From observations to 3D forecasts:
Data assimilation for high resolution lakes monitoring

Thèse N° 9475

Présentée le 24 mai 2019

à la Faculté de l'environnement naturel, architectural et construit
Laboratoire de physique des systèmes aquatiques - Chaire Margaretha Kamprad
Programme doctoral en génie civil et environnement

pour l'obtention du grade de Docteur ès Sciences

par

Theo BARACCHINI

Acceptée sur proposition du jury

Prof. T. Kohn, présidente du jury
Prof. A. J. Wüest, Dr D. Bouffard, directeurs de thèse
Dr Ph. Y. Chu, rapporteur
Dr J. Verron, rapporteur
Prof. M. Holzner, rapporteur

2019

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

À mes parents

Acknowledgements

Reflecting on my years spent at EPFL, I would like to take the next few lines to acknowledge the numerous people that were around me and who played a role in the completion of this research.

To my supervisors Johny and Damien, I would like to express my profound gratitude. Thank you for your guidance and support. I am especially thankful for the trust you have placed in me and the freedom I had during this thesis. Aside from your scientific and academic teachings, you offered me a great gift: a new vision and understanding on something I saw and appreciated for my entire life, the lake. To me, that means a lot.

To my insightful mentors, Stef Hummel and Rob Uittenbogaard from Deltares, and Philip Chu from NOAA. I am grateful for your invaluable advices, for always being available any time I needed, and for the fabulous work accomplished.

To my friends in the lab: Vivi, Camille, Shubham, Tania, Hannah, Sébi, Tomy, Isabel, Oscar, Hugo, Love, Bieito, Robert, Natacha and Lara. How can I summarize everything with a few words? Satellite? Coffee? These two hold in quite a few stories already. You created a fantastic atmosphere for both working and not working. Thank you for having made this thesis an outstanding human and academic experience.

It has been ten years since I started EPFL (I know). I cannot help but think about how it started, when I did not know anything about science. For their invaluable patience and help in my studies, I would like to thank some old classmates: Guigus, you taught me how to flip over an addition of fractions, Cédric, you taught me what the trigonometric circle was, Valeria, you taught me that velocity is distance over time, and Océane, you taught me how to project forces. You have helped me a lot in a time I needed it the most. I have not forgotten, thank you. To all my great friends and classmates: Baptiste, Pierre, Margaux, Adri, Fred, Flo, Arthur, Billoutte, Adeline, Akio, thank you for the truly memorable times we had. EPFL was not the same after you left.

To Adrien, my friend. You helped me a lot, you challenged me, you even got me started on lakes. You set some foundations for this thesis, which I always pictured with you. Adri, mon ami, you will be missed, but you will never be forgotten.

To my grand-parents, my warmest gratitude. You hosted me during all those years and you made me think about the real sense of what I was doing. In the end, *Meteolakes* partly exists thanks to you. Thank you for taking care of me.

To my sisters Lea and Sara, and to my brother Gian Marc. I have to say, sometimes I wonder if I am the elder brother, you are examples to me. Thank you for inspiring me, for your unwavering support, and for being here.

To Paola, a special thank you for your everlasting support in this thesis, and beyond. I learnt so much from you during those years. It's been a beautiful adventure, and I cannot wait for the ones to come.

I have often been thinking about this moment, about how I could thank, you, my parents, the true visionaries behind this work and my life. The more I think, the more I come to this conclusion: it is simply impossible, really. To you I would like to say one thing: the greatest challenge of my life is still ahead, the one I would be the proudest of, my greatest aspiration. And that is to become, one day, a parent like you.

Pully, le 22 Février 2019

Abstract

Lake mesoscale processes have no secrets for artists, they have some for scientists and policy-makers. In the introduction to this dissertation, we saw that poets and painters depict the lake with rich variability. This vision conflicts with conventional approaches inferring the lake's bio-physical status from a few in-situ measurements with limited spatio-temporal coverage. Yet it reflects upon true physical processes, capable of disrupting the misleading lentic nature of lakes, the essential ecosystem services they provide, and even citizens' safety. Recent overarching policies now aim at securing those services, using novel approaches like numerical simulations and remote sensing observations. Harnessing the potentiality of in-situ, remote sensing data and model simulations, this thesis developed an end-to-end framework delivering reliable synoptic lake information, at high spatial and temporal resolution.

In lakes, three-dimensional hydrodynamic models are the only information source capable of resolving transport and mixing, and forecasting their dynamics. However, they still rely on large observational datasets for their complex parameterizations and for constraining their uncertainties. This thesis addressed such challenges by (i) implementing an automated model calibration framework alleviating the need for expert knowledge, and by (ii) developing a data assimilative scheme capable of reducing and quantifying model uncertainties by incorporating satellite and in-situ data. Both yielded remarkable results: the former returned parametric values diminishing models Root Mean Square Error by up to 47 %, while the latter cut it further down by 54 %. Furthermore, the data assimilation enhanced the spatial coherence and magnitude of imperfectly resolved physical processes, and provided system uncertainties. Finally, this thesis delivered a practical outcome of its findings by developing an online pre-operational three-dimensional lake monitoring and forecasting system: *meteolakes.ch*. For two years, *Meteolakes* has been disseminating spatially explicit real-time lake information and data products to more than hundred thousand end-users. This pioneering platform has been featured in numerous media (newspapers, radio, television), public events, museum exhibitions, and benefited scientific, lake professionals, and public communities. A pinnacle of this research has been its early-warning and forecasting capabilities, which anticipated numerous mesoscale physical phenomena in Lake Geneva, such as upwellings, gyres, and strong currents. Two of those events, which impacted public and commercial activities, are illustrated in this dissertation. From observations and models to societal benefit, we created here a long value chain for water management.

At the crossroads of scientific, computational and observational advances, this research paves the route for understanding lakes' delicate imbalance. By producing data society can use, it opens novel frontiers for interdisciplinary research on previously elusive lake physical processes, and their implications on the everyday life of people.

Keywords

Lakes monitoring and forecasting, three-dimensional hydrodynamics, model calibration, data assimilation, Ensemble Kalman Filter, operational systems, mesoscale processes, Lake Geneva, *Meteolakes*.

Résumé

Les processus lacustres à méso-échelle n'ont pas de secrets pour les artistes, ils en ont quelques-uns pour les scientifiques et responsables politiques. Dans l'introduction de cette thèse, nous avons vu que les peintres et les poètes représentent le lac au travers d'une riche hétérogénéité. Cette vision s'oppose aux approches conventionnelles, qui déduisent son état biophysique à partir de quelques mesures in-situ peu représentatives. Pourtant, ce regard artistique reflète des processus physiques réels, capables de perturber la nature lentique trompeuse de ces eaux, d'interrompre les services écosystémiques essentiels qu'elles fournissent, pouvant aller jusqu'à compromettre la sécurité des habitants. Des politiques publiques actuelles visent alors à sécuriser ces services, au travers de nouvelles approches comme les simulations numériques ou la télédétection. En exploitant le potentiel des mesures in-situ, de télédétection, et des modèles numériques, cette recherche a développé une solution de bout-en-bout, délivrant des informations lacustres synoptiques à haute résolution spatiale et temporelle.

Dans les lacs, les modèles hydrodynamiques tridimensionnels sont la seule source d'information capable de résoudre entièrement la structure thermique et courantologie des eaux, ainsi que de prédire leur dynamique. Ils nécessitent cependant des observations pour leurs paramétrisations complexes et pour restreindre leurs incertitudes. Cette thèse a traité ces aspects par (i) l'implémentation d'un système de calibration automatique limitant ainsi le besoin d'expertise en modélisation, et (ii) par le développement d'une méthode d'assimilation de données satellites et in-situ permettant de quantifier et de réduire les incertitudes. Des résultats remarquables ont été obtenus : dans le premier cas, les nouvelles valeurs paramétriques ont permis de réduire jusqu'à 47 % l'erreur quadratique moyenne, tandis que dans le second, celle-ci a encore été diminuée de 54 %. De plus, le système d'assimilation a permis d'accroitre la cohérence spatiale et la magnitude de phénomènes physiques imparfaitement modélisés, ainsi que de renseigner sur l'incertitude du système. Finalement, cette recherche a mis en pratique ses résultats à travers l'implémentation d'une plateforme en ligne pré-opérationnelle permettant le suivi et la prédiction tridimensionnelle de divers lacs : *meteolakes.ch*. Depuis deux ans, *Meteolakes* dissémine en temps réel des informations lacustres spatialement explicites à plus d'une centaine de milliers d'utilisateurs. Cette plateforme novatrice a été mise en vedette à travers divers media (journaux, télévision, radio), expositions, événements publics et a bénéficié à diverses communautés scientifiques, professionnelles et publiques. Un point culminant de cette recherche a été sa capacité prévisionnelle et son système d'alerte, qui ont été mis à l'épreuve en anticipant de multiples processus physiques à méso-échelle (remontées d'eau, tourbillons, courants de tempête) ayant eu lieu dans *le Léman*. Deux de ces phénomènes, dont l'impact sur les activités commerciales et publiques fut considérable, sont présentés dans cette dissertation. À partir d'observations et modèles, jusqu'aux bénéfices sociétaux, nous apportions ainsi une chaine de valeur significative pour le management des eaux.

Au carrefour des avancées scientifiques, computationnelles et observationnelles, cette recherche ouvre la voie vers une compréhension synoptique des lacs. En générant des données à directe valeur sociétale, ce travail offre de nouvelles perspectives à des études interdisciplinaires des phénomènes physiques lacustres encore mal connus, et à leurs impacts sur le quotidien des populations concernées.

Mots-clés

Surveillance et prévision des lacs, hydrodynamique tridimensionnelle, calibration de modèles, assimilation de données, filtre de Kalman Ensemble, systèmes opérationnels, processus à meso-échelle, *le Léman*, *Meteolakes*.

Zusammenfassung

Die mesoskaligen Prozesse eines Sees bergen für Künstler keine Geheimnisse, für Wissenschaftler und Politiker jedoch schon. In der Einleitung dieser Dissertation wurde erläutert, dass Dichter und Maler den See mit grosser und reicher Vielfalt darstellen. Diese Wahrnehmungen stehen im Gegensatz zu konventionellen Ansätzen, die den biophysikalischen Status eines Sees aus wenigen in-situ Messungen mit begrenzter räumlicher und zeitlicher Abdeckung ableiten. Diese vielfältigen Strukturen reflektieren jedoch echte physikalische Prozesse, die in der Lage sind die Charakteristik stehender Gewässer und deren wichtigen ökologischen Funktionen zu beeinflussen und sogar die Sicherheit der Anwohner zu gefährden. Neuere übergeordnete Richtlinien zielen heute darauf ab, Monitoringdaten durch neuartige Ansätze wie numerische Simulationen und Fernerkundungsbeobachtungen abzusichern. Diese Dissertation nutzte das Potenzial von in-situ Fernerkundungsdaten und Modellsimulationen und entwickelte eine End-to-End-Plattform, welche zuverlässige synoptische Informationen von Seen mit hoher räumlicher und zeitlicher Auflösung liefert.

In Seen sind dreidimensionale hydrodynamische Modelle die einzige Informationsquelle, die Transport und Mischung fein auflösen und deren Dynamik vorhersagen können. Sie sind jedoch immer auf große Beobachtungsdatensätze angewiesen, um komplexe Parametrisierungen vorzunehmen und Unsicherheiten einzuschränken. Diese Dissertation befasste sich mit diesen Herausforderungen, indem (i) ein automatisches Modellkalibrierungssystem implementiert wurde, das den Bedarf an Expertenwissen mindert, und (ii) ein Datenassimilationsschema entwickelt wurde, das Modellunsicherheiten durch Einbezug von Satelliten- und in-situ Daten reduzieren und quantifizieren kann. Beide Vorgehensweisen erzielten bemerkenswerte Ergebnisse: Erstere lieferten Parameterwerte, welche den mittleren quadratischen Fehler (Root Mean Square Error) der Modelle um bis zu 47% minderte, während letztere Anwendung den Fehler um weitere 54% reduzierte. Darüber hinaus steigerte die Datenassimilation die räumliche Kohärenz und die Größenordnung unvollständig aufgelöster physikalischer Prozesse und sorgte für eine Schätzung der Systemunsicherheiten. Abschliessend erbrachte die vorliegende Arbeit, durch die Implementierung ihrer Erkenntnisse, die Entwicklung einer dreidimensionalen Online-Plattform als Überwachungs- und Vorhersagesystems für Seen: *meteolakes.ch*. Seit zwei Jahren bietet *Meteolakes* räumlich explizite Echtzeit-Information und Daten für Seen an über hunderttausend Endnutzern an. Diese wegweisende Plattform fand in zahlreichen Medien (Zeitungen, Radio, Fernsehen), öffentlichen Veranstaltungen, Museumsausstellungen hohe Anerkennung und kam Wissenschaftlern, professionellen Nutzern und der Allgemeinheit zugute. Ein Höhepunkt dieser Forschung sind die Frühwarn- und Prognosefähigkeiten, die zahlreiche physikalische mesoskalierte Phänomene im Genfersee vorwegnahmen, beispielsweise Auftriebszonen, grosse horizontale Wirbel und starke Strömungen. Zwei dieser Ereignisse, die sich auf öffentliche und kommerzielle Aktivitäten auswirkten, werden in dieser Dissertation dargestellt. Von Beobachtungen und Modellen bis zum gesellschaftlichen Nutzen haben wir eine lange Wertschöpfungskette für das Management von Seen geschaffen.

An der Schnittstelle von Wissenschaft, numerischen Simulationen und Monitoring ebnen die Fortschritte dieser Forschung den Weg, um das empfindliche Gleichgewicht der Seen zu verstehen. Durch die Nutzung der von der Gesellschaft produzierten Daten, können sich neue Grenzen eröffnen für die interdisziplinäre Erforschung bisher schwer greifbarer physikalischer Prozesse des Sees und deren Auswirkungen auf den Alltag der Menschen.

Stichworte

Überwachung und Vorhersage von Seen, dreidimensionale Hydrodynamik, Modellkalibrierung, Datenassimilation, Kalman-Filterung, Betriebssysteme, mesoskalige Prozesse, Genfersee, *Meteolakes*.

Sommario

I processi lacustri a mesoscala non hanno secreti per gli artisti, ma ne hanno qualcuno per gli scienziati e i responsabili politici. Nell'introduzione, si è rilevato che i pittori e i poeti rappresentano il lago attraverso una ricca eterogeneità. Questa visione si oppone agli approcci convenzionali, qui deducono il suo stato bio-fisico a partire da alcune misure realizzate in-situ e poco rappresentative. Tuttavia, questa visuale artistica riflette dei processi fisici reali, capaci di perturbare la natura lentica ingannevole di queste acque, d'interrompere i servizi ecosistemici essenziali che esse forniscono, fino ad una possibile messa in pericolo degli abitanti. Delle politiche pubbliche attuali cercano quindi di rendere affidabili questi servizi, attraverso dei nuovi approcci come le simulazioni numeriche o la telerilevazione. Utilizzando sia il potenziale delle misure in-situ che la telerilevazione e i modelli numerici, questa ricerca ha sviluppato una soluzione da un capo all'altro, fornendo delle informazioni lacustre sinottiche di elevata risoluzione spaziale e temporale.

Nei laghi, i modelli tridimensionali sono la sola sorgente d'informazione capace di risolvere completamente i processi di agitazione e di trasporto e inoltre di predire la dinamica delle acque. Essi richiedono tuttavia delle osservazioni per le loro parametrizzazioni complesse e per ridurre le loro incertezze. Questa tesi ha trattato questi aspetti con (i) l'implementazione di un sistema di calibraggio automatico limitando quindi la necessità di conoscenze in modellizzazione, e (ii) attraverso lo sviluppo di un metodo di assimilazione di dati satellitari e in-situ che permettono di quantificare e di ridurre le incertezze. Dei risultati ragguardevoli sono stati ottenuti: nel primo caso, dei nuovi valori parametrici hanno consentito di ridurre del 47% l'errore quadratico medio, mentre nel secondo caso questo errore è stato ridotto ulteriormente del 54%. Inoltre il sistema di assimilazione a consentito di aumentare la coerenza spaziale et la magnitudine dei fenomeni fisici non perfettamente modellizzati ed anche d'indicare l'incertezza del sistema. Infine, questa ricerca ha messo in pratica i risultati ottenuti attraverso l'implementazione di una piattaforma web preoperativa che consente la sorveglianza e la predizione tridimensionale di differenti laghi: *meteolakes.ch*. Da due anni, *Meteolakes* distribuisce in tempo reale delle informazioni lacustri spazialmente esplicite a qualche centinaia di migliaia di utilizzatori. Questa piattaforma innovatrice è stata messa in valore in diversi media (giornali, televisioni, radio), esposizioni, manifestazioni pubbliche e ha arrecato un beneficio a differenti comunità scientifiche, professionali e pubbliche. Un punto culminante di questa ricerca è stato la sua capacità previsionale e il suo sistema di allerta, che sono stati convalidati all'occasione di numerosi processi fisici a mesoscala (risalita delle acque, vortici, correnti di tempesta) che hanno avuto luogo sul *Lemano*. Due fra questi fenomeni, che hanno avuto un impatto considerevole sulle attività commerciali e pubbliche, sono illustrati in questa dissertazione. Partendo da osservazioni e modelli fino ai benefici per la società, noi forniamo un seguito importante di valori per il management delle acque.

All'incrocio dei progressi scientifici, computistici e osservativi, questa ricerca apre la strada verso una comprensione sinottica dei laghi. Generando dei dati di chiaro valore socialitario, questo lavoro offre delle nuove prospettive per degli studi interdisciplinari dei fenomeni fisici lacustri ancora mal conosciuti e dei loro impatti sul quotidiano delle popolazioni implicate.

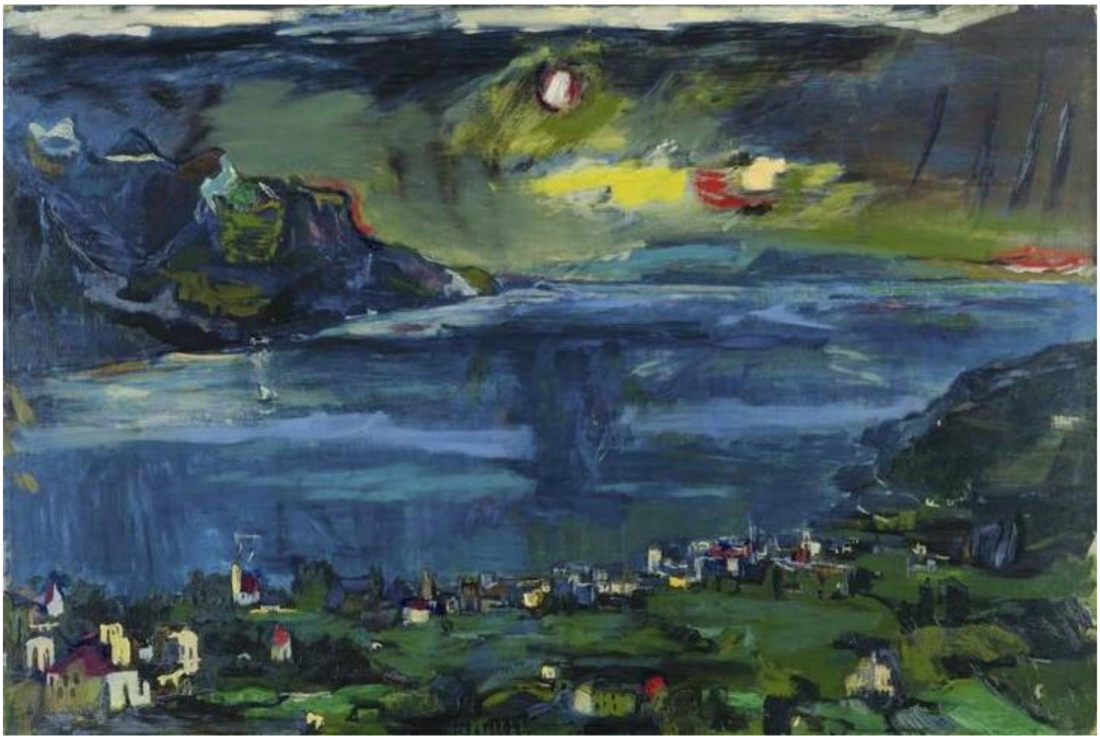
Parole chiave

Sorveglianza e previsione dei laghi, idrodinamica tridimensionale, calibraggio dei modelli, assimilazione dei dati, filtro di Kalman Insieme, sistemi operativi, processi a mesoscala, *il Lemano*, *Meteolakes*.

Contents

Prologue

Ce que je vois par la fenêtre…



Oskar Kokoschka, *Le Léman I*, 1923

Il y avait autrefois, sous mes fenêtres, tout un domaine de vignes : il n'y a plus aujourd'hui qu'un terrain assez dégarni, avec des jardins, des plantages, et, dans le bas, deux ou trois maisons pas méchantes, de construction assez récente, et puis…

Et puis plus rien qu'une vaste surface plate qui mène le regard directement jusque chez les Savoyards, un vaste espace sans relief qui est à la fois au-dessous et au-dessus de vous, qui est en même temps horizontal et vertical, qui est bleu comme le ciel, qui est brun comme un champ qu'on vient de labourer, qui est violet et tout uni ou bien encore marbré de taches : le lac que j'entrevois par les trous qu'il y a dans le feuillage d'un cognassier et entre les deux longs bras parallèles d'un cèdre, vu ainsi par petits morceaux et inséré dans les branchages comme des vitres de couleur.

Tout le lac, ce grand réflecteur, qui nous renvoie redoublée la lumière du jour et la chaleur de l'astre, les jours où il est luisant comme une feuille de fer-blanc, les jours aussi où il scintille de mille petits feux à la crête de chaque vague, comme des copeaux allumés.

C'est ce qu'on voit immédiatement au-dessous de soi quand on est tourné vers le sud, mais dès que le regard change de direction, tout se transforme. […] Les grandes montagnes, tachées de blanc, les grandes montagnes dans le ciel et qui semblent elles-mêmes un peu de ciel tombé, les hauts rochers sous cette voûte, et c'est comme si la voûte, à cette place, avait croulé : […] la percée de la vallée du Rhône d'où on le voit sortir, le Rhône, et qui fait une barre jaune qui s'avance très loin dans le lac.

[…] On voit au bord de l'eau les vieilles maisons du port de Pully. Puis vient une place sur le rivage, que rien ne marque pour les yeux, mais qui l'est singulièrement pour les oreilles, sitôt que l'été est venu. C'est le point de départ de tout un feu d'artifice de rires et de cris qui montent jusqu'en haut du mont : les enfants qui se baignent, et quand le bateau à vapeur arrive, les cris redoublent, parce que le bateau à vapeur passe ici tout près de la rive. […] Ce rivage qui tantôt projette une pointe, tantôt se retire, forme un golfe, et on voit dans l'eau du golfe un bateau de pêcheur et un pêcheur dedans. Les peupliers fument encore jaune : c'est le printemps.

C. F. Ramuz

Chapter 1 Introduction

I

What I see through the window…

1.1 The spectacular in the seemingly mundane

Like Charles Ferdinand Ramuz, a century before me, I too have been looking through the window. And like him, from my house in my hometown Pully, I too have gazed at this vast expanse. A vast expense of no landforms, horizontal and vertical at the same time, blue, brown, purple, plain or stained. The lake, a magnifying reflector of the light and heat of the star, the spectacular in the seemingly mundane.

The essence of the poet's and writer's words are well depicted in the opening artwork by Oskar Kokoschka[1]. A panorama of a stunning amplitude. Coloured patches, current patterns, weather effects, atmospheric reflections and dark static shades, all united to illustrate the heterogeneous nature of those waters. Both tumultuous and lentic at the same time. The lake, a system of stasis and flows.

Does such variability only arise from the visionary minds of poets and artists? Or is it a representation of true physical processes, whose different manifestations reflect the interactions of local meteorological variables and local hydro-morphology with the dynamics of the stratified waters of the lake? In a sense, this is an all-underlying question this thesis will address.

Implicitly conveyed through the prologue, *le Léman*[2] has long been a source of inspiration for a myriad of artists. Likewise, it has been a catalyst for scientists, and perhaps for lakes' science itself. In the XIXth century, on the shores of *le Léman*, François-Alphonse Forel coined the term limnology, defining it as "the oceanography of lakes" (Forel, 1892). The scientist and father of limnology already highlighted the unique yet extremely diverse nature of lakes, stating that the singularity of a lake is analogous to that of an island; the former being separated from the ocean by lands while the latter is isolated from continental lands by the ocean.

Such diversity – between and within lakes – requires novel approaches. New strategies to decipher those entangled dynamics acting at a broad range of spatio-temporal scales. Provided limnology stems from Greek λίμνη, *limne*, "lake", λόγος, *logos*, "study" and the name "*Léman*", refers to an Indo-European root, also meaning "lake" (Bergier, 2009), what is better than *The Lake*, to study lakes – this thesis is indeed about Lémanology and its variability.

1.2 A challenge of our time

Children bathing, vapour boats, fishermen. In the prologue, C. F. Ramuz further shows the predominant place of the lake for its surrounding inhabitants. Lakes indeed play an essential role for the economic development and wellbeing of our societies. Yet, they are threatened both locally through human influence and globally through climate and environmental changes. Half of the world's population lives within 3 km of freshwater bodies (Kummu et al., 2011). Ensuring water services through the monitoring and forecasting of lake ecosystems

[1] An Austrian artist who felt in love with the Lemanic region, living by the lake - its dearest wish - until its death.
[2] Known as Lake Geneva internationally, I refer to its local and authentic name in the introductory and concluding chapters.

reactions to anthropogenic pressures, global climate forcing, and to extreme weather events are among the great challenges of our time (United Nations Sustainable Development Goal 6).

In lakes, some of those reactions can seriously compromise the vital ecosystem services they provide. Among those are drinking water supply, food resources, heat supply and discharge, commercial navigation, recreational activities and tourism. In some cases, they can even threaten citizens' security (e.g. storm navigation hazards, harmful algal blooms). The close monitoring and forecasting of lakes status would then allow for a timely and informed decision making, to mitigate the negative effects of a variety of societally important issues.

The need for such capabilities has been well illustrated through recent studies. For instance, O'Reilly et al. (2015) showed that on a global scale, Lake Surface Water Temperature (LSWT) is warming at an alarming rate, higher than that of the atmosphere and oceans. Adrian et al. (2009) showed that even small changes to some physical characteristics, such as the LSWT, can deeply alter both the ecological and physical dynamics due to the non-linear nature of these systems; hence portraying lakes as "sentinels of climate change". Moreover, some studies found that extreme weather events are expected to become more frequent with climate change (Woolway et al., 2018). Thus highlighting the prime importance of preventive measures to mitigate the heavy economical and potential human losses lake reactions can generate from such events. The proper understanding of those processes and the timely dissemination of reliable information can only be achieved through new integrated data-model monitoring programs (Hering et al., 2015).

1.2.1 Monitoring lakes: a paradigm shift

For decades, lakes' biophysical status has been inferred from low frequency in-situ observations at a limited number of spatial locations. While for centuries, artists highlighted the strong spatio-temporal variabilities lakes are subject to. Passing on, in a sense, a synoptic view of the system. The message conveyed through the prologue is clear; both O. Kokoschka and C. F. Ramuz make use of colours and patterns to characterize the lake and emphasize, namely, what we call mesoscale processes.

In lakes, mesoscale processes – the size of several kilometres and smaller than the basin scale – disrupt the lentic nature of the system, thereby influencing biogeochemical processes (MacIntyre and Melack, 1995) and its water quality. "The Rhône, a yellow bar stepping far forward into the lake" (cf. the prologue), illustrates a river intrusion phenomenon, the words "marbled with spots", can be attributed to inhomogeneous glacial riverine particles and chlorophyll-a (or other constituent/pollutants) distributions. Along with upwellings, gyres, storm and their associated local strong currents, those are all mesoscale phenomena, which have long been insufficiently documented and monitored. This research gap is a result of the relatively transient nature of those processes, their sub-basin scale spatial extent, and the narrow spatio-temporal coverage of traditional in-situ monitoring, which fails, due to a lack of representativeness, in assessing those critical but often subtle changes (Kiefer et al., 2015). This change in paradigm has brought to the fore the need for new monitoring programs (Hering et al., 2015), using novel approaches, combining numerical simulations and remote sensing observations (Vörösmarty et al., 2015).

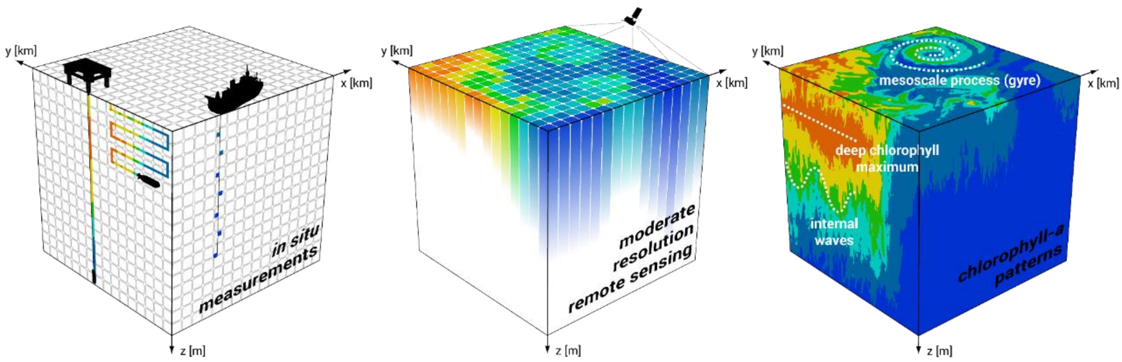


Figure 1.1: In-situ measurements (left), remote sensing observations (center) and model simulation (right) [Odermatt & Brockmann GmbH, Zurich].

Remote sensing observations, as well as one- and three-dimensional hydrodynamic models indeed addressed some of the spatial and temporal coverage limitations (Figure 1.1). However, perhaps due to the misleading definition of lakes as lentic systems, hydrodynamics studies often focused on the vertical structure of lakes and long time scales. Few have studied the aforementioned mesoscale processes, whose temporal scales are significantly shorter and their horizontal spatial variability often underestimated. Continuously improving spatial and temporal resolution, remotely sensed observations now provide an essential source of information, but it remains fundamentally two-dimensional. Key in grasping such variability are three-dimensional hydrodynamic models, the only information source capable of solving the large variety of spatio-temporal scales involved in local to basin-scale lake dynamics. Those models, however, have drawbacks: complex parameterizations requiring expert knowledge and numerous observations, expensive computations, and unavoidable deviations (Lahoz et al., 2010). This thesis addresses such challenges with the aim of exploiting a fundamental property of models: their ability to forecast.

1.2.2 The quiet revolution of numerical forecasting

In a *Nature* publication, Bauer et al. (2015) describe numerical weather forecasting as a "quiet revolution". A quiet revolution in that advances in numerical weather forecasting did not receive the same attention than fundamental physics breakthroughs. Yet, their societal impact is among the greatest of any domain of physical science and their computational complexity is comparable to the simulation of the human brain or the evolution of the early Universe – performed daily in several operational centres. The authors identify three key aspects from which the largest advances in predictive skills have been made, which are also the greatest challenges for the future: (i) physical representation (parameterization), (ii) ensemble forecasting and (iii) model initialization. Lakes hydrodynamics forecasting shares similar scientific, technical and societal challenges – and perhaps a revolution to come –, with one major difference: due to the less chaotic nature of lakes dynamics, the first two aspects prevail. Model physical representation and ensemble forecasting are both addressed in this thesis.

Physical representation – The purpose of a lake hydrodynamic model is to numerically solve various one- and three-dimensional, time-dependent, non-linear differential equations for water volumes (continuity equation), momentum (Reynolds-averaged Navier-Stokes (RANS) equations), and tracer masses (transport equation), using a spatio-temporal discretization. This discretization results from the mathematical intractability of deriving analytical solutions for some of those equations (Bauer et al., 2015). It is an approximation, which divides resolved from unresolved scales of motion. Nonetheless, because unresolved processes still influence larger scales, these interactions need to be parameterized. This parameterization also allows the description of other system properties, and is a fundamental stage in model development. Inferring some of those values, also referred to as calibration, is non-trivial, as a number of them cannot be directly observed, depend indeed on the computational grid, or are related to aggregated processes (Madsen, 2003). Traditionally achieved by trial-and-error procedures, calibration is a complex and time-consuming task that discourages non-expert users, distracts resources from model-based system understanding, and relies on the expertise of the modeller (Afshar et al., 2011; Fabio et al., 2010; Madsen, 2000). From those difficulties arose the need to automate this crucial phase (Solomatine et al., 1999).

Over the last decades, automated computer-based calibration frameworks have been developed in various fields of environmental sciences, such as surface water hydrology (Bárdossy and Singh, 2008; Hendrickson et al., 1988; Johnston and Pilgrim, 1976; Solomatine et al., 1999) and 1D/2D hydrodynamics (Afshar et al., 2011; Fabio et al., 2010; Gaudard et al., 2017), to infer unknown model parameters. However, it has been done at a significant computational cost. The computational burden associated with 3D hydrodynamic models limited the use of such frameworks and few authors developed parameter search and optimization techniques adapted to the high dimensionality of said systems. The second chapter of this thesis thereby aims at breaking down such task, using open-source tools, towards more open and reproducible applications for a wider range of users. It does so by presenting a systematic methodology to overcome the high computational costs associated with computer-based calibrations of 3D hydrodynamic models by proposing an adaptive and efficient calibration framework, tailored to the lake calibration data available to the user.

Ensemble forecasting – The need for propagating an ensemble of differing model trajectories, to provide reliable forecasted systems dynamics, probably finds its origins in the early formulation of the chaotic property. Henri Poincaré stated that "small differences in the initial conditions [may] produce very great ones in the final phenomena. A small error in the former will produce an enormous error in the latter. Prediction becomes impossible […]" (Poincaré, 1908). This statement was formulated for highly chaotic systems, such as atmospheric dynamics or the distribution of minor planets on the Zodiac, but the consequence is of prime importance for the forecasting of lakes dynamics: the need for encapsulating the evolution of uncertainties with the evolving system state. Since purely statistical methods were inadequate to quantify forecast uncertainties of complex non-linear systems, ensemble approaches have been proposed and represent a significant accomplishment in physical sciences (Bauer et al., 2015).

Ensemble forecasting consists in generating a number of model realizations from stochastic perturbations applied to model forcing conditions, physical processes, initial conditions or a combination of the three. Provided lakes dynamics are driven by external forcing (mainly atmospheric), and ultimately reach an equilibrium with their surroundings, those perturbations are applied to the model meteorological forcing. The entire set of model realizations is then propagated, taking observational information into account. This observational information is essential to improve the model representativeness and reduce its uncertainty. Models are indeed imperfect representations of physical processes and deviations are unavoidable due to uncertainties in processes, forcing and observations (Lahoz et al., 2010). For lakes, only the combination of remote sensing observations, numerical simulations and in-situ measurements can provide an adequate understanding of the system, reliable forecasts and associated uncertainties, at the large variety of spatio-temporal scales involved. Once a model is calibrated, this combination is achieved by data assimilation (DA), which is realized in the third chapter of this thesis.

1.3 Data assimilation

In 1809, the mathematician Carl Friedrich Gauss stated that "[…] since our observations are nothing more than approximation to the truth, […] we need a suitable combination of all observations and theory to approximate as much as possible the truth." DA is this combination, and although he was referring to planet orbits, his statement holds fundamentally also true for lakes data and models.

DA is the process by which the model of an evolving system is corrected by incorporating observations of the real system. It is an effective approach to blend observational data into model simulations (Bannister, 2017; Li et al., 2008) to improve both short-term forecasts and past (reanalysis) products (Hawley et al., 2006). One of its fundamental property is to take observation (e.g. instrument accuracy, representativeness) and model (e.g. in processes, forcing, initial conditions) uncertainties into account (Lahoz et al., 2010) to provide the analysis with corrected errors (Kourzeneva, 2014). In other words, DA aims at finding an optimal combination between the various sources of information, weighting each of them based on their error statistics.

1.3.1 The science of compromises

The multiple methods proposed for DA mainly fall into two categories: (i) variational (e.g. 3D-VAR, 4D-VAR) and (ii) sequential methods (e.g. Kalman Filtering, Particle Filtering). For variational methods, the optimization of the model states (e.g. lake temperature in this case) is based on the minimization of a cost function. Variational methods are popular in meteorological forecasting (Rawlins et al., 2007). Some drawbacks include reduced flexibility, complicated implementation of time-varying model parameters, or difficult *a priori* estimation and propagation of the background error covariance. The latter is a core element of the DA problem (Bertino et al., 2007) as it provides the spatial and multivariate structure of the analysis increments (Cao et al., 2010).

Sequential methods are robust techniques for DA, used in a broad range of applications. For linear dynamics and measurement processes with Gaussian error statistics, the Kalman Filter (Kalman, 1960) is an optimal sequential DA algorithm. It is able to estimate a non-static background error covariance based on system dynamics. However, most processes observed in nature, such as hydrodynamics, are non-linear. The analytical solution provided by the Kalman Filter can therefore not be derived to compute the posterior distribution of simulated variables. To overcome this limitation, variants exist such as the Extended Kalman Filter (EKF), consisting in a linearization of the model in the neighbourhood of the current estimate of the state vector. This linearization can lead to complicated calculations for systems with high dimensionality, the integration and propagation of the error covariance results in a significant computational demand (Gillijns et al., 2006). Linearization is done using first-order Taylor expansion, which implies a closure at the second-order moments. For highly non-linear systems this can result in an improper estimation of the state vector or covariance matrices and can therefore lead to quick divergence and instability (Moradkhani et al., 2005; Nakamura et al., 2006).

In order to cope with non-linearity and obtain a full representation of the posterior distribution, other statistical methods, such as Particle Filters, have been developed (Carpenter et al., 1999). The Particle Filter is a solution following a Darwinian-like process of survival of the fittest. Particle Filters do not need any assumption for the state variable distribution (e.g. Gaussian) and can deal with non-linear observation models as well. The updates are applied on particle weights rather than the state variable, which results in less numerical instabilities for process-based models (van Leeuwen, 2009; Liu et al., 2012; Moradkhani et al., 2005). A major drawback is the particle depletion, requiring complex resampling algorithms. Moreover, it is less computationally efficient than other approaches (e.g. Ensemble Kalman Filtering) due to the need for a high number of particles (in the order of tens of thousands). Despite its advantages, the use of the Particle Filter as an assimilation method in oceanography and limnology is limited due to its computational cost.

Overall, DA for inland waters is still in its infancy, and only a limited number of studies conducted DA experiments tailored to lakes (Kourzeneva, 2014; Stroud et al., 2009, 2010; Yeates et al., 2008; Zhang et al., 2007). Owing to the large heterogeneity found in lake dynamics, direct application of experiments designed for oceanography has been limited. This is partly due to the different scales involved and observations available. No consensus has hence been made on the right compromise (i.e. methods and algorithms) to optimally combine lake information sources.

1.3.2 The DA paradox

Bertino et al. (2007) mentioned a paradox in data assimilation, stating that computationally simple DA methods can become extremely complex engineering tasks. This occurs when the inconsistency between the stochastic and the physical model becomes relevant. Rudimentary DA algorithms (e.g. Optimal Interpolation) are computationally attractive but the estimates lack physical properties. The modeller therefore has to ensure that the optimal estimate is a valid physical solution, a complicated undertaking, particularly in the context of operational forecasting. A complexity balance between the stochastic and physical model must be found, else the compromise becomes a compromission. This thesis found that such balance is well encompassed in the Ensemble Kalman Filter (EnKF).

Ensemble Kalman Filtering – The EnKF (Evensen, 2003) is an efficient solution for non-linear dynamics and systems with high dimensionality (Crow, 2003; Reichle et al., 2002b, 2002a). It is able to account for both model and observational uncertainties and has been successfully applied to numerous problems, mainly in oceanography and atmospheric sciences (Eknes and Evensen, 2002; Evensen, 1994; Mao et al., 2009; Natvik and Evensen, 2003). The EnKF non-linearly propagates a finite ensemble of model trajectories instead of using a linearized equation for the error covariance. It is more robust than some variants (e.g. Extended Kalman Filter, Reichle et al. (2002a) and more flexible to obtain system covariances, a demanding task in DA (De Lannoy et al., 2007b). In the EnKF, error covariances are estimated dynamically from a small ensemble of model trajectories accounting for the physics of the model and grasping the essential parts of the error structure (Reichle et al., 2002b). In lake hydrodynamics, boundary conditions, especially the air-water heat and momentum budget, still contain a large uncertainty, which decreases the performance of any perfectly calibrated model. The differing model trajectories are therefore generated by introducing stochasticity in the model atmospheric forcing.

In a Bayesian inference setting, non-theoretical DA experiments blending massive datasets of both in-situ and remote sensing measurements into large three-dimensional hydrodynamic lake models are still scarce, if not non-existent. Using an EnKF, this thesis developed one such approach in its third chapter. Furthermore, considering that a sizeable proportion of the cost of producing a model forecast is associated with DA, the entire framework has been implemented with the constraints of near real-time operations in view (fourth chapter).

1.4 Novel frontiers: beyond scientists

An essential part of environmental science is to provide guidance for decision-making (Lahoz et al., 2010). This requires the timely delivery of reliable and comprehensive information, beyond the scientific community.

Three-dimensional model simulations are tools of particular importance, as they are the only source of information capable of providing a complete spatio-temporal coverage of the lake. Notwithstanding, they still require data to be calibrated (Chapter 2) and to have their deviations constrained (Chapter 3). A current challenge is the timely dissemination of lake information derived from in-situ measurements, remote sensing observations and model simulations. The wide-spread adoption of such integrated data-model approaches is still limited, partly as a result of the complexity of three-dimensional models, and their tedious parameterization, which restricted their spread beyond the expert user. Furthermore, when the effort to develop such models was made, the use of its output was often limited to the modellers themselves. Few authors proposed solutions to disseminate their results to a wider audience, thereby promoting more interdisciplinary contributions from various scientific but also public and policy-making communities. The timely and comprehensive dissemination of direct observations and numerical simulations is achieved by operational forecasting systems.

