InundatEd-v1.0: A ~~Large-scale Flood Risk Modeling System on~~HAND-based flood risk modeling system using a ~~Big-data –~~ Discrete Global Grid System ~~Framework~~

Chiranjib Chaudhuri[1], Annie Gray[1], and Colin Robertson[1]

[1]Wilfrid Laurier University, Department of Geography and Environmental Studies, Waterloo, Canada

Email: chiranjibchaudhuri@gmail.com

Keywords: Flood modeling system, Height Above Nearest Drainage, Discrete Global Grid System, IDEAS, Web-GIS, R/Shiny, Manning’s Equation, Regional Regression.

Abstract

Despite the high historical losses attributed to flood events, Canadian flood mitigation efforts have been hindered by a dearth of current, accessible flood extent/risk models and maps. Such resources often entail large datasets and high computational requirements. This study presents a novel, computationally efficient flood inundation modelling framework (“InundatEd”) using the height above nearest drainage-based solution for Manning’s equation, implemented in a big-data discrete global grid systems-based architecture with a web-GIS platform. Specifically, this study aimed to develop, present, and validate InundatEd through binary classification comparisons to recently observed flood events. The framework is divided into multiple swappable modules including: GIS pre-processing; regional regression; inundation model; and web-GIS visualization. Extent testing and processing speed results indicate the value of a DGGS-based architecture alongside a simple conceptual inundation model and a dynamic user interface.

Introduction:

~~Globally from 1994 to 2013 flood events accounted for 43% of recorded natural disasters (Centre for Research on the Epidemiology of Disasters, 2016). Flooding is responsible for one third of natural disaster costs in Europe (Albano, Sole, Adamowski, Perrone, & Inam, 2018), while in Canada mean annual losses of $1-2 billion (CAD) are attributed to flood disasters (Oubennaceur et al., 2019). A 2013 flood in southern Alberta, costing over 1.7 billion dollars (CAD) in insured property damages, is the most expensive natural disaster in Canadian history (Stevens & Hanschka, 2014). Rapid economic development and urbanization during the last few decades – particularly urban development in close proximity to Canadian waters following population expansions of the 1950s-1960s – have increased the amount of exposure and in turn the economic damages of flood events (Robert et al., 2003), making the availability of accurate, timely, and detailed flood information a critical information need (Pal, 2002).~~

~~Mitigating the considerable economic impact of flood events; the design of effective emergency response measures; the sustainable management of watersheds and water resources; and flood risk management, including the process of public flood risk education, have long been informed by the~~ The practice of flood modelling, which aims to understand, quantify, and represent the characteristics and impacts of flood events across a range of spatial and temporal scales, has long informed the sustainable management of watersheds and water resources including flood risk management (Handmer, 1980; Stevens & Hanschka, 2014; Teng et al., 2017, 2019; Towe et al., 2020). Flood modelling research has increased in response to such factors as predicted climate change impacts (Wilby & Keenan, 2012) and advancements in computer, GIS (Geographic Information Systems), and remote sensing technologies, among others (Kalyanapu, Shankar, Pardyjak, Judi, & Burian, 2011; Vojtek & Vojteková, 2016; Wang & Cheng, 2007).

Flood inundation modelling approaches can be broadly divided into three model classes: empirical~~;~~ (Schumann et al., 2009; Smith, 1997); hydrodynamic~~;~~ (Brunner, 2016, DHI, 2012); and simplified/conceptual~~.~~ (L'homme et al., 2008, Néelz & Pender, 2010). Empirical methods entail direct observation through methods such as remote sensing, measurements, and surveying, and have since evolved into statistical methods informed by fitting relationships to empirical data. Hydrodynamic models, incorporating three subclasses ~~(~~, viz; one-dimensional~~,~~ (Brunner, 2016; DHI, 2003), two-dimensional (DHI, 2012; Moulinec et. al., ~~,~~2011), and three-dimensional (Prakash et. al., 2014; Vacondio et. al., 2011), consider fluid motion in terms of physical laws to derive and solve equations. The third model class, simple conceptual, has become increasingly well-known in the contexts of large study areas, data scarcity, and/or stochastic modeling and encompasses the majority of recent developments in inundation modelling practices~~.~~ (Teng et. al. 2017). Relative to the typically complex hydrodynamic model class, simple conceptual models simplify the physical processes and are characterized by much shorter processing times (Teng et al., 2017, 2019). A class of model which uses the output of a more complex model as a means of calibrating a relatively simpler model is also gaining popularity (Oubennaceur et al., 2019). While each class has contributed substantially to the advancement of flood risk mapping and forecasting practices, a consistent barrier has been the trade-off between computer processing time and model complexity (Neal, Dunne, Sampson, Smith, & Bates, 2018), especially with respect to two-dimensional and three-dimensional hydrodynamic models, which entail specialized expertise to derive and apply physical and fluid motion laws, require adequate data to resolve equations, and the computational resources to process the equations. Neal et al. (2018) summarized the proposed solutions to such challenges as relating to 1) modifications to governing equations or 2) code parallelization, with the latter informing the method proposed in Oubennaceur et al. (2019). With respect to 2D/3D hydrodynamic model code parallelization, Vacondio et al. (2017) listed two approaches: classical (multi-treading or Open Multi-Processing and Message Passing Interface) and Graphics Processing Units (GPUs). The GPU-accelerated method has been shown to decrease execution times, while avoiding the use of supercomputers, for high-resolution, regional-scale flood simulations (e.g., Ferrari et al. (2020), Vacondio et al. (2017), Wang & Yang (2020), and Xing et al. (2019)). However, the GPU-accelerated method is still limited in terms of the hardware requirement (specialized graphics cards), the use of uniform and/or non-uniform grids (Vacondio et al. (2017)), and the need for specific, specialized modelling programs to handle the input data required to solve complex hydrodynamic equations. ~~The ongoing development~~

Several studies have introduced generic modelling frameworks that aim to provide robust flood risk estimates with relatively little configuration. Winsemius et al. (2013) for example developed GLOFRIS, a global-scale flood risk modelling framework comprised of global forcing data, a global hydrological model, a flood routing model, and an inundation downscaling model. While capable of providing flood risk at virtually any location on earth, the modelling framework is fixed to the existing datasets and models used, which have significant uncertainty at the scales considered. At a more local scale, Jamali et al. (2018) introduces a flexible flood inundation model that integrates a 1D hydraulic model with a simple GIS-based flood inundation approach. However, this loosely coupled approach still requires specification of a standalone hydraulic model for each location at which it is implemented. There has been a recent stream of research aiming to develop simple conceptual inundation models ~~offers~~that preserve both the generality of GLOFIS and the specificity of more local-scale models. Such simple conceptual inundation models offer another potential avenue to handle limitations such as computation requirements and data scarcity~~,~~ ~~allowing~~. In turn, areas and scales poorly served by standard hydrodynamic ~~modeling, to~~ modelling may be provided with up-to~~-~~ date flood extent maps~~ and provided with platforms with~~. Platforms through which the public can view and interact with the ~~simulated floods~~flood extent maps may also be developed (Tavares da Costa, 2019). One such simple conceptual inundation model is the flood model based on Height Above Nearest Drainage (HAND) (Liu et. al 2018). Zheng et al. (2018) estimated the River Channel Geometry and Rating Curve Estimation Using HAND which gained interest from the community, industry, and government agencies. Afshari et al. (~~2017~~2018) showed that, while HAND-based flood predictions can overestimate flood depth, this method provides fast and computationally light flood simulations suitable for large scales and hyper-resolutions. Although simple conceptual models using such methods as linear binary classification and Geomorphic Flood Index (Samela et al., 2017, 2018) have been, and continue to be, developed, the combination of simple conceptual flood methods with big-data approaches remains largely uninvestigated (Tavares da Costa, 2019).

Recent advances in big data architectures may hold potential to retain enough model complexity to be useful while providing computational speedups that support widespread and system agnostic model development and deployment. There is an increasing need for examination of the potential of decision-making through data-driven approaches in flood risk management and investigation of a suitable software architecture and associated cohort of methodologies (Towe et al., 2020).

Discrete global grid systems (DGGS) are emerging as a data model for a digital earth framework (Craglia et al. 2012; Craglia et al., 2008). One of the more promising aspects of DGGS data models to handle big spatial data is their ability to integrate heterogeneous spatial data into a common spatial fabric. This structure is suitable for rapid model developments where models can be split into unit processing regions. Furthermore, with the help of DGGS the model can be ported to a decentralized big-data processing system and many computations can be scaled for millions of unit regions. The Open Geospatial Consortium adopted a DGGS Abstract Specification in late 2017 and work is currently underway to develop standards for DGGS specification as a core geospatial data model (OGC, 2017). This is the first use of a DGGS for flood modelling we are aware of.

The Integrated Discrete Environmental Analytics System (IDEAS) is a recently developed DGGS-based data model and modelling environment which implements a multi-resolution hexagon tiling data structure within a hybrid relational database environment (Robertson, Chaudhuri, Hojati, & Roberts, 2020). Notably, and in contrast to previous systems, the only special installation entailed by the DGGS-based data spatial model is a relational database. As such, DGGS-based data model can be ported to any software-hardware architecture as long as it supports a relational database system. The system exploits the hardware capability of the database itself which can potentially incorporate the following: GPU(s), distributed storage, and a cloud database.

In this paper we employ the IDEAS framework for the efficient computation, simulation, analysis, and mapping of flood events for risk mitigation in a Canadian context. As such, the novelty of this study is twofold: 1) the contribution of the new DGGS-based big spatial data model to the field of flood modelling, and 2) the presentation of a web-interface which lets users compute the inundation on the fly based on input discharge for select Canadian regions where flood risk maps are either not publicly available or do not exist. Moreover, the properties and structure of the DGGS-based spatial data model address a number of challenges and limitations faced by previous flood modelling approaches in the literature. For instance, it is modular, making it easy to switch between Regional Flood Frequency Analysis (RFFA)-based, HAND-based, or alternative models without sacrificing the consistency of the framework. Likewise, the method by which Manning's n is calculated can be easily interchanged. Another novel aspect of this framework is the incorporation of Land Use Land Cover data in the estimation of the roughness coefficient Manning's n instead of a constant value or a channel-specific value of Manning's n as is typically used (Afshari et al., 2017; Zheng et al., 2018). In terms of the tradeoff between model complexity and computation power, the IDEAS framework uses an integer-based addressing system which makes it orders of magnitude more efficient than that of other, more traditional spatial data models (i.e, raster, vector) (Mahdavi-Amiri et. al. 2015; Li & Stefanakis, 2020; Robertson et al., 2020). This, in turn, benefits any and all spatial computations associated with flood modelling. Finally, whereas most major spatial computations entail specialized software/code, in the DGGS-based method the spatial relationship is embedded in the spatial-data model itself. Thus, the spatial relationships need not be considered beyond the use of certain rules of the spatial-data model. The overall efficiency and versatility provided by a DGGS framework can benefit the field of flood risk mapping, which uses the spatial distribution of simulated floods to identify vulnerable locations.

Access to flood risk maps can build the capacity of individuals to make informed and sustainable investment and residence decisions in an age of climate concern and environmental change (Albano et al., 2018). The current state of public knowledge of flooding risks is unsatisfactory, with an estimated 94% of 2300 Canadian respondents in highly flood-prone areas lacking awareness of the flood-related risks to themselves and their property, per a 2016 national survey (Calamai & Minano, 2017; Thistlethwaite, Henstra, Brown, & Scott, 2018; Thistlethwaite, Henstra, Peddle, & Scott, 2017). Calls for better transparency and access to reliable flood risk maps and data with which to improve public awareness and understanding of flood risks is in line with a contemporary trend toward more open and reproducible environmental models (Gebetsroither-Geringer, Stollnberger, & Peters-Anders, 2018). There is an opportunity to utilize big data architectures and recent developments in flood inundation modelling and risk assessment technologies to make flood risk information, based on best flood modelling practices, more accessible.

The aim of this paper is threefold: 1) propose a simple conceptual inundation model implemented in big-data architecture; 2) test the model and its results through comparison to known extents of previous flood events; and 3) present the resultant flood maps via an open source, interactive web application.

2. Methods

2.1 Overview

The modelling component of InundatEd incorporated four general stages: 1) GIS pre-processing; 2) flood frequency analysis and regional regression; 3) the application of the catchment integrated

Manning's Equation; 4) upscaling the model to a discrete global grid systems data model. Sections 2.2.1 to 2.2.4 describe stages 1-4 respectively.

The second component of InundatEd's development was the design of a Web-GIS interface, described in Section 2.3, which liaises with and between the big data architecture, the flood models' outputs as defined by user inputs, and FEMA's Hazus depth-damage functions (Nastev & Todorov, 2013) (Section S1). Section 2.4 subsequently links the Web-GIS interface conceptually to previous sections by providing a summary of InundatEd's system structure and its operation. Finally, simulated flood extents using InundatEd's methodology were compared to the extents of observed, historical flood extent polygons within the Grand River watershed and the Ottawa River watershed, provided respectively by the Grand River Conservation Authority and Environment Canada. The comparison and testing process is described in Section 2.5.

