

Summary

The manuscript reports the results of a lake model intercomparison study (including 8 different models) focused in a shallow lake in Germany. Several important issues were addressed as the impact of the representation of bottom sediments or lake depth definition. The overall motivation of the work is scientifically sound, and the results can have a significant impact on the community by pointing out processes that are important to represent when modelling lakes in different applications, such as numerical weather prediction. In general the manuscript is well written and organized, but the interpretation of some results seems inconsistent. Therefore, I recommend the manuscript to be accepted after those inconsistencies have been addressed by the authors.

General comments

Pag. 4003;L3-5: How were the initial conditions prescribed for the simulations/models with active bottom sediments ?

Pag. 4007;L3-6: The results in Fig. 3 do not seem to support this statement – pointing to the cold bias due to neglecting bottom sediments by SimStrat and LAKEoneD. CLM4 and MINLAKE96 also have a cold bias in the second period and they represent bottom sediments.

Pag. 4009;L10-19: These results seem to contradict the last sentence of the abstract: “while the effect of heat flux to bottom sediments can become significant for bottom temperatures, it has no important influence on the surface temperature”. However, in this section it is estimated that “neglecting sediments may account for at least 50% of the seasonal surface temperature difference error in k-e models”. Please clarify these statements.

Pag. 4010;L26:26: “These results also suggest that the “optimal depth” delivering the most realistic surface temperature is model-dependent”: This sentence seems to contradict the interpretation of table 4 (pag 4010;L23:34 “The data of Table 4 do not allow a conclusion...”). Looking at whole period in Table 4, the lower DM (absolute) is for the simulations with the local depth in all models. However, the differences between the different simulations are very small (and most likely statistically insignificant). Therefore, these results would suggest that for these types of shallow lakes there is a reduced sensitivity of lake depth and an “optimal depth” derived from a single model could be applied to other models.

Pag. 4011;L26:28: “During autumn FLake and Hostetler, along with other models, reproduced well the almost homogeneous thermal structure (weak stratification) developed due to convection (not shown).” : These results are presented in table 5, where FLake and Hostetler models have DM in the second period (for the reference run) of 1.36 and 1.56, respectively. These systematic errors are quite large when compared with the mean observations of 0.26. On the other hand, LAKE shows reduced DM. These results suggest that FLake and Hostetler models still have a strong stratification in the second period.

Different models used in different experiments: Along section 5 the results of different experiments do not include all the models (e.g. zero heat flux :MINLAKE96 is missing (table 3); different lake depths: only 3 models (table 4)). This makes the discussion sometime confusing. Not all the models performed all the sensitivity simulations? If this was the case, I understand

the doing those extra simulations can be time consuming, but it would make the discussion of the results more clear. I leave this suggestion to the author's consideration.

Abstract: The abstract mentions the turbulent fluxes error between 17-28 Wm², but it should also include that the residuals of the heat balance, based on observations, fall in the same range.

Minor changes

Pag. 3995;L11/12: "reproducing temperature well" : "accurately" instead of "well"

Pag. 3996;L12: "Global Coupled Models (GCMs)": more often used as: General Circulation Model or Global Climate Model.

Pag.3997:L3: "Given a number" to "Given the number"

Pag.4005: The use of the term "DM", for the differences of modelled and observed means is not very common. I would suggest using systematic bias, or just bias.

Pag.4009;L10: "(Fig.3)" should be (Table 3)

Fig.4: The overlap of 9 curves with daily time series make the results difficult to analysis. I would suggest applying some smoothing (maybe a 5/10 days running mean).

Pag:4017L20: "was also very model-dependent" - remove "very"