1.4.1 Operational systems

Lakes operational systems aim at providing users with real-time spatial information on the bio-physical status of the water body. The constraints associated with operational systems are numerous; data downlinks, pre-processing, ingestion in model forecasting computations, results analysis, information dissemination, etc., all have to be performed within tight production schedules. Such constraints impose hard limits on system complexity and design, with the major obstacle being reliability and stable operations.

As of today, a limited number of operational systems exist with an integrated data-model approach to monitor inland waters. Among those, the Great Lakes Operational Forecast System (GLOFS) provides nowcast and forecast guidance of various physical characteristics, such as water levels, temperature, currents, for the five North American Great Lakes (Chu et al., 2011). Other platforms have been developed such as the online lake modelling tool FLake-Global, a platform for the one-dimensional estimation of temperature and mixing conditions in any shallow freshwater lake at seasonal scale (Kirillin et al., 2011), or the three-dimensional monitoring and forecasting tool WIS-CAST, applied to a mid-sized lake for a duration of 3 months (Kimura and Wu, 2018). Nevertheless, the number of online lake operational systems remains very limited, particularly when compared to the widespread use and development of meteorological and coastal systems.

1.4.2 The *Meteolakes* platform: a successful pilot

This thesis finally aims at developing an online platform capable of delivering spatially explicit near real-time and short-term forecasts of lake hydrodynamic properties beyond scientists. A pre-operational system, called *Meteolakes*, has been developed. For more than two year, *Meteolakes* has provided spatio-temporal lake temperature and currents to more than hundred thousand visitors and has been featured in numerous local media, public events, and museum exhibitions. This has been achieved thanks to the development of a comprehensive, responsive and scalable online interface capable of openly distributing data, model results and data products: *meteolakes.ch*. This platform, detailed in the fourth chapter of this thesis, aims at bridging the gap between operational limnology and the various lake interest groups.

Impacts of *Meteolakes* have been observed on different communities: for scientists, it provided guidance in the design and planning of field campaigns, which is of particular importance for monitoring the transient dynamics of mesoscale phenomena such as upwellings and gyres. For lake-related professionals, it benefited fishermen, beach websites, rescue organisations, drinking water intakes operators, etc. Finally, hundreds of citizen have been using *Meteolakes* on a daily basis for recreational activities (e.g. navigation, swimming) and to discover how dynamic lake ecosystems are.

1.5 Dissertation objectives and outline

At the crossroads of scientific, computational and observational advances, this thesis aims at providing an end-to-end framework for producing and providing reliable data scientists and society can use. Specifically, synoptic lake physical information covering the wide range of spatio-temporal scales of inland water dynamics. It does so by addressing key challenges in three-dimensional hydrodynamic model development and data assimilation, up to the dissemination of information for interdisciplinary applications. Gaps in hydrodynamic model calibration and data assimilation are bridged through the development of open-source tools, new frameworks and methods tailored to lakes and their data sources. Finally, a solution to the lack of a flexible platform for real-time dissemination of lake information is provided.

The thesis is articulated around the following structure:

- ***Chapter 2*** covers the design and parameterization of three-dimensional lake hydrodynamic models. It proposes a parameterization framework with open-source tools to tackle parameters calibration, a complex but necessary phase of model development. Here, an automated calibration procedure, tailored to the observational data available to the user, is presented to leverage a time-consuming part of model development requiring extensive expert knowledge.

- ***Chapter 3*** proposes a data assimilation experiment to further improve the representativeness of lake models by combining numerical simulations, in a Bayesian inference setting, with satellite remote sensing and in-situ observations. This second study builds on the results of Chapter 2, as a well-calibrated model is a fundamental prerequisite to any DA experiment. Its goal is to quantify and reduce system uncertainties. Such property is of paramount importance in the operational context, to provide reliable short-term forecasts and accurate reanalysis products, for adequate water management and studies.

- ***Chapter 4*** aims at disseminating the results generated by the former chapters. It is the last step towards an end-to-end lake online service and infrastructure for sharing relevant lake information. Platform design, real-time system operations and its comprehensive and dynamic user interface, *meteolakes.ch*, are presented. *Meteolakes* ultimately aims at fostering the study of mesoscale physical phenomena, which have previously been elusive, and at building preparedness to natural and man-made lake hazards. It does so by providing, in near real-time, a combination of observations and 3D model results. The goal of this chapter is to benefit both environmental research, public and policy making communities. Its relevance is illustrated through two examples of notable mesoscale physical phenomena, including a strong upwelling and a storm event with localized fast currents, which impacted commercial and public activities, in and around the lake.

- Finally*,* ***Chapter 5*** synthesizes the thesis as a whole. It revisits the theoretical framework and outcomes in discussing future avenues for research and for the management of lakes.

Chapter 2
Automated calibration of 3D lakes hydrodynamic models using an open-source data assimilation platform

Theo Baracchini[1], Stef Hummel[2], Martin Verlaan[2,3], Andrea Cimatoribus[4], Alfred Wüest[1, 5], and Damien Bouffard[5]

[1] *Physics of Aquatic Systems Laboratory (APHYS) – Margaretha Kamprad Chair, ENAC, EPFL, Lausanne, 1015, Switzerland.*

[2] *Deltares, Delft, 2600, The Netherlands.*

[3] *Mathematical Physics, Electrical Engineering, Mathematics and Computer Science (EEMCS), TU Delft, Delft, 2600, The Netherlands.*

[4] *Ecological Engineering Laboratory (ECOL), ENAC, EPFL, Lausanne, 1015, Switzerland.*

[5] *Eawag, Swiss Federal Institute of Aquatic Science and Technology, Kastanienbaum, 6047, Switzerland.*

Abstract

The importance of the spatial and temporal heterogeneity in the distribution of all lake trophic levels is now largely recognised. Understanding its dynamics is crucial to provide scientifically credible information for ecosystem management. In this context, three-dimensional hydrodynamic models are a key information source capable to adequately assess critical but often subtle changes in lake dynamics occurring at all spatio-temporal scales. However, these models require time-consuming calibrations, often carried out either by trial and error, which strongly depends on the expertise of the modeller, or computationally heavy methods. In this study, we present a relatively simple and computationally inexpensive automated calibration framework using open-source tools. This approach is demonstrated on two different lakes (Lake Geneva and Greifensee) with an extensive dataset of field profiles of temperature and currents. Results indicate that the approach reduced the MAE of the models by up to ~50% over the baseline, while also highlighting that its success is linked to characteristics of the systems and observational data. Impacts of various parameters on both heat and momentum transfer are presented as well as their dependence on the observational setup. This study provides a computationally efficient and adaptive approach for a time-consuming task in hydrodynamic model development. The method, tailored to the calibration data available to the user, aims at (i) reducing the time spent on calibration by experts, and (ii) making three-dimensional lake modelling accessible to a broader range of users.

2.1 Introduction

Lakes play an essential role for the economic development and wellbeing of our societies. Securing sustainably ecosystem services provided by surface waters at regional to global scale, brings the need for monitoring programs to the fore (Hering et al., 2015). Traditionally, lake monitoring has been carried out by infrequent in-situ observations, which is often not enough to adequately assess critical but often subtle changes (Kiefer et al., 2015). Besides the temporal issue, the lateral variability is also lacking and new approaches including remote sensing, observations and numerical simulations are therefore required (Vörösmarty et al., 2015). Additionally, few have studied processes with horizontal gradients such as upwellings, gyres, or river intrusions, whose temporal scales are significantly shorter and their spatial variability often underestimated. This inhomogeneity can influence biogeochemical processes through horizontal advection and their associated net vertical transport (MacIntyre and Melack, 1995), hence playing an important role for the spatial structure of lake water quality. Key in grasping such variability are 3D hydrodynamic models, the only information source capable of solving entirely the spatio-temporal scales involved in local to basin-scale lake dynamics.

To accurately and reliably reproduce observed natural processes, a first and crucial phase in model development is its adequate parameterization. Calibration of the unknown model parameters is non-trivial. For instance, a number of them cannot be directly observed, are related to aggregated processes (Madsen, 2003), or depend on the model grid resolution. Hence, they cannot be determined from physical characteristics of the examined basin. This is a complex and time-consuming task that discourages non-expert and distracts resources from model-based system understanding. Traditionally, calibration is achieved by trial-and-error procedures. The latter being inefficient and complex, as well as partly subjective and relying on the experience of the modellers (Afshar et al., 2011; Fabio et al., 2010; Madsen, 2000). Those difficulties highlight the need to automate this crucial phase (Solomatine et al., 1999).

Calibration (parameter optimization) is an inverse problem that consists in minimizing a single-value cost/objective function that expresses the goodness-of-fit between computed and observed variables in numerical form (Fenicia et al., 2007) to identify parameters not known *a priori*. In the trial and error approach, the modeller adjusts repeatedly the model parameters, until an acceptable error/mismatch with respect to observations is achieved. The non-linearity and dimensionality of hydrodynamic models renders this task demanding, as changes in some parameters are often compensated by others (Bárdossy and Singh, 2008). Using limited human intervention, automated calibration, however, adjusts model parameters by solving an optimization problem (Solomatine et al., 1999), whose goal is to minimize the cost/objective function by comparing the simulated values with observations of the state of the system.

Over the last decades, in various field of environmental sciences such as surface water hydrology (Bárdossy and Singh, 2008; Hendrickson et al., 1988; Johnston and Pilgrim, 1976; Solomatine et al., 1999), and 1D2D hydrodynamics (Afshar et al., 2011; Fabio et al., 2010; Gaudard et al., 2017), computer-based automated calibration frameworks have been developed to infer unknown model parameters. However, it has been done at a significant computational cost. The computational burden associated with 3D hydrodynamic models limited the use of such frameworks. Indeed, few developed parameter search and optimization techniques adapted to the high dimensionality of 3D hydrodynamic systems.

In this study, we present the results of a new and relatively simple coupling of the Delft3D-FLOW hydrodynamic modelling suite to the open-source data-assimilation and calibration platform OpenDA. Delft3D-FLOW is an open-source multi-dimensional hydrodynamic simulation software with numerous successful applications in coastal, river, estuarine and lake domains. OpenDA is an open-source generic calibration and data assimilation environment (El Serafy et al., 2007), which can be coupled with limited software development to any model without having to change the model code. In the scope of this study, OpenDA has been extended to support Delft3D-FLOW through its file-based black-box wrapper approach. OpenDA has been successfully applied in data assimilation of current and salinity profiles (El Serafy et al., 2007), for flood forecasting purposes (Weerts et al., 2010), in calibrating the regional tidal prediction of the Singapore regional model (Kurniawan et al., 2010), or for tidal sensitivity analysis (Kurniawan et al., 2011), but not yet applied to 3D lake hydrodynamic modelling.

Parameter estimation in this kind of application became an expert-user problem. With this study, we aim at breaking it down, using open-source tools, towards more open and reproducible applications for a wider range of users. To do so, we present a systematic methodology to overcome the high computational costs associated with computer-based calibrations of 3D hydrodynamic models by proposing an adaptive and efficient calibration framework, tailored to the available calibration data, hence alleviating the time-consuming trial and error approach. Additionally, we present the limitations inherent to such an approach, given by the dependency on the type of observational data, its accuracy and its spatio-temporal density.

This framework is applied to calibration using temperature and flow-velocities of two different lake systems (both in terms of size and dynamics) using the Delft3D-FLOW hydrodynamic model coupled with the OpenDA assimilation platform. The methodology is introduced in Sections 2.2 and 2.3. The former presents the hydrodynamic model, study sites, computational domains, observations and their uncertainties. The latter details the different steps involved in the estimation process, including choice of model base and calibration parameters, software coupling and optimization algorithm. For both case studies, results (Section 2.4) show a statistical improvement over the baseline models. Section 2.4 only details the results for Lake Geneva (large-scale system), the analysis for Greifensee is available in the appendix. Finally, Section 2.5 contains the results analysis and discussion. Conclusions are summarised in Section 2.6.

2.2 Methods

2.2.1 Study sites

Lake Geneva – Lake Geneva is the largest freshwater lake of Western Europe (580 km^2 of surface area and 89 km^3 of volume, residence time of 11.4 years, Figure 2.1). Located between Switzerland and France (46.458° N, 6.528° E) at an altitude of 372 m, it is a warm-monomictic lake with its deepest seasonal mixing occurring in late winter (late February/early March). In addition to the mild climate, its maximum depth reaching 309 m prevents it from freezing in winter. The lake is thermally stratified from spring to late autumn; however, its complete overturn usually occurs only once per 5 to 10 winters on average.

Now mesotrophic, Lake Geneva is subject to strong variations of light penetration over the year. The Secchi depth typically ranges from 3.6 to 14 m and time series of Secchi depth measurements have to be used to force the model in addition to the meteorological data.

Greifensee – Greifensee is a small (8.5 km^2) lake located in the area of Zurich (Switzerland, Figure 2.1). The maximal and mean depth are 30 and 18 m, respectively. During the second half of the 20th century, Greifensee experienced strong eutrophication. As for Lake Geneva, the recovery of the lake started since strict phosphorus regulations were implemented, but even today, the lake is still considered eutrophic. This results in a shallower Secchi depth compared to Lake Geneva throughout the year. Monthly measurements, ranging from 1.5 to 8.7 m, are used to drive the model.

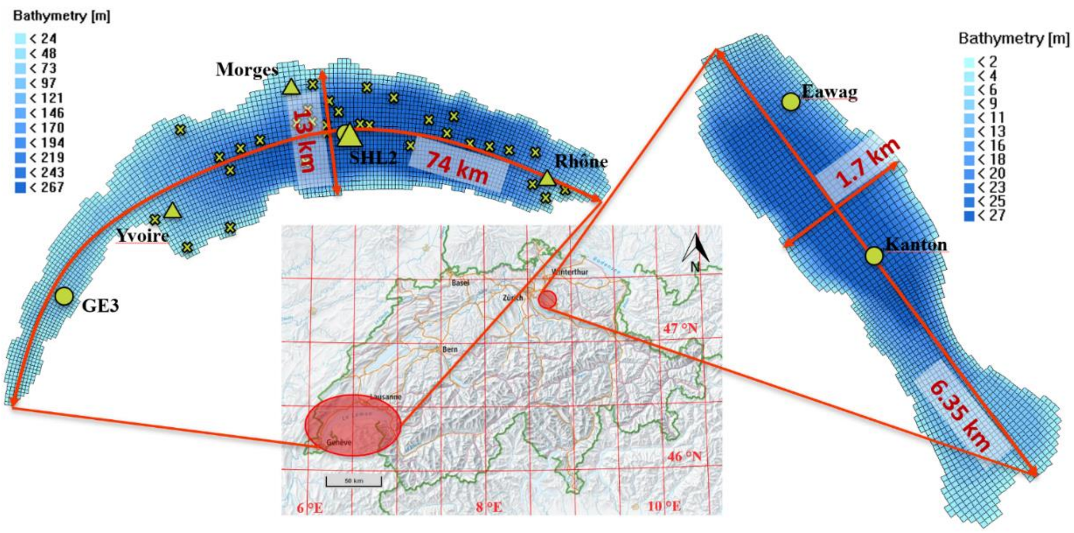


Figure 2.1: Study sites, morphology, model grid, and in-situ monitoring locations (circles are permanent stations, crosses mark single profiles, and triangles indicate current meter (ADCP) locations) [map.geo.admin.ch].

2.2.2 Delft3D-FLOW model setup

Delft3D-FLOW (tag 7426) is an open-source hydrodynamic simulation software for integrating coupled systems of differential equations for water volumes (continuity equation), linear momentum (Reynolds-averaged Navier-Stokes (RANS) equations), and tracer masses (transport equation) driven by atmospheric forcing. A detailed description of the model, its equations and numerical schemes can be found in the manual (Deltares, 2015).

In this study, the z-layers (e.g. horizontal layers) scheme is used since it is the only one capable of reproducing the lake stratification with the given steep basin morphologies. 50 layers are used for Greifensee, while 100 are used for Lake Geneva. The layers are unevenly distributed, with thinner surface layers (up to 20 cm thick right at the surface to a few meters in the deep hypolimnion). Computational grids (horizontal and vertical) are too coarse and time steps too long to resolve turbulence at small scales. The length scales of the turbulent processes are therefore of “sub-grid” scale and turbulence closure models, like the κ-ϵ model, have to be considered. A time-step of 1 minute for Greifensee and 2 minutes for Lake Geneva, maintain model stability. Simulations are initialized (uniformly horizontally) from an in-situ temperature profile taken at the deepest location of the lake in January (2014 for Greifensee, 2015 for Lake Geneva), when both lakes are partially or fully mixed.

For atmospheric forcing, we use MeteoSwiss COSMO-1 and COSMO-2 reanalysis products, consisting of seven meteorological variables on a regular 1.1 to 2.2 km grid with hourly resolution. The variables include: solar radiation, wind direction and speed, relative humidity, cloud cover, pressure, and air temperature from their physical atmospheric model tailored with data assimilation to the Alpine region. As for computation consideration; for both lakes, a single model run requires ~1 day on 3 Xeon Broadwell cores for Greifensee to ~2 days on 4 cores for Lake Geneva, indicating the paramount importance of the computational efficiency of the algorithms.

2.2.3 Assimilation platform OpenDA

OpenDA is a generic open-source data assimilation platform resulting from merging of Costa ((van Velzen and Verlaan, 2007) and DATools (El Serafy et al., 2007; Weerts et al., 2010). OpenDA is designed to be coupled with a wide range of physical process models, by making use of a set of interfaces describing the interaction between models, observations and algorithms (Deltares, 2019). Its generic and flexible interfacing protocol allows the use of existing calibration and data assimilation algorithms with any new model whose OpenDA interface is implemented accordingly.

There was no existing interface for temperature and velocity fields with the z-layers coordinate system of Delft3D-FLOW. For the purpose of this study, an OpenDA interface has been developed, as shown in the following sections, using the black-box wrapper. The coupling is a file-based approach, implemented by adding JAVA classes to allow the reading and writing of Delft3D-FLOW input-output files by OpenDA. While this method is less computationally efficient since it adds an input-output overhead, it is relatively simple and does not require modifying the model code. This interface also adds a layer allowing the conversion of some parameters, such as treating both horizontal (background viscosity and diffusivity) parameters and bottom drag coefficients (one for each direction). The entire code can be found on GitHub (https://github.com/OpenDA-Association/OpenDA).

2.2.4 Monitoring data

A key aspect in parameterization problems is the calibration data and its quality (Madsen, 2003). Hydrodynamic measurements can be demanding and costly to obtain. Moreover, in the case of 3D models the spatial variability of the lake has to be validated. This requires having multi-site (at least at two different locations), and eventually multi-variable time series of observations. In this study, we consider in-situ temperature and flow-velocity measurements collected from both, short-term field studies (single profiles) and permanent monitoring stations at various frequencies (6-hourly to monthly). The summer thermocline has a sharp temperature gradient that serves as a wave-guide for the propagation of internal waves (Bouffard et al., 2012). Moreover, Lake Geneva is wide enough for Earth-rotation affected internal waves (Bohle-Carbonell, 1986; Lemmin et al., 2005) and Kelvin / Poincaré waves are periodically observed (Bouffard and Lemmin, 2013). While basin-scale internal waves account for some of the spatial variability and dynamics observed, its calibration requires continuous high-frequency measurements covering a large spatial domain. This study here considers long-term monitoring data, and focuses on seasonal to interannual scales, hence basin-scale internal waves are not directly addressed. The ability of the model to reproduce the dynamics of these waves is only considered through spectral analysis in the discussion section.

Greifensee – The available data for Greifensee stems from two permanent field stations sampling temperature over the water column at different frequencies (Figure 2.1). The northern station (Eawag) has 321 profiles measured at a 6-hourly interval during summer 2014 with a maximum depth of 17 m, while the central station (Canton) has 12 monthly profiles over the same year with a maximum depth of 30 m (lake's deepest location).

Lake Geneva – The available dataset for Lake Geneva consists of 90 temperature profiles over the water column at 32 different locations (Figure 2.1) sampled during the years 2015 and 2016. While most of the locations have been sampled only once (black crosses in Figure 2.1), monthly to bi-monthly measurements are available at two stations (GE3 and SHL2, black circles in Figure 2.1). Some results will therefore focus on those stations. The entire dataset has been used for calibration.

In addition to temperature data, flow velocities are measured at four locations (triangles, Figure 2.1) with Acoustic Doppler Current Profilers (ADCPs, Cimatoribus et al., 2017). This data consists in continuous (over up to six months) flow velocities measurements at various depths (mainly in the hypolimnion) nearshore and offshore. Since the aim was to calibrate the model's mechanical energy content, only the current speed is considered and the directional flow component is discarded. This is also a consequence of the uncertainties inherent to wind forcing and monitoring data described in the next section, hence a perfect model match with data cannot realistically be expected. Finally, this high-frequency dataset has been down-sampled to 6-hour intervals, while the higher frequency variability will be considered in the data uncertainty definition to account for unavoidable model phase shifts.

Data uncertainty – In addition to model uncertainties (e.g. initial conditions, physical processes and their parameterizations, approximations of the system) and forcing errors, measurement data is subject to systematic and random errors (Bárdossy and Singh, 2008). Uncertainty of the information has to be considered in the calibration framework. Increasing the amount of data does not necessarily result in better parameter estimations (Sorooshian et al., 1993). Indeed, the quality of the data (e.g. instrument accuracy, measurements close to detection limits, good spatial coverage of the observations) is of prime importance.

In this study, we define the observational uncertainty as the maximum value of the two following elements: (1) the instrument precision (CTD or ADCP), and (2) the temporal dynamic variability at the measurement location. While the first is a fixed value (here, 0.002 °C or 0.8 cm/s), the latter is obtained from either model simulation results (in the case of temperature) or measurements (flow velocities) by computing the standard deviation of the observed variable over a period of ±6 to ±12 hours (depending on the lake and the period of basin-scale modes). The latter especially affects temperature uncertainty at the thermocline level and flow velocities but does not influence the seasonal scale hydrodynamics, which this study focus on. The goal of this procedure is to avoid affecting the cost function (Eq. (2.1)) too much when an event (e.g. internal wave) is not phased correctly, since this could result from uncertainties elsewhere (e.g. in forcing variables, which also come from a model). Absolute values of flow velocities are also averaged over the hour.

Lake hydrodynamics can be understood in terms of kinetic and potential (thermal) energy budgets, both requiring to be calibrated. Yet, kinetic and thermal dynamics do not have a first order reaction to changes in all parameters. The heat flux determines the surface boundary structure, while kinetic energy determines diffusion in the stratified parts. By selecting the appropriate set of parameters to calibrate, the two calibrations can be decoupled. In the case of Lake Geneva, observed flow velocities are most of the time close to their measurement accuracy (especially in the hypolimnion). This results in a much lower impact on the cost function compared to temperature measurements, whose accuracy is order of magnitudes higher than their standard deviation (see cost function Eq. (2.1)). Therefore, calibration of temperature and flow velocities will have to be decoupled to assess their influence on parameters and the performance of the calibration. While a single calibration is presented for Greifensee, two sequential setups are performed in the case of Lake Geneva. Flow-velocity calibration builds on the results obtained by temperature calibration. To avoid having the second calibration deviate too much and destroy the optimization for temperature, weak parameter constraints have been enabled (detailed in Section 2.3.1).

2.3 Calibration

When dealing with optimization schemes, global and local methods can generally be distinguished (Madsen, 2000). While global methods (e.g. particle filters, population-evolution-based search algorithms, genetic algorithms) are more robust in finding the global minimum in the parameter space of the cost function, they are computationally expensive (Fabio et al., 2010), thereby preventing their use for 3D hydrodynamic models and limited computing resources. Local methods on the other hand are more efficient, at risk of falling into local minima and having their outcome dependent on the initial setup.

2.3.1 Calibration algorithm

DUD (Doesn't Utilize Derivatives) is a derivative-free OpenDA algorithm for parameter estimation capable of coping with non-linear models (Ralston and Jennrich, 1978). DUD can be compared to a Gauss-Newton method as it transforms the non-linear least square problem to minimize into a linear one. However, instead of doing so by computing its derivatives (approximating its tangent function), it uses an affine function for the linearization (Ralston and Jennrich, 1978). This is interesting for objective functions in the form of a sum of squares, as Gauss-Newton based algorithms are faster (Bard, 1970; Box, 1966). Additionally, Gauss-Newton algorithms are widely used and have similarities with robust estimation methods (Beaton and Tukey, 1974) as well as with maximum likelihood estimation (Bradley, 1973; Charnes et al., 1976; Nelder and Baker, 2006; Ralston and Jennrich, 1978). A competitive advantage of DUD is its performance with respect to the low number of model iterations it needs, which is a requirement for the models used in this study.

DUD calibration is an iterative process, whose optimum is reached when specified convergence criterions are met. The minimized cost function is of the following form:

$$J = \Sigma_t \frac{(y_0(t) - y_m(t))^2}{\sigma_0^2} \tag{2.1}$$

With $y_0(t)$ the observation at time t, $y_m(t)$ the modelled value at time t, and σ_0 the observational uncertainty.

The choice of this algorithm has been motivated by its ability to deal with non-linear models while being relatively computationally inexpensive (usually, less than 10 model runs are required to reach convergence) in comparison to other algorithms. Some of its capabilities have been demonstrated by (Garcia et al., 2015) in similar problems. DUD can be used with or without user-defined constraints. In this study, constraints are used and detailed in the appendix. Finally, for the current speed calibration of Lake Geneva, the weak parameter constraint is enabled. This setting adds a penalty when the optimization changes a parameter away from its initial guess.

The entire procedure has been first tested through a twin-experiment with a small-scale lake (25 vertical z-layers, 339 horizontal grid points). The model was quick to converge back to its initial parameters, which generated the observations used in the experiment.

2.3.2 Calibration parameters

In 3D hydrodynamics and specifically to Delft3D-FLOW, some parameters play an essential role in the spatial variability of basin dynamics and mesoscale processes (e.g. up-/downwellings), yet they cannot be obtained with *a priori* knowledge. Those parameters are related to momentum and heat transfer, deep mixing, internal wave propagation and dampening, etc., such as diffusivities and viscosities. Some of them are grid-dependent (Toffolon and Rizzi, 2009). The approach presented in this study aims at focusing on quantities that are complicated to interpret with *a priori* knowledge. Its goal is to minimize Delft3D-FLOW model parameterization uncertainty while maximizing the calibration performance. Parameters with a direct interpretation, such as the Secchi depth, will still have to be user-defined. Our strategy is to focus on variables controlling the energy pathway (heat and momentum). Such parameters are described in Table 2.1, while our choice is motivated in the following sub-sections.

Table 2.1: Delft3D-FLOW calibration parameters acting on the energy pathway.

Parameter	Description	Influence on
$C_{10@0.5}$ $[-]$	Wind drag coefficient at 10 m above water and at speed of 0.5 m/s	Temperature/flow velocity
$z_{0,bottom}$ $[m]$	Bottom hydraulic roughness height	Flow velocity
C_{Dalton} $[-]$	Evaporative heat flux coefficient	Temperature
$C_{Stanton}$ $[-]$	Convective heat flux coefficient	Temperature
DV_H $[m^2 \cdot s^{-1}]$	Background horizontal viscosity and diffusivity	Temperature/flow velocity
DV_V $[m^2 \cdot s^{-1}]$	Background vertical viscosity and diffusivity	Temperature/flow velocity

Surface forcing and energy source – The thermal- and fluid-dynamics of a lake are mainly driven by the interactions with the atmosphere (Imberger and Hamblin, 1982) and only to a minor part by the Coriolis effect. The proper heat and mechanical energy transfer is thereby of paramount importance for the performance of the model, especially for surface and mixed layer processes. In that regard, the wind drag coefficient acts on heat and momentum transfer into the water. This parameter (often a function of wind speed) can be challenging to determine. In addition, changes in the wind drag coefficient can compensate inadequacy in the wind forcing, which are common in almost all topographies. Its characterization has been the aim of several studies (Wüest and Lorke, 2003) in both limnology and oceanography. In this paper, the wind drag coefficient at low wind speeds (0-5 m/s, most common winds in the considered region) will be calibrated.

Viscosities and diffusivities – Lakes are strongly stratified during summer, which results in deep waters being largely isolated from shallow waters. Moreover, large lakes are often horizontally heterogeneous which leads to spatially variable forcing. To encompass the range of system dynamics, parameters acting on both deep-water processes and horizontal transport have to be calibrated. Diffusion and dissipation related values will tend to act on such deep mixing, thermocline depth, flow velocities, internal waves and horizontal processes. In Delft3D-FLOW, background values for both vertical and horizontal diffusivities and viscosities can be specified by the user.

They can hence be of particular importance in the case of strongly stratified flows, as the layers may behave as if they were decoupled.

Both horizontal parameters depend on the grid size (*a priori* knowledge from the model manual suggests values ranging from 1-10 m^2/s for grids of ~10 m and 10-100 m^2/s for grids of 100 m or more). It is also specified (Deltares, 2015) that both coefficients are calibration parameters and have to be determined during the calibration phase. For 3D simulations, turbulent vertical eddy viscosity and diffusivity are computed from a turbulence closure model. In the κ-ϵ model, the eddy diffusivity is derived from the eddy viscosity. The applied vertical viscosity is the max of the user-defined background value, ν_V^{back}, and the calculated value from the turbulence model ν_{3D}, (Eq. (2.2)). The same applies for the vertical diffusivity. Since those values, derived from the turbulence model, are based on system dynamics, we do not consider them as candidates for calibration. Horizontal parameters will be considered as identical (diffusivity = viscosity) and be calibrated by the algorithm:

$$\nu_V = \nu_{mol} + max(\nu_{3D}, \nu_V^{back}) \tag{2.2}$$

$$\nu_H = \nu_{SGS} + \nu_V + \nu_H^{back} \tag{2.3}$$

with ν_{mol} for the kinematic viscosity of water, ν_{3D} is computed by the turbulence closure model, ν_V^{back} is the background vertical viscosity, and ν_{SGS} indicates the viscosity obtained from the sub-grid scale model.

Bottom boundary and energy sink – Finally, the main energy sink in lakes is by frictional dissipation at the bed. The bottom drag coefficient acts on this other boundary dynamics and the associated dampening of near-sediment flows and internal waves. Calibration of this parameter is also in accordance with similar problems found in oceanography. Specifically, we calibrate the bed roughness height through the z_0 parameter, which can also be a function of the size of the computational grid.

Other parameters and sensitivity analysis – In addition, we recommended an analysis to assess the sensitivity of the cost-function of the calibration, as it will determine which parameters are estimable (Skahill and Doherty, 2006). This is of particular importance to find the right trade-off between having enough parameters to lower the cost function to the extent it could be lowered with more parameters, while still conserving uniqueness in their estimation and maintaining a reasonable runtime (more parameters will require the DUD algorithm to perform more model runs). Within the Delft3D-FLOW modelling environment and specific to the two lakes studied, sensitivity analysis highlighted another parameter of significant importance for the aforementioned subdomains. The Dalton number (c_e), which act on the forced evaporative heat flux (Eq. (2.5)), which is part of the total heat budget:

$$Q_{tot} = Q_{sw} + Q_{alw} - Q_{lw} - Q_{ev} - Q_{conv} \tag{2.4}$$

$$Q_{ev,forced} = L_V \rho_a c_e U_{10} \cdot (q_s - q_a) \tag{2.5}$$

with L_V the latent heat of water, ρ_a the density of air, U_{10} the wind speed at 10 m above water level, q_s and q_a the specific humidity of saturated and remote air (10 m above water level), respectively. The c_e value will not affect the thermocline depth but plays an important role in the modelling of the mixed layer temperature, especially during the stratified period. It is worth noting that the Dalton number, whose signature is only visible in temperature changes, is kept fix in the second (current speed) calibration. This is especially relevant since there are no flow velocity measurements in the mixed layer.

2.4 Results

In this section, quantitative and qualitative results for both lakes are presented. Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Taylor Diagrams (Taylor, 2001) are used as indicator of the calibration framework performance. Profiles over the water column and time-series at multiple depth are used to illustrate the calibration of Lake Geneva. For Greifensee, the same is available in the appendix.

Greifensee has been calibrated with temperature profiles at two locations. Improvements in terms of MAE are at 51%, reaching a final MAE of 1.4 °C. A summary of RMSE and MAE values can be found in the appendix. Relative to Greifensee, Lake Geneva's improvements are lower. The main difference comes from a better initial model for Lake Geneva. The model reached a MAE of 1.12 °C for the temperature calibration and 2.5 cm/s for the flow velocity calibration. MAEs, RMSEs and improvements are summarized in Table 2.2.

Figure 2.2, shows the improvements brought by the automated calibration through Taylor diagrams. In the case of Greifensee, the centered Root Mean Square Difference (cRMSD) is reduced, while the correlation increased. However, no significant changes in the temperature standard deviation are achieved. For Lake Geneva, the calibration improvements are minimal in terms of cRMSD and correlation. However, changes in flow velocity standard deviation are substantial, now significantly closer to the observed flow variance.

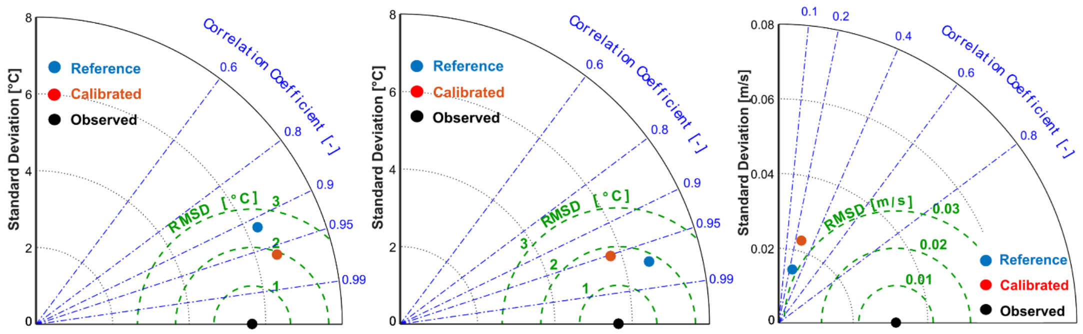


Figure 2.2: Taylor diagrams showing the part of cRMSD attributable to variance or pattern correlation (Taylor, 2001). The radial distance from the origin (bottom-left) defines the standard deviation. The radial extent from observations (black dot) to model values (red and blue dots) corresponds to the cRMSD. The azimuthal position provides the correlation coefficient. Taylor diagrams for Greifensee (left), Lake Geneva temperature calibration (center) and Lake Geneva current speed calibration (right).

Table 2.2: Summary of the calibration performance (MAE and RMSE) for Lake Geneva.

	Initial MAE	Initial RMSE	Final MAE	Final RMSE	**MAE improvement**	**RMSE improvement**
Greifensee	2.9 [°C]	3.6 [°C]	1.4 [°C]	1.9 [°C]	51 [%]	47 [%]
Lake Geneva	1.34 [°C] 2.7 [cm/s]	1.95 [°C] 3.9 [cm/s]	1.11 [°C] 2.5 [cm/s]	1.80 [°C] 3.6 [cm/s]	16.9 [%] 7.8 [%]	7.7 [%] 7.5 [%]

Parametric evolution is significant (Table 2.3). All parameters increased except the background horizontal viscosity and diffusivity, which approached its lower bound (set at 4.5 x 10^{-10} m^2/s, see appendix). The largest change observed is in the hydraulic roughness, which went from 1 cm to 1.1 m. The Dalton number is of greater influence during the stratified period, starting around May. It is worth remembering that the Dalton number has only been calibrated in the temperature calibration experiment. Finally, the wind drag coefficient at low wind speeds increased.

Table 2.3: Parameters evolution after temperature calibration at the western station (GE3) and central station (SHL2) of Lake Geneva.

Parameters	Initial value	Final value
$C_{10@0.5}$ $[-]$	9.8 x 10^{-3}	4.1 x 10^{-2}
$z_{0,bottom}$ $[m]$	1.0 x 10^{-2}	1.1 x 10^{-0}
C_{Dalton} $[-]$	1.3 x 10^{-3}	2.5 x 10^{-3}
DV_H $[m^2 \cdot s^{-1}]$	1.0 x 10^{-5}	1.6 x 10^{-9}

Two stations are selected for a detailed analysis of the temperature calibration, GE3 and SHL2 (Figure 2.1). These are the only stations to have monthly temperature profiles over the entire calibration period of two years. Improvements at SHL2 are significant (17 % MAE reduction), reaching a final MAE of 0.10 °C. Due to its shallower depth, GE3 benefited more from the calibration, MAE is reduced by 42 % (final MAE of 0.085 °C). For flow velocities, reductions in MAE for SHL2, Rhône, Morges, Yvoire are 7 %, 9 %, 15 % and 3 %, respectively. Selected results of three out of four stations are shown hereafter.

Profiles – Figure 2.3 provides a comparison of vertical temperature profiles at SHL2 for different times over the two years simulation. Refinements of the temperature modelling are reached over all seasons. As expected, most of the improvements come from better mixed layer modelling and more accurate thermocline depth.

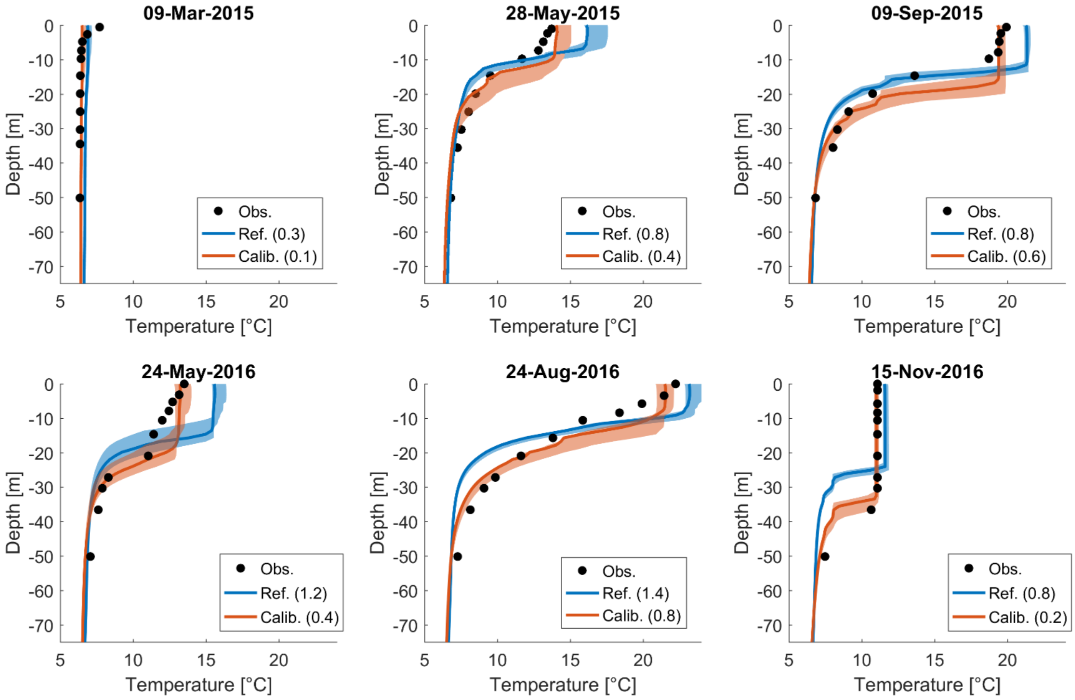


Figure 2.3: Temperature profiles at 10 a.m. at the SHL2 station. Shaded areas represent the modelled temperature variability over ±36 hours. Numbers in parenthesis correspond to the MAE in [°C].

Figure 2.4 displays vertical velocity profiles for the Morges and Yvoire stations at the beginning, middle and end of an ADCP measurement campaign. Overall, the calibrated model shows higher velocities than the reference run, mainly in the epilimnion. This is more in accordance with the observed flow velocities. In the hypolimnion, low flow velocities (< 2 cm/s) are recorded. While the reference model had a relatively low mismatch in this region, the observations are now mostly within the variability (over basin-scale wave periods) of the calibrated model.

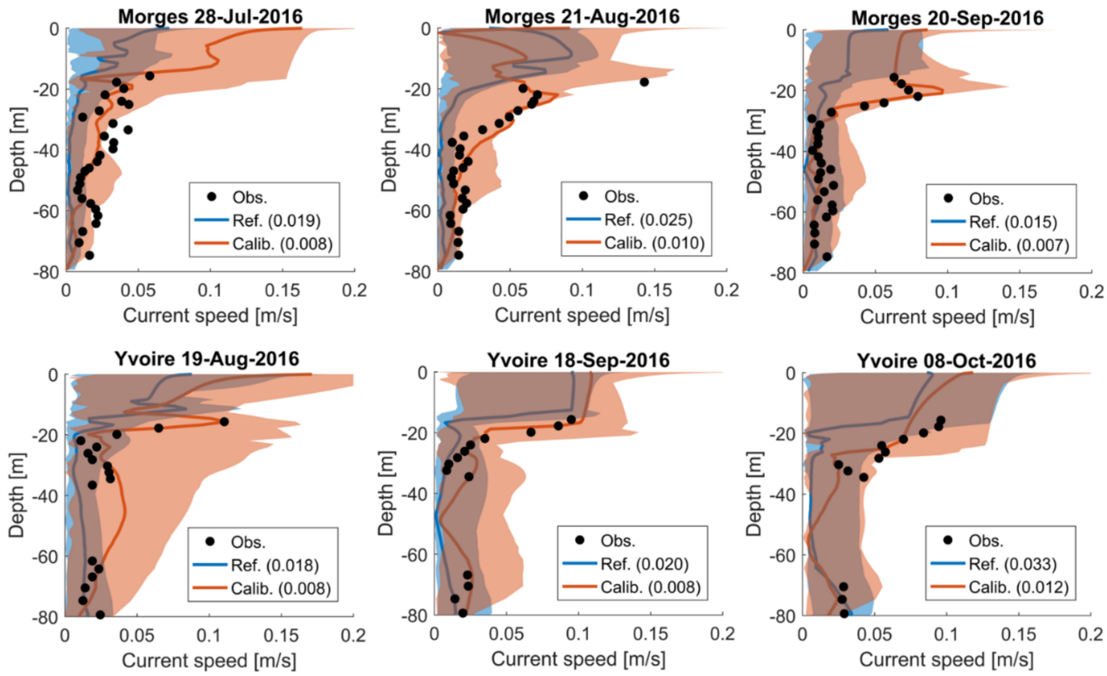


Figure 2.4: Selected velocity profiles at specific times at the Morges and Yvoire stations (Figure 2.1). Numbers in parenthesis correspond to the MAE in [m/s]. Shaded areas represent the modelled velocity variability over ±36 hours.