2.2. Modelling

2.2.1 – Stage 1: GIS Pre-processing

The following GIS input data were obtained from Natural Resources Canada for the Grand River and Ottawa River watersheds and cropped to their respective drainage areas of 6,800 square kilometres (Li et al., 2016) and 146,000 square kilometers (Nix, 1987): Digital Elevation Models (Canada Centre for Mapping and Earth Observation, 2015); river network vector shapefiles (Strategic Policy and Innovation Centre, 2019); and Land Use Land Cover (LULC) (Canada Centre for Remote Sensing, 2019). Figure 1 shows the input Digital Elevation Model with elevation values given in metres, and the dams and gauging stations used in this study. The resolution of the DEM and LULC data is 30m x 30m. The vertical accuracy of the DEM is 0.34 m ± 6.22 m, i.e., 10 m at the 90% confidence level~~.~~ (Beaulieu & Clavet, 2007). The vertical datum used is the Canadian Geodetic Vertical Datum of 2013 (CGVD2013). The stations used for station-level discharge comparison are labeled in Figure 1. The uncertainty in the vertical dimension affects the slopes of individual pixels, the upslope contributing area, and can potentially affect the quality of extracted hydrologic features (Lee et al., 1992, 1996; Liu, 1994; Ehlschlaeger and Shortridge, 1996). Hunter and Goodchild (1997), while investigating the effect of simulated changes in elevation at different levels of spatial autocorrelation on slope and aspect calculations, indicated the importance of a stochastic understanding of DEMs. The Monte Carlo method (Fisher 1991) could potentially shed some light on this kind of uncertainty. However, in our case it was beyond the focus of our study and we considered the vertical uncertainty small enough to not affect our large-scale flood modeling simulations. The remaining GIS input data is shown in Supplementary Figure S1. Very small networks, independent of the higher-order channels, were deleted from both regions. ArcGIS Desktop's Raster Calculator tool was used to burn the river network vector into the DEM to ensure the consistency of the river network between the dem delineated and observed. TauDEM (Terrain Analysis Using Digital Elevation Models) (Tarboton, 2005), an open-source tool for hydrological terrain analysis, was then used to determine drainage directions and drainage accumulation (Tarboton & Ames, 2004) within the watersheds of interest. Each watershed's drainage network was then established in TauDEM by defining a minimum threshold of two square kilometres on the contributory area of each pixel for the Grand River watershed and ten square kilometres for the Ottawa River watershed. Separately, a value of Manning's n was determined for each 30 x 30 metre pixel of the study areas based on land use/ land cover attributes (~~Comber & Wulder, 2019).~~Brunner, 2016). To this end, the input LULC classes (Canada Centre for Remote Sensing, 2019) within the study watersheds were mapped to the nearest class of the similar land cover classes documented in Chow (1959, Table 5-6) and Brunner (2016, Figure 3-19), from which the respective values of Manning's n were used. Table 1 provides the utilized input LULC classes, their respective description provided by NRCAN, and the employed n values. Height Above Nearest Drainage (HAND) (Rahmati, Kornejady, Samadi, Nobre, & Melesse, 2018; Garousi-Nejad, Tarboton, Aboutalebi, & Torres-Rua, 2019) was also calculated in TauDEM with reference to the DEM and derived drainage network. Figure 2a provides a visual overview of this stage of the modelling component.

2.2.2. Stage 2: Regional Regression and Flood Frequency Analysis

Perhaps one of the most popular methods of flood frequency analysis is the index flood approach - a regional regression model based on annual maximum discharge data (Dalrymple, 1960; Hailegeorgis & Alfredsen 2017). A variant of the index flood approach, which entails flood frequency analysis, has been employed to understand the characteristics of flood behavior at the global level (Smith et. al., 2014). At regional scale Burn 1997 has discussed the catchment procedure essential to undertake the flood frequency analysis. Faulkner et. al. (2016) devised the procedure to estimate the design flood levels using the available station data. Regional hydrological frequency analysis at ungauged sites is also studied by few researchers (Desai and Ouarda 2021).

The index flood approach was used to derive the discharges by return period at sub-catchment outlets. The model includes two sections: a) a relationship between index flood and contributory upstream area for each hydrometric station and each subcatchment outlet (regional regression); and b) a flood frequency analysis to estimate the quantile values of the departures, with a departure defined as discharge at given station divided by the index flood of that same station). The index flood approach entails the following assumptions: a) the flood quantiles at any hydrometric site can be segregated into two components – an index flood and regional growth curve (RGC); b) the index flood at a given location relates to the (sub)catchment characteristics via a power-scaling equation, either in a simpler case which considers only upstream contributory area or in a more complex case which incorporates land use/ land cover, soil, and climate information; and c) within a homogeneous region the departure/ratio between the index flood and discharge at hydrometric sites yields a single regional growth curve which can relate the discharge and return period (Hailegeorgis & Alfredsen, 2017).

Per assumption a) (the flood quantiles at any hydrometric site can be segregated into two components – an index flood and regional growth curve (RGC)), the index flood at each hydrometric station is required. To this end, annual maximum discharge values (m^3s^{-1}) were extracted within R (R Core Team, 2019) at hydrometric stations maintained by Environment Canada within the Grand River and Ottawa River watersheds (HYDAT) (Hutchinson, 2016). Only stations with a period of record >= 10 years of annual maximum discharge (England et al. (2018); Faulkner, Warren, & Burn (2016)) were maintained (n = 32 and n = 54 respectively for the Grand River watershed and the Ottawa River watershed). The minimum, median, and maximum periods of record for the Grand River watershed were 12, 50, and 86 years, respectively. Periods of record for the Ottawa River watershed ranged from a minimum of 10 years to a maximum of 58 years with a median of 36 years. A median annual maximum discharge value ($\tilde{Q}$) was then calculated for each hydrometric station. As discussed in Hailegeorgis & Alfredsen (2017), although the index flood is generally the sample mean of a set of annual maximum discharge values, index floods have also been evaluated based on the sample median (eg. Wilson et al., 2011) at the suggestion of Robson & Reed (1999). Finally, the index flood values (Q̃) were used to normalize the observed annual maximum discharge values (Q) at their respective station, resulting in a set of values designated as Qi, such that Qi = Q/ Q̃.

With respect to regional regression and assumption b) of the index flood method, a generalized linear model was applied to relate $\log_{10}$ transformed Q̃ values to $\log_{10}$ transformed upstream area values at each hydrometric station. The generalized linear model assumed an ordinary least squares error distribution. The results of the generalized linear model for each watershed allowed for the calculation of previously unknown Q̃ values for each subcatchment outlet. In a more complex model (Fouad et. al. 2016), other catchment characteristics such as land use/land cover, geology, etc. could be used. However, in the case of the proposed model the correlations between the calculated and observed index floods, on the sole basis of discharge records and a linear model relating upstream area, were high as discussed in the Results section. Thus, the simpler method was used to estimate index floods and to relate index flood to contributory area at hydrometric stations and subcatchment outlets. Thus, the regional regression model derived a relationship between index flood (Q̃) and upstream contributory area for each hydrometric station s or sub-catchment outlet. The relationship between index flood at station i or at a subcatchment outlet ($\widetilde{Q^s}$) (median of annual maximum discharge) and upstream contributory area (A_s) is given by:

$$\tilde{Q}^s = aA_s^c \quad (1)$$

where a is the index flood discharge response at a unit catchment outlet (or at a hydrometric station) and c is the scaling constant. We took the logarithm of Equation (1) on both sides - a procedure ~~used in~~ noted in Hailegeorgis & Alfredsen (2017) as used in Eaton, Church, & Ham (2002) - yielding a linear relationship which was solved using the Ordinary Least Squares approach (Haddad et al. (2011).

With respect to assumption c) of the index flood method, which assumes that a regional growth curve can be applied to a homogenous area as outlined above, we attempted to fit a distribution to the ratio of the annual maximum discharge values at each station to the corresponding index flood. Hailegeorgis and Alfredsen (2017) discussed a regionalization procedure which ensures the homogeneity of the station-level data over any region. However, due to the limited availability of the discharge data we avoided such sub-sampling and carried out the index flood method at the entire watershed scale (Faulkner, Warren, & Burn 2016). This, however, has impacted the upper quantiles of the flood estimation when comparing to the station level data

(Section 3.1). ~~The~~A fundamental step of the analysis process is the selection of a suitable probability distribution model~~—~~, a common tool in hydrologic modelling studies~~ (Langat et~~. The model should~~ al., 2019; Singh, 2015) for use in a watershed where the flow has been modified due to human impact – whether via development of built up areas, agriculture, road building, resource extraction activities such as forestry and mining, or flow abstraction in terms of dams and weirs is a fundamental step of the analysis process and must~~ account for ~~disturbance-related ~~changes to the flow's extreme value characteristics ~~of ~~in response to such factors as urbanization, agriculture, resource extraction, or the ~~flow~~operation of dams and weirs. Sometimes, natural hydrologic peaks, such as the spring freshet, are exacerbated by antecedent conditions such as large snowpacks and frozen soils, resulting in substantial flood events. While solutions to this problem have been proposed in the literature, artificial abstraction fundamentally changes the extreme value characteristics of the flow, thereby hindering the usability of most distributional forms (Kamal et. al. 2017).

Many researchers have tried to address this problem by putting explicit assumptions on types of non-stationarity affecting the river discharge and are able to devise a closed mathematical formulation which enables the parametric distributions to handle such non-stationarity. However, such methods typically entail knowledge of the specific design return periods of individual flood prevention structures (Salas & Obeysekera, 2014), many of which are absent in our case. To circumvent this problem, we used a non-parametric approach for the regional growth curve (RGC), which requires no fundamental sample characteristics. Thus, modified flood records and limited information notwithstanding, flood frequency estimation is possible using the index flood approach. Per assumption c) of the index flood method, a log-spline non-parametric approach was taken to model a RGC (Stone, Hansen, Kooperberg, & Truong, 1997) for each study watershed. Specifically, the index flood values ($\tilde{Q}$) were used to normalize the observed annual maximum discharge values (Q) at their respective station ($Q_i = Q/\tilde{Q}$). The Q_i values (n= 1487 and n = 1248 for the Ottawa River watershed and the Grand River watershed, respectively) were then fitted to a logspline distribution for their respective watershed. The discharge quantiles (Q_r) were extracted for the following return periods (T, years): 1.25, 1.5, 2.0, 2.33, 5, 10, 25, 50, 100, 200, and 500. The return periods were first converted to a cumulative distribution function:

Finally, flood quantile estimations were calculated for each return period as shown below:

$$Q_T^i = \widetilde{Q^i} q_T \quad (2)$$

such that T is a specified return period in years; Q_T^i is a quantile estimate of discharge for the specified return period T (years) at a specified station i (or a subcatchment outlet); $\widetilde{Q^i}$ is the "index flood" at the same station i (or at the same sub-catchment outlet); i = 1,2,…,N where N =32 for the Grand River watershed or N= 54 for the Ottawa River watershed; and q_T is the regional growth curve as described above. Figure 2b provides a visual accounting of the regional regression and flood frequency analysis methodology described in this section.

Some of the limitations of this framework include the long-term flow records and homogenous stations required for the creation of regional regression models. A dearth of long-term data affects flood magnitude computations specifically for the upper quantiles (5T rule, Section 3.1).

2.2.3 Stage 3: Catchment Integrated Manning's Equation

Manning's formula (Song et. al., 2017) is widely used to calculate the velocity and subsequently the discharge of any cross-section of an open channel. The Manning's equation is given in SI units by:

$$Q = \frac{1}{n} R_h^{\frac{2}{3}} A\, S^{\frac{1}{2}} \quad (3)$$

such that Q is discharge in cubic metres per second, A represents the cross-sectional area, n is a roughness coefficient, R_h is the hydraulic radius, and S represents slope (fall over run) along the flow path. Despite its widespread use, robustness, and relative ease of use, Manning's Equation has an inherent problem which comes from the uncertain orientation of cross-sections. To mitigate this problem, we integrated Manning's Equation along the drainage lines within the catchment, accounting for the slope of each grid cell to yield bed area and derived the stage-discharge relationship. This strategy uses hydrological terrain analysis, discussed previously in Section 2.2.1, to determine the Height Above Nearest Drainage (HAND) of each pixel (Rodda, 2005; Rennó et al., 2008). The HAND method determines the height of every grid cell to the closest stream cell it drains to. In other words, each grid cell's HAND estimation is the water height at which that cell is immersed. The inundation extent of a given water level can be controlled by choosing all the cells with a HAND less than or equal to the given level. The water depth at every cell can then be calculated as the water level minus the HAND value of the corresponding cell. The relevance of HAND to the field of flood modelling has been demonstrated in the literature (Rodda, 2005, Nobre et al., 2016). Its documented use notwithstanding, HAND's potential applications to the depiction of stream geometry information and to the investigation of stage-discharge connections have not been well investigated. Hydraulic methods of discharge calculation typically entail hydraulic parameters derived from the known geometry of a channel. The knowledge of a channel's cross sectional design is a requirement for many one-dimensional flood routing models, for instance the one-dimensional St. Venant equation (Brunner, 2016). ~~The~~ Even though the use of DEM interpolated bathymetry, as used by our method, induces error in the modelling of flood inundation, it is a necessity in the absence of bathymetry data. There are several instances in literature (Sanders, 2007) where the DEM interpolated bathymetry has been tested in place of actual bathymetry for hydrodynamic flood modelling. Furthermore, the requirement of the cross-section being perpendicular to the flow direction makes it an implicit problem and also dependent on the choice of cross-section position as well as the distance at which the points are taken on the cross-section. In the current practice of hand designing it makes it subjective and draws substantial uncertainty in the inundation simulation. Alternatively, HAND-based models do not explicitly solve the Manning's equation at individual cross-section, but rather solve for a catchment averaged version of it, by considering a river as a summation of infinite cross-sections. As such, the inherent uncertainty is avoided. However, the simplistic HAND-based model struggles to simulate proper inundation extent in case of complex conditions such as meandering main channels and confluences (Afshari et. al. 2017). This model doesn't capture the dynamic flow characteristics such as backwater effects created by flood mitigation structures. Furthermore, the large flood depth and low flow velocity in the natural rivers makes the river subcritical on many occasions, specifically for large floodplains where the water slows down significantly. This causes the backwater effect very far upstream of the flooding locations which is not simulated in HAND based methods. Therefore, users have to be cautious in such cases.

