cyanobacterial toxins are hepatotoxins (which affect the digestive tract, liver and at some extent lungs in humans), neurotoxins (brain is affected), cytotoxins (which can cause widespread necrotic injury in mammals, e.g. liver, kidneys, lungs, spleen, intestine), dermatotoxins and irritant toxins (Codd 2000). In recognition to this problem, the World Health Organisation (WHO) first published in 1998 a

provisional drinking-water guideline value of 1 μg l^{-1} for one very common cyanotoxin, microcystin-LR (MCYST-LR) (WHO, 1999), from which various countries (France included) derived national guidelines and restrictions on the amount of cyanotoxins in water when directly related to human activities (Ibelings et al. 2015). These restrictions may lead to bathing-bans or withdrawal prohibition and can cause economic losses when the water body is subject to touristic or recreational activities.

The occurrence of HABs has been documented by the scientific community for over a century (Francis, 1878), but during the last decades climate change and nutrient enrichment have increased the frequency of HABs development on a global scale (Hallegraeff 1993, Paerl & Paul 2012, Ho et al. 2019). Cyanobacteria in particular, present a higher optimum temperature (>25 °C) when compared to other phytoplankton groups, and the general view is that they will benefit of higher water temperatures and enhanced stratification, with a surface layer favourable to their dominance (Elliott 2012, Yan et al. 2017).

An increase in the occurrence of cyanobacteria blooms is documented at various latitudes and under different lake morphologies. Lake Erie (USA) and Lake Taihu (China) are two of the most relevant basins worldwide to have recently experienced several strong algal blooms. Europe makes no exception to this process and even in France various lakes are experiencing strong eutrophication conditions marked with HABs (Briand et al. 2002, Latour et al. 2007).

Climate change is therefore expected to further deteriorate the ecological status of a number of lakes worldwide that already suffer from eutrophication. In fact, lakes have experienced considerable warming along the past decades (O'Reilly et al. 2015, Schmid et al. 2014, Schneider & Hook 2010, Piccolroaz et al. 2020), sometimes even accelerated in respect to the surrounding areas (Schneider et al. 2009).

Lakes and reservoirs represent 3.7% of the Earth's non-glaciated continental area (Verpoorter et al. 2014). The global areal extent of lakes and impoundments is dominated by millions of water bodies smaller than 1 km^2 (Downing et al. 2006). However, their role in climate change studies has often been overlooked in favour of deeper larger water bodies. Urban lakes are among these small water bodies. They are generally man-made water bodies, often formed after excavation or mining, especially in alluvial plains (Tavernini et al. 2009). Most of them are small and shallow and they frequently do not present any inlet or outlet, being mostly fed by groundwater. In 2013, the European demand for sand and gravel generated an estimated annual turnover of 15 billion euros and employment for 200 000 people. Because of the demand for excavation materials, they are a category of water bodies growing in number and importance. The formation of gravel-pit lakes can affect the connected ground-water and increase evaporation. Once created they have to be correctly managed (Søndergaard et al. 2018).

With the advance of urbanization, the presence of aquatic environments has become a key feature for the improvement of life quality in the urban landscape (Frumkin et al. 2017, van den Bosch & Sang 2017). Urban lakes grant valuable ecosystem services and contribute to the preservation of biodiversity (Frumkin et al. 2017, Hill et al. 2017, Hassall 2014, Higgins et al. 2019). They also grant recreational and educational activities and contribute to a more pleasant landscape. However, as they are often located in vast metropolitan areas, they are particularly affected by high loads of nutrients and pollutants coming from the watershed. In addition,

they are characterized by particular nutrient sources such as waterbirds droppings or sediment release that can ultimately concur in causing eutrophication (Klimaszyk & Rzymski 2016, Naselli-Flores 2008).

The recognition of eutrophication as one of the main issues worldwide affecting water resources and the evidence of the on-going climate change made the modelling of thermal and biological processes in water bodies of great scientific interest. Aquatic ecological models can be used to test the current knowledge on the biogeochemical cycle, and to provide information for management policies in water resources. However, due to the complexity of the interactions between hydrodynamic and biogeochemical processes, numerous factors interplay to create the conditions for phytoplankton blooms. In turn, this makes the modelling of these processes extremely challenging. Stratification of the water column, light availability, resuspension of nutrients from sediments, and recycling of nutrients by bacterial and phytoplankton communities, wind-induced mixing and the morphology of the water body are some of the elements that interplay in the occurrence of HABs.

To meet the need for simulations of biological processes, a vast range of different biogeochemical models arose during the last two or three decades (Mooij et al. 2007), often with similar formulations. For instance, the Ecobas Register of Ecological Models (see: `http://ecobas.org/www-server/index.html`) indicated already in 2012 that "more than 100 aquatic models have been in existence in the past two decades, many of which have similar levels of ecological complexity and intent in terms of simulating selected components of aquatic ecosystems" (Trolle et al. 2012). Ecological models strongly evolved since their first development, going from simplified "Nutrient-Phytoplankton-Zooplankton-Detritus" (NPZD) models to complex modular models that divide phytoplankton into functional groups, that allow the user for very simple or very complex configurations depending on his needs (Vinçon-Leite & Casenave 2019).

The outcomes and formulations of aquatic ecological libraries, however, are not always innovative and often not deeply reliable (Anderson 2005, Li & Hipsey 2013, Shimoda & Arhonditsis 2016). The main reason suggested by different authors for these evidences is a general lack of data when coming to phytoplankton modelling. Such data sets traditionally come from periodic field campaigns and laboratory analysis. They give a reliable picture of the seasonal dynamics of a water body, but are too coarse to capture short-term blooms, the related sudden variations in the external forcing and the diurnal cycle.

Furthermore, biogeochemical models often rely on semi-empiric equations that involve a number of parameters to be set and calibrated. The complex system of interactions between biogeochemical processes usually forces these models to be highly parameterized. Most parameters are usually difficult to measure or estimate through laboratory experiments, and, even though estimates for their values can be found in scientific literature, their range of variability is often wide (e.g. Hipsey et al. 2013, Fenocchi et al. 2019). Furthermore, the use of literature values might not be adapted to all applications. In order to optimize model results automated calibration strategies are needed, but their implementation is often hindered by observation datasets that are too coarse for their application. The optimization of aquatic ecological models, in relation to their structure and parameters, remains an issue only partially resolved.

Finally, the kinetics of primary production and of the biogeochemical cycle depend on the hydrodynamic state of the environment under consideration, for instance in terms of turbulence, water temperature, salinity, bottom friction and water column stratification. In order to thoroughly simulate the complexity of the biogeochemical cycle in aquatic ecosystems, ecological models need to be coupled to a one-, two- or three-dimensional hydrodynamic driver.

Harmful algal blooms and cyanobacterial blooms are at the core of the international French-Chinese project "Analysis of Numerical Simulations of Water Ecosystem in Response to anthropogenic environmental changes" (ANSWER), aimed at a deeper understanding of lake ecological functioning and more specifically of the drivers triggering cyanobacterial blooms. The ANSWER project is funded by the French National Research Association (ANR) and the Natural Science Foundation of China (NSFC), and gathers various researchers with different expertises: from applied biology to environmental engineering and numerical methods. This PhD thesis is an integral part of the ANSWER project, and aims at contributing through the use of mathematical models to the simulation of cyanobacterial bloom dynamics in the context of shallow urban lakes.

The main objective of this work is to make use of modelling tools to simulate the thermal, phytoplankton and cyanobacteria dynamics on different time-scales, and to identify and analyse key elements that are of strategic importance for lake ecosystem modelling in the context of climate change. This thesis focuses on the implementation and the analysis of two coupled three-dimensional (3D) hydrodynamic and biogeochemical models on a study site equipped with automated *in situ* monitoring. The two biogeochemical models are based on different approaches. The first one, Delft3D-BLOOM, assigns biomass concentration based on a specific competition principle between phytoplankton groups, taking into account external constraints. The second model is the Aquatic EcoDynamics (AED2) library, whose coupling with the hydrodynamic driver TELEMAC3D was recently done by Energie de France (EDF), and which is based on process-based mass-balance equations.

The models are applied on Lake Champs-sur-Marne, a small and shallow urban lake located in the East of Paris that suffers from strong cyanobacterial blooms. This work benefits from the extensive dataset available for the study site, that includes high-frequency measurements of relevant physical and biological variables as well as periodical surveys, and aims at highlighting its importance for lake ecosystem modelling. The exploitation of the high-frequency dataset, in particular, is crucial to this work as it allows an in depth calibration and validation of the models while taking into account sudden changes and diurnal cycles.

In particular, three key elements for the implementation and exploitation of aquatic ecological models are addressed in this work: (i) the use of a 3D hydrodynamic modelling approach to characterize the thermal and stratification dynamics in shallow lakes and highlight its spatial heterogeneity and long term evolution, (ii) the high level of parametrization of deterministic biogeochemical models and the importance of automated techniques to address their calibration and improve their results, and (iii) different approaches to model the biogeochemical cycle and different grades of coupling between hydrodynamic and biogeochemical models, and their performances on different time-scales. Each of these elements is developed in this PhD thesis in a dedicated chapter.

Manuscript structure

This manuscript is divided into eight chapters. The bibliographic review (chapter 2) describes the main physical and biogeochemical dynamics of lake ecosystems highlighting their relation with anthropogenic stressors, and discusses the main characteristics, strengths and drawbacks of aquatic ecosystem modelling. Chapter 3 describes the study site and the tools used in this work: from the measuring instrumentation to the hydrodynamic and biogeochemical models. The data set collected and used for this work, as well as its processing, are presented in chapter 4.

Chapter 5 discusses the exploitation of a 3D hydrodynamic model to assess the long-term evolution of the thermal dynamics in a small and shallow polymictic lake, with the objective of highlighting the relation between climate change, water temperature, stratification and biological productivity.

Chapter 6 addresses the issue of parameter estimation in the context of complex ecological models through automated calibration strategies, and an extensive datasets that includes high-frequency measurements. Continuous *in situ* high-frequency measurements are exploited to improve the results of a complex biogeochemical model through an innovative methodology for parameter estimation: Approximate Bayesian Computation with random forest. This is done here focusing on short-term simulation (weeks), in order to train the model to rapid dynamics that could not be captured nor validated with traditional low-frequency periodical sampling.

Chapter 7 describes the set up of a model configuration capable of correctly reproducing the thermal and phytoplankton dynamics in a small urban lake on a short- (weeks) and mid- (seasonal) term, with particular regards to cyanobacteria and total chlorophyll concentration, in the case of a recently coupled three-dimensional hydrodynamic/biogeochemical model. This allows to analyse in closer detail the approach implemented for the coupling of the two models, particularly with regard to the feedbacks between the models, the effects of different formulations for phytoplankton growth and nutrient mineralization on model results, as well as the spatial distribution of relevant physical and biological variables.

Eventually, chapter 8 closes this manuscript with the conclusions and perspectives.

Chapter 2

Bibliographical review

In this chapter, the most relevant issues concerning lake ecosystems are explored in the light of scientific literature, and the role of mathematical models in improving our global understanding of such ecosystems, in relation to the anthropogenic stressors affecting them, is outlined.

The chapter is divided into two main sections. In section 2.1, the main traits of the physiscal and biological processes in freshwater ecosystems are presented, and the effects of eutrophication and climate change on lakes are analysed focusing in particular on shallow urban water bodies.

The second section (2.2) revises the current knowledge on the modelling of aquatic ecosystems. Hydrodynamic modelling and biogeochemical modelling of lakes are reviewed separately. Eventually, the limits and drawbacks in the use of coupled hydrodynamic-aquatic ecological models are highlighted in relation to parameter calibration and data availability.

2.1 Lake ecosystems

Lakes worldwide vary greatly in size, depth, morphology and origin, and a precise definition that clearly distinguishes them from marshes, ponds or lagoons is not trivial. Following the definition given by Forel (1901), a lake is "a body of standing water occupying a basin and lacking continuity with the sea". While Timms (1992), in order to avoid confusion with marshes, defines lakes as "areas where vegetation does not protrude above the water surface", while "swamps are areas where vegetation dominates the surface". Lakes can be further categorized based on their origin, size, depth, or mixing dynamics.

Lakes are integral features of the hydrological system. Figure 2.1 shows the main features of the hydrological cycle. In this context, lakes interact directly with atmospheric water, surface water and groundwater (Reynolds 2006). Lakes and reservoirs represent 3.7% of the Earth's non-glaciated continental area (Verpoorter et al. 2014). Small lakes, that is lakes with surface smaller than 1 km^2 (Downing et al. 2006), are the dominant part of this global areal extent (Downing et al. 2006). Lakes are defined as shallow if light can potentially penetrate to its bottom (Meerhoff & Jeppesen 2009).

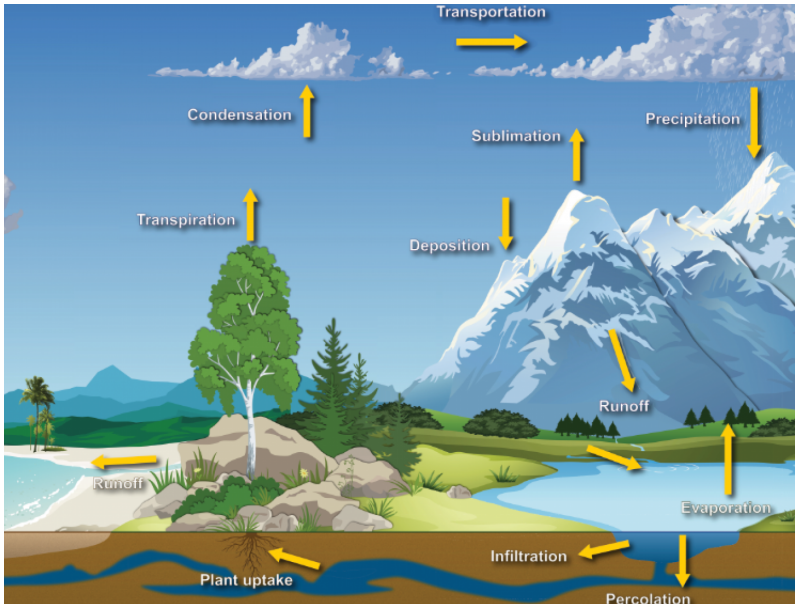


Figure 2.1 – A sketch of the hydrological cycle in a natural ecosystem (source: NOAA, `https://www.noaa.gov/education/resource-collections/freshwater/water-cycle`).

2.1.1 Physical processes in shallow lakes

2.1.1.1 Thermal stratification and mixing regime

In freshwaters, where the effect of salinity can be neglected, water density depends on temperature. It has a maximum at 4°C and decreases with higher (or lower) temperatures. Density gradients are usually extremely small over the water column, but have great importance (Reynolds 2006). Together with pressure gradients and turbulence, they determine the patterns of currents (Reynolds 2006). Taking under examination a water column with a verticallly uniform temperature above 4°C, when a net positive heat flux is entering the water body, the surface layer gets heated and its density decreases. If the density of surface water decreases, either a density gradient is built up or an existing gradient is strengthened, leading to thermal stratification. On the other hand, if the density of surface water increases, becoming higher than that of the deeper layers, the particles start to sink originating convection and mixing of the surface layer (Reynolds 2006).

Direct thermal stratification (or simply stratification) occurs during warm seasons when a positive vertical temperature gradient is created, inducing a negative density profile. During winter, stratification can also occur under particularly cold weather conditions, when the surface layers reach temperatures below 4°C (e.g. in the case of ice cover), inducing negative gradients for both temperature and density. Winter stratification (or inverse stratification) is typical of cold climates but is less frequent in temperate climates.

During thermal stratification, the water column can be divided into three main layers (Fig. 2.2): the epilimnion (the upper, warmer layer), the hypolimnion (the deeper, colder portion), and the metalimnion (the middle layer with the higher

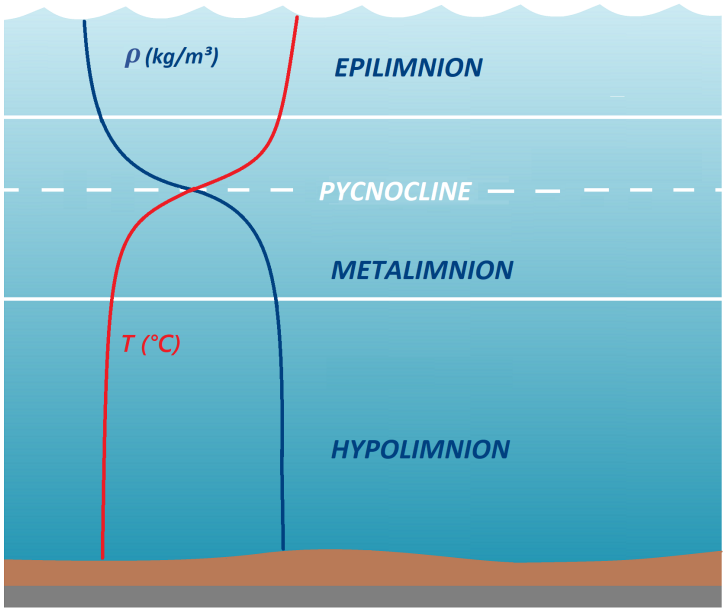


Figure 2.2 – Sketch of a stratified water column. Red line represents water temperature; blue line represents water density. Adapted from de Carvalho Bueno (2019).

temperature and density gradients). The thermocline is the depth corresponding to the maximum water temperature gradient, it lies in the metalimnion and in freshwater bodies it coincides with the pycnocline (maximum density gradient).

Most lakes alternate periods of stratification and periods of mixing. Based on the frequency of the mixing of the water column, lakes can be divided into five main categories: oligomictic lakes, monomictic lakes, dimictic lakes, polymictic lakes and amictic lakes (Lewis 1983). Oligomictic lakes are deep water bodies in temperate or warm climates that can remain stratified for a few consecutive years, where a complete overturn of the water column occurs only occasionally. Monomictic and dimictic lakes mix once and twice every year, respectively. Polymictic lakes are usually shallow water bodies that keep alternating, throughout the year, periods of thermal stratification and periods of complete mixing. Finally, amictic lakes are usually arctic lakes constantly covered by ice.

Water depth deeply influences the hydrodynamics of a water body, particularly in terms of thermal stratification dynamics. In temperate climates, if the water body is sufficiently deep the water column generally remains stratified approximately between spring and autumn, while in shallower water bodies thermal stratification can be quickly broken down by rapid variations in the meteorological conditions such as strong wind, precipitation or temperature drops. Shallow lakes in temperate climates are therefore polymictic (Lewis 1983, Condie & Webster 2002, Soulignac et al. 2017, Martinsen et al. 2019), as they keep alternating periods of complete mixing with thermal stratification.

Daily stratification events are often observed in shallow water bodies: the water column stratifies during daytime and breaks down at night due to convective mixing

(Martinsen et al. 2019). However, under favourable conditions, stratification can persist in shallow lakes for three to four consecutive weeks, or even more, before a complete overturn (e.g. Soulignac et al. 2017). This turns, between early spring and late autumn, into a constant alternation between periods of mixing, during which convective and turbulent fluxes dominate vertical transport, and periods of stable stratification, during which vertical mass transfer is inhibited.

Stratification induces a separation between the sediment and the surface layers, influencing the distribution of nutrients, oxygen and biomass over the water column. During stratification, deoxygenation can occur in the hypolimnion as the bacterial decomposition of the organic matter is oxygen-consuming and exchanges with the upper reaerated layers are inhibited by density gradients. Furthermore, due to the desoxygenation of the lake bottom layers, nutrients (phosphate in particular) are released from the sediment. During long periods of stratification, anoxic conditions and accumulation of nutrients might occur.

In shallow water bodies, the sediment is easily reached by wind-induced mixing (Bachmann et al. 2000). When mixing occurs, the replenishment of the whole water column with the nutrients released during previous stratification can act as an important internal nutrient source (Song et al. 2013, Wilhelm & Adrian 2008). The alternation between periods of stratification and mixing can result as an important internal source of nutrients.

Thermal stratification has been extensively studied throughout the 20th century, especially in the case of deep monomictic or dimictic lakes (e.g. Ullyott & Holmes 1936, Dake & Harleman 1969, Idso 1973, Imberger 1985, Boehrer & Schultze 2008, Sahoo et al. 2016, Liu et al. 2019). A greater attention has been drawn to the stratification dynamics in shallow polymictic water bodies during the past decade (Woolway et al. 2014, Hadley et al. 2014, Yang et al. 2018, Martinsen et al. 2019). In the recent years, an increasing number of papers started to analyse the impact of climate change on the thermal dynamics of such water bodies (Hadley et al. 2014, Sahoo et al. 2016, Magee & Wu 2017, Moras et al. 2019, Zhang et al. 2020, e.g.). Among shallow water bodies however, urban lakes have traditionally been mistakenly considered as completely mixed reactors (McEnroe et al. 2013, Soulignac et al. 2017).

Only a few studies addressed the spatial distribution of thermal stratification in shallow water bodies, showing that bathymetry and vegetation can originate strong spatial differences in the stratification dynamics (Vilas et al. 2017, Torma & Wu 2019). In shallow water bodies, bathymetric variations have a greater relative importance, and even small gradients in the bottom elevation can induce a complete switch in the local currents and stratification dynamics.

2.1.1.2 Wind-induced friction and horizontal patterns

Wind influences strongly the dynamics of a lake. It induces currents, affects the heat exchanges at the air-water interface and, especially in shallow water bodies, it can cause stratification breakdown, turbulent mixing and sediment resuspension.

The friction of wind over the water surface drags the surface layer causing motion. Due to the turbulent nature of the motion of fluids, the wind field acting on a certain region at the mesoscale, even when uniform, can cause strong local variability at a finer scale. The friction of wind induces ripples on the water surface, whose

amplitude depends on wind velocity, increasing in turn the shear stress of wind on the water itself. The dragging action of wind can originate spatial heterogeneity in floating tracers concentration, by accumulating them near shore. This is the case for near-surface phytoplankton biomass (World Health Organization 2003).

Under thermal stratification, the presence of a wind field can originate internal waves, that can be detected also in small and shallow lakes (Pannard et al. 2011, Soulignac et al. 2017). In shallow water bodies, strong winds can quickly breakdown thermal stratification if they overcome the buoyancy forces originated by the density gradient. The subsequent turbulence can easily reach the bottom of the water column, causing the resuspension of the sediment and of the nutrients and organic matter encapsulated in it. Bathymetric variations can induce spatial variability in the distribution of the sediment concerned by the action of the wind (Bachmann et al. 2000). Furthermore, spatial patterns in the wind field acting on a water body originated by the sheltering of the surrounding can have relevant effects even on small lakes, locally altering current patterns (Kimura et al. 2016).

2.1.2 Biological processes in lakes

2.1.2.1 Freshwater phytoplankton assemblage

Aquatic ecosystems are among the most important sources of biodiversity worldwide. For instance, freshwater systems were estimated to contain about 10% of all animal species on Earth (Balian et al. 2008).

With the word phytoplankton we refer to "collective of organisms that are adapted to spend part or all of their lives in apparent suspension in the open water of the sea, of lakes, ponds and rivers" (Reynolds 2006); phytoplankton is the autotrophic part of the plankton and a major primary producer of organic carbon in marine as well as in inland waters (Reynolds 2006). Including oceans, phytoplankton account for 1% of the photosynthetic biomass on Earth, but are nevertheless responsible for nearly 50% of global net primary production and are the primary energy source for aquatic ecosystems (Winder & Sommer 2012).

Phytoplankton are extremely diverse small organisms (usually 0.2 to 200 μm) that comprise both eukaryotes and prokaryotes; most of them are unicellular organisms, some of which form colonies and thus reach sizes visible to the naked eye. It has been estimated that more than 4000 species of phytoplankton exist in the marine ecosystem and a comparable amount in freshwater (Reynolds 2006). Due to this great diversity, phytoplankton species are often grouped into macro-categories, based either on taxonomic similarities or particular functional traits (Reynolds et al. 2002).

Phytoplankton have very low or no mobility. They are transported by the motion of water, even though some of them present flagella or specific gas vesicles that allow small migrations towards light (such as cyanobacteria). Their growth depends on several factors, among which the most important ones are light and carbon dioxide availability, water temperature and presence of nutrients like P, N and Si in biologically available forms (e.g. NO_3^-, NH_4^+ and PO_4^{3-}).

The major phytoplankton taxonomic groups present in freshwater are Cyanobacteria, Chlorophyta (Green algae), Bacillariophyta (Diatoms), Cryptophyta, Pyrrhophyta (dinoflagellates), Chrysophyta, Euglenophyta, Rhadiophyta and Dinophyta (Sheath & Wehr 2003, Watson et al. 2015).

The taxonomic composition of phytoplankton communities varies with nutrient levels according to the taxa nutrient uptake, growth, and respiration rates (Watson et al. 1997).

Under favourable circumstances in terms of light, nutrients and temperature, most phytoplankton groups can proliferate. The relation between algal growth rate and water temperature has been extensively studied through laboratory experiments (Lürling et al. 2013, Sherman et al. 2016, Paerl & Otten 2013). Some typical traits can be distinguished among taxonomic groups. Figure 2.3 shows the dependence of the growth rate from water temperature for green algae, dinoflagellates, diatoms and cyanobacteria. Diatoms generally show maximal growth rates at the lowest temperature of the lot (around 15°C); dinoflagellates follow, with slightly higher optimal temperatures (around 20°C), while green algae and cyanobacteria show similar optimal temperatures, close to 30°C (Lürling et al. 2013).

In an ecosystem where multiple species are present phytoplankton species are in competition for the available resources. The environmental conditions in terms of nutrients, temperature and light availability, together with the specific adaptation of each phytoplankton group, determine the composition of the phytoplankton assembly at a given time. As the environmental conditions evolve over the year, a succession of dominant species can develop in the phytoplankton community. However, due to the multiple and sometimes contrasting factors at interplay, the outcome of multispecies competition remains extremely difficult to predict (Huisman & Weissing 2001).

An attempt to describe phytoplankton succession in lakes with a step by step conceptual model, the PEG (Plankton Ecology Group) model, was made in the 1980s (Sommer et al. 1986). According to this model, during winter phytoplankton growth is energy-limited (by light and temperature). A first algal proliferation, also called bloom, dominated by diatoms or Cryptophyceae occurs during spring as water temperature increases and the water column stratifies. The spring bloom then rapidly ceases due to grazing, sedimentation and nutrient depletion. Follows a "clear-water" phase of equilibrium during which nutrients can be accumulated in the water body, and that persists until inedible algal species develop in large numbers. Under low grazing pressure and non-limiting nutrient concentrations summer phytoplankton taxa start developing: Cryptophyceae or inedible green algae at first, substituted later by cyanobacteria (possibly N-fixing). The model, here briefly resumed, showed some weaknesses when tested against shallow polymictic water bodies. Their unpredictability was linked to the mechanisms of wind-induced mixing that might alter through sediment resuspension the seasonal patterns of transparency and nutrient concentrations.

The importance of spring blooms for phytoplankton succession is clear both in marine and in freshwater ecosystems (Zhao et al. 2013, Spilling et al. 2018, Browning et al. 2020, Sommer et al. 1986, Reavie et al. 2016). In deep lakes, spring blooms start with the onset of thermal stratification (therefore, with the transition from strong to weak turbulent mixing) that allows algae to remain in the trophic zone for longer amounts of times (Bleiker & Schanz 1989, Peeters et al. 2007, Kienel et al. 2017); however, stratification might not be a triggering factor in shallower lakes where light can easily reach the bottom layer (Kong et al. 2021).

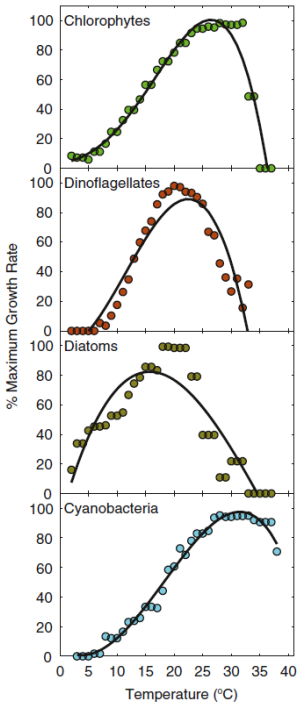


Figure 2.3 – Comparison of the temperature dependence of the growth rate for green algae (Chlorophytes), dinoflagellates, diatoms and cyanobacteria (source: Paerl & Otten (2013)).

2.1.2.2 Cyanobacteria

Cyanobacteria is a phylum belonging to the domain of bacteria. They are ancient prokaryote organisms, generally unicellular, which have adapted to colonize water environments at every latitude. Cyanobacterium-like fossils are among the oldest forms of life ever found on Earth (dating back to 3500 million years ago (Schopf 1993)).

They are autotrophs and synthesize the needed molecules through photosynthesis. They are unicellular organisms and can survive as solitary free-living cells. However, they often organize in colonies visible to the naked eye. During strong blooms they can form floating mats covering large portions of the water surface (Newton et al. 2011). Cyanobacteria contain photosynthetic pigments such as chlorophyll-a. Aside chlorophyll, they also contain accessory pigments such as phycocyanin, allophycocyanin and phycoerythrin that are specific to cyanobacteria. Namely, phycocyanin is a pigment-protein complex that operates together with chlorophyll in order to increase the efficiency of chlorophyll's oxygen production under low light conditions, and is specific to cyanobacteria (Zeece 2020). It absorbs red-orange wavelengths and gives cyanobacteria their characteristic blue-green colour (cyanobacteria are also known as blue-green algae). The presence of phycocynanin is a characteristic of cyanobacteria and it can be used by optical sensors to detect their presence and as a proxy for their abundance.

Some species of cyanobacteria have mobility capabilities, that help them to move towards light sources. However, cyanobacteria are unicellular organisms that do

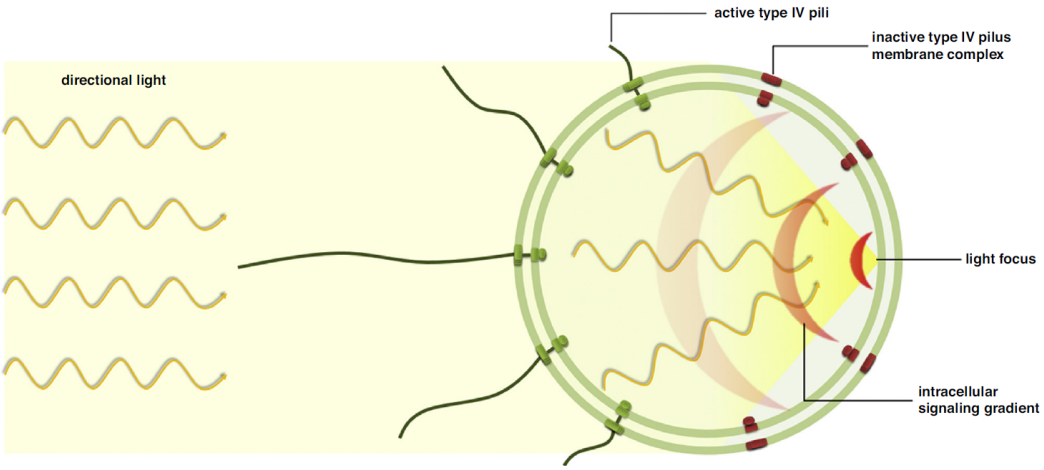


Figure 2.4 – Model for directional light sensing in the spherical unicellular cyanobacterium Synechocystis. The pili that remain active are at the side of the cell facing the light source, leading to movement towards the light (Schuergers et al. 2017).

not present flagella, and the way these movements are performed remains partially unclear (Schuergers et al. 2017).

Traditionally, cyanobacteria buoyancy is explained through the presence of gas vesicles, hollow protein-based structures that are found in various planktonik microorganisms, among which cyanobacteria. These gas vesicles are impermeable to liquid water, but highly permeable to gases and air. Their relative gas content can be regulated, allowing aquatic microbes to perform vertical migrations (Walsby 1994). With respect to other "passive" phytoplankton groups, this is a key advantage in order to reach the upper water layers where light penetrates.

However, it has also been suggested that their motion can sometimes derive from type IV pili, particular protein fibers exported by the polymerization pilin subunits, and possibly also slime extrusion. Cyanobacterial mobility is partially controlled by light; for instance, the spherical synechocystis cells (which presents type IV pili) is suggested to act as a microscopic eyeball. As shown in Figure 2.4, light impacting the cell is focused on the opposite side, which becomes four times brighter than the incoming light intensity. The sharp focal point is sensed by photoreceptors, which then transduce the signal to particular regulators (Schuergers et al. 2017). However, the link between environmental conditions and the mobility apparatus (gas vescicles or type IV pili) remains uncertain.

The greatest concern around cyanobacteria derives from their ability to produce toxins that are harmful for human health. These toxins present health hazards via several exposure routes. Several cases of animal poisoning and human health incidents associated with the toxins are detected every year worldwide, and have raised the profile of cyanobacterial blooms in the issues affecting water quality.

Some of the most common toxic cyanobacteria found in fresh water globally are *Microcystis* spp., *Cylindrospermopsis raciborskii*, *Planktothrix rubescens* and *agardhii*, *Synechococcus* spp., *Gloeotrichia* spp., *Anabaena* spp., *Lyngbya* spp., *Aphanizomenon* spp., and *Nostoc* spp. (World Health Organization 2003, Newton et al. 2011).

However, toxicity cannot be excluded for further species and genera.

Cyanobacteria can form a great variety of secondary metabolites, which exhibit various types of biological or biochemical activities, some of which have been identified as potent toxins (cyanotoxins). In terms of their toxicological target, cyanobacterial toxins are hepatotoxins (which in human affect the digestive tract, liver and at some extend lungs), neurotoxins (brain is affected), cytotoxins (which can cause widespread necrotic injury in mammals, e.g. to liver, kidneys, lungs, spleen, intestine), dermatotoxins and irritant toxins (Codd 2000). The main toxins involved in cyanobacteria blooms are summed up in Figure 2.5.

Neurotoxins act by blocking neuronal signal transmission (Zaccaroni & Scaravelli 2008). The neurotoxins in cyanobacteria include anatoxin-a and homoanatoxin-a, which are postsynaptic cholinergic nicotine agonists and neuromuscular blocking agents. These alkaloids cause staggering, gasping, and muscle fasciculations in animals and possibly death by respiratory arrest. Anatoxin-a(s), an organophosphate neurotoxin, causes hypersalivation in animals with death due to the inhibition of acetylcholinesterase. Saxitoxins (carbamate alkaloid neurotoxins) are well-known in a marine context as products of dinoflagellates. The same family of neurotoxins can be produced by a range of freshwater cyanobacteria and about 20 saxitoxin variants have been described to date. Despite their high toxicity, neurotoxins are only occasionally responsible of human intoxication, while they are more commonly dangerous to animals, that can drink polluted water (Zaccaroni & Scaravelli 2008).

Microcystins are probably the most prevalent cyanotoxin in the environment and they are present in high amounts in cyanobacterial biomass (Bláha et al. 2009). They are cyclic heptapeptides and are hepatotoxic to animals, with external signs of poisoning including weakness, pallor, cold extremities, heavy breathing, recumbency, vomiting, and diarrhoea. Liver cells are particularly susceptible to damage by microcystins. Disruption of liver structure and function occurs with haemorrhage into the liver and death by respiratory arrest (Codd 2000). Microcystins have also shown to be tumor-promoting agents, increasing the incidence of hepatic tumors in humans as well (Zaccaroni & Scaravelli 2008).

It is not possible to distinguish toxic and non-toxic genotypes within a species through the microscope (Kurmayer & Christiansen 2009). Furthermore, toxins production varies in time. The conditions triggering toxins production are still poorly understood by the scientific community (Catherine et al. 2013). It is suggested that maximum toxin production simply occurs when growth conditions are particularly favourable (Sivonen 1996), and that it is increased by exposure to zooplankton (Jang et al. 2003).

Focusing on microcystin-LR, the most common and most toxic of microcystins, the World Health Organization (WHO) derived a value of tolerably daily intake (TDI) for the assessment of risks related to human health. The TDI for microcystin-LR is 0.04 μg per kg of bodyweight per day and was used by the WHO to set the guidance value for the maximal acceptable concentration of microcystin-LR in drinking water, 1 μg L^{-1} (World Health Organization 1999).

Under low-N conditions, some groups of cyanobacteria also present the ability to fix atmospheric nitrogen. Nitrogen fixation is the reduction of dissolved N_2 gas to biologically available ammonium that can be used for growth and reproduction when the latter is not directly available. These species can play a key role in the

Cyanotoxins	LD_{50} (i.p. mouse)[b] of pure toxin (µg/kg)	Taxa known to produce the toxin(s)	Mechanism of toxicity
Protein phosphatase blockers (cyclic peptides with the amino acid ADDA)			
Microcystins in general (~60 known congeners)	45–>1000	*Microcystis*, *Planktothrix*, *Oscillatoria*, *Nostoc*	all block protein phosphatases by covalent binding and cause haemorrhaging of the liver; cumulative damage may occur
Microcystin-LR	60 (25–125)	*Anabaena*, *Anabaenopsis*	
Microcystin-YR	70	*Hapalosiphon*	
Microcystin-RR	300–600		
Nodularin	30–50	*Nodularia spumigena*	
Neurotoxins			
Anatoxin-a (alkaloid)	250	*Anabaena*, *Oscillatoria*, *Aphanizomenon*, *Cylindrospermum*	blocks post-synaptic depolarization
Anatoxin-a(s) (unique organophosphate)	40	known only from two species of *Anabaena*	blocks acetylcholinesterase
Saxitoxins (carbamate alkaloids)	10–30	*Aphanizomenon*, *Anabaena*, *Lyngbya*, *Cylindrospermopsis raciborskii*	block sodium channels
Cytotoxin			
Cylindrospermopsin (alkaloid)	2100 in 1 day 200 in 5–6 days	*Cylindrospermopsis raciborskii*	blocks protein synthesis; substantial cumulative toxicity

[a] derived from Turner et al., 1990; Kuiper-Goodman et al., 1999; Sivonen & Jones, 1999.

[b] LD_{50} = lethal $dose_{50}$ (the dose of a chemical that will, on average, kill 50% of a group of experimental animals); i.p. = intraperitoneal.

Figure 2.5 – Main toxins produced by cyanobacteria (source: World Health Organization 2003).

nutrient cycle in aquatic ecosystems, especially under N-limiting conditions. The majority of cyanobacterial taxa that form blooms in freshwater lakes are capable of fixing N, including *Anabaena, Aphanizomenon, Cylindrospermopsis, Gloeotrichia* and *Nostoc* (Cottingham et al. 2015).

Moreover, some cyanobacterial taxa are able to access pools of P in the sediments and bottom waters that are not generally available to other phytoplankton, similar to other mobile organisms that move nutrients from the sediments into the water column, such as benthic-feeding fish.

Cyanobacterial blooms can therefore influence lake nutrient cycling, its resilience, and cause regime shifts by tapping into pools of nitrogen and phosphorus not usually accessible to phytoplankton. In particular, cyanobacteria can increase available pools of the limiting nutrients nitrogen and phosphorus.

2.1.3 Anthropogenic stressors

2.1.3.1 Climate change

Climate change is globally considered as a major concern in policy making and development planning. In the last few decades, it has been widely studied by the scientific community and by international organizations, and its relation with human activities and greenhouse gases emissions is a factual evidence (Jamet & Corfee-Morlot 2009, Gosling et al. 2011, IPCC 2014). In the *Fifth Assessment Report* published by the Intergovernmental Panel on Climate Change (IPCC), it is stated that "warming of the climate system is unequivocal"and that "human influence on the climate system is clear, and recent anthropogenic emissions of greenhouse gases are the highest in history, and recent climate changes have had widespread impacts on human and natural systems" (IPCC 2014).

The public awareness of the existence of climate change and of its possible consequences started raising in particular after the publishing of the first IPCC assessment report (IPCC 1990). However, the idea that the global climate might be sensitive to the concentration of gases inducing a greenhouse effect dates back to the 19th century. It was first suggested in 1824 that the atmosphere itself could have a warming effect on the Earth by acting like a greenhouse. A few decades later, John Tyndall noticed that atmospheric variations of the amount of some active components (such as H_2O and CO_2) could have provoked "all the mutations of climate which the researches of geologists reveal" (Le Treut et al. 2007). At the end of the 19th century, it was suggested by Svante Arrhenius that a 40% variation of atmospheric CO_2 concentration might trigger glacial advances and retreats. Such hypothesis would have been confirmed later on by paleo-climatic research (Le Treut et al. 2007).

The link between human activities and climatic changes was first made in the 20th century. In particular, in 1938 G. S. Callendar suggested a warming effect of fossil fuel emissions on the Earth's climate: "as man is now changing the composition of the atmosphere at a rate which must be very exceptional on the geological time scale, it is natural to seek for the probable effects of such a change" (Callendar 1938, Le Treut et al. 2007). Finally, in the 1970s the main factors in climate change science were fully assessed, and all the main greenhouse gases (i.e. CO_2, CH_4, NO_2

and CFCs) were identified (Le Treut et al. 2007).

The general impacts of climate change on the Earth system and on human activities have been widely studied throughout the past couple of decades (e.g. Gosling et al. 2011, Chapman et al. 2017, Cronin et al. 2018, Venäläinen et al. 2020). The impact of climate change on aquatic ecosystems is of particular interest because of their vital importance for human society. A wide scientific production demonstrated that global warming is having strong effects on their biogeochemical and physical dynamics (Paerl & Huisman 2008, Schneider & Hook 2010, Anneville et al. 2013, Schmid et al. 2014, O'Reilly et al. 2015, Piccolroaz et al. 2020).

The scientific production dealing with lakes and climate change has traditionally focused mainly on large deep water bodies. Climate change is analysed either through series of observations, when sufficiently long, or through modelling approaches. Scientific papers are focused either on the temporal evolution of the thermal structure of water bodies or on the impact of climate change on the planktonic community. Rarely the two issues are addressed simultaneously.

Climate change influences the thermal structure of freshwater bodies mainly through warmer water temperatures and stronger thermal stratification and, when present, ice-cover duration. The great interest in analysing water temperature variations lies in the importance of water temperature for both the overall hydrodynamics of lakes and for their ecology. Furthermore, lakes sometimes show greater warming rates than those of the surrounding air temperature, and can be regarded as "sentinels" of climate change (Adrian R. et al. 2009).

Large water bodies often show a long history of measurements that can be analysed over the long term for trends. This is the case, for instance, of the Laurentian Great Lakes and of various European alpine lakes (Livingstone 2003, Austin & Colman 2007, Dobiesz & Lester 2009, Anneville et al. 2013). For instance, through an analysis of a 100-years long series of data, Austin & Colman (2007) pointed out that surface water temperature in Lake Superior increased at a considerable rate of 0.27°C per decade until 1980, with a consistent shift afterwards (to 1.1°C per decade). Summer water temperatures of lakes Ontario and Erie were found to be warming at a slightly lower rate, of 0.84°C and 0.48°C per decade, respectively, between 1968 and 2002 (Dobiesz & Lester 2009). Similar studies can be found for various other freshwater bodies in the world (Livingstone 2003, Mooij et al. 2005, Anneville et al. 2013).

Series of *in situ* observations long enough to carry out climate change trend analysis are not common. Recently, data from satellites have been widely used to widen the spectrum of study sites in climate change analysis (Livingstone 2003, O'Reilly et al. 2015, Sharma et al. 2015, Pareeth et al. 2016, Winslow et al. 2018). In particular, summer lake water temperature was analysed at a global scale (235 water bodies) using satellite data collected between 1985 and 2009 (O'Reilly et al. 2015). Most sites showed significant warming trends, with only a few instances of negative significant trends. The warming of water temperature was found to be geographically highly variable, with a global mean of 0.34°C per decade and maximum values around 1.3°C per decade O'Reilly et al. (2015).

Climate change impacts the stratification dynamics. The evolution of thermal stratification has been studied through *in situ* measurements and modelling ap-

proaches. Climate change generally induces stronger thermal stratification for longer periods. In deep lakes that stratify stably during summer, surface temperatures usually show greater warming trends than hypolimnetic temperatures. With higher epilimnion temperatures, the vertical density gradient increases leading to stronger stratification profiles (Livingstone 2003, Vinçon-Leite et al. 2014, Butcher et al. 2015).

The onset of summer stratification arrives earlier in the year, while its breakdown is delayed (Peeters et al. 2007). In shallow water bodies, an analogous definition of onset and breakdown of stratification is not possible, due to their polymictic character. However, it was recently shown that climate change induces stronger and longer thermal stratification in shallow lakes as well (Hadley et al. 2014, Magee & Wu 2017, Moras et al. 2019).

Water bodies located at high latitudes at the transition between temperate and polar climates, and subject to seasonal freezing and inverted thermal stratification might experience longer ice-free seasons in the next decades. In some cases they even might transition to a complete absence of ice cover (Sharma et al. 2019).

The strong changes induced in the thermal regime of lakes can heavily affect their ecological state. Climate change, in combination with additional nutrient loads coming from various human activities can impact the biogeochemical cycle as well as the composition of the species present in a water body. As a result of accelerated and generalized eutrophication, the occurrence of harmful algal blooms has sensibly increased during the past decades, and this trend is predicted to continue in the future (Hallegraeff 1993, Paerl & Paul 2012). Numerous studies also suggest that HABs are more and more likely to be dominated by cyanobacteria (Elliott 2012, Yan et al. 2017).

2.1.3.2 Eutrophication

Eutrophication can be a natural phenomenon; it can be generally defined as the process by which a water body becomes enriched in dissolved nutrients that stimulate the growth of aquatic plants, usually resulting in the depletion of oxygen. Currently, the term has been used to identify the cultural or accelerated eutrophication of lakes, rivers, estuaries, and marine waters, caused by anthropogenic pressure and non-natural nutrient enrichment (Anderson et al. 2002, Le Moal et al. 2019).

Eutrophication is strongly accelerated by certain anthropogenic stressors. Human society has influenced water resources deeply and in numerous way, from river rectification to dam building. One of the most direct impacts of human activities on water quality is the release of sewage, agricultural, industrial and urban wastewater, and the diffusion of agricultural fertilizers and chemicals in water bodies via water run-off (Codd 2000).

The state of eutrophication of a water body can be described through the trophic state index (Carlson 1977), that divides lakes into three categories based on their nutrient content and biological productivity: oligotrophic, mesotrophic or eutrophic (sometimes hypereutrophic) lakes.

The nutrients causing eutrophication are phosphorus (P) and nitrogen (N), that are the two main nutrients necessary for the growth of phytoplankton. The composition of phytoplankton in terms of these nutrients follows a given ratio N:P.

For marine ecosystems this ratio is considered to be N:P $\sim$ 16 (Redfield ratio), while it is more variable for freshwaters. Generally, phosphorus is the least abundant among the two and is often considered to be the limiting nutrient of phytoplankton in lakes (Granéli et al. 2008).

Because of the general belief that phosphorus limits primary production in lakes, its concentration was chosen by various countries as the main (sometimes, the only) indicator of the ecological state of water bodies. During the end of the 20th century, most nutrient-reducing policies have therefore focused on the control of anthropogenic phosphorus release (Poikane et al. 2019). There are evidences, however, that the monitoring of phosphorus alone is not a sufficient criteria to determine management policies at the watershed level, and that nitrogen limitation occurs more frequently than it is generally expected (Abell et al. 2010, Poikane et al. 2019).

The N:P ratio in a water body often reflects the source of nutrients. The ratio tends to be high in oligotrophic lakes that receive nutrients from natural watersheds, while in mesotrophic and eutrophic lakes incoming nutrients derive from various sources that have lower average N:P ratios (Downing & McCauley 1992). A water body can generally be considered Phosphorus-limited if N:P>15:1, while it is likely to be nitrogen limited if N:P<7:1 (Abell et al. 2010).

2.1.3.3 Algal blooms

Nutrient-rich ecosystems are usually biologically productive. Harmful algal blooms (HABs) are the uncontrolled and rapid proliferation of potentially toxic algae, producing toxic or harmful effects on people, animals or, more generally, to the entire ecosystem itself. An escalation in the occurrence of algal blooms was registered in the past decades both in coastal areas and inland waters (Anderson et al. 2002, Heisler et al. 2008, Paerl & Paul 2012), making HABs a major issue in marine, brackish, and freshwater systems worldwide (Hallegraeff 1993, Watson et al. 2015). HABs are usually more severe and frequent in eutrophic waters, but they can also occur intermittently in less productive systems, especially those with some degree of human impact such as acidification or organic loading (Watson et al. 2015, Carey et al. 2012).

Uncontrolled algal blooms represent a threat for the structure and biodiversity of aquatic ecosystems, for drinking and irrigation water supplies, recreational activities, tourism, fisheries, and the agriculture and power industries. HAB species include representatives from most major taxonomic groups: Cyanobacteria, Chlorophyta, Diatoms (Bacillariophyta), Chrysophyta, Euglenophyta, Rhadiophyta and Dinophyta (Hallegraeff 1993, Watson et al. 2015).

When HABs contain species able to produce toxins, they become a threat directly for human health. Relatively few freshwater species can produce harmful toxins, and all of those known to date are cyanobacteria. Roughly 50 cyanobacteria species have been proved to produce such toxins, however, this number is likely to be underestimated (Watson et al. 2015). Because of their potential toxicity, cyanobacteria have become one of the main concerns in water resources management worldwide (Paerl & Huisman 2008, Paerl & Paul 2012, Wagner & Erickson 2017). Figure 2.6 shows some intense algal bloom occurred in Lake Erie in September 2017.

In temperate climates, oligotrophic environments support minimal phytoplankton

Figure 2.6 – Photos of a harmful algal bloom occurred in Lake Erie in September 2017 (source: NOAA, `https://oceanservice.noaa.gov/facts/hab-solutions.html`).

biomass and diversity and are usually dominated by Cryptophyta or by cyanobacterial or chlorophyte picoplankton. Mesotrophic systems show a higher diversity. Most algal taxonomic groups are represented over the growing season (e.g. diatoms, green algae, Cryptophyta, Dinophyta and Cyanobacteria). Finally, eutrophic and hypereutrophic lakes can sustain very high algal biomass, and are often dominated by very few taxa, such as Cyanobacteria, diatoms or dinoflagellates (Watson et al. 1997).

Under favourable circumstances in terms of light, nutrients and temperature, most phytoplankton groups can proliferate causing HABs. In freshwater ecosystems, HABs are mostly lead by cyanobacteria or, less frequently, by diatoms or green algae. Dinoflagellates are mostly dominant in marine waters (salinity $> 5\%$), and usually a minor component of phytoplankton assemblies in freshwater (Sheath & Wehr 2003). In marine ecosystems, dinoflagellates can develop blooms with significant impacts on both economy and human health as they are able to induce human intoxication (via vectors such as molluscs) having lethal effects in some cases. However, toxic dinoflagellate species have not been recorded in freshwater ecosystems to date (Fachini & Vasconcelos 2006).

HABs are symptomatic of ecosystem imbalances. Their recent augmentation is generally linked to the expanding human footprint on natural ecosystems, and to the deep environmental changes caused by anthropogenic stressors such as urbanization, cultural nutrient enrichment and climate change (Paerl & Huisman 2008, Sommer & Lengfellner 2008, Wagner & Adrian 2009). Watershed development, including changes in land-use, deforestation, and rerouting of rivers, can modify the timing and volumes of compounds entering a water body (e.g., nutrients, sediment, and pollutants), and alter its species composition by introducing aquatic invasive species. Furthermore, the timing and magnitude of hydrologic events (e.g., severe storms causing flooding or higher spring runoff, extended periods of low or no inflow) can also impact the incoming discharges in a water body (Watson et al. 2015). Climate-related changes can also affect the duration of ice cover and of the growing season, as well as the stratification and mixing dynamics, with effects on the circulation patterns, on the spatial distribution of dissolved and particulate materials and on the resuspension of sediment-released nutrients (Watson et al. 2015, Song et al. 2013).

An augmentation in the frequency of blooms dominated by cyanobacteria has been observed globally during the last couple of decades (Huisman et al. 2018).

Cyanobacteria have high optimum temperatures (>25 °C) and therefore benefit of warmer water temperatures induced by climate change (Elliott 2012). Cyanobacterial blooms may even locally increase water temperatures through the intense absorption of light. For instance, the temperatures of surface blooms in the Baltic Sea and in Lake IJsselmeer (NL) can be around 1.5°C above those of the surrounding ambient waters (Paerl & Huisman 2008). The effect of raising CO_2 atmosphere concentrations is likely to be negated by the increased bloom activity, which enhances CO_2 consumption and elevates pH levels (Paerl & Paul 2012). Furthermore, cyanobacteria are advantaged under stratified conditions due to their motion capability that allows them to float towards the surface layer, where higher water temperatures and more light are available (Carey et al. 2012, O'Neil et al. 2012).

Spring blooms are thought to be favoured by climate change, mainly through earlier stratification and warmer water temperatures (Peeters et al. 2007, Kienel et al. 2017). However the impact of climate change on these events is rarely addressed. Most climate change impact studies that deal with the ecological state of freshwater bodies focus on toxic cyanobacterial summer blooms, and do not take into account the influence of early blooms on the subsequent phytoplankton dynamics, including cyanobacteria (Huber et al. 2008).

2.1.3.4 Water bodies in urban areas

Lakes and reservoirs represent 3.7% of the Earth's non-glaciated continental area (Verpoorter et al. 2014). The global areal extent of lakes and impoundments is dominated by millions of water bodies smaller than 1 km^2 (Downing et al. 2006). However, small water bodies are often overlooked in scientific literature. For instance, the Europaen Water Framework Directive sets to 0.5 km^2 of surface area the limit to include a lake in their statistics (Lyche Solheim et al. 2019).

With the advance of urbanization, the presence of small aquatic environments has become a key feature for the improvement of life quality in the urban landscape (Frumkin et al. 2017, van den Bosch & Sang 2017). An ever-increasing number of small water bodies is either incorporated into the expanding metropolitan areas or is artificially created for aesthetic or recreational purposes.

Urban lakes are generally man-made water bodies, often formed after excavation or mining, especially in alluvial plains (Tavernini et al. 2009). In 2013, the European demand for sand and gravel generated an estimated annual turnover of 15 billion euros and employment for 200000 people (Søndergaard et al. 2018). Most urban lakes are small and shallow and they frequently do not present any inlet or outlet, being mostly fed by groundwater. These lakes provide inhabitants with various ecosystem services that can sensibly increase life quality. They grant recreational and educational activities and they contribute to a more pleasant landscape (Naselli-Flores 2008). Furthermore, they are an important source of biodiversity in strongly anthropogenized environments (Frumkin et al. 2017, Hill et al. 2017, Hassall 2014, Higgins et al. 2019).

Despite their value, urban water bodies are often in a poor ecological state (Moore et al. 2003, Noble & Hassall 2015, Søndergaard et al. 2018). Temperatures in urban areas are higher than the surrounding rural areas. The combined effect of climate change and urbanization can strongly increase the heat stress on ecosystems

located in metropolitan areas (Chapman et al. 2017). Furthermore, due to the land use of their surroundings, these water bodies are affected by high loads of nutrients and pollutants coming from the watershed (Putt et al. 2019). Moreover, wind induced mixing is often sufficient to regularly refill the water column with nutrients recycled in the sediment (Bachmann et al. 2000, Song et al. 2013). The frequent lack of inlets and outlets makes the water in these lakes rather stagnant. The combination of nutrient enrichment and high retention times lead to eutrophic conditions and favour harmful algal blooms. In particular, stagnant waters can be especially favourable to cyanobacterial growth.

However, urban lakes cannot be simply considered as completely mixed reactors (McEnroe et al. 2013). Their hydrodynamics is far more articulate than it might appear. They are polymictic water bodies with complex patterns of mixing and stratification, highly dependent on local meteorological conditions. The thermal regime is among the main drivers of phytoplankton growth and harmful algal blooms in lakes (Wilkinson et al. 2016, Paerl & Huisman 2008). As all aquatic ecosystems, urban lakes are subject to climate change. However, its impact on these ecosystems still needs to be assessed.

More generally, the role of small lakes in climate change studies has so far been overlooked. On the first hand, they might be relevant on a global scale in elemental budgets, such as the carbon budget (Mendonça et al. 2017), and should be taken into account in large-scale climate change analysis (Downing et al. 2006). On the other hand, the impact of climate change on small and shallow water bodies has rarely been quantitatively assessed, even though they play an important role for biodiversity and are prone to harmful algal blooms (Biggs et al. 2016, Wilkinson et al. 2020).

To preserve the ecological value of small urban lakes, it is crucial to better understand how these ecosystems are evolving over time and how they react to changes in external forcing. The thermal regime and the hydrodynamics of lakes are among the main drivers of phytoplankton growth and harmful algal blooms (e.g. Wilkinson et al. 2016, Paerl & Huisman 2008) and are intimately related to eutrophication. It is therefore crucial to evaluate how climate change is impacting the thermal regime in urban lakes.

2.2 Modelling of aquatic ecosystems

Lakes and, more in general, all aquatic ecosystems are under great climatic and anthropogenic pressure. The demand of model simulations for current and future scenarios in environmental management is ever-increasing.

A model is any conceptual or mathematical representation of a system that serves to understand and quantify it (Torres & Santos 2015). Mathematical models describe a system through a set of equations. Broadly speaking, mathematical models can be divided into two main categories: empirical and mechanistic models (Biggs et al. 2016, Vinçon-Leite & Casenave 2019). Empirical models are data-driven and derive from observations and probability considerations. They mainly consist in statistical relationships (transfer functions) between the predictor and the response variables of interest. Mechanistic models are process-based and entail a physical interpretation and a mathematical description of the processes of interest (Vinçon-Leite & Casenave 2019). Both empirical and mechanistic approaches have been widely used for the modelling of aquatic ecosystems.

Aquatic ecological models are extremely valuable to test our knowledge and understanding of the relevant processes in the biogeochemical cycle. They are of particular interest for the scientific community and for policy makers as tools to simulate eutrophication, harmful algal blooms and their sensibility to environmental changes. However, models can never take into account every feature of the real system and, even though accurate, they always give an approximation of reality that introduces a certain rate of uncertainty (Uusitalo et al. 2015).

The construction of an aquatic ecological model requires a strongly multidisciplinary approach. State variables describing the motion of fluids, such as water temperature, density, and the velocity field are usually obtained through process-based hydrodynamic models. On the other hand, the processes relevant to the biogeochemical cycle and biomass production can either be modelled through data-driven empirical approaches, or through process-based kinetics and mass balances that entail a wide biological and chemical knowledge. Finally, programming skills are needed for the actual implementation of the model and the coupling between hydrodynamic and ecological models.

However, the study of hydrodynamics and biogeochemistry proceeded separately and on different timelines. As a result, hydrodynamic and ecological models were developed independently from one another. However, to acurately represent the ecosystem dynamic, these two models need to be coupled, with the ecological model representing the biochemical reaction processes and the hydrodynamic driver simulating currents, water temperature, and managing the transport and diffusion of the biochemical variables. The hydrodynamic driver is usually process-based, and its spatial dimension can vary from zero (box models, representing water temperature) to three (3D hydrodynamic models). The coupled ecological model can be either a statistical/empirical model or a mechanistic one (Vinçon-Leite & Casenave 2019).

The following sections describe the main characteristics of hydrodynamic and ecological models, focusing on their approaches, interactions, and their current limits and range of applicability. Concerning hydrodynamic models, only the mechanistic approach will be discussed, while the state of the art of a wider range of ecological models will be reviewed.

2.2.1 Hydrodynamic modelling

2.2.1.1 Governing equations

Hydrodynamics is the branch of physics that deals with the motion of fluids. The mathematical description of fluids motion was consolidated throughout centuries of studies and research. The next paragraphs give a brief insight on the theoretical and mathematical background that constitutes the basis of most hydrodynamic models, while pointing out its strengths and weaknesses.

Raynolds-Averaged Navier-Stokes equations
In physics, the motion of a viscous fluid is described by a system of partial differential equations known as the Navier-Stokes equations, that express the mass (equation (2.1)) and momentum (equation (2.2)) conservation for a control volume in a Newtonian fluid. Following the Einstein summation convention (with i=1,2,3 and j=1,2,3), their differential form can be written as:

$$\frac{\partial \rho}{\partial t} + \frac{\partial (\rho u_i)}{\partial x_i} = 0 \tag{2.1}$$

$$\underbrace{\rho \left(\frac{\partial u_i}{\partial t} + \frac{\partial}{\partial x_j}(u_i u_j) \right)}_{\text{Inertial forces}} = - \underbrace{\frac{\partial p}{\partial x_i}}_{\text{Pressure forces}} + \underbrace{(\lambda + \mu) \frac{\partial}{\partial x_i} \left(\frac{\partial u_j}{\partial x_j} \right) + \mu \frac{\partial^2 u_i}{\partial x_j^2}}_{\text{Viscous forces}} + \underbrace{\rho f_i}_{\text{External forces}} \tag{2.2}$$

where u is the fluid velocity, ρ is its density, μ and λ are respectively the dynamic and bulk viscosity, p is pressure, g is the acceleration of gravity and f generally represents external forces.

