Response to Reviewer 2:

We thank the anonymous reviewer 2 for taking the time to proofread and provide insightful comments on the manuscript. We have done our best to address all comments in the revised version to improve the overall quality of the manuscript in line with the reviewer's recommendations. Our responses are reported below in blue.

The Flake lake model within the ORCHIDEE land surface model has been updated by the Authors to better simulate winter ice cover. Using MODIS albedo data and the Great Lakes Ice Cover fraction dataset, the authors calibrated and validated a new lake albedo parameterization that accounts for partial ice cover. This update significantly improved the simulation of ice phenology for 200 lakes of various sizes, as reported in the Global Lake and River Phenology database. The improvements were notable across all lake sizes, with the largest and deepest lakes showing the greatest reduction in error for ice cover duration. The study underscores the importance of considering partial ice cover to accurately model lake ice.

Overall, the work presents a useful contribution to large-scale modelling in partially ice-covered regions but does leave some lack of clarity on how this performs in fully ice-covered regions (e.g. by implying Lake Erie did not improve as much as the rest due to the more uniform ice cover). Are the 200 lakes this is tested on in the warmer regions that have partial ice cover? Or the full GLRP, which includes many northern lakes with full ice cover? And are the SYKE lakes in southern Finland? Or in the north with solid ice covers? While acknowledging this is a brief communication, some important details have been omitted, and once addressed, this paper could be considered for publication.

Thanks for your constructive remarks. We agree that the GLRP and the SYKE databases are not well described. We added more details in the revised version and added a figure (figure S3) in the supplementary document to map the 200 lakes that are used in the evaluation of the lake ice phenology. Because of missing data over our study period (2008-2018), only 200 lakes among the 857 reported in GLRP could be used for the evaluation. These lakes are mainly located in the northeastern part of the US and Canada and in Scandinavia. Even if their location does not cover the whole circumpolar region, they cover a large diversity of atmospheric conditions with lakes presenting a complete ice cover during the cold season to lakes never freezing entirely.

Line 73: Hydrolakes – While agreed that it is the best option available, the limitations in lake depths should be noted here as outlined in Messager et al., 2016 (estimating the depths based on limited information in some locations), and what the implications are for your study.

We agree that the mean depth provided by HydroLakes is subject to large uncertainties. In order to better address these limitations, we added a sentence in the presentation of the database to explain how the lake depth was estimated and the model errors linked to

depth errors are also mentioned in the discussion section. See Lines 73-74 and Lines 315-317 of the revised document.

Line 90: Perhaps comment on why grouping the lake into similar depth tiles rather than just run for each 500m pixel? What if there is fractional ice coverage within each depth grouping?

In this work, we wanted to evaluate the ORCHIDEE-FLake model as it was developed for the modeling of the energy, water and carbon fluxes at the global scale. In this version, the model is forced with atmospheric reanalysis at 0.25° resolution. At such a large scale, lakes for the most part, are not represented explicitly and only lake tiles or effective lakes are represented in the horizontal model grid (as it is done for the vegetation). We have chosen to consider only three effective lakes differing by their depth in each grid cell. Effective depth is given by the mean average of all lakes located in the grid cell, clustered in three categories (shallow, medium and deep). The model evaluation is done by comparing the observed lake to its corresponding effective lake in the grid. We agree that our methodology was not clear in the first version of our paper. We revised therefore the presentation of the model (see Lines 146-152) and the Model experiments section (Lines 248 to 263)

Also is the MODIS water mask not a viable option rather than using a vegetation index?

The MODIS water mask is static (Feb 2000) and based only on the SRTM water mask, based on surface elevation data. Such data can be used to map surface water but don't indicate if the water is covered by aquatic vegetation which has a very different albedo compared to water. This is why we tried to use NDVI data to remove water vegetation, to assess the free water albedo. This has been done in the earlier phase of the work but was finally not kept further. Therefore, we decided to remove this point in the description of the albedo processing. See the slight modifications in section 2.1.2.

Line 132: is the medium category is missing here?

No, this is a mistake from our side; the medium category is the same as the intermediate category. We corrected the word in Line 127.

Line 180: "In Flake, the minimum and maximum albedos are equal to 0.1 and 0.6 respectively, for both snow and ice. They were revised by Bernus and Ottlé (2022) following Semmler et al. (2012) and Pietikaïnen et al. (2018), and set to 0.3-0.5 and 0.77-0.87"

These ice albedos are a lot lower than what have been published for lakes in Central Ontario. Are those albedo values to represent a broken ice cover only? or only during the melt season?