Time-series – Figure 2.5 provides time-series of temperature at GE3 station for various depths. Unlike for Greifensee, the model performance is better over the entire range of vertical layers. The thermocline depth, located between 20 and 30 m, is subject to the highest temporal variability, shown by the shaded red and blue areas (see appendix). At those depths, when an observation does not precisely match the modelled value, it is still located within the interval representing the variability of basin-scale internal waves. Additionally, model match and calibration performance are consistent over the two years.

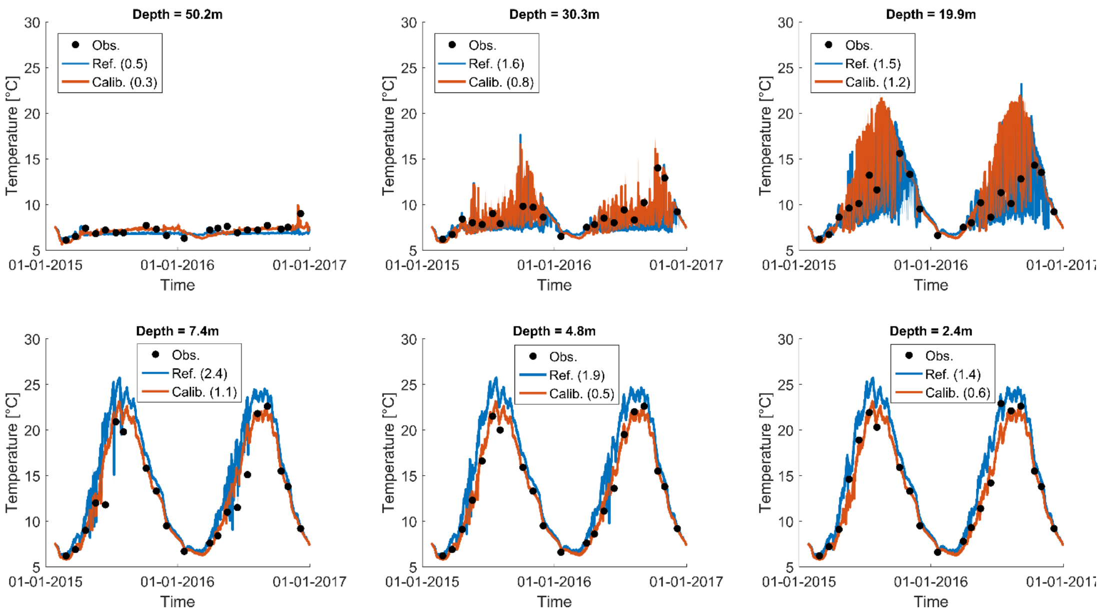


Figure 2.5: Temperature time-series at station GE3 (Figure 2.1). Numbers in parenthesis correspond to MAE [°C]. Shaded areas represent the modelled velocity variability over ±2 vertical layers.

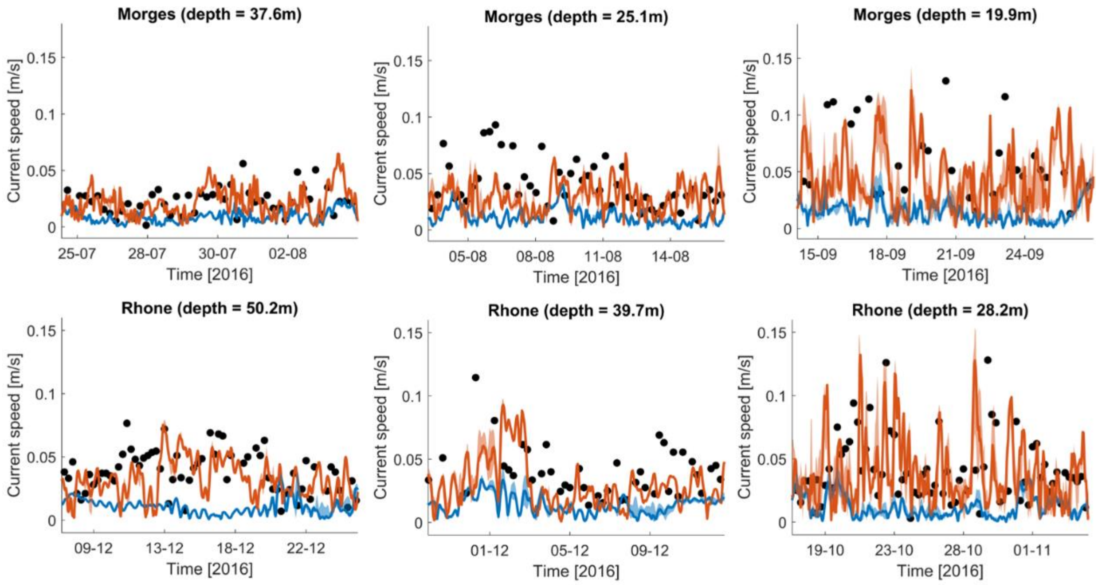


Figure 2.6: Selected flow velocities time-series at the Morges and Rhône stations (Figure 2.1) at various depths and periods. Red lines are the calibrated model, blue the reference, and black dots are observations. Shaded areas represent the modelled velocity variability over ±2 vertical layers.

Figure 2.6 shows time-series of flow velocities at two stations over the various depths and measurement periods they cover. As indicated above, the main improvement is the increase in average magnitude and amplitude of the currents, which could also be seen in the Taylor diagrams though the increase in standard deviation (Figure 2.2). Surface layers benefited the most from the calibration. Indeed, in those regions, the reference model failed to reproduce values high enough to match the data recorded by ADCP measurements. The calibrated model still failed at reproducing some high velocities, such as the one measured at the Rhône station at 39.7 m around December 1st, 2016. The overall event however has been captured as shown by the significant increase in modelled velocity around that date.

2.5 Discussion

The improvements brought by the OpenDA automated calibration are substantial for both lakes studied. For the temperature calibration, results are consistent in both lakes. Largest gains in modelling accuracy are obtained in the mixed layer and at the thermocline depth. Those regions are the most sensitive to the initial parametric setup and where the largest discrepancies are observed before calibration. For a calibration of basin-scale internal waves, continuous measurements over the entire year (or entire calibration period) at high frequency are needed at several locations. In that regard, a powerspectra of flow velocities in the hypolimnion at the central part of the lake (SHL2) is provided in Figure 2.7. The spectra shows a peak near the 12-13 hours mark in both data and models. Such a period is most likely the signature of Poincaré waves and is consistent with the findings of (Lemmin et al., 2005). Figure 2.7, therefore, highlights the capability of the model for reproducing internal waves dynamics.

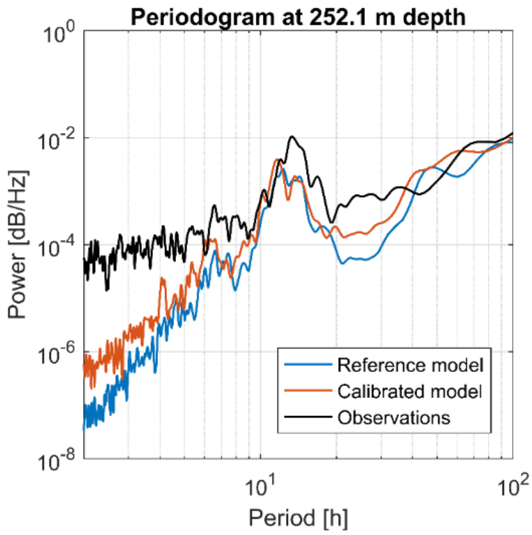


Figure 2.7: Periodogram of modelled and observed flow velocities at 252 m depth at SHL2 (centre of the lake, Figure 2.1).

The calibration performance of Lake Geneva was lower relative to Greifensee. This is mainly related to the depth of the lake. Deeper lakes are less sensitive to the initial parametric setup and typically have a better initial performance. Indeed, Lake Geneva's depth of 309 m makes it more sensitive to initial conditions, since mixing will not reach the bottom layers at short time-scales. However, model layers in the hypolimnion are less coupled to atmospheric model forcing and are therefore less dynamic.

At thermocline depth, measurements show significant variabilities due to basin-scale internal oscillations. Data-model mismatches can be observed, yet modelled internal wave oscillations are compatible in amplitude with observed ones. However, a phase shift is most likely present. The variability of the system is reproduced by the model, which can also be seen in the Taylor diagrams showing comparable model-data standard deviations. Such behaviour is also consistent with the uncertainty given to the observation in Section 2.2.4. Indeed, the goal was not to penalize excessively the model for being out of phase with its basin-scale internal waves since this could be the result of various other elements (such as inaccuracies in the wind forcing).

In terms of heat transfer, the model performance is most sensitive to the Dalton number during the stratified period, where it will be able to influence the most the temperature of the mixed layer. This parameter is hence of particular importance when calibrating a model using temperature data. Sensitivity analysis showed that a similar parameter, the Stanton number, which allows a calibration of the convective heat flux, did not have the same impact on model results.

The bottom boundary roughness length (z_0) suggested for Lake Geneva by the algorithm after calibration (1.1 m) is significantly higher than its baseline value (0.01 m) and what has been assumed in a previous study of Lake Geneva by Umlauf and Lemmin (2005). However, this value cannot be directly compared to its physical counterpart, due to effects of the computational grid. This is especially the case with z-layers, some of them reaching a thickness larger than 6 meters at the bottom boundary. This staircase effect is further exacerbated by the horizontal grid size of 450 meters. Those factors combined prevent solving the bottom roughness length and relating it to the physics of the bottom log-profile. Both the temperature and flow velocity calibration required an increased value of that parameter.

The wind drag coefficient at low wind speeds ($C_{10@0.5}$) increased. The end value differs from what has traditionally been found for oceans. It is consistent with findings from Wüest and Lorke (2003) for lakes, a value tainted, however, by large uncertainties. The argument is that at low wind speeds the atmosphere can reach unstable conditions, thus creating more turbulent exchange. It is worth pointing out that for Greifensee, the wind drag after calibration is significantly lower than its baseline. Assessing modelled currents of the system showed that the value suggested by a unique temperature calibration is most likely too low. This parameter is particularly sensitive to ADCP data, which emphasizes the need to have both temperature and current speeds for a complete system calibration.

Background horizontal parameters, capable of adding dampening internal motion or adding mixing through additional diffusion, were reduced close to their imposed lower bound (cf. appendix). The Delft3D-FLOW model of Lake Geneva hence did not require additional viscosity and diffusivity to the horizontal values computed from the state of the system.

Finally, the Dalton and Stanton numbers are parameters acting only on the heat transfer, therefore requiring temperature data to be calibrated. For the same reason they have been removed (considered as fixed parameters) from the flow velocity-based calibration. For the lakes studied in this paper, prior sensitivity analysis showed that the Dalton number, related to the evaporative heat flux, turned out to be of stronger influence than the Stanton number. However, that may not be the case for every system, some being more sensitive to convective heat flux. A sensitivity analysis tailored to each lake is therefore necessary before any calibration. For both lakes, larger values after calibration indicated a need for a larger heat loss through evaporation. This parameter, notably sensitive to temperature data from CTD profiles, contributed significantly in regulating the mixed layer temperature.

In terms of hypolimnetic flow velocities, both computed and measured values are close to the measurement accuracy of 0.8 cm/s (Teledyne RDI Workhorse Sentinel). Compared to temperature observations, their impact on the cost function is therefore almost negligible. This is a consequence of the temperature measurement error being orders of magnitude smaller relative to their values. Such discrepancy hinders the use of different kinds of data simultaneously with this methodology and hence a separated calibration had to be carried out for both types of observational data. Such limitation goes along the previous discussion item that uneven data density and coverage (spatially and temporally) will bias the cost function towards specific space-time locations. The same holds true for data uncertainty. It is worth noting that the temperature calibration did not degrade significantly after the calibration based on flow velocity. This is both a result of keeping the Dalton parameter constant and enabling the weak parameters constraint (Section 2.3.1) of the DUD algorithm.

However, the two-step calibration procedure allows for a comprehensive understanding of the role of parameters on the energy pathway and type of data required for an optimal calibration of seasonal hydrodynamic models. For instance, in addition to requiring homogeneous and continuous data in space and time, ADCP observations in surface layers would have been better suited for a calibration since the flow velocities there are significantly higher than the measurement accuracy of the instrument.

Overall the procedure has been found to be computationally efficient and relatively simple to use or expand to other systems and models by non-expert users. Significant time savings are achieved with such automated approaches, which allow the end-user to focus on model fine-tuning. The entire source code of OpenDA and developments made in this study are available on GitHub at https://github.com/OpenDA-Association/OpenDA.

2.6 Conclusion

Model calibration of computationally expensive 3D hydrodynamic model has been a challenge in physical limnology and oceanography. Despite the increase in computing capabilities, no simple solution and tools have been proposed for the automated parameterization of 3D hydrodynamic models using local computers with limited parallelism. Calibration of such systems has remained a task performed by trial-and-error requiring expert knowledge. Since lakes provide essential ecosystem services, the monitoring capabilities at previously unresolved (by traditional measurements) locations and spatio-temporal scales offered by those models has become of prime importance. This study shows that a relatively simple coupling approach of existing open-source tools provides significant improvements to the parameterization of three-dimensional inland-water models without user intervention.

Using temperature and flow velocity measurements at various locations and frequencies, two lake systems of different scale and morphology have been calibrated successfully. Insights were gained on the data required and importance of specific parameters for the dynamics in different seasons at various depths. In particular, the evaporative heat flux plays a decisive role in mixed layer temperature dynamics during the stratified period. The wind drag coefficient became system-specific and affected both thermocline depth as well as surface flow velocities. Such influences indicate that calibration of heat and momentum transfer, dissipation and background horizontal processes matter. Overall, we found that current measurements contain decisive information about deep-water dynamics, while temperature observations rather provide a good description of processes occurring in surface layers. The combination of both data type is thereby of prime importance for a more complete understanding of the system.

This study presents an adaptive approach, tailored to lakes and models of various scales and morphologies, with minimal development required. Compared to previous manual calibrations, better results are obtained without user intervention, hence minimizing the need for knowledge of the physical process and the modelling tools. In particular, the latter result has been achieved by keeping computational cost reasonably low: only 6 to 12 runs are required, on local workstations.

Such solution aims at facilitating three-dimensional modelling of lake ecosystems, rather than achieving the best possible calibration. Fine-tuning is a step that remains necessary, based on the number, quality and diversity of validation data available. As lakes are considered sentinels for climate and catchment changes (Adrian et al., 2009; Shimoda et al., 2011; Wagner and Adrian, 2009), a better modelling and monitoring of lakes' dynamical responses towards external influences will contribute to a more sustainable management and to securing their essential ecosystem services.

Code and data availability

Software – The source code and documentation of the numerical model (Delft3D-FLOW) and data assimilation platform (OpenDA) developed in and for this study can be accessed and downloaded on their online repositories at https://oss.deltares.nl/web/delft3d/source-code and https://github.com/OpenDA-Association/OpenDA.

Data – The authors are grateful to the following institutions that provided the data used in this paper: to Dr. Francesco Pomati and the Phytoplankton Ecology Lab in the department of Aquatic Ecology at Eawag for the temperature dataset at the Eawag station of Greifensee, to Amt für Abfall, Wasser, Energie und Luft (AWEL) for CTD profiles from Greifensee at the Kanton station, the Federal Office of Meteorology and Climatology (MeteoSwiss) for spatiotemporal meteorological data, and the SECOE Direction Générale de l'Eau du Canton de Genève (CH) for in-situ temperature data on Lake Geneva at GE3. In-situ data at SHL2 as well as Secchi disk measurements in Lake Geneva were provided by the *Commission International pour la Protection des Eaux du Leman* (CIPEL) and the Information System of the *SOERE OLA* (http://si-ola.inra.fr), INRA, Thonon-les-Bains. All this data cannot be published as it belongs to their aforementioned owners, it is not the property of the authors of this study. It can nonetheless be requested by contacting its respective institution. Any other data used in this study is property of the Physics of Aquatic Systems Laboratory at EPFL and can be obtained by contacting Prof. Alfred Johny Wüest (alfred.wueest@epfl.ch).

Appendix

Table S2.1: Parameters constraints defined in the calibration algorithm.

Parameter	Description	Initial value	Lower bound	Upper bound
$C_{10@0.5}\ [-]$	Wind drag coefficient at 10 m and 0.5 m/s	9.8 x 10^{-3}	6.6 x 10^{-5}	1.1 x 10^{-1}
$z_0\ [m]$	Bottom hydraulic roughness height	1.0 x 10^{-2}	1.0 x 10^{-3}	6.7 x 10^{-0}
$C_{Dalton}\ [-]$	Evaporative heat flux coefficient	1.3 x 10^{-3}	0.0 x 10^{-0}	5.0 x 10^{-2}
$DV_H\ [m^2 \cdot s^{-1}]$	Background horizontal viscosity and diffusivity	1.0 x 10^{-5}	4.5 x 10^{-10}	x 10^{+2}

A. Results Greifensee

Most parameters evolved significantly. While the wind drag coefficient at low wind speeds has been reduced, all other three parameters increased (Table S2.2).

Table S2.2: Parameters evolution for Greifensee calibration

Parameters	Initial value	Final value
$C_{10@0.5}\ [-]$	9.8 x 10^{-3}	5.5 x 10^{-5}
$z_{0,bottom}\ [m]$	1.0 x 10^{-2}	4.9 x 10^{-2}
$C_{Dalton}\ [-]$	1.3 x 10^{-3}	3.3 x 10^{-3}
$DV_H\ [m^2 \cdot s^{-1}]$	0.0 x 10^{-0}	2.8 x 10^{-0}

The low wind drag value is contrary to what has been suggested by Wüest and Lorke (2003) for small-scale hydrodynamics in lakes. The bottom hydraulic roughness z_0 increased from 1cm to ~5 cm, which corresponds to values observed in the literature. The latter contributes to increase the dissipation of energy. The same is true for the horizontal viscosity and diffusivity. Finally, mixed layer temperature has been greatly improved by an increase in the Dalton number, enhancing the evaporative heat flux.

B. Greifensee profiles

Figures S2.1 and S2.2 show the improvements through profiles at different times for the two locations. Substantial improvements are made in the mixed layer as well as for the thermocline depth over the entire year. However, the calibration has not been successful in the deep water. In terms of stations, better results were obtained for the northern (Eawag) location (Figure S2.1). The latter, in combination with the poorer deep water performance, are both related to a limitation of the method discussed in the next section.

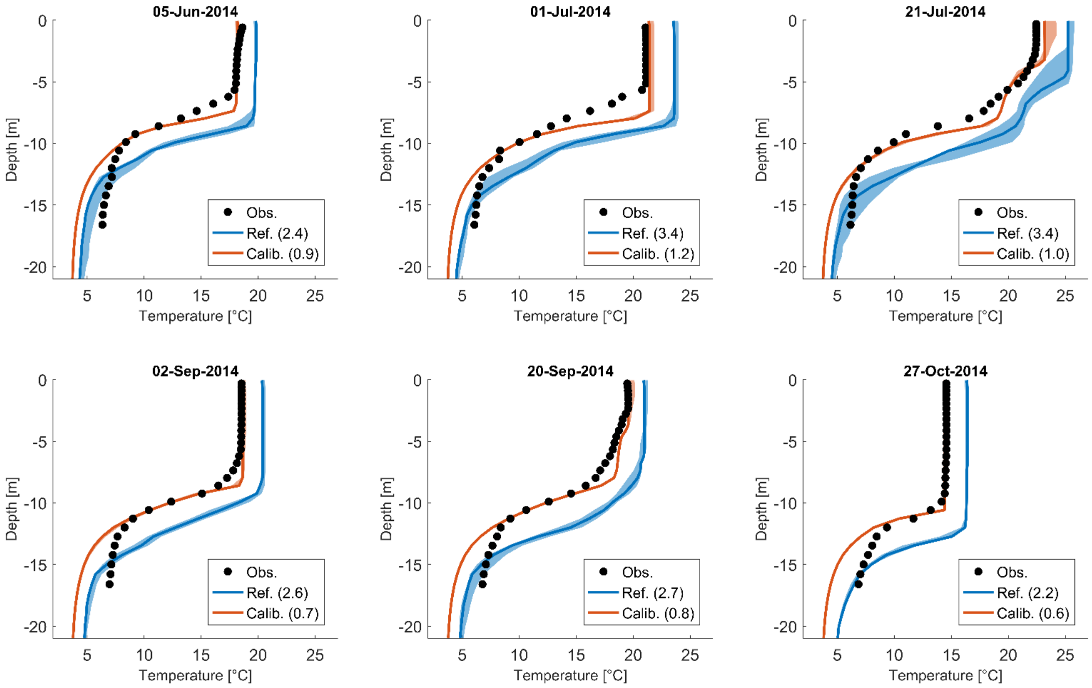


Figure S2.1: Temperature profiles at specific times at the high-frequency northern station (Eawag) with MAEs. Shaded areas represent the modelled temperature variability over ±12 hours.

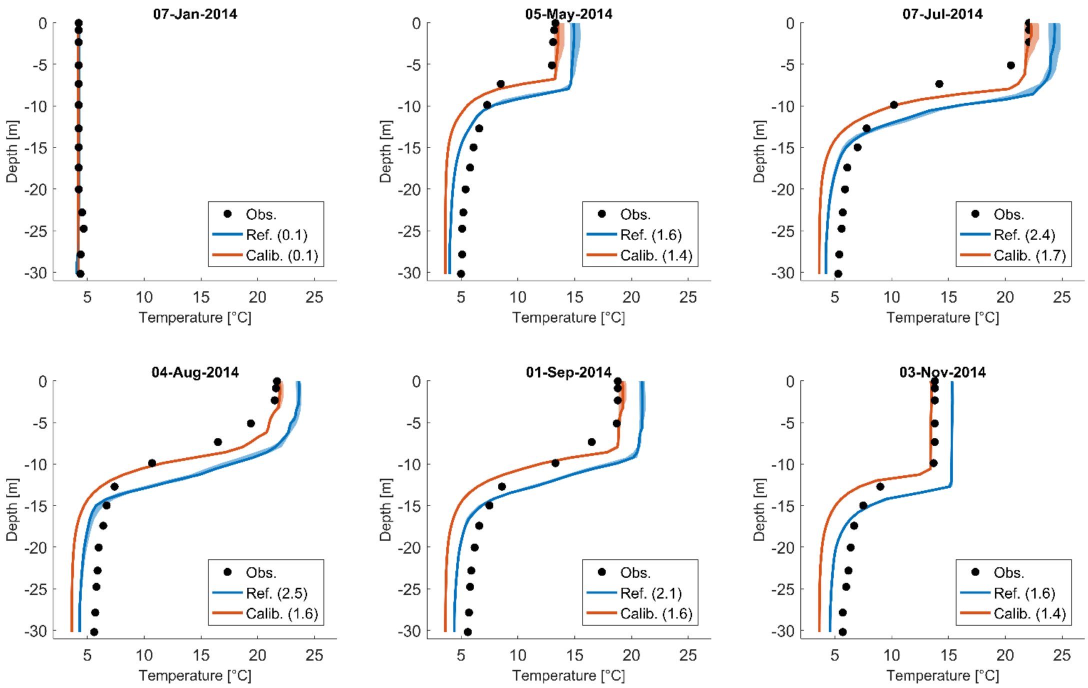


Figure S2.2: Temperature profiles at specific times at the central station (Kanton) with MAEs. Shaded areas represent the modelled temperature variability over over ±12 hours.

C. Greifensee time-series

Figures S2.3 and S2.4 provide a different perspective on the results by showing temperature time-series at various depth for both stations. The improvements are overall substantial, with a MAE reduction of up to 5 °C at the thermocline level (9.2 m depth). Additionally, most of the observations collected in the thermocline are within the modelled temperature variability over the internal waves period, which highest at the thermocline depth as has to be expected.

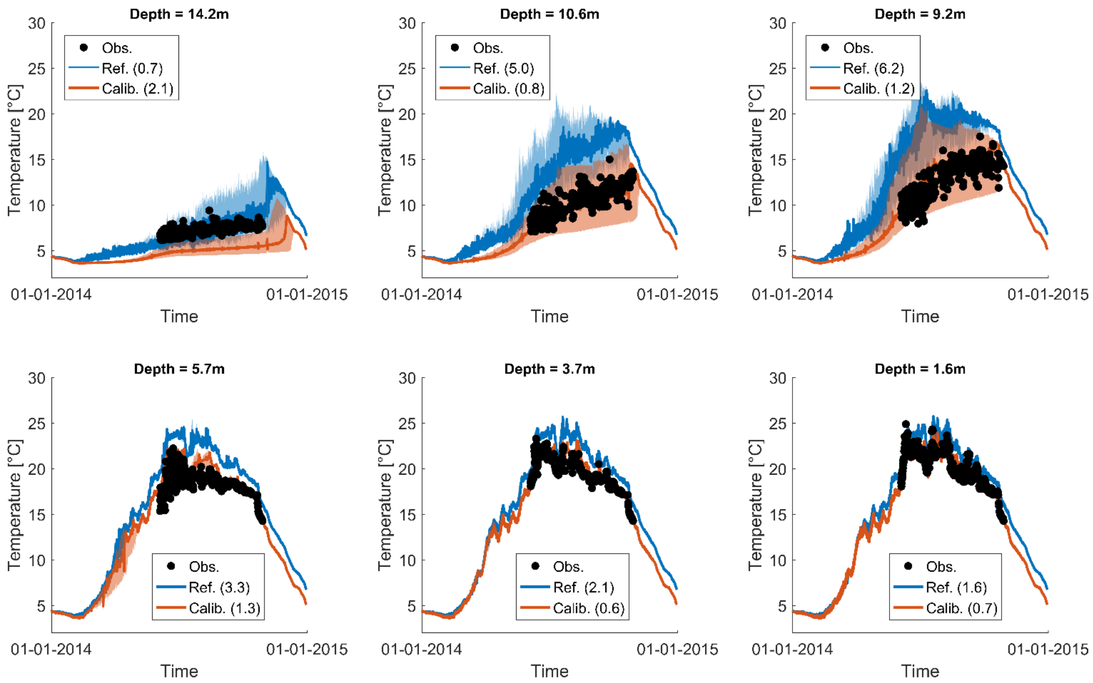


Figure S2.3: Time-series of temperature at various depth for the northern high-frequency station (Eawag) with MAEs. Shaded areas represent the modelled temperature variability over ±2 vertical layers.

The northern station (Eawag) benefited, once again, the most from the calibration. Bottom temperatures saw no improvements as they remain constant over the year while observations slightly increase. While this may hint at need for additional mixing or turbulence calibration, it is most likely a limitation of the method, discussed in the next section.

Finally, the model performance in the surface waters is greatly improved, reaching a MAE around half a degree. This is especially important considering the strong temporal variability in those shallow depths.

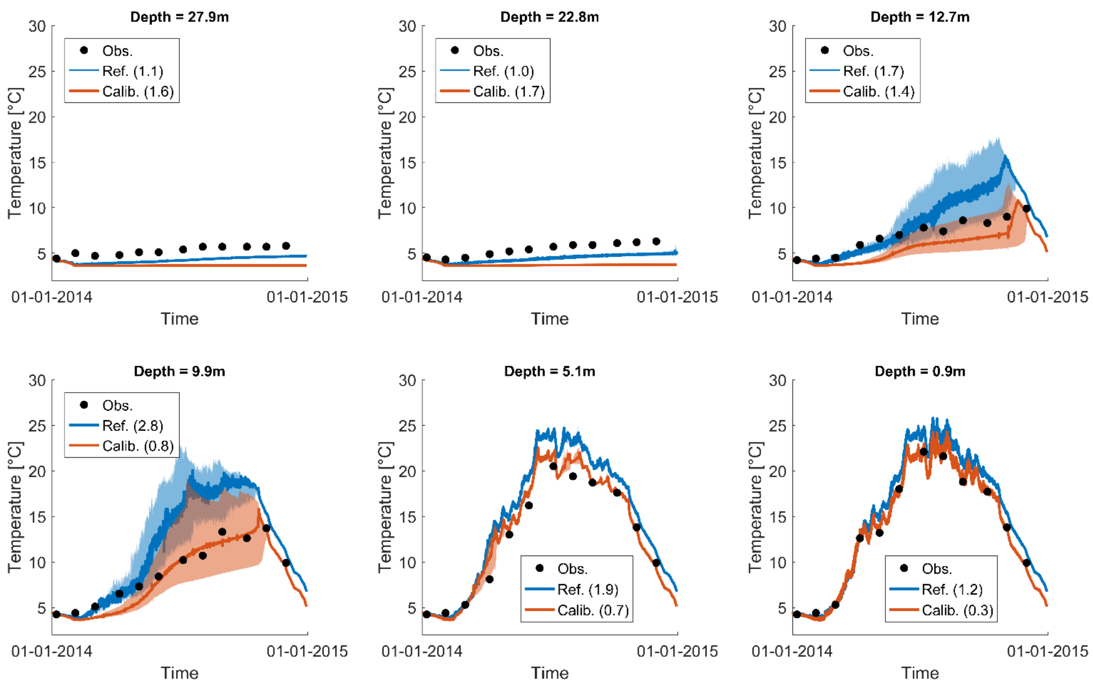


Figure S2.4: Time-series of temperature at various depth for the central station (Kanton). Shaded areas represent the modelled temperature variability over ±2 vertical layers.

D. Greifensee – discussion

The significant decrease in wind drag, which hit its constrained lower bound, combined with the increase in bottom drag and horizontal viscosity/diffusivity all seem to converge towards the same hypothesis: a need to decrease the energy content in the lake. This behaviour may be a reaction of the parameters to compensate for a problem in the forcing. Mainly an inaccurate and excessive wind intensity provided by the COSMO atmospheric model over Greifensee. This variable is known to be hard to model and is subject to inaccuracies. Indeed, better modelling results were obtained with a reduction in wind speed of up to 30%. Additionally, this may be the result of a limitation in the procedure, as only temperature data was used for the calibration. Further analysis of the currents, which reached lowed values, seemed to indicate that the wind drag has been reduced too much. ADCP data would be needed to properly assess such parameter and the overall dynamics of the system.

The Dalton number, a multiplicative coefficient influencing the evaporative heat flux, has a strong influence on the mixed layer temperature. Its calibration accounted for most of the improvements observed at shallow depths. This is especially visible during the second part of the year.

Finally, the low performance of the deep-water calibration is a consequence of the measurement model. The lower density of measurements near the bottom and below the thermocline resulted in a weak influence on the cost function in comparison to the numerous observations located in the mixed layer. While slightly less observations were available at the thermocline depth as well, their influence is high since the strong gradient in temperature can lead to large model-observation discrepancies, whose impact on the cost function is therefore stronger. The same stands horizontally: while still providing significant improvements at the central station (Kanton) with monthly measurements, the high-frequency observations of the northern station, and hence its higher shear number of data points, lead to a calibration more focused on that area.

Although this highlight a certain sensitivity to the measurement setup, those results show that the calibration algorithm is successful in matching model to data and that it is up to the user to define the location of focus by providing an adapted setup.

E. Additional results Lake Geneva

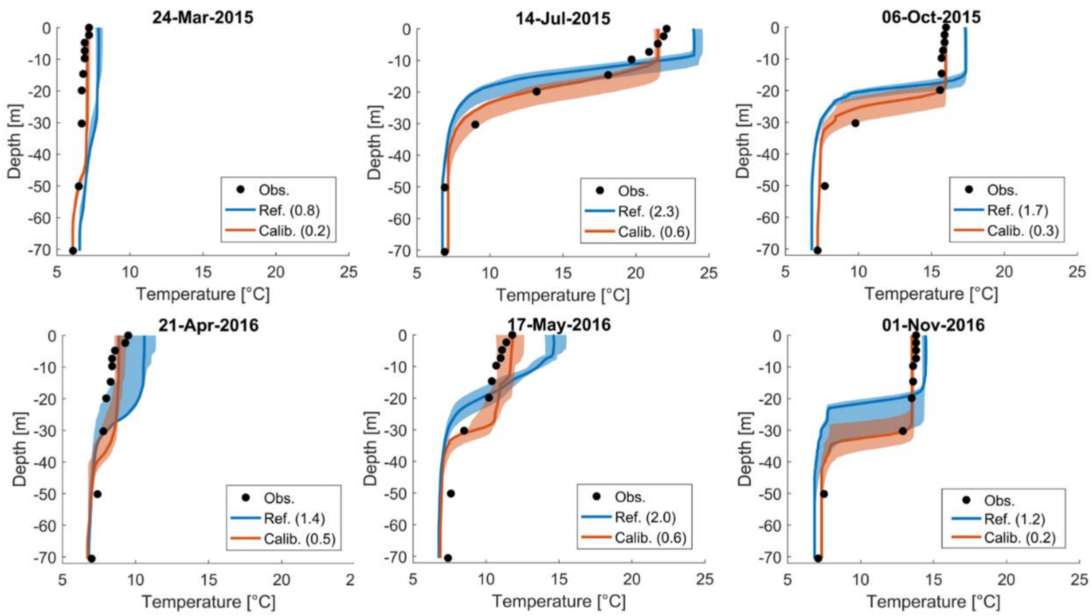


Figure S2.5: Temperature profiles at 10 a.m. at the small basin station (GE3) with MAEs. Shaded areas represent the modelled temperature variability over ±36 hours.

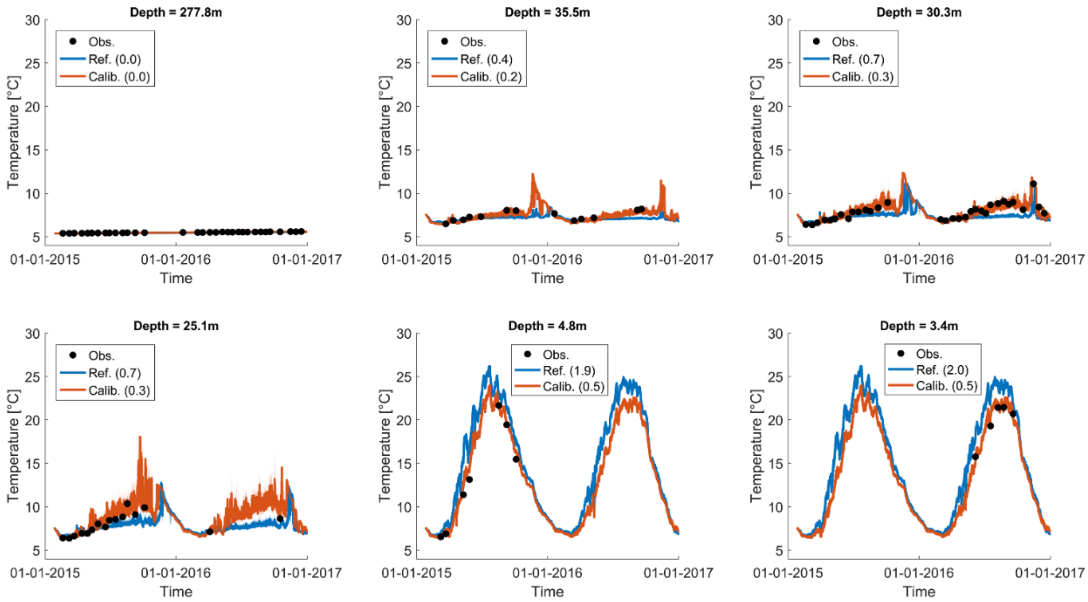


Figure S2.6: Time-series of temperature at various depth for the main basin station (SHL2). Shaded areas represent the modelled temperature variability over ± 2 vertical layers.

Authors contribution

TB, DB, AW designed the procedure and TB carried it out. MVL helped TB in the OpenDA interface implementation, AC collected most of the ADCP data and DB part of it. TB prepared the manuscript with contributions from all co-authors.

Acknowledgements

The authors would like to thank Stef Hummel (Deltares) for his help implementing the coupling between Delft3D-FLOW and OpenDA. This project was supported in part by the European Space Agency's Scientific Exploitation of Operational Missions element (CORESIM contract No.: AO/1-8216/15/I-SBo).

References

Afshar, A., Kazemi, H. and Saadatpour, M.: Particle swarm optimization for automatic calibration of large scale water quality model (CE-QUAL-W2): application to Karkheh Reservoir, Iran, Water Resources Management, 25(10), 2613–2632, doi:10.1007/s11269-011-9829-7, 2011.

Bard, Y.: Comparison of gradient methods for the solution of nonlinear parameter estimation problems, SIAM Journal on Numerical Analysis, 7(1), 157–186, doi:10.1137/0707011, 1970.

Bárdossy, A. and Singh, S. K.: Robust estimation of hydrological model parameters, Hydrology and Earth System Sciences, 12(6), 1273–1283, 2008.

Beaton, A. E. and Tukey, J. W.: The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data, Technometrics, 16(2), 147–185, doi:10.1080/00401706.1974.10489171, 1974.

Bohle-Carbonell, M.: Currents in Lake Geneva, Limnology and Oceanography, 31(6), 1255–1266, doi:10.4319/lo.1986.31.6.1255, 1986.

Bouffard, D. and Lemmin, U.: Kelvin waves in Lake Geneva, Journal of Great Lakes Research, 39(4), 637–645, doi:10.1016/j.jglr.2013.09.005, 2013.

Bouffard, D., Boegman, L. and Rao, Y. R.: Poincaré wave-induced mixing in a large lake, Limnology and Oceanography, 57(4), 1201–1216, doi:10.4319/lo.2012.57.4.1201, 2012.

Box, M. J.: A comparison of several current optimization methods, and the use of transformations in constrained problems, The Computer Journal, 9(1), 67–77, doi:10.1093/comjnl/9.1.67, 1966.

Bradley, E. L.: The equivalence of maximum likelihood and weighted least squares estimates in the exponential family, Journal of the American Statistical Association, 68(341), 199–200, doi:10.1080/01621459.1973.10481364, 1973.

Charnes, A., Frome, E. L. and Yu, P. L.: The equivalence of generalized least squares and maximum likelihood estimates in the exponential family, Journal of the American Statistical Association, 71(353), 169–171, doi:10.1080/01621459.1976.10481508, 1976.

Cimatoribus, A., Lemmin, U. and Barry, D. A.: Velocity and temperature data from Lake Geneva (CH), collected during year 2016, [online] Available from: https://doi.org/10.5281/zenodo.1109794, 2017.

Deltares: Delft3D-FLOW User Manual: Simulation of multi-dimensional hydrodynamic flows and transport phenomena, [online] Available from: https://oss.deltares.nl/documents/183920/185723/Delft3D-FLOW_User_Manual.pdf, 2015.

Deltares: The OpenDA Data-assimilation Toolbox, [online] Available from: www.openda.org, 2019.

El Serafy, G. Y., Gerritsen, H., Hummel, S., Weerts, A. H., Mynett, A. E. and Tanaka, M.: Application of data assimilation in portable operational forecasting systems—the DATools assimilation environment, Ocean Dynamics, 57(4–5), 485–499, doi:10.1007/s10236-007-0124-3, 2007.

Fabio, P., Aronica, G. T. and Apel, H.: Towards automatic calibration of 2-D flood propagation models, Hydrology and Earth System Sciences, 14(6), 911–924, doi:10.5194/hess-14-911-2010, 2010.

Fenicia, F., Savenije, H. H. G., Matgen, P. and Pfister, L.: A comparison of alternative multiobjective calibration strategies for hydrological modeling, Water Resources Research, 43(3), doi:10.1029/2006WR005098, 2007.

Garcia, M., Ramirez, I., Verlaan, M. and Castillo, J.: Application of a three-dimensional hydrodynamic model for San Quintin Bay, B.C., Mexico. Validation and calibration using OpenDA, Journal of Computational and Applied Mathematics, 273, 428–437, doi:10.1016/j.cam.2014.05.003, 2015.

Gaudard, A., Schwefel, R., Vinnå, L. R., Schmid, M., Wüest, A. and Bouffard, D.: Optimizing the parameterization of deep mixing and internal seiches in one-dimensional hydrodynamic models: a case study with Simstrat v1.3, Geoscientific Model Development, 10(9), 3411–3423, doi:10.5194/gmd-10-3411-2017, 2017.

Hendrickson, J. D., Sorooshian, S. and Brazil, L. E.: Comparison of Newton-type and direct search algorithms for calibration of conceptual rainfall-runoff models, Water Resources Research, 24(5), 691–700, doi:10.1029/WR024i005p00691, 1988.

Hering, D., Carvalho, L., Argillier, C., Beklioglu, M., Borja, A., Cardoso, A. C., Duel, H., Ferreira, T., Globevnik, L., Hanganu, J., Hellsten, S., Jeppesen, E., Kodeš, V., Solheim, A. L., Nõges, T., Ormerod, S., Panagopoulos, Y., Schmutz, S., Venohr, M. and Birk, S.: Managing aquatic ecosystems and water resources under multiple stress — An introduction to the MARS project, Science of The Total Environment, 503–504, 10–21, doi:10.1016/j.scitotenv.2014.06.106, 2015.

Imberger, J. and Hamblin, P. F.: Dynamics of lakes, reservoirs, and cooling ponds, Annual Review of Fluid Mechanics, 14, 153–187, 1982.

Johnston, P. R. and Pilgrim, D. H.: Parameter optimization for watershed models, Water Resources Research, 12(3), 477–486, doi:10.1029/WR012i003p00477, 1976.

Kiefer, I., Odermatt, D., Anneville, O., Wüest, A. and Bouffard, D.: Application of remote sensing for the optimization of in-situ sampling for monitoring of phytoplankton abundance in a large lake, Science of The Total Environment, 527–528, 493–506, doi:10.1016/j.scitotenv.2015.05.011, 2015.

Kurniawan, A., Ooi, S. K., Gerritsen, H. and Twigt, D.: Calibrating the regional tidal prediction of the Singapore Regional Model using OpenDA, in Proc. 9th Int. Conf on Hydroinformatics, Tianjin, pp. 7–11., 2010.

Kurniawan, A., Ooi, S. K., Hummel, S. and Gerritsen, H.: Sensitivity analysis of the tidal representation in Singapore Regional Waters in a data assimilation environment, Ocean Dynamics, 61(8), 1121–1136, doi:10.1007/s10236-011-0415-6, 2011.

Lemmin, U., Mortimer, C. H. and Bäuerle, E.: Internal seiche dynamics in Lake Geneva, Limnology and Oceanography, 50(1), 207–216, doi:10.4319/lo.2005.50.1.0207, 2005.

