Overall, this paper does not present any breakthrough insights or new ideas; however, it does present a useful analysis of a fairly unique data set – three paired lakes in the same ecoregion for which over 100 years of data are available, and for which weather forcing is similar but differences in lake morphometry are clear. The modeling approach itself is not particularly novel, but, to my knowledge, it has not previously been applied to this rich dataset. The principal study question regarding the relative influence of morphometry vs. climate forcing is of active interest and is elucidated by application to lakes with such rich data sets. Accordingly, I recommend publication with minor revisions.

The paper is generally well written and is a logical follow-on to an earlier paper in HESS (Mage et al., 2016, HESS 20:1681, doi:10.5194/hess-20-1681-2016). There are a few items that may require further attention, as described below; however, the paper is generally acceptable. Therefore, I recommend publishing subject to minor revisions as described below.

The submitted paper appears to have gone through an extensive and chequered review process. In the latest iteration, one reviewer was generally favorable and the other rather negative as to publication. The negative review (Reviewer #3) was largely based on critique of the selection of a 1-D model for the comparison. I do not find this to be a compelling reason to reject the paper. Indeed, 1-D (vertical) models may be preferable when looking to isolate the general impacts of external forcing on temperature trends. The potential problems with 1-D models are (1) influence of tributary inflows and (2) representation of mixing due to wind fetch. Item (1) is not a big issue for these lakes, which are natural lakes with large groundwater inputs. Item (2) is a potential concern, but it all depends on how well the 1-D model represents wind driven eddy diffusivity. In my opinion, this aspect of the paper would be fully acceptable with the addition of a few lines that address the details of how wind-driven mixing is addressed in the 1-D representation. The authors' response to Reviewer #3 that 3-D models are computationally expensive is correct, but not sufficient to answer this criticisim.

The three lakes included in this study have somewhat similar perimeter to area ratios, so the major obstacles to the 1-D approach of comparing lakes with very different wind fetch to area ratios will not be encountered here. The lakes in question are natural, seepage-dominated lakes, so enhanced diffusion due to inflow temperatures should not be a major issue. The authors should be clear that the results may not apply to individual lakes elsewhere depending on their specific configuration.

Wind-driven mixing is a bigger issue for 1-D models. Because DYRESM documentation is not readily available (see below) the authors should expand a small amount on this issue. Specifically, some notes on how wind-driven eddy diffusivity is represented should be supplied, along with any validation data to confirm the 1-D representation. I am more familiar with Hostetler-based 1-D lake models[1,2] in which wind-driven eddy diffusivity is estimated as a function of 2 m windspeed, the Brunt-Väsäilä frequency implied by the lake density-gradient, and the Ekman decay as a function of latitude[1]. These 1-D formulations generally require an expression of "enhanced" diffusivity to account for sources of turbulence not represented in the base 1-D formulation. Some additional explanation of why the 1-D

[1] Hostetler SW, Bartlein PJ. (1990) Simulation of lake evaporation with application to modeling lake level variations of Harney-Malheur Lake, Oregon. *Water Resour Res* 26: 2603-2612, doi:10.1029/WR026i010p02603.
[2] Subin ZM, Riley WJ, Mironov D. (2012) An improved lake model for climate simulations: Model structure, evaluation, and sensitivity analyses in CESM1. *J Adv Model Earth Syst* 4: M02001, doi:10.1029/2011MS000072, 2012

formulation is appropriate would be useful here (for instance, on p. 5 of the current draft). (See also Fang and Stefan[3] for arguments in favor of the 1-D approach.)

On the other hand, 3-D lake models are indeed computationally expensive, and the extra precision does not necessarily lead to greater accuracy. A 3-D model requires data at multiple points in space and time for calibration. When these data are lacking calibration of a 3-D model is often not well constrained and subject to over-fitting. Thus, a 1-D model can be preferable for answering questions about long-term trends and forcing factors.

