



Article **Probabilistic Forecasts of Flood Inundation Maps Using Surrogate Models**

Andre D. L. Zanchetta ^{1,*} and Paulin Coulibaly ^{1,2,3}

- ¹ Department of Civil Engineering, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4L7, Canada
- ² School of Geography and Earth Sciences, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4L7, Canada
- ³ United Nations University Institute for Water, Environment, and Health, Hamilton, ON L8P 0A1, Canada
- * Correspondence: dellalia@mcmaster.ca or adlzanchetta@gmail.com; Tel.: +1-289-489-7247

Abstract: The use of data-driven surrogate models to produce deterministic flood inundation maps in a timely manner has been investigated and proposed as an additional component for flood early warning systems. This study explores the potential of such surrogate models to forecast multiple inundation maps in order to generate probabilistic outputs and assesses the impact of including quantitative precipitation forecasts (QPFs) in the set of predictors. The use of a *k*-fold approach for training an ensemble of flood inundation surrogate models that replicate the behavior of a physics-based hydraulic model is proposed. The models are used to forecast the inundation maps resulting from three out-of-the-dataset intense rainfall events both using and not using QPFs as a predictor, and the outputs are compared against the maps produced by a physics-based hydrodynamic model. The results show that the *k*-fold ensemble approach has the potential to capture the uncertainties related to the process of surrogating a hydrodynamic model. Results also indicate that the inclusion of the QPFs has the potential to increase the sharpness, with the tread-off also increasing the bias of the forecasts issued for lead times longer than 2 h.

Keywords: rapid flood forecasting; flood inundation; flash flood; surrogate model; machine learning; ensemble forecasting

1. Introduction

Changes in land cover related to urbanization and an expected higher frequency and intensity of extreme rainfall events driven by climate change are mechanisms that are assumed to lead to an increase in the occurrence of flash flood events in different cities in the upcoming years, a trend already reported worldwide in the literature [1–3]. To reduce the impact caused by flash floods in terms of property damage and loss of lives, forecasting centers are established and flood early warning systems are implemented to support decision makers with information regarding the potential occurrence, location, and intensity of hazardous inundation conditions [4]. The closure of roads, evacuation of buildings, and interruption of mass transportation vehicles are examples of important preventive actions that can be taken in the imminence of urban floods if a timely and informative warning of an upcoming flooding event is available.

Forecasts produced by early warning systems are usually based on hydrographs issued for specific point locations of an open channel, which are extremely important for identifying scenarios of river overflow. However, the absence of flood inundation maps forecasted in real time can limit the ability of decision makers to take informed actions due to the importance of spatio-temporal data for first responders. Using conventional two-dimensional (2D) and quasi-2D hydraulic models based on physical representation of the water flow is considered the most accurate approach for simulating the development of flood inundations, especially for flashy catchments, given the relevance of momentum in



Citation: Zanchetta, A.D.L.; Coulibaly, P. Probabilistic Forecasts of Flood Inundation Maps Using Surrogate Models. *Geosciences* **2022**, *12*, 426. https://doi.org/10.3390/ geosciences12110426

Academic Editors: Md Jahangir Alam, Monzur Imteaz, Abdallah Shanbleh and Jesus Martinez-Frias

Received: 19 October 2022 Accepted: 18 November 2022 Published: 21 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). the water movement [5]. Such simulations require solving large sets of intercorrelated Saint-Venant equations, which leads to extensive computational demands that limit the real-time execution of the hydraulic models as part of operational flood forecasting chains. Instead, such hydrodynamic models are usually executed "offline", when processing time is not a constraint, and the generated maps are used for the establishment of flood risk maps [6,7], which are valuable resources for the management of catchments. While new hydrodynamic models are being proposed to explore growing sources of processing power, such as graphic processor units and cloud computing [8–10], adopting such new technologies by members of established forecasting centers may be challenging considering the need to migrate already-implemented models and implement potentially demanding structural changes in the data system.