MacIntyre, S. and Melack, J. M.: Vertical and horizontal transport in lakes: linking littoral, benthic, and pelagic habitats, Journal of the North American Benthological Society, 14(4), 599–615, doi:10.2307/1467544, 1995.

Madsen, H.: Automatic calibration of a conceptual rainfall–runoff model using multiple objectives, Journal of hydrology, 235(3), 276–288, 2000.

Madsen, H.: Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives, Advances in water resources, 26(2), 205–216, 2003.

Nelder, J. A. and Baker, R. J.: Generalized linear models, in Encyclopedia of Statistical Sciences, edited by S. Kotz, C. B. Read, N. Balakrishnan, B. Vidakovic, and N. L. Johnson, John Wiley & Sons, Inc., Hoboken, NJ, USA., 2006.

Ralston, M. L. and Jennrich, R. I.: Dud, a derivative-free algorithm for nonlinear least squares, Technometrics, 20(1), 7, doi:10.2307/1268154, 1978.

Shimoda, Y., Azim, M. E., Perhar, G., Ramin, M., Kenney, M. A., Sadraddini, S., Gudimov, A. and Arhonditsis, G. B.: Our current understanding of lake ecosystem response to climate change: What have we really learned from the north temperate deep lakes?, Journal of Great Lakes Research, 37(1), 173–193, doi:10.1016/j.jglr.2010.10.004, 2011.

Skahill, B. E. and Doherty, J.: Efficient accommodation of local minima in watershed model calibration, Journal of Hydrology, 329(1–2), 122–139, doi:10.1016/j.jhydrol.2006.02.005, 2006.

Solomatine, D. P., Dibike, Y. B. and Kukuric, N.: Automatic calibration of groundwater models using global optimization techniques, Hydrological Sciences Journal, 44(6), 879–894, doi:10.1080/02626669909492287, 1999.

Sorooshian, S., Duan, Q. and Gupta, V. K.: Calibration of rainfall-runoff models: Application of global optimization to the Sacramento Soil Moisture Accounting Model, Water Resources Research, 29(4), 1185–1194, doi:10.1029/92WR02617, 1993.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, Journal of Geophysical Research: Atmospheres, 106(D7), 7183–7192, doi:10.1029/2000JD900719, 2001.

Toffolon, M. and Rizzi, G.: Effects of spatial wind inhomogeneity and turbulence anisotropy on circulation in an elongated basin: A simplified analytical solution, Advances in Water Resources, 32(10), 1554–1566, doi:10.1016/j.advwatres.2009.08.001, 2009.

Umlauf, L. and Lemmin, U.: Interbasin exchange and mixing in the hypolimnion of a large lake: The role of long internal waves, Limnology and Oceanography, 50(5), 1601–1611, doi:10.4319/lo.2005.50.5.1601, 2005.

van Velzen, N. and Verlaan, M.: COSTA a problem solving environment for data assimilation applied for hydrodynamical modelling, Meteorologische Zeitschrift, 16(6), 777–793, doi:10.1127/0941-2948/2007/0241, 2007.

Vörösmarty, C. J., Hoekstra, A. Y., Bunn, S. E., Conway, D. and Gupta, J.: Fresh water goes global, Science, 349(6247), 478–479, doi:10.1126/science.aac6009, 2015.

Wagner, C. and Adrian, R.: Exploring lake ecosystems: hierarchy responses to long-term change?, Global Change Biology, 15(5), 1104–1115, doi:10.1111/j.1365-2486.2008.01833.x, 2009.

Weerts, A. H., El Serafy, G. Y., Hummel, S., Dhondia, J. and Gerritsen, H.: Application of generic data assimilation tools (DATools) for flood forecasting purposes, Computers & Geosciences, 36(4), 453–463, doi:10.1016/j.cageo.2009.07.009, 2010.

Wüest, A. and Lorke, A.: Small-scale hydrodynamics in lakes, Annual Review of Fluid Mechanics, 35(1), 373–412, doi:10.1146/annurev.fluid.35.101101.161220, 2003.

III

Chapter 3
Data assimilation of in-situ and satellite remote sensing data to 3D hydrodynamic lake models

Theo Baracchini[1], Philip Y. Chu[2], Jonas Šukys[3], Gian Lieberherr[4], Stefan Wunderle[4], Alfred Wüest[1, 5], and Damien Bouffard[5]

[1] *Physics of Aquatic Systems Laboratory (APHYS) – Margaretha Kamprad Chair, ENAC, EPFL, Lausanne, 1015, Switzerland.*

[2] *Great Lakes Environmental Research Laboratory, NOAA, Ann Arbor, MI 48108, USA.*

[3] *Eawag, Swiss Federal Institute of Aquatic Science and Technology, Systems Analysis, Integrated Assessment and Modelling, Dübendorf, 8600, Switzerland.*

[4] *Oeschger Centre for Climate Change Research, Institute of Geography, University of Bern, Bern, 3012, Switzerland.*

[5] *Eawag, Swiss Federal Institute of Aquatic Science and Technology, Surface Waters – Research and Management, Kastanienbaum, 6047, Switzerland.*

Abstract

The understanding of lakes physical dynamics is crucial to provide scientifically credible information for ecosystem management. We show how the combination of in-situ data, remote sensing observations and three-dimensional hydrodynamic numerical simulations is capable of delivering various spatio-temporal scales involved in lakes dynamics. This combination is achieved through data assimilation (DA) and uncertainty quantification. In this study, we present a flexible framework for DA into lakes three-dimensional hydrodynamic models. Using an Ensemble Kalman Filter, our approach accounts for model and observational uncertainties. We demonstrate the framework by assimilating in-situ and satellite remote sensing temperature data into a three-dimensional hydrodynamic model of Lake Geneva. Results show that DA effectively improves model performance over a broad range of spatio-temporal scales and physical processes. Overall, temperature errors have been reduced by 54 %. With a localization scheme, an ensemble size of 20 members is found to be sufficient to derive covariance matrices leading to satisfactory results. The entire framework has been developed for the constraints of operational systems and near real-time operations (e.g. integration into *meteolakes.ch*).

3.1 Introduction

The management of aquatic systems is a complex challenge including many stakeholders pursuing sometimes contradictory objectives. This becomes even more complex in view of climate change, affecting both heat and hydrology in the watershed of lakes. There is thereby an urgent need to provide accurate information of the lake hydrodynamics.

Traditionally, perhaps due to the misleading definition of lakes as lentic systems, hydrodynamics studies focused on the one-dimensional vertical structure of lakes using in-situ measurements with limited spatial and temporal coverage (Kiefer et al., 2015). Yet, the lentic definition of lakes is misleading at short time scale. Dynamical processes such as wind-induced upwellings, rivers intrusion and gyres strongly disrupt the spatial homogeneity and quiet nature of the systems and ultimately affect lakes biogeochemistry (MacIntyre and Melack, 1995). Remote sensing observations, as well as one- and three-dimensional hydrodynamic models, addressed some of the spatial and temporal coverage limitations.

While three-dimensional hydrodynamic models are the only source of information capable of simulating the multi-scale temporal and spatial 3D lake dynamics, measurements remained essential to properly calibrate models and improve their representativeness. Indeed, model deviations are unavoidable, due to uncertainties in processes, forcing and observations (Lahoz et al., 2010), which have to be taken into account. Remotely sensed observations provide another essential source of information with continuously improving spatial and temporal resolution. Yet, this information remains fundamentally 2D. Ultimately, only the combination of remote sensing, numerical simulations and in-situ measurements can overcome the large variety of spatio-temporal scales involved in lakes dynamics and hence provide an adequate understanding of the system. This combination is achieved by data assimilation (DA).

DA is an effective approach to blend observational data into model simulations (Bannister, 2017; Li et al., 2008). Defined as the process by which the model of an evolving system is corrected by incorporating observations of the real system, DA improves both short-term forecasts and past model reanalysis (Hawley et al., 2006). A fundamental property of DA is to take observation (e.g. instrument accuracy, representativeness) and model (e.g. in processes, forcing, initial conditions) errors into account (Lahoz et al., 2010) and to provide the analysis with corrected errors (Kourzeneva, 2014). Those are crucial elements for parameter inference, monitoring and forecast reliability.

Multiple methods have been proposed for DA, among those, the Ensemble Kalman Filter (EnKF, Evensen, 2003). The EnKF has been successfully applied to numerous problems with non-linear dynamics, mainly in oceanography and atmospheric sciences (Eknes and Evensen, 2002; Evensen, 1994; Mao et al., 2009; Natvik and Evensen, 2003). It was found to be an efficient tool for non-linear problems with high dimensionality (Crow, 2003; Reichle et al., 2002b, 2002a), computing system error statistics based on system dynamics. But those methods have rarely been applied to lakes and DA for inland waters is still in its infancy. The different scales involved, and therefore observations available, with the large heterogeneity found in lake dynamics limited the direct application of experiments designed for oceans. For instance, Zhang et al. (2007) assimilated current measurements into a two-dimensional circulation model of Lake Michigan, where current updates are calculated by kriging interpolation. Yeates et al. (2008) used a pycnocline filter that assimilated thermistor data into a three-dimensional model of a stratified lake, to negate numerical diffusion driving model predictions off-course. Stroud et al. (2009) assimilated satellite images into a two-dimensional sediment transport model of Lake Michigan, using direct insertion and a kriging-based approach, effectively reducing model forecast errors. Later on they used an EnKF and Smoother (Stroud et al., 2010) using similar data and model when a large sediment plume was observed after a major storm event. Results obtained were better relative to standard approaches (a static model, and a reduced rank square-root Kalman filter). Finally, Kourzeneva (2014) used an Extended Kalman Filter to assimilate lake surface temperature into a one-dimensional two-layers freshwater lake model, leading to massive improvements over the free model run. To our knowledge, this is the first DA experiment that blends massive datasets of both in-situ and remote sensing observations into a three-dimensional hydrodynamic model with high dimensionality.

The aim of this study is to provide a flexible framework, in a Bayesian inference setting, capable of updating and improving model states while taking into account the uncertainty of both the modelled system and observational data. Here, we present a novel DA experiment with EnKF tailored to lakes and observations using an open-source hydrodynamic model and assimilation platform. This approach uses a new file-based coupling recently developed for OpenDA and Delft3D-FLOW with z-layer support (Chapter 2). Delft3D-FLOW is an open-source multi-dimensional hydrodynamic simulation software with numerous successful applications in coastal, river, estuarine and lake domains. OpenDA is an open-source generic DA environment (El Serafy et al., 2007), used in various calibration and DA experiments (El Serafy et al., 2007; Weerts et al., 2010; Kurniawan et al., 2011), but not yet to 3D lakes hydrodynamic modelling with DA. Our methodology is tested on a large French-Swiss lake (Lake Geneva) with in-situ temperature measurements and Lake Water Surface Temperature (LSWT) retrieved from satellite data (AVHRR). The choice of testing a first DA of surface temperature on Lake Geneva was motivated by recent studies concluding that data from space-borne medium resolution radiometers specifically tailored to Lake Geneva (Oesch et al., 2005) could potentially be assimilated to numerical models (Oesch et al., 2008). Furthermore, Chapter 2 proposed a calibrated model and framework for Lake Geneva. This first step being an absolute requirement for DA. Here, LSWT and in-situ data are blended into such model, to expand its monitoring capabilities of physical phenomena. The latter is achieved by considering the stochasticity of the system and an EnKF algorithm to update model results. This procedure is expected to benefit both environmental research and operational monitoring and forecasting of mid-size and large lakes, with impacts on a broad diversity of societally important issues.

The study is articulated according to the following: Section 3.2, data and methods, describes the study site, model, tools and data used. This includes measurements retrieval and processing chain as well as the quantification of their uncertainty. Although part of the methods, the data assimilation algorithm and its configuration are provided in a different section (Section 3.3) due to their central role in the study. Noise generation, number of ensembles, and localization scheme are discussed in this section. Sections 3.4 and 3.5 consist of respectively the presentation and discussion of results. Finally, perspectives and conclusion are given in the final section.

3.2 Data and Methods

In this section we describe the various components used in the DA experiment, the challenges associated with high frequency and resolution measurements, modelling datasets, and their errors definition, which previously hampered the application of such systems.

3.2.1 Study Site

Lake Geneva (locally known as Le Léman) is a perialpine lake located between Switzerland and France (46.458 °N, 6.528 °E) at an altitude of 372 m (Figure 3.1). It is the largest freshwater lake of Western Europe (surface area and volume of respectively 580 km^2 and 89 km^3) with a retention time of 11.4 years. Due to relatively mild winter temperatures and its large depth of 309 m, complete deep convective mixing only occurs every 5 to 10 winters (Schwefel et al., 2016). The lake is composed of two parts: the large eastern basin (Grand Lac), with maximum depth of 309 m, mean depth of 160 m, mean width of 10 km in which gyres are frequently observed, and the Petit Lac, the narrow and shallow western basin (maximum depth of 70 m, mean width of 4.5 km). The centres of the two basins are some 30 km apart, which defines the cut-off distance of the EnKF (more details in Section 3.3). The surrounding topography is mountainous, mainly in the Southeast, hence affecting the wind circulation above the basin. Lake Geneva is mesotrophic, with strong variation in turbidity and light penetration depth over the year (ranging from 3.6 to 14 m).

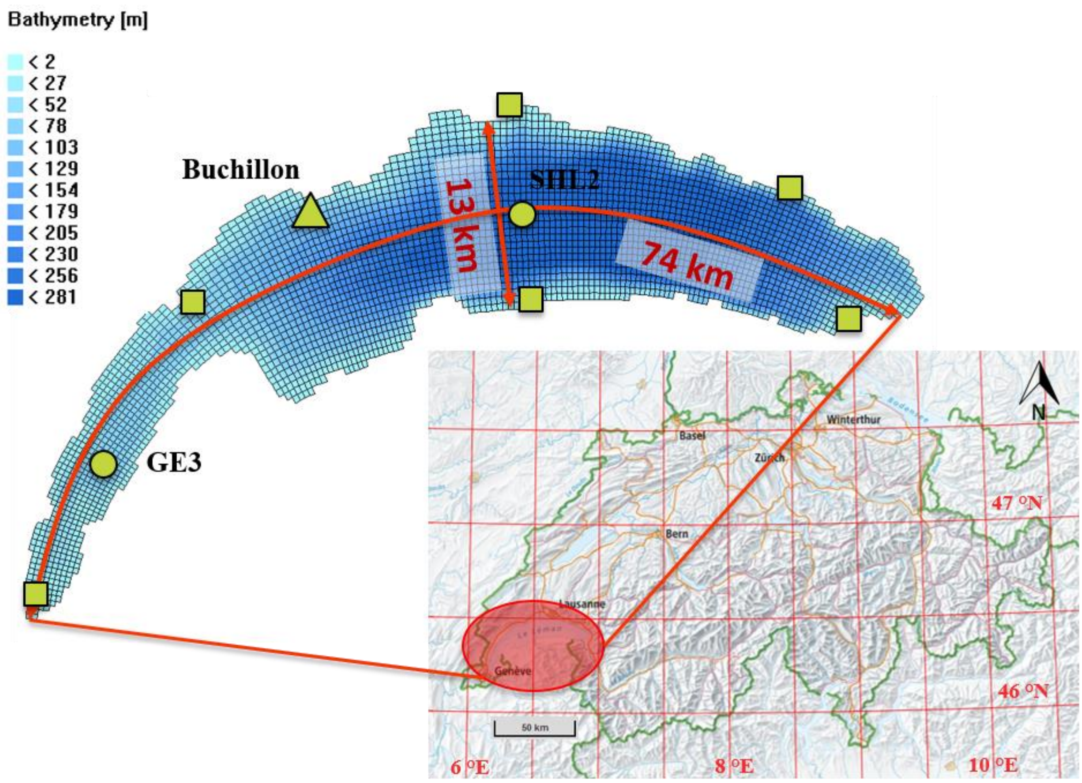


Figure 3.1: Lake Geneva location, computational grid and bathymetry. Circles are in-situ measurements. The triangle indicates the AVHRR validation station. Squares are selected sampling locations used to generate the wind fields of the COSMO-E products.

3.2.2 Model setup

The primary purpose of a hydrodynamic model is to solve various one- and three-dimensional (1D, 3D), time-dependent, non-linear differential equations related to hydrostatic free-surface flows in a computational grid. Various modelling suites have been developed to solve those equations accounting for momentum (Reynolds-averaged Navier-Stokes (RANS)), and fluid mass (continuity), as well as heat and mass transfer. The open source Delft3D-FLOW software is used in this study.

Delft3D-FLOW Numerical model – Delft3D-FLOW is an open-source hydrodynamic modelling suite developed by Deltares, Netherlands. Initially designed for coastal regions and estuaries, it has been expanded to rivers and lakes. A detailed model description of the equations and numerical schemes (conjugate gradient solver) can be found in the manual (Deltares, 2015).

We stress again that a fundamental prerequisite to any DA experiment is a calibrated model. Improper physical parameters could lead to strong discontinuities followed by waves (assimilation shocks) leading to spurious behaviours (Anderson et al., 2000). Assimilated variables could then, for example, go back to their pre-assimilated value. Lake Geneva's model has been extensively studied and calibrated (explicit optimization method by residuals minimization) in the previous chapter. This model consists of 100 unevenly distributed vertical layers, with thinner layers at the top (from 20 cm at the surface to several meters in the hypolimnion). Due to the steep bathymetry of Lake Geneva, we use the z-coordinate system (layers are horizontal) to avoid strong numerical diffusion and excessive artificial mixing. A computational time-step of 2 minutes is specified for the 450 m horizontal grid size to maintain model stability with the κ-ϵ turbulence closure model (Goudsmit et al., 2002). This turbulence closure model accounts for unresolved mixing at sub-grid scales. As initial conditions, the model is initialized (uniformly horizontally) from an in-situ temperature profile taken at the deepest location of the lake in January, when the lake is partially mixed. We consider a simulation period of one year, thereby covering the entire range of seasonal stratification dynamics.

The dynamics of a lake is mainly driven by interactions with the atmosphere and dissipation at the bed. As boundary forcing, we use MeteoSwiss COSMO-1 reanalysis products from their atmospheric model tailored to the Alpine region. They consist of various meteorological variables on a regular 1.1 km grid with hourly resolution. Seven of those variables are used in this study: solar radiation, wind direction and intensity, relative humidity, cloud cover, pressure, and air temperature.

Lake Geneva is subject to strong variations in turbidity which affect the stratification mainly in early summer. Monthly time-series of Secchi depth observations have therefore been used in the forcing.

On computation requirement, a single deterministic one-year model run for Lake Geneva without DA, requires up to 3 days of computing time on a single Xeon Broadwell core.

3.2.3 Assimilation platform

OpenDA is an open interface standard. It provides access to a set of open-source tools allowing the integration of arbitrary numerical models and observations through calibration and data assimilation algorithms. Its goal is to minimize algorithmic development by promoting the exchange of software solutions among researchers and users (Deltares, 2019, www.openda.org).

An OpenDA interface has recently been developed for the z-layer Delft3D-FLOW using the black-box wrapper (file-based) approach (Chapter 2). This interface has been further expanded for DA in this study. Additions include extended modifications of the Delft3D-FLOW model-definition file, model forcing files (on an equidistant grid only) for OpenDA's noise models, and a support for localization, which allows to limit the area of influence of an observation. The entire source-code is available on GitHub (https://github.com/OpenDA-Association/OpenDA).

3.2.4 Monitoring data

Role of data accuracy – Key in any DA problem is the observational data and its quality (Madsen, 2003). 3D-models require an especially large amount of data to assess and validate their horizontal variability. Remote sensing observations are therefore considered together with vertical in-situ profiles to constrain the system over the surface and depth. Errors are present in the system through its initial conditions, physical processes, approximations, and forcing (Bárdossy and Singh, 2008). Observations of the true system also require quantifying their uncertainties, as measurements are always an imperfect and incomplete representation (Bertino et al., 2007). This is particularly important as it defines how reliable an observation is and therefore how the model states are corrected. Injecting data with incorrect measurements error distribution into a good model could depreciate its relevance to the point where assimilation estimates are worse than the non-assimilative solution or the observations. The opposite holds true and model forecast would still be unreliable.

Lake in-situ data – The dataset consists of 31 temperature profiles over the water column at two locations of Lake Geneva (GE3 and SHL2, Figure 3.1) sampled during year 2017. Profiles are collected at a monthly (GE3) to bi-monthly (SHL2) rhythm. Uncertainty of in-situ temperature profiles is defined as the maximum value of the instrument precision (0.1 °C) and temporal variability at the measurement location. Reasons for the latter are twofold: first, some in-situ profiles did not have their exact collection time recorded; second, this study does not focus on reproducing short-term dynamics such as basin-scale internal waves and thermocline oscillations. The standard deviation of preliminary modelling results is computed over a time window to account for this variability. The temporal variability window is defined by the period of basin-scale internal oscillations (48 hours). This procedure allows to limit physical discontinuities created by the EnKF updates from specific physical processes (i.e. internal waves) which are not the focus of this study.

The Buchillon station (Figure 3.1), consisting of a mast measuring various atmospheric and hydrodynamic properties in real-time, has been used for the validation of AVHRR data detailed below. Of relevance for this study is a thermistor located at 1 m water depth representing the bulk LSWT.

AVHRR LSWT – The space-borne Advanced Very High Resolution Radiometer (AVHRR) sensor has been selected for its high temporal (up to 10 overpasses per day) and moderate spatial (1 km) resolution. We consider it to be the right trade-off for lakes; between the high spatial but low temporal resolution of Landsat 8 (90 m every 2 weeks) and the low spatial but high temporal one of SEVIRI (3 km every 15 min). The access to the AVHRR data was facilitated by a direct downlink and processing chain from the University of Bern. We describe below, and in the appendix, how AVHRR can be used for DA in lakes.

The AVHRR LSWT retrieval process, with locally adapted Split window coefficients for Lake Geneva, is described in Lieberherr and Wunderle (2018) and Lieberherr et al. (2017). Only pixels with quality levels higher than 3 (Lieberherr and Wunderle, 2018) are considered for the next sections.

An extensive description of the filtering of the data is available in the appendix. Overall, out of the 3372 AVHRR images of Lake Geneva available for 2017, 124 satisfy the selection criteria (see appendix A). This data is relatively evenly spread from February to October, with a maximum frequency of 1 image per 24 hours. Very few images are available in January, November and December due to bad weather conditions or cloud cover. The average lake coverage of those images is of 51 %.

3.3 Data assimilation

A short summary for popular DA approaches, their benefits and limitations is provided in the introductory chapter (Chapter 1). We briefly mention the Extended Kalman Filter (EKF) and Particle Filter, which are popular sequential algorithms for DA. The EKF is a variant of the Kalman Filter for non-linear dynamics and consists in a linearization of the model in the neighbourhood of the current estimate of the state vector. For highly non-linear systems this can result in an improper estimation of the state vector or covariance matrices and can therefore lead to quick divergence and instability (Moradkhani et al., 2005; Nakamura et al., 2006). The Particle Filter can cope with non-linearities and obtain a full representation of the posterior distribution but its computational cost (i.e. high number of particles required) limited its use with three-dimensional hydrodynamic models (more details in Chapter 1). For its flexibility and affordable computational cost, we further focus on the EnKF.

3.3.1 Ensemble Kalman Filter

The EnKF is an attractive alternative for non-linear dynamics and systems with high dimensionality. Reichle et al. (2002a) found that the EnKF is more robust than the EKF while being more flexible to obtain system covariances, a core element of the DA problem (Bertino et al., 2007). Indeed, whereas careful estimation of covariances often required a lot of effort (De Lannoy et al., 2007b), in the EnKF they are derived dynamically from a small ensemble of model trajectories (and therefore take into account the physics of the model), which grasps the essential parts of the error structure (Reichle et al., 2002b). The EnKF only considers a sample of the state variable to represent the processes modelled. The covariance matrix becomes a sampled covariance matrix and predictive probability density functions of the state vectors are approximated by Monte-Carlo simulations (Nakamura et al., 2006). It non-linearly propagates a finite ensemble of model trajectories instead of using a linearized equation for the error covariance, no computation of derivatives is required. The EnKF still considers a linear correction procedure, and assumes Gaussian distributions of the random variables. When this is not the case, the filter still produces a variance minimizing solution, though not being the optimal estimate (Bertino et al., 2007).

We develop below the fundamentals behind the algorithm. We first define the true model state (corresponding to the actual physical state of the lake) vector *x* of a system at time *t* as x_t (in our case temperature for the entire 3D model grid), $\mathcal{M}$ a non-linear operator, η some process noise, and u a forcing vector (here meteorological forcing) for a time t. The state propagation equation yields:

$$x_t = \mathcal{M}_t(x_{t-1}, u_{t-1}) + \eta_{t-1} \tag{3.1}$$

In this study, the noise is added in the forcing term u. The noise term is then dropped in the notation of (3.2) as it is included in the forcing. The state space vector, noted $\hat{x}$, is an approximation (done by the hydrodynamic model DELFT3D-FLOW) of the true state *x*. The forecast state (the input information for DA at time *t*) is defined by $\hat{x}^f$ and the analysis state obtained after DA as $\hat{x}^a$. The model propagation equation now reads:

$$\hat{x}_t^f = \mathcal{M}_t(\hat{x}_{t-1}^a, u_{t-1}) \tag{3.2}$$

As we do not measure the true state of the system (*x*), the observation (*y*) equation is defined by the following, with *H* an operator relating the system state to the observation and *ε* some measurement noise:

$$y_t = H_t(x_t) + \varepsilon_t \tag{3.3}$$

With the observation prediction given by:

$$\hat{y}_t = H_t(\hat{x}_t^f) \tag{3.4}$$

Note that in our case, we directly observe what we compute (i.e. surface temperature and profiles at computed grid points), thereby in this study *H* is an identity matrix. The resulting data assimilation estimate of the state vector ($\hat{x}^a$), which will be used in the next cycle as restart condition, is given by:

$$\hat{x}_t^a = \hat{x}_t^f + K_t(y_t - \hat{y}_t) \tag{3.5}$$

That last equation (3.5) is a central concept of DA, it introduces the weighting factor *K*, also referred to as Kalman gain. The Kalman gain can be viewed as a balance of the model and observation uncertainties, together with the error correlation of all the elements of the state vector. It aims at minimizing the error covariance of the state estimate during the analysis time (eq. (3.5)). It is defined as:

$$K_t = P_t^f H_t^T \left(H_t P_t^f H_t^T + R_t\right)^{-1} \tag{3.6}$$

with R_t the measurement error covariance matrix (in this study we assume no cross correlation between observation errors, hence R_t is diagonal and determined from the uncertainty of the measurements (Section 3.2.4)) and P^f the *a priori* state error covariance matrix. Error covariance is a key component of DA. The EnKF is able to compute a time-varying covariance error based on the dynamics of the system. This is a critical property when considering variables with short decorrelation spatio-temporal scales (Kuragano and Kamachi, 2000). In addition to the probability density function of the state (when in the presence of process noise), covariances estimation is achieved considering ensembles members. For an ensemble of forecasts (*j = 1, …, N*), each subject to a disturbance (e.g. in model processes, forcing or initial conditions), *P* is obtained from:

$$P_t^f = \frac{1}{N-1}\sum_{j=1}^{N}(x_{t,j}^f - \bar{x}_t^f)(x_{t,j}^f - \bar{x}_t^f)^T \tag{3.7}$$

From (3.7) we can conclude that the error spreading pattern across the domain is indeed derived from the ensemble members in a systematic way. This is not the case for some variational methods such as 3D-VAR, where the statistics are considered isotropic with little variation over time. In the EnKF each ensemble member is then updated individually (based on (3.5)). The state average over the ensemble provides the *a posteriori* state estimate. Additionally, in contrary to the Extended (or traditional) Kalman Filter, there is no need to propagate the state covariance, nor to estimate the initial state covariance and model error covariance matrices. The EnKF only uses the first and second moments to construct the probability density functions, it cannot assure higher order statistics by opposition to the Particle Filter (Nakamura et al., 2006).

The EnKF is widely used for large systems with uncertain initial states and variants are still being developed to leverage its limitations (Hoel et al., 2016). Several authors (Bertino et al., 2007; Evensen, 1994; Verlaan and Heemink, 2001) found better performance for highly non-linear systems in comparison to the EKF. This approach can accommodate massive datasets, missing observations, can incorporate correlated non-linear and error measurement models. Moreover, the ensembles are computationally easily parallelizable. Models with high-dimensionality are well suited for this type of assimilation, which requires a relatively low number of ensemble members to produce stable and accurate results (detailed in the results and discussion sections). We used this algorithm for the results presented in this study.

3.3.2 System setup

The aim of this section is to detail the various properties of the EnKF and DA setup, that are specific to this study.

Stochasticity and noise – The performance of a DA experiment strongly depends on the characterization of uncertainties (van Velzen and Verlaan, 2007). The hydrodynamics are modelled with deterministic equations. Their initial conditions, in the case Lake Geneva, only play a limited role on basin-scale dynamics over long periods (months, years). Yet, boundary conditions, especially, the air-water heat and momentum budget, still contain a large uncertainty that decrease the performance of any theoretically perfectly calibrated model. To overcome this issue, we added stochasticity to the system through its forcing. More specifically the East (u-direction) and North (v-direction) components of the wind velocity. These variables, coming from MeteoSwiss COSMO-1 reanalysis products with DA, are known to be most inaccurate and influential boundary forcing over lakes.

The addition of stochasticity to the deterministic model is done with OpenDA's noise model, which adds spatio-temporally correlated noise to the wind fields. This noise model, distributing the noise based on correlation scales derived from a distance- dependent function decaying to 0, requires three quantities (for both the u- and v-directions): (i) the wind standard deviation, (ii) the wind spatial correlation scale, and (iii) the temporal correlation scale. They are obtained from an analysis of the COSMO-E (ensemble) products over the entire year 2017. COSMO-E probabilistic products are derived from a 21-ensemble forecast on a 2.2 km grid and contain information on the variability of the computed atmospheric variables. The wind standard deviation is hence obtained taking the mean COSMO-E standard deviation of every pixel over the lake for the studied period. The spatio-temporal correlation scales are obtained from computing the cross-correlations of six fictive stations around the lake, as shown by Figure 3.1. The cross-correlation of a station with itself provides the temporal correlation scale, while the cross-correlation among stations allows to determine the spatial correlation scale. Table 3.1 summarizes the noise-model parameters aforementioned.

Table 3.1: Summary of the noise model parameters.

Parameter	Description	Value (u-direction, v-direction)
$\sigma\ [m \cdot s^{-1}]$	Wind standard deviation	1.11, 1.10
$\rho_L\ [m]$	Wind spatial correlation	20'000, 30'000
$\rho_t\ [h]$	Wind temporal correlation	5.67, 6.67

State variables – OpenDA has recently been updated to support three Delft3D state variables (Chapter 2), namely water levels, temperatures, and flow velocities. In this study, only temperatures are updated by the EnKF.

Ensembles – The EnKF operates using a statistical sample of the state of the system. The ensemble size (*N*) is often determined heuristically and must be a balance between a good representation of the state space and acceptable computation time. The errors in the solution pdf will approach zero at the rate 1/sqrt(*N*) (Evensen, 2003). A preliminary study showed that a satisfying compromise is obtained with 20 ensemble members. The choice for a small number of ensembles is further motivated by future use for operational purposes. More details and an ensemble size assessment are presented it the results section.

Localization scheme – It has been mentioned before (Section 3.3.1) that the covariance matrix links every domain point with each other. Covariances are derived from the ensemble members. A limitation of a small ensemble size is possible spurious correlations (Evensen, 2009) resulting in artefacts over long distances from the observation location. In such cases, when model spatial extent is large (an observation usually only influences its near vicinity, it has limited influence for greater spatial extensions (Stanev et al., 2011)), a localization scheme has to be applied. Such scheme has therefore been implemented in OpenDA, which collaterally also aims at reducing computation cost of the analysis time. This localization allows to define a cut-off distance, based on a Gaspari-Kohn function (an isotropic distance-based function decaying to 0 at a defined cut-off value), to limit the area of influence of an observation. This function ensures a smooth transition between a full and non-update for better model stability. Effectively, this removes long-range spurious correlations by scaling the size of the observation covariance matrix.

In this study, a cut-off distance of 15 km is defined. This distance is based on the two in-situ stations spacing and the radius of their associated basin gyres (Petit Lac and Grand Lac). This is further motivated by the fact that such distance allows to cover the entire interior of the basin by an update of in-situ data. Due to the significant depth of the lake, dynamics at deeper locations are less variable, hence their correlations at longer distances are easier to estimate. Regarding the LSWT, as it is partly the result of surface heat fluxes, its spatial structure is also expected to be correlated, to some extent, at relatively large spatial scales. Finally, as a result of the coarse vertical resolution of the in-situ profiles, we didn't define a different vertical localization scheme than the horizontal one in the vertical direction.

3.4 Results

In this section, we present both quantitative and qualitative results of the DA experiment. As mentioned in Section 3.2.4, the DA run consisted of the assimilation of 128 AVHRR LSWT images and 31 in-situ profiles over the entire year of 2017. Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and a Taylor diagram (Taylor, 2001) are used as benchmark indicators. Direct model comparisons with satellite images and in-situ profiles are provided to visualize the benefits of the approach both for surface and deep water dynamics. Implications of the DA for physical phenomena are presented.

Table 3.2: Summary of the data assimilation performance (MAE and RMSE).

	Control run	DA run	**Improvement** [%]
MAE [°C]	1.49	0.60	**60**
RMSE [°C]	2.07	0.95	**54**

Table 3.2, providing MAEs and RMSEs before and after DA, indicates significant improvements over the baseline simulation. RMSE and MAE values are reduced by 54 and 60 %, respectively. The discrepancy between the two indicate some occasional large data-model mismatch, which affects the RMSE more heavily. The Taylor diagram (Figure 3.2), displays large improvements in centered Root Mean Square Difference (RMSD), correlation and standard deviation.

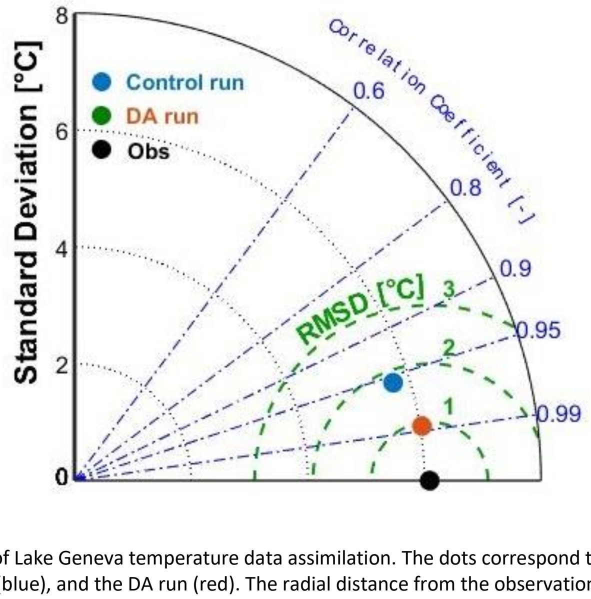


Figure 3.2: Taylor diagram of Lake Geneva temperature data assimilation. The dots correspond to the observations (black), the control run without DA (blue), and the DA run (red). The radial distance from the observations is the centred root mean squared difference, the radial distance from the origin defines the standard deviation, and the azimuthal position is the correlation coefficient.

Surface assimilation and physical processes – The benefit of DA is shown with four examples on Figure 3.3 comparing LSWT from AVHRR measurements, with LSWT from the control run model and DA experiment. We first highlight (top panels) that DA assimilation can perform correctly even in case of missing observations over the lake surface. The state covariance matrix could update the model in areas where no data was present. Model accuracy is thereby improved at basin-scale rather than at observation locations. This is particularly relevant as large lakes are often partly cloudy. The second example demonstrates the potential of DA to correct the state variable – a cold bias in the present case – while maintaining the coherent structure of the complex spatial thermal gradient (Figure 3.3, second row). The third example shows gyre-like flow structures. Such rotating structures are difficult to observe from AVHRR LSWT data (third row of Figure 3.3) partly due to its limited spatial resolution and its weak signature at the surface. However, a gyre created by an N/N-E wind on August 12th is better visible in the model results (clockwise in the western part of the main basin and counter-clock wise in the centre). In that case, the DA updated the LSWT while keeping the physical structure and flow spatial coherence of the control run. Finally, the lowest panels in Figure 3.3 show how DA improve observations and future quantification of transient upwelling. While the upwelling in the Petit-Lac was partially already caught by the control run, the DA allowed adjusting its intensity and extent. Another similar case is presented in appendix B.

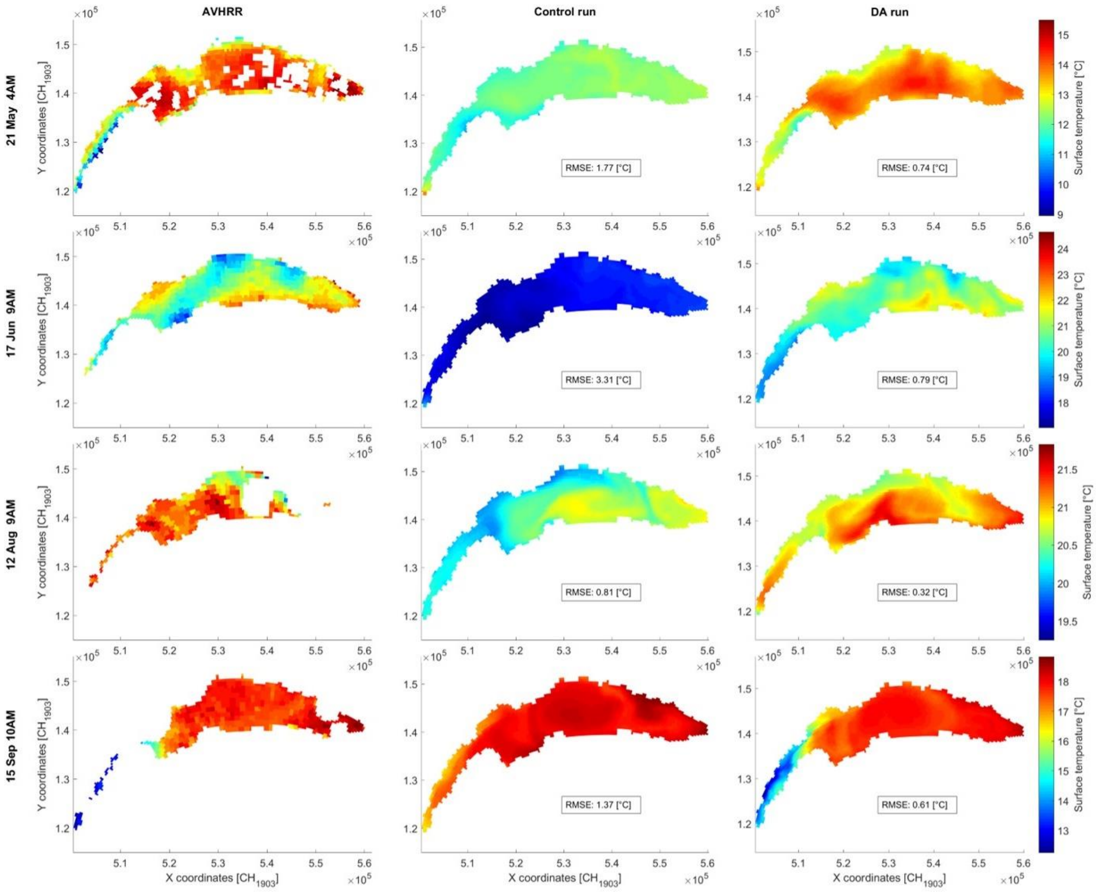


Figure 3.3: Surface temperature comparison of the AVHRR observations (left column), control run (central column), and DA run (right column) at selected analysis times (four rows) of 2017. The first row highlights the assimilation of sporadic data and the second row of complex surface patterns. The third row is an example for gyre phenomena, and the fourth row of an upwelling event.

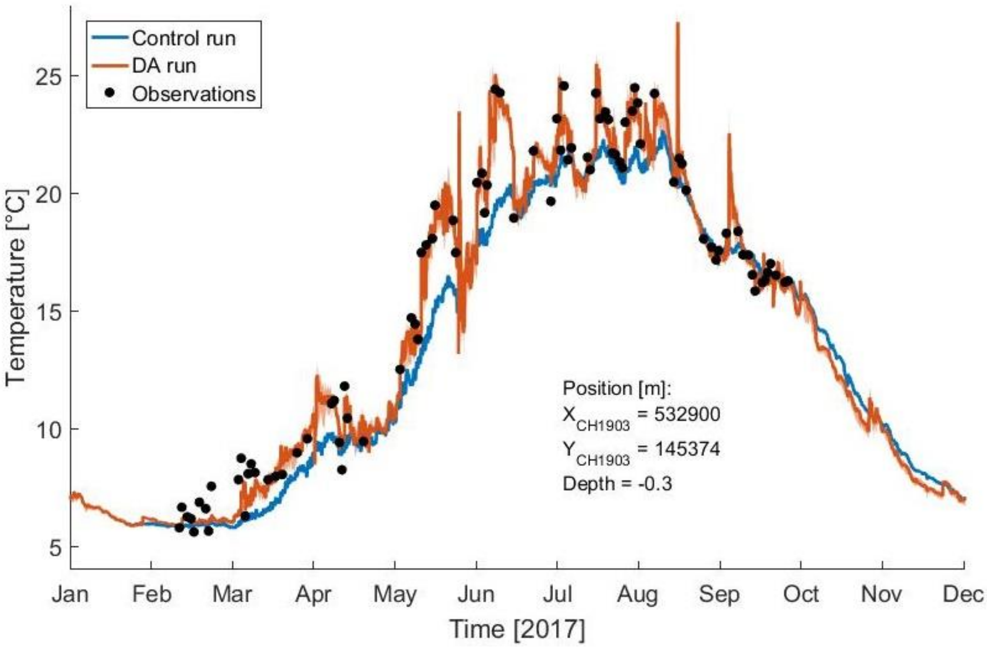


Figure 3.4: Time-series of the LSWT in the centre of the lake. The red line corresponds to the mean of the ensemble, and the red shaded area to the ensemble spread, while the blue line marks the control run, and the black dots the AVHRR observations.

