Response Letter for:

"A framework for ensemble modelling of climate change impacts on lakes worldwide: the ISIMIP Lake Sector"

Submitted to:
Geoscientific Model Development

Malgorzata Golub[1], Wim Thiery[2], Rafael Marcé[3,4], Don Pierson[1], Inne Vanderkelen[2], Daniel Mercado-Bettin[3,4], R. Iestyn Woolway[5], Luke Grant[2], Eleanor Jennings[6], Benjamin Kraemer[7], Jacob Schewe[8], Fang Zhao[8,9], Katja Frieler[8], Matthias Mengel[8], Vasiliy Yu. Bogomolov[10,11,12], Damien Bouffard[13], Marianne Côté[14], Raoul-Marie Couture[14,15], Andrey V. Debolskiy[16,12], Bram Droppers[17], Gideon Gal[18], Mingyang Guo[19], Annette B.G. Janssen[17], Georgiy Kirillin[20], Robert Ladwig[21], Madeline Magee[22], Tadhg Moore[23,24], Marjorie Perroud[25], Sebastiano Piccolroaz[26,27], Love Raaman Vinnaa[13], Martin Schmid[13], Tom Shatwell[28,20], Victor M. Stepanenko[16], Zeli Tan[29], Bronwyn Woodward[30], Huaxia Yao[31], Rita Adrian[20,32], Mathew Allan[33,34], Orlane Anneville[35], Lauri Arvola[36], Karen Atkins[37], Leon Boegman[38], Cayelan Carey[24], Kyle Christianson[39], Elvira de Eyto[40], Curtis DeGasperi[41], Maria Grechushnikova[16], Josef Hejzlar[42], Klaus Joehnk[43], Ian D. Jones[44], Alo Laas[45], Eleanor B Mackay[46], Ivan Mammarella[47], Hampus Markensten[48], Chris McBride[34], Deniz Özkundakci[34], Miguel Potes[49,50], Karsten Rinke[28], Dale Robertson[51], James A. Rusak[31], Rui Salgado[52], Leon van der Linden[53], Piet Verburg[54], Danielle Wain[55], Nicole K Ward[24], Sabine Wollrab[20], Galina Zdorovennova[56]

[1] Uppsala University, Department of Ecology and Genetics, Uppsala, Sweden
[2] Vrije Universiteit Brussel, Department of Hydrology and Hydraulic Engineering, Brussels, Belgium
[3] Catalan Institute for Water Research (ICRA), Girona, Spain
[4] University of Girona, Girona, Spain
[5] University of Reading, Department of Meteorology, Reading, UK

[6] Dundalk Institute of Technology, Centre for Freshwater and Environmental Studies, Dundalk, Ireland
[7]University of Konstanz, Limnological Institute, Konstanz, Germany
[8] Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Potsdam, Germany
[9] Key Laboratory of Geographic Information Science (Ministry of Education), School of Geographic Sciences, East China Normal University, Shanghai, China.
[10] Institute of Monitoring of Climatic and Ecological Systems of the Siberian Branch of the Russian Academy of Sciences, Tomsk, Russia
[11] Tomsk State University, Tomsk, Russia
[12] Obukhov Institute for Atmospheric Physics, Russian Academy of Science, Moscow, Russia
[13] Eawag: Swiss Federal Institute of Aquatic Science and Technology, Surface Waters - Research and Management
[14] Université Laval, Department of chemistry, Quebec, Canada
[15] Center for Northern Study, Quebec, Canada
[16] Lomonosov Moscow State University, Moscow, Russia
[17] Wageningen University & Research, Water Systems and Global Change Group, Wageningen, the Netherlands
[18] Kinneret Limnological Laboratory, Israel Oceanographic & Limnological Research, Migdal, Israel
[19] Purdue University, Department of Earth, Atmospheric and Planetary Sciences, West Lafayette, IN, USA
[20] Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany
[21] University of Wisconsin-Madison, Center for Limnology, Madison, WI, USA
[22] Wisconsin Department of Natural Resources, Madison, WI, USA
[23] Dundalk Institute of Technology, Department of Applied Sciences, Dundalk, Ireland
[24] Virginia Tech, Department of Biological Sciences & Forest Resources & Environmental Conservation, Blacksburg, Virginia, USA
[25] University of Geneva, Institute for Environmental Sciences, Geneva, Switzerland
[26] University of Trento, Department of Civil, Environmental and Mechanical Engineering, Trento, Italy
[27] École Polytechnique Fédérale de Laussane, Physics of Aquatic Systems Laboratory (APHYS)-Margaretha Kamprad Chair, Lausanne, Switzerland
[28] Helmholtz Centre for Environmental Research - UFZ, Department Lake Research, Magdeburg, Germany
[29] Pacific Northwest National Laboratory, Richland, WA, USA
[30]The University of Western Australia: Perth, WA, AU
[31] Dorset Environmental Science Centre, Ontario Ministry of Environment, Conservation and Parks, Dorset, ON, Canada
[32] Freie Universität Berlin; Berlin, Germany
[33] Waikato Regional Council, Hamilton, New Zealand
[34] University of Waikato, Environmental Research Institute, Hamilton, New Zealand
[35] INRAE, Université Savoie Mont-Blanc, UMR CARRTEL, France
[36] Lammi Biological Station, University of Helsinki, Finland
[37] Tahoe Environmental Research Center, University of California, Davis, Davis, USA

