

An automated calibration framework and open source tools for 3D lake hydrodynamic models

Theo Baracchini^a, Stef Hummel^b, Martin Verlaan^{b,c}, Andrea Cimattorus^d, Alfred Wüest^{a,e}, Damien Bouffard^{e,*}

^a Physics of Aquatic Systems Laboratory (APHYS), ENAC, EPFL, Lausanne, 1015, Switzerland

^b Deltares, Delft, 2600, the Netherlands

^c Mathematical Physics, Electrical Engineering, Mathematics and Computer Science (EEMCS), TU Delft, Delft, 2600, the Netherlands

^d Ecological Engineering Laboratory (ECOL), ENAC, EPFL, Lausanne, 1015, Switzerland

^e Eawag, Swiss Federal Institute of Aquatic Science and Technology, Surface Waters – Research and Management, Kastanienbaum, 6047, Switzerland

ARTICLE INFO

Keywords:

Auto-calibration
OpenDA
Delft3D-FLOW
Parameter estimation
Observational uncertainty
Model performance evaluation

ABSTRACT

Understanding lake dynamics is crucial to provide scientifically credible information for ecosystem management. In this context, three-dimensional hydrodynamic models are a key information source to assess critical but often subtle changes in lake dynamics occurring at all spatio-temporal scales. However, those models require time-consuming calibrations, often carried out by trial-and-error. Through a new coupling of open source software, we present here a flexible and computationally inexpensive automated calibration framework. The method, tailored to the calibration data available to the user, aims at (i) reducing the time spent on calibration, and (ii) making three-dimensional lake modelling accessible to a broader range of users. It is demonstrated for two different lakes (Lake Geneva and Greifensee) with an extensive multi-variable observational dataset. Models mean absolute errors are reduced by up to ~50% over the baseline. Guidelines on heat and momentum transfer parameters are given with their dependence on the observational setup.

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 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, which in addition to being inefficient and complex, is partly subjective and relies 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 implementation step (Solomatine et al., 1999).

Calibration (parameter optimization) is an inverse problem to minimize a cost/objective function that expresses the goodness-of-fit between simulated and observed variables in numerical form (Fenicia

* Corresponding author. Eawag, Seestrasse 79, Kastanienbaum, 6047, Switzerland.

E-mail address: damien.bouffard@eawag.ch (D. Bouffard).

<https://doi.org/10.1016/j.envsoft.2020.104787>

Received 29 May 2019; Received in revised form 10 April 2020; Accepted 22 June 2020

Available online 7 August 2020

1364-8152/© 2020 Elsevier Ltd. All rights reserved.

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).

Over the last decades, in various fields of environmental sciences, such as surface water hydrology (Ahmadi et al., 2014; Bárdossy and Singh, 2008; Gallagher and Doherty, 2007; Hendrickson et al., 1988; Hughes et al., 2014; Johnston and Pilgrim, 1976; Solomatine et al., 1999; Vidal et al., 2007; Wang and Brubaker, 2015), and 1D or 2D lake hydrodynamics (Afshar et al., 2011; Fabio et al., 2010; Gaudard et al., 2017), a number of computer-based automated calibration frameworks have been developed such as PEST, DREAM, GLUE. However, it has been done at a significant computational cost. The computational burden associated with 3D hydrodynamic models limited the use of such frameworks. In this situation, computational overhead is a major concern, and while various solutions have been proposed to tackle complex hydrologic models (Li et al., 2011; Rouholahnejad et al., 2012; Wu and Liu, 2012; Zhang et al., 2013), only a few developed parameter search and optimization techniques adapted to the high dimensionality of 3D lake hydrodynamic systems. We argue that this gap prevents the widespread of 3D hydrodynamic modelling for lake research.

In this study, we present a new and 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. 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), and for tidal sensitivity analysis (Kurniawan et al., 2011), but to our knowledge, not yet applied to 3D lake hydrodynamic modelling. In the scope of this study, an interface for OpenDA has been developed to support Delft3D-FLOW through its file-based black-box wrapper.

Parameter estimation in this kind of application became a time-consuming expert-user problem. With this study, we propose a solution to break down this task, 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 a flexible and efficient calibration framework, tailored to the available calibration data and its uncertainty, hence alleviating the time-consuming trial-and-error approach. Our approach uses a relatively simple and computationally inexpensive algorithm, DUD (Doesn't Utilize Derivatives, Ralston and Jennrich, 1978), for parameter inference. 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 and its performance are evaluated through the calibration of two different lake systems (both in terms of size and dynamics), using temperature and current observations. This study does not only provide a tool for modellers to automatize tedious calibration tasks but is also a necessary prerequisite for research aiming at hydrodynamic processes understanding or even improving model accuracy through data assimilation and uncertainty quantification. For instance, a data assimilative system updating model states artificially is counterproductive when it is not able to reproduce the physics of the observed system resulting from a

bad parameterization. The approach presented here already benefited various studies and systems, ranging from theoretical articles assessing the assimilation of satellite and in-situ observations (Baracchini et al., 2020a), up to 3D online lake operational systems (Baracchini et al., 2020b, 2019).

The paper is structured as follows: the methodology is introduced in Sections 2 and 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 4) show a statistical improvement over the baseline models. Section 4 details the results for Lake Geneva (large-scale system), whereas the Greifensee results are available in the appendix. Finally, Section 5 contains the results analysis and discussion before conclusions are summarized in Section 6.

2. Materials and methods

2.1. Study sites

Lake Geneva - Lake Geneva is the largest freshwater lake of Western Europe (580 km² of surface area, 89 km³ of volume, and residence time of 11.4 years, Fig. 1). Located between Switzerland and France (46.4° N, 6.5° 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 and almost homothermal in late winter, although complete deep-mixing usually occurs only once per 5 to 10 winters on average. Now mesotrophic, light penetration in Lake Geneva varies strongly over seasons and Secchi depth typically ranges from 3.6 to 14 m. Hence, time series of measured Secchi depth are used to force the model, in addition to the meteorological data.

Greifensee - Greifensee is a small (8.5 km² of surface area, 0.15 km³ of volume and residence time of 1.1 year, Fig. 1) lake located in the area of Zurich (47.2° N, 8.4° E, Switzerland). The maximal and mean depth are 30 and 18 m, respectively. During the second half of the 20th century, Greifensee experienced strong eutrophication. Although, recovery started since strict phosphorus regulations were implemented, the lake is still considered eutrophic. This results in shallower light penetration compared to Lake Geneva throughout the year. Monthly measured Secchi depth, ranging from 1.5 to 8.7 m, are used to drive the model.

2.2. Delft3D-FLOW model setup

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

In this study, the z-layers (e.g. fixed vertical layers) scheme is used, since it is capable of reproducing the lake stratification with the given steep basin morphologies (for such steepness, σ -layer models are subject to strong numerical diffusion). Fifty and hundred layers are used for Greifensee and Lake Geneva, respectively. The layers are unevenly distributed, with thinner surface layers (down to 20 cm thick right at the top) to a few meters in the deep hypolimnion. The horizontal grid sizes are 450 m (274'701 cells) for Lake Geneva and 60 m (70'232 cells) for Greifensee. Computational grids (horizontal and vertical) are too coarse and time steps too long to resolve small-scale turbulence. The length scales of the turbulent processes are therefore of "sub-grid" scale and closure models, like the k - ϵ model, have to be considered. Time-steps of 1 min for Greifensee and 2 min for Lake Geneva maintain model stability. Simulations are initialized (uniformly horizontally) from an in-

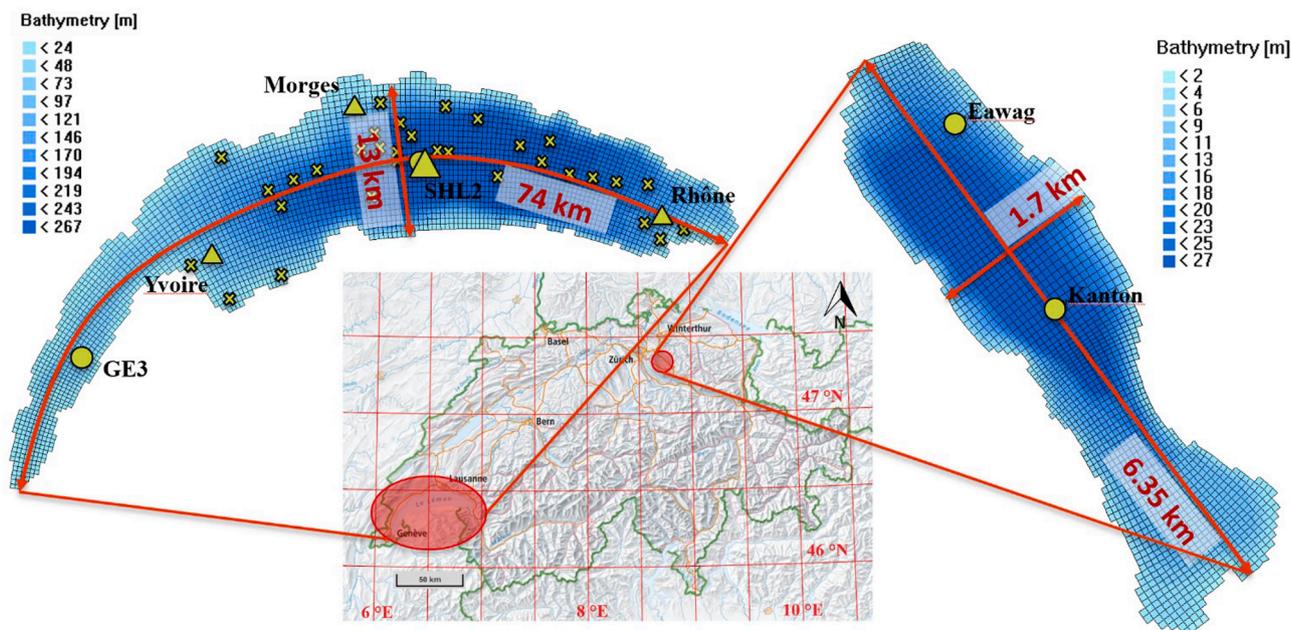


Fig. 1. Study sites, morphology, model grid and in-situ monitoring locations (circles are permanent stations, crosses mark single profiles, and triangles indicate ADCP current meter locations). Map courtesy of <https://map.geo.admin.ch>.

situ temperature profile taken at the deepest location 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 (MeteoSwiss, 2019a, 2019b) reanalysis products, consisting of seven meteorological variables on regular 1.1–2.2 km grid and hourly resolutions. The variables include: solar radiation, wind direction and speed, relative humidity, cloud cover, pressure, and air temperature from the COSMO atmospheric model with data assimilation tailored to the Alpine region. A single model iteration (two years for Lake Geneva and one for Greifensee) requires ~1 day of computation on three Xeon Broadwell cores for Greifensee to ~2 days on four cores for Lake Geneva, indicating the paramount importance of the computational efficiency of the calibration scheme.