The benefit from DA is also evident when looking at the temporal evolution of LSWT (Figures 3.4 and 3.5). In Figures 3.4 and 3.5, the AVHRR LSWT is again compared at two locations with the simulations with and without DA. The observed summer strong temporal variability with bi-weekly temperature variability exceeding 5 °C is not well resolved in the control run (Figure 3.4). It is however much better reproduced in the DA run. The warming phase also benefits significantly from the assimilation. The control run surface temperature started to increase in the second part of March, while the warming occurred early March in the observations and DA. Both models are in good agreement during the cooling phase from August to the end of the year. While few observations were available during this late period, not much improvements are obtained for the baseline, which was already accurate. Overall, every point of the DA run is close to or at least within the ~1 °C uncertainty of the AVHRR observations (see Section 3.2.4 for more information on data uncertainty).

Similar conclusion arises from Figure 3.5, which provides a close-up on time-series of the summer period in the western basin (Petit-Lac). Ensemble spread is smaller during the period of strongest stratification from late July to late August. Overall, the model uncertainty arising from perturbed wind fields reaches 2 °C in the summer, when it is the highest and 1 °C on average. Major upwellings in June and July are caught by both model runs, although the intensity is too weak in the control run. Again, data-model discrepancies and temperature variability are largest from late May to early August. Starting in August, the models with and without DA exhibit similar dynamics, both close to the observations.

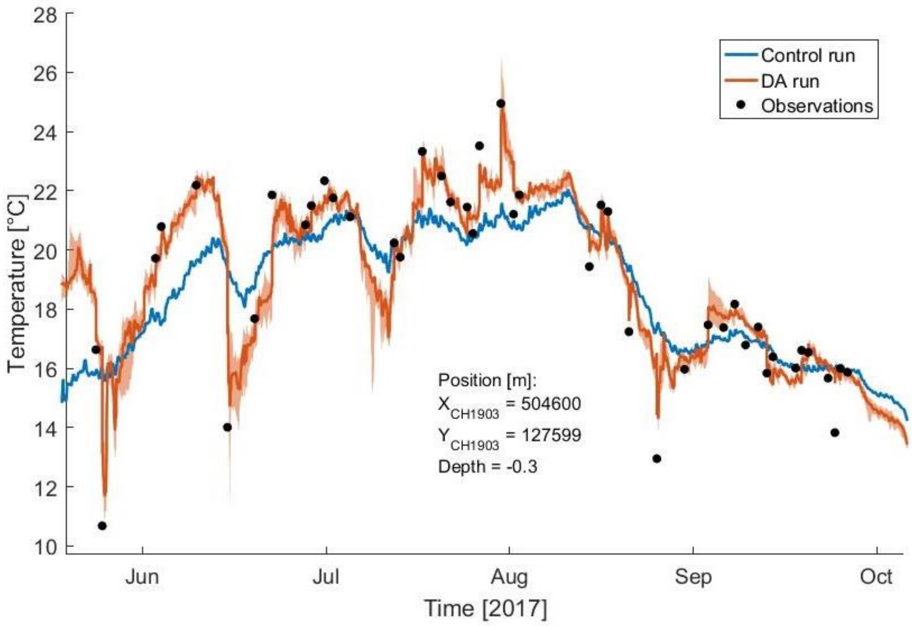


Figure 3.5: Zoomed time-series of LSWT in the centre of the western basin (Petit Lac). The red line corresponds to the mean of the ensemble, the red shaded area to the ensemble spread, the blue line to the control run, and the black dots to AVHRR observations.

We further compared the upwelling of September 15th with river temperature data with a model surface grid point located 3 km away (Figure 3.6). The upwelling has indeed been observed in the lake outflow, dropping from 21 °C to 12 °C in a matter of 6 days. The figure shows that the control run underestimated the upwelling by 5 °C, while the DA run underestimated it by ~2.5 °C. The AVHRR observation is 1.5 °C warmer than the river temperature. Figure 3.6 also allows to see that the model doesn't suffer from spurious behaviour after an assimilation. Model shocks are not observed and numerical equilibrium is reached in a sub-daily time frame.

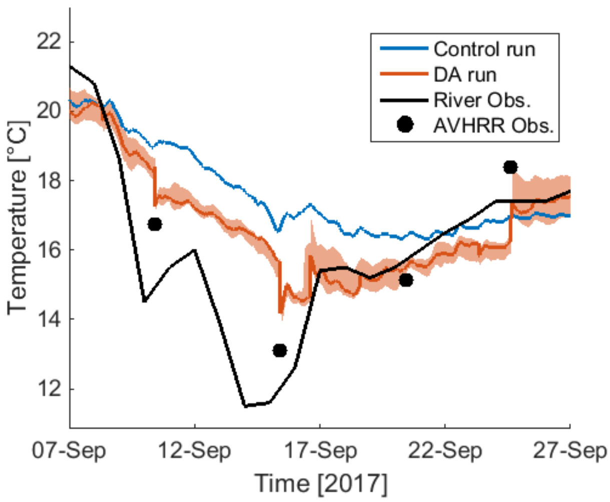


Figure 3.6: Close-up on the upwelling event of mid-September. River temperature from the lake outlet in Geneva is added as comparison. AVHRR data (black dots), control run (blue) and DA run (red) correspond to a surface pixel 3 km from the outflow.

Deep-water assimilation – We finally investigate how the vertical structure and sub-surface dynamics are affected by the DA. Figure 3.7 provides a comparison of the DA performance over depth with in-situ data instead of AVHRR measurements. Overall, for both stations significant improvements are obtained over the entire water column and throughout the year. Major improvements are observed at the thermocline depth, correctly represented in the DA experiment. Its strong vertical gradient significantly benefited from the assimilation of temperature profiles. The warm bias between 5 and 25 m depth, resulting in an overestimation of the mixed layer depth in the control run, is effectively eliminated.

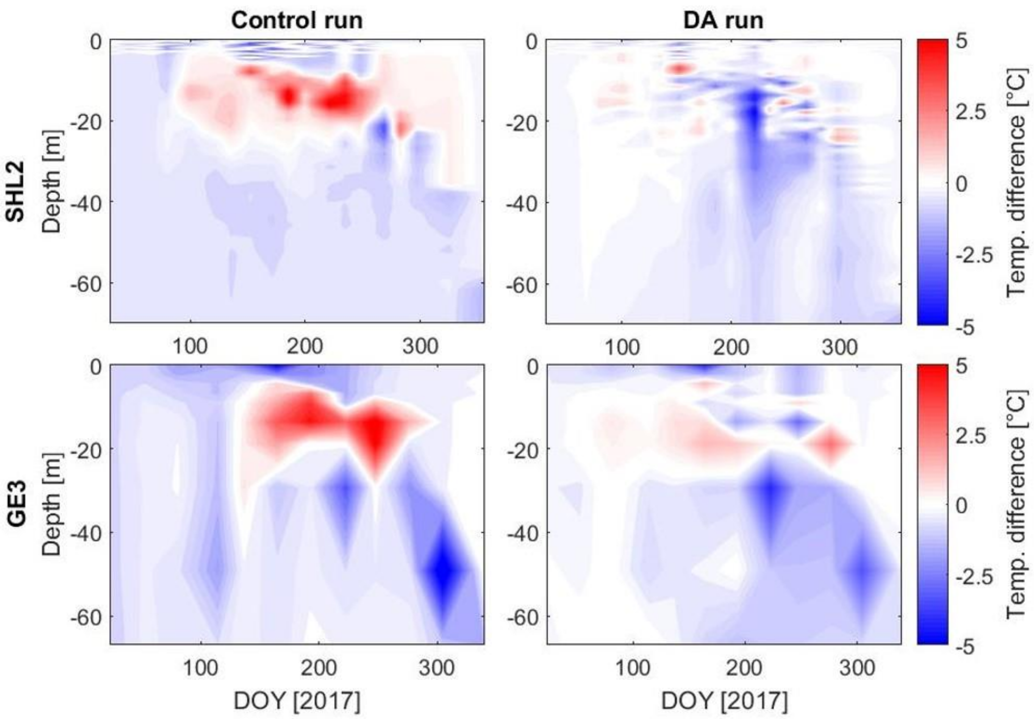


Figure 3.7: Evolution of the deep-water temperature without and with DA. The left column shows the differences of control runs minus in-situ observations; the right column shows the differences of DA runs minus in-situ observations. Upper row corresponds to the centre of the main basin (SHL2, Figure 3.1), and lower row to the Petit Lac (GE3, Figure 3.1).

Ensemble member size – Finally, we evaluated the EnKF ensemble size needed by a convergence analysis. A period of 1.5 month, from June to mid-July with a spin-up time of 2 weeks (without DA), is selected for assessment. This period of weak spring thermal stratification has been selected, as it is the time of the year with the most complex and broadest range of dynamics (Figure 3.4).

The results indicate that for an increasing number of ensembles (*N*) the analysis error decreases. Table 3.3 provides RMSEs and MAEs for different ensemble sizes. Figure 3.8 provides similar results while differentiating between assimilated data sources. We conclude that the major gains are achieved with 10 ensemble members. For in-situ data only, 20 ensembles seem to be the sweet-spot. Due to the much larger amount of AVHRR observations (i.e. one image provides thousands of observations since it covers the entire spatial extent of the computational grid), the red (AVHRR) and black (all measurements) lines are confounded. Finally, assessment of the ensemble spread showed that few gains in second-order moments were found with larger ensemble sizes. Indeed, in the scope of this study, the additional benefits for $N > 20$ are limited. At this stage, the 1 °C uncertainty of the RS images might become a limiting factor hindering further improvements.

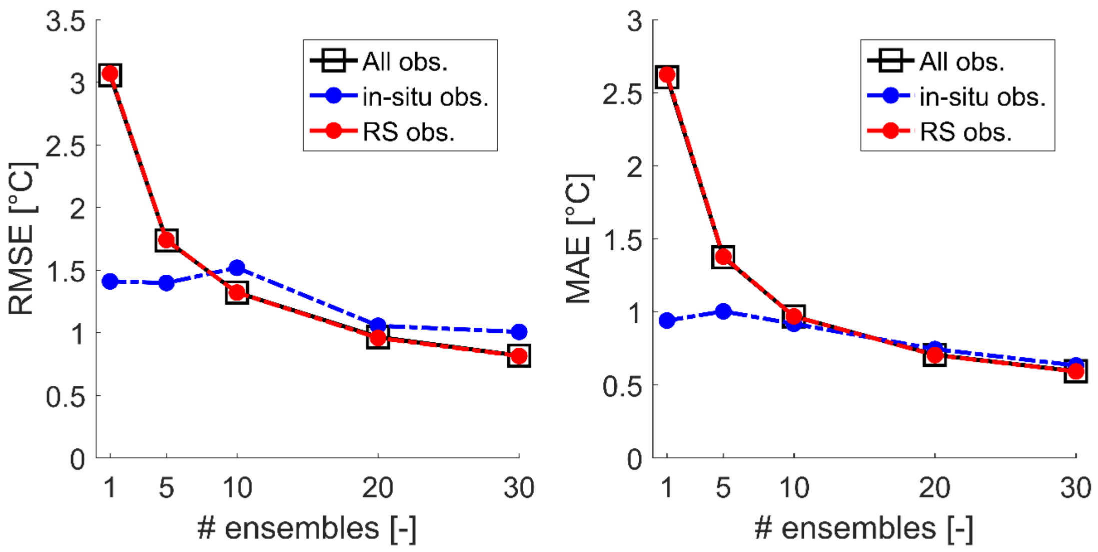


Figure 3.8: Data assimilation performance as a function of ensemble size. The dashed blue line corresponds to the error with respect to in-situ observations only, the red line the same with respect to LSWT only, and the black one the model error with respect to both observation sources.

With a view towards DA for operational lakes forecasting systems, and the computational constraints associated with real-time hydrodynamics, we conclude that 20 members provide a satisfactory compromise for the system considered in this study.

Table 3.3: DA performance (MAE and RMSE) for various ensemble sizes.

# ensembles	**Control run**	**5**	**10**	**20**	**30**
MAE [°C]	2.61	1.37	0.97	0.71	0.59
RMSE [°C]	3.06	1.74	1.32	0.96	0.82

3.5 Discussion

The DA framework has brought significant improvements to the hydrodynamics of Lake Geneva. It demonstrated its effectiveness to improve various model-forecasted meso- to large-scale thermal features. The combination of both in-situ measurements and remote sensing observations allowed constraining the 3D thermal structure of the model throughout the water column.

Surface time-series (Figure 3.4) indicated that spring / early summer observations play a key role in improving the model performance during the warming period (Kourzeneva, 2014). This allows for an adequate modelling of the lake warming, with significant implications expected for water quality models and the typical spring phyto-plankton blooms. Later in the year, late-spring and summer, the AVHRR data revealed high variability tempera-ture dynamics (e.g. upwellings) which are not reproduced by the control run. It is the time when the largest ensemble spread is observed, which indicates that the summer LSWT is sensitive to changes in wind patterns. Additionally, ensemble spread stemming from spatio-temporally correlated noise applied to the wind fields indi-cate that the model is sensitive to the changes in this forcing function. The effects on model outputs are correctly described and the uncertainty arising from this perturbation ranged from 1 °C on average with peak values at 2 °C.

Similar conclusion can be drawn for sub-surface thermal dynamics. Figure 3.7 indicates that data-model mis-matches in the mixed layer appeared as the lake started to warm and the thermocline formed. Compared to the control run, the DA run exhibited a more accurate warming phase and vertical temperature gradient during the stratified period.

Overall, the performance of the EnKF has been notable in a broad range of scenarios. Figure 3.3 showed that even with complex observational patterns, filter updates were performed with different amplitudes at each spatial location. Those spatially varying updates often are in agreement with the physical processes governing the hydrodynamics of the lake. Also, in the case of incomplete and sporadic data, the EnKF updates behaved well and good combinations of data and system dynamics were found. Some authors (De Lannoy et al., 2007b) found that when the update is performed through the covariance propagation (in case of missing observation), the *a posteriori* state might not be correct and counteract the updates in the surrounding locations. This behaviour has not been observed in the presented hydrodynamics of Lake Geneva. This indicates that the covariance matrices were well estimated from the ensemble members and their physical dynamics. The non-static covariance matrix derived from the EnKF allows longer-term studies, such as over the entire year, with complex changes in the thermal structure of the waterbody. Time-varying covariance error estimates for 3D models is a complex task in DA. Analysis updates were not intense nor frequent enough to cause model shocks or solver failure. This would have a minimal impact on the surface layers, since such corrections would not be persistent due to the rather variable nature of surface layers and sensitivity to atmospheric forcing. However, more issues would arise from model shocks in the deep water, which could trigger movements of large water volumes. Since in-situ profiles have a much lower uncertainty than AVHRR observations (<0.1 °C vs 1 °C, Section 3.2.4), intense state updates are more likely, however they have not been observed in this study and no model solver failure arose from the EnKF updates. After significant updates, the model generally recovered in a sub-daily time frame. Increasing the observational frequency (here limited to one satellite observation per day) would increase the likelihood of encountering model shocks, as equilibrium adjustment may not be reached between updates. Higher computational cost also weights in the data quality/quantity compromise, particularly when considering near real-time systems.

Physical processes – Figure 3.3 showed that various physical processes, such as upwellings and gyres, are better resolved with the use of EnKF. Phenomena such as upwellings typically occur more prominent at the beginning or end of the season, when stratification is weaker. The better identification of such processes is of prime importance for various water quality aspects (e.g. heat extraction (Gaudard et al., 2018), wastewater discharge, water intakes, etc.). Yet the magnitude of such events has rarely been quantified, due to difficulties with their large-scale identification. Through the combination of remote sensing observations and three-dimensional hydrodynamic modelling, we open new possibilities for monitoring of such phenomena. In this study we found that upwellings are better reproduced both in intensity and spatial extent. Comparing temperature measurements with a surface model grid point 3 km away from the outflow showed good agreements after DA. An underestimation of the upwelling of 2.5 °C after DA is observed (compared to 5 °C with the control run). Most of this remaining difference can be attributed to the satellite underestimating the event as well (by 1.5 °C, with an uncertainty of 1 °C) and the remoteness and depth (surface) of the pixel compared. For Lake Geneva, this is of particular interest when an upwelling occurs in the western basin (Petit-Lac), dropping the outflow temperature for millions of downstream residents. In terms of gyres, those structures are repeatedly observed in Lake Geneva (Bouffard et al., 2018; Kiefer et al., 2015). Because of the Coriolis force, subsequent strong up-/down-lifts of the thermocline occur, which structure the lateral dispersion of primary productivity (Soomets et al., 2019).

Ensemble size – Among the various ensemble sizes, assessed for this study (Figure 3.8), we found that relatively small ensemble sizes (~20) are large enough to derive suitable time-varying covariances and error spreading patterns. This is particularly important in the presence of variables with short decorrelation time and spatial scales. Studies indicated that relatively small ensembles fail at accurately estimating the small correlation patterns of remote observations (Houtekamer and Mitchell, 2001). The localization scheme implemented (Section 3.3.2), defining a cut-off radius around each observation, allows to circumvent this limitation. Houtekamer and Mitchell (2001) found that for an increasing ensemble size, the optimal cut-off value increases as well. Larger ensemble sizes not only restrain the underestimation of ensemble spread and accuracy, but also allow the use of more remote observations. For DA experiments with limited data, larger ensemble sizes may be a requirement to maximize observational coverage.

Limitations and perspectives – A main limitation of the EnKF is the Gaussian assumption, which in the case of large data-model mismatches, could have led to artefacts and unrealistic *a posteriori* state values. This has not been observed in this analysis with the provided noise definition and observational stochastic setup. Furthermore, while we did not study systematically the physics after each analysis step, we think the method can still be used for the study of physical processes, provided the user assesses the intensity of those physical discontinuities. Out of the 152 assimilations, only 8 created some numerical instabilities in the model, though small enough to prevent solver failure.

Other difficulties arise in the presence of bias, where Kalman Filtering performs suboptimal corrections (Dee and Da Silva, 1998), as observations and model are assumed unbiased. Solutions for dealing with biases in EnKF may become necessary (De Lannoy et al., 2007a). In the present approach, however, occasionally occurring model biases have been effectively handled by the update. The DA model did not drift back to its biased or control run state. We believe that this is a result of the adequate initial parameterization of the model (Chapter 2). This further highlights the crucial importance of accurate model calibration and formulation before proceeding with DA experiments. It is worth noting that the EnKF is able to provide updates also to parameters and forcing conditions, which in some cases, may provide more persistent improvements (for example when time-varying parameters are needed).

This DA experiment is fairly time-consuming from a computational aspect. For instance, it took nearly one month to compute the present setup on a dual Intel Xeon E5-2697v4 processors machine with 256 GB of memory, generating close to a terabyte of data. While the analysis time for in-situ data has been reasonable (~1 hour), the immense amount of observations generated by an AVHRR image (entire coverage of the surface layer of the computational grid) brought the analysis time up to 3 hours for a single image. This is largely due to the current lack of multi-core support for the analysis step. A multi-core local analysis has been implemented in the scope of this study for multi-variables (e.g. temperature with water levels and or flow velocities) assimilations, but gains can further be obtained from a local analysis based on the observation localization scheme.

3.6 Conclusion

For managerial and scientific purposes, new monitoring and forecasting tools, covering wide ranges of spatio-temporal scales, are of great interest. The coverage of such scale breadth of inland waters is achieved by combining three information sources, namely (i) in-situ measurements, (ii) remote sensing observations and (iii) model simulations. With data assimilation (DA), optimal combinations can be achieved and valorised.

For several decades, DA has been an expanding field in oceanography and atmospheric sciences, yet its presence in limnology remained limited. In this study, we propose a flexible framework and tools to blend real data into model simulations tailored to lakes. We applied this method to Lake Geneva using large datasets consisting of a three-dimensional hydrodynamic model, AVHRR lake surface temperature and in-situ profiles over an entire year. Results demonstrated the effectiveness of DA as significant gains were obtained for both the surface and deep-water dynamics over a well-calibrated baseline. We showed that both data types (in-situ and remote sensing) are important to constrain the entire spatial extent (horizontal and vertical) of the model. Results also indicate that AVHRR data is a valid RS data source for DA into lake hydrodynamics, provided that observational error and uncertainties are well defined.

In that regard, the use of an Ensemble Kalman Filter (EnKF) allowed to handle non-static covariance estimation, a key element of any DA problem. Additionally, it is able to account for the uncertainties of each data source. Those are essential elements influencing DA performance (Qi et al., 2014b). We found that the ensemble size played an important role in reducing model errors. To keep their number limited, a localization scheme has been implemented, hence circumventing the estimation of improper small correlations at large distances (Houtekamer and Mitchell, 2001). In that regard, while the EnKF adds computational complexity to the problem, it is capable of estimating dynamically the stochastic model based on the physical properties of the system. This is well encompassed by the paradox defined by (Bertino et al., 2007), stating that simple DA methods become complicated engineering tasks, when the inconsistency between the stochastic and the physical model becomes relevant. Due to the flexibility of the tools developed and used, we expect this procedure to be transferable to other lakes and hydrodynamical models with relatively limited development.

To conclude, this method has been designed with the constraints and challenges of near real-time applications in view. Implications of DA in the operational context are significant to provide robust and timely short-term forecasts, accurate reanalysis products, and uncertainties for reliable water management. Over the last decades, the number of remote sensing products available grew rapidly, however they have hardly been used in the operational context in an optimal way (de Rosnay et al., 2013). The timely retrieval of RS products requires interdisciplinary efforts to ensure robustness and the proper error definition of the data, which hinders the development of such operational systems (van Velzen and Verlaan, 2007). In this study we provided an example of how the entire chain, from the satellite to the assimilation into the model, can be performed, with limited field infrastructures. More concretely, we expect the findings of this study to be directly applicable to existing lake forecasting platforms, such as the one for Lake Geneva (*meteolakes.ch*). Impacts of such implementation are expected at scientific, governmental and public level.

Code and data availability

Software – The source code and documentation of the numerical model (Delft3D-FLOW) and data assimilation platform (OpenDA) developed in and for this study can be accessed and downloaded on their online repositories at https://oss.deltares.nl/web/delft3d/source-code and https://github.com/OpenDA-Association/OpenDA.

Data – The authors are grateful to the following institutions that provided the data used in this paper: the Federal Office of Meteorology and Climatology (MeteoSwiss) for meteorological data, the Département de l'environnement, des transports et de l'agriculture (*DETA*) du Canton de *Genève* for in-situ data on Lake Geneva at GE3 and the Federal Office of the Environment (FOEN) for the river data temperature in the outlet of Lake Geneva. In-situ data at SHL2 as well as Secchi disk measurements in Lake Geneva were provided by the *Commission International pour la Protection des Eaux du Leman* (CIPEL) and the Information System of the *SOERE OLA* (http://si-ola.inra.fr), INRA, Thonon-les-Bains. This data cannot be published as it belongs to their aforementioned owners, it is not the property of the authors of this study. It can nonetheless be requested by contacting its respective institution. Any other data used in this study is property of the Physics of Aquatic Systems Laboratory at EPFL and can be obtained by contacting Prof. Alfred Johny Wüest (alfred.wueest@epfl.ch).

Appendix

A. AVHRR validation

Validation – AVHRR data were validated for Lake Geneva by comparing in-situ data from the Buchillon station to the remotely sensed-derived skin temperature. Analysis of the data and comparison with both radiometric and in-situ observations at Buchillon showed that quality flags are not a sufficient measure to reliably quantify the accuracy of the AVHRR images. Indeed, we observed strong fluctuations reaching up to ±3 °C between skin and bulk temperature especially during daytime where a micro-stratification establishes in the surface layer (Gentemann et al., 2003). Skin and bulk temperature becomes similar under windy or convective conditions. Skin to bulk corrections were developed in oceanography as a function of the wind intensity (Minnett et al., 2011). Yet, lakes are much more quiet environment and parameterization should also take into account convective processes (Bouffard and Wüest, 2019). Such parameterization is unfortunately lacking for Lake Geneva and we initially instead selected night-time/early morning images where surface convective cooling reduces the skin to bulk difference. However, comparison with field data showed that this is still not reliable enough. The discrepancy is indeed strongly linked to day-night cycles, but those are also season-dependent. Therefore, no specific satellite overpass can be selected for the entire computational time (one year). To ascertain that images portray an accurate representation of the lake bulk LSWT, the thermistor at 1 m depth (recording at a 1 hour interval) has been used for a direct comparison with the space-borne AVHRR data. Considering that the Buchillon station is close (80 m) to the shore, its position has been shifted 2 km south to avoid land boundary contamination. Finally, the average of a 3x3 pixel window of the satellite image centred on the south-shifted Buchillon station coordinates is used as comparison point with the field data. Only images with an absolute deviation with respect to the bulk water lower than 1 °C are retained for further assimilation. The 1 °C threshold will also define the AVHRR observational uncertainty needed for the EnKF. Outlier pixel values colder than 4 °C and warmer than 28 °C are removed. Finally, to avoid assimilating observations at a too high frequency (or too close in time), which can result in physical discontinuities and destruction of model processes (model not reaching equilibrium between assimilations), the maximum frequency of satellite images is limited to 1 per 24 hours. The screened images are then mapped to the computational grid.

This procedure aims at bypassing the skin to bulk temperature effect, while ensuring best data quality for assimilation. This procedure assumes horizontal uniformity over the lake area (i.e. atmospheric effects are assumed to be the same over the entire domain), and may be sensitive to local cloud patches.

B. Additional results

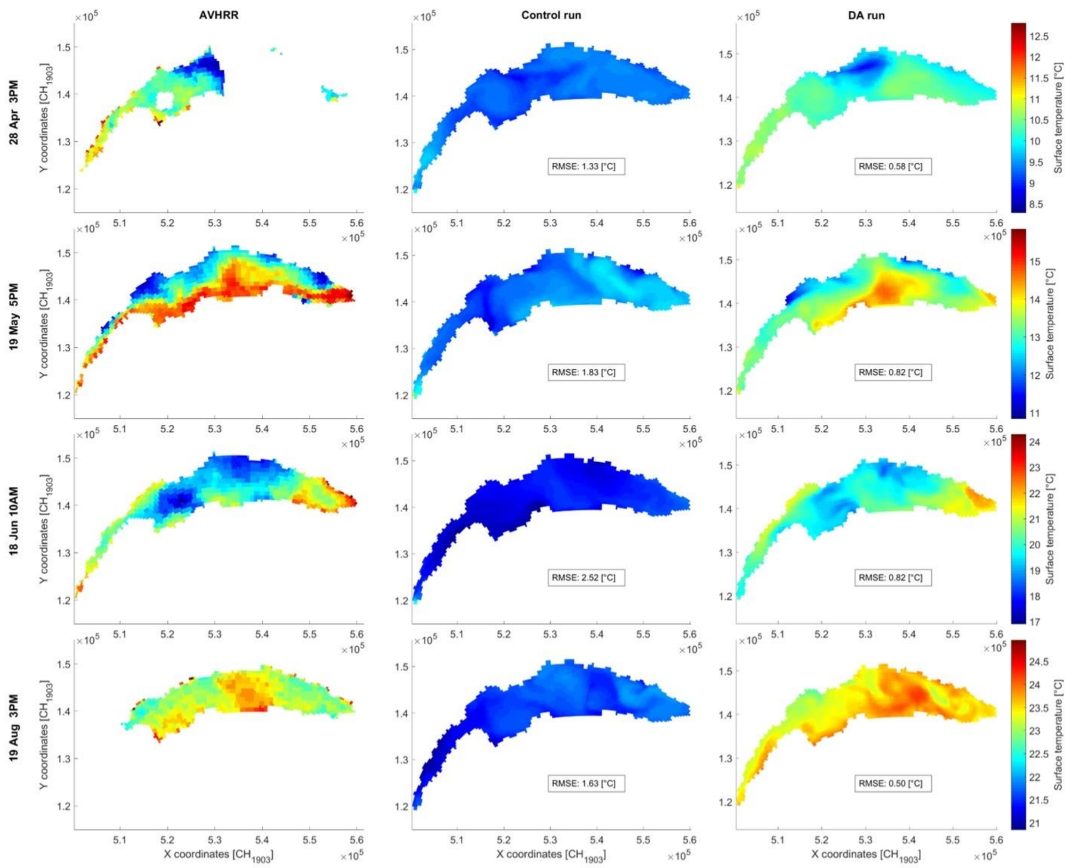


Figure S3.1: Surface temperature comparison of the AVHRR observations (left column), control run (central column), and DA run (right column) at selected analysis times (four rows) of 2017. The first row highlights the assimilation of sporadic data and the second row of complex surface patterns. The third row is an example of upwelling phenomena, and the fourth row of gyre-like structures.

Authors contribution

TB, DB, AW, PC designed the procedure and TB carried it out. PC and JS helped TB in the data assimilation implementation, GL and SW retrieved and processed the raw AVHRR data from satellites. TB prepared the manuscript with contributions from all co-authors.

Acknowledgements

The authors would like to thank Stef Hummel (Deltares) and Martin Verlaan (TU Delft / Deltares) for their help implementing the coupling between Delft3D-FLOW and OpenDA. This project was supported by the European Space Agency's Scientific Exploitation of Operational Missions element (CORESIM contract No.: AO/1-8216/15/I-SBo) and by EPFL.

References

Anderson, L. A., Robinson, A. R. and Lozano, C. J.: Physical and biological modeling in the Gulf Stream region, Deep Sea Res. Part Oceanogr. Res. Pap., 47(10), 1787–1827, 2000.

Bannister, R. N.: A review of operational methods of variational and ensemble-variational data assimilation, Q. J. R. Meteorol. Soc., 143(703), 607–633, doi:10.1002/qj.2982, 2017.

Bárdossy, A. and Singh, S. K.: Robust estimation of hydrological model parameters, Hydrol. Earth Syst. Sci., 12(6), 1273–1283, 2008.

Bertino, L., Evensen, G. and Wackernagel, H.: Sequential data assimilation techniques in oceanography, Int. Stat. Rev., 71(2), 223–241, doi:10.1111/j.1751-5823.2003.tb00194.x, 2007.

Bouffard, D., Kiefer, I., Wüest, A., Wunderle, S. and Odermatt, D.: Are surface temperature and chlorophyll in a large deep lake related? An analysis based on satellite observations in synergy with hydrodynamic modelling and in-situ data. Remote Sens. Environ., 209, 510–523, https://doi.org/10.1016/j.rse.2018.02.056, 2018.

Bouffard, D. and Wüest, A.: Convection in lakes, Annu. Rev. Fluid Mech., 51(1), 189–215, doi:10.1146/annurev-fluid-010518-040506, 2019

Carpenter, J., Clifford, P. and Fearnhead, P.: Improved particle filter for nonlinear problems, IEE Proc. - Radar Sonar Navig., 146(1), 2, doi:10.1049/ip-rsn:19990255, 1999.

Crow, W. T.: Correcting land surface model predictions for the impact of temporally sparse rainfall rate measurements using an Ensemble Kalman Filter and surface brightness temperature observations, J. Hydrometeorol., 4(5), 960–973, doi:10.1175/1525-7541(2003)004<0960:CLSMPF>2.0.CO;2, 2003.

De Lannoy, G. J. M., Reichle, R. H., Houser, P. R., Pauwels, V. R. N. and Verhoest, N. E. C.: Correcting for forecast bias in soil moisture assimilation with the Ensemble Kalman Filter, Water Resour. Res., 43(9), W09410, doi:10.1029/2006WR005449, 2007a.

De Lannoy, G. J. M., Houser, P. R., Pauwels, V. R. N. and Verhoest, N. E. C.: State and bias estimation for soil moisture profiles by an Ensemble Kalman Filter: effect of assimilation depth and frequency, Water Resour. Res., 43(6), W06401, doi:10.1029/2006WR005100, 2007b.

Dee, D. P. and Da Silva, A. M.: Data assimilation in the presence of forecast bias, Q. J. R. Meteorol. Soc., 124(545), 269–295, doi:10.1002/qj.49712454512, 1998.

Deltares: Delft3D-FLOW User Manual: Simulation of multi-dimensional hydrodynamic flows and transport phenomena, [online] Available from: https://oss.deltares.nl/documents/183920/185723/Delft3D-FLOW_User_Manual.pdf, 2015.

Deltares: The OpenDA Data-assimilation Toolbox, [online] Available from: www.openda.org, 2019.

Eknes, M. and Evensen, G.: An Ensemble Kalman Filter with a 1-D marine ecosystem model, J. Mar. Syst., 36(1–2), 75–100, doi:10.1016/S0924-7963(02)00134-3, 2002.

El Serafy, G. Y., Gerritsen, H., Hummel, S., Weerts, A. H., Mynett, A. E. and Tanaka, M.: Application of data assimilation in portable operational forecasting systems—the DATools assimilation environment, Ocean Dyn., 57(4–5), 485–499, doi:10.1007/s10236-007-0124-3, 2007.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res., 99(C5), 10143, doi:10.1029/94JC00572, 1994.

Evensen, G.: The Ensemble Kalman Filter: theoretical formulation and practical implementation, Ocean Dyn., 53(4), 343–367, doi:10.1007/s10236-003-0036-9, 2003.

Evensen, G.: The ensemble Kalman filter for combined state and parameter estimation, IEEE Control Syst., 29(3), 83–104, doi:10.1109/MCS.2009.932223, 2009.

Gaudard, A., Weber, C., Alexander, T. J., Hunziker, S. and Schmid, M.: Impacts of using lakes and rivers for extraction and disposal of heat, Wiley Interdiscip. Rev. Water, 5(5), e1295, doi:10.1002/wat2.1295, 2018.

Gentemann, C. L., Donlon, C. J., Stuart-Menteth, A., and Wentz, F. J.: Diurnal signals in satellite sea surface temperature measurements, Geophys. Res. Lett., 30(3), 1140, doi:10.1029/2002GL016291, 2003.

Gillijns, S., Mendoza, O. B., Chandrasekar, J., De Moor, B. L. R., Bernstein, D. S. and Ridley, A.: What is the Ensemble Kalman Filter and how well does it work?, in American Control Conference, 2006, pp. 6–pp, IEEE., 2006.

Goudsmit, G.-H., Burchard, H., Peeters, F. and Wüest, A.: Application of k-ϵ turbulence models to enclosed basins: The role of internal seiches, J. Geophys. Res. Oceans, 107(C12), 23-1-23–13, doi:10.1029/2001JC000954, 2002.

Hawley, N., Johengen, T. H., Rao, Y. R., Ruberg, S. A., Beletsky, D., Ludsin, S. A., Eadie, B. J., Schwab, D. J., Croley, T. E. and Brandt, S. B.: Lake Erie hypoxia prompts Canada-U.S. study, Eos Trans. Am. Geophys. Union, 87(32), 313, doi:10.1029/2006EO320001, 2006.

Hoel, H., Law, K. J. H. and Tempone, R.: Multilevel ensemble Kalman filtering, SIAM J. Numer. Anal., 54(3), 1813–1839, doi:10.1137/15M100955X, 2016.

Houtekamer, P. L. and Mitchell, H. L.: A sequential Ensemble Kalman Filter for atmospheric data assimilation, Mon. Weather Rev., 129(1), 123–137, doi:10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2, 2001.

Kalman, R. E.: A new approach to linear filtering and prediction problems, J. Basic Eng., 82(1), 35, doi:10.1115/1.3662552, 1960.

Kiefer, I., Odermatt, D., Anneville, O., Wüest, A. and Bouffard, D.: Application of remote sensing for the optimization of in-situ sampling for monitoring of phytoplankton abundance in a large lake, Sci. Total Environ., 527–528, 493–506, doi:10.1016/j.scitotenv.2015.05.011, 2015.

Kourzeneva, E.: Assimilation of lake water surface temperature observations using an extended Kalman filter, Tellus Dyn. Meteorol. Oceanogr., 66(1), 21510, doi:10.3402/tellusa.v66.21510, 2014.

Kuragano, T. and Kamachi, M.: Global statistical space-time scales of oceanic variability estimated from the TOPEX/POSEIDON altimeter data, J. Geophys. Res. Oceans, 105(C1), 955–974, doi:10.1029/1999JC900247, 2000.

Kurniawan, A., Ooi, S. K., Hummel, S. and Gerritsen, H.: Sensitivity analysis of the tidal representation in Singapore Regional Waters in a data assimilation environment, Ocean Dyn., 61(8), 1121–1136, doi:10.1007/s10236-011-0415-6, 2011.

Lahoz, W., Khattatov, B. and Menard, R., Eds.: Data Assimilation, Springer Berlin Heidelberg, Berlin, Heidelberg., 2010.

van Leeuwen, P. J.: Particle Filtering in geophysical systems, Mon. Weather Rev., 137(12), 4089–4114, doi:10.1175/2009MWR2835.1, 2009.

Li, Z., Chao, Y., McWilliams, J. C. and Ide, K.: A three-dimensional variational data assimilation scheme for the regional ocean modeling system, J. Atmospheric Ocean. Technol., 25(11), 2074–2090, doi:10.1175/2008JTECHO594.1, 2008.

Lieberherr, G. and Wunderle, S.: Lake surface water temperature derived from 35 years of AVHRR sensor data for European lakes, Remote Sens., 10(7), 990, doi:10.3390/rs10070990, 2018.

Lieberherr, G., Riffler, M. and Wunderle, S.: Performance assessment of tailored split-window coefficients for the retrieval of lake surface water temperature from AVHRR satellite data, Remote Sens., 9(12), 1334, doi:10.3390/rs9121334, 2017.

Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., van Dijk, A. I. J. M., van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O. and Restrepo, P.: Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities, Hydrol. Earth Syst. Sci., 16(10), 3863–3887, doi:10.5194/hess-16-3863-2012, 2012.

MacIntyre, S. and Melack, J. M.: Vertical and horizontal transport in lakes: linking littoral, benthic, and pelagic habitats, J. North Am. Benthol. Soc., 14(4), 599–615, doi:10.2307/1467544, 1995.

Madsen, H.: Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives, Adv. Water Resour., 26(2), 205–216, 2003.

Mao, J. Q., Lee, J. H. W. and Choi, K. W.: The Extended Kalman Filter for forecast of algal bloom dynamics, Water Res., 43(17), 4214–4224, doi:10.1016/j.watres.2009.06.012, 2009.

Minnett, P. J., Smith, M. and Ward, B.: Measurements of the oceanic thermal skin effect, Deep Sea Res. Part II Top. Stud. Oceanogr., 58(6), 861–868, doi:10.1016/j.dsr2.2010.10.024, 2011.

Moradkhani, H., Hsu, K.-L., Gupta, H. and Sorooshian, S.: Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, Water Resour. Res., 41(5), W05012, doi:10.1029/2004WR003604, 2005.

Nakamura, K., Higuchi, T. and Hirose, N.: Sequential data assimilation: information fusion of a numerical simulation and large scale observation data., J UCS, 12(6), 608–626, 2006.

Natvik, L.-J. and Evensen, G.: Assimilation of ocean colour data into a biochemical model of the North Atlantic, J. Mar. Syst., 40–41, 127–153, doi:10.1016/S0924-7963(03)00016-2, 2003.

Oesch, D., Jaquet, J. -M., Klaus, R. and Schenker, P.: Multi-scale thermal pattern monitoring of a large lake (Lake Geneva) using a multi-sensor approach, Int. J. Remote Sens., 29(20), 5785–5808, doi:10.1080/01431160802132786, 2008.

Oesch, D. C., Jaquet, J.-M., Hauser, A. and Wunderle, S.: Lake surface water temperature retrieval using Advanced Very High Resolution Radiometer and Moderate Resolution Imaging Spectroradiometer data: Validation and feasibility study, J. Geophys. Res., 110(C12), C12014, doi:10.1029/2004JC002857, 2005.

Qi, L., Ma, R., Hu, W. and Loiselle, S. A.: Assimilation of MODIS chlorophyll-a data into a coupled hydrodynamic-biological model of Taihu Lake, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 7(5), 1623–1631, doi:10.1109/JSTARS.2013.2280815, 2014.

Rawlins, F., Ballard, S. P., Bovis, K. J., Clayton, A. M., Li, D., Inverarity, G. W., Lorenc, A. C. and Payne, T. J.: The Met Office global four-dimensional variational data assimilation scheme, Q. J. R. Meteorol. Soc., 133(623), 347–362, doi:10.1002/qj.32, 2007.

Reichle, R. H., Walker, J. P., Koster, R. D. and Houser, P. R.: Extended versus Ensemble Kalman Filtering for land data assimilation, J. Hydrometeorol., 3(6), 728–740, doi:10.1175/1525-7541(2002)003<0728:EVEKFF>2.0.CO;2, 2002a.

Reichle, R. H., McLaughlin, D. B. and Entekhabi, D.: Hydrologic data assimilation with the Ensemble Kalman Filter, Mon. Weather Rev., 130(1), 103–114, doi:10.1175/1520-0493(2002)130<0103:HDAWTE>2.0.CO;2, 2002b.

de Rosnay, P., Drusch, M., Vasiljevic, D., Balsamo, G., Albergel, C. and Isaksen, L.: A simplified Extended Kalman Filter for the global operational soil moisture analysis at ECMWF, Q. J. R. Meteorol. Soc., 139(674), 1199–1213, doi:10.1002/qj.2023, 2013.

Schwefel, R., Gaudard, A., Wüest, A. and Bouffard, D.: Effects of climate change on deepwater oxygen and winter mixing in a deep lake (Lake Geneva): Comparing observational findings and modeling, Water Resour. Res., 52(11), 8811–8826, doi:10.1002/2016WR019194, 2016.

Soomets, T., Kutser, T., Wüest, A. and Bouffard, D.: Spatial and temporal changes of primary production in a deep peri-alpine lake, Inland Waters, 1-12, https://doi.org/10.1080/20442041.2018.1530529, 2019.

Stanev, E. V., Schulz-Stellenfleth, J., Staneva, J., Grayek, S., Seemann, J. and Petersen, W.: Coastal observing and forecasting system for the German Bight, Ocean Sci., 7(5), 569–583, doi:10.5194/os-7-569-2011, 2011.

Stroud, J. R., Lesht, B. M., Schwab, D. J., Beletsky, D. and Stein, M. L.: Assimilation of satellite images into a sediment transport model of Lake Michigan, Water Resour. Res., 45(2), W02419, doi:10.1029/2007WR006747, 2009.

