At the beginning, we would like to thank Referee #1 for useful comments which helped us improve the manuscript even further. Our response is as follows.

General comments

The manuscript has been significantly improved based on the previous reviews. However, I still think it is a bit of a mixture of topics that usually should not be combined in a single paper and much of which have relatively limited scientific novelty. But I think it is up to the journal editor to decide whether they think it is appropriate for the journal.

The manuscript in my view consists of three topics:

a) How to derive the forcing for a lake model if limited meteorological data is available (explicitly no direct observations of the radiative fluxes)? A viable way for this is now clearly described in the paper, but I don't think there is much novelty included in this approach.
b) The development of a new lake model. It remains unclear, however, which properties of the new model justify its presence besides the substantial number of already existing models. The manuscript focuses on the integration of the heat flux procedure from a), but this could have been easily integrated in any already existing lake model instead. Furthermore, due to the mixture of topics, the structure of the paper is suboptimal for a methods paper introducing a new model.
c) The numerical experiment looking at the model performance as a function of the simulation length. This is actually an interesting numerical experiment, and I am not aware that I have seen it been done elsewhere. However, it takes a disproportionate fraction of the paper, if the focus of the paper should be the introduction of a new model (as implied by the title). Furthermore, there is almost no guidance for the reader to understand the purpose and the implications of this numerical experiment. Why is this analysis performed? What hypotheses should be investigated with this approach? What general take-home messages do result from this analysis? What are the implications for the application of this and other lake models?

The leading goal of this paper was to develop a model which can be used as a black box with as little input data as possible. This can be especially useful, for example, for scientist from other fields. We do understand how the derivation of the forcing may be seen as a separate topic, but from the standpoint of creating a practical model we consider it to be a component of the model itself. Therefor we consider it is justified to unite these two topics (a and b).

Furthermore, GMD guidelines for model description papers say that "*Model description papers are comprehensive descriptions of numerical models which fall within the scope of GMD. The papers should be detailed, complete, rigorous, and accessible to a wide community of geoscientists. In addition to complete models, this type of paper may also describe model components and modules, as well as frameworks and utility tools used to build practical modelling systems, such as coupling frameworks or other software toolboxes with a geoscientific application.*" It is our opinion that our model falls into this category.

Regarding the evaluation part (c) - we took into account that it is most common, and even required, to include it in the model description paper. We didn't consider it as extensive as to require a separate paper.

We do recognize that the purpose of the said numerical experiment was not well communicated to the reader. Thank you for pointing that out. The results of this analysis are to show the model ability to provide quality short term prognosis and the rate of the result deterioration with the increase of the simulation length. The current text is now modified to address this.

In addition, the two following **major points** should be considered:

[1] The new comparison of the model with GOTM and SCHISM could be useful. However, it needs a description how the other two models were parameterized (could be in the supplement). Was exactly the same approach used for all three models (e.g., same initial profile, same meteorological forcing)? The text is a rather unclear about that.

A description of the SCHISM and GOTM model parametrization has been added in the Appendix B.

All three simulations are started with the same initial lake temperature profile. In SIMO the meteorological forcing is calculated using solely data measured at one point next to the lake. Apart from the measured air temperature and wind data (GOTM simulation) and measured air temperature (SCHISM simulation), meteorological forcing for GOTM and SCHISM was modeled with the atmospheric Weather Research and Forecasting (WRF) model (Skamarock and Klemp, 2008). As the SCHISM model is 3D (while SIMO is 1D), it requires atmospheric forcing above entire lake area (not only in a single point), thus, input data for SIMO and SCHISM could not be exactly identical. In both GOTM and SCHISM simulations, freshwater was assumed. Also, due to consistency, in both model runs the same k-ε turbulence closure scheme of Rodi (1984) was employed. Finally, both models were initialized with the lake temperatures observed at 1 January 2019 (same as SIMO).

This was already stated in the text, but it has now been slightly modified to better convey this information.

Also, I have no experience with SCHISM, but there must be clearly something wrong with the model setup. Otherwise the surface temperature couldn't possibly be systematically off by 5 °C for the entire second half of the year.

