



HAL
open science

Variance Based Sensitivity Analysis of FLake Lake Model for Global Land Surface Modeling

A Bernus, C Ottlé, N Raoult

► **To cite this version:**

A Bernus, C Ottlé, N Raoult. Variance Based Sensitivity Analysis of FLake Lake Model for Global Land Surface Modeling. *Journal of Geophysical Research: Atmospheres*, 2021, 126 (8), pp.e2019JD031928. 10.1029/2019jd031928 . hal-03300813

HAL Id: hal-03300813

<https://hal.science/hal-03300813>

Submitted on 27 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Variance Based Sensitivity Analysis of FLake Lake Model for Global Land Surface Modeling

**Key Points:**

- FLake model parameters sensitivities on lake surface temperature and sensible and latent heat fluxes are analyzed
- Parameter sensitivities to surface variables vary with depth, climate, and season
- Respective roles of depth, fetch, surface albedo, light extinction, and thermocline shape coefficients are quantified

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

A. Bernus,
anthony.bernus@lsce.ipsl.fr

Citation:

Bernus, A., Ottlé, C., & Raoult, N. (2021). Variance based sensitivity analysis of FLake lake model for global land surface modeling. *Journal of Geophysical Research: Atmospheres*, 126, e2019JD031928. <https://doi.org/10.1029/2019JD031928>

Received 31 OCT 2019
 Accepted 13 NOV 2020

Author Contributions:

Conceptualization: C. Ottlé
Data curation: A. Bernus
Formal analysis: A. Bernus, C. Ottlé
Funding acquisition: C. Ottlé
Investigation: A. Bernus
Methodology: A. Bernus, C. Ottlé
Project Administration: C. Ottlé
Resources: C. Ottlé
Software: A. Bernus
Supervision: C. Ottlé
Validation: A. Bernus
Visualization: A. Bernus, N. Raoult
Writing – original draft: A. Bernus, C. Ottlé
Writing – review & editing: A. Bernus, C. Ottlé, N. Raoult

© 2021. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

A. Bernus¹ , C. Ottlé¹ , and N. Raoult¹ 

¹Laboratoire des Sciences du Climat et de l'Environnement, IPSL, CEA-CNRS-UVSQ, Gif-sur-Yvette, France

Abstract Given the ever increasing spatial resolution of climate models and the significant role of lakes on the regional climate, it becomes important to represent water bodies in climate models. Such developments have started in the IPSL (Institut Pierre Simon Laplace) climate model and its land surface component, ORganizing Carbon and Hydrology In Dynamic Ecosystems (ORCHIDEE), with the Freshwater Lake model, FLake. To answer the questions raised by these new developments, such as the lake differentiation and related model parameters, we analyze spatial distributions of lake characteristics in the whole world to perform a global sensitivity analysis of the FLake parameters. As a result, three different climates and four lake depth configurations were selected as test cases. The Sobol method as sensitivity analysis based on variance decomposition was chosen to rank parameters impact on the model output, that is, lake surface water temperature, latent and sensible heat fluxes. We focus on the 11 parameters of the FLake model, which are the lake depth, the albedo and light extinction coefficient of water, snow, and ice respectively, the fetch, and the relaxation coefficient of the thermocline shape factor. The results show different sensitivity features according to the lake type and climate. The dominant role and time varying contribution of the lake depth, radiative parameters (albedo, light extinction coefficient) and thermocline relaxation coefficient linked to the atmospheric conditions, were clearly highlighted. These findings will lead us to distinguish between different lake categories in each grid cell of ORCHIDEE in the future implementation.

Plain Language Summary Lakes are often neglected in climate modeling because their spatial extent does not exceed 4% of the Earth land area. But given the ever increasing spatial resolution of climate models and the significant role of lakes on the regional climate, it becomes important to represent water bodies in climate models. Such developments have started in the French IPSL climate model with the implementation of a one-dimensional freshwater lake model (FLake). To answer the questions raised by these new developments, such as the lake differentiation and related model parameters, we analyze spatial distributions of lake characteristics in the whole world to perform a global sensitivity analysis of the model parameters. Our results show the respective roles of the depth, fetch, thermocline, sediment, and radiative (albedo and light extinction coefficient) parameters in different climate conditions and at different time scales. Some avenues for further implementation and model parameters calibration are finally given.

1. Introduction

Lakes are often neglected in climate modeling because their spatial extent at global scale does not exceed 3.7% of the Earth's non glacial land area (Verpoorter et al., 2014). However, larger concentrations can be found in some regions. For example, Telmer and Costa (2007) studying two regions of 200 by 200 km in Canada and in Brazil, found that lakes cover respectively 33% and 19% of the land surface. In such conditions, lakes may have real impact on the atmospheric exchanges and consequently on weather prediction and climate modeling. Actually, the specific properties and physics of lakes (i.e., high heat capacity, low albedo and surface roughness, and water-ice transition phases) explain their major role on local meteorology and boundary layer stratification, as well as their influence on cloud formation (Laird et al., 2009). This has been demonstrated in numerous studies at global scale (Bonan, 1995; Dutra et al., 2010; Krinner, 2003; Mironov et al., 2010) and more specifically on the climate of regions where large fractions of land are covered by lakes. For example, Chuang and Sousounis (2003) showed the impact of the American/Canadian great lakes on the regional climate. They especially explained how processes associated with

these lakes deepen surface pressure, enhance cyclones, and consequently generate or strengthen a warm front. Song et al. (2004) and Thiery et al. (2015) studied the eastern African lakes and their influence. They notably showed the cooler effect of lakes during the day, the warmer one during the night, and how lakes strengthen precipitation. Even though the impacts on the energy and water cycles are more significant at regional scales, the importance of lakes on climate prediction at larger scales has also been shown by various authors (e.g., Bonan, 1995 or Le Moigne et al., 2016). As an example, using the CNRM-CM5 climate model, Le Moigne et al. (2016) showed the substantial role of lakes on the simulation of regional surface energy budgets and the reduction of some air temperature biases mainly over the region of the Canadian Great Lakes.

The role of lakes and especially of thermokarst lakes on the methane global cycle and its climate feedbacks is also an important question. Bastviken et al. (2004) pointed out the important lake contribution to the non-anthropogenic methane emissions and consequently on climate. They estimated that between 8% and 16% of non-anthropogenic methane emissions come from lakes, and West et al. (2016) specified that the larger contribution arises from shallow and eutrophic lakes.

Given these features and the continuous increased spatial resolution of general circulation models, many efforts have been pursued during the last decade to include lake models in numerical weather prediction (NWP) and climate models. A good trade-off between numerical cost and the correct representation of the surface main processes is offered by one dimensional (1-D) approaches. Between the crude representation of bulk models representing lakes as a single water column completely mixed (Ljungemyr et al., 1996) and multi-layer approaches resolving the transport equations for the turbulence kinetic energy and its dissipation rate (Stepanenko, 2005), conceptual models such as the FLake model (Mironov, 2008) appear to be a good compromise able to simulate the lake water surface temperature (LWST) and the thermocline profile with a two-layer resolution of the energy budget.

FLake is a well-tested model and its ability to predict the temperature structure of various lakes has been demonstrated in numerous studies (Voros et al., 2010; Woolway & Merchant, 2019). FLake has been implemented in numerous climate and NWP models. To cite but a few, FLake is part of the German COSMO and TERRA land component (Mironov et al., 2010), the European ECMWF Earth System Model and its HTESSEL land surface component (Balsamo et al., 2012; Dutra et al., 2010), the UK Met Office Unified Model and the JULES Land simulator (Rooney & Bornemann, 2013), and the French CNRM-CM5 and its SURFEX interface (Le Moigne et al., 2016; Salgado & Le Moigne, 2010). All these works have demonstrated the added value of the FLake model in better simulating regional weather and climates.

When implemented at large scale, a lake model requires the definition of a lake fraction in each grid cell of the land surface model. Generally, the grid cell size is a few tens of kilometers for a climate model. Such information can be offered by existing global databases generally derived from remote sensing data. These databases provide water bodies at a very fine spatial resolution, according to the decametric resolution of optical and microwave sensors. As an example, the product of Pekel et al. (2016), derived from Landsat imagery, provides the water extent of each water body with a resolution of 30 m, during 32 years 1984–2015 and at a monthly time scale.

In addition to determining the different types of water surfaces that need to be considered, the definition of the model parameters necessary to initialize and run the model is an important task. In this aim, the prior model sensitivity analysis (SA) is a powerful tool that has to be conducted on the basis of previous knowledge about the spatial/temporal variabilities of lake features. Previous studies based on FLake model have shown the dominant role of the lake depth (Balsamo, 2010) and of the extinction coefficient of solar radiation (Heiskanen et al., 2015). The lake bathymetry has definitely a strong impact on the temporal dynamics of its temperature and capacity to mix and freeze. This property has been used by Balsamo (2010) to derive successfully, lake depth by optimization of LWST with observations. More recently, Layden et al. (2016) showed that not only lake depth but also snow and ice albedos and light extinction coefficient could be optimized in the same way globally, with space thermal infrared measurements. Although these studies have demonstrated the dominant role of lake depth and radiative properties on LWST, in other words their sensitivities, to our knowledge, no quantitative SA of FLake output variables has been conducted until now.

The work presented in this study aims to fill this gap. It is part of a community project devoted to the inclusion of water bodies in the IPSL climate model and more precisely in its land surface component: the ORCHIDEE land surface model. Previous studies to represent lakes in ORCHIDEE were focused either on the atmospheric impacts with no connections with the other hydrological processes (Krinner, 2003; S. Peng, personal communication, 2015) or on their impacts on the water flows via the routing hydrological scheme (d'Orgeval et al., 2008; X. Zhou, personal communication, 2019). In our future developments, we aim to represent the lakes' water and energy budgets explicitly by the addition of a lake tile in the ORCHIDEE Multi-Energy Budget version under development at IPSL. For this purpose, the FLake model has been chosen to resolve the surface energy budget on the water fraction of each grid mesh. However, before running the model, a number of parameters need to be prescribed and calibrated. To help this implementation in ORCHIDEE, we performed a preliminary SA of the FLake model in standalone mode. SA is an effective tool used to identify the dominant processes and understand their dynamics. It allows to rank the parameters and to determine which ones should require the largest attention in the following calibration step (Saltelli, 2008).

