Authors' responses to comments of Referee 1.

Referee 1 – Page 3, line 10. Daily precipitation was used in driving the model, while the other six datasets were put into the model as hourly resolution. This sounds strange to me. Are different climate variables allowed to put into the model with different temporal resolution?

Authors' response. The seven climatic parameters used to drive the models are grouped into three input datasets: a *meteo_file*, which contains wind speed, air pressure, air temperature, cloud cover and relative humidity data; a *swr_file* in which shortwave radiation data are stored and a *precip_file* where precipitation data are located. Within the same dataset, the parameters must have the same time-resolution, but it is possible each of the three datasets to differ in time-resolution. GOTM model allows to set a factor that converts the unit of measurement used in the the *precip_file* input (in our case mm/day) into the unit of measurement used in GOTM for precipitation (m/s). This possibility gave us the chance to use the most suitable time resolution for precipitation in our study, since no weather station around Lake Erken measured precipitation on hourly basis. For our long-term simulations we presented in our paper, we assume a constant water level. Therefore, precipitation had only minor effects on the model output.

Referee 1 – Page 4, line 30. Why the measured water temperatures with 30 minutes resolution were averaged to daily, not the hourly mean values for the model calibration? In this way, the diurnal variation of the water temperature is missing. Could you give an explanation here?

Authors' response. This is a good point and we are aware that using hourly values for model calibration would have taken into account the diurnal variation of water temperature. Our choice to average 30 minutes water temperature to daily values have been made by the fact that a calibration using hourly values was computationally too intensive. We set ACPy to run 10000 simulations to obtain the best parameter set. We calibrated the model using a daily water temperature dataset of 94244 data points. This process takes ~24 hours using daily values. The use of hourly data for model calibration would have been a very time-consuming process. In addition, most of the metrics of change in thermal structure used in our paper were most conveniently calculated using mean daily data. Therefore, we felt that it would be most appropriate to develop model calibration based on mean daily output.

Referee 1 – Page 5, line 3. I am afraid the wind factor of 1.28 is a little bit high, since wind is measured in or quite close to the lake (based on Figure 1). Could you explain why you use such a high wind factor here?

Authors' response. There are two possible explanations here. First, the dominant wind speed (ws) direction is along the longest east-west fetch of Lake Erken that is ~10 km as opposed to the north – south fetch that is only 2-3 km. The 1D model input for wind is only a mean velocity and does not account for the effects of fetch. Given that wind is often blowing along the longest fetch that would have that would have the greatest effect on the measured temperature measurements used for calibration at the Eastern end of the lake, it is reasonable to expect an elevated wind factor. Secondly, it is actually the wind speed cubed that is used in the model equations that effect turbulent mixing. Under variable and gusty conditions cubing the mean hourly wind speed calculated by our data logger measuring at 1 minute intervals

$$\left(\frac{\sum_{60}^{1} ws}{60}\right)^3$$

may underestimate the true effects of wind which would more properly be calculated as the mean of of all cubed wind speed measurements made during the hour.

$$\left(\frac{\sum_{60}^{1} ws^3}{60}\right)$$

This effect would also result in an elevated wind factor.

Referee 1 – Page 6, line 1. how did you define the thermocline depth in the study? As I know, there are two ways in defining the thermocline depth in rLakeAnalyzer (i.e. seasonal=TRUE/FALSE). The results, from the two approaches, are different (see " Read, J. S., Hamilton, D. P. P., Jones, I. D., Muraoka, K., Winslow, L. A., Kroiss, R., Wu, C. H. & Gaiser, E. (2011). Derivation of lake mixing and stratification indices from high resolution lake buoy data. Environmental Model ling and Software 26:1325 1336 ")").

Authors' response. We did not specify which condition I used to define thermocline depth in our R code. However, not specifying any condition as we did gives the same result of the condition "seasonal = TRUE". Yours is a valuable observation and we will better state this in our methodology.

Referee 1 – Page 9, line 4. As stronger evidence for such changing trend, could you also use the measured water temperature to do a Mann Kendall test? In the paper, all the statistical test s are based on the simulated temperature, it is better to prove the simulated trend also based on the temperature, it is better to prove the simulated trend also based on the measured values. If it takes you so much time to do this work for all the three cases (i.e whole lake, epilimnion and hyplimnion), I recommended to test the observed trend for the summer epilimnion because the simulated temperatures of the layer significantly increased in the whole period.

