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], Rosalie Bruel[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, DRAS, Service ECOAQUA, ~~DRAS,~~ 3275 Route Cézanne, 13100 Aix-en-Provence, France

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

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. We combined numerical modelling and satellite thermal data to create a regional dataset (LakeTSim: Lake Temperature Simulations) of long-term water temperatures for 401 French lakes in order to tackle the scarcity of in situ water temperature. The dataset consists of daily epilimnion and hypolimnion water temperatures for the period 1959-2020 simulated with the semi-empirical OKPLM (Ottosson-Kettle-Prats Lake Model) and the associated uncertainties. Here, we describe the model and its performance. Additionally, we present the uncertainty analysis of simulations with default parameter values (parametrized as a function of lake characteristics) and calibrated parameter values, along with the analysis of the sensitivity of the model to parameter values and biases in the input data. Overall, the 90% confidence uncertainty range is largest for hypolimnion temperature simulations with a median of 8.5 ℃ and 2.32 ℃ respectively with default and calibrated parameter values. There is less uncertainty associated with epilimnion temperature simulations with a median of 5.42 ℃ and 1.85 ℃, respectively before and after parameter calibration. This dataset provides over six decades of epilimnion and hypolimnion temperature data, crucial for climate change studies at a regional scale. It will help provide insight into the thermal functioning of French lakes and can be used to help decision-making and 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).

It is thus crucial to closely evaluate water temperature trajectories over the entire water column in space and time when assessing the impact of climate change on lake ecosystems. However, the lack of data coverage, both spatially and temporally, makes it difficult to accurately characterize lakes thermal response to climate change and to identify warming trends (Gray et al., 2018). Indeed, long-term datasets of in situ temperatures are usually scarce and mostly limited to large lakes (Layden et al., 2015). Moreover, sampling frequency and temporal coverage of

93   in situ water temperature varies greatly from one lake to the next, from a few years (Sharma et al., 2015) up to
94   decades (Piccolroaz et al., 2020; Rimet et al., 2020).

95   Due to the difficulties in setting up conventional (i.e., in situ) monitoring programs , often expensivetied to e.g.,
96   costs, governance and intercalibration, the coupling of modelling and satellite remote sensing data has become
97   fundamental in the field of limnology (Nouchi et al., 2019). Modelling provides means to interpolate both temporal
98   and spatial gaps. It thereby allows us to acquire information about surface water temperatures, which are globally
99   the focus of lake climate change studies, and deep-water temperatures which are as critical though often
100   disregarded in this context (however see Pilla et al., 2020). Several numerical models that vary in complexity exist
101   for conducting water temperature simulations, the most accurate being deterministic or process-based models.
102   Nevertheless, regional or global deterministic modelling efforts over long periods are usually hindered by the lack
103   of sufficiently detailed input data (e.g., meteorological and field data) to run the models (Kim et al., 2021). For
104   practical and operational purposes, simpler models (semi-empirical, statistical or hybrid physical-statistical based
105   models) with less requirements for forcing data, have been mostly applied to assess the impact of climate change
106   on lake ecosystems and study them (Piccolroaz et al., 2020; Toffolon et al., 2014; Sharma et al., 2008). Long-term
107   simulations across a considerable number of lakes are made possible with this type of models, enabling the
108   detection of trends in time series data that are not achievable with shorter datasets (Gray et al., 2018).

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

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

132 changes in the mixing regime among others. These indicators will contribute to assess the impact of climate change

133 on lakes thermal functioning and its influence on the biological community structure and trophic webs.

3. Data and methodology

3.1. The software suite ALAMODE

136 The simulations, sensitivity and uncertainty analysis presented in this paper were made using the software suite

137 ALAMODE (A LAke MODElling project). ALAMODE (Danis, 2020) is a software suite developed in pPython

138 3 by the Pôle R&D Ecosystèmes Lacustres (ECLA) and SEGULA Technologies to facilitate the realization of

139 simulations of lakes and the management of related information. It comprises multiple modules and packages

140 designed for lake and tributary modelling, as well as for processing the data necessary to operate these models.

141 These packages include OKPLM (Ottosson-Kettle-Prats Lake Model), CUSPY (Calibration, Uncertainty analysis

142 and Sensitivity analysis in PYthon), TMOD (Temperature MODelling), GLMtools (General Lake Model tools),

143 "tributary", TINDIC (Temperature INDICators) and ALAPROD (ALAMODE-Production). OKPLM (Prats-

144 Rodríguez and Danis, 2023b) is used to simulate epilimnion and hypolimnion water temperatures in lakes while

145 CUSPY (Prats-Rodríguez and Danis, 2023a) is used for model parameters estimations and conducting uncertainty

146 and sensitivity analyses. TMOD is used for managing the T-MOD database designated to facilitate the realization

147 and consultation of simulations. GLMtools is used to conduct lake hydrodynamic simulations using the one-

148 dimensional hydrodynamic General Lake Model (Hipsey et al., 2019) while "tributary" is used for the estimation

149 of flow and temperature of lake tributaries. The package TINDIC is used for calculating temperature indicators

150 from model simulations. Finally, ALAPROD integrates all the functionalities to produce simulations into a single

151 package: simulation of stream water temperature, of lake hydrodynamics and temperature, and of stream flow rate.

152 It also includes sensitivity and uncertainty analysis features. The functionalities of these packages can be accessed

153 either by using each package separately or by utilizing the ALAPROD package, which depends on the TMOD

154 database and requires access to it.

155 At present, only the ALAMODE packages related to the main functionalities used in this work are publicly

156 available (see Code availability section): the simulation of lake temperatures using the Ottosson-Kettle-Prats Lake

157 Model (Prats & Danis, 2019), implemented in the package OKPLM, and the sensitivity and uncertainty analysis

158 tools in the CUSPY package. We used ALAPROD to access the functionalities of both packages.

3.2. The OKP Lake Model description

160 The OKPLM (Ottosson-Kettle-Prats Lake Model) (Prats & Danis, 2019) is a two-layer semi-empirical data model

161 adapted from Kettle et al (2004) for the epilimnion module and Ottosson & Abrahamsson (1998) for the

162 hypolimnion module. The modifications proposed in Prats & Danis (2019) consisted mainly of simplifying the

163 mixing algorithm used in Ottosson & Abrahamsson (1998) using a basic stability condition, whereas for the

164 epilimnion module a sinusoidal fit to average daily solar radiation was used instead of the theoretical clear-sky

165 radiation. The OKPLM also runs on weekly and monthly frequencies. The regionalization of the parameters of the

166 model mainly depends on the geographical and morphological properties of the lake (maximal depth, volume,

167 surface area, latitude and altitude). The model requires few meteorological forcing data: solar radiation and air

168 temperature.

169 The model calculates water temperature as follows:

170 $$T_{e,i} = A + Bf(T_{a,i}^*) + CS_i \tag{1}$$

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

173 $$T_{a,i}^* = T_{a,i} - MAAT \tag{2}$$

174 Where $T_{a,i}$ is air temperature (°C) and $MAAT$ is the annual mean air temperature (°C). The smoothing function
175 $f(*)$ is such that it gives greater weight to the nearest observations and the weights decrease exponentially. It is
176 defined as:

177 $$f(T_{a,1}^*) = T_{a,1}^* \tag{3}$$

178 $$f(T_{a,i}^*) = \alpha T_{a,i}^* + (1 - \alpha)f(T_{a,i-1}^*) \tag{4}$$

179 where α is the smoothing factor. When $\alpha = 1$ there is no smoothing, while the smoothing increases with the
180 decrease in the value of α.

181 $$T_{h,i} = A \cdot D + E \cdot g(T_{e,i}) \tag{5}$$

182 where T_h is the hypolimnion temperature (°C), D and E are calibration parameters and $g(T_{e,i})$ is an exponential
183 smoothing as follows:

184 $$g(T_{e,1}) = T_{e,1} \tag{6}$$

185 $$g(T_{e,i}) = \beta T_{e,i} + (1 - \beta)g(T_{e,i-1}) \tag{7}$$

186 where β is the exponential smoothing factor. As for α, there is no smoothing for $\beta = 1$ and the smoothing increases
187 as β approaches zero.

188 In ALAPROD, OKPLM can be run in two modes: the "default" mode where model parameter values for each
189 lake are estimated using the parameterization presented in Prats & Danis (2019), and the "calibrated" mode where
190 model parameters are calibrated individually for each lake by using in situ temperature measurements. The default
191 parameterization was obtained by using the individually calibrated parameter values to fit appropriate expressions
192 as a function of the characteristics of lakes. In the epilimnion module model parameter values A, B, C, and α are
193 estimated based on lake characteristics (i.e., latitude, altitude, surface area, volume, and depth). These equations
194 were determined using robust regressions and Landsat infrared data (median skin temperatures) from 1999 to 2016
195 of French lakes to estimate mean surface temperatures (Prats et al., 2018). In contrast, for the hypolimnion module,
196 parameter values E and β were derived as a function of lake depth and lake type using temperature profile data
197 from 357 lakes; β can have a value of 1 ($E > 0.95$) or 0.13 ($E \leq 0.95$). The parameter D was assigned a constant
198 value of 0.51.

199 The parametrization of the OKPLM parameters as presented in Prats & Danis (2019) is as follows:

200 $$A = 39.9 - 0.484L_{Lat} - 4.52 \times 10^{-3}L_{Alt} - 0.167lnL_A \tag{8}$$

201 where L_{Lat} is lake latitude (°N), L_{Alt} is lake altitude (m) and L_A is lake surface area (m^2).

$$B = 1.058 - 0.0010L_{Dmax} \tag{9}$$

where L_{Dmax} is lake maximal depth (m).

$$C = 1.12 \times 10^{-3} - 3.62 \times 10^{-6}L_{Alt} \tag{10}$$

$$E = e_1 + \frac{1-e_1}{1+\exp[e_3(e_2-lnL_D)]} \tag{11}$$

where e_1, e_2 and e_3 are coefficients with respective values of 0.10, 2.0, -1.8 for natural lakes and 0.49, 1.7, -2.0 for artificial lakes (reservoirs, gravel pits, ponds and quarry lakes) and L_D is lake mean depth (m).

$$\alpha = \exp(0.52 - 3.0 \times 10^{-4}L_{Alt} + 0.25lnL_A - 0.36lnL_V) \tag{12}$$

where L_V is lake volume (m^3).

3.3. 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 referred to as "massifs" and covering the French Alps, Pyrenees and Corsica mountainous regions. In order to use S2M meteorological data over each lake an extraction of certain topographic classes is necessary. These include elevation, aspect and slope, which represent the spatial variability over "massifs". On average, a massif corresponds to a mountainous region of about 1000 km^2 over which meteorological conditions are considered homogeneous at a given elevation range. Two types of S2M reanalysis simulations exist for each elevation range, one at flat terrain and the other with 8 aspects at 2 different slope angles. For this study, this information (elevation, slope, aspect) was extracted from a Digital Elevation Model (BD Alti, IGN) for each lake over its drainage basin, combined into zones corresponding to S2M topographic classes. We considered a zero slope and average daily data for each study site.

