https://doi.org/10.5194/essd-2022-457
Preprint. Discussion started: 14 February 2023

A long-term dataset of simulated epilimnion and hypolimnion temperatures in 401 French lakes (1959-2020)

Najwa Sharaf[1,2], Jordi Prats[3], Nathalie Reynaud[1,2], Thierry Tormos[1,4], Tiphaine Peroux[1,2], Pierre-Alain Danis[1,4]

[1]Pôle R&D Ecosystèmes Lacustres (ECLA), OFB-INRAE-USMB, Aix-en-Provence, France
[2]INRAE, Aix Marseille Univ, RECOVER, Team FRESHCO, 3275 Route Cézanne, 13182 Aix-en-Provence, France
[3]Segula Technologies, C. Calàbria 169, 08015 Barcelona, Spain
[4]OFB, Service ECOAQUA, DRAS, 3275 Route Cézanne, 13100 Aix-en-Provence, France

Correspondence to: Sharaf Najwa (najwa.sharaf@inrae.fr), Tormos Thierry (thierry.tormos@ofb.gouv.fr), Reynaud Nathalie (nathalie.reynaud@inrae.fr)

https://doi.org/10.5194/essd-2022-457
Preprint. Discussion started: 14 February 2023

Open Access Earth System Science Data Discussions

1. Abstract

Understanding the thermal behavior of lakes is crucial for water quality management. Under climate change, lakes are warming and undergoing alterations in their thermal structure, including surface and deep-water temperatures. These changes require continuous monitoring due to the possible major ecological implications on water quality and lake processes. With the scarcity of long-term in situ water temperature datasets, we present a regional long-term water temperature dataset (LakeTSim: Lake Temperature Simulations) produced over 401 French lakes by combining numerical modelling and satellite thermal data. The dataset consists of daily epilimnion and hypolimnion temperatures for the period 1959-2020 simulated with the semi-empirical OKPLM (Ottosson-Kettle-Prats Lake Model). We also describe this model and its performance. We present the uncertainty analysis of simulations with default (parametrized with satellite thermal data over all lakes and in situ measurements) and calibrated (with in situ temperature measurements for each lake) model parameters as well as the sensitivity analysis of the latter. Overall, the 90% confidence uncertainty range is largest for hypolimnion temperature simulations with a median of 8.5 ºC and 2.32 ºC respectively with default and calibrated parameter values. There is less uncertainty associated with epilimnion temperature simulations with a median of 5.42 ºC and 1.85 ºC before and after parameter calibration. This dataset will help provide insight into the thermal functioning of French lakes. It provides over six decades of epilimnion and hypolimnion temperature data, crucial for climate change studies at a regional scale. The dataset will also be of great advantage for decision making by stakeholders.

2. Introduction

Lakes, both natural and artificial (i.e., reservoirs and gravel pits) are sentinels of environmental change and provide important services such as access to drinking water, hydropower production, recreation and fisheries (Adrian et al., 2009). Under climate change and anthropogenic pressures, many lakes are warming and consequently experiencing changes to their biophysicochemical structure and function that are leading to services being compromised (Janssen et al., 2021).

In lakes, water temperature is an essential parameter regulating processes such as the functioning of trophic webs, oxygen conditions, the physical structure of the water column as well as the biogeochemistry (Yang et al., 2018). Under warming, historical records and future projections demonstrate that for lakes, alterations in the thermodynamic functioning including warmer temperatures and shifts in mixing regimes already took place and are expected to persist in the future (Shatwell et al., 2019; Woolway and Merchant, 2019). In this context, they are undergoing shorter periods of ice cover and longer, more stable periods of thermal stratification (Woolway et al., 2022). These alterations could have considerable ecological implications for the biological communities (Lind et al., 2022; Havens and Jeppesen, 2018). For instance, worldwide studies have shown that the expansion of toxic cyanobacterial blooms is linked to warming (Griffith and Gobler, 2020). Other responses include species reduced body size (Daufresne et al., 2009), changes in thermal habitat and shifts in species seasonality (Kharouba et al., 2018).

For assessing the impact of climate change on lake ecosystems it is thus crucial to closely evaluate water temperature trajectories over the entire water column in space and time. However, long-term datasets of in situ temperatures are usually scarce and mostly limited to large lakes (Layden et al., 2015). Moreover, the sampling

https://doi.org/10.5194/essd-2022-457
Preprint. Discussion started: 14 February 2023

of water temperature differs in terms of approach and frequency, from decades (Piccolroaz et al., 2020) to a few years (Sharma et al., 2015), thus rendering it challenging to investigate warming trends.

Due to the difficulties in conventional in situ monitoring, which is often expensive, the coupling of modelling and satellite remote sensing data has become fundamental in the field of limnology (Nouchi et al., 2019). Modelling provides means to interpolate both temporal and spatial gaps. It thereby allows us to acquire information about surface water temperatures, which are globally the focus of lake climate change studies, and deep-water temperatures which are as critical though often disregarded in this context. Several numerical models that vary in complexity exist for conducting water temperature simulations, the most accurate being deterministic or process-based models. Nevertheless, regional or global deterministic modelling efforts over long periods are usually hindered by the lack of sufficiently detailed input data (e.g., meteorological and field data) to run the models (Kim et al., 2021). For practical and operational purposes, simpler models (semi-empirical, statistical or hybrid physical-statistical based models) with less requirements for forcing data, have been mostly applied to assess the impact of climate change on lake ecosystems and study them (Piccolroaz et al., 2020; Toffolon et al., 2014; Sharma et al., 2008). For conducting long-term simulations over a considerable number of lakes, this type of models is especially useful for detecting trends in time series, which with short datasets is not accurately achievable (Gray et al., 2018).

The performance of numerical models depends highly on the calibration of their parameters as well as on the quality of the input data. Satellite remote sensing is an effective way to monitor surface water temperature on a synoptic scale (Schaeffer et al., 2018; Sharaf et al., 2019) and provide a complementary source of data to in situ measurements for model calibration or validation purposes (Allan et al., 2016; Babbar-Sebens et al., 2013). In particular, thermal infrared sensors onboard the Landsat satellites are very adequate for retrospective analysis of surface water temperature with a spatial resolution adapted for small to medium size lakes and reservoirs at a bimonthly acquisition frequency. Landsat 4 and 5 TM (Thematic Mapper), 7 ETM+ (Enhanced Thematic Mapper) and 8 TIRS (Thermal InfraRed Sensor) provide surface temperature data at spatial resolutions of 120, 60 and 100 m respectively. Landsat series records of surface water temperature can be used to validate 3D hydrodynamic models when in situ measurements are scarce (Sharaf et al., 2021) and to spatially assess the quality and suitability of aquatic habitat for biological communities (Halverson et al., 2022). Although, satellite thermal data is limited to the surface, its integration into model calibration could improve the accuracy of simulations over the surface layer and the water column (Javaheri et al., 2016).

