

Large-scale flood models in watersheds with several lakes and reservoirs

Gaia Olcese¹, Christopher Sampson¹, Guéno   Chon  ², Pascale Biron², and Thomas Buffin-B  langer³

¹Fathom

²Concordia University

³Universit   du Qu  bec    Rimouski

September 14, 2020

Abstract

A series of recent flood events in Canada affecting areas around lakes and reservoirs have highlighted the need to explicitly represent such features in large scale flood models. Water level fluctuations in lakes are traditionally modelled using detailed hydrological models designed – as far as possible – to represent the actual physical processes that take place. This approach, while appropriate for local-scale studies in data-rich areas, is not applicable for large-scale flood modelling where data availability for model calibration and validation is often severely limited. This paper explores two methodologies, one statistical and one physically based, designed to approximately predict the increase in the water level of lakes in Quebec (Canada) using only limited morphological information about the lakes and the estimated discharge entering the water body during a flood event. Of the two methods, the statistical approach proved to be the most applicable to a large-scale modelling framework as it exhibited lower errors whilst being considerably easier to implement in a semi-automated modelling chain.

Introduction

Over recent years interest in large-scale flood modelling has grown due to the increase in computational capacity and availability of remotely-sensed terrain data sets (Alfieri et al., 2013; Dottori et al., 2016; Sampson et al. 2015; Wing et al. 2017; Winsemius et al., 2013). Historically, the vertical accuracy of large-scale terrain data sets has proven to be one of the most significant obstacles to obtaining accurate flood projections (Schumann 2014). Recent improvements to the wider accessibility of high-quality terrain data sets at large scales, such as the LiDAR-rich US National Elevation Dataset or the rapidly improving LiDAR coverage in Quebec with 1-m Digital Elevation Models (DEMs) freely available, have permitted the development of such models at national scales (Wing et al., 2017; Chon   et al., in review). When built with high quality input data, national scale flood models have been shown to demonstrate levels of skill approaching those of local scale models (Wing et al., 2017), and even where input data are less detailed they remain a useful starting point for the scoping of more detailed strategic and local-scale flood risk assessments. Due to the lack of accessible information on lakes and reservoirs and the complexity and heterogeneity of the physical processes involved, these models do not usually consider the effect of lakes during flood events and their skill in such areas remains poorly understood (Sampson et al., 2015).

With nearly 900,000 lakes covering more than 10 hectares, Canada accounts for 62% of the world’s lakes, a legacy of glaciers’ scouring action and their subsequent melting (Lo  c et al. 2016). Recent flood events, such as the spring floods of 2017 and 2019 caused not only rivers but also lakes to overflow in the province of

Quebec. In 2019, these inundations caused major flood stage to be recorded at 6 locations and middle flood stage at 12 locations, including the Lake of the Two Mountains (Lac des Deux Montagnes) and Lake Louise, damaging 2,341 homes and forcing around 1,200 residents to evacuate (Floodlist.com 2019). It is therefore unsurprising that the need for a more thorough understanding of lake water levels at large-scale has emerged in this context.

The literature currently provides various approaches to tackle the challenge of modelling water level stages in lakes. Previous studies focused on modelling the hydrological water balance of water basins including lakes (Setegn et al. 2008) or on identifying trends in the water level variability in a specific lake (Jöhnk et al. 2004). Other studies focused on the long-term prediction of changes in the water level using artificial intelligence methods (Altunkaynak 2006; Buyukyildiz et al. 2014; Khan & Coulibaly 2006; Piaseck et al. 2018) or on real time monitoring via satellite observations (Crétaux et al. 2011). Detailed hydrological models of lakes were developed in data-rich areas, considering riverine inflow, precipitation on the lake surface, evaporation and riverine outflow (Gibson et al. 2006). In other cases, spatially distributed hydrologic models were used for flood event simulation over basins with a complex system of reservoirs (Montaldo et al. 2004) and flood routing methods were applied to evaluate the effect of large artificial reservoirs (Gioia et al. 2016; Lee et al. 2001). However, no studies focused on analysing the impact of extreme flows on the increase of water level in both natural lakes and reservoirs and the consequential flood that could occur on the lakeshore.