In this study, we therefore show the results of a global sensitivity analysis of the FLake model designed with this aim of its forthcoming global scale implementation. Using three representative climates, we therefore focus on three of its main output variables at the atmosphere interface, that is, LWST and the turbulent latent (H_L) and sensible (H_S) heat fluxes. The goal of our study is then to answer the following questions:

- What are the main characteristics of lakes at a global scale?
- How do we define the study cases and the parameter space for this sensitivity analysis? Given the parameter space, which parameters are the most influential?
- How should model implementation be done to account for this diversity of lake properties in parameterization?

2. Materials and Methods

2.1. FLake Model

FLake (Mironov, 2008) is a 1-D thermodynamic lake model developed for NWP purposes. It is a bulk model capable of predicting the vertical temperature structure and mixing conditions in lakes, given the meteorological conditions at the atmosphere interface (incoming radiation, air temperature and humidity, and wind speed). The water temperature profile is represented by a single mixed layer with a uniform temperature above a thermocline. Extra modules are implemented to model the snow, ice, and sediment profile temperatures into specific layers. The sediment layer allows for the representation of the heat transfer at the lake bottom-sediment interface, which is not negligible in the case of shallow lakes. The structure of the lake thermocline is parameterized using the concept of self-similarity. Therefore, the depth-temperature relationship depends on a shape factor, which is resolved at each time step according to the boundary conditions. The same approach of "assumed shape" is used to represent the snow, ice and sediment layer temperature profiles (see for examples Figures 3 and 4 in Mironov, 2008 for a representation of the depth-temperature relationships).

The resolution of the bulk energy budgets of the mixed layer and thermocline allows to calculate the model prognostic variables, that is, the mixed-layer temperature and depth, the free water bottom temperature, the thermocline shape factor, the ice layer top temperature, and ice thickness. The resolution of the energy budget of the mixed layer allows to estimate the latent and sensible heat fluxes at the water-atmosphere interface.

When a sediment layer is accounted, two more variables are predicted: the attenuation depth of the annual thermal wave and the temperature at that depth. The bottom temperature and the total depth of the sediment layer are prescribed. If the snow module is activated, the snow temperature at the atmosphere interface and the snow layer thickness are calculated, based on the bulk resolution of the heat budget of the snow layer. It should be noted that the model resolves only the energy budget equations and that the water balance is not solved. That means that the water volume is kept constant in time, the lake depth and surface extent are therefore input parameters of the model. In addition, two important radiative parameters (surface albedo and light extinction coefficient both for water, snow

and ice) and the surface fetch involved in the calculation of the surface fluxes need to be prescribed to run the model.

FLake does not model the hypolimnion, the layer under the thermocline which is present or may appear seasonally in stratified lakes and where the water density is the highest with a constant temperature near of 4°C. To run the model, we add a false bottom by shortening the lake depth of deep lakes to 50 m. This allows us to get around the modelization of the hypolimnion. Also, the absence of 2-D and 3-D processes (e.g., lake currents) has shown the limits of FLake's performances in certain cases. For example, FLake fails to capture the typical pattern of springtime warming in the deep North American Great Lakes (Martynov et al., 2010). However, such examples remain specific and the overall performance of FLake is promising, especially given its low numerical cost.

For the implementation of FLake in the ORCHIDEE land surface model, a multi-tiled version of the model is used, for which separate energy budgets can be calculated for each type of surface and especially for water bodies. In order to prepare the lake model developments and the necessary step of parameter calibration, a global SA has been designed and performed.

2.2. Sobol Method

SA studies the variation of the model simulated variables according to the variations of some of its parameters. Here, we perform a global SA where each parameter varies over the entire range of variation. As such, we get information for the whole parameter space compared to local SA techniques where we would only get information around one point. Also, the parameters in a global SA can vary simultaneously, whereas this is not the case for local SA experiments where the parameters vary one by one (Saltelli, 2008).

First, we need to define the model inputs, parameters, forcing, and outputs not to mix up these concepts. The inputs are the external data needed to run the model. These can take the form of parameters, which describe explicitly the system (here the lake). Alternatively, forcing variables are inputs that provide the boundary conditions of the model (here the interface between the lake and the atmosphere). The atmospheric forcing inputs are the meteorological variables used to drive the model. These variables are not taken into account during the SA. Finally, the outputs are the variables that the model calculates.

To quantify the model sensitivity, we used the Sobol method. The main principle is to decompose the variance of the considered outputs into variances explained by the variation of each input. During this process, the parameter sampling should be independent from each other.

Briefly, the method is decomposed into four steps. The first one is to list the parameters and define their range of variation. Each parameter will be associated with a random variable uniformly distributed on each interval. The second is to consider the parameter as random variable and to sample the intervals (with Monte Carlo method for example) in a way to allow the variance decomposition estimation. The third step is then to perform the simulations needed and to choose a scalar output; this can either be the model output directly if it is already a scalar or a value calculated from the model output to create a scalar, for example, the Root Mean Square Error between modeled time series and observations. The final step consists in the analysis of the model output variance according to the parameter ranges.

This method is widely used in geophysical sciences and especially by the land surface model community. For example, Xu et al. (2012) applied this method to study the parameter sensitivity of the Common Land Surface Model (CLM). Li et al. (2013) also performed a sensitivity analysis on CLM for six different output variables using several SA methods including the Sobol one. In the same way, Zheng et al. (2019) studied the sensitivity of evapotranspiration and runoff to the Noah-MP land surface model parameters, and Guerrero et al. (2017) studied the sensitivity of the surface fluxes calculated by the Canadian Small Lake Model to its input parameters, in the specific case of a small lake in Canada. To understand how the model output variance can be decomposed with the Sobol approach, we first introduce some definitions. If we define a random variable (or stochastic variable) for each parameter with a uniform probability distribution on each interval, we can represent a model under a functional form:

$$Y = f(\mathbf{X}) = f(X_1, \dots, X_p) \quad (1)$$

where Y is the stochastic variable corresponding to the model output that we want to study and $\mathbf{X} = (X_1, \dots, X_p)$ is the parameter set with p parameters. \mathbf{X} is a random vector with a uniform distribution on their interval definition. As mentioned previously, Y is a scalar and so f is either the model output or the composition between the model and a post-processing metric.

We define $V(Y)$, the variance of the random variable Y , $E(Y)$, the expected value of Y , and $E(Y|X_i)$, the condition expected value of Y given X_i .

Sobol' demonstrated that the function f can be decomposed as in Equation 2. All the terms in this equation are orthogonal (namely the integral of the multiplication of two different terms is null), and the integrals of f_i over X_i interval, f_{ij} over X_i and X_j interval, etc, are null. This decomposition exists and is unique if all the random variables are independent.

$$f(\mathbf{X}) = f_0 + \sum_{i=2}^p f_i(X_i) + \sum_{i=1}^p \sum_{j>i}^p f_{ij}(X_i, X_j) + \dots + f_{12\dots p} \quad (2)$$

It can be shown that the terms of Equation 2 can be written using Equation 3 with $I = \{i_1, i_2, \dots, i_s\}$ be a set of s subscripts and \hat{I} be the concatenation of these subscripts, that is, $i_1 i_2 \dots i_s$:

$$f_{\hat{I}}(X_{i_1}, \dots, X_{i_s}) = E(Y | X_{i_1}, \dots, X_{i_s}) - \sum_{\{Z \in \mathcal{P}(I) | Z \neq I\}} f_{\hat{Z}} \quad (3)$$

where $\mathcal{P}(I)$ denotes the power of set of I , that is, all subsets of I , including I itself and the empty set. By convention, if I or Z is \emptyset , then $f_{\hat{I}}$ or $f_{\hat{Z}}$ is equal to f_0 .

This formula is coherent with f_0 being constant and equal to the expectation values of the output. The first term of Equation 3, $E(Y | X_{i_1}, \dots, X_{i_s})$ is a conditional expected value, and the last term removes all inferior orders so that it only accounts for the iteration between the s parameters used.

From Equation 2, the total variance of function f , $V(Y)$, is decomposed into component variances from individual parameters and their interactions.

Each variance component is the variance of the terms of Equation 2. If the variance of the sum of all the terms is the sum of the variance of each term, it is because each term is orthogonal with each other.

$$V(Y) = \sum_{i=1}^p V_i + \sum_{i=1}^p \sum_{i<j}^p V_{ij} + \dots + V_{12\dots p} \quad (4)$$

where $V_{\hat{I}} = V(E(Y | X_{i_1}, \dots, X_{i_s})) - \sum_{\{Z \in \mathcal{P}(I) | Z \neq I\}} V_{\hat{Z}}$.

From the previous equation, each term can be divided by the total variance to create the so-called Sobol indices. Each term measures the sensitivity of the output Y for each parameter X_i or the interactions between several and is written:

$$S_{\hat{I}} = \frac{V_{\hat{I}}}{V(Y)} \quad (5)$$

The first-order effect indices S_i denotes the sensitivity resulting from the main effect of individual parameter X_i . The second-order effect indice S_{ij} indicates the sensitivity resulting from the interaction of two parameters X_i and X_j . The higher order effect indices are not often calculated for their high numerical cost and their low utility. Another indice can be developed: the total effect sensitivity indice. It represents the main effect of X_i and all its interactions with the other parameters. This indice is defined as:

$$S_{T_i} = 1 - \frac{V_{1\dots i-1, i+1, \dots, p}(E(Y | X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_p))}{V(Y)} \quad (6)$$

All the Sobol indices have values between 0 and 1, and the sum is equal to 1 whereas the sum of the total indices may be superior to 1 if the variance Y linked to parameter interactions is not null.

In our case, the variable Y evolves with time. A preliminary analysis could be to apply the Sobol method on the Y averaged over a time period, typically one year. However, the different mechanisms of the study model varying temporally, this analysis would not be sufficient. In order to explore parameter sensitivities at various time scales, Lamboni et al. (2011) proposed the generalized Sobol indices defined by:

$$GS_i = \frac{\sum_{t=1}^N V(Y(t))}{\sum_{t=1}^N V(Y(t))} S_i(t) \quad \text{and} \quad GS_{T_i} = \frac{\sum_{t=1}^N V(Y(t))}{\sum_{t=1}^N V(Y(t))} S_{T_i}(t) \quad (7)$$

where $\sum_{t=1}^N V(Y(t))$ is the variance of Y over a time period of length N and $V(Y(t))$ the variance of Y at the time step t .