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 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 using existing calibration and data assimilation algorithms with any model whose OpenDA interface is implemented accordingly.

There was no existing interface for temperature and current fields with the z-layers coordinate system of Delft3D-FLOW. For the purpose of this study, an OpenDA interface has been developed using the black-box wrapper. The coupling is a file-based approach, implemented by adding JAVA classes to allow reading and writing 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. The black-box wrapper approach can be used to couple OpenDA with any other model. 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) as one. The entire code can be found on GitHub (<https://github.com/OpenDA-Association/OpenDA>).

2.4. Monitoring data

A key aspect in model parameterization is the quality of calibration data (Madsen, 2003). In-situ observations are often time demanding and costly. Moreover, in the case of 3D models, the spatial variability of the lake has to be validated. This requires multi-site (at least two different locations) and ideally multi-variable time series of observations. For hydrodynamic models, both heat and momentum input from the atmosphere have to be calibrated. This requires in-situ information on temperature and flows. In this study, we consider in-situ data of temperature and current measurements from both, short-term field studies (single profiles) and permanent monitoring stations at various frequencies (6-hourly to monthly). 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 (Fig. 1) sampling temperature over the water column at different frequencies. The northern station (Eawag) has 321 profiles during summer 2014, measured at a 6-hourly interval to a maximum depth of 17 m, while the central station (Canton) has 12 monthly profiles over the same year to a maximum depth of 30 m (deepest location). The model is initialized from a temperature profile sampled at the center of the lake. For initial conditions, horizontally temperature fields are set homogeneous and flow velocities are set to zero.

Lake Geneva – The available dataset for Lake Geneva consists of 90 temperature profiles over the water column at 32 different locations (Fig. 1) sampled during 2015 and 2016. While most of the locations have been sampled only once (black crosses in Fig. 1), monthly to bi-monthly measurements are available at two stations (GE3 and SHL2, black circles in Fig. 1). Some results will therefore focus on those two stations. The entire dataset has been used for calibration.

In addition to temperature data, currents are measured at four locations (triangles, Fig. 1) with Acoustic Doppler Current Profilers (ADCPs, Cimattoribus et al., 2017). This data consists in continuous (over up to six months) ADCP 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 components are 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 dynamic forcing cannot realistically be expected. Finally, this high-frequency dataset has been down-sampled to 6-h intervals, while the higher frequency variability will be considered in the data uncertainty definition to account for unavoidable model phase shifts. The Lake Geneva model is initialized from the output of a previously running model, its spin-up period is thereby not included in this analysis.

Data uncertainty - In addition to model uncertainties (e.g. initial conditions, physical processes and their parameterizations (Chen et al., 2019), approximations of the system) and forcing errors, measured data is also subject to systematic and random errors (Bárdossy and Singh, 2008). Hence, uncertainty of the information must be considered in the calibration framework. Increasing the amount of data does not necessarily result in better parameter estimations (Sorooshian et al., 1993; Sun and Bertrand-Krajewski, 2012). 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⁻¹), 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 h (depending on the lake size and the period of basin-scale modes). The latter especially affects temperature observational 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. (1)) too much when a short-term event (e.g. internal wave) is not phased correctly, which could result from uncertainties elsewhere (e.g. in forcing variables, which also come from a model). Absolute values of currents are also averaged hourly.

Calibration procedure - Lake hydrodynamics can be understood in terms of kinetic and potential (thermal) energy budgets, both require calibration. Yet, kinetic and thermal dynamics do not show a first order reaction to changes in all parameters. The heat fluxes predominantly determine the surface boundary structure, while turbulent diffusion in the stratified volume depends mainly on the kinetic energy. By selecting the appropriate set of calibration parameters, the two calibrations can be decoupled. In the case of Lake Geneva, observed currents are often 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. (1)). Therefore, calibration of temperature and currents 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 steps are performed in the case of Lake Geneva. Flow 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 3.1). This sequential approach is valid for calibration of lake hydrodynamics at seasonal scales for which the density stratification is mainly vertical. At hourly time scales, horizontal density gradients also drive local momentum exchanges that are not necessarily calibrated with such an iterative approach. This is also likely the case in complex multi-basin systems.

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. This study hence focuses on the latter approach.

3.1. Calibration algorithm

DUD 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 criteria (specified in the Appendix) are met. The minimized cost function is of the following form:

$$J = \sum_r \frac{(y_o(t) - y_m(t))^2}{\sigma_0^2} \quad (1)$$

With $y_o(t)$ 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 computationally relatively 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) for similar problems. DUD can be used with or without user-defined constraints. In this study, constraints are used and detailed in Table A1 (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. This computationally light twin-experiment has been included in the OpenDA GitHub repository to facilitate the understanding of the entire procedure and tools.

3.2. Calibration parameters

In 3D hydrodynamic models, some parameters cannot be obtained with *a priori* knowledge. Some of them are furthermore grid-dependent as for instance the horizontal diffusivity. Those parameters are perfect candidates for automatic algorithm calibration limiting tedious and subjective parameter hand-tuning. The approach presented in this study aims thus at focusing on quantities that are too complex 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 light attenuation or 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 1, while our choice is motivated in the following sub-sections.

Surface forcing and energy source - The thermal- and fluid-dynamics of a lake are mainly driven by the atmosphere (Imberger and Hamblin, 1982). The accuracy of the heat and momentum transfer at the air-water interface is thereby of paramount importance for the performance of the model, especially for surface and mixed layer processes. The wind drag coefficient, which acts on heat and momentum transfer into the water, is a function of wind speed and can be challenging to determine. Its characterization has been the aim of several studies in both limnology and oceanography (Wüest and Lorke, 2003) and functional parameterization can be found in the literature. Yet, the wind is rarely measured directly over the lake but at nearby official meteorological stations and therefore only approximately represent the wind field over the lake. Consequently, the physical parameterization of the wind drag coefficient needs to be adapted for specific inadequacy in the wind forcing. This is particularly true at low wind speed when local topography affects the wind field. In this paper, the wind drag coefficient at low wind speeds ($0\text{--}5\text{ m s}^{-1}$), most common winds in the considered region, is calibrated.

Viscosities and diffusivities - The summer strong stratification, typical for lakes, prevents vertical over horizontal transport, which is characterized at small scales by a strong anisotropy between the vertical and horizontal viscosity (and diffusivity). Toffolon and Rizzi (2009) showed that the ratio between vertical and horizontal viscosity could affect the lake response to wind by enhancing horizontal circulation or upwelling. In Delft3D-FLOW, as in many other 3D models, background values for both vertical and horizontal diffusivities and viscosities can be specified by the user. Horizontal parameters depend on the grid size (*a priori* knowledge from the model manual suggests values ranging from $1\text{ to }10\text{ m}^2\text{ s}^{-1}$ for grids of $\sim 10\text{ m}$ and $10\text{--}100\text{ m}^2\text{ s}^{-1}$ 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, vertical eddy viscosity and eddy diffusivity are computed from a turbulence closure model. In the $\kappa\text{-}\epsilon$ 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 a 3D turbulence model ν_{3D} (Eq. (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, referred to as DV_H in Table 1 and in the following sections) and be calibrated by the algorithm.

$$\nu_V = \nu_{mol} + \max(\nu_{3D}, \nu_V^{back}) \quad (2)$$

$$\nu_H = \nu_{SGS} + \nu_V + \nu_H^{back} \quad (3)$$

with ν_{mol} for the kinematic viscosity of water, ν_{3D} is computed by the

Table 1
Delft3D-FLOW calibration parameters acting on the energy pathway.

Parameter	Description	Influence on	Calibrated in
$C_{10@0.5}$ [-]	Wind drag coefficient at 10 m above water and at speed of 0.5 m s^{-1}	Temperature/ flow velocity	All setups
z_0 [m]	Bottom hydraulic roughness	Flow velocity	All setups
C_{Dalton} [-]	Evaporative heat flux coefficient	Temperature	Only for temperature
$C_{Stanton}$ [-]	Convective heat flux coefficient	Temperature	Not calibrated
DV_H [$\text{m}^2 \cdot \text{s}^{-1}$]	Background horizontal viscosity and diffusivity	Temperature/ flow velocity	All setups
DV_V [$\text{m}^2 \cdot \text{s}^{-1}$]	Background vertical viscosity and diffusivity	Temperature/ flow velocity	Not calibrated

vertical turbulence closure model, ν_V^{back} is the background vertical viscosity, and ν_{SGS} indicates the viscosity obtained from the sub-grid scale (Horizontal Large Eddy Simulation methodology) 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 by dampening of near-sediment flows and internal waves. The bed roughness height, z_0 , is used as a calibration parameter controlling the boundary layer assuming a logarithmic law-of-the-wall. While z_0 is a parameter that should be directly derived from physical observations such as Acoustic Doppler Current Profiler (ADCP), this parameter not only contain physical interpretation but also model error at the boundaries, which depend among others on the horizontal and vertical grid size. The steep slopes around Lake Geneva are challenging to model and lead to staircase effects with artificial turbulence, which can be damped by increasing the bottom drag coefficient. While the bottom drag coefficient should be evaluated at each grid point, we did not have enough observations and assumed here a horizontally homogeneous parameter.

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 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 sub-domains. The Dalton number (C_{Dalton}) acts directly on the forced evaporative heat flux (Eq. (5)), which is part of the total heat budget (Eq. (4)):

$$Q_{tot} = Q_{sw} + Q_{alw} - Q_{hw} - Q_{ev} - Q_{conv} \quad (4)$$

$$Q_{ev,forced} = L_V \rho_a C_{Dalton} U_{10} (q_s - q_a) \quad (5)$$

With Q_{sw} indicating the incident shortwave radiations, Q_{alw} the incident longwave radiations, Q_{hw} the back radiations, Q_{ev} the evaporative heat flux (composed of the free and forced components), and Q_{conv} the convective heat flux. L_V is 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 ambient air (10 m above water level), respectively. The C_{Dalton} 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 fixed in the second (current speed) calibration. This is especially relevant since there are no flow measurements in the mixed layer.

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 performance. Profiles over the water column and time-series at multiple depth are used to illustrate the model calibration for Lake Geneva. For the model of Greifensee, the same is available in the appendix.

The Greifensee model has been calibrated with temperature profiles at two locations. Improvements in terms of MAE are at 51%, reaching a final MAE of $1.4\text{ }^\circ\text{C}$. A summary of RMSE and MAE values can be found in Table 2. Relative to Greifensee, the Lake Geneva model improvements are lower. The main difference comes from a better initial model for Lake Geneva. The model reached a MAE of $1.12\text{ }^\circ\text{C}$ for the temperature calibration and 2.5 cm s^{-1} for the flow calibration. MAEs, RMSEs and their improvements are summarized in Table 2.