In situ temperature profiles, geographical and morphological data of the study sites were initially extracted from the PLAN_DEAU database. The extracted data was then incorporated into the T-MOD database, with the aim of simplifying the process of simulations and accessing information about the characteristics of the simulated lakes. Both databases are managed by INRAE (l'Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement) and Pôle R&D ECLA ("ECosystèmes Lacustres") in Aix-en-Provence, France. The geomorphological data consisting of maximal depth, volume, surface area, latitude and altitude were extracted for

237   401 lakes. In situ temperature profiles were extracted for 170 lakes over different depths. Depending on each lake,

238   the number of years with samples could vary from 1 to 12 with a number of samples ranging between 1 and 10 per

239   year.

3.4. Lake simulations

241   For this study, we considered 401 lakes (Figure 1) located in Metropolitan France monitored according to the

242   Water Framework Directive (WFD). Here we refer to lakes as natural lakes, reservoirs, gravel pits and other

243   artificial lakes (e.g., ponds and quarry lakes). The present lake dataset includes epi- and hypolimnion temperature

244   simulations for 54 natural lakes, 302 reservoirs, 7 gravel pits and 38 other artificial lakes (Figure 2). The lakes

245   characteristics range between 0 and 2279.7 m for altitude, 0.8 and 309.7 m for maximal depth, 0.08 and 577.12 km^2

246   for surface area and 5×10^4 and 8.9×10^{10} m^3 for volume.

247   The OKPLM was run for each lake using either "default" ~~and~~ or "calibrated" parameters, and either ~~with two~~

248   ~~sources of meteorological data,~~ SAFRAN ~~and~~ or S2M ~~over specific sets of lakes~~meteorological data. Specifically,

249   "~~Calibrated~~calibrated" model parameters were adopted when in situ temperature ~~measurements~~ profiles along the

250   water columns were available from the RCS/RCO (Réseau de Contrôle de Surveillance–/–Réseau de Contrôle

251   Opérationel, French networks for WFD) monitoring; these temperature profiles were then transformed and used

252   as epilimnion and hypolimnion temperatures. F~~f~~or those lakes, calibration parameters (A, B, C, D, E, α and β) are

253   lake-specific and determined using the lake-specific temperature profiles. ~~e~~Conversely, "default" parameters were

254   used for the rest of the lakes as well as when bathymetry data necessary for the transformation of temperature

255   profiles into epilimnion and hypolimnion temperatures were not available. ~~with insufficient in situ data;~~ In this

256   case, the values of the parameters were estimated according to equations (8) to (12).

257   SAFRAN data were used for most of the lakes except for a few lakes at higher altitudes. Indeed, S2M data are

258   more representative of mountainous meteorological conditions than SAFRAN data and were thus used, when

259   possible, for simulating the water temperature in lakes situated at altitudes higher than 900 m. For some lakes, it

260   was not possible to use S2M data, either because their drainage basins are not entirely part of a massif ($n = 1$), or

261   because they are located in massifs that are not covered by the S2M reanalysis dataset ($n = 18$). Among the total

262   number of study sites ($n = 401$), the model was forced using SAFRAN and S2M meteorological data respectively

263   for 210 and 21 lakes with "default" model parameters, and for 164 and 6 lakes with "calibrated" model parameters.

264   The geomorphological characteristics of the simulated lakes with each of the abovementioned configurations are
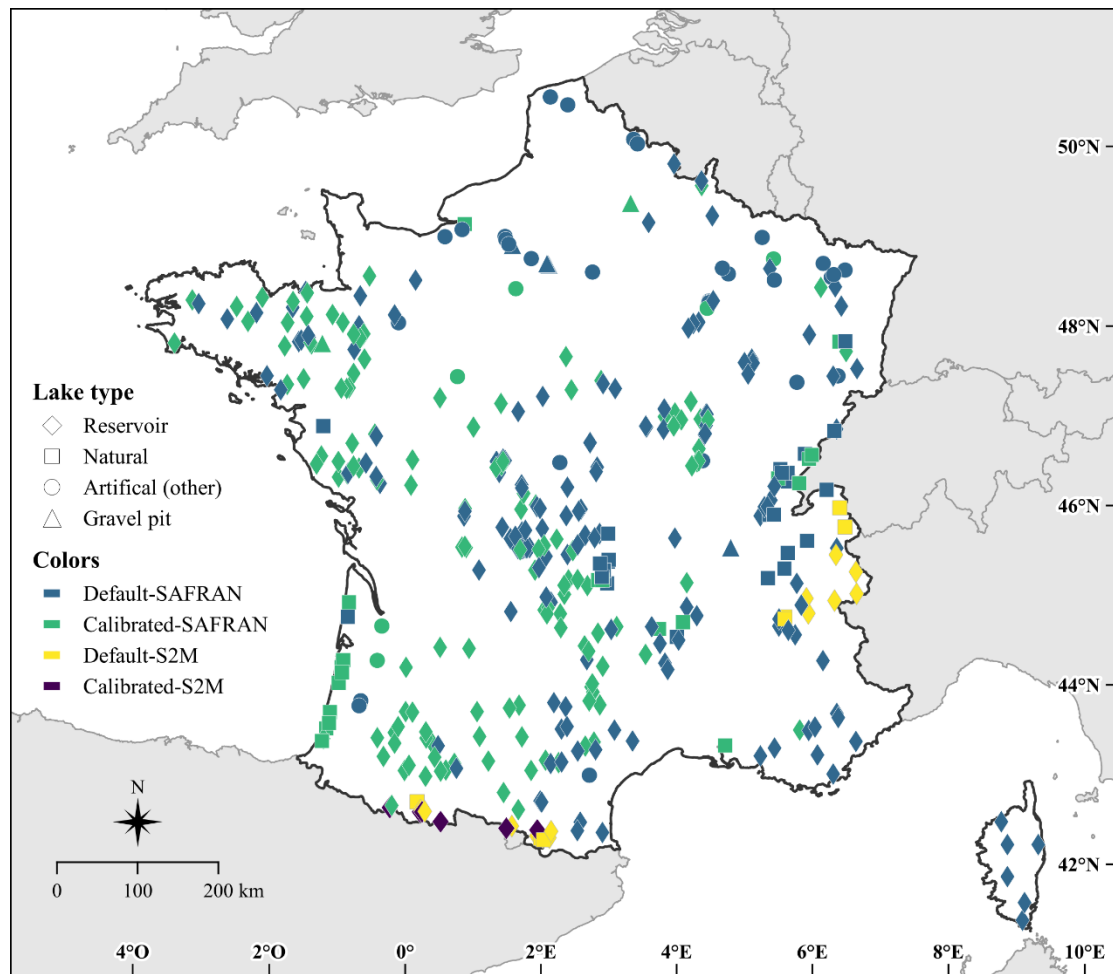
265   shown in Table 1.

266

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. The "other" artificial lakes consist of ponds and quarry lakes.
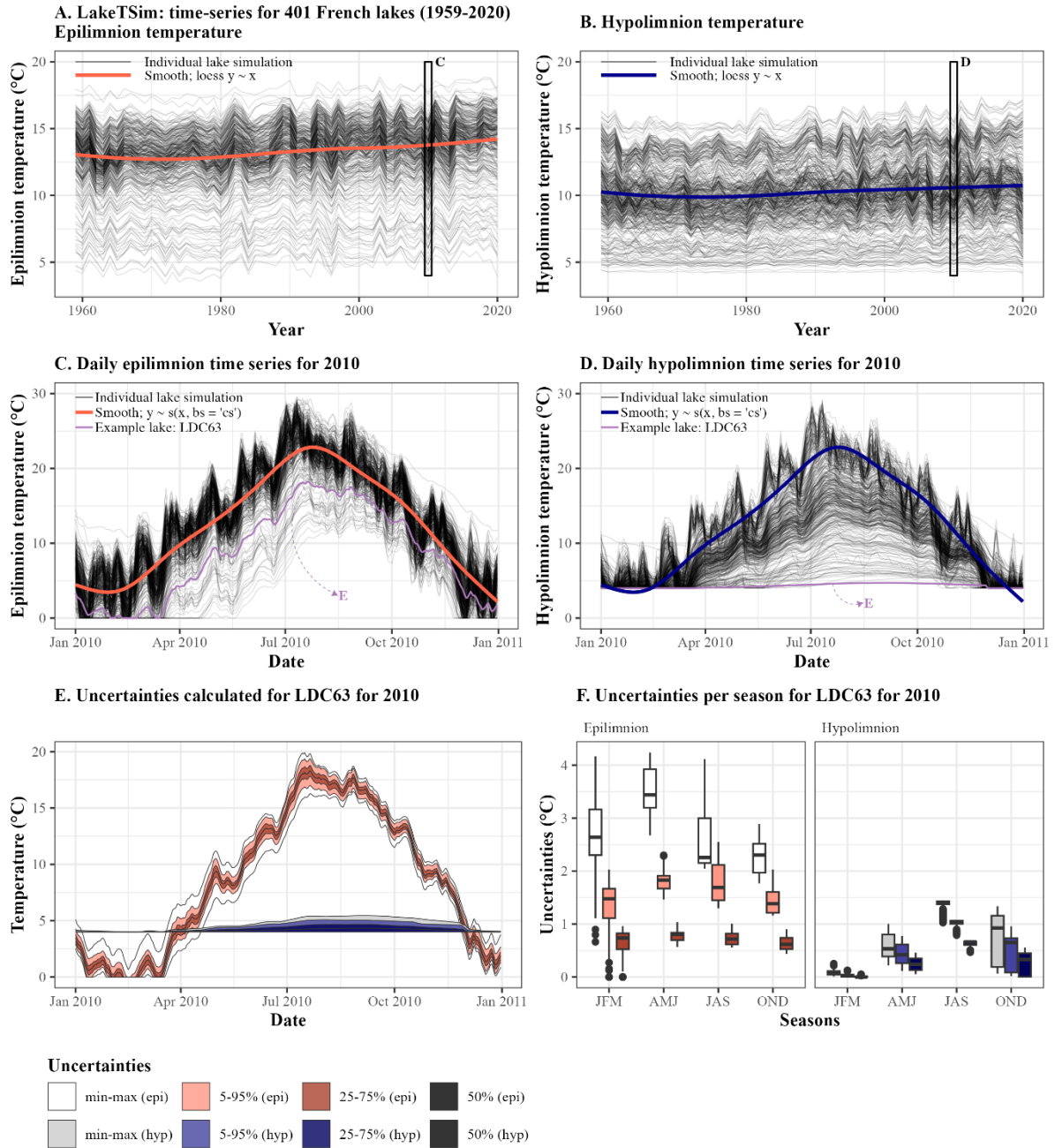
Figure 2: Presentation of the LakeTSim data. (A) Epilimnion and (B) hypolimnion mean annual temperatures, with average trend across lakes shown with a smooth spline. (C) Daily epilimnion temperature per lake in the dataset for 2010, with smooth spline and the time series for one lake (LDC63) highlighted. (D) Daily hypolimnion temperature per lake in the dataset for 2010, with smooth spline and the time series for one lake (LDC63) highlighted. LDC63 is the code for Lake Chauvet, a natural lake (45.46 °N, 2.83 °E) located at 1167 m asl, with a surface area of 0.51 km^2, a volume of 17.41 10^6 m^3, and a maximum depth of 66.8 m. The simulation for LDC63 was conducted resorting to SAFRAN data and was run with the "calibrated" mode. (E) Uncertainties were calculated per lake and per day and are shown here daily for LDC63, in 2010, for both the epilimnion (epi) and the hypolimnion (hyp). (F) Uncertainties are shown here seasonally for LDC63, in 2010, for both the epilimnion (epi) and the hypolimnion (hyp). JFM corresponds to January-February-March, AMJ corresponds to April-May-June, JAS corresponds to July-August-September and OND corresponds to October-November-December.