Several workarounds were proposed for making use of the valuable outputs produced by hydraulic models already implemented and validated. Usually, such approaches involve the steps of (1) pre-simulation (offline) of a variety of realistic rainfall-runoff scenarios, (2) identification of empirical relationships between the inputs used in the simulation and the output maps generated, and (3) inexpensive application of such relationships in real time (online) as hydrological observations or forecasts become available. In this context, Bhola et al. [11] and Crotti et al. [12] proposed a database-based approach in which pre-recorded simulated inundation maps can be retrieved through comparisons between their antecedent hydrographs and discharge timeseries forecasted by hydrological models. Despite its efficiency, the approach has limited potential to extrapolate (or interpolate) predictions for scenarios outside the records in the database. Alternatively, different machine learning techniques to surrogate hydraulic models have been explored through a variety of approaches; however, the high dimensionality of 2D inundation maps is a challenging aspect of such methods. Usually, 2D inundation maps are composed of thousands of grid cells to which a water depth value is assigned individually. The higher the number of individual values to be predicted, the higher the complexity of the datadriven model tends to be, which leads to a higher chance of the models trained to be overfit giving a limited dataset [13]. One approach to overcome this high-dimensionality issue is to set up multiple lower-dimensional machine learning models for individual [14] or spatially close 2D cells [15], which has the drawback of requiring the training and maintenance of a potentially high number of independent models, as one model is needed for each flood-prone point or region. Alternatively, the use of hybrid tools in which a clustering model is used to reduce the dimensionality of the maps has been proposed. Chang et al. [16], for example, combined the potential of self-organizing maps (SOMs) to cluster highly dimensional records and nonlinear autoregressive recurrent networks with exogenous inputs (NARX) to successfully generate multi-step regional flood inundation maps. The method was adapted to predict floods in urban areas caused by the overflow of sewer systems [17,18] and by the river overflow in a flashy catchment [19].

While multiple approaches have been proposed for rapidly producing flood inundation maps, to the best of the authors' knowledge, only results for deterministic forecasts were reported in the present literature despite the recognized importance of representing prediction uncertainties [20–22]. In this context, the surrogating of a 2D hydraulic model, as well as any other data-driven technique, has the drawback of having additional sources of uncertainties derived both from the process of abstracting the complex mechanisms of surface water flow and from the finite amount of data used for training.

In this study, surrogate models with hybrid NARX + SOM structures are trained and set up to reproduce the forecasting of ensemble inundation maps in an operational scenario. For the estimation of the uncertainties associated with the model-surrogating step, we propose a *k*-fold ensemble approach for the segmentation of a pre-simulated dataset of flood inundation maps used for model training. The surrogate models are used to produce ensemble forecasts, which are converted into probability distributions. The outputs are assessed both using and neglecting precipitation forecasts issued by numerical weather models for three intense rainfall events observed in the Don River Basin, Toronto, Canada, two of which are further analyzed as study cases. The outputs of the surrogate models are compared and a discussion about (1) the resulting uncertainty range and (2) the impact of including forecasted precipitation in the set of predictors to the model's outputs is presented.

2. Study Area

The Don River Basin, located in the Greater Toronto Area, Ontario, Canada (Figure 1a), has a total area of approximately 350 km², baseflow of approximately 4.5 m³/s, and land cover predominantly characterized by urban infrastructure (Figure 1b). The catchment is managed by the Toronto and Region Conservation Authority (TRCA). The high level of soil imperviousness, the channelization of large portions of the Don River and its tributaries, and a smooth relief result in a scenario of high propensity to flash flood [23], as also observed in other urban catchments surrounding the Great Lakes [24]. The response time of the catchment is in the order of 2.5 to 3 h and the southern region of the catchment recurrently reaches river overbank conditions, which results in significant socio-economic impacts mainly related to the inundation of high-traffic areas [25]. In this context, two points of interest (POIs) are taken into consideration: POI 1 refers to the location in which an urban train became stranded during the historical flood of July 2013, while POI 2 refers to a point on Bayview Avenue usually closed due to floods (Figure 1c). Properly predicting the occurrence/absence and the intensity of inundation scenarios in such locations is of particular interest for decision makers to drive the choice of taking (or not) actions with significant impact to local commuters, such as temporarily stopping train services (for POI 1) and blocking access to affected streets (POI 2).



Figure 1. Representation of the Don River Basin in terms of (**a**) its location, (**b**) its land coverage, and (**c**) its region prone to flash floods. Adapted from [19].