[38] Queen's University, Department of Civil Engineering, Kingston, ON, Canada
[39] University of Colorado, Institute of Arctic and Alpine Research, Boulder, CO, USA
[40] Marine Institute, Furnace, Newport, Co. Mayo, F28X252, Ireland.
[41] King County Department of Natural Resources and Parks, Seattle, WA, USA
[42] Biology Centre of the Czech Academy of Sciences, Institute of Hydrobiology, České Budějovice, Czech Republic
[43] CSIRO Land and Water, Canberra, Australia
[44] University of Stirling, Biological and Environmental Sciences, Stirling, UK
[45] Estonian University of Life Sciences, Chair of Hydrobiology and Fishery, Institute of Agricultural and Environmental Sciences, Tartu, Estonia
[46] UK Centre for Ecology & Hydrology, Lake Ecosystems Group, Bailrigg, Lancaster, LA1 4AP, United Kingdom
[47] University of Helsinki, Institute for Atmospheric and Earth System Research / Physics, Helsinki, Finland
[48] Swedish University of Agricultural Sciences, Department of Aquatic Sciences and Assessment, Sweden
[49] Universidade de Évora, Instituto de Ciências da Terra – ICT (Polo de Évora), Instituto de Investigação e Formação Avançada (IIFA), Évora, Portugal
[50] Universidade de Évora, Earth Remote Sensing Laboratory (EaRSLab), Instituto de Investigação e Formação Avançada (IIFA), Évora, Portugal
[51] U.S. Geological Survey, Upper Midwest Water Science Center, Madison, WI, USA
[52] Universidade de Évora, Instituto de Ciências da Terra, Departamento de Física, Escola de Ciências e Tecnologia, Évora, Portugal
[53] South Australian Water Corporation, Adelaide, Australia
[54] National Institute of Water and Atmospheric Research, Hamilton, New Zealand
[55] 7 Lakes Alliance, Belgrade Lakes, Maine, USA 04901
[56] Northern water problems Institute Karelian Research Centre of RAS, Petrozavodsk, Russia

Correspondence to: Wim Thiery (wim.thiery@vub.be)

Abstract

We would like to thank all reviewers for dedicating their time to improving this manuscript. We have applied substantial changes to our manuscript following the reviewers' feedback.

The response letter contains references to modifications in the manuscript as suggested by the reviewers. Page (P) and line (L) references are written short-hand as PiLj to mark the start of the location where text changes have been implemented. For quotations, the following convention denotes modifications in the original manuscript: new text. Quoted reviewer comments are written in *italics* and quotations for presenting text containing modifications are indented.

Reviewer 1 – Peter Düben

R1C1

In general, can you maybe also comment on the computational cost and the size of the datasets somewhere? If those simulations are cheap, you should also state this somewhere.

Thank you for your time to review our manuscript. We have added this relevant information as follows to P27L16, Section 3.8, "Output data format", to give examples on the storage requirements and compute times for providing global simulations under this protocol.

> This diversity of GCM input datasets, emissions scenarios, lake models and their output variables means that the total ensemble of impact simulations under the Lake Sector requires considerable storage space, and that appreciable computing resources should be anticipated by potential future collaborators. For example, the global lake simulations for ISIMIP2b take up 14TB of storage space. This means that applications with simulations under multiple GCMs, lake models and scenarios for a given variable will require high-performance computing resources. For running simulations, computing times may vary depending on the scale of one's contribution. On the one hand, simulating a local, calibrated lake with FLake for a single scenario and GCM combination may take seconds on a laptop, but, on the other hand, global simulations from CLM4.5 for one such scenario and GCM combination will require several weeks using 144 compute cores on a high-performance computer, substantiating both computational costs and resources for dataset storage. These technical prerequisites, in addition to individual model feasibility issues for local versus global domain simulations, explain the discrepancy in model availability across the ISIMIP2 local and global simulations.

R1C2

It is a bit difficult to understand whether the paper is summarizing results for ISIMIP 1, 2, 3 or all from the abstract. This should be stated explicitly, early in the paper.

We have adapted the abstract as follows to describe this explicitly in our objective ("Here we…") statement at P4L1.

> Here, we describe a simulation protocol developed by the Lake Sector of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) for simulating climate change impacts on lakes using an ensemble of lake models and climate change scenarios for ISIMIP phases 2 and 3.

Likewise, P6L25 in the introduction clarifies that this paper describes both phases 2 and 3.

> Here, we describe the protocol for the global- and local-scale intercomparison of lake model simulations completed for the second phase of ISIMIP (ISIMIP2), as well as the extensions to this protocol that have been implemented for the new phase three simulations (ISIMIP3).

R1C3

P8L5: "with the aim of conducting innovative science" should be rephrased.

We have removed this phrase. The sentence (P8L17) now reads as follows.

> Second, the Global Ecological Observatory Network (GLEON, …) started in 2005, with the aim of sharing and interpreting lake data to understand, predict, and communicate the role and responses of lakes in a changing global environment.

R1C4

P8L14: "were simulated for existing lakes in the local domain and "representative" lakes" is rather cryptic and should be rephrased.

We have adapted P9L1 with more text and references to the sections which elaborate on these differences.

> Climate change impacts were simulated after model calibration for specific lakes in the local domain (see Section 3.1), and for "representative" lakes without calibration in the global domain for each lake-containing grid cell in the ISIMIP global grid (see Section 3.2).

R1C5

P8L20: "for this first phase of lake sector simulations" I thought this would be about ISIMIP 2 and 3?

Thank you, we have adapted P9L8 as follows.

> This assumption allowed us to evaluate lake thermal structure based on meteorological forcing data only and was judged acceptable for the existing phases of ISIMIP.

R1C6

P9L18: "is expanded", can you be more specific?

The final number of local lakes is not yet known as we are currently still accepting new lakes for the ISIMIP3 round. We have adapted P10L11 to state that "the number of lakes has increased".