These generalized Sobol indices are very interesting because they allow to show synthetic information at different time scales for example monthly or annual. In the case of a lake model, because of the water phase changes, density and thermal properties, we expect the radiative and convective processes to have time varying relative weights, and so, the influence of the parameters to change with time. Therefore, in this study, the Sobol indices have been also analyzed at different time scales from the annual to the hourly one.

To calculate the model output variances and to derived the Sobol indices, we follow the method presented by Saltelli et al. (2010). Then, we define two matrices \mathbf{A} and \mathbf{B} with p columns and N lines where N is the sample size. The component a_{ij} represents the value of the parameter i for the simulation j .

We also use the notation $[\mathbf{B}]_{1 \leq k \leq p, j}$ for the line j of the matrix \mathbf{B} , k is the column index. Each line corresponds to one model simulation. We also define \mathbf{A}_i , equal to matrix \mathbf{A} where the i th column has been replaced by the corresponding i th column of the \mathbf{B} matrix. To generate these matrices, we use the quasi Monte Carlo method. The sampling is generated by low discrepancy series and here we use the Sobol ones (Sobol, 1990). Then, the estimation of the Sobol indices is done with the following equations:

- For the first order:

$$S_i = \frac{1}{N} \sum_{j=1}^N f([\mathbf{B}]_{1 \leq k \leq p, j}) (f([\mathbf{A}_i]_{1 \leq k \leq p, j}) - f([\mathbf{A}]_{1 \leq k \leq p, j})) \quad (8)$$

- For the total order:

$$S_i = \frac{1}{2N} \sum_{j=1}^N (f([\mathbf{A}]_{1 \leq k \leq p, j}) - f([\mathbf{A}_i]_{1 \leq k \leq p, j}))^2 \quad (9)$$

Although Sobol method has intensive computational requirements, its sensitivity indices have been shown to be more effective than other approaches in capturing the interaction effects of several parameters for highly nonlinear models (Tang et al., 2007). Overall computing the first-order and total-order sensitivity indices requires:

$$N_{sim} = N \times (p + 2) \quad (10)$$

with N_{sim} is the number of simulations, N is the sample size, and p is the number of parameters.

2.3. Experimental Design

The experiments were driven by two items: our final objective to model all the different kinds of lakes in ORCHIDEE, from the boreal thermokarst shallow lakes to the large deep lakes of temperate and tropical regions, and the knowledge of the existing parameter database to estimate the parameter uncertainties.

In order to do so, we separately apply the Sobol method on LWST, H_L , and H_S variables, calculated by FLake over a set of 12 representative lakes ranging different depths and generic climates. Therefore, Y in the previous section is alternatively LWST, H_L , and H_S .

The Sobol method is applied separately for each test case, considering their specific parameter intervals and meteorological forcing. We choose several cases to reflect the diversity in climate and lake size found around the world. We therefore decided to first explore various global databases of lake depth and extent in

order to select the representative lake depths and extents of the global distribution on which to perform the SA. Then, for each representative case, we define the uncertainty of each parameter, based on the available datasets and known literature, to create the intervals and hence the parameter space used for the study.

2.3.1. Representative Study Cases

We used the HydroLAKES database (Messenger et al., 2016) to investigate the global distribution of natural lakes and human-made reservoirs in extent and lake depth. This database provides polygons of water bodies with surface areas of at least 10 ha and their main morphometric features including average lake depth and volume.

We analyzed the distribution of the average lake depth of each lake weighted by their surface extent (see Figure S1). In this figure, a mode around 3 m corresponding to shallow lakes (lakes with depth inferior to 5 m represent about 30% of the total lake area) can be observed, as well as an important contribution of the very deep lakes to this distribution (about 30% of which will be all simulated by FLake with the recommended prescribed depth of 50 m). These features have been already discussed in various works (Messenger et al., 2016; Pekel et al., 2016). Given these features, we decided to retain three typical lake depths in the following SA: 3 and 50 m, and we chose an intermediate case at 25 m to represent the case of medium lake size.

2.3.2. Climate and Meteorological Conditions

Lakes are present in various regions under diverse climatic conditions. We therefore decided to perform our SA on the main three climates (boreal, temperate, and tropical) of the Earth. In order to get realistic but generic meteorological forcing and to explore the sensitivities at different time steps, we use the WFDEI atmospheric reanalysis (Weedon et al., 2014) provided at a spatial resolution of 0.5° and 3-hourly time step, to force our model. We linearized the atmospheric forcing to get a time step of 30 min in order to model the diurnal cycle and assess the model sensitivities at hourly scale.

Three meteorological time series were extracted for a 5-year period (2000–2004) to represent the meteorological conditions of three regions of the world influenced by lakes: Central Siberia, Central Europe, and East Africa. These points were selected to get representative meteorological conditions over our representative lakes in these three main types of climate.

2.3.3. FLake Parameter Space

The first step in a SA is the definition of the parameter space, that is, the study parameters and their range of variation. In our case, it is based on the parameter uncertainties. As previously noted, we want to study the parameter sensitivity in three climates for three typical lake depths. Since the sediment layer has proved to be influential for shallow lakes, we propose to study 12 representative cases resulting from the combination of the three typical climates (boreal, temperate, and tropical) and the four lake types (shallow, shallow with sediments, intermediate, and deep).

The selection of model parameters was done based on previous studies, notably ones using LWST to calibrate the model parameters (Layden et al., 2016). We retain the usual input parameters of FLake (lake depth, extinction coefficient and albedo both for free water, ice and snow, fetch, sediment depth, and bottom temperature) and add the relaxation constant governing the time evolution of the thermocline shape factor, whose value is highly uncertain (Salgado & Le Moigne, 2010; Layden et al., 2016; D. Mironov and G. Kirillin, personal communication, 2019). Thus, the vector \mathbf{X} of the method section corresponds to a vector of random variables of nine-dimensions (when the sediment layer is not activated) and 11 dimensions in the other cases.

The ranges of variation that have been prescribed in the SA were defined according to the following approach. When FLake is run at global scale, these parameters are prescribed to default values, constant with time and space except for lake depth and fetch for which global maps of lake extent and lake depth are available. We therefore assumed that the uncertainties are mostly the result of the non-accounted spatial or temporal variabilities. Then, we defined the parameter ranges of variation from the analysis of existing databases providing such information, as explained in the following. The final ranges of variation are given in Table 1 for the three kinds of lakes (shallow, intermediate, and deep).

Table 1
Parameter Ranges of Variation Used in the SA of the FLake Model

Parameter	Symbol		Shallow (3 m)	Intermediate (25 m)	Deep (50 m)
Lake depth (m)	D		[0.5, 5.5]	[22.5, 27.5]	[47.5, 52.5]
Fetch (m)	F		[0, 1300]	[0, 10600]	[0, 26, 600]
Light extinction coefficient of water (m^{-1})	κ_w			[0.225, 2.435]	
Light extinction coefficient of snow (m^{-1})	κ_s			[15, 25]	
Light extinction coefficient of ice (m^{-1})	κ_i			[8.4, 17.1]	
Water Albedo	α_w			[0.025, 0.175]	
Snow Albedo	α_s			[0.4, 0.8]	
Ice Albedo	α_i			[0.4, 0.65]	
Relaxation coefficient (log)	C_{relax}			[2, 5]	
Sediment ^a :					
bottom temperature (K)	T_H	boreal		[271.2, 273.2]	N/A
		temperate		[282.7, 284.1]	N/A
		tropical		[295.0, 296.8]	N/A
depth (m)	H		[3, 10]		N/A

^aTwo different types of shallow lakes are considered: with a sediment layer and without. These parameters are used when the sediment layer is activated.

- Depth: We use the HYDROWEB database (Crétau et al., 2011) available at <https://www.theia-land.fr> to estimate the range of uncertainty of our three categories of lakes. We estimate them as the annual lake depth amplitude. HYDROWEB provides surface water levels (and not lake depth) from satellite altimetry on large lakes and some virtual stations on the biggest rivers on the Earth. Time series of lake water levels are provided for 160 lakes, and we have used these data to estimate the maximal annual amplitude of the water level variations. By subtracting the annual minimum level to the annual maximum level, we can assess the lake depth amplitude. The results show that the annual amplitude is quite variable between lakes, varying between 20 cm and 19.5 m (after filtering the inconsistent data), due to the variability of the water balance processes. Given that our perspective in the future is to model the mass balance, we decided to take as uncertainty range, the average annual amplitude, which is equal to 5 m

- Light extinction coefficients: for free water, the database ILEC (ILEC: World Lake Database, International Lake Environment Committee Foundation, 1999) has been chosen to estimate the uncertainty of the light extinction coefficient, following the work of Layden et al. (2016). ILEC provides *in situ* measurements of the Secchi depth on 150 lakes all over the world. The Secchi depth can be related to the lake turbidity with empirical formula. We used the equation proposed by Poole and Atkins (1929) to derive the extinction coefficient. This formulation, which relates linearly the Secchi depth to the inverse of the extinction coefficient, has been derived from numerous measurements. The analysis of the data distribution allows us to estimate a single range for our study cases equal to $(0.23\text{--}2.44) m^{-1}$, which corresponds to the first and third quartiles of the distribution. As a point of comparison, the values suggested in FLake vary from $0.2 m^{-1}$ (for a very clear lake) to $2 m^{-1}$ (very turbid case) (Mironov et al., 2010). For the ice, the range of variation of the light extinction coefficient was provided by Launiainen and Cheng (1998) following the FLake documentation. The blue ice is considered as the most transparent with a value of $8.4 m^{-1}$ and the white one as the most opaque with a coefficient of $17.1 m^{-1}$. For snow, the range of variation was set to $(15\text{--}25) m^{-1}$ depending on its liquid water content with the dry snow considered as the most opaque

- Albedos: For free water and in absence of global databases, we refer to literature values and essentially to McMahon and Moore (2017). The final interval has been set to (0.025–0.175), which can be compared to the default value of 0.07 used in FLake (the value commonly used for the ocean albedo). For snow and ice, since the default values proposed by Mironov et al. (2012) (i.e., a single and common interval equal to [0.1–0.6]) was quite different from the values extracted from literature review, we decided to follow Le Moigne et al. (2016) for snow and Ebert and Curry (1993) for the ice. Therefore, the final ranges of variations were set to (0.4–0.8) for snow and (0.4–0.65) for the ice albedo
- Fetch: The fetch length is the length of water over which a given wind has blown without encountering obstacles. Over lakes, it can be estimated from the lake area assuming a given shape. We use the HydroLAKES database (Messenger et al., 2016) to estimate the fetch value for each lake referenced. Assuming a circle shape, the fetch length is derived from the circle diameter. The fetch intervals are centered on the mean diameter for each kind of lake (shallow, intermediate, and deep). Then, we add the fetch error which has been derived from an ellipse construction. In this way, the fetch can be a value between the two ellipse axes, both determined from the surface area and shoreline length for each lake, depending on the wind direction. Finally, the error is approached by the mean difference of the two axes for each lake distribution (shallow, intermediate, and deep).
- Thermocline shape factor: As already noted by Layden et al. (2016) and Salgado and Le Moigne (2010), the relaxation coefficient, which is part of the prognostic equation of the thermocline shape factor, is very uncertain. The default value is 3×10^{-3} , but values from 10^{-2} to 10^{-5} were adjusted for small to deep lakes. We kept this range of variation in our SA following Layden et al. (2016).
- Sediment parameters: when the sediment layer is activated, two other parameters are required and therefore added in the SA. They are the sediment layer depth and bottom temperature. The range of variation of the depth has been set to the interval (3–10) m (see <http://www.flake.igb-berlin.de/usefulhints.shtml>), and the bottom temperature has been derived from the annual air temperature averaged over a 15-year period (1990–2004) for our three climates following the advice of Golosov and Kirillin (2010). The range of variations were derived from the interannual variability calculated over the same time period.