Table 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	1.95 °C 3.9 cm s ⁻¹	1.11 °C 2.5 cm s ⁻¹	1.80 °C 3.6 cm s ⁻¹	16.9 % 7.8 %	7.7 % 7.5 %

Fig. 2, shows the improvements brought by the automated calibration through Taylor diagrams. In the case of Greifensee, the centred 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.

Parametric evolution is significant (Table 3). All parameters increased except the background horizontal viscosity and diffusivity, which approached its lower bound (set at $4.5 \times 10^{-10} \text{ m}^2 \text{ s}^{-1}$, see appendix). The largest change observed is in the hydraulic roughness, which went from 1 cm to 1.1 m in the case of Lake Geneva while remaining small (4.9 cm) in Greifensee. The Dalton number is of greater influence during the stratified period, starting around May. It is worth to remember that the Dalton number has only been calibrated in the temperature calibration experiment. Finally, the wind drag coefficient at low wind speeds increased.

Two stations are selected for a detailed analysis of the temperature calibration, GE3 and SHL2 (Fig. 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 – Fig. 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.

Fig. 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 currents than the reference run, mainly in the epilimnion. This is more in accordance with the observed flows. In the hypolimnion, low flows (<2 cm s⁻¹) are

Table 3
Parameters evolution after calibration of Lake Geneva.

Parameters	Initial value	Value after temperature calibration	Final value
$C_{10@0.5} [-]$	9.8×10^{-3}	4.1×10^{-2}	4.2×10^{-2}
$z_{0,bottom} [m]$	1.0×10^{-2}	7.2×10^{-1}	1.1×10^{-0}
$C_{Dalton} [-]$	1.3×10^{-3}	2.5×10^{-3}	2.5×10^{-3}
$DV_H [m^2 \cdot s^{-1}]$	1.0×10^{-5}	4.5×10^{-10}	1.6×10^{-9}

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.

Time-series – Fig. 5 provides temperature time-series at GE3 station for various depths. Unlike for Greifensee, the model performance is better over the entire range of vertical layers. The thermocline, located between 20 and 30 m depth, 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.

Fig. 6 shows time-series of flow velocities at two stations over 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 (Fig. 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 right before December 1st, 2016. The overall event however has been captured as shown by the significant increase in modelled velocity around that date.

For a calibration of basin-scale internal waves, continuous measurements over the entire calibration period (ideally an entire year) at high frequency are needed at several locations. In that regard, power

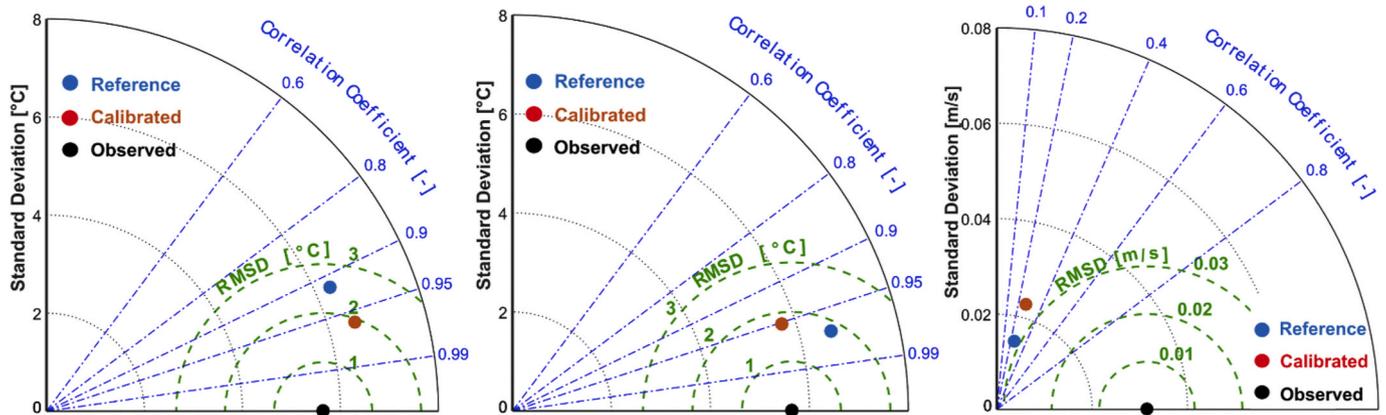


Fig. 2. Taylor diagrams showing the part of cRMSD attributable to variance or pattern correlation (Taylor, 2001). The radial distance from the origin (bottom-left) defines the standard deviation. The radial extent from observations (black dot) to model values (red and blue dots) corresponds to the cRMSD. The azimuthal position provides the correlation coefficient. Taylor diagrams for Greifensee (left), Lake Geneva temperature calibration (centre) and Lake Geneva current calibration (right). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

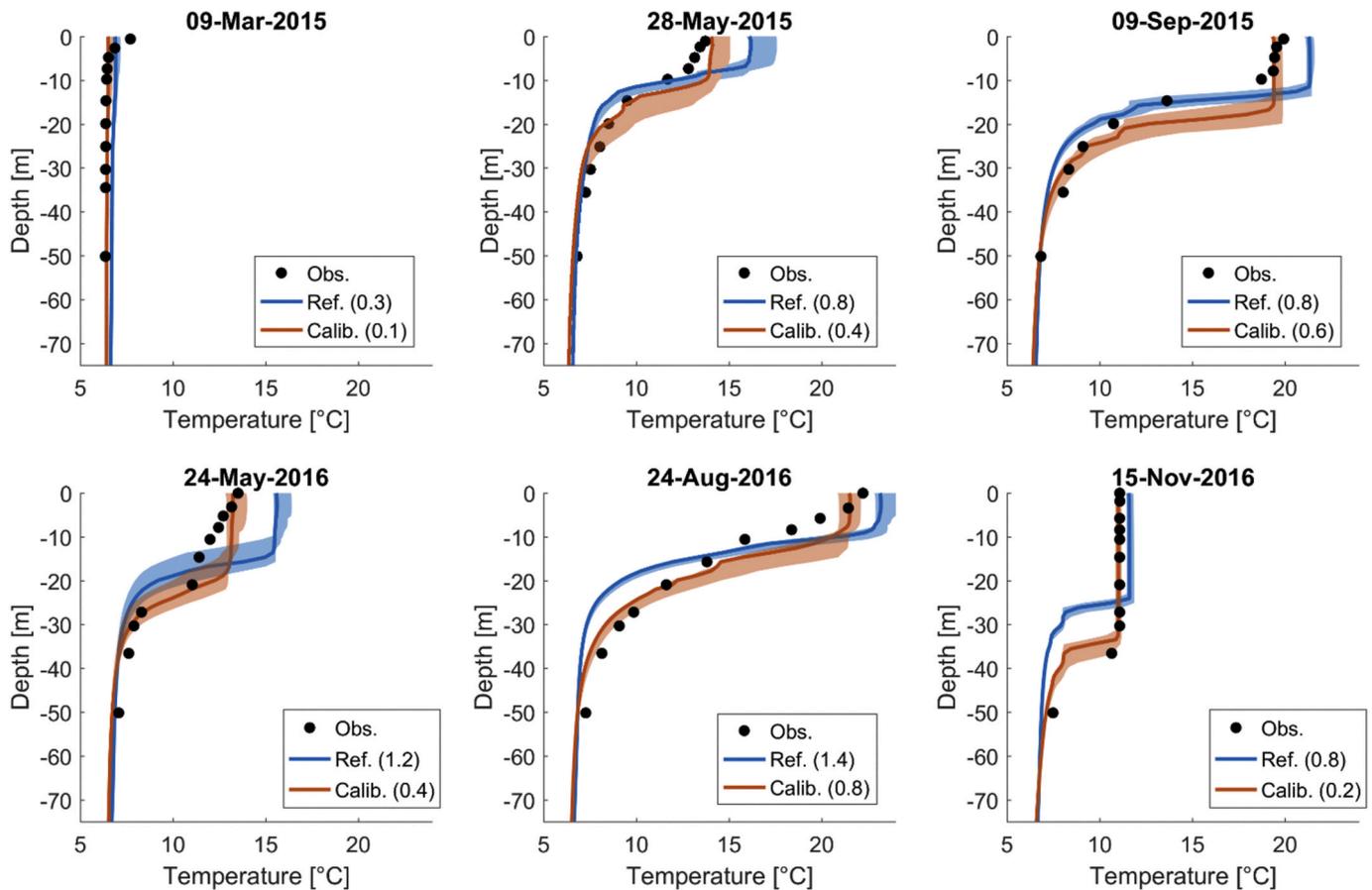


Fig. 3. Temperature profiles at 10 am at SHL2 station. Shaded areas represent the modelled temperature variability over ± 36 h. Numbers in parenthesis correspond to the MAE in $^{\circ}\text{C}$.

spectra of currents in the hypolimnion at the central part of the lake (SHL2) are provided in Fig. 7. The spectra show a peak near the 12–13 h mark (inertial period) in both data and models. Such a period is most likely the signature of Poincaré waves and is consistent with the findings of (Lemmin, 2020; Lemmin et al., 2005). Fig. 7, therefore, highlights the capability of the model for reproducing internal waves dynamics.

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 parameters and where the largest discrepancies are observed before calibration.

The calibration performance of the Lake Geneva model 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 before calibration. 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 the atmospheric 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 attributed to the observation in Section 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 influencing factors (such as inaccuracies in the wind forcing phase).

