An automatic lake-model application using near real-time data forcing: Development of an operational forecast workflow (COASTLINES) for Lake Erie

Shuqi Lin[1], Leon Boegman[1], Shiliang Shan[2], Ryan Mulligan[1]

[1]Department of Civil Engineering, Queen's University, Kingston ON Canada K7L 3N6

[2]Department of Physics and Space Science, Royal Military College of Canada, Kingston ON Canada K7K 7B4.

Correspondence to: Shuqi Lin (shuqi.lin@queensu.ca)

Abstract. For enhanced public safety and water resource management, a three-dimensional operational lake hydrodynamic forecasting system COASTLINES (Canadian cOASTal and Lake forecastINg modEl System) was developed. The modelling system is built upon the Aquatic Ecosystem Model (AEM3D) model, with predictive simulation capabilities developed and tested for a large lake (i.e., Lake Erie). The open-access workflow derives model forcing, code execution, post-processing, and web-based visualization of the model outputs, including water level elevations and temperatures, in near real-time. COASTLINES also generates 240-h predictions using atmospheric forcing from 15 km and 25 km horizontal-resolution operational meteorological products from the Environment Canada Global Deterministic Forecast System (GDPS). Simulated water levels were validated against observations from 6 gauge stations, with model error increasing with forecast horizon. Satellite images and lake buoys were used to validate forecast lake surface temperature and the water column thermal stratification. The forecast lake surface temperature is as accurate as hindcasts, with a root-mean-square-deviation <2℃. COASTLINES predicted storm-surges and up-/down-welling events that are important for coastal flooding and drinking water/fishery management, respectively. Model forecasts are available in real-time at https://coastlines.engineering.queensu.ca/. This study provides an example of the successful development of an operational forecasting workflow, entirely driven by open-access data, that may be easily adapted to simulate aquatic systems or to drive other computational models, as required for management and public safety.

1 Introduction

Lakes hold a large proportion of the global surface freshwater, which supports biodiversity and supplies water resources for drinking, transportation, and recreation. However, anthropogenic stressors are causing significant changes in the properties of lakes, such as rapid warming of surface water (O'Reilly et al., 2015), large seasonal water level fluctuations (Gronewold and Rood, 2019), increased frequency of extreme events (Saber et al., 2020) and severe water quality issues such as oxygen depletion (Rowe et al., 2019; Scavia et al., 2014) harmful algal blooms (Brookes and Carey, 2011; Watson et al., 2016). Effort has focussed on investigating the long-term responses of physical processes in lakes to climate change (O'Reilly et al., 2015; Woolway and Merchant, 2019; Jabbari et al., 2021), but improving lake monitoring and developing short-term forecast models, to predict the occurrence of extreme events is also necessary (Woolway et al., 2020). The biogeochemical cycles in lakes are complex and often regulated by physical forcing; therefore, the first step to model and forecast water quality issues, such as harmful algal blooms (Paerl and Paul, 2012; O'Neil et al., 2012) and hypoxia (Rao et al., 2008; Rao et al., 2014) is the development of accurate hydrodynamic models.

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. These range from one-dimensional (1D) models such as DYRESM (Antenucci and Imerito, 2000), Simstrat (Gaudard et al., 2017), and GLM (Hipsey et al., 2014), to three-dimensional (3D) models such as Delft3D (Lesser et al., 2004), FVCOM (Chen et al., 2012; Rowe et al., 2019) ELCOM (Hodges et al., 2000). Several of these hydrodynamic models may be coupled to biogeochemical models to allow for prediction of water quality. In the case of hindcast applications, the complex and time-consuming setup and calibration procedure, of these models, can result in a significant time lag (months to years) between when a project is initiated and when the model results are communicated to stakeholders. This delay severely limits the utility of computational models for policy and management decision making. For better application of these powerful computational tools, the ability for rapid monitoring and simulation forecasts should be established.

In addition to the significant effort required to setup and calibrate models, other hurdles exist such as data-sharing agreements between the agencies collecting forcing/validation data and those running the models. For example, the US National Oceanic and Atmospheric Administration (NOAA) Great Lakes Coastal Forecasting System (Chu et al., 2011; Anderson et al., 2018), is a comparatively large-budget multi-institutional (NOAA-GLERL and U. Michigan-CIGLR) project that predicts water levels, temperature profiles, currents, and wave heights over a 120-h timeframe in the five Laurentian Great Lakes and connecting channels, using FVCOM on a 3D unstructured grid with 30-2000 m horizontal resolution. 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. These latter applications employ a data sharing agreement between Swiss Federal Institute of Aquatic Science and Technology (EAWAG), École Polytechnique Fédérale de Lausanne (EPFL) and MeteoSwiss.

Owing to the present online proliferation of near real-time lake observation data (e.g., National Data Buoy Center (NDBC; https://www.ndbc.noaa.gov/); Great Lakes Observation System (GLOS; https://www.glos.us/)) and

high-resolution meteorological forecasts (e.g., Global Deterministic Prediction System, GDPS; https://dd.weather.gc.ca/model_gem_global/; High Resolution Rapid Refresh, HRRR; https://rapidrefresh.noaa.gov/hrrr/) the data-sharing agreements and managed data transfer protocols are no longer required. When coupled with recent dramatic improvements in workflow efficiency (e.g., Gaudard et al. (2019); Baracchini et al. (2020b)) near real-time inclusion of forcing from meteorological forecasts allow for the development of specific simulations tailored to meet diverse lake-management needs (e.g., prediction of coastal flooding, spill modelling, fish habitat, beach closures, and optimization of treatment or source water monitoring strategies).

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. We provide an overview of the COASTLINES system, including model implementation for Lake Erie (Section 2 Data and methods) and the accuracy of COASTLINES in forecasting water levels and temperature fields over timescales of 24-h and 240-h (Section 3 Results). In Section 4 Discussion, the predictive ability of COASTLINES for decision making is showcased through prediction of hydrodynamic events associated with fish kills, hypoxia near a drinking water intake and coastal flooding from a storm surge. We also discuss the relative advantages of COASTLINES in comparison to other model products, including bias and uncertainty.

2 Data and methods

2.1 Study site

Lake Erie, the shallowest lake of the Great Lakes with a mean depth of 19 m. Lake-wide hydrodynamics predominantly exhibits free surface and current oscillations from the 14-h barotropic seiche (Hamblin, 1987; Boegman et al., 2001). The lake morphometry consists of distinct, yet interconnected western, central, and eastern basins (Fig. 1), each with its own water quality concerns: The 11-m deep western basin is typically well mixed and has frequent harmful algae blooms related to climate-driven meteorological forcing (Michalak et al., 2013). The ephemeral stratification in late summer (Loewen et al., 2007) regulates vertical biogeochemical fluxes (Boegman et al., 2008). The 20-m deep central basin is prone to large-scale hypolimnetic hypoxia (Scavia et al., 2014). Hydrodynamics are governed by an internal Poincaré wave (Bouffard et al., 2012; Valipour et al., 2015) and a bowl-shaped depression of the summer thermocline, which influence the oxygen budget (Beletsky et al., 2012; Bouffard et al., 2014). The 65-m deep eastern basin has nearshore water quality concerns from *cladophora* (Higgins et al., 2006) and ecosystem engineering by dreissenid mussels (Hecky et al., 2004). Hydrodynamics of this region are controlled by the coastal internal Kelvin wave (Valipour et al., 2019).

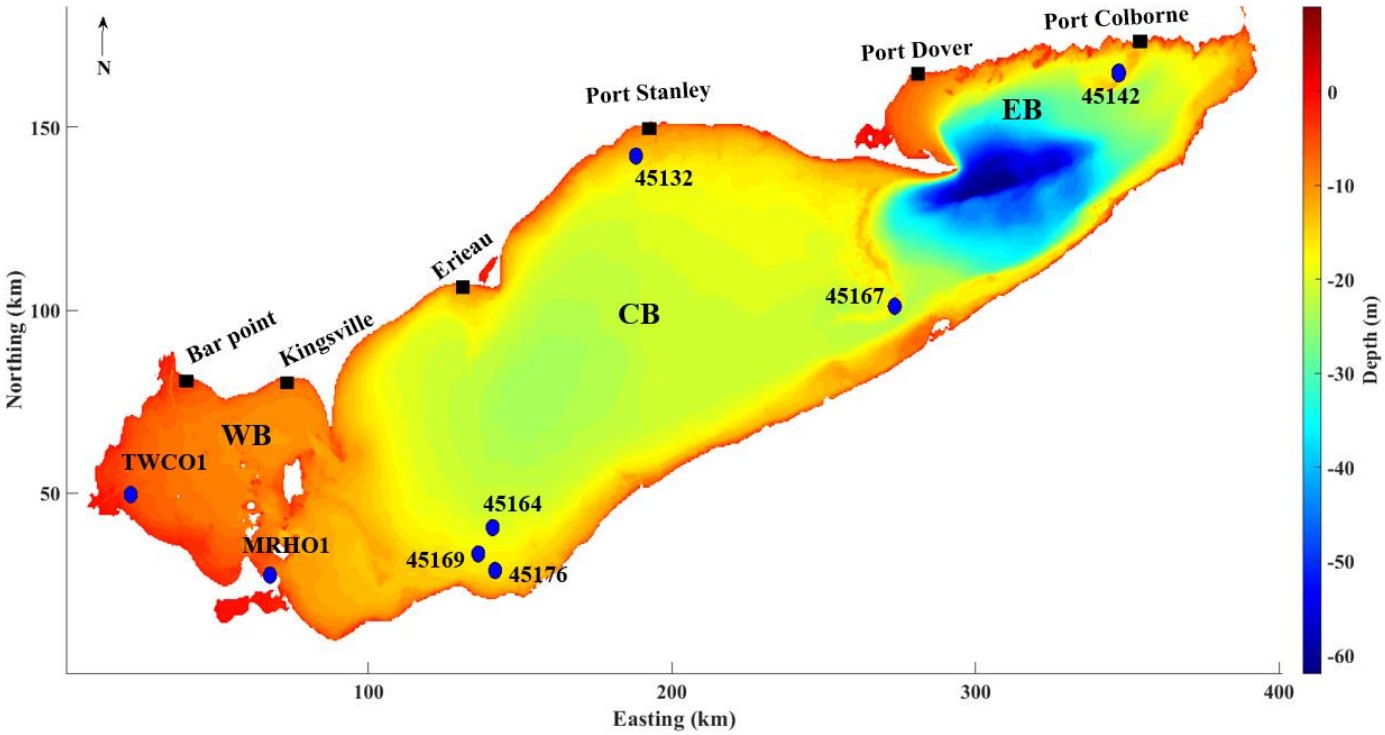


Fig.1 Map of Lake Erie showing the bathymetric depths and observation sites. The bathymetric map is at the resolution of the 500 m grid applied in the model. The western, central, and eastern basins are labeled as WB, CB, and EB, respectively. Blue circles indicate lake buoys and black squares indicate water level gauges.

2.2 Model description

COASTLINES applies the three-dimensional Aquatic Ecosystem Model (AEM3D, version 1.1.1, HydroNumerics Pty Ltd.). This model solves the unsteady 3D Reynolds-averaged Navier-Stokes equations for incompressible flow employing the Boussinesq and hydrostatic approximations. Momentum advection is based on the Euler-Lagrange method with a conjugate-gradient solution for the free-surface height (Casulli and Cheng, 1992), and a conservative ULTIMATE QUICKEST discretization scheme for advection of scalars (Leonard, 1991). AEM3D is a parallel version of the commonly applied Estuary and Lake Computer Model (ELCOM; (Hodges et al., 2000). ELCOM has been applied to Lake Erie to simulate currents and seasonal circulation (León et al., 2005), the internal Poincaré (Valipour et al., 2015) and Kelvin waves (Valipour et al., 2019), ice cover (Oveisy et al., 2012) and the response of the thermal structure, in Lake Erie, to air temperature and wind speed changes (Liu et al., 2014). ELCOM has been coupled with the biogeochemical CAEDYM model to simulate Lake Erie phytoplankton and nutrients (León et al., 2011), and the response of hypoxia (Bocaniov and Scavia, 2016) and algae blooms (Scavia et al., 2016) to nutrient load reductions. Recent applications of AEM3D include a study of the water level in Lake Arrowhead, California (Saber et al., 2020), ice cover in Lake Constance (Caramatti et al., 2019) and pollutant transport in Lake St. Clair (Madani et al., 2020).

2.3 Model setup and meteorological forcing variables

To adequately resolve the coastal boundary layer (~ 3 km width (Rao and Murthy, 2001)) and basin-scale internal waves (Poincaré (16.8 h) and Kelvin waves), the bathymetry of Lake Erie (https://www.ngdc.noaa.gov/mgg/greatlakes/erie.html) was discretized into a 500 m × 500 m horizontal grid, which is ~10 % of the internal Rossby radius (Schwab and Beletsky, 1998). The lake was discretized into 45 vertical layers, with fine resolution (0.5 m) through the surface layer, metalimnion and bottom of the central basin, and coarse layers (5 m) through the hypolimnion of the deeper eastern basin to the maximum depth of 64 m. The model was 'cold started' on April 8, 2020 (day of year (day) 99) with an initial temperature field spatially interpolated from observed water temperatures at stations 45142 and MHRO1; a time when spring turnover causes thermal stratification to be minimal. The model time step is dt = 300 s to satisfy the CFL = $\sqrt{2}$ (Courant-Friedrichs-Lewy) condition for internal waves (Hodges et al., 2000).

The model is forced by the surface meteorology (wind speed, wind direction, air temperature, shortwave solar radiation, relative humidity, air pressure, and net longwave radiation), with net longwave radiation being computed internally within AEM3D from cloud cover and the modelled surface temperature. In order to address the spatial variability of meteorological conditions across the lake, the computational domain was forced with meteorological data on horizontal grids at 15 km (https://dd.weather.gc.ca/model_gem_global/15km/) and 25 km (https://dd.weather.gc.ca/model_gem_global/25km/) resolution using meteorological forecasts from the Environment and Climate Change Canada Global Deterministic Forecast System (GDPS). This resulted in 31 and 23 meteorological sections for the 15 km and 25 km models, respectively. Wind speed, wind direction, air temperature, relative humidity, air pressure, dew point, and cloud cover are direct outputs from GDPS, with solar radiation calculated based on dew point and air pressure ((Meyers and Dale, 1983); Appendix C. in (Gaudard et al., 2019)). The meteorological forecast has an output timestep of 3-h and a forecast length of 240 h. The .GRIB2 meteorological data were retrieved with the 'urllib' library in Python and formatted into AEM3D input files using the nctoolbox in MATLAB.

In this pilot application, the Lake Erie inflows and outflows, which roughly balance, are neglected, however evaporation and precipitation are accounted for in the water balance. Over short timescales (<10 days), the contributions from evaporation and precipitation to water level change are minor, with water level oscillations resulting from storm surges and surface seiches (Trebitz, 2006).

2.4 Observations, implementation, and model validation

The water levels and temperatures simulated by COASTLINES were validated using both in situ and satellite observations. Near real-time water level data was used from six stations along the Canadian coastline, which reported hourly observations (Bar Point, Kingsville, Erieau, Port Stanley, Port Dover, and Port Colborne; Fig. 1; Table 1), retrieved from Fisheries and Oceans Canada (https://marees.gc.ca/eng/find/zone/44). The data are parsed using the 'BeautifulSoup' library in Python and saved as .csv files, to be read with MATLAB for model validation. The observations showed higher fluctuations in the western (Bar Point and Kingsville) and eastern (Port Dover and

Port Colborne) basins (Fig. 1). Thus, we quantify the water level forecast capability and uncertainty in terms of the Root Mean Square Deviation (RMSD) and Relative Error (RE):

$$RMSD = [\frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2]^{1/2}, \quad (1)$$

$$RE = 100\frac{RMSD}{log.\ mean(daily\ range)}, \quad (2)$$

where x_i and y_i (i = 1, 2, 3, ... N) are the model and observed water level timeseries and N is the number of samples. 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 can characterize the forecast bias, regardless of the instantaneous water level position. Here,~~The~~ forecast uncertainty ~~of forecast we mentioned following~~ is in the evaluation statistic from combining forecast dates – not actual uncertainty in an individual forecast~~indicates the error valuation based on statistical RMSD and RE, not the actual uncertainty in an individual forecast~~.