Here we present on a regional scale, a long-term dataset, LakeTSim (Lake Temperature Simulations), of daily epilimnion and hypolimnion temperature simulations for the period 1959-2020 over 401 French lakes monitored under the Water Framework Directive (WFD) including natural and artificial lakes, reservoirs and gravel pits. We present the OKPLM (Ottosson-Kettle-Prats Lake Model) used to produce water temperature simulations and its performance. Further, we provide the uncertainty analysis of simulations with default (parametrized with satellite thermal data over an entire set of lakes) and calibrated (with in situ temperature measurements for each lake) model parameter values as well as the sensitivity analysis for the latter. The goal of publishing this dataset is to provide new insight about surface and deep-water temperatures of lakes in France especially for those that are not monitored regularly through conventional methods. This long-term dataset is valuable for developing temperature indicators for identifying warming trends, extreme events and possible changes in the mixing regime

https://doi.org/10.5194/essd-2022-457
Preprint. Discussion started: 14 February 2023

among others. These indicators will contribute to assess the impact of climate change on lakes thermal functioning and its influence on the biological community structure and trophic webs.

3. Data and methodology

3.1. The OKP Lake Model description

The OKPLM (Ottosson-Kettle-Prats Lake Model) is a two-layer semi-empirical data model adapted from Kettle et al (2004) for the epilimnion module and Ottosson & Abrahamsson (1998) for the hypolimnion module. It was further modified in Prats & Danis (2019) and used to simulate daily epilimnion and hypolimnion temperatures of 401 French lakes. These modifications consisted mainly of simplifying the mixing algorithm used in Ottosson & Abrahamsson (1998) using a basic stability condition whereas for the epilimnion module a sinusoidal fit to average daily solar radiation was used instead of the theoretical clear-sky radiation. The OKPLM also runs on weekly and monthly frequencies. The regionalization of the parameters of the model mainly depends on the geographical and morphological properties of the lake (maximum depth, volume, surface area, latitude and altitude). The model requires few meteorological forcing data: solar radiation and air temperature.

The model calculates water temperature as follows:

$$T_{e,i} = A + Bf\left(T_{a,i}^{*}\right) + CS_i \quad (1)$$

where T_e is the epilimnion temperature (°C), i is the day number, A, B and C are calibration parameters, S is the solar radiation (W m^{-2}) and $f(*)$ is an exponential smoothing function with $T_{a,i}^{*}$ defined as:

$$T_{a,i}^{*} = T_{a,i} - MAAT \quad (2)$$

where $T_{a,i}$ is air temperature (°C) and $MAAT$ is the annual mean air temperature (°C).

$$T_{h,i} = A \cdot D + E \cdot g\left(T_{e,i}\right) \quad (3)$$

where T_h is the hypolimnion temperature (°C), D and E are calibration parameters and $g\left(T_{e,i}\right)$ is an exponential smoothing of T_e.

The OKPLM is integrated into a Python 3 package, "ALAPROD" (A LAke MODElling project-PRODuction) which for the present study was used to simulate epilimnion and hypolimnion water temperatures (Danis, 2020). ALAPROD is part of a software environment called ALAMODE. In addition to lake water temperature, this package can also be used to make simulations of stream water temperature, hydrodynamics and stream flow rates. In this package OKPLM can be run in two modes: the "default" mode where model parameters use the parameterization presented in Prats & Danis (2019), and the "calibrated" mode where model parameters are calibrated individually for each lake by using in situ temperature measurements. The "default" parameterization provides expressions of the model parameters as a function of lake characteristics (latitude, altitude, surface, volume, depth). The expressions for epilimnion module parameters were derived using surface temperatures estimated from Landsat infrared data acquired between 1999 and 2016 over French lakes (Prats et al., 2018), while the parameterization of hypolimnion parameters was derived from temperature profile data of 357 lakes.

https://doi.org/10.5194/essd-2022-457
Preprint. Discussion started: 14 February 2023

3.2. Input data

The OKPLM was forced with two sources of meteorological data extracted from the SAFRAN (Système d'Analyse Fournissant des Renseignements Adaptés à la Nivologie) analysis system (Durand et al., 1993) and the S2M (SAFRAN–SURFEX/ISBA–Crocus–MEPRA) meteorological reanalysis (Vernay et al., 2015, 2022).

The SAFRAN system provides meteorological variables at an hourly time step estimated through interpolation and assimilation processes with an 8 km square grid. Average daily data from the nearest grid cell was selected for each study site. The difference in altitude between the study site and the grid cell was accounted for by applying an adiabatic elevation correction on air temperature.

The S2M model chain combines the SAFRAN meteorological analysis and the SURFEX/ISBA–Crocus snow cover model including MEPRA **(**Modèle Expert d'Aide à la Prévision du Risque d'Avalanche). It is more adapted to mountainous regions as it has a spatial definition where spatial heterogeneity is taken into consideration. The S2M reanalysis uses a vertical resolution of 300 m and is the result of simulations performed over mountainous zones called "massifs" each corresponding approximately to an average surface of 1000 km^2. These massifs represent the spatial variability of processes in mountainous regions. Average daily data was used for each study site.

In situ temperature profiles, geographical and morphological data of the study sites were extracted from the PLAN_DEAU database managed by INRAE (l'Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement) and R&D consortium ECLA (ECosystèmes LAcustres) at Aix-en-Provence, France.

3.3. Lake simulations

For this study, we considered 401 lakes (Figure 1) located in Metropolitan France monitored according to the Water Framework Directive (WFD). Here we refer to lakes as natural lakes, reservoirs, artificial lakes and gravel pits. The present lake dataset includes 54 natural lakes, 302 reservoirs, 38 artificial lakes and 7 gravel pits that have characteristics ranging between 0 and 2279.7 m for altitude, 0.8 and 309.7 m for maximum depth, 0.08 and 577.12 km^2 for surface area and 5×10^4 and 8.9×10^{10} m^3 for volume.