This study sought to address this knowledge gap and derive a methodology that could approximately define the water level increase in lakes and reservoirs due to an extreme event with a specific probability of occurrence, and thus delineate the flood prone area in the surroundings. Ideally this method should be applicable to different types of water bodies, including natural lakes and artificial reservoirs. Since the final purpose of such a methodology is to be applied in the framework of large-scale flood simulations, the information required for each lake cannot be extensive and has to be easily available in a semi-automated way at national scales.

Methodology

Two approaches were considered to assess this problem: a physically based approach and a statistical approach.

Physical based model

The initial premise of the physically based approach was to develop a model that could simulate the fluctuations in the water level using an inflow peak hydrograph and remotely-sensed morphological characteristics of the lake as inputs. It is known from compensating reservoir modelling that lakes and reservoirs usually have an attenuation effect on the inflow hydrograph, causing a reduction of the peak discharge value and a release of the water volume during a longer period (United States National Resources Conservation Service, National Engineering Handbook. Section 630, Hydrology. Chapter 17, Flood Routing April 2014). The scale of the attenuation depends on the characteristics of the lake or reservoir, such as its storage capacity and the geometry of the spillway (Gioia 2016). To simulate this behaviour, it is possible to proceed with what is called flood routing, using as inputs the inflow discharge, the volume of water retained in the lake and a relationship between the variation of the water level and the outflow discharge.

A water balance equation is used to link the increase in the surface water elevation to the difference between the volume entering the lake and the volume leaving it at each time step.

$$\frac{S_2 - S_1}{t} = \frac{Q_{in1} + Q_{in2}}{2} - \frac{Q_{out1} + Q_{out2}}{2} \quad (1)$$

Where S stands for storage volume, t the timestep used for the calculations, Q_{in} is the inflow discharge, Q_{out} is the outflow discharge and the subscripts 1 and 2 represent different times (t). The water level at each

time step is calculated using the following equation.

$$\overline{H_2 = H_1 + \frac{(Q_{in1} - Q_{out1}) t}{A}} \quad (2)$$

Where H_1 represents the water level increase at t_1 , H_2 the water level at t_2 and A is the lake area. This equation assumes the lake area as constant: the lake is assumed to have a cylinder-type shape, in which the area is not a function of the water level.

There are various equations that can be used to compute the outflow discharge. In this study the rectangular weir equation was used. This choice was driven by the simplicity of the equation, which requires only the weir width to be derived from remotely sensed data.

$$\overline{Q_{out,i} = \mu L \sqrt{2gh_i} h_i} \quad (3)$$

μ is a shape coefficient, L represents the weir width (assumed as the width of the downstream river channel), g is the acceleration of gravity and h_i the height over the weir (equal to the water level increase). The subscript i refers to the time step: the outflow discharge is calculated at each time step as a function of the varying height over the weir, derived at the previous time step using equation (2). The shape coefficient was initially assumed as equal to 0.5; this value was identified as the value producing the smallest errors by some calibration tests. The parabolic weir equation was also tested: it proved to be less effective whilst also requiring more detailed information about the spillway.

The ideal test case to validate this model with observed data would be a lake with three gauges providing time series (instant values) of the water level and the upstream and downstream discharge. This way the inflow hydrograph of an event can be used as input for the model and the computed outputs can be compared with observed records. The lake should be small enough to be influenced by the inflow hydrograph in terms of fluctuations in the water level – a very large lake’s water elevation won’t vary during a short single event – and without a dam or any regulation device that could influence drastically the water level and the discharge downstream. Unfortunately, we were unable to identify such an ideal test case in either Canada or the United States, where most medium size lakes are dammed and/or do not have a gauging station upstream. Despite this, three test cases were used to compare the physically-based model output with water level observations: lake Maskinongé, lake Brulé in Quebec and Waterbury reservoir in Vermont. For lake Maskinongé and Waterbury reservoir three gauging stations were available, although both water bodies are dammed. There is no gauging station downstream Lake Brulé and the main inflow is influenced by a dam (barrage Ludger). Different peak hydrographs isolated from gauging stations upstream the lakes were used as inputs for the model, after being appropriately scaled to the watershed area at the lake’s outflow. This procedure allows to consider the inflow from ungauged tributaries.

$$\overline{Q_{in} = Q_{in,recorded} * F} \quad (4)$$

$$\overline{F = \frac{\text{Downstream watershed area}}{\text{Upstream watershed area}}} \quad (5)$$

After validation of the model framework on these case studies (see section 3.1.1), the next step was to test the model with synthetic hydrographs (necessary as the inflows to most lakes are ungauged) in order to produce water level frequency curves.