Stroud, J. R., Stein, M. L., Lesht, B. M., Schwab, D. J. and Beletsky, D.: An Ensemble Kalman Filter and smoother for satellite data assimilation, J. Am. Stat. Assoc., 105(491), 978–990, doi:10.1198/jasa.2010.ap07636, 2010.

Šukys, J. and Kattwinkel, M.: SPUX: scalable particle markov chain monte carlo for uncertainty quantification in stochastic ecological models, Adv. Parallel Comput., 159–168, doi:10.3233/978-1-61499-843-3-159, 2018.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res. Atmospheres, 106(D7), 7183–7192, doi:10.1029/2000JD900719, 2001.

van Velzen, N. and Verlaan, M.: COSTA a problem solving environment for data assimilation applied for hydrodynamical modelling, Meteorol. Z., 16(6), 777–793, doi:10.1127/0941-2948/2007/0241, 2007.

Verlaan, M. and Heemink, A. W.: Nonlinearity in data assimilation applications: a practical method for analysis, Mon. Weather Rev., 129(6), 1578–1589, doi:10.1175/1520-0493(2001)129<1578:NIDAAA>2.0.CO;2, 2001.

Weerts, A. H., El Serafy, G. Y., Hummel, S., Dhondia, J. and Gerritsen, H.: Application of generic data assimilation tools (DATools) for flood forecasting purposes, Comput. Geosci., 36(4), 453–463, doi:10.1016/j.cageo.2009.07.009, 2010.

Yeates, P. S., Imberger, J. and Dallimore, C.: Thermistor chain data assimilation to improve hydrodynamic modeling skill in stratified lakes and reservoirs, J. Hydraul. Eng., 134(8), 1123–1135, 2008.

Zhang, Z., Beletsky, D., Schwab, D. J. and Stein, M. L.: Assimilation of current measurements into a circulation model of Lake Michigan, Water Resour. Res., 43(11), W11407, doi:10.1029/2006WR005818, 2007.

Chapter 4
Meteolakes: an online 3D monitoring and forecasting platform for lakes

Theo Baracchini[1], Alfred Wüest[1,2], Damien Bouffard[2]

[1] *Physics of Aquatic Systems Laboratory (APHYS) – Margaretha Kamprad Chair, ENAC, EPFL, 1015 Lausanne, Switzerland.*

[2] *Eawag, Swiss Federal Institute of Aquatic Science and Technology, Surface Waters – Research and Management, 6047 Kastanienbaum, Switzerland.*

Abstract

Monitoring and forecasting ecosystem reactions to local anthropogenic pressures, global climate changes and extreme weather events, are great challenges of environmental management. Three-dimensional models allow resolving transport and mixing processes involved in lake dynamics, and forecast critical events with economical and human impacts. However, these models are complex to implement, and their use is often limited to modellers and scientists.

Here, we present *Meteolakes*, an online platform disseminating lake observations and three-dimensional numerical simulations in near real-time with short-term forecasts. Since summer 2016, the platform provided lake information to more than hundred thousand users around Lake Geneva. This article details the design and operations of such platform and its products. Applications of *Meteolakes* are demonstrated by two illustrative examples: an upwelling and a storm event with strong currents, which significantly affected the lake status and some citizens. Impacts of the data platform are observed at three different levels: for scientists it provides guidance in the design and analysis of field experiments and timely observations of transient physical processes. For governmental agencies and professionals, it helps in the planning and decision-making of water and resources management. Finally, for the public, *Meteolakes* is used for recreational activities but also to sensitise how dynamic and delicate lake ecosystems are.

4.1 Introduction

Adequate monitoring and reliable forecasting of ecosystem reactions to local anthropogenic pressures, global climate forcing, and to extreme weather events, are great challenges for environmental management. In lakes, some of those reactions can compromise crucial ecosystem services, such as drinking water quality, food resources, heat supply/discharge, or even the safety of citizens (e.g. storm navigation hazards, harmful algal blooms). The close monitoring and forecasting of lake physical status would allow mitigating the negative effects of a variety of societally important issues.

Most of the water quality issues (such as harmful algal bloom, hypoxia, or pollution) are related to the physical state of the lake. As a matter of fact, even small changes to some physical characteristics, such as the Lake Surface Water Temperature (LSWT), can alter both the ecological and physical dynamics due to its nonlinear nature (Adrian et al., 2009). Moreover, these processes are often subject to strong spatial variability, hence suggesting that traditional in-situ monitoring, may lack of representativeness due to its limited spatial coverage (Kiefer et al., 2015). Phenomena such as upwellings, gyres, inhomogeneous chlorophyll-a distributions, local high currents, are insufficiently documented and monitored. Yet they disrupt the spatial homogeneity and lentic nature of the systems, ultimately influencing biogeochemical processes in lakes at short time scales (MacIntyre and Melack, 1995) and its water quality. This brought to the fore the need for new monitoring programs (Hering et al., 2015), using novel approaches combining numerical simulations and remote sensing observations (Vörösmarty et al., 2015).

The adoption of one-dimensional hydrodynamic models is growing at a rapid pace (e.g. Gaudard et al., 2017; Peeters et al., 2007; Schwefel et al., 2016; Straile et al., 2010), while the one of three-dimensional models beyond the expert user has been limited, due to their complexity and their tedious calibration. Yet the latter is the only source of information capable of resolving physical processes at the large variety of spatio-temporal scales involved in lake dynamics. Some solutions have therefore been proposed to foster their spread, such as the second chapter of this thesis. An alternative is to ensure the reuse of data through the web-based dissemination of models outputs. Indeed, when the effort to develop such models is made, the use of its output is often limited to the modellers themselves. Few authors proposed solutions to disseminate their results to a wider audience, thereby promoting more interdisciplinary contributions from various scientific but also public and policy-making audiences.

Because of uncertainties in forcing data, initial conditions and process understanding, model deviations with respect to the real systems are unavoidable (Lahoz et al., 2010). Those uncertainties have to be taken into account through integrated data-model approaches (Chapter 3 of this thesis) to provide a reliable understanding of the system. The results of the combination of direct observations and numerical simulations can be disseminated in a timely and comprehensive manner by operational forecasting systems.

As of today, a limited number of operational systems with an integrated data-model approach exist to monitor inland waters. Among those, the Great Lakes Operational Forecast System (GLOFS) provides nowcast and forecast guidance of various physical characteristics, such as water levels, temperature, currents, for the five North American Great Lakes (Chu et al., 2011). Other platforms have been developed such as the online lake modelling tool FLake-Global, a platform for the one-dimensional estimation of temperature and mixing conditions in any shallow freshwater lake at seasonal scale (Kirillin et al., 2011), or the three-dimensional monitoring and forecasting tool WIS-CAST, applied to a mid-sized lake for a duration of three months (Kimura and Wu, 2018). Nevertheless, the number of online lake operational systems remains very limited, particularly when compared to the widespread use and development of meteorological and coastal ocean systems. For small to mid-sized lakes, such frameworks are virtually non-existent, due to the limited amount of accurate remote sensing products available for small scale and optically complex waters (Qi et al., 2014a), the requirement for detailed morphological information, and most importantly the need for real-time and high-resolution meteorological forcing (Kamarainen et al., 2009; Read et al., 2014).

In this study, we present a near real-time monitoring and forecasting system for lakes with its online platform, *meteolakes.ch*. *Meteolakes* processes and disseminates past, present and future spatial information on lakes biophysical status derived from in-situ measurements, space-borne remote sensing observations and numerical simulations to create a synoptic view of the basin dynamics. Operational since 2016, the system provides open access to lake environmental information to hundreds of daily visitors and aims at enhancing recreational activities, hazard warnings, risk assessment, decision-making and improving our understanding of the interactions between physical and biological lake processes. Impacts of the platform are observed in different communities. For scientists, it guides in the design and planning of field campaigns (e.g. glider deployments, Forrest et al., 2018), this is of particular importance for monitoring the transient dynamics of phenomena such as upwellings, gyres, calcite precipitation (Nouchi et al., 2019), which are difficult to observe using in-situ observations without *a priori* knowledge. For lake-related professionals, various use cases have been identified among fishermen, beach managers, rescue organisms, drinking water intake operators and engineering companies. Finally, hundreds of citizens use *Meteolakes* on a daily basis for recreational activities (e.g. navigation, swimming) and can discover how dynamic lake ecosystems are.

For this purpose, we designed an online data platform, including an Application Programming Interface (API), and real-time data processing chain, to distribute key hydrodynamic variables, with a complete spatio-temporal coverage of the lake, in near real-time with short-term forecasts. This article provides an overview of *Meteolakes*, its technical design and implementation, observational and modelling components. Finally, data products are showcased through two examples of notable meso-scale physical phenomena, including a strong upwelling and a storm event with localized fast currents, which affected commercial and recreational activities in and around the lake.

The publication is organized as follows: Section 4.2 provides an overview of the system components, data, tasks automation, and interfaces. Products and applications are described in Section 4.3. Section 4.4 discusses the relevance of such products for the scientific, professional, and general public. Finally, future activities and conclusion are given in the final section.

4.2 System overview & implementation

Meteolakes combines lake hydrodynamic simulations, in-situ measurements and remote sensing observations in real-time from multiple sources. This section provides overview of the data composition and acquisition scheme, the computational framework, key modelling components and the online interface.

4.2.1 Platform components

Figure 4.1 depicts an outline of system processes and key tasks, software and hardware components. Detailed information on each element are provided in the following sub-sections. Four Swiss lakes are currently available on *Meteolakes*: Lake Geneva, Lake Zurich, Lake Biel, and Greifensee. For the sake of simplicity, this publication focuses on Lake Geneva only.

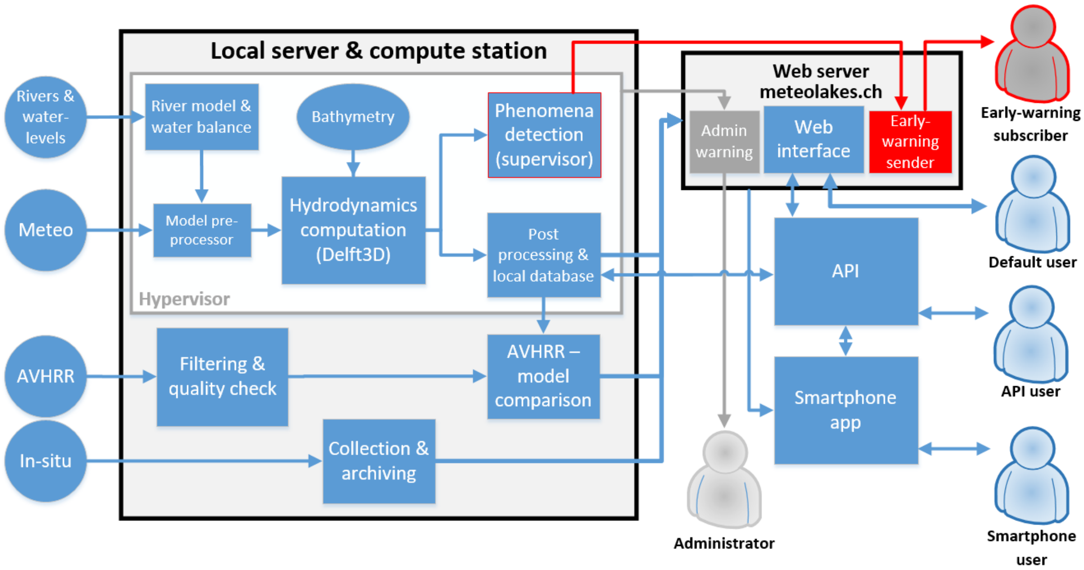


Figure 4.1: Flowchart of system processes and overview of key tasks, software and hardware components. Circles depict data and arrows represent information flows. Bold black frames are computing facilities, and filled rectangles are tasks or software elements. The grey frame (hypervisor) represents a monitoring system for events status warning.

4.2.1.1 Study site

Lake Geneva (locally Le Léman) is the largest freshwater lake of Western Europe (depth, surface area and volume of respectively 309 m, 580 km^2 and 89 km^3). It is located between Switzerland and France (46.458 °N, 6.528 °E), in the perialpine region at an altitude of 372 m. The hydraulic retention time is 11.4 years, with its main tributary, the Rhône River located at the eastern end (Figure 4.2), accounting for 75 % of the water flow entering the lake (average (1986-2013) of 184 m^3/s). The Dranse (France) is the second largest tributary with an annual average flow of 20 m^3/s accounting for 8 % of water inflow. Two largest remaining inflows (Venoge and Aubonne, Figure 4.2) have been retained in the model. Combined, they account of 4 % (10 m^3/s) of water input into the lake. Neglected inflows and diffuse sources have been accounted for through a rescaling of the selected afore-mentioned four inflows (more details in Section 4.2.1.2). Lake Geneva has one outflow, located in Geneva (western end). The outflow is dam-operated to regulate the lake water level. The lake is composed of two main basins: the Grand Lac (main basin, with maximum depth of 309 m, mean depth of 160 m, mean width of 10 km in which gyres are frequently observed) and the Petit Lac (small western basin, maximum depth of 70 m, mean width of 4.5 km). Due to its large depth and mild winter temperatures, complete deep convective mixing occurs only every 5 to 10 winters in Lake Geneva (Schwefel et al., 2016). Average wind speeds above the lake are in the range of 1 to 2 m/s and current speeds reach ~0.1 m/s on windy days (with wind speeds > 5 m/s).

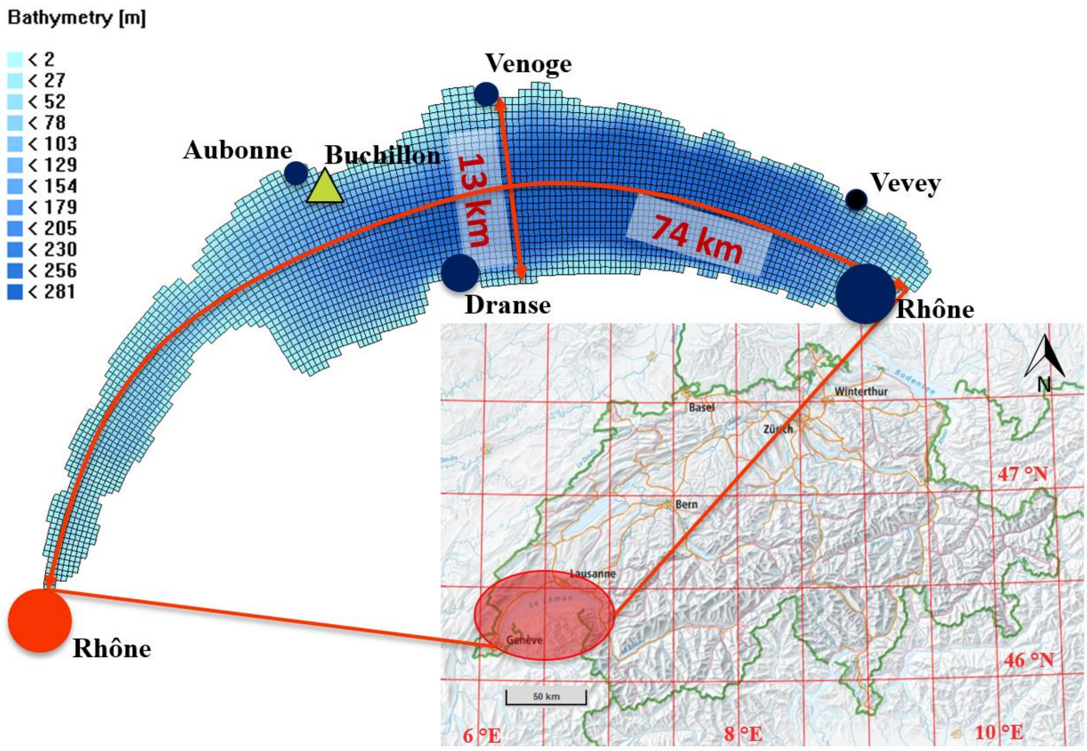


Figure 4.2: Computational grid and bathymetry of Lake Geneva. The yellow triangle corresponds to the in-situ mast. The four blue dots are the inflows modelled (dot radius, not proportional, indicates the flow contribution), the red dot is the outflow. The black dot (Vevey) corresponds to a harbour, which suffered damage during the storm illustrated in the results.

4.2.1.2 Hydrodynamics computations

Delft3D-FLOW – The Delft3D-FLOW hydrodynamic numerical model has been used for this data platform. Delft3D-FLOW is an open-source modelling suite developed by Deltares, Netherlands. While it has been initially intended for coastal and estuarine waters, it has been further applied to rivers and lakes. Among others, the implementation of its z-coordinates scheme enabled its use for basins with steep bathymetry, such as Lake Geneva. Detailed model description and its equations can be found in the Delft3D-FLOW manual (Deltares, 2015).

In the case of Lake Geneva, this proven model has been extensively calibrated and validated and we refer the reader to the two previous chapters for detailed descriptions. The model has a 450 m horizontal grid resolution and 100 unevenly distributed (from 20 cm at the surface to several m in the hypolimnion) vertical z-layers (layers are horizontal and do not follow the lake bed). To maintain model stability with the κ-ϵ turbulence closure model (Goudsmit et al., 2002) for the selected grid size, a computational time-step of 1 min is used. The model has initially been started from an in-situ temperature profile taken at the deepest location in January 2015 (when the lake was partially mixed) and has been running continuously since that time. The model is forced by river data and two-dimensional meteorological data.

Meteo forcing – MeteoSwiss COSMO-1 and COSMO-E products (MeteoSwiss, 2019a, 2019b) are used as meteorological forcing. They consist of various variables, derived from their atmospheric model tailored to the Alpine region. Seven variables are extracted from those files (by the model pre-processor, Figures 4.1 and 4.3): wind speed, wind direction, solar radiations, air pressure, cloud cover, relative humidity, and air temperature. Those variables are stored in daily netCDF files with a spatial domain covering entire Switzerland. COSMO-1 products are provided on a 1.1 km grid with hourly resolution. COSMO-E provide forecasts of the state of the atmosphere over 120 hours, at a reduced spatial resolution (hourly and 2.2 km grid). Model hindcasts are forced with COSMO-1 reanalyses, while daily forecasts are forced by COSMO-E forecasts.

Rivers forcing and water levels – River data consists mainly of flow and temperature data collected at 10 min intervals by the Federal Office of the Environment (FOEN). Flow and temperature are provided in real-time for the Rhône River (inlet and outlet, Figure 4.2). Lake water levels are collected ~4 km east from the Buchillon station, along the shore at 10 min interval. Finally, the data (FOEN) also contains 4.5 days forecasts of the Rhône River inflow based on MeteoSwiss COSMO-E products.

After the daily download from the FOEN FTP server, the data passes various integrity checks. The filtering includes removal of duplicates, removal of incoherent observations, and inference of missing data. We then estimate the outflow over 4.5 days. This is done by singular spectrum analysis forecasting (Marques et al., 2006) using the first four principal components. When both the lake outflow and inflow from the Rhône River measurements and forecasts are available for the computational period, the contributions from the remaining rivers are generated by the river model. Two methods have been implemented in *Meteolakes*: the first one considers a constant lake volume and the difference of the outflow with the Rhône inflow is spread among the remaining three tributaries. The second method, currently used on the platform, aims at reproducing the water level variations of the lake and therefore requires additional data (water levels measurements). In this method a Gaussian filter is first applied to the observed water level time-series to remove the high-frequency signature from surface seiches. The water levels time-series then undergoes singular spectrum analysis forecasting to generate the future 4.5 days needed. For each time-step *i*, the missing flow ($Q_{i,m}$) is given by:

$$Q_{i,m} = Q_{i.out} - Q_{i,Rhône(in)} + A(H)\frac{dH_i}{dt_i} \qquad (4.1)$$

With $Q_{i,out}$ the outflow from Lake Geneva (Figure 4.2), $Q_{i,Rhône(in)}$ the inflow from the Rhône, *A* the surface area of the lake as a function of its water level *H*, and *dH* the change in lake water level during time-step dt_i. The remaining flow $Q_{i,m}$ is then distributed among the three remaining rivers, based on their flow contribution over the past 20 years. Additional filters ensure the remaining flow is coherent. Finally, river temperature is needed to drive the model. For the Rhône, real-time measurements are used, however for the remaining rivers or Rhône temperature forecasts, a physical model is operated. From the work of Toffolon and Piccolroaz (2015), it is possible to estimate the river temperature as a function of air temperature and discharge. This is achieved with the following equations:

$$\frac{dT_{river}}{dt} = \frac{1}{\delta}\left\{a_1 + a_2T_{air} - a_3T_{river} + \theta\left[a_5 + a_6\cos\left(2\pi\left(\frac{t}{t_y} - a_7\right)\right) - a_8T_{river}\right]\right\}, \qquad (4.2)$$

$$\delta = \theta^{a_4}, \qquad (4.3)$$

$$\theta = \frac{Q_{river}}{\bar{Q}_{river}}, \qquad (4.4)$$

where the eight parameters, a_1-a_8, are obtained through calibration. Written in this form, they avoid having to specify explicitly all geometrical characteristics of the river and specific heat inputs (Toffolon and Piccolroaz, 2015). T_{air} is the air temperature obtained from COSMO-E products, *t* and t_y are the time and the duration of a year (in the same units as *t*), respectively, and $\theta = Q_{river}/\bar{Q}_{river}$ is the dimensionless discharge. This entire processing is monitored by the hypervisor (Figure 4.1), which warns the administrator via emails when encountering missing or incoherent data, and what has been done to circumvent the problem.

4.2.1.3 Remote sensing monitoring

The space-borne Advanced Very High Resolution Radiometer (AVHRR) sensor has been selected as operational remote sensing data source. Its moderate spatial (1 km) and high temporal (10 overpasses per day) resolution enables the real-time monitoring of LSWT and meso-scale to basin-scale lake dynamics. The partnering Oeschger Centre at the University of Bern, has an operational data downlink from satellites, which facilitated the access to the data. The AVHRR LSWT retrieval process, with locally adapted Split window coefficients for Lake Geneva, is described in Lieberherr and Wunderle (2018) and Lieberherr et al. (2017).

The AVHRR information is then retrieved via FTP to *Meteolakes* local database for additional screening and comparison with the model. The filtering on *Meteolakes* local server (Figure 4.1) includes removing pixels with quality levels lower than 4 (Lieberherr and Wunderle, 2018). Data is then mapped to the model grid, compared with its corresponding model surface temperature field, and uploaded via FTP to *Meteolakes* web server. The AVHRR skin temperature is directly compared with the model bulk temperature and no skin-to-bulk conversion is applied in this current version due to the non-trivial parameterization when both wind and convective processes play significant role in the near surface turbulence. Differences due to skin effects can thereby be expected. The update of the local database and comparison with the model is performed every three hours (Figure 4.3).

4.2.1.4 In-situ monitoring

A small nearshore permanent mast, located in shallow waters off the town of Buchillon (Figure 4.2), has been instrumented with various meteorological and in-water sensors. Those include thermistors located at 1 m and 35 m depth, two radiometers (Heitronics Pyrometer KT15II) measuring lake skin temperature, an ADCP, oxygen and CO_2 loggers. Meteorological observations include air temperature, solar radiations, wind speed and direction, humidity, dew point, and air pressure. Most of the records measured by those sensors are publicly shared on *meteolakes.ch*.

A programmable Campbell Scientific controller with data-logger and a GPRS module ensures the real-time transmission of the data. This controller can be accessed through the LoggerNet software suite from any computer allowing real-time broadcast and changes in the setup of the instruments. The data collection and update of the local database is performed on an hourly basis (Figure 4.3) and provided online on *meteolakes.ch*. Radiometric data is further transmitted to the Oeschger Centre at the University of Bern in order to improve the satellite-based skin temperature processing.

4.2.2 Computational framework

Meteolakes model computations and data processing is performed on a local server (Figure 4.1). The machine comprises two Intel Xeon E5-2697v4 with 256 GB of error-correcting code random access memory, and a RAID-1 configured storage solution. While the system has been built to be easily computationally distributed, so far all numerical simulations are performed on this machine.

The three-dimensional hydrodynamic simulations are run on a daily basis (Figure 4.3). Every day, the system recomputes the hydrodynamics of the lake starting from the previous Sunday at 00h00, using COSMO-1 gridded surface meteorological reanalysis and tributaries observations. Initial conditions are generated on Sundays by a computation comprising the entire previous week and using reanalysis forcing. Forecasts from the previous day are overwritten the next day by the new nowcast cycle. Those daily computations then perform a 4.5 days hydrodynamic forecast using the COSMO-E products and river model (Section 4.2.1.2). As shown by Figure 4.3, not all lake models require river data to run.

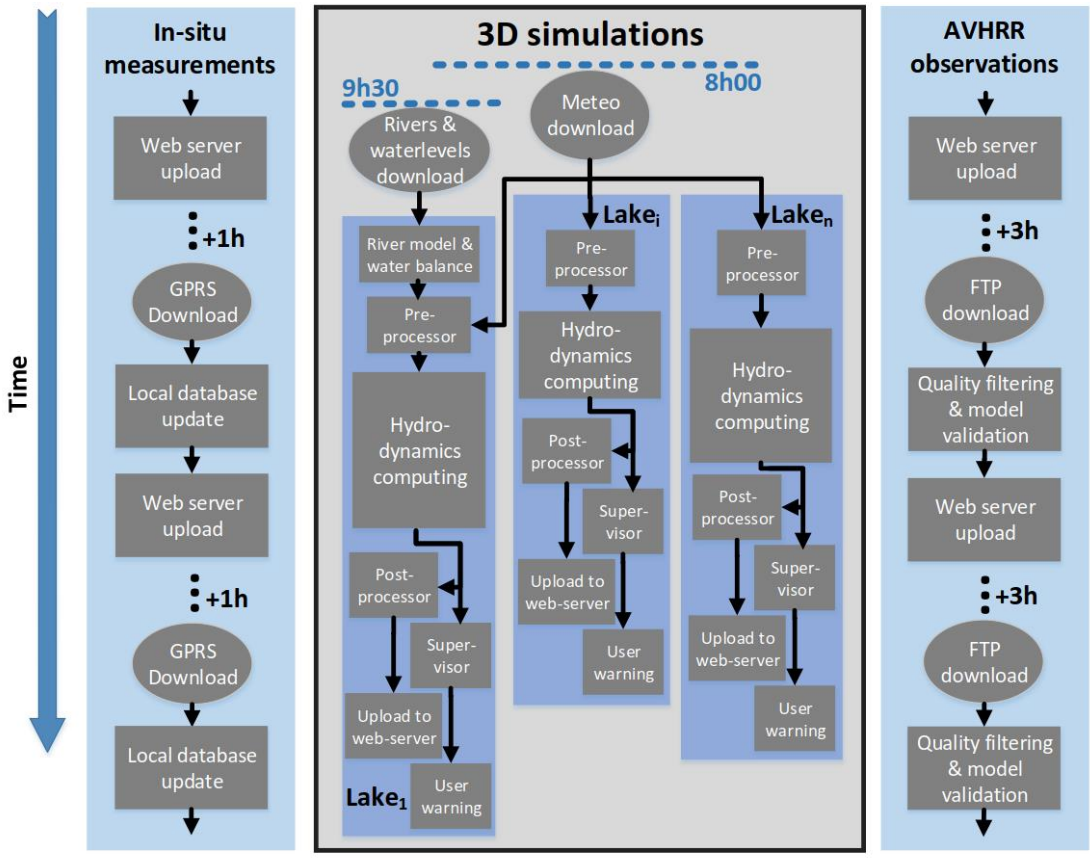


Figure 4.3: Daily workflow and automated main tasks performed by the background system (currently the local workstation). Routines for the numerical simulations (centre square), in-situ measurements (left) and remote sensing data (right) are schematized. In the numerical simulations, the asynchronous models computations is emphasized (river data is available later, therefore lakes requiring river data start their computations later). In-situ and remote sensing data is recursive (routines triggered every 1 to 3 hours). The varying size of some boxes (e.g. pro-processor, hydrodynamics computing) indicates different lake-specific computational complexities. The supervisor is *Meteolakes* early-warning system.

The automation of the processing tasks related to the numerical modelling is performed by PowerShell scripts triggered by the Windows Task Scheduler at ~8h00 in the morning (Figure 4.3, middle panel). The river model, model pre-/post-processors, and supervisor are coded in MATLAB and are called by the PowerShell scripting. The pre-processor formats river and meteorological data into Delft3D-FLOW input files. The post-processor creates enlightened netCDF and CSV model output files by extracting and saving only specific fields (e.g. temperature, flow velocity, grid information) at a reduced spatial resolution to optimize storage load and the online user experience. Finally, the supervisor analyses model results to detect notable physical phenomena. Those phenomena include upwelling flow (by k-means clustering), high surface flow velocities, and out-of-bounds cold/warm waters at various depths. The supervisor acts as an early warning-system by sending information (emails with text and image detailing the intensity and location of the event) to a list of subscribers via the *meteolakes.ch* web server. As shown by Figure 4.1, all those tasks are monitored by the hypervisor, which notifies the system administrator by email about eventual issues (e.g. missing data files, computational errors) in the model processing chain.

A python software has been developed to download and process the AVHRR data (Figure 4.3, right panel). This code runs as background process (daemon mode) and performs the tasks displayed in Figure 4.3 at user-selected frequency (three hours in this case). Tasks include a download from the Oeschger Centre FTP server and update of the local database, filtering of satellite images based on pixel quality flags, mapping to the model grid, comparison with corresponding (online or offline) model results, and FTP upload to the *meteolakes.ch* web server.

Automation of the in-situ data download is handled by the Campbell Scientific LoggerNet software (Figure 4.3, left panel). The local database is updated every hour. Additional python scripts triggered by the Windows Task Scheduler archive the in-situ data once a day. Figure 4.3 illustrates the scheduling of those tasks.

4.2.3 Web interface

Meteolakes comprises two main pathways for data dissemination. The majority of users interact with *Meteolakes* through its online interface, *meteolakes.ch*. Users with specific needs however can make use of the API. Both interfaces are described in the next sub-sections.

4.2.3.1 Online web-application

Meteolakes online interface is built around Web 2.0 concepts. It particularly emphasizes user-data interaction, intuitive information exchange through simple and responsive designs. The surfer becomes an active web user, able to visualize and interact with scientific data without requiring any particular scientific or technical background. Moreover, *meteolakes.ch* web interface has been built with platform and display scalability in mind, for a seamless experience on a broad range of devices (smartphone, computer, and tablet).

The AngularJS open-source JavaScript–based front-end framework is used to facilitate the web application development. We make use of various existing libraries and protocols such as the Leaflet map API, asynchronous server-client data transfers, for a spatially enabled and responsive content. To reduce computational load on the server, the vast majority of *Meteolakes* processing and rendering is performed client-side.

Apache Cordova mobile application development framework allowed wrapping up the CSS, HTML and JavaScript code into a packaged Android application. This enabled a distribution on the Google Play store, without the need to develop a truly native mobile application using platform-specific APIs. *Meteolakes* Android app can be downloaded at the following link: https://play.google.com/store/apps/details?id=ch.epfl.meteolakes&hl=fr_CH

The results stored on the web server, which are processed by the web interface or mobile application, are in text-file format. Upon user request, additional results can be displayed (e.g. temperature and flow velocity at different depth). Those are provided by the API directly from model results files located on the local server (Figure 4.1).

4.2.3.2 Application Programming Interface

Following users demands for raw data access and additional spatial information, we developed an API. *Meteolakes* API is built using the open-source Node.js runtime environment. Node.js allows the execution of JavaScript code outside the browser; in our case, it runs server-side directly on the *Meteolakes* compute and local server (Figure 4.1).

Data requests are made by generating URL links, which are interpreted by the server running the Node.js application. The developed back-end application runs as background task, constantly listening to a port on which the requests are made. When receiving a request, the server will extract in the model netCDF output files the desired dataset and send it back to the user. The data is sent in formatted CSV files. Due to the relatively large size model files can have, we decided to have the API directly interact with the local server computing the hydrodynamic models rather than interacting with *meteolakes*.ch web server, as the latter would require having significantly more file transfers. Detailed explanations on how to use the API and available data can be found directly on *meteolakes.ch* at the following link: http://meteolakes.ch/#!/data

4.3 Products and application examples

In this section, we provide an overview of *Meteolakes* online products. We start by describing the various interface elements, followed by examples on how such elements can be used for the practical assessment of physical phenomena. Two illustrative examples are presented, an upwelling and a storm event with localized high current speeds.

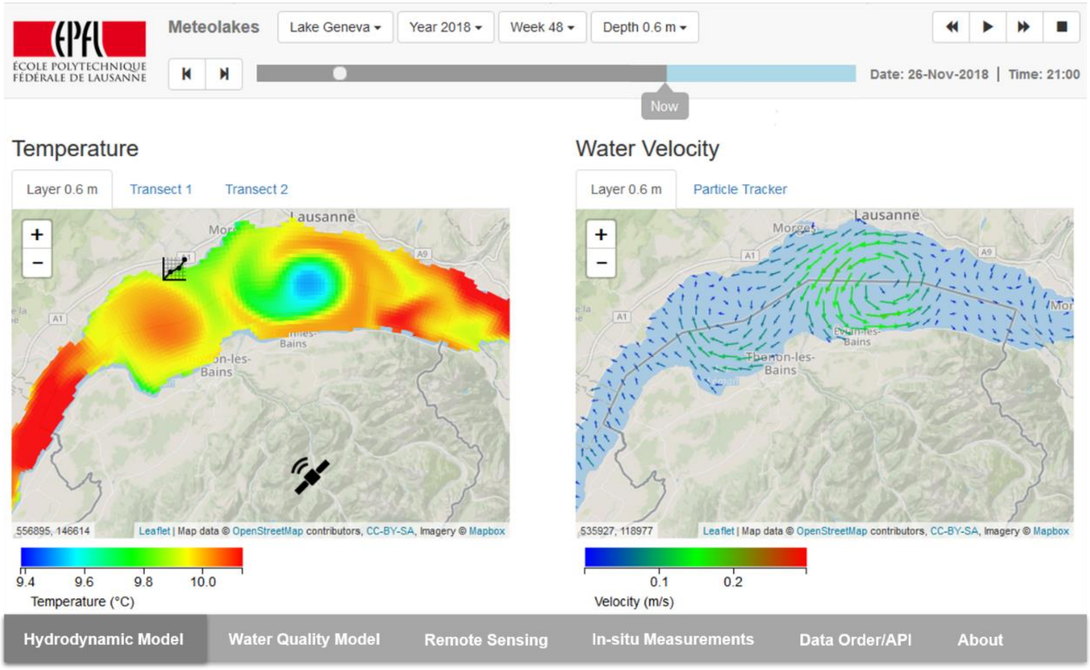


Figure 4.4: Online interface as viewable on *meteolakes.ch*. Simulated temperature (left map) and currents (right map) in Lake Geneva at 0.6 m depth for the time as indicate on the top right. The time-slider is blue for forecasts and grey for the past.

An illustration of *Meteolakes* homepage is presented in Figure 4.4. Two animated maps show the LWST of Lake Geneva on the left and currents on the right, computed by the Delft3D hydrodynamic model for the time indicated in the top panel. Several drop-down menus are available in the upper navigation bar. One can select the lake (currently four), the year (starting in 2009), the week and depth of interest. The hydrodynamic interface displays weekly periods. For each moment, the absolute time and position in the time-frame slider are provided. The time-slider is blue for forecasts and grey for the past. In addition to basic functionalities such as zoom and displacement, the maps can be clicked to obtain time-series at selected locations of interest. The temperature map has two additional tabs, providing a visualization of temperature over depth at pre-defined transects along the main axes of the lake. They allow the visualization of stratification and mixing. The flow velocity map has a secondary tab enabling a particle-tracking mode. When this mode is selected, the user has the possibility to release particles online at any spatial location (on the horizontal plane and over depth), and to follow their trajectories. Particles are treated as passive tracers, they are only advected horizontally by the user-defined model layer and are not impacted by vertical mixing nor settling. Finally, at the bottom of the page (Figure 4.4), a navigation bar allows to access the in-situ measurements, the remote-sensing validation, the API documentation and various additional system information. The water quality tab is addressed in the discussion section.

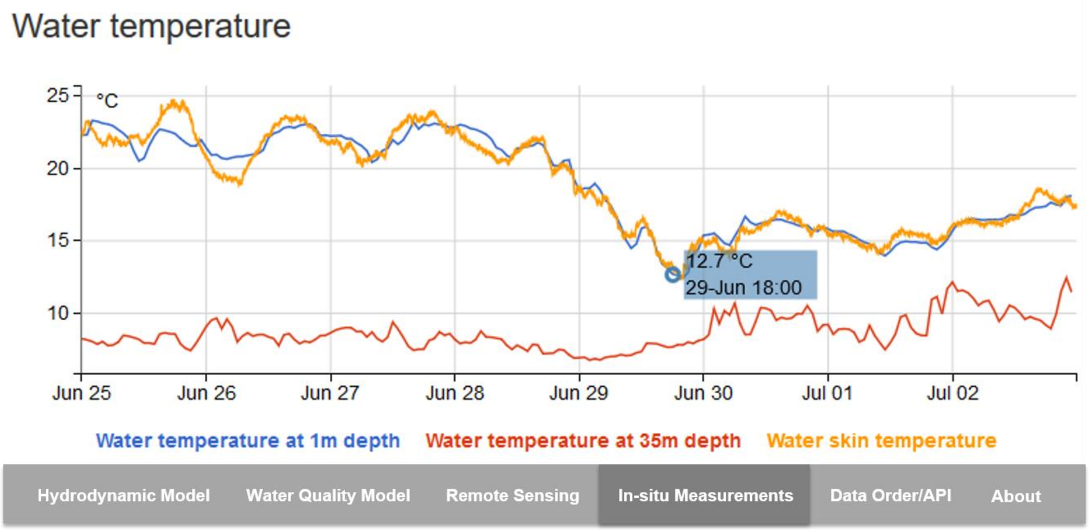


Figure 4.5: Example of in-situ data as displayed in the "In-situ Measurements" tab. Water temperature at 1 m depth (blue), at 35 m depth (red) and radiometric skin temperature measurements (orange) at the Buchillon station (Figure 4.2) in June/July 2017.

Figures 4.5 and 4.6 are examples of respectively in-situ measurements and remote sensing observations as displayed on *meteolakes.ch*. Figure 4.5 shows the water temperature at 1 m depth (blue), 35 m depth (red) and radiometric measurements of skin temperature (orange) measured at the Buchillon station. For in-situ measurements, the user has the possibility to define a period of interest (starting late 2016), this data is available in near real-time (maximum delay of one hour). Seven additional atmospheric variables are available (not shown here). Remote sensing observations and a comparison with the modelled LWST are available in near real-time through the "*Remote Sensing*" tab (Figure 4.6). It is worth noting that although the AVHRR data (upper right plot) has been filtered based on its quality flags (Section 4.2.1.3), no skin-to-bulk conversion is applied and the data is directly compared with model surface bulk temperature (upper left plot). Figure 4.6 provides a spatial (lower left plot) and temporal (lower right plot) overview of model deviation with respect to the observed skin temperature. The temporal evolution is shown by displaying the median difference (blue dot), along with the 10th and 90th percentiles of those offsets (green bars) for each image-model comparison. It is possible to cycle through this temporal comparison to visualize each satellite image and corresponding model snapshot.

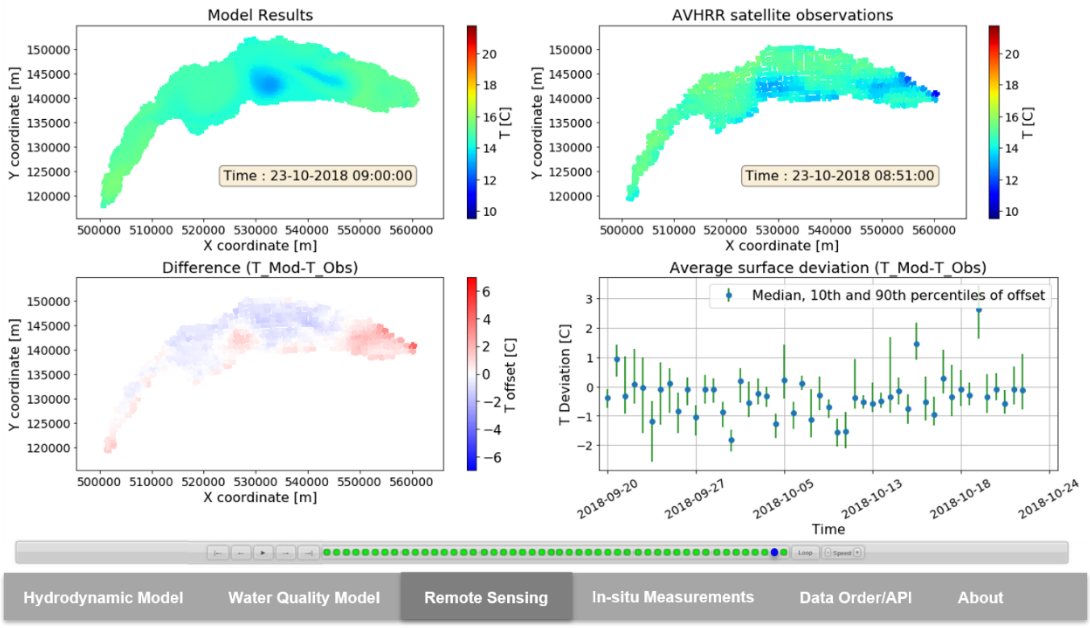


Figure 4.6: "*Remote Sensing*" tab: Model results (upper-left) are compared in real-time with AVHRR skin-temperature observations (upper-right). Bottom-left plot is the difference between the two datasets and bottom-right plot is the temporal evolution of those differences (here for one month prior to the shown snapshot).

Upwelling event – On 29th June 2017, a large upwelling had been forecasted by *Meteolakes* for the western basin of Lake Geneva. Figure 4.7 shows modelled temperature before (left plots) and during (right plots) the event. These figures show a rather uniform and warm (21 °C) LSWT on 30th June with a stratified profile and a thermocline around 15 m depth. Less than two days later, the surface shows a large horizontal thermal gradient resulting from westerly wind events with LSWT 10 °C colder on the western part of the lake compared to the main basin. The bottom-right plot indicates that the stratification has been broken in the western basin, with a full upwelling of hypolimnetic water up to the surface. The signature of the upwelling and following basin-scale internal waves oscillations are also evident at the Buchillon station (Figure 4.5) from in-situ measurements located some 10 km away from the main upwelling zone. There, LSWT drops on June 29th immediately followed by a temperature rise both at the surface and at 35 m depth due to gravitational adjustment through the propagation of internal Kelvin waves (Bouffard and Lemmin, 2013). Such full upwelling did not only affect the lake but also the downstream water as evidenced by the measured and modelled temperature at the outlet of Lake Geneva (Figure 4.8). Both model and observation showed a 12°C drop in temperature in the downstream water over 2 days.

