Anonymous Referee 1

General comments

The manuscript entitled "Simulations of future changes in thermal structure of Lake Erken: Proof of concept for ISIMIP2b lake sector local simulation strategy" aims to assess the impacts global warming on the thermal characteristics of Lake Erken. Since future projections of global warming are at a daily time step, the authors first analyses the need to disaggregate the input variables to the hourly time step. The manuscript is well written. The topic is scientifically relevant.

Response: We thank the Referee 1 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

Line 70-71: "It is the lake's relatively shallow depth and large surface area, which leads to large inter-annual variability in the timing and patterns of thermal stratification." Why is this? Perhaps explain in one or two sentences how this works and why this is different for relatively deep lakes or lakes with a small surface area.

Response: Mixing and stratification change in response to lake morphometry. Shallow lakes have lower heat storage, responding more directly to short-term variations in the weather conditions and heat can be transferred through the water column by wind mixing (Magee and Wu, 2017). However, deep lakes required greater wind speeds to complete the mix. Large surface areas or fetch increase the effects of mixing and vertical transfer of heat to the bottom (Rueda and Schladow, 2009).

Changes in manuscript: P3 L74-76.

Line 138-139: "More detailed description of the GRNN methods and models are given in the supplementary material to this paper." I was hoping to find equations on how the GRNN model calculates hourly estimations based on daily input, however, I could not find a detailed description of the GRNN methods in the supplementary materials.

Response: GRNN description was added in the supplement material section S1.1.

Changes in supplement: P1 L9-34.

Line 158: "Schmidt stability", perhaps give a definition or equation of the Schmidt stability.

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.

Line 170-174: "Air temperature, short-wave radiation, relative humidity and wind speed were temporarily disaggregated into hourly values from mean daily data, using the GRNN models. A database was constructed using 8 years of measurements. From this whole set of data, the first 5-years of data, that is, from 2008 to 2012, were used for training, and 3-years of data from 2013 to 2015 were used for validating the results obtained." This sentence was confusing. After reading the methods section I first assumed this was about the calibration/validation of GOTM. However, later I realized it was about the calibration/validation of GRNN. I would expect these sentences in the methods section. Moreover, it would be good to mention clearly that there are two types of calibration/validation: that of GOTM and that of GRNN.

Response*:* It has been moved from Results section 3.1. Hourly meteorological modelling to Material and Methods section 2.5. Temporal disaggregation of meteorological forcing data.

Changes in manuscript: P5 L147-154.

Line 192-193: "Temperature simulations for the validation period were more accurate (average RMSE of 0.66 ^{o}C and NSE of 0.97) than for the calibration period (average RMSE of 0.95 ^{o}C and NSE of 0.94), but in both periods the model performance was considered acceptable." I would expect that the validation period would be less accurate than the calibration period. Therefore, my first thought was then that perhaps the legend was swapped between calibration and validation. Yet, the authors later mention that this is "due the higher variability in observed water temperature during the long calibration period." (Line 284-285). Then the question raises, which data set is more representative? Was the high variability during the calibration period actually quite normal and the validation period exceptionally uniform? And what does this mean for the validity of the output?

Response: Water temperature simulations were apparently more accurate for the validation period (2015-2016) than for the calibration period (2006-2014), which may appear unusual, but is due to the higher variability in observed water temperature during the longer calibration period. Years with a longer duration of stratification and stronger stability, generally had higher simulation errors. Half of the eight-year calibration period exhibited these conditions, while the two-years used for validation both exhibited shorter duration of stratification and weaker stability.

	year	RMSE (^{o}C)			thermal stratification			Schmidt stability (J m^{-2})
		24h met	1h met	synthetic 1h met	duration (days)	onset	loss	
Calibration	2007	0.58	0.59	0.83	23	176	230	17.42
	2008	**1.42**	**1.13**	**1.04**	**103**	**124**	**227**	**31.52**
	2009	0.75	0.68	0.63	69	122	242	35.17
	2010	**1.10**	**0.92**	**0.99**	**111**	**139**	**254**	**80.77**
	2011	**0.92**	**0.79**	**0.81**	**90**	**152**	**252**	**43.77**
	2012	0.71	0.66	0.77	38	141	244	32.98
	2013	**1.42**	**1.52**	**1.08**	**124**	**129**	**259**	**79.48**
	2014	0.83	0.73	0.79	55	137	263	52.40
Validation	2015	0.59	0.66	0.65	71	162	240	17.60
	2016	0.69	0.73	0.71	67	173	239	47.25

Changes in manuscript: P14 L433-438.

Line 202: "As would be expected the simulations of bottom temperature were slightly less accurate" Why would this be expected?

Response: Higher errors were found at the lowest depth (15 m depth), part of this might have been caused by the presence of internal seiches in lake Erken which cannot be reproduced by 1D models such as GOTM."