The conceptual framework for implementing HAND to estimate the channel hydraulic properties and rating curve is as follows: for any reach at water level h, all the cells with a HAND value $<$ h compose the inundated zone F(h), which is a subarea of the reach catchment. The water depth at any cell in the inundated zone F(h) is the difference between the reach-average water level h and the HAND of that cell, $HAND_c$, which can be represented as: depth = $HAND_c$-h. Since a uniform reach-average water level h is applied to check the inundation of any cell within the catchment, the inundated zone F(h) refers to that reach level. The water surface area of any inundated cell is equal to the area of the cell A_c. This case study uses 30 metre x 30 metre grid cells, thus in this case $A_c = 900\ m^2$. The channel bed area for each inundated cell is given by

$$A_s = A_c\sqrt{(1 + slope^2)} \quad (4)$$

where slope is the surface slope of the inundated pixel expressed as rise over run or inverse tangent of the slope angle. This equation approximates the surface area of the grid cell as the area of the planar surface with surface slope, which intersects with the horizontal projected area of the grid cell. The flood volume of each inundated pixel at a water depth of h can be calculated as $V_c(h)=A_c(h-HAND_c)$. If the reach length L is known, the reach-averaged cross section area for each pixel is given by $A_i=V_c/L$. Similarly, the reach-averaged cross section wetted perimeter for each inundated pixel $P_i(h)= A_s/L$. Therefore, the hydraulic radius for each inundated pixel is given by $R_i=A_i/P_i$. Therefore, we can estimate the reach-averaged cross-section area $A = \cancel{\sum_i A_i,} \sum_i A_i,$ perimeter $P = \cancel{\sum_i P_i,} \sum_i P_i,$ and hydraulic radius R= A/P for the entire flooded area. We compared the composite Manning's n (Chow, 1959; Flintham & Carling, 1992; Pillai, 1962; Tullis, 2012) from 7 different methods: the Colebatch method; the Cox method; the Horton Method; the Krishnamurthy Method; the Lotter method; the Pavlovskii Method; and the Yen Method (McAtee, 2012). More details about these methods are in the supplementary Section S2 of this paper.

Thus the discharge Q(h) corresponding to inundation height can be computed by the Manning's equation and given by:

$$Q(h) = \frac{1}{n}R^{\frac{2}{3}}AS^{\frac{1}{2}} \quad (6)$$

where S is the slope of the river and n is the composite Manning's roughness coefficient. Figure 2c displays the sequence of methods outlined for the Catchment Integrated Manning's Equation method.

2.2.4 Stage 4: Upscaling and Data Conversion

The proposed InundatEd inundation model simulates the flood-depth distributions for each catchment independently. This makes this model suitable to be ported to a DGGS-based data model and processing system. Following the GIS preprocessing, done in TauDEM as discussed in Section 2.2.1, the required data was converted to a DGGS representation, as outlined in Robertson et al., (2020). Supplementary Figure S2 for raster input data (S2a), polygon (vector) input data (S2b), and network (directional polyline vector) input data (S2c). For raster data (S2a), the bounding box is used to extract a set of DGGS cells, and then for each DGGS cell's centroid the raster value is extracted. To convert polygon data to a DGGS data model, we sample from its interior and its boundary separately using uniform sampling. Then each sample point is converted into DGGS cells based on its coordinates and stored into IDEAS data model by aggregating both sets of DGGS cells (Figure S2b). The same process for the border extraction is applied to the polylines and networks, however with network data the order of the cells is also stored as a flag to use in directional analysis (Figure S2c). Following conversion, the data was ported to a 40-node IBM Netezza Database for subsequent calculations. General, systematic limitations of the InundatEd IDEAS-based inundation model are discussed in Section 3.1.

2.3 Web-GIS Interface

The R/Shiny platform and the R-Studio development environment were used to design the user interface and server components of an online web application, allowing users to query and interact with the inundation model. Features of R specific to InundatEd's modelling workflow were its support of the Hazus damage functions and its support for DGGS spatial data. Shown in Figure 3a, the InundatEd user interface offers widgets for the following user inputs: address (text); discharge (slider); and return period (drop down), as well as tabs for viewing interactive graphs. The InundatEd user interface also features an interactive map which leverages the Leafgl R package (Appelhans & Fay, 2019) for seamless integration with the DGGS data model. Users may click on the map to obtain point-specific depth information.

2.4 InundatEd Flood Information System – System Structure Summary

Figure 3b displays the overall system structure and linkages for the InundatEd flood information system. GIS input data, as discussed in Section 2.2, were staged, pre-processed, and ported to the database. Data querying was used to compute 'in-database' inundation (flood depth) and related damages (methods outlined in Section 2.1) in response to user interface inputs to the R/Shiny UI.

2.5 Flood Data Comparison and Model Testing

2.5.0 Study Areas

As preliminary testing domains, we created flood inundation models for the Grand River Basin and Ottawa River Basin respectively, both located in Ontario, Canada. Each basin has experienced historical flooding and have implemented varying measures of flood control. Table 2 shows different salient characteristics of these catchments. For the purposes of graphing and discussion of station-specific period of record (number of years with a recorded annual maximum discharge) on theoretical vs estimated flood quantiles, two stations from each study watershed were selected, one each for high period of record and low period of record. For the Grand River watershed, stations 02GA003 and 02GA047 were selected for high and low period of record, respectively. For the Ottawa River watershed, stations 02KF006 and 02JE028 were selected, respectively. "Theoretical quantiles" are here defined as the quantiles generated by our model based on the logspline fit, which incorporates annual maximum discharge values from multiple stations across each study watershed (Section 2.2.2 and Figure 3). In contrast, "estimated quantiles" are here defined as the flood quantiles calculated simply by extracting the quantiles for the desired return periods from the raw annual maximum discharge values observed at the hydrometric station of interest.

<u>2.5.1. Ottawa River Watershed</u>

Four flood extent polygons (FEPs) provided by Natural Resources Canada (Natural Resources Canada, 2018, 2020) from the May-June 2019 flood season were used as "observed" floods to test the model outputs for the Ottawa River watershed. Each FEP represented a previously digitized floodwater extent at a specified date/time.

A second criterion for selection was that the hydrometric station(s) intersected by the FEP provided discharge data for the FEP's respective datetime. Two hydrometric stations which met both criteria were selected: 02KF005 and 02KB001. The following procedure was followed for each FEP using the corresponding hydrometric station (02KF005 or 02KB001), the station level index flood ($\tilde{Q}$, previously calculated during Section 2.2.2), and the observed discharge (Q_{obs}). In both cases, the logspline fit for the Ottawa River watershed, previously generated during Section 2.2.2, was also used.

The observed discharge (Q_{obs}) was divided by the corresponding hydrometric station's index flood ($\tilde{Q}$) ($Q_i = Q_{obs} / \tilde{Q}$) The cumulative probability of Q_i was then converted to a return period.

To generate each simulated flood for comparison to its observed counterpart, the methodology outlined in Sections 2.2.2 and 2.2.3 was repeated with the four new return periods appended to the original list of return periods in Section 2.2.2. Table 3 lists each FEP, the corresponding intersected hydrometric station, the period of record used for each station to calculate $\tilde{Q}$, the observed discharge, the resultant cumulative probability value, and the final return period used to generate each simulated flood.

2.5.2. Grand River Watershed

Regulatory floodplain extent data (the greater of RP=100 or discharge from Hurricane Hazel, "observed" flood extent) was obtained from the Grand River Conservation Authority (GRCA) (Grand River Conservation Authority, 2019). However, analysis revealed that, at most hydrometric stations in the Grand River watershed, the 100-year return period yielded higher discharge values relative to the "Hurricane Hazel" storm. Thus, the 100-year return period could be used. The estimated flood extent for RP=100 was generated per sections 2.2.1-2.2.3. Table S1 provides a discharge comparison between the 100-year return period and the regulatory storm.

2.5.3. Flood Extent Comparisons

For both the Grand River watershed and the Ottawa River watershed, only those subcatchments in close proximity to the observed flood extent polygons were retained for visualization purposes. To this end, a criterion was applied to subcatchments in the Grand River watershed requiring an intersection with the observed flood polygon of >= 20% of the subcatchment's area. For the Ottawa River watershed, due to the use of station-specific observed discharges, an additional criterion was applied: that a given subcatchment intersects with a network line with contributory upstream area >= 80% and contributory upstream area <= 120% of the observed upstream area of the hydrometric station (02KF005 or 02KB001). Table S2 provides by-subcatchment areas of the observed flood extent polygons whose subcatchments were eliminated based on the 20% intersection threshold. Per Table S2, one excluded subcatchment (10505) had an intersection value >= 20%, attributable in part to the presence of a tributary along which it was not expected that the return period would be properly scaled but which intersected the subcatchment. Additionally, due to the pluvial nature of the flooding in that subcatchment, it was once again expected that the return period as a function of the river discharge would not be properly scaled without the presence of a hydrometric station to provide discharge information.

Binary classification metrics have been used to compare between observed and simulated floods in cases where the focus is on extent, not depth (eg Papaioannou et al., 2016; Wing et al., 2017; Chicco & Jurman, 2020). A binary classification (or 2x2 contingency) method was used to compare the simulated flood extent rasters to the extents of their observed counterparts, whereby a confusion matrix was generated for each subcatchment. Multiple accuracy measures were calculated from the contingency tables to support the evaluation of the flood model, including: True Positive Rate (TPR). True Negative Rate (TNR), Accuracy, Matthews Correlation Coefficient (MCC) (Chicco & Jurman, 2020; Esfandiari et al., 2020; Rahmati et al., 2020), and the Critical Success Index (CSI) (e.g., Papaioannou et al., 2016; Stephens & Bates, 2015). Both the CSI and the MCC have been used in the context of flood model validation. The Critical Success Index (CSI) is defined as:

$$CSI = \frac{TP}{TP + FN + FP} \quad (7)$$

The Matthews Correlation Coefficient (MCC) is defined as:

$$MCC = \frac{TP\ x\ TN - FP\ x\ FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (8)$$

such that TP = true positive, TN = true negative, FP = false positive, and FN = false negative.

3. Results and Discussion

3.1 Model Processes and DGGS

Intermediate model outputs for the Grand River and Ottawa River watersheds - Height Above Nearest Drainage, delineated river networks, and Manning’s n- are displayed in Figure S3. Figure 4 visualizes results for the Grand River watershed and for the Ottawa River watershed for the following method components: calculation of hydrometric station upstream (contributory) area; index flood regression as represented by the correlation of logged index discharge and logged upstream area; and flood frequency as represented by discharge against a Gumbel transformed return period (years), for the stations respectively representative of high and low observations. Figures 4a and 4b plot the log of calculated upstream area against the log of observed upstream area, yielding respective Pearson correlation coefficients of 0.99 and 0.63 for the Grand River and Ottawa River watersheds. The relatively weak correlation of the Ottawa River watershed arose primarily from the limited resolution (number of decimal places in lat-long) of the station location information; incorrect reporting of station locations and/or their drainage area (Environment Canada reported the drainage area as 0 for multiple stations); and sometimes wrongly snapping stations to the tributaries rather than to the main river, particularly in cases involving a wide river channel or braided river. However, this does not affect the model itself, as we have used the station-specific drainage areas reported by Environment Canada to create the regional regression model. With respect to regional regression, Figure 4c visualizes the relationship between predicted index flood discharge and contributory upstream area, at individual hydrometric stations, for the Grand River and Ottawa River watersheds ($R = 0.83$ and 0.95, respectively).The regional growth curves for both the Grand River watershed and the Ottawa River watershed are shown in Figure 4d. To compare the proposed approach of using log-spline distribution against a traditional parametric distribution we fitted a Generalized Extreme Value (GEV) distribution to the RGC (Supplementary Figure S4). With respect to the log-spline RGCs, AIC values of 1861.69 and 867.69 and (-2)(logliklihood) values of 1826.04 and 809.26 were reported for the Grand River watershed and Ottawa River watershed respectively. The log-spline (-2)(logliklihood) values were lower than their GEV counterparts (1837.56 and 880.12) for both watersheds. For the Ottawa River watershed, the log-spline AIC value, 867.69, was also lower than that of its GEV counterpart (886.12). Furthermore, the use of the log-spline distribution allows for a consistent method which can be applied readily across any watershed without careful calibration of the distribution function. Thus, the log-spline distribution was used for the regional growth curves. The lower values of the normalized discharge shown in Figure 4d for higher return periods (2-3) for the Ottawa River watershed suggest relatively more structural alterations within the watershed, for instance flood control and dams, than the Grand River watershed (Ottawa Riverkeeper, 2020). The Grand River watershed yielded relatively higher values of normalized discharge (>3) at higher return periods in Figure 4d. Figure 5 shows the comparison of estimated flood quantiles against theoretical flood quantiles at an individual station from each study watershed. The stations - 02GA034 of the Grand River watershed and 02KF001 of the Ottawa River watershed (Figure 1)- were selected due to their long “discharge counts”, referring to the number of years for which an annual maximum discharge was recorded at each station. Specifically, station 02GA034 (5a) yielded a discharge count of 101 and station 02KF001 (5b) yielded a discharge count of 84. Return periods (T, years) have been converted in terms of the Gumbel reduced variable as follows:

$$Gumbel = -ln\left[ln\left(\frac{T}{T-1}\right)\right] \quad (9)$$

The dotted lines on Figures 5a and 5b represent the 5T threshold - the return period limit beyond which flood simulations can not be reasonably estimated. The 5T threshold requires that, for the reasonable estimation of a quantile for a desired return period T, there be at least 5T years of data (Hailegeorgis & Alfredsen, 2017; Jacob et al., 1999). As expected, the theoretical and estimated return periods are comparable for low return periods. However, and as shown in Figure 5, the theoretical and estimated quantiles deviate at lower RP values than the 5T threshold for both stations. This disagreement between the theoretical and estimated quantiles recalls the assumption of homogeneity for each watershed (Burn, 1997) - estimations of higher return periods, considering the 5T rule, would require more observations. However, further sub-sampling the stations into regional homogeneous groups would have reduced the data quantity substantially for each group.

3.2 Web-GIS Interface

A pre-alpha version of the InundatEd app is available at https://spatial.wlu.ca/inundated/. Source code for the most recent version of InundatEd will be publicly available on GitHub (Spatial Lab, 2020). The use of R/Shiny to develop InundatEd and its provision on GitHub encourages transparency, ongoing development, and response to user feedback and preferences.