During calibration and parametrization of models, the GOTM model was set with finer vertical grid resolution, while in SCHISM the vertical resolution was coarser because it is a 3D model that consumes a lot of resources and time, and the calculations are more expensive. Likewise, the problem with SCHISM is that it covers elements in a large size range (from 1 to 200 m^2) and thus covers a relatively large CFL range (because it uses time implicit integration), which necessarily means that it has problems with either diffusion or numerical errors. In the end, the SCHISM model could perhaps be improved in some respects, but this would require very long time integrations (due to small elements and time steps), which was not the main goal of this work, but the development of a new model. Also in this case, vertical processes in the lake (because we are looking at long time scales) are more important than horizontal processes (advection), which is better resolved in the GOTM model.

[2] The model does not, as far as I understand, consider the changing lake area with depth. This leads by definition to either a wrong net heat flux at the lake surface or wrong temperatures in the lake, because the ratio of surface area to volume is different in the model than in the real lake. This should somewhere be mentioned as a limitation of the model.

It would be easy to include the changing of the lake area with depth in the equations/model. However, considering that more often than not, the bathymetry of the lakes is not available, as well as our goal to keep the model as simple as possible and limit the input data, we decided to use the constant area assumption. This explanation has been added in the text (at the beginning of chapter 3):

Considering that more often than not, the lake bathymetry is not available, as well as our goal to keep the model as simple as possible and limit the input data, it is assumed that the water body has a constant horizontal cross-sectional area (which can be of any shape).

[3] The manuscript is rather lengthy and contains a large number of figures. It also seems occasionally repetitive. There is certainly potential to make it more concise without losing relevant information.

We are aware that the manuscript is rather lengthy, but as it encompasses all the aspects of the model plus its evaluation we find it rather hard to make it much shorter. The evaluation segment may seem a bit repetitive at first glance, as the figures for the short term sensitivity analysis and the long term simulation qualitatively resemble, however, they convey different aspects of the model performance.

Minor comments

[1] Line 45: I would not consider GLM (Hipsey et al., 2019) as a two-layer model

Thank you for noticing, this is absolutely true. The text has been changed to:

Energy budget-based models assume series of well-mixed (sometimes just two, namely, the epilimnion and hypolimnion), and they use the kinetic energy produced by wind shear on the surface to account for the mixing dynamics within these two layers and/or to estimate the depths of these layers (e.g., Bell et al., 2006; Mironov et al., 2010; Hipsey et al., 2019).

[2] Line 54: The first author's name is Råman Vinnå.

Thank you for pointing this out. The mistake has been corrected.

[3] Line 175: Eddy diffusivity is certainly not negligible in the hypolimnion of these lakes or small lakes in general. Observations in much smaller lakes typically show vertical diffusivities in the hypolimnion on the order of 10^-6 m2/s, which is one order of magnitude larger than thermal diffusivity. One of the first famous studies to investigate this is Powell and Jassby (1975, https://doi.org/10.4319/lo.1975.20.4.0530), who investigated Castle lake (0.2 km2). Other examples are Soppensee (0.23 km2; Vachon et al. 2019, https://doi.org/10.1002/lno.11172), or two small lakes (0.25 and 0.05 km2) in the Canadian Experimental Lakes Area (Quay et al., 1980; https://doi.org/10.4319/lo.1980.25.2.0201). There are certainly many more examples

available in literature. Mixing in the interior of lakes can sometimes go down to molecular levels, but basin-scale mixing almost never does.

We really appreciate the mentioned references. The text has been changed to:

Although Sun et al. (2007) suggest that for shallow lakes (less than 50 m deep), the turbulent thermal conductivity is negligible, this is not in accordance with findings of numerous other studies which suggest that the turbulent thermal conductivity can be much larger than the molecular thermal conductivity even for shallow lakes (eg. Jassby and Powell, 1975, Quay et al., 1980. Vachon et al., 2019). It should be kept in mind that these studies often determine the turbulent diffusion coefficient based on measured change rate of lake water temperature vertical distribution, which means that the contribution of all mixing processes is included (i.e. shear-induced turbulence, breaking internal waves, boundary layer turbulence). However, the mixing processes and their contribution to turbulent mixing may differ from lake to lake. In the present study, turbulent thermal diffusion was taken into account using Eq. (3).

[4] Line 231: The description of D is rather confusing and disagrees with that in the original paper of Winslow et al. (2001).

Maybe it was formulated a bit complicated, but the description did agree with the original paper of Winslow et al. (2001). We tried to make the text less confusing.