Eight in situ lake buoys, distributed over the nearshore areas of the three basins (Fig. 1; Table 1), provided near real-time model validation data through the Great Lakes Observing System (GLOS: https://www.glos.us/) and National Data Buoy Center (NDBC: https://www.ndbc.noaa.gov/) portals. For each station, the text-based NDBC observations are parsed using the 'BeautifulSoup' Python library, and the GLOS observations are retrieved using 'webdriver' from the 'selenium' Python library. All the lake buoy observations are saved as .csv files and read into MATLAB for post-processing. Attempts to retrieve missing variables would result in run-time errors.

The lake buoys are deployed from April or May through mid-October, spanning the spring/fall turnover and seasonal summer stratification periods. However, due to COVID-19 related delays in instrument deployments in 2020, only two buoys located offshore of Cleveland, near the water intake crib (station 45176 and station 45164), were equipped with thermistor chains to monitor temperature profiles. The other six buoys provide air and lake surface temperature as well as wind speed and direction observations for hydrodynamic and meteorological forecast validation. Satellite-based observations of lake surface temperature were obtained from the Great Lakes Surface Environmental Analysis (GLSEA2), which is derived from NOAA CoastWatch AVHRR (Advanced Very High-Resolution Radiometer) imagery and updated on NOAA GLERL website (https://coastwatch.glerl.noaa.gov/erddap/files/GLSEA_GCS/). GLSEA2 produced daily observations, with 2.6 km resolution, from the cloud-free portions of the satellite images (Schwab et al., 1999). The netCDF data are retrieved using the 'BeautifulSoup' library and 'webdriver' from 'selenium'.

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

$$MBD = 100\ \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)}{\frac{1}{N}\sum_{i=1}^{N} y_i} \quad (3)$$

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.

Table 1
Details of field stations with water level gauges and lake buoys.

Station	Parameter	Sampling interval (min)	Depth of measurement (m)
Bar Point	Water level	60	Surface
Kingsville	Water level	60	Surface
Erieau	Water level	60	Surface
Port Stanley	Water level	60	Surface
Port Dover	Water level	60	Surface
Port Colborne	Water level	60	Surface
TWCO1	Temperature	10	Surface
45005	Temperature	10	Surface
45176	Temperature	10	1, 3, 4, 6, 7, 9, 10, 12, 14, 15
45169	Temperature	30	surface
45164	Temperature	60	1, 2, 4, 6, 8 10
45132	Temperature	60	Surface
45167	Temperature	10	Surface
45142	Temperature	60	Surface

2.5 System operation

The COASTLINES operational forecast system is run on a local server supported by Queen's University ITS (Kingston, Canada). The COASTLINES workflow is presented in Fig. 2. The system consists of input data acquisition and preparation, 24-h hydrodynamic simulations, 240-h hydrodynamic simulations, validation against in situ observations, and uploading the model forecasts and validation to the web platform. Given that the standard deviations of meteorological forecast variables increase with forecast lead time (Buehner et al., 2015), we performed separate 24-h and 240-h forecast simulations each day. The model advances every day according to the 24-h forecast simulation and terminates by generating 're-start' files. These files are then used to hot-start the 240-h forecast simulation and the 24-h simulations for the next day. The input files for the 240-h forecast simulations are iteratively replaced by the new 240-h meteorological forecast generated each day. The 24-h and 240-h forecast model outputs are compared against observations to evaluate the forecast performance against forecast lead time.

The long-term stability of employing daily 'hot' restarts can be seen in a comparison between simulated temperature profiles from a continuous run and that from stitching together the 24-h hot-start simulations (Appendix A; Fig. A1). At present, the initial water level cannot be modified using the AEM3D re-start files. Therefore, to account for long term drift in surface water level, we used real time gauge observations as the datum point for water level forecasts (automatically performed by MATLAB in post processing) and only consider errors resulting from simulation of storm surges and seiches, as opposed to those from seasonal changes in mean lake level. Automation of the processing tasks in the workflow is performed by Python scripts triggered by the Windows Task Scheduler every 24-h at midnight. The online meteorological forecast data are retrieved from GDPS once updated at 3 am EST. Forcing variables are then formatted in MATLAB, called by the Python scripts once the meteorological

forecast data has been retrieved. The AEM3D pre-compiled executable is then run as a black-box code, triggered by Python. The 24-h and 240-h simulations take 0.5 h and 4 h to complete, respectively. The observed data, including water levels from gauge stations, water temperatures from lake buoys and satellite images are scraped with Python at 8 am, followed by post-processing in MATLAB to validate model output, calculate statistical metrics (RMSD, MBD) and generate figures. The results are exported to Google sheets and published to the COASTLINES website (e.g., Appendix B). The authors (supervisors of COASTLINES) and Queen's ITS monitor forecast results and maintain system operation.

Global coverage of the GDPS forecasts enable this operational system to be readily implemented at other sites where lake bathymetry, boundary flows and in-situ validation data are available. The workflow may be easily modified to include additional meteorological forecasts or other black-box hydrodynamic drivers (e.g., HRRR and Delft3D, respectively (Rey and Mulligan, 2021)). This would require simple modification of the COASTLINES MATLAB-based write statements to meet the formatting requirements of a particular driver. ~~Although the code in the system for formatting input files of hydrodynamic simulation is specific for AEM3D models, it can be easily modified to generate other formats of input files within MATLAB.~~

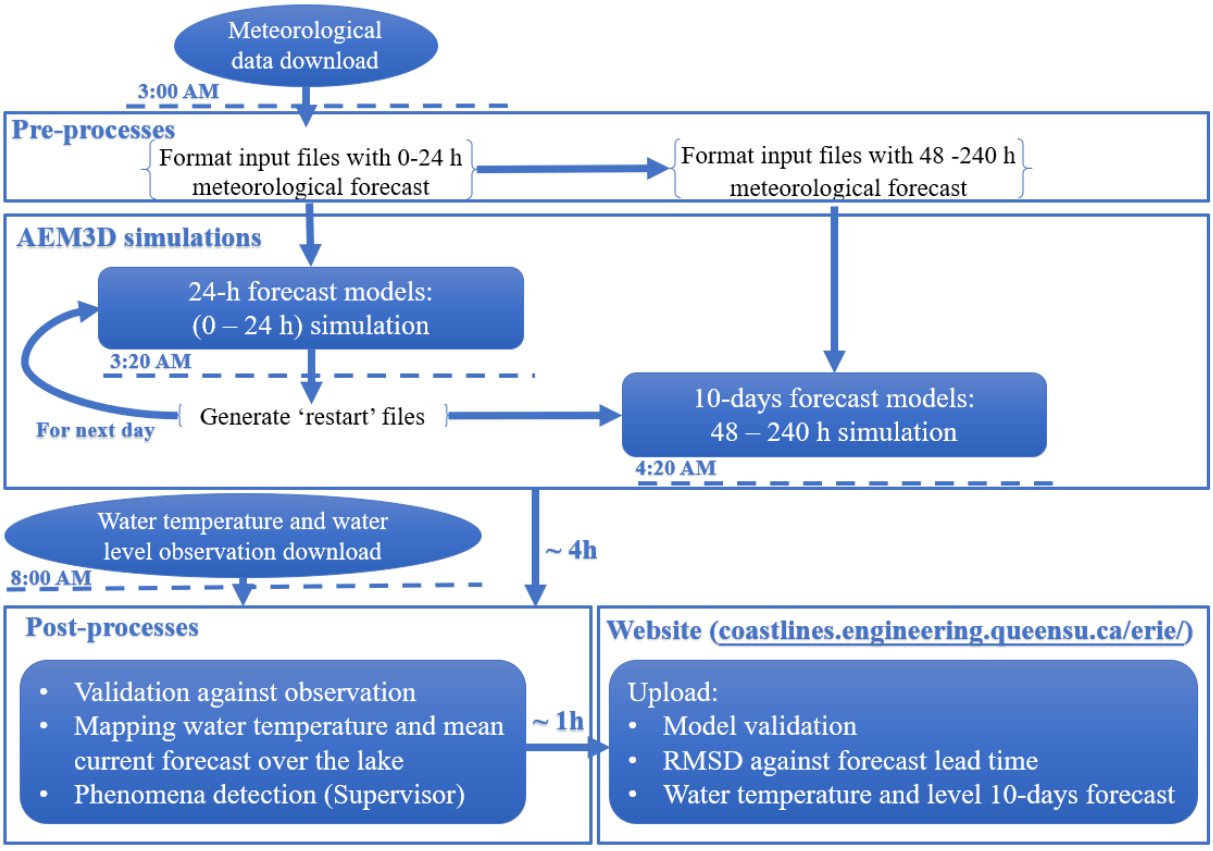


Fig. 2 Daily workflow and automated processes in the COASTLINES operational system as performed on the local server.

3 Results

The COASTLINES water level and temperature forecasts have been operational since April and July 2020, respectively. The 24-h and 240-h forecast water levels from the 15 km and 25 km resolution models were validated against real-time gauge station observations. 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.

3.1 Water level

The Relative Error (RE) of the forecast water level increases with forecast time when averaged over April to September 2020; the 24-h forecast error being ~ 40% at all six gauge stations (Fig. 3 a, c, e, g, i, k). Given the large water level fluctuation at Port Colborne (Fig. 3 l), the 240-h forecast RE is highest at this station, exceeding 70% (Fig. 3 k). Of the six gauge stations reported in this study, those at the western (Bar Point and Kingsville) and eastern (Port Dover and Port Colborne) ends of Lake Erie longitudinal axis had the largest water level fluctuations, resulting from the predominant south-westerly winds generating strong wind set-up and surface seiches (Fig. 3 b, d, f, h, j, l). The lognormal means of the daily range in water level at the six gauge stations are 0.21 cm (Bar Point), 0.16 cm (Kingsville), 0.07 cm (Erieau), 0.10 cm (Port Stanley), 0.15 cm (Port Dover), 0.17 cm (Port Colborne). The 24-h forecasts show qualitative agreement with observations in phase and magnitude (Fig. 4). The 24-h forecasts reproduce the dramatic surface seiches induced by westerly winds > 15 m s^{-1} (Fig. C2) on day 251 (RMSD < 0.1 cm), especially the obvious water level fluctuations at stations in the western and eastern basins (Fig. 4 a, b, e). However, the prediction of water level at Bar Point showed large bias (Fig. 4 f), with the model overestimating the water level fluctuation. The uncertainty in the model forecast, which increased with the range of the daily fluctuation, was captured by the ~~ensemble~~ 24-h forecast RE over April to September (the shaded areas in Fig. 4). Overall, the confidence interval of the 24-h forecast included most of the discrepancies between the observations and the model results.

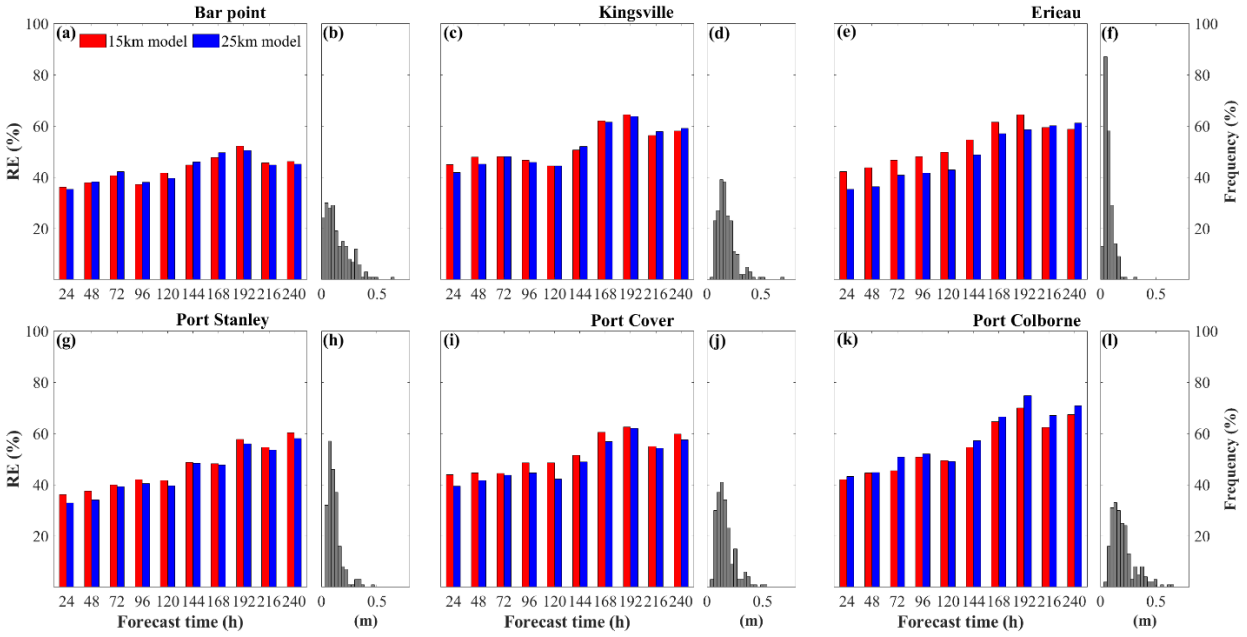


Fig. 3 Relative error (RE) in water level predictions against forecast time at six stations (a, c, e, g, i, k). Panels (b, d, f, h, j, l) are the corresponding frequency distribution of lognormal means of the daily water level fluctuation range (x-axes, unit in meter) at Bar Point, Kingsville, Erieau, Stanley, Port Dover, Port Colborne, respectively.

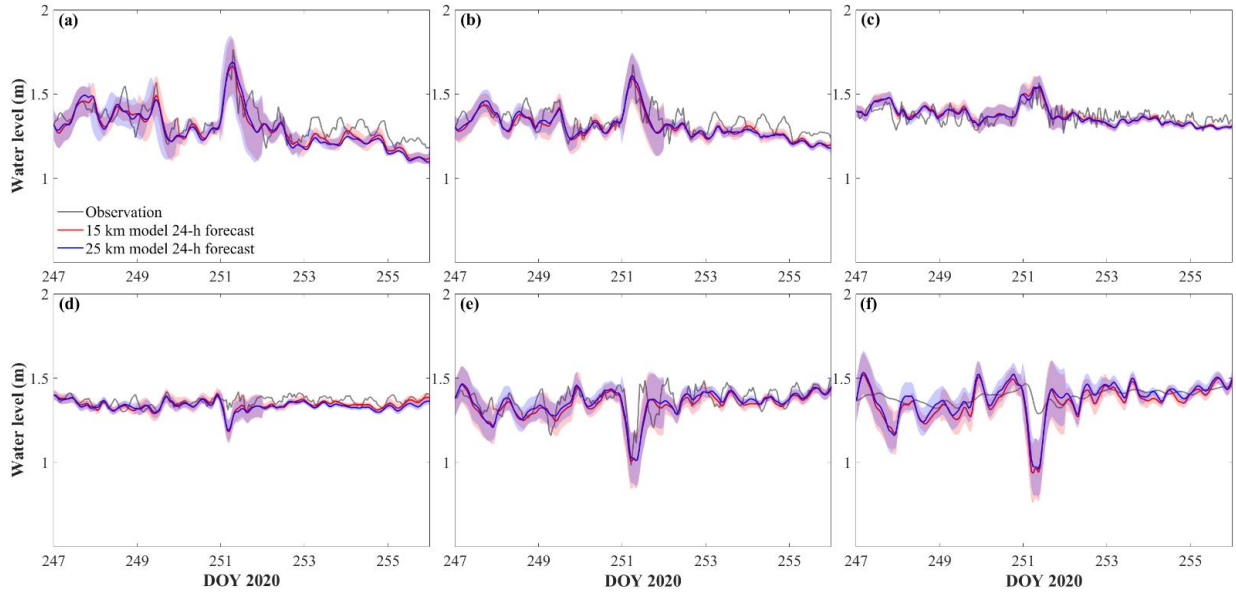


Fig. 4 Comparison between observed and stitched 24-h forecast modeled water level at (a) Port Colborne, (b) Port Dover, (c) Port Stanley, (d) Erieau, (e) Kingsville, and (f) Bar Point. The shaded areas show the confidence interval of the 15 km model (red shading) and the 25 km model (blue shading), as given by the ~~ensemble~~ 24-h RE in Fig. 3.