3.3 Model Testing

Of the binary comparison results for the 7 composite Manning’s n methods listed in Section 2.2.3, the Krishnamurthy method yielded the highest median CSI values (Table S3 for the Grand River watershed and Table S4 for the Ottawa River watershed). As such, it was selected for further visualization and discussion.

The following return periods (in years) were observed for FEPs intersecting hydrometric station 02KF005 in the Ottawa River watershed: 26.5, 16.52, and 25.96. Additionally, a return period of 42.69 years was observed for a FEP intersecting hydrometric station 02KB001 in the Ottawa River watershed. The 100-year return period was tested for the Grand River watershed. Binary classification results for the Grand River watershed are shown in Figure 6 for four comparison metrics: Critical Success Index, Matthews Correlation Coefficient, True Positive Rate, and True Negative Rate. Figure 7 presents Critical Success Index and Matthews Correlation Coefficient results for the four Ottawa River watershed cases, with True Positive and True Negative results presented in Supplementary Figure S5. Table 4 lists the number of subcatchments evaluated, the median CSI, and the median MCC for each of the 5 test return periods. The median values of additional metrics are provided in Table S5.

The median CSI values ranged from 0.581 to 0.849 (Table 4), with both of those values coming from the Ottawa River watershed (return periods 42.69 and 26.5, respectively). The median MCC values ranged from 0.743 (Ottawa RP 42.69) to 0.888 (Ottawa RP 26.5). The median CSI and MCC values for the Grand River watershed were 0.741 and 0.844, respectively. The results reported herein are comparable to, and in some cases exceed, previously published binary classification results. ~~For instance, Wing et al. (2017) achieved CSI values of 0.552 and 0.504 for a 100-year return period flood model of the conterminous United States at a 30m resolution. With~~For instance, with respect to the MCC, an urban flood model produced by Rahmati et al. (2020) provided an MCC value of 0.76 when compared to historical flood risk areas. Esfandiari et al. (2020) compared two flood simulations: a HAND-based flood model and a model which combined HAND and machine learning to observe flood extents, resulting in a range of MCC values from ~0.77 to ~0.85. Bates et al. (2021) achieved CSI values of 0.69 and 0.82 for a 100-year return period flood model of the conterminous United States at a 30m resolution. It must be noted that direct comparisons between the works listed here and this study must be viewed with caution, due to differences in methodologies, assumptions, data sources, data availability, and return periods between the studies. Furthermore, the extent comparison scores are not necessarily objective measures of performance of the simulation model. They can vary depending on the severity of the flood, catchment characteristics, and quality of the benchmark data (Mason et. al. 2009, Stephens et al., 2014, Wing et. al. 2021).

Additionally, the median F_1 score (Chicco & Jurman, 2020) for the Grand River watershed was 0.85. The median F_1 scores for Ottawa River watershed return periods 26.5, 16.52, 25.96, and 42.69 were 0.96, 0.95, 0.95, and 0.94 respectively. Such results are approximately in line with

Pinos & Timbe (2019), who achieved F_1 values from 0.625 to 0.941 for 50-year RP floods using a variety of 2D dynamic models. Afshari (~~2017~~2018) achieved F_1 values from 0.48 - 0.64 for the 10-year, 100-year, and 500-year return periods when comparing a HAND-based simulation against a HEC-RAS 2D control. Lim & Brandt (2019) determined that low-resolution DEMs are capable of yielding relatively high comparison metrics (e.g. F_1 values approximately >= 0.80) in situations where Manning's n varies widely over space. The connection between high values of Manning's n and flood overestimation (false discovery) was also discussed. The Grand River watershed yielded a median False Discovery Rate (FDR) of 0.117, and the four Ottawa River watershed cases yielded respective median FDRs of 0.019, 0.01, 0.006, and 0.44 for the evaluated subcatchments. The moderately high FDR value of 0.44 for the 42.69-year return period and the observed overestimation of flood extent (discussed below) may be a result of high local Manning's n values. In addition, the influences of flat terrain (Lim & Brandt, 2019) and anabranch must be considered as it can disrupt the assumption of a single drainage direction for each pixel during sub-catchment delineation. Additional factors potentially influencing the overestimation are the problems inherent to HAND-based modeling, as discussed in section 2.2.3. The topography of the area of the Ottawa River watershed wherein the extent comparisons were made is relatively flat with multiple anabranches and thus can lead to chaotic network delineation. Although attempts were made in this model to counter this impact and avoid slope values of 0 (the burning of the polyline network into the DEM, Section 2.2.1 and Figure 2a), the use of the Manning's equation was still compromised in certain areas and likely had a negative impact on the resultant flood simulations.

As noted in Lim & Brandt (2019), the reliability of the observed flood extent polygons also merits comment. In this case study, the observed FEPs for the Ottawa River watershed were originally digitized from remotely sensed data and thus carry forward the errors and uncertainties from prior processing. The Grand River watershed's 100-year return period extent was also generated outside of this study and potentially carries multiple sources of error and uncertainty. However, evaluation of the exact extent to which errors present in the observed flood extent polygons could have impacted the binary classification results was not an objective of this study.

Figure 8 visualizes the 100-year return period simulated flood for the Grand River watershed. Inset maps are provided which highlight one subcatchment with a high CSI (A, CSI= 0.77) and two subcatchments with low CSIs (B, CSI =0.17 and 0.22). The simulated flood shown in Figure 8A compares very well to the extent of its observed counterpart, consistent with the relatively high CSI value. Notably, three hydrometric stations are located within the Figure 8A subcatchment: 02GA014, 02GA027, and 02GA016. Per the methods in Section 2.2.2, station 02GA014 yielded a period of record of 54, 02GA027 yielded an insufficient (<10) period of record, and station 02GA016 yielded a period of record of 58. The presence of the two hydrometric stations with considerable periods of record likely strengthened the regional regression of the area and contributed to the success of the simulated flood shown in Figure 8A. In contrast, within the low-CSI (0.17 and 0.22) subcatchments shown in Figure 8B, the simulation considerably overestimated the extent of the 100-year return period flood. The overestimation of the flood extents observed in Figure 8B can likely be attributed, at least in part, to the following: a) multiple upstream and downstream dams (Grand River Conservation Authority, 2000) and b) the channel meanders - as discussed previously, the simple HAND-based model employed here is not robust against channel complexities nor flow control structures such as dams. It must be recalled here that the modular nature of the InundatEd model allows for the "swapping" of various flood modelling methods, and thus could easily accommodate, for instance, shallow water equations. It is also possible to include such operations in future versions of the model by either modifying the DEM values to reflect flood control structures or by offsetting the discharge of the catchment based on structure storage.

With respect to the Ottawa River watershed, Figure 9 highlights subcatchments whose comparison between observed and simulated flood extents yielded low (A: CSI= 0.13) , moderate (B: CSI = 0.66 and D: CSI =0.65) and high (C: CSI = 0.87) CSI values.

Figure 9A shows the simulated and observed flood extents for return period 25.69. Two main factors influencing the low CSI are readily apparent. The first is that the observed FEP appears "cut off", not extending through most of the subcatchment. It is possible that the flood in the remainder of the sub-catchment was simply not digitized during the observed FEP's generation, especially given the subcatchment's position. However, of the area of the subcatchment intersected by the observed FEP, the simulated flood has considerably underestimated the observed flood extent. Figure 9B shows the extent comparison of the 42.69 -year return period in a subcatchment of moderate CSI (0.66). Figure 9C illustrates a subcatchment of high CSI (0.87), characterized by an overall underestimation in flood extent, barring a slight overestimation in one area. Figure 9D (CSI = 0.65) shows a mixture of overestimation and underestimation.

Although the results for both the Grand River watershed and the Ottawa River watershed suggest substantial agreement between the respective observed and simulated flood extents, a number of considerations, including input data characteristics and metric bias, require that the presented results be taken with caution and, in some cases, offer clear paths for improvement. With respect to input data, the simulated floods presented within this case study are limited by the initial use of a 30m x 30 DEM raster. As concluded by Papaioannou et al. (2016), floodplain modelling is sensitive to both the resolution of the input DEM and to the choice of modelling approach. Additionally, and as discussed in Section 2.2.3, there are some inherent limitations of the HAND-based modeling approach.

Overall, the results indicated that the current iteration of the InundatEd flood model was reasonably successful on the basis of moderate-high MCC values ~~and direct~~indirect comparisons against the observed flooding extents. However, any weight assigned to this claim must, in addition to the previously discussed caveats, recall that only extent and not depth was compared between the observed and simulated floods. The use of the DGGS big-data architecture provides a promising foundation for further work, such as the incorporation of the impacts of flood control structures, on the InundatEd model.

3.4 Model Performance

~~Supplementary Figure S7 contrasts~~There is a distinct contrast of runtimes ~~using~~between the DGGS method ~~against~~and those using a~~-~~ traditional, raster-based method for sub-catchments within the Grand River Watershed (n= 306 for each method) during the generation of respective RP 100 flood maps. ~~To account for the substantial difference between the~~The DGGS ~~runtime range~~based storing and ~~that~~processing method is an order of ~~its~~magnitude faster than processing the HAND and catchment boundaries using raster ~~counterpart, we added 4 seconds to DGGS runtime in Figure S7~~and vector format. The mean runtime using the DGGS method (0.23 seconds) was significantly lower than the mean runtime using the raster-based method (3.98 seconds) at both the 99% confidence intervals ($p < 2.2e-16$). Thus, the efficiency of the proposed inundation model -coupled with a big-data Discrete Global Grids Systems architecture- is demonstrated with respect to processing times with limited input data. As the IDEAS framework and the InundatEd flood modelling method continue to develop, processing time benchmarks could be established to track and evaluate the model's robustness against increasing complexity (e.g., the integration of hydrological processing algorithms) and to facilitate comparisons with other inundation models.

3.5 Conclusions

We have tested a novel flood modelling and mapping system, implemented within a DGGS-based big data platform. In many parts of the world, including Canada, the widespread deployment of detailed hydrodynamic models has been hindered by complexities and expenses regarding input data and computational resources, especially the dichotomy between processing time and model complexity. This research proposes a novel solution to these challenges. First, we demonstrated the development of a flood modelling framework in a Discrete Global Grid Systems (DGGS) data model and the presentation of the models' outputs via an open-source R/Shiny interface robust against algorithm modifications and improvements. The DGGS data model efficiently integrates heterogeneous spatial data into a common framework, rapidly develops models, and can scale for thousands of unit processing regions through easy parallelization. ~~Second, the use of the catchment-integrated Manning's equation avoids high-uncertainty river cross-sections and produces physically justified flood inundation extents.~~Second, the computational framework has been implemented using a regional dataset over locations and at scales which have not been studied before. We successfully demonstrated the merit of the HAND-based inundation modelling to emulate the observed flooding extent for several historical and design floods. Third, DGGS-powered analytics allow users to quickly visualize flood extents and depths for regions of interest, with reasonable alignment with observed flooding events. Finally, we believe our flood-inundation estimation method can address situations where good quality data is scarce and/or there are insufficient resources for a complex model. To apply the model in a real time environment we would need a discharge forecasting model or have real-time discharge data at the catchment outlet, which could be used to compute the flood inundation using the pre-computed stage-discharge relationship and inundation model.

4. References

Afshari S, Ahmad A. Tavakoly, Mohammad Adnan Rajib, Xing Zheng, Michael L. Follum, Ehsan Omranian, Balázs M. Fekete (2018). Comparison of new generation low-complexity flood inundation mapping tools with a hydrodynamic model. *Journal of Hydrology, 556*, 539-556, ISSN 0022-1694, https://doi.org/10.1016/j.jhydrol.2017.11.036.

Albano, R., Sole, A., Adamowski, J., Perrone, A., & Inam, A. (2018). Using FloodRisk GIS freeware for uncertainty analysis of direct economic flood damages in Italy. *International Journal of Applied Earth Observation and Geoinformation*, *73*, 220–229. ~~https://doi.org/10.1016/j.jag.2018.06.019~~https://doi.org/10.1016/j.jag.2018.06.019

Appelhans, T., & Fay, C. (2019). leafgl: Bindings for Leaflet.glify. R package version 0.1.1.

Attari, M., & Hosseini, S. M. (2019). A simple innovative method for calibration of Manning's roughness coefficient in rivers using a similarity concept. *Journal of Hydrology*, *575*, 810–823.

~~https://doi.org/10.1016/j.jhydrol.2019.05.083~~https://doi.org/10.1016/j.jhydrol.2019.05.083

Bates, P. D., Quinn, N., Sampson, C., Smith, A., Wing, O., Sosa, J., et al. (2021). Combined modeling of US fluvial, pluvial, and coastal flood hazard under current and future climates. *Water Resources Research*, 57, e2020WR028673. https://doi.org/10.1029/2020WR028673

Brunner, G. W. (2016). *HEC-RAS River Analysis System 2D Modelling User's Manual Version 5.0.* (Report Number CPD-68A). US Army Corps of Engineers Hydrologic Engineering Center.

Burn, D. H. (1994). Hydrologic effects of climatic change in west-central Canada (1994). *Journal of Hydrology, 160(1-4),* 53-70. ISSN 0022-1694. https://doi.org/10.1016/0022-1694(94)90033-7.

Formatted: Indent: Left: 0 cm

Burn, D. H. (1997). Catchment similarity for regional flood frequency analysis using seasonality measures. Journal of Hydrology 202 (1997) 212–230.

Formatted: Indent: Left: 0 cm

Burrell, B., & Keefe, J. (1989). Flood risk mapping in ~~new brunswick~~New Brunswick: A decade review. *Canadian Water Resources Journal*, *14*(1), 66–77. https://doi.org/10.4296/cwrj1401066

Calamai, L., & Minano, A. (2017). Emerging trends and future pathways: A commentary on the present state and future of residential flood insurance in Canada. *Canadian Water Resources Journal*, *42*(4), 307–314. https://doi.org/10.1080/07011784.2017.1362358

Canada Centre for Mapping and Earth Observation (2015). *Canadian Digital Elevation Model, 1945-2011 (Record ID 7f245e4d-76c2-4caa-951a-45d1d2051333).* [Data set]. Natural Resources Canada. Retrieved from https://open.canada.ca/data/en/dataset/7f245e4d-76c2-4caa-951a-45d1d2051333#wb-auto-6

Canada Centre for Remote Sensing (2019). *2015 Land Cover of Canada (Record ID 4e615eae-b90c-420b-adee-2ca35896caf6)*. [Data set]. Natural Resources Canada. Retrieved from https://open.canada.ca/data/en/dataset/4e615eae-b90c-420b-adee-2ca35896caf6

Centre For Research On The Epidemiology Of Disasters – CRED (2015). *"The human cost of natural disasters" – 2015: A global perspective*. CRED: Brussels. Accessed from https://www.cred.be/index.php?q=HCWRD.