The model results from the synthetic hydrographs had to be validated against observed water level fluctuations in the lakes (section 3.1.3). Observed values were derived from a subset of water level measuring gauging stations with time records longer than 25 years in Quebec (33 stations) and the physically-based model was then tested on the gauges that also had synthetic discharge values available (31 stations). The

maximum annual fluctuations were initially derived as a difference between the recorded water levels and the mean at the corresponding station. However, because water level gauges are not available for most lakes, the final testing phase used water surface elevation derived from LiDAR as the baseline elevation to which water level increases were applied (on a subset of 23 stations at which all the necessary information was available). For those lakes with gauges, analysis shows that the average error between recorded mean water level and LiDAR was approximately 0.50 m and the median error was about 0.25 m (see Table 1 in supplementary information). A large portion of this error is driven by a small number of reservoirs that are likely to be affected by a strong seasonal regulation. Removing these stations from the analysis would significantly reduce the average difference between LiDAR and mean water level to a mean error of 0.25 m, but would also not be representative of an error affecting a non-negligible portion of lakes. Since the available LiDAR imagery is constantly increasing and will represent the main source to derive water level data at larger scale, the decision was to keep using the LiDAR elevation as a reference. The values for each lake were fitted with an appropriate distribution to extract values at different return periods (20, 100 and 350 years).

The inputs required to run the model for each lake are the inflow discharge, the lake area and the outflow channel width. The discharge was derived from the distributed hydrological model HYDROTEL (Fortin et al. 1995; 2001) for three return periods of interest while the channel widths were manually measured in QGIS for the different lakes in question. The time to concentration of the inflow hydrographs was set to a fixed value of 200 hours after performing a sensitivity analysis on the model.

Statistical model

In contrast to a physically based methodology, a statistical approach focuses on analysing the water level fluctuations at the available gauging stations across the region in order to identify the driving factors that determine the nature of water level increases. This is done by analysing the recorded time series with a probabilistic distribution and linking the results with observable characteristics of the lakes, in order to identify a statistical model that can be used at ungauged locations. The analysis focused on finding plausible linear regressions that could link the water level increases to different variables, such as lake area, upstream drainage area and peak discharge. To explore all the different possibilities the analysis was assessed in three steps: single variable regression analysis, multivariable regression analysis, and multivariable regression analysis with variable transformation. Several interaction terms were considered, in order to identify a statistically significant relationship.

Results

3.1 Physical based model

3.1.1 Validation of the model with recorded hydrographs

The results obtained by running the model with recorded hydrographs show that even though the water level variation doesn't always match the pattern of the recorded fluctuations, the peak water level can be reproduced with an accuracy varying between 0.05 and 0.25 m (Figure 1). These results were positive, especially considering that all the test cases were influenced to varying degrees by a dam downstream that imposed a regulating effect on the lake outflow. These preliminary results suggested that this approach, even if very simplistic, could produce sensible outputs for specific events and thus led the research toward testing it with synthetic hydrographs (for ungauged rivers) in order to produce water level frequency curves.

[Figure 1]

3.1.2 Definition of an appropriate time to concentration

During the initial model runs, times to concentration derived from Fathom's global flood model (Sampson et al. 2015) were used. The model uses the velocity method (United States National Resources Conservation

Service, National Engineering Handbook. Section 630, Hydrology. Chapter 15, Time of Concentration) to calculate the time to concentration as a sum of the travel time in shallow concentrated flow and the travel time in open channel flow. The travel time is derived using the longest flow path from the point of interest and an average velocity derived using Manning’s coefficient. This produced a substantial underestimation of the observed water levels. To understand the reason of this behaviour, several experiments were undertaken to help understand model parameter sensitivity. All the following tests were performed using the discharge estimated for a return period of one hundred years.