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 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 modelling study of Lake Geneva by Umlauf and Lemmin (2005) using σ -coordinates. Both the temperature (first step) and currents (second step) calibration required an increased z_0 value. We argue that the discrete stair-steps representing the bottom in the case of Lake Geneva lead to excess turbulence that has to be artificially dampened. The corresponding bottom roughness value can therefore not directly be compared to its physical counterpart but is rather a model-technical interpretation. This is especially the case with z-layers, some of them reaching a thickness larger than 6 m at the bottom boundary hence not fitting the real topography (Chen, 2004). This staircase effect is further exacerbated by the horizontal grid size of 450 m. Those factors combined prevent solving the bottom roughness length and relating it to the physics of the bottom log-profile, these values are therefore not true representations of this parameter and the current version of the model cannot be used to study bottom boundary layer in Lake Geneva. Yet, we stress that the kinetic energy remains well captured with the present

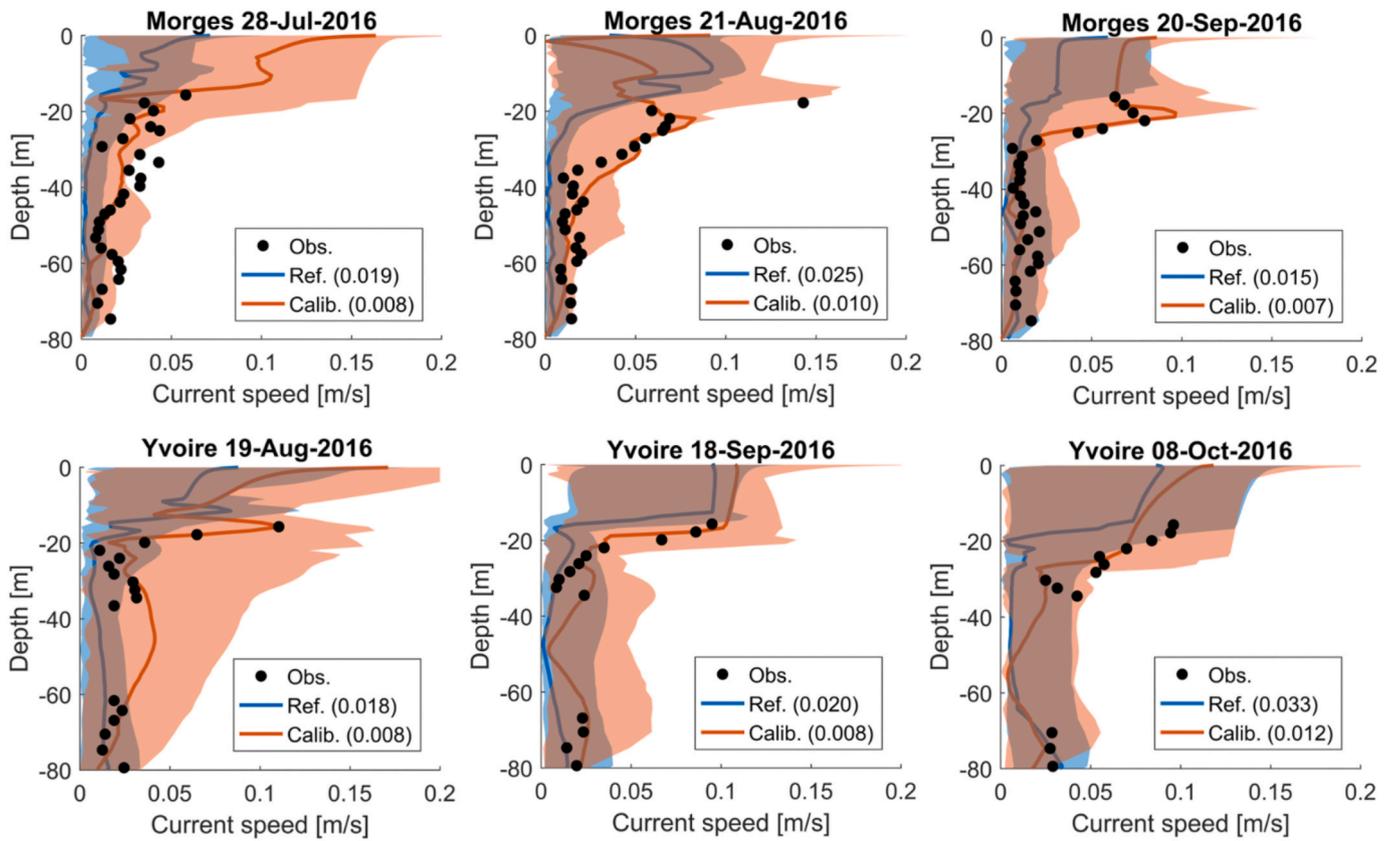


Fig. 4. Selected velocity profiles at specific times at the Morges and Yvoire stations (Fig. 1). Numbers in parenthesis correspond to the MAE in $[m s^{-1}]$. Shaded areas represent the modelled velocity variability over ± 36 h.

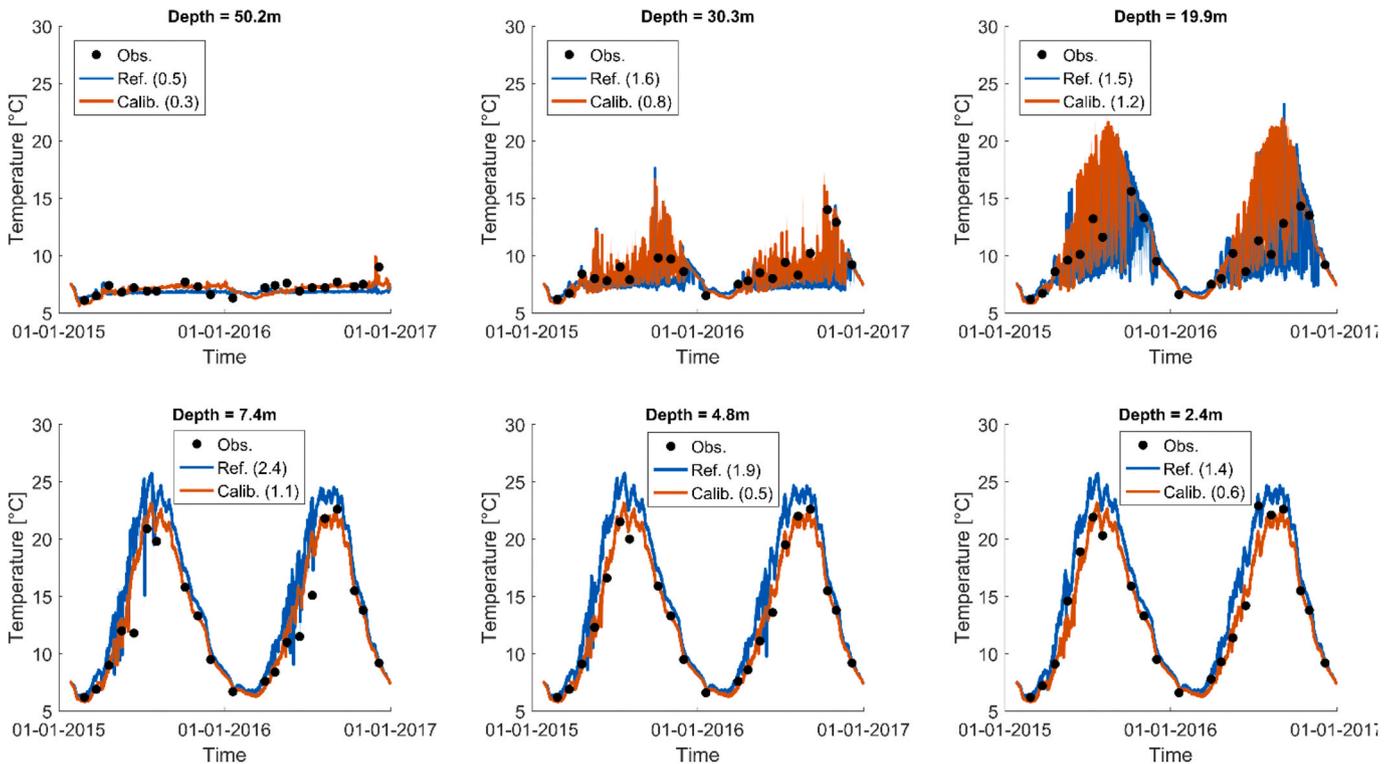


Fig. 5. Temperature time-series at station GE3 (Fig. 1). Numbers in parenthesis correspond to MAE $[^{\circ}C]$. Shaded areas represent the modelled temperature variability over ± 2 vertical layers.

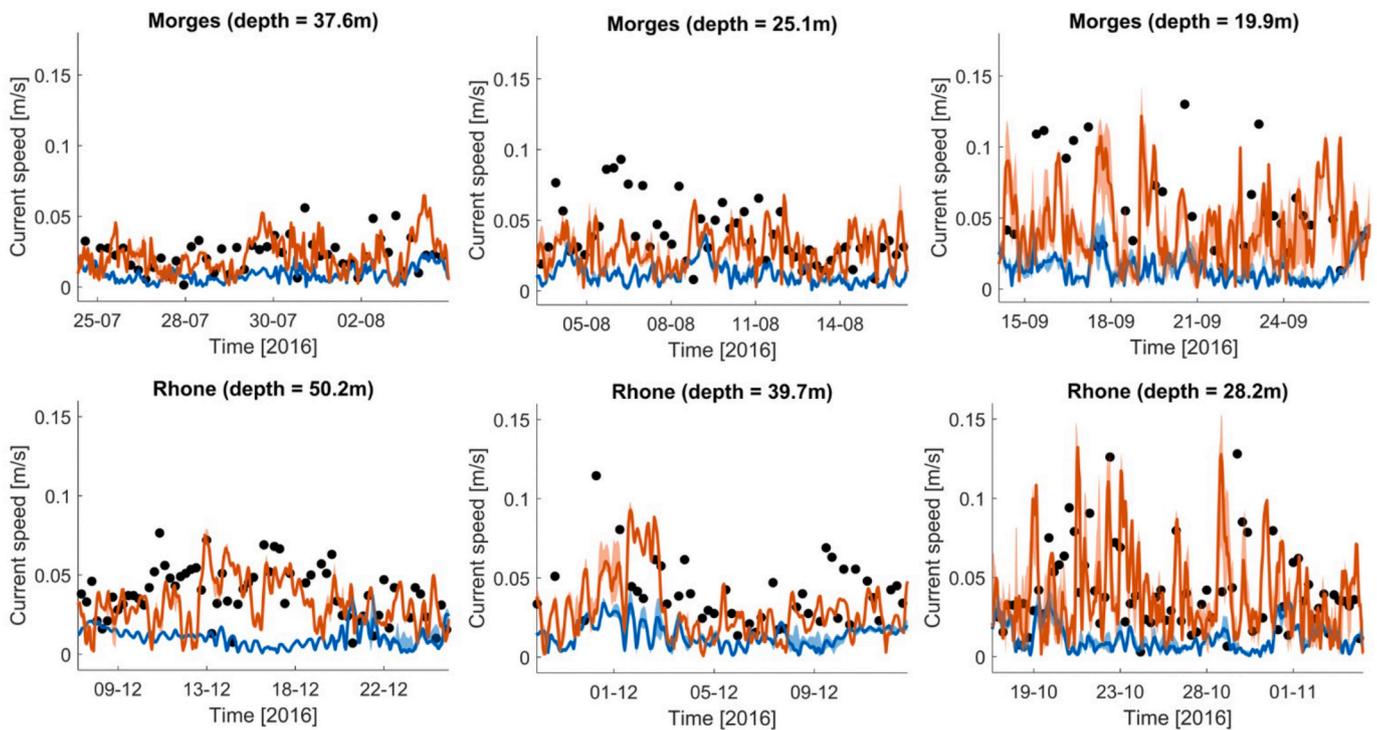


Fig. 6. Selected currents time-series at the Morges (upper row) and Rhône stations (lower row; Fig. 1) for various depths and periods. Red lines are the calibrated model, blue the reference, and black dots are observations. Shaded areas represent the modelled current variability over ± 2 vertical layers. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

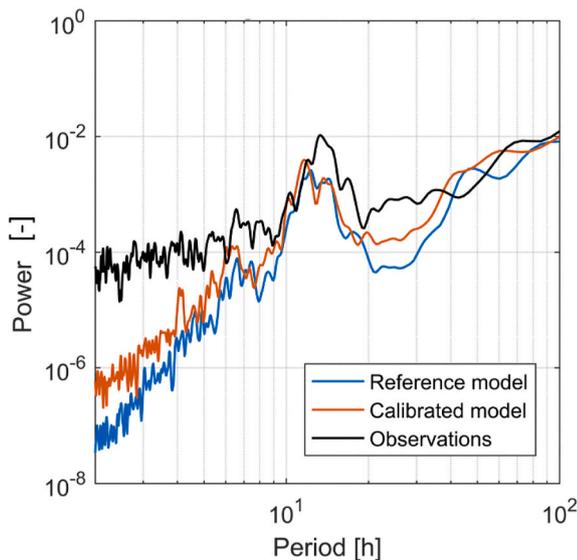


Fig. 7. Periodogram of modelled and observed currents at 252 m depth (over the 2 years computed) at SHL2 (centre of the lake, Fig. 1).

model as shown with the periodogram in Fig. 7. The power spectra of deep current extracted from the model matches the observed one for periods above ~ 0.5 h. An unrealistic damping at the boundaries would have led to smaller values in the power spectra for the basin-scale internal waves and the background circulation.

