

We thank the reviewer for the valuable suggestions and feedback to improve the manuscript. We have provided point-wise responses to the reviewer’s comments in blue text, followed by our replies in black text.

1) Model time step and computational comparison

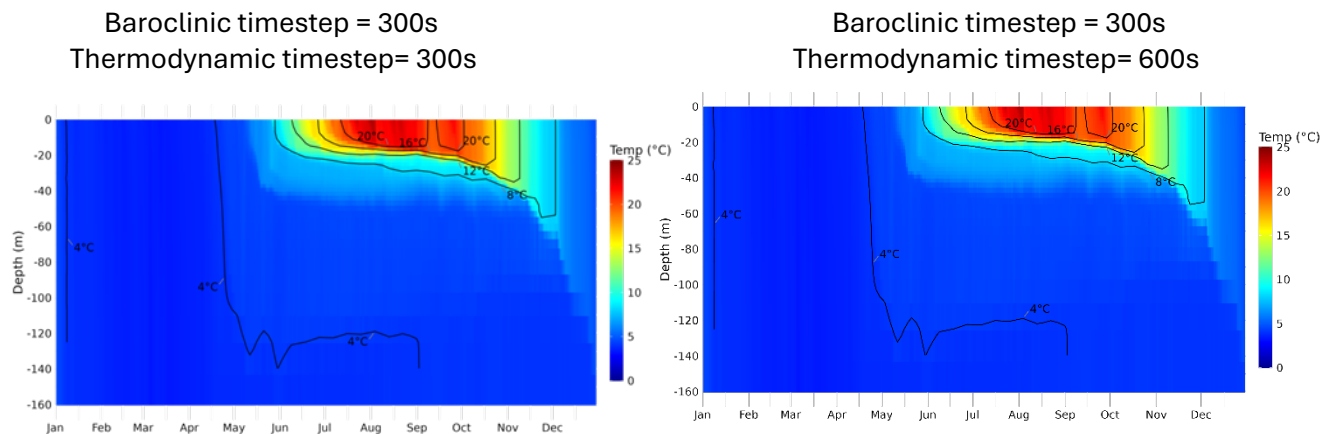
The time step used in the model configuration appears relatively large. As stated in the manuscript, LMH-FVCOM (with 100 m–2.5 km resolution) uses a time step of less than 10 s. Have the authors conducted any time-step sensitivity experiments to evaluate its impact on model stability and accuracy?

Thank you for this insightful comment. We acknowledge that the time step used in our configuration is larger than that typically used in the LMH-FVCOM model. The difference primarily arises from the numerical formulation and grid structure of the two models.

LMH-FVCOM model employs an unstructured triangular grid and a mode-splitting time-integration scheme, which requires relatively small-time steps to satisfy the Courant–Friedrichs–Lewy (CFL) stability criterion, particularly in regions with very fine spatial resolution (e.g., ~100 m). Consequently, FVCOM simulations commonly use time steps on the order of a few seconds to maintain numerical stability.

In contrast, MOM6 uses a structured finite-volume discretization on an Arakawa C-grid and incorporates split-explicit time stepping and separates barotropic–baroclinic timesteps from tracer advection, thermodynamics, mixing, and coupled ocean biogeochemistry. The grid resolution used in this study (kilo meter scale) is also coarser than the finest FVCOM resolution mentioned above. These factors allow MOM6 to operate with larger time steps while still satisfying stability constraints and maintaining numerical accuracy.

Sensitivity experiments were conducted by setting the baroclinic timestep at 300s and varying the thermodynamic time steps between 300s and 600s to check model stability and impact on vertical thermal structure. Both the configurations remained stable and resulted in similar vertical thermal structure (at an offshore location in southern Lake Michigan; shown in Figure 4) as shown below. So, in the current study, we adopted a baroclinic timestep of 300s and thermodynamic timestep of 600s.