The left-hand side of equation (2.2) represents inertial forces, the first addendum on the right-hand side represents the pressure forces, the final one represents the external forces applied to the fluid (e.g. the gravitational forces), and the remaining terms represent the viscous forces.

However, the mathematical understanding of this complex system of equations is still incomplete, as it has not yet been proven whether smooth solutions of the Navier-Stockes equations always exist in three dimensions. Hydrodynamic model therefore rely on approximated solution (or, in strongly simplified cases, even analytical solutions) obtained through numerical methods under certain assumptions and hypothesis.

Namely, in the applications of interest for this work, it is reasonable to assume (i) an incompressible Newtonian fluid with (ii) a vertical hydrostatic pressure distribution and to (iii) neglect density variations in the gravity term of the momentum equation (i.e. Boussinesq approximation, Deltares (2014)). Under the first hypothesis, equations (2.1) and (2.2) become respectively:

$$\frac{\partial u_i}{\partial x_i} = 0 \tag{2.3}$$

$$\frac{\partial u_i}{\partial t} + \frac{\partial}{\partial x_j}(u_i u_j) = -\frac{1}{\rho}\frac{\partial p}{\partial x_i} + \nu \frac{\partial^2 u_i}{\partial x_j^2} + f_i \tag{2.4}$$

and the hydrostatic pressure (p) can be expressed as a function of atmospheric pressure (p_{atm}), water density (ρ) and free surface elevation (Z_s):

$$p = p_{atm} + \rho g(Z_s - z) \tag{2.5}$$

Furthermore, in order to overcome the difficulty of describing the chaotic turbulent nature of fluid motion, Osborne Reynolds suggested to decompose turbulent instantaneous quantities into the sum of their time-averages and fluctuating turbulent quantities. Following the Reynolds decomposition the dependent variables in (2.4) can be decomposed into a mean ($\overline{x}$) and a fluctuating (x') part:

$$u_i = \overline{u}_i + u', \qquad p = \overline{p} + p' \tag{2.6}$$

and the momentum conservation equation becomes:

$$\frac{\partial \overline{u}_i}{\partial t} + \frac{\partial}{\partial x_j}(\overline{u_i u_j}) = -\frac{1}{\rho}\frac{\partial \overline{p}}{\partial x_i} + \nu \frac{\partial^2 \overline{u}_i}{\partial x_j^2} + f_i \tag{2.7}$$

From the properties of a Reynolds operator (see, for instance, Alfonsi 2009), we moreover have:

$$\overline{u_i u_j} = \overline{(\overline{u}_i + u_i')(\overline{u}_j + u_j')} = \overline{\overline{u}_i \overline{u}_j} + \overline{\overline{u}_i u_j'} + \overline{u_i' \overline{u}_j} + \overline{u_i' u_j'} = \overline{u}_i \overline{u}_j + \overline{u_i' u_j'} \tag{2.8}$$

After substituting in (2.7), we obtain:

$$\frac{\partial \overline{u}_i}{\partial t} + \frac{\partial}{\partial x_j}(\overline{u}_i \overline{u}_j) = -\frac{1}{\rho}\frac{\partial \overline{p}}{\partial x_i} + \nu \frac{\partial^2 \overline{u}_i}{\partial x_j^2} - \frac{\partial}{\partial x_j}(\overline{u_i' u_j'}) + f_i \tag{2.9}$$

The Reynolds-stress term ($\tau_{ij} = \overline{u_i' u_j'}$) raises from the Reynolds averaging process and incorporates the effects of turbulent stresses.

Finally, we obtain the Reynolds-Averaged Navier-Stokes equations (RANS) for mass (equation (2.10)) and momentum (equation (2.11)) conservation, that are at the core of most 3D hydrodynamic models (de Goede 2020):

$$\frac{\partial \overline{u}_i}{\partial x_i} = 0 \tag{2.10}$$

$$\frac{\partial \overline{u}_i}{\partial t} + \frac{\partial}{\partial x_j}(\overline{u}_i \overline{u}_j) = -\frac{1}{\rho}\frac{\partial \overline{p}}{\partial x_i} + \nu \frac{\partial^2 \overline{u}_i}{\partial x_j^2} - \frac{\partial}{\partial x_j}\tau_{ij} + f_i \tag{2.11}$$

However, system (2.10) and (2.11) is not a closed system in terms of the four dependent variables $\overline{u}_i$ and $\overline{p}$, as the Reynolds-stress term introduces six additional independent unknowns. The closure of the system is obtained through the implementation of specific turbulence models that express the Reynold-stress term as a function of the mean field and/or other variables (Alfonsi 2009).

Heat exchange at the water surface

Lakes are integral features of the hydrological system (Fig. 2.1). They interact directly with atmospheric water, surface water and groundwater (Reynolds 2006). The areal extent of lakes bottom is considerably larger than their surface area. Their

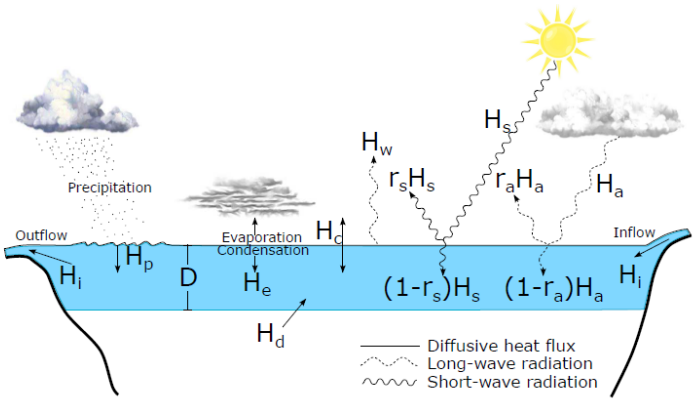


Figure 2.7 – Heat fluxes that affect water surface temperature (source: Piccolroaz et al. 2013).

interface with external systems is much larger through the bottom than through the surface. Groundwater infiltration, runoff discharge and evaporation are major components in lakes heat and mass balances. However, their contribution is highly uncertain (Winter 1981, Trask et al. 2017). In particular, due to their uncertainty, exchanges of mass and energy through the bottom are often neglected in linmological studies (Krabbenhoft et al. 1990, Shaw et al. 1990).

For this reason, lakes bottom is often assumed to be impermeable and adiabatic. In this case, surface water temperature is the result of the incoming and outgoing heat fluxes at the air-water interface, that include direct exchanges with atmosphere and deeper waters, latent and sensible energy transfers via inflows and outflows and evaporation (Fig. 2.7). The complete heat budget at the water-air interface for the calculation of the net heat flux (H_{net}) is given by (Piccolroaz et al. 2013):

$$H_{net} = (1 - r_s)H_s + (1 - r_a)H_a + H_W + H_e + H_c + H_p + H_i + H_d \tag{2.12}$$

where H_s is the net shortwave solar radiation, H_a is the net longwave radiation emitted from the atmosphere toward the lake and H_W is that emitted from the water itself, with r_s and r_a appropriate reflection coefficients; H_e represents the latent heat fluxes due to evaporation and condensation, H_c is the convection heat transfer, H_p and H_i take into account precipitation and inlet/outlet effects respectively and H_d represents the heat exchanges with deeper waters. Fluxes are considered positive when entering the surface layer and negative otherwise.

Some of the terms included in the complete heat budget (2.12) are difficult to measure or retrieve. In order to meet the needs of the user, most hydrodynamic models include various simplified formulations of the heat budget at the water surface.

Advection and diffusion

The dispersion of tracers occurs through advection and diffusion processes. Water temperature (T) is often treated as a tracer in hydrodynamic models. The heat entering the water body through the heat-budget (2.12) is transferred within the water body through an advection-diffusion equation:

$$\frac{\partial T}{\partial t} + u_i \frac{\partial T}{\partial x_i} = \frac{\partial}{\partial x_i}\left((\nu_M + \nu_T)\frac{\partial T}{\partial x_i}\right) + S \tag{2.13}$$

where S represents the source/sink term, and ν_M and ν_T are respectively the molecular and turbulent diffusion coefficients. Equation (2.13), here written for water temperature, applies to any tracer in a hydrodynamic model.

2.2.1.2 Model dimension

Hydrodynamic models are used in extremely diverse contexts and with very different objectives. Depending on the context, the spatial dimension of the model might vary. Zero-dimensional (0D) models can be used to focus on the heat budget at the air-water interface and simulate the average surface temperature evolution of a water body (Piccolroaz et al. 2013). 1D models simulate the evolution of currents and the distribution of water temperature over the water column, but do not give any information on heterogeneities in the horizontal plane. Bi-dimensional (2D) models reproduce spatial heterogeneities but do not resolve the differences on the vertical axis. Three-dimensional (3D) models take into account both the horizontal and vertical variabilities, and their mathematical approach entails the least grade of approximation.

The computational cost of hydrodynamic models increases with their spatial dimension. However, the computational cost of 3D models has been strongly reduced in recent years. Despite their computational cost being higher, 3D model have the considerable advantage to be as accurate as 1D models in the simulation of the vertical dynamics, while reproducing spatial variability as well (e.g. Polli et al. 2019).

2.2.1.3 Applications in climate change studies

Modelling approaches to climate change studies have several advantages. Provided that observations are available for their calibration and validation, models can be used to simulate the behaviour of a water body outside of the monitoring period as far as meteorological input data are available. This opens to an extremely wide range of applications. Our knowledge of water bodies with only short series of observations can be extended both in the past, when long-term meteorological data are available, and in the future, if models are forced with climate change scenarios. Depending on the input data, their response to changes in the climate can be reconstructed in the past, and simulated in the future.

Such modelling approaches have grown substantially in the last decade (e.g. Vinçon-Leite et al. 2014, Piccolroaz et al. 2018, Shatwell et al. 2019, Piccolroaz et al. 2020). A large portion of the dedicated scientific production focuses on large and deep lakes (Perroud & Goyette 2010, Vinçon-Leite et al. 2014, Piccolroaz et al. 2018). However, the interest in smaller and shallower lakes has risen in the recent years (Magee & Wu 2017, Moras et al. 2019, Zhang et al. 2020, Stetler et al. 2020). Most publications make use of 0D (Piccolroaz et al. 2018) or 1D models (e.g. Vinçon-Leite et al. 2014, Magee & Wu 2017, Moras et al. 2019, Soares et al. 2021) to simulate lakes dynamics under climate change, or of satellite measurements to reconstruct their long-term surface temperature evolution (e.g. Zhang et al. 2020, Stetler et al. 2020). 1D models have the great advantage of a low computational cost, with the advantage of being very accurate, especially in terms of water temperature vertical distribution (Piccolroaz et al. 2013, Polli et al. 2019). However, such models are

inherently incapable of directly resolving a number of physical processes crucial for lakes dynamics that take place on the horizontal plane. These processes can only be directly taken into account by 3D models, and are necessarily strongly approximated in 0D or 1D models (Romero et al. 2002, Piccolroaz et al. 2013, Xue et al. 2017).

2.2.2 Ecological modelling

Ecology, as the study of the interactions between organisms and their surrounding ecosystems, is a relatively new term when compared to other classical physical, biological or chemical disciplines. The idea of modelling ecological processes through a mathematical approach draws back to the 1920s, with the Lotka-Volterra equation for prey and predator interactions. However, ecological modelling finds a considerable acceleration only starting from the 1970s, due to: (i) the development of computer technology and computational capacity, (ii) the urgency linked to some environmental problems such as eutrophication and (iii) a deeper knowledge of the relationships between ecological properties and environmental factors.

Triggered by compelling issues such as eutrophication and harmful algal blooms, the request for simulating ecological and biological processes in water environments is ever increasing. To meet the need for simulations of biological processes, a vast range of different biogeochemical models has been developed in the last two to three decades, often with very similar formulations (Mooij et al. 2007, Trolle et al. 2012). Mooij et al. (2007) suggested that, this way, a high effort is misused in the process of 'reinventing the wheel'. The Ecobas Register of Ecological Models (`http://ecobas.org/www-server/index.html`) indicated, in 2012, the existence of more than 100 aquatic models, often with similar levels of ecological complexity and modelling objectives (Trolle et al. 2012). This trend was accompanied by a strong increase in the annual number of scientific publications dealing with the modelling of aquatic ecosystems (Trolle et al. 2012, Vinçon-Leite & Casenave 2019), that confirms the strong interest for ecological modelling found in the scientific community, and the need of ecological simulations to validate management policies in water resources.

The following sections contain a brief review of the main traits and approaches in aquatic ecological modelling.

2.2.2.1 Empirical models

Aquatic ecological modelling developed mainly in relation to eutrophication and eutrophication management policies. The main objective of an aquatic ecological model is therefore to reproduce the most relevant processes of the biogeochemical cycle, relative to the variables of interest.

Two main approaches can be distinguished in aquatic ecological modelling: the empirical one and the mechanistic one. Empirical models are data-driven and derive from observations through probability considerations. They mainly consist in statistical relationships (transfer functions) between the predictor and the response variables of interest. These models require very little understanding of the actual physical processes happening in the modelled system (Vinçon-Leite & Casenave 2019). Their formulation is avoided and the descriptive power of the model is transferred to statistical considerations. These models are quite generic and are often applied to groups of lakes. Their great advantage is their simplicity of set up.

They can be built using few data of sporadic time series of sampling. However, their results bear a high rate of uncertainty (Vinçon-Leite & Casenave 2019).

During the last couple of years, empirical models based on machine learning techniques (such as neural network or random forests) also found a considerable development, due to their capability of predicting highly non-linear and complex responses (Fornarelli et al. 2013). Contrary to the more traditional statistical approaches, models based on machine learning need a very high amount of data to reduce their uncertainty, and benefit greatly from the availability of high-frequency *in situ* data. These approaches showed some promising results especially in short term forecasting of harmful algal blooms, and could be of particular value for the water resources management in recreational areas (Vinçon-Leite & Casenave 2019).

2.2.2.2 Mechanistic models

Mechanistic models are process-based and entail a physical interpretation and a mathematical description of the driving processes of the ecosystem. They usually consist of a set of differential equations and mass balances, solved through numerical methods, that describe in detail the biogeochemical cycle (Vinçon-Leite & Casenave 2019).

One of the first mechanistic ecological models was proposed by Chen (1970), with the objective of addressing eutrophication through computer simulations. In this conceptualization, the aquatic system is divided into four major interrelated components:

1. Abiotic substances (H_2O, CO_2, NO_3, PO_4, Fe and so on)
2. Producers
3. Consumers
4. Decomposers

In the model, the variation of the phytoplankton concentration P_i (where i stands for a specific algal group) over time is described by:

$$\frac{d(\overline{V}P_i)}{dt} = T + AE\frac{dP_i}{dx} + (\mu_i - r - s)\overline{V}P_i - \frac{1}{Y_z}gp_{fi}z\overline{V} \tag{2.14}$$

where V is the volume of the considered element, T is the total advective mass transfer, E is the effective diffusion coefficient, μ_i is the specific growth rate of the algal group, A is the area of the element, x is a distance (this model was developed for rivers, in the case of lakes it would become the vertical length of the element), r is the percentage of mass lost by respiration per unit of time, s is the percentage of mass settling per unit time, Y_z the yield coefficient of zooplankton, g the specific growth rate of zooplankton, p_{fi} the preference factor for the algal group, and z is the zooplankton biomass concentration.

The model also includes limiting factors for growth, as the growth rate of phytoplankton depends on environmental conditions and may vary in time. The limiting-factor are here modelled through specific coefficients $k_{1,2,3,4}$:

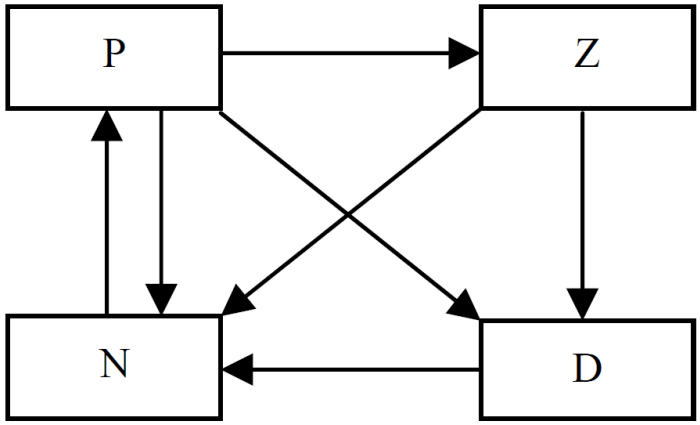


Figure 2.8 – Schematic representation of NPZD models (source: Li & Hipsey 2013).

$$\mu_i = \hat{\mu}\left(\frac{L}{k_1 + L}\right)\left(\frac{C}{k_2 + C}\right)\left(\frac{N}{k_3 + N}\right)\left(\frac{P}{k_4 + P}\right) \tag{2.15}$$

where L, C, N and P, are, respectively, light intensity, carbon dioxide, nitrate and phosphate concentrations. Similarly, a series of differential equations was formulated to describe the changing rates of zooplankton, fish, detritus, BOD, nutrients and DO. This model achieves a high level of complexity in its description of the interactions in an ecosystem and still constitutes the core of the current generation of mechanistic aquatic biogeochemical models (Shimoda & Arhonditsis 2016).

Throughout the decades, the structure of ecological models has become more complex. However, most of the current dynamic ecological models are still rooted in the early ideas of modified prey-predator models and 'Nutrient-Phytoplankton-Zooplankton-Detritus' (NPZD) models (Li & Hipsey 2013). In NPZD models, similarly to what proposed by Chen (1970), the biogeochemical cycle is modelled through four main interacting players (Fig. 2.8). Pure NPZD models can be considered of medium complexity (Heinle & Slawig 2013). They have already been widely applied for aquatic ecosystem research in both regional and global modelling studies (e.g. Anderson 2005, Hipsey et al. 2013), and they still represent a valid choice for specific modelling tasks.

Modern ecological models are dynamic. They are built with a modular structure that allows the user to achieve the desired level of complexity in its configuration. These complex dynamic models include a high number of variables and processes and are designed to be used as a virtual reality for the simulation of the whole aquatic ecosystem, from microbial decomposition to primary production, predation and vegetation (Mooij et al. 2010, Vinçon-Leite & Casenave 2019). Various complex dynamic models are currently available. They differ in terms of the hydrodynamic driver, but share a similar overall modular structure. The models CAEDYM, Aquatic EcoDynamics (AED), PCLake and DELWAQ are among the most used in scientific literature (Mooij et al. 2010, Vinçon-Leite & Casenave 2019).

Figure 2.9 shows the typical set of modules available in a current complex

dynamic ecological model (namely, AED) and their interactions. These models look at the ecosystem as a whole. Little remains of the simplicity of NPZD models, with their four compartments here divided into sub-modules. The grade of complexity can then be chosen by the user according to the data and knowledge available for a given application.

In each module, the variables are usually subject to a mass balance of the following type (for the generic variable C, of concentration $[C]$):

$$\frac{d[C]}{dt} = \sum_i f^{\,C}_{prod_i} - \sum_j f^{\,C}_{cons_j} \tag{2.16}$$

where $f^{\,C}_{prod}$ and $f^{\,C}_{cons}$ respectively represent all fluxes producing and consuming C. Each separate flux constituting the producing and consuming terms are modelled specifically according to the process they represent, often through modified Michaelis-Menten kinetics:

$$f^{\,C}_{prod_i} = R_{prod_i} \frac{[C]}{K_{prod_i} + [C]} \theta^{T-20}_{prod_i} [C] \tag{2.17}$$

where: R_{prod_i} is the maximum rate of production (consumption), K_{prod_i} is the half-saturation constant and θ_{prod_i} is a factor for temperature dependence. The set of differential equations such as (2.16) are solved through numerical methods. The number of fluxes taken into account in the mass balances depends on the activated modules in the configuration.

Some modules appear to be more commonly included in the applications of these models for eutrophication modelling.

Oxygen

In complex dynamic models, oxygen dynamics are generally modelled through a dedicated module. Oxygen concentration is usually modelled as a function of exchanges with atmosphere, consumption from respiration, production from photosynthetic organisms and exchanges with the sediment. When sediment interaction with the water column is not explicitly simulated, oxygen flux through the sediment can still be included, through the sediment oxygen demand (Mooij et al. 2010, Vinçon-Leite & Casenave 2019).

Inorganic nutrients

Inorganic nutrients such as nitrate (NO_3^-) and phosphate (PO_4^{3-}) are directly available for primary producers. They are therefore crucial for eutrophication modelling. In complex dynamic models, they are usually separated into different sub-modules for nitrogen (NH_4^+ and NO_3^-), phosphorus and silica.

Phosphorus is largely believed to be the main limiting nutrient for phytoplankton growth (Granéli et al. 2008, Vinçon-Leite & Casenave 2019). For this reason restoration policies of eutrophic environments focused for decades on phosphorus reduction. Phosphorus concentration was often used as the only indicator of the ecological state of rivers and water bodies (Poikane et al. 2019). However, nitrogen limitation occurs quite frequently (Abell et al. 2010, Poikane et al. 2019). In the nitrogen module, nitrate (NO_3^-) and ammonium (NH_4^+) are usually separated and

subject to the processes of nitrification and denitrification.

Organic matter

Organic matter constitutes the main pool for dead biomass. Most dynamic models distinguish between dissolved and particulate organic matter (Vinçon-Leite & Casenave 2019). Organic matter gets mineralized by bacteria and partially refilling the inorganic nutrient pools. When bacteria modules are not included, mineralization is usually modelled through dedicated Michaelis-Menten style equations. In shallow water bodies, wind-induced sediment resuspension is quite frequent and the replenishment of the water column with nutrients mineralized in the sediment can become an important internal source of nutrients (Song et al. 2013).

Phytoplankton

The phytoplankton module is at the core of many applications of ecological models (Vinçon-Leite & Casenave 2019). In most complex dynamic models the phytoplankton community is divided into different groups. The phytoplankton groups available are customizable through a series of parameters that define their characteristics in terms of growth, mortality, and temperature, light and nutrient requirements.

Traditionally, the phytoplankton community is divided into taxonomic groups that separate phytoplankton into filogenetic classes (e.g. diatoms, cyanobacteria or green algae). However, different species belonging to the same class might have characteristics that are radically different. The concept of functional groups was introduced by Reynolds et al. (2002), and aims at resolving the problems arising from filogenetic classification, by regrouping phytoplankton into new cross-species categories with similar adaptative responses to environmental characteristics (Reynolds et al. 2002, Shimoda & Arhonditsis 2016). Models purely based on functional grouping are promising but still under development (Anderson 2005, Shimoda & Arhonditsis 2016, Di Maggio et al. 2016). However, a mixed approach is sometimes used, where functional traits are introduced in the definition of phytoplankton taxonomic groups, in order to better model their responses to environmental changes.

The phytoplankton succession in a natural environment is the result of competition for resources. However, multispecies competition is often regarded as fundamentally unpredictable (Huisman & Weissing 2001). In most multispecies models, the competition for resources is mimicked through the definition for each group of different characteristics and requirements that lead, under the same environmental conditions, to different limiting factors. The model BLOOM (Deltares 2014) includes a competition principle and constitutes an exception. The competition principle is based on the ration between the growth rates and the resource requirements for each phytoplankton group, in relation with local environmental conditions (Los 2009). The competition principle works through an optimization algorithm, according to which at each time step the biomass is distributed among the phytoplankton groups and net primary production is maximized consistently with the environmental conditions and existing biomass levels.

Cyanobacteria

According to numerous studies climate change and eutrophication promote harmful algal blooms and cyanobacteria blooms in particular (Paerl & Huisman 2008, Wagner

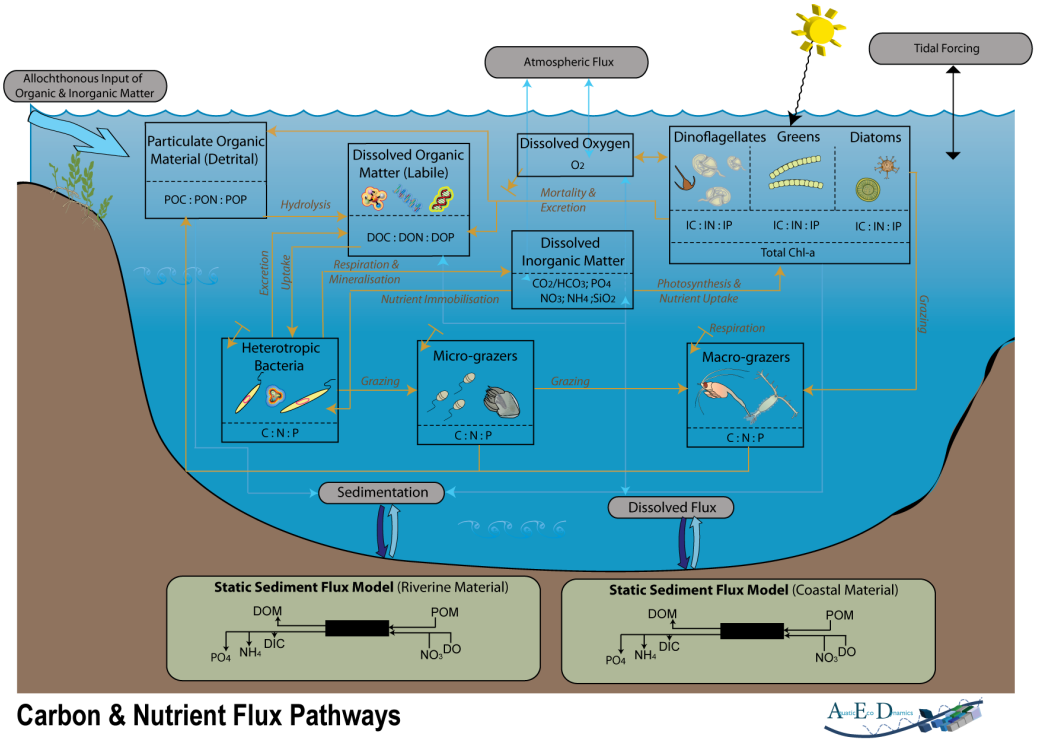


Figure 2.9 – Schematic representation of dynamic models, specifically for the model Aquatic EcoDynamics (source: Hipsey et al. 2013).

& Adrian 2009, O'Neil et al. 2012, Paerl 2014). Cyanobacteria are therefore an important feature of aquatic ecological models. They show some peculiar traits that distinguish them from other phytoplankton groups and that are accounted for in most dynamic ecological models. Various models (e.g. CAEDYM, BLOOM or AED) include options for nitrogen-fixing cyanobacteria. Their buoyancy capabilities as well as their toxin production are not explicitly modelled in most ecological models, due to the poor theoretical understanding of these process (Walsby 1994, Schuergers et al. 2017, Sivonen 1996, Catherine et al. 2013).

2.2.3 Applicability of ecological models: the need for data

A high number of aquatic ecological models are available, with a great variety of approaches and complexity. The choice of the model to use should depend on the focus of the study and on the available knowledge and data. With little knowledge about the ecosystem, empirical models should be preferred (Vinçon-Leite & Casenave 2019). Models based on machine learning need a high amount of data to be effective, but require little knowledge of the ecosystem. Finally, the choice should fall upon mechanistic models if sufficient knowledge about the key processes of the ecosystem is available (Vinçon-Leite & Casenave 2019). Once a model is chosen, data are required for the calibration of the model parameters, as well as for validation.

While empirical models require only a few parameters, mechanistic models are often over-parametrized (Reichert & Omlin 1997, Makler-Pick et al. 2011, Rigosi et al. 2011). Parameter values for mechanistic models are often deduced from

the literature (Romero et al. 2004, Gal et al. 2009, Lürling et al. 2013, Fenocchi et al. 2019). However, values found in literature might be derived from controlled environments (e.g. laboratory experiments or mesocosms) or from different study site, and might not be appropriate for the ecosystem under consideration. A great range of values might be found in literature for the same parameter (Fenocchi et al. 2019). Furthermore, in most studies the calibration of the parameters is done without any sensitivity analysis through a trial and error procedure. Automated calibration is only rarely applied and usually confined to 0D or 1D models (Vinçon-Leite & Casenave 2019).

Results from ecological models often bear a high rate of uncertainty (Arhonditsis et al. 2006, Vinçon-Leite & Casenave 2019). This is caused by three main reasons: (i) the complexity of the biogeochemical cycle and the still incomplete theoretical understanding of some of its key processes, (ii) the difficulties in parameters calibration, and (iii) the lack of data to perform calibration and validation.

Empirical models usually require only a few parameters to be calibrated, however uncertainty is inherent to their empirical, data-driven approach. On the contrary, in the case of process-based models, the general lack of calibration and validation strongly contributes to their uncertainty. Traditionally, data relevant for ecological models (e.g. nutrient and phytoplankton concentration) derive from field sampling and subsequent laboratory analysis. These analysis are accurate but require certain amount of time to be carried out. Field trips might also be problematic for remote study sites. Such data are therefore traditionally taken on a monthly or weekly basis. This low frequence does not facilitate the implementation of automated calibration strategies, and does not allow to validate the modelling of the processes occurring at a lower frequency than that of the sampling.

High-frequency *in situ* measuring systems can measure relevant physical and chemical variables such as water temperature, pH, oxygen and dissolved nutrient concentration. Furthermore, through optical sensors, they can give reliable estimates of total chlorophyll and phycocyanin concentration, and in some cases also of the phytoplankton assembly (Tran Khac et al. 2018). These instruments as well, however, need a correct calibration and validation that can be obtained through sampling and laboratory analysis. Accompanied by a traditional field monitoring, high-frequency *in situ* monitoring can give new and important insights on the ecology of aquatic ecosystems. Furthermore, consistent high-frequency data could help reducing models uncertainty by applying sensitivity analysis, automated calibration and more accurate validation.

2.2.4 Parameter optimization

The biogeochemical cycle is extremely complex. Its mathematical description often involves parameters that need to be empirically estimated. In the effort of including all of its processes and interactions, mechanistic ecological models are often highly complex and over-parameterized (Reichert & Omlin 1997, Makler-Pick et al. 2011, Rigosi et al. 2011, Vinçon-Leite & Casenave 2019, Luo et al. 2018). Most parameters are difficult to measure directly through observations or laboratory experiments. Reference values can be found in scientific literature but are uncertain and often have a wide range of variability (Makler-Pick et al. 2011, Fenocchi et al.

2019). The uncertainty on the values of these parameters affects models reliability.

For these reasons, calibrating and validating an ecological model is an important task. However, some studies showed that a proper calibration stage or a sensitivity analysis are not commonly performed (roughly in half of the publications found in scientific literature), and that automated calibration strategies are rarely employed, in favour of trial and error procedures (e.g. Arhonditsis & Brett 2004, Shimoda & Arhonditsis 2016).

In the case of trial and error calibrations, their outcome is highly dependent on the modeller's skills and knowledge (Luo et al. 2018). Automated calibration can reduce models uncertainty and simultaneously allow to carry out a sensitivity analysis of the model parameters. However, it is rarely applied for ecological models: very few papers make use of automated calibration for mechanistic ecological models. In literature, automated calibration is only applied on 0D or 1D models, mostly through optimization or Monte Carlo and Bayesian inference (Vinçon-Leite & Casenave 2019).

Automated calibration strategies are generally computationally costly. They often need a large number of model runs and their computational cost increases with the number of parameters to be estimated, hindering their application to complex ecological models. Furthermore, in limnological studies, data traditionally come from field campaigns that, even though regular, lead at best to sparse datasets that are not adapted for automated calibration strategies.

A vast range of approaches and techniques are available for automated parameter optimization, from Newton and quasi-Newton algorithms, to genetic algorithms (e.g. Particle Swarm Optimization) and Bayesian parameter inference (Mahevas et al. 2019).

Classic Bayesian parameter inference is often problematic for complex mechanistic models. For such models, the likelihood function is analytically intractable and its evaluation through computational methods is extremely computationally demanding. Approximate Bayesian Computation (ABC) is an innovative and promising technique for parameter inference, rooted in Bayesian statistic, that has the great advantage of bypassing the computation of the likelihood function. ABC exploits a large number of model runs obtained through user-defined prior distributions of the model parameters, as a training data set to approximate of the parameters posterior probability distribution function. The approximation of the posterior distribution can be obtained through various methodologies, among which the use of machine learning techniques was recently proposed and appears to be particularly advantageous (Raynal et al. 2019).

Chapter 3

Materials and methods

This chapter describes the materials and methods used in this PhD thesis. First, the study site, the measuring system, the field campaigns and the data collected and used in this work are presented in section 3.1. Following, in section 3.2, the four models deployed in this work (namely, Delft3D-FOW, Delft3D-BLOOM, TELEMAC3D and Aqutic Ecodynamics) are described. The description of the models is limited to their most relevant characterisitcs in relation to their applications in this thesis, and covers: the computational domain built for the study site, the structure implemented for the models, and a mathematical description of the processes of relevance for this work.

3.1 Study site and data set

In the following sections, I will present: (i) the study site Lake Champs-sur-Marne, (ii) the measuring instrumentation deployed on the lake, as well as (iii) the meteorological data used in the modelling of the study site. The data set collected on the study site during my PhD is presented in chapter 4.

3.1.1 Study site

The study site is Lake Champs-sur-Marne. It is a sand-pit lake located in the eastern suburbs of the Great Paris region (latitude: 48°51'50" N, longitude: 2°35'52" E), next to the Marne River (see Figure 3.1). It is a small and shallow water body with a surface of 0.12 km^2, mean depth of 2.5 m and maximum depth of around 3.5 m. The bathymetry of the lake was measured in 2016 by the research team at LEESU through an echo-sounder. The southern part of the lake is the deepest one, while depth decreases below two meters around the small island and in the northern part of the lake. Lake Champs-sur-Marne has no inflow or outflow and is fed primarily by groundwater and occasionally by rainfall runoff. Its water level is therefore mainly influenced by the presence of the Marne River, that flows from East to West right north of the pond, and varies weakly during the year, with monthly oscillations lower than 0.2 m on average.

The lake was originated in the 1940s by excavation and represents now a valuable and demanded recreational area for the population of the surrounding departments. However, it suffers from strong eutrophic conditions and experiences severe harmful

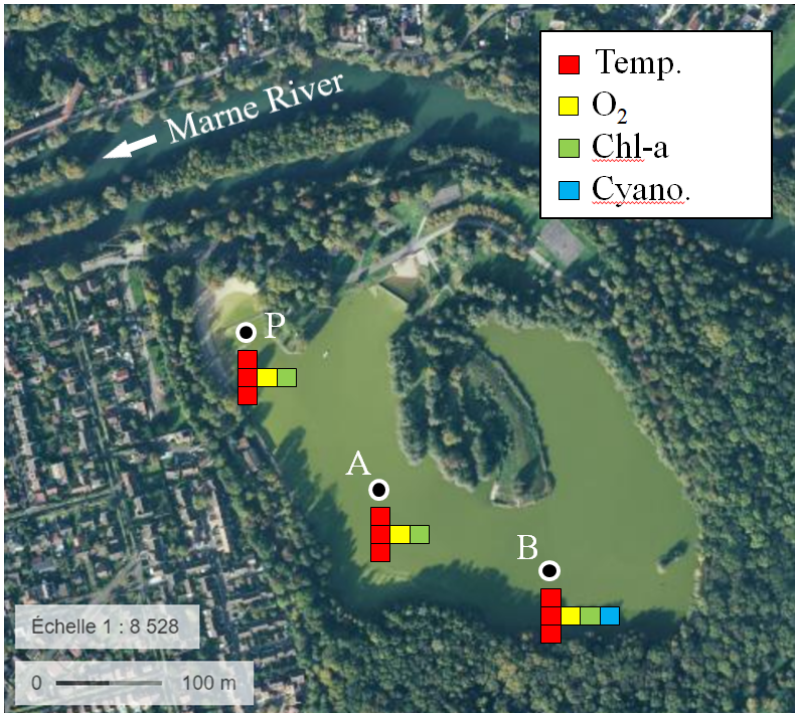


Figure 3.1 – Satellite picture of lake Champs-sur-Marne (source: *geoportail.fr*) and sketch of the measuring system at the three locations (A, B, P). Water temperature (red squares) is measured at the surface, bottom and middle layers, while the remaining variables are only measured at the middle layer.

algal blooms, especially between early spring and autumn. These blooms are often dominated by potentially toxic species of cyanobacteria. This regularly leads to a bathing ban and to restrictions in the access to the lake.

Given its shallowness, lake Champs-sur-Marne is polymyctic and its thermal behavior is strongly influenced by meteorological conditions. Between Spring and Autumn, periods of stable thermal stratification that can last up to three or four continuous weeks alternate with complete mixing and overturn of the water column.

3.1.2 Measuring instrumentation

Lake Champs-sur-Marne is monitored by the Water, Environment and Urban Systems Laboratory (Laboratoire Eau Environnement et Systèmes Urbains: LEESU) through sampling and laboratory analysis, profiling, and the deployment of automated high-frequency *in situ* measurements. The survey carried out during this PhD was funded by the ANR projects OSS-Cyano and ANSWER. The field campaigns are organised with a different frequency depending on the season: the survey is monthly from the months of October to March, and biweekly from April to September.

The lake is monitored at three different locations (Fig. 3.1). Sites A and B are representative of the deepest portion of the study site, while site P is located close to the bathing area. The dataset collected during each field campaign at the three locations includes: water depth and Secchi depth measurements, profiles of relevant variables and water samples.

The profiles are taken via two separate instruments. The multi-parametric probe Seabird (SBE 19™) gathers profiles of water temperature, conductivity, dissolved oxygen, pH and conductivity; the spectrofluorometer Fluoroprobe (BBE Moldaenke™) uses optical sensors to record profiles of the abundance of four phytoplankton groups (green algae, diatoms, cryptophyceae and cyanobacteria). Eventually, the AlgaeTorch profiler (BBE Moldaenke™) was also occasionally employed on the study site. The AlgaeTorch gives measures of total chlorophyll and cyanobacteria chlorophyll concentration.

Water samples are taken at the three sites at a 1.5 m depth via a Niskin bottle, to allow a direct comparison with the high-frequency data recorded at the same depth. The samples are subsequently analysed in laboratory, to obtain the concentration values for: the main inorganic nutrients (NO_3, PO_4, N_{tot}, P_{tot}), total chlorophyll and, starting from 2019, dissolved organic carbon (DOC). Finally, starting from 2019, the samples are also analysed for phytoplankton species identification and abundance.

The automated high-frequency instruments are installed on the study site at the same three locations shown in Figure 3.1. Each measuring site is equipped with sensors at three depths: below the surface at 0.5 m depth, at the middle of the water column at 1.5 m and close to the bottom at 2.5 m. The structure of the measuring system is shown in Figure 3.2 and is repeated at all locations: water temperature is measured at all three depths, while at the middle of the water column a multi-parameters sensor is able to capture also electrical conductivity, dissolved oxygen, Chlorophyll-a and (only at site B) phycocyanin concentration (see Figure 3.1). Measurements started in spring 2015 at site A, followed closely by those at site B (December 2015), while the third measuring site (site P) was set up in March 2017. A detailed description of the automated measuring system can be found in Tran Khac et al. (2018). In addition, a sensor measuring nitrate at high-frequency (i.e. ten minutes) was installed at site B in winter 2019.

3.1.3 Organisation of the field campaigns and data collection

The monitoring of Lake Champs-sur-Marne by LEESU started in May 2015 with the installation at measuring site A of *in situ* sensors at the surface, middle and bottom layers, recording data at high-frequency. The sensors at sites B and P were installed respectively in December 2015 and in March 2017, and a near-real time automated transmission system was activated in spring of 2016 for all the data recorded at the middle layer at site B. High-frequency *in situ* measurements are accompanied by periodic field campaigns, that started for sites A and B in September 2015, and in April 2017 for site P.

During the field campaigns, for the three measuring sites shown in Figure 3.1, the following tasks are performed: (i) maintenance, cleaning and data retrieval of the high-frequency sensors, (ii) deployment of the profiling instruments (Seabird and BBE Fluoroprobe, described in section 3.1.2), (iii) measurement of the water column and Secchi depths, and (iv) water sampling through a Niskin bottle. Namely, water samples were collected starting from June 2017. Further details on the data set and its collection are given in chapter 4.

A few photos taken on Lake Champs-sur-Marne during my PhD are presented in

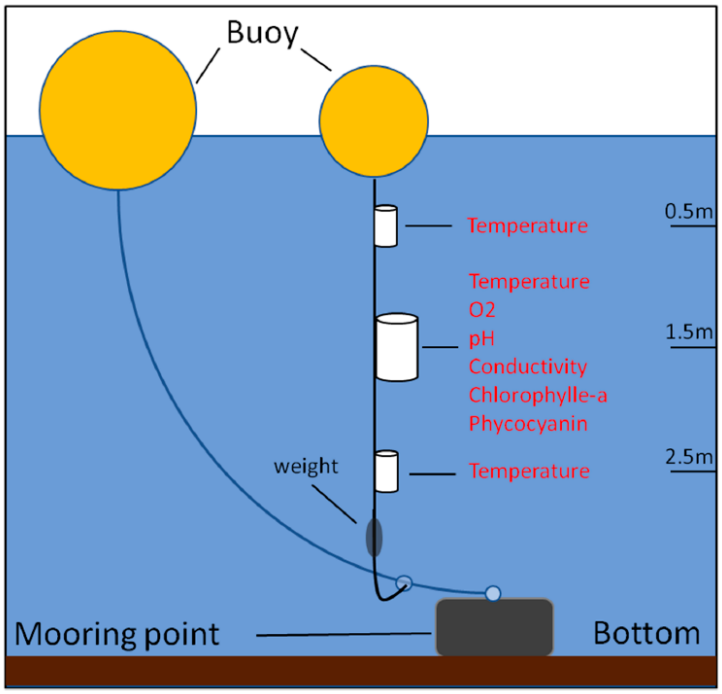


Figure 3.2 – Sketch of the automated *in situ* measuring system at site B (from: Tran Khac et al. 2018).

Figure 3.3 to better describe the study site itself and the measuring instrumentation used during the field campaigns.

3.1.4 Meteorological data

The thermal and ecological dynamics of lakes depend greatly on the meteorology, in particular for a small and shallow water body such as Lake Champs-sur-Marne. In this work, meteorological data are retrieved here from two different sources: (i) the Meteofrance meteorological station at the Orly airport, which is the closest meteorological station to the study site and is located roughly 20 km far South-West of Lake Champs-sur-Marne, and (ii) the spatialized SAFRAN reanalysis, produced by Meteofrance.

Meteorological data from the Orly Meteofrance station were downloaded in terms of relative humidity [-], air temperature [°C], net solar radiation [J s^{-1} m^{-2}], sky cloudiness [-], as well as wind speed [m s^{-1}] and direction [°N] between the years 2015 and 2019. All data were downloaded with a hourly time step.

The second source of meteorological data is the SAFRAN meteorological reanalysis, produced by Meteofrance. The Système d'Analyse Fournissant des Renseignements Atmosphériques à la Neige (SAFRAN) is a spatialized meteorological analysis system that offers, at present, validated meteorological data between 1959 and 2017 over the entire French territory (Durand et al. 1993).

SAFRAN is part of the SAFRAN-ISBA-MODCOU chain of reanalysis that covers the hydrological cycle over France, from meteorology to snow and ice formation to hydrology, respectively (Habets et al. 2008). SAFRAN integrates spatialized

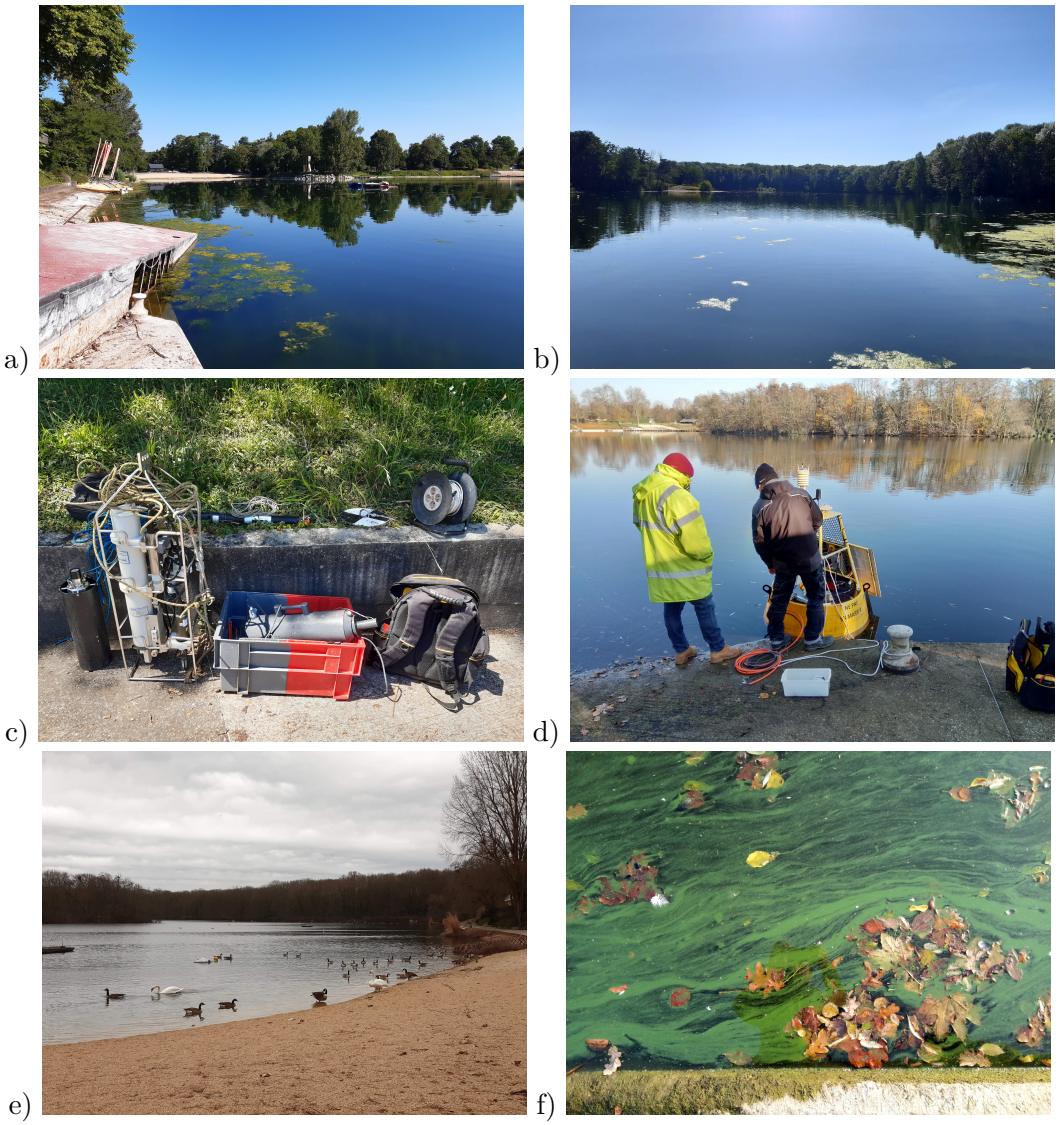


Figure 3.3 – Photos of Lake Champs. Panel a: view from the western shore of the lake towards site P; panel b: view from the western shore of the lake towards site B; panel c: field campaign instrumentation; panel d: a maintenance campaign in December 2019; panel e: an example of the bird population in Lake Champs-sur-Marne; panel f: a thick phytoplankton bloom in October 2018.

data from meteorological models with various sources of observations through data assimilation techniques, in order to create a consistent and spatially detailed record of meteorological data over the French territory. Its outcomes have been thoroughly validated against observed series (Quintana-Seguí et al. 2008), and tested as inputs to hydrological models (Raimonet et al. 2017). The data are spatialized on a regular square grid (8 km between each cell centre) that covers the entire French Territory. The location of Lake Champs-sur-Marne falls midway on the axis connecting the centres of SAFRAN cells number 1457 (North of the lake) and 1566 (South of the lake). The average of these two cells was therefore considered representative of the conditions over the study site.

Data were downloaded from the SAFRAN suite for the complete available period, that is between 1959 and 2017, in terms of air temperature [°C], specific humidity [-], solar radiation (direct and diffused) [W m^{-1}] and wind speed [m s^{-1}]. All these variables are well reproduced by SAFRAN (Quintana-Seguí et al. 2008). Data were downloaded at a hourly time step.

SAFRAN Specific humidity (SH) was converted into relative humidity (RH) through the following formula:

$$RH = 100 \cdot \frac{w}{w_s} \approx 100 \cdot \frac{SH}{w_s} \tag{3.1}$$

where w is the mixing ratio of water with dry air [kg kg^{-1}], the subscript s stands for saturation conditions and SH is the specific humidity, numerically very close to the mixing ratio value. The saturation mixing ratio can be calculated as follows:

$$w_s = \frac{R_a}{R_v} \cdot \frac{e_s}{p_{atm} - e_s} \tag{3.2}$$

where the atmospheric pressure (p_{atm}) was considered to be constant and equal to the global average: p_{atm} =1013 hPa. The ratio between the air and vapor ideal gas constants (R_a and R_v, respectively) is equal to 0.622. The partial vapour pressure at saturation (e_s) is temperature dependent and can be estimated (in hPa) through the Magnus equation:

$$e_s = 6.1094 \cdot \exp\left(\frac{17.625 \cdot T}{T + 243.04}\right) \tag{3.3}$$

where T is air temperature, measured in °C. The numerical coefficients in Eq. (3.3) are issued from Alduchov & Eskridge (1997).

Wind direction data were not available from the SAFRAN reanalysis. They were therefore downloaded over the same period of time from the MétéoFrance station located in Trappes (ID: 78621001, roughly 60 km West of the study site) at a daily time step through the INRAE CLIMATIK platform (`https://intranet.inrae.fr/climatik/`, in French) managed by the AgroClim laboratory of Avignon, France.

3.2 The models

Two coupled hydrodynamic - biogeochemical 3D models were applied to the study site, with the objective to simulate its thermal and planktonic dynamics. Namely, these models are: (i) the coupled hydrodynamic and biogeochemical models FLOW and BLOOM of the Delft3D modelling suite, and (ii) the TELEMAC3D hydrodynamic model coupled with the biogeochemical model Aquatic EcoDynamics 2 (AED2).

This section gathers a general description of the models, in relation to their implementation for the study site. However, the models have been applied in several studies that differ in terms of objectives, computational effort and time-horizon. For this reason, their configuration is not unique but might vary slightly between different applications.

Some of the specificities related to each single application are already highlighted in the following sections, and are further detailed in the chapters 5, 6 and 7, which are dedicated to each specific application.

3.2.1 Delft3D-FLOW

Delft3D-FLOW is a well known and robust hydrodynamic model that has been applied in various contexts, from estuaries to rivers, lakes and reservoirs (Chanudet et al. 2012, McCombs et al. 2014, Wahl & Peeters 2014, Piccolroaz et al. 2019). In particular, it has proved to correctly reproduce the stratification dynamics in shallow water bodies (Soulignac et al. 2017). FLOW is the hydrodynamic module of the broader Delft3D modelling suite, an open source software developed by Deltares (`https://oss.deltares.nl/web/delft3d`) that allows to investigate hydrodynamics, sediment transport and morphology and water quality for fluvial, estuarine and coastal environments. Version 4.01.01.rc.03 (2014) was used in this work.

Delft3D-FLOW was used in two main applications in this work: for the long-term reconstruction of the thermal dynamics of the study site (described in chapter 5), and for the automated calibration of a complex biogeochemical model (see chapter 6). The model set up differs slightly between the two applications, in particular in terms of the computational domain and of the heat-exchange module.

The specificities of the model configurations used for each of the above-mentioned studies are detailed in chapters 5 and 6. The following paragraphs describe the main characteristic of the hydrodynamic model itself, in relation to its set up for the study site.

3.2.1.1 Numerical aspects and computational domain

Delft3D-FLOW solves the Reynolds averaged Navier-Stokes equations for an incompressible fluid under the shallow water and the Boussinesq assumptions. The time integration of the partial differential equations is done through an Alternate Direction Implicit method (Deltares 2014, Leendertse 1967). For the spatial discretization of the horizontal advection terms the Cyclic scheme was selected (Stelling & Leendertse 1992). The computation is based on finite volumes. Detailed information on the numerical methods implemented in FLOW for the integration of

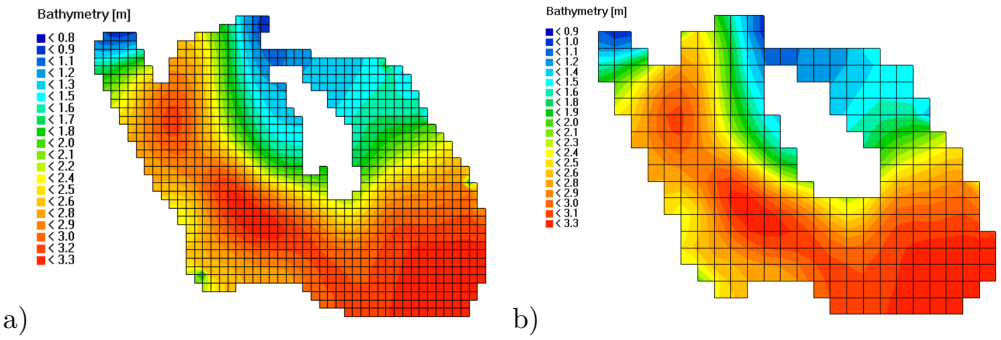


Figure 3.4 – Comparison of the two computational grids implemented in Delft3D-FLOW for Lake Champs-sur-Marne. Panel a: finer 10 m × 10 m grid; panel b: coarser 20 m × 20 m grid. Bathymetric data are shown through a colour chart.

the momentum and mass conservation equations can be found in the user manual (Deltares 2014).

The three dimensional computational domain for the study site is based on a Cartesian grid on the x-y plan and on a regular discretization of the vertical axis.

The horizontal mesh is built through the dedicated tools implemented in the Delft3D modelling suit. Namely, two computational grids with different definitions were built and used in different applications of the model. Namely, the less refined mesh is constituted of 255 20 m × 20 m cells, and was used for the long-term climate change study, in order to save computational time; this application is described in chapter 5. The other mesh is constituted of 813 10 m × 10 m square cells and is used in all other applications of the Delft3D suite. The bathymetry of the study site was interpolated from *in situ* measurements taken with echo-sounder by the research team at LEESU in 2016. The two computational grids are shown in Figure 3.4.

The discretization of the vertical axis was done through a series of 12 fixed parallel horizontal layers of 30 cm thickness (the so-called Z-model). When compared to the σ-model (i.e. layers with thickness varying according to the bathymetry), it is generally accepted that horizontal layers help avoiding artificial mixing, improving model results in terms of thermal stratification (Hodges 2014).

3.2.1.2 Turbulence closure

In most applications of 3D hydrodynamic models, the computational grid is too coarse and the time step too large to resolve the turbulent scales of motion: the turbulent processes are "sub-grid". In FLOW, the vertical turbulent eddy viscosity and diffusivity are directly computed through a turbulence closure model:

$$\nu_V = \nu_{mol} + max\left(\nu_{3D}, \nu_{3D}^{back}\right) \tag{3.4}$$

$$D_V = \nu_{mol}/\sigma_{mol} + max\left(D_{3D}, D_{3D}^{back}\right) \tag{3.5}$$

where ν_V and D_V are respectively the vertical eddy viscosity and diffusivity (m^2 s^{-1}), ν_{mol} is the molecular viscosity, σ_{mol} is the Prandtl-Schmidt number (0.7), ν_{3D}

is the eddy viscosity computed by the turbulence model, and ν_V^{back} and D_V^{back} are the background values for vertical eddy viscosity and diffusivity. In this work, the k-ε turbulence closure model was activated. The background values were set to zero [m^2 s^{-1}] for vertical viscosity and diffusivity.

The horizontal turbulent eddy viscosity and diffusivity are modelled as:

$$\nu_H = \nu_V + \nu_H^{back} \tag{3.6}$$

$$D_H = D_V + D_H^{back} \tag{3.7}$$

where ν_H and D_H are respectively the horizontal eddy viscosity and diffusivity, and ν_H^{back} and D_H^{back} are the background values for horizontal eddy viscosity and diffusivity. Eddy viscosity and diffusivity are usually much larger on the horizontal than on the vertical axis ($D_H \gg D_V$ and $\nu_H \gg \nu_V$, Deltares 2014). Because of the "sub-grid" nature of turbulence, the background values for horizontal eddy viscosity and diffusivity in eq. (3.6) and (3.7) have to be estimated on the basis of grid size (Okubo 1971, Deltares 2014, Piccolroaz et al. 2019).

For this reason different values were implemented for the constant horizontal background eddy viscosity and diffusivity. Namely, after Soulignac et al. (2017) and according to the grid size (Okubo 1971), they were set to 0.01 m^2 s^{-1} for the coarser grid (see chapter 5), and to 0.0025 m^2 s^{-1} for the finer grid (application in chapter 6).

3.2.1.3 Heat exchange with the atmosphere

Coupling the 3D model with climatic variables is crucial in order to correctly simulate the stratification and thermal dynamics in a water body. Different heat-flux models are implemented in FLOW to model energy exchanges at the air-water interface. They differ in terms of the number of meteorological variables required and therefore in the complexity of their formulations. In FLOW, the heat budget at the air-water interface is computed by taking into account the net incident solar radiation (Q_s), the heat losses due to back radiation (long wave, Q_b) and evaporation (latent heat flux, Q_e), and the sensible convective heat flux (Q_c). The various models differ in the way such terms are calculated or approximated (Deltares 2014). The total upward heat flux through the air-water interface (Q) is therefore:

$$Q = -Q_s + Q_b + Q_e + Q_c \tag{3.8}$$

Among the heat-flux models available in FLOW, the "Ocean" model and Murakami's model are the most complex ones, and in respect to equation (3.8) they present the lower number of approximated terms (Murakami et al. 1985, Deltares 2014). Both these models were used in this work, for different applications.

The Ocean model is the more widely used in scientific literature (Chanudet et al. 2012, McCombs et al. 2014, Soulignac et al. 2017, Piccolroaz et al. 2019). Here, it was used for the study described in chapter 6. It requires as inputs time series of relative humidity [-], air temperature [°C], net solar radiation [J s^{-1} m^{-2}], sky cloudiness [-], as well as wind speed [m s^{-1}] and direction [°N]. The Dalton and Stanton numbers can be set by the user. After (Soulignac et al. 2017), their values

were respectively set to 0.0015 and 0.00145. The "Ocean" model represents the go-to choice when all the required variables are available or can be easily estimated.

For the long-term climate change study (see chapter 5 for further details), the Murakami's model was used. This choice was motivated by the unavailability of sky cloudiness data covering the complete simulation period, and by the high uncertainty that affects ground-based cloud cover observations in any case (Silva & Souza-Echer 2016, Zelinka et al. 2017). As the implementation of the Murakami's model in FLOW does not require such data, it was preferred for this study to the "Ocean" model. Namely, Murakami's model requires as input time series of relative humidity [-], air temperature [°C], net solar radiation [J s^{-1} m^{-2}], wind speed [m s^{-1}] and direction [°N].

Both heat-exchange models explicitly require values for Secchi depth (H_S), which is the parameter that defines water transparency. It is correlated with the penetration of solar radiation in water through the light extinction coefficient ($\gamma = 1.7/H_S$, Poole & Atkins 1929) and therefore plays an important role on the stratification of the water column. *In situ* data show that the Secchi depth in Lake Champs-sur-Marne can vary suddenly between 0.8 m and the whole water column depth according to phytoplankton bloom events. For this reason, its value was calibrated for each application of Delft3D-FLOW, in order to optimize results in terms of thermal stratification.

Following Soulignac et al. (2017), a constant wind drag coefficient was applied to Lake Champs-sur-Marne. Its value was calibrated to 0.0013. Bottom roughness was computed through Chézy's formulation with the default value for the Chézy coefficient of 65 $m^{1/2}$ s^{-1}.

Finally, in all applications of FLOW, evaporative mass flux is neglected and water volume and depth are therefore considered as constant.