The consequences of upwellings for the lake ecosystem and the downstream water remain poorly investigated. One reason is the difficulty to conduct specific field measurements to track transient processes such as the illustrated event. Open access forecasting system such as *Meteolakes* can help planning field measurements and thereby improve the understanding of such processes.

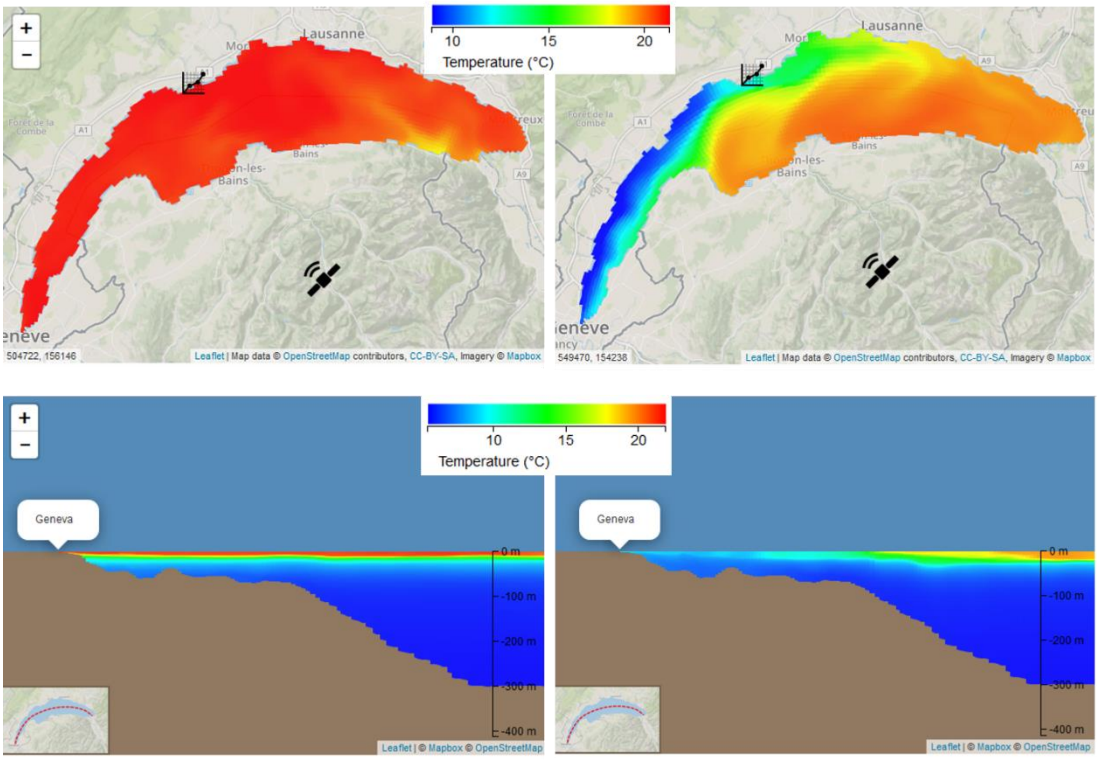


Figure 4.7: Upwelling in late June 2017 as displayed by *meteolakes.ch*. Left column on 28th June 2017 at 15h00, right column on 30th June 2017 at 9h00. The upper row shows the surface temperature and the lower row temperatures along a transect centred on the western basin. The graph icon (upper row) corresponds to the location of the in-situ station Buchillon.

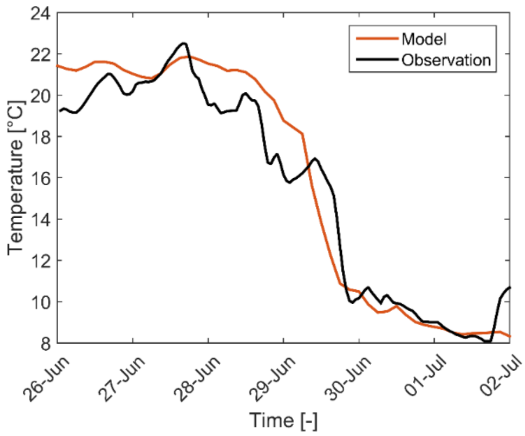


Figure 4.8: Time-series of the late June 2017 upwelling event. River Rhône temperature from the lake outflow in Geneva (black line) is compared to the temperature of a near-outlet model surface grid point (red line).

Storm-induced lake currents – Lake forecasting systems are also relevant for predicting the effects of extreme storms on the lake dynamics. The sudden high-speed winds event from 6th August 2018 is an illustrative example for the interest of such predictive approach. Figure 4.9 illustrates the surface currents as a result of a spatially localized North-Eastern coastal wind event. Timing was critical as the storm lasted less than three hours, as shown by the plots before (upper-left), during (main plot) and after (upper right) the event. Surface flows three hours before the event were in the range of 0.2 to 0.3 m/s. During peak intensity at 18h00, currents reached 0.8 m/s in the red patch of Figure 4.9, before returning to 0.2 to 0.3 m/s three hours later. The inset in Figure 4.9 shows the devastating wind waves in Vevey Harbour on the North-Eastern shore (Figure 4.2).

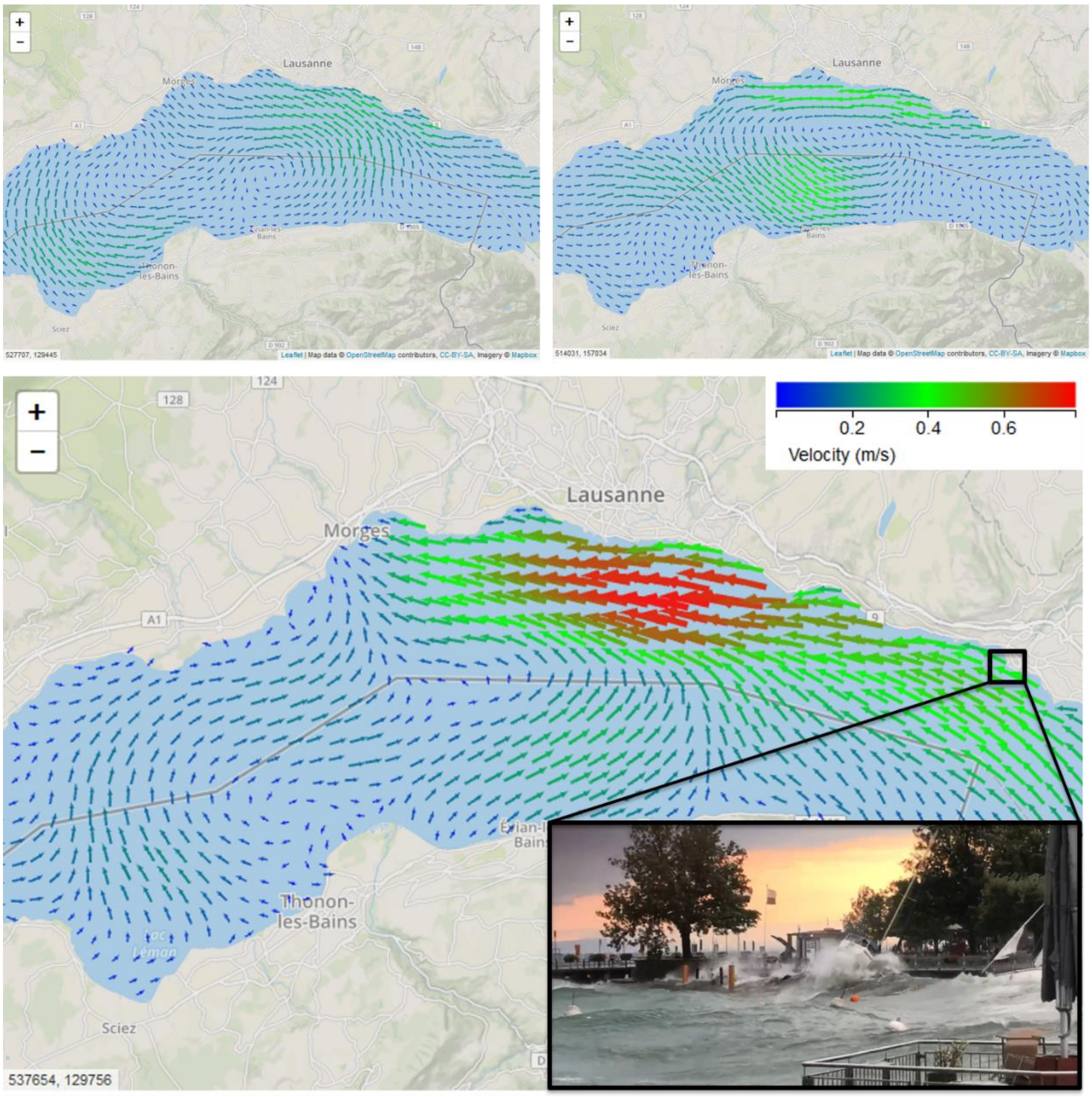


Figure 4.9: Surface water velocity field of the main basin from *meteolakes.ch* for the 6th August 2018 at 15h00 (upper left), 18h00 (lower plot) and 21h00 (upper right). The inset image (extracted from a footage by Dominique Fabien Rimaz) shows two boats smashed against the shore in Vevey Harbour (Figure 4.2) during the event.

4.4 Discussion

Platform acceptance – Since its creation in summer 2016, *Meteolakes* has provided lake information to more than 110'000 users. Its daily frequentation keeps growing, with a 2018 summer daily average of 330+ visitors. The platform demonstrated its effectiveness in disseminating both three-dimensional model results and lake observations from satellites and in-situ sensors. Analysis of logging data, showed that 51 % of logs come from smartphones, 41 % from computers and 8 % from tablets with 78 % of the traffic originating from Switzerland. This device log distribution indicates that we created a successfully scalable and responsive interface to distribute lake data. Among those visits, we found that the majority are citizens interested in lake temperature and currents for recreational activities (navigation, swimming) or for discovering how dynamic lake ecosystems are. Lake professionals, such as fishermen, contribute also to a regular stream of visits, independent of the season. They are interested in zones with strong temperature gradients, which are expected to have higher nutrient loads, and possibly more fishes (Olin et al., 2002). Beach managers are displaying lake temperature at specific coordinates on their website. Drinking water intakes operators for assessing risks of harmful algae contamination (Soontiens et al., 2018) by looking at the thermocline position. Such a model allows for identifying areas warranting the deployment of in-situ sensors (Baschek et al., 2017). In that regard, *Meteolakes* benefited various research groups around Lake Geneva: the system provided *a priori* knowledge for deployments of a Rockland Scientific

Slocum glider to study the horizontal variability resulting from gyre structures (Forrest et al., 2018), and helped in identifying the source and the spatial distribution of a calcite precipitation event, which has been observed by satellite (Nouchi et al., 2019). For the latter, model results combined with the implemented particle tracker found that the event originated from the Rhône River inflow. We also believe that lake physical information can be fed back to weather models. It is known that surface processes determine lower atmosphere conditions (Entekhabi et al., 1999; Koster and Suarez, 1995; Shukla and Mintz, 1982). This is particularly true for large lakes, where sensible and latent heat fluxes are expected to influence the local atmospheric boundary layer and where only few or no field observations are available to constrain weather models. The accurate characterization of the lower atmosphere above lakes is critical in some cases, such as in close proximity of an airport, as on Lake Geneva. Such integrated approaches, using coupled environmental model components, are key to more accurate prediction and warning of severe weather impacts (Lewis et al., 2017). Finally, *Meteolakes* could be used as impacts forecasting tool for downstream activities.

Integration of diverse data sources – We found that the combination of multiple data sources is of prime importance to assess the large variety of processes observed in lakes hydrodynamics. Real-time in-situ measurements provide a direct observation of the system, but their spatial extent is severely limited. They also enhance the processing of remote sensing data by providing data on critical atmospheric and lake variables, which can affect space-borne observations (e.g. wind speed, humidity). Real-time remote sensing observations of skin LSWT, AVHRR in this case, cover the entire surface of the basin and allow a validation of the model deviations. However their temporal resolution is limited, particularly in regions with frequent cloud coverage and they lack information in the vertical dimension. For *Meteolakes*, we observed on average deviations of real-time modelled surface bulk temperature from observed skin temperature in the range of ± 1.5 °C. We consider those deviations small, particularly when considering that a skin-to-bulk conversion is not applied. Most importantly, we found that surface patterns with significant spatial heterogeneities are well resolved by the model. This also indicates that AVHRR LSWT is an adequate source of information to monitor the surface temperature of Lake Geneva (Oesch et al., 2008; Chapter 3 of this thesis). Finally, three-dimensional model simulations are the only information source capable to solve the processes occurring at the large range of spatio-temporal scales in lakes dynamics, but they require the two former sources of information to be calibrated and to have their deviations constrained. In that regard, we found the Delft3D-FLOW model of Lake Geneva, calibrated in the second chapter, accurate when run in real-time. Observed deviations were in the usual range as those in the aforementioned chapter, which used an extensive dataset of 90 temperature and 4 current speed profiles to calibrate it over a period of two years (mean absolute error of 1.1 °C and 2.5 cm/s). Since the introduction of the platform in 2016, this model enabled the real-time study and forecasting of various physical phenomena. We illustrated two of them in this study.

Upwelling events – Upwellings are physical processes that have typically been elusive due to the lack of adequate monitoring tools for mesoscale processes. This is mainly a consequence of the transient nature of such events, which make them difficult to observe through in-situ campaigns without *a priori* knowledge on their spatial and temporal extent. Real-time three-dimensional hydrodynamic modelling opens new opportunities for monitoring the magnitude and spatial extent of those events. The upwelling of 29th June 2017 has been of strong intensity, which resulted in its featuring in various local newspapers. We thereby assessed how *Meteolakes* is capable to capture such events and found that the real-time model simulations displayed online were in good agreement with in-situ measurements. The in-situ station at Buchillon, measuring temperature at two depths in real-time, showed that both thermistor records (at 1 m and 35 m depth) were close to overlapping (Figure 4.5). This indicates a well-mixed water column. Moreover, in-situ temperature measurements in the Rhône River, taken in close proximity to the outlet of the lake and being indicative of the lake surface temperature in the Geneva region, showed a good match with the modelled temperature from a neighbouring model grid point (Figure 4.8). The intensity and timing of the upwelling were well resolved by the model. A drop of ~11 °C in a single day is observed in both data sources. Higher frequency variations in the measured river temperature time-series (black line of Figure 4.8) are not present in the modelled temperature. This is most likely a result of the output frequency available online (three hours, while measurements are hourly) and riverine heterogeneities. Clear day-night cycles can be found in the river data, indicating the signature of other processes (the station is located ~1 km downstream from the lake). Looking at the meteorological forcing, which created the upwelling event, we found that it is the result of relatively long-lasting winds blowing from South-West.

Overall, voluminous upwellings typically occur when the stratification is weak, in spring or autumn. The proper quantification of such events is of prime importance for a number of societally relevant aspects (e.g. water quality, algal blooms, heat extraction and discharge, etc.). Practically, with the *Meteolakes* platform, a diagnostic system to automatically detect and warn user was developed. The method is based on a k-means clustering method, splitting the lake surface temperature into two clusters. When the centroid difference of those cluster is larger than 4 °C, and alert is triggered. In the case of Lake Geneva, the early-warning system developed allowed us (and the subscribers) to closely monitor such events during the course of two years (2017 and 2018), by being warned about possible occurrences up to four days in advance. Using a similar approach (i.e. comparing a lake model grid point with outlet river temperature measurements), we found that the few upwellings that occurred during 2017-2018 have been well forecasted by *Meteolakes*. For Lake Geneva, when such events occur in the western basin, forecasting its magnitude is of particular interest as temperature is dropped for millions of downstream residents.

Storm events – An extreme storm of rare occurrence, with winds exceeding 25 m/s, created significant damage in North-Eastern coastal areas of Lake Geneva in the evening of 6th August 2018. The event has been relatively localized, with only the North-Eastern shore being impacted. Numerous rescue operations have been conducted because of the difficult navigation conditions. Surface current speeds reached values up to 0.8 m/s as visible on *Meteolakes* (Figure 4.9), a rare situation for Lake Geneva. While we were not able to deploy instruments to quantitatively assess the event, presence in close proximity and *a posteriori* analysis of online content (local news, weather bulletins) and discussion with rescue teams, indicated that *Meteolakes* captured both the critical timing and exceptional intensity of the event. We therefore believe that such platform can be used as preventive measure to mitigate the heavy economical and potential human losses extreme weather events like this can create. This is of particular importance considering that the frequency and severity of extreme weather are increasing with climate change (Coumou and Rahmstorf, 2012; Hansen et al., 2012).

Two physical phenomena are described in this study, yet other processes with strong implications on the biogeochemistry of the lake are well visible on the platform. For instance, gyres have been found to structure the lateral dispersion of primary productivity (Soomets et al., 2019) as a result of the Coriolis force and subsequent strong up-/down lifts of the thermocline. Those structures are repeatedly observed in Lake Geneva (Bouffard et al., 2018; Kiefer et al., 2015). We refer the reader to *meteolakes.ch* in order to visualize two of those events: (1) November 3rd, 2015 (week 45 of 2015 on *Meteolakes*), and (2) December 3rd, 2018 (week 48 of 2018), visible in Figure 4.4. In that regard, we found that these recurrent gyre systems are the response of a prevalent North-Eastern wind and that they decay within weeks after those wind patterns dissipate.

Perspectives – The implementation of new lakes aside, future developments are primarily focused on adding two key functionalities: water quality information and forecasts uncertainties. Water quality modelling requires the development of an ecological model building on the results of the hydrodynamics. This includes modelling and forecasting nutrient loads from rivers but also waste water treatment plants all around the lake. In that regard, a prototype of this functionality is already available online as experimental feature for Lake Geneva, under the "Water Quality" tab of the interface (Figure 4.4). Hypoxia and algal bloom detection by the early-warning system (supervisor) are also supported in the experimental version.

Forecast uncertainties and more robust hindcasts are achieved by implementing a data assimilative operational model. Upgrading operational systems is needed to make optimal use of satellite-based surface information (Drusch et al., 2009) and in-situ data. The application of advanced data assimilation techniques in the operational context with data from various sources across multiple spatio-temporal scales is rare (van Velzen and Verlaan, 2007). It is however needed to effectively quantify and reduce the uncertainties of products capable of providing actionable guidance to enable risk-based decision-making (Coccia and Todini, 2011; Pappenberger et al., 2007, 2008; Thielen et al., 2009; Weerts et al., 2011). In that regard, a DA experiment directly transferable to *Meteolakes* has already been designed and successfully applied to Lake Geneva in the third chapter of this thesis. Chapter 3 has been built with a particular focus on near real-time operations, and is tailored to the data sources presented here and their uncertainties. Its implementation towards an operational status for *Meteolakes* is ongoing.

4.5 Conclusion

Lakes are undergoing various stresses, both from global and local changes, seriously compromising the ecosystem services they provide. Additionally, considering that a large part of humanity lives near freshwater bodies (Kummu et al., 2011), lakes reactions to extreme weather events can have devastating economical and human costs. New monitoring and forecasting tools, covering wide ranges of statio-temporal scales, are thereby of prime importance for both policy-/decision-making and scientific communities. The coverage of those scales is only achieved by the combination and timely distribution of three information sources: (i) in-situ measurements, (ii) remote sensing observations and (iii) model simulations.

To address this issue, we developed *Meteolakes*: an online platform capable of providing near real-time and short-term forecasts of lake hydrodynamic properties for lakes. For more than two years, *Meteolakes* has provided spatio-temporal lake temperature and currents, which are important determinants of water quality, to more than 110'000 visitors and has been featured in numerous local media, public events, and museum exhibitions. Our goal has been to bridge the gap between operational limnology and the various interest groups on lakes. This has been achieved thanks to the development of a comprehensive and responsive online interface capable of openly distributing data, model results and data products. This interface and its technology are scalable to a broad range of devices and can be accessed at *meteolakes.ch*.

In this study, we documented the technical design and functioning of such integrated data-model platform, and presented the products it generates. We found that the platform is capable of revealing and forecasting mesoscale phenomena, such as upwellings and localized fast surface currents following a storm event. More than the high scientific value, the anticipation of those processes has a direct practical and economical relevance for the effective management of lakes.

Meteolakes has set good foundations to provide representative real-time monitoring and forecasts of lakes hydrodynamic conditions. The focus for future work is on the optimal combination of data and models, and integrated bio-physical modelling approaches. The vision is therefore for a future lake operational system providing robust high-resolution system with forecast uncertainties and correction of model deviations for monitoring the hydrodynamics and water quality of inland waters. At the current stage, *Meteolakes* provides a new tool with well-established foundations for research and pre-operational applications expandable to the majority of Swiss lakes, and beyond.

Authors contribution

TB, DB and AW designed the platform and contributed to its dissemination. TB implemented and maintained the system operational since 2016. TB prepared the manuscript with contributions from all co-authors and DB and AW supported editing the publication.

Acknowledgements

Data - The authors are grateful to the different institutions that provided the data used in this paper: the Federal Office of Meteorology and Climatology (MeteoSwiss) for meteorological data, the Federal Office of the Environment (FOEN), for the river data and water levels of Lake Geneva and the Oeschger Centre of the University of Bern for the AVHRR data. All this data is provided in real-time by those organisations and are not the property of the authors of this study.

The authors would like to thank Guillaume Ulrich, Stephane Restani, Roman Zoller, Martin Zoller and Daniel Forte-Marques for their helpful contributions to the development of *Meteolakes*. We acknowledge Sébastien Lavanchy, for his work on the Buchillon mast infrastructure and maintenance. Finally, we would like to thank the users for the outstanding array of feedback we received and their helpful suggestions. It strongly contributed to the expansion and maintenance of *Meteolakes*. With this article, we want to honour our young colleague and dear friend, Adrien Gaudard, who was involved in an earlier development of the hydrodynamic model for Lake Geneva, and who unexpectedly passed away during preparation of this publication. This project was supported by the European Space Agency's Scientific Exploitation of Operational Missions element (CORESIM contract No.: AO/1-8216/15/I-SBo).

References

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. and Winder, M.: Lakes as sentinels of climate change, Limnology and Oceanography, 54(6part2), 2283–2297, doi:10.4319/lo.2009.54.6_part_2.2283, 2009.

Baschek, B., Schroeder, F., Brix, H., Riethmüller, R., Badewien, T. H., Breitbach, G., Brügge, B., Colijn, F., Doerffer, R., Eschenbach, C., Friedrich, J., Fischer, P., Garthe, S., Horstmann, J., Krasemann, H., Metfies, K., Merckelbach, L., Ohle, N., Petersen, W., Pröfrock, D., Röttgers, R., Schlüter, M., Schulz, J., Schulz-Stellenfleth, J., Stanev, E., Staneva, J., Winter, C., Wirtz, K., Wollschläger, J., Zielinski, O. and Ziemer, F.: The Coastal Observing System for Northern and Arctic Seas (COSYNA), Ocean Science, 13(3), 379–410, doi:10.5194/os-13-379-2017, 2017.

Bouffard, D. and Lemmin, U.: Kelvin waves in Lake Geneva, Journal of Great Lakes Research, 39(4), 637–645, doi:10.1016/j.jglr.2013.09.005, 2013.

Bouffard, D., Kiefer, I., Wüest, A., Wunderle, S. and Odermatt, D.: Are surface temperature and chlorophyll in a large deep lake related? An analysis based on satellite observations in synergy with hydrodynamic modelling and in-situ data, Remote Sensing of Environment, 209, 510–523, doi:10.1016/j.rse.2018.02.056, 2018.

Chu, P. Y., Kelley, J. G. W., Mott, G. V., Zhang, A. and Lang, G. A.: Development, implementation, and skill assessment of the NOAA/NOS Great Lakes Operational Forecast System, Ocean Dynamics, 61(9), 1305–1316, doi:10.1007/s10236-011-0424-5, 2011.

Coccia, G. and Todini, E.: Recent developments in predictive uncertainty assessment based on the model conditional processor approach, Hydrology and Earth System Sciences, 15(10), 3253–3274, doi:10.5194/hess-15-3253-2011, 2011.

Coumou, D. and Rahmstorf, S.: A decade of weather extremes, Nature Climate Change, 2, 491, 2012.

Deltares: Delft3D-FLOW User Manual: Simulation of multi-dimensional hydrodynamic flows and transport phenomena, [online] Available from: https://oss.deltares.nl/documents/183920/185723/Delft3D-FLOW_User_Manual.pdf, 2015.

Drusch, M., Scipal, K., de Rosnay, P., Balsamo, G., Andersson, E., Bougeault, P. and Viterbo, P.: Towards a Kalman Filter based soil moisture analysis system for the operational ECMWF Integrated Forecast System, Geophysical Research Letters, 36(10), doi:10.1029/2009GL037716, 2009.

Forrest, A., Sepúlveda Steiner, O. and Wüest, A.: An underwater glider for measuring turbulence in Lake Geneva, actuEPFL [online] Available from: https://actu.epfl.ch/news/an-underwater-glider-for-measuring-turbulence-in-6/, 2018.

Gaudard, A., Schwefel, R., Vinnå, L. R., Schmid, M., Wüest, A. and Bouffard, D.: Optimizing the parameterization of deep mixing and internal seiches in one-dimensional hydrodynamic models: a case study with Simstrat v1.3, Geoscientific Model Development, 10(9), 3411–3423, doi:10.5194/gmd-10-3411-2017, 2017.

Goudsmit, G.-H., Burchard, H., Peeters, F. and Wüest, A.: Application of k-ϵ turbulence models to enclosed basins: The role of internal seiches, Journal of Geophysical Research: Oceans, 107(C12), 3230, doi:10.1029/2001JC000954, 2002.

Hansen, J., Sato, M. and Ruedy, R.: Perception of climate change, Proceedings of the National Academy of Sciences, 109(37), E2415–E2423, doi:10.1073/pnas.1205276109, 2012.

Hering, D., Carvalho, L., Argillier, C., Beklioglu, M., Borja, A., Cardoso, A. C., Duel, H., Ferreira, T., Globevnik, L., Hanganu, J., Hellsten, S., Jeppesen, E., Kodeš, V., Solheim, A. L., Nõges, T., Ormerod, S., Panagopoulos, Y., Schmutz, S., Venohr, M. and Birk, S.: Managing aquatic ecosystems and water resources under multiple stress — An introduction to the MARS project, Science of The Total Environment, 503–504, 10–21, doi:10.1016/j.scitotenv.2014.06.106, 2015.

Kamarainen, A. M., Yuan, H., Wu, C. H. and Carpenter, S. R.: Estimates of phosphorus entrainment in Lake Mendota: a comparison of one-dimensional and three-dimensional approaches: Approaches for estimating entrainment, Limnology and Oceanography: Methods, 7(7), 553–567, doi:10.4319/lom.2009.7.553, 2009.

Kiefer, I., Odermatt, D., Anneville, O., Wüest, A. and Bouffard, D.: Application of remote sensing for the optimization of in-situ sampling for monitoring of phytoplankton abundance in a large lake, Science of The Total Environment, 527–528, 493–506, doi:10.1016/j.scitotenv.2015.05.011, 2015.

Kimura, N. and Wu, C. H.: Using a nowcasting system to better understand lake water dynamics, Lakes & Reservoirs: Research & Management, 23(4), 367–380, doi:10.1111/lre.12239, 2018.

Kirillin, G., Hochschild, J., Mironov, D., Terzhevik, A., Golosov, S. and Nützmann, G.: FLake-Global: Online lake model with worldwide coverage, Environmental Modelling & Software, 26(5), 683–684, doi:10.1016/j.envsoft.2010.12.004, 2011.

Kummu, M., de Moel, H., Ward, P. J. and Varis, O.: How close do we live to water? In: A global analysis of population distance to freshwater bodies, edited by M. Perc, PLoS ONE, 6(6), e20578, doi:10.1371/journal.pone.0020578, 2011.

Lahoz, W., Khattatov, B. and Menard, R., Eds.: Data Assimilation, Springer Berlin Heidelberg, Berlin, Heidelberg., 2010.

Lewis, H. W., Castillo Sanchez, J. M., Graham, J., Saulter, A., Bornemann, J., Arnold, A., Fallmann, J., Harris, C., Pearson, D., Ramsdale, S., Martínez de la Torre, A., Bricheno, L., Blyth, E., Bell, V., Davies, H., Marthews, T. R., O'Neill, C., Rumbold, H., O'Dea, E., Brereton, A., Guihou, K., Hines, A., Butenschon, M., Dadson, S. J., Palmer, T., Holt, J., Reynard, N., Best, M., Edwards, J. and Siddorn, J.: The UKC2 regional coupled environmental prediction system, Geoscientific Model Development Discussions, 11(1), 1–42, doi:10.5194/gmd-11-1-2018, 2018.

Lieberherr, G. and Wunderle, S.: Lake surface water temperature derived from 35 years of AVHRR sensor data for European lakes, Remote Sensing, 10(7), 990, doi:10.3390/rs10070990, 2018.

Lieberherr, G., Riffler, M. and Wunderle, S.: Performance assessment of tailored split-window coefficients for the retrieval of lake surface water temperature from AVHRR satellite data, Remote Sensing, 9(12), 1334, doi:10.3390/rs9121334, 2017.

MacIntyre, S. and Melack, J. M.: Vertical and horizontal transport in lakes: linking littoral, benthic, and pelagic habitats, Journal of the North American Benthological Society, 14(4), 599–615, doi:10.2307/1467544, 1995.

Marques, C. A. F., Ferreira, J. A., Rocha, A., Castanheira, J. M., Melo-Gonçalves, P., Vaz, N. and Dias, J. M.: Singular spectrum analysis and forecasting of hydrological time series, Physics and Chemistry of the Earth, Parts A/B/C, 31(18), 1172–1179, doi:10.1016/j.pce.2006.02.061, 2006.

MeteoSwiss: COSMO-1 – high-resolution forecasts for the Alpine region, [online] Available from: https://www.meteoswiss.admin.ch/home/measurement-and-forecasting-systems/warning-and-forecasting-systems/cosmo-forecasting-system/cosmo-1-high-resolution-forecasts-for-the-alpine-region.html (Accessed 21 February 2019a), 2019.

MeteoSwiss: COSMO-E – probabilistic forecasts for the Alpine region, [online] Available from: https://www.meteoswiss.admin.ch/home/measurement-and-forecasting-systems/warning-and-forecasting-systems/cosmo-forecasting-system/cosmo-e-probabilistic-forecasts-for-the-alpine-region.html (Accessed 21 February 2019b), 2019.

Nouchi, V., Kutser, T., Wüest, A., Müller, B., Odermatt, D., Baracchini, T. and Bouffard, D.: Resolving biogeochemical processes in lakes using remote sensing, Aquatic Sciences, 81(2), 27, doi:10.1007/s00027-019-0626-3, 2019.

Oesch, D., Jaquet, J. -M., Klaus, R. and Schenker, P.: Multi-scale thermal pattern monitoring of a large lake (Lake Geneva) using a multi-sensor approach, International Journal of Remote Sensing, 29(20), 5785–5808, doi:10.1080/01431160802132786, 2008.

Olin, M., Rask, M., Ruuhljärvi, J., Kurkilahti, M., Ala-Opas, P. and Ylönen, O.: Fish community structure in mesotrophic and eutrophic lakes of southern Finland: the relative abundances of percids and cyprinids along a trophic gradient, Journal of Fish Biology, 60(3), 593–612, doi:10.1111/j.1095-8649.2002.tb01687.x, 2002.

Pappenberger, F., Beven, K., Frodsham, K., Romanowicz, R. and Matgen, P.: Grasping the unavoidable subjectivity in calibration of flood inundation models: A vulnerability weighted approach, Journal of Hydrology, 333(2–4), 275–287, doi:10.1016/j.jhydrol.2006.08.017, 2007.

Pappenberger, F., Bartholmes, J., Thielen, J., Cloke, H. L., Buizza, R. and de Roo, A.: New dimensions in early flood warning across the globe using grand-ensemble weather predictions, Geophysical Research Letters, 35(10), L10 404, doi:10.1029/2008GL033837, 2008.

Peeters, F., Straile, D., Lorke, A. and Ollinger, D.: Turbulent mixing and phytoplankton spring bloom development in a deep lake, Limnology and Oceanography, 52(1), 286–298, doi:10.4319/lo.2007.52.1.0286, 2007.

Qi, L., Hu, C., Duan, H., Barnes, B. and Ma, R.: An EOF-based algorithm to estimate chlorophyll-a concentrations in Taihu Lake from MODIS land-band measurements: implications for near real-time applications and forecasting models, Remote Sensing, 6(11), 10694–10715, doi:10.3390/rs61110694, 2014.

Read, J. S., Winslow, L. A., Hansen, G. J. A., Van Den Hoek, J., Hanson, P. C., Bruce, L. C. and Markfort, C. D.: Simulating 2368 temperate lakes reveals weak coherence in stratification phenology, Ecological Modelling, 291, 142–150, doi:10.1016/j.ecolmodel.2014.07.029, 2014.

Schwefel, R., Gaudard, A., Wüest, A. and Bouffard, D.: Effects of climate change on deepwater oxygen and winter mixing in a deep lake (Lake Geneva): Comparing observational findings and modeling, Water Resources Research, 52(11), 8811–8826, doi:10.1002/2016WR019194, 2016.

Soomets, T., Kutser, T., Wüest, A. and Bouffard, D.: Spatial and temporal changes of primary production in a deep peri-alpine lake, Inland Waters, 9: x1–x12, doi:10.1080/20442041.2018.1530529, 2019.

Soontiens, N., Binding, C., Fortin, V., Mackay, M. and Rao, Y. R.: Algal bloom transport in Lake Erie using remote sensing and hydrodynamic modelling: Sensitivity to buoyancy velocity and initial vertical distribution, Journal of Great Lakes Research, 45, doi:10.1016/j.jglr.2018.10.003, 2019.

Straile, D., Kerimoglu, O., Peeters, F., Jochimsen, M. C., Kümmerlin, R., Rinke, K. and Rothhaupt, K.-O.: Effects of a half a millennium winter on a deep lake - a shape of things to come?, Global Change Biology, 16(10), 2844–2856, doi:10.1111/j.1365-2486.2009.02158.x, 2010.

Thielen, J., Bartholmes, J., Ramos, M.-H. and de Roo, A.: The European Flood Alert System – Part 1: Concept and development, Hydrology and Earth System Sciences, 13(2), 125–140, doi:10.5194/hess-13-125-2009, 2009.

Toffolon, M. and Piccolroaz, S.: A hybrid model for river water temperature as a function of air temperature and discharge, Environmental Research Letters, 10(11), 114011, doi:10.1088/1748-9326/10/11/114011, 2015.

van Velzen, N. and Verlaan, M.: COSTA a problem solving environment for data assimilation applied for hydrodynamical modelling, Meteorologische Zeitschrift, 16(6), 777–793, doi:10.1127/0941-2948/2007/0241, 2007.

Vörösmarty, C. J., Hoekstra, A. Y., Bunn, S. E., Conway, D. and Gupta, J.: Fresh water goes global, Science, 349(6247), 478–479, doi:10.1126/science.aac6009, 2015.

Weerts, A. H., Winsemius, H. C. and Verkade, J. S.: Estimation of predictive hydrological uncertainty using quantile regression: examples from the National Flood Forecasting System (England and Wales), Hydrology and Earth System Sciences, 15(1), 255–265, doi:10.5194/hess-15-255-2011, 2011.

Chapter 5 Conclusion and outlook

V

What we have seen through this study…

5.1 Looking behind

Artists were right. Through colours and patterns, C. F. Ramuz, O. Kokoschka, and many others, captured true physical processes appearing at mesoscale. In the introduction to this thesis, we emphasized the importance of those processes, capable of affecting ecosystem services provided by lakes (e.g. drinking water supply, fish resources, heat supply and discharge, commercial navigation, recreational activities), and even citizens' safety (e.g. storm navigation hazards, harmful algae blooms). Consequently, a number of recent political measures aim at securing sustainably such essential services, at regional to global scale. Among those, the EU Water Framework Directive (2000/60/EC), the EU Drinking Water Directive (98/83/EC), EU Bathing Water Quality Directive (2006/7/EC), or more recently the United Nation's Post-2015 Development Agenda (UNEP/UN-Habitat/WHO, 2015; UN Water, 2015). These overarching policies have brought to the fore the need for new monitoring programs (Hering et al., 2015), using novel approaches like Earth-system simulations and remote sensing observations (Vörösmarty et al., 2015).

This thesis developed one such approach – from early scientific challenges to practical application, and beyond.

To provide insights on transient – yet impactful – mesoscale phenomena and their strong spatial variability, traditional in-situ measurements are not enough (Kiefer et al., 2015). Remote sensing observations addressed some of the spatial coverage limitations, but their temporal resolution is limited and they remain essentially two-dimensional. The only information source capable of resolving entirely the spatio-temporal scales involved in lake dynamics are three-dimensional hydrodynamic models. Such models, however, still rely on large observational datasets for their complex parameterizations and for constraining their uncertainties. Motivated by the weaknesses and lack of representativeness of each data source, this thesis aimed at providing reliable synoptic lake information, at high spatial and temporal resolution, by harnessing the combined potential of each of those sources.

The major outcome is an end-to-end framework for monitoring and forecasting lake dynamics. Using and developing open-source tools, special emphasis has been put on fostering the adoption of three-dimensional lake hydrodynamic models and improving their accuracy. This has been achieved in two stages: (i) by implementing an automated model calibration framework, and (ii) by developing a data assimilation (DA) scheme capable of reducing and quantifying model uncertainties by incorporating satellite lake surface temperature and in-situ temperature profiles. While the former facilitates a time-consuming part of model development - traditionally requiring expert knowledge -, the latter aims at optimally blending the three information sources for inland waters, trusting each of them based on their uncertainties. Both yielded remarkable results: the automated calibration returned parametric values diminishing models Root Mean Square Error by up to 47 %, while the DA cut it further down by 54 %. Furthermore, the DA experiment - the first of its kind for lakes - enhanced the spatial coherence and magnitude of imperfectly resolved physical processes, and provided system uncertainties. Homogeneous data density (in space and time) and dataset uncertainties has been identified as a main limitation of the calibration framework, while for DA it is the Gaussian assumption. The latter still provided an efficient variance minimizing solution without creating physical inconsistencies nor numerical instabilities. Finally, this thesis provided a practical outcome of its findings by deploying an online pre-operational three-dimensional lake monitoring and forecasting system: *meteolakes.ch*.

For more than two years, *Meteolakes* has provided spatially explicit real-time lake information and data products to more than hundred thousand end-users in Switzerland – Europe's water castle. This pioneering platform has been featured in numerous media (newspapers, radio, television), public events, and museum exhibitions. Heretofore, it benefited various communities, such as scientists for the timely planning of in-situ campaigns (e.g. glider deployment, Forrest et al., 2018) and study of transient lake processes (e.g. calcite precipitation, Nouchi et al., 2019). Its application programming interface allowed various beach managers to broadcast lake temperature at specific points of interest, fishermen to identify on a daily basis zones with strong temperature gradients, water intake operators and engineering companies to assess the position of the thermocline, etc. One of the most successful examples of this system, and thesis, has been demonstrated by its early-warning and forecasting capabilities, which anticipated numerous mesoscale physical phenomena in *le Léman*, such as upwellings, gyres, and high currents. Two of those events, which impacted public and commercial activities, are illustrated in this dissertation. Such capabilities, of predicting the outcome of an event to provide guidance in decision making, are not only one of the goals of environmental sciences, but also a touchstone of scientific knowledge and understanding (Bauer et al., 2015).

5.2 Looking beyond

In the prologue to this thesis, C. F. Ramuz described the lake as "a magnifying reflector of the light and heat of the star", which in a sense, is true. Likewise, lakes are reflectors of their watersheds, and therefore of human activities. Mirrors reflecting global changes, but also local anthropogenic ones: more than the sentinels of climate change (Adrian et al., 2009) alluded to in the introduction, they are sentinels of environmental change.

Consequently, future challenges to fully understand and mitigate reactions to external pressures, require integrated strategies. Coupled systems accounting for biological dynamics and human influences (e.g. nutrient loads, pollutants input), in light of delivering a complete bio-physical understanding of the system and ecological guidance. Some of the first steps towards such a coupled approach have been undertaken in this study, and *Meteolakes* is being augmented with a preliminary water quality model for forecasting algae blooms and hypoxic events.

In meteorological and oceanographic forecasting, current DA algorithm mostly rely on variational or hybrid (ensemble-variational) algorithms. The next decade is likely to see purely ensemble-based data assimilative systems, using, for instance, Ensemble Kalman Filtering (Bauer et al., 2015). This study developed such a method, tailored to lakes hydrodynamics and the constraints associated with the tight production schedules of operational systems. In *Meteolakes*, all data pipelines are operational and coherent data is available in real-time. The next step is in the near real-time implementation of the proposed DA framework, to provide lake forecast uncertainties and correct model deviations from satellite observations or in-situ field stations. These capabilities will elevate our approach at the forefront of environmental operational systems.

Increase in model spatial resolution is also expected to be seen in the future. Along with DA, those computationally expensive solutions require new approaches to computing. With Moore's law (the principle describing the exponential growth of computer performance since the 1970s) coming to an end (Waldrop, 2016), a change in paradigm regarding hardware is needed (Bauer et al., 2015). Largest gains are now to be found in code scalability to better exploit parallelism and heterogeneous chip architectures, the hardware solution to the software problem is obsolete. This study has been conducted with a focus on computational efficiency, and further hinted at critical locations where code and algorithmic improvements are to be made.

The present period is of fundamental importance for how inland waters monitoring and lake science will evolve. Building strong foundations and tools, this research paves the route for understanding lakes' delicate imbalance, thus opening new frontiers for interdisciplinary research on previously elusive lake physical processes, and their implications on the everyday life of people. The better understanding and anticipation of future environmental changes, and the mitigation of their effects, are some of the major challenges of our time. Ultimately producing data society can use, such elements and potential are well encompassed in this thesis.