Furthermore, the comparison of processor hours between the two models (lines 554–560) is not meaningful if different time steps are used. A fair comparison would require simulations with similar time-step settings. The authors are encouraged to reconsider this comparison or perform additional experiments.

This is an important point, thanks for the comment. We agree that a strictly controlled comparison of computational performance would ideally require simulations with identical numerical settings, including comparable time-step configurations. However, the allowable time step in the two hydrodynamic models is strongly constrained by the numerical formulation and grid structure, as explained in the previous response.

Because the stable time step is intrinsically linked to each model's numerical formulation and grid structure, using identical time steps across the two models would not reflect their respective stable configurations. Our intent was not to provide a strict algorithmic benchmark, but rather to discuss the numerical characteristics of the two modeling approaches. Lines 554-560 have been revised to read:

“The allowable time step in hydrodynamic models is strongly constrained by the numerical formulation and grid structure. MOM6 uses a structured finite-volume discretization on an Arakawa C-grid with a split-explicit time-stepping approach that separates barotropic and baroclinic modes. This formulation allows MOM6 to maintain numerical stability with larger baroclinic time steps while preserving the simulation accuracy. In contrast, LMH-FVCOM employs an unstructured grid with a mode-splitting time-integration scheme, which requires relatively smaller time steps to satisfy the Courant–Friedrichs–Lewy (CFL) stability criterion, particularly in regions with very fine nearshore resolution. This reflects the much shorter time steps required to maintain stability with high nearshore resolution and the added computational complexity of finite volume calculations (Zhang et al., 2023). As a result, MOM6 regional configurations can operate with larger time steps while maintaining stable integrations, which can lead to improved computational efficiency compared to models that require much smaller time steps to satisfy stability constraints.”

2) Formatting issue

Please correct the font inconsistency on line 134.

This is fixed in the revised version.

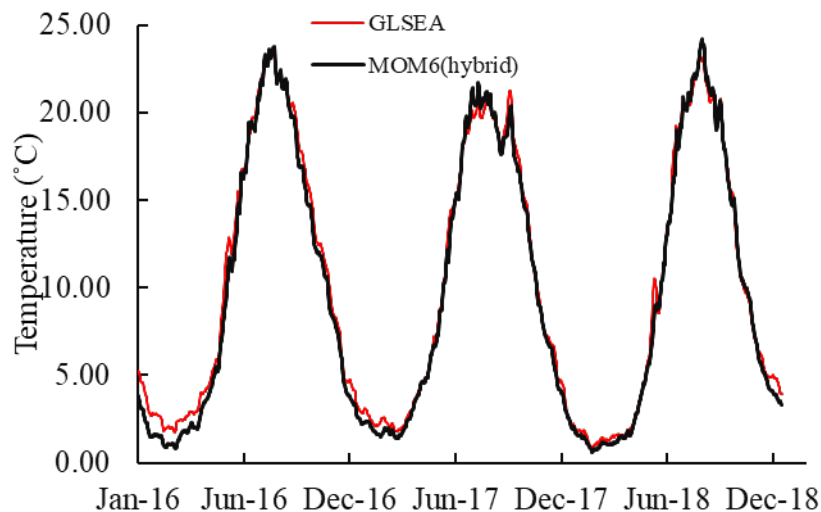
3) Spin-up time

The specified spin-up duration appears short for a lake-scale circulation model. The authors should justify the chosen spin-up time, ideally by presenting LST evolution plots or comparisons with observational products to demonstrate model stabilization.