The OKPLM was run using "default" and "calibrated" parameters with two sources of meteorological data "SAFRAN" and "S2M" over specific sets of lakes. Among the total number of study sites (n=401), the model was forced using SAFRAN and S2M meteorological data respectively for 210 and 21 lakes with "default" model parameters, and for 164 and 6 lakes with "calibrated" model parameters. The geomorphological characteristics of the simulated lakes with each of the abovementioned configurations are shown in Table 1. "Calibrated" model parameters are adopted when in situ temperature measurements are available; conversely, "default" parameters are used. S2M data are more representative of mountainous meteorological conditions than SAFRAN data and were thus used for simulating the water temperature in lakes situated at altitudes higher than 900 m.

https://doi.org/10.5194/essd-2022-457
Preprint. Discussion started: 14 February 2023

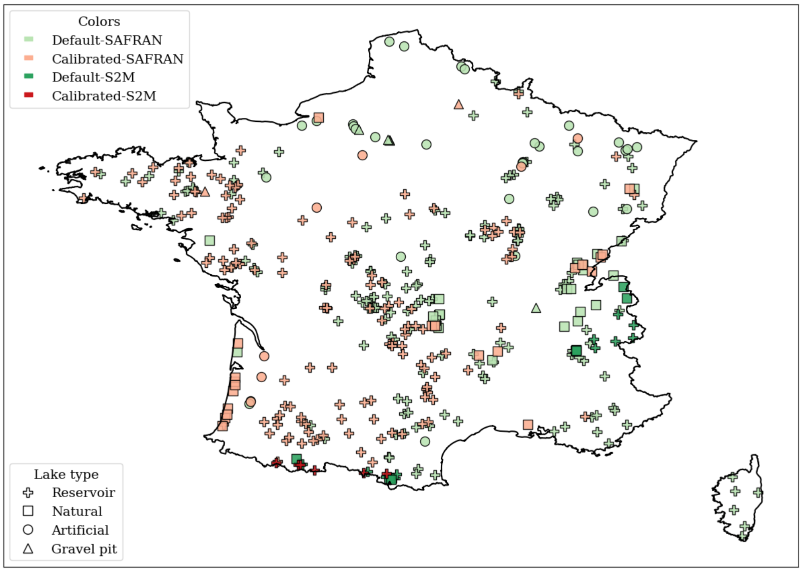

Figure 1 : Location and lake type of the 401 French lakes simulated with the OKPLM in "default" and "calibrated" modes, with SAFRAN and S2M meteorological data for the period 1959-2020.

Table 1: Characteristics of the lakes simulated with the OKPLM in "default" and "calibrated" modes with SAFRAN and S2M meteorological data for the period 1959-2020; *n* represents the number of lakes.

Variables	**Minimal - Maximal range**			
Model parameters	**Default**		**Calibrated**	
Meteorological data	**SAFRAN**	**S2M**	**SAFRAN**	**S2M**
n	210	21	164	6
Altitude (m)	1 - 1753	916 - 2213	0 - 2279.7	1577.5 - 2172.5
Latitude (°N)	41.47 - 50.87	42.55 - 46.21	42.88 - 49.87	42.65 - 42.86
Longitude (°E)	-3.90 - 9.48	0.08 - 6.94	-4.24 - 6.96	-0.33 - 1.92
Maximal depth (m)	0.8 - 309.7	10.3 - 180	1.2 - 124	49 - 112
Surface area (km^2)	0.08 - 577.12	0.11 - 6.52	0.29 - 57.57	0.45 - 1.21
Volume (m^3)	5×10^4 – 8.9×10^{10}	51.4×10^4 - 33.32×10^7	12.9×10^4 - 49.88×10^7	72.7×10^5 - 68.6×10^6

3.4. Calibration, uncertainty and sensitivity analysis

The initial assessment of the quality of OKPLM simulations described in the previous section has been completed with a sensitivity and uncertainty analysis. For calibration and uncertainty analysis we used the

https://doi.org/10.5194/essd-2022-457
Preprint. Discussion started: 14 February 2023

package "CUSPY" (Calibration, Uncertainty analysis and Sensitivity analysis in PYthon), which is part of the software environment "ALAMODE" (Danis, 2020) and acts as an interface to the package "pyemu" (White et al., 2016).

Parameter values were calibrated for lakes with enough available in situ data (temperature profiles and bathymetry). Parameter values were calibrated using the Gauss-Levenberg-Marquardt algorithm and Tikhonov regularization (White et al., 2020). In addition to the calibrated parameter values, the calibration process also provided posterior parameter uncertainty and composite scaled sensitivities. Composite scaled sensitivities (CSS) indicate the quantity of information provided by each parameter and the sensitivity of the model to them (Ely, 2006).

The uncertainty of the simulations (calibrated and default) was analyzed using Monte Carlo simulations. For each lake, 100 Monte Carlo simulations were carried by randomly selecting the value of the model parameters. Two parameters, at_factor and sw_factor, multiplying the meteorological input, were added to account for possible uncertainties in input data. For default simulations, the a priori distribution of the parameters was assumed to follow a normal distribution with the average value and lower and upper bounds shown in Table 2. The ranges for parameters A, B and C were estimated as four times the standard deviation of the residuals of the formulas used to estimate them according to Prats & Danis (2019). For D, E and β, given their higher uncertainty, the full 0-1 range was explored. For $MAAT$, at_factor and sw_factor, reasonable ranges (±10%) were chosen to account for meteorological data uncertainty (measurement error, errors in regionalization, etc.). For calibrated simulations, the distribution of the parameters was obtained from the calibration results.

Table 2: Characteristics of the a priori distributions of the model parameters. Parameters with tilde indicate parameter values estimated for a particular lake according to the regionalization formulas by Prats & Danis (2019).

Parameter	Average value	Lower bound	Upper bound
$\boldsymbol{A}$	$\tilde{A}$	$\tilde{A} - 2 \cdot 0.74$	$\tilde{A} + 2 \cdot 0.74$
$\boldsymbol{B}$	$\tilde{B}$	$\tilde{B} - 2 \cdot 0.08$	$\tilde{B} + 2 \cdot 0.08$
$\boldsymbol{C}$	$\tilde{C}$	$\tilde{C} - 2 \cdot 0.004$	$\tilde{C} + 2 \cdot 0.004$
$\boldsymbol{D}$	$\tilde{D}$	0	1
$\boldsymbol{E}$	$\tilde{E}$	0	1
$\boldsymbol{\alpha}$	$\tilde{\alpha}$	0	$\tilde{\alpha} + 2 \cdot 0.08$
$\boldsymbol{\beta}$	$\tilde{\beta}$	0	1
$\boldsymbol{MAAT}$	$\widetilde{MAAT}$	$\widetilde{MAAT} - 2 \cdot 0.5$	$\widetilde{MAAT} + 2 \cdot 0.5$
$\boldsymbol{at\_factor}$	1	0.9	1.1
$\boldsymbol{sw\_factor}$	1	0.9	1.1

https://doi.org/10.5194/essd-2022-457
Preprint. Discussion started: 14 February 2023

4. Model performance

The performance of the OKPLM was assessed in Prats & Danis (2019) by comparing its performance to two other often-applied models in lake studies, air2water and FLake. The air2water model is a semi-empirical model used to calculate the epilimnion temperature of temperate lakes (Toffolon et al., 2014). Flake is a one-dimensional (1D) hydrodynamic lake model for simulating temperature vertical profiles and mixing conditions in lakes (Mironov, 2008). To assess their performances, both models were run with default parameter values between 1999 and 2016 over a set of 409 French lakes of different types (reservoirs, natural lakes, artificial lakes and gravel pits) with temperature measurements including five lakes with continuous profile measurements. Meteorological forcing (SAFRAN) consisted of air temperature for the air2water model in addition to solar radiation, vapor pressure, cloud cover and wind speed for Flake.

