

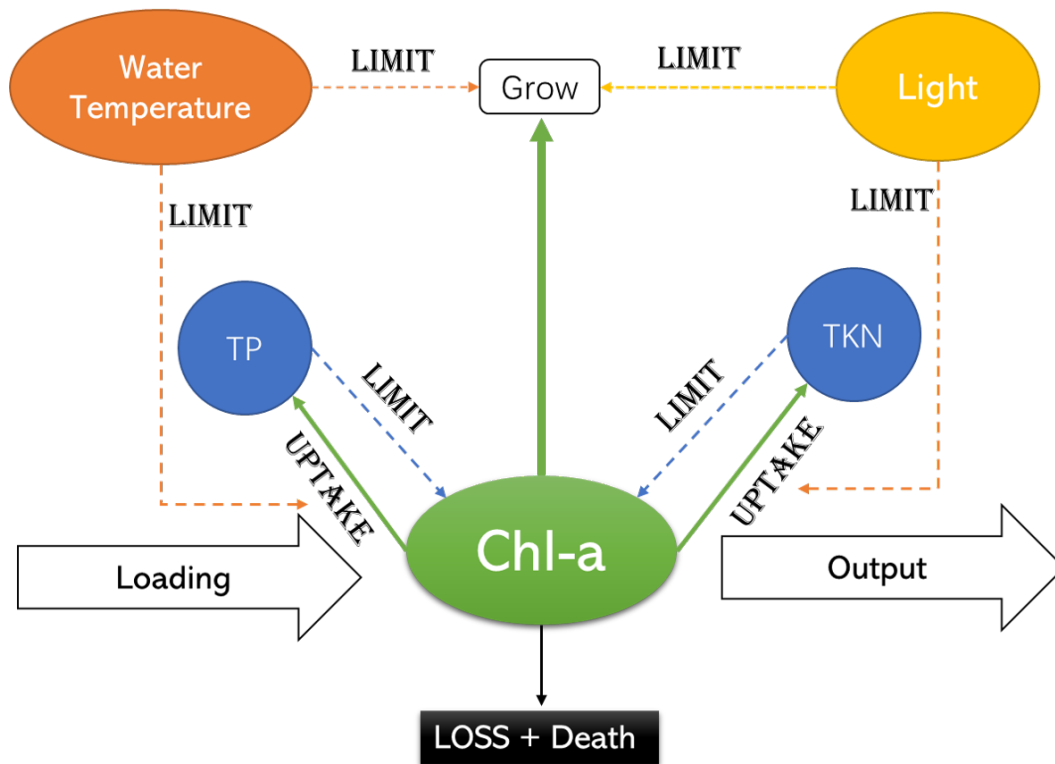
A Python based model to simulate the lake algal bloom, a case study in Lake Erie

Xinchen He¹

¹Department of Civil and Environmental Engineering, University of Massachusetts, Amherst, MA, USA

Key Points:

- Nutrients, temperature and light in this model were the main stressors on algal growth.
- Chlorophyll-a in the Lake Erie during algal bloom season in 2018 was simulated.
- The model shared a similar temporal variation trend with the observed values.



Corresponding author: Xinchen He, xinchenhe@umass.edu

Abstract

Lake algal bloom is one of the major environmental issues in the world. To help on algal bloom management, an numerical model based on Python3 was developed to simulate the mean chl-a variation in a lake ecosystem. The model took the tributary loading, nutrient limitation, temperature and light effects on algal growth, uptake and death into consideration and quantified them with time based ordinary system equation systems. To test the mode ability, I built and validated the model with the nutrients, chl-a, hydrodynamic and meteorological data from the Lake Erie to simulate the chl-a variation in the algal bloom season in 2018 (May to Oct). The bias of the model was significant (Nash–Sutcliffe Efficiency coefficient of the chl-a was -4.44)., but the model outputs showed a similar variation pattern with the trend of the observed results. To make the model better, I proposed several thoughts on finding reliable validation data source and calibration techniques at the last section. Generally speaking, the model still has a large space to improve and requires continuing study to make the model more 'real'.