2.3.4. Simulations Protocol

To generate the FLake simulations for the 12 study cases (three climates and four lake types), we have chosen to sample the parameter space with a uniform distribution using a quasi Monte Carlo algorithm based on low discrepancy Sobol series. The number of simulations have been defined following Equation 10 with N set to 1000. As a result, for the different lake cases, 11,000 (when 9 parameters are sampled) or 13,000 simulations (in the sediment cases with 11 parameters) were generated. In comparison, Rosolem et al. (2012) performed 45,000 simulations for 42 parameters and Zhang et al. (2013) generated 60,000 for 28 parameters.

Each simulation was generated on a fifteen-year period (1990–2004). Ten years of spin-up are necessary and sufficient to reach model equilibrium and get rid of incorrect initial conditions. Hence the SA has been performed on the last 5 years of the simulation (2000–2004). The variance analysis and Sobol decomposition have been performed on the three main output variables of the model: LWST, H_L , and H_S . Since the conclusions are generally the same for these three variables, we generally plot only the ones concerning LWST, the others are given in the supplementary document.

3. Results

The model parameter sensitivities are shown in the following for LWST, H_L , and H_S . The sensitivities were studied at annual, monthly and daily scales through the Generalized Sobol Indices (GSI) and the daily Sobol Indices (First order and Total order). We recall that all the Sobol indices are in the range (0, 1), that the sum among the nine or the 11 parameters (sediment case) is equal to 1 and the larger the value, the greater the sensitivity. Since these indices are relative to the total variance of the studied variable, it is important to look first to the simulated variances of our main surface variables of interest, that is, LWST, H_L , and H_S .

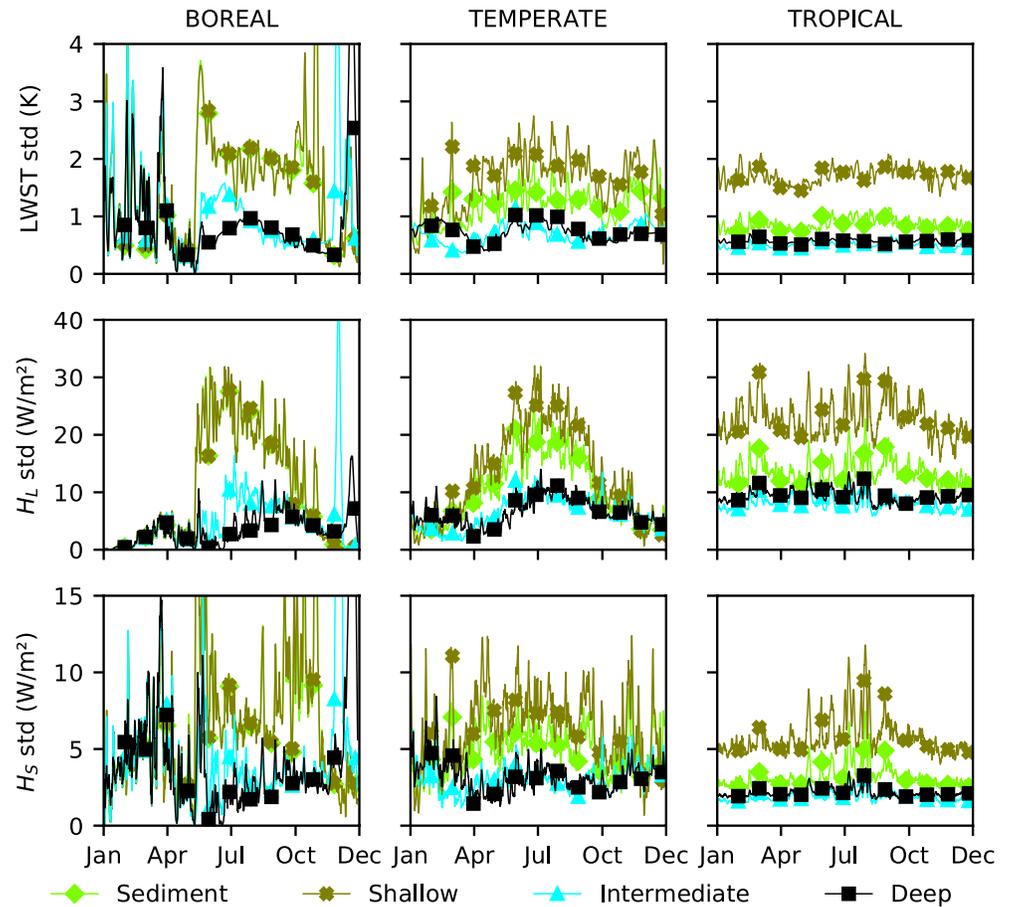


Figure 1. Total variances of LWST, H_L , and H_S simulated by our ensemble of simulations for the four studied lake types over the Year 2001. Each column corresponds to one of the three climate conditions. The time series are calculated at half-hourly time step and smoothed over a 1 day averaging window.

The experiments were conducted over a 5-year study period. However, for sake of clarity, and because the results do not differ from one year to the other, we only plot the results obtained for the year 2001 in this section.

3.1. Total Variances of Lake Surface Variables

Figure 1 shows the total variances of LWST, H_L , and H_S generated by our ensemble of simulations for our 12 representative cases. For each case, the total variances are calculated at a time step of 30 min with a smoothing average of 1 day. The figure shows clearly that the shallower the lake, the larger the variances. Indeed, LWST variances for shallow lakes are more than twice those of deep lakes and can exceed 3 K in summer boreal conditions.

For climates with a strong seasonality, for example, boreal, the variances are larger in summer. In addition, for the intermediate/deep lakes in boreal conditions, large peaks are observed in fall and winter. These peaks can be explained by the larger variability of the frozen conditions among the ensemble of simulations.

The impact of the sediment layer on LWST is larger under tropical climates and during summer in temperate regions. The addition of the sediment layer results in the variances halving. This effect is similar to the effect of increasing the lake depth.

For the simulated variances of H_L and H_S (Figure 1), the same features can be seen. However, larger values are obtained for H_L compared to H_S . This also results in a more pronounced seasonal cycle for H_L in summer for the temperate and boreal climates.

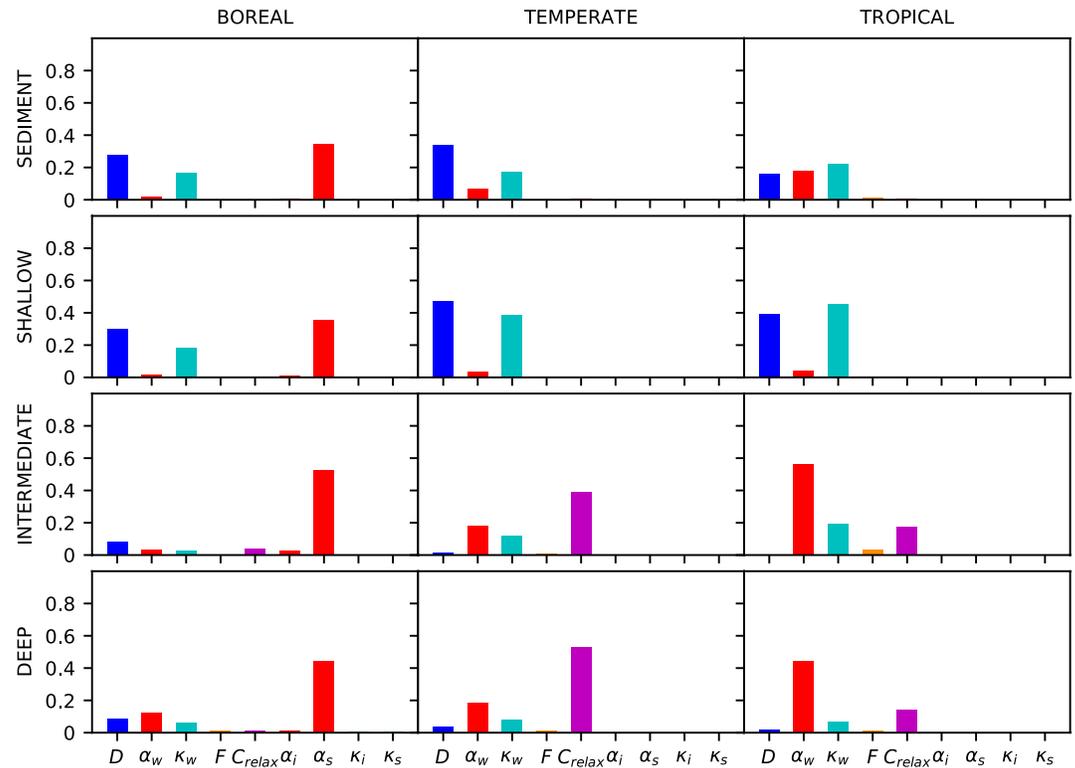


Figure 2. The first order generalized Sobol indices are presented in vertical bar graphs. Columns (respectively lines) are related to climate conditions (respective lake types). Each bar represents the parameter contribution to the LWST variance for Year 2001.

3.2. Generalized Sobol Indices

Figure 2 shows the GSI for each of the nine parameters shared by all the model configurations. The GSI are calculated for LWST at annual time scale. The relative importance of each parameter changes depending on lake type and climate.