3.2.2 Delft3D-BLOOM

In the Delft3D modelling suite, BLOOM is part of water quality module (WAQ), and is designed to model phytoplankton growth through a specific competition principle among different phytoplankton functional groups.

The model BLOOM was used for one of the applications in this work: the implementation of Approximate Bayesian Computation for the automated calibration of a complex biogeochemical model (described in detail in chapter 6). The following paragraphs illustrate the main characteristic of the biogeochemical model itself, in relation to its set up for the study site.

3.2.2.1 General structure of the model

Differently from other hydro-biogeochemical models, BLOOM is not fully coupled with its hydrodynamic driver FLOW. The hydrodynamic simulation is run previously, independently from the biological counterpart. BLOOM shares with FLOW the computational grid and uses hydrodynamic results in terms of velocity fields and other environmental state variables such as water temperature for transport and growth. Biological processes therefore do not have any feedback on the hydrodynamics.

In BLOOM, the mass conservation of the substances included in the physical and biogeochemical processes is modelled on the basis of the advection-diffusion

equation. For each state variable and computational cell, the advection-diffusion equation is represented through:

$$M_i^{t+\Delta t} = M_i^t + \Delta t \cdot \left[\left(\frac{\Delta M_i}{\Delta t} \right)_{Tr} + \left(\frac{\Delta M_i}{\Delta t} \right)_{P} + \left(\frac{\Delta M_i}{\Delta t} \right)_{S} \right] \tag{3.9}$$

With:

M_i^t: mass at the beginning of a time step.

$M_i^{t+\Delta t}$: mass at the end of a time step.

$\left(\frac{\Delta M_i}{\Delta t}\right)_{Tr}$: changes by transport.

$\left(\frac{\Delta M_i}{\Delta t}\right)_{P}$: changes by physical or biogeochemical processes.

$\left(\frac{\Delta M_i}{\Delta t}\right)_{S}$: changes by sources (e.g. river discharges or waste loads).

Δt: time step duration of Delft3D-WAQ.

The different terms in eq. (3.9) are calculated based on the processes and substances activated in BLOOM, while the transport term is derived from the previously run hydrodynamic simulation (e.g. in terms of velocity field and turbulent dispersion). Point-source terms are not present in Lake Champs-sur-Marne and were therefore not modelled.

A high number of processes from the biogeochemical cycle are included in BLOOM, which is built with a modular structure that allows the user to activate or deactivate them. Namely, following the nomenclature used in BLOOM, its structure is based on a few very general "groups" to be chosen by the user. Each group then incorporates various substances, and for each substance various processes can be taken into account. A complex series of interdependences among the processes, that mirrors the complexity of a real ecosystem, is implemented in BLOOM and is automatically managed through the Delft3D graphical user interface.

A sketch of the substances included in the configuration implemented for Lake Champs-sur-Marne is shown in Figure 3.5. In the Figure, the substances are not strictly organized according to the groups available in Delft3D-BLOOM, but are rather arranged in order to show their main interactions. The activated groups activated in Delft3D-BLOOM are:

- Oxygen-BOD
- Particulate inorganic matter
- Dissolved inorganic matter (nutrients)
- Organic matter (Particulate and dissolved)
- Algae

Table 3.1 – Modules and variables activated in the configuration of the biogeochemical model Delft3D-BLOOM.

Group	**Substances**
Oxygen-BOD	Dissolved oxygen
Particulate and dissolved inorganic matter	Inorganic matter (IM1)
	Ammonium
	Nitrate
	Ortho-phosphate
	Adsorbed ortho-phosphate (AAP)
	Dissolved Silica
	Opal-Si
Organic matter	POC, fractions 1,2,3,4
	PON, fractions 1,2,3,4
	POP, fractions 1,2,3,4
	DOC
	DON
	DOP
Algae	Cyanobacteria
	Freshwater diatoms
	Freshwater flagellates
	Green algae

BLOOM includes a set of 23 phytoplankton groups. Four of these were included in the configuration for Lake Champs-sur-Marne, based on the available observations. Namely they are: cyanobacteria, green algae, flagellates and diatoms. Exchanges with atmosphere are activated, while benthic processes in the sediment layer are not taken into account. A complete list of the groups and substances implemented for the modelling of Lake Champs-sur-Marne is given in table 3.1.

Processes in BLOOM are highly parametrized. Default values of the parameters are derived from a years-long calibration process operated by the developers (Deltares 2018). However, such parameters still need to be calibrated on the specific study site. As it will be further explained in the following sections, BLOOM parameters were first calibrated by trial and error, while afterwards an automated calibration methodology is applied. This procedure is detailed in chapter 6.

A complete description of the groups and substances implemented in BLOOM can be found on the Delft3D-WAQ user manual (Deltares 2018). Only the mathematical representation of the processes more relevant for phytoplankton growth and organic matter mineralization will be briefly described in the next two sections.

3.2.2.2 Mathematical representation of phytoplankton growth

In BLOOM, phytoplankton dynamics are simulated through a competition principle that is able to choose, at every time step and each grid point, the best adapted phytoplankton group consistently with the available resources and the existing biomass level (Los 2009). This is a particular characteristic of BLOOM and it differentiates it from most other ecological models for aquatic ecosystems.

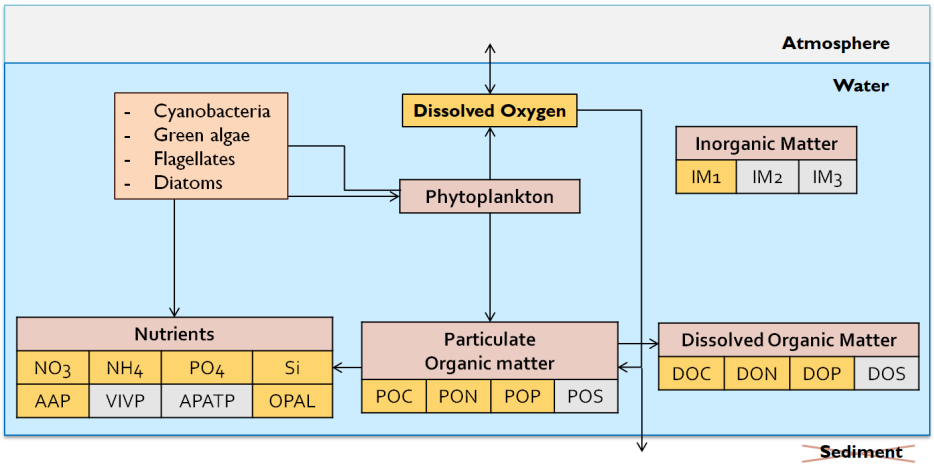


Figure 3.5 – Sketch of the substances activated in Delft3D-BLOOM and their main interactions.

The competition principle is based on the ratios between growth rates and nutrient demand. These ratios are calculated at all time steps for each activated phytoplankton group included in the model configuration. The ratios are then compared: the greater is the ratio, the more competitive a group is under the given conditions. The biomass is subsequently allocated among the different functional types through a specific linear programming algorithm, in order to maximize the total biomass. This way BLOOM is able to distribute biomass dynamically among the activated algal groups over time and space.

More precisely, the linear programming algorithm implemented in BLOOM optimizes total biomass while respecting four constraints in terms of nutrients, energy, growth and mortality.

Nutrient constraints

The solution of the linear programming algorithm must satisfy the following nutrient balance:

$$C_{tot,k} = C_k + \sum_{i=1}^{N} R_{k,i/C} \times C_{alg,i} - C_{min,k} \tag{3.10}$$

where $C_{tot,k}$ stands for the total available concentration of the nutrient k, C_k is the dissolved form of nutrient k, $R_{k,i/C}$ is the ratio between the nutrient k and carbon content in algal group i, $C_{alg,i}$ is the biomass concentration for algal group i, C_{min} indicates a threshold for uptake by algae, and the index k runs over the number of activated nutrients. The ratios $R_{k,i/C}$ (e.g. N:C, P:C and Si:C), specific to each phytoplankton group, therefore play a significant role in the allocation of biomass done by the competition principle.

Energy constraint

Phytoplankton growth depends on light availability. The penetration of solar radiation from the surface through the water column depends on the concentration

of dissolved and suspended matter in the water. Light extinction is modelled in BLOOM through a series of light-extinction coefficients (e_i), specific to each algal group i. Luminous energy becomes limiting for a phytoplankton group when the total extinction (e_{tot}) exceeds the maximum ($e_{max,i}$) at which growth is just balanced by respiration and mortality (Deltares 2018). Furthermore, the total extinction coefficient cannot be lower than a certain minimum value ($e_{min,i}$) given by the background extinction of light due to water and to the minimum algae concentration defined by the user. For each phytoplankton group i, the following constraint has therefore to be satisfied in order to have a net growth:

$$e_{min,i} \leq e_{tot} \leq e_{max,i} \tag{3.11}$$

with:

$$e_{tot} = \sum_{i=1}^{N} (e_i \cdot C_{alg,i}) \tag{3.12}$$

where $C_{alg,i}$ is the concentration of algal group i.

Growth and mortality constraints

The maximal growth allowed by the optimization algorithm for the algal group i is subject to the following constraint:

$$C_{M,alg,i}^{\,t+\Delta t} \leq C_{alg,i}^{\,t} \cdot \exp(R_{g,i} \cdot Ef_i - R_{r,i}) \cdot \Delta t \tag{3.13}$$

where $C_{M,alg,i}^{\,t+\Delta t}$ is the maximum concentration of algal group i at the end of a time step, $C_{alg,i}^{\,t}$ is the concentration of algal group i at the previous time step, $R_{g,i}$ is its potential specific growth rate, $R_{r,i}$ is its specific respiration rate, Ef_i is the light efficiency factor of phytoplankton i, and Δt is the BLOOM time step.

Similarly to growth, the mortality of a phytoplankton group is also constrained in order to prevent a complete removal within a single time step (Deltares 2018). The maximum biomass decrease of an algal group is obtained when there is no production, but only mortality. In this case the constraint reads:

$$C_{m,alg,i}^{\,t+\Delta t} = C_{alg,i}^{\,t} \cdot \exp(R_{m,i} \cdot \Delta t) \tag{3.14}$$

where $C_{m,alg,i}^{\,t+\Delta t}$ is the minimum concentration of algal group i at the end of a time step, $C_{alg,i}^{\,t}$ is the concentration of algal group i at the previous time step, $R_{m,i}$ is its potential specific mortality rate, and Δt is the BLOOM time step.

Potential growth and mortality rates therefore influence the outcome of the competition principle, which ultimately distributes the biomass among the phytoplankton groups at each time step. Either a linear or an exponential law can be chosen in BLOOM for the dependence of growth and mortality rates from water temperature. The latter was implemented in our configuration for all phytoplankton groups:

$$R = k_{0^\circ C}\ \theta^T \tag{3.15}$$

where R is here the potential growth (respectively, mortality) rate, $k_{0^\circ C}$ is the growth (respectively, mortality) rate at 0°C, θ is a coefficient for temperature dependence, and T is water temperature.

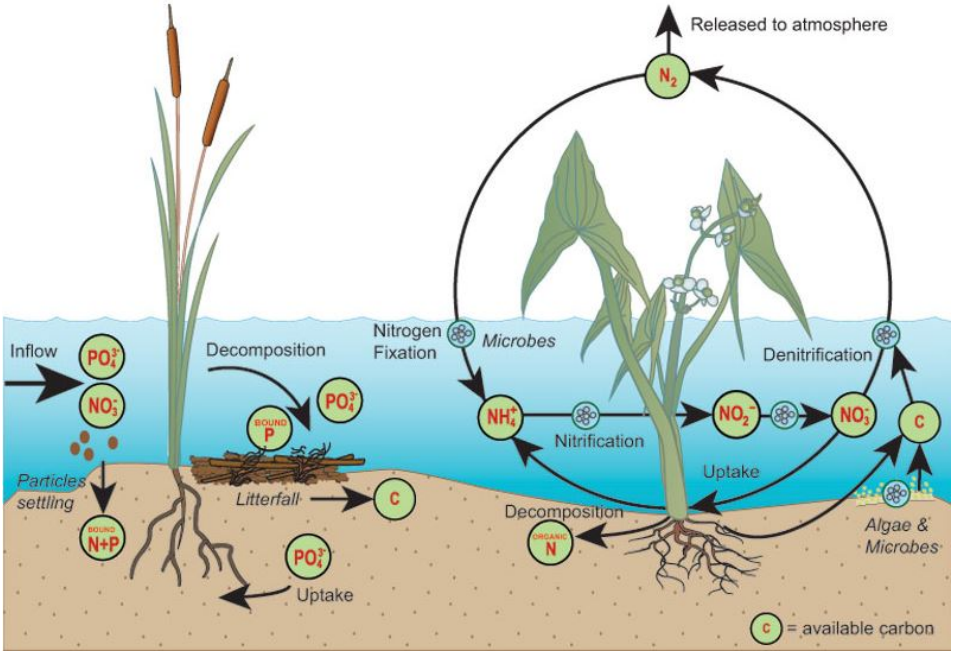


Figure 3.6 – Sketch of the nitrogen and phosphorus cycles in aquatic ecosystems (adapted from: Kadlec & Knight 1996).

Furthermore, another peculiar trait of the model BLOOM is the concept of "phenotypes". Each algal group is modelled at each time step as one of three phenotypes, depending on the external conditions, or, more precisely, on the current limiting factor. Algal groups can therefore be (i) energy limited (limitation by light), (ii) nitrogen limited, or (iii) phosphorus limited. The phenotypes of a single algal group respond to the same mathematical laws, but might be parametrized differently from one another. Further details on how the four constraints of the optimization algorithm are implemented in order to take into account for the possible transitions between phenotypes can be found on the BLOOM user manual (Deltares 2018).

3.2.2.3 Nutrient cycles and organic matter mineralization

Three nutrient cycles are considered in BLOOM: nitrogen, phosphorus and silicon, while the carbon cycle is indirectly modelled through the mass balance of all components containing organic carbon. Some processes in these cycles are of particular importance for primary production in aquatic ecosystems, and were therefore included in the configuration implemented for Lake Champs-sur-Marne: (i) nitrification, i.e. the microbial oxidation of ammonium (NH_4) to nitrate (NO_3, more readily available for uptake by phytoplankton), (ii) denitrification, the reduction of nitrate into elemental nitrogen under anoxic conditions, and (iii) adsorption of dissolved phosphate (mainly ortho-phosphate) onto suspended sediment. Figure 3.6 shows a sketch of the nitrogen and phosphorus cycles in aquatic ecosystems.

In a water body without any inlet or outlet as Lake Champs-sur-Marne, dissolved and particulate organic matter is mainly produced when algae and higher plants die off. The microbial decomposition of organic matter into its basic inorganic compounds

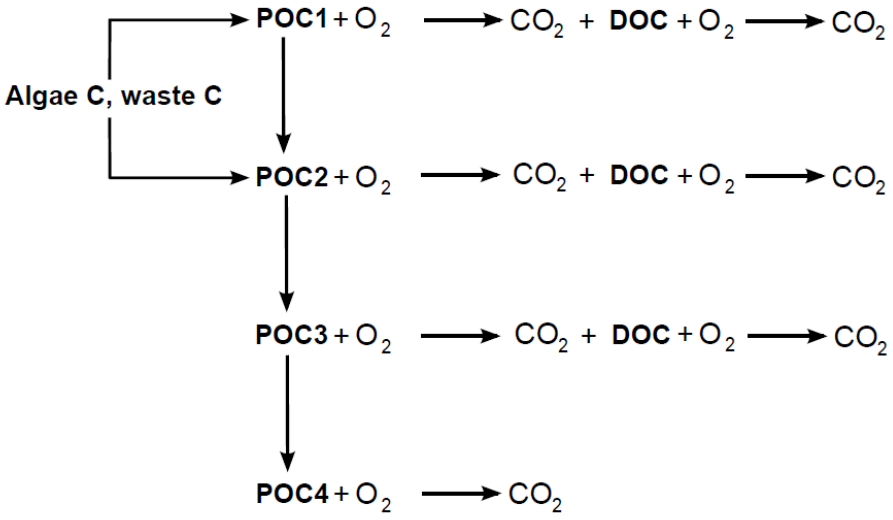


Figure 3.7 – Phases of the mineralization process as modelled in BLOOM, for the case of POC (source: Deltares 2018).

such as carbon dioxide, ammonium, phosphate and sulfide is called mineralization, and is modelled in BLOOM as a first order kinetic (Deltares 2018). During the decomposition process, while dissolved organic matter is quickly mineralized, the particulate fraction is gradually converted into forms that are more resistant to microbial breakdown (Deltares 2018). For this reason, the particulate fraction of organic matter is divided in BLOOM into five categories:

- POX_1: fast decomposing fraction
- POX_2: medium-slow decomposing fraction
- POX_3: slow decomposing fraction
- POX_4: particulate refractory fraction
- POX_5: organic matter contained in stems and roots

where POX generically stands for particulate organic nitrogen (PON), phosphorus (POP) and carbon (POC). Only fraction form one to four are considered in the configuration implemented for the study site. Figure 3.7 shows the phases of the degradation process as modelled in BLOOM for the case of POC.

3.2.3 TELEMAC 3D

TELEMAC3D is an open-source three-dimensional hydrodynamic model developed by Electricité de France (EDF) and a group of partners (further information at: `http://www.opentelemac.org/index.php/presentation?id=18`). It can be downloaded, together with its two- and one-dimensional counterparts (TELEMAC2D and Mascaret, respectively) from: `http://www.opentelemac.org/`. TELEMAC3D has been applied and validated in river and coastal environments as well as lagoons or reservoirs (e.g. Laurent et al. 2010, Villaret et al. 2013, Feng & Jodeau 2016, Cooper et al. 2017).

Its code solves through numerical methods based on finite elements the three-dimensional free surface Navier-Stokes equations under the Boussinesq, hydrostatic pressure and incompressible fluid assumptions (Telemac modelling system 2016). A detailed description of the model's mathematical and computational approach can be found in Hervouet (2007) and in the user manual (Telemac modelling system 2016).

TELEMAC3D has been coupled with the biogeochemical library Aquatic EcoDynamics library (AED2) by Laboratoire Nationale d'Hydraulique et Environnement (EDF R&D) and Laboratoire Hydraulique Saint-Venant (EDF R&D / ENPC / CEREMA) enabling the modelling of the biogeochemical cycle in the Telemac-Mascaret system. However, the coupling between the models is recent (started in 2015), and the works on the coupling technique remain in progress. This has influenced the implementation both of the hydrodynamic and biogeochemical models. For example, at present, the Z-model, the phytoplankton uptake of Silica or the fixation of Nitrogen are not currently implemented in the coupled version of the models. The present study represents one of the first tests of the coupled models in a natural environment.

In the coupling, the hydrodynamic model manages the main "physical" processes (e.g. transport, velocity field and heat exchange with the atmosphere), while AED2 simulates the biogeochemical cycle. Differently from the Delft3D suite, where the hydrodynamic simulations are independent from the results of the biogeochemical model, TELEMAC3D and AED2 are fully coupled. This means that the equation from both models are computed at each time step, evolving together. As it will be further explained in the following sections, this implies a potential feedback of AED2 on the hydrodynamic results. For this reason, the general description of the two models given in the following sections cannot be completely disentangled.

3.2.3.1 Numerical aspects and computational domain

In TELEMAC3D, the Navier-Stokes equations are solved in three subsequent stages. First (advection step), the advected velocity are estimated taking into account only the convective terms in the momentum equations (see eq. (2.11)). The resulting velocity field is then used to obtain the new velocity field taking into account both diffusive and source terms in the momentum equations (diffusion step). Finally (propagation step), water depth is computed from the vertical integration of the continuity and momentum equations (see equations (2.10) and (2.11)), only including the pressure-continuity terms (Telemac modelling system 2016).

The numerical scheme chosen here for the advection step is the MURD-PSI

(Multidimensional Upwind Residual Distribution - Positive Streamwise Implicit) with an explicit time discretization. This configuration is suitable for tracers as it is conservative and monotonic, and introduces limited numerical diffusion (Telemac modelling system 2016, Malcherek 2000). The default implicit method was used for the diffusion step. All linear systems associated with the numerical resolution of the Navier-Stokes equations are addressed with the default conjugate gradient method, except for the propagation step, for which the Generalised Minimum RESidual (GMRES) method was activated.

The open-source software BlueKenue™ (Canadian Hydraulic Centre 2011) was used to create the 2D triangular mesh used both in TELEMAC3D and AED2. The triangular mesh is shown in Fig. 3.8, and is built with an average distance between the nodes of 20 m, and a refined zone around the narrower portion of the water body.

The bathymetric data available form the echo-sounder campaign of 2016 were here translated according to the local elevation and used to represent the bottom of the lake (see Fig. 3.8). The raw data were interpolated through the BlueKenue™ software, and a constant elevation of 39.5 m a.s.l. was imposed along all the solid boundaries of the model.

Eventually, the 3D computational domain is obtained from the 2D mesh in Figure 3.8, through the discretization of the vertical axis, done by TELEMAC3D. Differently from the Delft3D configuration, where the Z-model was implemented for the discretization along the vertical axis, for TELEMAC3D/AED2 the standard σ-transformation is activated. The Z-model is available in TELEMAC3D. However, at present, this option is not yet correctly coupled with the biogeochemical model AED2. Some preliminary tests were run on the stand-alone version of TELEMAC3D in order to test the effect of the two discretization models on model results in terms of thermal stratification, and did not show considerable differences in the water temperature results.

Finally, the grid shown in Figure 3.8 is composed of 404 nodes (661 triangular elements), and 10 σ-layers of varying thickness along the vertical axis.

3.2.3.2 Turbulence closure

A mixing length turbulence closure model was activated for all applications with TELEMAC3D. Preliminary tests showed that the k-ε turbulence model was considerably more diffusive than the mixing length model. In a shallow lake such as Lake Champs-sur-Marne, the mixing length model performed better in terms of thermal stratification, especially when it was coupled with a dumping function. Mixing length models express turbulent viscosity (ν) as a function of the mean velocity gradient and of the mixing length (L_m) (Telemac modelling system 2016):

$$\nu = L_m^2 \sqrt{2 D_{ij} D_{ij}} \tag{3.16}$$

where the strain rate D_{ij} is expressed a function of the mean velocities of the fluid ($\overline{U}_{i,j}$):

$$D_{ij} = \frac{1}{2} \left(\frac{\partial \overline{U}_i}{\partial x_j} + \frac{\partial \overline{U}_j}{\partial x_i} \right) \tag{3.17}$$

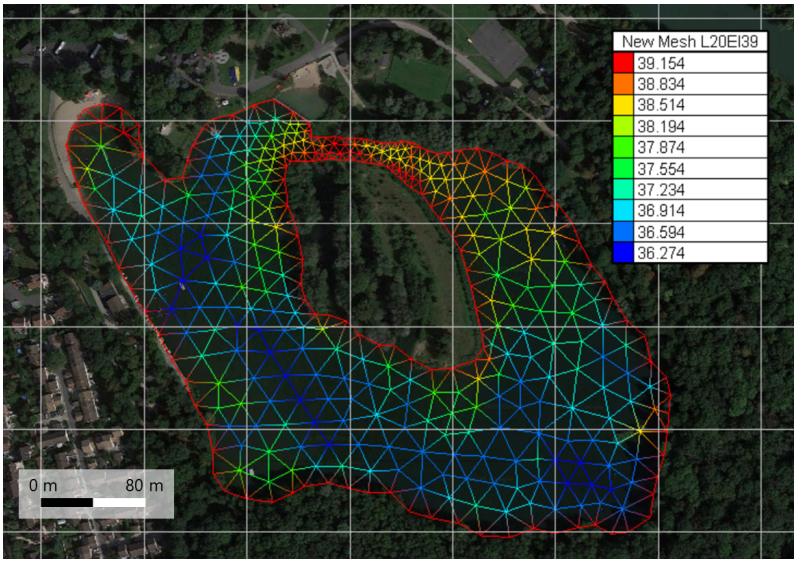


Figure 3.8 – Example of the computational domain for Lake Champs-sur-Marne in TELEMAC3D and AED2 (mean distance between elements: 20 m). The colour chart shows the elevation of each point of the grid in meters above sea level.

For the definition of the mixing length, the formulation proposed by Nezu and Nakagawa was activated (Nezu & Nakagawa 1994, Telemac modelling system 2016).

Damping functions can be introduced into a mixing length model expressly to take into account the reduced mixing under stratified conditions. In the presence of vertical density gradients, the exchanges of mass and momentum are hindered (respectively enhanced) by water column stability (instability) (Telemac modelling system 2016). The Richardson's number ($R_i = \frac{g\nabla\rho}{\rho\nabla u^2}$) is a local property of the fluid and represents the ratio between the buoyancy and the flow shear terms. It is commonly used to quantify the effects of the gravity terms in the turbulent power balance. Damping functions are usually formulated depending on its local value. Namely, Viollet's formulation was adopted here (Viollet 1988, Telemac modelling system 2016):

$$f(R_i) = (1 + 14R_i)^{0.75} \tag{3.18}$$

In TELEMAC3D, activating the mixing length model for vertical turbulence implies the use of constant values for horizontal viscosity and diffusivity. For this reason, the value of the horizontal viscosity and diffusivity of velocities were considered as parameters to be calibrated based on model results and grid size (Okubo 1971).

3.2.3.3 Heat exchange with the atmosphere

The effects of wind and meteorological conditions at the air-water interface were activated. The TELEMAC system offers two models of different complexity to account for the heat exchange at the air-water interface. The simpler one is a linearised formulation to be calibrated based on the available data. The more

comprehensive formulation was activated in this work. It takes explicitly into account the following terms:

$$Q = -Q_s + Q_b + Q_e + Q_c + Q_p \tag{3.19}$$

where, using the same notation used in section 3.2.1.3, Q is the total upward heat flux through the air-water interface, Q_s is the net incident solar radiation, Q_b is the effective back radiation, and Q_e, Q_c and Q_p are the heat transfer due to evaporation, convection and precipitation, respectively (Bourban & Huang 2018).

The meteorological variables needed to simulate the heat exchanges at the air/water interface are: relative humidity [-], air temperature [°C], cloud cover [-], net solar radiation [J s^{-1} m^{-2}], wind speed [m s^{-1}] and direction [°N]. The contribution of precipitation was neglected in this work both in terms of energy and mass.

TELEMAC allows to take into account the fact that meteorological forcing data might not be available *in situ* and might be recorded far from the study site. This is done through a wind correction function:

$$f(U_2) = b(1 + U_2) \tag{3.20}$$

where U_2 is the wind speed at 2 m altitude (converted from wind speed at 10 m altitude through: $U_2 = 0.85U_{10}$), and b is a parameter to be calibrated by the user (the "coefficient to calibrate the atmosphere-water heat exchange"), which constitutes the main parameter to calibrate for the whole heat exchange module (Telemac modelling system 2016).

Similarly to Delft3D-FLOW, the light extinction coefficient (γ) in TELEMAC3D stand-alone is calculated based on a constant Secchi depth value defined by the user ($\gamma = 1.7/H_S$, Poole & Atkins 1929). However, when the coupling with AED2 is activated, this is not the case any more. As it will be further explained in the following sections, AED2 estimates the water light extinction coefficient dynamically based on the concentration of the various dissolved and particulate fractions activated and on their specific light extinction coefficient values. In such case the computation of the light extinction coefficient is therefore transferred to the biogeochemical library, and its calibration is indirectly obtained by setting the values of the specific light extinction coefficients.

3.2.4 Aquatic EcoDynamics

This section contains a general description of the mathematical approach at the basis of the model, and of the configuration set up for Lake Champs-sur-Marne. As it will be further detailed in chapter 7, AED2 was applied focusing on different time scales: from short-term simulations of a few weeks, to a year-long simulation. Such applications required slightly different configurations. However, the following paragraphs will be as general as possible. When necessary, additional and specific details will be given for each application in chapter 7.

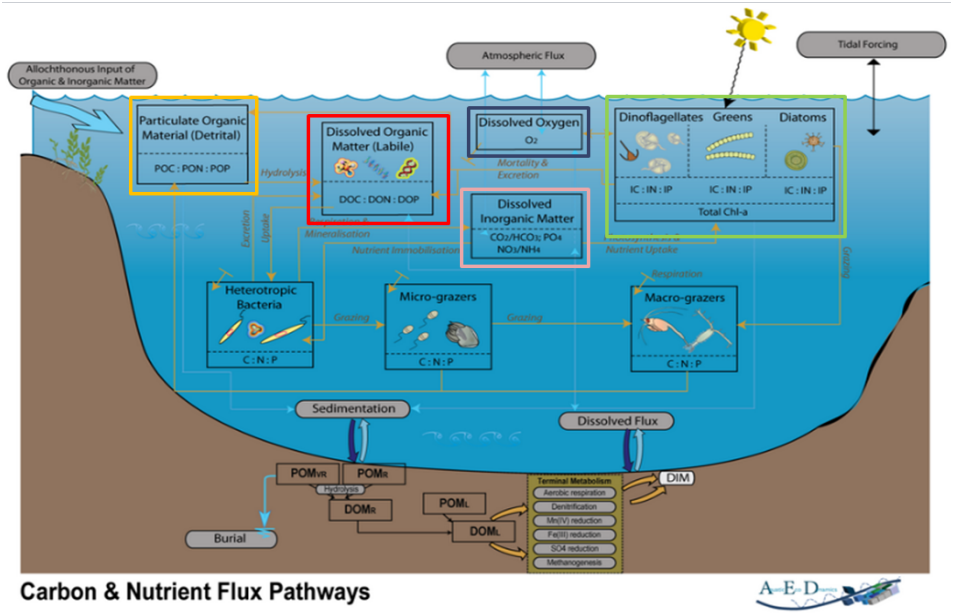


Figure 3.9 – Sketch of the some of the most commonly used modules in AED2. Those framed with colours are activated in the configuration for Lake Champs-sur-Marne (adapted from: Hipsey et al. 2013).

3.2.4.1 General structure of the model

Aquatic EcoDynamics is a modular biogeochemical library, developed by the University of Western Australia (`https://aquatic.science.uwa.edu.au/`), that potentially allows the user to take into account all the processes playing a role in the biogeochemical cycle in an aquatic ecosystem, from benthic fluxes and microbial decomposition to primary production, micro- and macro-grazing, and macrophytes.

It solves a series of ordinary differential equations that model cycles of the main nutrients (N, P and Si), oxygen and carbon, including phytoplankton and zooplankton groups. When coupled with a hydrodynamic model, AED calculates these functions at each time step, passing the calculated terms to the hydrodynamic counterpart, which handles transport.

Its main modules are similar to those discussed for Delft3D-BLOOM. They are shown in Figure 3.9, where those activated for this work are framed in colours. In Figure 3.9, the model variables are not strictly grouped as in the AED2 modules listed above. For instance, particulate and dissolved fractions of organic matter are represented separately in Figure 3.9, but are actually part of the same "organic matter" module in AED2. In Figure 3.9, the model variables are reorganised based on their roles in the biogeochemical cycle, in order to show their interdependences.

Namely, the modules activated are:

- Dissolved Oxygen
- Particulate organic matter
- Dissolved organic matter

- Dissolved inorganic matter (Carbon, Phosphorus and Nitrogen)
- Phytoplankton

As for Delft3D-BLOOM, benthic processes in the sediment layer are not explicitly taken into account.

According to the modules activated in the configuration for Lake Champs-sur-Marne, the total concentration of five main chemical elements must remain constant in the system at each time step: oxygen, carbon, nitrogen, phosphorus and silica. Their total amount is distributed among multiple biological and chemical compounds, whose concentration varies in time and space. AED2 calculates the evolution in time of all the variables activated (e.g. the concentration of biological or chemical compounds) based on a series of mass balances, generally expressed through partial differential equations. The variations of their concentrations due to transport (advection or diffusion), are handled by the hydrodynamic model.

A detailed description of all the equations used in the model can be found in the user manual (Hipsey et al. 2013). Here, the case of nitrogen is briefly described as an example: this will allow to illustrate the formulation of important processes such as organic matter mineralization and phytoplankton uptake and growth. The total nitrogen concentration (TN) is given by:

$$TN = NO_3 + NH_4 + DON + PON + \sum_{i}^{N_{PHY}} PHY_{N_i} \tag{3.21}$$

where DON and PON are respectively the dissolved and particulate fraction of organic nitrogen, PHY_{N_i} is the nitrogen concentration in the i-th phytoplankton group, N_{PHY} is the total number of active phytoplankton groups, and the zooplankton contribution was neglected.

Each compound in eq. (3.21) is modelled through a mass balance that takes into account the different processes influencing its concentration. For instance, the mass balances for the two main nitrogen-based inorganic nutrients, ammonium and nitrate (expressed through their concentrations $[\mathrm{NH_4}]$ and $[\mathrm{NO_3}]$) are:

$$\frac{d[NH_4]}{dt} = f_{sed}^{NH_4} - f_{nitrif}^{NH_4} + f_{miner}^{DON} - \sum_{i}^{N_{PHY}} p_{NH_4}^{i} \cdot f_{uptake}^{PHY_{Ni}} \tag{3.22}$$

$$\frac{d[NO_3]}{dt} = -f_{sed}^{NO_3} + f_{nitrif}^{NH_4} - f_{denit}^{NO_3} - \sum_{i}^{N_{PHY}} p_{NO_3}^{i} \cdot f_{uptake}^{PHY_{Ni}} \tag{3.23}$$

where p is a preference factor for the substance under consideration specific to each phytoplankton group, and all fluxes (f_X^Y) represent the contribution of process X to variable Y. Furthermore, the mass balance for dissolved organic nitrogen (DON) is:

$$\frac{d[DON]}{dt} = f_{decom}^{PON} + f_{sed}^{DON} - f_{miner}^{DON} - \sum_{i}^{N_{PHY}} f_{excr}^{PHY_{Ni}} \tag{3.24}$$

with similar symbols meaning. The fluxes in equations (3.22), (3.23) and (3.24) show us the processes considered in AED2 for the concentration of three variables examined here ($\mathrm{NH_4}$, $\mathrm{NO_3}$ and DON), and how they might affect multiple variables. Namely,

such processes are: sediment release or demand, nitrification of NH_4, denitrification of NO_3, mineralization of DON, decomposition of PON, uptake and excretion by phytoplankton. Their formulation is often based on Michaelis-Menten-like kinetics. This is the case, for instance, of the nitrification flux:

$$f_{nitrif}^{NH_4} = R_{nitrif} \frac{[O_2]}{K_{nitrif} + [O_2]} \theta_{nitrif}^{T-20} [NH_4] \tag{3.25}$$

where R_{nitrif} represents the maximum nitrification rate, K_{nitrif} its half-saturation constant, and θ_{nitrif} is a factor for temperature (T) scaling.

An equivalent formulation is also adopted for the fluxes of dissolved organic matter mineralization and particulate organic matter decomposition, that intervene in eq. (3.22) and (3.24). In the case of nitrogen, the mineralization flux of DON is:

$$f_{miner}^{DON} = R_{miner} \frac{[O_2]}{K_{miner} + [O_2]} \theta_{miner}^{T-20} [DON] \tag{3.26}$$

where R_{miner} represents the maximum mineralization rate, K_{miner} its half-saturation constant, and θ_{miner} is a factor for temperature (T) scaling. The formulation for PON decomposition into DON is equivalent and can be found in Hipsey et al. (2013).

3.2.4.2 Mathematical representation of phytoplankton growth

Similarly to what discussed for ammonium, nitrate and DON, phytoplankton concentration is also modelled through mass balances. Namely, in AED2 the concentration of each phytoplankton group is split into the contributions of the basic chemical elements composing the group under consideration (i.e. C, N, P and, possibly, Si). A mass balance is implemented for each element contribution. For the case of carbon:

$$\frac{d[PHY_{C_i}]}{dt} = f_{uptake}^{PHY_{C_i}} - f_{excr}^{PHY_{C_i}} - f_{mort}^{PHY_{C_i}} - f_{resp}^{PHY_{C_i}} - f_{sett}^{PHY_{C_i}} \tag{3.27}$$

where [PHY_{C_i}] is the carbon concentration of the i-th phytoplankton group. The fluxes f respectively represent the following processes: uptake, excretion, mortality, respiration, sedimentation, and the zooplankton contribution was neglected. Similar partial differential equations are implemented for the phytoplankton concentrations of the remaining elements (N, P, Si). The expressions of the excretion, mortality, respiration and settling fluxes can be found in Hipsey et al. (2013). The carbon phytoplankton uptake flux embodies the photosynthetic processes and is therefore described here in more detail.

Phytoplankton carbon uptake defines phytoplankton growth and depends, for each algal group, on the local environmental conditions in terms of water temperature and nutrient and light availability. A temperature scaling function is always taken into account, while only one limiting factor is considered relatively to light and nutrients. The carbon uptake is modelled through the following equation:

$$f_{uptake}^{PHY_{C_i}} = R_{growth}^{PHY_i}[PHY_{Ci}] \underbrace{(1 - k_{pr}^{PHY_i})}_{\substack{\text{photorespiratory} \\ \text{loss}}} \underbrace{\Phi_{tem}^{PHY_i}(T)}_{\substack{\text{temperature} \\ \text{scaling}}} \underbrace{\Phi_{str}^{PHY_i}(T)}_{\substack{\text{metabolic} \\ \text{stress}}}$$
$$\times \min\Big\{\underbrace{\Phi_{light}^{PHY_i}(I)}_{\substack{\text{light} \\ \text{limitation}}}, \underbrace{\Phi_{N}^{PHY_i}(NO_3, NH_4, PHY_{N_i})}_{\substack{\text{N} \\ \text{limitation}}}, \underbrace{\Phi_{P}^{PHY_i}(PO_3, PHY_{P_i})}_{\substack{\text{P} \\ \text{limitation}}}, \underbrace{\Phi_{Si}^{PHY_i}(Si)}_{\substack{\text{Si} \\ \text{limitation}}}\Big\} \tag{3.28}$$

where $[PHY_{Ci}]$ is the carbon concentration of phytoplankton group i, and $R_{growth}^{PHY_i}$ is the maximum growth rate of phytoplankton group i at 20°C. Specific functions (here generically labelled Φ) are implemented in the model to calculate each limiting factor, as well as the photorespiratory loss, the metabolic stress and the temperature scaling. Their formulation is available in the user manual (Hipsey et al. 2013). The nitrogen and phosphorus fluxes due to phytoplankton uptake (that appear in equations (3.22) and (3.23)) are simply modelled as a constant fraction of the phytoplankton carbon uptake.

3.2.4.3 Dynamic calculation of the light extinction coefficient

The light extinction coefficient plays an important role in the availability of light over the water column. Incident shortwave radiation is supplied to AED2 by the hydrodynamic driver. For primary production, AED2 converts the incident shortwave radiation into its photosynthetically active portion (PAR, roughly 45%). PAR than penetrates into the water column according to the Beer-Lambert law (Hipsey et al. 2013), which can be expressed as:

$$ln(I_z) = ln(I_0) - K_d z \tag{3.29}$$

where I_0 is PAR at the surface, I_z is the downwelling irradiance at depth z, and K_d is the light extinction coefficient (Weiskerger et al. 2018).

The light extinction coefficient also plays an important role in the vertical distribution of heat, and therefore for thermal stratification, which is managed by the hydrodynamic model TELEMAC3D. Through AED2, the light extinction coefficient is calculated dynamically based on the concentrations of phytoplankton and dissolved and particulate fractions of organic carbon.

Namely, it is computed through the following equation:

$$K_d = K_w + K_{e,DOC}DOC + K_{e,POC}POC + \sum_{i}^{N_{PHY}} K_{e_i}PHY_{C_i} \tag{3.30}$$

where K_w is the base light extinction coefficient associated with water, and $K_{e,DOC}$, $K_{e,POC}$ and K_{e_i} represent respectively the specific light extinction coefficient for dissolved organic carbon, particulate organic carbon and for the phytoplankton group i.

In terms of thermal stratification, the dynamic calculation of K_d induces a considerable feedback of AED2 on the outcomes of the hydrodynamic model, which is absent in the partial coupling of the models from the Delft3D suite.

Chapter 4

Collected data set

This chapter describes the data set collected on the study site used in this work. Section 4.1 describes my participation to the field campaigns and sums up the complete data set available for the study site used in this work. Section 4.2 describes the high-frequency data were post-processed. Finally, the main physical and biological characteristics of Lake Champs-sur-Marne are described, through the available data, in section 4.3.

4.1 Field work and available data set

The measuring instruments and the field campaigns are described in sections 3.1.2 and 3.1.3. The monitoring of Lake Champs-sur-Marne includes *in situ* high-frequency measurements, profiling and water samples analysis from the field campaigns. Tables 4.1, 4.2 and 4.3 sum up the data collected on the study site at sites A, B, and P by LEESU between 2015 and 2020. In particular, table 4.1 is relative to the measurements recorded at high-frequency, table 4.2 to the measurements recorded through profiling instruments and Secchi disk, and table 4.3 covers the analysis done on the water samples.

My participation to the field campaigns and to the data collection started in November 2017 and continued until April 2021. However, during the whole period of my PhD, the monitoring of the study site was shared with the technical staff at LEESU, with other fellow PhD students and occasionally with interns. My contribution to the collection of the data concerned the participation to the field campaigns and the storage, post processing and analysis of the collected data. Laboratory analysis of the water samples were carried out by specialized researchers at LEESU. The water samples were analysed for total chlorophyll, phycocyanin, total nitrogen, total phosphorus, nitrate and phosphate concentrations, dissolved organic carbon and algal species identification. Further details on the water samples analysis (e.g. location and depth of the water samples, period during which the analysis were carried out) can be found in table 4.3.

Table 4.1 – List of variables recorded at high-frequency at the three measuring sites (A, B and C), depth and date of installation of the corresponding sensor. All variables were measured with a five minutes frequency until March 2018, and with a ten minutes frequency since.

		Variables measured at high-frequency					
Site	Depth	Temperature	Oxygen	Total chl.	Phycocyanin	Conductivity	NO_3
A	0.5 m	✓					
	1.5 m	✓	✓	✓		✓	
	2.5 m	✓					
Start:		May 2015					
B	0.5 m	✓					
	1.5 m	✓	✓	✓	✓	✓	✓
	2.5 m	✓					
Start:		Dec. 2015					Nov. 2018
P	0.5 m	✓					
	1.5 m	✓	✓	✓		✓	
	2.5 m	✓					
Start:		Mar. 2017					

Table 4.2 – List of variables recorded *in situ* during the field campaigns with the two profiling instruments, location and date of first utilisation.

	Seabird					
Site	Depth	Temperature	Oxygen	pH	Conductivity	**Start**
A	✓	✓	✓	✓	✓	Sept. 2015
B	✓	✓	✓	✓	✓	Sept. 2015
P	✓	✓	✓	✓	✓	Apr. 2017

	Fluoroprobe					
Site	Total chl.	Green algae	Cyanobacteria	Diatoms	Cryptophyceae	**Start**
A	✓	✓	✓	✓	✓	Sept. 2015
B	✓	✓	✓	✓	✓	Sept. 2015
P	✓	✓	✓	✓	✓	Apr. 2017

Table 4.3 – List of variables analysed from the water samples; location and depth of the water samples and date of first laboratory analysis for each variable.

		Water samples							
Site	Depth	Total N	Total P	NO_3	PO_4	Total chl.	Phyco.	DOC	Alg. Count.
A	1.5 m	✓	✓		✓	✓	✓		
B	1.5 m	✓	✓	✓	✓	✓	✓	✓	✓
P	1.5 m	✓	✓		✓	✓	✓		
Start:		Jun. 2017						Sep. 2018	Apr. 2019

4.2 High-frequency data treatment

The data set available (resumed in tables 4.1, 4.2 and 4.3) covers a wide range of variables crucial for the biogeochemical cycle, and was extremely important to analyse and understand the thermal and biogeochemical dynamics of the study site. Furthermore, the high-frequency data set was particularly exploited in this work for the calibration and validation of the deployed models, and, more generally, to test their performances. Before doing so, the high-frequency observations were verified by comparison with the other data sources coming from the field campaigns. In the case of persistent discrepancies, the stock of unreliable high-frequency data was detected and was either discarded or adjusted through correction factors (as described for total chlorophyll later in this section).

The raw high-frequency data set presents considerable scattering. For this reason, such data needed to be smoothed and cleaned from the outliers originated by periodic sensors maintenance during field campaigns. Furthermore, when the data needed to be compared with model outputs, they were further averaged in order to match the model output time step. Among the high-frequency data set, water temperature is the variable presenting the lowest scattering. For this reason, such data were not smoothed but only cleaned from the outliers. The latter were defined as sudden water temperature variations ($> 1°C$) over the 10 minutes separating two successive measurements, and consequently erased. All other variables were smoothed with a four-hours moving average window, and the outliers originated during the field campaigns were detected and manually erased from the data set. The high-frequency data set smoothed and cleaned from the outliers is shown in Figure 4.1 for site B between 2016 and 2019, in terms of water temperature, total chlorophyll, cyanobacteria concentration (in the raw measuring unit), dissolved oxygen and nitrate (NO_3). In Figure 4.1, data are only shown for site B.

Starting from 2017, total chlorophyll measurements were available from three different data sources: high-frequency observations, BBE Fluoroprobe periodic profiles, and analysis of water samples. The comparison of the three data set showed a general underestimation by the chlorophyll fluerescence *in situ* sensors, in particular during cyanobacteria dominated blooms. In order to address this issue, the available data set was used to implement correction factors to be applied to the high-frequency measurements. Such correction factors evolved through time together with the expansion of the available data set.

A first correction factor ($Chl_{corr} = 1.93 \cdot Chl_{raw}$) was initially implemented in 2018 using the BBE Fluoroprobe profile series, which are available since 2015 on the study site. A linear regression model was set up between total chlorophyll data from the Fluoroprobe profiles and the corresponding high-frequency measurements. For both data sources, the data used to build the linear model are relative to 1.5 m of depth at site B. Only a few data from water samples analysis were available at this time, and were therefore not included in this initial analysis. This correction factor was used for short-term applications (see chapters 6 and 7). Later, in 2020, a multi-linear regression model was built exploiting the complete available data set in terms of chlorophyll and phycocyanin fluorescence and laboratory analysis, with the objective to achieve a more robust correction of the high-frequency total chlorophyll measurements. In this thesis, such model was applied to the data of the year 2019,

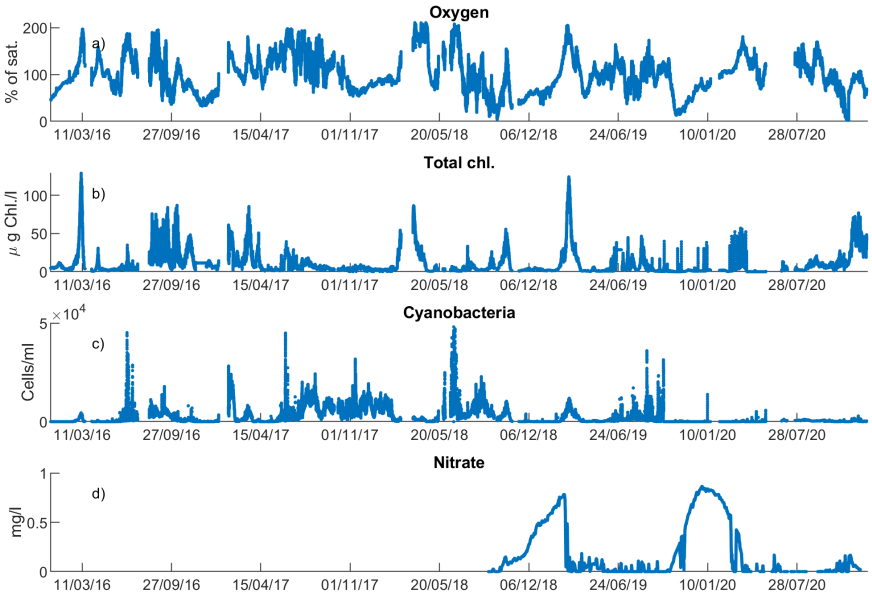


Figure 4.1 – High-frequency data set at site B between 2016 and 2020, smoothed and cleaned from the outliers for dissolved oxygen (panel a), total chlorophyll (panel b), cyanobacteria (panel c) and nitrate concentrations (panel d).

used for the seasonal application described in chapter 7.

High-frequency measurements of cyanobacteria are given by the sensor in cells/ml. Their conversion into chlorophyll equivalents is done exploiting the BBE Fluoroprobe profiles and data from laboratory analysis. Similarly to what was discussed above for total chlorophyll, the conversion factor evolved through time together with the expansion of the available data set. A proportional factor of 0.001 was initially implemented for the short-term simulation of 2018 using the BBE Fluoroprobe profile series to convert high-frequency measurements into equivalent of chlorophyll [μg Chl./l]; this is also the same conversion factor implemented in the two BBE Moldaenke™profilers used on the study site (Fluoroprobe and AlgaeTorch). In 2020, a multi-linear regression model was built exploiting profiles and laboratory analysis and was applied to the high-frequency data of cyanobacteria for the year 2019.

The result of the treatment and correction described above are shown for total chlorophyll and cyanobacteria chlorophyll in panels b and c of Figure 4.2, for the year 2019, along with the corresponding laboratory analysis from water samples and BBE Fluoroprobe profiles measurements. The Figure also shows the comparison between high-frequency nitrate data and water samples laboratory analysis (panel d), as well as high-frequency data for surface and bottom water temperatures (panel a). All the data in Figure 4.2 are relative to measuring site B. Figure 4.2b shows a very good agreement between the corrected high-frequency data of total chlorophyll and both the water samples analysis and the Fluoroprobe profiles. The only exception being the Fluoroprobe profile of 13/03/2019. Similar considerations can be made for cyanobacteria (Figure 4.2c). Eventually, high-frequency observations of nitrate also

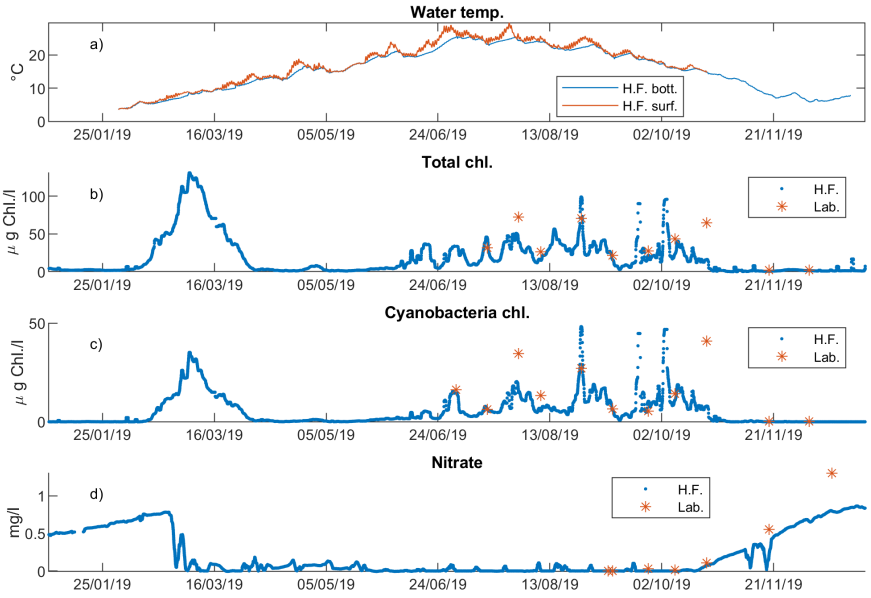


Figure 4.2 – Treated and corrected high-frequency data set at site B for the year 2019 for: water temperature (panel a), total chlorophyll (panel b), cyanobacteria (panel c) and nitrate (panel d) concentrations. When available, data from water sample analysis and periodic Fluoroprobe profiles are also shown.

match very closely with data from water samples (Figure 4.2d).

4.3 Physical and biogeochemical characteristics of the study site

The data shown in Figure 4.2 can be further analysed in order to highlight the main hydrodynamic and biogeochemical traits characterizing the study site. Exploiting the data from 2019, the following paragraphs contain a general description of the overall characteristics of the study site in terms of thermal, phytoplankton and nutrient dynamics.

Lake Champs-sur-Marne is a polymictic water body that alternates mixing and periods of strong and stable thermal stratification (i.e. thermal stratification lasting for at least two consecutive days without breakup at night). Figure 4.2a shows bottom and surface water temperature at site B for the year 2019. Strong differences in water temperature between the surface and the bottom layers (respectively located at 0.5 and 2.5 m of depth) can be seen between the months of March and October. Maximum temperature differences between the surface and bottom layers can reach up to roughly 6°C and stable thermal stratification can last up to two or three consecutive weeks.

The phytoplankton and nitrate dynamics observed in 2019 are exemplary of the general evolution of phytoplankton biomass on the study site along the span of a whole year. The lake is usually characterized by very low phytoplankton biomass

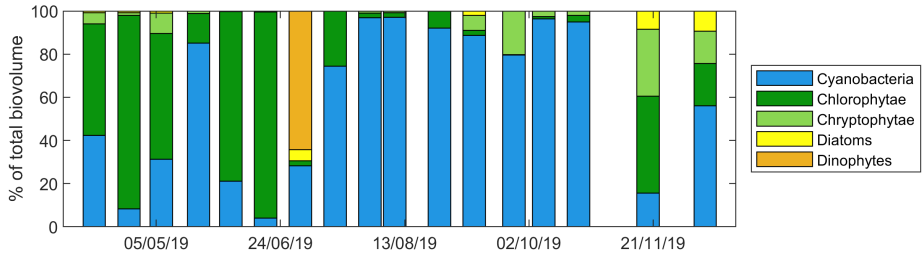


Figure 4.3 – Evolution of the phytoplankton assembly in Lake Champs-sur-Marne for the year 2019, after the identification analysis from water samples carried out by Nicolas Clercin.

during the colder winter months, which are accompanied by an increase in nitrate concentration. With increasing water temperature and day length, a first and strong phytoplankton bloom usually occurs in early Spring. During this bloom the highest phytoplankton concentrations are recorded. The bloom usually spans three or four weeks and consumes the available nitrate. During the following months until the end of the growing season (i.e. around the end of October), the biomass on the study site is therefore controlled by nitrogen and light availability (nitrogen- and light-limitation). Such behaviour can be seen in panels b, c and d of Figure 4.2 for 2019. A similar pattern was also observed in 2016, 2018 and 2020.

Even though this pattern characterizes four of the six years during which Lake Champ-sur-Marne was monitored by LEESU, interannual variability is present. For instance, the data collected in 2021 showed a considerably different pattern compared to the one so-far described. The data for 2021 were not fully analysed and will not be discussed in detail. However, they were characterized by a strong and persistent algal bloom between the months of January and February, that caused a weaker nitrate accumulation. This could have strong effects on the following phytoplankton succession. However, further analysis should be carried out in order to understand the factors that triggered this anticipated winter bloom.

The evolution of the phytoplankton assembly was analysed through laboratory analysis of water samples starting from April 2019 by Nicolas Clercin. Preliminary results of this analysis are shown in Figure 4.3 for the months between April and November of 2019, and show the distribution of biomass among the different phytoplankton classes present on the study site. Five main phytoplankton classes were detected: chlorophytes, cyanobateria, cryptophytes, diatoms and dinophytes. In 2019, cyanobacteria were the dominant group in the months from July to October, while green algae were dominant roughly between April and June. During 2019, the main species detected for each phytoplankton class shown in Fig. 4.3 are: *Sphaerocystis* and *Volvox* (chlorophytes), *Aphanizomenon* (cyanobacteria), *Ceratium* (dinophytes), *Cryptomonas* (cryptophytes) and *Aulacoseira* (diatoms).

Broadly speaking, in Lake Champs-sur-Marne, blooms dominated by cyanobacteria are observed not only during the summer months, but are also frequent during colder months (e.g. March, February, October and November). The main cyanobacteria observed on the study site are *Aphanizomenon* and *Microcystis*.

Chapter 5

Hindcast of the thermal response of a small and shallow lake to climate change

In this chapter the results of the long-term climate change impact study carried out on the study site are presented. The aim of this analysis is to hindcast the long-term dynamics of a small and shallow urban lake, such as Lake Champs-sur-Marne, in order to test the influence of climate change on such ecosystems, in relation to water temperature, stratification dynamics and potential primary production.

To do so, the thermal dynamics of a small urban lake over the past six decades was reconstructed (from 1960 to 2017) through the 3D hydrodynamic model Delft3D-FLOW, forced with meteorological data from the SAFRAN reanalysis.

A series of indices are defined to characterize and analyse the time and spatial evolution of the thermal regime of the study site in terms of stratification dynamics and impact on cyanobacteria growth. The presence of long-term trends and the evolution of spatial heterogeneity of these indices are also assessed.

The results presented in this chapter were published in the scientific journal Earth System Dynamics (Piccioni et al. 2021, DOI: `https://doi.org/10.5194/esd-12-439-2021`).

5.1 Model configuration

The hydrodynamics of the study site were simulated with the FLOW module of the Delft3D modelling suite (Deltares 2014), as described in section 3.2.1.

The bathymetry and the two-dimensional mesh of the domain representing the study site are shown in Fig. 3.4b. The surface of the lake is divided in 255 20 m $\times$ 20 m cells. The Z-model was implemented for the discretization of the vertical axis, with 12 fixed parallel horizontal layers of 30 cm thickness. Turbulent eddy viscosity and diffusivity were computed through the k-ε turbulence closure model. Background values were set to zero [$m^2\ s^{-1}$] for vertical viscosity and diffusivity, while they were set to 0.01 $m^2\ s^{-1}$, after Soulignac et al. (2017) and according to the grid size, for horizontal viscosity and diffusivity. Bottom roughness was computed through Chézy's formulation with the default value for the Chézy coefficient of 65 $m^{1/2}\ s^{-1}$.

The computation of the heat exchange at the air-water interface is done through Murakami's model (Murakami et al. 1985). It requires, as inputs, time series of relative humidity [-], air temperature [°C], net solar radiation [J s^{-1} m^{-2}], wind speed [m s^{-1}] and direction [°N], as well as constant values for sky cloudiness [-] and Secchi depth [m]. Meteorological data used for this study come from the SAFRAN metorological reanalysis and are described in section 3.1.4.

Finally, evaporative mass flux is here neglected and water volume and depth are therefore considered as constant. This assumption makes it possible to analyse exclusively the impact of changes in the climatic forcing.

5.2 Indices for the characterization of the lake thermal regime

The thermal regime of the lake is assessed directly through the analysis of model results in terms of water temperature and through a series of indices that explore the phenology of stratification and highlight the relation between temperature and cyanobacteria production, which are described hereafter. All indices are computed both on an annual and on a seasonal basis, according to the following definitions for the four seasons: (i) January, February and March (winter), (ii) April, May, June (spring), (iii) July, August, September (summer), (iv) October, November, December (autumn).

5.2.1 Stratification indices

In order to thoroughly characterize the phenology of stratification in Lake Champs-sur-Marne, two indices for the stability of the water column have been calculated: the Schmidt stability index and an index based on temperature difference between surface and bottom layers. The Schmidt stability index is a parameter often used in limnological studies to estimate the resistance of a water body to mixing, and therefore its stability. It has been extensively used in scientific literature to describe the strength of stratification in lakes and, more recently, to analyse its evolution over time in relation to climate change (Vinçon-Leite et al. 2014, Niedrist et al. 2018, Kraemer et al. 2015, Livingstone 2003) and algal blooms (Wagner & Adrian 2009). The Schmidt stability index (S) represents the amount of work per unit area that would be required to mix the lake water column at one time instant. It has been here calculated following Idso's formulation (Idso 1973), in which the vertical axis z is considered positive downwards from the surface to the maximum lake depth z_M [m]:

$$S = \frac{g}{A_0} \int_0^{z_M} (z_v - z)(\rho_i - \rho_v) A(z) dz \qquad [J\ m^{-2}] \tag{5.1}$$

where:

$$z_v = \frac{1}{V} \int_0^{z_M} z A(z) dz \tag{5.2}$$

is the depth of the center of volume of the lake, ρ_v [kg m^{-3}] is water density at the depth of the centre of volume z_v, ρ_i is the mean uniform density, g [m s^{-2}] is the acceleration of gravity, V [m^3] and A_0 [m^2] are respectively the volume and the

surface area of the lake, and $A(z)$ is the surface of the horizontal section of the lake at depth z. Computed for each time step, the Schmidt stability can also be averaged over each year or season.