Timeseries validations for the 240-h model forecast (Fig. 5) include confidence intervals from the ~~ensemble~~ RE (Fig. 3). As shown, the forecast began 6 days in advance of the large surface seiche event on day 251 and predicted

the seiche to crest at Port Colborne 1-2 h ahead of the observations, and to trough at Kingsville 1-2 h behind the observations (Fig. 5 a, c). Damping of the seiche oscillations (~144 hours in the future) was excessive, with the water levels being underestimated and the phase shifted by approximately 12 hours (Fig 5. a, b). Despite the wide confidence intervals, due to the increasing RE with forecast time, large bias existed after the seiche event (forecast time >168 hours). When the forecast was initiated close to the event (3 days before), the prediction of seiche phase was more accurate (Fig. 5 d, e, f); however, the seiche decay still had a 12-h phase shift. The discrepancies in seiche amplitude (< 0.1 m) were within the confidence intervals of the models.

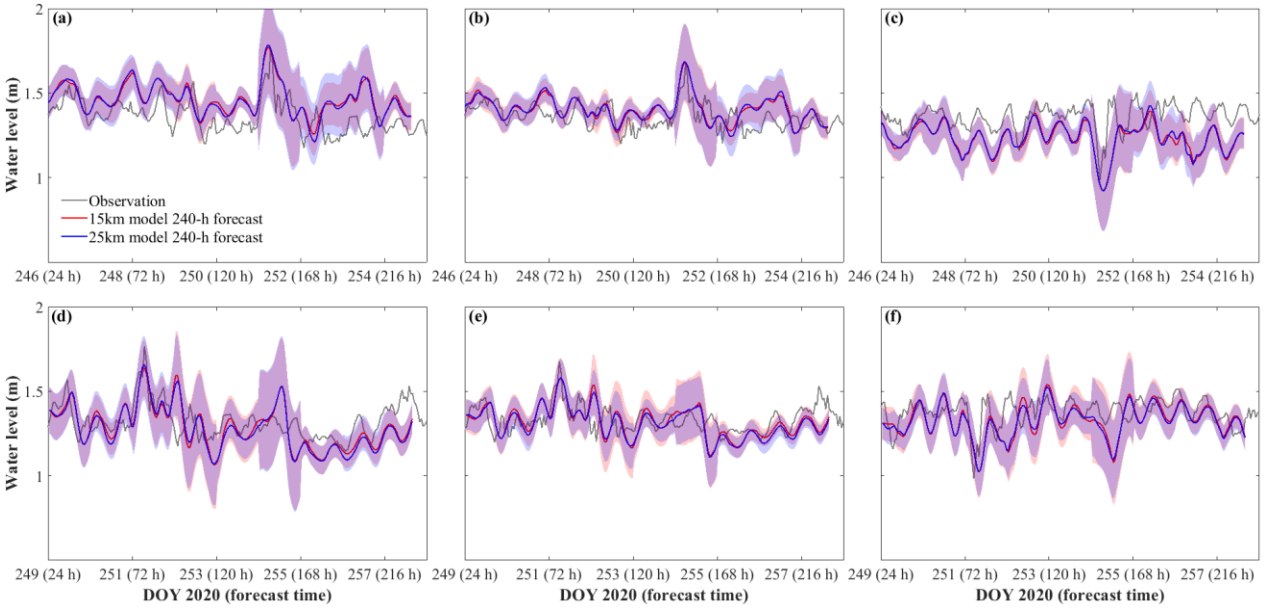


Fig. 5 Comparison between the observed water level and 240-h forecast hot-started on day 245 (a, b, c) and day 248 (d, e, f) at Port Colborne, Port Dover, and Kingsville, respectively. The shaded areas show the confidence interval of the 15 km model (red shading) and the 25 km model (blue shading), as given by the ~~ensemble~~ 240-h RE in Fig. 4.

3.2 Water temperature

3.2.1 Lake surface temperature

Using satellite-based and lake buoy-based observations, we evaluated the lake surface temperature forecast (Fig. 6). The 24-forecast captured the seasonal variation of lake surface temperature, particularly the rapid increase in temperature on days 180-190, and the gradual decrease in temperature after day 240; at all eight stations. However, the forecast overestimated the lake surface temperature in July by 3-5 ℃ (days 180-210), particularly at STN 45167 and 45142. Due to the 3-h output interval associated with the meteorological forecast data, the forecast model was insensitive to temperature fluctuations over shorter timescales, as recorded by the lake buoys, and it underestimated the sudden decrease in temperature near day 220 and 255 at STN 45005.

Overall, the t-MBD and t-RMSD, over these eight stations, were ~6% and 1.4 ℃ (15 km model) and ~5% 1.3 ℃ (25 km model) for the 24-h forecast, respectively (Table 2). The average s-MBD and s-RMSD over the 50 days from July-September were ~4% and 1.2 ℃, respectively, for both 15 km and 25 km resolution models.

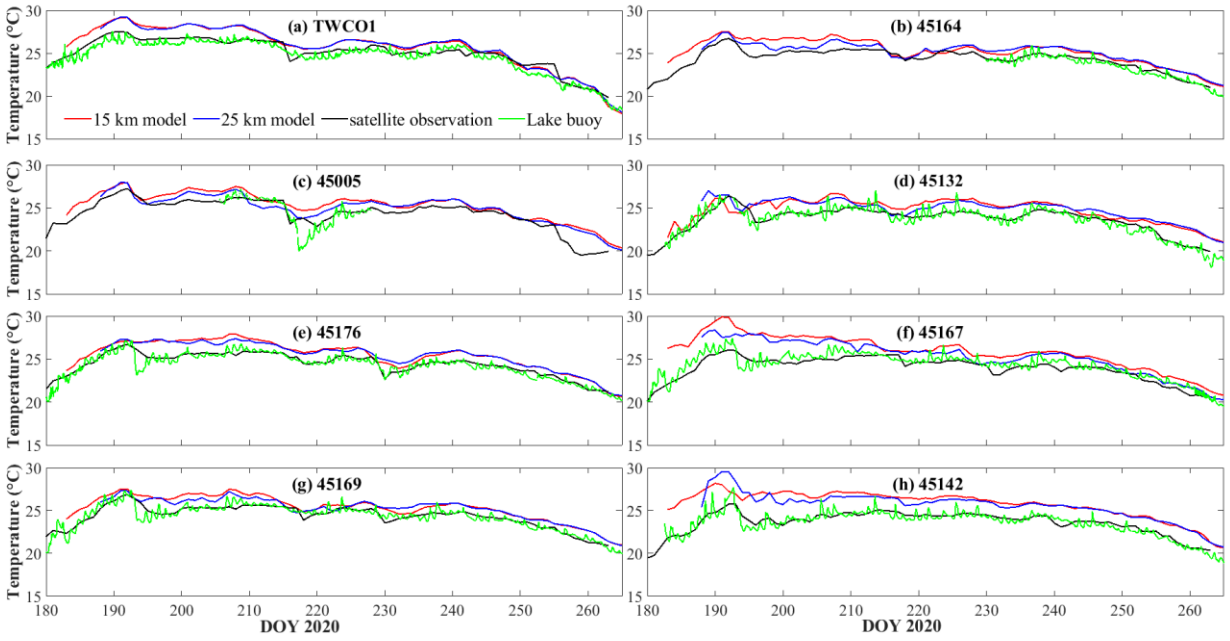


Fig. 6 Comparison between the stitched 24-h forecast and observed lake surface temperature at 8 stations (a) TWCO1, (b) 45164, (c) 45005, (d) 45132, (e) 45176, (f) 45167, (g) 45169, and (h) 45142. The green lines are timeseries observations from lake buoys, the black lines are daily observations derived from satellite imagery.

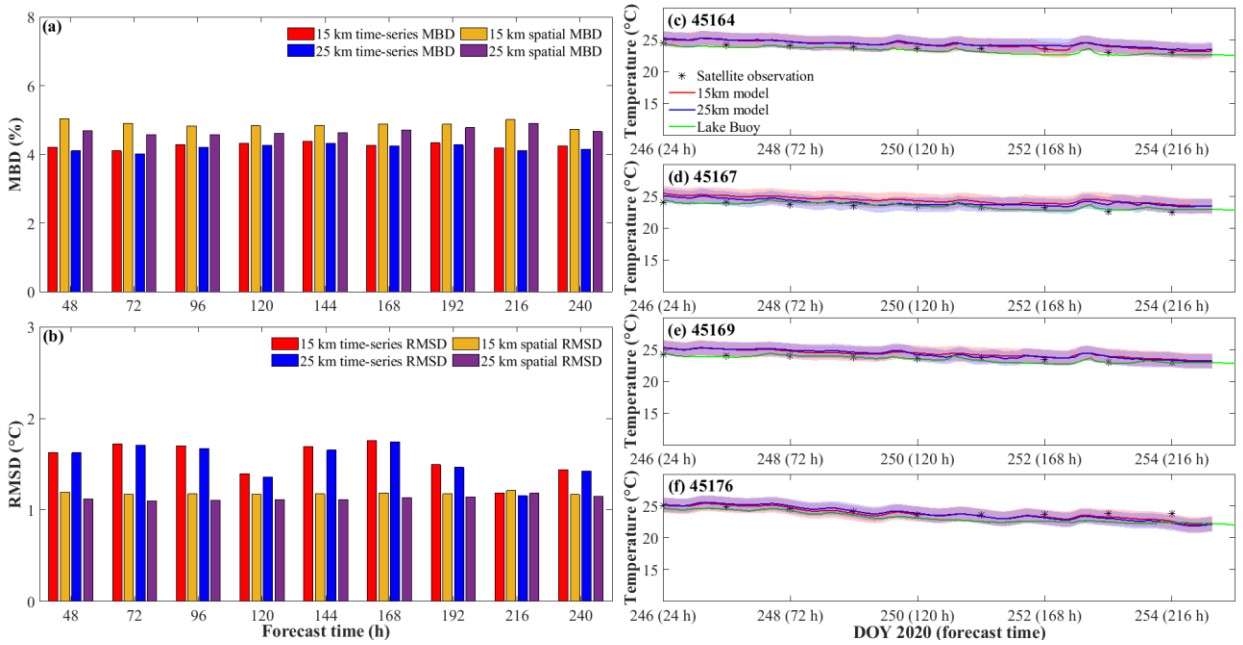


Fig. 7 (a) Mean-Bias Deviation (MBD) against forecast time; (b) Root-Mean-Square Deviation (RMSD) against forecast time. (c-f) Timeseries of 240-h forecast and observed lake surface temperature at stations 45164, 45167, 45169, 45176, respectively, and daily averaged satellite lake surface temperature (black asterisks). The confidence interval (shaded areas) in (c-f) represents the uncertainty of the 240-h forecast model through the timeseries RMSD with the forecast time (panel b).

The 240-h forecast MBD and RMSD, surprisingly, do not show an increase in error with forecast time (Fig. 7 a, b). Both t-MBD and s-MBD, over the 240-h forecast, are ~4-5%, with s-MBD 0.5-1% higher than t-MBD. Although

both 240-h s- and t-RMSD are under 2 ℃, the t-RMSD show the error with forecast time to be higher than s-RMSD. Both timeseries observations from lake buoys and daily averaged observations from satellite imagery fall into the forecast confidence interval based on the 240-h t-RMSD (Fig 7 c-f).

Spatial comparisons of satellite-based observations to the 24-h, 48-h, 120-h, 168-h surface temperature forecasts illustrate that the forecast system (with 15 km meteorological data) predicted the cooler water mass along the northwest shoreline of the central basin with a cold bias ~ 2 ℃ (Fig. 8); this may be up-welling hypolimnetic water (see following Discussion 4.2). The model also predicted lower surface temperatures in coastal regions of the western basin with a cold bias ~2 ℃ (Fig. 8 m-t); the bias presumably was induced by neglecting riverine inflows (e.g., Detroit River and Maumee River; see also Discussion 4.3), which are typically near the air temperature and several degrees warmer than the lake surface (Wang and Boegman, 2021). Further comparisons between model predictions and satellite-based observations of lake surface temperature can be found in the Supporting material (Fig. D1-2).

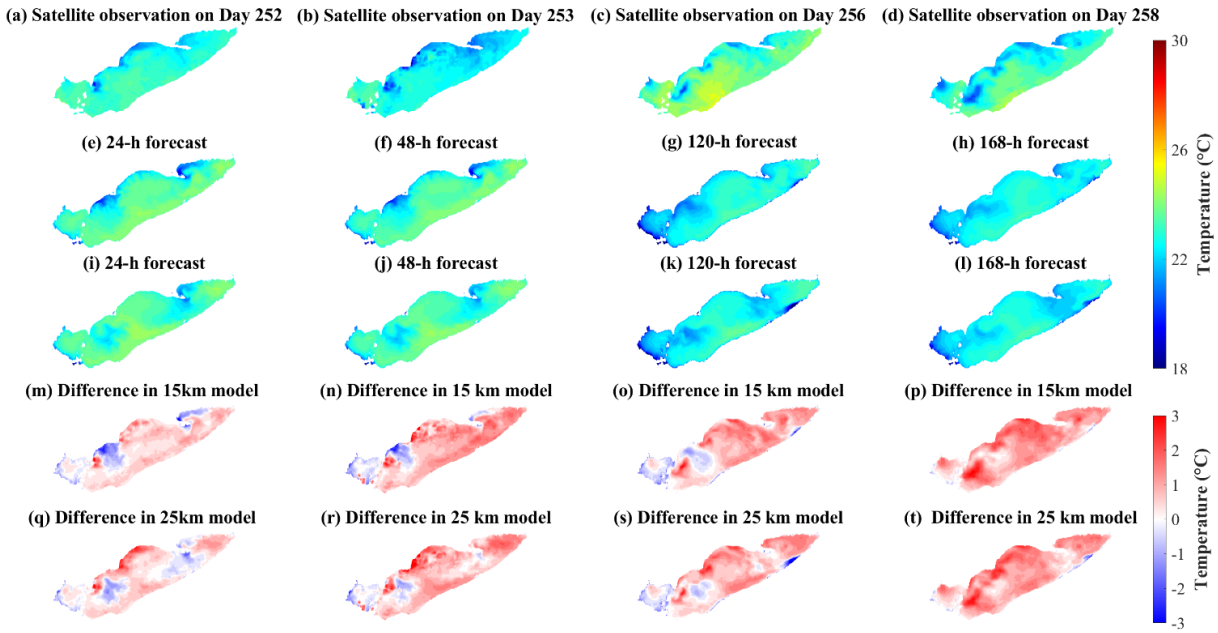


Fig. 8 Comparison of lake surface temperature from (a-d) satellite observations, (e-h) 15 km model forecast, and (i-l) 25 km model forecast during late summer. The models were hot-started on day 251. The difference between observations and models are shown in (m-t).