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 (km²)	0.08 - 577.12	0.11 - 6.52	0.29 - 57.57	0.45 - 1.21
Volume (m³)	$5\times10^4 - 8.9\times10^{10}$	$51.4\times10^4 - 33.32\times10^7$	$12.9\times10^4 - 49.88\times10^7$	$72.7\times10^5 - 68.6\times10^6$

268

3.5. Calibration, uncertainty and sensitivity analysis

270 Calibration, uncertainty and sensitivity analyses of model parameters were carried out using the package "CUSPY"
271 (Calibration, Uncertainty analysis and Sensitivity analysis in PYthon), which is part of the software suite
272 "ALAMODE" (Danis, 2020) and acts as an interface to the package "pyemu" (White et al., 2016, 2020). In addition
273 to model parameters, sensitivity analysis was extended to encompass forcing parameters (*MAAT, at_factor,*
274 *sw_factor*) as they provide information about the degree of sensitivity exhibited by model parameters in response
275 to biases in the forcing data.

276 Parameter values were calibrated for lakes with ~~enough~~ available in situ data (temperature profiles and
277 bathymetry). Parameter values were calibrated using the Gauss-Levenberg-Marquardt algorithm and Tikhonov
278 regularization (White et al., 2020), and the squared sum of residuals as objective function. In addition to the
279 calibrated parameter values, the calibration process also provided posterior parameter uncertainty and composite
280 scaled sensitivities. Composite scaled sensitivities (*CSS*) indicate the quantity of information provided by each
281 parameter and the sensitivity of the model to them (Ely, 2006). The parameters with higher *CSS* values will have
282 a greater impact on the resulting simulation compared to those with low *CSS* values. To determine the *CSS* values
283 for each parameter, the Dimensionless Scaled Sensitivities (*DSS*) are used. *DSS* indicate how important an
284 observation or how sensitive a simulated equivalent of an observation is in relation to the estimation of a parameter.
285 Further information on these statistical measures is available in Hill (1998) and Poeter & Hill (1997). The
286 dimensionless scaled sensitivity for *i* and *j*, *i* being one of the observations and *j* being one of the parameters, is
287 calculated as:

288 $$DSS_{i,j} = \left[\frac{\partial y'_i}{\partial b_j}\right] b_j w_i^{1/2} \tag{13}$$

10

289 where y_i' is the simulated value associated with the *i*th observation, b_j is the *j*th estimated parameter, $\frac{\partial y_i'}{\partial b_j}$ is the

290 sensitivity of the simulated value associated with the *i*th observation and w_i is the weight of the *i*th observation

291 calculated based on the inverse of the variance-covariance matrix of the observation errors.

292 The *CSS* for parameter *j* is calculated from *DSS* as follows:

293 $$CSS_j = \left[\frac{\sum_{i=1}^{ND} (DSS_{ij})^2 |_{\boldsymbol{b}}}{ND} \right]^{1/2} \qquad (14)$$

294 where ND is the number of observations and $\boldsymbol{b}$ is a vector of parameters values.

295 The uncertainty of the simulations (calibrated and default) was analyzed using Monte Carlo simulations. For each
296 lake, 100 Monte Carlo simulations were carried by randomly selecting the value of the model parameters. Two
297 parameters, at_factor and sw_factor, multiplying the meteorological input, were added to account for possible
298 uncertainties in input data. For default simulations, the a priori distribution of the parameters was assumed to
299 follow a normal distribution with the average value and lower and upper bounds shown in Table 2. The ranges for
300 parameters A, B and C were estimated as four times the standard deviation of the residuals of the formulas used to
301 estimate them according to Prats & Danis (2019). The parameters D, E and β, are expected to lie in the range 0-1
302 for mathematical and physical reasons. However, their respective values are highly interdependent and are difficult
303 to identify. Given their higher uncertainty, the full 0-1 range was explored. For $MAAT$, at_factor and sw_factor,
304 reasonable ranges (±10%) were chosen to account for meteorological data uncertainty (measurement error, errors
305 in regionalization, etc.). For calibrated simulations, the distribution of the parameters was obtained from the
306 calibration results.

307 In this study, the non-parametric Kendall's tau coefficient (significance level at 5%) was used to identify statistical
308 associations between uncertainty values and *CSS* in respect to lake geomorphological characteristics (maximal
309 depth, volume, surface area, latitude and altitude).

310

311

312

313

314

315

316

317

318

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

Parameter	Average value	Lower bound	Upper bound
A	$\hat{A}$	$\hat{A} - 2 \cdot 0.74$	$\hat{A} + 2 \cdot 0.74$
B	$\hat{B}$	$\hat{B} - 2 \cdot 0.08$	$\hat{B} + 2 \cdot 0.08$
C	$\hat{C}$	$\hat{C} - 2 \cdot 0.004$	$\hat{C} + 2 \cdot 0.004$
D	$\hat{D}$	0	1
E	$\hat{E}$	0	1
α	$\hat{\alpha}$	0	$\hat{\alpha} + 2 \cdot 0.08$
β	$\hat{\beta}$	0	1
$MAAT$	$M\hat{A}AT$	$M\hat{A}AT - 2 \cdot 0.5$	$M\hat{A}AT + 2 \cdot 0.5$
at_factor	1	0.9	1.1
sw_factor	1	0.9	1.1

319

4. Model performance

320

321 The performance of the OKPLM was assessed in Prats & Danis (2019) by comparing its performance to two other
322 often-applied models in lake studies, air2water (the 4-parameter version) and FLake. The air2water model is a
323 semi-empirical model used to calculate the epilimnion temperature of temperate lakes (Toffolon et al., 2014;
324 Piccolroaz et al., 2013). FLake is a one-dimensional (1D) hydrodynamic lake model for simulating temperature
325 vertical profiles and mixing conditions in lakes (Mironov, 2008). To assess their performances, the three models
326 were run between 1999 and 2016 over two sets of French lakes of different types (reservoirs, natural lakes, ponds,
327 quarry lakes and gravel pits): a group of five lakes with continuous profile measurements, and a group of 404 lakes
328 with less frequent temperature measurements. The performance assessment was limited to the period of 1999-2016
329 due to the availability of water temperature data (in situ and satellite) during that specific timeframe. The scarcity
330 of in situ water temperature measurements before 1999 applies to the entire set of lakes. However, it is important
331 to note that long-term in situ water temperature data is available for a few large lakes, which was used to assess
332 the performance of the three models (Prats & Danis, 2015). The OKPLM was run with the "default" parameter
333 values given by the parameterization in Prats & Danis (2019). The air2water parameter values were obtained as a
334 function of lake depth from the parametrization presented in Toffolon et al. (2014), based on data from 14 lakes
335 around the globe. In this case, the air2water model parameters were not calibrated due to the fact that the percentage
336 of missing data within the LakeSST dataset employed in Prats & Danis (2019), exceeded 97% for most lakes.
337 Beyond this threshold of 97% missing data, the performance of the calibrated 4-parameter version of the air2water
338 model was found to be unsatisfactory (Piccolroaz, 2016). However, When when evaluating the model performance

with the set of five lakes with continuous data, air2water was ~~also~~ run using parameter values calibrated for the individual lakes available data. FLake does not have calibration parameters. Meteorological- forcing (SAFRAN) consisted of air temperature for the air2water model; solar radiation, vapor pressure, cloud cover and wind speed for FLake; and air temperature and solar radiation for the OKPLM.

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. Although when using calibrated parameter values for air2water, median RMSE was below 1 ºC in most cases. 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

Overall, for both simulations with default and calibrated model parameters, uncertainty was higher for hypolimnion temperature compared to epilimnion temperature especially in reservoirs (Figure 3). In default simulations, the uncertainty of simulated epilimnion temperatures showed a clear and strong relation with lake maximal depth (Figure 3, Table 3). On one hand, maximal depth had the highest Kendall's tau value of 0.64 (p-value < 0.0001), indicating a strong positive correlation with uncertainty followed by volume with a Kendall's tau of 0.59 (p-value < 0.0001). Uncertainty increased with maximal depth and volume in particular for lakes with depths greater than 10 m and volumes greater than 10^6 m^3 (Figure 3). Overall, lakes with higher maximal depths have higher volumes and are located at greater altitudes (Figures A1-A2 in Appendix A). On the other hand, moderate significant correlations were identified with surface area, altitude and latitude (Table 3). Lakes with larger surface areas and higher altitudes tend to have higher uncertainties whereas lakes located at higher latitudes tend to have lower uncertainties (Figure A3 in Appendix A). The latter can be linked to the fact that more shallow lakes are located at higher latitudes (Figure A1 in Appendix A). For default simulations of hypolimnion temperatures, uncertainty was maximal for lakes with depths around 10 m. Kendall's tau values revealed a moderate significant correlation between hypolimnion temperature uncertainty and altitude (-0.45, p-value < 0.0001). The decrease in uncertainties with altitude can be related to the fact that lakes situated at very high altitudes are mostly deep. Further, in the present dataset, lakes with higher maximal depths occur as altitude increases (Figure A1-A2 in Appendix A).