The wind drag coefficient at low wind speeds ($C_{10@0.5}$) increased by the calibration. The end value differs from what has traditionally been found for oceans. It is however consistent with findings from Wüest and Lorke (2003) for lakes. The argument is that at low wind speeds laminarization of the flow leads to an u^{-1} behaviour of the drag coefficient. 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 currents 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) for the flow-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 significantly larger heat loss by evaporation. This is likely a consequence of inaccuracies in the model forcing, such as in the short-wave radiations or the wind fields, a quantity known as complex to model above lake surfaces (e.g. only few field observations available). This parameter, notably sensitive to temperature data from CTD profiles, contributed significantly in regulating the mixed layer temperature.

In terms of hypolimnetic currents, both computed and measured values are close to the measurement accuracy of 0.8 cm s^{-1} (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 currents. This is both a result of keeping the Dalton parameter constant and enabling the weak parameters constraint (Section 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 there the flows are significantly higher than the measurement accuracy of the instrument.

Overall the procedure has been found to be computationally efficient (6–12 model iterations needed for the models considered here) and relatively simple to use or expand to other systems and models by non-expert users. Significant timesaving is 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>.

6. Conclusion

Model calibration of computationally expensive three-dimensional (3D) hydrodynamic models 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 parallelization. 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 presents an efficient and flexible auto-calibration solution by coupling open-source tools, thereby providing significant improvements to the parameterization of three-dimensional inland-water models without user intervention.

Using temperature and current 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 flows. 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 provide a rather 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, tailored to lakes and models of various scales and morphologies, provides better results than previous manual calibrations, 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. As lakes are considered sentinels for climate and catchment changes (Shimoda et al., 2011; Wagner and Adrian, 2009), a better modelling and monitoring of dynamical responses to external influences will contribute to a more sustainable management and to securing essential ecosystem services of lakes. Such a solution aims at facilitating their modelling for a wider spread adoption of complex 3D hydrodynamic models.

Software 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>. Both software are developed by Deltares in the Netherlands (P.O. Box 177, 2600 MH Delft, The Netherlands, +31 (0)88 335 8273). Delft3D-FLOW tag 7426 is used. It is developed in FORTRAN and is available for both Windows and Linux operating systems. OpenDA v2.4.5 is developed in Java and is platform-independent.

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 their Cantonal 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. All these data can be requested by contacting their aforementioned owners, as it is not the property of the authors of this study. In-situ data at SHL2 as well as Secchi disk measurements in Lake Geneva were provided by the *Commission Internationale pour la Protection des Eaux du Léman* (CIPEL) and the Information System of the SOERE OLA, INRA, Thonon-les-Bains. It can be downloaded from their online platform: <http://si-ola.inra.fr>. The ADCP data collected by Andrea Cimattoribus used in this study can be downloaded from the Zenodo archive: <https://zenodo.org/record/1109794#.XM1duKQ68UE>. Any other temperature profile and ADCP measurement has been collected by the Physics of Aquatic Systems Laboratory at EPFL and has been openly published on the following Zenodo archive: <https://zenodo.org/record/3716076>.

Author contributions

TB, DB, AW designed the procedure and TB carried it out. MVL and SH 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.

Declaration of competing interest

None.

Acknowledgements

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).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.envsoft.2020.104787>.
Appendix

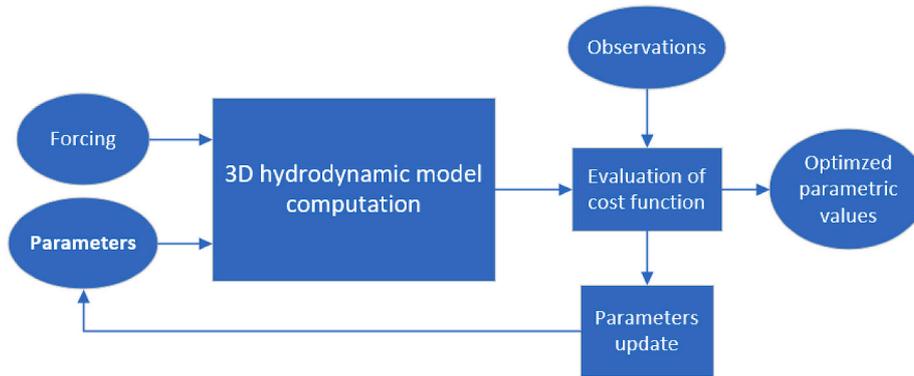


Fig. A1. Flowchart of the calibration procedure. The forcing contains the atmospheric data driving the Delft3D hydrodynamic models. 7 atmospheric variables provided on a 1×1 km grid at hourly time-step are used. The models are run over 1–2 years to cover the entire range of lake dynamics (stratified and mixed regimes). Parameters are the user-defined model parameters adjusted by OpenDA (here the wind drag coefficient, evaporative heat flux coefficient, bottom drag coefficient and horizontal background diffusivity/viscosity). Observations consist in temperature and ADCP profiles collected at various locations in the lake. After a model run a cost function is computed to assess data-model mismatch, depending on the results, OpenDA will generate a new set of parameter values and recompute the model. This iterative process is repeated until an optimized set of parameters is found. In this study, calibrating 4 parameters required 6 to 12 iterations.

Table A1
 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^{-1}	9.8×10^{-3}	6.6×10^{-5}	1.1×10^{-1}
z_0 [m]	Bottom hydraulic roughness height	1.0×10^{-2}	1.0×10^{-3}	6.7×10^{-0}
C_{Dalton} [-]	Evaporative heat flux coefficient	1.3×10^{-3}	0.0	5.0×10^{-2}
DV_H [$m^2 \cdot s^{-1}$]	Background horizontal viscosity and diffusivity	1.0×10^{-5}	4.5×10^{-10}	$1.5 \times 10^{+2}$

The stopping criteria for the DUD algorithm, specified in the OpenDA algorithm configuration XML file, are set to the following values. The outerloop has a maximum number of 10 iterations, and a maximum absolute and relative difference between the costs of two best parameter estimates of 0.01. The innerloop/linesearch has a maximum of 5 iterations and a maximum relative step size of 10. If the linesearch produces estimates with a larger cost, 3 iterations are allowed before searching in opposite direction, with a backtracking with a step size shortening factor of 5.

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 A2).

Table A2
 Parameters evolution for Greifensee calibration

Parameters	Initial value	Final value
$C_{10@0.5}$ [-]	9.8×10^{-3}	5.5×10^{-5}
$z_{0, bottom}$ [m]	1.0×10^{-2}	4.9×10^{-2}
C_{Dalton} [-]	1.3×10^{-3}	3.3×10^{-3}
DV_H [$m^2 \cdot s^{-1}$]	0.0×10^{-0}	2.8×10^{-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 1 cm to ~5 cm, which corresponds to observed values from literature. The latter contributes to increase the dissipation of kinetic 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.

Table A3

Parameters evolution over the iterations for the temperature calibration of Lake Geneva.

Iteration	$DV_H [m^2 \cdot s^{-1}]$	$z_0 [m]$	$C_{Dalton} [-]$	$C_{10@0.5} [-]$
1	1.0×10^{-5}	1.0×10^{-2}	1.3×10^{-3}	9.8×10^{-3}
2	$1.2 \times 10^{+1}$	1.0×10^{-2}	1.3×10^{-3}	9.8×10^{-3}
3	1.0×10^{-5}	2.4×10^{-0}	1.3×10^{-3}	9.8×10^{-3}
4	1.0×10^{-5}	1.0×10^{-2}	4.3×10^{-3}	9.8×10^{-3}
5	1.0×10^{-5}	1.0×10^{-2}	1.3×10^{-3}	4.4×10^{-2}
6	4.5×10^{-10}	7.2×10^{-1}	2.5×10^{-3}	4.2×10^{-2}

Table A4

Parameters evolution over the iterations for the flow velocity calibration of Lake Geneva.

Iteration	$DV_H [m^2 \cdot s^{-1}]$	$z_0 [m]$	$C_{Dalton} [-]$	$C_{10@0.5} [-]$
1	4.5×10^{-10}	7.2×10^{-1}	–	4.2×10^{-2}
2	5.5×10^{-4}	7.2×10^{-1}	–	4.2×10^{-2}
3	4.5×10^{-10}	1.7×10^{-2}	–	4.2×10^{-2}
4	4.5×10^{-10}	7.2×10^{-1}	–	1.9×10^{-1}
5	2.5×10^{-9}	2.1×10^{-0}	–	6.7×10^{-2}
6	1.1×10^{-9}	1.2×10^{-0}	–	5.3×10^{-2}
7	7.0×10^{-10}	9.4×10^{-1}	–	4.7×10^{-2}
8	5.6×10^{-10}	8.2×10^{-1}	–	4.5×10^{-2}
9	5.1×10^{-10}	7.7×10^{-1}	–	4.3×10^{-2}
10	4.3×10^{-10}	6.9×10^{-1}	–	4.1×10^{-2}
11	1.6×10^{-9}	1.1×10^{-0}	–	4.2×10^{-2}

Table A5

Parameters evolution over the iterations for the temperature calibration of Greifensee.