3.2.2 Thermal structure

The 3D AEM3D model structure applied in COASTLINES enables the prediction of the thermal profiles in the lake. On 15 Jun. 2020 (day 168), a rapid drop (~ 6℃) in surface temperature, was recorded by the thermistor at STN 45176, and predicted by the stitched 24-h COASTLINES model (15 km meteorological input) (Fig. 9 a, b). The timing and intensity of this up-welling event were accurately forecast, but before and after the upwelling event, the mixed layer depth was modelled to be deeper than observed; perhaps a result of spurious numerical diffusion resulting from the

thermocline swashing along the stair-step z-level grid at the lake perimeter. The 240-h forecast model was not yet operational at this time.

Both the 240-h 15 km and 25 km resolution forecasts predicted the down-welling event on 11 Jul. 2020 (day 193) at STN 45176 (Fig. 10). The forecasts were hot-started 7 days before the event (day 187), successfully predicting when warm surface water down-welled toward the bed displacing the thermocline (Fig. 10 b, c), but the 15 km resolution underestimated the intensity of down-welling, predicting thermocline recovery on day 193. The forecast hot-started 5 days before the event (day 189) gave a more accurate prediction with the down-welling persisting over 2 days (Fig. 10 d, e) – as observed (Fig. 10 a).

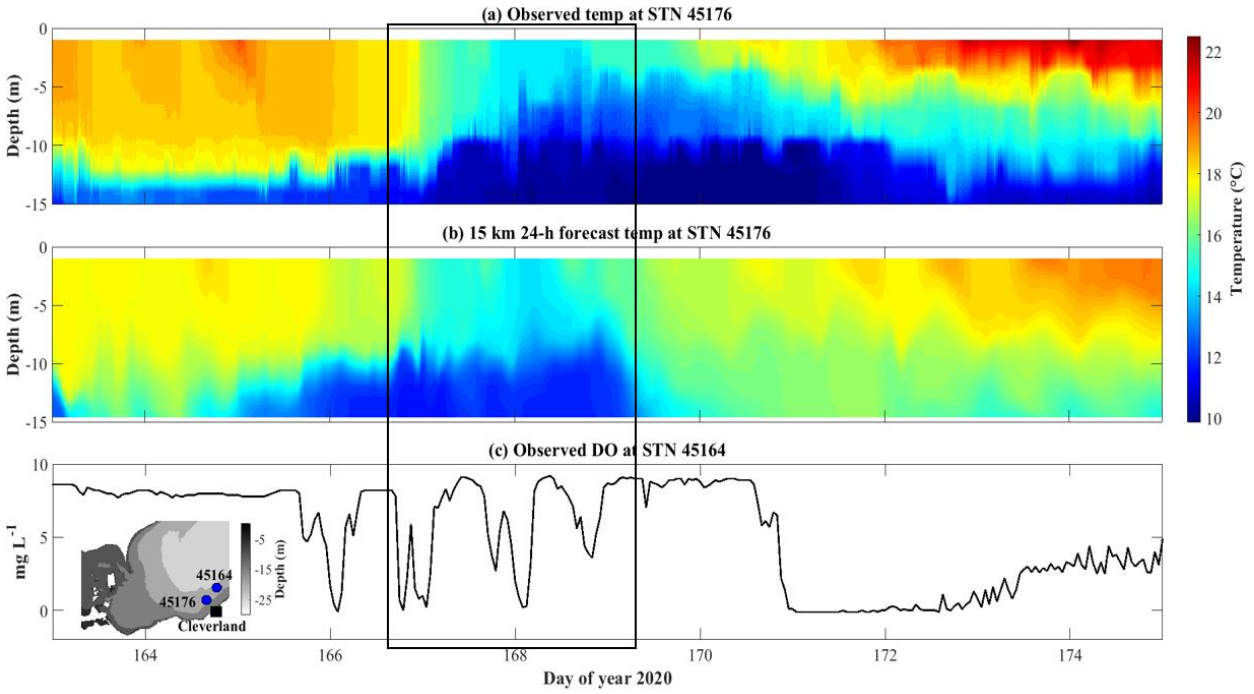


Fig. 9 Temperature profile comparisons between (a) observations and (b) stitched daily 24-h forecasts from the 15 km resolution model at station 45176. (c) Observed dissolved oxygen concentration at station 45164 from lake buoy (https://www.glos.us/). The inset image shows the bathymetry and locations of lake buoys. The black square indicates the timing of the up-welling event.

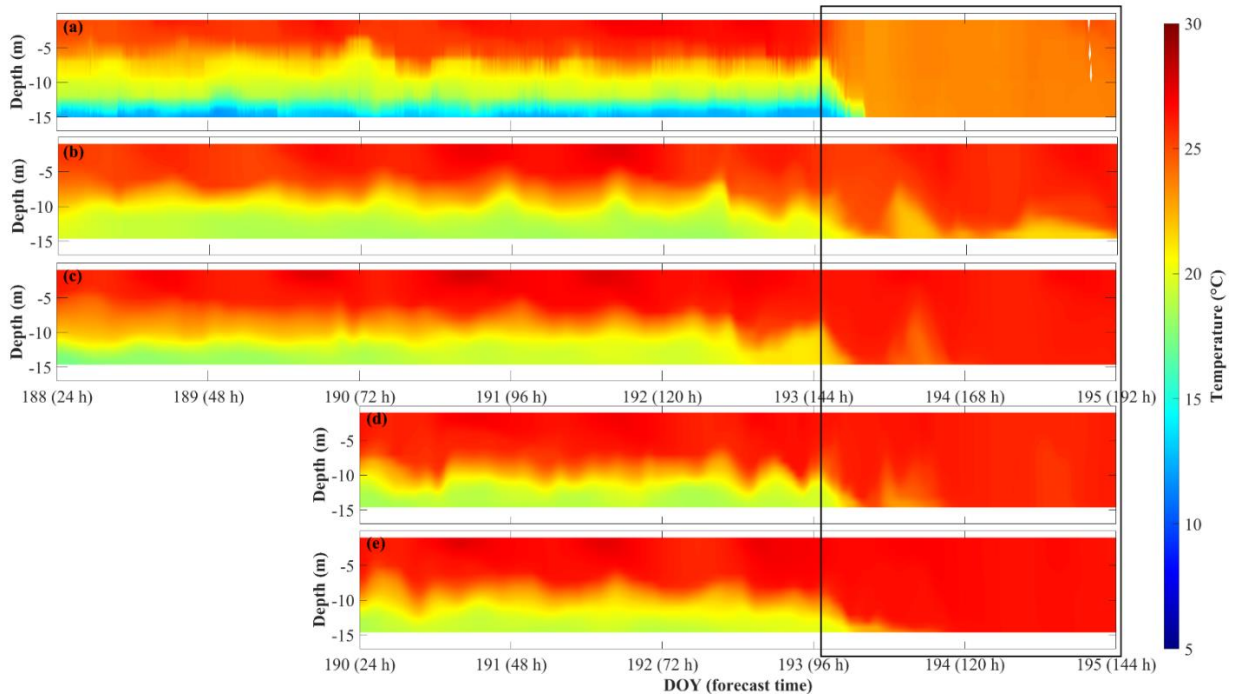


Fig. 10 Comparisons of (a) observed temperature profile, (b, d) 240-h 15 km resolution modeled, and (c, e) 240-h 25 km resolution modeled temperature profiles at STN 45176. The forecast models were hot-started on day 187 (b, c), and day 189 (d, e). The black square indicates the down-welling event.

Table 2
Statistical measures of t-MBD (Mean-Bias Deviation) and t-RMSD (Root-Mean-Square Deviation) between the 24-h forecast model and observations of water temperature.

Station	RMSD (°C)		MBD (%)	
	15 km model	25 km model	15 km model	25 km model
45176	2.6	2.6	6.8	6.8
45164	1.8	2.1	2.2	2.3
45132	1.5	1.5	5.5	5.7
45142	2.4	2.1	9.9	8.8
45167	1.2	1.1	4.6	4.0
45169	1.3	1.2	4.7	4.6
TWCO1	1.0	1.0	3	1.9
45005	1.2	1.1	8.2	7.9

4 Discussion

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

The central basin of Lake Erie is vulnerable to hypoxia in summer from near-bed thermal stratification and the relatively large ratio of sediment area to hypolimnetic volume (Bouffard et al., 2013; Nakhaei et al., 2021). Associated fish kill events (10s of thousands) are regularly reported, including an event on north shore of the central basin in the late summer of 2012, which was presumably was caused by up-welling of cold anoxic water from the hypolimnion (MOE, 2012; Rao et al., 2014). Similarly, 1000s of freshwater drum were killed in a rapid warming event (~5 °C /week) in the western basin in 2020 (https://www.13abc.com/content/news/Hundreds-of-dead-fish-

wash-up-in-Sandusky-Bay-571025541.html). Similarly, shoreward advection of hypoxic water, from up-welling or internal waves also adversely affects source water quality at drinking water intakes (https://epa.ohio.gov), whereby high Fe and Mn or low pH, associated with hypoxic water requires adjustment of treatment processes. This is a particular issue along the Ohio coast of the central basin (Ruberg et al., 2008; Rowe et al., 2019).

The ability to predict these movements of hypolimnion water would aid management of both Lake Erie fisheries and drinking water treatment. Here, we test the ability of the model to predict up-welling of cold bottom water in the region where the fish kill was observed in 2012. On days 249-253, 2020 (Fig. 8) strong southwesterly winds (~ 12 m s^{-1}; Fig. C2) were modelled and observed to create up-welling along the north shore, as expected from Ekman drift of the surface layer (Jabbari et al., 2019). The upwelled cold hypolimnetic water is shown near the coast of Erieau in the satellite observations and the 15 km resolution model (Fig. 8 a, b, e, f). The depth-averaged water temperature and current circulation in the forecast shows up-welling to persist for several days (Fig. 11), with cold hypolimnetic water accumulating along north shore and strong eastward currents along the northern shoreline of the east-central basin. The up-welling region matched that shown in a 2013 hindcast simulation (Valipour et al., 2019), revealing the hotspots of vertical transport of nutrients and anoxic hypolimnetic water.

Another up-welling event occurred near the Cleveland drinking water intake crib on days 167-170 (Fig. 9). This event was accompanied by simultaneous ~8 mg L^{-1} oscillations in the observed dissolved oxygen concentration (Fig. 9 c) at STN 45164 (~20 km away from STN 45176), followed by the dissolved oxygen concentration becoming locally hypoxic (< 2 mg/L) for ~2 days. The COASTLINES model predicted this event (section 3.2.2), which would have provided sufficient notice for drinking water plane operators to implement the additional treatment required for hypoxic water (Rowe et al., 2019). Future work, using the coupled iWaterQuality biogeochemical module (formerly CAEDYM) could extend COASTLINES to forecast water quality in Lake Erie (León et al., 2011), including dissolved oxygen concentrations and formation of algae blooms (Bocaniov et al., 2020).

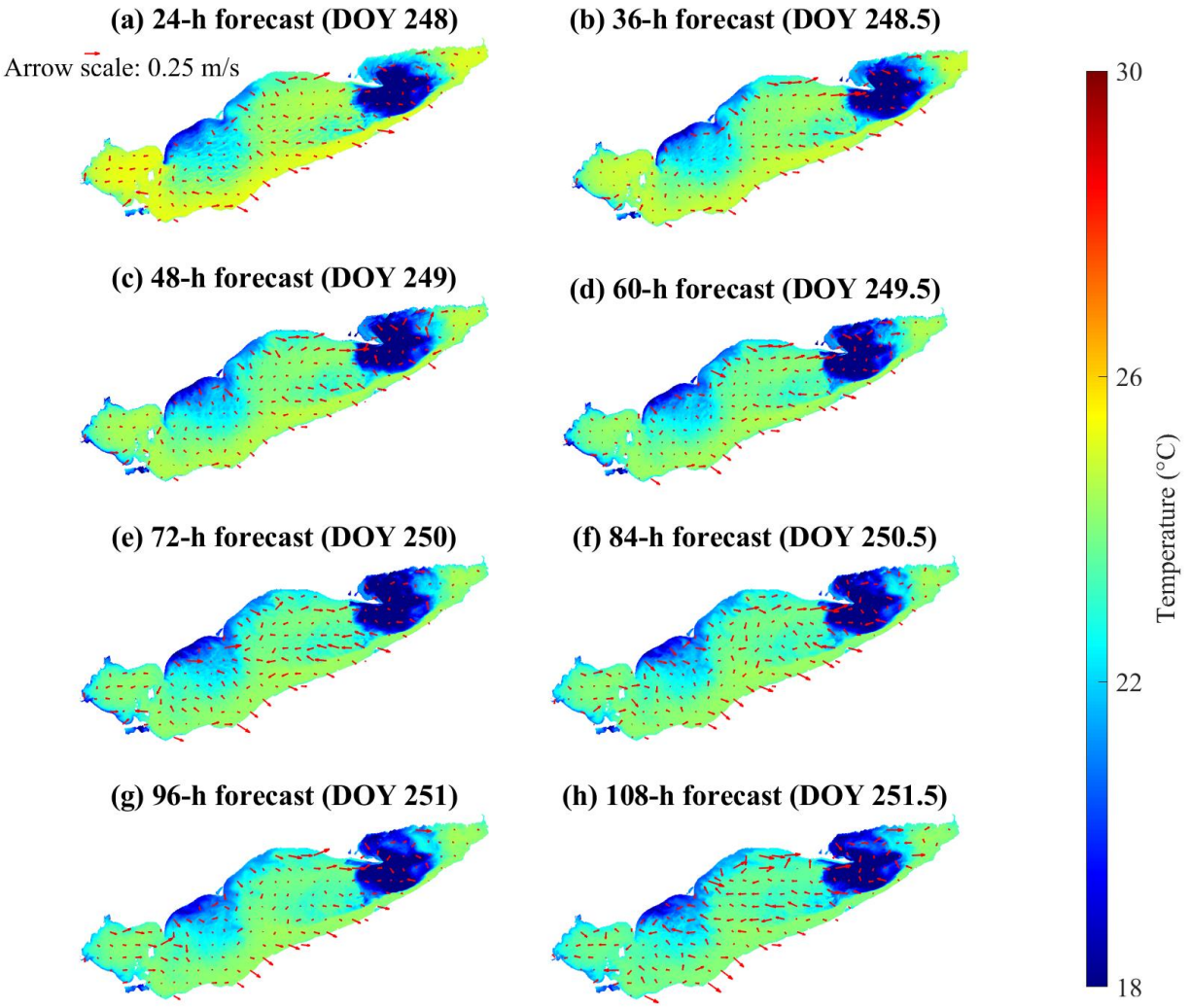


Fig. 11 Color maps showing the forecast depth-averaged temperature throughout the lake. The red arrows represent forecast depth-averaged currents. The model results are from the 240-h forecast model hot-started on day 247.