The OKPLM, air2water and FLake simulations were assessed through comparison to in situ measurements. For epilimnion temperatures, the average discrepancies calculated between OKPLM simulations and observations remained below 2 °C in most cases, in contrast to the air2water and Flake models. The performance comparison between the OKPLM, air2water and FLake yielded respectively median RMSE's (Root Mean Square Error) of 1.7, 2.3 and 2.6 °C calculated between simulations and observations of epilimnion water temperature. For hypolimnion temperatures, the median RMSEs by lake type obtained with OKPLM simulations remained below 2 °C, except for gravel pits (RMSE = 2.7 °C) and reservoirs (RMSE = 2.3 °C), whereas FLake yielded a median RMSE of 3.3 °C. For the epilimnion, the differences between the RMSE of lake types were not significant. In terms of depth, discrepancies between epilimnion temperature simulations with the OKPLM and measurements were highest for lakes with a depth > 10 m and for ponds around 1 m deep. The OKPLM simulations were also evaluated seasonally, in particular during summer and winter. The model simulated temperatures well with a median RMSE of 1.4 and 1.6 °C in summer and winter respectively.

5. Uncertainty analysis

The uncertainty analysis revealed that, overall, for both simulations with default and calibrated model parameters; uncertainty was higher and recurrent for hypolimnion temperature compared to epilimnion temperature especially in reservoirs (Figure 2). In default simulations, the uncertainty of simulated values showed a clear relation with lake maximal depth (Figure 2). For epilimnion temperature, uncertainty increased with maximal depth in particular for lakes with depths greater than 10 m. For hypolimnion temperature, uncertainty was maximal for lakes with depths around 10 m.

After calibration, there was an important reduction in simulation uncertainty. For default simulations of epilimnion temperature the median of the 90% confidence uncertainty range was 5.42 ºC, while after calibration it was 1.85 ºC. For hypolimnion temperature, the median of the 90% confidence uncertainty range of default simulations was 8.5 ºC, while it was 2.32 ºC after calibration. However, many reservoirs with depths greater than 8 m still had a much greater uncertainty (uncertainty range > 4 ºC) than the rest of lakes after calibration.

https://doi.org/10.5194/essd-2022-457
Preprint. Discussion started: 14 February 2023

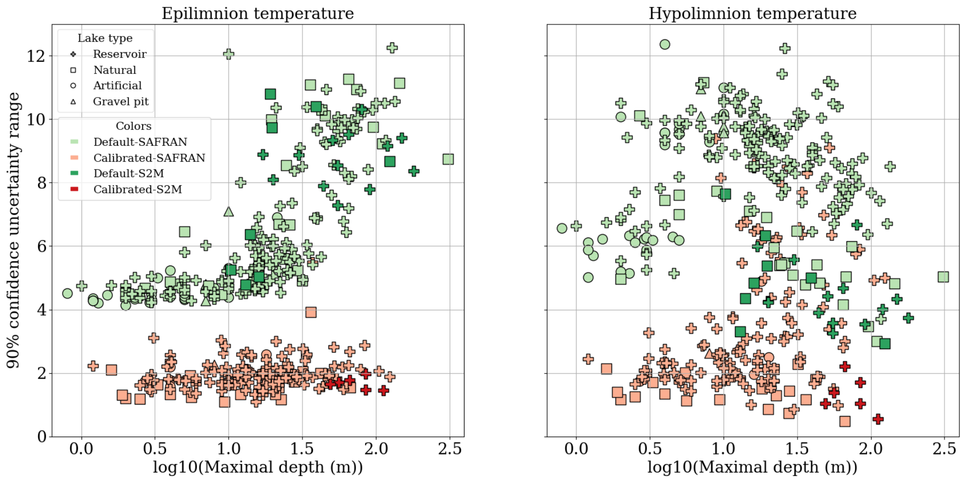

Figure 2: Average 90% confidence uncertainty range for epilimnion and hypolimnion temperatures in calibrated (*n* = 170) and default (*n* = 231) simulations for the period 1959-2020.

6. Sensitivity analysis

The parameter to which the model was most sensitive was the parameter *C* (Figure 3), which multiplies solar radiation in Eq. (1). The CSS for *C* were an order of magnitude greater than for the next parameters with highest CSS, the parameter *α* and *at_factor*, both influencing the effect of air temperature on simulated water temperature. Other parameters to which the model was somewhat sensitive were *E*, *B* and *β*. The model was quite insensitive to *sw_factor*, *MAT* and *A*. The parameter *D*, with CSS several orders of magnitude smaller than the other parameters, was unidentifiable.

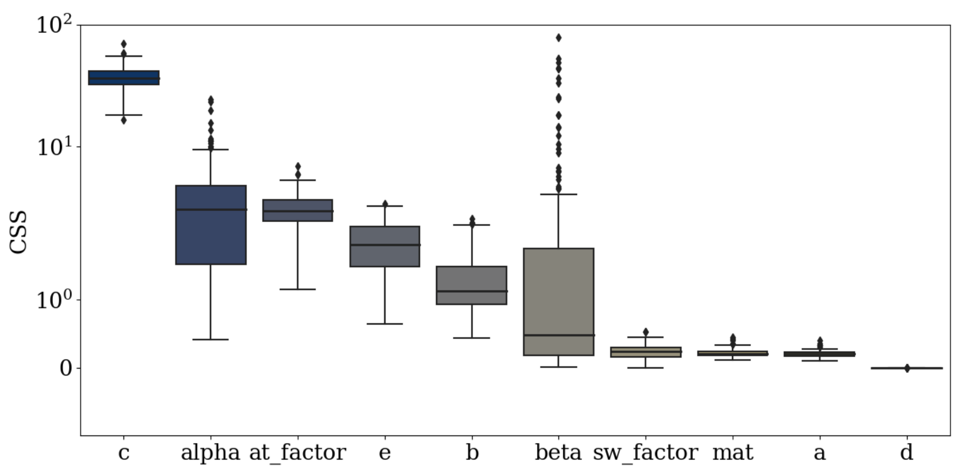

Figure 3: Composite scaled sensitivities (CSS) for each parameter. The boxplots indicate the distribution of CSS between the simulations calibrated for different lakes. The y-axis is in logarithmic form.

https://doi.org/10.5194/essd-2022-457
Preprint. Discussion started: 14 February 2023

The model tended to be more sensitive to the parameter values in the case of reservoirs than in the case of natural lakes (Figure 4). Some parameters (α, β) also showed a dependency on lake depth. The increase of model sensitivity to the parameter α with depth can be related with the increase in uncertainty with depth in the default simulations. In the case of the parameter β, CSS were mostly low, with a median value of 0.49. However, for some reservoirs and artificial water bodies CSS could attain very high values.