R1C7

P9L24: 0.5/0.5: You are using different ways to write ½ degree resolution throughout the paper. Maybe you can state once that you are working on a lon/lat grid with ½ degree grid-spacing and then remove the information from the rest of the paper?

We have reduced the heterogeneity in describing the ISIMIP spatial resolution and have removed unnecessary repetitions of this information. The ISIMIP spatial resolution is now mentioned appropriately in the following sections.

In the abstract (P4L4):

> The protocol prescribes lake simulations driven by climate forcing from gridded observations and different Earth system models under various representative greenhouse gas concentration pathways (RCP), all consistently bias-corrected on a 0.5° x 0.5° global grid.

Initially when describing representative lakes in the global domain (P10L16; Section 3.2):

> Lake simulations in the global domain considered a single, lake in each grid cell that contains lakes in the 0.5° x 0.5° ISIMIP global grid.

This also involved removing repetitions of “0.5° grid cells/pixels” throughout the text. We rather refer to the 0.5° x 0.5° spatial resolution of the ISIMIP global grid as “the ISIMIP grid scale”.

R1C8

P10L22: “for which the original data was remapped from 30’’ to match” This needs more information.

We have fully substituted this paragraph describing the representative lakes in ISIMIP3. As well, much of the detailed information behind lake mapping at this stage has been moved to Section 3.5.1 “Bathymetry”.

R1C9

P13L13: “the lake layers can be defined according to water volume” No sure what this means.

We have expanded on this description for clarity (P14L4).

> MyLake runs at a daily time step using regularly spaced water layers whose vertical resolution is defined by the user. Different versions of the open-source code have been applied to simulate algal blooms (Moe et al. 2016), $CO_{2(g)}$ and $CH_{4(g)}$ (Kiuru et al. 2019), internal phosphorus loads (Markelov et al. 2019) and light attenuation dynamics (Pilla and Couture, 2021).

R1C10

P13L20: “suite of Swiss lakes” Can you be more specific? What kind of predictions, how many lakes?

We have updated P14L14 as follows.

> Simstrat […] and is operationally applied to provide near-real time, open access simulation output of the thermal structure and ice cover of all natural Swiss lakes and lake basins greater than 1 km^2 and a growing number of reservoirs and small lakes (Gaudard et al., 2019).

R1C11

P14L15: Simstrat (v2.1.2) was described above, not Simstrat v1.4, or am I missing something?

We have clarified this entry (P15L16) for the earlier version of Simstrat used in the global simulations.

> **Simstrat-UoG v1** is based on Simstrat v1.4 and is therefore an earlier version of the model described above. This version uses an earlier snow and ice formulation from Patterson and Hamblin (1988).

R1C12

P20L20: "^-0424" is not a very nice format.

The equation now reads as follows.

$$K_d = 1.1925*\max(mean\_depth,1)^{-0.424}$$

R1C13

P24L17: "robust fit for all eight local lake models was found" What does this mean?

P28L12 has been adjusted for clarity.

> Based on the simulation data from 62 lakes, all eight local lake models were calibrated with a multi-model mean RMSE of 1.50°C that ranged from 0.98°C (air2water6par) to 2.41°C (FLake, Table 3).

R1C14

P25L16: "calibrated parameters c" This should be rephrased.

This sentence has been removed as its paragraph's content was found to be redundant and is otherwise communicated in Section 4.3.1, "Model response to observational vs simulated forcing data".

R1C15

P29L24: "the first stage of simulations undertaken by the ISIMIP Lake Sector" But this was about ISIMIP 2 and not 1?

P36L8 has been adapted for clarity. Note that at the time of ISIMIP1 ('fast track'), the ISIMIP lake sector did not yet exist.

> Here, we have described the protocol of the Lake Sector in ISIMIP2 and ISIMIP3, which includes the simplifying assumption that hydrologic inputs from the lake watershed had minimal effects on the simulated thermal structure.

R1C16

P29L27: "it will clearly not be the case for" should be rephrased

P36L10 has been rephrased as follows.

> While this is a reasonable assumption for lake hydrodynamic simulations, it will not be sufficient for simulations of lake biogeochemistry and ecology that strongly depend on the nutrient inputs from the lake watershed.

R1C17

P30L2: "Within the ISIMIP framework, the simulated climate change impacts are (inter-) comparable with results from 13 other sectors, supporting cross-sectoral aggregation of impacts (Vanderkelen et al., 2020)." I do not understand this sentence.

P36L12 has been reworked.

> Under the ISIMIP framework's provision of consistent climate forcing datasets and scenarios, the climate change impacts simulated in the Lake Sector are comparable with simulation results from other ISIMIP sectors, supporting cross-sectoral assessments of climate change impacts (Lange et al., 2020; Vanderkelen et al., 2020; Thiery et al., 2021).

R1C18

Table 1: "L" and "G" should be defined in the caption.

The footnotes describing these terms have been moved to the caption of this table. The caption now reads as follows.

Table 1. **Overview on lake impact models participating in the Lake Sector of ISIMIP2a/b. For “Spatial Domain”, L defines local or site-specific and G defines global simulations.**

R1C19

Table 2: What does the dot in brackets stand for? Maybe?

The brackets indeed stand for optional input variables. We have clarified this in the table caption.

R1C20

Table 3: “Calibrated parameter name” should be rephrased. Maybe “Names of model parameters that need to be calibrated”

Thank you, we have adapted this column header as recommended but substituted “need to” for “can”.

R1C21

Table 3: “Cross-site calibrated parameters summary statistics” Can you explain exactly in the caption what the two numbers are standing for and what the numbers in the brackets are?