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*,6. doi: 10.1186/s12864-019-6413-7.

Chow, V.T. (1959). *Open-channel hydraulics. McGraw-Hill.*

Comber, A., & Wulder, M. (2019). Considering spatiotemporal processes in big data analysis: Insights from remote sensing of land cover and land use. *Transactions in GIS, 23(5),* 879–891. https://doi.org/10.1111/tgis.12559

Craglia, M., de Bie, K., Jackson, D., Pesaresi, M., Remetey-Fülöpp, C., Wang, C., et al. (2012). Digital Earth 2020: Towards the vision for the next decade. *Int. J. Digital Earth, 5(1)*,4-21. 10.1080/17538947.2011.638500

Craglia, M., Goodchild, M.F., Annoni, A., Câmara, G., Gould, M.D., Kuhn, W., et al. (2008). Next-Generation Digital Earth (Editorial). *Int. J. Spat. Data Infrastruct. Res.*, *3*,146-167. 10.2902/1725-0463.2008.03.art9.

Dalrymple, T. (1960). *Rep. No Water Supply Paper 1543-A*. U.S. Geological Survey, Reston, VA, U .S.

Eaton, B., Church, M., & Ham, D. (2002). Scaling and regionalization of flood flows in British

Formatted: Indent: Left: 0 cm

Columbia, Canada. *Hydrol. Processes 16*, 3245–3263,

DHI, 2003. MIKE 11-A Modelling System for Rivers and Channels - User Guide. DHI, p. 430.

DHI, 2012. MIKE 21-2D Modelling of Coast and Sea. DHI Water & Environment Pty Ltd.

Ehlschlaeger, C., and Shortridge, A., (1996) Modeling Elevation Uncertainty in Geographical Analyses, , Proceedings of the International Symposium on Spatial Data Handling, Delft, Netherlands, 9B.15-9B.25.

England, J.F., Jr., Cohn, T.A., Faber, B.A., Stedinger, J.R., Thomas, W.O., Jr., Veilleux, A.G., Kiang, J.E., & Mason, R.R., Jr. (2018). Guidelines for determining flood flow frequency - Bulletin 17C (ver. 1.1). U. S. Geological Survey Techniques and Methods, book 4, chap. B5, 148 p.https://doi.org/10.3133/tm4B5.

~~http://dx.doi.org/10.1002/hyp.1100.~~

Environment and Climate Change Canada (2019). *An Examination of Governance, Existing Data, Potential Indicators and Values in the Ottawa River Watershed. I*SBN: 978-0-660-31053-4

Esfandiari, M., Abdi, G., Jabari, S., McGrath, H., & Coleman, D. (2020). Flood hazard risk mapping using a pseudo supervised random forest. *Remote sensing, 12(19),* 1-23. DOI: 10.3390/rs12193206

Faulkner, D., Warren, S., & Burn, D. (2016). Design floods for all of Canada. *Canadian Water Resources Journal, 41(3)*, 398-411. 10.1080/07011784.2016.1141665.

Ferrari, A., Dazzi, S., Vacondio, R., & Mignosa, P. (2020). Enhancing the resilience to flooding induced by levee breaches in lowland areas: a methodology based on numerical modelling. *Nat. Hazards Earth Syst. Sci., 20*, 59–72.

Fisher, P. F. (1991). First experiments in viewshed uncertainty: the accuracy of the viewshed area, Photogramm. Eng. Rem. S., 57, 1321– 1327.

Flintham, T. P., and Carling, P. A. (1992), "Manning's n of Composite Roughness in Channels of Simple Cross Section." In Channel Flow Resistance: Centennial of Manning's formula, B. C. Yen, ed., Water Resource Publications, Highlands Ranch, CO (1992) pp. 328–341.

Fouad, G., Skupin, A., & Tague, C. L. (2016). Regional regression models of percentile flows for the contiguous US: Expert versus data-driven independent variable selection. *Hydrology and Earth Systems Sciences Discussions, 17*, 1-33. 10.5194/hess-2016-639.

Garousi-Nejad, I., Tarboton, D.G., Aboutalebi, M., & Torres-Rua, A. F. (2019). Terrain analysis enhancements to the Height Above Nearest Drainage flood inundation mapping method. *Water Resources Research*, *55*, 7983-8009. https://doi.org/10.1029/2019WR024837

Gebetsroither-Geringer, E., Stollnberger, R., & Peters-Anders, J. (2018). Interactive Spatial Web-Applications as New Means of Support for Urban Decision-Making Processes. In *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (Vol. 4, pp. 59–66). Delft, The Netherlands. https://doi.org/10.5194/isprs-annals-IV-4-W7-59-2018

Goteti, Gopi (2014). hazus: Damage functions from FEMA's HAZUS software for use in modeling financial losses from natural disasters. R package version 0.1. Retrieved from https://CRAN.R-project.org/package=hazus

Grand River Conservation Authority (2019). *Regulatory Floodplain* [Data set]. Grand River

Conservation Authority. https://data.grandriver.ca/downloads-geospatial.html

Grand River Conservation Authority (2014). *Grand River Watershed Water Management Plan Executive Summary - March 2014*. Cambridge, ON. Retrieved from https://www.grandriver.ca.

Grand River Conservation Authority (2000). *Dams*. [Data set]. Grand River Conservation Authority. https://data.grandriver.ca/downloads-geospatial.html.

Haddad, K., Rahman, A., & Kuczera, G. (2011). Comparison of Ordinary and Generalised Least Squares Regression Models in Regional Flood Frequency Analysis: A Case Study for New South Wales. *Australasian Journal of Water Resources, 15(1),* 59-70, doi: 10.1080/13241583.2011.11465390

Handmer, J. W. (1980). Flood hazard maps as public information: An assessment within the context of the Canadian flood damage reduction program. *Canadian Water Resources Journal*, *5*(4), 82–110. https://doi.org/10.4296/cwrj0504082

Hailegeorgis, T. T., & Alfredsen, K. (2017). Regional flood frequency analysis and prediction in ungauged basins including estimation of major uncertainties for mid-Norway. *Journal of Hydrology: Regional Studies, 9*, 104-126.

Hunter, G. and Goodchild, M., (1997), Modeling the Uncertainty of Slope and Aspect Estimates Derived From Spatial Databases, Geographical Analysis, Vol. 29, No. 1, p. 35-49.

Hutchinson, David. (2016). HYDAT: An interface to Canadian Hydrometric Data. R package version 1.0. [GitHub Repository]. Retrieved from https://github.com/CentreForHydrology/HYDAT.git

Jacob, D., Reed, D. W., & Robson, A. J. (1999). *Choosing a Pooling Group Flood Estimation Handbook.* Institute of Hydrology, Wallingford, U. K.

Jamali, B., Löwe, R., Bach, P. M., Urich, C., Arnbjerg-Nielsen, K., & Deletic, A. (2018). A rapid urban flood inundation and damage assessment model. *Journal of Hydrology*, *564*, 1085-1098. https://doi.org/10.1016/j.jhydrol.2018.07.064.

Juraj M., Cunderlik, T., & Ouarda, B. M. J. (2009). Trends in the timing and magnitude of floods in Canada, Journal of Hydrology, Volume 375, Issues 3–4, 2009, Pages 471-480,ISSN 0022-1694, https://doi.org/10.1016/j.jhydrol.2009.06.050.

Kalyanapu, A. J., Shankar, S., Pardyjak, E. R., Judi, D. R., & Burian, S. J. (2011). Assessment of GPU computational enhancement to a 2D flood model. *Environmental Modelling and Software*, *26*(8), 1009–1016. https://doi.org/10.1016/j.envsoft.2011.02.014

Kamal, V., Mukherjee, S., Singh, P. *et al.* (2017). Flood frequency analysis of Ganga river at Haridwar and Garhmukteshwar. *Appl Water Sci* **7,** 1979–1986. https://doi.org/10.1007/s13201-016-0378-3

Kaur, B., Shrestha, N. K., Daggupati, P., Rudra, R. P., Goel, P. K., Shukla, R., Allataifeh, N. (2019). Water Security Assessment of the Grand River Watershed in Southwestern Ontario, Canada. *Sustainability, 11(7).* doi: http://dx.doi.org/10.3390/su11071883.

Langat, P.K., Kumar, L., & Koech, R. (2019). Identification of the Most Suitable Probability Distribution Models for Maximum, Minimum, and Mean Streamflow. *Water*, *11*, 734.

L'homme J., P. Sayers, B. Gouldby, P. Samuels, M. Wills, J. (2008) Mulet-Marti Recent development and application of a rapid flood spreading method P. Samuels, S. Huntington, W. Allsop, J. Harrop (Eds.), Flood Risk Management: Research and Practice, Taylor &

Francis Group, London, UK,

Li, M., Stefanakis, E. (2020) Geospatial Operations of Discrete Global Grid Systems—a Comparison with Traditional GIS. *J geovis spat anal* **4,** 26. https://doi.org/10.1007/s41651-020-00066-3

Lim, N. J., & Brandt, S. A. (2019). Are Feature Agreement Statistics Alone Sufficient to Validate Modelled Flood Extent Quality? A Study on Three Swedish Rivers Using Different Digital Elevation Model Resolutions. *Mathematical Problems in Engineering, 2019,* 9816098. https://doi.org/10.1155/2019/9816098.

Liu, Y.Y., Maidment, D.R., Tarboton, D.G., Zheng, X., and Wang, S.. (2018). A CyberGIS Integration and Computation Framework for High-Resolution Continental-Scale Flood Inundation Mapping. Journal of the American Water Resources Association 54(4): 770–784. https://doi.org/10.1111/1752-1688.12660.

Lee, J., Snyder, P., Fisher, P., (1992), Modeling the Effect of Data Errors on Feature Extraction From Digital Elevation Models, Photogrammertic Engineering and Remote Sensing, Vol. 58, No. 10, p. 1461-1467.

Lee, J., (1996), Digital Elevation Models: Issues of Data Accuracy and Applications, Proceedings of the Esri User Conference, 1996.

Liu, R., (1994), The Effects of Spatial Data Errors on the Grid-Based Forest management Decisions, Ph.D. Dissertation, State University Of New York College Of Environmental Science and Forestry, Syracuse, NY, 209 pp.

Li, Z., Huang, G., Wang, X., Han, J., Fan, Y. (2016). Impacts of future climate change on river discharge based on hydrological interference: a case study of the Grand River Watershed in Ontario, Canada. *Science of the Total Environment, 548-549, 198-210.* https://doi.org/10.1016/j.scitotenv.2016.01.002.

Mahdavi-Amiri, A., Alderson, T., & Samavati, F. (2015). A Survey of Digital Earth. *Computers & Graphics, Volume 53, Part B,* Pages 95-117, ISSN 0097-8493, https://doi.org/10.1016/j.cag.2015.08.005.

Mason, D. C., Bates, P. D., & Dall' Amico, J. T. (2009). Calibration of uncertain flood inundation models using remotely sensed water levels, Journal of Hydrology, Volume 368, Issues 1–4, Pages 224-236, ISSN 0022-1694, https://doi.org/10.1016/j.jhydrol.2009.02.034.

McAtee, K. (2012). *Introduction to Compound Channel Flow Analysis for Floodplains*. SunCam. https://s3.amazonaws.com/suncam/docs/162.pdf.

Moulinec, C., Denis, C., Pham, C.T., Rouge, D., Hervouet, J.M., (2011). TELEMAC: an efficient hydrodynamics suite for massively parallel architectures. Comput. Fluids 51 (1), 30e34.

Nastev, M., & Todorov, N. (2013). Hazus: A standardized methodology for flood risk assessment in Canada. *Canadian Water Resources Journal*, *38*(3), 223–231. https://doi.org/10.1080/07011784.2013.801599.

Natural Resources Canada (2018). *Floods in Canada -Archive (Record ID 74144824-206e-4cea-9fb9-72925a128189).* [Data set]. Natural Resources Canada. Retrieved from https://open.canada.ca/data/en/dataset/74144824-206e-4cea-9fb9-72925a128189.

Natural Resources Canada (2020). *Flood in Canada Product Specifications.*

Neal, J., Dunne, T., Sampson, C., Smith, A., & Bates, P. (2018). Optimisation of the two-

Formatted: Left, Indent: Left: 0 cm, Hanging: 0.85 cm, Space Before: 0 pt, After: 0 pt

Field Code Changed

Formatted: Hyperlink, Font: Arial, 11 pt, No underline, Font color: Auto

Formatted: Indent: Left: 0 cm, First line: 0 cm

dimensional hydraulic model LISFOOD-FP for CPU architecture. *Environmental Modelling and Software*, *107*(May), 148–157. ~~https://doi.org/10.1016/j.envsoft.2018.05.011~~https://doi.org/10.1016/j.envsoft.2018.05.011

Néelz S., Pender G. (2010). Benchmarking of 2D Hydraulic Modelling Packages DEFRA/Environment Agency, UK.

Nix, G. A. (1987). Management of the Ottawa River Basin. *Water International*, *12(4)*, 183-188.

Nobre, A. D., Cuartas, L. A., Momo, M. R., Severo, D. L., Pinheiro, A., & Nobre, C. A. (2016). HAND contour: A new proxy predictor of inundation extent. *Hydrological Processes 30(2)*: 320-333. doi: 10.1002/hyp.10581

OGC. (2017). Topic 21: Discrete Global Grid Systems Abstract Specification. Retrieved July 26, 2019, from http://docs.opengeospatial.org/as/15-104r5/15-104r5.html

Ottawa Riverkeeper. (2020). *Dams*. [Website] Accessed at https://www.ottawariverkeeper.ca/home/explore-the-river/dams/.