Authors' response. Even though Lake Erken has a relatively long measured water temperature record compared to other lakes, there are still significant data gaps within the dataset. There were several years with no (or very few) measured temperature before the deployment of the automatic floating station in 1988. There are significant data gaps in Erken temperature record after 1988 as well, during the maintenance/failures of the floating station for example. Since our trend analysis is based on seasonal means, performing a trend analysis on measured water temperature with several missing data would have made our results unrealistic. Having such data gaps in our water temperature record is actually the main reason why we developed the approach described in this study in order to get a more consistent and reliable water temperature historical record using a hydrodynamic model.

Referee 1 – Page 12, line 7. I am confused here, you said that the summer epilimentic temperature significantly increased for the whole period, but not significantly increased in two sub intervals? To me, it sounds like a paradox. Please check it.

Authors' response. When Mann-Kendall test is performed on the two sub-intervals (1961-1987 and 1988-2017) of summer epilimnetic temperature, positive trends are detected but they are not significant. This means that the two sub-intervals are too short to detect a significant trend. Indeed, when the trend test is performed on the entire study period (1961-2017) the summer epilimnetic temperature shows a significant increasing trend. From our results, we can infer that summer epilimnetic temperature was subjected to a slower but more stable warming compared to, for example, spring and autumn epilimnetic temperature, which showed a more abrupt increase in water temperature in the most recent sub-interval (1988-2017).

Referee 1. Also, as shown in Blenckner 2002, Lake Erken is always ice covered for the whole winter and the ice melts between March and early May. It is a weak point to use GOTM, without an ice module, to simulate such a lake with a long ice duration. I suggest adding some sentences, in this part, to clarify this limitation. Considering the future model development, it is a valuable work to include ice part into GOTM which could also be added into the Discussion.

Authors' response. GOTM developers are currently working on integrating GOTM with an ice module, but this was not available for this work. The GOTM model used for the simulations documented here did not have a functioning ice model, but instead cut off surface heat exchange when the simulated surface water temperature became negative. This provided a very simple way to make continuous simulations that include freezing conditions that would normally lead to the formation of ice. However, the temperature profiles during winter were not realistic, and could not be used for model calibration. This can be seen in figures 1-2 (below) where a comparison between simulated an observed water temperature at 1m and 15 m depth is reported for year 2009. At 1 m depth, simulated and observed temperature are rather similar throughout the entire year. However, at 15 m depth, the model does not take into account the heat loss from sediment during ice-cover, which cause an increase in bottom water temperature. During winter, there is a clear mismatch between simulated and observed water temperature. For this reason, all data collected between 1 December - 31 March are excluded from the temperature data used for model calibration and only data between 1 April and 30 November are used for model calibration. Yours is a valuable comment and we will better clarify this limitation in our Discussion.