The revised caption now states that “Summary statistics for calibrated parameters and models’ performance represent the cross-site mean, median [minimum-maximum].”

R1C22

Table 4: “Highlighted” I guess this refers to the grey colour?

We have clarified this as follows.

Highlighted columns (grey) represent variables available for at least half of the lake models.

R1C23

Figure 1 A: Why is there more than one colour for the dots?

The markers for locating lakes in the local domain simulations are visualized with semi-transparent markers. We have clarified this in the caption of the figure:

Figure 1: Map of lakes at local (A) and global (B) scales participating in the ISIMIP2a/b Lake Sector. In panel a, the local lake sites are visualised through semi-transparent markers, hence darker markers highlight locations where several lakes are located close to each other.

R1C24

Figure 2: The panels are too small, and A, B and C are not visible.

Thank you for this suggestion, we have updated this figure.

R1C25

Figure 3: I do not understand the sub-plot on the bottom left.

The figure and caption have been adapted as follows to show the updated mapping in ISIMIP3. As this no longer includes an inset describing reservoir area expansion, this is no longer unclear.

Figure 3. Maps of lake area (a), lake volume (b) and mean lake depth (c) used in the ISIMIP3 simulation round. In the input data for ISIMIP3 simulations, a single lake is assigned to each grid cell. However, here we show a modified version of the dataset to delineate large lakes in the global map. The datasets are derived from HydroLAKES (Messager et al., 2018) and GLOBathy (Khazaei et al., 2022) datasets using the ISIMIP3 lake mapping methods described in “Code and data availability”.

Reviewer 2 – Bertrand Guenet

R2C1

First, I found a bit misleading to mix results from ISIMIP2 and the new protocol of ISIMIP3. To improve this, I propose first to change the title of the paper into something like "The ISIMIP Lake Sector: an ensemble modelling of climate change impacts on lakes worldwide. Results from the last simulations and framework for the next step". I also propose to better separate in particular in section 3 what is coming from ISIMIP2 and what will be done in ISIMIP3. Maybe a dedicated section for ISIMIP 3 will help the reader to make the difference between what is done what will be done. Another option could be to split the paper into 2 parts to separate both but I let the decision to the editor.

Thank you for your much-appreciated review of our manuscript.

In response to this suggestion, we modified the manuscript on several locations.

First, our objective statement in the introduction is expanded on starting at P6L25 by better describing our intention to review the existing phase two protocol alongside how this protocol has progressed for phase 3 on a topic-by-topic basis.

> Here, we describe the protocol for the global- and local-scale intercomparison of lake model simulations completed for the second phase of ISIMIP (ISIMIP2), and the extensions to this protocol that have been implemented for the new phase three simulations (ISIMIP3). The evolution of the modelling protocol from ISIMIP2 to ISIMIP3, as well as the rationale for these advancements, will be described in individual sections related to the experimental setup of the Lake Sector, such as in changes to lake model forcing datasets and background information on lake mapping.

Second, for sections that explain protocol differences or consistencies between ISIMIP2 and ISIMIP3, we have updated the topical sentences for each paragraph to emphasize the phase(s) that they focus on.

Section 3.2, P10L23:

> In the global domain of ISIMIP2, generic lakes in each grid cell used average lake depth and surface area information derived from a rasterized version of the Global Lake and Wetland Database (Lehner and Doll, 2004).

Section 3.5, P20L23:

In ISIMIP2 and ISIMIP3, to account for variations in individual lake responses to meteorological drivers (Kraemer et al., 2015; Shatwell et al., 2019; Heiskanen et al., 2015), there were only two types of data needed by the lake models: a description of the lake bathymetry and information on the lake water transparency, which are necessary for estimating the diffuse attenuation coefficient of incoming shortwave radiation.

Section 3.5.1, P21L10:

For global lake simulations in ISIMIP2, the bathymetry of the representative lakes in each grid cell was derived from a rasterized version of the Global Lake and Wetland Database (Lehner and Doll, 2004; Toptunova, 2019).

Section 3.5.1, P21L21:

In the ISIMIP3 global lake simulations, we selected a representative lake for each grid cell from the 1.4 million lakes included in the HydroLAKES shapefiles (Messager et al., 2016).

Section 3.5.3, P23L15:

To simplify lake simulations, the water balance and water inputs and withdrawals were not considered in ISIMIP2 and ISIMIP3.

We believe that these change now better explain the differences between simulation rounds and prefer this option over splitting the paper into two parts for describing ISIMIP2 and ISIMIP3. The latter would mean that a reader has to scroll throughout much of the manuscript to make comparisons between how, for example, lake input datasets for bathymetry are handled differently between ISIMIP2 and ISIMIP3. This also does not work for sections where the protocol between the two phases has not advanced, such as for the water balance or water transparency parameterization. Likewise, we prefer to not change the original title, but we now clearly state the relation of the paper to the two simulation rounds in the abstract of the paper:

> Empirical evidence demonstrates that lakes and reservoirs are warming across the globe. Consequently, there is an increased need to project future changes in lake thermal structure and resulting changes in lake biogeochemistry in order to plan for the likely impacts. Previous studies of the impacts of climate change on lakes have often relied on a single model forced with limited scenario-driven projections of future climate for a relatively small number of lakes. As a result, our understanding of the effects of climate change on lakes is fragmentary, based on scattered

studies using different data sources and modelling protocols, and mainly focused on individual lakes or lake regions. This has precluded identification of the main impacts of climate change on lakes at global and regional scales and has likely contributed to the lack of lake water quality considerations in policy-relevant documents, such as the Assessment Reports of the Intergovernmental Panel on Climate Change (IPCC). Here, we describe a simulation protocol developed by the Lake Sector of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) for simulating climate change impacts on lakes using an ensemble of lake models and climate change scenarios for ISIMIP phases 2 and 3. The protocol prescribes lake simulations driven by climate forcing from gridded observations and different Earth system models under various representative greenhouse gas concentration pathways (RCP), all consistently bias-corrected on a 0.5° x 0.5° global grid. In ISIMIP phase 2, 11 lake models were forced with these data to project the thermal structure of 62 well-studied lakes where data were available for calibration under historical conditions, and using uncalibrated models for 17,500 lakes defined for all global grid cells containing lakes. In ISIMIP phase 3, this approach was expanded to consider more lakes, more models, and more processes. The ISIMIP Lake Sector is the largest international effort to project future water temperature, thermal structure, and ice phenology of lakes at local and global scales and paves the way for future simulations of the impacts of climate change on water quality and biogeochemistry in lakes.

R2C2

You compared model metrics in table 3 (RMSE and R2) but the models can be adjusted on a different number of parameters. As a consequence, the model with more parameters to adjust can better explore the space and at the end have better RMSE or R2 because of an overparameterization. I suggest to used metrics that take the number of adjusted parameters into account such as the BIC.

We agree with the Reviewer that using an information criterion such as BIC or similar is good practice when comparing models' performance in multi-model studies. A prerequisite for this is that the dataset used for model calibration and model performance assessment is the same. However, in this specific case the different models rely on different vertical discretizations of the lake, ranging from models with 1 layer to models with tens of layers. Therefore, different models rely on different quantities and types of data, which prevents us from assessing a model performance ranking based on the use of an information criterion. Similarly, the model performance statistics presented in Table 3 should be interpreted in a critical way and cannot be used for direct model ranking only based on comparing the values of these performance metrics. Indeed, in some cases these metrics are calculated considering the whole temperature profiles (e.g., GOTM) but in other cases only consider lake surface water temperature (air2water). Expanding on the Reviewer's comment, we added a clarification on this aspect at P29L3 of the revised manuscript.

It should be noted, however, that inter-model performance comparisons are difficult here. Due to the diverse discretization of lake temperature profiles across models, each model is being evaluated on a derivation of available lake measurements. Therefore, the observations used as a reference in the performance metrics are different across models.

R2C3

P4 line 16-17 "…lakes are among the most anthropogenically altered ecosystems on Earth…" This kind of statement is a bit weird I would prefer to read specific examples such as eutrophication, change in the water regimes etc.

This was indeed an overly general statement. We have therefore restructured the introductory paragraph for greater clarity (P4L16), which includes removing this sentence.

> There are over 117 million lakes on Earth covering only 3% of the land surface (Verpoorter et al., 2014), yet freshwater ecosystems in general host 10% of Earth's known animal species (Reid et al., 2018). Many lakes provide ecosystem services to their local communities for drinking water, fisheries and transportation, and the number of services provided by lakes has been shown to decrease with deteriorating lake health (Janssen et al., 2021). As well, lakes are effective as local indicators for both environmental changes at the watershed scale and as "sentinels of climate change" in that they buffer synoptic-scale variability but incorporate information on seasonal cycling, inter-annual variability and long-term changes in lower atmospheric conditions. Therefore, studying lake impacts across scales is an important field of research for disentangling the global impacts of climate change from the other anthropogenic pressures that climate change interacts with. However, estimates of historical and future lake responses to climate change have, until recently, largely been carried out as site-specific studies with different goals, data and modelling protocols, which complicates the generalization of simulated impacts at regional and global scales.

R2C4

P8 line 14: Maybe worth to better define what is a representative lake.

We have adapted the text here in response to R1C4, which makes a likewise critique of the wording in this sentence. As this paragraph introduces our methods in Section 3: "Experimental setup", we do not use it to deeply elaborate on the meaning of a representative lake. However, we clarify its meaning in the first paragraph of Section 3.2, which is referenced at the first instance of the "representative lake" term after the abstract (P10L17).

> For a given grid cell, such a lake is termed "representative" because it is assumed to represent real lakes bound by its coordinates by sampling their bathymetric information to perform uncalibrated lake model simulations. The background data and sampling methods for generating representative lakes has evolved from ISIMIP2 to ISIMIP3.

R2C5

Section 3.2: In general when the ISIMIP3 protocol is presented it would be nice to have few lines to explain what is the rational behind the modifications from ISIMIP2 to ISIMIP3.

We have made a number of efforts in the text include the rationale for these progressions in simulation protocol from ISIMIP2 to ISIMIP3.

In Section 3.2, we describe the benefit of improved lake mapping established in ISIMIP3 relative to ISIMIP2 (P11L13):

> With this methodology, the 41449 generic lakes in ISIMIP3 represent true lakes in a more realistic way than for generic lakes defined in ISIMIP2 (see Section 3.5.1).

In Section 3.4.3, we elaborate that, in ISIMIP3a, the addition of a historical, counter-factual forcing dataset allows for climate impact attribution (P19L23):

> Models driven by the counterfactual climate and other historical human pressures provide a baseline to compare with simulations forced by the observational climate forcing to determine climate change impacts, paving the way for IPCC Working Group II style impact attribution (Cramer et al., 2014).

Again in Section 3.4.3, we describe reasoning for the GCMs selected from CMIP6 for the updated input datasets used in ISIMIP3 (P20L9):

> Like in ISIMIP2, the GCMs chosen for ISIMIP3 were constrained by data availability, yet they are also a subset of better-performing models relative to the entire CMIP6 ensemble and they contain structurally independent model components (Lange, 2021).