Water resistance to mixing as estimated by the Schmidt stability index is closely correlated to temperature stratification. However, universal thresholds for the onset and breakdown of stratification are difficult to define based on this index and cannot be found in the literature, especially for shallow polymictic lakes. In order to assess the succession of stratification events in a polymictic water body, after Kerimoglu & Rinke (2013) and Magee & Wu (2017), the lake was considered to be stably stratified during a day if the minimum temperature difference between the surface and bottom layers (ΔT) is greater than 1°C. This allows to identify all stably stratified days (SSD) and to compute their total number over a year (annual SSD), or over a season (seasonal SSD), as defined in section 5.2.

5.2.2 Growth rate and growing degree days

Changes in the thermal regime might impact primary production. Here, we make use of two indices as proxies of the potential growth of phytoplankton species: the thermal growth rate (GR) and the growing degree days (GDD). Under the assumption of nutrient and light availability, phytoplankton growth rate can be modelled, for different species, as a function of temperature (T) as in Bernard & Rémond (2012). $\forall\ T \in [T_{min}, T_{max}]$:

$$GR(T) = \frac{k_{opt}(T - T_{max})(T - T_{min})^2}{(T_{opt} - T_{min})[(T_{opt} - T_{min})(T - T_{opt}) - (T_{opt} - T_{max})(T_{opt} + T_{min} - 2T)]} \tag{5.3}$$

where k_{opt} is the optimal growth rate, T_{min} the minimal temperature, T_{opt} the optimal temperature and T_{max} the maximal temperature. The model parameters were calibrated by You et al. (2018) through experimental data to describe the response to water temperature of *Microcystis aeruginosa*, a species of cyanobacteria present in Lake Champs-sur-Marne and often dominant in freshwater bodies globally. The same values are used in this work:

$$k_{opt} = 0.74\ \mathrm{d}^{-1},\ \ T_{min} = 0°C,\ \ T_{opt} = 27.5°C,\ \ T_{max} = 38.4°C. \tag{5.4}$$

Microcystis aeruginosa is thought to be favoured by the warmer water temperatures induced by climate change. However, the curve obtained from eq. (5.3) and (5.4) (shown in figure 5.1), is here more generally intended to be representative of the typical thermal response of cyanobacteria with high optimum temperature. Mean annual and seasonal (according to section 5.2) growth rates are here calculated through eq. (5.3) using simulated surface water temperature, and analysed over time and space respectively in sections 5.4 and 5.5.

The growing degree days are a weather based indicator for biological growth, widely used in the field of agronomy. Based on air temperature, it gives an estimate of the rate of development and of the span of the growing season for terrestrial plants and insects. It is a useful indicator capable to link global warming and biology (Grigorieva et al. 2010, Schlenker et al. 2007). Approaches based on GDD have

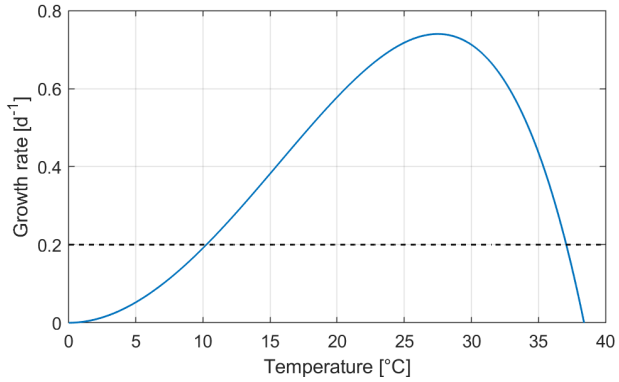


Figure 5.1 – Thermal growth rate calculated after equation (5.3). The horizontal dashed line for GR=0.2 d^{-1} meets the curve at the temperature limits for the calculation of the GDD (10°C and 37°C, respectively).

been increasingly applied to phytoplankton communities and fisheries (e.g. Gillooly 2000, Neuheimer & Taggart 2007, Ralston et al. 2014, Dupuis & Hann 2009), in order to correlate water temperature and phytoplankton growth while taking into account interannual variability. After Dupuis & Hann (2009), GDD were calculated as follows:

$$GDD(t) = \sum_{i=t_0}^{t} a_i \cdot (T_i - T_{base}) \cdot \Delta t, \quad \text{with } a_i = \begin{cases} 1 & \text{if } T_{base} < T_i < T_{sup} \\ 0 & \text{elsewhere} \end{cases} \tag{5.5}$$

where t is the time (here in days) with t_0 the reference day to start the calculation, Δt is the time step (equal to 1 day), T_i is the daily average of the modelled surface water temperature of day i and T_{base} (respectively T_{sup}) is a physiological threshold below which (respectively above which) growth does not occur. Compared to the formulation found in Dupuis & Hann (2009), an upper limit for growth was introduced here (T_{sup}) to take into account high temperature stress. Our focus here is, as for the GR, on cyanobacteria. After Thomas et al. (2016) and based on the latitude of the study site, we set the base temperature at 10°C and the upper limit for growth at 37°C. This results in considering, for the calculation of the GDD, only temperatures that yield to a GR above 0.2 d^{-1} (see figure 5.1).

GDD can be calculated on an annual or a seasonal basis by adjusting the values of t_0 and t. Annual GDD are calculated from the first of January until the 31st of December. Seasonal GDD are obtained according to the definitions of section 5.2.

5.3 Model calibration and validation

Delft3D-FLOW stands on a robust mathematical and physical structure and only few parameters have to be calibrated. In this work, only those directly involved in the heat-flux model and in the wind module were calibrated: the Secchi depth [m], the mean cloud cover [-] and the wind drag coefficient [-]. In order to get a first estimate for the sky cloudiness parameter, cloud cover data from the MétéoFrance station in Trappes (ID: 78621001) were averaged over the calibration period. The

wind drag coefficient was calibrated in order to take into account the presence of tall trees surrounding the contour of the lake, locally reducing wind speed.

The calibration of the FLOW model was done by trial and error, based on high-frequency water temperature data at the surface, middle and bottom layers (0.5, 1.5 and 2.5 m depth, respectively). Model results were compared to water temperature at the three layers and two different locations (A and B). The root mean square error (RMSE) was calculated to evaluate model performances. For this purpose, high-frequency data were first averaged every hour to match the model output time step and cleaned from the outliers originated by periodic sensor maintenance. The latter were defined as sudden water temperature variations ($> 1°C$) over the 10 minutes separating two successive measurements, and consequently erased.

The model was calibrated on the year 2016 and validated on two other periods: from May to December 2015, and during the whole year 2017. Field values for the Secchi depth in Lake Champs-sur-Marne vary between 0.5 and 3 m; using this range, the Secchi depth parameter was calibrated and finally set to 1.2 m. Based on meteorological data downloaded from the Orly meteorological station, a mean value for sky cloudiness was set to 80%, and a uniform wind drag coefficient was set to 0.005 [-].

Model performance during calibration and validation is shown in figure 5.2 relatively to site A. Parity diagrams between observed and simulated water temperature are plotted for the surface, middle and bottom layers (see panels a, b and c, respectively) and show an excellent agreement between observations and model results. A slight underestimation of surface water temperature can be noticed for the surface layer during the colder winter months, as well as a slight overestimation of the highest values of water temperature by the model, especially for the middle and surface layers (see also Fig. 5.2d). However, overall model performances are satisfactory for all three layers, with RMSE values between simulated and observed water temperature of 0.85°C, 0.78°C and 0.81°C at site A during calibration, respectively for the surface (0.5 m), middle (1.5 m) and bottom (2.5 m) layers. Model results are spatially robust and satisfactory also for the validation periods, with similar RMSE values for sites A (surface:1.0°C, middle:0.96°C and bottom:0.96°C) and B (surface:1.0°C, middle:0.96°C, bottom:0.99°C).

Furthermore, the observed (blue) and simulated (orange) temperature difference between the surface and bottom layers is plotted in figure 5.2e, with a dashed lined representing the 1°C threshold for the definition of the SSD. Panels f and g of figure 5.2 show the succession of stable stratification events as defined in section 5.2.1 calculated through observations and model simulations, respectively. Some discrepancy is present, notably in spring 2016, which can be explained by a slight overestimation of surface temperature, combined with the threshold effect of the definition of SSD. However, the model correctly captures the succession of stable stratification events both in terms of frequency and timing over the considered three-years period.

Overall, the model results fit very well the high-frequency water temperature data, and accurately reproduce the water temperature dynamics, including the diurnal cycle, as well as the stratification regime.

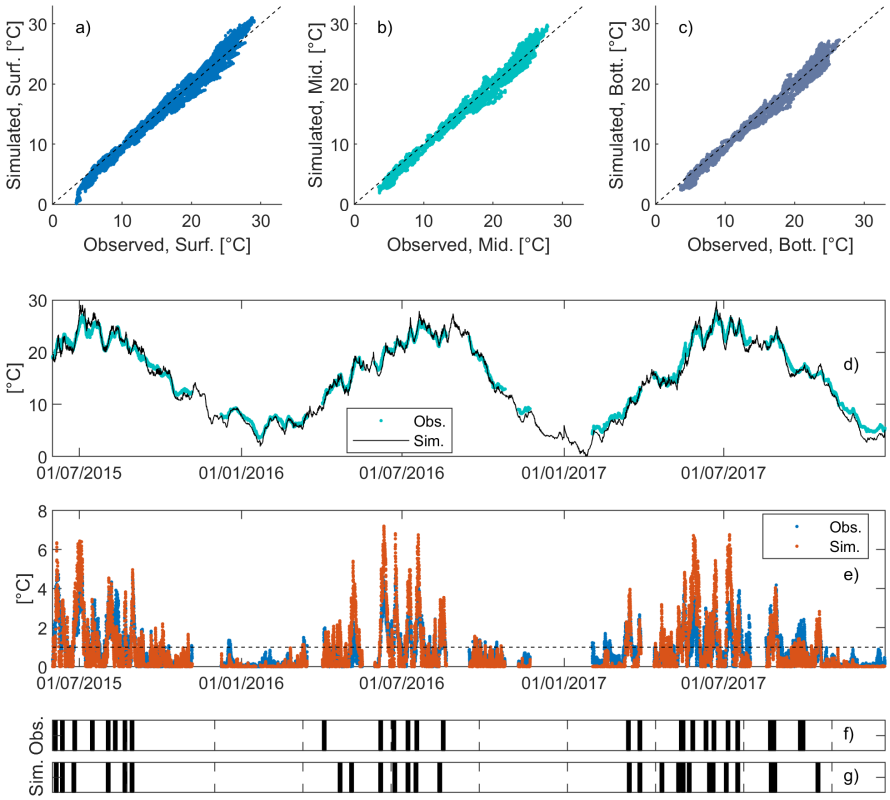


Figure 5.2 – Model performance during validation at site A. Panels a, b and c: parity diagrams between simulations and observations for the surface, middle and bottom layers, respectively. Panel d: visual comparison of simulated and observed water temperature at the middle layer. Panel e: modelled (orange) vs. observed (blue) temperature difference between surface and bottom layer and relative comparison between the timing of observed and modelled stable stratification events (panels f and g, respectively).

5.4 Long-term trend analysis

In the present study we hindcast the long-term dynamics of a small and shallow urban lake between 1960 and 2017, in order to test the influence of climate change on such ecosystems.

5.4.1 Approach

The long-term hydrodynamic simulation starts on the first of January 1960. No data were available to set the initial conditions of the model, neither in terms of water temperature, nor in terms of current velocities. However, the model is strongly driven by the meteorological data and the influence of the initial condition vanishes after only a few days (Piccolroaz et al. 2019). Indeed, small perturbations in water temperature initial conditions (± 2°C) were tested and resulted to vanish in 5 to 7 days. The model was therefore initialized with water at rest and with a uniform water temperature of 7°C, the average of the water temperature recorded on the lake on the first of January in 2016, 2017, 2018 and 2019. Model results are stored at a hourly time step on every element of the mesh.

Model results at site A are analysed on an annual and seasonal basis for long-term trends, in terms of water temperature (averaged over the water column) and through the indices defined in section 5.2. The presence of long-term trends is tested (with a threshold for significance $\alpha = 0.05$) through the Mann-Kendall test (Mann 1945, Kendall 1975), a non-parametric test for the individuation of overall monotonic trends performed here through the MATLAB software (Burkey 2020). The Mann-Kendall test is often preferred to simple linear regression in the analysis of meteorological and hydrological time series, as it does not require any assumption on the distribution of the analysed dataset (Tímea et al. 2017, Wang et al. 2020). Once a trend is detected, its strength is evaluated through the Sen's slope estimator, that uses a linear model to evaluate the intensity of the trend (Sen 1968).

Meteorological forcing is crucial for this work, as it drives the hydrodynamic model and represents the only source of variability in our modelling configuration. The presence of long-term trends in the meteorological dataset was also evaluated by applying the Mann-Kendall test and the Sen's slope estimator to their annual averages.

5.4.2 Trends on meteorological input data

Annual averages of the SAFRAN dataset used as input to the Delft3D model were calculated from 1960 to 2017 and tested with the Mann-Kendall test. Strongly significant monotonic trends ($p \ll 0.05$) were found for the air temperature, solar radiation and wind speed, as shown in Fig. 5.3. The Sen's slope estimator was used to test the intensity of the significant monotonic trends. Air temperature displays a considerable warming trend of 0.3°C dec^{-1}; solar radiation also shows a significant increasing trend, with an overall intensity of 3.5 W m^{-2} dec^{-1}. Wind speed decreases quite sharply over time, at an overall estimated rate of 0.2 m s^{-1} dec^{-1}. While the increase in air temperature appears extremely linear (see Fig. 5.3a), a sharp shift in the behaviour of both solar radiation and wind speed appears around the year 1988 (Fig. 5.3b and c, respectively). A change-point detection was therefore performed on

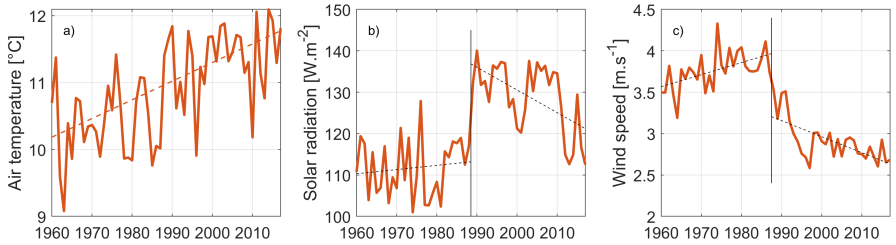


Figure 5.3 – Annual averages of the three meteorological variables which exhibit significant monotonic trends, that is a) air temperature, b) solar radiation, c) wind speed. The relative overall trend intensity has been evaluated through Sen's slope estimator for air temperature (orange dashed line, panel a) whereas a piecewise trend has been calculated after change-point detection for solar radiation and wind speed (black dashed lines, panels b and c).

the latter two series, and showed for both variables the existence of two significant sub-trends separated by a drastic shift towards the end of the 1980s. Both variables are characterized by a mild increase until 1987 (1988 for solar radiation), followed by a considerable decrease until the end of the available series. However, despite this piecewise linear behaviour, the presence of overall monotonic increasing (for solar radiation) or decreasing (for wind speed) trends is confirmed by the very low p-values obtained for these variables through the Mann-Kendall test.

Finally, no significant trend was found for relative humidity and wind direction. The two variables appear to be stationary, the former fluctuating around an annual average of roughly 80% and the latter around an annual prevailing wind direction of 200°N (South-West). Three of the five meteorological variables forcing the hydrodynamic model were therefore characterized by strongly significant monotonic trends along the past six decades, confirming changes in the climate of the region around the study site.

5.4.3 Trends on simulated indices

Long-term monotonic trends have been researched at site A on an annual and seasonal basis for: mean water temperature (vertically averaged), number of stably stratified days (SSD), mean Schmidt stability index, mean growth rate (GR) and growing degree days (GDD). Figure 5.4 shows all the significant monotonic trends found from this analysis. On an annual basis, the Mann-Kendall test highlighted the presence of strongly significant increasing trends ($p \ll 0.05$) for all variables.

Mean annual water temperature shows a very sharp warming tendency of 0.6°C dec^{-1} (see Fig. 5.4a), even greater than what was found for air temperature (0.3°C dec^{-1}). The Pearson correlation coefficient (r) was calculated between water temperature and the five meteorological input variables in terms of annual averages in order to explain this behaviour. Modelled water temperature is strongly correlated with air temperature, solar radiation and wind speed, with correlation coefficients of 0.8 for solar radiation and air temperature and -0.9 for wind speed. Water temperature shows significant increase during all seasons, with higher slopes during spring and summer (0.8 and 0.7°C dec^{-1}, respectively), and a lower yet considerable intensity during autumn and winter (respectively 0.4 and 0.5°C dec^{-1}).

The warming trend is accompanied by reinforced stratification. An increase in water column stability is highlighted on an annual basis by both stratification indices: the annual number of SSD increased on average of around two days per decade, while the Schmidt stability index increased of 0.9 J m^{-2} dec^{-1} (Fig. 5.4b and c, respectively). Despite a warming trend being present in all seasons, both stratification related indices show significant increasing trends only during winter (1 d dec^{-1} and 0.4 J m^{-2} dec^{-1}) and spring (sharper trends of 1.8 d dec^{-1} and 2.6 J m^{-2} dec^{-1}, for the seasonal SSD and the Schmidt index, respectively). Furthermore, the number of stable stratification events (i.e. the count of the slots of consecutive SSD during a year) was calculated to characterize the frequency of stable stratification. It did not show significant trends over time, varying between a minimum value of 8 to a maximum of 16 around an overall average of 12 stable stratification events. Similarly, the duration of the longest stable stratification event (i.e. the longest slot of consecutive SSD in a year) did not show significant trends, but a high interannual variability. It varies around an average value of 11 d, between a minimum value of 5 d and a maximum of 15 d.

The analysis of the growing degree days and of the mean growth rate shows the progressive improvement of conditions for cyanobacteria. The pattern of the mean annual GR is highly correlated to that of water temperature and shows a significant trend of 0.02 d^{-1} dec^{-1} (black line in Fig. 5.4d). However, the stronger intensity of the trend for the GR during spring (0.03 d^{-1} dec^{-1}) indicates an amplified effect of water temperature on the potential growth of cyanobacteria during this season. Annual GDD (see figure 5.4e) shows a considerable increasing rate of 157°C d dec^{-1}, with a strong shift around the year 1989. This behaviour cannot be regarded as linear and is highly influenced by the piece-wise behaviour of mean annual solar radiation and wind speed. However, it corroborates the idea of a greater amount of thermal energy reaching the ecosystem, at different rates but consistently throughout the four seasons.

The changes in the meteorological forcing clearly had an impact on the dynamics of the study site. The lake has sensibly warmed, its tendency to thermal stratification has increased, and the thermal conditions for cyanobacterial growth have improved. Spring shows the sharpest trends for all indices, and might ultimately be the season suffering the strongest changes in terms of primary production and algal blooms.

5.5 Spatial analysis

5.5.1 Approach

The long-term evolution and the spatial variability of the thermal regime of Lake Champs-sur-Marne was further analysed exploiting the three-dimensional model simulations. Mean annual surface water temperature, annual SSD, mean annual GR and annual GDD were computed on the whole computational domain, with the objective of investigating the relation between climate change and time evolution of the spatial distribution of these variables. For each variable x, the overall mean annual value x_m (averaged over the complete domain) and the deviation from the mean value $(x - x_m)$ have then been computed. In order to quantify the spatial heterogeneity of these variables, the probability distribution of the deviation from

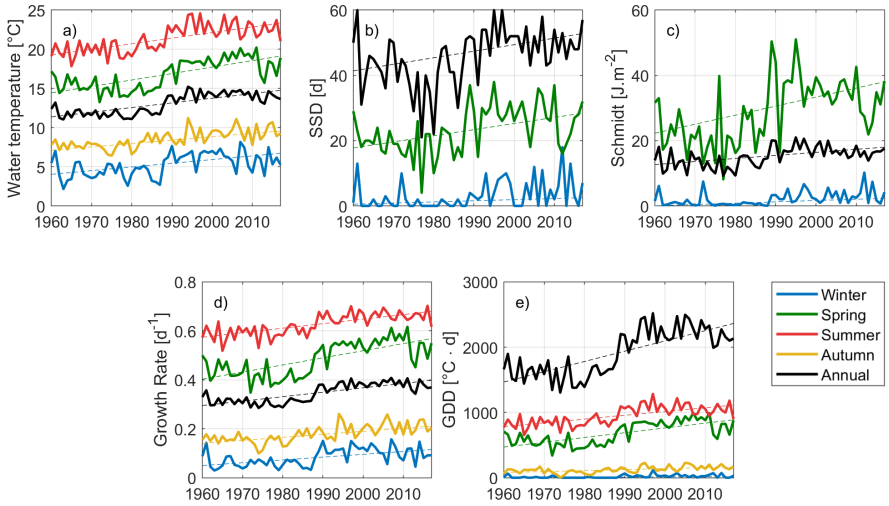


Figure 5.4 – Statistically significant climate change trends at monitoring site A for the five indices, both on an annual (black) and seasonal (other colours) basis. a) Water temperature (averaged on the water column); b) Number of stably stratified days (SSD); c) Schmidt stability. d) Growth rate; e) Growing degree days (GDD). Blue lines represent the winter season, green lines represent spring, red lines are for summer trends and yellow lines for autumn; black lines represent annual values.

the mean value of each variable was finally calculated on the computational domain and fitted, for each year, with a non-parametric Kernel probability distribution through the Matlab *pdf* function. The resulting probability density function (PDF) was plotted over time as a heat map and the mean value as a simple line plot. This allows to visualize both the time and the spatial evolution of the variable under consideration, by looking at the mean value and at the range of values characterized by a non-zero probability.

During stably stratified periods, cyanobacteria are favoured over other algal groups because of their ability to move within the water column and possibly float towards the water surface (Humphries & Lyne 1988, Wagner & Adrian 2009, You et al. 2018). For this reason, the spatial analysis of the GR and GDD was completed, by calculating these two indices only on stable stratified days during each year. The obtained GR were further averaged for each cell over the local number of stably stratified days. Cells that showed an annual number of SSD<10 where discarded from this analysis. Finally, the resulting modified indices were analysed over space and time as described above by using a non-parametric Kernel probability distribution as an approximation of the PDF for each simulated year.

5.5.2 Analysis of model results

Lakes are not spatially homogeneous systems. Heterogeneity can be generated by the interplay between bathymetric and morphological features, or by particular meteorological conditions, especially in terms of wind direction.

In order to quantify the level of spatial variability in the lake, the deviations

between local annual values (calculated for each computational cell) and their overall annual mean value (calculated on the complete domain) were calculated and fitted with a probability density function (PDF). As shown in figure 5.5a (top panel), mean annual surface water temperature is rather uniform over the study site. The difference between the maximum and minimum values if of roughly 0.1°C (around 1% of variability relative to the overall mean) and does not vary substantially over time. During the first half of the simulation period, and in particular between the years 1967 and 1987, the width for the PDFs (i.e. the domain on which PDFs are greater than 0) is narrower, reflecting a higher annual spatial uniformity than what can be observed after 1990. After 1990, the width of PDFs is indeed wider, with only a few exceptions where the PDFs are on the contrary quite sharp. This change in the spatial distribution of annual surface water temperature before and after 1990 is accompanied by a sharp increase in the overall mean value (bottom panel in Fig. 5.5a), which is indeed greater (around 14.5°C) after 1990 than before (around 12°C).

The annual number of SSD shows greater spatial heterogeneity (see Fig. 5.5b). The difference between the maximum and minimum values of SSD varies between approximately 45 and 90 days. The spatial heterogeneity is mainly induced by bathymetry. Stable stratification only occurs in the deeper portion of the basin, while the northern part of the study site, namely the portion with depth lower than 1.8 m (see Fig. 3.1b), remains constantly mixed according to our definition of SSD. The PDF is dissymmetric, with the most probable value for the annual SSD higher than the overall annual mean, by 10 to 15 days. As for the surface water temperature, the spatial heterogeneity of SSD is higher after 1990 than before. In fact, a rather high correlation is present between the spatial distribution of mean annual surface water temperature and SSD. The correlation coefficient between the two variables in each simulated year varies between 0.4 and 0.8, with an overall mean of 0.62 and p-values always lower than the threshold for significance (p=0.05). This suggests that surface water temperature tends to be slightly warmer in areas characterized by longer periods of stable stratification.

The thermal growth rate and the GDD were analysed over the domain during stably stratified days, which are particularly favourable to the growth of cyanobacteria.

The thermal GR shows a low spatial heterogeneity that does not vary over time, as confirmed both by the PDF in figure 5.6a (top panel) and by the maps in figure 5.6b. The difference between minimum and maximum values for the GR is around 0.03 d^{-1} (around 5% of the overall mean value), always rather centred around the overall mean. Calculated during stratification, the overall mean thermal GR takes high values (around 0.6 d^{-1}), comparable to those found at site A for the summer season (see the bottom panel of fig. 5.6a and fig. 5.4d).

The overall mean annual GDD increases over time (bottom panel of Figure 5.6c), from around 400 d·°C before 1980 to 650 d·°C after. The PDF of the GDD displays a clear increase in spatial heterogeneity (Fig. 5.6c). Its range increases substantially starting from the 1980s, roughly doubling: from 100 d·°C before 1980 to around 200 d·°C afterwards. This is due to the concurring effects of warmer water temperature and higher number of stably stratified days in the calculation of the GDD as defined in section 5.2.2. In particular, part of the heterogeneity is induced by shallow areas of the water body that only account for a low number of SSD

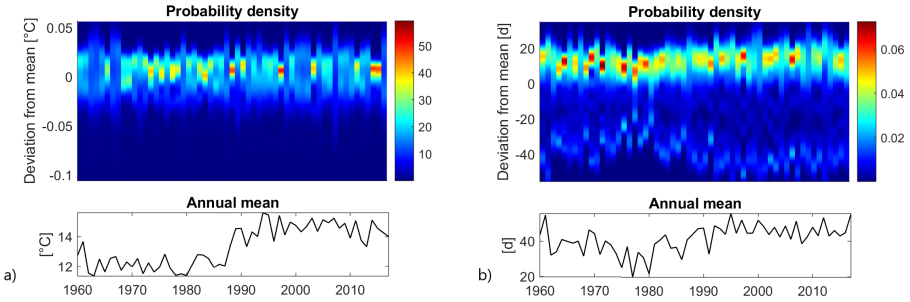


Figure 5.5 – Top panels: Time evolution of the probability density function of the anomalies (i.e. the spatial deviations of a variable to its annual mean over the lake). Bottom panels: Time evolution of the annual mean calculated over the lake. a) Mean annual surface water temperature; b) annual SSD.

and therefore for low values of GDD. The corresponding computational cells, not affected by stable stratification during the 1960s, are evermore likely to show stable stratification in the 2000s (see the maps in fig. 5.6d). However, the maps for the years 2017 and 2005 also show a high heterogeneity in the deeper part of the water body.

5.6 Discussion

In the present chapter, the thermal regime of a shallow urban lake was reconstituted over six decades (between 1960 and 2017) with a 3D thermal-hydrodynamic model. Simulation results were analysed over time (for long-term monotonic trends), and space (for spatial heterogeneity), through a series of indices that characterize the stratification and highlight the relation between temperature and cyanobacteria production.

Meteorological forcing data

The model was forced with data from the SAFRAN meteorological reanalysis. Air temperature and solar radiation showed increasing monotonic trends (0.3°C dec^{-1} and 3.5 W m^{-2} dec^{-1}, respectively), while wind speed showed a decreasing monotonic trend of -0.2 m s^{-1} dec^{-1}. A shift was observed during the studied period around 1987, especially in the data series of solar radiation and wind speed, and was confirmed by a change-point detection analysis. The existence of such shift in global climate during the 1980s has been highlighted by a number of studies using different data sources (Reid et al. 2016, Mariani et al. 2012, Gallagher et al. 2013).

Climate change in the Paris region has been assessed in literature mainly in terms of air temperature (Perrier et al. 2005, Lemonsu et al. 2013). Compared to our result, a milder increasing trend of 0.1°C dec^{-1} was found based on ground measurements, between 1900 to 1987, with a steeper increment of 0.7°C dec^{-1} later on until 2005 (Perrier et al. 2005). Similarly, we also find a steeper trend of 0.55 °C dec^{-1} on the years from 1987 to 2005. Less information can be found in literature for solar radiation and wind speed. A decrease in wind speed on land was found over Europe since 1980 (around -0.1 m s^{-1} dec^{-1}) as part of a large-scale

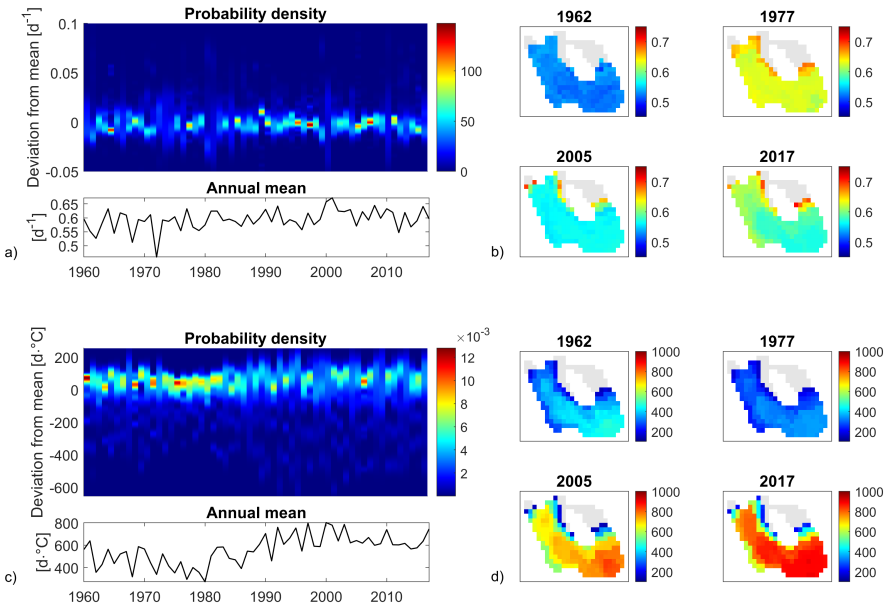


Figure 5.6 – Spatial analysis of stratification. a) Probability density function (PDF) for mean GR during stratification over the computational domain and over time; b) Four examples of spatial distribution for mean GR during stratification over the lake; c) PDF for GDD during stratification over the computational domain and over the years; d) Four examples of spatial distribution for annual GDD during stratification over the lake. Grey cells in panels b and d do not stratify longer than 10 days over a year.

analysis of observations in the northern hemisphere (Vautard et al. 2010). At a global scale, an overall decreasing trend in wind speed was found over land in the period 1985-2015 through meteorological reanalysis, principally over Europe, India and western Africa (Torralba et al. 2017). In South-East China, in the Lake Chaohu region, a strong decline in wind speed (China Meteorological station) was also found in the period 1980-2016 (Zhang et al. 2020). An overall increase in surface solar radiation was recently found for Europe between 1983 and 2015, specifically of 3 W m^{-2} dec^{-1} for western Europe (Pfeifroth et al. 2018).

Meteorological reanalyses usually cover multi-decadal periods and have the great benefit of being spatialized over vast portions of the globe. Even though their use in limnological studies is quite recent, they have already been used to simulate water temperature (Layden et al. 2016, Piccolroaz et al. 2020), stratification dynamics (Frassl et al. 2018) and phytoplankton distribution (Soulignac et al. 2018). As shown in this work, their use as external forcing to thermal-hydrodynamic models can yield, provided that observations are available for calibration and validation, to accurate simulations of the behaviour of water bodies even in the absence of local meteorological observations. This could open to a great range of applications in limnology and paleolimnology (Jenny et al. 2016, Maier et al. 2019). The proposed methodology allows to thoroughly reconstruct the behaviour of any water body both in time and space, independently of its proximity to meteorological stations. This is particularly interesting for small or remote water bodies that often lack long-term measurements.

Water temperature and stratification

Based on the 3D model results found for Lake Champs-sur-Marne, long-term trends were analysed in detail at site A. Significant increasing trends were detected for water temperature both on an annual and seasonal basis. The highest seasonal warming was found during spring and summer (0.8 and 0.7°C dec^{-1} respectively). These trends are particularly intense and could have strong impacts on the ecosystem under examination. In particular, the intensity of these trends is greater than that suggested for summer water temperature in a large-scale analysis (0.53°C dec^{-1}) for lakes with similar changes in the meteorological forcing (O'Reilly et al. 2015). Furthermore, mean annual depth-averaged water temperature also increased at a considerable rate of 0.6°C dec^{-1}, greater than the rate found for air temperature, a behaviour also highlighted for other water bodies (Austin & Colman 2007, Schneider et al. 2009). The piecewise linear behaviour of mean annual water temperature, is induced by that of solar radiation and wind speed. In fact, similarly to what was found by Magee & Wu (2017), mean annual water temperature was highly correlated (i.e. $|r| > 0.8$) with air temperature, solar radiation and wind speed. This suggests that meteorological variables have additive effects that concur to enhance the response of the dependent variables. These effects might be particularly intense for wind over small and shallow lakes, due to their low volume to surface ratio.

Both stratification related indices (SSD and Schmidt stability) showed a significant mean annual increasing trend (3 d dec^{-1} and 1 J m^{-2} dec^{-1}, respectively). Similar values were recently found for shallow water bodies in other long-term studies (Magee & Wu 2017, Moras et al. 2019). However, despite a strong augmentation in water temperature, stratification did not show a significant increase during summer.

In shallow polymictic lakes the water column is mixed frequently also during the warmer seasons. Summer surface and bottom water temperature increased at a very similar rate (0.7°C dec^{-1}) in the study site, resulting in small changes in Schmidt stability and number of SSD. This result marks a strong difference with the behaviour of deeper monomictic or dimictic lakes, where the summer Schmidt stability often shows an increasing trend (e.g. Niedrist et al. 2018, Flaim et al. 2016), but it is not uncommon for shallow water bodies, where Schmidt stability can even show decreasing summer trends (Fu et al. 2020).

Stratification induces a separation between the sediment and the surface layers, influencing the distribution of nutrients and biomass over the water column. During stratification, due to the desoxygenation of the lake bottom layers, nutrients (phosphate in particular) are released from the sediment. In polymictic water bodies, when mixing occurs, a replenishment of the whole water column with the nutrients released during previous stratification has been observed (Song et al. 2013, Wilhelm & Adrian 2008). In Lake Champs-sur-Marne, neither the frequency nor the duration of the stable stratification events show a significant trend during the past decades. However, with a mean value of 12 annual separated stable stratification events, lasting up to two consecutive weeks, the replenishment of the water column with nutrients is ensured. The multiple pulses associated with the alternation between mixing and stratification events are an important internal source of nutrients, especially in a lake such as the study site, whose water inflow is limited to underground waters.

The thermal regime was further characterized over the computational domain by analysing the spatial distribution of surface water temperature. While annual averages of surface water temperature are rather uniform over the domain, with around 0.1°C of difference between maximal and minimal values, the bathymetric variations induced greater variability in the distribution of SSD. The stratification regime drastically changes between the deeper portion of the water body and the shallower northern part. According to the definition of the SSD, stable stratification never occurs in cells with water depth lower than 1.8 m. In shallow water bodies, even small bathymetric variations can cause drastic differences in the thermal regime. Different regimes of mixing and stratification between shallower and deeper areas can result in considerable differences in the spatial distribution of nutrients, with effects on bloom initiation and phytoplankton growth, as well as on the resulting oxygen concentration. However, spatial heterogeneity of the mixing and stratification regime inside a water body is rarely addressed in scientific literature, especially with regard to small and shallow lakes (e.g. Bachmann et al. 2000).

Indices for primary production

The thermal regime is a key factor in the regulation of the biogeochemical cycle and in the development of algal blooms. The worldwide intensification of harmful algal blooms over the past decades (Paerl & Huisman 2008, Paerl & Paul 2012, Wagner & Erickson 2017) is often associated with climate change and nutrient enrichment (Zou et al. 2020, Huisman et al. 2018).

Due to their potential toxicity, cyanobacteria are of particular concern in freshwater management. Warmer water temperature can favour their growth because of their high optimal temperatures. However, they can proliferate under a wide range of temperatures (Lürling et al. 2013, Carey et al. 2012). The expression of the

growth rate proposed by Bernard & Rémond (2012) (see eq. 5.3) accounts for this dependence from water temperature. Based on this expression, the mean annual thermal growth rate of cyanobacteria showed a significant increasing monotonic trend of 0.02 d^{-1} dec^{-1}. Compared to the initial annual value of roughly 0.3 d^{-1} at the beginning of the 1960s, this results in a considerable +40% increase in the growth rate at the end of the sutdied period. Significant trends were also found during the four seasons, the highest being during spring (0.03 d^{-1} dec^{-1}, or +45% of the initial value). The growing degree days (GDD) of cyanobacteria were analysed here for a range of temperatures comprised between 10°C and 37°C, corresponding to thermal growth rates higher than 0.2 d^{-1}. However, given the temperate climate of the region under examination, the upper limit for growth did not have any effect on the results, whilst it could be an important parameter for species with lower optimum temperatures such as diatoms.

Whereas the growth rate gives an estimation of the mean value of cyanobacteria growth, that can be computed on a seasonal and an annual basis, the GDD is a cumulative index that gives a measure of the amount of time and degrees available during a year for photosynthetic growth. Originating from the field of agronomy and forestry, it represents a "thermal time" and is considered as a better descriptor of vegetal phenology than the simple Julian days (McMaster & Wilhelm 1997). Under an appropriate temperature range, it can be considered as representative for organism developmental time (Dupuis & Hann 2009). The highest trend for GDD was found on an annual basis (157 d·°C dec^{-1}), denoting that the temperatures favourable to cyanobacteria growth are more and more frequently reached. Seasonal trends varied greatly in intensity. The highest was found for spring (73 d·°C dec^{-1}) and represents, relative to the values in the early 1960s, a substantial increase of 90% during the six decades under consideration. The trends found for winter and autumn are mild but denote an increased tendency to overpass the base temperature during these two seasons, and therefore a dilatation of the season favourable to cyanobacteria growth.

Harmful algal blooms and phytoplankton dynamics depend on factors such as the settling or buoyancy rate of phytoplankton, the availability of nutrients over the water column, which can be enhanced by the release from the sediment, and the resuspension of particulate organic matter. In polymictic water bodies, the processes of sedimentation and resuspension are strongly influenced by the alternation between mixing and stratification (Song et al. 2013). Because of their ability to migrate within the water column, stratified environments are favourable to cyanobacteria (e.g. Carey et al. 2012, Aparicio Medrano, Uittenbogaard, Van de Wiel, Dionisio Pires & Clercx 2016). The increase of water temperature and of stable stratification could concur resulting in frequent cyanobacteria blooms. However, stratified conditions do not occur uniformly. The calculation of the thermal GR and of the GDD quantifies the potential effect of water temperature on cyanobacteria growth, under the hypothesis of nutrient and light availability. Their calculation during stratification allows to address the combined effect of water temperature during a particularly favourable environmental conditions.

During stratification cyanobacteria GR was characterized by high values (around 0.6 d^{-1}), with a variability quite uniform over time of ±5% over the study site. These values are comparable, or even higher (until the 1990s) than those obtained

during the summer season. The GDD give a deeper insight on the interplay between temperature and stratification. The strong augmentation in the overall mean value of GDD during stratification confirms a concurring positive effect of the increase of water temperature and of the duration of stable stratification on the growth of cyanobacteria. Moreover, the greater spatial variability of GDD values during the second half of the simulation indicates that some parts of the lake will be more affected than others by the variation of water temperature and stratification. In particular we observe the development over time of certain areas in the study site, especially the deeper part, with very high values of GDD under stratified conditions, and that are therefore particularly favourable to cyanobacteria dominance and bloom initiation.

The combination of increasing trends for water temperature, stable stratification and the widening of the growing season can favour the occurrence of harmful cyanobacterial blooms (Winder & Sommer 2012, Jones & Brett 2014, Noble & Hassall 2015). If these trends are confirmed, during the decades to come cyanobacteria could become the dominant species in the study site, seriously affecting the lake ecological network and its biodiversity (Rasconi et al. 2017, Toporowska & Pawlik-Skowronska 2014).

Model-based approach

Through our modelling approach it was possible to reconstruct the thermal dynamics of a small and shallow lake and to thoroughly analyse its evolution over time and space. The use of an extensive data set of high-frequency observations allowed to test the model not only against the general seasonal water temperature pattern, but also against daily and sub-daily dynamics of stratification and mixing, at two locations. Other works have focused on the hindcast of lakes thermal regime, successfully reconstructing their dynamics in order to analyse their evolution over time (e.g. Magee & Wu 2017, Moras et al. 2019, Zhang et al. 2020, Stetler et al. 2020). Most of these studies, however, make use of a 1D approach. By means of a 3D model it is possible to aggregate information on both time and space (horizontal and vertical) through the use of appropriate indices. Our work demonstrates that even on a small water body spatial variations can be important, and that their influence on the thermal and biological regime must be considered. It provides additional evidence that supports the hypothesis of a positive effect of climate change over cyanobacteria blooms.

Hydro- and thermal dynamics are at the core of the biogeochemical cycles, influencing transport, sediment resuspension, organic matter mineralization in addition to primary production. In this work, we focus on water temperature, quantifying its impact for stratification and primary production. The proposed methodology allows to focus solely on the role of the meteorological forcing, addressing its direct impact on the thermal regime and on primary production. However, other factors could have even a stronger impact: nutrient and light limitation or grazing could offset temperature-derived advantages (Elliott et al. 2006). These factors are not taken into account in this work, since it is focused on the impact of climate change from a thermal standpoint, all other factors being equal. This work opens to a wide range of additional analysis and further research. In particular, the coupling with a biogeochemical model could give further insight on the impact of climate change

on the ecological state of a water body. Such a study, however, would introduce additional sources of uncertainties, especially regarding the evolution of nutrient sources over time and could only be profitably if performed after a thorough analysis of the hydrodynamic and thermal regime.

5.7 Conclusion

In this work, the long-term thermal regime of a shallow urban lake is reconstructed through model simulations from 1960 to 2017. A series of indices are proposed with the objective of thoroughly describing the thermal regime of shallow water bodies, in relation with stratification dynamics and cyanobacterial production. The meteorological data set is derived from the SAFRAN reanalysis and shows a significant increase in air temperature and solar radiation and a significant decrease in wind speed, with a regime shift in the late 1980s. Simulation results show that small urban lakes react rapidly and strongly to external meteorological conditions, with only limited resilience to climatic shifts. The additive effect of increasing solar radiation and air temperature and decreasing wind speed acts on different terms of the heat budget at the lake surface, enhancing the changes found in the lake. The mean water warming of 0.6°C dec^{-1} represents an increase of 32% in water temperature values between 1960 and 2017 and is much stronger than the air warming (0.3°C dec^{-1}, i.e. an increase of 18% during the same period). The impact on stratification and cyanobacteria production is even more alarming, with an increase of over 30% of the stability indices and over 60% of the growing degree days during the six past decades. Spring shows the sharpest trends in terms of water temperature, water column stability (Schmidt and SSD) and growing degree days, and might ultimately be the season suffering the strongest changes in terms of primary production and algal blooms. The spatial heterogeneity found for thermal stratification and growing degree days might also concur to locally create conditions particularly favourable for cyanobacteria blooms. These tendencies could favour early phytoplankton blooms (during late winter or spring) and contribute to the proliferation of cyanobacteria, and ultimately to the degradation of the whole aquatic ecosystem. Our results highlight the importance of a three-dimensional approach to thoroughly infer the dynamics of a water body. Horizontal patterns can be particularly strong for shallow lakes due to the relative importance of bathymetric variations.

Small and shallow lakes are extremely widespread ecosystems. Our results suggest that such systems experience considerable thermal stress caused by climate change and that, in nutrient-enriched systems, cyanobacteria dominance could become a widespread issue in the future decades.

Chapter 6

Automated calibration of a complex ecological model

In this chapter, an innovative method for automated parameter inference is proposed and applied on the complex biogeochemical model Delft3D-BLOOM. This method (called ABC-RF with SA in the sequel) is based on the Approximate Bayesian Computation (ABC) and combines the machine learning technique called "Random Forest" (RF) with a sensitivity analysis (SA) of the model parameters. Three target variables are considered in this calibration procedure: total chlorophyll, phycocyanin and dissolved oxygen, and 133 parameters are calibrated. The method ABC-RF with SA is first validated through an application on simulated observations, in order to test its capability to reproduce both the simulated observations and the parameters values. Then it is applied on a real observation dataset of the lake Champs-sur-Marne. The standard ABC method and the ABC-Random Forest (ABC-RF) method are also applied for comparison.

6.1 Model configuration

A 3D hydrodynamic model was set up for Lake Champs-sur-Marne with the Deltf3D-FLOW module. The bathymetry was interpolated from *in situ* measurements. The horizontal mesh is composed of 813 10 m $\times$ 10 m square cells. Twelve horizontal layers with a fixed thickness of 27 cm were used for the discretization of the vertical axis. The k-ε turbulence closure model was used for the computation of turbulent eddy viscosity and diffusivity. Background values for horizontal viscosity and diffusivity were set to 0.0025 $m^2\ s^{-1}$, according to literature values (Soulignac et al. 2017) and to the grid size. Background values were set to zero [$m^2\ s^{-1}$] for vertical viscosity and diffusivity. The heat budget at the air-water interface was computed through the Ocean model. It requires as inputs time series of relative humidity [-], air temperature [°C], net solar radiation [$J\ s^{-1}\ m^{-2}$], sky cloudiness [-], as well as wind speed [$m\ s^{-1}$] and direction [°N], that were downloaded for this study from the Meteofrance station at the Orly airport as described in section 3.1.4.

The module BLOOM makes use of the results from the FLOW simulation (e.g. in terms of velocity field, friction, diffusivity, water temperature), but is run separately. Four main modules are activated in the configuration set up for BLOOM

in this study: oxygen and Biological Oxygen Demand (BOD), dissolved inorganic matter, organic matter and algae. Each module contains numerous variables, that are listed in table 3.1. In particular, the phytoplankton module includes four algal groups commonly present in Lake Champs-sur-Marne: green algae, diatoms, flagellates and cyanobacteria. In the biogeochemical cycle, all the variables activated are interdependent through a large number of processes, simulated by BLOOM. A complete description of these processes can be found in the related manuals (Deltares 2018). Biogeochemical models often include a large number of parameters, possibly site-dependent. In our case study, the processes and variables activated lead to a set of 144 editable parameters.

6.2 Formulation of the calibration problem

In this work, we are interested in the automated calibration of the complex biogeochemical model BLOOM on the case of the Lake Champs-sur-Marne. The objective is to find one or several sets of model parameters that lead to simulated values of the variables of interest that are close to the observed data on a chosen period of time.

Processes in BLOOM are highly parametrized. The default values of the parameters are derived from a years-long calibration process operated by the developers (Deltares 2018). However, most of the parameters in the model remain site-dependent and still need to be calibrated case by case. The calibration of the parameters in BLOOM was performed in two steps. First, a trial and error calibration was implemented to obtain a first estimate of the parameters values adapted to the study site for the simulated period. The results from this initial calibration are described in section 6.3, and were used as a starting point to implement, in a second step, the automated calibration methodology described in sections 6.4 and 6.5.

Among the 144 parameters included in the BLOOM configuration presented in section 6.1, 114 were selected to be included in the calibration, together with 19 initial conditions. Ultimately, this leads to 133 model parameters and initial conditions considered for calibration purposes. The remaining values for parameters and initial conditions were either considered to be known with a sufficiently low uncertainty or showed from previous tests to have a negligible influence on the model outcomes and are therefore not included in the calibration.

A period of 16 days of high-frequency observations going from the 25th of July to the 10th of August 2018 was chosen for the automated calibration of the biogeochemical model. The variables of interest considered for the calibration are total chlorophyll, phycocyanin and dissolved oxygen. These variables are highly representative of biological processes in aquatic ecosystems. Total chlorophyll is a proxy for total biomass growth and is the variable on which most alert guidelines for harmful algal blooms monitoring are based; phycocyanin, being a pigment proper of cyanobacteria can be considered as an indicator of their presence. Finally, the concentration of dissolved oxygen, especially in an eutrophic environment, can be considered as one of the result of the processes of growth, mortality, organic matter decomposition and nutrient recycling.

In order to make the data set suitable for the application of the automated calibration, raw measurements of the three variables needed to be converted in the

appropriate units used in model simulations, namely: [g C m^{-3}] for cyanobacteria and [g O m^{-3}] for oxygen concentration, while total chlorophyll is one of the outputs of the model. Observed dissolved oxygen concentration was converted from percentage of saturation into [g O m^{-3}] using the empirical equation proposed by Weiss (1970), together with high-frequency water temperature data. Phycocyanin was first converted into equivalent of chlorophyll through a conversion factor deduced by comparison with monthly profiles taken *in situ* with the BBE FluoroProbe profiler, and finally to carbon content using the stoichiometric ratio C/Chl. value of 0.03, typically found in scientific literature. Furthermore, profiles taken on the 25th of July with the BBE FluoroProbe were used to set and validate the initial conditions of the model in terms of [μg Chl-a l^{-1}].

The methodology used for the automated calibration are based on a recently developed approach combining Approximate Bayesian Computation (ABC) and Random Forests (RF) denoted ABC-RF in the sequel. This approach is described in section 6.4. In section 6.5, we then introduce a calibration procedure that is based both on the ABC-RF and on a sensitivity analysis of the model outputs.

6.3 Short-term trial and error calibration

The hydrodynamic model Delft3D-FLOW was calibrated based on high-frequency water temperature values at the surface and bottom layers at site B. Three paramters were included in the calibration: the Secchi depth, and the Stanton and Dalton numbers. They were respectively set to 1 m, 0.00145 [-] and 0.0015 [-]. RMSE values were satisfying for both the surface and bottom layers (respectively: 0.6 and 0.5 °C, see Figure 6.1a).

Delft3D-BLOOM was first calibrate through trial and error procedure by comparison with high-frequency observations of dissolved oxygen, total chlorophyll and cyanobacteria concentration. This was done to understand the behavior of the model when its parameters are varied, to detect parameters with very low influence on the models results relatively to the target variables and ultimately to gain the expertise necessary to set up and implement the automated calibration through Approximate Bayesian Computation. Most of the parameters included in the model configuration were varied. However, only a reduced number of the model parameters was actually calibrated. Such parameters deal directly with phytoplankton growth and mortality and are listed in table 6.1, together with their calibrated values.

The main objective of this initial calibration was to tune the model parameters to roughly reproduce the timing of growth and mortality recorded by the high-frequency observations, focusing in particular on total chlorophyll data. This was done to ensure that the model was capable of reproducing such dynamics in a short period of time of only a few days, before the application of the automated calibration methodology.

Model results after this preliminary calibration are shown in panels b, c and d of Figure 6.1. The model correctly reproduces the timing of the chlorophyll peak and the overall dynamics of growth and mortality. However, the results are not satisfactory: total chlorophyll (Fig. 6.1c) is strongly overestimated, as are oxygen consumption (Fig. 6.1b) and cyanobacteria concentration (Fig. 6.1d). Given the parameters in table 6.1 and the environmental conditions, the competition principle

Table 6.1 – Values of the parameters resulting after the trial and error calibration of the model Delft3D-BLOOM. In the first column, the different phenotypes of the four activated phytoplankton groups are indicated as: E (energy limited phenotype), P (phosphorus limited phenotype), and N (phosphorus limited phenotype).

	Growth rate [1 d^{-1}]	Temperature coeff. growth	Mortality rate [1 d^{-1}]	Temperature coeff. mort.	N/C	P/C	Chl-a/C
Diat. (E)	0.45	1.06	0.035	1.08	0.21	0.018	0.04
Diat. (P)	0.45	1.06	0.045	1.085	0.188	0.0113	0.025
Flag. (E)	0.45	1.06	0.035	1.08	0.275	0.018	0.029
Flag. (P)	0.45	1.06	0.045	1.08	0.275	0.018	0.029
Greens (E)	0.45	1.06	0.035	1.08	0.275	0.0238	0.033
Greens (N)	0.45	1.06	0.045	1.085	0.175	0.015	0.025
Greens (P)	0.45	1.06	0.045	1.085	0.2	0.0125	0.025
BG (E)	0.3	1.06	0.035	1.08	0.225	0.0188	0.033
BG (N)	0.3	1.06	0.045	1.085	0.125	0.0188	0.02
BG (P)	0.3	1.06	0.045	1.085	0.15	0.015	0.02

at the basis of primary production in BLOOM does not manage to distribute the biomass correctly, at least in terms of cyanobacteria.

Results from this preliminary calibration tests show that the model is able to simulate rapid dynamics of growth and mortality spanning a few days, and that its results still need to be improved through further calibration. Eventually, this configuration was used as a starting point for the implementation of the Approximate Bayesian Computation.

6.4 Approximate Bayesian Computation

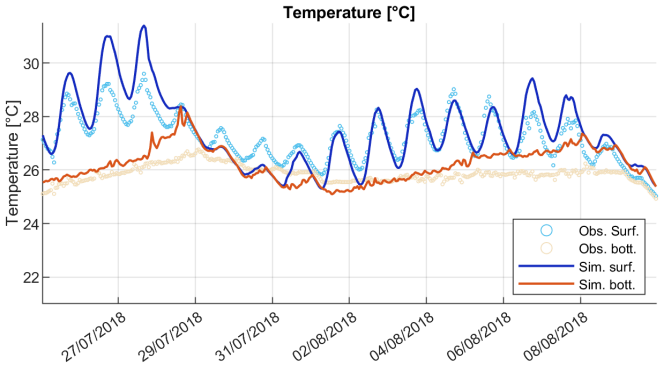
6.4.1 General framework

Approximate Bayesian Computation (ABC) is a class of computational methods rooted in Bayesian statistics first proposed by Beaumont in 2002 (Beaumont et al. 2002). It allows for parameter inference without the need to explicitly compute the likelihood function (Sunnåker et al. 2013). Developed in the field of population genetics, it has quickly grown as a solid alternative to likelihood-based methods for model calibration and it has already been applied in evolutionary biology and ecology (Csilléry et al. 2010). Given a model $M(x, \theta)$ where x are the variables and θ the parameters, and a data set D that is composed of the observed values of the variables x, the posterior probability of the model parameters can be obtained through the Bayes' theorem:

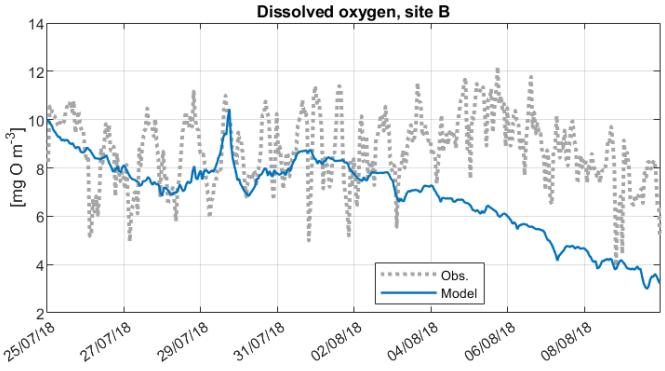
$$\pi(\theta|D) = \frac{\pi(D|\theta)\pi(\theta)}{\pi(D)} \tag{6.1}$$

where $\pi(\theta|D)$ is the conditional probability of the model parameters given the observations D (the posterior probability), $\pi(D|\theta)$ is the conditional probability of the observations given the parameter values (the likelihood function), $\pi(\theta)$ is the prior distribution of θ, and $\pi(D)$ is the marginal probability of the data. The marginal

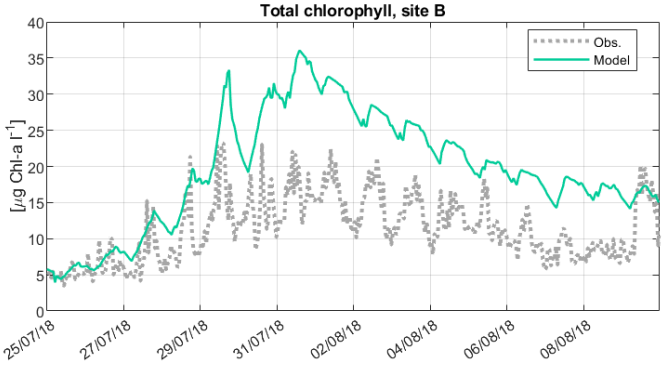
a)



b)



c)



d)

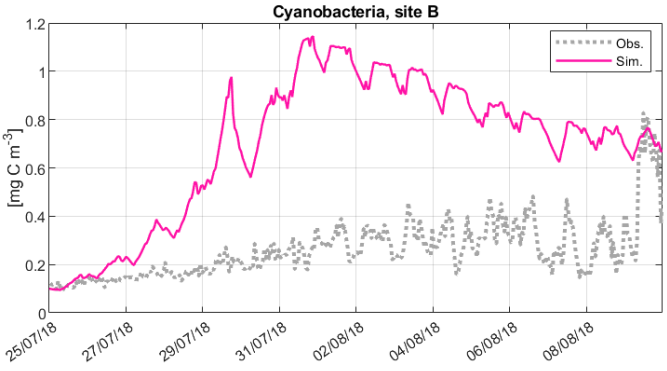


Figure 6.1 – Comparison of model results (solid lines) and high-frequency observations (dotted lines) for water temperature (a), dissolved oxygen (b), total chlorophyll (c) and cyanobacteria (d) concentration relatively to the calibration period in the summer of 2018. In panel a, blue lines represent surface water temperature, and red lines bottom water temperature; in the remaining panels, both observations and model results are relative to the middle layer of the water column at site B.

probability can be considered as a normalizing constant and is often neglected in applications where model intercomparison is not involved. In Bayesian inference, the desired posterior probability can therefore be described through the prior distribution and the likelihood function.

The evaluation of the likelihood function, however, represents the real challenge. It is analytically intractable for most applications, and its estimation through standard computational methods (such as Markov Chain Monte Carlo algorithms) is computationally highly expensive, especially when the dimension of the considered model grows (Csilléry et al. 2010, Sunnåker et al. 2013). The idea at the core of ABC is to bypass the explicit evaluation of the likelihood function, directly obtaining an approximation of the posterior probability distribution.

To do so, prior probability distributions are first defined for the model parameters. The model M is then used to generate a large set of simulations by randomly sampling the parameters values, according to their priors (van der Vaart et al. 2015). The posterior distributions can eventually be estimated through the application of a rejection algorithm or of machine learning techniques.

6.4.2 Standard ABC

In its standard form, the ABC usually consists of three main steps: (i) sample parameter values from the prior distribution, (ii) obtain a simulated data set $\hat{D}$ by running the model M for each sampled value of θ, and (iii) retain a fraction of the sampled parameters by comparison of $\hat{D}$ with the observed data D, in order to approximate their posterior distribution (Burr & Skurikhin 2013). Namely, the rejection algorithm discards a sampled parameter set if the distance (ρ) between $\hat{D}$ and the observations D exceeds a certain threshold (ε). However, as the dimensionality of the data increases, the probability of satisfying the acceptance condition decreases. Both observations (D) and simulated data ($\hat{D}$) are therefore replaced with a set of lower dimensional summary statistics ($S(D)$, $S(\hat{D})$) (Stumpf 2014), that will be more accurately described in section 6.4.4. The rejection criteria therefore writes:

$$\rho(S(D), S(\hat{D})) \geq \varepsilon \tag{6.2}$$

When repeated a sufficient number of times, the outcome of this procedure is a sample of parameter values approximately distributed according to the desired posterior distribution (Sunnåker et al. 2013). Note that a large number of sampled parameters θ and therefore an equal number of simulations $\hat{D}$ are necessary to obtain a robust posterior distribution.

The formulation for the distance ρ, the threshold ε and the summary statistics are choices that the modeller has to take when applying this procedure.

6.4.3 ABC random forest

In its standard form, ABC retrieves the posterior parameters distribution starting from the set of model simulations through a rejection algorithm (Burr & Skurikhin 2013). This entails the definition of a tolerance level separating acceptance from rejection. However, such threshold is arbitrary and should be calibrated for each particular application (Simola et al. 2021). To overcome this issue, Raynal et al.

(2019) proposed to substitute the distance-based rejection algorithm with a machine learning technique, namely the random forests (RF). At the expense of introducing a few parameters defining the structure of a RF, this allows to overcome the definition of the tolerance level.

A regression tree is a structure made of binary nodes, that are of two types: the internal nodes and the terminal nodes (the leaves). It can be automatically built by iteratively dividing a training dataset into subsets of increasing uniformity, until a certain condition is satisfied. Namely, the process of growth of the tree continues until all terminal nodes either (a) have less than n data points (with n possibly equal to 1), or (b) are "pure", that is all elements in a node have (almost) the same outcome.