Iteration	$DV_H [m^2 \cdot s^{-1}]$	$z_0 [m]$	$C_{Dalton} [-]$	$C_{10@0.5} [-]$
1	0.0×10^{-0}	1.0×10^{-2}	1.3×10^{-3}	9.8×10^{-3}
2	1.0×10^{-0}	2.5×10^{-0}	1.3×10^{-3}	9.8×10^{-3}
3	0.0×10^{-0}	1.0×10^{-2}	4.0×10^{-3}	9.8×10^{-3}
4	0.0×10^{-0}	1.0×10^{-2}	1.3×10^{-3}	3.0×10^{-4}
5	0.0×10^{-0}	4.9×10^{-2}	3.3×10^{-3}	5.5×10^{-5}
6	2.8×10^{-0}	4.9×10^{-2}	3.3×10^{-3}	5.5×10^{-5}

B. Greifensee profiles

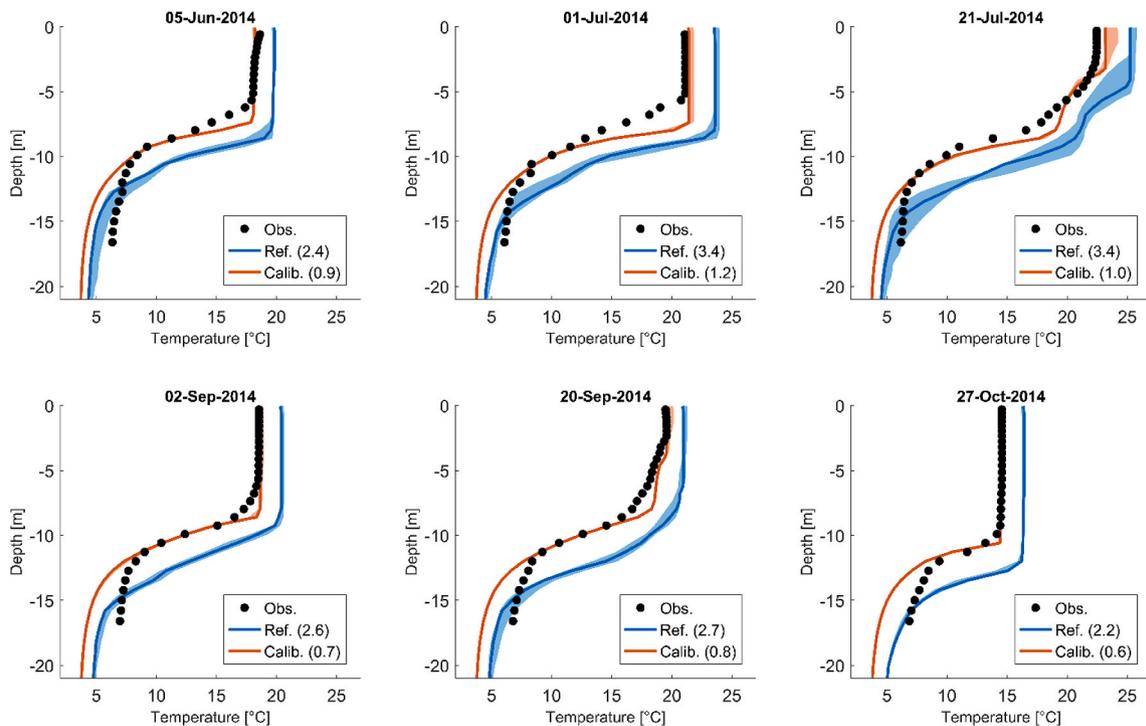


Fig. B1. Temperature profiles at specific times at the high-frequency northern station (Eawag) with MAEs in parenthesis. Shaded areas represent the modelled temperature variability over ± 12 h.

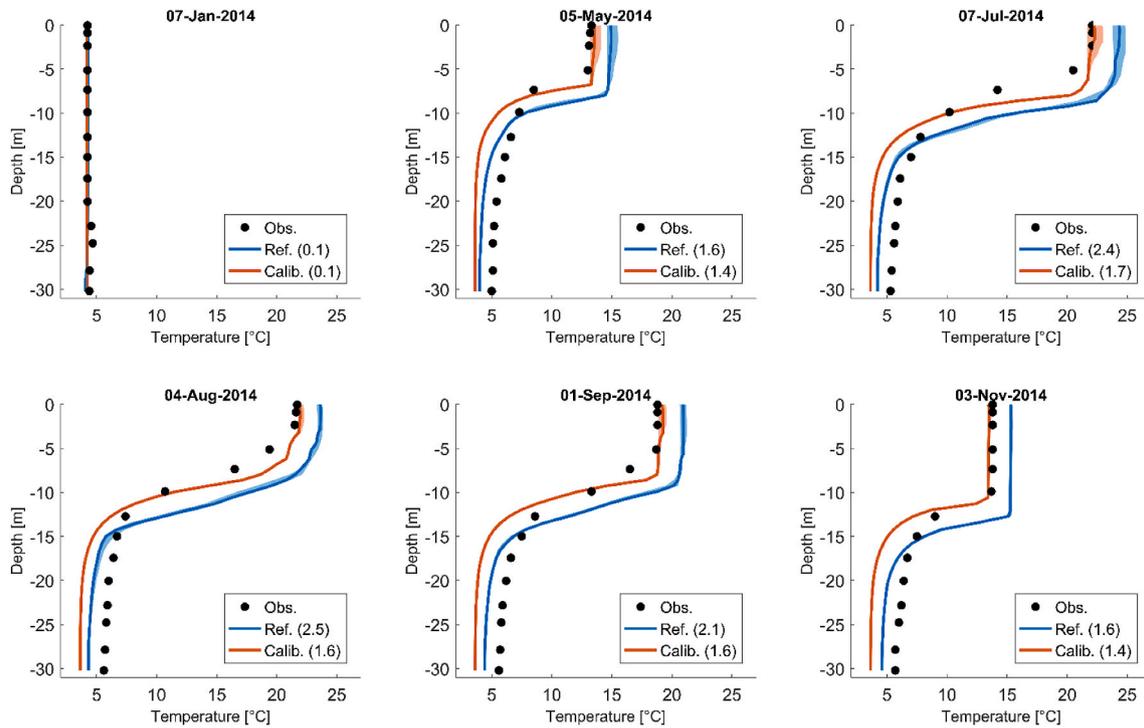


Fig. B2. Temperature profiles at specific times at the central station (Canton) with MAEs in parenthesis. Shaded areas represent the modelled temperature variability over ± 12 h.

Figs. B1 and B2 show the improvements through profiles at different times for the two locations in Greifensee. 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 (Fig. B1). The latter, in combination with the poorer deep-water performance, are both related to a limitation of the method discussed in the next section.

C. Greifensee time-series

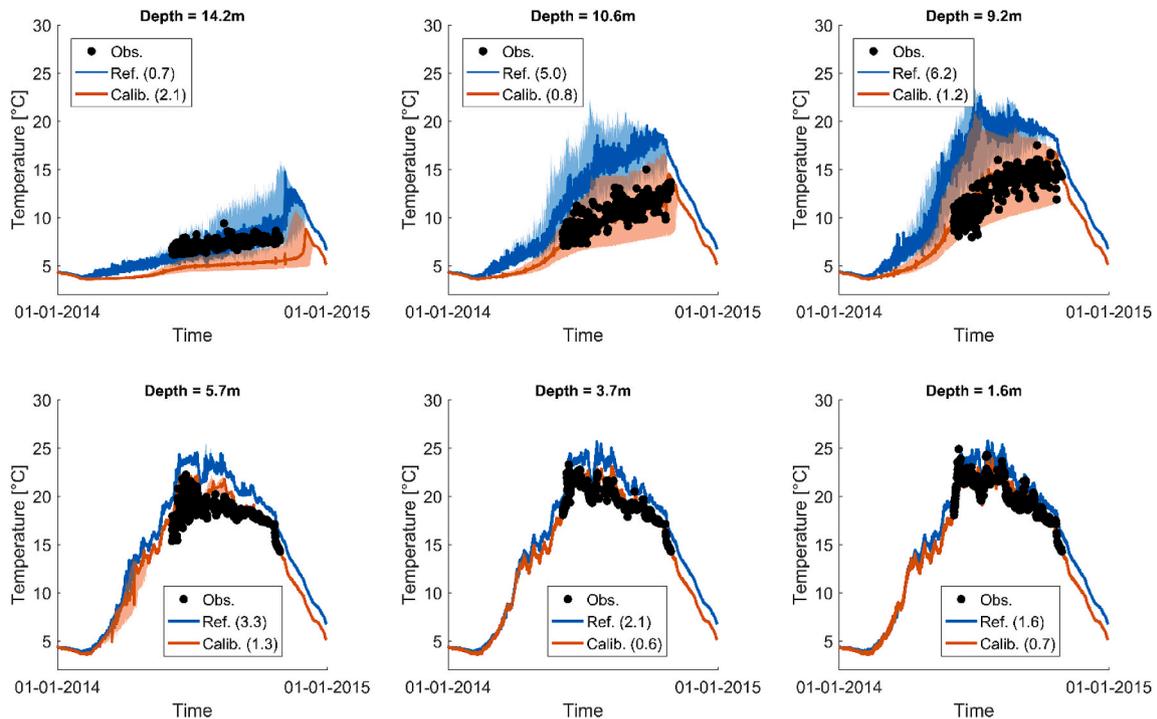


Fig. C1. Time-series of temperature at various depth for the northern high-frequency station (Eawag) with MAEs in parenthesis. Shaded areas represent the modelled temperature variability over ± 2 vertical layers.

Figs. C1 and C2 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 periods, which is highest at the thermocline depth as has to be expected.

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 increase slightly. 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.