4.2 Prediction of storm surge events for public safety

Due to its shallowness and long fetch aligned with the predominant southwest winds (Hamblin, 1979), Lake Erie has the largest daily range of water level amongst the Great Lakes (Trebitz, 2006); these water level fluctuations are mainly due to storm surges and surface seiches (Mortimer, 1987). In every month of 2020, Lake Erie set new mean water level records (https://www.tides.gc.ca/C&A/bulletin-eng.html), causing the shoreline to be exposed to high risk from erosion and flooding and making the shoreline communities susceptible to costly damage and economic loss (e.g., https://www.lowerthames-conservation.on.ca/flood-forecasting/flood-notices/). Given the ability of COASTLINES to predict water level fluctuations induced by storm surges and seiches (Fig. 3, 5), we tested the ability of the model to act as a coastal flooding warning system. Due to the unpredictability and severity of water

level fluctuations in Lake Erie, there is currently a need to improve short-term water level forecasts and water level warning systems (Gronewold and Stow, 2014). This would assist early decision making during natural hazards (Gronewold and Rood, 2019).

We forecast the storm event on 15 Nov. 2020, which generated a wind-induced storm surge (~1-1.5 m) in the eastern basin with associated strong surface currents (Fig. 12). The inset image, taken during the event, shows the coastal flooding from this event. COASTLINES successfully predicted the high-water level phase at Port Dover 72 hours in advance, but underestimated the water level increase by 0.5 m. The hot-start forecast 24 hours in advance was more accurate in predicting the water level prediction, with a difference <0.5 m from the observations (Fig. 12 d). Note that both forecasts missed the small (~0.5 m) seiche before the significant increase at the end of day 320, presumably due to the low temporal resolution of the meteorological forecast input or local topography near the gauge. The overall wind-induced tilt of the free surface was less from the 72 hours hot-start, relative to the 24 hour hot start (Fig. E1), which predicted a larger local storm surge (Fig. 12d).

The impacts of coastal flooding could be improved by including simulation of wind-waves through enabling the coupled surface wave model SWAN (Booij et al., 1999) in AEM3D. Similarly coupled Delft3D-SWAN models have recently been applied in the development of a real-time predictive system for the coastal ocean and large estuaries (Rey and Mulligan, 2021).

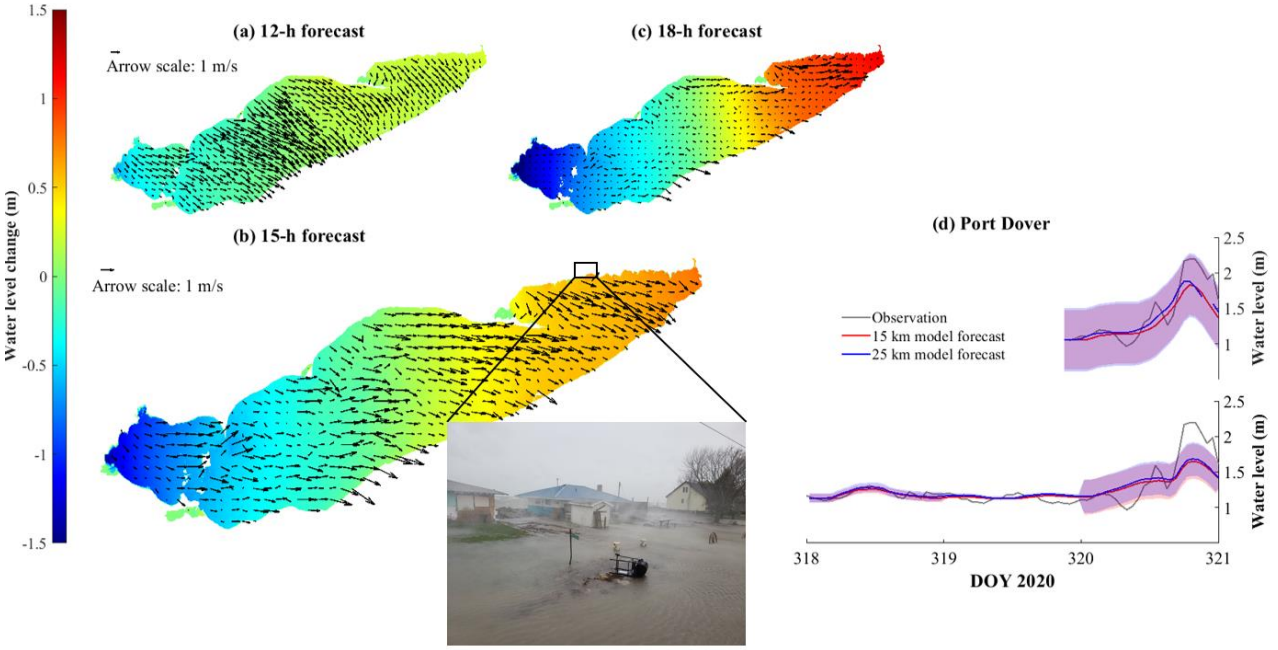


Fig. 12 Color maps showing the water level change compared to Nov 15th 00h from (a) 12 h, (b) 15 h, and (c) 18 h forecasts from 15 km resolution model. The black arrows are depth-averaged mean current fields. Panel (d) shows a comparison between forecast and observed water level at Port Dover. The upper panel shows the forecast hot-started on 15 Nov. 2020 (day 320), and the lower panel shows the forecast hot-started on 12 Nov. 2020 (day 317). The shaded region indicates the confidence interval. The inset image (extracted from a footage by J. Homewood from Lower Thames Valley Conservation Authority) shows the flooding induced by the dramatic water level increase during this event. The two cottages shown in the images were destroyed later in the afternoon.

4.3 Bias and uncertainty

The AEM3D model (formerly ELCOM) employed in COASTLINES has shown skill in temperature hindcasts in the Great Lakes with RMSD ~ 0.9 – 3 ℃ in Lake Erie (Liu et al., 2014; Oveisy et al., 2012) and 1.5 – 1.9 ℃ in Lake Ontario (Paturi et al., 2012). Similarly, the 24-h COASTLINES forecast predicted water temperatures with an average s-RMSD and t-RMSD < 2 ℃ at the surface (Table 2). Therefore, the forecasts are within ~1 ℃ RMSD in comparison to hindcasts, showing sufficient model skill for predictive simulations to aid lake management (e.g., movements of hypoxic water, fish thermal habitat, etc.). The accuracy of the COASTLINES forecasts may result from the high spatial resolution and coverage of meteorological forecast compensating for the inherent inaccuracies in the weather forecast data. Errors in forcing data may be compensated for using data assimilation (Baracchini et al., 2020b). In the hindcast models, Liu et al. (2014) applied uniform Lake Erie meteorological forcing over 4 zones and Valipour et al. (2019) utilized 6 zones, each spanning ~100 km. These included land-based observations, when there was no available lake buoy data, which induces error, especially in large shallow lakes (Hamblin, 1987). The comparatively high-resolution GDPS meteorological forecast was four to five times higher in horizontal resolution than used in the hindcast simulations, improving the representation of regional meteorological and climatological conditions.

Compared to other operational lake forecast systems, the 240-h COASTLINES forecast is longer (e.g., GLCFS forecasts 120 hours and meteolakes.ch forecasts 108 hours) and is the only one forced with open-access meteorological data that has global coverage. The GLCFS provided 48-h water level forecasts with RMSD ~0.12 m at the Buffalo gauge and ~0.14 m at the Toledo gauge, corresponding to RE ~ 60% and 51%, respectively (O'connor et al., 1999; Trebitz, 2006); using the older 4 km grid Princeton Ocean Model implementation, as opposed to the newer unstructured grid FVCOM GLCFS. COASTLINES gives better 48-h water level forecast performance (RE ~ 40 %) at six gauge stations. For temperature, benefitting from a smaller domain, finer resolution meteorological input (~2.2 km) and data assimilation, the 4.5-day lake surface temperature predicted by meteolakes.ch has a RMSD = 0.8 ℃ (Baracchini et al., 2020b), whereas COASTLINES predicted the 120-h (5 d) lake surface temperature with RMSD ~ 1.7 ℃. 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.

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. (2020b) reduced the RMSE temperature simulation of Lake Geneva from ~ 2 ℃ to ~ 1 ℃ by employing a sequential data assimilation routine; this would correspond to a <5% improvement in simulation of Lake Erie summer surface temperature. Before implementing data assimilation in our system, the limitations of such a scheme must be considered: (*i*) As we are performing forecasts, not hindcasts, we are unable to assimilate observations during model runtime (e.g., as done in the NCEP North American Regional Reanalysis: NARR);~~The lack of observations in the future, makes data assimilation impossible for adjusting forecasts;~~ ~~(*ii*) data assimilation is computationally intensive,~~

~~required ~1 month of computational time (Baracchini et al., 2020a), clearly not an option for operational forecasting); and~~ (~~*ii*~~) Sequential assimilation could be employed to nudge the initial conditions for the 24-h runs with real-time-observed data. This could be achieved by modifying the binary AEM3D restart files using model specific -read/write statements in our Python workflow (e.g., from aem3d_restart_v3_type.f90), followed by smoothing (e.g., with a Kalman filter). 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.~~

The errors induced by hot-starting were shown to be negligible (Fig. 4, 6, 7 (a, b), Fig. A1). However, uncertainty from boundary conditions, especially the meteorological forcing, may introduce error. The 23 to 31 meteorological zones from the forecast wind field provides spatial variability required to simulate the mean surface circulation (Laval et al., 2003), water level (Trebitz, 2006), and thermocline motions (Valipour et al., 2015; Valipour et al., 2019). However, the 3-h time interval between GDPS forecast dataset updates is much less than the 10-min interval associated with meteorological data observed at lake-buoys (typically one to six) used to drive hindcasts (e.g., León et al. (2005)) and so the coarse temporal resolution in GDPS forecast may alias temporal events, such as wind gusts (Fig. C1). This is of particular concern in large shallow lakes, such as Lake Erie, where winds play the dominant role in driving hydrodynamics. The rapid response of the water level to windstorms (Hamblin, 1987) could result in the effects of aliasing and forecast error being passed to the water level, leading to the growth of RE against forecast time (Fig. 3).The meteorological forecast from the 15 km and 25 km GDPS models did not show discrepancies (Fig. C2-5) and the evaluation metrics indicate that forecast results were largely insensitive to the meteorological inputs in Lake Erie (Fig. 3, 7). However, the 15 km model better predicted the mesoscale upwelling event (Fig. 8, 9, D2). The 24-h air temperature and wind speed forecasts had ~ 1.5 °C and ~ 2 m s^{-1} RMSD, respectively. However, bias in the 240-h forecast increases with forecast time (Buehner et al., 2015). The 168-h forecast meteorological data overestimated wind speeds by up to 10 m s^{-1} (Fig. C4), and bias in the air temperature forecast (Fig. C5) may cause the consistent warm bias (up to 3°C) in forecast lake surface temperature (Fig. 8). These errors may be corrected through real time calibration using data assimilation (Baracchini et al., 2020a, b). The growing bias in air temperature, with forecast time, does not affect the lake surface temperature (Fig. 7), presumably owing to the buffer effect of surface mixing layer (Schertzer et al., 1987).

Neglecting the inflows and outflows in the predictive simulation could induce bias in the forecast. The overestimation of water level fluctuation range near Bar Point (Fig. 4f) may result from neglecting the large Detroit River inflow, which regulates the seiche magnitude. The inflows also adjust more rapidly to air temperatures compared to deep lake waters. Thus, the up to 2 °C cold bias in coastal regions of the western basin (Fig. 8 m-t, Fig. D2) could be induced by neglecting the heated flux from two major inflows (i.e., Detroit River and Maumee River) of Lake Erie.

In addition to inaccuracy in initial and boundary conditions, the discrepancies in simulating temperature profiles forecast may result from numerical diffusion arising due to the discrete nature of the vertical and horizontal grids. The simulated thermocline depth is overestimated (Fig. 9, 10), as occurred in applications of ELCOM with

both higher (Nakhaei et al., 2019) and lower resolution (Paturi et al., 2012). COASTLINES has the potential to predictively simulate mesoscale physical processes, such as Kelvin waves (Bouffard and Lemmin, 2013) and nearshore-offshore exchange (Valipour et al., 2019); however, model performance is poor in nearshore areas, where topographic features remain poorly resolved.

5 Conclusions

We developed an operational forecast system COASTLINES, using the Windows Task Scheduler, Python-based data scrapping/formatting, and MATLAB data processing scripts, to automate application of a black-box hydrodynamic driver (AEM3D) to Lake Erie as an operational forecast tool. The resulting real-time and predictive lake modelling system used meteorological forecasts to generate 240-h forecasts of the lake surface level and 3D temperature and current fields on a 500 m × 500 m (horizontal) × ~ 1 m (vertical) grid, compares model output with near real time observations and publishes the model output on a web-based platform.

The favorable agreement between forecast model results and observed physical variables (e.g., water level RE ~ 40 % and temperature t-RMSD and s-RMSD < 2 ℃) in Lake Erie demonstrates the ability of the forecast system to make predictions of hydrodynamic processes on time horizons up to 240-h that are as accurate as traditional hindcast simulations using directly observed meteorological forcing. This enables the near real-time updates to the web platform to be used as a communication tool that rapidly disseminate forecast results to managers and stakeholders. Examples include >24-h prediction of: (*i*) up- and down-welling events leading to fish kills; (*ii*) up-welling events transporting hypoxic water to a drinking water intake; and (*iii*) coastal flooding events from storm surges.

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. To facilitate further development of open-access predictive modelling systems, agencies are encouraged to share model validation observations, in real-time, through organizations such as GLEON (www.gleon.org) and GLOS (www.glos.us). This will enable extension of COASTLINES to include prediction of the physical-biogeochemical variables that drive sediment transport, hypoxia, and harmful algal blooms.

Code and data availability.

The observed data and meteorological forcing used in this study is openly accessible online as cited in the text. The COASTLINES model output is archived on the server and can be obtained by contacting the corresponding author. The Python and MATLAB scripts as well as the timeline set in Windows Task Scheduler are archived in the Scholars Portal Dataverse (https://doi.org/10.5683/SP2/VTN7WC, Lin, 2021). The AEM3D executable was used as a black box hydrodynamic transport code. The executable used in COASTLINES is available for a nominal license

fee from from Hydronumerics (https://www.hydronumerics.com.au/). The AEM3D source code was not modified in this application but is available with permission from Hydronumerics.

Author contributions.

The concept of the COASTLINES workflow was designed by LB, SL, SS, and RM, and SL carried them out. SL developed the model code and performed the simulations. All authors contributed to the validation of the model and interpretation of the results. SL wrote the manuscript with contributions from LB, SS, and RM.

Acknowledgements.

This project was funded by the Dean's Research Fund from the Faculty of Engineering and Applied Science at Queen's University. Computational support was provided by Alexander Rey and FEAS-ITS. LB thanks Damien Bouffard for discussions during visits to EAWAG, which inspired this research. James Homewood, from the Lower Thames Valley Conservation Authority (LTVCA) providing footages of the storm event on Nov. 15th, 2020.

Reference

Anderson, E. J., Fujisaki-Manome, A., Kessler, J., Lang, G. A., Chu, P. Y., Kelly, 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, 10.3390/jmse6040123, 2018.

Antenucci, J., and Imerito, A.: The CWR dynamic reservoir simulation model DYRESM, Science Manual, The University of Western Australia, Perth, Australia, 2000.

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.

Beletsky, D., Hawley, N., Rao, Y. R., Vanderploeg, H. A., Beletsky, R., Schwab, D. J., and Ruberg, S. A.: Summer thermal structure and anticyclonic circulation of Lake Erie., Geophys. Res. Lett., 39, 10.1029/2012GL051002, 2012.

Bocaniov, S. A., and Scavia, D.: Temporal and spatial dynamics of large lake hypoxia: Integrating statistical and three-dimensional dynamic models to enhance lake management criteria, Water Resour. Res., 52, 4247-4263, 10.1002/2015WR018170, 2016.