With such a process, we can build a regression tree to get an estimation of the value of $\theta_i \in \mathbb{R}$, the i^{th} component of the model parameters vector θ. This tree will be trained on a training set of k summary statistics $X^{(k)}$ (see section 6.4.4), which are computed from the set of model simulations. Once the tree is trained, we can apply it on the observed dataset D and get the estimated value of the parameter θ_i.

A random forest consists in aggregating (or bagging) randomized regression trees. A large number of trees (n_{tree}) are trained each on a different bootstrap subsample taken from the complete available training dataset. Furthermore, only a subset of m_{try} summary statistics among the k available are randomly considered at each node for splitting (Raynal et al. 2019). The estimations obtained by the n_{tree} regression trees can be treated and used to obtained a posterior probability distribution for the parameters θ_i. Eventually, once the random forest is grown, different choices can be made for the inference of the parameter value (see section 6.4.6). For example, the final calibrated value of θ_i can be determined by averaging all the n_{tree} predictions obtained in the random forest, or by taking the most probable value from the posterior distribution.

In the present study, we tested a first calibration procedure that relies on the assumption that the model parameters can be considered independent of each other. In that case, the ABC-RF method is applied separately once for each parameter of the model. This way, one RF is built for each parameter and the associated approximate posterior distributions can be plotted, from which an estimate value of the parameters are deduced (see section 6.4.6). The different steps of this calibration procedure are summarized in Algorithm 1. In the sequel, this calibration procedure will be referred as ABC-RF.

For the implementation of the ABC-RF, we make use of a specific R package (*regAbcrf*), developed by Raynal et al. (Raynal et al. 2019) for ABC-RF parameter inference. The configuration of the ABC-RF only needs the definition of two main arguments: (i) the number of trees to grow in the random forest (n_{tree}), and (ii) the number of variables among which to choose for splitting at each node (m_{try}). For the former, different values have been tested (see section 6.6.1) while the latter was set to its default value, which is the maximum between 1 and the number of variables divided by 3.

Algorithm 1: ABC-RF

Data: N: number of parameters to be calibrated
M: number of summary statistics
X_k: summary statistics ($k = 1{:}M$)

Result: estimated value $\hat{\theta}_i$ of θ_i for $i = 1{:}N$

for *i=1:N* **do**
  Application of ABC-RF to estimate the parameter θ_i from the set of summary statistics $\{X_k,\, k = 1 : M\} \rightarrow$ approximate posterior distribution of θ_i
  Determination of the estimated value $\hat{\theta}_i$ of θ_i from the approximate posterior distribution of θ_i
end

6.4.4 Prior distributions and summary statistics

The number of model runs used here to test ABC (standard and RF) on the Delft3D-BLOOM was set to 30000. The model was run 30000 times drawing parameters randomly from the user-assigned prior distributions. Depending on previous knowledge, for each parameter either a uniform or a gamma distribution was chosen. Gamma distributions were defined for most of the parameters, using their default value as mean of the distribution and setting a standard deviation of 20%. The choice of the gamma distribution is motivated by the fact that, unlike normal distributions, it is defined only on positive values through the equation:

$$f(x) = \frac{x^{k-1} e^{-\frac{x}{\theta}}}{\Gamma(k)\theta^k} \tag{6.3}$$

where k and θ can be deduced once the mean value (m) and the standard deviation (σ) are defined by the user, through:

$$m = k \times \theta, \qquad \sigma^2 = k \times \theta^2 \tag{6.4}$$

A uniform density function was assigned to the parameters (such as the initial conditions) for which very little information was available from literature or previous studies, through the following equation:

$$f(x) = \frac{1}{b-a} \mathbb{I}_{[a,b]}(x) \tag{6.5}$$

where a and b represent the lower and upper bounds, respectively. The values of the bounds a and b were set separately for the parameters based on the knowledge from previous trial and error calibrations. For parameters with default values very close to 0, the upper limit was set to 1.

Figure 6.2 shows, as an example, the prior distributions for four of the model parameters included in the calibration routine. Gamma distributions were assigned to the two parameters in panels a and b (*TcPMxGRE_N* and *PPMaxGRE_N*, respectively the temperature coefficient for growth, and the growth rate for green

algae under nitrogen limiting conditions), while uniform distributions were chosen for the two parameters in Fig. 6.2c and 6.2d.

In the case two parameters depend directly on each other's value through a known mathematical equation, a prior probability distribution was assigned to one of the two parameters, and the remaining one was computed through the appropriate equation. This is the case, for instance, of the parameters in panels c and d of Figure 6.2, namely: the fraction of POC1 converted into DOC ($b\_poc1doc$), and the fraction POC1 converted to POC2 ($b\_poc1poc2$), whose sum should be equal to 1. A uniform probability distribution was assigned to the parameter $b\_poc1doc$; consequently, the remaining one also varied through a uniform distribution, via the equation: $b\_poc1poc2 = 1 - b\_poc1doc$.

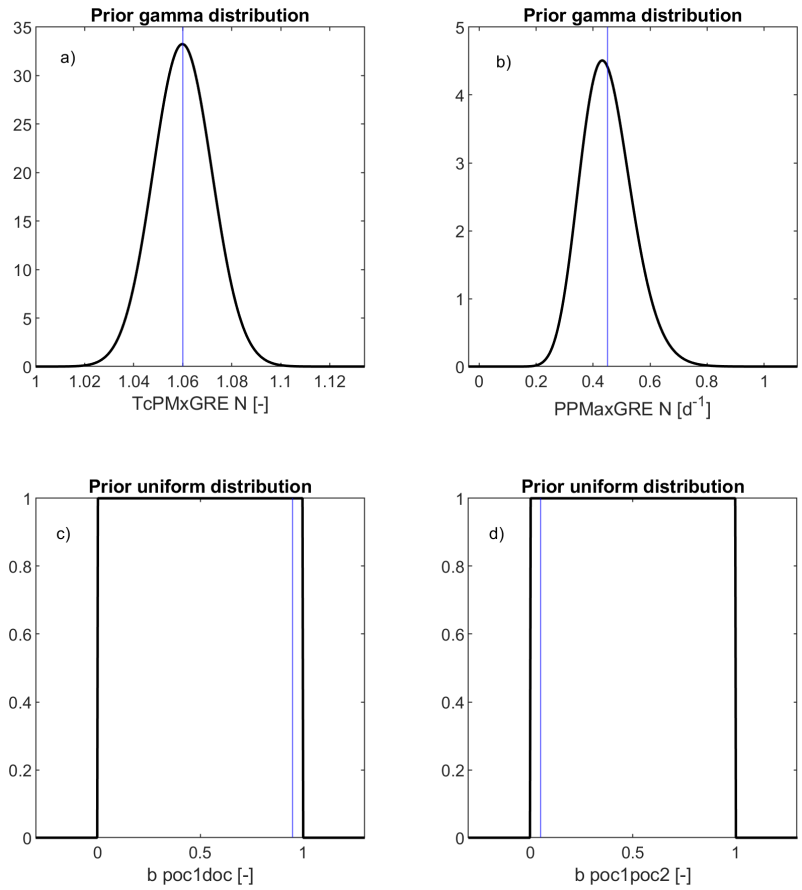


Figure 6.2 – Examples of *a priori* distributions for four parameters. Panals a and b illustrate the gamma distribution, while panels c and d illustrate the uniform distribution. Blue vertical line represent the reference values for the parameters around which the distributions are built.

The use of summary statistics constitutes one significant advantage of ABC. The most relevant aspects and traits of model runs and observations are summarized into a set of user-chosen low-dimensional summary statistics, in order to gain computational time and avoid the so-called curse of dimensionality (Prangle 2015). Each of the 30000 model runs originated by the model M according to the prior distributions is resumed through the summary statistics, and eventually stored in the

so-called *reference table*, together with the corresponding parameter set. Ultimately, the reference table constitutes the training data set on which ABC is applied.

The definition of the summary statistics is therefore crucial, as they replace the model runs in the calibration procedure. Summary statistics should minimize information loss and maximize dimension reduction (Csilléry et al. 2010). However, their choice is also closely correlated to the processes subject of the study. Here, we are mainly interested in the time evolution of a phytoplankton community, in particular in terms of total population, cyanobacterial population and oxygen consumption and production. Summary statistics are calculated on the complete set of model runs as well as on the observed data for the three variables considered for calibration: total chlorophyll, cyanobacteria and oxygen concentrations. Two different summary statistics were tested: (i) the normalized square of residuals between each model run ($\hat{D}$) and the observation series (D) and (ii) the normalized mean square error (NMSE) between $\hat{D}$ and D. The summary statistics are used here similarly to objective functions to be minimized. They have the particularity to be dependent on the observation series D. This choice will be discussed in section 6.8.4.

The normalized square of residuals (R) is defined as follows:

$$R = \sum \frac{\int (\hat{D} - D)^2 dt}{\int D^2 dt} \tag{6.6}$$

and was numerically approximated by use of numerical quadrature (function *integrate.xy* of the R package *sfsmisc*).

The normalized mean square error (NMSE) was computed as defined in (Poli & Cirillo 1993), and normalized via the product between the mean of model runs $\bar{\hat{D}}$ and that of the observations $\bar{D}$:

$$NMSE = \frac{1}{N} \sum_i (\hat{D}_i - D_i)^2 \cdot \left(\bar{\hat{D}} \cdot \bar{D} \right)^{-1} \tag{6.7}$$

6.4.5 Preselection of a subset of simulations

The ABC-RF method can be applied to the whole set of 30000 simulations or on a subset of simulations. Such subset can be chosen in different ways. Here, our objective is to minimize the error between model results and observation series. For this reason, we decided to select the simulations that are the closest to the observations. The preselection of a subset of simulations also reduces the dimension of the reference table and, consequently, the computational time. For this purpose, we calculated for each simulation the sum of the summary statistics (that is the sum of the R or NMSE values of the variables): this value will be called "total R" or "total NMSE" in the sequel. The set of simulations with the smallest values of total R or total NMSE were selected for the subset.

6.4.6 Selection of parameter value from posterior distribution

The calibrated values of the parameters have to be determined from the obtained posterior probability distributions. These distributions might present various local

maxima which makes the selection of an optimal value for the parameters difficult. Three options were therefore considered and tested in this study. Given an approximate posterior distribution $\pi(\theta)$ for a parameter θ, the calibrated value $\hat{\theta}$ of the parameter can be chosen as follows:

• **option** P_{max}: the calibrated value is the one corresponding to the maximal value of the approximate posterior probability distribution:

$$\hat{\theta} = argmax(\pi) \tag{6.8}$$

• **option** P_{med}: the parameter is the median of the approximate posterior probability distribution:

$$\hat{\theta} = median(\pi) \tag{6.9}$$

• **option** $P_{mix,k}$: it is a compromise between the two first options. Depending on a criterion, $\hat{\theta}$ will be equal either to the most probable value (option P_{max}) or the median (option P_{med}):

$$\hat{\theta} = \begin{cases} argmax(\pi) & \text{if } max(\pi) > \frac{k}{\theta_{max} - \theta_{min}} \\ median(\pi) & \text{elsewhere} \end{cases} \tag{6.10}$$

where θ_{min} and θ_{max} are the lower and upper bounds of the support of the approximate posterior distribution π (which is determined numerically and has therefore a finite support), and k is a constant value that has to be chosen. In this study, we will test the values $k = 2, 3$.

The option P_{max} is well adapted to the case where the function π is sharp, whereas the option P_{med} is more suitable for flat functions (see Figure 6.6 as an example). The option $P_{mix,k}$ introduces a threshold to switch between the two first options depending on the shape of the function π.

6.5 Combining ABC random forest and sensitivity analysis

If the model parameters can be considered independent from each other, the ABC-RF can be applied to each parameter indepently from each other as it is done in algorithm 1. However, the value of some of the model parameters might have a non negligible influence on the remaining ones. In order to take into account the possible mutual influence, we set up a different calibration procedure, that includes a sensitivity analysis of the model outputs to the model parameters.

6.5.1 General procedure

Before applying the ABC-RF to the model parameters, a sensitivity analysis is performed using the set of 30000 simulations that is available. This allows to estimate which parameters have the greatest influence on the outcomes of the simulations and to sort the parameters in a chart from the most to the least important. We can then apply the ABC-RF method in a sequence starting from the most important one, adding at each iteration the previously calibrated parameters to the set of

summary statistics. The different steps of this calibration procedure are summarized in Algorithm 2 and will be discussed in the paragraphs 6.5.2 and 6.5.3.

In the sequel, this calibration procedure will be referred as ABC-RF with SA or ABC-RF SA.

Algorithm 2: ABC-RF with SA

Data: N: number of parameters to be calibrated
M: number of summary statistics
X_k: summary statistics ($k = 1{:}M$)
$S_{i,k}$: sensitivity indices ($i = 1{:}N, k = 1{:}M$)

Result: estimated value $\hat{\theta}_i$ of θ_i for $i = 1{:}N$

1 Sorting the parameters according to the values of the sensitivity indices $S_{i,k}$ $\rightarrow$ vector σ of sorted sensitivity indices
2 **for** *i=1:N* **do**
3 Selection of a subset Y of summary statistics X_k
4 Application of ABC-RF to estimate the parameter $\theta_{\sigma(i)}$ from the set of summary statistics $\left\{Y, \theta_{\sigma(1:i-1)}\right\}$ $\rightarrow$ approximate posterior distribution of $\theta_{\sigma(i)}$
5 Determination of the estimated value $\hat{\theta}_{\sigma(i)}$ of $\theta_{\sigma(i)}$ from the approximate posterior distribution of $\theta_{\sigma(i)}$
6 **end**

6.5.2 Sensitivity indices and sorting

Performing a standard sensitivity analysis (based on Sobol or FAST methods) directly to our set of 30000 simulations was not possible because the model parameters are possibly not independent from one another. To overcome this issue, the methodology proposed in (Veiga et al. 2009) for models with correlated inputs was adopted. Following this methodology, a sensitivity index was calculated for each of the 133 parameters and for each summary statistic (namely the R and the NMSE values) of the three variables (chlorophyll-a, phycocyanin and dissolved oxygen). These indices represent the sensibility of the summary statistic of the variable to the variation of the parameter. For each summary statistic and each parameter, three sensitivity indices have therefore been obtained, one for each variable.

For a given summary statistic, two options were then tested to sort the parameters based on the sensitivity indices obtained for the 133 parameters and the three variables (step 1 of Algorithm 2):

- **option "Max"**: sort the parameters following the value of the greatest sensitivity index among the three;

- **option "Sum"**: sort the parameters following the value of the sum of the three sensitivity indices.

6.5.3 Selection of a subset of summary statistics

In Algorithm 2, at each iteration of the loop on the model parameters to be calibrated, we can choose a subset of summary statistics on which the calibration will be applied (step 3 of Algorithm 2). If the parameters are sorted following the values of the sum of the sensitivity indices, we chose to apply the calibration on the whole set of summary statistics at each iteration. In the case where the parameters are sorted following the value of the greatest sensitivity index, two options were considered: the calibration was performed (i) either on the whole set of summary statistics, (ii) or only on the summary statistic of the variable for which the sensitivity index of the current parameter is the greatest.

6.6 Validation of the methodology

6.6.1 Preliminary tests

The calibration methodology was first applied to a set of simulated observations, that is some data issued from a model simulation. The use of simulated observations instead of real observations ensures the existence of a known parameter set with which the model will reproduce the data correctly. This allows to test the capability of the calibration methodology to reproduce both the simulated observations and the parameter value in an ideal case where the model is exact.

Here, the simulation with the lowest total NMSE (i.e. the closest one to the real observations, see section 6.4.5) among the 30000 model runs was selected, and its results in terms of total chlorophyll, dissolved oxygen and cyanobacteria concentration were used as simulated observations. Namely, the best simulation is simulation number 4022, and was discarded from the reference table before the application of the calibration methodology.

In order to choose the main characteristics of the ABC-RF (e.g. the number of simulations constituting the training data set, and the number of trees used to build the random forests), a series of preliminary tests were performed. Such tests only regarded the classic ABC-RF formulation (see Algorithm 1). Namely, the tests investigated the influence on the calibration outcomes of: the number of simulations, the randomness inherent to the ABC-RF procedure, the use of a preselected subset of simulations (according to section 6.4.5), the number of trees in the random forests, and the different options for the selection of the parameters values from the posterior distribution (see section 6.4.6). For these preliminary tests, only one summary statistic was considered (NMSE). The tests are detailed here-after.

In order to test the influence of the number of simulations used in the ABC-RF, various calibrations were carried out using subsets of increasing size of the complete set of simulations. Namely, the number of simulations was varied between 2000 and 30000, with a 2000 step. The subsets were chosen in two ways. Either the elements of each subset were chosen randomly or only the best simulations (in terms of values of total NMSE) were selected as proposed in section 6.4.5. For these tests, the number of trees was set to 500, and each calibration was performed ten times (with the same parameters configuration) to test the variability of the results inherent to the randomness of the method.

Table 6.2 – Summary of the calibration runs performed for the preliminary tests, the validation on simulated observations, and the application on real data.

a) **Preliminary tests**

Method of calibration	**Summary statistic**	**Training dataset size**	**Simulation preselection**	**Number of trees**	**Repetition number**	**Parameter value selection**
ABC-RF	NMSE	2000:2000:30000	random & closest	500	10	P_{max}, P_{med}, $P_{mix,2}$, $P_{mix,3}$
ABC-RF	NMSE	2000:2000:30000	random & closest	250:250:2000	1	P_{max}, P_{med}, $P_{mix,2}$, $P_{mix,3}$

b) **Application on simulated and real observations with (i) preselection of the closest simulations for the training dataset, (ii) 500 trees per random forest, (iii) no repetition**

Method of calibration	**Summary statistic**	**Training dataset size**	**Parameter value selection**	**Sensitivity index based sorting**	**Variables used in calib.**	**Number of calib. runs**
with the three variables						
ABC-RF	NMSE & R	10000 & 25000	P_{max}, P_{med}, $P_{mix,2}$, $P_{mix,3}$	-	-	16
ABC-RF SA	NMSE & R	10000 & 25000	P_{max}, P_{med}, $P_{mix,2}$, $P_{mix,3}$	Max	One	48
				Max	All	
				Sum	All	
with only real chlorophyll data						
ABC-RF SA	R	10000 & 25000	P_{max}, P_{med}, $P_{mix,2}$, $P_{mix,3}$	Max=Sum	One=All	8

Similarly, eight calibrations were carried out with an increasing value of number of trees. The number of trees was varied from 250 to 2000 with a 250 step and tested for all simulation subsets. Each calibration was performed only one time.

For each of these tests, once the posterior distribution was obtained for each model parameter, the four options defined in section 6.4.6 (options P_{max}, P_{med}, $P_{mix,2}$ and $P_{mix,3}$) were applied. This provided several sets of estimated parameter values. For each of these estimated parameter sets, the model was then run. The so-obtained simulations were finally compared with one another by calculating the total NMSE between model results and simulated observations.

A summary of all the calibration runs performed for the preliminary tests is given in Table 6.2a.

The most relevant results of the tests that investigate the influence on the calibrated model output of (i) the number of simulations used as training dataset, (ii) the uncertainty deriving from the inherent randomness of the ABC-RF, and (iii) the use of preselected simulations as training dataset are highlighted in Fig. 6.3. The Figure shows the evolution of the total NMSE between calibrated model outcomes and simulated observations, according to the size of the simulations subset used for the calibration. Panel a), is relative to subsets with randomly selected simulations, while panel b) is relative to preselected simulations (see section 6.4.5). For each subset size, ten calibration runs were carried out. The resulting values of total NMSE are plotted in Fig. 6.3 as a series of box-plots. On the box-plots, the red "minus" and "plus" symbols indicate the median and the mean values of the ensemble, respectively; the bottom and top edges of the box respectively mark the 25th and 75th percentiles. Eventually, the maximal and minimal values of the ensemble are marked through the whiskers extent. Fig. 6.3 shows the calibration results obtained with the option $P_{mix,2}$ only; similar results have been obtained for the other options of parameters values selection.

The results in Fig. 6.3 show that in the case of randomly selected simulations

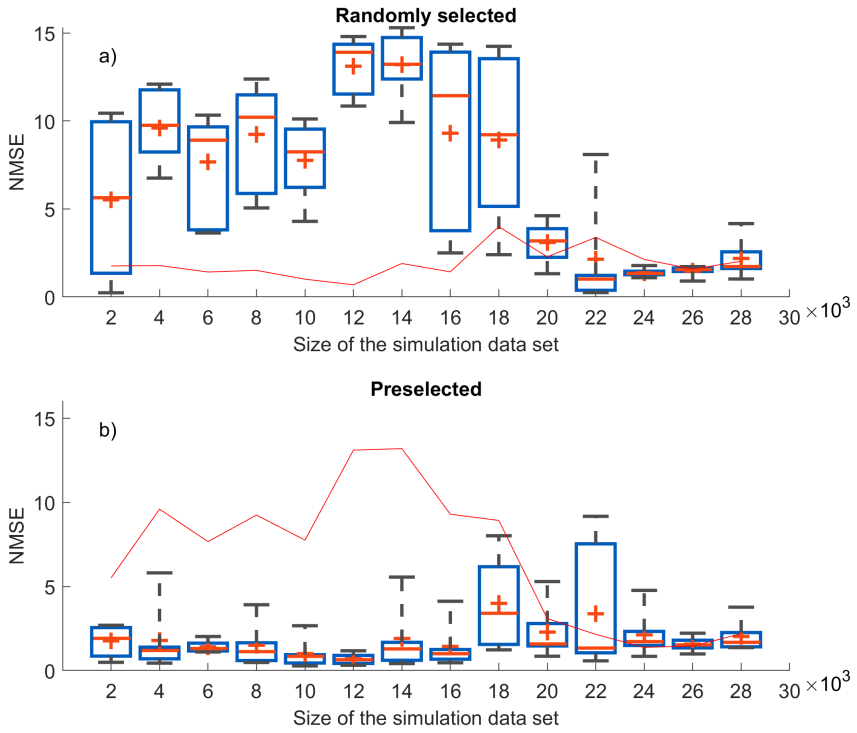


Figure 6.3 – Total NMSE between the simulated observations and the model results calibrated through the ABC-RF calibration, according to the size of the simulations training subset. Panel a): randomly selected subsets; panel b): subsets of preselected simulations. The box-plots represent relevant statistical characteristics of the ensembles of ten equivalent runs of ABC-RF calibration. The red "minus" and "plus" symbols indicate the median and the mean values of the ensemble, respectively. The bottom and top edges of the boxes respectively mark the 25th and 75th percentiles, and the maximal and minimal values of the ensemble are marked through the whiskers extent. Eventually, the red solid line in panel a) represents, for a direct comparison, the mean values of the box-plots in panel b). Vice-versa for the red solid line in panel b). The number of trees was set to 500 and the option for parameters values selection is $P_{mix,2}$.

(panel a) both the average error and the variability of the calibration runs (i.e. the extent of the box-plots) decrease strongly for subsets of at least 20000 simulations. In particular a minimum of both the total NMSE value and the variability was detected for the case of 24000 simulations. Panel b) of the same figure, shows that the pre-selection of the best simulations largely improves the results of the calibration procedure, both in terms of average total NMSE and of variability. In the case of preselected subsets, both the average error and the variability appear to be rather independent from the number of simulations used for the calibration, as they do not vary strongly with the increase of the subset size (except for the cases of 18 and 22 thousand simulations where the mean values and the variability are greater, perhaps because of the randomness of the method and of the number of repetitions not high enough). Minima can be found for ten and twelve thousand simulations. With randomly selected subsets, in order to obtain performances comparable to the case of preselected subsets in terms of total NMSE and variability, at least 20000 simulations are needed as training dataset.

The tests carried out on the number of trees used to build the random forests did not show any particular influence on the calibration outcomes, that did not consistently improve as the number of trees was increased, neither in terms of total NMSE nor in terms of variability.

The options for the choice of the parameters values from the posterior distribution did influence the outcomes of the model. However, in this regard the preliminary tests did not show clear and concluding results, and it was not possible to identify one option that performed consistently better than the others for the three variables at the same time. For this reason, all four options (P_{max}, P_{med}, $P_{mix,2}$ and $P_{mix,3}$) will be tested in the subsequent applications.

To conclude, the tests described above indicate that, for the model under consideration, a training set of at least 24000 randomized simulations is required in order to minimize the error and the variability of the ABC-RF methodology. In the case of preselected simulations, comparable results can be obtained with a smaller set of around 10000 simulations. The number of trees didn't show a strong effect on the outcomes of the calibration procedure, and none of the four options implemented for the choice of the parameters values could be preferred over the others.

6.6.2 Application on simulated observations

Following the results of the preliminary tests presented in section 6.6.1, the ABC-RF (algorithm 1) and the ABC-RF with SA (algorithm 2) were tested and compared under two configurations: using a subset of either (i) 10000 or (ii) 25000 preselected simulations. In both cases, the number of trees used to build the random forests was set to 500, and both R and NMSE were tested as summary statistic. In the case of ABC-RF with SA, the three possible combinations of parameters sorting options (section 6.5.2) and options of selection of a subset of summary statistics (section 6.5.3) were tested. Furthermore, the four options for the choice of the parameters values (options P_{max}, P_{med}, $P_{mix,2}$ and $P_{mix,3}$) described in section 6.4.6 were also examined for each calibration.

The combination of all the methodologies (ABC-RF and ABC-RF with SA), configurations and above-described options results, for each summary statistic, in

a set of eight calibration runs with the ABC-RF, and 24 calibration runs with the ABC-RF with SA (i.e., 64 calibration runs overall) that are summarized in Table 6.2 b). Model simulations were then carried out with the 64 estimated parameters sets, and their outcomes were compared to the appropriate set of observations through the value of total R or total NMSE, coherently with the choice of the summary statistic.

The ABC-RF (algorithm 1) and ABC-RF with SA (algorithm 2) were first tested using the set of simulated observations (as described in section 6.6.1).

In that case, the estimated parameter values were also compared with the known parameter values used to generate the simulated observations. To do so, the error (e) between calibrated and known parameters was calculated, normalized over the range of variability allowed for each parameter, and converted into a percentage:

$$e = \frac{|\theta_{calib} - \theta_{true}|}{\theta_{max} - \theta_{min}} \cdot 100 \tag{6.11}$$

where θ_{calib} is the value of the calibrated parameter, θ_{true} is the known parameter value used to generate the simulated observations, and θ_{max} and θ_{min} are the values of θ above and below which the prior distribution takes values smaller than 0.05.

Broadly speaking, the results of the calibration procedure in terms of model outcomes were similar whether R or NMSE are used as a summary statistic. The overall best calibration being obtained with R as summary statistic, results using NMSE will not be discussed.

The eight best calibration runs (in terms of total R value) for the ABC-RF with SA with R as summary statistic are listed in table 6.3, along with the two best calibration runs for the ABC-RF without SA. In the table, for each calibration run are given the number of preselected simulations (10000 or 25000), the name of the method (ABC-RF or ABC-RF with SA), the options used for the calibration, and the value of the total R. The definition of the different options of the ABC-RF with SA are given in section 6.5.2, 6.5.3 and 6.4.6. The calibration runs are sorted according to the total R value.

The best overall calibration run (the one with the lowest value of total R) is obtained with the larger set of 25000 simulations. Its total R value (0.057), is sensibly lower than all the remaining calibration runs. However, the remaining seven calibration runs presented in table 6.3 for the ABC-RF with SA also show good model performances. In particular, the use of the smaller preselected training data set does not deteriorate model performances. Notably, the second best overall calibration (total R=0.109) is obtained with the 10000 preselected simulations.

The implementation of the sensitivity analysis in the ABC-RF methodology strongly improves model outcomes when compared both to the ABC-RF. All calibration runs included in table 6.3 that make use of the SA show total R values five to ten times lower than those obtained through the classic ABC-RF. The implementation of the SA in the ABC-RF framework therefore appears to be crucial for the application of ABC to a complex process-based model.

The two best calibration runs are plotted for the three variables of interest in Fig. 6.4, along with the simulated observations. Total chlorophyll (Fig. 6.4a) shows the strongest differences among model runs. The best overall calibration run (red line) follows very closely the simulated observations, correctly reproducing the daily

Table 6.3 – Application to the simulated observations. List of the calibration runs with the lowest total R values for the ABC-RF with SA (eight best runs) and without SA (two best runs), sorted according to the total R value. The characteristics of each calibration run are detailed in terms of: size of training data set, calibration method, sorting of the sensitivity indices, variables used in the calibration, and option for selecting the parameter value (see sections 6.5.2, 6.5.3 and 6.4.6).

Training data set size	Method of calibration	Sensitivity index based sorting	Variables used in calib.	Parameter value selection	Total R
25000	ABC-RF SA	Max	One	P_{max}	0.057
10000	ABC-RF SA	Sum	All	$P_{mix,3}$	0.109
25000	ABC-RF SA	Sum	All	P_{med}	0.115
10000	ABC-RF SA	Max	All	P_{max}	0.125
10000	ABC-RF SA	Sum	All	$P_{mix,2}$	0.127
25000	ABC-RF SA	Sum	All	P_{max}	0.137
10000	ABC-RF SA	Max	All	$P_{mix,2}$	0.137
25000	ABC-RF SA	Max	All	P_{max}	0.138
25000	ABC-RF	-	-	P_{med}	0.594
10000	ABC-RF	-	-	P_{med}	0.671

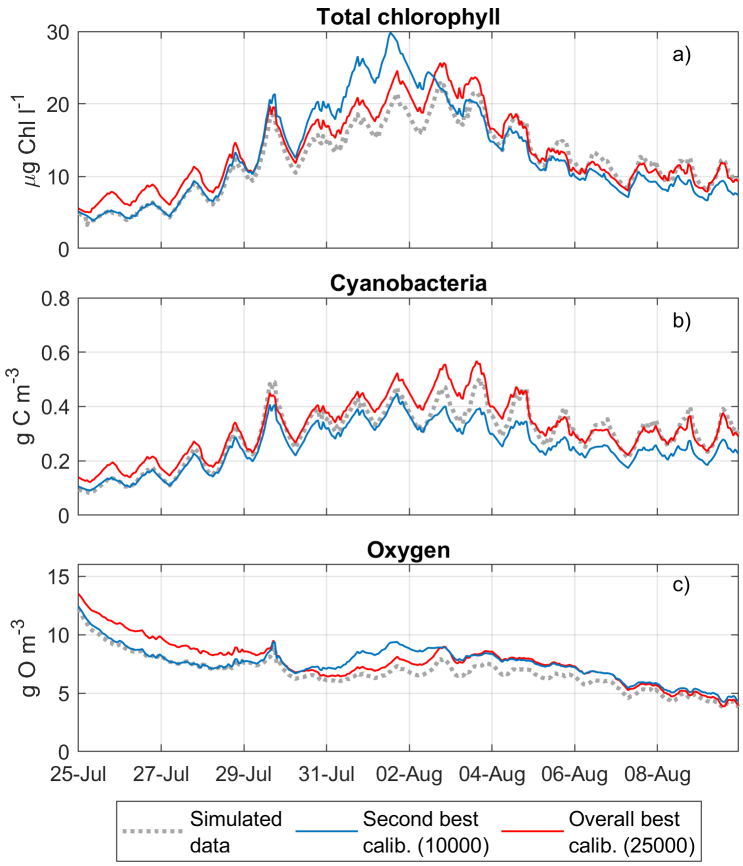


Figure 6.4 – Time series of total chlorophyll (a), cyanobacteria (b) and dissolved oxygen (c) concentrations for the two best calibration runs and the simulated observations (dashed grey lines). The best overall calibration run (red lines) is obtained with a set of 25000 simulations, while the second best (blue lines) with a set of 10000 preselected simulations.

oscillations. The second best calibration (blue line) shows an anticipated and slightly overestimated peak of chlorophyll.

Both the calibration runs plotted in the Figure give good results in terms of cyanobacteria (Fig. 6.4b), with a slight underestimation of the second best calibration (blue line) in the last days of the bloom event, and oxygen concentration (Fig. 6.4c).

As shown by the attributes featured in Table 6.3, the two best calibration runs plotted in Fig. 6.4 are obtained through different configurations of the ABC-RF with SA. Namely, the overall best calibration run, which is performed on the set of 25000 simulations, was obtained by sorting the parameters based on the value of the greatest sensitivity index among the three variables of interest (option "Max" in section 6.5.2), and by calibrating each parameter only using the summary statistic of the corresponding most influential variable (see section 6.5.3). The second overall best calibration (subset of 10000 simulations) is obtained by using the opposite options for the parameters sorting and the selection of summary statistics used for the calibration (see sections 6.5.2 and 6.5.3). The options for the selection of the parameter values from the posterior distribution are also different for these two calibrations runs (P_{max} and $P_{mix,3}$).

The use of simulated observations assures the existence of a known parameter set that yields to a correct reproduction of the data. The difference between the calibrated parameters sets and the original one of the two best calibration runs (i.e. those plotted in Fig. 6.4) have been quantified by means of the error e defined in equation (6.11). Panels a and c of Fig. 6.5 show the values of the error e for the parameters of the two best calibration runs: the second best calibration (obtained with a subset of 10000 simulations, Fig. 6.5a), and the best overall calibration (obtained with a subset of 25000 simulations, Fig. 6.5c).

The error bars in panels a and c of Fig. 6.5 are sorted according to the sensitivity indices calculated for the parameters in the corresponding calibration run (see section 6.5.2 and table 6.3 for details). Only the 30 most relevant parameters (in terms of values of sensitivity index) are shown. The sensitivity indices used for sorting the parameters are plotted through a heat chart in figures 6.5b (second best calibration run), and 6.5d (overall best calibration run). Their values are here normalized over the sum of the indices and converted into a percentage of importance.

As shown by figures 6.5b and 6.5d, the order of importance of the parameters varies only slightly between the two calibration runs. In particular, concerning the ten most important parameters, the main differences between the two calibration runs are the permutation of the second and third most important parameters and a ninth parameter, which changes from a flagellate's growth rate (for the run with 10000 preselected simulations), to the Chl:C ratio for flagellates (for the run with 25000 simulations). Panels b and d also show that the percentage of importance of the parameters decreases rapidly: for both calibration runs it is around 20% for the most important parameter, and drops under 5% after the parameter in tenth position.

In terms of error, the two calibration runs from Fig. 6.5 show different behaviours. The second best run (panel a) has rather low errors for the ten most important parameters, smaller than 25%. The errors grow over 40% only for parameters with very low impact on the model outcomes.

On the other hand, the overall best calibration run (panel c), shows considerable

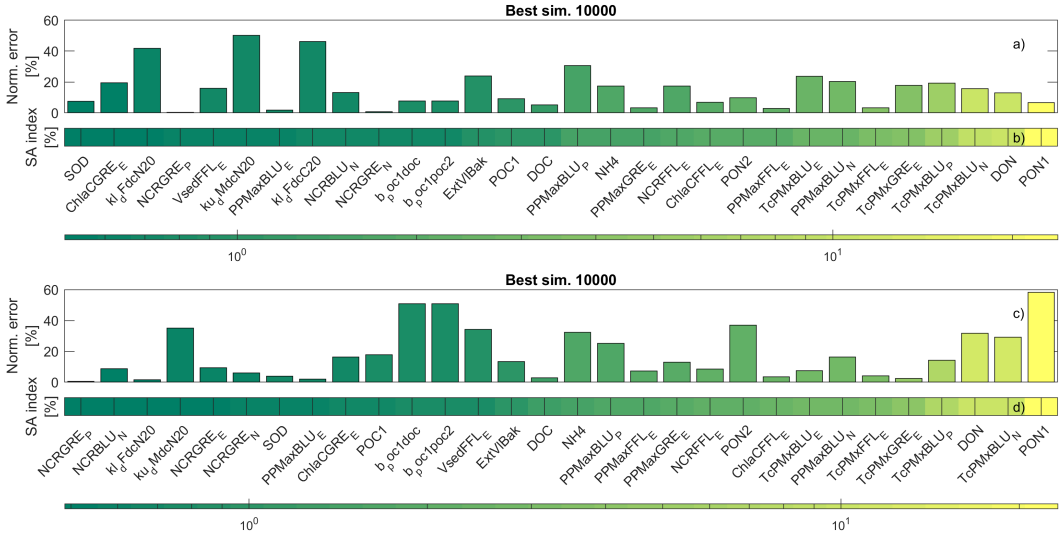


Figure 6.5 – Normalized error e (see equation (6.11)) between the estimated parameters set and the one used to generate the simulated observations (a and c) and associated color chart showing the sensitivity indices for the 30 most influential parameters (b and d). Panels a) and b) are relative to the best calibration run with a subsample of 10000 simulations. Panels c) and d) are relative to the best calibration run with 25000 simulations. Both color charts are plotted with a logaritmic scale.

errors for the most important parameters (around 60% for the most important one, and around 30% for the second and third). Low errors are then found for the remaining parameters up to the tenth position. After the tenth position, the errors increase without a specific pattern, similarly to what found in panel a. Despite such large errors in the estimation of the most relevant parameters, the ABC calibration represented in Fig. 6.5c eventually manages to give the overall best results.

The application of ABC-RF with SA influences the parameters posterior distributions. As the number of summary statistics used for the calibration increases at each iteration of the algorithm 2, the resulting posterior distributions tend to be smoother and less irregular compared to those obtained through the classic ABC RF, for which the same summary statistics are used for the calibrations of all the parameters, that are independent from each other. This smoothing effect increases with the number of iterations of algorithm 2, but can already be noticed after only a few iterations and therefore on most of the parameters.

Figure 6.6 shows, as an example, the posterior distribution estimated by a run of the ABC-RF (panels a and c) and of the ABC-RF with SA (panel b and d) for the two parameters, namely the seventh (panels a and b) and the 80th (panels c and d) in order of importance according to the sensitivity analysis shown in Fig. 6.5d. Already at the seventh iteration, the posterior distribution appears to be sensibly smoother in the case of the ABC with SA (Fig. 6.6b). This is even more striking for parameters further on in the calibration loop, such as that of figures 6.6c and 6.6d.

In particular, such smoothing has a strong impact in the choice of the parameter values. In Fig. 6.6a, at least four parameter values sensibly different correspond to local maxima with similar probability density. In Fig. 6.6b, even though local

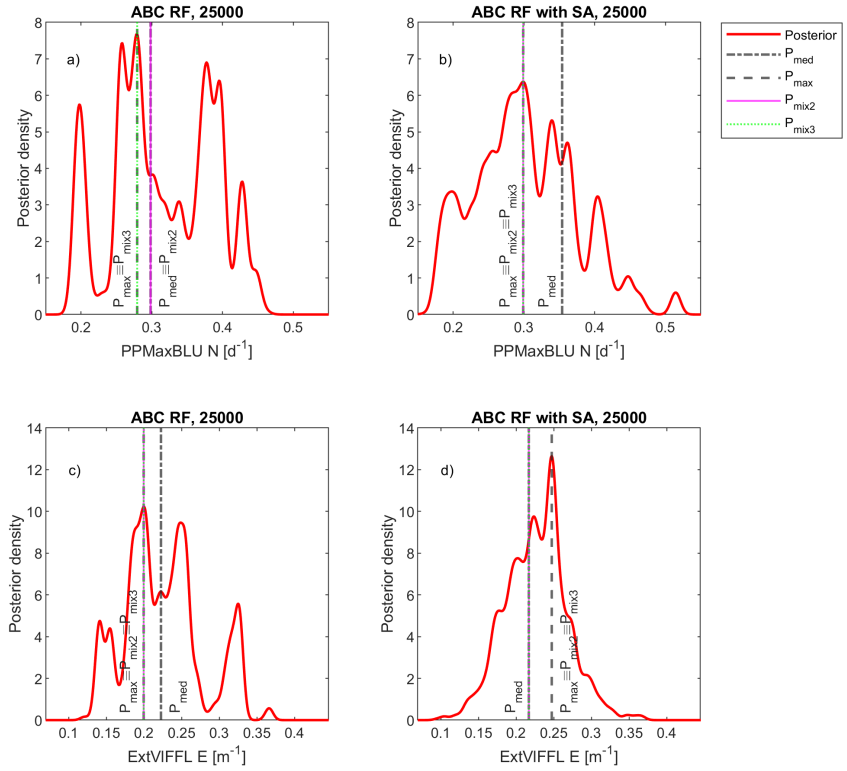


Figure 6.6 – Examples of posterior probability density for two parameters: PPMaxBLU N (growth rate of cyanobacteria under Nitrogen-limiting conditions, panels a and b) and ExtVIFFL E (specific light extinction coefficient for flagellates under light-limiting conditions, panels c and d). The corresponding values for P_{max}, P_{med}, $P_{mix,2}$, $P_{mix,3}$ are also compared. Panels a and c are relative to the aplication of the ABC-RF to a subset of 25000 simulations. Panels b and d are relative to the ABC-RF with SA, applied to a subset of 25000 simulations.

maxima are still present, the most probable parameter value is more simply detectable from the estimated posterior distribution.

The shape of the posterior distribution also affects the outcomes of the four options described in section 6.4.6. For the posterior distribution of Fig. 6.6b (ABC-RF with SA), $P_{mix,2}$ and $P_{mix,3}$ both coincide with P_{max}, while in the case of the classic ABC-RF (Fig. 6.6a) $P_{mix,3}$ coincides with P_{max}, and $P_{mix,2}$ with P_{med}.

6.7 Application on real data

After validation of the calibration procedures on the simulated observations, the ABC-RF (algorithm 1) and ABC-RF with SA (algorithm 2) were tested using the set of real data as observations. The 64 calibration runs described in section 6.6.2 in the case of simulated observations (and summarized in Table 6.2 b)) were applied to the real data set.

Finally, we tested the ABC-RF with SA using only one of the three variables (i.e. as if observations were recorded for only one variable). To do so, we decided to focus on total chlorophyll, one of the variables most commonly measured in the framework of freshwater ecological studies. This was done with two main objectives: (i) to assess the quality of the calibration when only one variable is targeted; (ii) to test the capacity of the calibrated model to simulate the two remaining variables when they are not included in the calibration. For this test, only the ABC-RF with SA was applied and the summary statistic relative to total chlorophyll was the only one used. Similarly to the previous calibration runs, two preselected subsets of 10000 and 25000 simulations were tested, but the selection was based on the total R values of the total chlorophyll a only. The four options described in section 6.4.6 for parameter values selection were considered. The two options for parameters sorting (see section 6.5.2) and the two options for selection of a subset of summary statistics (see section 6.5.3) being the same when only one variable is considered, we eventually performed eight calibration runs that are summarized in Table 6.2b.

Similarly to the application to the simulated observations, also when applied to the real observations the two summary statistics (R and NMSE) lead to similar results. As the best overall calibration is obtained with R in the application to real observations, only the results obtained with this summary statistic will be presented in the following.

Table 6.4 sums up the best calibration runs obtained with the ABC-RF with SA (eight runs) and the ABC-RF without SA (two runs) applied to the real observations, using R as summary statistic. In the table, the calibration runs are sorted according to the total R value, which is also given in the table, as well as the characteristics of each calibration run. The R value for the closest simulation to the real observations (simulation number 4022, see section 6.6.1) is also provided as a benchmark.

As it was already highlighted in section 6.6.2, the use of the sensitivity analysis through the implementation of the ABC-RF with SA (algorithm 2) is crucial for making the ABC suitable to a complex model such as the one under examination. Indeed, only the calibration runs with ABC-RF with SA lead to values of total R smaller than the one of the best simulation (simulation number 4022). The best calibration runs of the ABC-RF without SA are reported in table 6.4 for the cases of 10000 and 25000 simulations subsets, and show considerable poor performances,

Table 6.4 – Application to the real observations. List of the calibration runs with the lowest total R values for the ABC-RF with SA (first eight runs) and without SA (first two runs), sorted according to the total R value. The characteristics of each calibration run are detailed in terms of: size of training data set, calibration method, sorting of the sensitivity indices, variables used in the calibration, and option for selecting the parameter value (see sections 6.5.2, 6.5.3 and 6.4.6). The total R value for the closest simulation to the simulated observations is also provided as a benchmark.

Training data set size	**Method of calibration**	**Sensitivity index based sorting**	**Variables used in calib.**	**Parameter value selection**	**Total** R
10000	ABC-RF with SA	Sum	All	$P_{mix,3}$	0.282
25000	ABC-RF with SA	Sum	All	P_{max}	0.327
10000	ABC-RF with SA	Sum	All	P_{max}	0.341
10000	ABC-RF with SA	Sum	All	$P_{mix,2}$	0.342
25000	ABC-RF with SA	Sum	All	$P_{mix,2}$	0.343
10000	ABC-RF with SA	Max	All	P_{max}	0.344
25000	ABC-RF with SA	Max	All	$P_{mix,2}$	0.345
10000	ABC-RF with SA	Max	All	$P_{mix,2}$	0.346
25000	ABC-RF	-	-	P_{med}	0.594
10000	ABC-RF	-	-	P_{med}	0.671
				Sim. number	
			Best simulation:	4022	0.368

with a total R around two times greater than the one of the best overall calibration run.

The best overall calibration run is here obtained with the smaller set of 10000 simulations, and through the implementation of algorithm 2 (ABC-RF with SA). With a total R value of 0.282, it is sensibly better than the other calibration runs. Namely, results from the second overall best calibration run, obtained through the larger set of 25000 simulations, lead to a total R value 16% greater (R=0.327). This total R value is 21% greater for the third best calibration run.

In terms of characteristics, the five best calibration runs are all obtained by sorting the parameters based on the value of the sum of the sensitivity indices over the three variables (option "Sum" in section 6.5.2). However, none of the four options for the selection of parameter values from the posterior distribution appears to be preferable compared to the others.

The best overall calibration run is shown in Figure 6.7 (red lines), along with the observation series (dashed grey lines) and the closest simulation to the observation from the training dataset (simulation number 4022, black lines). When compared to simulation 4022, model performances improve after the calibration with the ABC RF and SA of around 24%. Namely, the peak of chlorophyll is more correctly modelled by the best calibration run, both in terms of timing and of maximal concentration; cyanobacteria concentration is also more precisely simulated, especially during the growing phase. Concerning oxygen concentration, the calibration of the initial condition allows to avoid underestimation in the second half of the simulations. After calibration with the ABC-RF with SA, the model seems therefore to be able to correctly reproduce the general behaviour for all three variables. However, the amplitude of the observed daily variations is strongly reduced in the model results, in particular in terms of total chlorophyll and oxygen concentration.

The ABC-RF with SA analysis was eventually applied using only the total

chlorophyll dataset. Several calibration runs with different options and configurations were performed as introduced at the beginning of the present section. The best calibration run, that was selected on the basis of the R value for total chlorophyll only, was obtained with a subset of 10000 preselected simulations. It is plotted in Fig. 6.7 (purple lines). In terms of total chlorophyll (panel a), the performance of this calibration run is better than the best calibration obtained on the complete dataset (red lines), with R values (of total chlorophyll) respectively equal to 0.0764 and 0.0929. However, the improvement obtained by focusing only on one variable is marginal (roughly 15%), especially when results in terms of the two remaining variables are considered. Panels b and c of Fig. 6.7 show how the phytoplankton growth is attributed to species other than cyanobacteria, and how erroneous anoxic conditions are simulated in this case by the model. Namely the total R of the simulation calibrated only on total chlorophyll is equal to 1.377, whereas that of the overall best calibration run is equal to 0.282. Finally, in this application the alternatives for parameter selection $P_{mix,2}$ and $P_{mix,3}$ always coincided with P_{max}, suggesting particularly peaked posterior distributions.

The calibration using only total chlorophyll data was not sensibly improved by the use of a larger subset of 25000 simulations, with R equal to 0.0761 for total chlorophyll and a total R on the three variables of 1.377.

6.8 Discussion

In this study, Approximate Bayesian Computation with random forest (ABC-RF) was tested for the calibration of a complex highly-parametrized biogeochemical model. The calibration procedure focuses on three variables: total chlorophyll, cyanobacteria and dissolved oxygen concentrations, that are particularly relevant in aquatic ecology and for the management of water resources.

6.8.1 A novel approach for parameter inference

Approximate Bayesian computation is a methodology that quickly became a standard technique for parameter inference (Beaumont et al. 2002, Toni et al. 2009, Raynal et al. 2019). Even though ABC theoretically enables Bayesian inference for models of almost arbitrary complexity (van der Vaart et al. 2015, Beaumont 2010, Csilléry et al. 2010), both deterministic and non-deterministic (Toni et al. 2009), the ensemble of ABC methodologies has not yet been tested against highly-parametrized deterministic models. To the best of our knowledge, our application to the biogeochemical model Delft3D-BLOOM is the first to test any ABC-based methodology for the calibration of a complex physically-based model with so many parameters to be calibrated.

Novelty of the work

To date, most applications of ABC for parameter inference in scientific literature either make use of its standard formulation, in which the posterior distribution is estimated through a rejection algorithm (e.g. Lagarrigues et al. 2015, van der Vaart et al. 2015, Dominguez Almela et al. 2020), or focus on methods that improve the robustness and efficiency of the rejection algorithm (e.g. Markov chain Monte Carlo

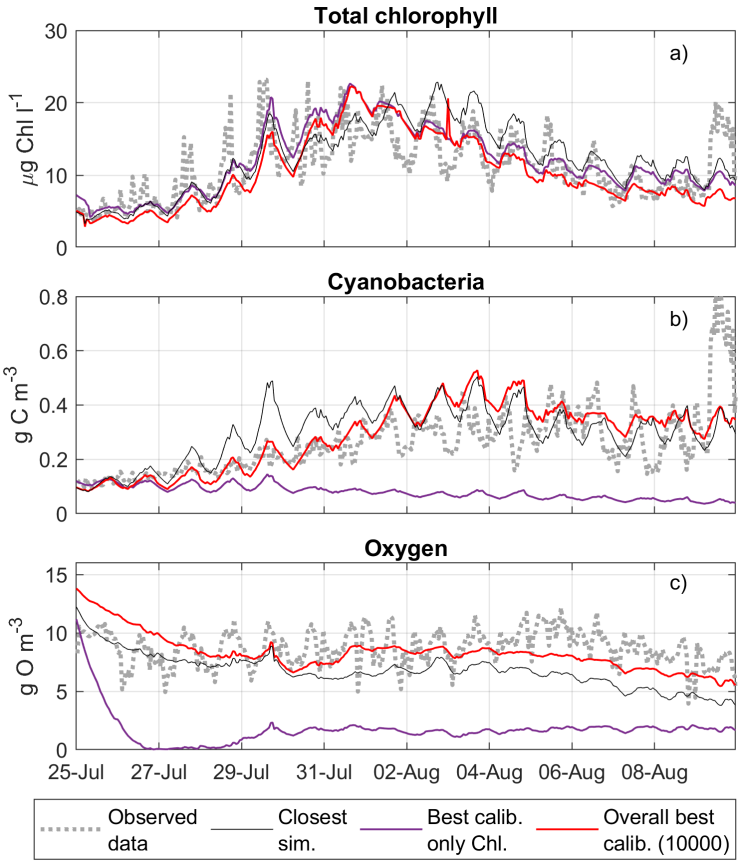


Figure 6.7 – Time series of total chlorophyll (a), cyanobacteria (b) and dissolved oxygen (c) concentrations for: real observations (dashed grey lines), closest simulation to the data (simulation n. 4022, black lines), best overall calibration run (red lines), and the best calibration considering only total chlorophyll (purple lines).

(Marjoram et al. 2003) or sequential Monte Carlo methods (Sisson et al. 2007, Toni et al. 2009, Beaumont et al. 2009)).

The ABC random forest, proposed in 2019, allows to surpass the rejection algorithm in favour of machine learning techniques (Raynal et al. 2019). Only a few applications of this recent methodology can be found at present in scientific literature outside of (Raynal et al. 2019) (e.g.(Chapuis et al. 2020, Estoup et al. 2018)). The ABC-RF was tested here in its original form as proposed by Raynal et al. (2019), and in a novel framework where results from a sensitivity analysis are integrated in the calibration procedure to take into account the mutual influence between the model parameters and their relative importance with respect to the model outputs.

Preliminary tests were carried out on the ABC-RF in order to define a robust configuration in terms of the size of the training data set and of the number of trees in the random forests. These tests show that at least 25000 simulations are needed as a training data set to obtain a good calibration of the three variables under consideration. A comparable value of 20000 simulations was found in a similar application for a model with less parameters (Lagarrigues et al. 2015). In (Estoup et al. 2018), where the authors applied the ABC-RF not for parameter inference, but for model choice, a dataset of 100000 simulations was used. In our application, we also found that the number of simulations can be reduced to 10000 without deteriorating the overall calibration performance, by preselecting the closest simulations to the observations. On the other hand, the test performed on the influence of the number of trees show that this variable has not a great impact on the calibration results: it was therefore set to its default value as preconized in (Estoup et al. 2018).

Due to the novelty of the methodologies, the ABC-RF and ABC-RF with SA were tested under different configurations. The two methodologies were subsequently tested against both simulated and observed data.

Independently from the size of the training dataset, the standard formulation of the ABC-RF was not successful in calibrating the model parameters. In this case the calibrated model performance would not justify the computational effort of the methodology. However, the calibration procedure benefits greatly from the coupling with the sensibility analysis proposed here through algorithm 2. Through the ABC-RF with SA, the performances of the calibrated model improve considerably, fully justifying the use of this methodology.

6.8.2 Computational effort

The stock of simulations

The main computational cost in ABC is the generation of the reference table. The preliminary tests showed that, for the model BLOOM, a set of at least 25000 simulations allows to drastically reduce both the overall error and the uncertainty of model outcomes. Depending on the model under consideration, this might be a relatively high number of model runs.

However, differently from other popular techniques for automated calibration (e.g. Newton or genetic algorithms), the computational cost of the methods based on ABC resides mainly in the generation of the simulation dataset. Once this task is

completed, the calibration itself is usually computationally inexpensive. This allowed for instance to carry out numerous calibration runs under different configurations. This represents a great advantage, especially in relation to a young methodology such as ABC that still lacks a structured working framework.

The computational impact of coupling ABC-RF and SA
With the implementation of algorithm 2 (ABC-RF with SA), the computation cost has increased: at each iteration the number of summary statistics is incremented, inducing a persistent increase of the computational time. Furthermore, as the parameters are interdependent in their determination, their computation cannot be parallelized. Broadly speaking, run on the same computer and with a set of 25000 simulations, the ABC-RF with SA can take up to ten times longer than the ABC-RF to complete the estimation of all 133 parameters.

In this respect, the use of a preselected set of simulations as training data set can considerably reduce the computational time required by the ABC-RF with SA. Our results showed that the use of a smaller set of preselected simulations (namely, the 10000 runs closest to the observation data set) did not deteriorate the outcomes of the calibration procedure. In our application, the use of preselected simulations reduced of about 2/3 the computational time of the ABC-RF with SA when compared to the larger set of 25000 simulations.

Finally, results from the sensitivity analysis could be exploited to select a reduced number of parameters to be included in the calibration, further reducing the computational cost of the ABC-RF with SA. Such approach was not yet tested in this work. In this framework, a threshold for the significance of the parameters should be set based on their sensitivity index, and the parameters with lower importance could be subsequently dismissed.

6.8.3 Parameter selection is improved by coupling ABC-RF and SA

The uncertainty in parameter selection
In Bayesian parameter inference, once the posterior probability is retrieved, the selection of the parameter value is not straightforward (Kruschke 2018). Parameters that are particularly consistent with the training data set will have peaked posterior distributions. However, the posterior distributions of some parameters might be uncertain, showing multiple local maxima or being particularly spread out over the parameter space (Kruschke 2018). The options $P_{mix,2}$ and $P_{mix,3}$ described in section 6.4.6 were designed to automatically discriminate between peaked and spread out distributions, facilitating the selection of the parameter value. Among the four options tested for the choice of the parameter values, the results from the application to both the simulated and real observations showed that option P_{med} was the less adapted one.

Through algorithm 2, new information is added at every iteration of the calibration routine (namely, the values of the previously calibrated parameters). After only a few iterations, this has a sharp effect on the shape of the posterior distributions that appear smoother when compared to the standard ABC-RF. This clearly reduces the uncertainty in parameter selection and ultimately represents a clear advantage

for the calibration procedure.

Equifinality

The advantage of reducing uncertainty in parameter selection, however, is negligible during the very first iterations of the algorithm, that deal with the most relevant parameters. This is evident when looking at Fig. 6.5a and 6.5c, that show the error between the real and the estimated parameters in the application to the simulated observations. For some parameters, a considerable error is done. Errors are very low for parameters with medium-high importance (roughly between the 3rd and the 10th positions), while they increase again for various parameters with a low sensibility index. Despite the strong differences between the parameter sets associated to the three model runs (the one originating the simulated observations, and the parameter sets of the two calibration runs in Fig. 6.5), their overall model performances are comparable. Tables 6.3 and 6.4 also show that different parameters sets, originated by calibration runs with different characteristics, can lead to comparable model performances.

The objective of this calibration work was not to retrieve the real values of the model parameters, but rather to identify some parameter sets that drive the model to match the observations in terms of the three variables of interest. Our results suggest that several distinct parameters sets can fit the data equally well. This is known as equifinality: because the model inputs are linked to each other through complex relationships in the model, different sets of model inputs might result in equally plausible behaviours and produce equivalent model outputs (Arhonditsis et al. 2008, Hipsey et al. 2020).

The non-uniqueness of the model solutions in complex ecological models is a known issue (Beck 1987, Arhonditsis et al. 2008, Hipsey et al. 2020). It derives from the fact that the dimensions of the observations D are much smaller than those of the state and parameter vectors (x, θ) (Beck 1987). Automated calibration procedures therefore seek to optimize some dynamics that are of a substantially higher order than what can be observed to describe the system (Beck 1987, Hipsey et al. 2020).

The possibility of running multiple calibration runs, thanks to the rather low computational cost of the ABC methodologies, once the reference table is built, allowed to highlight the existence of multiple parameters sets with equivalent model outputs. This could be ultimately related to the complex structure of the model rather than to the proposed calibration methodology, partly questioning the idea of seeking the optimal parameter set for a biogeochemical model.

6.8.4 A specific working framework

The ABC-RF was applied here to a specific working framework, particularly in terms of the available data set, the parameters to be calibrated and the summary statistics.

Use of high-frequency data

High-frequency data of three variables relevant to aquatic ecology and environmental modelling were available for this work. This allowed to concentrate the computational

effort on a period of roughly two weeks, relatively short for the typical aquatic ecological modelling applications, that often span a few months at least. The main objective was here to test the ability of a complex biogeochemical model to correctly reproduce a specific bloom event, possibly discriminating the biomass between cyanobacteria and other algal species.

Events of this temporal span are often completely missed by traditional limnological monitoring, that is based on periodical sampling or profiling. However, they are extremely important: on the first hand, to test the formulations of biogeochemical models in between subsequent field campaigns, and, on the other hand, for the management of water resources in general and of our study site in particular, where bathing bans need to be issued rapidly in case of presence of cyanobacteria. Reliable short-term model simulations could represent a great advantage in this regard.

Furthermore, the choice of a short simulation period also allowed to contain the computational cost of each model simulation, lightening greatly the application of ABC-based methods from a computational point of view.

Choice of the parameters

A set of 133 parameters are considered in this work for calibration. Namely, they include 114 model parameters and 19 initial conditions. The model configuration implemented for this study includes a total number of parameters that is considerably higher. Previous trial and error tests showed that a large portion of these parameters had a low influence on model results, at least in terms of the three target variables. This could be the case, for instance, of parameters involved in processes not directly affecting the target variables. Three main physical processes were targeted in this calibration: algal physiology (e.g. growth, mortality and sedimentation), oxygen consumption, and nutrient and organic matter evolution. The 114 parameters included in the calibration were selected based on their physical meaning and direct association with processes of interest, and on previous tests that gave a first estimate of their importance for the model.

The choice of including some initial conditions in the calibration is motivated by the fact that all of the 34 variables listed in table 3.1 need to be initialized. However, some of them, such as the four fractions of particulate organic matter, are extremely difficult to measure or estimate, despite their importance in the model. The presence of nutrients in readily available forms or in less accessible compounds clearly influences the outcomes of the model in terms of phytoplankton dynamics and, subsequently, in terms of oxygen concentration.