Keywords: Lake algal bloom, numerical model, Python3, Lake Erie

1 Introduction

Algal bloom is one of the global environmental issues due to its negative effects on ecosystem and human society (Glibert et al., 2018). The excessive algal growth, especially the *Microcystis spp.* in the cyanobacterial community, can produce toxins which are detrimental for animals' health, including human. For instance, the cyanobacterial algal bloom in Lake Erie, America. In August 2014, the water supply to 600,000 people in Toledo was shut down for two days from Lake Erie (Steffen et al., 2017). The bloom events were frequently found on the shallow lakes in tropic and temperate regions where the flow disturbance is weak (Huang et al., 2016) and the light, temperature is optimal for algal growth (Gomes et al., 2015). It is believed that nutrients (nitrogen, phosphorus) from fertilizers and the wastewater are the main factors to the algal bloom (Paerl et al., 2016). Other than that, global warming was also proposed to be responsible for increasing algal bloom in global scale (Paul, 2008).

It is essential to manage and control the algal blooms in the lake. By using the data of the algal bloom stressors: nutrients, temperature, light, zooplankton (Chouvelon et al., 2019), sediment and other physiochemical parameters, many researchers build various numerical models to simulate the variation of the algal biomass (mostly indicated by chlorophyll-a). Many fantastic software could do that and showed good performances in the real environmental management studies, such as EFDC (Environmental Fluid Dynamic Code), Delft3D, DHI Mike (Maintained by Danish Hydraulic Institute). For example, (Kim et al., 2017) applied EFDC to build an algal bloom forecasting model at the lower Han River, Korea, which showed a good performance comparing with the monitoring data collected from 2013 to 2015. However, most of them are business software with expensive price. Moreover, source codes of the free algal bloom models were mostly written in FORTRAN or Microsoft C++, which are cryptic to read and configure.

In this context, we designed a lake algal bloom model based on Python3 (SciPy, NumPy, Pandas needed to be installed), in which nutrients limitation, temperature effects and light intensity to control the algal growth. More importantly, the model only requires discharge, Total Kjeldahl Nitrogen (TKN), Total Phosphorus (TP), temperature and light intensity data to support the model building. We apply the model on the Lake Erie to model the chl-a variation in 2018 algal bloom season (May-Oct). Because it is a prototype of a lake algal bloom model, there remains a large space to make the model more 'real'. The source code will be uploaded on my GitHub and GitLab repository, welcome to add your thoughts and contribute to the model.

2 Methods and Materials

2.1 Research Area

Lake Erie is the fourth-largest lake (by surface area) of the five Great Lakes in North America. The northern part belongs to Canada and the southern part locates in the United States. According to the description from the United States Environmental Protection Agency (<https://www.epa.gov/greatlakes/lake-erie>), the surface area of the lake is around $25,874\text{km}^2$, and the average volume is 480km^3 . The increasing nutrients loading from the fertilizer, wastewater and tributaries caused increasing algal bloom occurrence in the lake. The main flow direction is from west to east. The main inflow tributaries of the Lake Erie are the Detroit River, which connects the Lake Huron and the Lake Erie. The Maumee River, Sandusky River, Portage River and Huron River also played an important role in nutrients loading to the lake (Maccoux et al., 2016). The main out-flow tributary is the Niagara River. From Fig.1, It is clear to observe that the algal bloom often occurs in western part, whereas it does not happen frequently the central and the eastern lake. Which attributes to the water depth in western area is shallower than the other area.

