RC1

Lin et describes and evaluates a forecasting system for predicting 3-D thermal structure of Lake Erie. The manuscript is an evaluation of a true forecasting system (i.e., it is evaluating a set of forecasts of the future rather than mimicking the forecasting process with historical data). It uses forecasted meteorology from Environment Canada Global Deterministic Forecast System to drive the model. The model used (AEM3D) had previously been applied to Lake Erie so the novelty of the paper is using the model with forecasted meteorology. I really appreciated the discussion sections on how the forecasting system could potentially be used to anticipate critical events for decision makers, emergency managers, and users of the lake. I visited the website for the project and it is indeed up-to-date. The paper evaluates the performance of the forecasts for a three-month period of time, although the manuscript discusses forecasts from outside this period. The authors highlight the automation provided by a Python script, but this seems to lack novelty (thought it appears to get the job done). The model is a standardly used model, downloading one meteorology source and converting to another is straightforward, and using a task manager to run a job is routine. The application to Lake Erie is new but the manuscript is introducing a named forecasting system (COASTLINES) with the only model development being a Python script used to execute the model and download observations (actually also Matlab scripts). Further, the code for running COASTLINES is not provided. Overall, the novelty of COASTLINES beyond the application to Lake Erie needs to be better motivated in the introduction.

> **Reply 1:** Thank you for the comments and advice. We have made the framework of COASTLINES open-source, by uploading all the code for running COASTLINES to the Dataverse (https://doi.org/10.5683/SP2/VTN7WC) (see Code and data availability). To apply a computational model to a lake requires laborious data sourcing and preparation, model setup and calibration. Thus, it is impossible to rapidly generate forecast output. The novelty of our study is to automate this lake-model application, using open data, enabling timely hydrodynamic forecasts that are communicated on a web platform. We agree that all the tools we have used are routine, but we have developed a workflow that combines them into a useful product and are the only operational forecast system that solely uses publicly available online data for model forcing. As we stated in the Introduction (lines 69-72),

"In the present study we developed and tested the COASTLINES (Canadian cOASTal and Lake forecastINg modEl System; https://coastlines.engineering.queensu.ca/) lake-model application workflow, that rapidly accesses near real-time online data (weather forecasts, water level and temperature observations) for automated model forcing, execution and validation. Hydrodynamic forecasts are automatically post-processed and posted on a web platform."

The framework of this operational forecast system can be widely applied in other water systems around the world owning to the spatial coverage of the meteorological data and proliferation of near real-time lake observation data. We have emphasized this point in the conclusion section (lines 490-493),

"This operational system shows the feasibility of applying freely available meteorological forecasts (e.g., GDPS, HRRR), in situ buoy data and satellite images to drive and validate any computational lake model (e.g., AEM3D, DELFT3D, GLM), without modifying the source code. The global coverage of the weather model allows generalization of model application to and lake or coastal domain."

The manuscript only uses observations to evaluate the model. It does not perform data assimilation as other forecasting system do. The workflow highlights the automation of the evaluation using the observations but does not offer a way that the evaluation is used to improve the model or the forecasts. Therefore, it isn't clear why the automated evaluation is necessary. It would be great to explore how a feedback between the evaluation and the forecasts can be developed. The forecasts lack a representation of uncertainty. Uncertainty in forecasts is increasingly the state-of-the-art. There is reference to uncertainty in the figures, but the manuscript does not describe how uncertainty is estimated. At minimum, the discussion needs to address the lack of uncertainty in the forecast and explore how it might be included in the forecasts.

Reply 2: We cannot use data assimilation for model forecasting, because we do not have observations in the future to assimilate. Rather, we could apply data assimilation to better calibrate the real-time model, from which we generate our forecasts. This has been done (e.g., (Baracchini et al., 2020a)) to reduce the RMSE temperature simulation of Lake Geneva from ~ 2 ℃ to ~ 1 ℃ by employing data assimilation that required ~ 1 month of computational time. As future work, we will improve the overall model calibration (which is not included in the COASTLINES forecast workflow) by employing the same OpenDA calibration (https://www.openda.org/). This approach would be consistent with our philosophy to develop modelling tools that can be universally applied to hydrodynamic source codes.