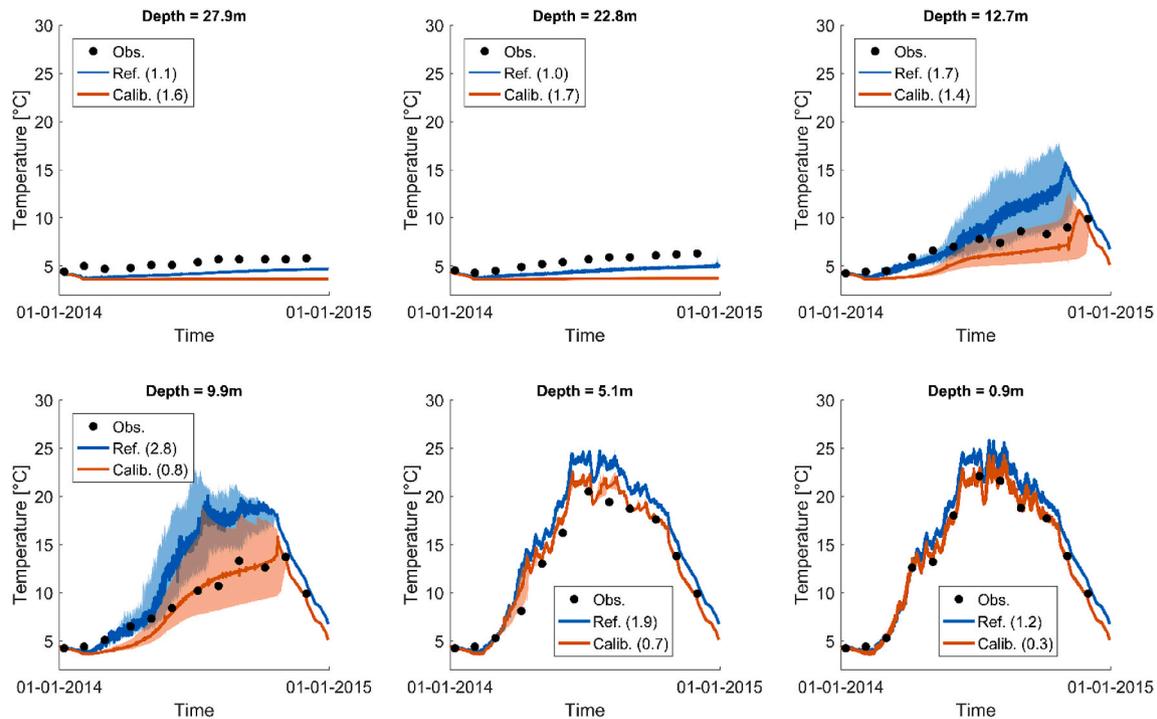


Fig. C2. Time-series of temperature at various depth for the central station (Canton) with MAEs in parenthesis. Shaded areas represent the modelled temperature variability over ± 2 vertical layers.

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. While 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 lower 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 observational setup. The lower density of measurements near the bottom and below the thermocline resulted in a weak influence on the cost function in comparison with 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 (Canton) with monthly measurements, the high-frequency observations of the northern station, and hence its higher number of data points, lead to a calibration more focused on that geographical location within the lake.

Although this highlights 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 Lake Geneva model results

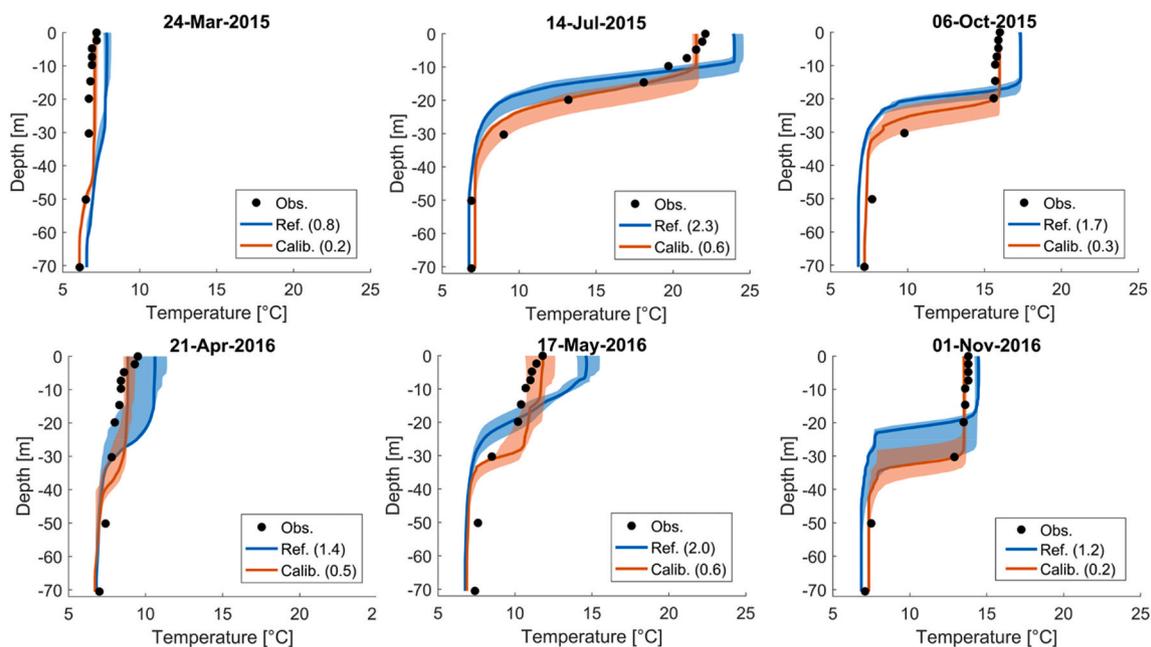


Fig. E1. Temperature profiles at 10 am at station GE3 (Fig. 1) with MAE in parenthesis. Shaded areas represent the modelled temperature variability over ± 36 h.

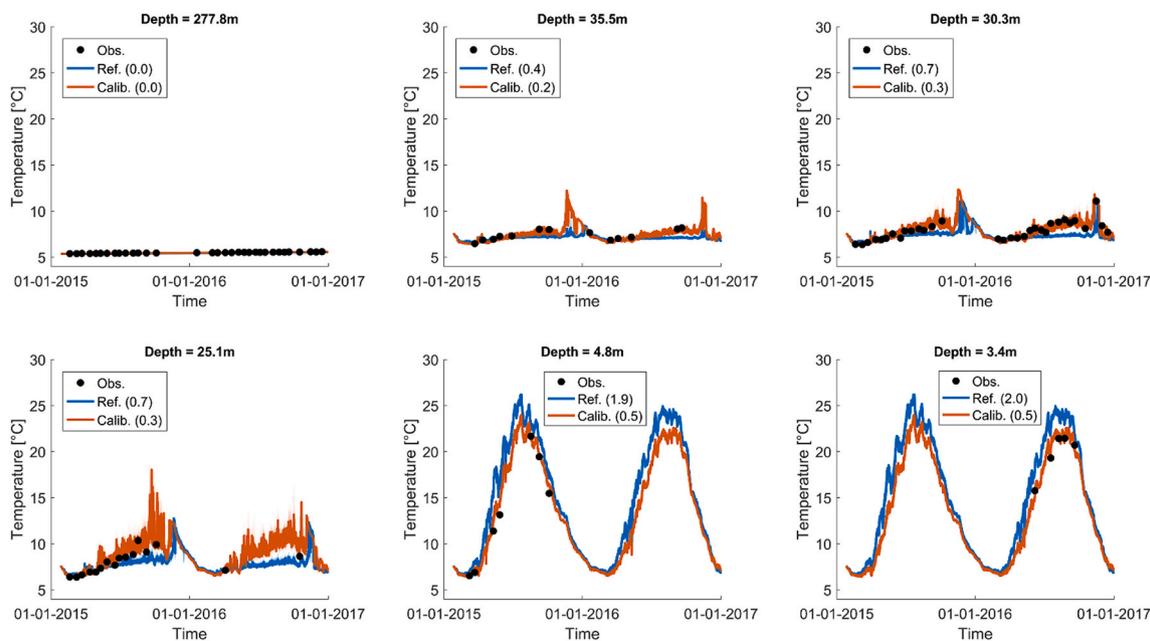


Fig. E2. Time-series of temperature at various depth for station SHL2 (Fig. 1) with MAE in parenthesis. Shaded areas represent the modelled temperature variability over ± 2 vertical layers.

References

Afshar, A., Kazemi, H., Saadatpour, M., 2011. Particle swarm optimization for automatic calibration of large scale water quality Model (CE-QUAL-W2): application to Karkheh Reservoir, Iran. *Water Resour. Manag.* 25, 2613–2632. <https://doi.org/10.1007/s11269-011-9829-7>.

Ahmadi, M., Arabi, M., Ascough, J.C., Fontane, D.G., Engel, B.A., 2014. Toward improved calibration of watershed models: multisite multiobjective measures of information. *Environ. Model. Software* 59, 135–145. <https://doi.org/10.1016/j.envsoft.2014.05.012>.

Baracchini, T., Bärenzung, K., Bouffard, D., Wüest, A., 2019. Le Lac de Zurich en Ligne - prévisions hydrodynamiques 3D en temps-réel sur meteolakes.ch. *Aqua & Gas - Fachzeitschrift für Gas, Wasser Abwasser* 99 (12), 24–29.

Baracchini, T., Chu, P.Y., Šukys, J., Lieberherr, G., Wunderle, S., Wüest, A., Bouffard, D., 2020a. Data assimilation of in situ and satellite remote sensing data to 3D hydrodynamic lake models: a case study using Delft3D-FLOW v4.03 and OpenDA v2.4. *Geosci. Model Dev. (GMD)* 13, 1267–1284. <https://doi.org/10.5194/gmd-13-1267-2020>.

Baracchini, T., Wüest, A., Bouffard, D., 2020b. Meteolakes: an operational online three-dimensional forecasting platform for lake hydrodynamics. *Water Res.* 172, 115529. <https://doi.org/10.1016/j.watres.2020.115529>.

Bard, Y., 1970. Comparison of gradient methods for the solution of nonlinear parameter estimation problems. *SIAM J. Numer. Anal.* 7, 157–186. <https://doi.org/10.1137/0707011>.

Bárdossy, A., Singh, S.K., 2008. Robust estimation of hydrological model parameters. *Hydrol. Earth Syst. Sci.* 12, 1273–1283.