Thanks for the thoughtful comment. Lakes Michigan and Huron are dimictic lakes, meaning they mix from top to bottom twice a year. These lakes experience an "overturn" in the spring and again in the fall as water temperatures pass through 4 ° C (the temperature of maximum density for freshwater). These seasonal overturns promote full water column mixing, which helps the system adjust relatively quickly to model forcing conditions. Considering the dimictic nature of the lakes and the associated

complete mixing during overturn periods, the “initial condition memory” for the lake is generally assumed to be ~1 year (Xue et al., 2017). This is a common practice for Great Lakes hydrodynamic models, as described in many recent studies (e.g. Anderson et al., 2018; Cannon et al., 2024; Xue et al., 2017)

Lake Surface Temperature (LST) time series including the spin-up period, 2016



From the time series, it is evident that during the spin-up period the model stabilizes and captures the seasonal cycle and its timing, indicating that a one-year spin-up period is appropriate. In addition to the justification for the spin-up period in Lines 161-164, the above figure with 2016 LST time series has been included in the Supplementary Material of the revised version.

[Additionally, the relatively poor performance in 2017 \(after Section 3.4\) may be linked to insufficient spin-up. Please clarify whether this was assessed.](#)

The evaluation of the spin-up period on LST (explained in the previous response), indicated that the one-year time period is sufficient to capture the seasonal cycle, timing, and magnitude of surface temperatures, showing good agreement with observations. Using a one-year spin-up for ice modeling is also common practice in the Great Lakes (e.g. Anderson et al., 2018).

In view of the above, the relatively poor performance of MOM6 in ice concentration (section 3.5) in 2017 is unlikely to be related to spin-up. Maximum errors were actually larger in winter 2018 than winter 2017, and the largest biases occurred during ice formation periods. This is a common issue in Great Lakes ice models (e.g. Anderson et al., 2018) and is often linked to missing physical processes (e.g. ice-wave interactions). As noted in the manuscript (lines 104-106), although ice salinity was set to zero to approximate freshwater conditions, the physical parameterizations such as rheology and ice strength were designed for sea ice and have not been modified for lake ice. We suspect that these missing physical processes are likely culprits for ice model mismatches across all years. Parameterizing a lake ice model within MOM6 is planned as a future application and is expected to improve the simulation of ice-related processes and associated surface conditions. It is also worth noting that ice simulations here are fairly strong across the board, even as compared to operational ice models (Anderson et al., 2018), which also struggled to reproduce ice cover in 2017.

4) Closed lateral boundaries and absence of freshwater fluxes

Lines 150–151 indicate that the model uses closed lateral boundaries and excludes evaporation and precipitation. This configuration will affect both radiation balance and water discharge.

Have the authors tested simulations including freshwater forcing? A comparison between the current setup and a freshwater-forced configuration would strengthen the manuscript. Please discuss the implications of excluding freshwater fluxes.

Thank you for this comment and for the opportunity to clarify the treatment of surface mass fluxes in our model configuration. The model configuration does include evaporation, precipitation, and the associated surface heat fluxes through atmospheric surface forcing. However, the model employs closed lateral boundaries and does not include lateral freshwater fluxes such as river inflows or outflows. As a result, it is not possible to maintain a stable long-term mean lake level without the lateral fluxes (rivers). To close the mass budget and prevent artificial drift at lake level, the model applies a spatially uniform correction to the surface freshwater flux that offsets the domain-averaged imbalance between evaporation and precipitation. This approach ensures that the total lake volume remains stable over long simulations. A similar approach is used in the LMH-FVCOM simulations with which we compare our results. Importantly, this correction only affects the lake-mean volume and does not influence the spatial patterns of lake surface height, which primarily reflect circulation-driven variability. We have added the text within quotes to section 2.1.3 Model domain, in our revised version to clarify our model configuration,

“For this initial configuration of MOM6-SIS2-LMH, similar to previous model simulations (Cannon et al., 2023), the lake domain is closed at its lateral and bottom boundaries, with no river inflows or groundwater exchange. Evaporation and precipitation are included through the atmospheric surface forcing; however, because lateral freshwater fluxes are not represented, a spatially uniform correction is applied to the net surface freshwater flux to prevent long-term drift in lake level.”

