Anonymous Referee 2

General comments

The article "Simulations of future changes in thermal structure of Lake Erken: Proof of concept for ISIMIP2b lake sector local simulation strategy" presents impacts of changing climate on lake water temperature. The article is of very high interest and very well written, the work is thoroughly executed and discussion is relevant. The authors used a hydrodynamic lake model GOTM with 4 GCM/RCMs, using Generalized Regression Artificial Neural Network to disaggregate daily climate into hourly data. The GOTM model was able to reproduce observed lake temperature data for current time period (8 years). The model was then executed with climate forcing data from 4 GCM/RCMs

Response: We thank the Referee 2 for the positive comments about the text. The paper was edited very carefully and modifications and improvements were made. Below, we address every comment and explain the corresponding changes in the manuscript.

Specific comments

I would recommend expanding on the Methods section to provide more information that is critical in understanding the study, its aims, and results. It is unclear why the authors chose to consider 2006-2099 as the future even though the period begins 13 years ago. It is also unclear why this full period is evaluated without any consideration of the changes that occur from 2006 to 2099 based on the trend analyses also included in the manuscript. It seems that changes that can occur during this "future" period are considered representative of changes that will occur by 2099. The averages from this 94-year period are compared to averages from a 30-year period of 1975-2005. The variability during a 30-year period and during a 94-year period with a significant trend is expected to differ and this affects the projected changes. A more typical approach in many climate impact studies is to select two 30-year periods, one that represents a current climate (reference period, e.g. 1981-2010) and one that represents a future climate (e.g., mid-century 2041-2010 or late century 2071-2100). Forcing data from the same climate model would then be used as model inputs for both time periods; the difference between these results would represent the projected impact. It is also not clear from the manuscript how were the reference period values calculated for calculation of anomalies from the respective GCM/RCMs during the reference time period.

The results for the mid-century and late century should be added to the manuscript to evaluate how the change progresses; alternatively, the current results can be replaced with the late century period as that seem to be the focus of the "proof of concept" study.

It is also important to include information on the variability of the simulated thermal indices due to the climate model selection, i.e. present information for all 4 GCM/RCMs for the reference time period. That can give indication to the significance of the projected impact.

Response:

Climate impact studies can be approached in two ways: (1) assessing the difference in mean lake conditions (for example, mean surface temperature) between the reference periods and both mid-century and late-century (Woolway and Merchant, 2019) or (2) long-term trend analysis (O'Reilly et al., 2015; Shatwell et al, 2019; Moras et al., 2019). The use of anomalies or absolute values in trend analysis does not change the value of the slope. The use of anomalies in frequency distribution figures provide an alternative method of comparing the changes simulated by the future climate scenarios. So we consider it appropriate to combine both approaches. We have also added information about the reference period and the same analysis has been made to the meteorological variables in order to understand the variability of the projected thermal metrics derived from GCMs.

Changes in manuscript: Material and methods: section 2.8 Statistical analysis P9 L255-279, Results: section 3.3 Climate data projections P11 L324-343, section 3.4 Long-term modelled changes in thermal stratification P11-14 L344-420 and section 3.5 Comparison between long-term thermal metrics derived from daily and hourly climate data P14 L421-428. Discussion: P16-18 L500-580.
Changes in supplement: Sections S3-S5.

The information on the GOTM model for Lake Erken is very limited; the methods section should be expanded to include more details on the model structure, e.g. vertical resolution, inflow and outflow from the lake, etc.

Response: The GOTM model version 5.1 was used in this study. The meteorological parameters for running the model were air temperature (^{o}C), wind speed (m s^{-1}), short-wave radiation (W m^{-2}), cloud cover (dimensionless, 0-1), relative humidity (%), atmospheric pressure (hPa) and precipitation (mm day^{-1}or mm $hour^{-1}$). Inflows and outflows were not included in this study, and water level was considered fixed in the simulations. This version of GOTM did not have the ability to simulate lake ice, so for this study the inverse stratification period was not analysed. Moras et al., (2019), has shown that despite this limitation, the mode is able to accurately simulate water temperature and the phenology of thermal stratification during the remainder of the year. The initial conditions for water temperature were derived from a measured vertical profile. GOTM was run at hourly model computational time step, and simulated water temperature was saved as daily mean values each 0.5 m (42 layers).
Changes in manuscript: P5-6 L155-163.