We have discussed the limitations of implementing data assimilation in current COASTLINES (lines 429-441)

"The uncertainty and bias in the COASTLINES forecast results from error induced by the initial conditions at each hot-start, error in the meteorological forecasts and error in the numerical methods. These errors could be reduced by improving model calibration through data assimilation. For example, (Baracchini et al., 2020a) reduced the RMSE temperature simulation of Lake Geneva from ~ 2 ℃ to ~ 1 ℃ by employing a data assimilation routine; this would correspond to a <5% improvement in simulation of Lake Erie summer surface temperature. Before implementing data assimilation, the limitations of such a scheme must be considered: (*i*) The lack of observations in the future, makes data assimilation impossible for adjusting forecasts; (*ii*) data assimilation is computationally intensive (Baracchini et al., 2020a) required ~1 month of computational time, clearly not an option for operational forecasting); and (*iii*) data assimilation requires modification of the source code, which is not consistent with our philosophy to develop modelling tools that can be universally applied. Rather, future work will focus on adding real time model calibration (e.g., (Gaudard et al., 2019)), which is not presently included in the COASTLINES forecast workflow. For example, (Baracchini et al., 2020a) employed OpenDA (https://www.openda.org/) as a black-box wrapper to calibrate DELFT3D for Lake Geneva. This approach can be adapted to any other model."

We have tracked model error with forecast horizon. The present evaluation provides the root-mean-squared-deviation (RMSD), relative error (RE), and mean bias deviation (MBD) of the forecast results which indicate the quantitative uncertainty in the prediction according to the horizon (Fig. 3 and 7(a)). The automated evaluation reveals the probabilistic nature of the prediction results. The automated evaluations included in the COSTLINES workflow provide a quantitative confidence interval in terms of the predictive horizon, presenting how the uncertainty grows through time (Fig. 3, 7) and what maximum and minimum limits we should expect.

In the section 2.4 (lines 152-155), we describe how we calculated RMSD and RE for water level predictions:

"RMSD is the absolute error of the model against the observation. The difference between the observed daily minimum and maximum value was defined as the daily water level fluctuation range, where RE is the ratio between the RMSD and lognormal mean of daily range over April to September 2020. Given that our study focusses on a 240-h forecast, RE is able to characterize the forecast bias, regardless of the instantaneous water level position."

And in lines 173-178, we describe how we calculate MBD for water temperature predictions,

"We quantified the temperature forecast capability using the statistical measures of RMSD (eq. 1) and Mean Bias Deviation (MBD):

$$MBD = 100\,\frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)}{\frac{1}{N}\sum_{i=1}^{N}y_i} \qquad (1)$$

For the spatial MBD and RMSD (s-MBD and s-RMSD), x_i and y_i are the model and observed temperature in each grid, and N is the total number of grids. For timeseries MBD and RMSD (t-MBD and t-RMSD), x_i and y_i are the model and observed temperature at each sample time, and N is the total number of samples."

As we stated in lines 219-223, we use ensembled and averaged RMSD, RE, and MBD to represent the confidence interval.

"The water level statistical metrics (RMSD and RE) were ensembled and averaged over April to September 2020. The 24-h and the 240-h forecast lake surface temperature and temperature profiles, from the models, were also validated against real-time lake buoy data and daily averaged satellite imagery. The timeseries and spatial MBD and RMSD (t-RMSD, t-MBD and s-RMSD, s-MBD) were ensembled and averaged over July to September 2020."

And the estimation of uncertainty has been included in the application of forecast (e.g., Fig. 4, 5, and Fig. 12(d)). The shaded areas in the panel represent the confidence interval, indicating the highest water level we should expect in terms of the forecast time scale, according to the ensembled and averaged RMSE, RE, and MBD.

The argument for why data assimilation is not necessary could be stronger. They argue that it would only potentially decrease the RMSE by ~half (0.7C) by citing another study that used data assimilation. However, how does the reader know whether this is a meaningful magnitude?

Reply 3: We agree. See Reply 2 above. We have added further info on the magnitude of improvement (lines 429-441).