In the current study, we have not conducted simulations including freshwater forcing. However, excluding these fluxes is not expected to substantially affect the large-scale circulation patterns or the seasonal thermal structure, which are the primary focus of this study, because the volume change due to river inflows or surface fluxes is small relative to the total lake volume. Inclusion of freshwater fluxes would be important for studies focusing on lake level variations, nutrient transport, or ice cover dynamics, and is a potential avenue for future model development.

We have added the following lines in the discussion section, 4.3 Strengths and limitations of MOM6-SIS2 for applications in Great Lakes,

“In the present study, simulations were conducted without freshwater forcing such as river inflows. Excluding these fluxes is not expected to substantially affect the large-scale circulation patterns or the seasonal thermal structure, which are the primary focus of this work. However, freshwater inputs are important drivers of nutrient loading and material transport, which are critical for ecological and biogeochemical processes. Therefore, incorporating freshwater forcing will be an important step in future model development aimed at coupled ecosystem and biogeochemical applications.”

5) Warm bias in nearshore regions (Figure 3)

In Figure 3, the warm bias is particularly pronounced in nearshore regions during summer. Is this related to insufficient vertical grid resolution near the lake boundary or limitations in VCS representation?

Please elaborate on the sensitivity of LST to the vertical coordinate system in these regions.

Additionally, improve the resolution and clarity of Figure 3.

This is an interesting comment to discuss, thanks for pointing out. Figure 3 shows the difference between MOM6 (hybrid) and satellite estimates (GLSEA) for winter and summer seasons and in comparison, to LMH-FVCOM. The vertical resolution of MOM6 is 2m from the surface to 30m depth, whereas for LMH-FVCOM it varies from 2.5cm in the nearshore regions to 14m in deeper regions. The warm bias in the nearshore areas in summer season is seen in both MOM6 and LMH-FVCOM which hints that the vertical resolution might not be a contributing factor. We suspect the biases in near shore regions are due to poor representation of near shore areas by the observational product, GLSEA. Previous studies by Schwab et al. (1992) and Trumpickas et al. (2015), in the region have indicated that although GLSEA provides an estimate of lake-wide mean surface temperatures, the observations are less reliable in nearshore regions. During post-processing, GLSEA pixels (taken on a 2.5 km scale) overlapping the shoreline are removed from the GLSEA water temperature estimate, decreasing the influence of nearshore areas on the estimate of the overall lake-wide mean surface water temperature. Because these shallow nearshore regions respond most quickly to atmospheric forcing, their exclusion leads to biases in “observational” products near the coast, resulting in low-biased temperature simulations in the winter and warm-biased simulations in the summer. This is apparent in all MOM6 and FVCOM simulations, with the exception of LMH-FVCOM simulation in western lake Michigan, which show a persistent (and erroneous) upwelling for most of the summer.

Impact of vertical coordinate system to LST

We have added the following lines in the discussion section, 4.2 Vertical Coordinate System Evaluation,

“Spatially, during summer, the z^* configuration is slightly warmer than the hybrid VCS (Figures 3 and S3), likely because the z^* VCS enhances upward heat transport. In contrast, the hybrid VCS better aligns with density surfaces, reducing spurious mixing and maintaining sharper stratification, which leads to relatively cooler LST. In nearshore regions, both VCS configurations exhibit a warm bias, which are likely associated with uncertainties in satellite-based estimates, including known biases in GLSEA products in coastal areas (Schwab et al., 1992; Trumpickas et al., 2015).”

A revised Figure 3 will be included in the revised version.

6) Performance of z^* VCS

In Subsection 3.4, the performance of the z^* VCS configuration is not sufficiently described. Please provide a clearer quantitative and qualitative assessment.