R2C6

P10 line 22-23: So at the end you may have situations with more than one lake within a grid cell. In this case how do you manage, do you "merge" the 2 lakes to have a single water body or do you have a more complex description of the sub grid heterogeneity.

The sections relevant to this question have been heavily reworked. This includes, i) only using Section 3.2 to define what a “representative lake” means, its function in the context of studying climate change impacts on lakes, and, briefly, how representative lakes are generated in ISIMIP2 and ISIMIP3, ii) a migration of the detailed methodologies for setting bathymetric information in our global lake simulations and representative lakes therein for both ISIMIP2 and ISIMIP3 toward Section 3.5.1, and iii) an updated explanation for how representative lakes are handled in ISIMIP3 in Section 3.5.1. We therefore refer to sections 3.2 and 3.5.1 for this answer.

R2C7

Section 3.3: It would be interesting to know what are the criteria to be included in the group of models. For instance, should a model pass a couple of benchmarks before being incorporated? Is it based on the representation of some key mechanisms?

We did not use any benchmarking to allow models to be used in the lake sector. Modelers were welcome to participate if they wanted. However, for some publications some of the lead authors decided not to incorporate the results from some models because limitations of the models. Since there is a description of each model in ISIMIP, one can decide which models are suitable for a particular analysis.

R2C8

Section 3.5: I am not really a specialist of lake model parametrization so I assume that it was done following state of the art methods.

Indeed, model parameterization is done as much as possible according to the ISIMIP protocol, which connects to the state-of-the-art methods for model parameterization.

IR2C9

P24 line 10: Maybe it worth reminding here that all information to download the data are in the code and data availability section.

Thank you for this suggestion. We have additionally merged and migrated separate instances of data/code availability to a dedicated section at the end of the paper. This includes moving a link for ISIMIP input data from Section 3.4.4. (now titled, “Climate data application” instead of “Climate data availability”; P20L13) to the “Code and data availability” section (P36L22).

R2C10

Table S2 and S3 for some models (GOTM, CLM Mylake there is no answer to the question "Was a spinup scenario used?". I guess that when it is not answered yes it means no but it should be clearly written to avoid confusion.

We have updated both tables to include the spin-up details of all missing models.

R2C11

Fig2: The letters are visible in the plots.

We have adapted the figure in response to R1C24.

Reviewer 3 – Willem van Verseveld

R3C1

While there is some overlap between lake models (ALBM, GOTM and Simstrat) for local and global simulations, it is not clear to me why not all 6 global lake impact models were also applied to the local domain and calibrated in this setting. This should be clarified in the paper. In fact, you could use the same model sets for both applications. This approach can also give useful information about the value of default/a-priori model parameters at the global scale when compared to local domain calibrated parameters and simulation results. In the Conclusions lines 9-12 states that reasonable parameter and coefficient values from the local domain were used in the global domain. How was this exactly done (it seems this is only described for GOTM)?

Thank you very much for the review of this study. The differences in model application across local and global domains exist because modelers could apply for contributing global and/or local runs freely. We did not impose any limitations concerning this. Applying the local and global simulation procedure has some notable differences, however. For example, the local simulations require calibration, and the global models are gridded. The latter implies that a lake model could be run embedded in a larger, gridded land surface model (e.g., LISSS within CLM4.5 or VIC-LAKE within VIC). In these cases, it is not straightforward to additionally apply the model also at the local scale without considerable extra technical work. Another issue is that the global simulations require much larger computational resources, which may explain why some local lake models were not applied at the global scale (see response to R1C1). These considerations may explain the discrepancy in models between the local and global scale, which we briefly describe in Section 3.8, "Output data format" (P28L3).

> […] These technical prerequisites, in addition to individual model feasibility issues for local versus global domain simulations, explain the discrepancy in model availability across the ISIMIP2 local and global simulations.

Regarding the use of locally calibrated parameter values in the global simulations, indeed, only GOTM attempted this. This was because GOTM contributed global simulations after its calibration in the local domain. The rest of the modelling groups did not have the same sequence of work. Instead, these modelers performed global simulations as soon as possible to adequately pace the contribution of the Lake Sector simulations relative other ISIMIP sectors that had already completed ISIMIP2a/b simulations. Our manuscript was indeed not clear on this distinction. We have therefore made several adaptations to clarify this.

Section 3.7, P26L13, where this application of locally calibrated parameters at the global scale is described for GOTM.

> For simulations in the global domain, most lake models used default parameter and coefficient values that were set according to previous experience with each model (see Table 1: "Key references"). Exceptionally, for GOTM, the average values of calibrated coefficients from the GOTM local lakes (Table 3) and default values for the coefficients that were not calibrated (Umlauf and Lemmin, 2005; Sachse et al., 2014) were used for all representative lakes in the global domain.

In our conclusions, P35L17, this sentence is clarified to better suggest that the application of locally calibrated values in the global simulations is a strong path forward but that it has not been fully achieved yet.

> In future global simulations, these locally derived parameter and coefficient values could improve the full ensemble of models that have so far been uncalibrated in their global domain applications.

R3C2

The water balance is not considered as part of the lake modelling efforts (paragraph 3.5.3). In lines 20-23 of paragraph 3.5.3 the explanation is that one should use caution for (only) seven lakes or reservoirs with large water level fluctuations (Table 1). If think the wrong Table is referenced here? Seven lakes or reservoirs are part of the local domain, what about the global domain? And what are large water level fluctuations (definition)? But even without large water level fluctuations, large input-output changes (inflow, outflow, precipitation, evaporation) because of climate change can have a significant impact on lake temperature (e.g. changes in residence time)? This part in the paper requires more explanation, either including appropriate references that confirm that the omission of water balance components has only a significant impact on lakes with large water level fluctuations, or rewrite this to a more cautious statement (this omission can have a broader impact).