Instead of putting the Python script in the Appendix, I recommend putting them in a repository like Zenado. That would allow someone to use the scripts without having to cut and paste from the Appendix. Furthermore, the manuscript highlights the Python code but also has a dependence on MatLab for foundational parts of the workflow (i.e., line 131-132 – the conversation of weather forecasts into the AEM3D input). Therefore, it is a Python + Matlab + Windows Schedule workflow. Additionally, the code that is provided is only for retrieving the observations. The code to run the forecasting system is missing. Overall, I do not think that the availability of code and model output meets GMD's standards.

> **Reply 4:** We agree and have posted the COASTLINES source code in the Dataverse (see Code and data availability). This includes the wrapping Python code, MATLAB scripts, and the timeline set in Windows task scheduler. The hydrodynamic driver is set as an executable black box and, while available form a third party, could be replaced with any executable hydrodynamic code.

The term hindcast is used throughout the manuscript but is not defined. It would be help to define exactly what a hindcast is.

> **Reply 5:** Thank you for the suggestion. We have added the definition in line 37-38:
> "Over the past several decades, many computer models have been applied to hindcast (running models forced with and validated against historically collected data) lake hydrodynamics to aid management."

Line 440 states "To facilitate further development of open-access predictive modelling systems", which is a major oversell of COASTLINES as being in the group of open-access predictive modeling. The hydrodynamic model requires a license and the code to run the forecasting system is not made available.

> **Reply 6:** The modelling system we developed has been placed in a data repository. The hydrodynamic driver (AEM3D) is treated as a black box (as often done with OpenDA for calibration, (Baracchini et al., 2020a)), is available from Hydronumerics and cold be replaced by any other executable hydrodynamic code.

Can the authors point to a manuscript that demonstrates that the re-start works or provide a figure that shows that a restarted run matches a run that was continuous (i.e., same length of simulation but without stopping and restarting)? Some hydrodynamic models are designed to be run from a cold-start and have internal variables that are not saved – thus preventing a true restart.

> **Reply 7:** To address your concern, we have added the temperature profile comparison of restarted run and continuous run in the Appendix A. It turns out that there is no obvious difference between these two modes (lines 207-209).

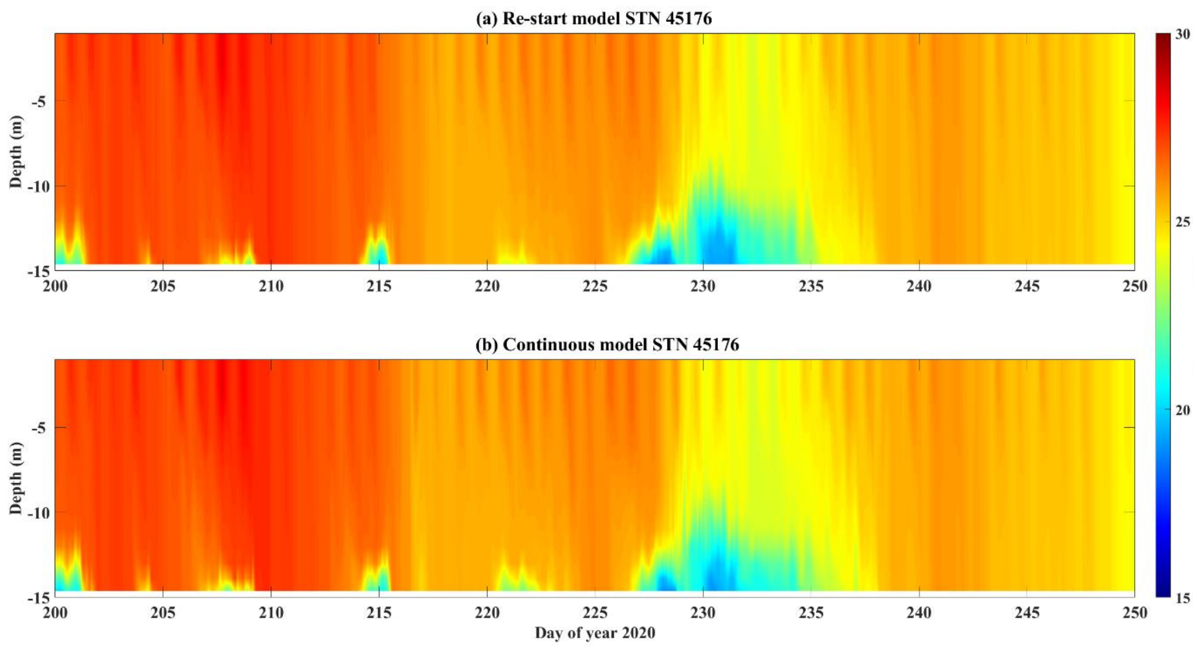

Fig. A1 Temperature profile comparisons between (a) stitched 24 h model run with re-start files, and (b) model run with continuous input files.