Some relevant parts should be moved from Results to Methods (e.g., the beginning of section 3.1 and 3.3. Also, the periods used in the calibration (training) & validation periods for GOTM and GRNN should be put into the context between these two models. It is not readily apparent from the manuscript.

Response:
The beginning of sections 3.1 and 3.3 have been moved to 2. Materials and Methods section 2.5 Temporal disaggregation of meteorological forcing data 2.8 Statistical analysis respectively.
The GRNN training and validation periods do not fit into GOTM calibration periods. Putting these periods in context does not produce significant changes in the GRNN models performance (see table below) but it would entail a high computational cost since changing the GRNN models would require all the GCM scenarios to be disaggregated a second time and all GCM scenarios to be run again using these alternative data.

	Air temperature (^{o}C)			Relative humidity (%)			Wind speed (m s^{-1})			Short-wave rad (W m^{-2})		
	BIAS	RMSE	NSE	BIAS	RMSE	NSE	BIAS	RMSE	NSE	BIAS	RMSE	NSE
Training: 2006-2014	0.00	0.32	1.00	0.00	0.96	1.00	-0.01	1.15	0.74	0.00	8.39	1.00
Validation: 2015-2016	0.03	0.70	0.95	0.44	2.09	0.69	-0.07	2.50	0.60	0.08	18.15	0.86
Training: 2008-2012	0.00	0.26	1.00	0.00	0.79	1.00	-0.01	1.06	0.78	0.00	6.35	1.00
Validation: 2013-2015	-0.06	0.32	0.94	0.34	1.02	0.69	-0.01	1.37	0.58	-0.04	8.20	0.87

Changes in manuscript: P5 L147-154 and P9 L255-263.

It should be emphasized that training the temperature disaggregation algorithm on the current diurnal patterns means those current patterns will be projected to the future time series and any potential changes in diurnal pattern from the changing climate are ignored.
Response: GRNNs proved to be an effective method to disaggregate daily GCM forcing to an hourly temporal resolution for different weather variables such as air temperature, short-wave radiation, etc. However, GRNNs require a training phase, in which the diurnal patterns to be learned are presented to the network from historical meteorological measurements, and therefore if there are future changes in diurnal patterns, these cannot be reproduced. In addition, there is a high computational cost of disaggregating and storing the long-term daily climate data into an hourly data set.
Changes in manuscript: P17 L539-545.

Technical comments
Increases are given to 0.01 ^{o}C – what is the accuracy of the measurement and of the simulations? Is this accuracy adequate?
Response: The accuracy of thermocouple sensor is approximately ± 0.1 ^{o}C and can at times be somewhat better than 0.1. The simulated water temperature is given with 7 decimals. So two decimal places in the GOTM model performance are adequate, and match the best expected performance of our monitoring data.
L 68 –dimictic?
Response: Change made.

L89 Mean sea level(,) pressure(,) relative humidity and precipitation were measured – missing commas?
Response: Change made.

Section 2.6. It would be useful to include model performance for other thermal indices used for evaluation of change, e.g. duration of thermal stratification
Response:
Response: The model performance for the duration, onset and loss of stratification has been added to the section 3.2 Lake Model performance (Table 4).

L 162: Schmidt's stability – needs a reference/ brief explanation
Response: The following Schmidt stability definition was added: resistance to mechanical mixing due to the potential energy inherent in the density stratification of the water column (Schmidt, 1928; Idso, 1973).
Changes in manuscript: P6 L190-191.