Bocaniov, S. A., Lamb, K. G., Liu, W., Rao, Y. R., and Smith, R. E.: High sensitivity of lake hypoxia to air temperatures, winds, and nutrient loading: Insights from a 3-D lake model, Water Resources Research, 56, e2019WR027040, 2020.

Boegman, L., Loewen, M. R., Hamblin, P., and Culver, D.: Application of a two-dimensional hydrodynamic reservoir model to Lake Erie, Canadian Journal of Fisheries and Aquatic Sciences, 58, 858-869, 2001.

Boegman, L., Loewen, M., Hamblin, P. F., and A., C. D.: Vertical mixing and weak stratification over zebra mussel colonies in western Lake Erie, Limnol. Oceanogr., 53, 1093-1110, 10.4319/lo.2008.53.3.1093, 2008.

Booij, N., Ris, R. C., and Holthuijsen, L. H.: A third-generation wave model for coastal regions 1. Model description and validation J. Geophys. Res. Oceans, 104, 7649–7666, 10.1029/98JC02622, 1999.

Bouffard, D., Boegman, L., and Rao, Y. R.: Poincaré wave–induced mixing in a large lake, Limnology and oceanography, 57, 1201-1216, 2012.

Bouffard, D., Ackerman, J. D., and Boegman, L.: Factors affecting the development and dynamics of hypoxia in a large shallow stratified lake: hourly to seasonal patterns, Water Resources Research, 49, 2380-2394, 2013.

Bouffard, D., and Lemmin, U.: Kelvin waves in Lake Geneva, J. Great Lakes Res., 39, 637-645, 10.1016/j.jglr.2013.09.005, 2013.

Bouffard, D., Boegman, L., Ackerman, J. D., Valipour, R., and Rao, Y. R.: Near-inertial wave driven dissolved oxygen transfer through the thermocline of a large lake, J. Great Lakes Res., 40, 300-307, 10.1016/j.jglr.2014.03.014, 2014.

Brookes, J. D., and Carey, C. C.: Resilence to blooms, Science, 334, 46-47, 10.1126/science.1207349, 2011.

Buehner, M., McTaggart-Cowan, R., Beaulne, A., Charette, C., Garand, L., Heilliette, S., Lapalme, E., Laroche, S., Macpherson, S. R., Morneau, J., and Zadra, A.: Implementation of deterministic weather forecasting systems based on ensemble–variational data assimilation at Environment Canada. Part I: the global system, Mon. Wea. Rev., 143, 2532-2559, 10.1175/MWR-D-14-00354.1, 2015.

Caramatti, I., Peeters, F., Hamilton, D., and Hofmann, H.: Modelling inter-annual and spatial variability of ice cover in a temperate lake with complex morphology, Hydrol. Process., 34, 691-704, 10.1002/hyp.13618, 2019.

Casulli, V., and Cheng, R.: Semi-implicit finite difference methods for three-dimensional shallow water flow. , Int. J. Numer. Methods Fluids 15, 629-648, 10.1002/fld.1650150602 1992.

Chen, C., Beardsley, R. C., Cowles, G., Qi, J., Lai, Z., Gao, G., Stuebe, D., Xu, Q., Xue, P., Ge, J., Ji, R., Tian, R., Huang, H., Wu, L., and Lin, H.: An unstructured grid, finite-volume community ocean model FVCOM user manual, SMAST/UMASSD Tech. Rep. 11-1101, 373 pp., Dartmouth, Mass., 2012.

Chu, P. Y., Kelley, J. G. W., Mott, G. V., Zhang, A., and Lang, G. A.: Development, implementation, and skill assessment of the NOAA/NOS Great Lakes Operational Forecast System, Ocean Dyn. , 61, 1305-1316, 10.1007/s10236-011-0424-5, 2011.

Gaudard, A., Schwefel, R., Vinnå, L. R., Schmid, M., Wüest, A., and Bouffard, D.: Optimizing the parameterization of deep mixing and internal seiches in one-dimensional hydrodynamic models: a case study with Simstrat v1.3, Geosci. Model Dev., 10, 3411-3423, 10.5194/gmd-10-3411-2017, 2017.

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.

Gronewold, A. D., and Rood, R. B.: Recent water level changes across Earth's largest lake system and implications for future variability, J. Great Lakes Res., 45, 1-3, 10.1016/j.jglr.2018.10.012, 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.
Hamblin, P. F.: Great Lakes storm surge of April 6, 1979, J. Great Lakes Res., 5, 312-315, 10.1016/S0380-1330(79)72157-5, 1979.
Hamblin, P. F.: Meteorological forecing and water level fluctuations on Lake Erie, J. Great Lakes Res., 13, 436-453, 10.1016/S0380-1330(87)71665-7, 1987.
Hecky, R. E., Smith, R. E. H., Barton, D. R., Guildford, S. J., Taylor, W. D., Charlton, M. N., and Howell, T.: The nearshore phosphorus shunt: a consequence of ecosystem engineering by dreissenids in the Laurentian Great Lakes, Can. J. Fish. Aquat. Sci., 61, 1285-1293, 10.1139/F04-065, 2004.
Higgins, S. N., Hecky, R. E., and Guildford, S. J.: Environmental controls of cladophora growth dynamics in eastern Lake Erie: Application of the Cladophora Growth Model (CGM), J. Great Lakes Res., 32, 629-644, 10.3394/0380-1330(2006)32[629:ECOCGD]2.0.CO;2, 2006.
Hipsey, M. R., Bruce, L. C., and Hamilton, D. P.: GLM - General Lake Model. Model overview and user information, Technical Manual, The University of Western Australia, Perth, Australia., 2014.
Hodges, B. R., Imberger, J., Saggio, A., and Winters, K. B.: Modeling basin-scale internal waves in a stratified lake, Limnology and Oceanography, 45, 1603-1620, 10.4319/lo.2000.45.7.1603, 2000.
Jabbari, A., Ackerman, J. D., Boegman, L., and Zhao, Y.: Episodic hypoxia in the western basin of Lake Erie, Limnology and Oceanography, 64, 2220-2236, 2019.
Jabbari, A., Ackerman, J. D., Boegman, L., and Zhao, Y.: Increases in Great Lake winds and extreme events facilitate interbasin coupling and reduce water quality in Lake Erie, Scientific Reports, 11, 5733, 10.1038/s41598-021-84961-9, 2021.
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.
Laval, B., Imberger, J., Hodges, B. R., and Stocker, R.: Modeling circulation in lakes: Spatial and temporal variations, Limnology and oceanography, 48, 983-994, 2003.
León, L. F., Imberger, J., Smith, R. E. H., Hecky, R. E., Lam, D. C. L., and Schertzer, W. M.: Modeling as a tool for nutrient management in Lake Erie: a hydrodynamics study, J. Great Lakes Res., 31, 309-318, 10.1016/S0380-1330(05)70323-3, 2005.
León, L. F., Smith, R. E. H., Hipsey, M. R., Bocaniov, S. A., Higgins, S. N., Hecky, R. E., Antenucci, J. P., Imberger, J. A., and Guildford, S. J.: Application of a 3D hydrodynamic-biological model for seasonal and spatial dynamics of water quality and phytoplankton in Lake Erie, J. Great Lakes Res., 37, 41-53, 10.1016/j.jglr.2010.12.007, 2011.
Leonard, B. P.: The ULTIMATE conservative difference scheme applied to unsteady one-dimensional advection., Comp. Methods Appl. Mech. Eng., 88, 17-74, 1991.
Lesser, G. R., Roelvink, J. V., Van Kester, J. A. T. M., and Stelling, G. S.: Development and validation of a three-dimensional morphological model. , Coastal Engineering, 51, 883-915, 10.1016/j.coastaleng.2004.07.014, 2004.
Liu, W., Bocaniov, S. A., Lamb, K. G., and Smith, R. E. H.: Three dimensional modeling of the effects of changes in meteorological forcing on the thermal structure of Lake Erie, J. Great Lakes Res., 40, 827–840, 10.1016/j.jglr.2014.08.002, 2014.
Loewen, M., Ackerman, J. D., and Hamblin, P. F.: Environmental implications of stratification and turbulent mixing in a shallow lake basin, Can. J. Fish. Aquat. Sci. , 64, 43-57, 10.1139/F06-165, 2007.
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.
Madani, M., Seth, R., León, L. F., Valipour, R., and McCrimmon, C.: Three dimensional modelling to assess contributions of major tributaries to fecal microbial pollution of lake St. Clair and Sandpoint Beach, J. Great Lakes Res., 46, 159-179, 10.1016/j.jglr.2019.12.005, 2020.
Meyers, T., and Dale, R.: Predicting daily insolation with hourly cloud height and coverage, Journal of Applied Meteorology and Climatology, 22, 537-545, 1983.
Michalak, A. a. M., Anderson, E. J., Beletsky, D., Boland, S., Bosch, N. S., Bridgeman, T. B., Chaffin, J. D., Cho, K., Confesor, R., Daloğlu, I., DePinto, J. V., Evans, M. A., Fahnenstiel, G. L., He, L., Ho, J. C., Jenkins, L., Johengen, T. H., Kuo, K. C., LaPorte, E., Liu, X., McWilliams, M. R., Moore, M. R., Posselt, D. J., Richards, R. P., Scavia, D., Steiner, A. L., Verhamme, E., Wright, D. M., and Zagorski, M. A.: Record-setting algal bloom in Lake

Erie caused by agricultural and meteorological trends consistent with expected future conditions, Proceedings of the National Academy of Sciences, 110, 6448-6452, 10.1073/pnas.1216006110, 2013.
MOE: Lake Erie fish kill incident on September 1, 2012. Summary Report.Ontario Ministry of the Environment Southwestern Region 14 ((Available from OntarioMinistry of Environment).), 2012.
Mortimer, C. H.: Fifty Years of Physical Investigations and Related Limnological Studies on Lake Erie, 1928–1977, Journal of Great Lakes Research, 13, 407-435, 10.1016/S0380-1330(87)71664-5, 1987.
Nakhaei, N., Boegman, L., Mehdizadeh, M., and Loewen, M.: Hydrodynamic modeling of Edmonton storm-water ponds, Environ. Fluid Mech., 19, 305-327, 10.1007/s10652-018-9625-5, 2019.
Nakhaei, N., Ackerman, J. D., Bouffard, D., Rao, Y. R., and Boegman, L.: Empirical modeling of hypolimnion and sediment oxygen demand in temperate Canadian lakes, Inland Waters, 1-17, 2021.
O'Reilly, C. M., Sharma, S., Gray, D. K., Hampton, S. E., Read, J. S., and Rowley, R. J., et al.: Rapid and highly variable warming of lake surface waters around the globe, Geophys. Res. Lett., 42, 10773-10781, 10.1002/2015GL066235., 2015.
O'connor, W. P., Schwab, D. J., and Lang, G. A.: Forecast verification for Eta Model winds using Lake Erie storm surge water levels, Weather and forecasting, 14, 119-133, 1999.
O'Neil, J. M., Davis, T. W., Burford, M. A., and Gobler, C. J.: The rise of harmful cyanobacteria blooms: The potential roles of eutrophication and climate change, Harmful Algae, 14, 313-334, 10.1016/j.scitotenv.2011.02.001, 2012.
Oveisy, A., Boegman, L., and Imberger, J.: Three-dimensional simulation of lake and ice dynamics during winter, limnol. Oceanogr., 57, 42-57, 10.4319/lo.2012.57.1.0043, 2012.
Paerl, H. W., and Paul, V. J.: Climate change: Links to global expansion of harmful cyanobacteria, Water Research, 46, 1349-1363, 10.1016/j.watres.2011.08.002, 2012.
Paturi, S., Boegman, L., and Rao, Y. R.: Hydrodynamics of eastern Lake Ontario and the upper St. Lawrence River, J. Great Lakes Res., 38, 194-204, 10.1016/j.jglr.2011.09.008, 2012.
Rao, Y. R., and Murthy, C. R.: Coastal boundary layer characteristics during summer stratification in Lake Ontario., J. Phys. Oceanogr., 31, 1088-1104, 10.1175/1520-0485(2001)031<1088:CBLCDS>2.0.CO;2, 2001.
Rao, Y. R., Hawley, N., Charlton, M. N., and Schertzer, W. M.: Physical processes and hypoxia in the central basin of Lake Erie, Limnol. Oceanogr., 53, 2007-2020, 10.4319/lo.2008.53.5.2007, 2008.
Rao, Y. R., Howell, T., Watson, S. B., and Abernethy, S.: On hypoxia and fish kills along the north shore of Lake Erie, J. Great Lakes Res., 40, 187-191, 10.1016/j.jglr.2013.11.007, 2014.
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.
Rowe, M. D., Anderson, E. J., Beletsky, D., Stow, C. A., Moegling, S. D., and Chaffin, J. D., et al: Coastal upwelling influences hypoxia spatial patterns and nearshore dynamics in Lake Erie., J. Geophys. Res. Oceans, 124, 10.1029/2019JC015192, 2019.
Ruberg, S. A., Guasp, E., Hawley, N., Muzzi, R. W., Brandt, S. B., and Vanderploeg, H. A., et al.: Societal benefits of the Real-time Coastal Observation Network (ReCON): Implications for municipal drinking water quality., Mar. Technol. Soc. J., 42, 103-109, 10.4031/002533208786842471, 2008.
Saber, A., James, D. E., and Hannoun, I. A.: Effects of lake water level fluctuation due to drought and extreme winter precipitation on mixing and water quality of an alpine lake, Case Study: Lake Arrowhead, California, Sci. Total Environ., 714, 10.1016/j.scitotenv.2020.136762, 2020.
Scavia, D., Allan, J. D., Arend, K. K., Bartell, S., Beletsky, D., Bosch, N. S., Brandt, S. B., Briland, R. D., Daloglu, I., DePinto, J. V., Dolan, D. M., and Evans, M. A. e. a.: Assessing and addressing the re-eutrophication of Lake Erie: Centralbasin hypoxia, J. Great Lakes Res., 40, 226-246, 10.1016/j.jglr.2014.02.004, 2014.
Scavia, D., DePinto, J. V., and Bertani, I.: A multi-model approach to evaluating target phosphorus loads for Lake Erie, J. Great Lakes Res., 42, 1139-1150, 10.1016/j.jglr.2016.09.007, 2016.
Schertzer, W. M., Saylor, J. H., Boyce, F. M., Robertson, D. G., and Rosa, F.: Seasonal Thermal Cycle of Lake Erie, Journal of Great Lakes Research, 13, 468-486, 10.1016/S0380-1330(87)71667-0, 1987.
Schwab, D. J., and Beletsky, D.: Propagation of kelvin waves along irregular coastlines in finite-difference models., Adv. Water Resour., 22, 239-245, 10.1016/S0309-1708(98)00015-3, 1998.
Schwab, D. J., Leshkevich, G. A., and Muhr, G. C.: Automated Mapping of Surface Water Temperature in the Great Lakes, J. Great Lakes Res., 25, 468-481, 10.1016/S0380-1330(99)70755-0, 1999.
Trebitz, A. S.: Characterizing seiche and tide-driven daily water level fluctuations affecting coastal ecosystems of the Great Lakes, J. Great Lakes Res., 32, 102-116, 10.3394/0380-1330(2006)32[102:CSATDW]2.0.CO;2, 2006.
Valipour, R., Bouffard, D., Boegman, L., and Rao, Y. R.: Near-inertial waves in Lake Erie, Limnol. Oceanogr., 60, 1522–1535, 10.1021/es301422r, 2015.