Although the model in general was not very sensitive to the values of the parameters most directly related with hypolimnion temperatures (D, E, β), the quality of hypolimnion temperature was greatly improved through calibration. This would seem to indicate that the quality of simulated hypolimnion temperature was improved through the improvement of epilimnion temperature simulations.

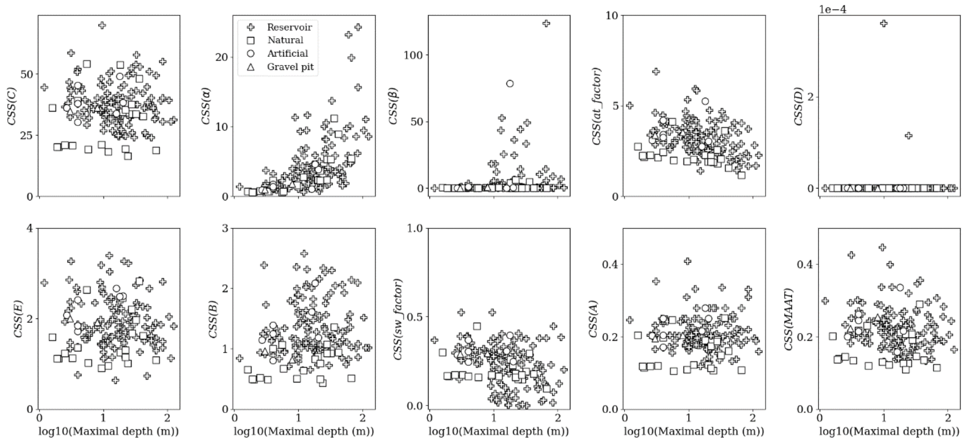

Figure 4 : Composite scaled sensitivities (CSS) for each model parameter as a function of maximal depth.

7. Discussion and implications

Lakes are undeniably changing under climate change and long-term future projections show that the shifts in ecosystem functioning will continue with aggravated alterations. In particular, given the key role of warming lake water temperature in regulating ecosystem processes, its warming has become a response that is crucial to monitor, explore and understand. Hence, the importance of developing or adopting approaches, such as numerical models, that will provide long-term information about water temperature and allow us to understand the thermal response of lakes to climate change.

Here we used a semi-empirical model, the OKPLM, to simulate six decades of epilimnion and hypolimnion water temperatures in French lakes. In comparison to similar models, overall, the OKPLM provides acceptable estimations of water temperatures, with better results for epilimnion temperatures. The values of the RMSEs provided in Prats & Danis (2019) and obtained between OKPLM simulations and observations are comparable to values found in studies applying complex hydrodynamic lake models (Read et al., 2014; Fang et al., 2012). The analysis revealed that the uncertainty associated with both epilimnion and hypolimnion temperature simulations was highly related to maximal lake depth. The uncertainty in hypolimnion simulations is more important and especially associated with reservoirs having maximal depths > 8 m. The calibration of model parameters significantly reduced the uncertainties yet, for hypolimnion temperatures, they remained considerably high and increased with depth especially in reservoirs.

The high levels of uncertainty found in reservoirs could be somewhat attributed to the lack of consideration of water level fluctuations in the model. In contrast to other waterbodies (e.g., natural lakes, artificial lakes and gravel pits) reservoirs experience significant variations in their water level, which influences the heat budget and hence their thermal regime. Therefore, even under similar meteorological conditions lakes and reservoirs could have different thermal behaviors (Nowlin et al., 2004). In reservoirs, the discharge depth is a driver of thermal structure. Deep discharges could contribute to warmer bottom waters (Carr et al., 2020) whereas in some cases if the reservoir is shallow or if the discharge depth is not deep, it could demonstrate lake-like thermal behavior. This does not necessarily mean that, in this case, the entire functioning of the reservoir resembles one of a natural lake; there are still differences to consider (Detmer et al., 2022).

The application of the OKPLM should be made with caution given its performance and depending on the objective of the study. The model does not take into account a complete set of meteorological forcing (e.g., with cloud cover, relative humidity and wind speed and direction) or other variables (e.g., inflow and outflow rates or water level fluctuations, inflow discharge depth and inflow temperature) that could influence the thermal structure of the ecosystem (Yang et al., 2020; Carr et al., 2020). Furthermore, the OKPLM was parametrized for a specific set of lakes with particular geomorphological characteristics. Thus, it would be advisable to apply the model over lakes with similar characteristics. If the aim is to conduct a long-term regional or global study for studying general patterns of climate change impacts over a large number of study sites, the utilization of semi-empirical models such as the OKPLM is the most suitable choice. Although complex, deterministic or process-based models provide a more accurate representation of thermal conditions, applying these models over several study sites and for long periods is usually hindered by the scarcity of the required input data. The increased complexity of these models (with reference to an increased number of model parameters) is beneficial for representing additional ecosystem processes. Yet the greater number of model parameters, increases the sensitivity of models and demands more calibration efforts (Lindenschmidt, 2006). Furthermore, a reduction in model errors is sometimes associated with an increased complexity in model structure; however, this is not always consistent since a complex model does not necessarily provide better estimations and thus lower errors than a simple model (Snowling and Kramer, 2001).

Our goal in publishing the present dataset is to expand knowledge about the water temperature of French lakes and provide data, with enough details and reliability, that it could be implemented in different studies where water temperature is implicated for understanding specific processes or interactions, in particular under climate change. Hence the significance of the present findings. The present study, making use of a semi-empirical model to provide long-term data about water temperature, was necessary for several reasons. Equipping a large number of lakes with thermal sensors is challenging and labor-intensive, it comes with a high financial cost that is often not available. Consequently, historical and even current water temperature datasets are often scarce, which can be problematic for studying the impact of climate change, as it requires high frequency data over a long duration of time for accurate analysis. In general, the higher the sampling frequency and duration, the better the data is suited to estimate or analyze specific processes or warming trends. The sampling frequency and length of a dataset have been shown to play a role in determining the accuracy of estimating warming trends where time

https://doi.org/10.5194/essd-2022-457
Preprint. Discussion started: 14 February 2023

series longer than 30 years seem to be the most appropriate (Gray et al., 2018). Although, the duration and frequency of a dataset have a major role in reflecting accurate representations, their influence is scarcely addressed when it comes to climate change studies related to warming trends in water temperature.