Thank you for pointing out the need for a clearer assessment of the z^* VCS configuration. However, the processes described in Section 3.4 primarily involve surface-driven dynamics (e.g., wind-driven circulation, nearshore upwelling, and seiche variability). Our analysis of lake-wide daily surface temperature time-series indicates minimal differences between the z^* and hybrid VCS (Figure 2b) configurations in representing these surface processes. Given this similarity, we used the hybrid VCS configuration as a representative example to evaluate MOM6 performance in capturing nearshore dynamics, and to compare against LMH-FVCOM and observations. Including parallel figures and detailed descriptions for the z^* configuration would largely duplicate the results without providing additional insight. To clarify this point, we have added the following statement at the beginning of Section 3.4:

“Both z^* and hybrid VCSs show comparable skill in simulating surface temperature variability (Figure 2), therefore, the hybrid configuration is presented here for clarity and brevity, as it adequately represents MOM6 behavior for the processes examined.”

7) Clarification of statement (lines 380–381)

The statement in lines 380–381 requires further explanation. Please elaborate on the physical reasoning or provide supporting evidence.

The physical reasoning behind this statement is that the coarser grid in MOM6-SIS2-LMH constrains the model’s representation of energetic variability, even though the large-scale qualitative flow patterns are well reproduced. We have added this clarification to the revised version.

“MOM6-SIS2-LMH successfully reproduces the observed bi-directional flow patterns at the Straits of Mackinac (UGLOS station shown in Fig. 1), capturing both the timing and directionality of the velocity fluctuations despite only having 1 km horizontal resolution (Fig. 10). However, it underestimates the magnitude of velocities throughout the time series. This suggests that, although MOM6-SIS2-LMH captures the qualitative dynamics of the bi-directional flow patterns, it does not fully resolve the energetic variability as seen in LMH-FVCOM. This underestimation arises primarily from the coarser horizontal resolution in MOM6-SIS2-LMH, which limits the model’s ability to represent small-scale turbulent and nearshore processes that contribute to energetic variability. In contrast, LMH-FVCOM, with its finer unstructured grid, better resolves these features and captures higher-frequency velocity fluctuations.”

8) Water level comparison

In my opinion, the water-level comparison (lines 399–409) is not meaningful due to the absence of freshwater fluxes in the model configuration. Without river inflow, precipitation, and evaporation, water-level validation lacks physical consistency. I recommend removing this comparison and Figure 12.

We understand your point regarding water level validation without freshwater fluxes. Importantly, this model parameterization only affects the lake-mean water level (i.e. the total lake volume is held

constant) and does not influence the temporal variability of water level fluctuations. The power spectral density shown in Figure 12 reflects variability across different frequencies driven primarily by atmospheric forcing and basin-scale circulation processes (i.e. seiches), rather than changes in the lake volume. Therefore, we believe that the comparison remains physically meaningful for evaluating the model's ability to reproduce observed water level variability across temporal scales.

For these reasons, we believe that retaining Figure 12 is appropriate, as it provides valuable information on the model's performance in capturing the frequency-dependent variability of lake surface height. It is also worth noting that these 6-month water level time series are detrended for spectral analysis, so seasonal trends in water levels driven by inflows, outflows, and evaporation (which have periods on the order of 12 months) are not meaningfully reflected in Figure 12.

9) Surface wind forcing (NARR data)