For shallow lakes (with and without sediments), the dominant effect of lake depth and water light extinction coefficient on LWST is clear. The influence of the light extinction coefficient of water is more important for temperate and tropical climates compared to the boreal one. Under boreal conditions, the snow albedo parameter dominates because of the frozen and snowy conditions present during several months of the year. The other parameters are not influential for these shallow lakes, except for the water albedo which appears to be more sensitive when the sediment layer is activated under temperate and tropical climates.

For intermediate/deep lakes, the parameter contributions are different. The relaxation time of the thermocline shape factor and water albedo (snow albedo for the boreal climate) dominates instead of the lake depth and water light extinction coefficient. Similarly, to shallow lakes, water albedo is most important in tropical climates where solar radiation is larger.

We further calculated the same indices on H_L and H_S (supplementary document: Figures S2 and S3). The same features are observed, with the exception of the fetch parameter that appears to be slightly sensitive in the case of deep lakes.

3.3. Sobol Indices at Daily Time Step

3.3.1. First Order Indices on LWST

Figure 3 shows the temporal series of the first order SI (for LWST), calculated at a time step of 30 min and smoothed over a daily window for each of the 11 model parameters and for each of the 12 lake cases.

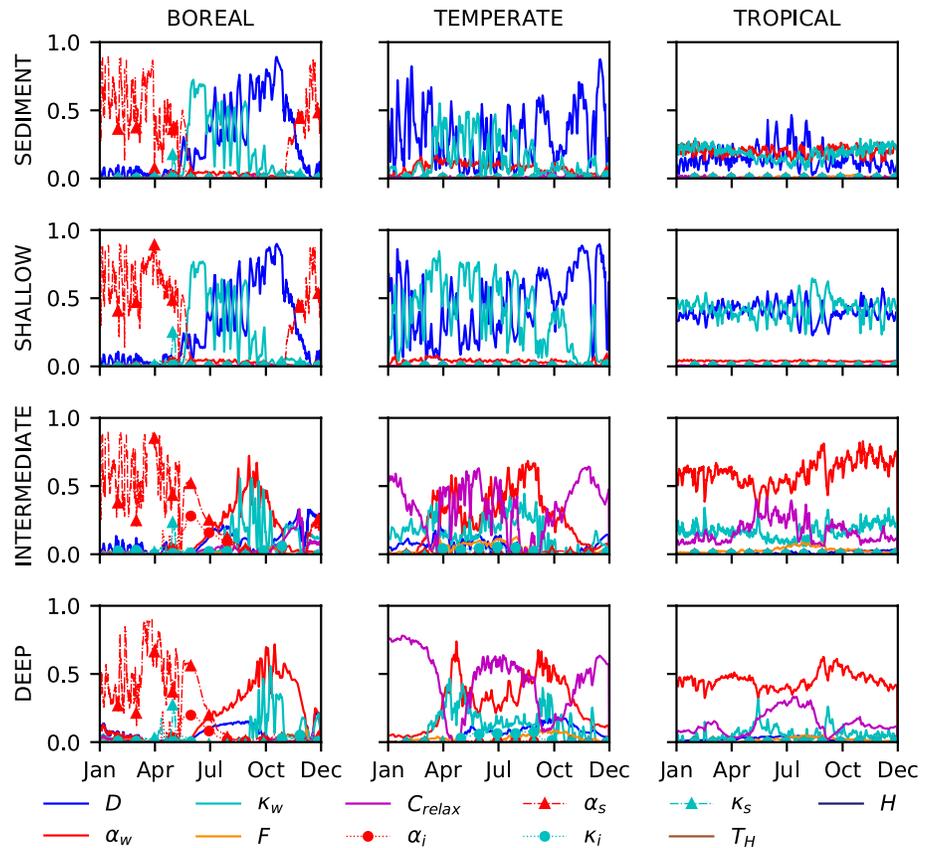


Figure 3. LWST first order Sobol indices are shown here for the 12 study cases and at a half-hourly time step (after application of a 1-day moving average). Columns (respectively lines) are related to climate conditions (respective lake types). Colors are related to model parameters, lines to free water, triangle markers to snow optical parameters and circles to the ice ones.

The noisy character of the curves is explained by the temporal variability of the meteorological conditions (cloudiness, wind speed, air temperature, and vapor pressure), which impact the total variance of LWST and the relative contribution of the lake physical processes. Despite these high frequency variations, the plots show the same main features already identified in the GSI plots (Figure 2); (i) the key role of the snow/ice radiative parameters in frozen conditions, (ii) the dominant influence of lake depth and water light extinction coefficient for shallow lakes, and (iii) the larger influence of fresh water albedo and relaxation coefficient for deeper lakes. The seasonal variations also bring additional information. An annual cycle is clearly highlighted, more important under boreal climates than tropical ones, because of the changing atmospheric conditions throughout the seasons.

For shallow lakes (both with and without sediment), the respective roles of lake depth and water light extinction coefficient appear to be complementary. During fall and winter (under unfrozen conditions), the lake depth has a larger role than water light extinction coefficient, whereas in summer, the water light extinction coefficient gains importance. In contrast, these two parameters have a more balanced influence under tropical conditions where there is no seasonal effect.

For intermediate/deep lakes, fresh water albedo and water light extinction coefficient have the same temporal evolution, which acts opposite to the evolution of the thermocline relaxation coefficient, especially under temperate and tropical climates. The temporal variation of the SI of the thermocline relaxation coefficient, which vanishes twice per year in spring and fall, is particularly remarkable. This is due to the time variation of the dominant physical processes that occur in the temperate and tropical lakes all along the year. Lakes show thermal stratification, which evolves seasonally according to the surface energy balance at the atmosphere interface because of the water density variation with temperature and

its anomaly at 4°C (larger density). The lake mixing resulting from the convection flows can occur at a number of times during the year depending mostly on lake depth and atmospheric conditions (such lakes are called holomictic). When the lake is mixed, the temperature is homogeneous, and the thermocline is non-existent. The shape factor will therefore be non-efficient and its impact null. On the contrary, when the lake is stratified (in summer or in winter for temperate lakes), the thermocline develops, and the shape factor plays a larger role.

As expected, the fresh water albedo plays a larger role when the solar radiation is larger, that is, for tropical climates and during summer for boreal and temperate climates. It seems also to play a larger role in shallow lakes when the sediment layer is activated, in temperate and tropical conditions. In fact, this feature highlights the lower role played by the water light extinction coefficient and the lake depth when the sediment processes are accounted for in shallow lakes.

The addition of a sediment layer with the energy transfers at the interface impacts the surface temperature because of the added constraint on the lake bottom temperature. As a result, the thermocline temperature is affected and the main parameters driving it (in the shallow lake case, water light extinction, and lake depth) see their influence decrease. The fresh water albedo which is the third ranked influential parameter on LWST, sees its sensitivity increasing.

In the case of frozen lakes, the ice and snow albedos are the dominant parameters because of their strong opacity, which do not allow the water below to play any role on the surface-atmosphere processes. The interplay of the ice/snow radiative parameters during the snow melting season in spring and for the intermediate/deep lakes is clearly seen on the plots and explained by the longer length of the ice melting period compared to that of shallow lakes. The lake depth is dominant during the fall season, since it greatly controls the lake capacity to freeze.

A synthetic picture of the parameter sensitivities all along the year is provided by the representation of the first order GSI calculated at monthly timescales (Figure 4). Indeed, the figure summarizes the features already observed at daily scale, that is, the dominant role of the snow/ice albedos in frozen conditions, of depth and light extinction coefficient for shallow lakes, and the larger contribution of water albedo and relaxation constant for deeper lakes. Their seasonality is clearly shown also, according to the seasonality of the dominant physical processes (see, e.g., the sensitivity drops of the relaxation constant during the turnover periods for the dimictic temperate lakes).

All these features suggest therefore some correlations between parameters, at least between lake depth and water light extinction coefficient in the case of shallow lakes that can be explored with the higher order Sobol indices.

3.3.2. Higher Order Indices on LWST

In Figure 5, the differences between the total order and first order SI are plotted for LWST. By subtracting the first order SI from the total order SI, we are left with the higher order SI, which quantify the interactions between the parameters. When two parameters dominate this difference, we can infer that the variance is influenced by the correlation between these parameters. As it can be seen on Figure 5, for shallow lakes, these differences are small (less than 0.2) and concerned only the lake depth and water light extinction coefficient. This means that a very small part of LWST variance is explained by the second order interactions between these two parameters. When the sediment layer is activated and in free water conditions, lake depth and water light extinction appear more correlated with second order effects explaining about half of the variance under temperate and tropical climates. The rest of the variance is explained by the higher order interactions between the sediment bottom temperature, the fetch, and in the case of tropical conditions, the relaxation constant (in the case of tropical conditions only). This could be the consequence of the thermal processes described previously and the impact of the sediment bottom temperature on the mixed layer temperature. The energy exchanges at the lake-sediment interface, driven partly by the imposed bottom sediment temperature, influence indirectly the weight of the other parameters influencing the depth of the mixed layer and of the thermocline like the fetch and the relaxation constant.