The manuscript highlights the use of the Environment Canada Global Deterministic Forecast System (GDPS) product. Could other freely available forecast model output be used? What about NOAA's Global Forecasting System or NOAA's Global Ensemble Forecasting system?

> **Reply 8:** Yes. In this study, the meteorological forecasts are provided by GDPS, but other weather forecast model with various horizontal resolutions can also be used (e.g., High Resolution Rapid Refresh (HRRR) from NOAA (lines 210-212), see also(Rey and

Mulligan, 2021)). We are presently developing COASTLINES for Lake Ontario with DELFT3D as the black box hydrodynamic code and NOAA forcing. These options will be included as switches in the COASTLINES workflow.

Specific comments

Line 47: Are there 1-D water temperature forecasting systems that can also provide context in the introduction and discussion?

Reply 9: As you suggested, we have added reference to a Simatrat application which is a near-realtime system (lines 55-56).

"Similarly, meteolakes.ch (Baracchini et al., 2020b) applies Delft3D for short-term 3D forecasts (4.5 days) of four Swiss lakes and simstrat.eawag.ch (Gaudard et al., 2019) applies Simstrat for near-realtime 1D simulation of 54 Swisss lakes. "

Some 1D models lake models (e.g., General Ocean Turbulence Model (GOTM, https://gotm.net), GLM (Hipsey et al., 2014), and Freshwater lake model (FLake, (Mironov, 2008)) have been incorporated with climate forecast system (e.g., Climate Forecast System (CFS) at US National Centers for Environmental Prediction (NCEP; Saha et al., 2014), European Centre for Medium-Range Weather Forecast (ECMWF; Johnson et al., 2019)) to conduct retrospective forecast of water temperature in the lakes (Mercado-Bettín et al., 2021;Lv et al., 2019), but none of them are operational forecasting systems. Also, these implementations require relatively complex processes of input data (climate forecast outputs) acquisition and re-analysis, hindering their usage in real-time. We also provided some comparison of the RMSD of LST in an 1D model (Freshwater Lake [FLake]) and in COASTLINES (lines 425-428)

"COASTLINES also outperforms 1D climatological hindcasts (e.g., Freshwater Lake (FLake)), with 2– 4 ℃ RMSD over a 120-h lake surface temperature forecast (Lv et al., 2019;Gu et al., 2015) and has similar error to the 3D Princeton Ocean Model (Kelley et al., 1998), with 0.6-0.9 ℃ mean absolute error in the 36-h lake surface temperature forecast at station 45005."

Line 63: I recommend starting a new paragraph here.

Reply 10: We started a new paragraph here as you suggested.

Line 95: Is there a specific version of the AEM3D that was used? Without that the forecasting system is not reproducible.

Reply 11: We are using AEM3D version 1.1.1. The information was supplied in line 108.

Line 119: what is CFL =?

Reply 12: We are sorry about the typo here. CFL = $\sqrt{2}$ (line 121)

Line 156: What happens when there are run-time errors?

Reply 13: We used the try/except function in Python to avoid the program accidently stopping due to run-time errors (See the code archived in the Dataverse). Also, the supervisor of the COASTLINES system routinely monitors the whole system for troubleshooting (lines 207-208).

"The authors (supervisors of COASTLINES) and Queen's ITS monitor forecast results and maintain system operation.**"**

Line 184: Is the restart file used to generate a 216-hr forecast since the first 24-hr have already been generated?

Reply 14: Yes. Your understanding is correct. Because 24h and 240h forecast are driven by same 10-day meteorological data, there is no need for repeating the first 24h forecast.

Line 189: The Windows Task Scheduler is another dependency of the forecasting system. Can the forecasting system only run in a Windows environment?