The effect of cloudiness is indirectly taken into account by introducing the factor ($1-\beta rh_{Tmax}$). This is based on the finding that the solar irradiation from sunrise, when minimum humidity is expected ($rh_{Tmin}\approx 1$), until the maximum daily air temperature (and minimum humidity rh_{Tmax}) is reached, is proportional to the decline of the relative humidity, $S_{surf\_Tmax} \propto (1-\beta rh_{Tmax})$. The factor $D=S_{surf}/S_{surf\_Tmax}$ is introduced to account for the surface solar irradiation from the moment when the air temperature reaches its daily maximum until sunset. D is calculated assuming that the air temperature reaches its maximum around 3pm

[5] Line 252: add units to K1 (lambda_e) and K3. Why not just use lambda_e in the equation instead of K1, as it is used in the equation for K2 anyway?

K1 was used as we wanted to keep the equation structure as shown in Henderson-Sellers (1986).

[6] Equation (22): I don't think there is a physical reason to assume that (1 - longwave albedo r) and the emissivity of the water surface (epsilon) are identical. Why not simply keep r and eps as two variables, they can still take the values 0.04 and 0.96 to keep the results the same.

The proposed change has been adopted.

[7] Line 290: typo in "heat"

Corrected.

[8] Equations (29) and (30): should these equations not have a negative sign (positive fluxes downward)?

That is true and has been corrected. Thank you for noticing. The mistake is present only in the text and not in the code.

[9] Line 347: I still think a time step of 1 hour is rather large for the interpolation and could lead to significant numerical errors. Maybe it is ok, but it should at least be checked for one example by how much the model output changes if a smaller time step is used. I disagree with the previous reply that the available time resolution of the meteo data precludes such an analysis. The main question is whether the mixing algorithm in the

lake model produces different results for a higher time resolution and that can also be tested with meteo forcing that is constant over each hour.

The one year simulation for Lake 12 was run for dt = 60s, 600s and 3600s. Surface temperature and performance measures are shown in the figure and table below. The improvement with dt=60s compared to dt=3600s is far from dramatic. Even further reduction of the time step to dt=60s even slightly worsens the performance compared to dt=600s.

Comparison of the near surface water temperature for the period from 01.01.-27.12.2019. for different integration time steps

Comparison of performance measures for the period from 01.01.-27.12.2019. for different integration time steps

Performance measure	Unit	Time step		
		3600 s	600 s	60 s
RMSE	° C	1,4803	1,4348	1,4352
Bias	° C	0,8504	0,7878	0,7891
MAE	° C	1,1847	1,1344	1,1375
MaxAE	° C	3,9641	3,9654	3,9654

The text has been modified to:

Considering the time resolution of the available input data, the model was run with a time step of one hour (runs with finer time steps were attempted, however the performance improvements were not significant).

[10] Line 396: Is “totally out of phase” the correct statement here?

The word “totally” has been removed.

[11] Line 405: Monomictic lakes can also mix at temperature significantly different from 4°C. It is therefore not necessary a correct assumption to initiate the model with a constant temperature of 4°C.

The temperature in the parenthesis was removed from the text. We agree that lakes don’t necessarily mix at this temperature, however in this case it is true (shown by measured data).

[12] Line 457: If the reason for the too high surface temperatures was an underestimation of turbulent mixing, there should also be an underestimation of deep water temperatures for the same reason. This is clearly not the case, as the bias is positive for all depths. The reason for the bias therefore must almost certainly be the heat budget, where either some incoming heat flux is systematically overestimated or some outgoing heat flux is underestimated. Unfortunately, it is not possible to make a full heat budget calculation since the model does not consider the lake bathymetry (see point [2] above), but it could be at least approximately estimated by how much the net heat input to the lake is overestimated. For example, a mean bias of 0.5 °C (as roughly estimated from Fig. 10a) over 40 m depth would correspond to an excess heat of 84 MJ/m2. If that excess heat results over a time of 30 days, the net heat flux at the surface is overestimated by about 32 W/m2.

Indeed, but we did mention in the text that the heat flux may be overestimated which may be one of the factors leading to the estimation in the epilimnion (line 469).

One more factor that we failed to mention is the neglecting of the tributary influence. This has now been added to the text.

[13] Figure 9: add units to those metrics that have units.

Units added to the figure.

[14] Figure 15: wrong caption

Corrected to:

Figure 1: Comparison of near surface water temperature for SIMO, GOTM and SCHISM for the period from 01.01.-27.12.2019.

[15] Line 580: straightforward

Corrected

[16] Figures A4 and A5: captions are exchanged?

This has been corrected. Thank you for noticing.