This dataset is very useful for climate change studies; it could be used for developing and analyzing several temperature indicators (e.g., annual or seasonal maximal and minimal temperature values, temperature exceeding certain thresholds with biological implications, etc.). Further, mixing and stratification dynamics are important to characterize as they drive lake biogeochemistry. Among other processes, they influence the distribution of nutrients, primary productivity and the composition of phytoplankton and zooplankton communities along the water column (Judd et al., 2005). With the LakeTSim dataset, it is possible to classify the mixing regime of lakes and investigate possible triggers of regime shifts.

8. Code availability

The respective codes for the "CUSPY" (Prats-Rodríguez and Danis, 2023a) and "OKPLM" (Prats-Rodríguez and Danis, 2023b) packages, which can be used to conduct sensitivity and uncertainty analysis and to run the OKP Lake Model, are available at https://github.com/inrae/ALAMODE-cuspy and https://github.com/inrae/ALAMODE-okp as well as ZENODO.

9. Data availability

The LakeTSim dataset (Sharaf et al., 2023) for epilimnion and hypolimnion water temperature simulations and supporting information are available at doi:10.57745/OF9WXR . The file "00_Data_description.txt" contains a description of the dataset. The geographical (longitude and latitude) and morphological (surface area, volume and maximum depth) data for the 401 lakes are presented in the file "01_Lake_data.txt" in addition to the name, type, altitude and the identification code for each lake. The data for daily epilimnion (tepi) and hypolimnion (thyp) temperatures simulated with the OKPLM are presented in text files available in the folders "02_LakeTSim_SAFRAN_OKPdefault_data", "03_LakeTSim_SAFRAN_OKPcalibrated_data", "04_LakeTSim_S2M_OKPdefault_data" and "05_LakeTSim_S2M_OKPcalibrated_data". Each file within these folders is named according to the identification code of the lake.

10. Conclusions

We present the LakeTSim dataset and the semi-empirical OKP Lake Model for simulating water temperature in Lakes. We applied the model over a set of 401 French lakes for the period 1959-2020 to derive daily simulations of epilimnion and hypolimnion water temperatures, here referred to as the LakeTSim dataset. Previous efforts to assess the model's performance show an overall acceptable representation of epilimnion and hypolimnion temperatures when compared to in situ measurements. The uncertainty analysis of simulations demonstrates that more uncertainty is associated firstly with default simulations, secondly hypolimnion compared to epilimnion temperatures and, thirdly deep lakes in particular reservoirs (maximal depth > 8 m). Although the calibration significantly decreases the uncertainties related to both the epilimnion and hypolimnion, in some cases they are still considerable for the latter. Based on these results and if enough observation data are available, optimally we recommend the use of the OKPLM over shallow lakes with calibrated model parameters. However, if applied in its default or even calibrated configuration over deep lakes, one should be aware of the presented limitations and address them in the analysis. The LakeTSim dataset is valuable for assessing the impact of climate change on

https://doi.org/10.5194/essd-2022-457
Preprint. Discussion started: 14 February 2023

lakes thermal functioning, which is often hindered by the lack of water temperature observations. The present dataset will provide new insights about the thermal behavior of French lakes. This will be of great advantage for stakeholders, as it should allow them to take better management strategies under climate change.

11. Author contributions

NS wrote the original manuscript with input from JP and PAD. NS, JP and PAD discussed the results. JP developed and carried out the implementation of the OKP Lake Model and the uncertainties computation in ALAMODE. JP and NS performed the simulations and provided uncertainty analysis results with SAFRAN and S2M data respectively. JP and NS implemented respectively the integration of SAFRAN and S2M data in ALAMODE. NS prepared the LakeTSim dataset. JP provided the uncertainty and sensitivity analysis. PAD designed, contributed and supervised the implementation of S2M data in ALAMODE for forcing the OKPLM when simulating high altitude lakes. PAD supervised the findings of this work. NR and TT supervised and contributed to the implementation of simulation results in the database. NR processed S2M data. NR and TT prepared the doi for the LakeTSim dataset. TP conducted the fieldwork for the monitoring, acquisition and verification of in situ temperature data. All authors reviewed, edited and approved the final paper.

12. Competing interests

The authors declare that they have no conflict of interest.

13. Acknowledgments

The authors thank Météo-France for providing SAFRAN and S2M meteorological data, Matthieu Vernay for his feedback on the utilization of S2M data and the “réseau lacs sentinelles” for providing bathymetry data.

14. Financial support

The authors were supported by OFB (Office Français de la Biodiversité), SEGULA Technologies, INRAE (Institut National de Recherche pour l’Agriculture, l’Alimentation et l’environnement) and Pôle ECLA (ECosystèmes LAcustres).

15. 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, Limnol. Oceanogr., 54, 2283–2297, https://doi.org/10.4319/lo.2009.54.6_part_2.2283, 2009.

Allan, M. G., Hamilton, D. P., Trolle, D., Muraoka, K., and McBride, C.: Spatial heterogeneity in geothermally-influenced lakes derived from atmospherically corrected Landsat thermal imagery and three-dimensional hydrodynamic modelling, Int. J. Appl. Earth Obs. Geoinf., 50, 106–116, https://doi.org/10.1016/j.jag.2016.03.006, 2016.

Babbar-Sebens, M., Li, L., Song, K., and Xie, S.: On the Use of Landsat-5 TM Satellite for Assimilating Water Temperature Observations in 3D Hydrodynamic Model of Small Inland Reservoir in Midwestern US, Adv. Remote Sens., 02, 214–227, https://doi.org/10.4236/ars.2013.23024, 2013.

Carr, M. K., Sadeghian, A., Lindenschmidt, K. E., Rinke, K., and Morales-Marin, L.: Impacts of Varying Dam Outflow Elevations on Water Temperature, Dissolved Oxygen, and Nutrient Distributions in a Large Prairie Reservoir, Environ. Eng. Sci., 37, 78–97, https://doi.org/10.1089/ees.2019.0146, 2020.

Danis, P.: Rapport d ’ avancement sur les outils de modélisation thermique, 1–8 pp., https://doi.org/hal-03253847, 2020.

Daufresne, M., Lengfellner, K., and Sommer, U.: Global warming benefits the small in aquatic ecosystems, Proc.

Natl. Acad. Sci., 106, 12788–12793, https://doi.org/10.1073/pnas.0902080106, 2009.

Detmer, T. M., Parkos, J. J., and Wahl, D. H.: Long-term data show effects of atmospheric temperature anomaly and reservoir size on water temperature, thermal structure, and dissolved oxygen, Aquat. Sci., 84, 1–13, https://doi.org/10.1007/s00027-021-00835-2, 2022.

Durand, Y., Brun, E., Merindol, L., Guyomarc'h, G., Lesaffre, B., and Martin, E.: A meteorological estimation of relevant parameters for snow models, Ann. Glaciol., 18, 65–71, https://doi.org/10.3189/S0260305500011277, 1993.