Oubennaceur, K., Chokmani, K., Nastev, M., Lhissou, R., & El Alem, A. (2019). Flood risk mapping for direct damage to residential buildings in Quebec, Canada. *International Journal of Disaster Risk Reduction*, *33*, 44–54. https://doi.org/10.1016/j.ijdrr.2018.09.007

~~Pal, K. (2002). Assessing Community Vulnerability to Flood Hazard in Southern Ontario. *Canadian Water Resources Journal*, *27 (2)*, 155-173.~~

Papaioannou, G., Loukas, A., Vasiliades, L., & Aronica, G. T. (2016). Flood inundation mapping sensitivity to riverine spatial resolution and modelling approach. *Nat Hazards*, *83*, S117-S132. 10.1007/s11069-016-2382-1.

Pillai, C. R. S. (1962), "Composite Rugosity Coefficient in Open Channel Flow." Irrigation and Power, J. Central Board of Irrigation and Power, New Delhi, India, Vol. 19, No. 3 (1962) pp. 174–189

Pinos, J., & Timbe, L. (2019). Performance assessment of two-dimensional hydraulic methods for generation of flood inundation maps in mountain river basins. *Water Science and Engineering*, *12(1)*, *11-18*.ISSN 1674-2370

Prakash, M., Rothauge, K., Cleary, P.W. (2014). Modelling the impact of dam failure scenarios on flood inundation using SPH. Appl. Math. Model. 38 (23), 5515e5534.

Rahmati, O., Kornejady, A., Samadi, M., Nobre, A. D., & Melesse, A. M. (2018). Development of an automated GIS tool for reproducing the HAND terrain model. *Environmental Modelling and Software*, *102*, 1–12. https://doi.org/10.1016/j.envsoft.2018.01.004

Rahmati, O., Darabi, H., Panahi, M., Kalantari, Z., Naghibi, S. A., Ferreira, C. S. S., et al. (2020). Development of novel hybridized models for urban flood susceptibility mapping. *Scientific Reports*, *10(1)*, *1-19*. DOI: 10.1038/s41598-020-69703-7.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rennó, C.D., Nobre, A.D., Cuartas, L.A., Soares, J.V., Hodnett, M.G., Tomasella, J., & Waterloo, M. (2008). HAND, a new terrain descriptor using SRTM-DEM; mapping terra-firme rainforest environments in Amazonia. *Remote Sensing of Environment 112*, 3469–3481.

~~Robert, B., Forget, S., & Rousselle, J. (2003). The Effectiveness of Flood Damage Reduction Measures in the Montreal Region. *Natural Hazards*, *28*, 367–385. Retrieved from https://journals.scholarsportal.info/pdf/0921030x/v28i2-3/367_teofdrmitmr.xml~~

Formatted: Indent: First line: 0 cm

Robertson, C., Chaudhuri, C., Hojati, M., & Roberts, S. (2020). An integrated environmental analytics system (IDEAS) based on a DGGS. *ISPRS Journal of Photogrammetry and Remote Sensing, 162,* 214-228.

Robson, A. J., & Reed, D. W. (1999). *Flood Estimation Handbook, vol. 3 Statistical Procedures for Flood Frequency Estimation.* Institute of Hydrology, Wallingford, UK.

Rodda, Harvey J. (2005). The development and application of a flood risk model for the Czech Republic. *Natural hazards 36(1-2)*: 207-220. doi: 10.1007/s11069-004-4549-4

Salas, J. D., & Obeysekera J. (2014). Revisiting the Concepts of Return Period and Risk for Nonstationary Hydrologic Extreme Events, *J. Hydrol. Eng,19,* 554-568.

Samela, C., Albano, R., Sole, A., & Manfreda, S. (2018). A GIS tool for cost-effective delineation of flood-prone areas. *Computers, Environment, and Urban Systems, 70,* 43-52. https://doi.org/10.1016/j.compenvurbsys.2018.01.013

Samela, C., Troy, T. J., & Manfreda, S. (2017). Geomorphic classifiers for flood-prone areas delineation for data-scarce environments. *Advances in Water Resources, 102,* 13-28. http://dx.doi.org/10.1016/j.advwatres.2017.01.007

~~Shahab Afshari, Ahmad A. Tavakoly, Mohammad Adnan Rajib, Xing Zheng, Michael L. Follum, Ehsan Omranian, Balázs M. Fekete (2018). Comparison of new generation low-complexity flood inundation mapping tools with a hydrodynamic model, Journal of Hydrology,Volume 556, 2018, Pages 539-556, ISSN 0022-1694, https://doi.org/10.1016/j.jhydrol.2017.11.036.~~

Sanders, B. F. (2007). Evaluation of on-line DEMs for flood inundation modeling. *Advances in Water Resources, 30 (8),* 1831-1843. ISSN 0309-1708. https://doi.org/10.1016/j.advwatres.2007.02.005.

Schumann, G., Bates, P.D., Horritt, M.S., Matgen, P., Pappenberger, F., 2009. Progress in integration of remote sensing-derived flood extent and stage data and hydraulic models. Rev. Geophys. 47 (4), RG4001.

Shitanshu Desai, Taha B.M.J. Ouarda,
(2021), Regional hydrological frequency analysis at
ungauged sites with random forest regression,
Journal of Hydrology, ~~Volume~~ *594,* ~~2021,~~ 125861, ISSN 0022-1694

Singh, Vijay P.(2015). *Entropy Theory in Hydrologic Science and Engineering.* McGraw-Hill Education: New York, Chicago, San Francisco, Athens, London, Madrid, Mexico City, Milan, New Delhi, Singapore, Sydney, Toronto. Accessed at: https://www.accessengineeringlibrary.com/content/book/9780071835466

Smith, L.C., 1997. Satellite remote sensing of river inundation area, stage, and discharge: a review. Hydrol. Process. 11 (10), 1427e1439

Smith, A., Sampson, C., and Bates, P. (2014), Regional flood frequency analysis at the global scale, *Water Resour. Res.*, 51, 539– 553, doi:10.1002/2014WR015814.

Song, S., Schmalz, B., Zhang, J. X., Li, G., & Fohrer, N. (2017). Application of modified Manning formula in the determination of vertical profile velocity in natural rivers. *Hydrology Research, 48(1),* 133–146. doi: https://doi.org/10.2166/nh.2016.131

Spatial Lab (2020). InundatEd. [Github Repository]. Retrieved from https://github.com/thespatiallabatLaurier/floodapp_public

Formatted: Underline, Font color: Custom Color(RGB(17,85,204))

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Indent: Left: 0 cm, First line: 0 cm

Stephens, E., & Bates, P. (2015). Assessing the reliability of probabilistic flood inundation model predictions. *Hydrol. Process. 29,* 4264–4283. doi: 10.1002/hyp.10451.

Stephens, E., Schumann, G. and Bates, P. (2014). Problems with binary pattern measures for flood model evaluation. Hydrol. Process., 28: 4928-4937. https://doi.org/10.1002/hyp.9979

Stevens, M. R., & Hanschka, S. (2014). Municipal flood hazard mapping: the case of British Columbia, Canada. *Natural Hazards*, *73*, 907–932. https://doi.org/10.1007/s11069-014-1117-4

Stone, C. J., Hansen, M. H., Kooperberg, C., & Truong, Y. K. (1997). Polynomial Splines and their Tensor Products in Extended Linear Modeling. *The Annals of Statistics*, *25*(4), 1371–1425.

Strategic Policy and Innovation Centre (2019). *Lakes, Rivers and Glaciers in Canada – CanVec Series – Hydrographic Features (Record ID 9d96e8c9-22fe-4ad2-b5e8-94a6991b744b).* [Data set]. Natural Resources Canada. https://open.canada.ca/data/en/dataset/9d96e8c9-22fe-4ad2-b5e8-94a6991b744b

Tarboton, D. G. (2005). Terrain Analysis Using Digital Elevation Models Version 5 [Computer Software]. Utah State University, Logan. Retrieved from http://hydrology.usu.edu/taudem/taudem5/downloads.html

Tarboton, D. G., & Ames, D. P. (2004). Advances in the mapping of flow networks from digital elevation data. *Bridging the Gap: Meeting the World's Water and Environmental Resources Challenges - Proceedings of the World Water and Environmental Resources Congress 2001*, *111*(435), 1–10. https://doi.org/10.1061/40569(2001)166

Tavares da Costa, R., Manfreda, S., Luzzi, V., Samela, C., Mazzoli, P., Castellarin, A., & Bagli, S. (2019). A web application for hydrogeomorphic flood hazard mapping. *Environmental Modelling & Software*, *118*, 172-186. https://doi.org/10.1016/j.envsoft.2019.04.010.

Teng, J., Jakeman, A. J., Vaze, J., Croke, B. F. W., Dutta, D., & Kim, S. (2017). Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. *Environmental Modelling and Software*, *90*, 201–216. https://doi.org/10.1016/j.envsoft.2017.01.006

Teng, J., Vaze, J., Kim, S., Dutta, D., Jakeman, A. J., & Croke, B. F. W. (2019). Enhancing the Capability of a Simple, Computationally Efficient, Conceptual Flood Inundation Model in Hydrologically Complex Terrain. *Water Resources Management*, *33*(2), 831–845. https://doi.org/10.1007/s11269-018-2146-7

Tharwat, A. (2018). Classification assesment methods. *Applied Computing and Informatics. https://doi.org/10.1016/j.aci.2018.08.003.*

Thistlethwaite, J., Henstra, D., Brown, C., & Scott, D. (2018). How Flood Experience and Risk Perception Influences Protective Actions and Behaviours among Canadian Homeowners. *Environmental Management*, *61*(2), 197–208. https://doi.org/10.1007/s00267-017-0969-2

Thistlethwaite, J., Henstra, D., Peddle, S., & Scott, D. (2017). *Canadian Voices on Changing Flood Risk: Findings from a National Survey*. Waterloo. Retrieved from https://uwaterloo.ca/climate-centre/sites/ca.climate-centre/files/uploads/files/canadian_voices_on_changing_flood_risk_fnl.pdf

Towe, R., Dean, G., Edwards, L., Nundloll, V., Blair, G., Lamb, R., Hankin, B., & Manson, S. (2020). Rethinking data-driven decision support in flood risk management for a big data age. *J Flood Risk Management*, *e12652*. https://doi.org/10.1111/jfr3.12652.

Tullis, B. P. (2012). *NCHRP Report 734 Hydraulic Loss Coefficients for Culverts.* Transportation Research Board, Washington, D. C. Accessed at:

https://www.nap.edu/read/22673/chapter/1.

Vacondio, R., Rogers, B., Stansby, P., Mignosa, P. (2011). SPH modeling of shallow flow with open boundaries for practical flood simulation. J. Hydraul. Eng. 138 (6), 530e541.

Vacondio, R., Palù, A., Ferrari, A., Mignosa, P., Aureli, F., & Dazzi, S. (2017). A non-uniform efficient grid type for GPU-parallel Shallow Water Equations models. *Environmental Modelling & Software 88*, 119-137

Veale, B., & Cooke, S. (2017). Implementing integrated watershed management: illustrations from the Grand River watershed. *International Journal of Water Resources Development, 33(3),* 375-392

Vojtek, M., & Vojteková, J. (2016). Flood hazard and flood risk assessment at the local spatial scale: a case study. *Geomatics, Natural Hazards and Risk*, *7*(6), 1973–1992. https://doi.org/10.1080/19475705.2016.1166874

Wang, L., & Cheng, Q. (2007). Design and implementation of a web-based spatial decision support system for flood forecasting and flood risk mapping. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 4588- 4591. https://doi.org/10.1109/IGARSS.2007.4423879

Wang, Y., & Yang, X. (2020). A Coupled Hydrologic–Hydraulic Model (XAJ–HiPIMS) for Flood Simulation. *Water, 12,* 1288. https://doi.org/10.3390/w12051288.

Werstuck, C., Coulibaly, P. (2017). Hydrometric network design using dual entropy multi-objective optimization in the Ottawa River Basin. *Hydrology Research, 48(6),* 1639-1651.

Wilby, R. L., & Keenan, R. (2012). Adapting to flood risk under climate change. *Progress in Physical Geography*, *36*(3), 348–378.

Wilson, D., Fleig, A. K., Lawrence, D., Hisdal, H., Pettersson, L. E., & Holmqvist. E. (2011). *A review of NVE's flood frequency estimation procedures.* Norwegian Water Resources and Energy Directorate Report no. 9.

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

Wing, O. E. J., Bates, P. D., Sampson, C. C., Smith, A. M., Johnson, K. A., & Erickson, T. A. (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.

Wing, O. E. J., Smith, A. M., Marston, M. L., Porter, J. R., Amodeo, M. F., Sampson, C. C., and Bates, P. D. (2021). Simulating historical flood events at the continental scale: observational validation of a large-scale hydrodynamic model, Nat. Hazards Earth Syst. Sci., 21, 559–575, https://doi.org/10.5194/nhess-21-559-2021.

Xing, Y., Liang, Q., Wang, G., Ming, X., & Xia, X. (2019). City-scale hydrodynamic modelling of urban flash floods: the issues of scale and resolution. *Natural Hazards, 96,* 473 - 496. https://doi.org/10.1007/s11069-018-3553-z

Zheng, X., Tarboton, D.G., Maidment, D.R., Liu, Y.Y., and Passalacqua, P. (2018). River Channel Geometry and Rating Curve Estimation Using Height above the Nearest Drainage. *Journal of the American Water Resources Association* 54(4): 785– 806. https://doi.org/10.1111/1752-1688.12661.