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 °C) than the rest of lakes after calibration. Additionally, reservoirs (and a few natural lakes) above 100 m in altitude showed the highest uncertainties in the simulation of epilimnion temperature.
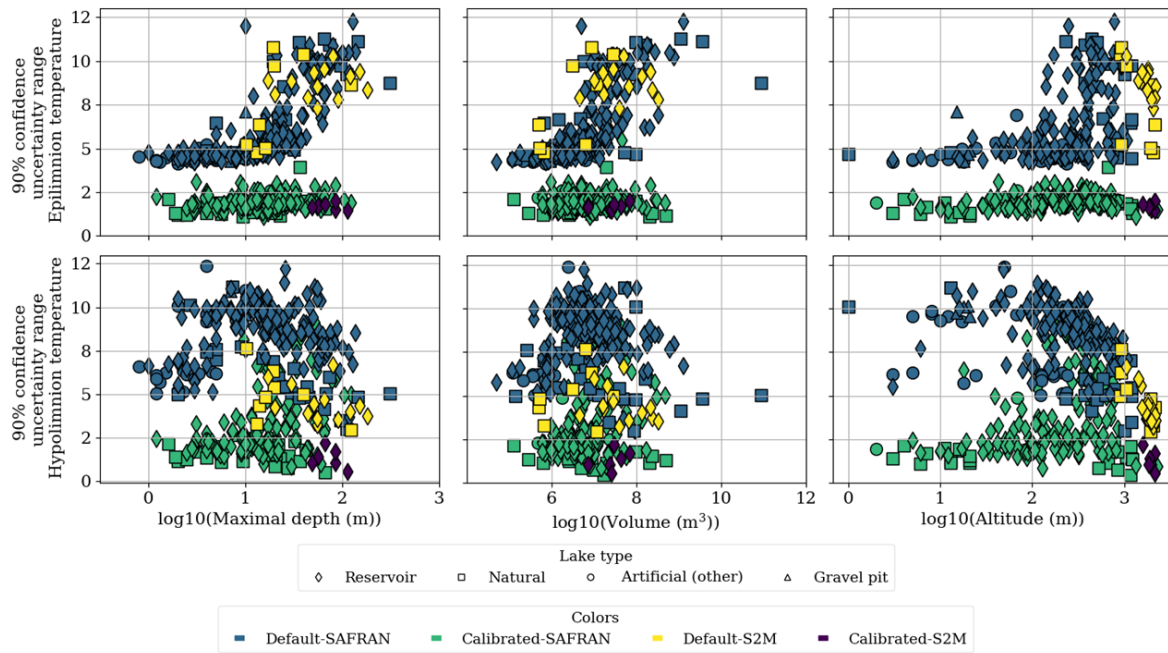

Figure 3: Average 90% confidence uncertainty range for epilimnion (top panel) and hypolimnion (bottom panel) temperatures in calibrated (*n* = 170) and default (*n* = 231) simulations for the period 1959-2020. **The "other" artificial lakes consist of ponds and quarry lakes.**

Table 3: Kendall's tau coefficients and *p*-values of average 90% confidence uncertainty range for epilimnion and hypolimnion temperatures obtained from default simulations (1959-2020) in respect to lakes geomorphological characteristics. For each lake, "Epilimnion uncertainty" and "Hypolimnion uncertainty" are defined as the average 90% confidence uncertainty range calculated as the difference between the 95th and 5th percentiles of the daily simulated epilimnion and hypolimnion water temperatures. The significance levels are represented as follows: *: $1.00e-02 < p$-value $\leq 5.00e-02$, **: $1.00e-03 < p$-value $\leq 1.00e-02$, ***: $1.00e-04 < p$-value $\leq 1.00e-03$, ****: p-value $\leq 1.00e-04$. Otherwise, correlations are not significant (*p*-value > 0.05).

	Maximal depth (m)	Surface area (km²)	Altitude (m)	Latitude (°N)	Volume (m³)
Epilimnion uncertainty	0.64****	0.31****	0.39****	-0.40****	0.59****
Hypolimnion uncertainty	-0.13**	0.05	-0.45****	0.03	-0.03

6. Sensitivity analysis

The parameter to which the model was most sensitive was the parameter *C* (Figure 4), 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*, *MAAT* and *A*. The parameter *D*, with *CSS* several orders of magnitude smaller than the other parameters, was unidentifiable.
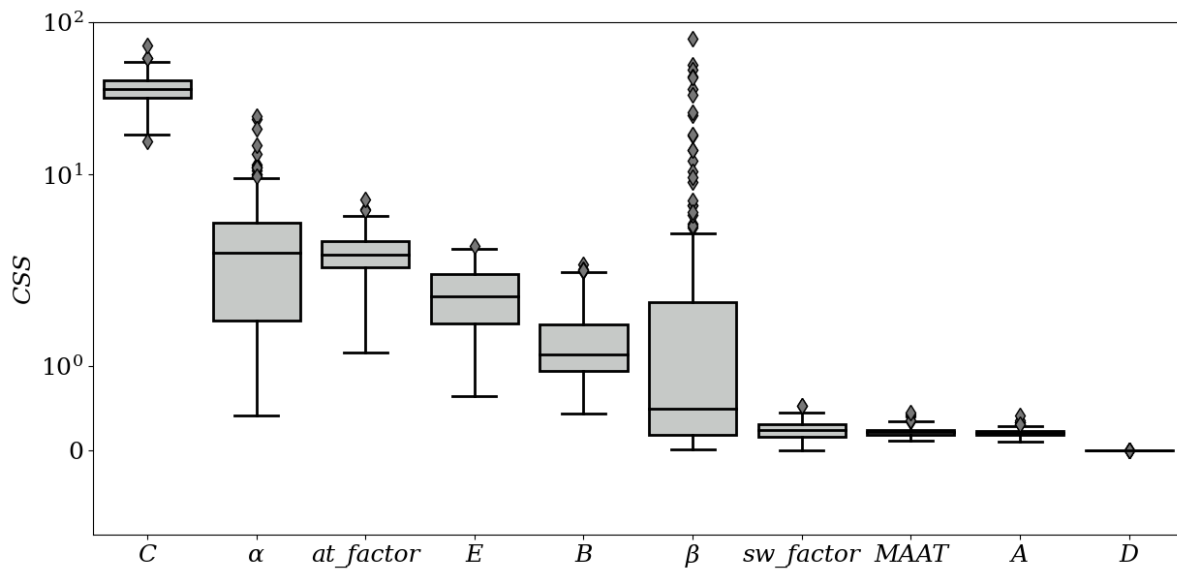
Figure 4: Composite scaled sensitivities (*CSS*) for each parameter. The boxplots indicate the distribution of *CSS* between the simulations calibrated for different lakes. The y-axis is in logarithmic form.

The model tended to be more sensitive to the parameter values in the case of reservoirs than in the case of natural lakes (Figure 5, Figures A4-A7 in Appendix A). Some parameters showed a dependency on lakes geomorphological characteristics. With the exception of a weak correlation with altitude (Kendall's tau = 0.18), there was no significant dependence between the parameter *C* and lakes geomorphological characteristics (Table 4, Figure A4 in Appendix A). The parameter α being parametrized as a function of lake volume, surface area and altitude reflects the thermal inertia of the lake. It showed a clear highly significant dependency primarily on lake depth (Kendall's tau = 0.47) followed by altitude (Kendall's tau = 0.4) and volume (Kendall's tau = 0.39) (Figure 5, Table 4). The increase of model sensitivity to the parameter α primarily with depth as well as altitude and volume propagated to the default simulations and explain the increased uncertainty with these same geomorphological characteristics in the default simulations. The parameter *at_factor*, was weakly but significantly correlated with all lakes geomorphological characteristics except for latitude with which no correlation was found (Figure 5, Table 4, Figures A4-A7 in Appendix A). *CSS* were mostly low for the parameter *β*, except for a few reservoirs and artificial lakes that scored very high *CSS* values. The sensitivity of *β* displayed a weak but significant correlation with lakes geomorphological characteristics, except for volume (Table 4).

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.

15

Table 4: Kendall's tau coefficients and *p*-values of *CSS* for model parameters values and drivers obtained from calibrated simulations (1959-2020) in respect to lakes geomorphological characteristics. The significance levels are represented as follows: *: 1.00e-02 < *p*-value ≤ 5.00e-02, **: 1.00e-03 < *p*-value ≤ 1.00e-02, ***: 1.00e-04 < *p*-value ≤ 1.00e-03, ****: *p*-value ≤ 1.00e-04. Otherwise, correlations are not significant (*p*-value > 0.05).

	Maximal depth (m)	Surface area (km²)	Altitude (m)	Latitude (°N)	Volume (m³)
CSS_A	0.02	-0.1	0.14**	-0.08	-0.07
CSS_B	0.09	-0.04	0.14**	-0.14**	0.02
CSS_C	-0.04	-0.09	0.18***	-0.05	-0.1
CSS_D	-0.12*	0.02	-0.14**	0.06	-0.1
CSS_E	-0.01	-0.001	0.02	0.0003	-0.03
CSS_α	0.47****	0.07	0.4****	-0.23****	0.39****
CSS_β	0.16**	-0.12*	0.22****	-0.19***	0.05
CSS_{at_factor}	-0.25****	-0.14**	-0.13*	0.04	-0.28****
CSS_{sw_factor}	-0.22****	-0.06	-0.14**	0.06	-0.2****
CSS_{MAAT}	-0.09	-0.13**	0.13*	-0.02	-0.15**

412
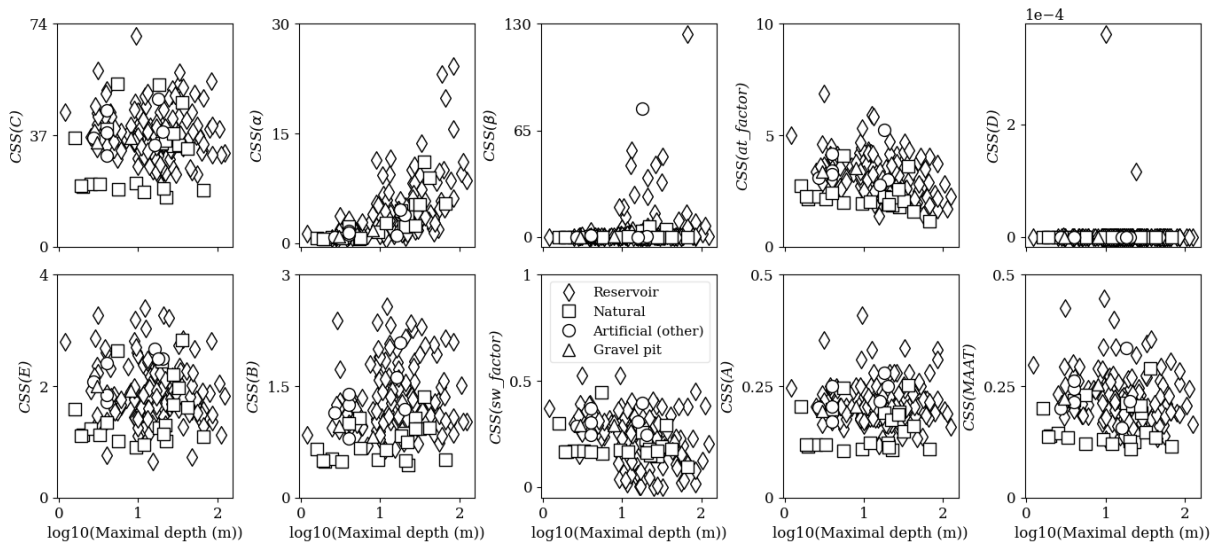
Figure 5 : Composite scaled sensitivities (*CSS*) for each model parameter as a function of maximal depth. The "other" artificial lakes consist of ponds and quarry lakes.