In our application, the available data did not allow to estimate without uncertainty the concentrations of the different fractions of nutrients as modelled by BLOOM, that were therefore included in the calibration. Furthermore, even when measurements are available, they might be affected by a grade of uncertainty that justifies their calibration. For instance, the scattering that characterizes high-frequency measurements of oxygen and cyanobacteria concentrations introduces a rate of uncertainty in the estimation of a single initial value. For this reason these two initial conditions were also included in the list of parameters to be calibrated.

The sensitivity analysis shows, however, that most of the parameters under consideration have a negligible influence on the model outcomes, with 10 (20, respectively) parameters accounting for around 70% (80%, respectively) of the

overall variability. From the sensitivity analysis, nitrogen is the most important nutrient in the system. In particular, its distribution among the dissolved and fast-decomposing particulate organic fractions was of particular importance for model results. The calibration of the initial conditions, when their values are uncertain, can sensibly improve model results, and, in the light of the model formulation, give new insights on the functioning of the system.

The physiological parameters with greater importance were those directly involved in the equation modelling growth (i.e. the coefficients for temperature dependence of growth and the potential growth rates).

Choice of the summary statistics

The choice of the summary statistics is crucial to ABC. To the best of our knowledge, in all applications of ABC, the summary statistics do not depend on the observations. Generally, the set of simulations is used by the ABC approaches to generate an inverse model which is intended to be applied to several sets of observations to estimate the associated parameters. This is a great advantage of the ABC approaches. However, this does not apply to any model. In our case for instance, the set of simulations depends on the meteorological conditions forcing the model, that are specific to the period under consideration. In such case, the inverse model generated by the ABC method is indeed specific and remains valid only for the period under consideration.

It was therefore possible in our case to use summary statistics that directly depend on the observations, such as R and NMSE, without loss of generality. In our application, the summary statistics ultimately serve as objective functions to be minimized through the ABC. Other options, independent from the observations, for the summary statistics were tested to describe the time series (e.g. series of successive means, splines projection coefficients). However, the use of error metrics as summary statistics proved to be the most effective.

Finally, this particular framework also justifies the use of a preselected subset of simulations to run the ABC, that we suggested. As the simulation subset can only be used to calibrate the model on a specific series of observations, it is possible to rule out simulations that deviated from the target behaviour without loss of generality.

6.8.5 Analysis of the performance of the model BLOOM

The model calibrated on the real observation data set reproduces the general behaviour of the observations very well in terms of all the three target variables on the selected period. However, the observations show strong daily cycles that are not fully reproduced by the model. This is probably due to the model structure, rather than to the calibration methodology.

On the one hand, complex biogeochemical models are conceived focusing on dynamics that span longer periods than the one simulated here (i.e., monthly to seasonal). Moreover, the configuration set up for this work, that is the ensemble of substances and processes activated in the model, might not be the optimal one.

Very complex and very simple models are available to describe the biogeochemical cycle in aquatic ecosystems, a system that cannot be fully validated due to the lack of data (Ward et al. 2013, Kriest 2017). A broad literature has already addressed their

advantages and drawbacks (e.g. Anderson 2005, Raick et al. 2006, Ward et al. 2013, Kriest 2017). In this respect, our configuration, even though complex, only describes a portion of the real natural ecosystem. For instance, benthic processes, macrophytes and zooplankton are not explicitly included. However, as in all modelling works, the challenge is to find the good level of complexity to explain the dynamics under consideration.

In this work, the comparison with three variables recorded at high-frequency shows that short-term phytoplankton bloom can be simulated by a model with relatively basic processes that can be easily measured (i.e. growth, mortality, nutrient uptake, oxygen production and organic matter decomposition). However, without additional data it is not possible to evaluate the correctness of other important simulated processes such as the mineralization of organic matter. This issue was highlighted to some extent by calibrating the model using only one variable: the total chlorophyll, which was chosen because of its importance for the management of aquatic ecosystems (World Health Organization 2003). After calibration with total chlorophyll data only, model results were only slightly improved in terms of the target variable in respect to the best calibration run using all the three variables available. On the other hand, the dynamics in terms of oxygen and cyanobacteria concentration were extremely inaccurate. This result highlights the importance of gathering the widest possible range of data to evaluate the performance of a model that aims at describing a complex natural system. Only the comparison with multiple variables make it possible to determine if the global functioning of the system under consideration is captured as a whole and not only in terms of a few single variables (e.g. Ward et al. 2013, Kriest 2017).

6.9 Conclusion

Biogeochemical models are often highly parametrized and complex. Their calibration is challenging and often overlooked in scientific literature. Our study shows that, among the various techniques available for automated calibration, ABC-RF can be successfully applied to calibrate a complex and highly-parametrized biogeochemical model. Our work focuses on a short-term algal bloom, an event that would possibly be missed by traditional periodical survey. After calibration, the model was able to reproduce biogeochemical dynamics that span a relatively short period. The overall phytoplankton growth and mortality were correctly simulated, as well as cyanobacteria and oxygen concentration. However, daily cycles and other dynamics at shorter time-scale were not fully reproduced.

To obtain such results, the coupling of ABC-RF with sensitivity analysis through algorithm 2 was crucial, so as the availability of high-frequency data. Indeed, the main computational effort required by ABC is for the generation of the reference table, that should be of at least 25000 simulations for the optimization of more than 100 parameters. The computational cost of the ABC algorithm itself (once the simulations were performed) can be reduced by preselecting 10000 simulations from the complete stock.

The summary statistics were defined here based on the modeller expertise, an approach shared with most applications of ABC so far (e.g. Lagarrigues et al. 2015, Dominguez Almela et al. 2020). This highlights the importance of the modeller's

experience and knowledge, that remains an essential feature in the Bayesian approach to parameter inference, and that should not be dismissed also when applying automated calibration strategies.

In optimization techniques such as local gradient-based methods or global optimization algorithms, the exploration of the parameters space depends on the initial conditions. This is not the case for ABC, where the parameters space is explored based on the prior distributions defined by the user. In this respect, ABC could also be a useful technique to define appropriate initial conditions for the application of other calibration algorithms (e.g. seeking the parameters optimum).

Finally, care should also be taken when approaching the calibration of a complex aquatic ecological model in evaluating the data available for the procedure. The use of multiple variables considerably improves the overall performance of the model.

Chapter 7

Implementation of a newly coupled hydro-biogeochemical model through short-term and seasonal simulations

This chapter describes the implementation and calibration of the recently coupled models TELEMAC3D and Aquatic EcoDynamics (AED2) on the study site. This work has two main objectives: (i) to test the coupling and the feedbacks between the two models in a natural ecosystem against *in situ* observations, and (ii) to analyse the performance of a second and different approach to biogeochemical modelling over different time-scales.

The applications presented in this chapter are among the first ones to test the coupling between this two models in a natural environment. In fact, the coding of the coupling between the models started in 2015, and is currently still evolving. All the results shown in this chapter are obtained through the version 8.1.2 of TELEMAC for Linux, which was released in October 2020 and includes the coupling with AED2 (`http://svn.opentelemac.org/svn/opentelemac/tags/v8p1r2`).

The implementation and calibration of the two coupled models for Lake Champs-sur-Marne proceeded through successive stages. First, the hydrodynamic model was tested in its stand-alone version in order to examine its ability to reproduce the dynamics of water temperature, mixing and stratification of a shallow polymictic lake. High-frequency temperature data allowed to carry out such tests on short-term simulations of two-weeks representative periods. Secondly, the coupling with the biogeochemical library was activated and a first calibration of the AED2 parameters was performed on short periods of roughly two weeks. Once a robust configuration was achieved, the coupled models were applied to a longer period and the parameters recalibrated, with the objective to simulate the seasonal phytoplanktonic dynamics recorded in the study site between late winter and summer.

Table 7.1 – Short-term calibration and validation periods, and list of variables used to evaluate model performances.

Year	Start	End	Use	H.F. data set	Layer	Site
2015	13 Jul.	27 Jul.	Valid.	Water temp.	Surf., Bott.	B
				Dissolved oxygen	Mid.	
				Total chl.	Mid.	
2016	14 Jul.	30 Jul.	Valid.	Water temp.	Surf., Bott.	B
				Dissolved oxygen	Mid.	
				Total chl.	Mid.	
2018	25 Jul.	8 Aug.	Calib.	Water temp.	Surf., Bott.	B
				Dissolved oxygen	Mid.	
				Total chl.	Mid.	
				Cyanobacteria	Mid.	

7.1 Short-term applications

In this section, the first set-up of the hydrodynamic and biogeochemical models on the study site is described. This work was carried out on short-term periods of roughly two weeks (described in section 7.1.1), in order to reduce the computational time while adjustng the models configurations. Namely, the hydrodynamic 3D model TELEMAC3D was first tested and calibrated in its stand-alone version. This is described in section 7.1.2. Afterwards, the coupling with the biogeochemical model AED2 was activated. The set up and the calibration of AED2 is described in section 7.1.3.

7.1.1 Simulation periods and dataset

The tests described in this section were carried out on three distinct periods of roughly two weeks in the Summer of the years 2015, 2016 and 2018. The three periods are described in table 7.1, together with the high-frequency data available to evaluate model performances. The three periods listed in table 7.1 were used for the calibration and validation of both TELEMAC3D in its stand-alone version (using only water temperature data, see section 7.1.2), and TELEMAC3D coupled with AED2 (using all the variables listed in table 7.1, see section 7.1.3).

The selection of the simulation periods was done based on two main factors: (i) all periods are characterised by the presence of thermal stratification, and (ii) all periods show an increase in total chlorophyll concentration, even though with different rates. The period going from the 25 of July to the 10th of August 2018 was used for the calibration of the model parameters. This is the same simulation period analysed in chapter 6 and used for the automated calibration of the Delft3D-BLOOM model. The remaining two periods were used for validation.

High-frequency observations of water temperature constituted the data set used to initialize and evaluate thermal performance of the model TELEMAC3D, both in its stand-alone and in its coupled version. The biogeochemical model was tested and initialized in terms of dissolved oxygen and total chlorophyll concentration based on the high-frequency observations recorded at site B for the calibration period, and at

Table 7.2 – Short-term simulations: initial conditions for the inorganic nutrients.

	Summer 2015 (validation)	Summer 2016 (validation)	Summer 2018 (calibration)
NH_4 [mmol m^{-3}]	8.0	8.0	2.0
NO_3 [mmol m^{-3}]	10.0	15.0	5.5
PO_4 [mmol m^{-3}]	2.3	2.3	0.3

site A for the two validation periods. Cyanobacteria concentration data were only available for the calibration period (i.e. 2018) via the phycocyanin sensor at site B.

Water samples were taken on Lake Champs-sur-Marne at site B on the initial day of the calibration period. The subsequent laboratory analysis allowed to directly set the initial condition of the three inorganic nutrients (ammonium, nitrate and phosphate, table 7.2), and to obtain the concentrations of total nitrogen and phosphorus (in dissolved inorganic and organic form) in the study site. This allowed to set the initial conditions for dissolved nitrogen and phosphorus, that were also used as a first approximation of the relative particulate forms.

Previous studies carried out at LEESU already analysed the two validation periods through model simulations with Delft3D-BLOOM (Plec et al. 2017). Such work was used as a reference to set the initial conditions of the three inorganic nutrients (ammonium, nitrate and phosphate), which are listed in table 7.2.

7.1.2 Set up of the stand-alone hydrodynamic model

This section describes the first set-up, calibration and validation of TELEMAC3D for the study site. The main objective is the calibration and validation of the thermo-hydrodynamic model in terms of water temperature and thermal stratification. This was done in order to set up a configuration for TELEMAC3D accurate enough to allow the activation of the coupling with the biogeochemical model.

7.1.2.1 Model configuration

The computational domain for the study site was built with the open-source software BlueKenue™ (Canadian Hydraulic Centre 2011), and is represented in Figure 7.1. It consists of a triangular grid with an average distance between the nodes of 20 m, and a refined zone around the narrower portion of the water body. The raw bathymetric data available from the 2014's echo-sounder campaign were here translated according to the local elevation. The mesh is composed of 404 nodes (661 elements), with 10 σ-layers for the discretization on the vertical axis and a uniform water level set at an elevation of 40 m a.s.l..

In the heat budget, the contribution of precipitation was neglected both in terms of energy and mass, while for evaporation only the mass flux was neglected. Preliminary tests showed that the thermal stratification of a polymyctic water body as the study site was better modelled through a mixing length turbulence model than through the k-ε turbulence model. After such preliminary tests, Nezu and Nakagawa's formulation of the mixing length model with Viollet's damping function

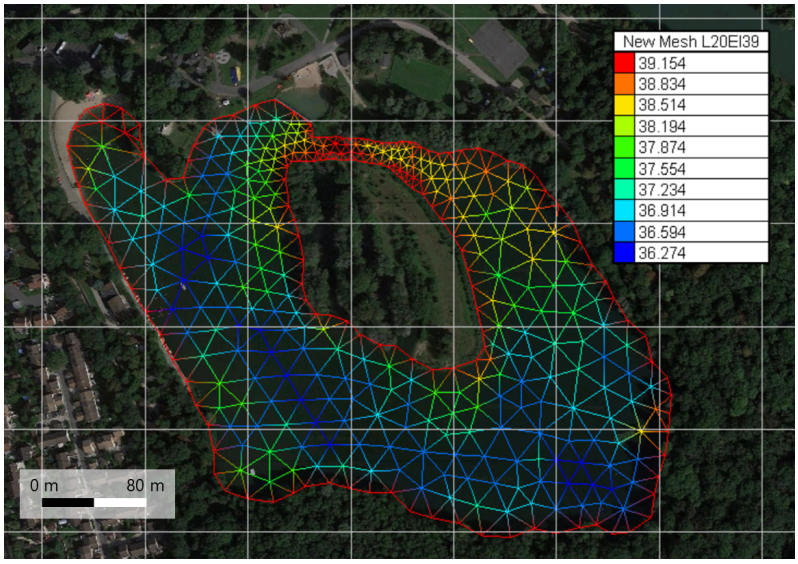


Figure 7.1 – Computational domain for Lake Champs-sur-Marne used for the first calibration and validation tests with the coupled TELEMAC3D - AED2. The triangular mesh is obtained through the software BlueKenue™ (Canadian Hydraulic Centre 2011). The mean distance between its elements is of 20 m and the bathymetric data were automatically interpolated by the software. The elevation of the lake bottom as interpolated by BlueKenue™ is represented through a colour chart. The mesh was refined around the narrower portion of the study site.

was implemented for vertical turbulence closure (Nezu & Nakagawa 1994, Viollet 1988, Telemac modelling system 2016). The molecular diffusivity of water is used on the vertical as a background value by the model, and was set to 10^{-6} m^2 s^{-1}. The choice of the mixing length model entails a constant model for horizontal turbulence.

Among the two heat-exchange models available in TELEMAC3D to compute the heat transfer at the air-water interface, the more detailed one was selected (see section 3.2.3.3). The model is forced with six meteorological variables: relative humidity [-], air temperature [°C], cloud cover [-], net solar radiation [J s^{-1} m^{-2}], wind speed [m s^{-1}] and direction [°N]. Their values were downloaded from the closest meteorological station, located at the Orly airport and described in section 3.1.4. Finally, the model was run with a 60 s time step, and its outputs were saved with a one-hour time step.

Data measured at site B were used to initialize the model in terms of water temperature. No data were available to set the initial velocity field, and the model was therefore initialized with water at rest.

7.1.2.2 Model calibration and validation

The calibration of TELEMAC3D was done by trial and error, based on high-frequency water temperature data at the surface and bottom layers (0.5 and 2.5 m depth, respectively), at measuring site B (see section 3.1.1). The root mean square error (RMSE) was calculated between model results and observations to evaluate model performances during both calibration and validation. For this purpose, high-

frequency data were first averaged hourly according the model output time step, and cleaned from the potential outliers originated by periodic sensor maintenance. The latter were defined as sudden water temperature variations (> 1°C) over the ten minutes separating two successive measurements, and consequently erased.

Preliminary tests allowed to set the model configuration described in section 7.1.2.1. The calibration of the model was further pursued focusing mainly on three parameters: (i) the constant horizontal diffusivity defining the horizontal turbulence model, (ii) water transparency (via the parameter Secchi depth [m]), and (iii) the coefficient b [-] of the wind correction function (see equation (3.20)) involved in the atmosphere-water heat exchange.

7.1.2.3 Model results

Multiple calibration runs were carried out varying the values for the Secchi depth and the coefficient b [-]. Field values for the Secchi depth in Lake Champs-sur-Marne are available since 2015. They vary between 0.5 and 3 m with a mean value of 1.17 m in July-August 2018; using this range, the Secchi depth parameter was calibrated and finally set to 0.85 m. The suggested range of variability for coefficient b is between 0.0017 and 0.0035 (Telemac modelling system 2016). Here, b was eventually set to 0.00135. The constant horizontal diffusivity was calibrated based on the computational mesh size and on model results, and was finally set to 0.1 m^2 s^{-1}.

Model performances were evaluated in terms of water temperature at the surface and bottom layers through the RMSE with observations at site B. RMSE values are presented in table 7.3. During the calibration period, the model correctly simulated water temperature both for the surface and bottom layers, with RMSE values of 0.6°C for both layers.

The first validation period (summer 2015) shows the lowest RMSE values (namely, 0.3°C for the surface layer and 0.6°C for the bottom layer). The highest error made by the model was detected for the bottom layer during the second validation period (2016, RMSE=1.4°C), while the surface layers was correctly simulated also in this case (RMSE=0.7°C).

The results of the calibrated model are shown in Figure 7.2 for the three simulation periods: panel c is relative to the calibration period, while panels a and b are relative to the two validation periods. Surface (blue) and bottom (red) water temperatures are plotted for both simulation results (solid lines) and high frequency observations (dotted lines). The Figure shows that surface water temperature is very accurately simulated by TELEMAC3D, as confirmed by the low RMSE values in table 7.3, while bottom water temperature is often overestimated.

This impacts the ability of the model to correctly reproduce the alternation between mixing and thermal stratification in a shallow water body. Through a constant Secchi depth value, the model fails to fully reproduce the strength of thermal stratification during the periods under consideration. Furthermore, exchanges with ground water are neglected in the model configuration but could also sensibly impact bottom water temperature in the real system. Their effect therefore is not taken into acount by the model. In particular, as shown in Figure 7.2b, for the validation period in 2016 surface water temperature is correctly simulated, while bottom water temperature is overestimated by the model. This causes an anticipated breakdown

Table 7.3 – RMSE values between TELEMAC3D stand-alone and high-frequency observations of water temperature for the three simulation periods.

	Summer 2015 (validation)	Summer 2016 (validation)	Summer 2018 (calibration)
RMSE surface [°C]	0.3	0.7	0.6
RMSE bottom [°C]	0.6	1.4	0.6

of thermal stratification when compared to high-frequency observations. However, daily stratification events (i.e. water column stratified during the day and overturned at night) are correctly reproduced by the configuration set up for TELEMAC3D. This could also be partially due to an excessive numerical diffusion introduced by the numerical schemes implemented in TELEMAC3D.

During the three simulation periods, the overall behaviour of water temperature in the study site is correctly reproduced by the model, with low RMSE values (generally < 1°C), close to those found in the literature. For this reason, the configuration described in this section was considered adapted to test the coupling with the biogeochemical model.

The calibration period (25/07/2018 – 10/08/2018) was also simulated with the hydrodynamic model Delft3D-FLOW in the context of the automated calibration of the biogeochemical model Delft3D-BLOOM (see chapter 6, and section 6.3 in particular). The outcomes of Delft3D-FLOW in terms of surface and bottom water temperature are shown in Figure 6.1a, and are repeated here in panel d of Figure 7.2 for an easier comparison with the results obtained with TELEMAC3D (Fig. 7.2c). The performances of the two models are very close for both the surface (RMSE values of 0.6°C for both models) and bottom (RMSE of 0.6°C for TELEMAC3D, and of 0.5°C for Delft3D) layers.

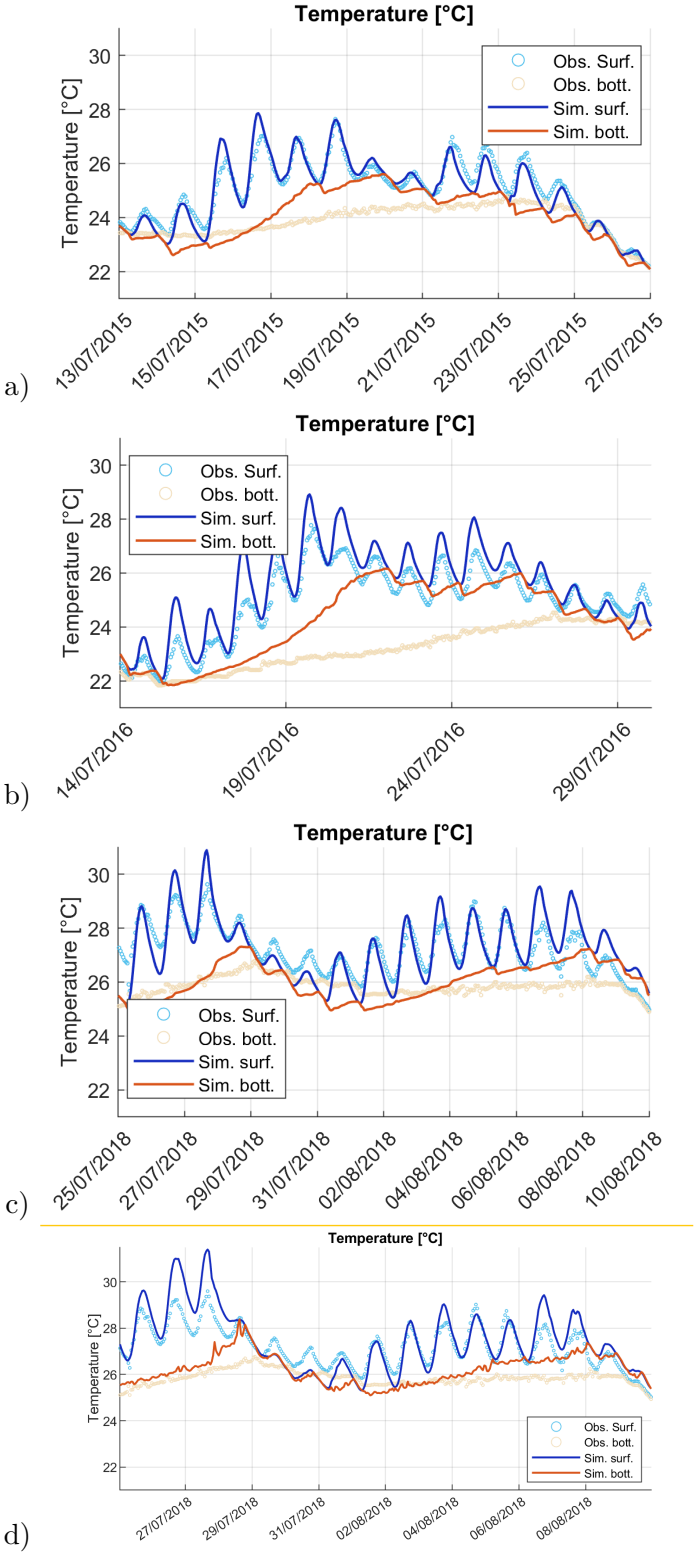


Figure 7.2 – Panels a, b and c: model results (solid lines) and high-frequency observations (dotted lines) of surface and bottom water temperature at site B during the calibration and validation periods of the TELEMAC3D hydrodynamic model. Panel c is relative to the calibration period, and panels a and b to the validation periods. Panel d is visually separated by an orange line and shows, for comparison, the results obtained with Delft3D-FLOW over the 2018 simulation period; this is the same graph shown in Figure 6.1a. Blue lines represent surface water temperature, red lines bottom water temperature.

7.1.3 Coupling with the biogeochemical model

The same simulation periods presented in section 7.1.1 were selected for the initial set up of the coupled models. In particular, the period in the Summer of 2018 (25/07/2018 − 10/08/2018) was used for calibration, and the remaining two simulation periods for validation.

Here the calibration only concerned the biogeochemical model parameters. The tests carried out for the calibration were done through trial and error, and involved variations both on the structure of the model (i.e. the activated modules) and on the parameters values.

7.1.3.1 Model configuration

Hydrodynamic model and computational domain
The configuration of the hydrodynamic model TELEMAC3D is the one described in section 7.1.2.1. The coupled models TELEMAC3D and AED2 share the same computational grid (see Figure 7.1), computational time step (60 s) and output time step (1 h).

Due to the dynamic calculation of the light extinction coefficient included in the coupled models (see section 3.2.4.3), here the Secchi depth does not need to be defined by the user.

Biogeochemical model
Aquatic EcoDynamics (AED2) is a modular biogeochemical library, whose grade of complexity can be set by the user through the activation of the desired modules. Its general description can be found in section 3.2.4. For the application described in this section, six modules were activated:

- Phytoplankton
- Oxygen
- Phosphorus
- Nitrogen
- Organic matter (particulate and dissolved)
- Sediment fluxes

Two phytoplankton groups typically present in the study site during the summer season are activated in the phytoplankton module: green algae and cyanobacteria. According to the modules activated, the 12 variables listed in table 7.4 are simulated by the model. Such configuration includes a total of 75 parameters.

7.1.3.2 Model calibration and validation

The calibration of the coupled TELEMAC3D and AED2 was based on high-frequency measurements at site B. Water temperature data at the surface and bottom layers were used to calibrate and validate the simulated thermal dynamics.

Table 7.4 – Modules and variables activated in the configuration of the biogeochemical model AED2 for the short-term tests.

Module	**Variables**
Oxygen	Dissolved oxygen
Phosphorus	Ortho-phosphate
Nitrogen	Ammonium
	Nitrate
Organic matter	POC
	PON
	POP
	DOC
	DON
	DOP
Phytoplankton	Cyanobacteria
	Green algae

The light extinction coefficient used in the hydrodynamic model is computed at each time step by AED2. Therefore, the feedback of the biogeochemical model on the hydrodynamic model was also evaluated. High-frequency data of dissolved oxygen, total chlorophyll and, when available (i.e. summer of 2018) phycocyanin, were used to calibrate and validate the biogeochemical model.

The root mean square error (RMSE) was calculated between model results and observations to evaluate model performances. For this purpose, high-frequency data were smoothed and cleaned from the outliers originated by periodic sensor maintenance as described in section 4.2, and averaged hourly to match the model output time step.

During the initial calibration tests, most of the 75 paramters activated in the biogeochemical model were varied in order to gain knowledge of their importance for the simulation results. Eventually, the calibration effort mainly focused on a reduced number of parameters, the most sensitive for the model in terms of the variables of interest. These parameters are: the phytoplankton potential growth rates at 20°C, the optimum and maximum temperatures, the half saturation constant for light limitation of growth and specific light attenuation coefficients, the sediment oxygen demand, and the specific light attenuation coefficients of particulate and dissolved organic carbon. Such parameters, listed in table 7.5, were calibrated through trial and error based on literature values (e.g. Hipsey et al. 2013, Fenocchi et al. 2019).

In the stand-alone version of TELEMAC3D, the light extinction coefficient K_d is calculated based on the Secchi depth (S) defined by the user through the equation:

$$K_d = 1.7/S \tag{7.1}$$

The calibration of the stand-alone version of TELEMAC3D (section 7.1.2) lead to an optimal value for the Secchi depth of 0.85 m (i.e., through equation (7.1), $K_{d_{T3D}} = 2\ \mathrm{m}^{-1}$).

When the coupling with AED2 is activated, K_d is calculated dynamically, through the following equation (see section 3.2.4.3 for further details):

Table 7.5 – List of the most sensitive AED2 parameters, and corresponding calibrated values (relatively to the short-term tests).

Parameter	**Value**	**Unit**
Sediment O_2 demand	100	mmol O m^{-2} d^{-1}
DOC light ext. coeff.	0.0003	m^2 $mmolC^{-1}$
POC light ext. coeff.	0.003	m^2 $mmolC^{-1}$
Water light ext. coeff.	0.5	m^{-1}

	Green algae	**Cyanobacteria**	**Unit**
Growth rate at 20°C	1.5	1.2	d^{-1}
Optimum water temp.	25	28	°C
Maximum water temp.	35	36	°C
Specific light ext. coeff.	0.005	0.005	m^2 mmol C^{-1}
Half saturation constant light limitation	10	10	μE m^{-2} s^{-1}

$$K_d = K_w + K_{e,DOC}DOC + K_{e,POC}POC + \sum_{i}^{N_{PHY}} K_{e_i}PHY_{C_i} \tag{7.2}$$

In this respect, the concentrations of phytoplankton and dissolved (DOC) and particulate organic carbon (POC) have a strong feedback on the hydrodynamic model, and in particular on the way heat is transferred over the water column. However, at this stage of the work with the coupled models TELEMAC3D and AED2, no data were available on the study site to set the initial conditions for particulate and dissolved organic carbon. Their initial concentrations were here estimated in order to obtain, through equation (7.2), an initial value of the light extinction coefficient of 2 m^{-1}, equal to the one used in the stand-alone version of TELEMAC3D. This was done based on the initial conditions set for the two phytoplankton groups, and on the values set for the specific light extinction coefficients (listed in table 7.5).

The obtained initial conditions values for dissolved and particulate organic carbon are, respectively, of 9.6 and 4.8 mg C l^{-3}, and were used for both the calibration and validation periods. DOC measurements were later available on the study site starting from the end of 2018. DOC concentration is discussed, for the year 2019, in section 7.2. The value estimated here for DOC is only slightly overestimated when compared to the values observed in the summer of 2019 (around 8 mg C l^{-3}).

Table 7.6 – RMSE values between observations and results of the coupled models TELEMAC3D and AED2; calibration period, summer 2018.

	Water Temp. [°C]	Dissolved Oxygen [g O m^{-3}]	Cyanobacteria [μg Chl l^{-1}]	Total Chl. [μg Chl l^{-1}]
Surface	0.7	-	-	-
Middle	0.6	1.8	3.1	3.7
Bottom	0.6	-	-	-

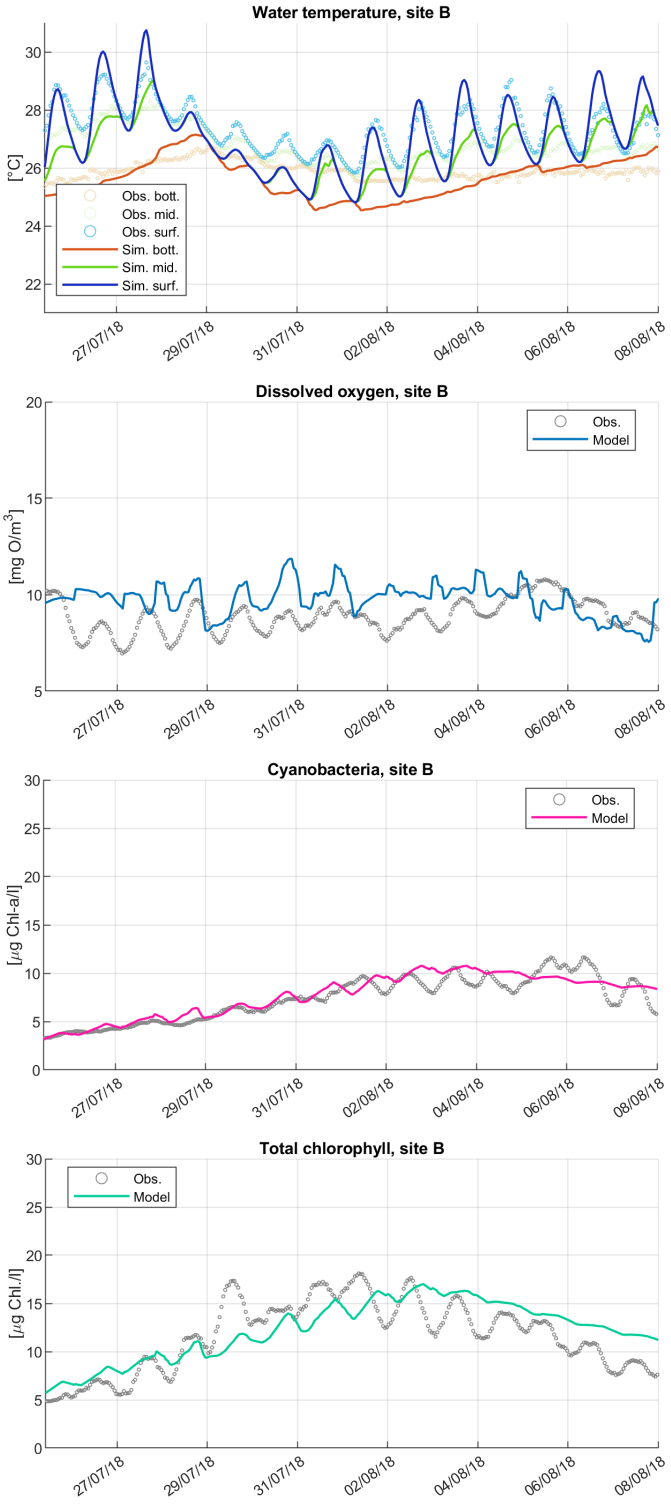
7.1.3.3 Model results

Calibration period

Numerous model runs were carried out on the calibration period to test the sensitivity of the AED2 parameters and calibrate their values. The results of the model during the calibration period are shown for measuring site B in Figure 7.3, along with high-frequency observations. RMSE values were calculated for water temperature at the surface and bottom layers, and for dissolved oxygen, cyanobacteria and total chlorophyll concentrations at the middle layer (7.6).

As shown by Figure 7.3a, surface and bottom water temperature are very well reproduced by the model, with RMSE values of respectively 0.7 and 0.6°C. Such values are very close to those found in section 7.1.2 for the hydrodynamic model alone, which were of 0.6°C for both the surface and bottom layers. Water temperature at the middle layer is also well reproduced by the model (green line in Fig. 7.3a), with an RMSE of 0.6°C. The slight underestimation of water temperature between the 28/07/2018 and the 31/08/2018 can be explained by the presence of strong peaks of wind speed relatively to the mean wind speed over the period. With the exception of the period between the 28th and the 31st of July, the timing and the values of the daily temperature maxima in the surface layer are correctly reproduced, as are the nighttime mixing events. Overall, the coupled models capture very well the alternation between stable stratification, mixing, and daily stratification.

Also the three remaining variables that could be compared with high-frequency data are well simulated by the model. The RMSE for dissolved oxygen is of 1.8 mg O m^{-3}. Figure 7.3b shows that the model reproduces correctly the overall order of magnitude of dissolved oxygen concentration, even though some discrepancies are present between observations and model results in the daily cycle. Such differences can be associated with the strongly stratified profile simulated by the model for dissolved oxygen (shown in Figure 7.4). Under such stratified conditions, a small error in the water depth at which model results are extracted can result in an amplified bias with high-frequency observations recorded at 1.5 m depth. Total chlorophyll is well captured by the model (Fig. 7.3d). The corresponding RMSE between model results and high-frequency observations is of 3.7 μg Chl. l^{-1}, or, in terms of relative RMSE (RRMSE): 32% of the average total chlorophyll observed over the period. RRMSE values lower than 100% are usually considered satisfactory. In particular, the growing phase, as well as the timing and the concentration of the peak of total chlorophyll are very well reproduced. During the descending phase, total chlorophyll concentration is slightly overestimated. Biomass is also correctly distributed among the two simulated phytoplankton groups. Cyanobacteria



a)

b)

c)

d)

Figure 7.3 – Model results (solid lines) and high-frequency observations (dotted lines) at site B for water temperature (a), dissolved oxygen (b), cyanobacteria (c), and total chlorophyll (d), relatively to the calibration period in the summer of 2018. In panel a, blue lines represent surface water temperature, green lines middle water temperature, and red lines bottom water temperature; in the remaining panels, both observations and model results are relatively to the middle layer of the water column at site B.

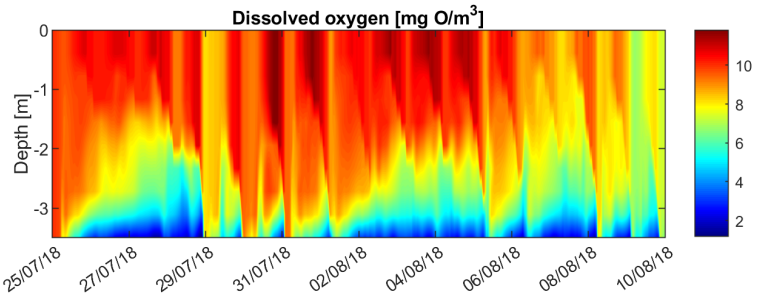


Figure 7.4 – Simulated dissolved oxygen concentration during the calibration period (summer 2018) along the water column at site B.

concentration is very well reproduced, with an RMSE value of 3.1 μg Chl. l^{-1}.

Validation periods

The calibrated model was then tested on the two validation periods: between the 13th and the 27th of July 2015, and between the 14th and the 30th of July 2016. Its results in terms of surface and bottom water temperature, dissolved oxygen and total chlorophyll concentrations are presented in Figure 7.5 for the first validation period in 2015, and in Figure 7.6 for the second validation period in 2016. The figures show both model results (solid lines) and high-frequency observations (dotted lines) at site A.

Figures 7.5a and 7.6a show that both surface and bottom water temperatures are well reproduced by the model. In particular, for the year 2016, both bottom and surface water temperatures are slightly better simulated when compared to the results obtained with the stand-alone version of the hydrodynamic model (Fig. 7.2b). Concerning the validation run in 2015 (Fig. 7.5), both total chlorophyll and dissolved oxygen concentrations are very well simulated by the model, even though the daily oscillation recorded by the high-frequency observation are not fully reproduced by the biogeochemical model.

The total chlorophyll concentration recorded at high-frequency during the validation period in 2016 is sensibly weaker than for 2015. In the validation run, total chlorophyll concentration is largely overestimated by the model with respect to the high-frequency observations (see Figure 7.6c). Conversely, observations of dissolved oxygen are well reproduced by the model (Figure 7.6b). The strong increase observed for dissolved oxygen during the simulation period might indicate that a stronger phytoplankton growth was actually occurring on the study site, without being correctly captured by the chlorophyll optical sensor. This hypothesis is confirmed by total chlorophyll data from two profiles recorded at site A during the simulation period, namely on the 18th and 27th of July 2016. The profiles were averaged between the 1 m and 2 m of depth in order to compare their data with the high-frequency observation at 1.5 m of depth. Such data are also plotted in panel c of Figure 7.6 and match closely with the simulation results.

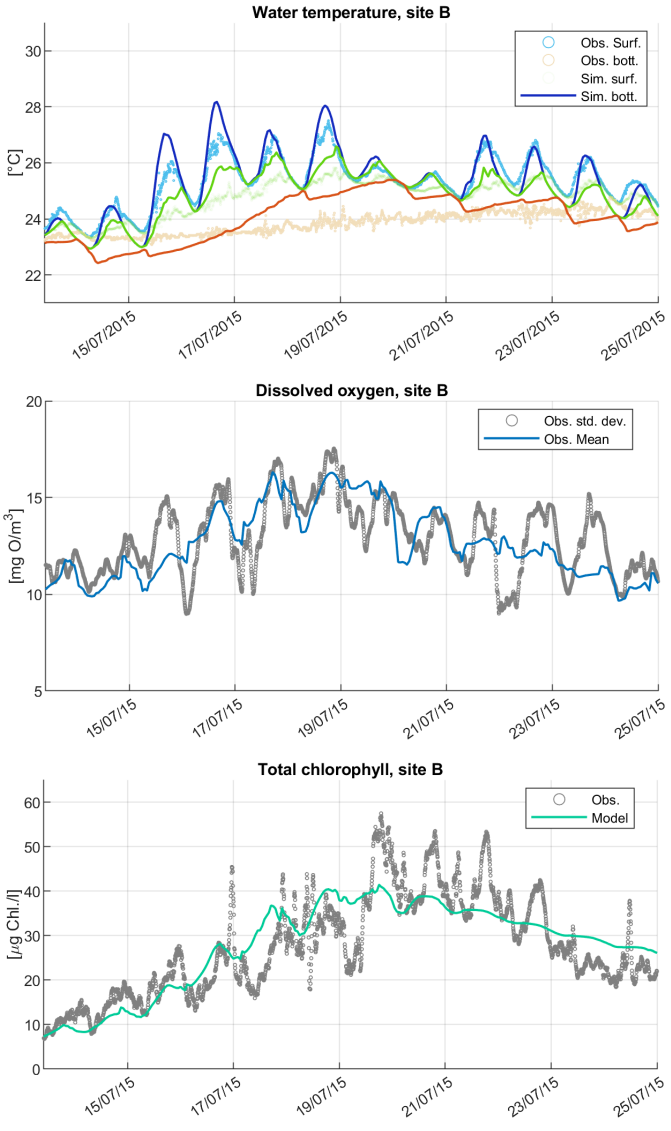
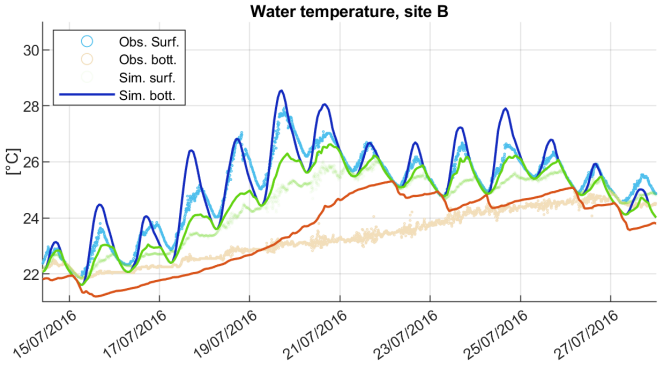
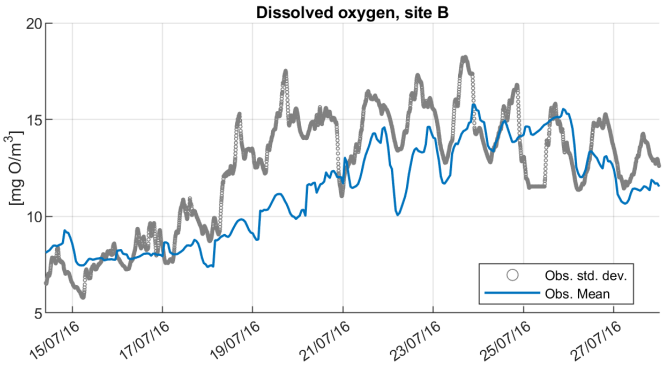


Figure 7.5 – Model results (solid lines) and high-frequency observations (dotted lines) at site A for water temperature (a), dissolved oxygen (b), and total chlorophyll (c), relatively to the first validation period in the summer of 2015. In panel a, blue lines represent surface water temperature, green lines middle water temperature, and red lines bottom water temperature; in the remaining panels, both observations and model results are relatively to the middle layer of the water column.

a)



b)



c)

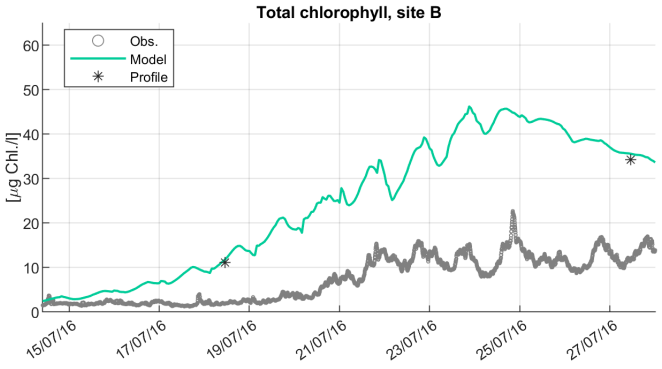


Figure 7.6 – Comparison of model results (solid lines) and high-frequency observations (dotted lines) at site A for water temperature (a), dissolved oxygen (b), and total chlorophyll (c), relatively to the second validation period in the summer of 2016. In panel a, blue lines represent surface water temperature, green lines middle water temperature, and red lines bottom water temperature; in the remaining panels, both observations and model results are relatively to the middle layer of the water column. Profiles data at 1.5 m depth are reported in panel c for total chlorophyll.

7.1.3.4 Spatial analysis

Model results from the calibration period were further analysed based on their spatial distribution over the study site. This analysis focused on two variables: surface total chlorophyll concentration, and water temperature difference between the surface and bottom layers (ΔT). The latter was calculated on the whole computational domain using the first and the last layers of the vertical σ discretization. Similarly to what discussed in chapter 3, the temperature difference between the lake surface and bottom layers is used here as a proxy of the strength of thermal stratification.

The results of this analysis are shown in Figure 7.7 for thermal stratification and in figure 7.8 for total chlorophyll. Both figures are constituted of 15 panels (from "a" to "o"), each of which represents the variable under consideration over the computational domain at midday (i.e. 12 a.m.) for all the 15 days of simulation. Each panel in figures 7.7 and 7.8 therefore represents one time step of model results. The panels in Figure 7.8 have different scale limits, as the main focus of this Figure is to discuss the spatial patterns, rather then to show the overall time-evolution of total chlorophyll (already presented in Figure 7.3d).

As shown in Figure 7.7, thermal stratification (i.e. ΔT>1°C, similarly to chapter 3) at midday is largely present on the study site during most of the simulation period. The overall ΔT at midday averaged spatially over the whole domain is comprised between 0.2°C (on simulation day five) and 1.2°C (on simulation day three). During all simulated days, thermal stratification is present at midday at least in the deeper south-eastern portion of the lake, the only exception being day four (Fig. 7.7d). Similarly to what was discussed in chapter 3, strong spatial variability in the distribution of water temperature difference is induced by the bathymetry. This is true both when thermal stratification is strong (panels a, b, c and m of Figure 7.7), and when the overall average ΔT is the lower (panels d, e, f and o).

The concentration of total chlorophyll on the surface layer shows a lower spatial variability than thermal stratification. The normalized standard deviation associated with the spatial distribution of surface total chlorophyll from the panels in Figure 7.8 is comprised between 0.6% and 5.8% of the corresponding instantaneous overall spatial average. The highest concentrations of total chlorophyll are usually found in the northern shallower and mixed part of the basin, that appears particularly favourable for the phytoplankton growth and accumulation in the model. The spatial distribution of total chlorophyll on the surface layer is mainly induced by the velocity field over the lake surface. During the simulation period, simulated surface horizontal currents are weak but present (around 1 cm/s on average). The prevailing wind direction during the simulated period is a northern wind, in particular for the simulation days one to three (panels a to c in Figure 7.8) and seven to 14 (panels g to n). This creates a pattern of surface currents similar to the one shown in Figure 7.9a for simulation day ten. Figure 7.9 represents the spatial distribution of surface total chlorophyll concentration at midday over the domain (as for Figure 7.8), together with the vectors representing the velocity field at the surface layer at the same time step. Only simulation days ten (panel a) and 15 (panel b) are represented in Figure 7.9. The value of the strongest vectors in Figure 7.9 is of the order of a few cm s^{-1}.

Figure 7.9a shows how the higher biomass produced in the northern portion of the lake gets transported following the simulated surface velocity field. During

the last day of the simulation, the prevailing wind direction changes to a South or South-West wind. This originates a different field of currents (see Figure 7.9b), that reinforces the accumulation of higher biomass towards the northern part of the lake.

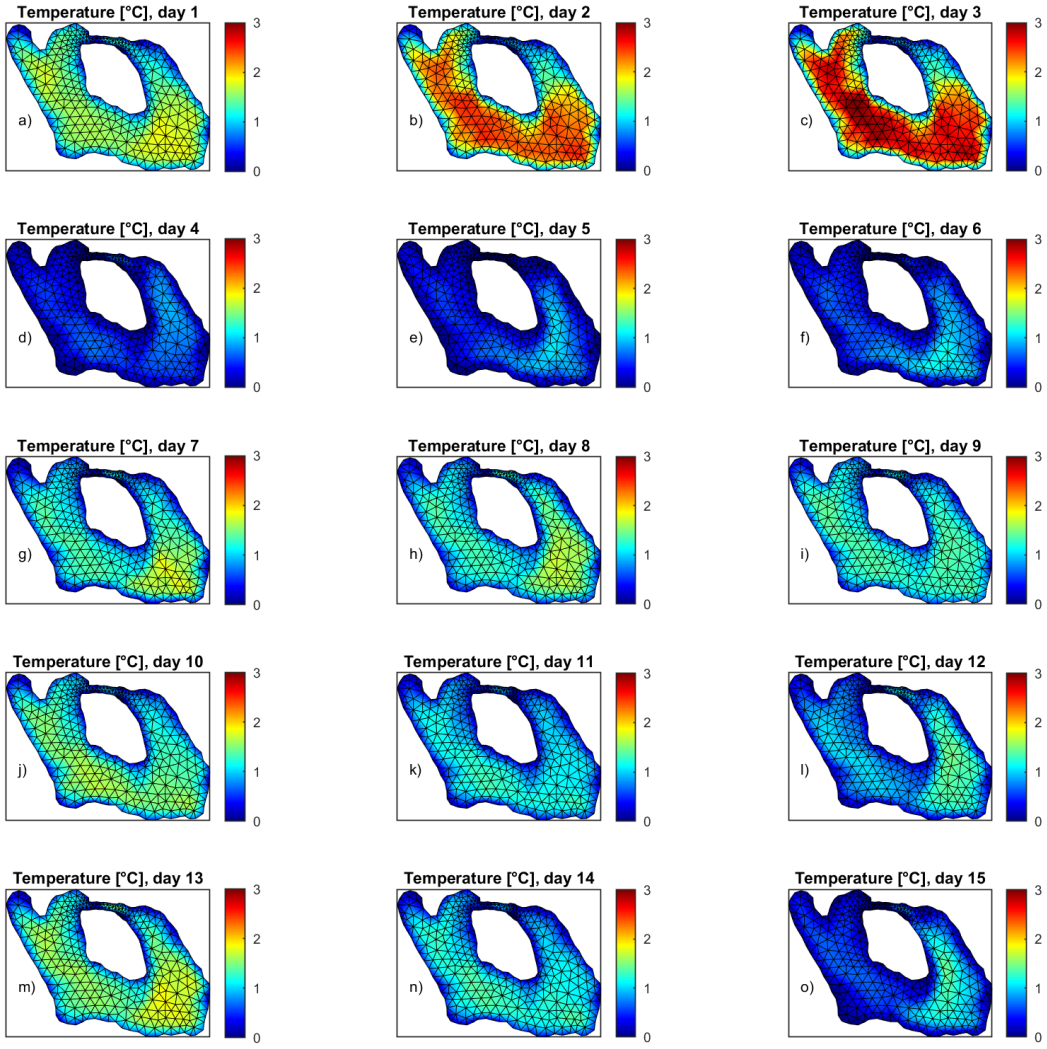


Figure 7.7 – Simulated spatial distribution of water temperature difference between the surface and bottom layers, during the calibration run in 2018. Each panel represents the temperature difference calculated from model results at midday (i.e. 12 a.m.) for each successive day in the simulation: from day one (panel a) to day 15 (panel o).

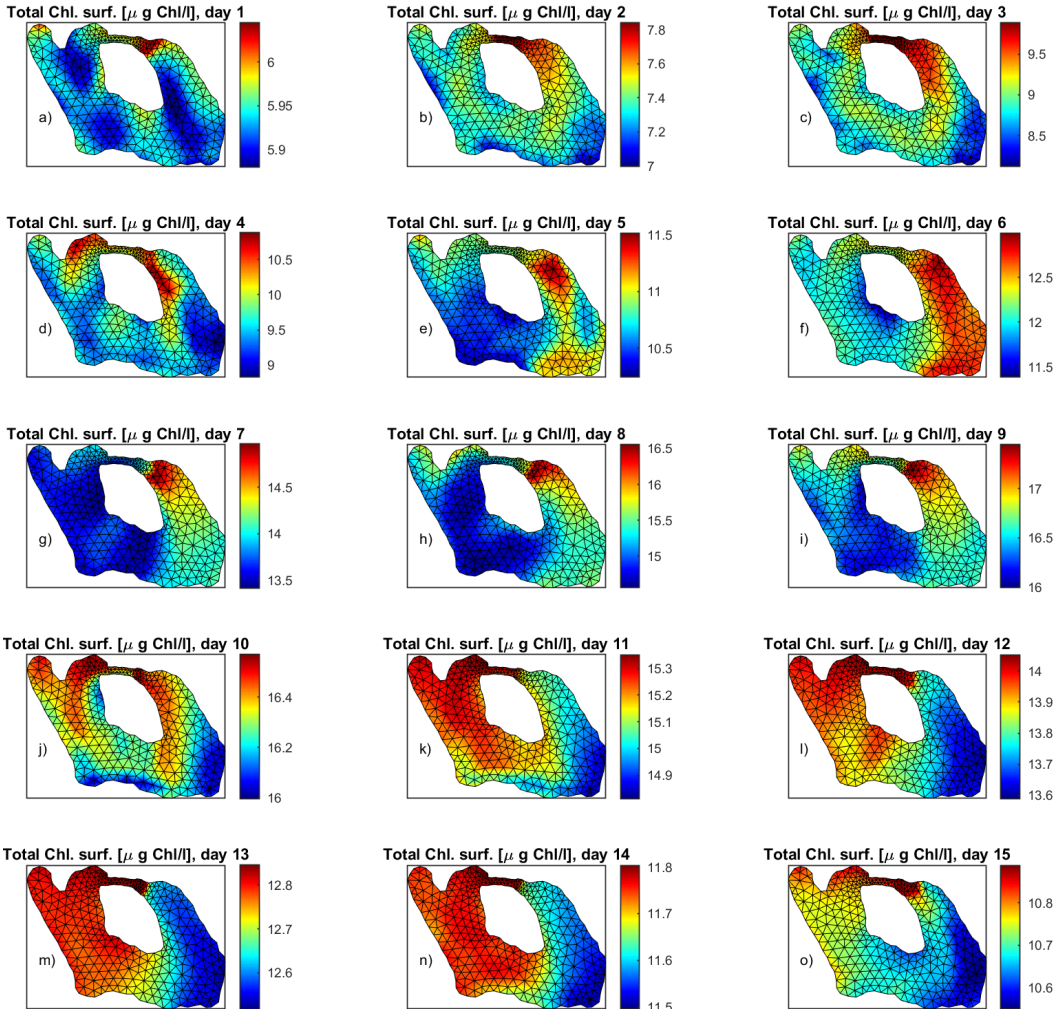


Figure 7.8 – Simulated spatial distribution of total chlorophyll concentration for the surface layer, during the calibration run in 2018. Each panel represents model results at midday (i.e. 12 a.m.) for each successive day in the simulation: from day one (panel a) to day 15 (panel o).

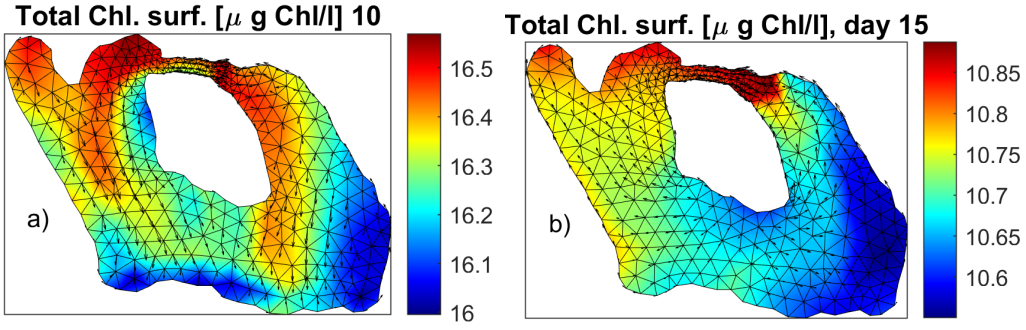


Figure 7.9 – Simulated spatial distribution of total chlorophyll concentration and velocity field for the surface layer at midday (i.e. 12 a.m.), during the calibration run in 2018. Panel a represents simulation day 10, and panel b represents simulation day 15. The value of the strongest vectors in the two panels is of the order of a few cm s^{-1}.

7.2 Seasonal application

After the short-term calibration runs, the coupled TELEMAC3D and AED2 were run over a longer amount of time, in order to test their ability to reproduce the seasonal dynamics in terms of water temperature and primary production. To do so, the configuration of the biogeochemical model was extended to species adapted to lower water temperatures and recalibrated.

For this application, the work was focused on the year 2019, for which an extensive data set is available both in terms of high-frequency observations and of laboratory analysis of water samples.

7.2.1 Model configuration

7.2.1.1 Hydrodynamic model and computational domain

The configuration of the hydrodynamic model TELEMAC3D is the same as the one described in section 7.1.2.1 in terms of turbulence closure, heat exchange model and numerical schemes. However, in this application the parameters of the heat exchange model were recalibrated, in order to match water temperature data during winter. The coupled models TELEMAC3D and AED2 share the same computational grid (see Figure 7.1), computational time step (60 s) and output time step, which was set to 4h for this application in order to reduce the size of the resulting files. Due to the dynamic calculation of the light extinction coefficient implemented in the coupled models, the Secchi depth does not need to be defined by the user.

7.2.1.2 Biogeochemical model

The configuration of the biogeochemical model implemented for the seasonal simulation of lake Champs-sur-Marne is largely based on that described in section 7.1.3.1, with the only changes being the addition of two phytoplankton groups to the configuration: diatoms and flagellates. They are introduced here as species respectively adapted to cold and medium temperatures, and lower light intensities than green algae and cyanobacteria, with the objective of reproducing the year-long evolution of total chlorophyll in the study site.

The uptake of Silica by phytoplankton groups is currently not implemented in the coupled TELEMAC3D and AED2 and couldn't therefore be activated for the diatoms group. For this reason, the AED2 module for Silica was not added to this configuration. The modules activated and the variables simulated by the model in this configuration are listed in table 7.7. Such configuration includes a total of 115 parameters.

7.2.2 Simulation period and dataset

The analysis of the data set collected on the study site (see chapter 4) showed that in Lake Champs-sur-Marne phytoplankton blooms do not occur only during the warmer summer months. On the contrary, the highest peak of total chlorophyll is often recorded in late winter or early spring. The simulation was therefore conducted with the configuration described in section 7.2.1 and table 7.7, from the beginning of February until the end of November of 2019. The simulation therefore covers

Table 7.7 – Modules and variables activated in the configuration of the biogeochemical model AED2 for the seasonal application in 2019.

Module	**Variables**
Oxygen	Dissolved oxygen
Phosphorus	Ortho-phosphate
Nitrogen	Ammonium
	Nitrate
Organic matter	POC
	PON
	POP
	DOC
	DON
	DOP
Phytoplankton	Cyanobacteria
	Green algae
	Flagellates
	Diatoms

11 months. The first six months (from the beginning of February until the end of July) were used for the calibration of the coupled models, while the remaining four months were run for validation.

The year 2019 was selected for the simulation because of the exhaustive data set available that covers: high-frequency observations of water temperature, total chlorophyll, dissolved oxygen, cyanobacteria and nitrate concentrations, as well as concentrations of total nitrogen, nitrate, total phosphorus, phosphate, and dissolved organic carbon obtained through the laboratory analysis of periodical *in situ* water samples.

Initial conditions

The amount of phytoplankton present on the study site at the beginning of the simulation is very low, with high-frequency observations of total chlorophyll close to zero (see Fig. 7.12a). In the model, the initial biomass was equallly dispatched among the four phytoplankton groups, so to match the total chlorophyll values recorded by the high-frequency sensor at site B. High-frequency observations were also used to initialize dissolved oxygen, nitrate and water temperature. As for the other simulations discussed in the previous sections, the hydrodynamic model was initialized with water at rest.

Two field campaigns were carried out on the 16th of January 2019 and on the 13th of February of 2019 and granted data for the concentrations of total nitrogen, nitrate, total phosphorus, phosphate, and dissolved organic carbon, which were used to initialize the remaining variables of the model. As the beginning of the simulation period falls midway between the two campaigns, their data were averaged to obtain the initial conditions.

The data from the field campaigns allowed to directly set the initial conditions for phosphate, and dissolved organic carbon. Values of total nitrogen and phosphorus allowed to obtain estimates for the remaining variables. Since Lake Champs-sur-

Marne has no tributaries, it was assumed that the inorganic fractions of dissolved and particulate phosphorus and nitrogen could be considered negligible compared to their organic forms. An estimate of dissolved organic phosphorus (DOP) was therefore obtained by subtraction of phosphate from total phosphorus concentration. The concentration for dissolved organic nitrogen (DON) was estimated on the basis of the DOP concentration through the Redfield ratio (i.e. N/P $\sim$ 16, Redfield 1934). The ammonium concentration was then calculated by subtracting nitrate and DON concentrations from the amount of total dissolved nitrogen.