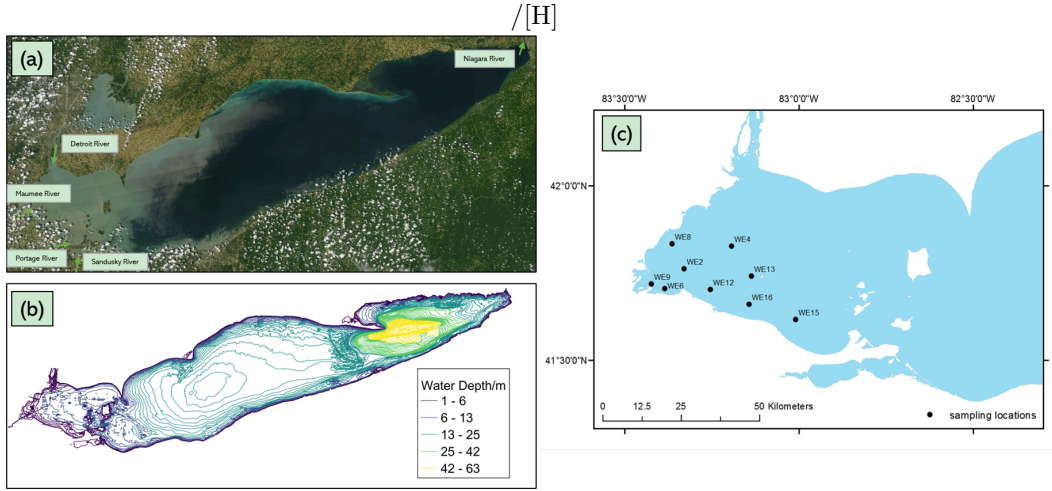


Figure 1. Lake Erie. (a) MODIS True color image of the Lake Erie image taken on 07/01/2018. (b) Water Depth contour map of the Lake Erie. (c) The location of the Lake Erie Monitoring stations, from (https://www.glerl.noaa.gov/res/HABs_and_Hypoxia/wle-weekly-current/)

2.2 Data Acquisition

The data we used in this paper contained two parts: Tributary discharge and water quality data, physical parameters from the stations in the lake (water quality, water temperature, light intensity, water levels). The discharge of the of the main tributaries we used were from National water Inventory System (NWIS), maintained by U.S. Geological Survey (USGS) (<https://waterdata.usgs.gov/nwis>). The water quality data of the tributaries were from the National Center for Water Quality Research, maintained by Heidelberg University (<https://ncwqr.org/monitoring/data/>). The monitoring data in the lake region were from the NOAA National Centers for Environmental Information, provided by Cooperative Institute for Great Lakes Research, University of Michigan and NOAA Great Lakes Environmental Research Laboratory. Because the record chl-a mon-

itoring data of the tributaries is unavailable, we used the lake monitoring data to represent the inflow or outflow chl-a concentration. Because our model time step is 1 day, we use second order polynomial interpolation to create daily series by using monthly or weekly sampling data.

2.3 Model Establishment

2.3.1 Model Setting

The time range of the model was from 05/01/2018 to 10/31/2018. The time step is 1 day. The initial volume was set as $480km^3$, initial TN, TP and chl-a concentration of lake Erie is 0.2 mg/L, 0.1 mg/L, 0.01 mg/L.

2.3.2 Volume Variation

$$V_{(t)} = \frac{V_{mean}}{H_{mean}} H_t + b = 2.7484 \times 10^9 \times H_t + b \quad (1)$$

$V_{(t)}$, m^3 , is the volume of the Lake Erie at t (time/date). H_t , *meter(IGLD85)*, denotes the water level observation value in Lake Erie. b , m^3 , is the correction item for approximating the volume, here we set it as 0 for simplicity. V_{mean} is the mean storage of the Lake Erie, H_{mean} is the mean water level in Lake Erie. Both mean values are from the EPA description page (<https://www.epa.gov/greatlakes/lake-erie>).

We assume that the surface area of the lake is constant and the water level variation has a robust positive relationship to the volume. According to EPA, the average of the volume is $480km^3$ and our water level average is 174.64 m. So the coefficient transformed from water level to the volume was set as $2.7484 \times 10^9 m^2$.