414 **7. Discussion and implications**

415 Lakes are undeniably changing under climate change and long-term future projections show that the shifts in
416 ecosystem functioning will continue with aggravated alterations (Woolway & Merchant, 2019). In particular, given
417 the key role of lake water temperature in regulating ecosystem processes, its warming has become a response that
418 is crucial to monitor, explore and understand. Hence, the importance of developing or adopting approaches, such
419 as numerical models, that will provide long-term information about water temperature and allow us to understand
420 the thermal response of lakes to climate change.

421 Here we used a semi-empirical model, the OKPLM, to simulate six decades of epilimnion and hypolimnion water

422 temperatures in French lakes. In comparison to similar models, overall, the OKPLM provides acceptable

423 estimations of water temperatures, with better results for epilimnion temperatures. The values of the RMSEs

424 provided in Prats & Danis (2019) and obtained between OKPLM simulations and observations are comparable to

425 values found in studies applying complex hydrodynamic lake models (Read et al., 2014; Fang et al., 2012). When

426 using the default parameter values, the uncertainty associated with epilimnion temperature simulations was

427 significantly related to all geomorphological characteristics however, it was especially strongly correlated to lake

428 maximal depth. In contrast, the uncertainty in the hypolimnion simulations had a significant correlation solely with

429 altitude and maximal depth. The importance of this correlation was especially noteworthy in the case of reservoirs

430 located in low-altitude regions where uncertainties were the lowest. While the association between hypolimnion

431 uncertainty and maximal depth exhibited only a weak correlation, the instances of highest uncertainties were

432 predominantly found in reservoirs having maximal depths around 10 m. The correlations found between lakes

433 geomorphological characteristics and simulations uncertainties suggests that there might be systematic biases in

434 the definition of model parameters or in the forcing data. The calibration of model parameters significantly reduced

435 the uncertainties yet, for hypolimnion temperatures, they remained considerably high and increased with depth

436 especially in reservoirs.

437 The high levels of uncertainty found in reservoirs could be somewhat attributed to the lack of consideration of

438 water level fluctuations in the model. In contrast to other lakes (e.g., natural lakes, artificial lakes and gravel pits)

439 reservoirs experience significant variations in their water level, which influences the heat budget and hence their

440 thermal regime. Therefore, even under similar meteorological conditions lakes and reservoirs could have different

441 thermal behaviors (Nowlin et al., 2004). In reservoirs, the discharge depth is a driver of thermal structure. Deep

442 discharges could contribute to warmer bottom waters (Carr et al., 2020) whereas in some cases if the reservoir is

443 shallow or if the discharge depth is not deep, it could demonstrate lake-like thermal behavior. This does not

444 necessarily mean that, in this case, the entire functioning of the reservoir resembles one of a natural lake; there are

445 still differences to consider (Detmer et al., 2022).

446 The application of the OKPLM should be made with caution given its performance and depending on the objective

447 of the study. The model does not take into account a complete set of meteorological forcing (e.g., with cloud cover,

448 relative humidity and wind speed and direction) or other variables (e.g., inflow and outflow rates or water level

449 fluctuations, inflow discharge depth and inflow temperature) that could influence the thermal structure of the

450 ecosystem (Yang et al., 2020; Carr et al., 2020). Furthermore, the OKPLM was parametrized for a specific set of

451 lakes with particular geomorphological characteristics. Thus, it would be advisable to apply the model over lakes

452 with similar characteristics. If the aim is to conduct a long-term regional or global study for studying general

453 patterns of climate change impacts over a large number of study sites, the utilization of semi-empirical models

454 such as the OKPLM is the most suitable choice. Although complex, deterministic or process-based models provide

455 a more accurate representation of thermal conditions, applying these models over several study sites and for long

456 periods is usually hindered by the scarcity of the required input data. The increased complexity of these models

457 (with reference to an increased number of model parameters) is beneficial for representing additional ecosystem

458 processes. Yet the greater number of model parameters, increases the sensitivity of models and demands more

459 calibration efforts (Lindenschmidt, 2006). Furthermore, a reduction in model errors is sometimes associated with

460   an increased complexity in model structure; however, this is not always consistent since a complex model does
461   not necessarily provide better estimations and thus lower errors than a simple model (Snowling and Kramer, 2001).

462   Our goal in publishing the present dataset is to expand knowledge about the water temperature of French lakes and
463   provide data, with enough details and reliability, that it could be implemented in different studies where water
464   temperature is implicated for understanding specific processes or interactions, in particular under climate change.
465   Hence the significance of the present findings. The present study, making use of a semi-empirical model to provide
466   long-term data about water temperature, was necessary for several reasons. Equipping a large number of lakes
467   with thermal sensors is challenging and labor-intensive, it comes with a high financial cost that is often not
468   available. Consequently, historical and even current water temperature datasets are often scarce, which can be
469   problematic for studying the impact of climate change, as it requires high frequency data over a long duration of
470   time for accurate analysis. In general, the higher the sampling frequency and duration, the better the data is suited
471   to estimate or analyze specific processes or warming trends. The sampling frequency and length of a dataset have
472   been shown to play a role in determining the accuracy of estimating warming trends where time series longer than
473   30 years seem to be the most appropriate (Gray et al., 2018). Although, the duration and frequency of a dataset
474   have a major role in reflecting accurate representations, their influence is scarcely addressed when it comes to
475   climate change studies related to warming trends in water temperature.

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

483   **8. Data usage**
484   The LakeTSim dataset comprises water temperature simulations for natural lakes ($n = 54$), reservoirs ($n = 302$),
485   gravel pits ($n = 7$), and other artificial lakes (e.g., ponds and quarry lakes, $n = 38$). The simulations are for both the
486   epi- and hypolimnion. Lakes that are fully mixed throughout the year (typically, shallower lakes) have the same
487   temperature value for both layers. More generally, the delta of temperature can be used to calculate mixing regimes
488   (Sharaf et al., in prep.).

489   The lakes in the dataset were selected because they are monitored as part of the European Water Framework
490   Directive (Directive 2000/60/EC). The majority of the 401 lakes are non-natural and some were only created after
491   1959 (i.e., the start of our simulations). We compiled the initial temporal gap filling related to the initial filling
492   years for 282 of these 347 non-natural lakes (269 reservoirs and 13 artificial lakes, Figure A8 in Appendix A) in
493   Table S1 (see Supplement) to be used as a companion dataset to LakeTSim. The filling years were sourced from
494   https://www.barrages-cfbr.eu for 179 of the lakes and from the PLAN_DEAU database for 103 of the lakes; the
495   information was not available for 33 reservoirs, 7 gravel pits and 25 other artificial lakes of the LakeTSim dataset.

496   The median filling date was 1962 and 67% of the lakes with known filling dates were filled by 1980. While the
497   complete simulations ranging from 1959 to 2020 can also be used as theoretical lake temperature for comparison

across similar periods, we recommend that users of LakeTSim data for reservoir and artificial lake simulations consider the initial filling dates provided in Table S1 to filter out years from the simulations during which lakes were not filled yet.

Additionally, users should be aware that some reservoirs might be drained completely at certain intervals (e.g., every 10 years) for maintenance and inspection purposes, and that this is not reflected in our dataset. Finally, as mentioned in the discussion, some of the lakes in the dataset experience artificial (e.g., in reservoirs) or natural (e.g., in some smaller ponds) water level fluctuations, and potential intermittent dry-periods lasting weeks or months; none of these hydrological processes are accounted for in the simulations.

9. 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.

10. 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 maximal 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 are located in two main folders: "02_Temperature_data" containing daily epilimnion (tepi) and hypolimnion (thyp) temperatures simulated with the OKPLM and "03_Uncertainty_data" containing daily tepi and thyp uncertainties. In each folder, the data for temperature simulations and their uncertainties are presented in text files available in the folders "00_LakeTSim_SAFRAN_OKPdefault_data", "01_LakeTSim_SAFRAN_OKPcalibrated_data", "02_LakeTSim_S2M_OKPdefault_data" and "03_LakeTSim_S2M_OKPcalibrated_data". The name of each file within these folders includes the identification code of the lake. From 2024, the data will be visible from a dashboard. The link to the dashboard will be accessible from data.ecla.inrae.fr.

11. 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 higher uncertainties are found for, by order of relative importance: (1) default simulations, (2) hypolimnion compared to epilimnion temperatures and, (3) deep lakes, in particular reservoirs (maximal depth greater than 10 m for epilimnion temperature and around 10 m for hypolimnion temperature simulated with default model parameters). Although the calibration significantly decreases the uncertainties related to both the epilimnion and hypolimnion, in some cases they are still considerable in the hypolimnion. Based on these results and if enough observation data are available, optimally we recommend the use of the OKPLM for shallow (maximal depth < 8

19

536   m) lakes with calibrated model parameters. However, if applied in its default or even calibrated configuration over

537   deep lakes, one should be aware of the presented limitations and address them in the analysis. The LakeTSim

538   dataset is valuable for assessing the impact of climate change on lakes thermal functioning, which is often hindered

539   by the lack of water temperature observations. The present dataset will provide new insights about the thermal

540   behavior of French lakes, which can provide useful context for stakeholders as they design management strategies

541   in a context of climate change.