7.2.3 Model calibration and validation

The calibration of the coupled models for the seasonal simulation was done by trial and error, based on high-frequency measurements at site B. Water temperature data at the surface and bottom layers were used to calibrate and validate the simulated thermal dynamics and evaluate the feedback of the biogeochemical model on the hydrodynamic model results induced by the dynamic calculation of the light extinction coefficient. High-frequency observations of dissolved oxygen, total chlorophyll, cyanobacteria, and nitrate were used to calibrate and validate the biogeochemical model.

The root mean square error (RMSE) was calculated between model results and observations to evaluate model performances during both calibration and validation. For this purpose, high-frequency data were treated and averaged every four hours, according to the model output time step, as described in chapter 4.

The parameters resulting from the short-term calibrations of the hydrodynamic (section 7.1.2) and biogeochemical (section 7.1.3) models were used here as a reference. The calibration effort was mainly focused on the biogeochemical model AED2. However, the set of parameters calibrated in section 7.1.2 for the hydrodynamic model, lead to an underestimation of water temperature during the winter months. For this reason a few parameters from the heat-exchange model of TELEMAC3D were here recalibrated.

Most of the AED2 parameters were varied in order to test their impact on the year-long dynamics of stratification, oxygen, phytoplankton and nutrients in the study site. After such tests and based on values found in similar applications in scientific literature (e.g. Hipsey et al. 2013, Fenocchi et al. 2019), only a reduced number of parameters were finally calibrated. They are listed in table 7.8, along with their calibrated values.

7.2.4 Model results

The results of the calibrated model are presented in this section, separately for water temperature (section 7.2.4.1) and biogeochemical variables (section 7.2.4.2).

7.2.4.1 Thermal dynamics

Water temperature is analysed here at site B for the complete simulation period (from February to November 2019), at the surface and bottom layers. Over the calibration period (i.e. between the months of February and July), the evolution of water temperature is correctly reproduced by the coupled models for both the surface

Table 7.8 – List of the calibrated parameters relatively to the seasonal application of the coupled TELEMAC3D and AED2, and corresponding calibrated values.

Model	Parameter name	Parameter value	Unit
T3D	Coefficient for atmospheric radiation	0.89	-
	Coefficient for surf. water rad.	0.93	-
	Coefficient for atm.-wat. heat exc.	0.0017	-
AED2	Sediment O_2 demand	200	mmol O m^{-2} d^{-1}
	Nitrification rate	0.05	d^{-1}
	DOC mineral. rate	0.01	d^{-1}
	POC mineral. rate	0.007	d^{-1}
	DON mineral. rate	0.01	d^{-1}
	PON mineral. rate	0.01	d^{-1}
	DOP mineral. rate	0.005	d^{-1}
	POP mineral. rate	0.005	d^{-1}
	DOM light ext. coeff.	0.002	m^2 $mmolC^{-1}$
	POM light ext. coeff.	0.002	m^2 $mmolC^{-1}$
	Water light ext. coeff.	0.5	m^{-1}

	Green algae	Cyanobacteria	Flagell.	Diat.	Unit
Growth rate at 20°C	1.4	1.2	1.6	3.7	d^{-1}
Temp. coeff. for growth	1.08	1.08	1.08	1.05	-
Standard water temp.	20	20	18	4	°C
Optimum water temp.	25	28	23	9	°C
Maximum water temp.	35	36	30	18	°C
Specific light ext. coeff.	0.005	0.005	0.01	0.005	m^2 mmol C^{-1}
Half saturation constant light limitation	28	25	28	10	μE m^{-2} s^{-1}

and bottom layers, as shown in Figure 7.10. The RMSE values with high-frequency measurements are rather low and comparable to those found with the hydrodynamic model Delft3D-FLOW in chapter 5. The RMSE is of 1.08°C for the surface layer and 0.91°C for the bottom layer.

Model performances in terms of water temperature deteriorate during the following period, and in particular between the months of October and November. The total RMSE calculated over the whole simulation period (i.e. calibration and validation) increases, especially for the bottom layer. In fact, for surface water temperature, high-frequency data at site B were available only until the 22nd of October 2019 and therefore model performances could not be compared to observations after this date. For the surface layer, the RMSE calculated during the simulated period where data were available is of 1.18°C. For the middle and bottom layers, data were available at site B until the end of the simulation. In this case the corresponding RMSE for the whole simulation period are, respectively, of 1.39°C and 1.49°C. This behaviour was further analysed by comparing observed and simulated bottom water temperature at site B with the air temperature values used to force the model. Such comparison is presented in Figure 7.11 and shows how, between the months of October and November, simulated water temperature follows closely the data of air temperature used to force the model.

Furthermore, the parity diagrams in panels a, b and c of Fig. 7.10 show a slight overestimation of the highest values of water temperature. Looking at figures 7.10d, 7.10e and 7.10f, the overestimation of water temperature by the model is present between the months of June and August, and is particularly strong for the bottom layer. During summer, the strength of thermal stratification is therefore not correctly simulated by the coupled models. This could be caused by an underestimation of the light extinction computed by AED2. To test this hypothesis, the light extinction coefficient (K_d) was calculated over the whole simulation period following equation (7.2) , and using the values listed in table 7.8 for the specific light extinction coefficients and model results for the concentrations of phytoplankton, DOC and POC. Finally, the light extinction coefficient was converted into its corresponding simulated Secchi depth S ($S = 1.7/K_d$, equation (7.1)).

The simulated and measured Secchi depth at site B are shown in Figure 7.10e. The simulated values match closely with the observetions, with only four main exceptions (namely in the months of April, June and October). During the months of July and August the simulated Secchi depth is comprised between 1.2 and 2 m. Such values match very well with the observations, but are considerably higher than the Secchi depth calibrated in the short-term application of the hydrodynamic model (section 7.1.2), which was of 0.85 m. Ultimately, the underestimation of summer thermal stratification could be linked not only with some simplifications made in the model configuration, such as the constant water level and the absence of exchanges with groundwater (as suggested in section 7.1.2.3), but also to an excessive numerical diffusion introduced by the computational schemes of the hydrodynamic model.

7.2.4.2 Biogeochemical model

The calibration of the model was done by trial and error, comparing its results with high-frequency observations of total chlorophyll, cyanobacteria, dissolved oxygen

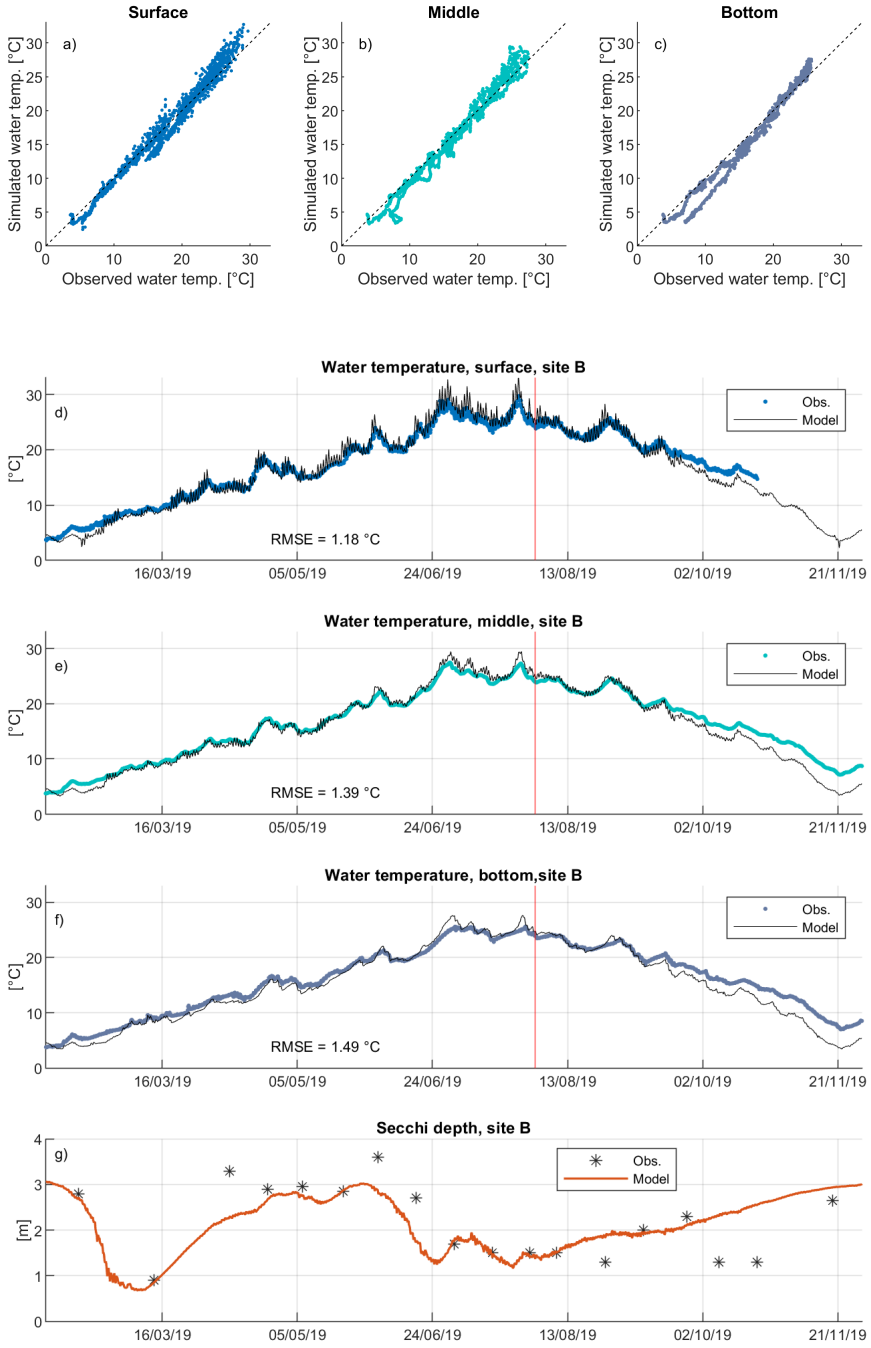


Figure 7.10 – Results of the nine months-long simulation in 2019 with TELEMAC3D/AED2. Top panels: parity diagrams for water temperature at the surface, middle and bottom layers (respectively panels a, b and c). Panels d, e and f: observations and simulation results in terms of water temperature respectively for the surface, middle and bottom layers. Panel g shows the simulated and observed values for Secchi depth. All figures are referred to measuring site B. The red vertical line in panels d, e and f separate the period used for calibration to the one run for validation.

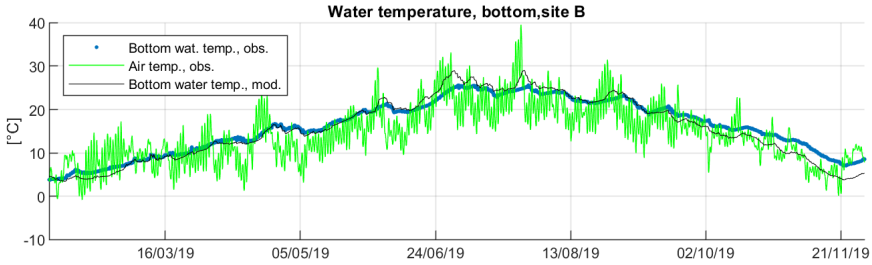


Figure 7.11 – Water temperature and air temperature data for the whole simulation period in 2019. The dotted blue line represents observed bottom water temperature at site B; black solid line represents simulated bottom water temperature at site B; green line represents the air temperature data used as input to the model.

and nitrate concentration.

The analysis of such data set, as described in chapter 4, shows the presence of a first strong peak of phytoplankton biomass around the beginning of March. It constitutes the strongest bloom of the simulated period and completely consumes the stock of nitrate present in the water column, influencing the subsequent availability of nutrients (and nitrate in particular) to sustain phytoplankton growth during the remaining growing season.

The results of the calibrated model are shown in Figure 7.12, in terms of total chlorophyll (panel a), cyanobacteria (panel b), dissolved oxygen (panel c) and nitrate concentration (panel d). In terms of total chlorophyll (Fig. 7.10a), the model reproduces correctly the overall behaviour recorded by the high frequency sensor. The first algal bloom is correctly simulated, both in terms of timing and intensity. Following, the model correctly reproduces the decrease of phytoplankton biomass, as well as the span and overall concentration magnitude of the phytoplankton during the summer months. Eventually, the end of the growing season is also well captured by the model around the end of October.

The year 2019 was not characterized by particularly strong cyanobacterial blooms. As shown by Fig. 7.10b, their maximal concentration reaches roughly 40 μg Chl. l^{-1} in four separate occasions: once during the late winter bloom, and the remaining times during sudden growth peaks between the end of August and the beginning of October. However, in the present configuration, the group representing cyanobacteria is adapted to warm water temperatures (i.e. optimum temperature of 28°C, see table 7.8), and is therefore not capable to reproduce their winter growth. In this configuration in fact, winter growth completely deputed to the diatoms group. Similarly to what was discussed for total chlorophyll, during Summer the model manages to correctly simulate the span of the growing season for the group of cyanobacteria, as well as their overall concentration magnitude. However, the model fails to reproduce the succession of short term peaks detected by the high-frequency observations.

The dynamics of dissolved oxygen simulated by the model fits very closely that recorded by the high-frequency measurements. Figure 7.8c shows that the model overestimates slightly the concentration of dissolved oxygen, in particular during

the colder months of the simulation: during the strong algal bloom of February and March, as well as during the month of November. In the remaining months of simulation dissolved oxygen concentration is correctly reproduced.

Figure 7.8d shows the comparison between high-frequency observations and model results at site B in terms of nitrate concentration. The model correctly reproduces the observed nitrate dynamics before and during the late-winter algal bloom. The initial increase in nitrate concentration is modelled here solely through the processes of mineralization of organic matter and nitrification of ammonium. The rapid consumption of all the available nitrate during the late winter algal bloom is also correctly simulated.

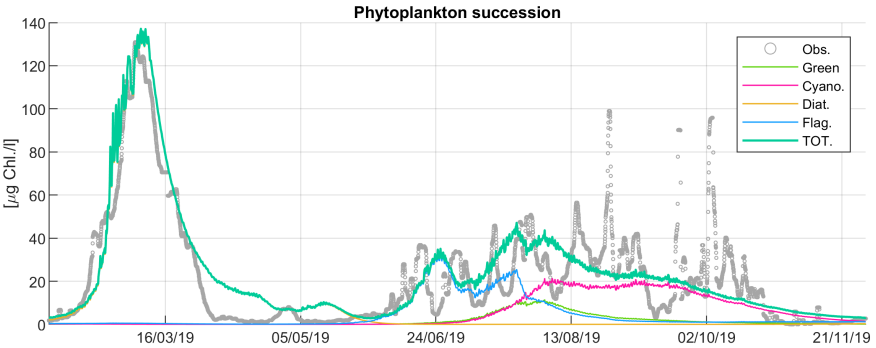
Right after the late-winter bloom, during the months of April and May, the phytoplankton observations are very low and correspond to an increase in the observed nitrate concentrations. The nitrate accumulated in the water column during this period is then quickly consumed at the beginning of the second blooming period around the month of June. The lake appears to be nitrogen limited from this point until the end of the growing season. These dynamics are not fully reproduced by the model. At the end of the late-winter bloom the simulated phytoplankton concentration is higher than the observed one, causing, in the model, a stronger consumption of nitrate by the phytoplankton, and ultimately delaying the simulated accumulation of nitrate. Such accumulation of nitrate is necessary to the model in order to sustain phytoplankton growth during the subsequent summer and autumn months. Eventually, at the end of the growing season, around the end of October, a strong increase of nitrate concentration is recorded in the study site, that returns to values similar to the initial ones before the end of December. The increase of nitrate concentration reproduced by the model shows a lower rate, roughly one third of the one of detected in the observations. In the model, such increase is obtained solely through the processes of organic matter mineralization and nitrification.

Eventually, the concentration of dissolved organic carbon simulated by the model was compared with data from the laboratory analysis of the periodic water samples taken at site B. Such comparison is shown in Figure 7.13. The increase of DOC observed on the study between the months of March and April, following the initial algal bloom is overestimated in the model simulation, as is the rate of mineralization of DOC and its decrease during the months of May and June. However, the overall concentration and dynamics of DOC is adequately simulated by the model.

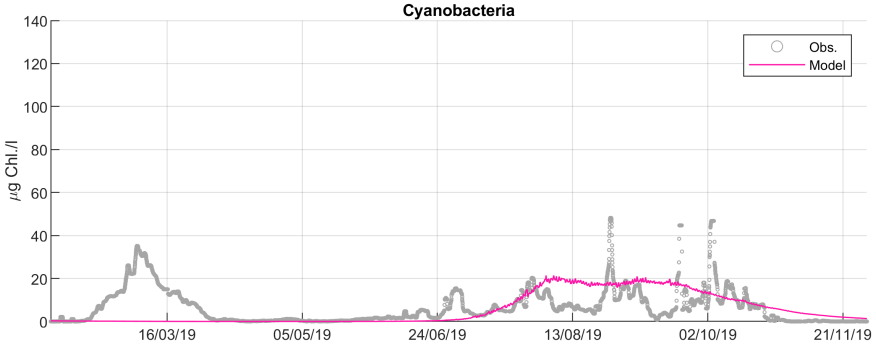
7.3 Discussion

The hydrodynamic model TELEMAC3D and the biogeochemical model Aquatic EcoDynamics (AED2) were implemented on Lake Champs-sur-Marne and simulation results. In their own fields of application, they are both well established models that, separately, have been employed in various contexts (e.g., for TELEMAC3D: Villaret et al. 2013, Feng & Jodeau 2016, Merkel 2019) (e.g., for AED2: Zhang et al. 2017, Fenocchi et al. 2019, Krinos et al. 2019). The coupling of the two models introduces the possibility of modelling the biogeochemical cycle directly through the TELEMAC modelling system. The coupling between the models is recent, and, to the best of my knowledge, no applications can be found in scientific literature. In this work, it was possible to test the behaviour of the coupled models in a natural

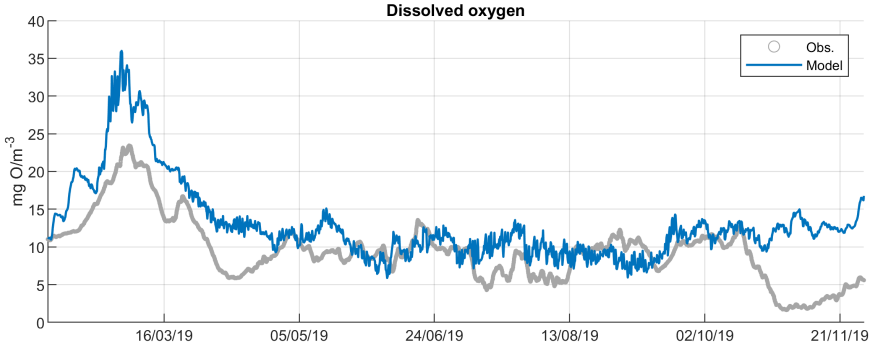
a)



b)



c)



d)

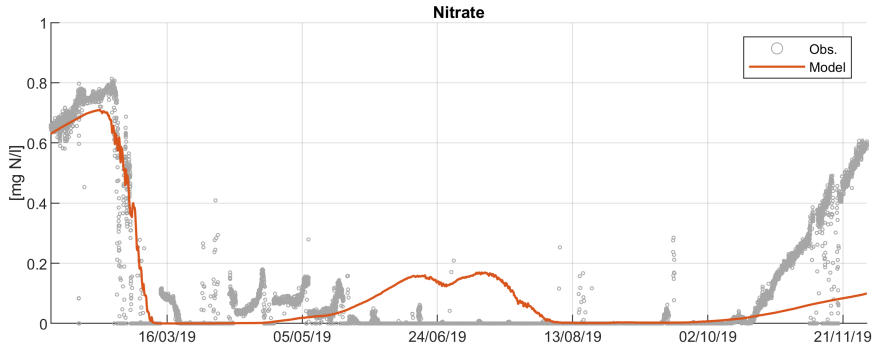


Figure 7.12 – Results of a nine-months long simulation in 2019 with TELEMAC3D/AED2. Coloured lines represent model results and grey dotted lines the observation series; all figures are referred to site B. Panel a: observed and simulated total chlorophyll and chlorophyll content of the four simulated phytoplankton groups; panel b: cyanobacteria concentration; panel c: dissolved oxygen concentration; panel d: nitrate concentration.

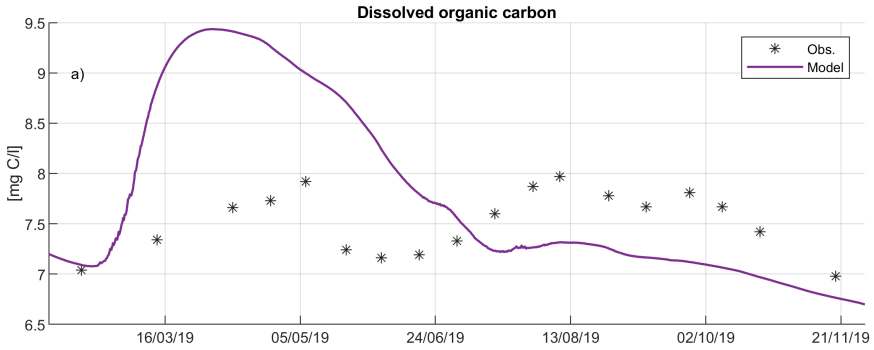


Figure 7.13 – Results of a nine-months long simulation in 2019 with TELEMAC3D/AED2in terms of dissolved organic carbon (DOC, purple line), and results of the laboratory analysis of periodic water samples for DOC.

lake ecosystem. The use of high-frequency data allowed to evaluate the performance of the models both on short-term and on seasonal dynamics.

7.3.1 Short-term simulations and the dynamic calculation of the light extinction coefficient

The stand-alone hydrodynamic model TELEMAC3D was calibrated and validated on short-term simulations in order to reproduce at best the thermal and stratification dynamics of a small and shallow water body such as Lake Champs-sur-Marne. The model performed well against high-frequency water temperature observations (RMSE generally lower than 1°C) and thermal stratification was overall well captured by the model (Fig. 7.2). However, the model shows the tendency to overestimate bottom water temperature (as in particular in the validation period during 2016) causing the simulated stratification in some instances to be weaker than the observed one. This can be linked, on the first hand, to an excessive numerical diffusion introduced by the computational schemes of the hydrodynamic model and, on the other hand, to some of the simplifications included in the model configuration. In particular, water level variations and exchanges with ground water are here neglected, but could also sensibly impact bottom water temperature in the real system.

The performance of TELEMAC3D during the short-term calibration period of 2018 was very similar to that of the hydrodynamic model Delft3D-FLOW during the same period (see section 6.3). This shows the soundness of the two hydrodynamic models, that manage to achieve similar performances in terms of thermal stratification in a shallow water body despite the differences in their mathematical structure and numerical approach (such as: the formulation of the heat-exchange model, or, more generally, the approach to the resolution of the systems of partial differential equations). Furthermore, the configurations implemented here for the two models differed in a few important aspects and a direct and generalized comparison of the models results is not straightforward. In particular, both the horizontal (Cartesian mesh for Delft3D, triangular mesh for TELEMAC3D) and vertical discretization (Z-model for Delft3D-FLOW, and σ-layers for TELEMAC3D), the turbulence clo-

sure model (k-ε for Delft3D-FLOW, and mixing length with damping function for TELEMAC3D), the calibrated value of the Secchi depth (1 m for Delft3D-FLOW, and 0.85 m TELEMAC3D), and the formulations of the heat-exchange models were different in the two approaches. In order to thoroughly compare the performance of the two models further tests would be needed, and should feature configurations as similar as possible, in terms at least of average distance among the nodes of the computational mesh, vertical discretization, and turbulence model.

The same short-term simulation periods were then simulated to first test the coupled models TELEMAC3D and AED2. In terms of surface and bottom water temperature, the coupled models showed performances comparable to those of the stand-alone TELEMAC3D during both calibration and validation. In the coupled models, the vertical distribution of water temperature is strongly influenced by the dynamic calculation of the light extinction coefficient, which is computed by AED2 at each time step. The values of the specific light extinction coefficients for phytoplankton and dissolved and particulate organic carbon, as well as their concentrations, are crucial for the thermal response of the models on the vertical axis. The feedback of AED2 on TELEMAC3D needs to be carefully taken into account when calibrating the heat-exchange budget at the water-air interface.

The dynamic calculation of the light extinction coefficient also influences light availability for biomass growth. In this respect, the parameter *half saturation constant for light limitation*, that acts on the light limitation factor for phytoplankton growth (see section 3.2.4.2, equation (3.28)), has a strong importance in the relation between light availability and phytoplankton biomass production. The configuration implemented for the short-term simulations included two phytoplankton groups: green algae and cyanobacteria. Such configuration, despite being basic, proved to be able to correctly simulate biomass growth and oxygen dynamics in a natural ecosystem over short periods spanning roughly two weeks during the summer season. During the simulation in 2016, the model overestimated observed high-frequency total chlorophyll, while correctly simulating dissolved oxygen concentraiton. The analysis of two profiles taken with the BBE Fluoroprobe indicates an underestimation of total chlorophyll by the optical sensor measuring at high-frequency, rather then an overestimation of phytoplankton concentration by the model. According to the profiles, at the begginning of the simulaiton the bloom was dominated by cyanobacteria, while green algae became dominant towards the end of the simulaiton. Cyanobacteria chlorophyll therefore seems to be hardly detected, in this instance, by the high-frequency intrument. This highlights the advantage of having access to multiple sources of observations in order to validate the data, and reinforces the idea, already presented in chapter 4 and applied here for the year 2019, of implementing regression models to correct, when necessary, the data series.

7.3.2 Spatial variability

The presence of spatial patterns was analysed for the short-term simulation in 2018, for thermal stratification and total chlorophyll. In the model, the strong horizontal patterns shown in Figure 7.7 for the difference between surface and bottom water temperature are induced by the bathymetry and are consistent with the results of the spatial analysis of stratification carried out with the model Delft3D-FLOW

in the context of the climate change analysis of chapter 5. In section 5.5.2, it was highlighted how the shallower northern portion (i.e. depth < 0.8 m) of the study site did not stratify (i.e. temperature difference between surface and bottom layers >1°C) for at least 24 consecutive hours (i.e. one stable stratified day, see for instance Figure 5.6). The same result was found here with TELEMAC3D. However, Figure 7.7, which plots model results at 12 a.m., shows how water temperature differences between the surface and bottom layers can be considerably greater than 1°C, even in some of the shallowest areas at least during the day.

Horizontal gradients in the distribution of biomass in the surface layer were weak but present on the study site. As shown in figures 7.8 and 7.9, total chlorophyll behaved similarly to an active tracer and its spatial distribution was mainly influenced by the velocity field. Extremely similar patterns were found for cyanobacteria. This shows how, in the present model configuration, some of the factors supposed to be important for biomass growth, and particularly favourable for that of cyanobacteria, are not fully reproduced by the model. Some of these factors were discussed in chapter 5 for cyanobacteria, and include their ability to migrate towards the surface of the water in stratified environments, and the process of nutrient accumulation in the sediment during stratification followed by a replenishment of the water column with such nutrients during subsequent mixing events. However, the latter process is only partially taken into account in this model configuration through the alternation between mixing and thermal stratification. The motion capabilities of cyanobacteria and biogeochemical processes taking place in the sediments are not explicitly considered.

From these first model applications, thermal stratification and mixing events seem to influence principally the vertical distribution of phytoplankton concentration, and less its horizontal patterns. In the short term simulations discussed in section 7.1.3, slightly higher concentrations of phytoplankton were simulated during stratification towards the surface layer, which is characterized by more penetrating light and warmer temperature. During subsequent overnight mixing, the phytoplankton concentration was levelled over the water column. Further tests would be necessary to analyse the importance of stratification and mixing for horizontal patterns and bloom initiation in models simulations, with a more complete model configuration possibly including the migration capabilities of cyanobacteria, their nitrogen-fixation abilities, and an explicit description of the sediment layer.

7.3.3 The coupling and the performances of the biogeochemical model

During the seasonal simulation of 2019, the coupled models reproduced well the overall water temperature dynamics at all three layers. The underestimation of water temperature found for the months of October an November, as well as the slight overestimation of summer water temperature, could be linked with the absence in the model configuration of exchanges with groundwater, that could somewhat moderate the seasonal variations of water temperature in the real system. Despite this underestimation of water temperature, the overall RMSE values between model and observations were lower than 1.5°C. Furthermore, starting in particular from the month of November until the end of January, very low phytoplankton biomass

is usually detected on the study site. In this respect, the bias found for water temperature during autumn should have a reduced impact on the outcomes of the coupled biogeochemical model.

The biogeochemical model reproduced well the biogeochemical cycle in the study site, both on a short term and on a seasonal term. The availability of an extensive high-frequency data set allowed to test the model thoroughly, against multiple variables that characterize the biogeochemical cycle exhaustively. On the short term, the calibrated model was able to reproduce adequately rapid phytoplankton and cyanobacteria growth and decline occurring over a period of only two weeks, with the implementation of only two phytoplankton groups (green algae and cyanobacteria).

For the seasonal simulation of 2019, the availability of additional measured variables, dissolved organic carbon (DOC) and nitrate in particular, allowed for a more accurate initialization of the model. This fact, together with the need to reproduce dynamics occurring at a wider range of water temperatures and daylight durations, lead to a recalibration of the coupled models. The introduction of two additional phytoplankton groups, adapted to cold (diatoms) and medium (flagellates) water temperatures, allowed to correctly reproduce the overall phytoplankton dynamics observed in the study site for the year 2019. Similarly, the overall dynamics of oxygen, nitrate, DOC and Secchi depth were adequately reproduced by the model for 2019.

Some discrepancies were detected when comparing model results to high-frequency observations. Notably, in the context of a seasonal simulation such as that of 2019, the model did not fully reproduce the sudden peaks of growth and decline observed during spring and summer. This could be explained by the absence, in the models configuration, of: (i) nitrogen-fixing cyanobacteria such as *Aphanizomenon*, both present on the study site, that should be advantaged under nitrogen-limited conditions, (ii) predation by zooplankton or competition for nutrients and light with other organisms, such as macrophytes, and (iii) by the absence in the model configuration of external nutrient sources, that might locally boost phytoplankton growth over a short period of time.

The model underestimates the accumulation of nitrate observed in the study site during the months of October and November 2019. This is partially due to the slight overestimation of the phytoplankton population in the study site during the same period, which consumes the mineralized nitrate at a faster rate. However, such accumulation is obtained in the model solely through the processes of mineralization of organic matter and nitrification of ammonium. This suggests that a non negligible external source of nutrients (N in particular), probably coming from underground water exchanges with the Marne river, droppings from the vast local population of waterbirds, or, in a smaller measure, rainfall run-off, is acting on the study site and should be estimated and introduced in the model configuration. Alternatively, the underestimation of nitrate during autumn could be simply originated by an underestimation of the mineralization rate of organic matter. In the latter case, however, if the mineralization rate was to be increased, other competitors (e.g. macrophytes) for nutrients, and nitrate in particular, should be introduced in the model configuration in order to prevent an overestimation of nitrate concentration during spring and summer.

In fact, since lake Champs-sur-Marne has no direct surface inlets, in the simu-

lations discussed in this chapter, no external sources of nutrients are introduced, and the mineralization of organic matter constitutes the only "source" of nutrients sustaining primary production during the entire simulation. The mathematical formulation of organic matter mineralization in AED2, substantially different from that of Delft3D-BLOOM (see chapter 3 for further details), was particularly adapted for such an approximation. In particular, on a seasonal term and using organic matter mineralization as the only source of nutrients, similar results could not be obtained with Delft3D-BLOOM. In deflt3D-BLOOM organic matter mineralization was not rapid enough for this purpose, and the inorganic nutrient supply to the system could not sustain phytoplankton growth over periods longer than roughly a month.

The coupling of Aquatic EcoDynamics with TELEMAC3D is recent and the work on the coupling remain in progress. For instance, during this PhD, two successive versions of the coupling between the models were tested, until version 8.1.2 was released in October 2020 and was used to obtain the results presented in this chapter. Some functions available in AED2 are currently not fully implemented in the coupling with TELEMAC3D, and have forced some of the choices made while building the model configuration: for instance, the choice of the σ-layers over the Z-model for the discretization of the vertical axis, and the non-activation of the Silica module despite the presence of the diatoms phytoplankton group. However, this model configuration has proved that the coupled models are stable, robust and that allow to take into account and simulate the main characteristics of primary production, cyanobacteria production and, more generally, of the biogeochemical cycle in a lake.

At this first stage of the utilization of the coupled models, their calibration was done by trial and error. However, thanks to a collaboration with Laboratoire Hydraulique Saint-Venant (LHSV) that granted access to computational server CRONOS owned by Elecrticité de France (EDF), it was possible to start the application of Approximate Bayesian computation (ABC, see chapter 5 for further details), for the automated calibration of the coupled models. The objective was to apply the same methodology developed in chapter 5 for the model Delft3D-BLOOM, for the calibration of the coupled TELEMAC3D and AED2. However, differently from what was done in chapter 5 for Delft3D-BLOOM, the calibration period is of six months (from the 1st of February 2019 to the 31st of July 2019) for TELEMAC3D/AED2.

Thanks to the access to the CRONOS computational server, the 30000 model simulations necessary to create the training data set for the ABC calibration were carried out. However, the calibration procedure through ABC is still to be analysed. This work is at present in an early stage but should be pursued through future research projects.

Chapter 8

Conclusions and perspectives

Lake ecosystems are vital for the biosphere, and are of crucial importance for human society. However, during the past decades, they have been deteriorating at an increasing pace, and the occurrence of harmful algal blooms has aggravated worldwide. This is principally caused by eutrophication and climate change. In this context, the importance of small urban lakes has been recognized only recently, and such ecosystems have so far received less attention by the scientific community in comparison with deeper and larger water bodies.

The response of lake ecosystems to changing climatic and eutrophication conditions can be simulated through coupled hydrodynamic and biogeochemical models. A vast range of models is available to address this issue, and they might differ strongly in terms of dimensionality (1D, 2D or 3D), coupling and formulation. Moreover, biogeochemical models are often complex and highly parametrized. Their calibration is a challenging task, further complicated by the low-frequency of traditional limnological monitoring, and is often overlooked in scientific literature.

This PhD thesis is an integral part of the ANSWER research project, funded by the French National Research Agency (ANR) and the Natural Science Foundation of China (NSFC), that seeks a deeper understanding of the triggering factors for algal blooms and cyanobacteria blooms in freshwater bodies, and gathers different expertises to carry out both laboratory experiments and numerical simulations.

In this framework, the main objectives of this PhD thesis are to make use of modelling tools to simulate the thermal, phytoplankton and cyanobacteria dynamics on different time-scales in a small and shallow urban lake, and to identify and analyse key elements that are of strategic importance for lake ecosystem modelling in the context of climate change. Through the use of two different coupled hydrodynamic and biogeochemical models, three crucial aspects of aquatic ecological modelling were analysed: (i) the hydrodynamic model and its ability to simulate the thermal dynamics of a small and shallow lake in a context of climate change, (ii) the high level of parametrization of deterministic biogeochemical models and the importance of automated techniques to address their calibration, and (iii) different formulations and approaches to the coupling between hydrodynamic and biogeochemical models, and the feedbacks between the models.

In particular, three key elements for the implementation and exploitation of aquatic ecological models are addressed in this work: (i) the use of a 3D hydrodynamic modelling approach to characterize the thermal and stratification dynamics in shallow

lakes and highlight its spatial heterogeneity and long term evolution, (ii) the high level of parametrization of deterministic biogeochemical models and the importance of sensitivity analysis and automated techniques to address their calibration and improve their results, and (iii) different approaches to model the biogeochemical cycle and different grades of coupling between 3D hydrodynamic and biogeochemical models.

The availability of an extensive data set of high-frequency measurements allowed the models to be tested on different time frames: from short-term (a few weeks), to seasonal-term and long-term (decades) simulations.

The aquatic ecological models used in this work are complex three-dimensional deterministic models constituted of a purely hydrodynamic model coupled with a biogeochemical model. Namely, the models used for this work are Delft3D-FLOW and Delft3D-BLOOM from the Delft3D modelling suite, and TELEMAC3D coupled with Aquatic EcoDynamics (AED2).

The hydrodynamic model Delft3D-FLOW was used to test the impact of climate change on a small urban lake, by reconstructing the long-term thermal regime of lake Champs-sur-Marne from 1960 to 2017. To do so, the SAFRAN meteorological reanalysis were used to force the model, and a series of indices describing the dynamics of stratification and cyanobacterial production were defined to characterize the thermal regime.

Small and shallow lakes are extremely widespread ecosystems. The results of this study suggest that such systems experience considerable thermal stress caused by climate change that could have favoured cyanobacteria dominance, especially in nutrient-enriched systems. More precisely, the results obtained in this work showed that climate change had a strong impact on the study site during the past six decades, and that small urban lakes react rapidly to the external meteorological conditions, with only limited resilience to climatic shifts. The additive effect of increasing solar radiation and air temperature and decreasing wind speed enhances the changes found in the lake. The simulated average annual water warming ($0.6°C.dec^{-1}$, or an increase of 32% in annual water temperature between 1960 and 2017) was stronger than that found for air temperature ($0.3°C.dec^{-1}$). The impact found for stratification and cyanobacteria production was even more alarming, with an increase of over 30% of the stratification indices and over 60% of the cyanobacteria average growing degree days during the six past decades. Spring shows the sharpest trends in terms of water temperature, water column stability and cyanobacteria growing degree days, and might ultimately be the season suffering the strongest changes. Eventually, the strong spatial heterogeneity found in the study site for thermal stratification and growing degree days might locally create conditions particularly favourable for cyanobacteria blooms. Horizontal patterns can be particularly strong for shallow lakes due to the relative importance of bathymetric variations, and can only be detected through a 3D approach.

The tendencies found for water temperature, stratification and cyanobacteria production could favour early phytoplankton blooms, contribute to the proliferation of cyanobacteria, and ultimately enhance the degradation of the whole aquatic ecosystem. The computation of the thermal growth rate and of the growing degree days for cyanobacteria demonstrated how the use of thermal-hydrodynamic simulations can allow to quantitatively estimate tendency to biomass production. The

approach in this study is general and could be extended to any water body, as long as meteorological forcing and data for calibration are available. It could also be used to obtain future projections of the thermal regime of the study site, if the model is forced with climate change scenarios.

Furthermore, by activating the coupling with a biogeochemical model, such study could constitute the basis for long-term quantitative simulations of the biogeochemical cycle. However, in order to do so, the uncertainty affecting the biogeochemical model parameters values, and the data describing the long-term evolution of the nutrient supplies to a water body should be carefully analysed and minimized.

In fact, deterministic biogeochemical models are often complex and highly parametrized, and their calibration is often overlooked in scientific literature. Automated calibration strategies can help estimating parameters values, therefore reducing the uncertainty of the outcomes of a model. In this work, an innovative technique for automated parameter inference (Approximate Bayesian Computation, or ABC) was tested on the biogeochemical model Delft3D-BLOOM. This technique (ABC-RF with SA) is based on the Approximate Bayesian Computation (ABC) wich is combined with the machine learning technique called "Random Forest" (RF) and a sensitivity analysis (SA) of the model parameters.

The results of this study show that, without a preliminary sensitivity analysis, the standard formulations of ABC and ABC-RF are not suited for the calibration of a complex biogeochemical model. However, the new algorithm proposed here (ABC-RF with SA) was successfully applied for the calibration of the Delft3D-BLOOM model. The use of high-frequency data allowed to focus the work on a short-term algal bloom, an event that would possibly be missed by traditional periodical survey, also reducing computational time. After the automated calibration, the overall phytoplankton growth and mortality were correctly simulated, as well as cyanobacteria and oxygen concentration. Eventually, this work confirms the importance and the advantages of implementing automated calibration strategies in the context of biogeochemical models, showing that this allows on the one hand to improve model results while reducing their uncertainty, and on the second hand to gain new insights both on the optimal values for the parameters, and on their importance relatively to the model results.

This study constitutes a first and successful application of the ABC-RF with SA for parameter inference, and opens to a wide range of possible future applications. The calibrated model could be tested on short-term extreme meteorological events, such as heat-waves. More generally, i the context of aquatic ecological modelling, ABC-RF with SA could be applied to different models, and for simulations covering a longer period of time. Furthermore, the parameters obtained through this calibration technique could be tested for validation over multiple periods of time. Finally, ABC could also be used to define appropriate initial conditions for the application of other calibration algorithms, such as local gradient-based methods or global optimization algorithms, with the objective to seek the parameters optimum.

The set up of a second coupled hydrodynamic / biogeochemical model for Lake Champs-sur-Marne, via TELEMAC3D and AED2, allowed to test their recent coupling on a natural ecosystem and to analyse a different mathematical approach from that of Delft3D-BLOOM. The use of high-frequency data was crucial to test the model both on short-term and seasonal simulations.

The results of this study showed that the coupling between the two models is robust and stable. Even though not all the features of the two models are currently implemented in their coupling, the biogeochemical cycle is correctly simulated by TELEMAC3D coupled with AED2, both on the short term and on the seasonal term. In particular, the results show a correct simulation of the light extinction coefficient and highlight the feedback of its dynamic calculation on the hydrodynamic model. Furthermore, the results of the seasonal simulations carried out with TELEMAC3D and AED2 show the importance of the modelling approach describing organic matter mineralization for simulating phytoplankton growth over an annual cycle, in particular in an ecosystem without direct surface inlets. More broadly, this also highlights the importance of testing different modelling approaches for improving simulation performances.

At this early stage, the calibration of the coupled models was done by trial and error. However, thanks to a collaboration with Laboratoire Hydraulique Saint-Venant (LHSV) that granted access to computational server owned by Elecrticité de France (EDF) its was possible to start the application of ABC-RF with SA, for the automation of the calibration of the model parameters. This work is still in an early stage and should be pursued in the future, via research collaborations between LEESU and LHSV.

This work highlights elements of strategic importance for the reliability of lake ecosystem modelling, especially in a context of climate change. Hydrodynamic models are at the basis of the modelling chain and need to be thoroughly calibrated and validated. They can be then used to gain new insights on the past or future evolution of a water body, and to quantitatively link thermal dynamics and primary production. In this regard, their performances are solid and the use of 3D models is beneficial for the identification of spatial heterogeneity also at the scale of a small and shallow urban lake. Similar long-term studies are far more challenging when the biogeochemical cycle is introduced in the modelling chain, as new sources of uncertainty are introduced (e.g. model parameters and nutrient input). This work shows the importance of the calibration step in biogeochemical modelling and highlights the potential of automated calibration strategies, in particular by proposing an innovative algorithm for Approximate Bayesian Computation and sensitivity analysis. Moreover, this work puts emphasis on the relevance of an extensive data set that includes laboratory analyses and continuous measurements. This is crucial to ensure that the collected data are reliable and that models are able to reproduce the biogeochemical cycle on different time-frames. Finally, the implementation of two different coupled 3D hydrodynamic/biogeochemical models allowed to analyse and highlight in closer detail the effect on model results of different approaches for hydrodynamic and biogeochemical modelling, as well as for the coupling between the models.

Managing the effects of anthropogenic stressors such as climate change and eutrophication on aquatic ecosystems will be one of the greatest challenges of the upcoming decades. This work has proven that, given the right forcing, models can be reliable tools to simulate the past, present and, possibly, the future conditions of aquatic ecosystems. This work also highlights the importance of an extensive data set to thoroughly assess a model performance, and the benefits of automated calibration strategies in improving model results and enhance their reliability.

Bibliography

Abell, J., Özkundakci, D. & Hamilton, D. (2010), 'Nitrogen and Phosphorus Limitation of Phytoplankton Growth in New Zealand Lakes: Implications for Eutrophication Control', *Ecosystems* **13**, 966–977.

Adrian R., O'Reilly C. M., Zagarese H., Baines S. B., Hessen D. O., Keller W., Livingstone D. M., Sommaruga R., Straile D., Van Donk E., Weyhenmeyer G. A. & Winder M. (2009), 'Lakes as sentinels of climate change', *Limnology and Oceanography* **54**(6part2), 2283–2297.

Alduchov, O. A. & Eskridge, R. E. (1997), Improved Magnus' form approximation of saturation vapor pressure, Technical Report DOE/ER/61011-T6, Department of Commerce, Asheville, NC (United States).

Alfonsi, G. (2009), 'Reynolds-Averaged Navier–Stokes Equations for Turbulence Modeling', *Applied Mechanics Reviews* **62**(040802).

Anagnostou, E., Gianni, A. & Zacharias, I. (2017), 'Ecological modeling and eutrophication—A review', *Natural Resource Modeling* **30**(3), e12130.

Anderson, D. M., Glibert, P. M. & Burkholder, J. M. (2002), 'Harmful algal blooms and eutrophication: Nutrient sources, composition, and consequences', *Estuaries* **25**(4), 704–726.

Anderson, T. R. (2005), 'Plankton functional type modelling: running before we can walk?', *J Plankton Res* **27**(11), 1073–1081.

Anneville, O., Beniston, M., Gallina, N., Gillet, C., Jacquet, S., Lazzarotto, J. & Perroud, M. (2013), 'Climate change and Lake Geneva: the evidence is here!', *Archives des Sciences* **66**, 157–172.

Aparicio Medrano, E., Uittenbogaard, R., Van de Wiel, B., Dionisio Pires, M. & Clercx, H. (2016), 'An alternative explanation for cyanobacterial scum formation and persistence by oxygenic photosynthesis', *Harmful Algae* **60**.

Aparicio Medrano, E., Van de Wiel, B., Uittenbogaard, R., Dionisio Pires, M. & Clercx, H. (2016), 'Simulations of the diurnal migration of Microcystis aeruginosa based on a scaling model for physical-biological interactions', *Ecological Modelling* **337**, 200–210.

Arhonditsis, G. B., Adams-VanHarn, B. A., Nielsen, L., Stow, C. A. & Reckhow, K. (2006), 'Evaluation of the current state of mechanistic aquatic biogeochemical

modeling: Citation analysis and future perspectives', *Environ. Sci. Technol.* **40**(21), 6547–6554.

Arhonditsis, G. B. & Brett, M. T. (2004), 'Evaluation of the current state of mechanistic aquatic biogeochemical modeling', *Marine Ecology Progress Series* **271**, 13–26.

Arhonditsis, G. B., Perhar, G., Zhang, W., Massos, E., Shi, M. & Das, A. (2008), 'Addressing equifinality and uncertainty in eutrophication models', *Water Resources Research* **44**(1). _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2007WR005862.

Austin, J. A. & Colman, S. M. (2007), 'Lake Superior summer water temperatures are increasing more rapidly than regional air temperatures: A positive ice-albedo feedback', *Geophysical Research Letters* **34**(6).

Bachmann, R., Hoyer, M. & Canfield, D. (2000), 'The Potential For Wave Disturbance in Shallow Florida Lakes', *Lake and Reservoir Management* **16**, 281–291.

Balian, E. V., Segers, H., Martens, K. & Lévéque, C. (2008), The Freshwater Animal Diversity Assessment: an overview of the results, *in* E. V. Balian, C. Lévêque, H. Segers & K. Martens, eds, 'Freshwater Animal Diversity Assessment', Developments in Hydrobiology, Springer Netherlands, Dordrecht, pp. 627–637.

Beaumont, M. A. (2010), 'Approximate Bayesian Computation in Evolution and Ecology', *Annual Review of Ecology, Evolution, and Systematics* **41**(1), 379–406.

Beaumont, M. A., Cornuet, J.-M., Marin, J.-M. & Robert, C. P. (2009), 'Adaptive approximate Bayesian computation', *Biometrika* **96**(4), 983–990. arXiv: 0805.2256.

Beaumont, M. A., Zhang, W. & Balding, D. J. (2002), 'Approximate Bayesian Computation in Population Genetics', *Genetics* **162**(4), 2025–2035. Publisher: Genetics Section: Investigations.
URL: *https://www.genetics.org/content/162/4/2025*

Beck, M. B. (1987), 'Water quality modeling: A review of the analysis of uncertainty', *Water Resources Research* **23**(8), 1393–1442. _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/WR023i008p01393.

Beck, R., Xu, M., Zhan, S., Liu, H., Johansen, R. A., Tong, S., Yang, B., Shu, S., Wu, Q., Wang, S., Berling, K., Murray, A., Emery, E., Reif, M., Harwood, J., Young, J., Martin, M., Stillings, G., Stumpf, R., Su, H., Ye, Z. & Huang, Y. (2017), 'Comparison of Satellite Reflectance Algorithms for Estimating Phycocyanin Values and Cyanobacterial Total Biovolume in a Temperate Reservoir Using Coincident Hyperspectral Aircraft Imagery and Dense Coincident Surface Observations', *Remote Sensing* **9**(6), 538. Number: 6 Publisher: Multidisciplinary Digital Publishing Institute.

Bedri, Z., Bruen, M., Dowley, A. & Masterson, B. (2011), 'A Three-Dimensional Hydro-Environmental Model of Dublin Bay', *Environmental Modeling and Assessment* **16**, 369–384.

Bernard, O. & Rémond, B. (2012), 'Validation of a simple model accounting for light and temperature effect on microalgal growth', *Bioresource Technology* **123**, 520–527.

Biggs, J., von Fumetti, S. & Kelly-Quinn, M. (2016), 'The importance of small waterbodies for biodiversity and ecosystem services: implications for policy makers', *Hydrobiologia* **793**.

Bleiker, W. & Schanz, F. (1989), 'Influence of environmental factors on the phytoplankton spring bloom in Lake Zürich', *Aquatic Science* **51**(1), 47–58.

Bláha, L., Babica, P. & Maršálek, B. (2009), 'Toxins produced in cyanobacterial water blooms – toxicity and risks', *Interdiscip Toxicol* **2**(2), 36–41.

Boehrer, B. & Schultze, M. (2008), 'Stratification of lakes', *Reviews of Geophysics - REV GEOPHYS* **46**.

Bourban, S. & Huang, F. (2018), *Introducing KHIONE – (Eulerian) Part I of the ice modelling component of TELEMAC, in the proceedings of the XXVth TELEMAC User Conference, Norwich, UK.*

Briand, J. F., Robillot, C., Quiblier-Llobéras, C., Humbert, J. F., Couté, A. & Bernard, C. (2002), 'Environmental context of Cylindrospermopsis raciborskii (Cyanobacteria) blooms in a shallow pond in France', *Water Research* **36**(13), 3183–3192.

Browning, T. J., Al-Hashem, A. A., Hopwood, M. J., Engel, A., Wakefield, E. D. & Achterberg, E. P. (2020), 'Nutrient regulation of late spring phytoplankton blooms in the midlatitude North Atlantic', *Limnology and Oceanography* **65**(6), 1136–1148.

Bruggeman, J. & Bolding, K. (2014), 'A general framework for aquatic biogeochemical models', *Environmental Modelling & Software* **61**(Supplement C), 249–265.

Burkey, J. (2020), 'Mann-kendall tau-b with sen's method (enhanced), matlab central !file exchange. retrieved february 12, 2020'.

Burr, T. & Skurikhin, A. (2013), 'Selecting Summary Statistics in Approximate Bayesian Computation for Calibrating Stochastic Models', *BioMed Research International* . ISSN: 2314-6133 Library Catalog: www.hindawi.com Pages: e210646 Publisher: Hindawi Volume: 2013.

Butcher, J. B., Nover, D., Johnson, T. E. & Clark, C. M. (2015), 'Sensitivity of lake thermal and mixing dynamics to climate change', *Climatic Change* **129**(1-2), 295–305.

Callendar, G. S. (1938), 'The artificial production of carbon dioxide and its influence on temperature', *Quarterly Journal of the Royal Meteorological Society* **64**(275), 223–240.

Canadian Hydraulic Centre (2011), *Blue Kenue reference manual*, Ottawa, Ontario, Canada.

Carey, C. C., Ibelings, B. W., Hoffmann, E. P., Hamilton, D. P. & Brookes, J. D. (2012), 'Eco-physiological adaptations that favour freshwater cyanobacteria in a changing climate', *Water Res* **46**(5), 1394–1407.

Carlson, R. (1977), 'A Trophic State Index for Lakes', *Limnology and Oceanography - LIMNOL OCEANOGR* **22**, 361–369.

Carmichael, W. W., Azevedo, S. M., An, J. S., Molica, R. J., Jochimsen, E. M., Lau, S., Rinehart, K. L., Shaw, G. R. & Eaglesham, G. K. (2001), 'Human fatalities from cyanobacteria: chemical and biological evidence for cyanotoxins.', *Environ Health Perspect* **109**(7), 663–668.

Catherine, Q., Susanna, W., Isidora, E.-S., Mark, H., Aurélie, V. & Jean-François, H. (2013), 'A review of current knowledge on toxic benthic freshwater cyanobacteria – Ecology, toxin production and risk management', *Water Research* **47**(15), 5464–5479.

Chanudet, V., Fabre, V. & van der Kaaij, T. (2012), 'Application of a three-dimensional hydrodynamic model to the Nam Theun 2 Reservoir (Lao PDR)', *Journal of Great Lakes Research* **38**(2), 260–269.

Chapman, S., Watson, J. E. M., Salazar, A., Thatcher, M. & McAlpine, C. A. (2017), 'The impact of urbanization and climate change on urban temperatures: a systematic review', *Landscape Ecol* **32**(10), 1921–1935.

Chapuis, M.-P., Raynal, L., Plantamp, C., Meynard, C. N., Blondin, L., Marin, J.-M. & Estoup, A. (2020), 'A young age of subspecific divergence in the desert locust inferred by ABC random forest', *Molecular Ecology* **29**(23), 4542–4558.

Chen, C. W. (1970), 'Concepts and Utilities of Ecologic Model', *Journal of the Sanitary Engineering Division* **96**(5), 1085–1097.

Codd, G. A. (2000), 'Cyanobacterial toxins, the perception of water quality, and the prioritisation of eutrophication control', *Ecological Engineering* **16**(1), 51–60.

Condie, S. & Webster, I. (2002), 'Stratification and Circulation in a Shallow Turbid Waterbody', *Environmental Fluid Mechanics* **2**, 177–196.

Cooper, A. J., Spearman, J. & Park, H. (2017), 'Validation of a TELEMAC-3D model of a seamount', p. 7.

Cottingham, K. L., Ewing, H. A., Greer, M. L., Carey, C. C. & Weathers, K. C. (2015), 'Cyanobacteria as biological drivers of lake nitrogen and phosphorus cycling', *Ecosphere* **6**(1), 1–19.

Cronin, J., Anandarajah, G. & Dessens, O. (2018), 'Climate change impacts on the energy system: a review of trends and gaps', *Climatic Change* **151**(2), 79–93.

Csilléry, K., Blum, M. G. B., Gaggiotti, O. E. & François, O. (2010), 'Approximate Bayesian Computation (ABC) in practice', *Trends in Ecology & Evolution* **25**(7), 410–418.

Cukier, R. I., Fortuin, C. M., Shuler, K. E., Petschek, A. G. & Schaibly, J. H. (1973), 'Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory', *J. Chem. Phys.* **59**(8), 3873–3878. Publisher: American Institute of Physics.

Dake, J. M. K. & Harleman, D. R. F. (1969), 'Thermal stratification in lakes: Analytical and laboratory studies', *Water Resources Research* **5**(2), 484–495.

de Carvalho Bueno, R. (2019), Wind-induced internal waves in closed basins: A laboratory experiment and field study, PhD thesis.

de Goede, E. (2020), 'Historical overview of 2D and 3D hydrodynamic modelling of shallow water flows in the Netherlands', *Ocean Dynamics* **70**.

Deltares (2014), *Delft3D-FLOW - Simulation of multi-dimensional hydrodynamic flows and transport phenomena, including sediments. User Manual Version: 3.15.30059*, Delft Hydraulics, Delft.

Deltares (2018), *D-Water Quality User Manual*, Delft Hydraulics, Delft.

Deng, J., Zhang, Y., Qin, B.-Q., Yao, X. & Deng, Y. (2017), 'Trends of publications related to climate change and lake research from 1991 to 2015', *Journal of Limnology* **76**.

Di Maggio, J., Fernández, C., Parodi, E. R., Diaz, M. S. & Estrada, V. (2016), 'Modeling phytoplankton community in reservoirs. A comparison between taxonomic and functional groups-based models', *Journal of Environmental Management* **165**, 31–52.

Diggle, P. J. & Gratton, R. J. (1984), 'Monte Carlo Methods of Inference for Implicit Statistical Models', *Journal of the Royal Statistical Society. Series B (Methodological)* **46**(2), 193–227. Publisher: [Royal Statistical Society, Wiley]. **URL:** *https://www.jstor.org/stable/2345504*

Dobiesz, N. & Lester, N. (2009), 'Changes in Mid-Summer Water Temperature and Clarity Across the Great Lakes between 1968 and 2002', *Journal of Great Lakes Research* **35**, 371–384.

Dominguez Almela, V., Palmer, S. C. F., Gillingham, P. K., Travis, J. M. J. & Britton, J. R. (2020), 'Integrating an individual-based model with approximate Bayesian computation to predict the invasion of a freshwater fish provides insights into dispersal and range expansion dynamics', *Biol Invasions* **22**(4), 1461–1480.

Downing, J. A. & McCauley, E. (1992), 'The nitrogen : phosphorus relationship in lakes', *Limnology and Oceanography* **37**(5), 936–945.

Downing, J. A., Prairie, Y. T., Cole, J. J., Duarte, C. M., Tranvik, L. J., Striegl, R. G., McDowell, W. H., Kortelainen, P., Caraco, N. F., Melack, J. M. & Middelburg, J. J. (2006), 'The global abundance and size distribution of lakes, ponds, and impoundments', *Limnology and Oceanography* **51**(5), 2388–2397.

Dupuis, A. P. & Hann, B. J. (2009), 'Warm spring and summer water temperatures in small eutrophic lakes of the Canadian prairies: potential implications for phytoplankton and zooplankton', *J Plankton Res* **31**(5), 489–502.

Durand, Y., Brun, E., Merindol, L., Guyomarc'h, G., Lesaffre, B. & Martin, E. (1993), 'A meteorological estimation of relevant parameters for snow models', *Annals of Glaciology* **18**, 65–71.

Elliott, J. A. (2012), 'Is the future blue-green? A review of the current model predictions of how climate change could affect pelagic freshwater cyanobacteria', *Water Research* **46**(5), 1364–1371.
URL: *http://www.sciencedirect.com/science/article/pii/S0043135411007901*

Elliott, J. A., Jones, I. D. & Thackeray, S. J. (2006), 'Testing the Sensitivity of Phytoplankton Communities to Changes in Water Temperature and Nutrient Load, in a Temperate Lake', *Hydrobiologia* **559**(1), 401–411.

Estoup, A., Raynal, L., Verdu, P. & Marin, J.-M. (2018), 'Model choice using approximate bayesian computation and random forests: analyses based on model grouping to make inferences about the genetic history of pygmy human populations', *Journal de la Société Française de Statistique* **159**(3), 167–190.

Fachini, A. & Vasconcelos, M. T. S. D. (2006), 'Enhancing diatom growth by using zeolites to change seawater composition', *Environ Sci Pollut Res Int* **13**(4), 238–241.

Fath, B., Scharler, U. & Jørgensen, S. (2012), Ecological modeling in environmental management: History and applications, *in* 'Treatise on Estuarine and Coastal Science'.

Feng, J. & Jodeau, M. (2016), 'Three-dimensional numerical modeling of sediment transport with TELEMAC-3D: validation of test cases', p. 9.

Fenocchi, A., Rogora, M., Morabito, G., Marchetto, A., Sibilla, S. & Dresti, C. (2019), 'Applicability of a one-dimensional coupled ecological-hydrodynamic numerical model to future projections in a very deep large lake (Lake Maggiore, Northern Italy/Southern Switzerland)', *Ecological Modelling* **392**, 38–51.

Flaim, G., Eccel, E., Zeileis, A., Toller, G., Cerasino, L. & Obertegger, U. (2016), 'Effects of re-oligotrophication and climate change on lake thermal structure', *Freshwater Biology* **61**(10), 1802–1814. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/fwb.12819.

Forel, F. A. (1901), *Handbuch der Seenkunde: allgemeine Limnologie*, J. Engelhorn, Stuttgart.

Fornarelli, R., Galelli, S., Castelletti, A., Antenucci, J. & Marti, C. (2013), 'An empirical modeling approach to predict and understand phytoplankton dynamics in a reservoir affected by interbasin water transfers', *Water Resources Research* **49**.