Additional comments that should be addressed include the following:

- Reviewer #2 correctly noted that humidity, cloud cover, solar radiation all influence lake response, in addition to surface area and air temperature. To this list, precipitation regime could also be added, as large direct precipitation inputs can have a significant impact on stratification stability. The authors added a reasonable discussion of these issues. However, I think the main point that should be added is that the study addresses three lakes in the same ecoregion with similar climate forcing, so differences in responses relate primarily to morphometry or general climate perturbations. Another potentially important factor is water clarity, which determines how solar radiation is vertically partitioned. The authors mention some of these issues in their revision, but should present more discussion. In particular, Table 1 presents Secchi depth as a constant for each of the lakes, but is there any evidence on how this may have changed over time?
- The DYRESM model is a useful formulation. However, it is also somewhat problematic as changes in the Australian scientific establishment have resulted in the deletion of most all links to the DYRESM code and documentation. It is not immediately clear how to obtain the code today. Authors should include a note about the availability of DYREMS – and, if possible, provide a link for access to the DYRESM code as adapted for ice cover.

Line-by-line specific comments:

10. (abstract) This posits an effect from "decreasing wind speed" – but is decreasing wind speed really a known for these lakes? If it is, it needs to be stated in the abstract.

13 (abstract) Make clear what is inferred from data vs. from modeling

17. (abstract) "larger lakes have more variability": This needs to be qualified as to what is mean by "larger lakes". You are comparing three relatively small lakes. The conclusions likely do not apply to Lake Superior.

p. 1, 24.Text states that land and ocean surface temperature anomaliies increased from 1850 to 2012. An anomaly is a departure from an expectation, so you need to state what basis is used for the anomaly assessment.

p.3, 26: "Large surface areas increase the effects of vertical wind mixing…" Isn't it a ratio of wind fetch to depth that is more important here?

---

[3] Fang X, Stefan HG. (1996) Development and validation of the water quality model MINLAKE96 with winter data, Project Report 390, St. Anthony Falls Laboratory, Univ. of Minnesota.

p.5,line 28: Provide a reference for the assumed value of Ksed.

p. 7, line 16: Text implies that meteorological data are entered into the model on a daily basis. Is this true? If so, explain how daily cycles of heating and cooling of the epilimnion and their effect on vertical stability are incorporated into the model.

p. 8, line 24: I agree with the prior reviewers in having some discomfort here about the discussion of adjusting "the minimum water level thickness" as a calibration parameter. This may be warranted in terms of finding an appropriate minimum thickness that correctly resolves the thermocline position, but needs to be better explained. If is also unclear why it is appropriate to choose "one minimum layer thickness" for all three lakes.

p. 9, line 25-29: Perturbation tests examine response to changes in air temperature and wind speed. These tests assume "the water balance is maintained." This seems unlikely if increased air temperature leads to increase ET. However, that is the nature of one-parameter perturbation tests – but, the limitation should be acknowledged.

p. 10, line 5: Air temperature "showed a significant change in slope" – based on what test at what significance level?

p. 10, line 15: For NS efficiencies, state the time basis (daily?)

p. 10, line 23: "lake 1980s" should presumably be "late 1980s."

Section 3.5: Presents surface heat fluxes. Are there any direct measurements to validate these estimates?

Section 3.6: Presents correlations between "lake pairs" for the energy balance. These are simulated results, right? If so, how much of the apparent correlation is due to model formulation vs. reality?

p.13, line 1: Discussion of stratification onset: Is this based on observations or modeled results? If modeled, how does the model compare to observations?

p. 15, line 24: Missing word in "correlations air temperature".

p. 16, line 3: Discusses results of Woolway et al. indicating that influence of air temperature increases was minimal relative to wind speed in determining the length of stratification period. Isn't this result dependent on the surface area-to-volume ratio of the specific lake in question? Also, the length of stratification may not be the key result as the onset of stratification may determine the relative temperatures in the epilimnion and hypolimnion over the summer.

p. 17, line 15-20: Discussion of heat fluxes: Is this entirely based on model results? Is there any observational evidence to confirm magnitude of heat fluxes?

Table 4: Estimated trends need a time dimension – e.g., express in per year metrics. Also, clarify what test is used to define significance of trend.

Table 5: Clarify that correlation is based on modeled results (?). If so, how do the modeled results compare to observations?

Figure 3: Results appear “cleaner” for Lake Wingra because it is shallow and more closely tied to air temperature. For Mendota and Fish Lake is there further insight to be gained by plotting simulated vs. observed temperature by depth range or by season? Are larger apparent discrepancies caused by mis-estimation of the depth of the thermocline? Also, put correlation coefficients on the X:Y plots.