List of tables:

Table 1. Values of Manning's n

NRCAN LULC Value	NRCAN Description	Manning's n
1	Temperate or sub-polar needleleaf forest	0.16
2	Sub-polar taiga needleleaf forest	0.16
5	Temperate or sub-polar broadleaf deciduous forest	0.16
6	Mixed forest	0.16
8	Temperate or sub-polar shrubland	0.1
10	Temperate or sub-polar grassland	0.035
12	Sub-polar or polar grassland-lichen-moss	0.035
13	Sub-polar or polar barren-lichen-moss	0.03
14	Wetland	0.1
15	Cropland	0.035
16	Barren lands	0.025
17	Urban	0.08
18	Water	0.04

Formatted Table

Table 2. Study Watershed Characteristics

Characteristic	**Grand River Watershed**	**Ottawa River Watershed**
Drainage Area (km^2)	6,800 (Li et al., 2016)	146,000 (Nix, 1987)
Elevation range (masl)	173-535 (Lake Erie Source Protection Region Technical Team, 2008)	430 – 20 (Nix, 1987)
Geologic characteristics	Underlain by groundwater-rich, fractured, porous limestone bedrock; surface geology characterized by glacial till and moraine complexes (Liel et al., 2016)	Incorporates the geological subdivisions St. Lawrence Lowlands, Grenville Province, Superior Province, and Cobalt Plate within the region of the Canadian Shield (Environment and Climate Change Canada, 2019)
Approximate Population size	985,000 (Grand River Conservation Authority, 2014)	> 2,000,000 (Environment and Climate Change Canada, 2019)
Land Use / Land Cover	43% agriculture; 26.92% range-	73% forested (Quebec); 85% mixed

Formatted Table

	grass and pasture; 12% forests; 9.29 % urban areas; 1.8% wetlands (Veale & Cooke, 2017)	and deciduous forest, 15% boreal (middle-south and northern regions, respectively) (Environment and Climate Change Canada, 2019); 6% farmland; <2% developed (Werstuck & Coulibaly, 2017)
Average Annual Precipitation (mm)	800-900 (Kaur et al., 2019)	840 (Werstuck & Coulibaly, 2017)
Temperature	8-10 ° C average annual; moderate-to-cool temperate (Kaur et al., 2019)	21 - -10 °C average daily (Werstuck & Coulibaly, 2017)

Table 3. Simulated Flood Generation – Ottawa River Watershed

Observed Flood Extent Polygon	**Observed Date and Time (UTC)**	**Intersected Hydrometric Station**	**Station Period of Record (years)**	**Index Flood (Q̃, m^3s^{-1})**	**Observed Discharge (m^3s^{-1})**	**Logspline fit observation count**	**Cumulative Probability Value**	**Period (years)**
FloodExtentPolygon_QC_LowerOttawa_20190429_230713.shp	2019/04/29 23:07:13	02KF005	38	3400	5790	1487	0.962	26.5
FloodExtentPolygon_QC_LowerOttawa_20190507_111329.shp	2019/05/07 11:13:29	02KF005	38	3400	5350	1487	0.939	16.52
FloodExtentPolygon_QC_LowerOttawa_20190513_225800.shp	2019/05/13 22:58:00	02KF005	38	3400	5570	1487	0.961	25.96
FloodExtentPolygon_QC_CentralOttawa_20190503_113004.shp	2019/05/03 11:30:04	02KB001	52	258	477	1487	0.977	42.69

Formatted Table

Table 4. Binary Comparison Results

Watershed	**Return Period (years)**	**Number of evaluated subcatchments**	**Median CSI**	**Median MCC**
Grand River	100	71	0.741	0.844
Ottawa River	26.5	17	0.849	0.888
Ottawa River	16.52	21	0.785	0.826
Ottawa River	25.96	22	0.803	0.852
Ottawa River	42.69	7	0.581	0.743

Formatted Table

List of Figures

Figure 1. GIS Input Data – Grand River Watershed (a) and Ottawa River Watershed (b) Topography. The maps are created in ArcGIS with the basemaps provided by © ESRI. The stations that are used later in Figure 5 comparison are labeled in the plot.

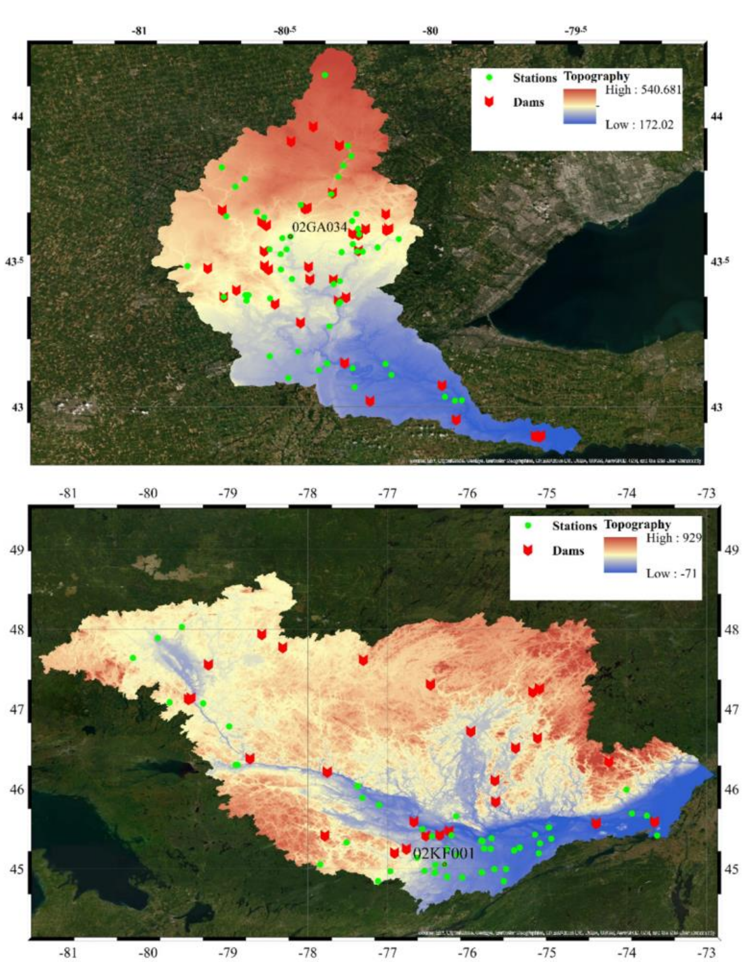


Figure 2. Flood model flowchart illustrating three sub-phases of overall modelling methodology: a) GIS Pre-processing; b) Flood frequency analysis and regional regression; and c) HAND-based solution of Manning's Equation

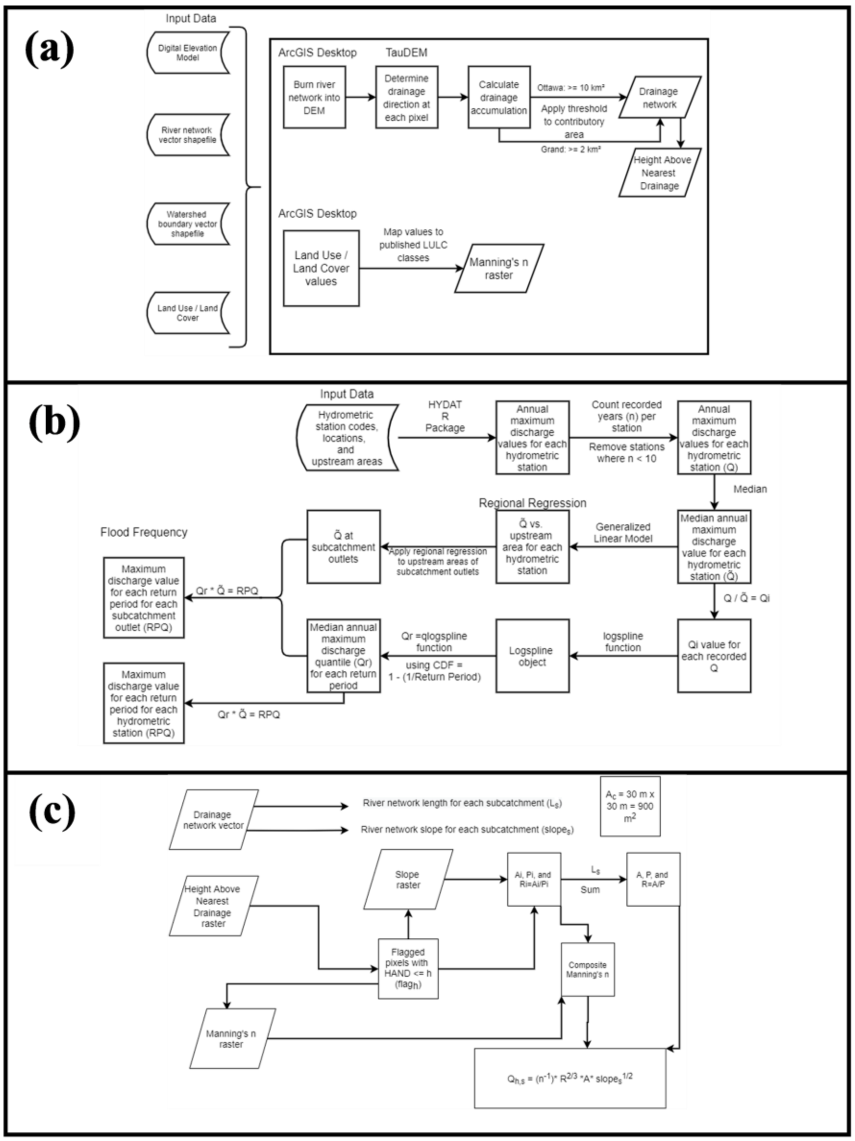
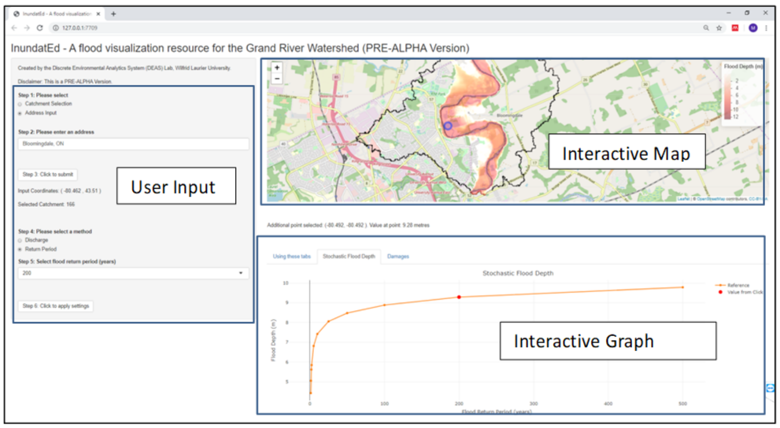
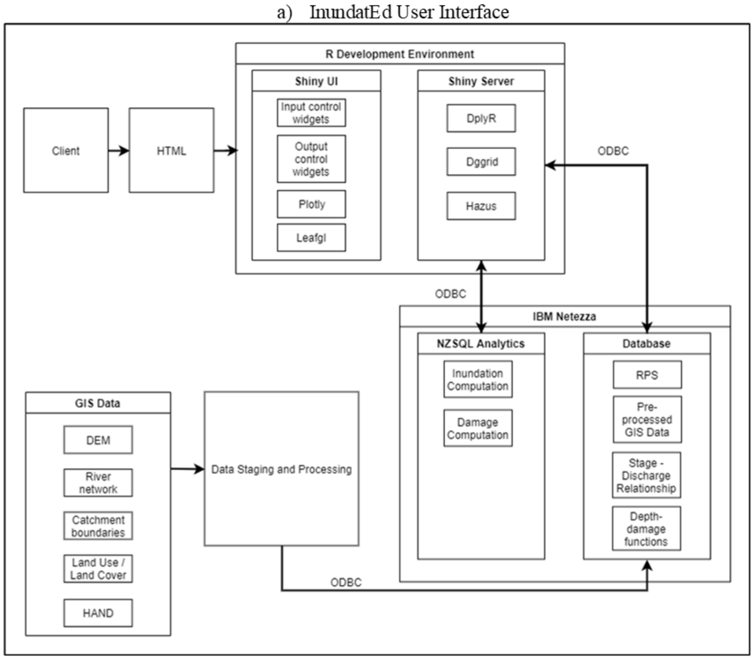


Figure 3. InundatEd User Interface (a) and System Diagram (b). The basemap is created in Leaflet using © OpenStreetMap contributors 2020. Distributed under a Creative Commons BY-SA License



a) InundatEd User Interface



b) InundatEd System Diagram

Figure 4. Flood frequency and regional regression plots

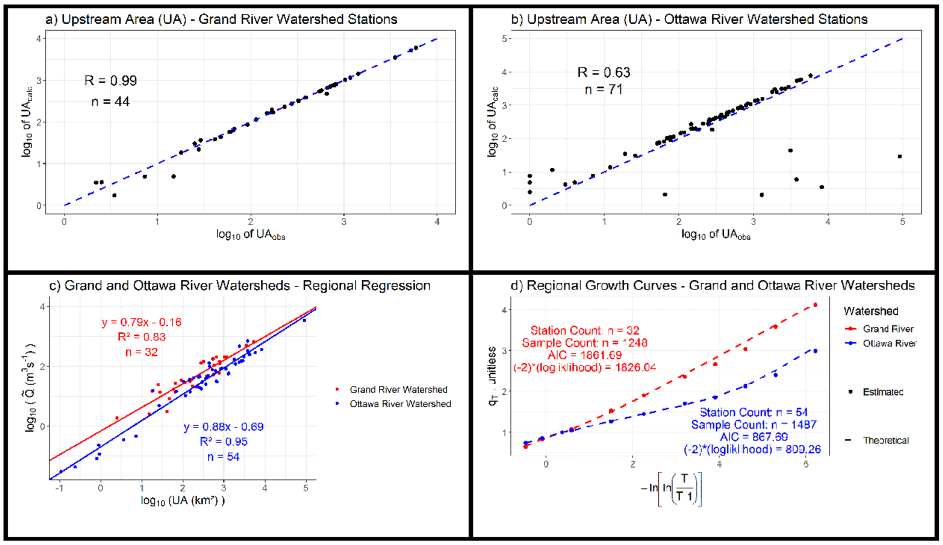
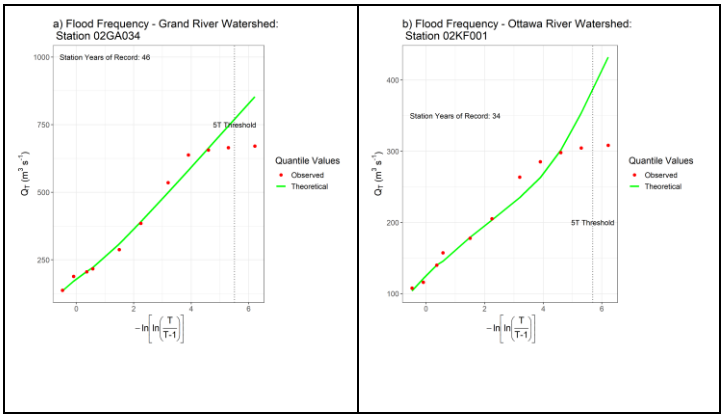