We have fixed the table reference to Table S1, which highlights seven reservoirs which could exceptionally be affected by this assumption. As well, we have reworked this section to describe some of the cases where omitting the water balance could affect the reliability of our simulations. We further clarify that this assumption is made for both ISIMIP2 and ISIMIP3, as this was not explicitly said in the original manuscript (P23L15).

> To simplify lake simulations, the water balance and water inputs and withdrawals were not considered in ISIMIP2 and ISIMIP3. The formulations of some lake models (e.g., air2water or FLake) do not explicitly include hydrological balances. For the rest of the models, the precipitation and evaporation component of water mass exchange was switched off (i.e., only heat exchange occurred) or compensated with a closure term (e.g., CLM4.5). This assumption allowed us to evaluate changes in lake thermal structure in the time frame of the ISIMIP2 and ISIMIP3 simulation periods.
>
> Regional studies assessing the hydrologic responses of lakes to an ensemble of future climate change scenarios show that our omission might variably affect lakes depending on lake type and future climate outcomes for seasonal drying and wetting (Hanson et al., 2021; Hunt et al., 2013). These studies found that drainage lakes in northern Wisconsin, US, which are hydrologically mediated by lake inflows and outflows, were projected to maintain stable water levels because of competing climatological factors that did not promote a clear drying trend. Under our omission of lake water balances, projections for such lakes could lose reliability where future climate conditions reduce watershed runoff. In the same region, seepage lakes with minimal surface water fluxes and a greater dependence on ground water inflows, however, were projected to significantly decrease in water level, especially in higher elevation regions near groundwater divides. These studies are

relevant for both our local and global lake simulations. For lakes in the local domain, despite accurate representations of historical changes in lake thermal structure (Table 3), the omission of a water balance could additionally affect the simulated climate change impacts in seven lakes and reservoirs with large water level fluctuations (Table S1), thus caution should be used when evaluating these results. For lake simulations in the global domain, this omission is yet another necessary trade-off between experimental complexity and spatial representativeness (see Section 3.2).

R3C3

It is stated that the global lake models were not calibrated because of lack of a global-scale data set of measured lake water temperatures. What about using other datasets like satellite based datasets for example for surface water temperature? Or are there any (planned) efforts to setup a central data repository to collect measured lake water temperature (and other variables) data, for example similar to the Global Runoff Data Centre (discharge data for hydrological applications)? Also, when the water balance is considered, additional data like water level, surface area dynamics etc. (for example from satellite data) could be considered for calibration/validation purposes. Would be good to add a section/alinea to the paper that considers some of the solutions/ideas from the authors for a of lack of a global-scale data set.

In the experience of some of our co-authors, remote sensing lake surface temperatures, if not corrected by in-situ temperatures, are not (yet) sufficiently accurate to be used for calibrating lake models, as the error of an uncalibrated model compared to in-situ observations is often smaller than that of uncalibrated remotely sensed surface lake temperatures. Remote sensing observations are also constrained to surface level measurements, which are less effective than in situ measurements for calibration.

For the rest, there are no current initiatives to set up further data collection processes in an attempt to calibrate the global lake simulations. However, we discuss the potential for developments here in response to R3C5.

R3C4

Generally a calibration strategy also includes the validation of the calibrated model. It is not clear if this was done, please describe this in more detail or explain why a validation was not carried out.

We have updated the text for clarity around this point (P25L11). For the most part, modelers have not followed a typical split-sample approach to calibration and validation. As explained in our addition to the manuscript Section 3.6 (“Calibration of local lake models in ISIMIP2a”) below, most rather choose to use the full set of available lake temperature measurements for each local lake for calibration. Therefore, for all models except for ALBM, performance metrics from calibrated lake simulations come from comparing the simulations against the calibration data.

Most lake models were calibrated with the full series of available measured observations. In this majority of cases, no data was withheld for an independent model validation. Considering the relatively short temporal extent of lake measurements, this was done to base parameter estimates on the full range of environmental conditions encountered during simulation for producing robust future projections (Larssen et al., 2007). This is justifiable given extensive research validating the performance of these models outside the calibration period (e.g. Stepanenko et al., 2013; Thiery et al., 2014) and arguments calling for skepticism of the split-sample approach to calibration and validation (Augusiak et al., 2014; Shen et al., 2022). Exceptionally, ALBM only used the full series of measured observations when the observations were shorter than five years. Where measurements exceeded five years, modelers running ALBM simulations opted for a split-sample approach to tuning their model and used the first five years of measurements for calibration.

R3C5

An extra section to describe future work (ongoing ISIMIP 3, possible solutions lack global dataset?) would be useful (see also the last specific comment below).

We have added Section 4.3, "Future work", to address this (P33L1).

4.3.1. Model response to observational vs simulated forcing data

In addition to simulations using ISIMIP2a forcing, the ALBM and FLake models were also used for simulations forced by EWEMBI observational data (1979-2016). This will allow for assessment of the difference in model output when used with observational forcing data compared to simulations with GCM forcings during the historical time period. Given that impacts under past and future climates are modelled with bias-adjusted GCMs, a comparison with simulations using the observed data used for bias correction will allow an assessment of how simulations forced with the GCM historical inputs compare with those forced using observed (historical) climate (see also Piccolroaz et al., 2018 for a similar analysis). This can give an estimate of the uncertainty in the ISIMIP GCM scenarios and the bias correction method. There are so far no studies for this application of the ISIMIP2a simulations, but the existing simulation outputs archive are publicly accessible and hold potential for further study.