2.3.3 Nutrient Balance

$$S_{N(t)} = \frac{d_N}{d_t} = \frac{\sum_i^n I_i c_{Ni} - \sum_j^m Q_j c_{Nj}}{V_t} - d_N \cdot c_{N(t)} - Upt_N \cdot c_{N(t)} \quad (2)$$

$$S_{P(t)} = \frac{d_P}{d_t} = \frac{\sum_i^n I_i c_{Pi} - \sum_j^m Q_j c_{Pj}}{V_t} - d_P \cdot c_P - Upt_P \cdot c_P \quad (3)$$

The nitrogen/Phosphorus concentration in the Lake is influenced by tributary loading, natural losing(settle, other chemical reactions), uptake by phytoplankton. $S_{N(t)}$ is the change rate of Nitrogen at time t . I_i is the inflow discharge of tributary i , c_{Ni} is the N concentration of the inflow tributary i . Q_j is the outflow discharge of tributary j . c_N is the mean N concentration in the Lake at time t . d_N is the overall loss of Nitrogen which are not attribute to biological process. For example, settling + chemical reactions. Upt_N is the uptake rate by algae.

2.3.4 Chlorophyll-a Balance

$$S_{chl(t)} = \frac{d_{chl}}{d_t} = \frac{\sum_i^n I_i c_{chli} - \sum_j^m Q_j c_{chlj}}{V_t} - R \cdot c_{chl} + \mu \cdot c_{chl} - d_{chl} c_{chl} \quad (4)$$

Chlorophyll-a is the indicator of algal biomass. The balance of chl-a depends on tributary loading, death, growth and natural loss. R is death rate, biological related, unit is d^{-1} . μ is growth rate of chl-a, unit is d^{-1} . d_{chl} is the natural loss rate chl-a. c_{chl} is the mean concentration at the lake.

2.3.5 Additional Processes

Nutrient's Limitation

$$F(N) = \frac{N}{k_{sn} + N} \quad (5)$$

$$F(P) = \frac{P}{k_{sp} + P} \quad (6)$$

$$F(N, P) = \frac{2}{\frac{N}{k_{sn} + N} + \frac{P}{k_{sp} + P}} \quad (7)$$

$F(N)$ or $F(P)$ is the nitrogen/phosphorus limitation on various algal activity. $F(N, P)$ describes the combined nutrient limitation on algal growth. N and P are the concentration of the Total Kjeldahl Nitrogen and Total Phosphorus respectively. k_{sn} is the half saturation concentration of N to algae. k_{sp} is the half saturation concentration of P to algae.

Temperature and Light's Limitation

$$F(T) = \exp\left(-\frac{2.3}{15}T - T_{opt}\right) F(L) = \frac{L}{L + k_L} \quad (8)$$

The temperature equation is from (Wang et al., 2016). T and L are temperature ($^{\circ}\text{C}$) and light intensity ($\mu\text{Em}^{-2}\text{s}^{-1}$). T_{opt} is the optimal temperature for algal growth. k_L is the half saturation of light intensity affecting algae growth.

Growth

$$\mu = \mu_{max} \cdot F(L) \cdot F(N, P) \quad (9)$$

μ_{max} is the maximum growth rate of chl-a, unit is d^{-1}

Uptake

$$Upt_N = Upt_{Nmax} \cdot F(N) \cdot F(T) \quad (10)$$

$$Upt_P = Upt_{Pmax} \cdot F(P) \cdot F(T) \quad (11)$$

Upt_{Nmax} is the maximum nitrogen uptake rate of nitrogen, unit is d^{-1} . Upt_{Pmax} is the maximum Phosphorus uptake rate of nitrogen, unit is d^{-1}

Death

$$R = R_{max} \cdot F(T) \cdot F(N, P) \quad (12)$$

R_{max} is the maximum death rate of the chl-a, unit is d^{-1}

2.3.6 Solution Technique and Parameters Setting

The model has 13 interacted ordinary differential equations. For the ODE system, Scipy.integrate.solve_ivp is a great package to solving those equations. https://docs.scipy.org/doc/scipy/reference/generated/scipy.integrate.solve_ivp.html#scipy.integrate.solve_ivp. The solution technique we choose is the Fourth Order Rungu-Kuta method. Which