Initially, the model was run for each lake keeping the same weir value (best estimate from GIS and remotely sensed data) and varying the time to concentration from 1 hour up to 600 hours (24 days), in order to evaluate model sensitivity to this variable. In some cases, the time to concentration had a big influence on the modelled water level, while in others it seemed to be relatively insensitive. In all cases the time to concentration showed an asymptotic trend. The asymptotic behaviour indicates that it is essential not to underestimate time to concentration, while overestimation will be less harshly penalised in terms of model performance. This is intuitively correct as water levels in lakes are naturally self-regulating, with outflow increasing as lake level increases until an equilibrium level is reached. The weir equation represents this, with discharge being proportional to $h^{3/2}$, where h is water height above the weir crest.

The other variable shown to have a strong influence on model behaviour is the weir width. To evaluate model sensitivity to this variable, the time to concentration value was held constant while weir width was varied across a wide range of values. Again, some test cases proved to be very sensitive to this variable while others exhibited minimal sensitivity, with water level increases remaining almost constant regardless of weir width.

Following the univariate analysis of time to concentration and weir width, the next step was to try and delineate the behaviour of these lakes and reservoirs using a bivariate analysis. A range of different simulations were therefore run for each lake, varying both the weir width and the time to concentration. Figure 2 represents an example of the results obtained for the Lake Massawippi (Quebec) station, representing the absolute error between the peak water level increase produced by the model and the maximum recorded water level increase (difference between annual maximum and annual mean).

[Figure 2]

From these results, it is possible to identify some general patterns across a subset of 31 water level measuring gauging stations with time records longer than 25 years, known lake area and synthetic discharge in Quebec. Overestimation typically occurs when the weir is narrow or when the time to concentration drastically increases, whilst it appears more difficult to provoke underestimation from the model. In most cases it is also possible to note that time to concentration maintains its asymptotic trend: once the inflow hydrograph has a long enough duration, the water level fluctuation stabilises and grows very slowly. Bigger lakes generally appear to be more sensitive to the time to concentration, and less to the weir width, while for smaller lakes the best estimation of the water level seems to be very dependent on a good estimate of the weir width whilst still requiring a long enough hydrograph. Unfortunately, it doesn’t seem possible to generalise overall behaviour in water levels as even lakes that seem to be similar in size and with a comparable inflow show different values of recorded water lake fluctuation. Since these analyses highlighted how an overestimation of the inflow duration shouldn’t heavily penalize the model performance, a fixed value of 200 hours was chosen for the time to concentration to use hereafter.

3.1.3 Validation of the model with synthetic hydrographs

The sensitivity tests were performed on a dataset of 31 stations, using average observed water levels as reference. From now on, for the actual validation, the model was run on a reduced subset of 23 stations (the ones that also have available LiDAR data). Running the model using synthetic hydrographs produced three simulated water level increases for the three examined return periods (20, 100 and 350 years). A plot with the results can be found in Figure 2 of supplementary materials.

[Table 1]

[Table 2]

To evaluate the performance of the physical model and determine if it is worth implementing it in the workflow of flood modelling, the results were compared to using a median GEV distribution for all the 23 stations. Table 1 summarises the error that would derive from applying a median GEV distribution to all the 23 lakes in the subset, while Table 2 shows how the physical model would perform in terms of bias and RMSE (Root Mean Square Error). By comparing the RMSE values to the standard deviation associated with the GEV distribution, it is possible to deduce that, although the physical based approach produces smaller errors than just referring to the average values predicted by a statistical analysis across all the stations, the difference in precision of the two methodologies is not substantial. The error varies from 0.58 to 0.71 m when using a fixed GEV distribution and from 0.49 to 0.63 m when using the physical model. Figure 3 presents examples where the physical model greatly underestimates lake levels (Lake Simon, station 040408), one where it is close to the GEV (Lac Barrière, station 040407) and one that overestimates the lake level (lac du Poisson Blanc, station 040602), whereas graphical results for all the stations are presented in Figure 1 of supplementary material.