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563
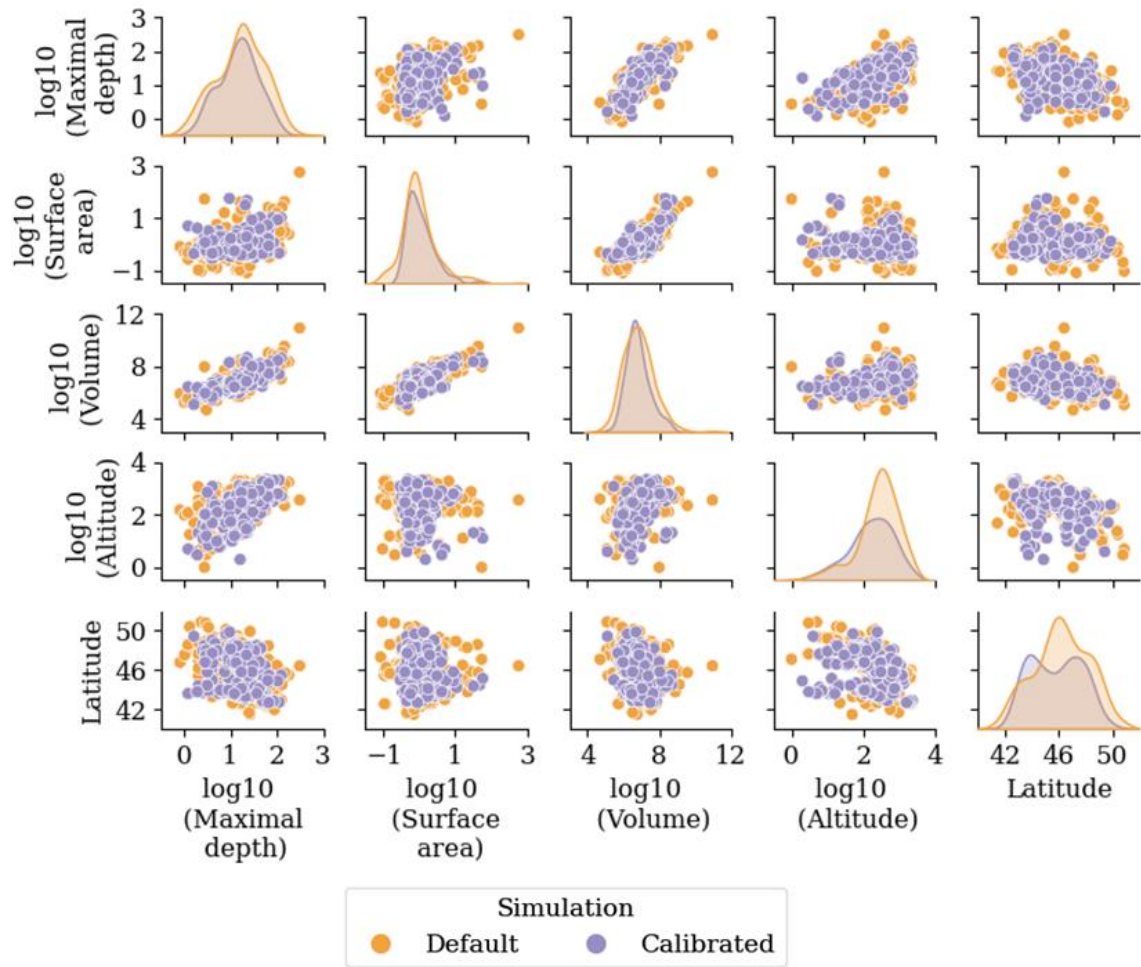
Figure A1: Scatter plots of lakes (n = 401) geomorphological characteristics: Maximal depth (m), Surface area (km^2), volume (m^3), altitude (m) and latitude (°N).
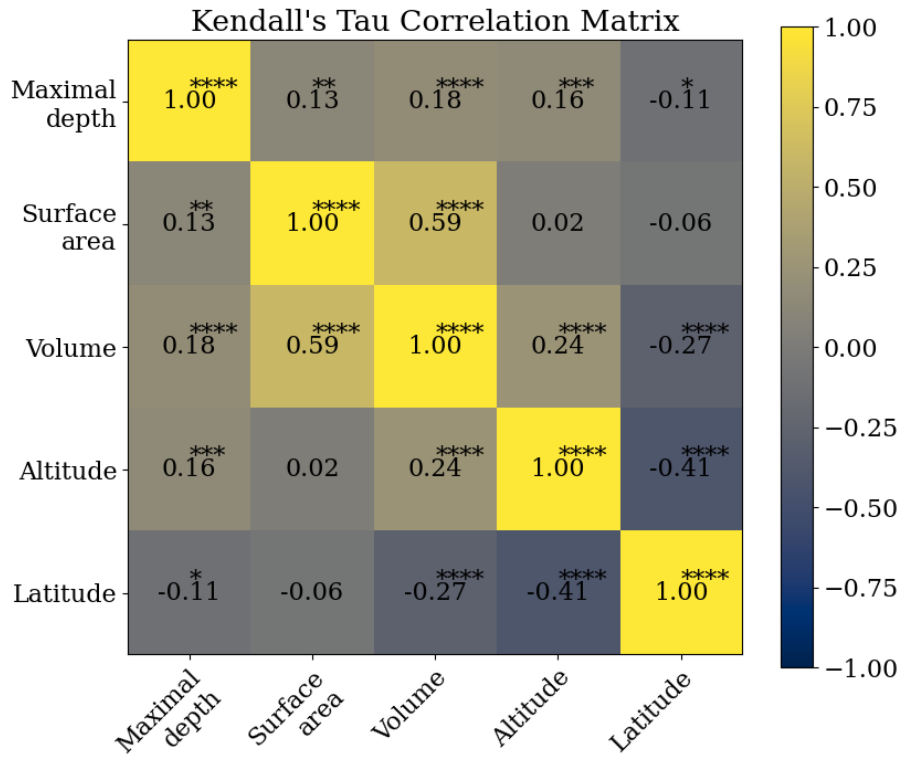
Figure A2: Kendall's tau correlation matrix of the geomorphological characteristics of lakes simulated in "default" mode ($n = 231$): Maximal depth (m), Surface area (km^2), volume (m^3), altitude (m) and latitude ($°$N). The significance levels are represented as follows: *: $1.00e\text{-}02 < p$-value $\leq 5.00e\text{-}02$, **: $1.00e\text{-}03 < p$-value $\leq 1.00e\text{-}02$, ***: $1.00e\text{-}04 < p$-value $\leq 1.00e\text{-}03$, ****: p-value $\leq 1.00e\text{-}04$. Otherwise, correlations are not significant (p-value > 0.05).
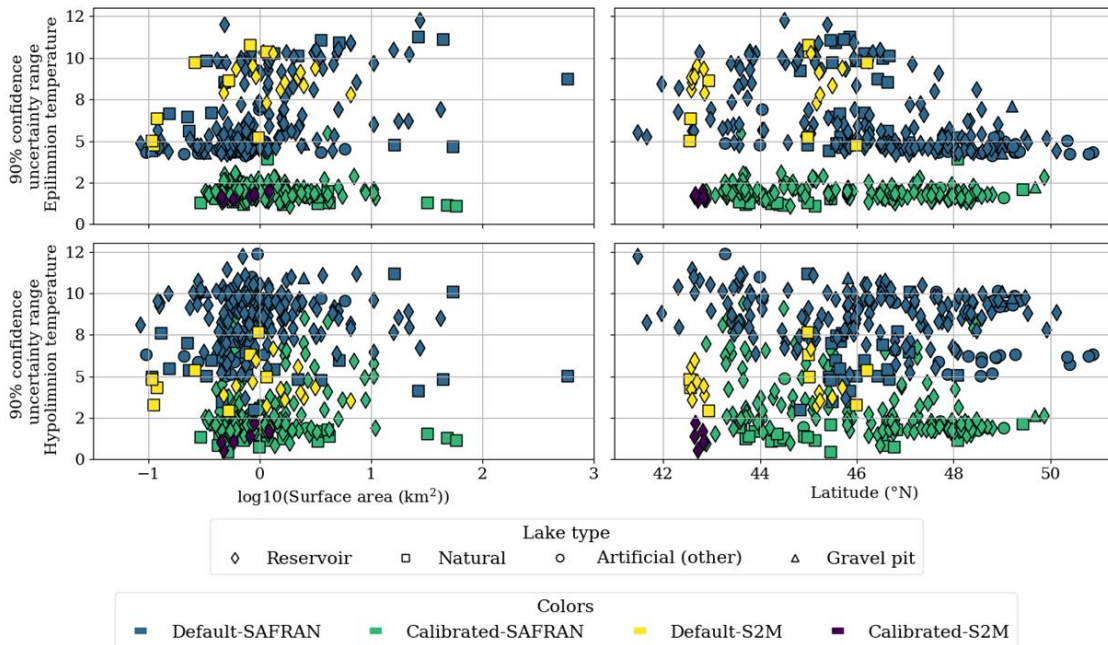
Figure A3: Average 90% confidence uncertainty range for epilimnion (top panel) and hypolimnion (bottom panel) temperatures in calibrated ($n = 170$) and default ($n = 231$) simulations for the period 1959-2020 as a function of surface area (km^2) and latitude ($°$N). The "other" artificial lakes consist of ponds and quarry lakes.
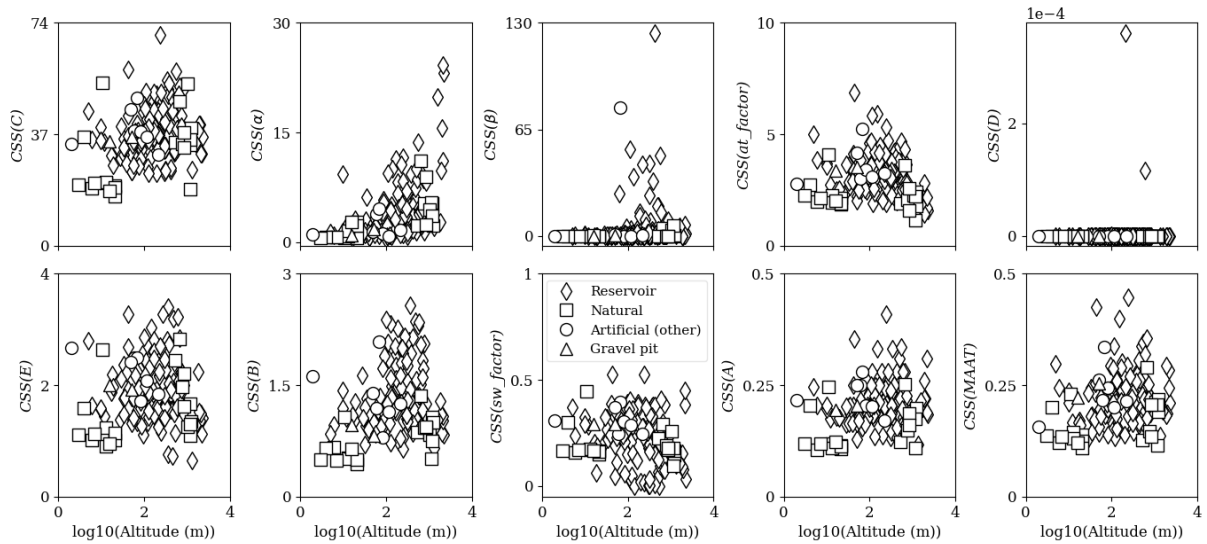
565