4.3.2. Lake hydrology and water quality assessments

Current lake modelling activities in ISIMIP are biased towards lake physics and concentrate mostly on water temperature and related variables like ice cover or stratification state. Lake managers require more than that and are usually highly interested in projections for water quantity (i.e. inflow discharges) and water quality and potential effects on the services lakes deliver. Future directions of the lake sector beyond ISIMIP3 are therefore seen in (i) linking the water sector (hydrological models) with the lake sector in order to integrate water quantity projections into lake simulations and (ii) adding water quality descriptors by biogeochemical modeling of lakes. Such modeling can make projections for the future development of ecosystem services and biodiversity in lakes in relation to climate change and socio-economic development. For climate change, such assessments can be directly built on the ISIMIP2 and ISIMIP3 simulations of the Lake

Sector but require linkage with the transport of water and nutrients from their catchments (Janssen et al., 2019). For that, nutrient transport models such as IMAGE-GNM (Beusen et al., 2015) or MARINA (Strokal et al., 2016) need to be aligned akin to the ISIMIP approach.

4.3.3. Global scale calibration and validation

Important steps can be made in the development of a global-scale data set for calibration and validation purposes. There are a few challenges to overcome in the future (Janssen et al., 2015). First, due to project-based research, long-term measurements are rare as often measurement campaigns stop when projects are over. Second, data is often locked within institutes, meaning that a consistent global database requires corporation between various parties. Similarly, in-situ data that have not been properly indexed and stored, sometimes referred to as "dark data", require rescue efforts to extend back our measurement period of lakes. Third, data is gathered inconsistently e.g. by using different methods, measuring over different periods, or collecting at different spatial scales. Remote sensing could overcome these issues to some extent, as they can provide long-term, global observations.

Remote sensing datasets for lakes are increasing (Dörnhöfer and Oppelt, 2016). Examples of already existing datasets are datasets for lake temperature (Sharma et al., 2015), ice phenology (Wang et al., 2021), and even biological indicators (Fang et al., 2022; Hou et al., 2022). A disadvantage is that remote sensing is limited to proxy values, which still require ground truthing by in situ monitoring data. Moreover, remote sensing performs variably depending on the measurement system, weather conditions and variable in assessment. While optical imagery is easily obscured by cloud cover, active microwave systems can be used in all-weather conditions for some variables such as ice cover (Kilic et al., 2018; Murfitt and Duguay, 2021). Therefore, satellite observations must be combined with highly spatiotemporally resolved in situ measurements from buoys, field sampling programs, and long-term monitoring networks (Rand et al., 2022). Specifically, in situ measurements are essential for observing lake processes below the water surface (such as stratification and mixing), to improve understanding of complex air–water energy fluxes (such as evaporation) and to maintain long-term perspectives that began prior to the advent of satellites and regardless of weather conditions that adversely impact some satellite measurements.. First attempts at such databases are for example the HydroLAKES database which already has water discharge into lakes (Messager et al., 2016).

R3C6

And, finally, briefly some differences between models (P26, lines 24-27) are mentioned. Are there more examples from his study? I think it could be useful to include this kind of information more extensively in the paper.

These brief analyses using GOTM are meant to exemplify the possibilities for undertaking studies using the ISIMIP2 Lake Sector simulations. We therefore choose not to provide any further analysis because the focus of our paper is to describe our experimental protocol. However, several published studies demonstrate the variety of potential use cases for lake simulations organized under the ISIMIP framework. In Section 4.2.1, we therefore added a paragraph dedicated to summarizing these efforts (P32L13).

> Existing studies applying these simulations demonstrate the many possibilities for exploring the impacts of climate change on lake physics under the ISIMIP protocol. ISIMIP simulations have been used in a first ever assessment of the global heat uptake by inland waters (Vanderkelen et al., 2020), a relevant addition to existing evaluations of Earth's global heat budget in its land, atmosphere and oceans. The ISIMIP Lake Sector database has also been used to assess present and future alterations of lake mixing regimes (Woolway and Merchant, 2019) and the shifts in lake stratification and their climatic drivers (Woolway et al., 2021). Finally, both event (Woolway et al., 2021, 2022) and trend attribution (Grant et al., 2021) of lake heatwaves and lake ice cover changes, respectively, have been undertaken using ISIMIP simulations in combination with global observational datasets to confirm the role of anthropogenic climate change in observed lake changes.

R3C7

P22 line 6-7: different objective funtions were used by the different models. Why was not the same objective function used for each model? I think using the same objective function is an important aspect of an ensemble modelling protocol.

Like in our response to R2C2, models represent vertical lake temperature profiles by uniquely discretizing lakes into a number of layers using their own model physics. With this in mind, the reference measurements used for calibrating a set of models to a given lake must be manipulated for compatibility with each lake model's layering of the water column. This means that the output of a single objective function is not directly comparable across models. Therefore, we leave this decision to the modelling groups to use their experience regarding what objective functions best tune their models.

R3C8

Table S2 and S3 seem to have missing information in some table cells (empty)

Thank you for noting, we have filled in these missing entries.

R3C9

P22 line 25: spinup periods were different. Bit similar to objective function, please state clearly why not the same spinup period was used (if applicable) for each model.

We refer to responses to R2C2 and R3C7 here.

R3C10

P25 line 17: Gao reference is missing.

We have removed this paragraph.

R3C11

Paragraph 4.1.2 Better move this part to an “Outlook or further work” section, this is not really a result, but part of possible further study.

We have created a new section to describe future possibilities with ISIMIP (see response to R3C5 and Section 4.3, “Future work”) and migrated this paragraph and subsection there.