[Figure 3]

A study of the error associated with this type of model was performed to check the assumption of homoscedasticity and potentially identify any influence of some specific variables on the performance of the model. The model was analysed in relation to several variables: lake area, watershed area, peak discharge, peak discharge times lake area, degree of regulation and surface area variation index. The degree of regulation (DOR) is an index designed to quantify hydrological alterations induced by dams (Mailhot *et al.* 2018) and should show if dams have an identifiable effect on the water level fluctuations in lakes. The surface area variation index refers to the increase in the extension of the lake’s surface area with level increase and thus considers the impact of topography on the process. The tests revealed the errors to be homoscedastic, being homogeneously distributed when plotted against the possible predictors, and there was no clear relationship between the residuals of the model and any of the analysed variables (plot in Figure 2 of supplementary materials).

3.2 Statistical model

3.2.1 Single variable regression analysis

The relationships between the 350-year water level (estimated using a GEV distributions fitted to gauged level data at 23 stations and a set of individual predictor variables were initially tested using single variable regressions. Note that these 23 stations are a subset of the 31 stations with at least 25 years of water level data, as for this analysis we also needed LiDAR data as well as HYDROTEL discharge data to be available. The first step consisted of identifying significant correlations between water level and any single variable such as lake area, upstream drainage area, peak discharge or outflow channel width. The results showed no statistically significant relationship existed between water levels and any single variable; performing standard transformations to the variables, such as applying a logarithm or square root, did not yield any improvements.

3.2.2 Multivariable regression analysis

A multi-variable regression analysis was undertaken to identify any significant relationship between the same 350-year water level variation and a range of sets of variables. In order to find a plausible regression model, different predictors were taken into consideration, as well as their possible interaction terms. To conduct the analysis in a systemic way and explore all the possible combinations of the predictors, a maximum of ten input variables have been considered: lake area, watershed area, outflow channel width, peak discharge (return period of 350 years) and their six associated interaction terms (i.e. lake area multiplied by watershed area, lake area multiplied by outflow channel width, etc.). Most of the regression models showing p-values lower than the threshold value of 0.05 are associated with high RMSE or very low values of adjusted R squared. Moreover, none of the regression models proved to be robust: by simply removing a few stations

from the sample and re-running the fitting we can obtain models that use completely different variables as predictors.

3.2.3 Multivariable regression analysis applying variable transformations

Two transformations were then applied to the predictors (logarithmic and square root), to identify any linear correlation between the water level above the LiDAR level and the transformed variables. The predictors considered by this analysis were lake area, upstream drainage area, outflow channel width and peak discharge with return period of 350 years. In this context it was observed that squaring the discharge values made a relevant difference in the predictive performance of the model, while performing the same transformation on the area and watershed areas was not as significant. The weir value was not identified as a relevant variable. The simplest significant model used as predictors the lake area and the square root of the peak discharge, and shows a RMSE of 0.48 m with an adjusted R squared of 0.5495. The p-values for both the predictors are in the order of 10^{-5} , showing a statistically significant correlation (Table 3, Figure 4).

[Table 3]

[Figure 4]

The error characteristics of this approach show the assumption of homoscedasticity to be valid in this model. Neither the error nor the absolute value of the error are shown to be linked with an increase of any of the predictors or other lake characteristics (plot in Figure 3 of supplementary materials).

An identical multivariable regression procedure was applied using return periods of 100 and 20 years. The results show that an equivalent model, with slightly modified coefficients, provides a good performance for a 100 year flow (RMSE = 0.48 m, adjusted R squared of 0.4849). However, the adjusted R squared dropped to 0.3574 for the 20-year events, indicating that the model was not able to predict water levels associated with more frequent events. We hypothesise that this is because higher frequency events are more readily controlled by engineered features leading to highly unpredictable water level behaviour.

Conclusions

Even though the physically based approach shows some predictive skill in estimating lake water level fluctuations, the small difference in precision when compared to using an average distribution inevitably leads to the question of whether it is worth implementing it in a large-scale modelling framework. Including it in the automated process of flood simulation and deriving all the data needed as input (especially the outlet channel width, which needs to be measured manually) would require a considerable amount of effort. Moreover, the results suggest that the physical model is not suitable to simulate the complexity of the processes that take place during flood routing of a streamflow in lakes. Although it performs reasonably well when accurate streamflow data is provided, it is not reliable enough when run with synthetic hydrographs across all Quebec. It is likely that similar findings would have been obtained in other geographical contexts.