694 Valipour, R., Rao, Y. R., León, L. F., and Depew, D.: Nearshore-offshore exchanges in multi-basin coastal waters:
695 Observations and three-dimensional modeling in Lake Erie, J. Great Lakes Res., 45, 50-60,
696 10.1016/j.jglr.2018.10.005, 2019.
697 Wang, Q., and Boegman, L.: Multi-Year Simulation of Western Lake Erie Hydrodynamics and Biogeochemistry to
698 Evaluate Nutrient Management Scenarios, Sustainability, 13, 7516, 2021.
699 Watson, S. B., Miller, C., Arhonditsis, G., Boyer, G. L., and al, e.: The re-eutrophication of Lake Erie: Harmful algal
700 blooms and hypoxia, Harmful Algae, 56, 44-66, 10.1016/j.hal.2016.04.010, 2016.
701 Woolway, R. I., and Merchant, C. J.: Worldwide alteration of lake mixing regimes in response to climate change.,
702 Nat. Geosci., 12, 271-276, 10.1038/s41561-019-0322-x, 2019.
703 Woolway, R. I., Kraemer, B. M., and Lenters, J. D., et al.: Global lake responses to climate change, Nat Rev Earth
704 Environ, 10.1038/s43017-020-0067-5, 2020.

705

Appendix A: Comparison of 24 h model run with re-start files and model run with continuous files.

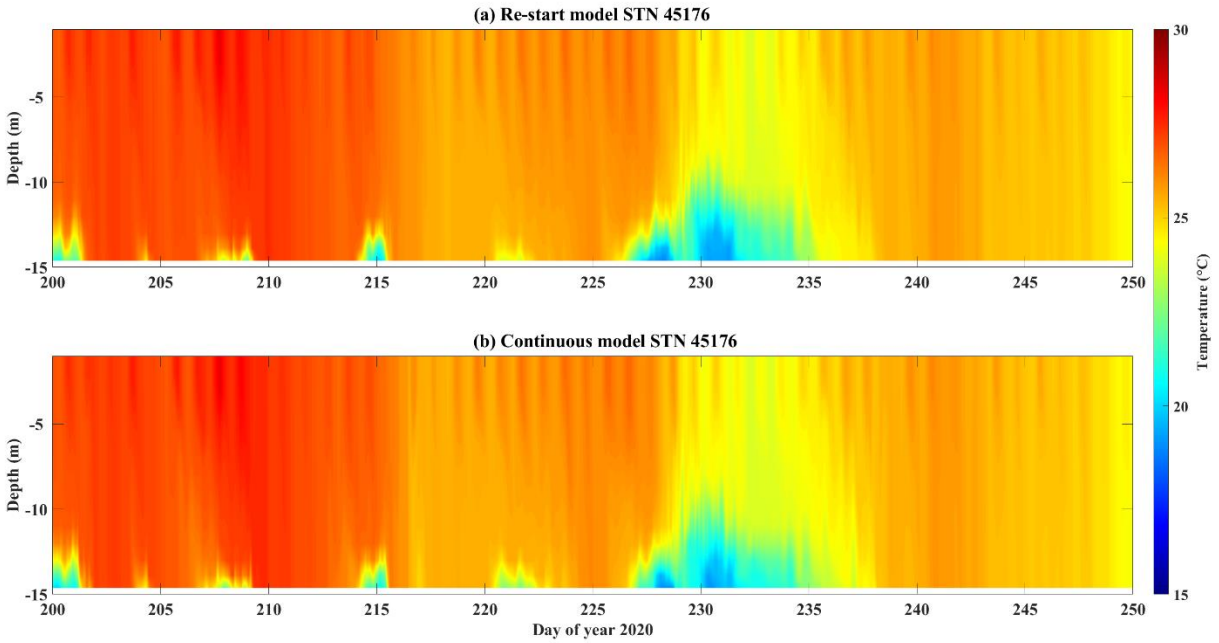


Fig. A1 Temperature profile comparison between (a) stitched 24 h model run with re-start files, and (b) model run with continuous input files.

Appendix B: COASTLINE website snapshot

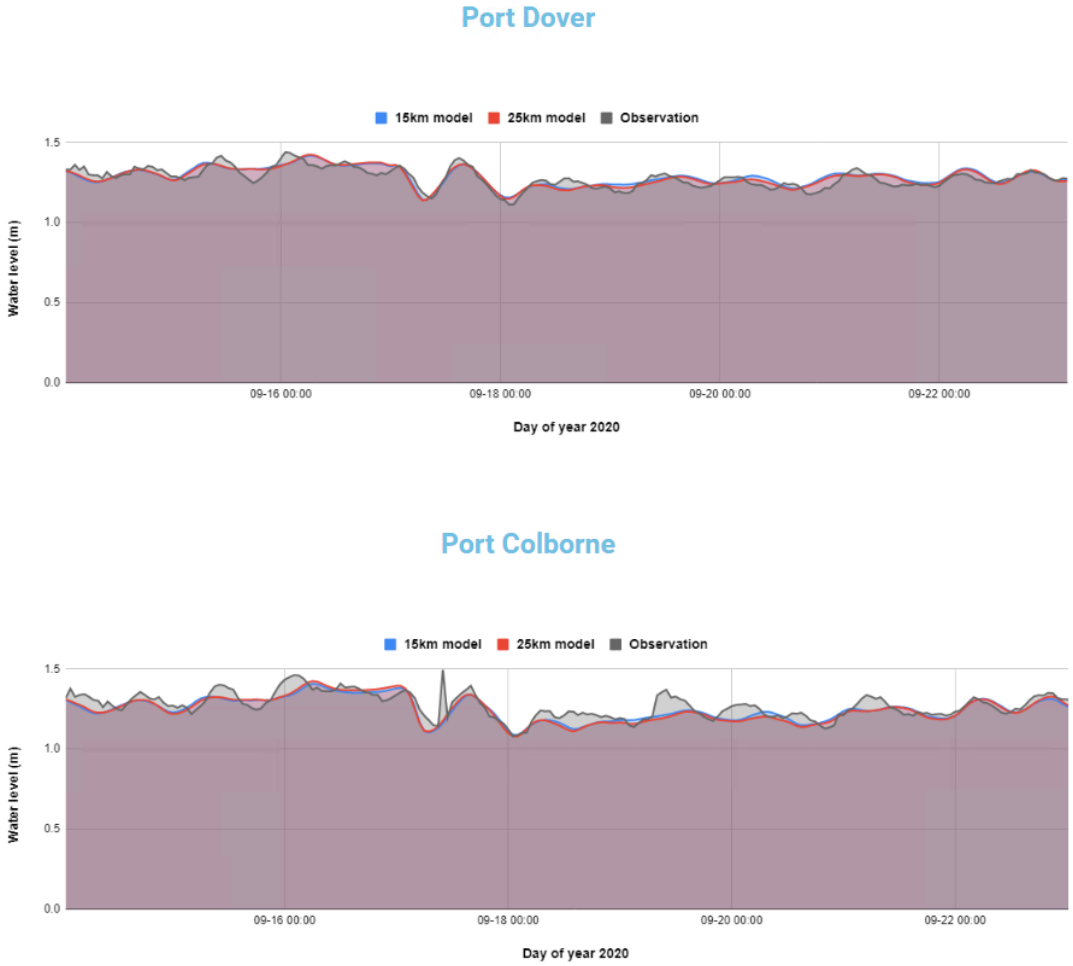


Fig. B1 Snapshot of water level forecast validation web page displayed on COASTLINES online platform: https://coastlines.engineering.queensu.ca/erie/water-level-forecast. Status on Sep 23rd, 2020.

Appendix C: Validation of meteorological input variables

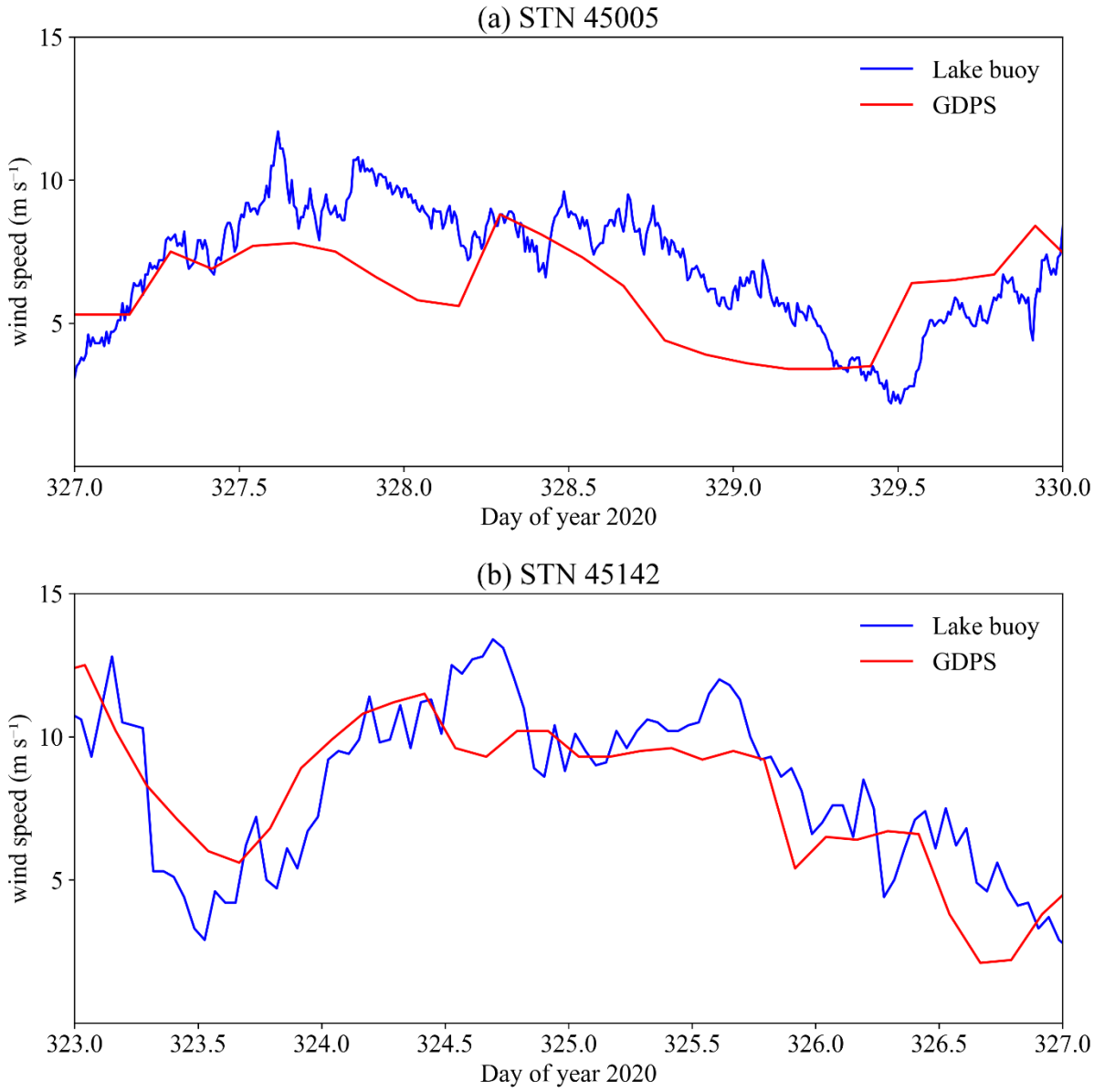
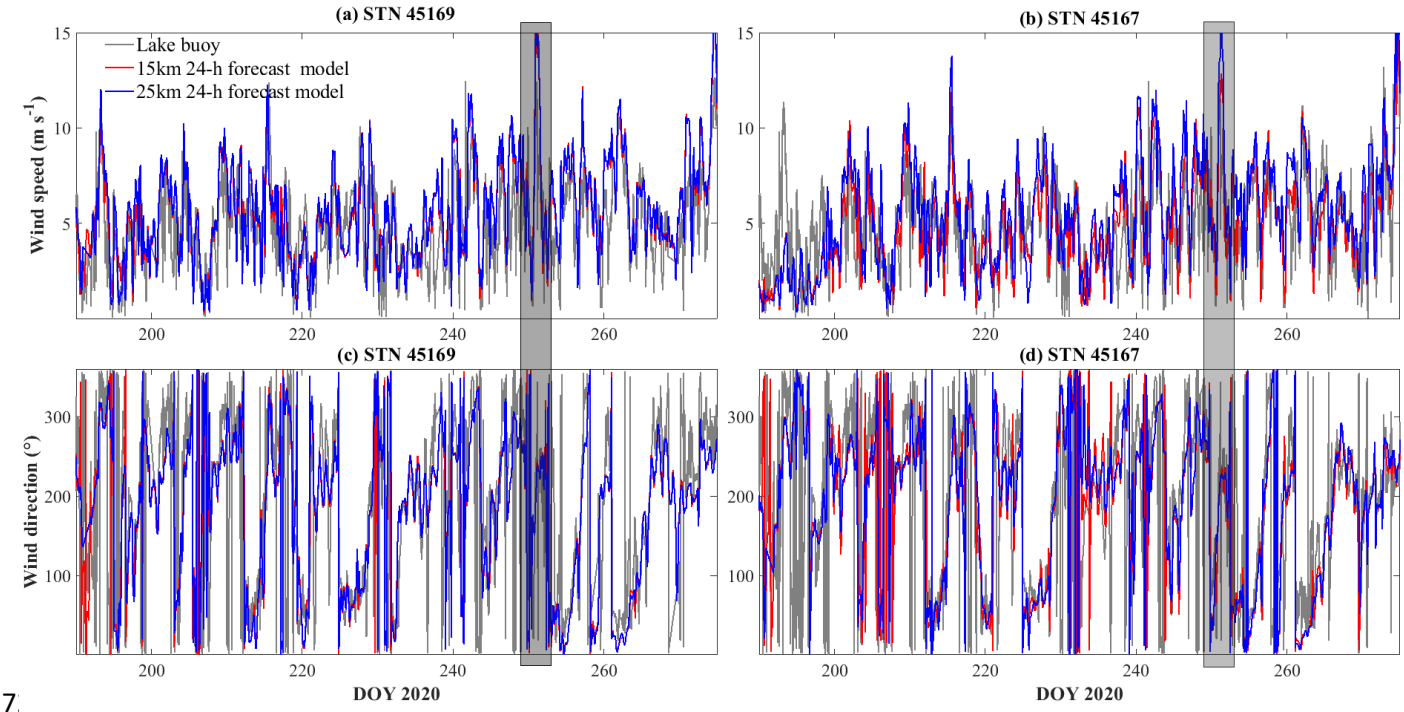
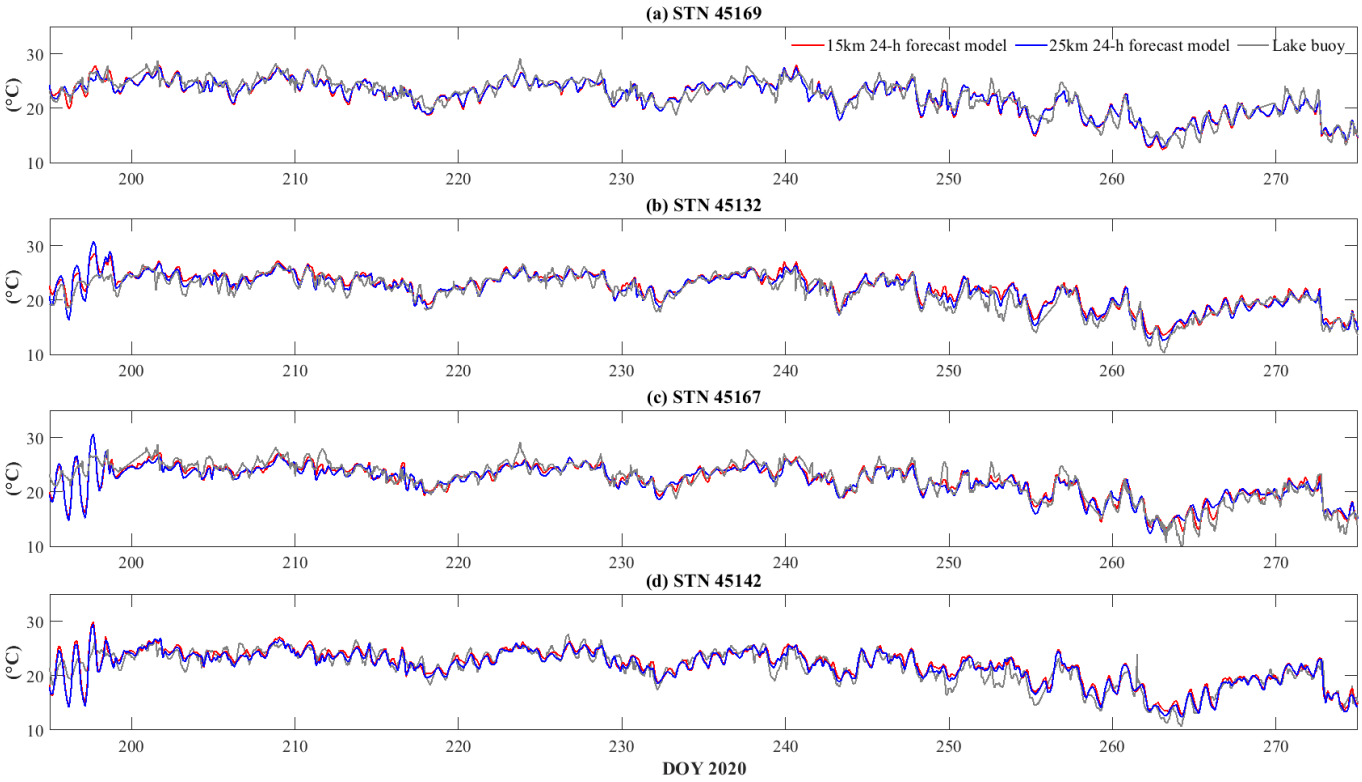