Figure 5. Theoretical Versus Estimated Flood Quantiles



Formatted Table

Figure 6. Binary Classification Results – Grand River Watershed

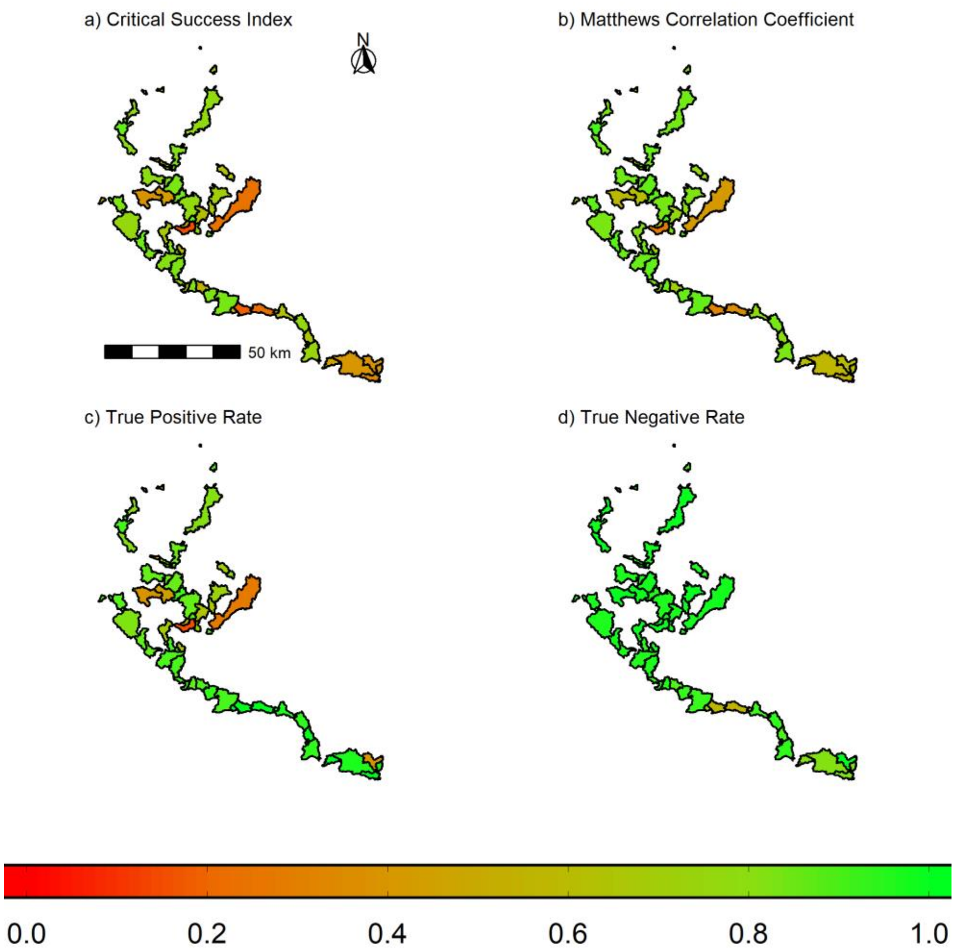


Figure 7. Binary Classification Results – Ottawa River Watershed

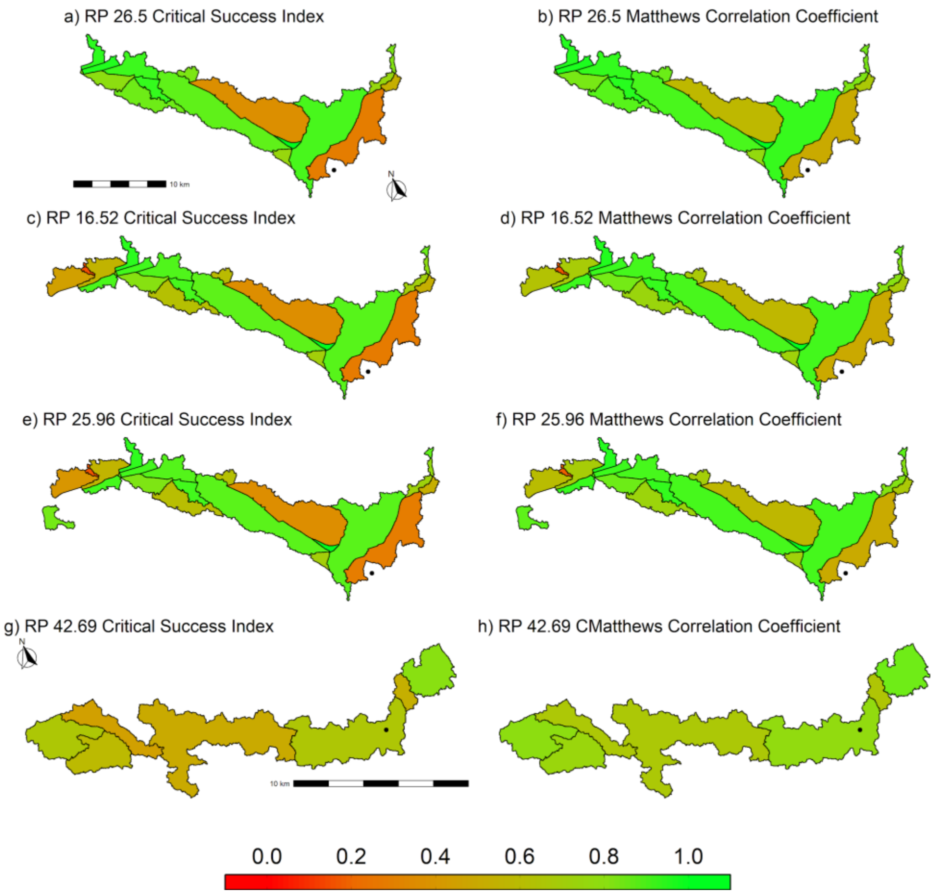
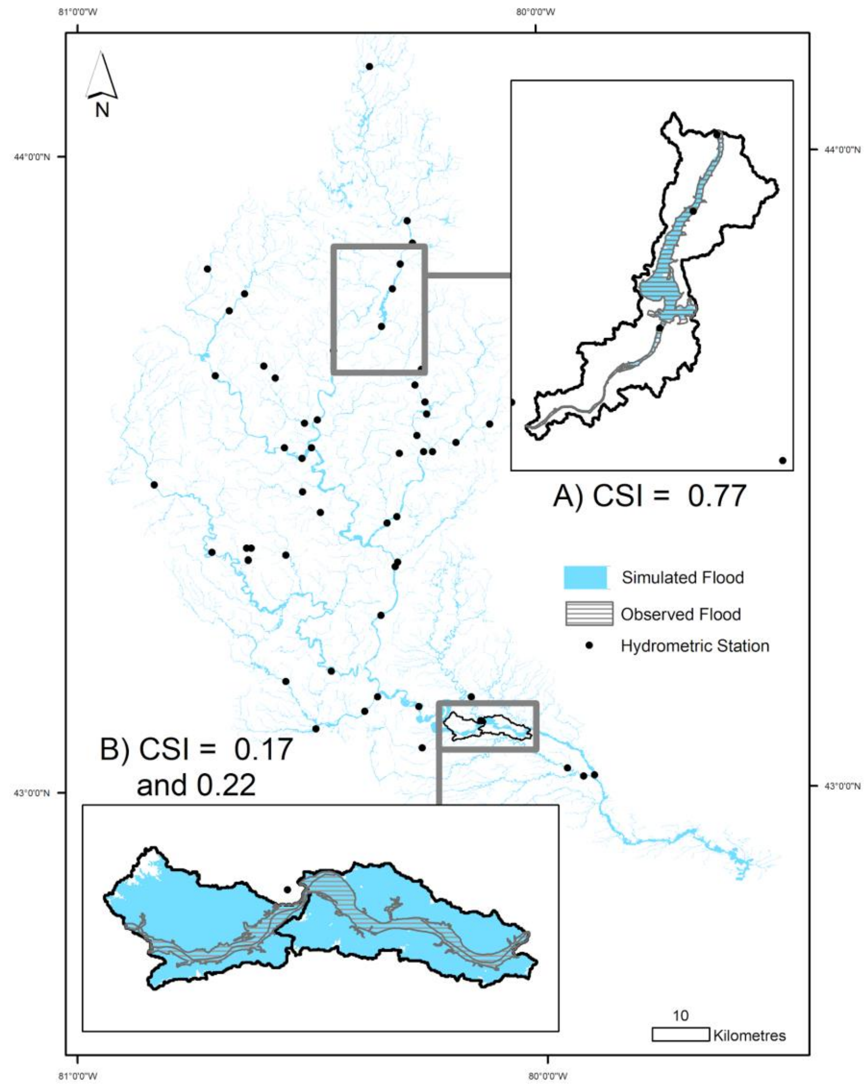
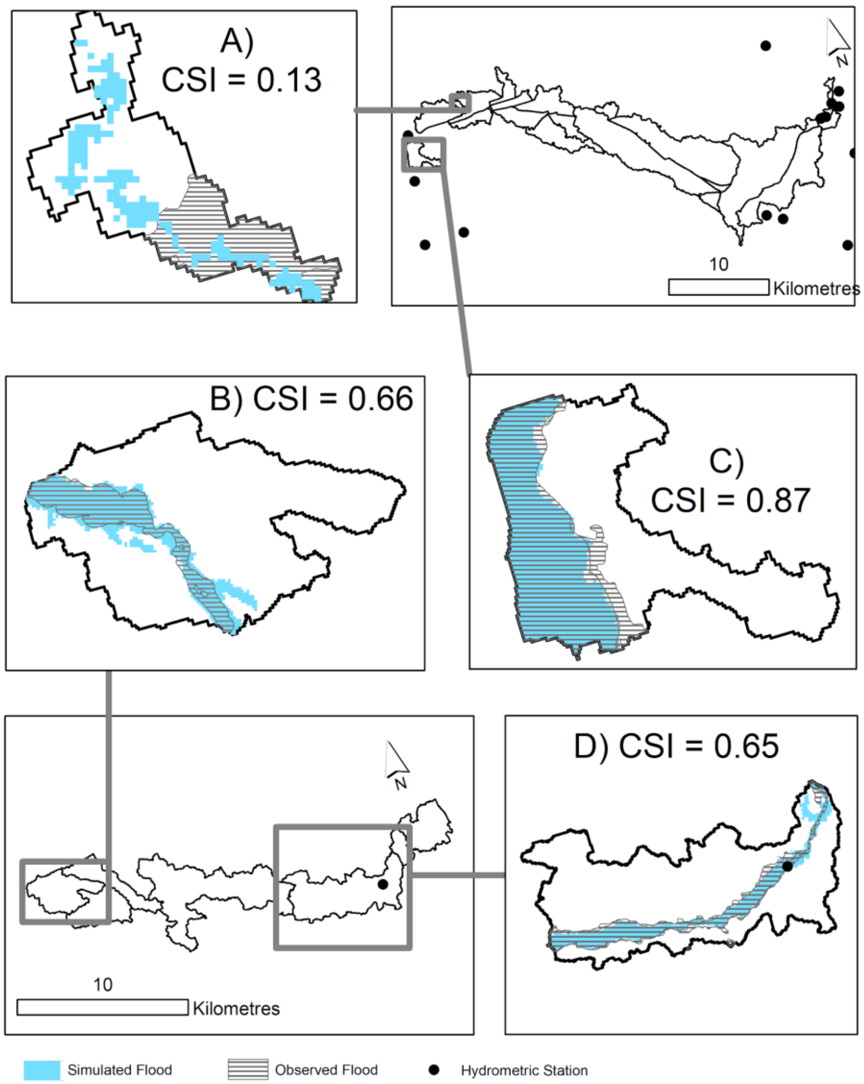


Figure 8. Simulated Flood and Insets – Grand River Watershed 100-Year Return Period



81°0'0"W
80°0'0"W
44°0'0"N
43°0'0"N
N
A) CSI = 0.77
Simulated Flood
Observed Flood
Hydrometric Station
B) CSI = 0.17
and 0.22
10
Kilometres

Figure 9. Observed and Simulated Flood Extents– Ottawa River Watershed



Acknowledgement:

Thank you, Majid Hojati and Amit Kumar, for assistance in GIS and software set up.

The flood extent products are derived from satellite images and ancillary data with a system developed and operated by the Strategic Policy and Innovation Sector of Natural Resources Canada

Funding

This work was funded by the Global Water Futures research programme under the Developing Big Data and Decision Support Systems theme.

Conflicts of interest/Competing interests

The authors declare that there are no competing interests.

Availability of data and material

Any data that support the findings of this study, not already publicly available, are available from the corresponding author, C. Chaudhuri, upon reasonable request.

Author Contribution

The idea behind this research was conceived, implemented, and written equally by all the authors.

Code availability

The current version of InnundatEd is available from the project GitHub website: https://github.com/thespatiallabatLaurier/floodapp_public. The exact version of the model used to produce the results used in this paper is archived on Zenodo (*10.5281/zenodo.4095618*)