Table 1. Parameters setting for the algal bloom model in Lake Erie

Parameter	Description	Unit	Value
μ_{max}	Maximum growth rate of chl-a	d^{-1}	1.28
T_{opt}	Optimal Temperature for algal growth	$^{\circ}C$	30
k_L	Half Saturation Light Intensity of algal growth	$\mu Em^{-2}s^{-1}$	300
k_{sn}	Half Saturation nitrogen concentration of algal growth	mg/L	0.2
k_{sp}	Half Saturation phosphorus concentration of algal growth	mg/L	0.015
d_n	Loss rate of nitrogen	d^{-1}	0.01
d_p	Loss rate of phosphorus	d^{-1}	0.01
Upt_{Nmax}	Maximum Uptake rate of nitrogen	d^{-1}	0.05
Upt_{pmax}	Maximum Uptake rate of phosphorus	d^{-1}	0.05
R_{max}	Maximum Death rate of algae	d^{-1}	1
d_{chl}	Loss rate of chl-a	d^{-1}	0.5

is commonly used by environmental scientists to solve the algorithm made by a ODE system. The model are controlled by 11 parameter listed in Table.1. The parameters values were mainly from my previous work with DHI Mike (He et al., 2020) and change the maximum growth rate, half saturation concentration for nitrogen and phosphorus, maximum death rate of algae, maximum uptake rate of nitrogen and phosphorus to limit the model output into the range of the observed values.

2.4 Performance Assessment

To evaluate the model performance, Nash Sutcliffe model Efficiency coefficient (NSE) was applied, which is a normalized statistic that determines the relative magnitude of the residual variance compared to the measured data variance. $0 < NSE < 1$, the closer to 1, the better the model can simulate the algal bloom process. $inf < NSE < 0$, means the model is unreliable. The mean value of the monitoring data from all the stations were used for validate the model output.

$$Nash = 1 - \frac{\sum_i^n (OBS_i - SIM_i)^2}{\sum_i^n (OBS_i - \overline{OBS})^2} \quad (13)$$

3 Results and Discussion

3.1 Monitoring Results

The daily monitoring data of water temperature, light intensity and chl-a from tributaries are unavailable, so we use the weekly sampling data to assimilate the daily data by interpolation and extrapolation. So the curve from the Figure 2 at the left is smoother than the right two figures.

Figure 2 shows the loading's from the tributaries. The daily loading of chl-a from the tributaries to the Lake Erie was 22.28 t/d averagely, ranges from 6.39 t/d, 42.61 t/d. From the Figure 2, the peak of the chl-a loading occurred at July 1st and September 1st. The TKN and TP daily loading of the tributaries to the Lake Erie were 1066 t/d and 174.91 t/d averagely. The TKN and TP fluctuated randomly and there is no significant peak periods detected.

As Figure 2 shows, both the water temperature and light intensity arrived at peak value at the end of July, the middle of the August and the beginning of September. Combine

the peak time with that of the chl-a loading, we could infer that during the starting days of September, the temperature and light intensity were high. It is possibly due to the high temperature and light,