The statistical approach on the contrary provides a lower RMSE than the one obtained using the physical based model and eliminates the need for measuring the outflow channel width for every lake, thus simplifying the process. This procedure can be easily implemented in a more extensive large-scale modelling framework to provide first-order approximations of water levels associated with extreme floods. These levels could be used as boundary conditions for two-dimensional hydraulic simulations of river flow into the lake, a very common situation in Canada but also in many other regions affected by the Laurentide or Scandinavian ice sheets, as well as to define flood prone areas around lakes where detailed hydrological models are not available.

Acknowledgements

This work was funded by the Ministry of Environment of Quebec (Ministère de l'Environnement et de la Lutte contre les changements climatiques).

References

- Alfieri L., Salamon P., Bianchi A., Neal J., Bates P., Feyen L. 2013. Advances in pan-European flood hazard mapping. *Hydrological Processes*, 28(13), 4067–4077. doi:10.1002/hyp.9947
- Altunkaynak A. 2007. Forecasting surface water level fluctuations of lake van by artificial neural networks. *Water Resour. Manage.* 21:399-408. doi: 10.1007/s11269-006-9022-6.
- Buyukyildiz, M., Tezel, G. & Yilmaz, V. 2014. *Water Resources Management* 28:4747. doi: 10.1007/s11269-014-0773-1.
- Choné, G., Biron, P.M., Buffin-Bélanger, T., Mazgareanu, I., Neal, J.C., Sampson, C.C. In review. An assessment of large-scale non-calibrated flood modelling based on LiDAR data. *Hydrological Processes*.
- Crétaux, J.-F., Jelinski, W., Calmant, S., Kouraev, A., Vuglinski, V., Bergé-Nguyen, M., Gennero M.-C., Nino F., Abarca Del Rio R., Cazaneve A, Maisongrande, P. 2011. SOLS: A lake database to monitor in the Near Real Time water level and storage variations from remote sensing data. *Advances in Space Research* 47(9), 1497–1507. doi: 10.1016/j.asr.2011.01.004.
- Dottori, F., P. Salamon, A. Bianchi, L. Alfieri, F. A. Hirpa, and L. Feyen (2016), Development and evaluation of a framework for global flood hazard mapping, *Adv. Water Resour.*, 94, 87–102. doi: 10.1016/j.advwatres.2016.05.002.
- Floodlist.com. (2019). Canada – Floods Damage Over 2,000 Homes in Quebec. [online] Available at: <http://floodlist.com/america/canada-floods-quebec-april-2019> [Accessed 16 Dec. 2019].
- Fortin, J. P., R. Moussa, C. Bocquillon, J. P. Villeneuve. 1995. Hydrotel, a Distributed Hydrological Model Compatible with Remote Sensing and Geographical Information Systems. *Revue Des Sciences de l'Eau* 8 (1): 97–124.
- Fortin, J.P, R. Turcotte, S. Massicotte, R. Moussa, J. Fitzback, J.-P. Villeneuve. 2001. Distributed Watershed Model Compatible with Remote Sensing and GIS Data. I: Description of Model. *Journal of Hydrologic Engineering* 6 (2): 91–99.
- Gibson J. J., Prowse T. D., Peters D. L. 2006. Hydroclimatic controls on water balance and water level variability in Great Slave Lake. *Hydrological Processes*, 20(19), 4155–4172. doi: 10.1002/hyp.6424
- Gioia, A., 2016. Reservoir routing on double-peak design flood. *Water*. 8. 553. doi: 10.3390/w8120553.
- Khan M. S., Coulibaly P. 2006. Application of support vector machine in lake water level prediction. *Journal of Hydrologic Engineering* 11(3) 199-205. doi: 10.1061/(ASCE)1084-0699(2006)11:3(199).
- Lee K.T., Chang C.H., Yang M.S., Yu W.S. 2001. Reservoir attenuation of floods from ungauged basins. *Hydrological Sciences Journal* Vol. 46, No. 3, pp. 349-362.
- Mailhot A., Talbot G., Ricard S., Turcotte R., Guinard K. 2018. Assessing the potential impacts of dam operation on daily flow at ungauged river reaches. *Journal of Hydrology: Regional Studies*. doi: 10.1016/j.ejrh.2018.06.006.
- Messenger M., Lehner B., Grill G., Nedeva I., Schmitt O. 2016. Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications* 7, 13603 (2016) doi:10.1038/ncomms13603
- Montaldo N., Mancini M., Rosso R. (2004). Flood hydrograph attenuation induced by a reservoir system: analysis with a distributed rainfall-runoff model. *Hydrological Processes*, 18(3), 545–563. doi:10.1002/hyp.1337
- Piasecki A., Jurasz J., Adamowski, J.F. 2018. Forecasting surface water-level fluctuations of a small glacial lake in Poland using a wavelet-based artificial intelligence method. *Acta Geophysica* (2018) 66: 1093. doi: 10.1007/s11600-018-0183-5.