Figure A4: Composite scaled sensitivities (*CSS*) for each model parameter as a function of altitude. The "other" artificial lakes consist of ponds and quarry lakes.
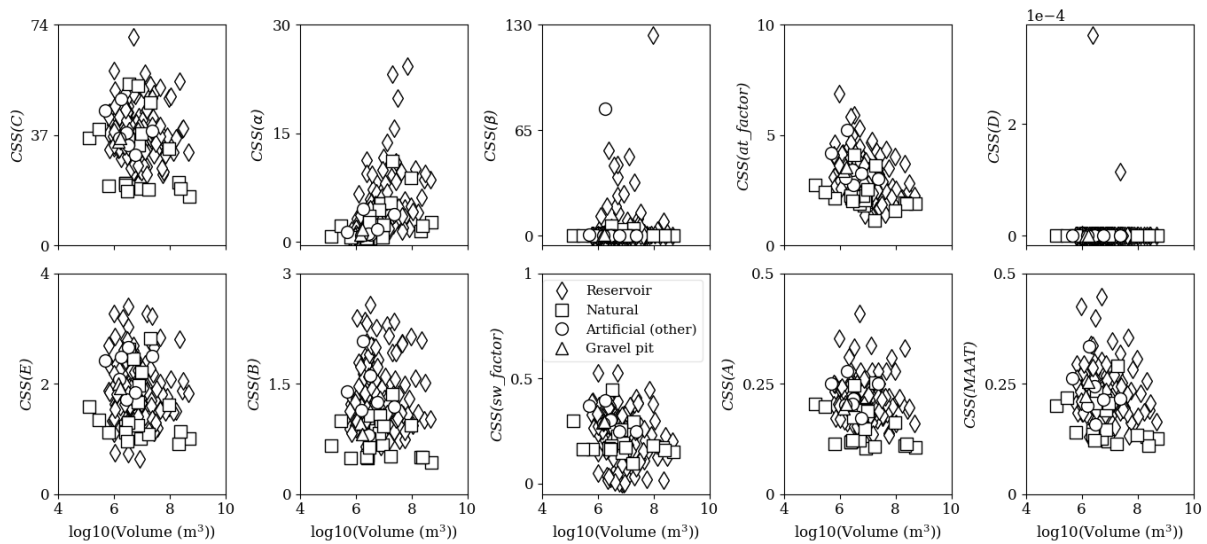

Figure A5: Composite scaled sensitivities (*CSS*) for each model parameter as a function of volume. The "other" artificial lakes consist of ponds and quarry lakes.
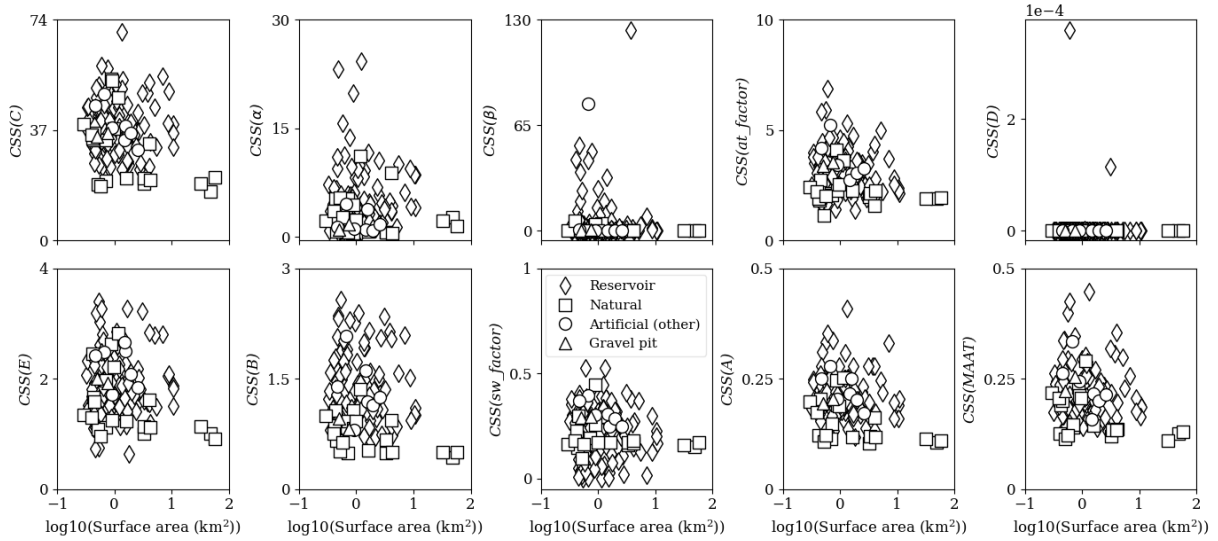
Figure A6: Composite scaled sensitivities (*CSS*) for each model parameter as a function of surface area. The "other" artificial lakes consist of ponds and quarry lakes.
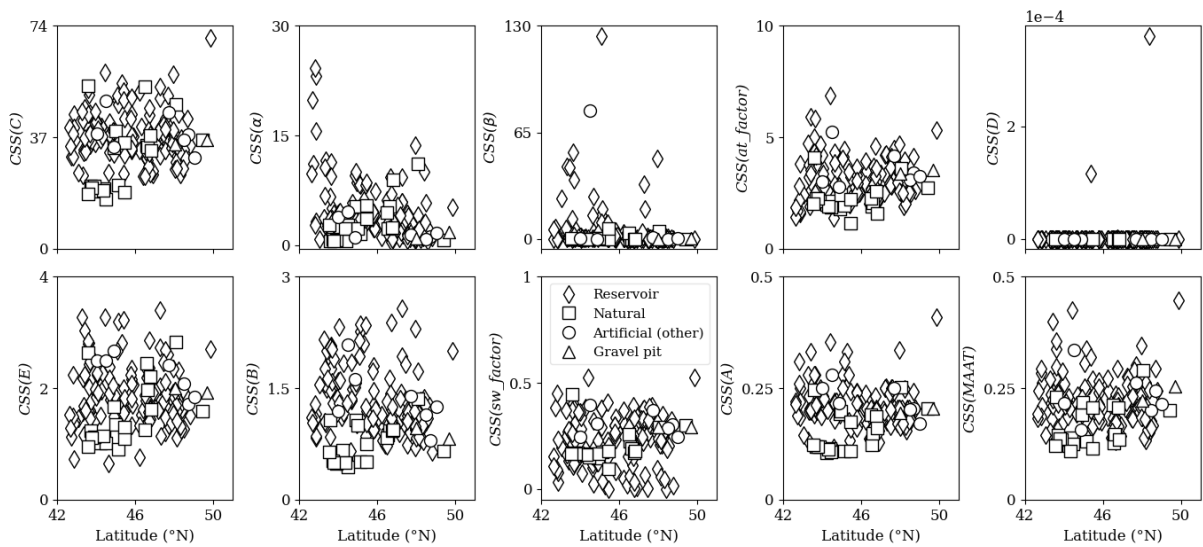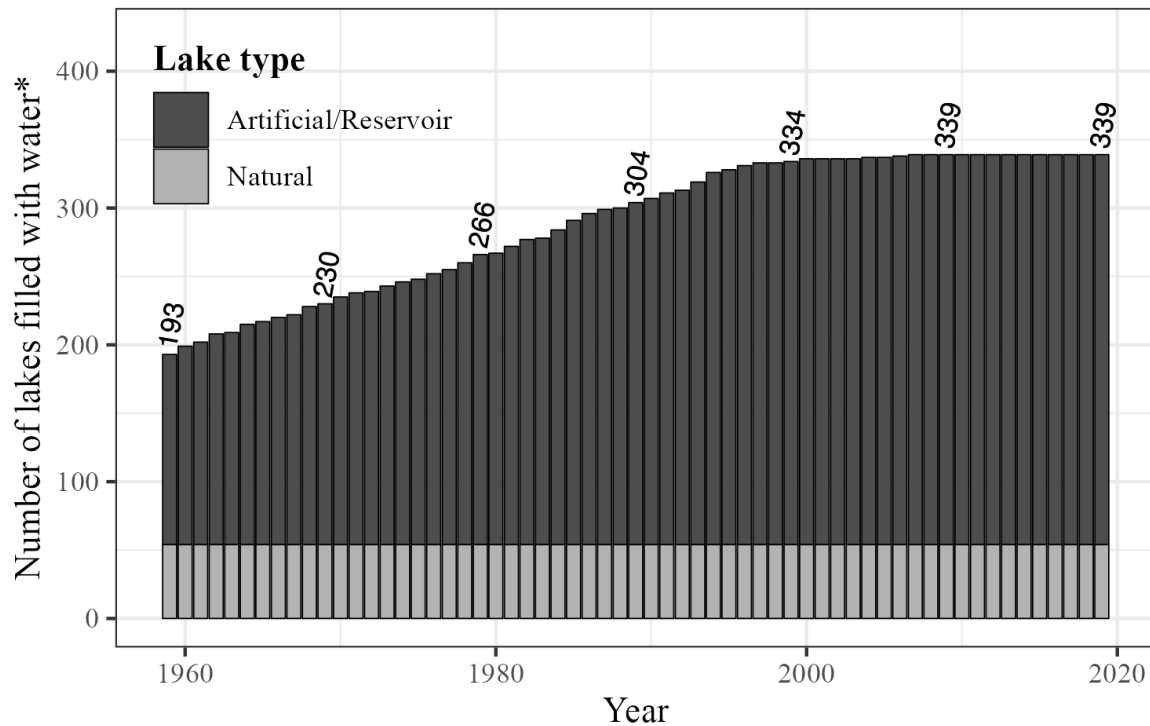

Figure A7: Composite scaled sensitivities (*CSS*) for each model parameter as a function of latitude. The "other" artificial lakes consist of ponds and quarry lakes.

* 7 gravel pits, 33 reservoirs & 25 other artificial lakes with no information on filling year.

Figure A8: Distribution of initial filling years for lakes (e.g., reservoirs, gravel pits, ponds and quarry lakes and other artificial lakes) of the LakeTSim dataset.

13. 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 and NS 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. RB proposed and contributed to the integration of the data consisting of initial filling dates of reservoirs and other artificial lakes in the manuscript. NR and TT supervised and contributed to the implementation of simulation results in the database. NR processed S2M data and compiled the data for initial filling years of reservoirs and other artificial lakes. NS, RB, and NR prepared the figures. 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.

14. Competing interests

The authors declare that they have no conflict of interest.

25

590 **17. References**

591 Adrian, R., O'Reilly, C. M., Zagarese, H., Baines, S. B., Hessen, D. O., Keller, W., Livingstone, D. M.,
592 Sommaruga, R., Straile, D., Van Donk, E., Weyhenmeyer, G. A., and Winder, M.: Lakes as sentinels of climate
593 change, Limnol. Oceanogr., 54, 2283–2297, https://doi.org/10.4319/lo.2009.54.6_part_2.2283, 2009.

594 Allan, M. G., Hamilton, D. P., Trolle, D., Muraoka, K., and McBride, C.: Spatial heterogeneity in geothermally-
595 influenced lakes derived from atmospherically corrected Landsat thermal imagery and three-dimensional
596 hydrodynamic modelling, Int. J. Appl. Earth Obs. Geoinf., 50, 106–116, https://doi.org/10.1016/j.jag.2016.03.006,
597 2016.

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

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

604 Danis, P.: Rapport d ' avancement sur les outils de modélisation thermique, 1–8 pp., https://doi.org/hal-03253847,
605 2020.