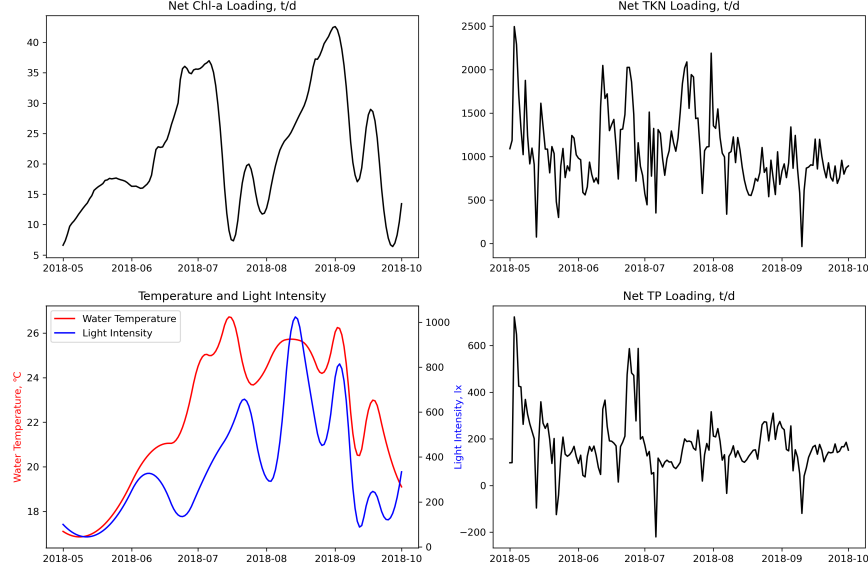


Figure 2. The nutrients, chl-a loading from tributaries to Lake Erie; The water temperature and light intensity observed in the Lake Erie

3.2 Modelling Results and Error Analysis

The Model results were presented in Figure 3. The NSE of chl-a, TKN and TP model outputs were -4.44, -11.20 and -11.56 respectively. From the NSE perspective, the model performances were not good. However, when looking through the model output and observation results, we found the model can capture the peak events and showed a similar trend with the observations temporal variation. Take the chl-a result as an example, both the model and the observations increases from May 21st to June 12nd and successfully capture the peak value at June 15th. Between June 12nd and Sep 4th, both of them shared similar trend and the peak value of the model output (0.024 mg/L) was close the the observed peak(0.029 mg/L).

Through checking the comparison plot, I found the chl-a modelling results were generally lower than the observed values, particularly during between June 12nd and Aug 6th. From the data credibility perspective, the validation data cannot represent the mean chl-a of the whole lake, because the sampling stations were located at western lake where algal biomass were significantly more than the other region as the MODIS image on Figure 2.1 shows. To take this factor into consideration, the model results may be close to the real mean chl-a concentration at Lake Erie.

But, from the model perspective. The deviation between the simulated and observed values of TKN and TP between June 12nd and Aug 6th was surprisingly opposite with that of chl-a. For instance, the modelling results of TKN were under 0.005mg/L but the observed results ranged from 0.10mg/L to 0.021mg/L. The situation was opposite when

it comes to TKN and TP, the model outputs were significantly larger than the observed results. I assume that is because, the uptake process of TKN and TP during this period in the model was weakened. Thus, if we need to calibrate the model and make it closer to the reality, we can start with tuning the parameters related to the uptake processes, such as Upt_{Nmax} or Upt_{Pmax} .