For deeper lakes, the correlations between the parameters become larger but are quite negligible in summer season. The differences between total and first order indices are much lower (generally less than 0.2), and

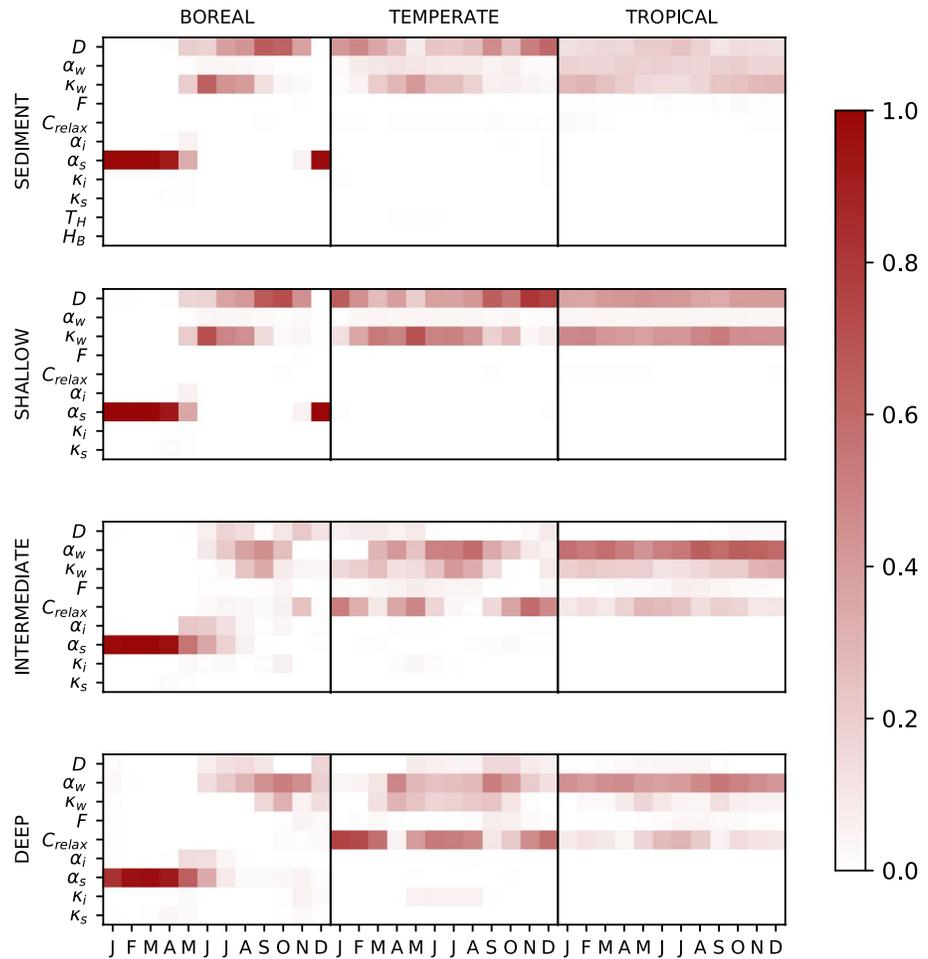


Figure 4. The first order GSI at monthly time step are shown for the 12 study cases. Each matrix column (respective line) corresponds to a month (respective parameter). Columns of layout (respective lines) are related to climate conditions (respective lake types). Color intensity is related to the parameter influence.

their behaviors suggest some correlations between water light extinction and relaxation constant on one side and between fresh water albedo and lake depth on the other side. For deep lakes, we also observe larger values in winter under boreal conditions. However, these values are simply due to the larger variability of the frozen conditions and reflect the combined impacts of lake depth, fresh water albedo, fetch and extinction coefficient on the lake stratification and frozen status.

3.3.3. Sensitivity on Surface Turbulent Fluxes

Figure 6 shows the first order Sobol indices calculated for H_L for our 12 study cases (the results for H_S are very similar and can be found in Figure S4). The same features already observed for LWST are visible here. Lake depth and water light extinction are the dominant parameters for shallow lakes in unfrozen periods, with the fresh water albedo playing a larger role in the case of the added sediment layer. In freezing conditions, the snow/ice albedos are the most influential as already noted for LWST. For deeper lakes, the fresh water albedo and shape factor relaxation coefficient are dominant. However, the fetch parameter appears more sensitive especially during the periods of lake turnover in spring and fall. This is explained by the dominant role of wind in the mixing processes during these periods. Concerning the parameter interactions, exactly the same features observed on LWST can be seen on the surface fluxes, as expected.

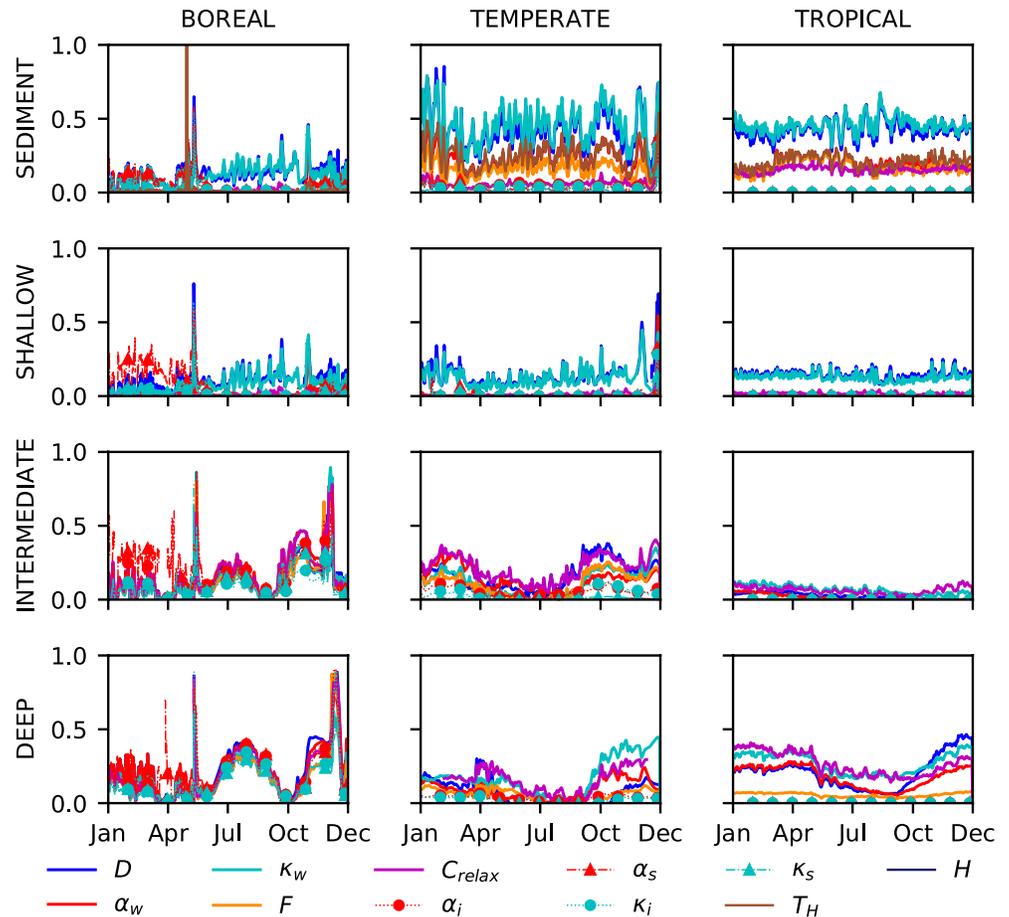


Figure 5. Differences between the LWST total and first order Sobol indices at a half hour time step (after application of a 1-day moving average) for the 12 study cases and for Year 2001. Columns (respective lines) are related to climate conditions (respective lake types). Colors are related to model parameters, lines to free water, triangle markers to snow optical parameters and circles to the ice ones.

4. Discussion

4.1. Respective Role of Model Parameters

First of all, our results confirmed the dominant influence of the lake depth and the varying influence of the other parameters on the thermal processes. Due to the seasonal and diurnal variations of the atmospheric conditions, the dominant parameters vary along the year. In frozen conditions, the snow and ice albedos were shown to play a key role on the surface temperature and fluxes as expected. Furthermore, the dominant parameters are not the same when shallow or deep lakes are considered.

For shallow lakes, in the case of free water, it has been shown that lake depth and water light extinction uncertainties explain the larger part of the variance of LWST and the surface fluxes. For intermediate/deep lakes, the total variance is more explained by the fresh water albedo and thermocline shape relaxation coefficient. The larger sensitivity of the surface fresh water albedo for deep lakes compared to shallow ones was not shown in previous studies. Although the final impact on LWST and fluxes is quite moderate (less than 1 K for LWST, 5 W/m² for H_s , and 15 W/m² for H_L , respectively), this parameter should receive some attention in the future when more measurements will be available.

The water light extinction coefficient is less influential on LWST and heat fluxes for deeper lakes. For these lakes, there is generally a deep mixed layer. In FLake, the water light extinction coefficient influences essentially the amount of solar energy that will be transferred in the thermocline below the mixing layer. If the mixing layer is deep, it will absorb all the solar energy no matter the value of the light extinction coefficient.

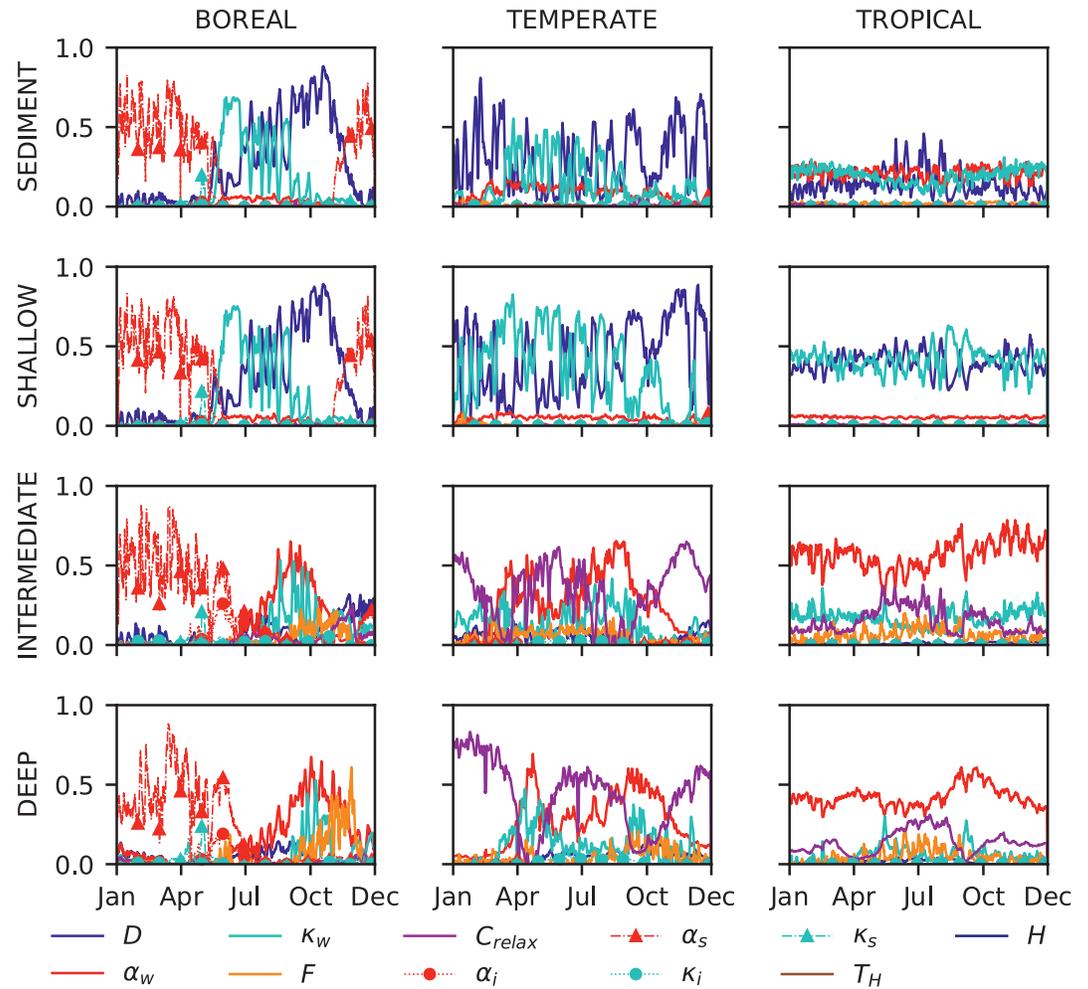


Figure 6. H_i first order Sobol indices are shown here at a half-hourly time step (after application of a 1-day smoothing average) for the 12 study cases. Columns (respective lines) are related to climate conditions (respective lake types). Colors are related to model parameters, lines to free water, triangle markers to snow optical parameters and circles to the ice ones.