Reply 15: It is straightforward to run the forecasting system operationally in Unix-like operating systems (e.g., Linux Mint, Ubuntu, macOS) by using the software utility *cron*, which is a time-based job scheduling tool that will run jobs/scripts at regular intervals.

Line 224: The sentence refers to the estimation of uncertainty in the model forecast but the manuscript lacks methods describing the uncertainty estimation process.

Reply 16: Sorry for the confusion here. We quantify the forecast uncertainty in terms of the Root Mean Square Deviation (RMSD), Relative Error (RE), and Mean Bias Deviation (MBD), and the calculation processes were mentioned in section 2. 4 (Eq. 1-3). The uncertainties we showed in the Fig. 3-5 are statistical metrics of water level RMSD and RE ensembled over April to September 2020, and the uncertainties we showed in the Fig. 7 are the statistical metrics of LST MBD and RMSD ensembled over July to September 2020 (lines 219-223).

Line 260: What forecast horizon do these numbers refer to?

Reply 17: These numbers refer to 24h forecast as we indicated in the Table 2. We also added this information here (lines 274-275).

Line 329: Readily what? (missing word)

Reply 18: We are sorry about the typo here. The sentence should be

"the objective of the present work is to develop a simple automated lake modelling system that can be readily operated to diverse field sites to suit management needs."

This sentence has bee removed in the revision.

Fig 2: What is Phenomena detection (Supervisor)? It is not mentioned anywhere in the manuscript.

Reply 19: The authors and Queen's ITS play the supervisor roles, maintaining system normal operation and detecting the phenomena of interest. We have added the explanation in lines 207-208: "The authors (supervisors of COASTLINES) and Queen's ITS monitor forecast results and maintain system operation."

Fig 2: The caption says "Daily Python workflow" but the text states that Matlab is also used.

Reply 20: We modified the caption and deleted "Python" here.

Reference

Baracchini, T., Hummel, S., Verlaan, M., Cimatoribus, A., Wüest, A., and Bouffard, D.: An automated calibration framework and open source tools for 3D lake hydrodynamic models, Environmental Modelling & Software, 134, 104787, 2020a.

Baracchini, T., Wüest, A., and Bouffard, D.: Meteolakes: An operational online three-dimensional forecasting platform for lake hydrodynamics, Water Res., 172, 10.1016/j.watres.2020.115529, 2020b.

Gaudard, A., Vinnå, L. R., Bärenbold, F., Schmid, M., and Bouffard, D.: Toward an open-access of high-frequency lake modelling and statistics data for scientists and practitioners. The case of Swiss Lakes using Simstrat v2.1, Geosci. Model Dev., 12, 3955-3974, 10.5194/gmd-2018-336, 2019.

Gu, H., Jin, J., Wu, Y., Ek, M. B., and Subin, Z. M.: Calibration and validation of lake surface temperature simulations with the coupled WRF-lake model, Climatic Change, 129, 471-483, 2015.

Kelley, J. G. W., Hobgood, J. S., Bedford, K. W., and Schwab, D. J.: Generation of Three-Dimensional Lake Model Forecasts for Lake Erie, Weather and Forecasting, 13, 659-687, 10.1175/1520-0434(1998)013<0659:GOTDLM>2.0.CO;2, 1998.

Lv, Z., Zhang, S., Jin, J., Wu, Y., and Ek, M. B.: Coupling of a physically based lake model into the climate forecast system to improve winter climate forecasts for the Great Lakes region, Climate Dynamics, 53, 6503-6517, 10.1007/s00382-019-04939-2, 2019.

Mercado-Bettín, D., Clayer, F., Shikhani, M., Moore, T. N., Frías, M. D., Jackson-Blake, L., Sample, J., Iturbide, M., Herrera, S., French, A. S., Norling, M. D., Rinke, K., and Marcé, R.: Forecasting water temperature in lakes and reservoirs using seasonal climate prediction, Water Research, 201, 117286, https://doi.org/10.1016/j.watres.2021.117286, 2021.

Mironov, D. V.: Parameterization of lakes in numerical weather prediction: Description of a lake model, DWD, 2008.

Rey, A., and Mulligan, R. P.: Influence of Hurricane Wind Field Variability on RealTime Forecast Simulations of the Coastal Environment, J. Geophys. Res. Oceans, 126, 10.1029/2020JC016489, 2021.