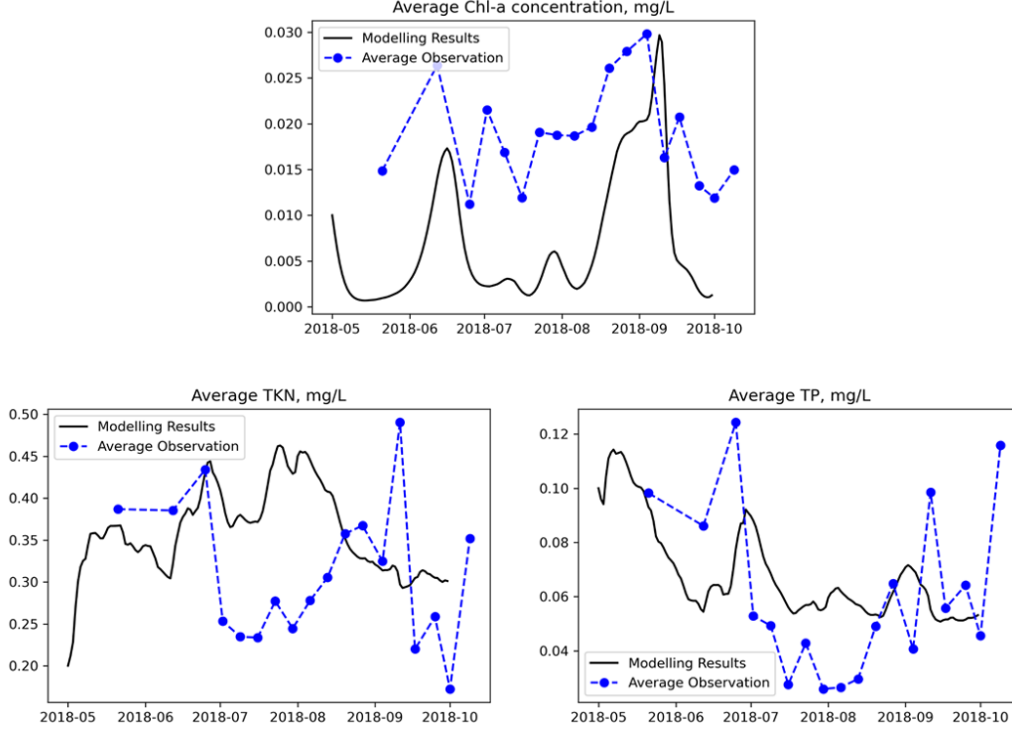


Figure 3. Comparison between observed results and Model outputs.

3.3 How to make the model better

Apparently, the bias of the model was significant but the model still showed the similar trend with the observed values and successfully captured the peak algal bloom in September. That means the model has potential to be a great tool to simulate algal bloom in lake ecosystem. To make it better, some thoughts were proposed below and welcome people who are interested in this model to continue.

To deal with the incredibility of the observed data, we can use the remote sensing technique to estimate the average chl-a concentration over the whole lake. Sensors like Sentinel-3 OLCI (Pinardi et al., 2015), MODIS Aqua (Shi et al., 2015), Landsat-8 (Page et al., 2018), etc. showed great performance detecting wide range of algal blooms. Applying the satellite derived data on both input (tributary loading) and output calibration and validation can significant enhance the model accuracy.

Calibration is an important procedure for modelling. It is an optimization problem to get minimum Nash–Sutcliffe Efficiency coefficient by changing the combinations of parameters. The model is like a dominoes, slightly modification of a parameter would bring changes of many processes. Directly doing the Monte Carlo simulation on the 11 parameters are computationally exhausting. The first step is trying to find which process caused the most errors, then test the sensitivity of the parameters controlling the process (Jiang

et al., 2018). Select 4 to 5 parameters then input them into a optimization function to get a best combination of parameters.

4 Conclusion

An algal bloom model based on Python3 was developed to simulate the mean chl-a variation in a lake ecosystem. The model took the tributary loading, nutrient limitation, temperature and light effects on algal growth, uptake and death into consideration and quantified them with time based ordinary system equation systems. To test the model ability, I built and validated the model with the nutrients, chl-a, hydrodynamic and meteorological data from the Lake Erie to simulate the chl-a variation in the algal bloom season in 2018 (May to Oct). The bias of the model was significant, but the model outputs showed a similar variation pattern with the trend of the observed results. To conclude, the model still has a large space to improve.

Acknowledgments

The data in Lake Erie we used contained two parts: Tributary discharge and water quality data, physical parameters from the stations in the lake (water quality, water temperature, light intensity, water levels). The discharge of the of the main tributaries we used were from National water Inventory System (NWIS), maintained by U.S. Geological Survey (USGS). The water quality data of the tributaries were from the National Center for Water Quality Research, maintained by Heidelberg University. The monitoring data in the lake region were from the NOAA National Centers for Environmental Information, provided by Cooperative Institute for Great Lakes Research, University of Michigan and NOAA Great Lakes Environmental Research Laboratory. I sincerely thank Dr. Guzman and all students in CEE-577 for creating such great environment to study water quality modelling.