Ely, D. M.: Analysis of Sensitivity of Simulated Recharge to Selected Parameters for Seven Watersheds Modeled Using the Precipitation-Runoff Modeling System, U.S. Geological Survey Scientific Investigations Rep, 21 pp., 2006.

Fang, X., Alam, S. R., Stefan, H. G., Jiang, L., Jacobson, P. C., and Pereira, D. L.: Simulations of water quality and oxythermal cisco habitat in Minnesota lakes under past and future climate scenarios, Water Qual. Res. J. Canada, 47, 375–388, https://doi.org/10.2166/wqrjc.2012.031, 2012.

Gray, D. K., Hampton, S. E., O'Reilly, C. M., Sharma, S., and Cohen, R. S.: How do data collection and processing methods impact the accuracy of long-term trend estimation in lake surface-water temperatures?, Limnol. Oceanogr. Methods, 16, 504–515, https://doi.org/10.1002/lom3.10262, 2018.

Griffith, A. W. and Gobler, C. J.: Harmful algal blooms: A climate change co-stressor in marine and freshwater ecosystems, Harmful Algae, 91, 101590, https://doi.org/10.1016/j.hal.2019.03.008, 2020.

Halverson, G. H., Lee, C. M., Hestir, E. L., Hulley, G. C., Cawse-Nicholson, K., Hook, S. J., Bergamaschi, B. A., Acuña, S., Tufillaro, N. B., Radocinski, R. G., Rivera, G., and Sommer, T. R.: Decline in Thermal Habitat Conditions for the Endangered Delta Smelt as Seen from Landsat Satellites (1985–2019), Environ. Sci. Technol., 56, 185–193, https://doi.org/10.1021/acs.est.1c02837, 2022.

Havens, K. and Jeppesen, E.: Ecological responses of lakes to climate change, Water, 10, 917, https://doi.org/10.3390/w10070917, 2018.

Janssen, A. B. G., Hilt, S., Kosten, S., de Klein, J. J. M., Paerl, H. W., and Van de Waal, D. B.: Shifting states, shifting services: Linking regime shifts to changes in ecosystem services of shallow lakes, Freshw. Biol., 66, 1–12, https://doi.org/10.1111/fwb.13582, 2021.

Javaheri, A., Babbar-Sebens, M., and Miller, R. N.: From skin to bulk: An adjustment technique for assimilation of satellite-derived temperature observations in numerical models of small inland water bodies, Adv. Water Resour., 92, 284–298, https://doi.org/10.1016/j.advwatres.2016.03.012, 2016.

Judd, K. E., Adams, H. E., Bosch, N. S., Kostrzewski, J. M., Scott, C. E., Schultz, B. M., Wang, D. H., and Kling, G. W.: A case history: Effects of mixing regime on nutrient dynamics and community structure in third sister lake, michigan during late winter and early spring 2003, Lake Reserv. Manag., 21, 316–329, https://doi.org/10.1080/07438140509354437, 2005.

Kettle, H., Thompson, R., Anderson, N. J., and Livingstone, D. M.: Empirical modeling of summer lake surface temperatures in southwest Greenland, Limnol. Oceanogr., 49, 271–282, https://doi.org/10.4319/lo.2004.49.1.0271, 2004.

Kharouba, H. M., Ehrlén, J., Gelman, A., Bolmgren, K., Allen, J. M., Travers, S. E., and Wolkovich, E. M.: Global shifts in the phenological synchrony of species interactions over recent decades, Proc. Natl. Acad. Sci. U. S. A., 115, 5211–5216, https://doi.org/10.1073/pnas.1714511115, 2018.

Kim, J., Seo, D., Jang, M., and Kim, J.: Augmentation of limited input data using an artificial neural network method to improve the accuracy of water quality modeling in a large lake, J. Hydrol., 602, 126817, https://doi.org/10.1016/j.jhydrol.2021.126817, 2021.

Layden, A., Merchant, C., and Maccallum, S.: Global climatology of surface water temperatures of large lakes by remote sensing, Int. J. Climatol., 35, 4464–4479, https://doi.org/10.1002/joc.4299, 2015.

Lind, L., Eckstein, R. L., and Relyea, R. A.: Direct and indirect effects of climate change on distribution and community composition of macrophytes in lentic systems, Biol. Rev., 1686, 1677–1690, https://doi.org/10.1111/brv.12858, 2022.

Lindenschmidt, K. E.: The effect of complexity on parameter sensitivity and model uncertainty in river water

https://doi.org/10.5194/essd-2022-457
Preprint. Discussion started: 14 February 2023

quality modelling, Ecol. Modell., 190, 72–86, https://doi.org/10.1016/j.ecolmodel.2005.04.016, 2006.

Mironov, D. V.: Parameterization of lakes in numerical weather prediction: Description of a lake model, Encyclopedic Dictionary of Archaeology, Offenbach, Germany: DWD, 2008.

Nouchi, V., Kutser, T., Wüest, A., Müller, B., Odermatt, D., Baracchini, T., and Bouffard, D.: Resolving biogeochemical processes in lakes using remote sensing, Aquat. Sci., 81, 1–13, https://doi.org/10.1007/s00027-019-0626-3, 2019.

Nowlin, W. H., Davies, J. M., Nordin, R. N., and Mazumder, A.: Effects of water level fluctuation and short-term climate variation on thermal and stratification regimes of a British Columbia reservoir and lake, Lake Reserv. Manag., 20, 91–109, https://doi.org/10.1080/07438140409354354, 2004.

Ottosson, F. and Abrahamsson, O.: Presentation and analysis of a model simulating epilimnetic and hypolimnetic temperatures in lakes, Ecol. Modell., 110, 233–253, https://doi.org/10.1016/S0304-3800(98)00067-2, 1998.

Piccolroaz, S., Woolway, R. I., and Merchant, C. J.: Global reconstruction of twentieth century lake surface water temperature reveals different warming trends depending on the climatic zone, Clim. Change, 160, 427–442, https://doi.org/10.1007/s10584-020-02663-z, 2020.

Prats-Rodríguez, J. and Danis, P.-A.: inrae/ALAMODE-cuspy: cuspy v1.0, https://doi.org/10.5281/ZENODO.7585606, 2023a.

Prats-Rodríguez, J. and Danis, P.-A.: inrae/ALAMODE-okp: okplm v1.0, https://doi.org/10.5281/ZENODO.7564750, 2023b.

Prats, J. and Danis, P. A.: An epilimnion and hypolimnion temperature model based on air temperature and lake characteristics, Knowl. Manag. Aquat. Ecosyst., 8, https://doi.org/10.1051/kmae/2019001, 2019.

Prats, J., Reynaud, N., Rebière, D., Peroux, T., Tormos, T., and Danis, P. A.: LakeSST: Lake Skin Surface Temperature in French inland water bodies for 1999-2016 from Landsat archives, Earth Syst. Sci. Data, 10, 727–743, https://doi.org/10.5194/essd-10-727-2018, 2018.