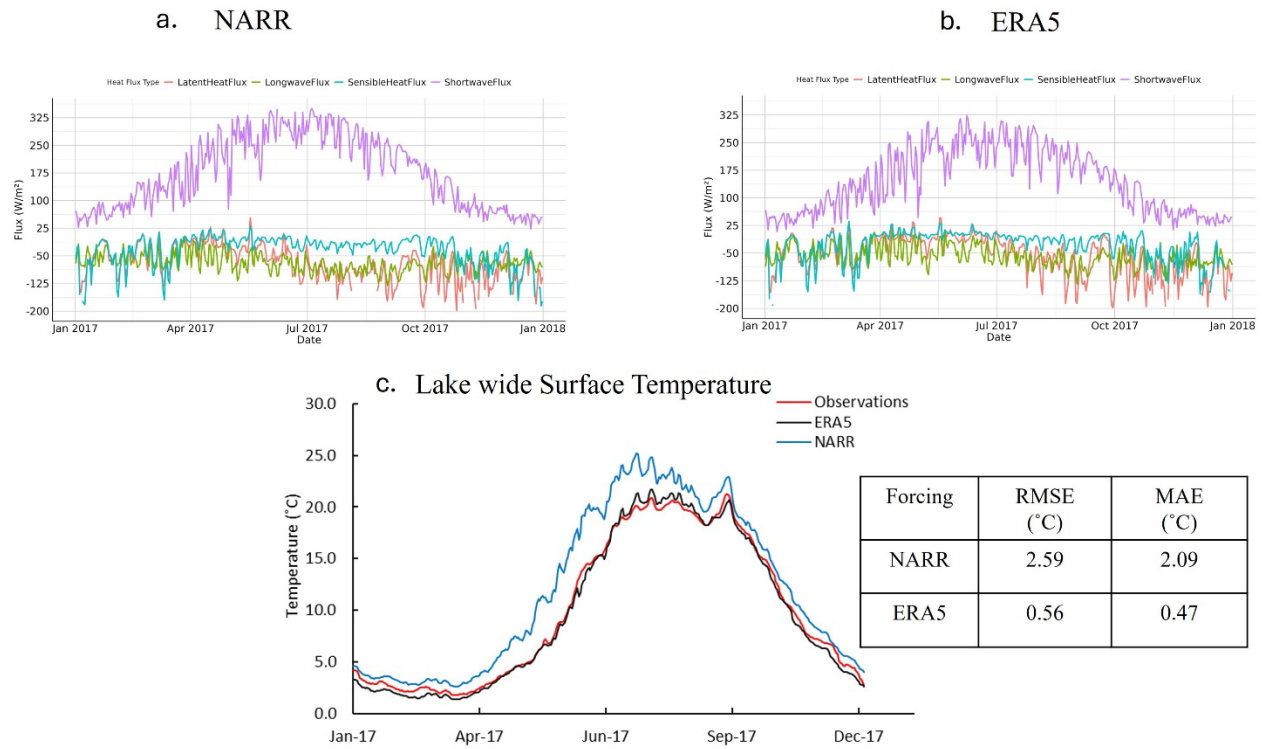
[Why was North American Regional Reanalysis \(NARR\) not tested as a surface wind forcing dataset? Please justify this choice and support the findings with appropriate discussion.](#)

Thank you for this important question. Experiments were also conducted using NARR as atmospheric forcing. It was observed that the NARR forcing exhibits a positive bias of approximately 30% in downward shortwave radiation, resulting in a consistent warm bias in lake-wide surface temperature (LST) during spring and summer. This bias affects the surface heat budget and results in reduced agreement with observations compared to ERA5-forced simulations.

In addition to these biases, we chose not to use NARR selectively (e.g., for wind forcing only) while using ERA5 for other variables, as this would introduce inconsistencies in the surface forcing fields. Air-lake interactions depend on the coupled influence of wind, heat, and moisture fluxes, which are internally consistent within a given reanalysis product. Mixing variables from different reanalysis datasets can disrupt this balance and lead to physically inconsistent surface fluxes.

For these reasons, ERA5 was used consistently for all atmospheric forcing variables in this study. Supporting analysis comparing NARR- and ERA5-forced simulations, including impacts on heat

budget components and LST for 2017, has been included in the Supplementary Material.



The following lines are included in section 2.1.4 Atmospheric forcing and initial conditions, to justify the forcing selection.