Line 349-351: "Combined these results suggest important changes in the factors affecting lake biogeochemistry directly through changes in temperature and indirectly by influencing the availability of light and nutrients." The presented results only indicate an increase in temperature and stratification period. Since the presented data does not show how this affects biogeochemistry and the availability of light and nutrients, could the authors be a bit more specific on this in the conclusion? How do the authors think/speculate it will change (perhaps refer to the introduction where a short explanation is already given)?

Response: As mentioned in the introduction the projected changes in thermal stratification can influence many aspects of the lake ecosystem. Increases in thermal stability and duration of stratification can intensify hypolimnetic oxygen depletion (Foley et al., 2012; Schwefel et al., 2016) and hence induce enhanced internal phosphorous loading (North et al., 2014), increase the release of dissolved iron and manganese from sediments (Schultze et al., 2017) and also increase methane emissions (Grasset et al., 2018). Warming lake temperature affects biological rates of metabolism, growth and reproduction (Rall et al., 2012) and can promote cyanobacterial blooms (Paerl and Paul, 2012). When coupled to a reduction in oxygen-rich water, warming water temperature leads to a lower fish populations (O'Reilly et al., 2003; Yankova et al., 2017). Increase in evaporation associated with warming can lead to declines in lake water level (Hanrahan et al., 2010) with implications for water security. So these changes are expected in Lake Erken.
The expected changes in the lake ecosystem caused by changes in thermal stratification have been moved from 1. Introduction to 4. Discussion section and so our conclusions can be more understandable.
Changes in manuscript: P18-19 L581-588.

Technical corrections

Figures in general; 1) it would be good to have comparable axes per figure. For example, figure 3a has a y-axis going from 0-1.2 ^{o}C, while figure 3b goes from 0-0.8 ^{o}C. I suggest that the authors uniform the axis and perhaps use the normalized RMSE to compare the different subfigures 2).From the figures caption, it is not always clear if the predicted output is with GOTM or with GRNN. Perhaps include this information in the figure's caption. General: sometimes I read "wind_factor" and sometimes "wind factor" without "_". Is there a difference in meaning?

Response:
Figure 3 has been removed because GOTM model performance had been shown in twice (Figure 3 and Table 4).
GRNN and GOTM has been added to the figure captions to indicate if models have been used to disaggregate meteorological forcing data or to simulated water temperature.
Wind factor is the meaning of the parameter wind_factor.

Line 246-247: "Simulated changes were generally slight less for the simulations driven by daily forcing data as shown by the figures in parentheses". Put a dot after parentheses and change "slight" to "slightly"
Response: Change made.

Line 284: "were more accurate than for the calibration period (2006-2014) due the higher variability in observed water temperature" add "to" after "due".
Response: change made.

Figure 2: 1) the caption says that validation is figure 2a, 2c, 2e and 2g, however, the title of the figures suggest that validation is figure 2b, 2d, 2f and 2h. This is confusing. 2) Perhaps include the words "observations", "daily data", "hourly data", "synthetic hourly data" on the left side of/ or under the figure. It is now quite a puzzle to find which subfigure tells what. 3) Perhaps also include a difference graph where the difference between "observations" and respectively "daily data", "hourly data", "synthetic hourly data" is shown. From figure 2, it is now hard to see the differences. (The same holds for figure 4, where it is hard to see the differences between historical and the rcp's)

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 S14 has been added to the supplementary material showing the differences between the historical and RCP 2.6 scenarios, and the historical and RCP 6.0 scenarios for the IPSL-CM5A-LR projection (when the lake model was forced at daily resolutions).

Figure 5 and 6: In figure 5i, the authors indicate the words “deeper” and “shallower” with arrows. This really increases the readability of that specific subfigure and the same would help the reader in all other subfigures.
Response: In Figures 6 and 7, the arrow and the words “deeper” and “shallower” have been removed for easy viewing of the figure. However, in the figure caption was added the meaning of values greater or less than 0 of each of the thermal indices: changes in thermal metrics greater than 0 show an increase and lower than 0 show a decrease.

References:
Idso, S. B.: On the concept of lake stability, Limnol. Oceanogr., 18, 681–683, 1973.
Magee, M. R., and Wu, C. H.: Response of water temperatures and stratification to changing climate in three lakes with different morphometry, Hydrol. Earth Syst. Sci., 21, 6253-6274, https://doi.org/10.5194/hess-21-6253-2017, 2017.
Rueda, F., and Schladow, G.: Mixing and stratification in lakes of varying horizontal length scales: Scaling arguments and energy partitioning, Limnol. Oceanogr., 54, 2003-2017, https://doi.org/10.4319/lo.2009.54.6.2003, 2009.
Schmidt, W.: Über Temperatur und Stabilitätsverhaltnisse von Seen, Geogr. Ann., 10, 145–177, 1928.