Sampson, C. C., A. M. Smith, P. D. Bates, J. C. Neal, L. Alfieri, and J. E. Freer (2015), A high-resolution global flood hazard model, *Water Resour. Res.*, 51, 7358–7381, doi:10.1002/ 2015WR016954.

Setegn S. G., Srinivasan R., Dargahi B. 2008. Hydrological Modelling in the Lake Tana Basin, Ethiopia Using SWAT Model. *The Open Hydrology Journal*. doi: 10.2174/1874378100802010049.

Schumann G. 2014. Fight floods on a global scale. *Nature* 507, 169 (2014) doi:10.1038/507169e

United States. National Resources Conservation Service. National Engineering Handbook. Section 630, Hydrology. Chapter 17, Flood Routing. Washington, D.C.: U.S. Dept. of Agriculture, Natural Resources Conservation Service, 2014.

United States. National Resources Conservation Service. National Engineering Handbook. Section 630, Hydrology. Chapter 15, Time of Concentration. Washington, D.C.: U.S. Dept. of Agriculture, Natural Resources Conservation Service, 2014. Wing, O. E. J., P. D. Bates, C. C. Sampson, A. M. Smith, K. A. Johnson, and T. A. Erickson (2017), Validation of a 30 m resolution flood hazard model of the conterminous United States, *Water Resour. Res.*, 53, 7968–7986, doi:10.1002/ 2017WR020917.

Winsemius, H. C., L. P. H. Van Beek, B. Jongman, P. J. Ward, and A. Bouwman (2013), A framework for global river flood risk assessments, *Hydrol. Earth Syst. Sci.*, 17, 1871–1892, doi:10.5194/hess-17-1871-2013.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

	GEV distribution		
	RP 20 years	RP 100 years	RP 350 years
Average (m)	0.91	1.11	1.27
Standard deviation (m)	0.58	0.65	0.71

	Physically-based model compared to observations		
	RP 20 years	RP 100 years	RP 350 years
Bias (m)	-0.02	-0.03	-0.04
RMSE [m]	0.49	0.56	0.63

Multivariable regression model			
RP 350 years			
Predictors used	p - values	RMSE	R squared adjusted
Lake area (km ²)	3.55e-05	0.4880	0.5495
Sqrt(Q ₃₅₀) (m ³ /s) ^{0.5}	6.32e-05		
Formula		<i>Water level (m)</i> = 0.716 − 0.0048 · A + 0.0593 * √Q ₃₅₀	
RP 100 years			
Predictors used	p - values	RMSE	R squared adjusted
Lake area (km ²)	1.41e-04	0.4771	0.4849
Sqrt(Q ₁₀₀) (m ³ /s) ^{0.5}	2.28e-04		
Formula		<i>Water level (m)</i> = 0.619 − 0.0042 · A + 0.0562 * √Q ₁₀₀	
RP 20 years			
Predictors used	p - values	RMSE	R squared adjusted
Lake area (km ²)	0.0014	0.4722	0.3574
Sqrt(Q ₂₀) (m ³ /s) ^{0.5}	0.0021		
Formula		<i>Water level (m)</i> = 0.515 − 0.0033 · A + 0.0357 * √Q ₂₀	