606 Daufresne, M., Lengfellner, K., and Sommer, U.: Global warming benefits the small in aquatic ecosystems, Proc.
607 Natl. Acad. Sci., 106, 12788–12793, https://doi.org/10.1073/pnas.0902080106, 2009.

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

611 Durand, Y., Brun, E., Merindol, L., Guyomarc'h, G., Lesaffre, B., and Martin, E.: A meteorological estimation of
612 relevant parameters for snow models, Ann. Glaciol., 18, 65–71, https://doi.org/10.3189/S0260305500011277,
613 1993.

614 Ely, D. M.: Analysis of Sensitivity of Simulated Recharge to Selected Parameters for Seven Watersheds Modeled
615 Using the Precipitation-Runoff Modeling System, U.S. Geological Survey Scientific Investigations Rep, 21 pp.,
616 2006.

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

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

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

625 Halverson, G. H., Lee, C. M., Hestir, E. L., Hulley, G. C., Cawse-Nicholson, K., Hook, S. J., Bergamaschi, B. A.,
626 Acuña, S., Tufillaro, N. B., Radocinski, R. G., Rivera, G., and Sommer, T. R.: Decline in Thermal Habitat
627 Conditions for the Endangered Delta Smelt as Seen from Landsat Satellites (1985–2019), Environ. Sci. Technol.,

628   56, 185–193, https://doi.org/10.1021/acs.est.1c02837, 2022.

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

631   Hill, M. C.: Methods and guidelines for effective model calibration: U.S. Geological Survey Water-Resources
632   Investigations Report 98-4005, 90 pp., https://doi.org/10.3133/wri984005, 1998.

633   Hipsey, M. R., Bruce, L. C., Boon, C., Busch, B., Carey, C. C., Hamilton, D. P., Hanson, P. C., Read, J. S., De
634   Sousa, E., Weber, M., and Winslow, L. A.: A General Lake Model (GLM 3.0) for linking with high-frequency
635   sensor data from the Global Lake Ecological Observatory Network (GLEON), Geosci. Model Dev., 12, 473–523,
636   https://doi.org/10.5194/gmd-12-473-2019, 2019.

637   Janssen, A. B. G., Hilt, S., Kosten, S., de Klein, J. J. M., Paerl, H. W., and Van de Waal, D. B.: Shifting states,
638   shifting services: Linking regime shifts to changes in ecosystem services of shallow lakes, Freshw. Biol., 66, 1–
639   12, https://doi.org/10.1111/fwb.13582, 2021.

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

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

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

650   Kharouba, H. M., Ehrlén, J., Gelman, A., Bolmgren, K., Allen, J. M., Travers, S. E., and Wolkovich, E. M.: Global
651   shifts in the phenological synchrony of species interactions over recent decades, Proc. Natl. Acad. Sci. U. S. A.,
652   115, 5211–5216, https://doi.org/10.1073/pnas.1714511115, 2018.

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

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

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

661   Lindenschmidt, K. E.: The effect of complexity on parameter sensitivity and model uncertainty in river water
662   quality modelling, Ecol. Modell., 190, 72–86, https://doi.org/10.1016/j.ecolmodel.2005.04.016, 2006.

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

665   Nouchi, V., Kutser, T., Wüest, A., Müller, B., Odermatt, D., Baracchini, T., and Bouffard, D.: Resolving
666   biogeochemical processes in lakes using remote sensing, Aquat. Sci., 81, 1–13, https://doi.org/10.1007/s00027-
667   019-0626-3, 2019.

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

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

673   Piccolroaz, S.: Prediction of lake surface temperature using the air2water model: Guidelines, challenges, and future
674   perspectives, Adv. Oceanogr. Limnol., 7, 36–50, https://doi.org/10.4081/aiol.2016.5791, 2016.

675   Piccolroaz, S., Toffolon, M., and Majone, B.: A simple lumped model to convert air temperature into surface water

676   temperature in lakes, Hydrol. Earth Syst. Sci., 17, 3323–3338, https://doi.org/10.5194/hess-17-3323-2013, 2013.

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

680   Pilla, R. M., Williamson, C. E., Adamovich, B. V, Adrian, R., Anneville, O., Chandra, S., Colom-Montero, W.,
681   Devlin, S. P., Dix, M. A., Dokulil, M. T., Gaiser, E. E., Girdner, S. F., Hambright, K. D., Hamilton, D. P., Havens,
682   K., Hessen, D. O., Higgins, S. N., Huttula, T. H., Huuskonen, H., Isles, P. D. F., Joehnk, K. D., Jones, I. D., Keller,
683   W. B., Knoll, L. B., Korhonen, J., Kraemer, B. M., Leavitt, P. R., Lepori, F., Luger, M. S., Maberly, S. C., Melack,
684   J. M., Melles, S. J., Müller-Navarra, D. C., Pierson, D. C., Pislegina, H. V, Plisnier, P.-D., Richardson, D. C.,
685   Rimmer, A., Rogora, M., Rusak, J. A., Sadro, S., Salmaso, N., Saros, J. E., Saulnier-Talbot, É., Schindler, D. E.,
686   Schmid, M., Shimaraeva, S. V, Silow, E. A., Sitoki, L. M., Sommaruga, R., Straile, D., Strock, K. E., Thiery, W.,
687   Timofeyev, M. A., Verburg, P., Vinebrooke, R. D., Weyhenmeyer, G. A., and Zadereev, E.: Deeper waters are
688   changing less consistently than surface waters in a global analysis of 102 lakes, Sci. Rep., 10, 20514,
689   https://doi.org/10.1038/s41598-020-76873-x, 2020.

690   Poeter, E. P. and Hill, M. C.: Inverse models: A necessary next step in ground-water modeling, Groundwater, 35,
691   250–260, 1997.

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

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

696   Prats, J. and Danis, P.-A.: Optimisation du réseau national de suivi pérenne in situ de la température des plans
697   d'eau : apport de la modélisation et des données satellitaires, [Rapport Rech. irstea, 93, https://doi.org/hal-
698   02602604, 2015.

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

701   Prats, J., Reynaud, N., Rebière, D., Peroux, T., Tormos, T., and Danis, P. A.: LakeSST: Lake Skin Surface
702   Temperature in French inland water bodies for 1999-2016 from Landsat archives, Earth Syst. Sci. Data, 10, 727–
703   743, https://doi.org/10.5194/essd-10-727-2018, 2018.

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

707   Rimet, F., Anneville, O., Barbet, D., Chardon, C., Crépin, L., Domaizon, I., Dorioz, J. M., Espinat, L., Frossard,
708   V., Guillard, J., Goulon, C., Hamelet, V., Hustache, J. C., Jacquet, S., Lainé, L., Montuelle, B., Perney, P., Quetin,
709   P., Rasconi, S., Schellenberger, A., Tran-Khac, V., and Monet, G.: The Observatory on LAkes (OLA) database:
710   Sixty   years   of   environmental   data   accessible   to   the   public,   J.   Limnol.,   79,   164–178,
711   https://doi.org/10.4081/jlimnol.2020.1944, 2020.

712   Schaeffer, B. A., Iiames, J., Dwyer, J., Urquhart, E., Salls, W., Rover, J., and Seegers, B.: An initial validation of
713   Landsat 5 and 7 derived surface water temperature for U.S. lakes, reservoirs, and estuaries, Int. J. Remote Sens.,
714   39, 7789–7805, https://doi.org/10.1080/01431161.2018.1471545, 2018.

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

718   Sharaf, N., Lemaire, B. J., Fadel, A., Slim, K., and Vinçon-Leite, B.: Assessing the thermal regime of poorly
719   monitored reservoirs with a combined satellite and three-dimensional modeling approach, Inl. Waters, 11, 302–
720   314, https://doi.org/10.1080/20442041.2021.1913937, 2021.

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

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

725 Sharma, S., Gray, D. K., Read, J. S., O'Reilly, C. M., Schneider, P., Qudrat, A., Gries, C., Stefanoff, S., Hampton,
726 S. E., Hook, S., Lenters, J. D., Livingstone, D. M., McIntyre, P. B., Adrian, R., Allan, M. G., Anneville, O., Arvola,
727 L., Austin, J., Bailey, J., Baron, J. S., Brookes, J., Chen, Y., Daly, R., Dokulil, M., Dong, B., Ewing, K., De Eyto,
728 E., Hamilton, D., Havens, K., Haydon, S., Hetzenauer, H., Heneberry, J., Hetherington, A. L., Higgins, S. N.,
729 Hixson, E., Izmest'eva, L. R., Jones, B. M., Kangur, K., Kasprzak, P., Köster, O., Kraemer, B. M., Kumagai, M.,
730 Kuusisto, E., Leshkevich, G., May, L., MacIntyre, S., Müller-Navarra, D., Naumenko, M., Noges, P., Noges, T.,
731 Niederhauser, P., North, R. P., Paterson, A. M., Plisnier, P. D., Rigosi, A., Rimmer, A., Rogora, M., Rudstam, L.,
732 Rusak, J. A., Salmaso, N., Samal, N. R., Schindler, D. E., Schladow, G., Schmidt, S. R., Schultz, T., Silow, E. A.,
733 Straile, D., Teubner, K., Verburg, P., Voutilainen, A., Watkinson, A., Weyhenmeyer, G. A., Williamson, C. E.,
734 and Woo, K. H.: A global database of lake surface temperatures collected by in situ and satellite methods from
735 1985-2009, Sci. Data, 2, 1–19, https://doi.org/10.1038/sdata.2015.8, 2015.

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

738 Snowling, S. D. and Kramer, J. R.: Evaluating modelling uncertainty for model selection, Ecol. Modell., 138, 17–
739 30, https://doi.org/10.1016/S0304-3800(00)00390-2, 2001.

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

743 Vernay, M., Lafaysse, M., Mérindol, L., Giraud, G., and Morin, S.: Ensemble forecasting of snowpack conditions
744 and avalanche hazard, Cold Reg. Sci. Technol., 120, 251–262, https://doi.org/10.1016/j.coldregions.2015.04.010,
745 2015.

746 Vernay, M., Lafaysse, M., Monteiro, D., Hagenmuller, P., Nheili, R., Samacoïts, R., Verfaillie, D., and Morin, S.:
747 The S2M meteorological and snow cover reanalysis over the French mountainous areas: description and evaluation
748 (1958-2021), Earth Syst. Sci. Data, 14, 1707–1733, https://doi.org/10.5194/essd-14-1707-2022, 2022.

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

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

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

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

759 Yang, K., Yu, Z., Luo, Y., Yang, Y., Zhao, L., and Zhou, X.: Spatial and temporal variations in the relationship
760 between lake water surface temperatures and water quality - A case study of Dianchi Lake, Sci. Total Environ.,
761 624, 859–871, https://doi.org/10.1016/j.scitotenv.2017.12.119, 2018.

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

764