These results are in line with the findings of Heiskanen et al. (2015). In their study, the same argument was given to explain why this coefficient is only influential when the lake is very clear and not turbid. This explains also the non-negligible correlations that we identified between the lake depth parameter and water light extinction coefficient, and the role of lake depth on the mixing layer thickness leads to secondary effects on the sensitivity of the water light extinction coefficient.

In this study, the model is not sensitive to the fetch, even for the deepest lakes where a larger parameter range is prescribed. Yet, the role of fetch is well-known on water mixing. This lack of sensitivity is because the effect of the fetch is not fully represented in FLake since it is a 1D model, and the large convection flows are not represented.

As with any sensitivity analysis experiment, these results are dependent on the definition of the parameter space, that is, the selected parameters and their range of variation. As an example, consider the water light extinction coefficient discussed above. If the minimum of the variation range was set to 1 m^{-1} instead of 0.2 m^{-1} , that is, excluding very clear cases, then, even in the case of shallow lakes, the SI values would be close to 0. In the same way, other internal parameters fixed constant could have been considered in this study, such as the relaxing constant for the wind mixed layer. Our strategy was to focus our analysis on the parameters that were already identified by FLake community as subject to uncertainties and

to define their range of variation according to the expected errors in a future implementation at global scale. As a consequence, our SA results are more relevant for the global application foreseen. Different applications of FLake would require dedicated SA focused on the uncertain parameters and output variables of interest.

In this work, we have also tried to assess the role of the sediment parameters (depth and bottom temperature) on LWST and surface fluxes. We showed that the influence of these two parameters is very negligible compared to the other model parameters at the first order. This is an advantage since these parameters are very uncertain. However, this result has to be taken carefully because the ranges of variation that we have prescribed may be underestimated, especially for the bottom temperature. In our work, and in absence of measurements, this temperature has been estimated from the annual air temperature following the advice of Golosov and Kirillin (2010). Such assumption, if not valid, could produce incorrect heating or cooling flux at the soil-sediment interface and lead to significant errors on the thermocline temperature and, to a lesser extent, on the surface variables.

4.2. Global Modeling Perspectives

The results obtained through this SA allowed us to answer our preliminary modeling issues. First, the analysis of lakes global features showed the spatial distribution of lakes and their main features according to spatial extent, seasonal dynamics and most importantly lake depth. From the HydroLAKES database, we have seen that deep lakes represent more than 30% of the total lake area globally and that shallow lakes with lake depth lower than 5 m represent about 30% of the distribution (this number could be however underestimated because of the limited resolution of the sensors used to map water bodies. As an example, this limit is equal to 10 ha for HydroLAKES (Messenger et al., 2016). Their spatial distribution told us that the density of shallow lakes is clearly dominant in the northern latitudes, whereas the deep lakes can be distributed all over the world. Given these different features and parameter sensitivities, it appears important to distinguish these two kinds of lakes during our implementation of FLake into ORCHIDEE.

Second, the larger sensitivity of shallow lakes to lake depth confirms also the added contribution of a continuous monitoring and in this objective, the importance to simulate the lake water budget and the relations with surface extent. In this perspective, the benefit of future altimetry missions, such as the SWOT satellite based on radar interferometry, should be of great interest to initialize and update regularly our parameterizations. SWOT will be able to provide surface extent and water levels of all the water bodies of size larger than 6 ha, at global scale and with a repetitivity of at least 21 days (Biancamaria et al., 2016). New products of lake turbidity and albedo are also coming out in the remote sensing community, such as those derived from Sentinel-2 instruments (<https://www.theia-land.fr/en/products/>), and they should be helpful to better assess the light extinction coefficient as well as the albedo and their respective ranges of variation. Data assimilation of multi-datasets (albedo, turbidity, LWST, surface water elevation, and surface extent) is therefore a promising track to calibrate and update regularly lake model parameters.

Third, this study highlighted as well, the contribution of the relaxation constant of the thermocline shape factor on the surface temperature of deep lakes. However, the dynamics of this shape factor is hard to model. This issue was already raised by different authors such as Salgado and Le Moigne (2010) who proposed a solution to limit its high frequency time variations. Another solution was proposed by Golosov et al. (2018) but more work is needed for calibration. As a first step, we will follow Salgado and Le Moigne (2010)'s approach, which is easier to implement, in order to avoid some of the unrealistic behaviors they revealed during the lake turnover periods. Further work is currently being undertaken by G. Kirillin (personal communication, 2019) to better parameterize the relaxation constant according to lake depth.

Finally, the notable impact of the sediment processes on the surface variables despite the low sensitivity of the related parameters was shown in this study. Given the large uncertainties associated with these parameters in the absence of information on the sediment characteristics, we will not activate this module in FLake in its global scale implementation.

5. Conclusion

To answer some questions raised by the foreseen implementation of the FLake lake model into ORCHIDEE, a SA of model parameters was applied to four different kinds of lakes under three different climates. The analysis was performed on the 11 parameters of FLake that will need to be prescribed at global scale (lake depth, albedo and light extinction both for fresh water, ice and snow, fetch, thermocline shape relaxation coefficient, sediment layer thickness and bottom temperature, as summarized in Table 1). The range of variation of each parameter has been carefully defined with the aim to represent the future uncertainties, keeping in mind that the lake water budget will be eventually resolved in our modeling.

With this work, we have shown the relevance of the SA methods to better understand model functioning and parameter influences. The added information brought by the analysis of SI time series has been also demonstrated here. The time varying sensitivities of the input parameters was clearly demonstrated, from the dominance of the snow/ice ones in frozen conditions, to the relative contribution of the other ones for free water according to lake depth, that is, larger sensitivity of lake depth and light extinction coefficient for shallow lakes whereas water albedo and thermocline relaxation constant are more influential for deeper lakes.

The analysis was limited to LWST and surface turbulent fluxes because these are the main variables seen by the atmosphere in a climate model. It has been fully informative to better understand the functioning of FLake model, its parameters and interactions. It allows us to start its implementation in ORCHIDEE with confidence and to identify the databases that will be used for its calibration. The altimetry and most recent multispectral space missions will be given first priority for that purpose.

Data Availability Statement

The Flake model used in this study was downloaded from <http://www.flake.igb-berlin.de/site/download>. The codes used to generate the results of the sensitivity analysis can be downloaded from <https://doi.org/10.5281/zenodo.3700135>

Acknowledgments

This work was essentially supported by the French National Space Agency (Center National d'Etudes Spatiales) through the TOSCA-SPAWET program in preparation to the SWOT space mission. CEA (Commissariat à l'Energie Atomique et aux énergies alternatives) is also acknowledged for funding part of A. Bernus PhD grant as well as the H2020 MULTIPLY project (Grant No. 687320) funding N. Raoult. The authors would like to thank Karine Pétrus for her preliminary works with FLake model in our team and Fabienne Maignan for her precious advices on coding issues. The authors are also grateful to Dr. D. Mironov and G. Kirillin for valuable comments and helpful discussions. The reviewers are also acknowledged for their constructive comments.

References

- Balsamo, G. (2010). Deriving an effective lake depth from satellite lake surface temperature data: A feasibility study with MODIS data. *Boreal Environment Research*, 15(2), 178–190.
- Balsamo, G., Salgado, R., Dutra, E., Boussetta, S., Stockdale, T., & Potes, M. (2012). On the contribution of lakes in predicting near-surface temperature in a global weather forecasting model. *Tellus A: Dynamic Meteorology and Oceanography*, 64(1), 15829. <https://doi.org/10.3402/tellusa.v64i0.15829>
- Bastviken, D., Cole, J., Pace, M., & Tranvik, L. (2004). Methane emissions from lakes: Dependence of lake characteristics, two regional assessments, and a global estimate: Lake Methane Emissions. *Global Biogeochemical Cycles*, 18(4). <https://doi.org/10.1029/2004gb002238>
- Biancamaria, S., Lettenmaier, D. P., & Pavelsky, T. M. (2016). The SWOT mission and its capabilities for land hydrology. *Surveys in Geophysics*, 37(2), 307–337. <https://doi.org/10.1007/s10712-015-9346-y>
- Bonan, G. B. (1995). Sensitivity of a GCM simulation to inclusion of inland water surfaces. *Journal of Climate*, 8(11), 2691–2704. [https://doi.org/10.1175/1520-0442\(1995\)008<2691:SOAGST>2.0.CO;2](https://doi.org/10.1175/1520-0442(1995)008<2691:SOAGST>2.0.CO;2)
- Chuang, H.-Y., & Sousounis, P. J. (2003). The impact of the prevailing synoptic situation on the lake-aggregate effect. *Monthly Weather Review*, 131(5), 990–1010. [https://doi.org/10.1175/1520-0493\(2003\)131<0990:TOTPS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<0990:TOTPS>2.0.CO;2)
- Crétau, J.-F., Arsen, A., Calmant, S., Kouraev, A., Vulginski, V., Bergé-Nguyen, M., et al. (2011). SOLS: A lake database to monitor in the near real time water level and storage variations from remote sensing data. *Advances in Space Research*, 47(9), 1497–1507. <https://doi.org/10.1016/j.asr.2011.01.004>
- d'Orgeval, T., Polcher, J., & de Rosnay, P. (2008). Sensitivity of the West African hydrological cycle in ORCHIDEE to infiltration processes. *Hydrology and Earth System Sciences*, 12(6), 1387–1401. <https://doi.org/10.5194/hess-12-1387-2008>
- Dutra, E., Stepanenko, V. M., Balsamo, G., Viterbo, P., Miranda, P., Mironov, D., & Schar, C. (2010). An offline study of the impact of lakes on the performance of the ECMWF surface scheme. *Boreal Environment Research*, 15(2), 100–112.
- Ebert, E. E., & Curry, J. A. (1993). An intermediate one-dimensional thermodynamic sea ice model for investigating ice-atmosphere interactions. *Journal of Geophysical Research*, 98(C6), 10085. <https://doi.org/10.1029/93jc00656>
- Saltelli, A. (Ed.). (2008). *Global sensitivity analysis: The primer* (OCLC: ocn180852094). John Wiley.
- Golosov, S., & Kirillin, G. (2010). A parameterized model of heat storage by lake sediments. *Environmental Modelling & Software*, 25(6), 793–801. <https://doi.org/10.1016/j.envsoft.2010.01.002>
- Golosov, S., Zverev, I., Shipunova, E., & Terzhevik, A. (2018). Modified parameterization of the vertical water temperature profile in the FLake model. *Tellus A: Dynamic Meteorology and Oceanography*, 70(1), 1–7. <https://doi.org/10.1080/16000870.2018.1441247>
- Guerrero, J.-L., Pernica, P., Wheeler, H., Mackay, M., & Spence, C. (2017). Parameter sensitivity analysis of a 1-D cold region lake model for land-surface schemes. *Hydrology and Earth System Sciences*, 21(12), 6345–6362. <https://doi.org/10.5194/hess-21-6345-2017>
- Heiskanen, J. J., Mammarella, I., Ojala, A., Stepanenko, V., Erkkilä, K. M., Miettinen, H., et al. (2015). Effects of water clarity on lake stratification and lake-atmosphere heat exchange. *Journal of Geophysical Research - D: Atmospheres*, 120(15), 7412–7428. <https://doi.org/10.1002/2014JD022938>