From observations and models, to societal benefit, we presented here a long value chain for water management and for an often overlooked element of the global water cycle: lakes.

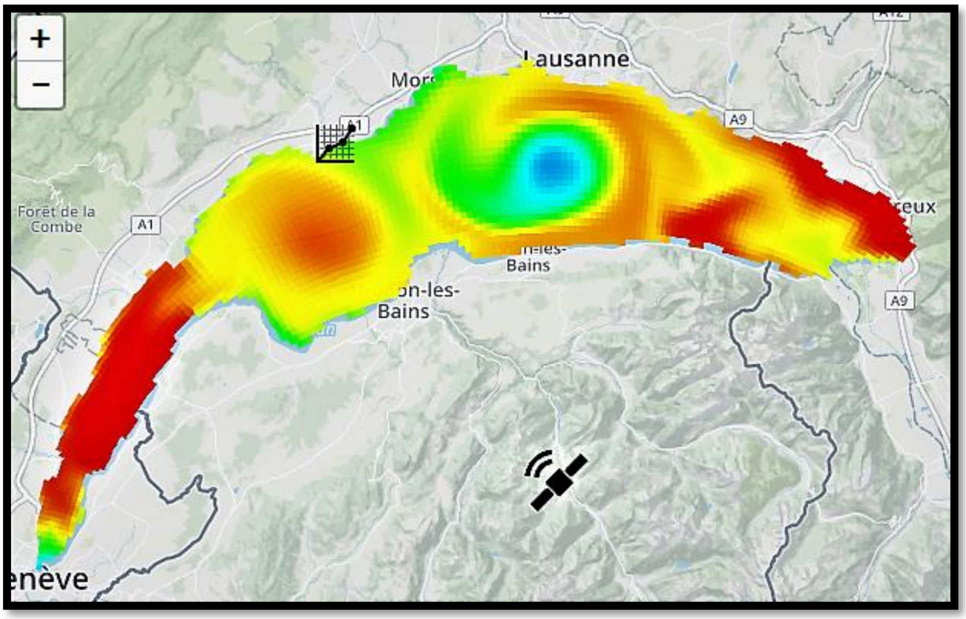
Epilogue

Lémancolie

What I now see, through the window…

As I write those words, I ask myself what I truly learned during those four years. Obviously, a lot. I think about all those times I have been looking through the window, at the lake, and I realize: I have never seen it twice the same. Perhaps a great gift this PhD offered me is better acuity, by an opening of the mind, at something I saw my entire life. Without taking away any of the magic and poetry from the lake it gave the ability to see that it is also scientifically impressive, beyond measure. The ability to see some spectacular, in the seemingly mundane.

I also think about the Meteolakes adventure. It has indeed been an adventure. The outstanding positive array of feedback I received from the people has been the true gift of this thesis to me. It made the many-evenings spent trying to unveil and share some of the mysteries of the lake – or rather fixing computers and bugs - worth it. In the end, I was the one to discover.



meteolakes.ch, 2018

References

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. and Winder, M.: Lakes as sentinels of climate change, Limnology and Oceanography, 54(6part2), 2283–2297, doi:10.4319/lo.2009.54.6_part_2.2283, 2009.

Afshar, A., Kazemi, H. and Saadatpour, M.: Particle swarm optimization for automatic calibration of large scale water quality model (CE-QUAL-W2): application to Karkheh Reservoir, Iran, Water Resources Management, 25(10), 2613–2632, doi:10.1007/s11269-011-9829-7, 2011.

Anderson, L. A., Robinson, A. R. and Lozano, C. J.: Physical and biological modeling in the Gulf Stream region, Deep Sea Research Part I: Oceanographic Research Papers, 47(10), 1787–1827, 2000.

Bannister, R. N.: A review of operational methods of variational and ensemble-variational data assimilation, Quarterly Journal of the Royal Meteorological Society, 143(703), 607–633, doi:10.1002/qj.2982, 2017.

Bard, Y.: Comparison of gradient methods for the solution of nonlinear parameter estimation problems, SIAM Journal on Numerical Analysis, 7(1), 157–186, doi:10.1137/0707011, 1970.

Bárdossy, A. and Singh, S. K.: Robust estimation of hydrological model parameters, Hydrology and Earth System Sciences, 12(6), 1273–1283, 2008.

Baschek, B., Schroeder, F., Brix, H., Riethmüller, R., Badewien, T. H., Breitbach, G., Brügge, B., Colijn, F., Doerffer, R., Eschenbach, C., Friedrich, J., Fischer, P., Garthe, S., Horstmann, J., Krasemann, H., Metfies, K., Merckelbach, L., Ohle, N., Petersen, W., Pröfrock, D., Röttgers, R., Schlüter, M., Schulz, J., Schulz-Stellenfleth, J., Stanev, E., Staneva, J., Winter, C., Wirtz, K., Wollschläger, J., Zielinski, O. and Ziemer, F.: The coastal observing system for northern and arctic seas (COSYNA), Ocean Science, 13(3), 379–410, doi:10.5194/os-13-379-2017, 2017.

Bauer, P., Thorpe, A. and Brunet, G.: The quiet revolution of numerical weather prediction, Nature, 525(7567), 47–55, doi:10.1038/nature14956, 2015.

Beaton, A. E. and Tukey, J. W.: The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data, Technometrics, 16(2), 147–185, doi:10.1080/00401706.1974.10489171, 1974.

Bergier, J.-F.: Léman, lac, in Dictionnaire historique de la Suisse (DHS). [online] Available from: http://www.hls-dhs-dss.ch/textes/f/F8657.php (Accessed 21 February 2019), 2009.

Bertino, L., Evensen, G. and Wackernagel, H.: Sequential data assimilation techniques in oceanography, International Statistical Review, 71(2), 223–241, doi:10.1111/j.1751-5823.2003.tb00194.x, 2007.

Bohle-Carbonell, M.: Currents in Lake Geneva, Limnology and Oceanography, 31(6), 1255–1266, doi:10.4319/lo.1986.31.6.1255, 1986.

Bouffard, D. and Lemmin, U.: Kelvin waves in Lake Geneva, Journal of Great Lakes Research, 39(4), 637–645, doi:10.1016/j.jglr.2013.09.005, 2013.

Bouffard, D. and Wüest, A.: Convection in lakes, Annual Review of Fluid Mechanics, 51(1), 189–215, doi:10.1146/annurev-fluid-010518-040506, 2019.

Bouffard, D., Boegman, L. and Rao, Y. R.: Poincaré wave-induced mixing in a large lake, Limnology and Oceanography, 57(4), 1201–1216, doi:10.4319/lo.2012.57.4.1201, 2012.

Bouffard, D., Kiefer, I., Wüest, A., Wunderle, S. and Odermatt, D.: Are surface temperature and chlorophyll in a large deep lake related? An analysis based on satellite observations in synergy with hydrodynamic modelling and in-situ data, Remote Sensing of Environment, 209, 510–523, doi:10.1016/j.rse.2018.02.056, 2018.

Box, M. J.: A comparison of several current optimization methods, and the use of transformations in constrained problems, The Computer Journal, 9(1), 67–77, doi:10.1093/comjnl/9.1.67, 1966.

Bradley, E. L.: The equivalence of maximum likelihood and weighted least squares estimates in the exponential family, Journal of the American Statistical Association, 68(341), 199–200, doi:10.1080/01621459.1973.10481364, 1973.

Cao, X.-Q., Zhang, W.-M., Song, J.-Q. and Zhang, L.-L.: Modeling background error covariance in variational data assimilation with wavelet method, in 2010 Third International Conference on Information and Computing, pp. 173–176, IEEE, Wuxi, TBD, China., 2010.

Carpenter, J., Clifford, P. and Fearnhead, P.: Improved particle filter for nonlinear problems, IEE Proceedings - Radar, Sonar and Navigation, 146(1), 2, doi:10.1049/ip-rsn:19990255, 1999.

Charnes, A., Frome, E. L. and Yu, P. L.: The equivalence of generalized least squares and maximum likelihood estimates in the exponential family, Journal of the American Statistical Association, 71(353), 169–171, doi:10.1080/01621459.1976.10481508, 1976.

Chu, P. Y., Kelley, J. G. W., Mott, G. V., Zhang, A. and Lang, G. A.: Development, implementation, and skill assessment of the NOAA/NOS Great Lakes Operational Forecast System, Ocean Dynamics, 61(9), 1305–1316, doi:10.1007/s10236-011-0424-5, 2011.

Cimatoribus, A., Lemmin, U. and Barry, D. A.: Velocity and temperature data from Lake Geneva (CH), collected during year 2016, [online] Available from: https://doi.org/10.5281/zenodo.1109794, 2017.

Coccia, G. and Todini, E.: Recent developments in predictive uncertainty assessment based on the model conditional processor approach, Hydrology and Earth System Sciences, 15(10), 3253–3274, doi:10.5194/hess-15-3253-2011, 2011.

Coumou, D. and Rahmstorf, S.: A decade of weather extremes, Nature Climate Change, 2, 491, 2012.

Crow, W. T.: Correcting land surface model predictions for the impact of temporally sparse rainfall rate measurements using an ensemble Kalman filter and surface brightness temperature observations, Journal of Hydrometeorology, 4(5), 960–973, doi:10.1175/1525-7541(2003)004<0960:CLSMPF>2.0.CO;2, 2003.

De Lannoy, G. J. M., Reichle, R. H., Houser, P. R., Pauwels, V. R. N. and Verhoest, N. E. C.: Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter, Water Resources Research, 43(9), doi:10.1029/2006WR005449, 2007a.

De Lannoy, G. J. M., Houser, P. R., Pauwels, V. R. N. and Verhoest, N. E. C.: State and bias estimation for soil moisture profiles by an ensemble Kalman filter: effect of assimilation depth and frequency, Water Resources Research, 43(6), W06401, doi:10.1029/2006WR005100, 2007b.

Dee, D. P. and Da Silva, A. M.: Data assimilation in the presence of forecast bias, Quarterly Journal of the Royal Meteorological Society, 124(545), 269–295, doi:10.1002/qj.49712454512, 1998.

Deltares: Delft3D-FLOW User Manual: Simulation of multi-dimensional hydrodynamic flows and transport phenomena, [online] Available from: https://oss.deltares.nl/documents/183920/185723/Delft3D-FLOW_User_Manual.pdf, 2015.

Deltares: The OpenDA Data-assimilation Toolbox, [online] Available from: www.openda.org, 2019.

Drusch, M., Scipal, K., de Rosnay, P., Balsamo, G., Andersson, E., Bougeault, P. and Viterbo, P.: Towards a Kalman filter based soil moisture analysis system for the operational ECMWF integrated forecast system, Geophysical Research Letters, 36(10), doi:10.1029/2009GL037716, 2009.

Eknes, M. and Evensen, G.: An ensemble Kalman filter with a 1-D marine ecosystem model, Journal of Marine Systems, 36(1–2), 75–100, doi:10.1016/S0924-7963(02)00134-3, 2002.

El Serafy, G. Y., Gerritsen, H., Hummel, S., Weerts, A. H., Mynett, A. E. and Tanaka, M.: Application of data assimilation in portable operational forecasting systems—the DATools assimilation environment, Ocean Dynamics, 57(4–5), 485–499, doi:10.1007/s10236-007-0124-3, 2007.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, Journal of Geophysical Research, 99(C5), 10143, doi:10.1029/94JC00572, 1994.

Evensen, G.: The ensemble Kalman filter: theoretical formulation and practical implementation, Ocean Dynamics, 53(4), 343–367, doi:10.1007/s10236-003-0036-9, 2003.

Evensen, G.: The ensemble Kalman filter for combined state and parameter estimation, IEEE Control Systems, 29(3), 83–104, doi:10.1109/MCS.2009.932223, 2009.

Fabio, P., Aronica, G. T. and Apel, H.: Towards automatic calibration of 2-D flood propagation models, Hydrology and Earth System Sciences, 14(6), 911–924, doi:10.5194/hess-14-911-2010, 2010.

Fenicia, F., Savenije, H. H. G., Matgen, P. and Pfister, L.: A comparison of alternative multiobjective calibration strategies for hydrological modeling, Water Resources Research, 43(3), doi:10.1029/2006WR005098, 2007.

Forel, F.-A.: Le Léman : monographie limnologique, F. Rouge, Lausanne., 1892.

Forrest, A., Sepúlveda Steiner, O. and Wüest, A.: An underwater glider for measuring turbulence in Lake Geneva, actuEPFL [online] Available from: https://actu.epfl.ch/news/an-underwater-glider-for-measuring-turbulence-in-6/, 2018.

Garcia, M., Ramirez, I., Verlaan, M. and Castillo, J.: Application of a three-dimensional hydrodynamic model for San Quintin Bay, B.C., Mexico. Validation and calibration using OpenDA, Journal of Computational and Applied Mathematics, 273, 428–437, doi:10.1016/j.cam.2014.05.003, 2015.

Gaudard, A., Schwefel, R., Vinnå, L. R., Schmid, M., Wüest, A. and Bouffard, D.: Optimizing the parameterization of deep mixing and internal seiches in one-dimensional hydrodynamic models: a case study with Simstrat v1.3, Geoscientific Model Development, 10(9), 3411–3423, doi:10.5194/gmd-10-3411-2017, 2017.

Gaudard, A., Weber, C., Alexander, T. J., Hunziker, S. and Schmid, M.: Impacts of using lakes and rivers for extraction and disposal of heat, Wiley Interdisciplinary Reviews: Water, 5(5), e1295, doi:10.1002/wat2.1295, 2018.

Gentemann, C. L., Donlon, C. J., Stuart-Menteth, A. and Wentz, F. J.: Diurnal signals in satellite sea surface temperature measurements, Geophysical Research Letters, 30(3), 1140, doi:10.1029/2002GL016291, 2003.

Gillijns, S., Mendoza, O. B., Chandrasekar, J., De Moor, B. L. R., Bernstein, D. S. and Ridley, A.: What is the ensemble Kalman filter and how well does it work?, in American Control Conference, 2006, pp. 6–pp, IEEE., 2006.

Goudsmit, G.-H., Burchard, H., Peeters, F. and Wüest, A.: Application of k-ϵ turbulence models to enclosed basins: The role of internal seiches, Journal of Geophysical Research: Oceans, 107(C12), 3230, doi:10.1029/2001JC000954, 2002.

Hansen, J., Sato, M. and Ruedy, R.: Perception of climate change, Proceedings of the National Academy of Sciences, 109(37), E2415–E2423, doi:10.1073/pnas.1205276109, 2012.

Hawley, N., Johengen, T. H., Rao, Y. R., Ruberg, S. A., Beletsky, D., Ludsin, S. A., Eadie, B. J., Schwab, D. J., Croley, T. E. and Brandt, S. B.: Lake Erie hypoxia prompts Canada-U.S. study, Eos, Transactions American Geophysical Union, 87(32), 313, doi:10.1029/2006EO320001, 2006.

Hendrickson, J. D., Sorooshian, S. and Brazil, L. E.: Comparison of Newton-type and direct search algorithms for calibration of conceptual rainfall-runoff models, Water Resources Research, 24(5), 691–700, doi:10.1029/WR024i005p00691, 1988.

Hering, D., Carvalho, L., Argillier, C., Beklioglu, M., Borja, A., Cardoso, A. C., Duel, H., Ferreira, T., Globevnik, L., Hanganu, J., Hellsten, S., Jeppesen, E., Kodeš, V., Solheim, A. L., Nõges, T., Ormerod, S., Panagopoulos, Y., Schmutz, S., Venohr, M. and Birk, S.: Managing aquatic ecosystems and water resources under multiple stress — An introduction to the MARS project, Science of The Total Environment, 503–504, 10–21, doi:10.1016/j.scitotenv.2014.06.106, 2015.

Hoel, H., Law, K. J. H. and Tempone, R.: Multilevel ensemble Kalman filtering, SIAM Journal on Numerical Analysis, 54(3), 1813–1839, doi:10.1137/15M100955X, 2016.

Houtekamer, P. L. and Mitchell, H. L.: A sequential ensemble Kalman filter for atmospheric data assimilation, Monthly Weather Review, 129(1), 123–137, doi:10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2, 2001.

Imberger, J. and Hamblin, P. F.: Dynamics of lakes, reservoirs, and cooling ponds, Annual Review of Fluid Mechanics, 14, 153–187, 1982.

Johnston, P. R. and Pilgrim, D. H.: Parameter optimization for watershed models, Water Resources Research, 12(3), 477–486, doi:10.1029/WR012i003p00477, 1976.

Kalman, R. E.: A new approach to linear filtering and prediction problems, Journal of Basic Engineering, 82(1), 35, doi:10.1115/1.3662552, 1960.

Kamarainen, A. M., Yuan, H., Wu, C. H. and Carpenter, S. R.: Estimates of phosphorus entrainment in Lake Mendota: a comparison of one-dimensional and three-dimensional approaches: Approaches for estimating entrainment, Limnology and Oceanography: Methods, 7(7), 553–567, doi:10.4319/lom.2009.7.553, 2009.

Kiefer, I., Odermatt, D., Anneville, O., Wüest, A. and Bouffard, D.: Application of remote sensing for the optimization of in-situ sampling for monitoring of phytoplankton abundance in a large lake, Science of The Total Environment, 527–528, 493–506, doi:10.1016/j.scitotenv.2015.05.011, 2015.

Kimura, N. and Wu, C. H.: Using a nowcasting system to better understand lake water dynamics, Lakes & Reservoirs: Research & Management, 23(4), 367–380, doi:10.1111/lre.12239, 2018.

Kirillin, G., Hochschild, J., Mironov, D., Terzhevik, A., Golosov, S. and Nützmann, G.: FLake-Global: Online lake model with worldwide coverage, Environmental Modelling & Software, 26(5), 683–684, doi:10.1016/j.envsoft.2010.12.004, 2011.

Kourzeneva, E.: Assimilation of lake water surface temperature observations using an extended Kalman filter, Tellus A: Dynamic Meteorology and Oceanography, 66(1), 21510, doi:10.3402/tellusa.v66.21510, 2014.

Kummu, M., de Moel, H., Ward, P. J. and Varis, O.: How close do we live to water? In: A global analysis of population distance to freshwater bodies, edited by M. Perc, PLoS ONE, 6(6), e20578, doi:10.1371/journal.pone.0020578, 2011.

Kuragano, T. and Kamachi, M.: Global statistical space-time scales of oceanic variability estimated from the TOPEX/POSEIDON altimeter data, Journal of Geophysical Research: Oceans, 105(C1), 955–974, doi:10.1029/1999JC900247, 2000.

Kurniawan, A., Ooi, S. K., Gerritsen, H. and Twigt, D.: Calibrating the regional tidal prediction of the Singapore Regional Model using OpenDA, in Proc. 9th Int. Conf on Hydroinformatics, Tianjin, pp. 7–11., 2010.

Kurniawan, A., Ooi, S. K., Hummel, S. and Gerritsen, H.: Sensitivity analysis of the tidal representation in Singapore Regional Waters in a data assimilation environment, Ocean Dynamics, 61(8), 1121–1136, doi:10.1007/s10236-011-0415-6, 2011.

Lahoz, W., Khattatov, B. and Menard, R., Eds.: Data Assimilation, Springer Berlin Heidelberg, Berlin, Heidelberg., 2010.

van Leeuwen, P. J.: Particle Filtering in geophysical systems, Monthly Weather Review, 137(12), 4089–4114, doi:10.1175/2009MWR2835.1, 2009.

Lemmin, U., Mortimer, C. H. and Bäuerle, E.: Internal seiche dynamics in Lake Geneva, Limnology and Oceanography, 50(1), 207–216, doi:10.4319/lo.2005.50.1.0207, 2005.

Lewis, H. W., Castillo Sanchez, J. M., Graham, J., Saulter, A., Bornemann, J., Arnold, A., Fallmann, J., Harris, C., Pearson, D., Ramsdale, S., Martínez de la Torre, A., Bricheno, L., Blyth, E., Bell, V., Davies, H., Marthews, T. R., O'Neill, C., Rumbold, H., O'Dea, E., Brereton, A., Guihou, K., Hines, A., Butenschon, M., Dadson, S. J., Palmer, T., Holt, J., Reynard, N., Best, M., Edwards, J. and Siddorn, J.: The UKC2 regional coupled environmental prediction system, Geoscientific Model Development, 11(1), 1–42, doi:10.5194/gmd-11-1-2018, 2018.

Li, Z., Chao, Y., McWilliams, J. C. and Ide, K.: A three-dimensional variational data assimilation scheme for the regional ocean modeling system, Journal of Atmospheric and Oceanic Technology, 25(11), 2074–2090, doi:10.1175/2008JTECHO594.1, 2008.

Lieberherr, G. and Wunderle, S.: Lake surface water temperature derived from 35 years of AVHRR sensor data for European lakes, Remote Sensing, 10(7), 990, doi:10.3390/rs10070990, 2018.

Lieberherr, G., Riffler, M. and Wunderle, S.: Performance assessment of tailored split-window coefficients for the retrieval of lake surface water temperature from AVHRR satellite data, Remote Sensing, 9(12), 1334, doi:10.3390/rs9121334, 2017.

Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., van Dijk, A. I. J. M., van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O. and Restrepo, P.: Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities, Hydrology and Earth System Sciences, 16(10), 3863–3887, doi:10.5194/hess-16-3863-2012, 2012.

MacIntyre, S. and Melack, J. M.: Vertical and horizontal transport in lakes: linking littoral, benthic, and pelagic habitats, Journal of the North American Benthological Society, 14(4), 599–615, doi:10.2307/1467544, 1995.

Madsen, H.: Automatic calibration of a conceptual rainfall–runoff model using multiple objectives, Journal of hydrology, 235(3), 276–288, 2000.

Madsen, H.: Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives, Advances in water resources, 26(2), 205–216, 2003.

Mao, J. Q., Lee, J. H. W. and Choi, K. W.: The extended Kalman filter for forecast of algal bloom dynamics, Water Research, 43(17), 4214–4224, doi:10.1016/j.watres.2009.06.012, 2009.

Marques, C. A. F., Ferreira, J. A., Rocha, A., Castanheira, J. M., Melo-Gonçalves, P., Vaz, N. and Dias, J. M.: Singular spectrum analysis and forecasting of hydrological time series, Physics and Chemistry of the Earth, Parts A/B/C, 31(18), 1172–1179, doi:10.1016/j.pce.2006.02.061, 2006.

MeteoSwiss: COSMO-1 – high-resolution forecasts for the Alpine region, [online] Available from: https://www.meteoswiss.admin.ch/home/measurement-and-forecasting-systems/warning-and-forecasting-systems/cosmo-forecasting-system/cosmo-1-high-resolution-forecasts-for-the-alpine-region.html (Accessed 21 February 2019a), 2019.

MeteoSwiss: COSMO-E – probabilistic forecasts for the Alpine region, [online] Available from: https://www.meteoswiss.admin.ch/home/measurement-and-forecasting-systems/warning-and-forecasting-systems/cosmo-forecasting-system/cosmo-e-probabilistic-forecasts-for-the-alpine-region.html (Accessed 21 February 2019b), 2019.

Minnett, P. J., Smith, M. and Ward, B.: Measurements of the oceanic thermal skin effect, Deep Sea Research Part II: Topical Studies in Oceanography, 58(6), 861–868, doi:10.1016/j.dsr2.2010.10.024, 2011.

Moradkhani, H., Hsu, K.-L., Gupta, H. and Sorooshian, S.: Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, Water Resources Research, 41(5), W05012, doi:10.1029/2004WR003604, 2005.

Nakamura, K., Higuchi, T. and Hirose, N.: Sequential data assimilation: information fusion of a numerical simulation and large scale observation data., J. UCS, 12(6), 608–626, 2006.

Natvik, L.-J. and Evensen, G.: Assimilation of ocean colour data into a biochemical model of the North Atlantic, Journal of Marine Systems, 40–41, 127–153, doi:10.1016/S0924-7963(03)00016-2, 2003.

Nelder, J. A. and Baker, R. J.: Generalized linear models, in Encyclopedia of Statistical Sciences, edited by S. Kotz, C. B. Read, N. Balakrishnan, B. Vidakovic, and N. L. Johnson, John Wiley & Sons, Inc., Hoboken, NJ, USA., 2006.

Nouchi, V., Kutser, T., Wüest, A., Müller, B., Odermatt, D., Baracchini, T. and Bouffard, D.: Resolving biogeochemical processes in lakes using remote sensing, Aquatic Sciences, 81(2), 27, doi:10.1007/s00027-019-0626-3, 2019.

Oesch, D., Jaquet, J. -M., Klaus, R. and Schenker, P.: Multi-scale thermal pattern monitoring of a large lake (Lake Geneva) using a multi-sensor approach, International Journal of Remote Sensing, 29(20), 5785–5808, doi:10.1080/01431160802132786, 2008.

Oesch, D. C., Jaquet, J.-M., Hauser, A. and Wunderle, S.: Lake surface water temperature retrieval using advanced very high resolution radiometer and Moderate Resolution Imaging Spectroradiometer data: Validation and feasibility study, Journal of Geophysical Research, 110(C12), doi:10.1029/2004JC002857, 2005.

Olin, M., Rask, M., Ruuhljärvi, J., Kurkilahti, M., Ala-Opas, P. and Ylönen, O.: Fish community structure in mesotrophic and eutrophic lakes of southern Finland: the relative abundances of percids and cyprinids along a trophic gradient, Journal of Fish Biology, 60(3), 593–612, doi:10.1111/j.1095-8649.2002.tb01687.x, 2002.

O'Reilly, C. M., Sharma, S., Gray, D. K., Hampton, S. E., Read, J. S., Rowley, R. J., Schneider, P., Lenters, J. D., McIntyre, P. B., Kraemer, B. M., Weyhenmeyer, G. A., Straile, D., Dong, B., Adrian, R., Allan, M. G., Anneville, O., Arvola, L., Austin, J., Bailey, J. L., Baron, J. S., Brookes, J. D., de Eyto, E., Dokulil, M. T., Hamilton, D. P., Havens, K., Hetherington, A. L., Higgins, S. N., Hook, S., Izmest'eva, L. R., Joehnk, K. D., Kangur, K., Kasprzak, P., Kumagai, M., Kuusisto, E., Leshkevich, G., Livingstone, D. M., MacIntyre, S., May, L., Melack, J. M., Mueller-Navarra, D. C., Naumenko, M., Noges, P., Noges, T., North, R. P., Plisnier, P.-D., Rigosi, A., Rimmer, A., Rogora, M., Rudstam, L. G., Rusak, J. A., Salmaso, N., Samal, N. R., Schindler, D. E., Schladow, S. G., Schmid, M., Schmidt, S. R., Silow, E., Soylu, M. E., Teubner, K., Verburg, P., Voutilainen, A., Watkinson, A., Williamson, C. E. and Zhang, G.: Rapid and highly variable warming of lake surface waters around the globe: global lake surface warming, Geophysical Research Letters, 42(24), 10,773-10,781, doi:10.1002/2015GL066235, 2015.

Pappenberger, F., Beven, K., Frodsham, K., Romanowicz, R. and Matgen, P.: Grasping the unavoidable subjectivity in calibration of flood inundation models: A vulnerability weighted approach, Journal of Hydrology, 333(2–4), 275–287, doi:10.1016/j.jhydrol.2006.08.017, 2007.

Pappenberger, F., Bartholmes, J., Thielen, J., Cloke, H. L., Buizza, R. and de Roo, A.: New dimensions in early flood warning across the globe using grand-ensemble weather predictions, Geophysical Research Letters, 35(10), L10 404, doi:10.1029/2008GL033837, 2008.

Peeters, F., Straile, D., Lorke, A. and Ollinger, D.: Turbulent mixing and phytoplankton spring bloom development in a deep lake, Limnology and Oceanography, 52(1), 286–298, doi:10.4319/lo.2007.52.1.0286, 2007.

Poincaré, H.: Science et méthode, Flammarion, Paris, 1908.

Qi, L., Hu, C., Duan, H., Barnes, B. and Ma, R.: An EOF-based algorithm to estimate chlorophyll-a concentrations in Taihu Lake from MODIS land-band measurements: implications for near real-time applications and forecasting models, Remote Sensing, 6(11), 10694–10715, doi:10.3390/rs61110694, 2014a.

Qi, L., Ma, R., Hu, W. and Loiselle, S. A.: Assimilation of MODIS chlorophyll-a data into a coupled hydrodynamic-biological model of Taihu Lake, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7(5), 1623–1631, doi:10.1109/JSTARS.2013.2280815, 2014b.

Ralston, M. L. and Jennrich, R. I.: Dud, a derivative-free algorithm for nonlinear least squares, Technometrics, 20(1), 7, doi:10.2307/1268154, 1978.

Rawlins, F., Ballard, S. P., Bovis, K. J., Clayton, A. M., Li, D., Inverarity, G. W., Lorenc, A. C. and Payne, T. J.: The Met Office global four-dimensional variational data assimilation scheme, Quarterly Journal of the Royal Meteorological Society, 133(623), 347–362, doi:10.1002/qj.32, 2007.

Read, J. S., Winslow, L. A., Hansen, G. J. A., Van Den Hoek, J., Hanson, P. C., Bruce, L. C. and Markfort, C. D.: Simulating 2368 temperate lakes reveals weak coherence in stratification phenology, Ecological Modelling, 291, 142–150, doi:10.1016/j.ecolmodel.2014.07.029, 2014.

Reichle, R. H., Walker, J. P., Koster, R. D. and Houser, P. R.: Extended versus ensemble Kalman filtering for land data assimilation, Journal of Hydrometeorology, 3(6), 728–740, doi:10.1175/1525-7541(2002)003<0728:EVEKFF>2.0.CO;2, 2002a.

Reichle, R. H., McLaughlin, D. B. and Entekhabi, D.: Hydrologic data assimilation with the ensemble Kalman filter, Monthly Weather Review, 130(1), 103–114, doi:10.1175/1520-0493(2002)130<0103:HDAWTE>2.0.CO;2, 2002b.

de Rosnay, P., Drusch, M., Vasiljevic, D., Balsamo, G., Albergel, C. and Isaksen, L.: A simplified extended Kalman filter for the global operational soil moisture analysis at ECMWF, Quarterly Journal of the Royal Meteorological Society, 139(674), 1199–1213, doi:10.1002/qj.2023, 2013.

Schwefel, R., Gaudard, A., Wüest, A. and Bouffard, D.: Effects of climate change on deepwater oxygen and winter mixing in a deep lake (Lake Geneva): Comparing observational findings and modeling, Water Resources Research, 52(11), 8811–8826, doi:10.1002/2016WR019194, 2016.

Shimoda, Y., Azim, M. E., Perhar, G., Ramin, M., Kenney, M. A., Sadraddini, S., Gudimov, A. and Arhonditsis, G. B.: Our current understanding of lake ecosystem response to climate change: What have we really learned from the north temperate deep lakes?, Journal of Great Lakes Research, 37(1), 173–193, doi:10.1016/j.jglr.2010.10.004, 2011.

Skahill, B. E. and Doherty, J.: Efficient accommodation of local minima in watershed model calibration, Journal of Hydrology, 329(1–2), 122–139, doi:10.1016/j.jhydrol.2006.02.005, 2006.

Solomatine, D. P., Dibike, Y. B. and Kukuric, N.: Automatic calibration of groundwater models using global optimization techniques, Hydrological Sciences Journal, 44(6), 879–894, doi:10.1080/02626669909492287, 1999.

Soomets, T., Kutser, T., Wüest, A. and Bouffard, D.: Spatial and temporal changes of primary production in a deep peri-alpine lake, Inland Waters, 9: x1-x12, doi:10.1080/20442041.2018.1530529, 2019.

Soontiens, N., Binding, C., Fortin, V., Mackay, M. and Rao, Y. R.: Algal bloom transport in Lake Erie using remote sensing and hydrodynamic modelling: Sensitivity to buoyancy velocity and initial vertical distribution, Journal of Great Lakes Research, 45, doi:10.1016/j.jglr.2018.10.003, 2019.

Sorooshian, S., Duan, Q. and Gupta, V. K.: Calibration of rainfall-runoff models: Application of global optimization to the Sacramento Soil Moisture Accounting Model, Water Resources Research, 29(4), 1185–1194, doi:10.1029/92WR02617, 1993.

Stanev, E. V., Schulz-Stellenfleth, J., Staneva, J., Grayek, S., Seemann, J. and Petersen, W.: Coastal observing and forecasting system for the German Bight, Ocean Science, 7(5), 569–583, doi:10.5194/os-7-569-2011, 2011.

Straile, D., Kerimoglu, O., Peeters, F., Jochimsen, M. C., Kümmerlin, R., Rinke, K. and Rothhaupt, K.-O.: Effects of a half a millennium winter on a deep lake - a shape of things to come?, Global Change Biology, 16(10), 2844–2856, doi:10.1111/j.1365-2486.2009.02158.x, 2010.

Stroud, J. R., Lesht, B. M., Schwab, D. J., Beletsky, D. and Stein, M. L.: Assimilation of satellite images into a sediment transport model of Lake Michigan, Water Resources Research, 45(2), W02419, doi:10.1029/2007WR006747, 2009.

Stroud, J. R., Stein, M. L., Lesht, B. M., Schwab, D. J. and Beletsky, D.: An ensemble Kalman filter and smoother for satellite data assimilation, Journal of the American Statistical Association, 105(491), 978–990, doi:10.1198/jasa.2010.ap07636, 2010.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, Journal of Geophysical Research: Atmospheres, 106(D7), 7183–7192, doi:10.1029/2000JD900719, 2001.

Thielen, J., Bartholmes, J., Ramos, M.-H. and de Roo, A.: The European flood alert system – Part 1: concept and development, Hydrology and Earth System Sciences, 13(2), 125–140, doi:10.5194/hess-13-125-2009, 2009.

Toffolon, M. and Piccolroaz, S.: A hybrid model for river water temperature as a function of air temperature and discharge, Environmental Research Letters, 10(11), 114011, doi:10.1088/1748-9326/10/11/114011, 2015.

Toffolon, M. and Rizzi, G.: Effects of spatial wind inhomogeneity and turbulence anisotropy on circulation in an elongated basin: A simplified analytical solution, Advances in Water Resources, 32(10), 1554–1566, doi:10.1016/j.advwatres.2009.08.001, 2009.

Umlauf, L. and Lemmin, U.: Interbasin exchange and mixing in the hypolimnion of a large lake: The role of long internal waves, Limnology and Oceanography, 50(5), 1601–1611, doi:10.4319/lo.2005.50.5.1601, 2005.

van Velzen, N. and Verlaan, M.: COSTA a problem solving environment for data assimilation applied for hydrodynamical modelling, Meteorologische Zeitschrift, 16(6), 777–793, doi:10.1127/0941-2948/2007/0241, 2007.

Verlaan, M. and Heemink, A. W.: Nonlinearity in data assimilation applications: a practical method for analysis, Monthly Weather Review, 129(6), 1578–1589, doi:10.1175/1520-0493(2001)129<1578:NIDAAA>2.0.CO;2, 2001.

Vörösmarty, C. J., Hoekstra, A. Y., Bunn, S. E., Conway, D. and Gupta, J.: Fresh water goes global, Science, 349(6247), 478–479, doi:10.1126/science.aac6009, 2015.

Wagner, C. and Adrian, R.: Exploring lake ecosystems: hierarchy responses to long-term change?, Global Change Biology, 15(5), 1104–1115, doi:10.1111/j.1365-2486.2008.01833.x, 2009.

Waldrop, M. M.: The chips are down for Moore's law, Nature, 530(7589), 144–147, doi:10.1038/530144a, 2016.

Weerts, A. H., El Serafy, G. Y., Hummel, S., Dhondia, J. and Gerritsen, H.: Application of generic data assimilation tools (DATools) for flood forecasting purposes, Computers & Geosciences, 36(4), 453–463, doi:10.1016/j.cageo.2009.07.009, 2010.

Weerts, A. H., Winsemius, H. C. and Verkade, J. S.: Estimation of predictive hydrological uncertainty using quantile regression: examples from the National Flood Forecasting System (England and Wales), Hydrology and Earth System Sciences, 15(1), 255–265, doi:10.5194/hess-15-255-2011, 2011.

Woolway, R. I., Simpson, J. H., Spiby, D., Feuchtmayr, H., Powell, B. and Maberly, S. C.: Physical and chemical impacts of a major storm on a temperate lake: a taste of things to come?, Climatic Change, 151(2), 333–347, doi:10.1007/s10584-018-2302-3, 2018.

Wüest, A. and Lorke, A.: Small-scale hydrodynamics in lakes, Annual Review of Fluid Mechanics, 35(1), 373–412, doi:10.1146/annurev.fluid.35.101101.161220, 2003.

Yeates, P. S., Imberger, J. and Dallimore, C.: Thermistor chain data assimilation to improve hydrodynamic modeling skill in stratified lakes and reservoirs, Journal of Hydraulic Engineering, 134(8), 1123–1135, 2008.

Zhang, Z., Beletsky, D., Schwab, D. J. and Stein, M. L.: Assimilation of current measurements into a circulation model of Lake Michigan, Water Resources Research, 43(11), W11407, doi:10.1029/2006WR005818, 2007.

THEO BARACCHINI CURRICULUM VITAE

Av. des Alpes 18 - 1009 Pully
+41 21 728 52 24 – +41 78 851 51 41
theo.baracchini@meteolakes.ch

Civil status : Single
Date of Birth : 15th Sep. 1989
Nationality : Swiss (VD/GR) - Italian

ACADEMIC CAREER AND DIPLOMAS

Swiss Federal Institute of Technology of Lausanne (EPFL) 2015 - 2019
Physics of Aquatic Systems Laboratory (APHYS)

PhD Student: *From observations to 3D forecasts: data assimilation for high resolution lakes monitoring.*
Developer & manager of the operational 3D forecasting platform: *meteolakes.ch*

University of Michigan (UMich) – Ecology and Evolutionary Biology 2014
Pascual Laboratory

Master Thesis: *Modelling cholera dynamics in North-East India and Bangladesh* (published [1]), 6/6

Swiss Federal Institute of Technology of Lausanne (EPFL) 2009 - 2014
Environmental Engineering and Science

Triple specialisation Master: *Monitoring and modelling of the environment (C); Water, soils, and ecosystems engineering (B); Area and cultural studies;* 146/120 ECTS, average of 5.7/6, *Award of excellence (top 5)*

Bachelor 1st and 2nd year, bachelor's general average (1st, 2nd & 3rd year) of 5.1/6

Technical university of Denmark (DTU) 2011 - 2012
Environmental & Management Engineering

Bachelor 3rd year, average of 10/12 (5.5/6), w*inner of the Student and Lecturer awards*

High school of Chamblandes, Pully - Baccalaureate in literature 2005 - 2008

PROFESSIONAL EXPERIENCE

- **IT manager, APHYS laboratory, EPFL, Lausanne** 2016 - 2019
- **PhD representative, Environmental Engineering Teaching & Aviseur committees, EPFL, Lausanne** 2016 - 2019
- **Scientific assistant, Applied hydroeconomics and alpine environmental dynamics lab, EPFL, Lausanne** 2015
  Creation & following of the course project on micro-irrigation improvements in dryland areas
- **Scientist, Cooperation & Development Center, EPFL, Lausanne** 2014-2015
  Modelling soil water dynamics of a wireless sensor network for efficient dryland agriculture in Burkina Faso [2]
- **Teaching-assistant, EPFL, Lausanne** 2012 - 2018
  Limnology, Environmental Engineering, Quantitative Methods, Soil Water Regime Management, Geomatics, Robotics for children
- **Associative experience, EPFL, Lausanne** 2012 - 2013
  Active member of World Engineers, department for internships in developing countries
- **Scientific internship of 3 months, Hydrique Ingénieurs, Lausanne** 2012
  Validation of a wastewater network model of the Riviera, field measurements, lab analysis
- **Military service, Bière, Birmensdorf, Bern, Geneva** 2008 - 2009
  Military obligations completed in the infantry (300 days) - tank rifleman and embassy guard

PUBLICATIONS

Published:

[1]. Baracchini, T., King, A. A., Bouma, M. J., Rodó, X., Bertuzzo, E., and Pascual, M.: “Seasonality in cholera dynamics: a rainfall-driven model explains the wide range of patterns in endemic areas,” *Advances in Water Resources*, 2016.

[2]. Bouleau, C. R., Baracchini, T., Barrenetxea, G., Repetti, A., Bolay, J.-C.: “Low-cost wireless sensor networks for dryland irrigation agriculture in Burkina Faso”, *Technologies for Development*, 2015.

[3]. Nouchi, V., Kutser, T., Wüest, A., Müller, B., Odermatt, D., Baracchini, T., and Bouffard, D.: “Resolving biogeochemical processes in lakes using remote sensing”, *Aquatic Sciences*, 2019.

Submitted & in preparation:

[1]. Baracchini, T., Hummel, S., Verlaan, M., Cimatoribus, A., Wüest, A., and Bouffard, D.: “Automated calibration of 3D lake hydrodynamic models using and open-source data assimilation platform”. In prep, 2019.

[2]. Baracchini, T., Chu, Y. P., Šukys, J., Lieberherr, G., Wunderle, S., Wüest, A., and Bouffard, D.: “Data assimilation of in-situ and satellite remote sensing data to 3D hydrodynamic lake models”. Submitted to *Geoscientific Model Development*, 2019.

[3]. Baracchini, T., Wüest, A., and Bouffard, D.: “*Meteolakes*: an online 3D monitoring and forecasting platform for lakes”. In prep, 2019.

LANGUAGES

French	Mother tongue	**Italian**	Oral: excellent, written: fair (B2)
English	Oral and written: excellent (C1)	**German**	Oral and written: fair (B1-B2)
Spanish	Oral: excellent, written: good (B2)	**Mandarin**	Learning (A1)

INFORMATICS & PROGRAMMING SKILLS

- Excellent knowledge in software and hardware informatics
- Good knowledge in web-development (JavaScript, HTML, CSS, website/server management) – cf. meteolakes.ch
- Excellent knowledge in MATLAB & R
- Advanced knowledge in JAVA, Python & Powershell
- Good knowledge of GIS software (QuantumGIS, ArcGIS)

INTERESTS

- Adventure motorcycling and engine mechanics
- Astrophysics
- Hardware informatics
- Beer lover
- Sailing
- Swimming (TriTeam-Pully)

Ce document a été imprimé au Centre d'impression EPFL,
imprimerie climatiquement neutre, certifiée myClimate.