References

- Chouvelon, T., Strady, E., Harmelin-Vivien, M., Radakovitch, O., Brach-Papa, C., Crochet, S., ... Chiffolleau, J.-F. (2019). Patterns of trace metal bioaccumulation and trophic transfer in a phytoplankton-zooplankton-small pelagic fish marine food web. *Marine Pollution Bulletin*, 146, 1013-1030. doi: <https://doi.org/10.1016/j.marpolbul.2019.07.047>
- Glibert, P. M., Berdalet, E., Burford, M. A., Pitcher, G. C., & Zhou, M. (2018). Introduction to the global ecology and oceanography of harmful algal blooms (geohab) synthesis. *Global Ecology and Oceanography of Harmful Algal Blooms*.
- Gomes, A., Azevedo, S., & Lurling, M. (2015). *Temperature effect on exploitation and interference competition among microcystis aeruginosa, planktothrix agardhii and, cyclotella meneghiniana* (Vol. 2015). doi: 10.1155/2015/834197
- He, X., Wang, H., Fan, L., Liang, D., Ao, Y., & Zhuang, W. (2020, 6). Quantifying physical transport and local proliferation of phytoplankton downstream of an eutrophicated lake. *Journal of Hydrology*, 585, 124796. doi: 10.1016/j.jhydrol.2020.124796
- Huang, J., Xi, B., Xu, Q., Wang, X., Li, W., He, L., & Liu, H. (2016). Experiment study of the effects of hydrodynamic disturbance on the interaction between the cyanobacterial growth and the nutrients. *Journal of Hydrodynamics*, 28, 411-422.
- Jiang, L., Li, Y., Zhao, X., Tillotson, M. R., Wang, W., Zhang, S., ... Pan, B. (2018, 5). Parameter uncertainty and sensitivity analysis of water quality model in lake taihu, china. *Ecological Modelling*, 375, 1-12. doi:

- 10.1016/j.ecolmodel.2018.02.014
- Kim, J., Lee, T., & Seo, D. (2017, 12). Algal bloom prediction of the lower han river, korea using the efdc hydrodynamic and water quality model. *Ecological Modelling*, 366. doi: 10.1016/j.ecolmodel.2017.10.015
- Maccoux, M. J., Dove, A., Backus, S. M., & Dolan, D. M. (2016, 12). Total and soluble reactive phosphorus loadings to lake erie. *Journal of Great Lakes Research*, 42. doi: 10.1016/j.jglr.2016.08.005
- Paerl, H. W., Scott, J. T., McCarthy, M. J., Newell, S. E., Gardner, W. S., Havens, K. E., ... Wurtsbaugh, W. A. (2016, 10). It takes two to tango: When and where dual nutrient (n amp; p) reductions are needed to protect lakes and downstream ecosystems. *Environmental Science Technology*, 50, 10805-10813. doi: 10.1021/acs.est.6b02575
- Page, B. P., Kumar, A., & Mishra, D. R. (2018). A novel cross-satellite based assessment of the spatio-temporal development of a cyanobacterial harmful algal bloom. *International Journal of Applied Earth Observation and Geoinformation*, 66, 69-81. doi: https://doi.org/10.1016/j.jag.2017.11.003
- Paul, V. J. (2008). Global warming and cyanobacterial harmful algal blooms. *Oxygen Transport to Tissue XXXIII*, 619(6), 239.
- Pinardi, M., Fenocchi, A., Giardino, C., Sibilla, S., Bartoli, M., & Bresciani, M. (2015, 4). Assessing potential algal blooms in a shallow fluvial lake by combining hydrodynamic modelling and remote-sensed images. *Water*, 7, 1921-1942. doi: 10.3390/w7051921
- Shi, K., Zhang, Y., Xu, H., Zhu, G., Qin, B., Huang, C., ... Lv, H. (2015, 6). Long-term satellite observations of microcystin concentrations in lake taihu during cyanobacterial bloom periods. *Environmental Science Technology*, 49, 6448-6456. (doi: 10.1021/es505901a) doi: 10.1021/es505901a
- Steffen, M. M., Davis, T. W., McKay, R. M., Bullerjahn, G. S., Krausfeldt, L. E., Stough, J. M. A., ... Johengen, T. H. (2017). Ecophysiological examination of the lake erie microcystis bloom in 2014: linkages between biology and the water supply shutdown of toledo, ohio. *Environmental Science Technology*, 51, acs.est.7b00856.
- Wang, H., Zhang, Z., Liang, D., Du, H., Pang, Y., Hu, K., & Wang, J. (2016, 7). Separation of wind's influence on harmful cyanobacterial blooms. *Water Research*, 98, 280-292. doi: 10.1016/j.watres.2016.04.037