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, Ecol. Modell., 291, 142–150, https://doi.org/10.1016/j.ecolmodel.2014.07.029, 2014.

Schaeffer, B. A., Iiames, J., Dwyer, J., Urquhart, E., Salls, W., Rover, J., and Seegers, B.: An initial validation of Landsat 5 and 7 derived surface water temperature for U.S. lakes, reservoirs, and estuaries, Int. J. Remote Sens., 39, 7789–7805, https://doi.org/10.1080/01431161.2018.1471545, 2018.

Sharaf, N., Fadel, A., Bresciani, M., Giardino, C., Lemaire, B. J., Slim, K., Faour, G., and Vinçon-Leite, B.: Lake surface temperature retrieval from Landsat-8 and retrospective analysis in Karaoun Reservoir, Lebanon, J. Appl. Remote Sens., 13, 1, https://doi.org/10.1117/1.jrs.13.044505, 2019.

Sharaf, N., Lemaire, B. J., Fadel, A., Slim, K., and Vinçon-Leite, B.: Assessing the thermal regime of poorly monitored reservoirs with a combined satellite and three-dimensional modeling approach, Inl. Waters, 11, 302–314, https://doi.org/10.1080/20442041.2021.1913937, 2021.

Sharaf, N., Prats, J., Reynaud, N., Tormos, T., Peroux, T., and Danis, P. A.: LakeTSim (Lake Temperature Simulations), Rech. Data Gouv, https://doi.org/doi:10.57745/OF9WXR, 2023.

Sharma, S., Walker, S. C., and Jackson, D. A.: Empirical modelling of lake water-temperature relationships: A comparison of approaches, Freshw. Biol., 53, 897–911, https://doi.org/10.1111/j.1365-2427.2008.01943.x, 2008.

Sharma, S., Gray, D. K., Read, J. S., O'Reilly, C. M., Schneider, P., Qudrat, A., Gries, C., Stefanoff, S., Hampton, S. E., Hook, S., Lenters, J. D., Livingstone, D. M., McIntyre, P. B., Adrian, R., Allan, M. G., Anneville, O., Arvola, L., Austin, J., Bailey, J., Baron, J. S., Brookes, J., Chen, Y., Daly, R., Dokulil, M., Dong, B., Ewing, K., De Eyto, E., Hamilton, D., Havens, K., Haydon, S., Hetzenauer, H., Heneberry, J., Hetherington, A. L., Higgins, S. N., Hixson, E., Izmest'eva, L. R., Jones, B. M., Kangur, K., Kasprzak, P., Köster, O., Kraemer, B. M., Kumagai, M., Kuusisto, E., Leshkevich, G., May, L., MacIntyre, S., Müller-Navarra, D., Naumenko, M., Noges, P., Noges, T., Niederhauser, P., North, R. P., Paterson, A. M., Plisnier, P. D., Rigosi, A., Rimmer, A., Rogora, M., Rudstam, L., Rusak, J. A., Salmaso, N., Samal, N. R., Schindler, D. E., Schladow, G., Schmidt, S. R., Schultz, T., Silow, E. A., Straile, D., Teubner, K., Verburg, P., Voutilainen, A., Watkinson, A.,

https://doi.org/10.5194/essd-2022-457
Preprint. Discussion started: 14 February 2023

Weyhenmeyer, G. A., Williamson, C. E., and Woo, K. H.: A global database of lake surface temperatures collected by in situ and satellite methods from 1985-2009, Sci. Data, 2, 1–19, https://doi.org/10.1038/sdata.2015.8, 2015.

Shatwell, T., Thiery, W., and Kirillin, G.: Future projections of temperature and mixing regime of European temperate lakes, Hydrol. Earth Syst. Sci., 23, 1533–1551, https://doi.org/10.5194/hess-23-1533-2019, 2019.

Snowling, S. D. and Kramer, J. R.: Evaluating modelling uncertainty for model selection, Ecol. Modell., 138, 17–30, https://doi.org/10.1016/S0304-3800(00)00390-2, 2001.

Toffolon, M., Piccolroaz, S., Majone, B., Soja, A. M., Peeters, F., Schmid, M., and Wüest, A.: Prediction of surface temperature in lakes with different morphology using air temperature, Limnol. Oceanogr., 59, 2185–2202, https://doi.org/10.4319/lo.2014.59.6.2185, 2014.

Vernay, M., Lafaysse, M., Mérindol, L., Giraud, G., and Morin, S.: Ensemble forecasting of snowpack conditions and avalanche hazard, Cold Reg. Sci. Technol., 120, 251–262, https://doi.org/10.1016/j.coldregions.2015.04.010, 2015.

Vernay, M., Lafaysse, M., Monteiro, D., Hagenmuller, P., Nheili, R., Samacoïts, R., Verfaillie, D., and Morin, S.: The S2M meteorological and snow cover reanalysis over the French mountainous areas: description and evaluation (1958-2021), Earth Syst. Sci. Data, 14, 1707–1733, https://doi.org/10.5194/essd-14-1707-2022, 2022.

White, J. T., Fienen, M. N., and Doherty, J. E.: A python framework for environmental model uncertainty analysis, Environ. Model. Softw., 85, 217–228, https://doi.org/10.1016/j.envsoft.2016.08.017, 2016.

White, J. T., Hunt, R. J., Fienen, M. N., Doherty, J. E., and Survey, U. S. G.: Approaches to highly parameterized inversion: PEST++ Version 5, a software suite for parameter estimation, uncertainty analysis, management optimization and sensitivity analysis:, U.S. Geological Survey Techniques and Methods 7C26, 52 pp., https://doi.org/10.3133/ tm7C26, 2020.

Woolway, R. I. and Merchant, C. J.: Worldwide alteration of lake mixing regimes in response to climate change, Nat. Geosci., 12, 271–276, https://doi.org/10.1038/s41561-019-0322-x, 2019.

Woolway, R. I., Sharma, S., and Smol, J. P.: Lakes in Hot Water : The Impacts of a Changing Climate on Aquatic Ecosystems, Bioscience, https://doi.org/10.1093/biosci/biac052, 2022.

Yang, K., Yu, Z., Luo, Y., Yang, Y., Zhao, L., and Zhou, X.: Spatial and temporal variations in the relationship between lake water surface temperatures and water quality - A case study of Dianchi Lake, Sci. Total Environ., 624, 859–871, https://doi.org/10.1016/j.scitotenv.2017.12.119, 2018.

Yang, K., Yu, Z., and Luo, Y.: Analysis on driving factors of lake surface water temperature for major lakes in Yunnan-Guizhou Plateau, Water Res., 184, 116018, https://doi.org/10.1016/j.watres.2020.116018, 2020.