L 231: this model handicap and any other should be described in section 3.2
Response: The GOTM model version 5.1 did not have the ability to simulate lake ice, so for this study the inverse stratification period was not analysed. Moras et al., (2019), has shown that despite this limitation, the mode is able to accurately simulate water temperature and the phenology of thermal stratification during the remainder of the year. A new GOTM model version 5.4 with ice-module was released after this project was submitted, allowing to evaluate the effect of the lack of ice module on the onset of the direct stratification. The onset of direct stratification was derived from simulations of water temperature with GOTM version 5.1 and 5.4 from 2006 to 2016. The RMSE between the onset of direct stratification from GOTM version 5.1 and 5.4 was 5.22 days showing a slight impact the lack of ice-module on the onset of the direct stratification.

onset of direct stratification	
GOTM v5.1	GOTM v5.4
2007-04-27	2007-04-16
2008-04-27	2008-04-26
2009-04-27	2009-04-27
2010-05-01	2010-05-13
2011-04-24	2011-04-25
2012-05-03	2012-05-01
2013-05-09	2013-05-09
2014-04-21	2014-04-21
2015-05-22	2015-05-22
2016-05-04	2016-05-03

Changes in manuscript: P6 L159-162.

L 312-314 However, the dominant wind (is) along : : : missing word?
Response: Change made.

L 322, 324 When GOTM was forcing with : : : forced?
Response: Change made.

L 350: it would be good to put the statement into context; what kind of changes can be expected with these increases in temperature?
Response: The expected changes in the lake ecosystem caused by an increase in water temperature have been moved from 1. Introduction to 4. Discussion section.

Figure 2 heading: figure shows calibration as plots a, c, e, and g, but the caption says these are validations. I would recommend including a similar plot but with model residuals (perhaps in Supplementary materials); that would make any differences much easier to see especially on the timing.
Response:
Figure 2 has been renumbered and a subtitle added to each subfigure.
Figure S7 has been added to the supplementary material showing the differences between simulated (when the lake model was forced with daily, hourly and synthetic hourly meteorological forcing data) and observed water temperature.

Figure 3: it would be helpful if the scale on y axis with the same units had the same range (a-d)
Figure 3 has been removed because GOTM model performance had been shown in twice (Figure 3 and Table 4).

References:
Idso, S. B.: On the concept of lake stability, Limnol. Oceanogr., 18, 681–683, 1973.
Moras, S., Ayala, A. I., and Pierson, D. C.: Historical modelling of changes in Lake Erken thermal conditions, Hydrol. Earth Syst. Sci., 23, 5001–5016, https://doi.org/10.5194/hess-23-5001-2019, 2019.
O'Reilly, C., Sharma, S., Gray, D. K., Hampton, S. E., Read, J. S., Rowle,y R. J., Schneider, P., Lenters, J. D., McIntyre, P.B., Kraemer, B. M., Weyhenmeyer, G. A., Straile, D., Dong, B., Adrian, R., Allan, M. G., Anneville, O., Arvola, L., Austin, J., Bailey, J. L., Baron, J. S., Brookes, J. D., de Eyto, E., Dokulil, M. T.,

Hamilton, D. P., Havens, K., Hetherington, A. L., Higgins, S. N., Hook, S., Izmest'eva, L. R., Joehnk, K. D., Kangur, K., Kasprzal, P., Kumagai, M., Kuusisto, E., Leshkevich, 20 G., Livingtone, D. M., McIntyre, S., May, L., Melack, J. M., Mueller-Navarra, D. C, Naumenko, M., Noges, P., Noges, T., North, R. P., Plisnier, P. D., Rigosi, A., Rimmer, A., Rogora, M., Rudstam, L. G., Rusak, J. A., Salmaso, N., Samal, N. R., Schindler, D. E., Schladow, S. G., Schmid, M., Schmidt, S. R., Silow, E., Soylu, M. E., Teubner, K., Verburg, P., Voutilainen, A., Watkinson, A., Wiliamson, C. E., and Zhang G.: Rapid and highly variable warming of lake surface waters around the globe, Geophys. Res. Lett., 42, 10773–10781, https://doi.org/10.1002/2015GL066235, 2015.

Schmidt, W.: Über Temperatur und Stabilitätsverhaltnisse von Seen, Geogr. Ann., 10, 145–177, 1928.

Shatwell, T., Thiery, W., and Kirillin, G.: Future projections of temperature and mixing regime of European temperate lakes, Hydrol. Earth Syst. Sci., 23, 1533–1551, https://doi.org/10.5194/hess–23–1533–2019, 2019.

Woolway, R.I., and Merchant, C.J.: Worldwide alteration of lake mixing regimes in response to climate change, Nature Geoscience, 12, 271–276, https://doi.org/10.1038/s41561–019–0322–x, 2019.