The value of 0.1 for ice is the one prescribed in FLake model (see documentation and related publication). Compared to what has been previously published, we also found that they were too low; this is why in our previous work referenced here (Bernus and Ottlé,

2022) it was set to 0.3. In this work, we observed values as low as 0.15 for dates where ice cover was still reported to be present. It was indeed during the melting period. This is why we decided to reduce the minimum value of the ice albedo to 0.15.

~ line 200 – I think it would be pertinent to show the results here. Also, add more explanation – if Lake Erie has the largest ice fraction, and the worst RMSE does that not indicate that your new method is not suitable for lakes with a full ice cover and should only be used on a partial ice cover? Or is the albedo selected based on the ice cover fraction? This needs some clarity.

Because of the format of the brief communication, we are limited in the number of figures and tables. We chose not to add too many details on the methodology that we have developed to assess the relationship between ice coverage and albedo and to focus our paper on the results obtained after implementation and calibration of the ice cover fraction into ORCHIDEE-FLake. All the results are available in the Master thesis of Z. Titus (in French) which could be referenced if you think it necessary.

Line 205: "Therefore, we compared these specific dates estimated from the MODIS albedo time series to those observed in the GLIC database and found that over the period 2010 - 2015, our method has an average error of 2 days for Lake Nilakka and 5 days for Lake Nasijarvi, the only two lakes we could formally identify in the HydroLAKES database."

Unclear – why are there Finnish lakes in the GLIC database? Clarify why only the 5 years? Can you not use maps to identify the other lakes in the database? The explanation here needs more clarity. And again, does this indicate that these lakes had partial ice cover? Or did the results improve for full ice cover here, but not for Lake Erie?

We are sorry, we meant the SYKE database and not the GLIC database. The GLIC data concerns only the Laurentian Great Lakes. Sorry for the confusion. It is corrected in the revised version.

Line 230: which lakes? Where? From what I recall the GLRP is not really global, it's mostly North American - are your test lakes mainly in Canada then? in regions with full ice covers? The work needs a map to show which lakes you are using.

We followed your suggestion and added a map as a Supplementary figure (Figure S3) to show the GLRP lakes used for the model validation (see also response to the general comments above.

Line 249: "This is particularly the case for Lake Erie in 2008 and 2015 or for Lake Huron in 2015 (not shown here)."

Consider showing the examples and why the warm/cold year performs well/not well. Perhaps in the supplemental section if length is an issue for the brief communication. Variability is important.

We fully agree that time variability is very important and finally replaced Figure 1 with a new figure showing the interannual variability and model performances over the studied period (2008-2018). Moreover, we added as supplementary Figure S2 the prior and remaining model errors for the five Great Lakes, for each studied year, highlighting with a specific color scale, the degree of coldness of the winter, from the warmest in dark red to the coldest in dark blue. We tried to better discuss in the results section 3.1, the performances of the model according to the atmospheric temperature conditions. We also transfer the previous Figure 1 showing the interannual mean average of the surface albedo to Figure S1 (supplementary document), in order for the reader to get a global picture of the model improvements. Finally, we added Table S3 in supplementary, to give more quantitative results of the comparison, with minimum, maximum, standard deviation and interquartile range of the annual distributions of the water albedo, observed and modeled. We hope that these changes make it easier to appreciate the added value of our developments.

Line 265: "Given that deep lakes present partial coverage most of the time and that the observations may not be representative of the whole lake, we decided, after various tests, to diagnose the ice phenology dates with the simulated surface temperature as was done in Bernus and Ottlé, 2022."

Which deep lakes have partial coverage? Where? How far north? this is something that needs to be clearly addressed throughout.

In ORCHIDEE-FLake, the effective lakes are clustered in 3 categories (shallow, medium and deep) according to the HydroLakes database as it is hopefully, better explained in Section 2.2.1 in the revised version. We chose to diagnose the start and end of the freezing season differently for the modeled deep lakes, based on the surface water temperature since we realized that the observed dates are local and are not representative of the whole lake. For the shallow and medium categories, the full ice coverage is generally observed and can be compared to the simulated ones. For the deep lakes (which represent only 10 lakes over the 200 simulated), the start and end of freezing are diagnosed in the model with the water temperature and the 0°C threshold. We better explain our methodology in the Methods section (Model experiments) and added this uncertainty issue in the discussion.

Figure 2: Do you mean the global river and lake ice database? Not the great lakes ice cover?

We are sorry, there was a typo in the legend of Figure 2. The data are coming from the GLRP database and not the GLIC one. We corrected this error in the revised version. Thanks again for your careful reading.