- Beaton, A.E., Tukey, J.W., 1974. The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics* 16, 147–185. <https://doi.org/10.1080/00401706.1974.10489171>.
- Box, M.J., 1966. A comparison of several current optimization methods, and the use of transformations in constrained problems. *Comput. J.* 9, 67–77. <https://doi.org/10.1093/comjnl/9.1.67>.
- Bradley, E.L., 1973. The equivalence of maximum likelihood and weighted least squares estimates in the exponential family. *J. Am. Stat. Assoc.* 68, 199–200. <https://doi.org/10.1080/01621459.1973.10481364>.
- Charnes, A., Frome, E.L., Yu, P.L., 1976. The equivalence of generalized least squares and maximum likelihood estimates in the exponential family. *J. Am. Stat. Assoc.* 71, 169–171. <https://doi.org/10.1080/01621459.1976.10481508>.
- Chen, C., Huang, J., Chen, Q., Zhang, J., Li, Z., Lin, Y., 2019. Assimilating multi-source data into a three-dimensional hydro-ecological dynamics model using Ensemble Kalman Filter. *Environ. Model. Software* 117, 188–199. <https://doi.org/10.1016/j.envsoft.2019.03.028>.
- Chen, X., 2004. Using a piecewise linear bottom to fit the bed variation in a laterally averaged, z-co-ordinate hydrodynamic model. *Int. J. Numer. Methods Fluid.* 44, 1185–1205. <https://doi.org/10.1002/flid.680>.
- Cimattoribus, A., Lemmin, U., Barry, D.A., 2017. *Velocity and Temperature Data from Lake Geneva (CH), Collected during Year 2016*.
- Deltares, 2019. *The OpenDA Data-Assimilation Toolbox*.
- Deltares, 2015. *Delft3D-FLOW User Manual: Simulation of Multi-Dimensional Hydrodynamic Flows and Transport Phenomena*.
- El Serafy, G.Y., Gerritsen, H., Hummel, S., Weerts, A.H., Mynett, A.E., Tanaka, M., 2007. Application of data assimilation in portable operational forecasting systems—the DATools assimilation environment. *Ocean Dynam.* 57, 485–499. <https://doi.org/10.1007/s10236-007-0124-3>.
- Fabio, P., Aronica, G.T., Apel, H., 2010. Towards automatic calibration of 2-D flood propagation models. *Hydrol. Earth Syst. Sci.* 14, 911–924. <https://doi.org/10.5194/hess-14-911-2010>.
- Fenicia, F., Savenije, H.H.G., Matgen, P., Pfister, L., 2007. A comparison of alternative multiobjective calibration strategies for hydrological modeling. *Water Resour. Res.* 43, W03434. <https://doi.org/10.1029/2006WR005098>.
- Gallagher, M., Doherty, J., 2007. Parameter estimation and uncertainty analysis for a watershed model. *Environ. Model. Software* 22, 1000–1020. <https://doi.org/10.1016/j.envsoft.2006.06.007>.
- García, M., Ramirez, I., Verlaan, M., Castillo, J., 2015. Application of a three-dimensional hydrodynamic model for San Quintin Bay, B.C., Mexico. Validation and calibration using OpenDA. *J. Comput. Appl. Math.* 273, 428–437. <https://doi.org/10.1016/j.cam.2014.05.003>.
- Gaudard, A., Schwefel, R., Vinnä, L.R., Schmid, M., Wüest, A., Bouffard, D., 2017. Optimizing the parameterization of deep mixing and internal seiches in one-dimensional hydrodynamic models: a case study with Simstrat v1.3. *Geosci. Model Dev.* (GMD) 10, 3411–3423. <https://doi.org/10.5194/gmd-10-3411-2017>.
- Hendrickson, J.D., Sorooshian, S., Brazil, L.E., 1988. Comparison of Newton-type and direct search algorithms for calibration of conceptual rainfall-runoff models. *Water Resour. Res.* 24, 691–700. <https://doi.org/10.1029/WR024i005p0691>.
- 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., Kodés, V., Solheim, A.L., Nöges, T., Ormerod, S., Panagopoulos, Y., Schmutz, S., Venohr, M., Birk, S., 2015. Managing aquatic ecosystems and water resources under multiple stress — an introduction to the MARS project. *Sci. Total Environ.* 503 (504), 10–21. <https://doi.org/10.1016/j.scitotenv.2014.06.106>.
- Hughes, J.D., Dutta, D., Vaze, J., Kim, S.S.H., Podger, G., 2014. An automated multi-step calibration procedure for a river system model. *Environ. Model. Software* 51, 173–183. <https://doi.org/10.1016/j.envsoft.2013.09.024>.
- Imberger, J., Hamblin, P.F., 1982. Dynamics of lakes, reservoirs, and cooling ponds. *Annu. Rev. Fluid Mech.* 14, 153–187.
- Johnston, P.R., Pilgrim, D.H., 1976. Parameter optimization for watershed models. *Water Resour. Res.* 12, 477–486. <https://doi.org/10.1029/WR012i003p0477>.
- Kiefer, L., Odermatt, D., Anneville, O., Wüest, A., Bouffard, D., 2015. 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. <https://doi.org/10.1016/j.scitotenv.2015.05.011>.
- Kurniawan, A., Ooi, S.K., Gerritsen, H., Twigt, D., 2010. Calibrating the regional tidal prediction of the Singapore Regional Model using OpenDA. In: *Proc. 9th Int. Conf on Hydroinformatics*, pp. 7–11. Tianjin.
- Kurniawan, A., Ooi, S.K., Hummel, S., Gerritsen, H., 2011. Sensitivity analysis of the tidal representation in Singapore Regional Waters in a data assimilation environment. *Ocean Dynam.* 61, 1121–1136. <https://doi.org/10.1007/s10236-011-0415-6>.
- Lemmin, U., 2020. Insights into the dynamics of the deep hypolimnion of Lake Geneva as revealed by long-term temperature, oxygen, and current measurements: dynamics of the deep hypolimnion of Lake Geneva. *Limnol. Oceanogr.* <https://doi.org/10.1002/lno.11441>.
- Lemmin, U., Mortimer, C.H., Bäuerle, E., 2005. Internal seiche dynamics in Lake Geneva. *Limnol. Oceanogr.* 50, 207–216. <https://doi.org/10.4319/lo.2005.50.1.0207>.
- Li, T., Wang, G., Chen, J., Wang, H., 2011. Dynamic parallelization of hydrological model simulations. *Environ. Model. Software* 26, 1736–1746. <https://doi.org/10.1016/j.envsoft.2011.07.015>.
- MacIntyre, S., Melack, J.M., 1995. Vertical and horizontal transport in lakes: linking littoral, benthic, and pelagic habitats. *J. North Am. Benthol. Soc.* 14, 599–615. <https://doi.org/10.2307/1467544>.
- Madsen, H., 2003. Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. *Adv. Water Resour.* 26, 205–216.
- Madsen, H., 2000. Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. *J. Hydrol.* 235, 276–288.
- MeteoSwiss, 2019a. COSMO-1 – high-resolution forecasts for the Alpine region (accessed 2.21.19). <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>.
- MeteoSwiss, 2019b. COSMO-E – probabilistic forecasts for the Alpine region (accessed 2.21.19). <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>.
- Nelder, J.A., Baker, R.J., 2006. Generalized linear models. In: Kotz, S., Read, C.B., Balakrishnan, N., Vidakovic, B., Johnson, N.L. (Eds.), *Encyclopedia of Statistical Sciences*. John Wiley & Sons, Inc., Hoboken, NJ, USA <https://doi.org/10.1002/0471667196.ess0866.pub2>.
- Ralston, M.L., Jennrich, R.I., 1978. Dud, a derivative-free algorithm for nonlinear least squares. *Technometrics* 20, 7–14. <https://doi.org/10.2307/1268154>.
- Rouholahnejad, E., Abbaspour, K.C., Vejdani, M., Srinivasan, R., Schulin, R., Lehmann, A., 2012. A parallelization framework for calibration of hydrological models. *Environ. Model. Software* 31, 28–36. <https://doi.org/10.1016/j.envsoft.2011.12.001>.
- Shimoda, Y., Azim, M.E., Perhar, G., Ramin, M., Kenney, M.A., Sadraddini, S., Gudimov, A., Arhonditsis, G.B., 2011. Our current understanding of lake ecosystem response to climate change: what have we really learned from the north temperate deep lakes? *J. Great Lake. Res.* 37, 173–193. <https://doi.org/10.1016/j.jglr.2010.10.004>.
- Skahill, B.E., Doherty, J., 2006. Efficient accommodation of local minima in watershed model calibration. *J. Hydrol.* 329, 122–139. <https://doi.org/10.1016/j.jhydrol.2006.02.005>.
- Solomatine, D.P., Dibike, Y.B., Kukuric, N., 1999. Automatic calibration of groundwater models using global optimization techniques. *Hydrol. Sci. J.* 44, 879–894. <https://doi.org/10.1080/02626669909492287>.
- Sorooshian, S., Duan, Q., Gupta, V.K., 1993. Calibration of rainfall-runoff models: application of global optimization to the Sacramento Soil moisture Accounting model. *Water Resour. Res.* 29, 1185–1194. <https://doi.org/10.1029/92WR02617>.
- Sun, S., Bertrand-Krajewski, J.-L., 2012. On calibration data selection: the case of stormwater quality regression models. *Environ. Model. Software* 35, 61–73. <https://doi.org/10.1016/j.envsoft.2012.02.007>.
- Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.: Atmosphere* 106, 7183–7192. <https://doi.org/10.1029/2000JD900719>.
- Umlauf, L., Lemmin, U., 2005. Interbasin exchange and mixing in the hypolimnion of a large lake: the role of long internal waves. *Limnol. Oceanogr.* 50, 1601–1611. <https://doi.org/10.4319/lo.2005.50.5.1601>.
- van Velzen, N., Verlaan, M., 2007. COSTA a problem solving environment for data assimilation applied for hydrodynamical modelling. *Meteorol. Z.* 16, 777–793. <https://doi.org/10.1127/0941-2948/2007/0241>.
- Vidal, J.-P., Moisan, S., Faure, J.-B., Dartus, D., 2007. River model calibration, from guidelines to operational support tools. *Environ. Model. Software* 22, 1628–1640. <https://doi.org/10.1016/j.envsoft.2006.12.003>.
- Vörösmarty, C.J., Hoekstra, A.Y., Bunn, S.E., Conway, D., Gupta, J., 2015. Fresh water goes global. *Science* 349, 478–479. <https://doi.org/10.1126/science.aac6009>.
- Wagner, C., Adrian, R., 2009. Exploring lake ecosystems: hierarchy responses to long-term change? *Global Change Biol.* 15, 1104–1115. <https://doi.org/10.1111/j.1365-2486.2008.01833.x>.
- Wang, Y., Brubaker, K., 2015. Multi-objective model auto-calibration and reduced parameterization: exploiting gradient-based optimization tool for a hydrologic model. *Environ. Model. Software* 70, 1–15. <https://doi.org/10.1016/j.envsoft.2015.04.001>.
- Weerts, A.H., El Serafy, G.Y., Hummel, S., Dhondia, J., Gerritsen, H., 2010. Application of generic data assimilation tools (DATools) for flood forecasting purposes. *Comput. Geosci.* 36, 453–463. <https://doi.org/10.1016/j.cageo.2009.07.009>.
- Wu, Y., Liu, S., 2012. Automating calibration, sensitivity and uncertainty analysis of complex models using the R package Flexible Modeling Environment (FME): SWAT as an example. *Environ. Model. Software* 31, 99–109. <https://doi.org/10.1016/j.envsoft.2011.11.013>.
- Wüest, A., Lorke, A., 2003. Small-scale hydrodynamics in lakes. *Annu. Rev. Fluid Mech.* 35, 373–412. <https://doi.org/10.1146/annurev.fluid.35.101101.161220>.
- Zhang, X., Beeson, P., Link, R., Manowitz, D., Izaurralde, R.C., Sadeghi, A., Thomson, A.M., Sahajpal, R., Srinivasan, R., Arnold, J.G., 2013. Efficient multi-objective calibration of a computationally intensive hydrologic model with parallel computing software in Python. *Environ. Model. Software* 46, 208–218. <https://doi.org/10.1016/j.envsoft.2013.03.013>.