Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC2
https://doi.org/10.5194/gmd-2021-34-RC2, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

# Comment on gmd-2021-34

Daisuke Tokuda (Referee)

---

Referee comment on "An automatic lake-model application using near-real-time data forcing: development of an operational forecast workflow (COASTLINES) for Lake Erie" by Shuqi Lin et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-34-RC2, 2021

---

**General comments:**

This study develops an operational forecast system with a three-dimensional lake hydrodynamic model and validated the performance of hindcast experiments in Lake Erie. To conduct a robust forecast, the study separated 24-h and 240-h forecast simulations, and update the restart file for 240-h forecast every day. The data retrieval, numerical simulation, and validation are automated. This study is a development of a forecast system rather than a new model, but it is still of importance from the perspective of research implementation to society. However, some information for long-term steady operation is lacking, and providing us with such information would be helpful to other models/system developers, which are listed in the following major comments.

**Specific comments:**

How did you set the initial value of the model? You mentioned that the model was 'cold started' on day 99 (line 116) and generates a restart file with a 24-h forecast simulation (line 185), but some other sentences imply that the model was initiated on another date (e.g., lines 288 and 299). I imagine that "initiated" means the model was just restarted in the consistent simulation (if so, only lead (forecast) time information is enough), but the authors are requested to clearly explain the system setup because such information is useful for other forecast systems without data assimilation.

What is the difference of the meteorological forcing data between 15km and 25km? If your

system can work in time with 15-km data, I think you can focus on the results of 15-km data only because the difference of results seems to be negligible. Or is there any known problem with 15-km data?

You showed one-year results, but I have an interest in the long-term stability of the system operation. Do you reset the initial condition every year?

## 2 Data and methods

### 2.2 Model description

Line 96: Which programming language is the AEM3D written in? You mentioned that the wrapper code is written in Python, but it has no advantage on computational efficiency. In addition, you pointed out that previous hydrodynamic models are difficult to apply to the hind- and forecast applications due to the computational cost in line 42, how did you solve the problem?

### 2.3 Model setup and meteorological forcing variables

Line133: I have three questions.

(1) The mass balance of lakes is described in + Precipitation – Evaporation + Riverine inflow – Riverine outflow ± Groundwater infiltration/seepage. Did you consider the groundwater component? If not, explicitly describe that assumption.

(2) In addition, is Precipitation – Evaporation balanced in Lake Erie? This budget controls the seasonal variability in water level, but mass imbalance may cause a problem in long-term operation.

(3) Even if the riverine in/outflows are balanced, can you ignore the effect on the fluid velocity field near the inlets and outlets? If you can, add the reference.

## 3 Results

You showed the confidence shade in the figures of time series, how did you do conduct ensemble simulations? This question is related to the reliability of the system if it is operated as a warning system to society.

### 3.2.1 Lake surface temperature

Line 273: Those results are interesting; longer forecast time does not increase the error for surface temperature (thermodynamics) even with consistent bias according to Fig. 6 but does the error for water level (hydrodynamics) according to Fig. 4 and 5. Can you discuss the difference? Comparison between the model bias and forecast time would be helpful in this respect.

Line 280: Why could you conclude the underestimation is due to ignoring river inputs? The underestimation occurs on the east side of the inlet from the Detroit River. Can you have a consistent discussion between line 223 and here?

Fig. 8: What is the main reason for the consistent underestimation in lake surface temperature?

## 4 Discussion

### 4.1 Bias and uncertainty

If the system developed in this study focuses on the forecast of some critical events like coastal up-welling and storm surge as discussed in the Sects. 4.1 and 4.2, could you move this Sect. 4.1 after the Sect. 4.3? The current Sect. 4.1 discussed the mean RMSD and compared it with a previous study incorporating data assimilation, but the data assimilation corrects only the initial condition of state variables in a model; not boundary conditions including meteorological forcing data. On the other hand, some of the critical events are caused by extreme atmospheric conditions in my understanding. So, can you discuss the model uncertainty and further improvement separating into initial-value and boundary-value problem?

**Technical corrections:**

**2.3 Model setup and meteorological forcing variables**

Line 116: How did you set the initial condition for the water level?

Line 119: What is "which is CFL = (Hodges et al., 2000)?

Line 121: "and net longwave radiation" (the former one) -> "and downward longwave radiation"? Because net (downward -upward) longwave radiation is calculated within the model as you mentioned.

**3.2.1 Lake surface temperature**

Line 252: Lake "S"urface temperature -> Lake "s"urface temperature

Line 288: needs a punctuation after "day 251".

**4.2 Prediction of coastal up-welling for fishery and drinking water management**

Fig. 11:

(1) Correct the caption of the colorbar (remains selected?), and the same problem happened in Fig. D1.

(2) Can you show the observation data?

4.3 Prediction of storm surge events for public safety

Fig. 12:

(1) Can you show the spatial distribution of the 24-h and 96-h forecast to compare with

the time series in (d)?

(2) According to (a) to (c), the forecasted water level is highly dependent on forecast time, and can you show the relationship between water level and forecast time? (It does not seem to be saturated even if the latest data is used according to (a).