“MOM6-SIS2-LMH experiments were also run using North American Regional Reanalysis (NARR; Mesinger et al., 2006), but ERA5 was ultimately selected due an overestimation of downward shortwave radiation in NARR (identified from preliminary investigations) that notably biased modeled LST. NARR-forced experiments exhibited a systematic positive bias in downward shortwave radiation (approximately 30%), which resulted in a consistent warm bias in lake-wide surface temperature, particularly during spring and summer. This bias propagated through the surface heat budget, degrading the agreement with observed temperature variability relative to ERA5-forced simulations. In contrast, ERA5 provided a more balanced representation of surface heat flux components and improved overall consistency with observed lake-wide thermal conditions. Therefore, ERA5 was selected as the primary atmospheric forcing dataset for this study.”

10) Future applications section

Please consider creating a dedicated “Future Applications” section (lines 575–600) to clearly articulate potential extensions and practical applications of the modeling framework.

We appreciate your comment. This is included in the revised version.

11) Conclusion – spatial specificity

In Point 1 of the conclusion, please specify that the stated performance primarily applies to offshore regions, as nearshore summer biases are comparatively large.

Thank you for this suggestion. We would like to clarify that the warm nearshore bias noted in summer is not solely attributable to model performance. As shown in the spatial LST maps, both MOM6-SIS2-LMH and LMH-FVCOM exhibit similar warm biases in nearshore regions, suggesting a common source of discrepancy. Additionally, the relatively low RMSE for MOM6 compared to LMH-FVCOM in reproducing daily lake-averaged LST indicates that the model performs well in capturing the LST.

We therefore attribute a significant portion of the nearshore bias to limitations in the observational dataset, particularly the known uncertainties in satellite-based LST product, GLSEA in nearshore areas.

We believe that (1) is correct as written, since biases in MOM6-SIS2 are similar (or better than) those estimated for LMH-MOM6. We have modified (1) to note that “lake-wide” parameters are in good-agreement and that all models showed biases consistent with satellite-derived observations.

(1) MOM6-SIS2 performs well for large freshwater systems, matching or outperforming unstructured grid models in simulating lake-wide SST, ice cover, and circulation patterns in the Great Lakes. Performance was especially strong for vertical thermal structure, for which simulated thermocline depths, turnover dates, and hypolimnetic temperatures showed excellent agreement with observational moorings in each lake. Both structured and unstructured grid models displayed biases in nearshore SST simulations consistent with uncertainties in satellite-derived observations.

References

Anderson, E. J., Fujisaki-Manome, A., Kessler, J., Lang, G. A., Chu, P. Y., Kelley, J. G. W., Chen, Y., and Wang, J.: Ice forecasting in the next-generation great lakes operational forecast system (GLOFS), *J. Mar. Sci. Eng.*, 6, 123, 2018.

Cannon, D., Wang, J., Fujisaki-Manome, A., Kessler, J., Ruberg, S., and Constant, S.: Investigating Multidecadal Trends in Ice Cover and Subsurface Temperatures in the Laurentian Great Lakes Using a Coupled Hydrodynamic–Ice Model, *J. Clim.*, 37, 1249–1276, <https://doi.org/https://doi.org/10.1175/JCLI-D-23-0092.1>, 2024.

Schwab, D. J., Leshkevich, G. A., and Muhr, G. C.: Satellite measurements of surface water temperature in the Great Lakes: Great Lakes Coastwatch, *J. Great Lakes Res.*, 18, 247–258, 1992.

Trumpickas, J., Shuter, B. J., Minns, C. K., and Cyr, H.: Characterizing patterns of nearshore water temperature variation in the North American Great Lakes and assessing sensitivities to climate change, *J. Great Lakes Res.*, 41, 53–64, 2015.

Xue, P., Pal, J. S., Ye, X., Lenters, J. D., Huang, C., and Chu, P. Y.: Improving the Simulation of Large Lakes in Regional Climate Modeling: Two-Way Lake–Atmosphere Coupling with a 3D Hydrodynamic Model of the Great Lakes, *J. Clim.*, 30, 1605–1627, <https://doi.org/https://doi.org/10.1175/JCLI-D-16-0225.1>, 2017.