Fig. C1 Comparisons of stitched GDPS wind forecast with 3 h delivery interval and lake buoy measured wind speed at (a) station 45005 (10 min sampling interval), and (b) station 45142 (1 h sampling interval). The wind gusts on day 327 at station 45005 and day 324 at station 45142 were missed by wind forecast.

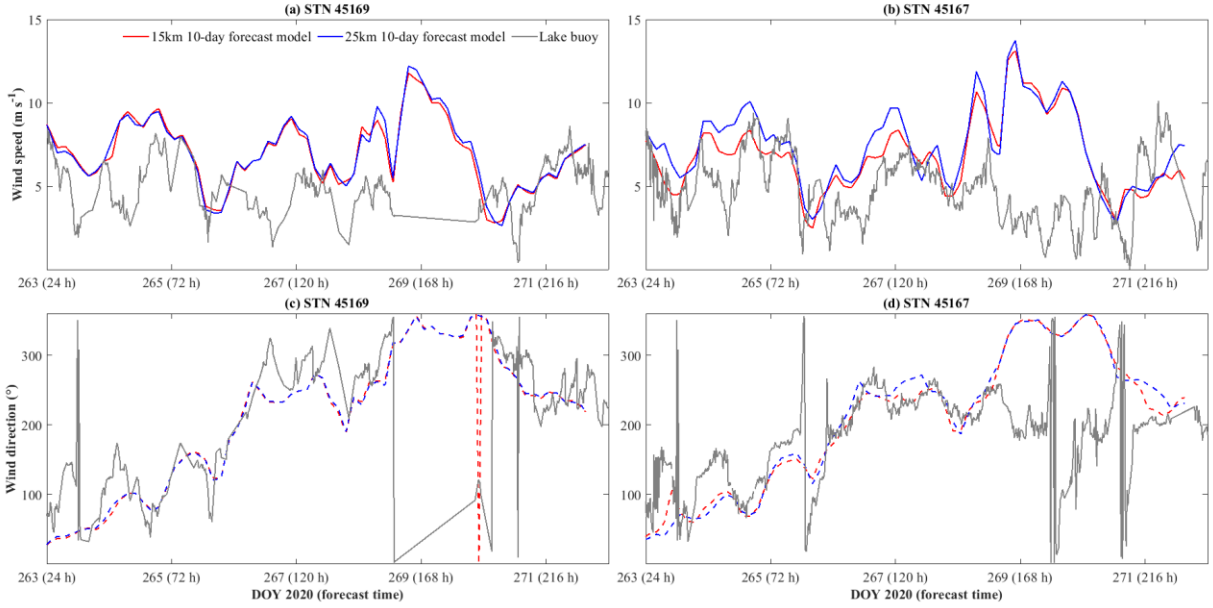


7

722 **Fig. C2 Comparisons of 24-h meteorological forecast and lake buoy observations of wind speed (a, b) and wind**
723 **direction (c, d). The gray rectangle indicates the storm that led to up-welling along northern shoreline on days**
724 **248-253.**



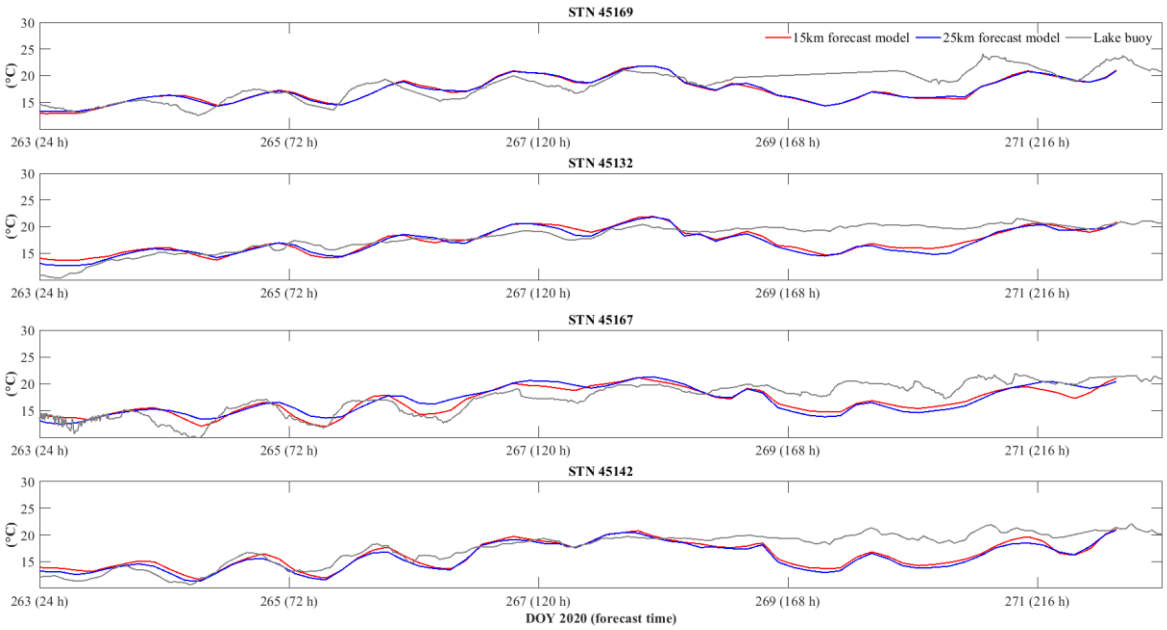
726 **Fig. C3 Comparisons of 24-h air temperature forecast and lake buoy observations of air temperature.**



727

728 **Fig. C4 Comparisons of 240-h meteorological forecast and lake buoy observations of wind speed (a, b) and wind**
729 **direction (c, d).**

730



731

732 **Fig. C5 Comparisons of 240-h air temperature forecast and lake buoy observations.**

Appendix D: Temperature validation against satellite observations

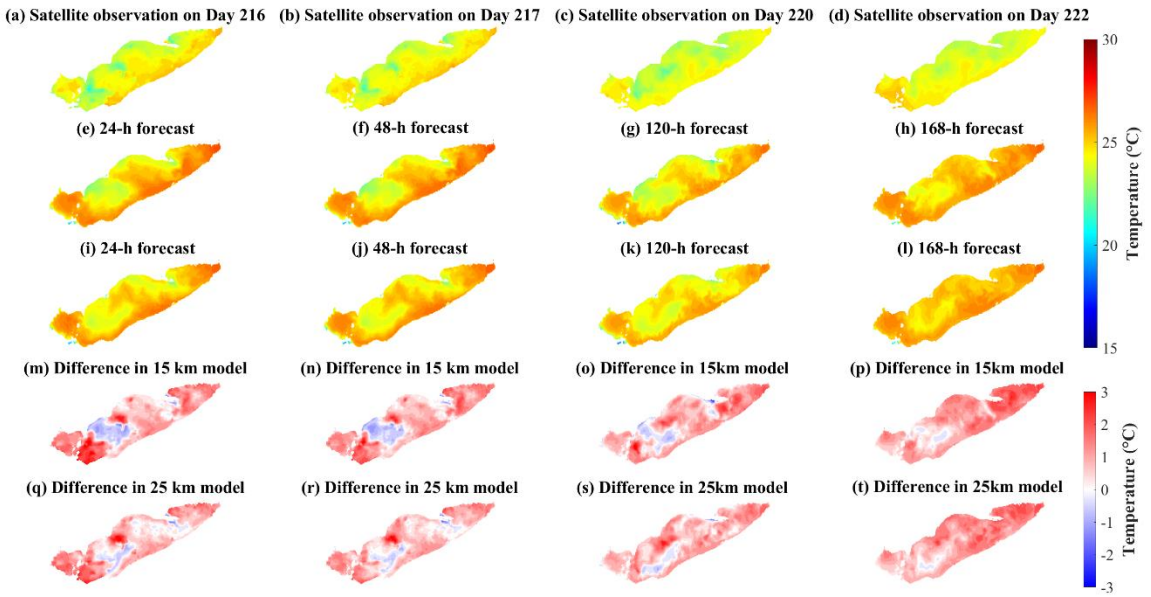


Fig. D1 comparisons of (a-d) satellite observations, (e-h) 15 km model 240-h forecast, and (i-l) 25 km model 240-h forecast during summer. The models were hot-started on Day 215. The difference between observations and models are shown in (m-t).

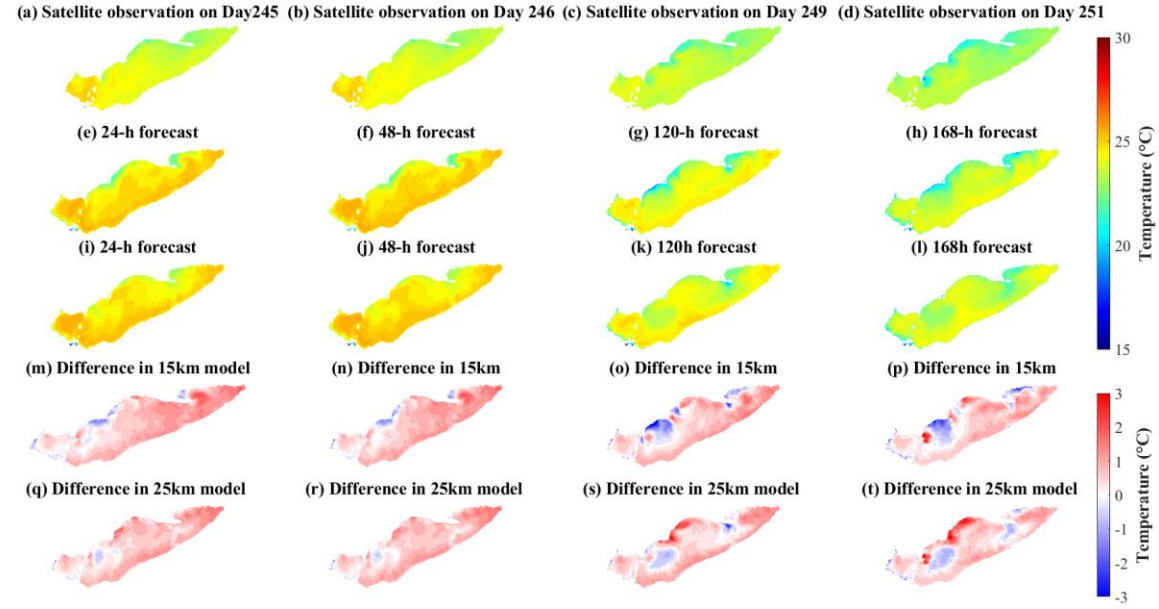


Fig. D2 comparisons of (a-d) satellite observations, (e-h) 15 km model 240-h forecast, and (i-l) 25 km model 240-h forecast during late summer. The models were hot-started on Day 244. The difference between observations and models are shown in (m-t).

Appendix E: Water level change during windstorm on Nov 15th, 2020

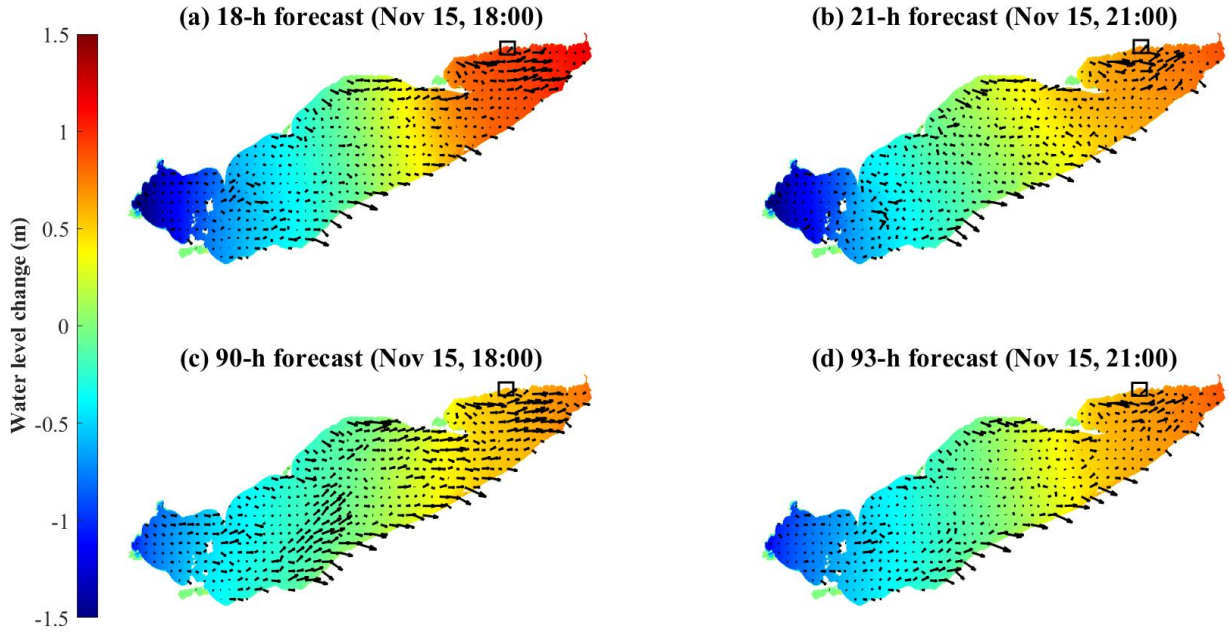


Fig. E1 Spatial distribution of water level change from forecasts hot-started on Nov 15th (a, b) and Nov 12th (c, d). The water level at Nov 15th 00:00 is the reference level. The black arrows are depth-averaged mean current fields. The black squares in the upper right corners of each map indicate the location of Port Dover (Fig. 12d).