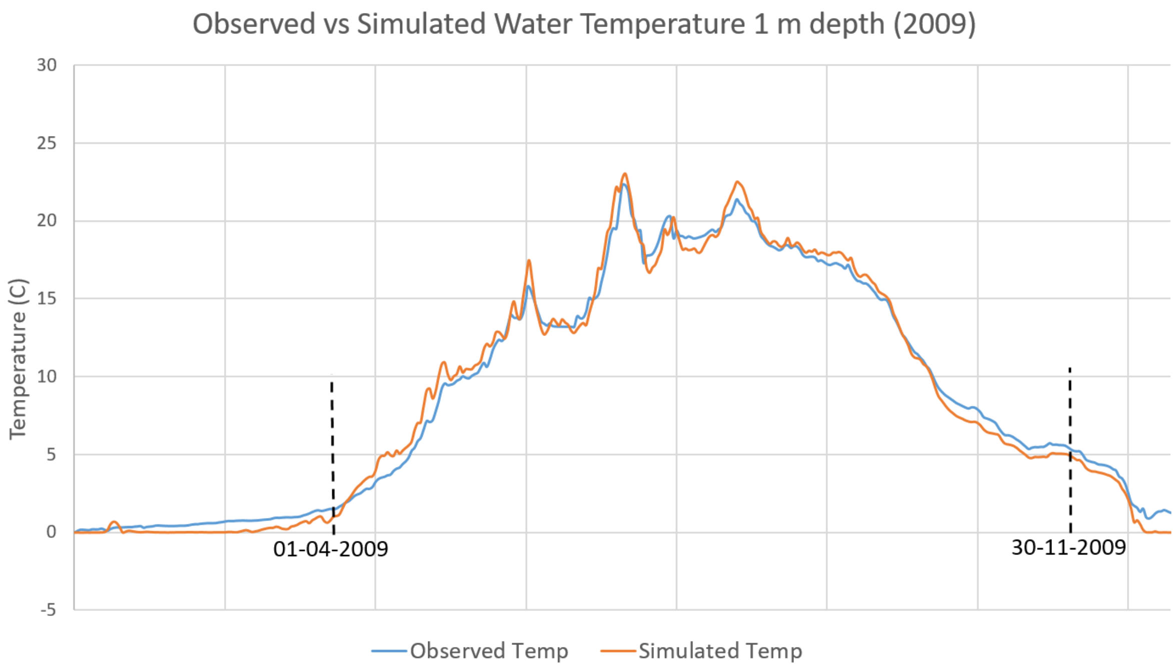

Figure 1

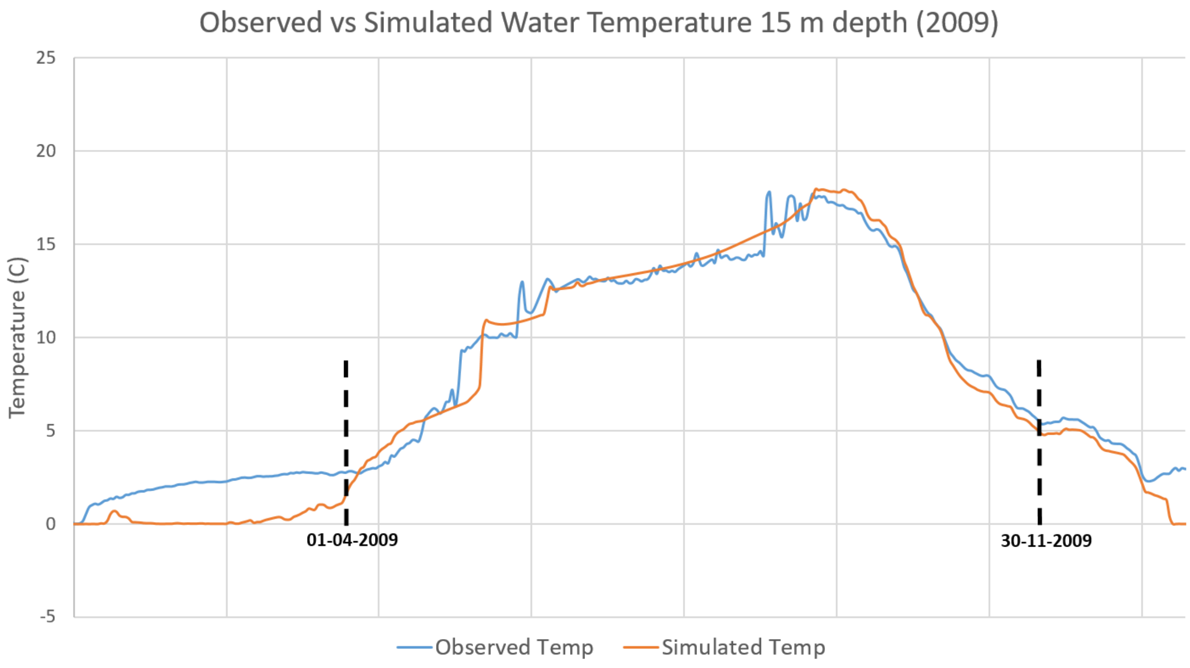
Observed vs Simulated Water Temperature 15 m depth (2009)
Temperature (C)
25
20
15
10
5
0
-5
01-04-2009
30-11-2009
Observed Temp
Simulated Temp

Figure 2