- ILEC. (1999). *World Lake database*. International Lake Environment Committee Foundation. Retrieved from <http://wldb.ilec.or.jp>
- Laird, N., Sobash, R., & Hodas, N. (2009). The frequency and characteristics of lake-effect precipitation events associated with the New York State Finger Lakes. *Journal of Applied Meteorology and Climatology*, 48(4), 873–886. <https://doi.org/10.1175/2008JAMC2054.1>
- Lamboni, M., Monod, H., & Makowski, D. (2011). Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. *Reliability Engineering & System Safety*, 96(4), 450–459. <https://doi.org/10.1016/j.res.2010.12.002>
- Launiainen, J., & Cheng, B. (1998). Modelling of ice thermodynamics in natural water bodies. *Cold Regions Science and Technology*, 27(3), 153–178. [https://doi.org/10.1016/S0165-232X\(98\)00009-3](https://doi.org/10.1016/S0165-232X(98)00009-3)
- Layden, A., MacCallum, S. N., & Merchant, C. J. (2016). Determining lake surface water temperatures worldwide using a tuned one-dimensional lake model (FLake, v1). *Geoscientific Model Development*, 9(6), 2167–2189. <https://doi.org/10.5194/gmd-9-2167-2016>
- Le Moigne, P., Colin, J., & Decharme, B. (2016). Impact of lake surface temperatures simulated by the FLake scheme in the CNRM-CM5 climate model. *Tellus A: Dynamic Meteorology and Oceanography*, 68(1), 31274. <https://doi.org/10.3402/tellusa.v68.31274>
- Li, J., Duan, Q. Y., Gong, W., Ye, A., Dai, Y., Miao, C., et al. (2013). Assessing parameter importance of the Common Land Model based on qualitative and quantitative sensitivity analysis. *Hydrology and Earth System Sciences*, 17(8), 3279–3293. <https://doi.org/10.5194/hess-17-3279-2013>
- Ljungemyr, P., Gustafsson, N., & Omstedt, A. (1996). Parameterization of lake thermodynamics in a high-resolution weather forecasting model. *Tellus*, 48(5), 608–621. <https://doi.org/10.1034/j.1600-0870.1996.t01-4-00002.x>
- Martynov, A., Sushama, L., & Laprise, R. (2010). Simulation of temperate freezing lakes by one-dimensional lake models: Performance assessment for interactive coupling with regional climate models. *Boreal Environment Research*, 15(1), 143–164.
- McMahon, A., & Moore, R. D. (2017). Influence of turbidity and aeration on the albedo of mountain streams. *Hydrological Processes*, 31(25), 4477–4491. <https://doi.org/10.1002/hyp.11370>
- Messenger, M. L., Lehner, B., Grill, G., Nedeva, I., & Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications*, 7(1). <https://doi.org/10.1038/ncomms13603>
- Mironov, D. (2008). Parameterization of lakes in numerical weather prediction. In Description of a lake model (Technical Report No. 11). Deutscher Wetterdienst.
- Mironov, D., Heise, E., Kourzeneva, E., Ritter, B., Schneider, N., & Terzhevik, A. (2010). Implementation of the lake parameterization scheme Flake into the numerical weather prediction model COSMOS. *Boreal Environment Research Publishing Board*, 15, 218–230.
- Mironov, D., Ritter, B., Schulz, J.-P., Buchhold, M., Lange, M., & MacHulska, E. (2012). Parameterization of sea and lake ice in numerical weather prediction models of the German weather service. *Tellus A: Dynamic Meteorology and Oceanography*, 64(1), 17330. <https://doi.org/10.3402/tellusa.v64i0.17330>
- Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, 540(7633), 418–422. <https://doi.org/10.1038/nature20584>
- Poole, H. H., & Atkins, W. R. G. (1929). Photo-electric measurements of submarine illumination throughout the year. *Journal of the Marine Biological Association of the United Kingdom*, 16(1), 297–324. <https://doi.org/10.1017/s0025315400029829>
- Rooney, G. G., & Bornemann, F. J. (2013). The performance of FLake in the met office unified model. *Tellus A: Dynamic Meteorology and Oceanography*, 65(1), 21363. <https://doi.org/10.3402/tellusa.v65i0.21363>
- Rosolem, R., Gupta, H. V., Shuttleworth, W. J., Zeng, X., & de Gonçalves, L. G. G. (2012). A fully multiple-criteria implementation of the Sobol' method for parameter sensitivity analysis. *Journal of Geophysical Research*, 117(D7). <https://doi.org/10.1029/2011JD016355>
- Salgado, R., & Le Moigne, P. (2010). Coupling of the Flake model to the surfex externalized surface model. *Boreal Environ Reseach Publishing Board*, 15, 231–244.
- Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., & Tarantola, S. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2), 259–270. <https://doi.org/10.1016/j.cpc.2009.09.018>
- Sandström, G. (2003). Impact of lakes and wetlands on boreal climate. *Journal of Geophysical Research*, 108(D16). <https://doi.org/10.1029/2002JD002597>
- Sobol, I. (1990). Quasi-Monte Carlo methods. *Progress in Nuclear Energy*, 24(1–3), 55–61. [https://doi.org/10.1016/0149-1970\(90\)90022-W](https://doi.org/10.1016/0149-1970(90)90022-W)
- Song, Y., Semazzi, F. H. M., Xie, L., & Ogallo, L. J. (2004). A coupled regional climate model for the Lake Victoria basin of East Africa. *International Journal of Climatology*, 24(1), 57–75. <https://doi.org/10.1002/joc.983>
- Stepanenko, V. (2005). Numerical simulation of the thermal regime of shallow water bodies. *Computational Technologies*, 10(1), 100–106.
- Tang, Y., Reed, P., Wagener, T., & van Werkhoven, K. (2007). Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. *Hydrology and Earth System Sciences*, 11(2), 793–817. <https://doi.org/10.5194/hess-11-793-2007>
- Telmer, K. H., & Costa, M. P. F. (2007). SAR-based estimates of the size distribution of lakes in Brazil and Canada: A tool for investigating carbon in lakes. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 17(3), 289–304. <https://doi.org/10.1002/aqc.854>
- Thiery, W., Davin, E. L., Panitz, H.-J., Demuzere, M., Lhermitte, S., & van Lipzig, N. (2015). The impact of the African great lakes on the regional climate. *Journal of Climate*, 28(10), 4061–4085. <https://doi.org/10.1175/JCLI-D-14-00565.1>
- Verpoorter, C., Kutser, T., Seekell, D. A., & Tranvik, L. J. (2014). A global inventory of lakes based on high-resolution satellite imagery. *Geophysical Research Letters*, 41(18), 6396–6402. <https://doi.org/10.1002/2014GL060641>
- Voros, M., Istvnovics, V., & Weidinger, T. (2010). Applicability of the Flake model to lake Balaton. *Boreal Environment Research*, 15(2), 245–254.
- Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., & Viterbo, P. (2014). The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. *Water Resources Research*, 50(9), 7505–7514. <https://doi.org/10.1002/2014WR015638>
- West, W. E., Creamer, K. P., & Jones, S. E. (2016). Productivity and depth regulate lake contributions to atmospheric methane. *Limnology & Oceanography*, 61(S1), S51–S61. <https://doi.org/10.1002/lno.10247>
- Woolway, R. I., & Merchant, C. J. (2019). Worldwide alteration of lake mixing regimes in response to climate change. *Nature Geoscience*, 12(4), 271–276. <https://doi.org/10.1038/s41561-019-0322-x>
- Xu, J., Shu, H., Jiang, H., & Dong, L. (2012). Sobol sensitivity analysis of parameters in the common land model for simulation of water and energy fluxes. *Earth Science Informatics*, 5(3–4), 167–179. <https://doi.org/10.1007/s12145-012-0105-z>
- Zhang, C., Chu, J., & Fu, G. (2013). Sobol's sensitivity analysis for a distributed hydrological model of Yichun River basin, China. *Journal of Hydrology*, 480, 58–68. <https://doi.org/10.1016/j.jhydrol.2012.12.005>
- Zheng, H., Yang, Z. L., Lin, P., Wei, J., Wu, W. Y., Li, L., et al. (2019). On the sensitivity of the precipitation partitioning into evapotranspiration and runoff in land surface parameterizations. *Water Resources Research*, 55(1), 95–111. <https://doi.org/10.1029/2017wr022236>