Francis, G. (1878), 'Poisonous Australian Lake', *Nature* **18**(444), 11–12.

Frassl, M. A., Boehrer, B., Holtermann, P. L., Hu, W., Klingbeil, K., Peng, Z., Zhu, J. & Rinke, K. (2018), 'Opportunities and Limits of Using Meteorological Reanalysis Data for Simulating Seasonal to Sub-Daily Water Temperature Dynamics in a Large Shallow Lake', *Water* **10**(5), 594.

Frumkin, H., Bratman, G. N., Breslow, S. J., Cochran, B., Kahn, P. H., Lawler, J. J., Levin, P. S., Tandon, P. S., Varanasi, U., Wolf, K. L. & Wood, S. A. (2017), 'Nature Contact and Human Health: A Research Agenda', *Environ. Health Perspect.* **125**(7), 075001.

Fu, H., Yuan, G., Özkan, K., Johansson, L. S., Søndergaard, M., Lauridsen, T. L. & Jeppesen, E. (2020), 'Seasonal and long-term trends in the spatial heterogeneity of lake phytoplankton communities over two decades of restoration and climate change', *Science of The Total Environment* **748**, 141106.

Gal, G., Hipsey, M., Parparov, A., Wagner, U., Makler, V. & Zohary, T. (2009), 'Implementation of ecological modeling as an effective management and investigation tool: Lake Kinneret as a case study', *Ecological Modelling* **220**, 1697–1718.

Gallagher, C., Lund, R. & Robbins, M. (2013), 'Changepoint Detection in Climate Time Series with Long-Term Trends', *J. Climate* **26**(14), 4994–5006.

Garnier, J. & Billen, G. (1993), 'Ecological interactions in a shallow sand-pit lake (créteil lake, france)', *Hydrobiologia* **275**, 97–114.

Gillooly, J. (2000), 'Effect of body size and temperature on generation time in zooplankton', *Journal of Plankton Research* **22**.

Gong, R., Xu, L., Wang, D., Li, H. & Xu, J. (2016), 'Water Quality Modeling for a Typical Urban Lake Based on the EFDC Model', *Environ Model Assess* **21**(5), 643–655.

Gosling, S., Warren, R., Arnell, N., Good, P., Caesar, J., Bernie, D., Lowe, J., van der Linden, P., O'Hanley, J. & Smith, S. (2011), 'A review of recent developments in climate change science. Part II: The global-scale impacts of climate change', *Progress in Physical Geography* **35**, 443–464.

Granéli, E., Weberg, M. & Salomon, P. S. (2008), 'Harmful algal blooms of allelopathic microalgal species: The role of eutrophication', *Harmful Algae* **8**(1), 94–102.

Grigorieva, E., Matzarakis, A. & De Freitas, C. (2010), 'Analysis of growing degree-days as climate impact indicator in a region with extreme annual air temperature amplitude', *Climate Research* **42**, 143–154.

Habets, F., Boone, A., Champeaux, J.-L., Etchevers, P., Franchisteguy, L., Leblois, E., Ledoux, E., Le Moigne, P., Martin, E., Morel, S., Noilhan, J., Quintana Seguí, P., Rousset, F. & Viennot, P. (2008), 'The SAFRAN-ISBA-MODCOU hydrometeorological model applied over France', *Journal of Geophysical Research* **113**.

Hadley, K. R., Paterson, A. M., Stainsby, E. A., Michelutti, N., Yao, H., Rusak, J. A., Ingram, R., McConnell, C. & Smol, J. P. (2014), 'Climate warming alters thermal stability but not stratification phenology in a small north-temperate lake', *Hydrological Processes* **28**(26), 6309–6319.

Hallegraeff, G. (1993), 'A Review of Harmful Algal Blooms and Their Apparent Global Increase', *Phycologia* **32**.

Hanna, M. (1998), 'Evaluation of Models Predicting Mixing Depth', *Canadian Journal of Fisheries and Aquatic Sciences* **47**, 940–947.

Hassall, C. (2014), 'The ecology and biodiversity of urban ponds', *Wiley Interdisciplinary Reviews: Water* **1**(2), 187–206.

Heil, C. A., Glibert, P. M. & Fan, C. (2005), 'Prorocentrum minimum (Pavillard) Schiller: A review of a harmful algal bloom species of growing worldwide importance', *Harmful Algae* **4**(3), 449–470.

Heinle, A. & Slawig, T. (2013), 'Internal dynamics of NPZD type ecosystem models', *Ecological Modelling* **254**, 33–42.

Heisler, J., Glibert, P., Burkholder, J., Anderson, D., Cochlan, W., Dennison, W., Gobler, C., Dortch, Q., Heil, C., Humphries, E., Lewitus, A., Magnien, R., Marshall, H., Sellner, K., Stockwell, D., Stoecker, D. & Suddleson, M. (2008), 'Eutrophication and Harmful Algal Blooms: A Scientific Consensus', *Harmful Algae* **8**(1), 3–13.

Henderson-Sellers, B. (1984), *Engineering limnology*, Pitman Advanced Pub. Program. Google-Books-ID: NeJRAAAAMAAJ.

Hervouet, J.-M. (2007), 'Hydrodynamics of Free Surface Flows: Modelling With the Finite Element Method', *Hydrodynamics of Free Surface Flows: Modelling with the finite element method* .

Higgins, S. L., Thomas, F., Goldsmith, B., Brooks, S. J., Hassall, C., Harlow, J., Stone, D., Völker, S. & White, P. (2019), 'Urban freshwaters, biodiversity, and human health and well-being: Setting an interdisciplinary research agenda', *WIREs Water* **6**(2), e1339.

Hill, M. J., Biggs, J., Thornhill, I., Briers, R. A., Gledhill, D. G., White, J. C., Wood, P. J. & Hassall, C. (2017), 'Urban ponds as an aquatic biodiversity resource in modified landscapes', *Global Change Biology* **23**(3), 986–999.

Hipsey, M. R., Bruce, L. C. & Hamilton, D. P. (2013), 'Aquatic Ecodynamics (AED) Model Library. Science Manual.'.

Hipsey, M. R., Gal, G., Arhonditsis, G. B., Carey, C. C., Elliott, J. A., Frassl, M. A., Janse, J. H., de Mora, L. & Robson, B. J. (2020), 'A system of metrics for the assessment and improvement of aquatic ecosystem models', *Environmental Modelling & Software* **128**, 104697.

Ho, J. C., Michalak, A. M. & Pahlevan, N. (2019), 'Widespread global increase in intense lake phytoplankton blooms since the 1980s', *Nature* **574**(7780), 667–670.

Hodges, B. (2014), Hydrodynamical Modeling, *in* 'Encyclopedia of Inland Waters'.

Huber, V., Adrian, R. & Gerten, D. (2008), 'Phytoplankton response to climate warming modified by trophic state', *Limnology and Oceanography* **53**(1), 1–13.

Huisman, J., Codd, G. A., Paerl, H. W., Ibelings, B. W., Verspagen, J. M. H. & Visser, P. M. (2018), 'Cyanobacterial blooms', *Nature Reviews Microbiology* **16**(8), 471.

Huisman, J. & Weissing, F. (2001), 'Fundamental Unpredictability in Multispecies Competition', *The American Naturalist* **157**, 488.

Humphries, S. E. & Lyne, V. D. (1988), 'Cyanophyte blooms: The role of cell buoyancy1', *Limnology and Oceanography* **33**(1), 79–91. _eprint: https://aslopubs.onlinelibrary.wiley.com/doi/pdf/10.4319/lo.1988.33.1.0079.

Ibelings, B. W., Backer, L. C., Kardinaal, W. E. A. & Chorus, I. (2015), 'Current approaches to cyanotoxin risk assessment and risk management around the globe', *Harmful Algae* **49**, 63–74.

Idso, S. B. (1973), 'On the concept of lake stability', *Limnology and Oceanography* **18**(4), 681–683.

Imberger, J. (1985), 'The diurnal mixed layer1', *Limnology and Oceanography* **30**(4), 737–770.

IPCC (1990), Climate change: The ipcc scientifi c assessment, Technical report, Cambridge, UK and New York, NY, USA.

IPCC (2014), Climate change 2014: Synthesis report. summary for policymakers, Technical report, Cambridge, UK and New York, NY, USA.

Jamet, S. & Corfee-Morlot, J. (2009), 'Assessing the Impacts of Climate Change: A Literature Review', *OECD, Economics Department, OECD Economics Department Working Papers* .

Jang, M.-H., Ha, K., Joo, G.-J. & Takamura, N. (2003), 'Toxin production of cyanobacteria is increased by exposure to zooplankton', *Freshwater Biology* **48**(9), 1540–1550. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1365-2427.2003.01107.x.

Jankowski, T., Livingstone, D. M., Bührer, H., Forster, R. & Niederhauser, P. (2006), 'Consequences of the 2003 European heat wave for lake temperature profiles, thermal stability, and hypolimnetic oxygen depletion: Implications for a warmer world', *Limnology and Oceanography* **51**(2), 815–819.

Jenny, J.-P., Francus, P., Normandeau, A., Lapointe, F., Perga, M.-E., Ojala, A., Schimmelmann, A. & Zolitschka, B. (2016), 'Global spread of hypoxia in freshwater ecosystems during the last three centuries is caused by rising local human pressure', *Global Change Biology* **22**(4), 1481–1489.

Jewel, A., Affan, M. & Khan, S. (2003), 'Fish Mortality Due to Cyanobacterial Bloom in an Aquaculture Pond in Bangladesh', *Pakistan Journal of Biological Sciences* **6**.

Jones, J. & Brett, M. T. (2014), Lake Nutrients, Eutrophication, and Climate Change, *in* B. Freedman, ed., 'Global Environmental Change', Handbook of Global Environmental Pollution, Springer Netherlands, Dordrecht, pp. 273–279.

Jørgensen, S. E. (2008), 'Overview of the model types available for development of ecological models', *Ecological Modelling* **215**(1), 3–9.

Jørgensen, S. E. & Bendoricchio, G. (2001), *Fundamentals of Ecological Modelling*, Elsevier. Google-Books-ID: hd7aPFIdvTcC.

Kaczmarek, Z., Strzepek, K. M., Somlyódy, L. & Priazhinskaya, V., eds (1996), *Water Resources Management in the Face of Climatic/Hydrologic Uncertainties*, Water Science and Technology Library, Springer Netherlands.

Kadlec, R. H. & Knight, R. L. (1996), Treatment wetlands, Lewis Publishers, New York, NY, USA.

Kangur, K., Kangur, A., Kangur, P. & Laugaste, R. (2005), 'Fish kill in Lake Peipsi in summer 2002 as a synergistic effect of a cyanobacterial bloom, high temperature, and low water level', *Proc. Estonian Acad. Sci. Biol. Ecol.* **54**, 67–80.

Kendall, M. (1975), *Rank correlation methods*, Rank correlation methods, Griffin, Oxford, England.

Kerimoglu, O. & Rinke, K. (2013), 'Stratification dynamics in a shallow reservoir under different hydro-meteorological scenarios and operational strategies', *Water Resour. Res.* **49**(11), 7518–7527.

Kienel, U., Kirillin, G., Brademann, B., Plessen, B., Lampe, R. & Brauer, A. (2017), 'Effects of spring warming and mixing duration on diatom deposition in deep Tiefer See, NE Germany', *J Paleolimnol* **57**(1), 37–49.

Kimura, N., Wu, C. H., Hoopes, J. A. & Tai, A. (2016), 'Diurnal Dynamics in a Small Shallow Lake under Spatially Nonuniform Wind and Weak Stratification', *Journal of Hydraulic Engineering* **142**(11), 04016047. Publisher: American Society of Civil Engineers.

Klimaszyk, P. & Rzymski, P. (2016), 'The complexity of ecological impacts induced by great cormorants', *Hydrobiologia* **771**(1), 13–30.

Kong, X.-Z., Seewald, M., Dadi, T., Friese, K., mi, C., Boehrer, B., Schultze, M., Rinke, K. & Shatwell, T. (2021), 'Unravelling winter diatom blooms in temperate lakes using high frequency data and ecological modeling', *Water Research* **190**, 116681.

Kozlowski, G. & Bondallaz, L. (2013), 'Urban aquatic ecosystems: Habitat loss and depletion of native macrophyte diversity during the 20th century in four Swiss cities', *Urban Ecosyst* **16**(3), 543–551.

Krabbenhoft, D. P., Bowser, C. J., Anderson, M. P. & Valley, J. W. (1990), 'Estimating groundwater exchange with lakes: 1. The stable isotope mass balance method', *Water Resources Research* **26**(10), 2445–2453.

Kraemer, B., Anneville, O., Chandra, S., Dix, M., Kuusisto, E., Livingstone, D., Rimmer, A., Schladow, S., Silow, E., Sitoki, L., Tamatamah, R., Vadeboncoeur, Y. & McIntyre, P. (2015), 'Morphometry and average temperature affect lake stratification responses to climate change: Lake Stratification Responses to Climate', *Geophysical Research Letters* **42**.

Kriest, I. (2017), 'Calibration of a simple and a complex model of global marine biogeochemistry', *Biogeosciences* **14**, 4965–4984.

Krinos, A. I., Farrell, K. J., Daneshmand, V., Subratie, K. C., Figueiredo, R. J. & Carey, C. C. (2019), 'Including variability in air temperature warming scenarios in a lake simulation model highlights uncertainty in predictions of cyanobacteria', *bioRxiv* p. 734285.

Kruschke, J. K. (2018), 'Rejecting or Accepting Parameter Values in Bayesian Estimation', *Advances in Methods and Practices in Psychological Science* **1**(2), 270–280. Publisher: SAGE Publications Inc.

Kurmayer, R. & Christiansen, G. (2009), 'The Genetic Basis of Toxin Production in Cyanobacteria', *frer* **2**(1), 31–50. Publisher: Freshwater Biological Association.

Lagarrigues, G., Jabot, F., Lafond, V. & Courbaud, B. (2015), 'Approximate Bayesian computation to recalibrate individual-based models with population data: Illustration with a forest simulation model', *Ecological Modelling* **306**, 278–286.

Landsberg, J. H., Hendrickson, J., Tabuchi, M., Kiryu, Y., Williams, B. J. & Tomlinson, M. C. (2020), 'A large-scale sustained fish kill in the St. Johns River, Florida: A complex consequence of cyanobacteria blooms', *Harmful Algae* **92**, 101771.

Latour, D., Salençon, M., Reyss, J.-L. & Giraudet, H. (2007), 'Sedimentary imprint of Microcystis Aeruginosa (Cyanobacteria) blooms in grangent reservoir (Loire, FRANCE)', *Journal of Phycology* **43**(3), 417–425. Publisher: Wiley.

Laurent, M., Gouze, E., Emile, R., Jean-Michel, H., Durand, N. & Pham, C.-T. (2010), *3D coupled modeling of hydrodynamics and water quality in the Berre Lagoon (France)*, Vol. 1.

Layden, A., MacCallum, S. & Merchant, C. (2016), 'Determining lake surface water temperatures worldwide using a tuned one-dimensional lake model (FLake, v1)', *Geoscientific Model Development* **9**, 2167–2189.

Le Moal, M., Gascuel-Odoux, C., Ménesguen, A., Souchon, Y., Étrillard, C., Levain, A., Moatar, F., Pannard, A., Souchu, P., Lefebvre, A. & Pinay, G. (2019), 'Eutrophication: A new wine in an old bottle?', *Science of The Total Environment* **651**, 1–11.

Le Treut, H., Somerville, R., Cubasch, U., Ding, Y., Mauritzen, C., Mokossit, A. Peterson, T. & Prather, M. (2007), *Historical Overview of Climate Change. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, UK and New York, NY, USA.

Lee, R., Biggs, T. & Fang, X. (2018), 'Thermal and Hydrodynamic Changes under a Warmer Climate in a Variably Stratified Hypereutrophic Reservoir', *Water* **10**, 1–24.

Leendertse, J. J. (1967), 'Aspects of a Computational Model for Long-Period Water-Wave Propagation:'.

Leendertse, J. J. & Liu, D. S.-K. (1975), 'A Three-Dimensional Model for Estuaries and Coastal Seas: Volume II, Aspects of Computation'.

Lemonsu, A., Kounkou-Arnaud, R., Desplat, J., Salagnac, J.-L. & Masson, V. (2013), 'Evolution of the Parisian urban climate under a global changing climate', *Climatic Change* **116**.

Lewandowska, A. & Sommer, U. (2010), 'Climate change and the spring bloom: a mesocosm study on the influence of light and temperature on phytoplankton and mesozooplankton', *MEPS* **405**, 101–111.

Lewis, W. M. J. (1983), 'A revised classification of lakes based on mixing', *Canadian journal of fisheries and aquatic sciences = Journal canadien des sciences halieutiques et aquatiques* .
URL: *https://agris.fao.org/agris-search/search.do?recordID=US201302129823*

Li, Y. & Hipsey, M. R. (2013), 'The importance of model structural complexity when simulating aquatic food webs'.

Lindeman, R. L. (1942), 'The Trophic-Dynamic Aspect of Ecology', *Ecology* **23**(4), 399–417.

Liu, M., Zhang, Y., Shi, K., Zhu, G., Wu, Z., Liu, M. & Zhang, Y. (2019), 'Thermal stratification dynamics in a large and deep subtropical reservoir revealed by high-frequency buoy data', *Science of The Total Environment* **651**, 614–624.

Livingstone, D. M. (2003), 'Impact of Secular Climate Change on the Thermal Structure of a Large Temperate Central European Lake', *Climatic Change* **57**(1), 205–225.

Los, F. H. (2009), Eco-hydrodynamic modelling of primary production in coastal waters and lakes using BLOOM., PhD thesis, Wageningen Universiteit.

Luo, L., Hamilton, D., Lan, J., McBride, C. & Trolle, D. (2018), 'Autocalibration of a one-dimensional hydrodynamic-ecological model (DYRESM 4.0-CAEDYM 3.1) using a Monte Carlo approach: simulations of hypoxic events in a polymictic lake', *Geoscientific Model Development* **11**(3), 903–913.

Lyche Solheim, A., Globevnik, L., Austnes, K., Kristensen, P., Moe, S. J., Persson, J., Phillips, G., Poikane, S., van de Bund, W. & Birk, S. (2019), 'A new broad typology for rivers and lakes in Europe: Development and application for large-scale environmental assessments', *Science of The Total Environment* **697**, 134043.

Lürling, M., Eshetu, F., Faassen, E. J., Kosten, S. & Huszar, V. L. M. (2013), 'Comparison of cyanobacterial and green algal growth rates at different temperatures', *Freshwater Biology* **58**(3), 552–559.

Magee, M. R. & Wu, C. H. (2017), 'Response of water temperatures and stratification to changing climate in three lakes with different morphometry', *Hydrology and Earth System Sciences* **21**(12), 6253–6274.

Mahevas, S., Picheny, V., Lambert, P., Dumoulin, N., Rouan, L., Soulié, J.-C., Brockhoff, D., Lehuta, S., Riche, R. L., Faivre, R. & Drouineau, H. (2019), 'A Practical Guide for Conducting Calibration and Decision-Making Optimisation with Complex Ecological Models', *Publisher: Preprints* .

Maier, D. B., Diehl, S. & Bigler, C. (2019), 'Interannual variation in seasonal diatom sedimentation reveals the importance of late winter processes and their timing for sediment signal formation', *Limnology and Oceanography* **64**(3), 1186–1199. _eprint: https://aslopubs.onlinelibrary.wiley.com/doi/pdf/10.1002/lno.11106.

Makler-Pick, V., Gal, G., Gorfine, M., Hipsey, M. R. & Carmel, Y. (2011), 'Sensitivity analysis for complex ecological models – A new approach', *Environmental Modelling & Software* **26**(2), 124–134.

Malcherek, A. (2000), 'Application of TELEMAC-2D in a narrow estuarine tributary', *Hydrological Processes - HYDROL PROCESS* **14**, 2293–2300.

Mann, H. B. (1945), 'Nonparametric Tests Against Trend', *Econometrica* **13**(3), 245–259.

Mariani, L., Parisi, S. G., Cola, G. & Failla, O. (2012), 'Climate change in Europe and effects on thermal resources for crops', *Int J Biometeorol* **56**(6), 1123–1134.

Marjoram, P., Molitor, J., Plagnol, V. & Tavaré, S. (2003), 'Markov chain Monte Carlo without likelihoods', *PNAS* **100**(26), 15324–15328. Publisher: National Academy of Sciences Section: Physical Sciences.

Martinsen, K. T., Andersen, M. R. & Sand-Jensen, K. (2019), 'Water temperature dynamics and the prevalence of daytime stratification in small temperate shallow lakes', *Hydrobiologia* **826**(1), 247–262.

McCombs, M. P., Mulligan, R. P. & Boegman, L. (2014), 'Offshore wind farm impacts on surface waves and circulation in Eastern Lake Ontario', *Coastal Engineering* **93**, 32–39.

McEnroe, N. A., Buttle, J. M., Marsalek, J., Pick, F. R., Xenopoulos, M. A. & Frost, P. C. (2013), 'Thermal and chemical stratification of urban ponds: Are they 'completely mixed reactors'?', *Urban Ecosyst* **16**(2), 327–339.

McMaster, G. S. & Wilhelm, W. W. (1997), 'Growing degree-days: one equation, two interpretations', *Agricultural and Forest Meteorology* **87**(4), 291–300.

Meerhoff, M. & Jeppesen, E. (2009), Shallow Lakes and Ponds, *in* G. E. Likens, ed., 'Encyclopedia of Inland Waters', Academic Press, Oxford, pp. 645–655.

Mendonça, R., Müller, R. A., Clow, D., Verpoorter, C., Raymond, P., Tranvik, L. J. & Sobek, S. (2017), 'Organic carbon burial in global lakes and reservoirs', *Nature Communications* **8**(1), 1694.

Merkel, U. H. (2019), 'Thermal Stratification in Small Lakes with TELEMAC-3D: Showcase "Lake Monsterloch"', *XXVIth TELEMAC-MASCARET User Conference, 15th to 17th October 2019, Toulouse* . Accepted: 2020-07-07T15:10:18Z.

Mooij, W., Hülsmann, S., de Senerpont Domis, L., Nolet, B., Bodelier, P., Boers, P., Dionisio Pires, M., Gons, H., Ibelings, B., Noordhuis, R., Portielje, R., Wolfstein, K. & Lammens, E. (2005), 'The impact of climate change on lakes in the Netherlands: A review', *Aquatic Ecology* **39**, 381–400.

Mooij, W. M., Janse, J. H., Domis, L. N. D. S., Hülsmann, S. & Ibelings, B. W. (2007), 'Predicting the effect of climate change on temperate shallow lakes with the ecosystem model PCLake', *Hydrobiologia* **584**(1), 443–454.

Mooij, W. M., Trolle, D., Jeppesen, E., Arhonditsis, G., Belolipetsky, P. V., Chitamwebwa, D. B. R., Degermendzhy, A. G., DeAngelis, D. L., Domis, L. N. D. S., Downing, A. S., Elliott, J. A., Fragoso, C. R., Gaedke, U., Genova, S. N., Gulati, R. D., Håkanson, L., Hamilton, D. P., Hipsey, M. R., Hoen, J., Hülsmann, S., Los, F. H., Makler-Pick, V., Petzoldt, T., Prokopkin, I. G., Rinke, K., Schep, S. A., Tominaga, K., Dam, A. A. V., Nes, E. H. V., Wells, S. A. & Janse, J. H. (2010), 'Challenges and opportunities for integrating lake ecosystem modelling approaches', *Aquat Ecol* **44**(3), 633–667.

Moore, J. W., Schindler, D. E., Scheuerell, M. D., Smith, D. & Frodge, J. (2003), 'Lake Eutrophication at the Urban Fringe, Seattle Region, USA', *ambi* **32**(1), 13–18.

Moras, S., Ayala, A. I. & Pierson, D. C. (2019), 'Historical modelling of changes in Lake Erken thermal conditions', *Hydrology and Earth System Sciences* **23**(12), 5001–5016.

Murakami, M., Oonishi, Y. & Kunishi, H. (1985), 'A numerical simulation of the distribution of water temperature and salinity in the Seto Inland Sea', *Journal of the Oceanographical Society of Japan* **41**(4), 213–224.

Naselli-Flores, L. (2008), Urban Lakes: Ecosystems at Risk, Worthy of the Best Care, Sengupta, M. and Dalwani, R., pp. 1333–1337.

Neuheimer, A. B. & Taggart, C. T. (2007), 'The growing degree-day and fish size-at-age: The overlooked metric', *Canadian Journal of Fisheries and Aquatic Sciences* **64**, 375–385.

Newton, R. J., Jones, S. E., Eiler, A., McMahon, K. D. & Bertilsson, S. (2011), 'A Guide to the Natural History of Freshwater Lake Bacteria', *Microbiol Mol Biol Rev* **75**(1), 14–49.

Nezu, I. & Nakagawa, H. (1994), *Turbulence in Open-Channel Flows*. Journal Abbreviation: Turbulence in Open Channel Flows Pages: 281 Publication Title: Turbulence in Open Channel Flows.

Niedrist, G., Psenner, R. & Sommaruga, R. (2018), 'Climate warming increases vertical and seasonal water temperature differences, and inter-annual variability in a mountain lake', *Climatic Change* .

Noble, A. & Hassall, C. (2015), 'Poor ecological quality of urban ponds in northern England: causes and consequences', *Urban Ecosystems* **18**, 649–662.

Nogrady, B. (2019), 'Mass fish deaths in Australia set to continue', *Nature* . Publisher: Nature Publishing Group.

Nuwer, R. (2020), '356 Elephants Dropped Dead. Did This Bacteria Poison Them?', *The New York Times* .
URL: *https://www.nytimes.com/2020/09/23/science/dead-elephants-botswana.html*

Okubo, A. (1971), 'Oceanic diffusion diagrams', *Deep Sea Research and Oceanographic Abstracts* **18**(8), 789–802.

O'Reilly, C. M. et al. (2015), 'Rapid and highly variable warming of lake surface waters around the globe', *Geophysical Research Letters* **42**(24), 10,773–10,781.

O'Neil, J. M., Davis, T. W., Burford, M. A. & Gobler, C. J. (2012), 'The rise of harmful cyanobacteria blooms: The potential roles of eutrophication and climate change', *Harmful Algae* **14**, 313–334.

Paerl, H. & Huisman, J. (2008), 'Blooms Like It Hot', *Science (New York, N.Y.)* **320**, 57–8.

Paerl, H. & Otten, T. (2013), 'Harmful Cyanobacterial Blooms: Causes, Consequences, and Controls', *Microbial ecology* **65**.

Paerl, H. W. (2014), 'Mitigating Harmful Cyanobacterial Blooms in a Human- and Climatically-Impacted World', *Life* **4**(4), 988.

Paerl, H. W. & Paul, V. J. (2012), 'Climate change: Links to global expansion of harmful cyanobacteria', *Water Research* **46**(5), 1349–1363.

Paerl Hans W. (2003), 'Nuisance phytoplankton blooms in coastal, estuarine, and inland waters1', *Limnology and Oceanography* **33**(4part2), 823–843.

Pannard, A., Beisner, B. E., Bird, D. F., Braun, J., Planas, D. & Bormans, M. (2011), 'Recurrent internal waves in a small lake: Potential ecological consequences for metalimnetic phytoplankton populations', *Limnology and Oceanography: Fluids and Environments* **1**(1), 91–109.

Pareeth, S., Salmaso, N., Adrian, R. & Neteler, M. (2016), 'Homogenised daily lake surface water temperature data generated from multiple satellite sensors: A long-term case study of a large sub-Alpine lake', *Scientific Reports* **6**(1), 31251.

Peeters, F., Straile, D., Lorke, A. & LIVINGSTONE, D. (2007), 'Earlier onset of the spring phytoplankton bloom in lakes of the temperate zone in a warmer climate', *Global Change Biology* **13**, 1898–1909.

Perrier, F., Le Mouël, J., Poirier, J. & Shnirman, M. (2005), 'Long-term climate change and surface versus underground temperature measurements in Paris', *Long-term climate change and surface versus underground temperature measurements in Paris* **25**(12), 1619–1631.

Perroud, M. & Goyette, S. (2010), 'Impact of warmer climate on Lake Geneva Water-temperature profiles', *Boreal Environment Research* **15**.

Pfeifroth, U., Sanchez-Lorenzo, A., Manara, V., Trentmann, J. & Hollmann, R. (2018), 'Trends and Variability of Surface Solar Radiation in Europe Based On Surface- and Satellite-Based Data Records', *Journal of Geophysical Research: Atmospheres* **123**(3), 1735–1754.

Piccioni, F., Casenave, C., Lemaire, B. J., Le Moigne, P., Dubois, P. & Vinçon-Leite, B. (2021), 'The thermal response of small and shallow lakes to climate change: new insights from 3D hindcast modelling', *Earth System Dynamics* **12**(2), 439–456. Publisher: Copernicus GmbH.
URL: *https://esd.copernicus.org/articles/12/439/2021/*

Piccolroaz, S., Amadori, M., Toffolon, M. & Dijkstra, H. A. (2019), 'Importance of planetary rotation for ventilation processes in deep elongated lakes: Evidence from Lake Garda (Italy)', *Scientific Reports* **9**(1), 1–11.

Piccolroaz, S., Healey, N. C., Lenters, J. D., Schladow, S. G., Hook, S. J., Sahoo, G. B. & Toffolon, M. (2018), 'On the predictability of lake surface temperature using air temperature in a changing climate: A case study for Lake Tahoe (U.S.A.)', *Limnol. Oceanogr.* **63**(1), 243–261.

Piccolroaz, S., Toffolon, M. & Majone, B. (2013), 'A simple lumped model to convert air temperature into surface water temperature in lakes', *Hydrology and Earth System Sciences* **17**(8), 3323–3338. Publisher: Copernicus GmbH.

Piccolroaz, S., Woolway, R. I. & Merchant, C. J. (2020), 'Global reconstruction of twentieth century lake surface water temperature reveals different warming trends depending on the climatic zone', *Climatic Change* .

Plec, D., Soulignac, F., Porto, L., Lemaire, B. J., Dubois, P., Saad, M., Huguenard, L., Bernard, C., Nascimento, N. d. O. & Vinçon-Leite, B. (2017), Modelling Short-term cyanobacterial Dynamics in a small urban Lake using a coupled hydrodynamicecological three-dimensional model, *in* 'SimHydro 2017: Choosing the right model in applied hydraulics, 14-16 June 2017, Nice'.
URL: *https://hal-enpc.archives-ouvertes.fr/hal-01692215*

Poikane, S., Kelly, M. G., Salas Herrero, F., Pitt, J.-A., Jarvie, H. P., Claussen, U., Leujak, W., Lyche Solheim, A., Teixeira, H. & Phillips, G. (2019), 'Nutrient criteria for surface waters under the European Water Framework Directive: Current state-of-the-art, challenges and future outlook', *Science of The Total Environment* **695**, 133888.

Poli, A. A. & Cirillo, M. C. (1993), 'On the use of the normalized mean square error in evaluating dispersion model performance', *Atmospheric Environment. Part A. General Topics* **27**(15), 2427–2434.

Polli, B. A., Bleninger, T., Polli, B. A. & Bleninger, T. (2019), 'Comparison of 1D and 3D reservoir heat transport models and temperature effects on mass transport', *RBRH* **24**. Publisher: Associação Brasileira de Recursos Hídricos.

Poole, H. H. & Atkins, W. R. G. (1929), 'Photo-electric Measurements of Submarine Illumination throughout the Year', *Journal of the Marine Biological Association of the United Kingdom* **16**(1), 297–324.

Prangle, D. (2015), 'Summary Statistics in Approximate Bayesian Computation', *arXiv:1512.05633 [math, stat]* . arXiv: 1512.05633.

Putt, A. E., MacIsaac, E. A., Herunter, H. E., Cooper, A. B. & Selbie, D. T. (2019), 'Eutrophication forcings on a peri-urban lake ecosystem: Context for integrated watershed to airshed management', *PLOS ONE* **14**(7), e0219241.

Quintana-Seguí, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L. & Morel, S. (2008), 'Analysis of Near-Surface Atmospheric Variables: Validation of the SAFRAN Analysis over France', *J. Appl. Meteor. Climatol.* **47**(1), 92–107.

Raick, C., Soetaert, K. & Grégoire, M. (2006), 'Model complexity and performance: How far can we simplify?', *Progress in Oceanography* **70**(1), 27–57.

Raimonet, M., Oudin, L., Thieu, V., Silvestre, M., Vautard, R., Rabouille, C. & Le Moigne, P. (2017), 'Evaluation of Gridded Meteorological Datasets for Hydrological Modeling', *J. Hydrometeor.* **18**(11), 3027–3041.

Ralston, D. K., Keafer, B. A., Brosnahan, M. L. & Anderson, D. M. (2014), 'Temperature dependence of an estuarine harmful algal bloom: Resolving interannual variability in bloom dynamics using a degree day approach', *Limnol. Oceanogr.* **59**(4), 1112–1126.

Rasconi, S., Winter, K. & Kainz, M. J. (2017), 'Temperature increase and fluctuation induce phytoplankton biodiversity loss – Evidence from a multi-seasonal mesocosm experiment', *Ecology and Evolution* **7**(9), 2936–2946. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ece3.2889.

Rast, W. & Thornton, J. A. (1996), 'Trends in eutrophication research and control', *Hydrological Processes* **10**(2), 295–313.

Raynal, L., Marin, J.-M., Pudlo, P., Ribatet, M., Robert, C. P. & Estoup, A. (2019), 'ABC random forests for Bayesian parameter inference', *Bioinformatics* **35**(10), 1720–1728. Publisher: Oxford Academic.

Reavie, E. D., Cai, M., Twiss, M. R., Carrick, H. J., Davis, T. W., Johengen, T. H., Gossiaux, D., Smith, D. E., Palladino, D., Burtner, A. & Sgro, G. V. (2016), 'Winter–spring diatom production in Lake Erie is an important driver of summer hypoxia', *Journal of Great Lakes Research* **42**(3), 608–618.

Redfield, A. C. (1934), *On the Proportions of Organic Derivatives in Sea Water and Their Relation to the Composition of Plankton*, University Press of Liverpool.

Reichert, P. & Omlin, M. (1997), 'On the usefulness of overparameterized ecological models', *Ecological Modelling* **95**(2), 289–299.

Reid, P. C., Hari, R. E., Beaugrand, G., Livingstone, D. M., Marty, C., Straile, D., Barichivich, J., Goberville, E., Adrian, R., Aono, Y., Brown, R., Foster, J., Groisman, P., Hélaouët, P., Hsu, H.-H., Kirby, R., Knight, J., Kraberg, A., Li, J., Lo, T.-T., Myneni, R. B., North, R. P., Pounds, J. A., Sparks, T., Stübi, R., Tian, Y., Wiltshire, K. H., Xiao, D. & Zhu, Z. (2016), 'Global impacts of the 1980s regime shift', *Global Change Biology* **22**(2), 682–703.

Reynolds, C. S. (2006), 'The Ecology of Phytoplankton'.

Reynolds, C. S., Huszar, V., Kruk, C., Naselli-Flores, L. & Melo, S. (2002), 'Towards a functional classification of the freshwater phytoplankton', *J Plankton Res* **24**(5), 417–428.

Reynolds Colin S. (1980), 'Phytoplankton assemblages and their periodicity in stratifying lake systems', *Ecography* **3**(3), 141–159.

Rigosi, A., Marcé, R., Escot, C. & Rueda, F. J. (2011), 'A calibration strategy for dynamic succession models including several phytoplankton groups', *Environmental Modelling & Software* **26**(6), 697–710.

Romero, J., Dallimore, C., Antenucci, J., Hamilton, D., Imberger, J., Horn, D. & Deen, A. (2002), 'Application of 1D and 3D Hydrodynamic Models Coupled to an Ecological Model to Two Water Supply Reservoirs'.

Romero, J. R., Antenucci, J. P. & Imberger, J. (2004), 'One- and three-dimensional biogeochemical simulations of two differing reservoirs', *Ecological Modelling* **174**(1), 143–160.

Rousseau, D. (2009), 'Paris temperatures: monthly means 1676 to 2008'.

Rousseau, D. (2013), 'Les moyennes mensuelles de températures à paris de 1658 à 1675 - annexe i'.

Rubin, D. B. (1984), 'Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician', *Ann. Statist.* **12**(4), 1151–1172. Publisher: Institute of Mathematical Statistics.

Sahoo, G. B., Forrest, A. L., Schladow, S. G., Reuter, J. E., Coats, R. & Dettinger, M. (2016), 'Climate change impacts on lake thermal dynamics and ecosystem vulnerabilities', *Limnology and Oceanography* **61**(2), 496–507.

Sayer, C., Roberts, N., Sadler, J., David, C. & Wade, P. (2009), 'Biodiversity changes in a shallow lake ecosystem: A multi-proxy palaeolimnological analysis', *Journal of Biogeography* **26**, 97–114.

Schlenker, W., Hanemann, W. M. & Fisher, A. C. (2007), 'Water Availability, Degree Days, and the Potential Impact of Climate Change on Irrigated Agriculture in California', *Climatic Change* **81**(1), 19–38.

Schmid, M., Hunziker, S. & Wüest, A. (2014), 'Lake surface temperatures in a changing climate: a global sensitivity analysis', *Climatic Change* **124**(1), 301–315.

Schneider, P. & Hook, S. J. (2010), 'Space observations of inland water bodies show rapid surface warming since 1985', *Geophysical Research Letters* **37**(22).

Schneider, P., Hook, S. J., Radocinski, R. G., Corlett, G. K., Hulley, G. C., Schladow, S. G. & Steissberg, T. E. (2009), 'Satellite observations indicate rapid warming trend for lakes in California and Nevada', *Geophysical Research Letters* **36**(22).

Schopf, J. W. (1993), 'Microfossils of the Early Archean Apex chert: new evidence of the antiquity of life', *Science* **260**, 640–646.

Schuergers, N., Mullineaux, C. W. & Wilde, A. (2017), 'Cyanobacteria in motion', *Current Opinion in Plant Biology* **37**, 109–115.

Sen, P. K. (1968), 'Estimates of the Regression Coefficient Based on Kendall's Tau', *Journal of the American Statistical Association* **63**(324), 1379–1389.

Sharma, S., Blagrave, K., Magnuson, J. J., O'Reilly, C. M., Oliver, S., Batt, R. D., Magee, M. R., Straile, D., Weyhenmeyer, G. A., Winslow, L. & Woolway, R. I. (2019), 'Widespread loss of lake ice around the Northern Hemisphere in a warming world', *Nature Climate Change* **9**(3), 227–231.

Sharma, S., Gray, D. K., Read, J. S., O'Reilly, C. M., Schneider, P., Qudrat, A., Gries, C., Stefanoff, S., Hampton, S. E., Hook, S., Lenters, J. D., Livingstone, D. M., McIntyre, P. B., Adrian, R., Allan, M. G., Anneville, O., Arvola, L., Austin, J., Bailey, J., Baron, J. S., Brookes, J., Chen, Y., Daly, R., Dokulil, M., Dong, B., Ewing, K., de Eyto, E., Hamilton, D., Havens, K., Haydon, S., Hetzenauer, H., Heneberry, J., Hetherington, A. L., Higgins, S. N., Hixson, E., Izmest'eva, L. R., Jones, B. M., Kangur, K., Kasprzak, P., Köster, O., Kraemer, B. M., Kumagai, M., Kuusisto, E., Leshkevich, G., May, L., MacIntyre, S., Müller-Navarra, D., Naumenko, M., Noges, P., Noges, T., Niederhauser, P., North, R. P., Paterson, A. M., Plisnier, P.-D., Rigosi, A., Rimmer, A., Rogora, M., Rudstam, L., Rusak, J. A., Salmaso, N., Samal, N. R., Schindler, D. E., Schladow, G., Schmidt, S. R., Schultz, T., Silow, E. A., Straile, D., Teubner, K., Verburg, P., Voutilainen, A., Watkinson, A., Weyhenmeyer, G. A., Williamson, C. E. & Woo, K. H. (2015), 'A global database of lake surface temperatures collected by in situ and satellite methods from 1985–2009', *Scientific Data* **2**(1), 150008.

Shatwell, T., Thiery, W. & Kirillin, G. (2019), 'Future projections of temperature and mixing regime of European temperate lakes', *Hydrology and Earth System Sciences* **23**(3), 1533–1551. Publisher: Copernicus GmbH.

Shaw, R. D., Shaw, J. F. H., Fricker, H. & Prepas, E. E. (1990), 'An integrated approach to quantify groundwater transport of phosphorus to Narrow Lake, Alberta', *Limnology and Oceanography* **35**(4), 870–886.

Sheath, R. G. & Wehr, J. D. (2003), 1 - INTRODUCTION TO FRESHWATER ALGAE, *in* J. D. Wehr & R. G. Sheath, eds, 'Freshwater Algae of North America', Aquatic Ecology, Academic Press, Burlington, pp. 1–9.
URL: *http://www.sciencedirect.com/science/article/pii/B9780127415505500027*

Sherman, E., Moore, J. K., Primeau, F. & Tanouye, D. (2016), 'Temperature influence on phytoplankton community growth rates', *Global Biogeochemical Cycles* **30**(4), 550–559.

Shiklomanov, I. A. (1998), 'World water resources, a new appraisal and assessment for the 21st Century'.

Shiklomanov, I. & Rodda, J. (2004), 'World Water Resources at the Beginning of the Twenty-First Century', **13**.

Shimoda, Y. & Arhonditsis, G. B. (2016), 'Phytoplankton functional type modelling: Running before we can walk? A critical evaluation of the current state of knowledge', *Ecological Modelling* **320**(Supplement C), 29–43.

Silva, A. A. & Souza-Echer, M. P. (2016), 'Ground-based observations of clouds through both an automatic imager and human observation', *Meteorological Applications* **23**(1), 150–157.

Simola, U., Cisewski-Kehe, J., Gutmann, M. U. & Corander, J. (2021), 'Adaptive Approximate Bayesian Computation Tolerance Selection', *Bayesian Analysis* **-1**(-1), 1–27. Publisher: International Society for Bayesian Analysis.

Sisson, S. A., Fan, Y. & Tanaka, M. M. (2007), 'Sequential Monte Carlo without likelihoods', *PNAS* **104**(6), 1760–1765. ISBN: 9780607208108 Publisher: National Academy of Sciences Section: Physical Sciences.

Sivonen, K. (1996), 'Cyanobacterial toxins and toxin production', *Phycologia* **35**(sup6), 12–24.

Soares, L. M. V., do Carmo Calijuri, M., das Graças Silva, T. F. & de Moraes Novo, E. M. L. (2021), 'Climate change enhances deepwater warming of subtropical reservoirs: evidence from hydrodynamic modelling', *Climatic Change* **166**(1), 21.

Sobol, I. M. (2001), 'Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates', *Mathematics and Computers in Simulation* **55**(1), 271–280.

Sommer, U., Gliwicz, Z. M., Lampert, W. & Duncan, A. (1986), 'The PEG-model of seasonal succession of planktonic events in fresh waters', *Archiv für Hydrobiologie* **106**(4), 433–471.

Sommer, U. & Lengfellner, K. (2008), 'Climate change and the timing, magnitude, and composition of the phytoplankton spring bloom', *Global Change Biology* **14**(6), 1199–1208.

Song, K., Xenopoulos, M. A., Buttle, J. M., Marsalek, J., Wagner, N. D., Pick, F. R. & Frost, P. C. (2013), 'Thermal stratification patterns in urban ponds and their relationships with vertical nutrient gradients', *Journal of Environmental Management* **127**, 317–323.

Soulignac, F., Danis, P.-A., Bouffard, D., Chanudet, V., Dambrine, E., Guénand, Y., Harmel, T., Ibelings, B. W., Trevisan, D., Uittenbogaard, R. & Anneville, O. (2018), 'Using 3D modeling and remote sensing capabilities for a better understanding of spatio-temporal heterogeneities of phytoplankton abundance in large lakes', *Journal of Great Lakes Research* **44**(4), 756–764.

Soulignac, F., Vinçon-Leite, B., Lemaire, B. J., Scarati Martins, J. R., Bonhomme, C., Dubois, P., Mezemate, Y., Tchiguirinskaia, I., Schertzer, D. & Tassin, B. (2017), 'Performance Assessment of a 3D Hydrodynamic Model Using High Temporal Resolution Measurements in a Shallow Urban Lake', *Environ Model Assess* **22**(4), 309–322.

Spilling, K., Olli, K., Lehtoranta, J., Kremp, A., Tedesco, L., Tamelander, T., Klais, R., Peltonen, H. & Tamminen, T. (2018), 'Shifting Diatom—Dinoflagellate Dominance During Spring Bloom in the Baltic Sea and its Potential Effects on Biogeochemical Cycling', *Front. Mar. Sci.* **5**.

Stelling, G. S. & Leendertse, J. J. (1992), Approximation of Convective Processes by Cyclic AOI Methods, ASCE, pp. 771–782.

Stetler, J. T., Girdner, S., Mack, J., Winslow, L. A., Leach, T. H. & Rose, K. C. (2020), 'Atmospheric stilling and warming air temperatures drive long-term changes in lake stratification in a large oligotrophic lake', *Limnology and Oceanography* .

Stumpf, M. P. (2014), 'Approximate Bayesian inference for complex ecosystems', *F1000Prime Rep* **6**.

Sunnåker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M. & Dessimoz, C. (2013), 'Approximate Bayesian Computation', *PLOS Computational Biology* **9**(1), e1002803. Publisher: Public Library of Science.

Søndergaard, M., Lauridsen, T. L., Johansson, L. S. & Jeppesen, E. (2018), 'Gravel pit lakes in Denmark: Chemical and biological state', *Sci. Total Environ.* **612**, 9–17.

Tan, Z., Yao, H. & Zhuang, Q. (2018), 'A Small Temperate Lake in the 21st Century: Dynamics of Water Temperature, Ice Phenology, Dissolved Oxygen, and Chlorophyll a', *Water Resources Research* **54**(7), 4681–4699. _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2017WR022334.

Tavernini, S., Nizzoli, D., Rossetti, G. & Viaroli, P. (2009), 'Trophic state and seasonal dynamics of phytoplankton communities in two sand-pit lakes at different successional stages', *Journal of Limnology* **68**(2), 217–228.

Telemac modelling system (2016), *TELEMAC3D User manual - v.7.2*.

Thomas, M. K., Kremer, C. T. & Litchman, E. (2016), 'Environment and evolutionary history determine the global biogeography of phytoplankton temperature traits', *Global Ecology and Biogeography* **25**(1), 75–86. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/geb.12387.

Timms, B. V. (1992), *Lake Geomorphology*, Gleneagles Pub.

Toni, T., Welch, D., Strelkowa, N., Ipsen, A. & Stumpf, M. P. (2009), 'Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems', *Journal of The Royal Society Interface* **6**(31), 187–202. Publisher: Royal Society.

Toporowska, M. & Pawlik-Skowronska, J. (2014), 'Four-year study on phytoplankton biodiversity in a small hypertrophic lake affected by water blooms of toxigenic Cyanobacteria', *Polish Journal of Environmental Studies* **23**(2). Publisher: -.

Torma, P. & Wu, C. (2019), 'Temperature and Circulation Dynamics in a Small and Shallow Lake: Effects of Weak Stratification and Littoral Submerged Macrophytes', *Water* **11**, 128.

Torralba, V., Doblas-Reyes, F. J. & Gonzalez-Reviriego, N. (2017), 'Uncertainty in recent near-surface wind speed trends: a global reanalysis intercomparison', *Environ. Res. Lett.* **12**(11), 114019.

Torres, N. V. & Santos, G. (2015), 'The (Mathematical) Modeling Process in Biosciences', *Front. Genet.* **6**. Publisher: Frontiers.

Townsend, D. W., Cammen, L. M., Holligan, P. M., Campbell, D. E. & Pettigrew, N. R. (1994), 'Causes and consequences of variability in the timing of spring phytoplankton blooms', *Deep Sea Research Part I: Oceanographic Research Papers* **41**(5), 747–765.

Tran Khac, V., Hong, Y., Plec, D., Lemaire, B. J., Dubois, P., Saad, M. & Vinçon-Leite, B. (2018), 'An Automatic Monitoring System for High-Frequency Measuring and Real-Time Management of Cyanobacterial Blooms in Urban Water Bodies', *Processes* **6**(2), 11.

Trask, J. C., Fogg, G. E. & Puente, C. E. (2017), 'Resolving hydrologic water balances through a novel error analysis approach, with application to the Tahoe basin', *Journal of Hydrology* **546**, 326–340.

Trolle, D., Hamilton, D. P., Hipsey, M. R., Bolding, K., Bruggeman, J., Mooij, W. M., Janse, J. H., Nielsen, A., Jeppesen, E., Elliott, J. A., Makler-Pick, V., Petzoldt, T., Rinke, K., Flindt, M. R., Arhonditsis, G. B., Gal, G., Bjerring, R., Tominaga, K., Hoen, J., Downing, A. S., Marques, D. M., Fragoso, C. R.,

Søndergaard, M. & Hanson, P. C. (2012), 'A community-based framework for aquatic ecosystem models', *Hydrobiologia* **683**(1), 25–34.

Tímea, K., Kovács-Székely, I. & Anda, A. (2017), 'Comparison of parametric and non-parametric time-series analysis methods on a long-term meteorological data set', *Central European Geology* **60**, 1–17.

Ullyott, P. & Holmes, P. (1936), 'Thermal Stratification in Lakes', *Nature* **138**, 971. **URL:** *http://adsabs.harvard.edu/abs/1936Natur.138..971U*

United Nations, Department of Economic and Social Affairs, P. D. (2014), World Urbanization Prospects: The 2014 Revision, Methodology Working Paper No. ESA/P/WP.238, Technical report.

Uusitalo, L., Lehikoinen, A., Helle, I. & Myrberg, K. (2015), 'An overview of methods to evaluate uncertainty of deterministic models in decision support', *Environmental Modelling & Software* **63**(Supplement C), 24–31.

van den Bosch, M. & Sang, A. o. (2017), '2096 - Urban Natural Environments As Nature Based Solutions for Improved Public Health – a Systematic Review of Reviews', *Journal of Transport & Health* **5**, S79.

van der Vaart, E., Beaumont, M. A., Johnston, A. S. A. & Sibly, R. M. (2015), 'Calibration and evaluation of individual-based models using Approximate Bayesian Computation', *Ecological Modelling* **312**, 182–190.

Vautard, R., Cattiaux, J., Yiou, P., Thépaut, J.-N. & Ciais, P. (2010), 'Northern Hemisphere atmospheric stilling partly attributed to an increase in surface roughness', *Nature Geoscience* **3**(11), 756–761.

Veiga, S. D., Wahl, F. & Gamboa, F. (2009), 'Local Polynomial Estimation for Sensitivity Analysis on Models With Correlated Inputs', *Technometrics* **51**(4), 452–463. Publisher: Taylor & Francis _eprint: https://doi.org/10.1198/TECH.2009.08124.

Venäläinen, A., Lehtonen, I., Laapas, M., Ruosteenoja, K., Tikkanen, O.-P., Viiri, H., Ikonen, V.-P. & Peltola, H. (2020), 'Climate change induces multiple risks to boreal forests and forestry in Finland: A literature review', *Global Change Biology* **26**(8), 4178–4196.

Verdonschot, R. C. M., Keizer-vlek, H. E. & Verdonschot, P. F. M. (2011), 'Biodiversity value of agricultural drainage ditches: a comparative analysis of the aquatic invertebrate fauna of ditches and small lakes', *Aquatic Conservation: Marine and Freshwater Ecosystems* **21**(7), 715–727. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/aqc.1220.

Verpoorter, C., Kutser, T., Seekell, D. A. & Tranvik, L. J. (2014), 'A global inventory of lakes based on high-resolution satellite imagery', *Geophysical Research Letters* **41**(18), 6396–6402.

Vidal, J.-P., Martin, E., Franchistéguy, L., Baillon, M. & Soubeyroux, J.-M. (2010), 'A 50-year high-resolution atmospheric reanalysis over France with the Safran system', *International Journal of Climatology* **30**(11), 1627–1644.

Vilas, M. P., Marti, C. L., Adams, M. P., Oldham, C. E. & Hipsey, M. R. (2017), 'Invasive Macrophytes Control the Spatial and Temporal Patterns of Temperature and Dissolved Oxygen in a Shallow Lake: A Proposed Feedback Mechanism of Macrophyte Loss', *Front. Plant Sci.* **8**.

Villaret, C., Hervouet, J.-M., Kopmann, R., Merkel, U. & Davies, A. G. (2013), 'Morphodynamic modeling using the Telemac finite-element system', *Computers & Geosciences* **53**, 105–113.

Vinçon-Leite, B. & Casenave, C. (2019), 'Modelling eutrophication in lake ecosystems: A review', *Science of The Total Environment* **651**, 2985–3001.

Vinçon-Leite, B., Lemaire, B. J., Khac, V. T. & Tassin, B. (2014), 'Long-term temperature evolution in a deep sub-alpine lake, Lake Bourget, France: how a one-dimensional model improves its trend assessment', *Hydrobiologia* **731**(1), 49–64.

Viollet, P. L. (1988), On the Numerical Modelling of Stratified Flows, *in* J. Dronkers & W. van Leussen, eds, 'Physical Processes in Estuaries', Springer, Berlin, Heidelberg, pp. 257–277.

Wagner, C. & Adrian, R. (2009), 'Cyanobacteria dominance: Quantifying the effects of climate change', *Limnology and Oceanography* **54**(6part2), 2460–2468.

Wagner, T. & Erickson, L. (2017), 'Sustainable Management of Eutrophic Lakes and Reservoirs', *Journal of Environmental Protection* **08**, 436–463.

Wahl, B. & Peeters, F. (2014), 'Effect of climatic changes on stratification and deep-water renewal in Lake Constance assessed by sensitivity studies with a 3D hydrodynamic model', *Limnology and Oceanography* **59**(3), 1035–1052.

Walsby, A. E. (1994), 'Gas vesicles', *Microbiology and Molecular Biology Reviews* **58**(1), 94–144. Publisher: American Society for Microbiology Journals Section: Research Article.
URL: *https://mmbr.asm.org/content/58/1/94*

Wang, F., Shao, W., Yu, H., Kan, G., He, X., Zhang, D., Ren, M. & Wang, G. (2020), 'Re-evaluation of the Power of the Mann-Kendall Test for Detecting Monotonic Trends in Hydrometeorological Time Series', *Front. Earth Sci.* **8**.

Ward, B. A., Schartau, M., Oschlies, A., Martin, A. P., Follows, M. J. & 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.

Watson, S. B., McCauley, E. & Downing, J. A. (1997), 'Patterns in phytoplankton taxonomic composition across temperate lakes of differing nutrient status', *Limnology and Oceanography* **42**(3), 487–495.

Watson, S. B. & Molot, L. (2013), Harmful Algal Blooms, *in* J.-F. Férard & C. Blaise, eds, 'Encyclopedia of Aquatic Ecotoxicology', Springer Netherlands, Dordrecht, pp. 575–596.

Watson, S., Whitton, B., Higgins, S., Paerl, H., Brooks, B. & Wehr, J. (2015), Harmful Algal Blooms, pp. 873–920.

Weiskerger, C. J., Rowe, M. D., Stow, C. A., Stuart, D. & Johengen, T. (2018), 'Application of the Beer–Lambert Model to Attenuation of Photosynthetically Active Radiation in a Shallow, Eutrophic Lake', *Water Resources Research* **54**(11), 8952–8962.

Weiss, R. F. (1970), 'The solubility of nitrogen, oxygen and argon in water and seawater', *Deep Sea Research and Oceanographic Abstracts* **17**(4), 721–735.

Wilhelm, S. & Adrian, R. (2008), 'Impact of summer warming on the thermal characteristics of a polymictic lake and consequences for oxygen, nutrients and phytoplankton', *Freshwater Biology* **53**(2), 226–237.

Wilkinson, A. A., Hondzo, M. & Guala, M. (2020), 'Vertical heterogeneities of cyanobacteria and microcystin concentrations in lakes using a seasonal In situ monitoring station', *Global Ecology and Conservation* **21**, e00838.

Wilkinson, A., Hondzo, M. & Guala, M. (2016), 'Effect of Small-Scale Turbulence on the Growth and Metabolism of Microcystis aeruginosa', *Advances in Microbiology* **06**(05), 351–367.

Winder, M. & Sommer, U. (2012), 'Phytoplankton response to a changing climate', *Hydrobiologia* **698**(1), 5–16.

Winslow, L. A., Leach, T. H. & Rose, K. C. (2018), 'Global lake response to the recent warming hiatus', *Environ. Res. Lett.* **13**(5), 054005.

Winter, T. C. (1981), 'Uncertainties in Estimating the Water Balance of Lakes1', *JAWRA Journal of the American Water Resources Association* **17**(1), 82–115.

Wipf, S., Stoeckli, V. & Bebi, P. (2009), 'Winter climate change in alpine tundra: plant responses to changes in snow depth and snowmelt timing', *Climatic Change* **94**, 105–121.

Woolway, R. I., Maberly, S. C., Jones, I. D. & Feuchtmayr, H. (2014), 'A novel method for estimating the onset of thermal stratification in lakes from surface water measurements', *Water Resources Research* **50**(6), 5131–5140.

World Health Organization (1999), Toxic cyanobacteria in water, Technical report, 11 New Fetter Lane, London, UK.

World Health Organization (2003), Guidelines for safe recreational water environments. volume 1, coastal and fresh waters, Technical report. Publisher: World Health Organization.

Xue, P., Pal, J. S., Ye, X., Lenters, J. D., Huang, C. & Chu, P. Y. (2017), 'Improving the Simulation of Large Lakes in Regional Climate Modeling: Two-Way Lake–Atmosphere Coupling with a 3D Hydrodynamic Model of the Great Lakes', *Journal of Climate* **30**(5), 1605–1627. Publisher: American Meteorological Society Section: Journal of Climate.

Yan, X., Xu, X., Wang, M., Wang, G., Wu, S., Li, Z., Sun, H., Shi, A. & Yang, Y. (2017), 'Climate warming and cyanobacteria blooms: Looks at their relationships from a new perspective', *Water Research* **125**(Supplement C), 449 – 457.

Yang, Y., Wang, Y., Zhang, Z., Wang, W., Ren, X., Gao, Y., Liu, S. & Lee, X. (2018), 'Diurnal and Seasonal Variations of Thermal Stratification and Vertical Mixing in a Shallow Fresh Water Lake', *J Meteorol Res* **32**(2), 219–232.

You, J., Mallery, K., Hong, J. & Hondzo, M. (2018), 'Temperature effects on growth and buoyancy of Microcystis aeruginosa', *J Plankton Res* **40**(1), 16–28. Publisher: Oxford Academic.

Zaccaroni, A. & Scaravelli, D. (2008), Toxicity of Fresh Water Algal Toxins to Humans and Animals, *in* 'NATO Security through Science Series C: Environmental Security', pp. 45–89. Journal Abbreviation: NATO Security through Science Series C: Environmental Security.

Zeece, M. (2020), Chapter Eight - Food colorants, *in* M. Zeece, ed., 'Introduction to the Chemistry of Food', pp. 313–344.

Zelinka, M. D., Randall, D. A., Webb, M. J. & Klein, S. A. (2017), 'Clearing clouds of uncertainty', *Nature Climate Change* **7**(10), 674–678. Number: 10 Publisher: Nature Publishing Group.

Zhang, L., Hipsey, M. R., Zhang, G. X., Busch, B. & Li, H. Y. (2017), 'Simulation of multiple water sources ecological replenishment for Chagan Lake based on coupled hydrodynamic and water quality models', *Water Science and Technology: Water Supply* p. ws2017079.

Zhang, X., Wang, K., Frassl, M. A. & Boehrer, B. (2020), 'Reconstructing Six Decades of Surface Temperatures at a Shallow Lake', *Water* **12**(2), 405.

Zhao, H., Han, G. & Wang, D. (2013), 'Timing and magnitude of spring bloom and effects of physical environments over the Grand Banks of Newfoundland', *Journal of Geophysical Research: Biogeosciences* **118**(4), 1385–1396.

Zou, W., Zhu, G., Cai, Y., Xu, H., Zhu, M., Gong, Z., Zhang, Y. & Qin, B. (2020), 'Quantifying the dependence of cyanobacterial growth to nutrient for the eutrophication management of temperate-subtropical shallow lakes', *Water Research* **177**, 115806.