The Don River Basin was selected as the study case as it may be considered a good representative of urban catchment prone to flash flooding, for which one (or more) hydro-

dynamic model(s) is already implemented and calibrated by the agency responsible for its management, but is so computationally expensive that it is not used in real-time applications (further information about the hydrodynamic model is provided in Section 3.1.2). Any other catchment meeting the data requirements and physical model availability and with similar characteristics could have been selected for this study.

3. Materials and Methods

3.1. Materials

3.1.1. Data

The catchment management agency maintains the four rain gauges (HY016, HY021, HY027, and HY036) and the stream gauge (HY019) considered in this study (Figure 1b). The observed timeseries of such gauges are made publicly available at a temporal resolution of 15 min. In this study, the historical data used spans from 2012 to 2020 due to the mutual data availability for the five gauges.

Quantitative precipitation forecasts (QPFs) from the Rapid Refresh system (RAP) [26] are included in the predictors of half of the data-driven models. Among the systems that produce QPF products covering the study area, RAP was selected for this study due to the large archive publicly available, with predictions issued as early as May 2012, and due to the hourly temporal resolution and hourly update rate, which are the closest available for the needs of a flash flood forecasting system [27].

Official point intensity-duration-frequency (IDF) curves are provided to the public by the governmental agency Environment and Climate Change Canada (ECCC) [28] for the Toronto Pearson International Airport, located near the study catchment. Such IDF is used in the design of synthetic storms.

3.1.2. Hydrodynamic Model

The catchment management agency developed a calibrated hydrological model of the Don River Basin in Storm Water Management Model (SWMM) [29] using the software PCSWMM [30]. The hydrological model was originally composed of 462 sub catchments: 2703 conduits (river or channel segments) that represent the water flow unidirectionally and do not count in a 2D hydraulic component to simulate flood inundation maps. This work used the modified version of the model described by Zanchetta and Coulibaly [19], in which a hydraulic flow surface component for the flood-prone area (region of Figure 1c) is included. The spatial resolution of the hydraulic surface flow component is in the order of 2 m, following the granularity of the digital elevation model (DEM) on which it is based and meeting the high degree of spatial discretization required for urban environments [31]. Such a model is hereafter referred to simply as "hydrodynamic model".

3.2. Methodology Overview

This work is organized in three major stages: the setup (offline), the emulation of the operational use (online), and the performance assessment of the surrogate models, as represented in Figure 2 and as further described in Sections 3.3–3.5.



Figure 2. Flowchart of the methodology used in this study.

3.3. Setting Up the Ensemble Surrogate Model System (Offline Stage)

This work follows a modified sequence of steps adopted by Zanchetta and Coulibaly [19] for setting up the surrogate models used. Initially, an extensive set of significant observed and synthetic rainfall events is established. For each such event, the hydrodynamic model is used to simulate the response hydrographs in the main inputs of the inundation area and the resulting inundation maps. A database is constructed in which each instant inundation map is stored with its antecedent simulated conditions. A representative subset of the records in the database is selected as the training/validation dataset, which is split into subsets of equal size using the *k*-fold approach. For each such subset, one hybrid surrogate model system, composed of pairs of one recurrent and one classifier network, is trained.

3.3.1. Establishing a Dataset of Significant Rainfall Events

We considered an observed rainfall event to be significant if the observed discharge captured by the gauge HY019 exceeded 2 times the baseflow (i.e., $9 \text{ m}^3/\text{s}$) within 3 h from a rainfall input recorded by at least one of the rain gauges. In order to capture eventual long-standing precipitation records and most parts of the recession curve, all rainfall and discharge data from 36 h centered around the discharge peak was considered as part of each event. To avoid the potential influence of rain-over-snow and snowmelt, only events occurring in the warm season of the years were considered.

Two sets of synthetic events were included in the dataset. Such augmentation is inspired by the work of Crotti et al. [12], which identified that using datasets composed of a hybrid of synthetical and historical pre-simulated events has the potential to improve the performance of offline 2D models.

The first set of synthetic events is generated by the perturbation of the observed rainfall events identified in the previous step through spatial random redistribution of the timeseries recorded by the rain gauges. Such a set simulates scenarios with the same rainfall intensity, which was observed as having the potential to result in different outcomes if they had different spatial configurations. Considering the small size of the catchment and the mutual proximity between the rain gauges, the rainfall timeseries are expected to be highly correlated during stratiform precipitation events and less correlated during convective events due to the limited coverage area of such types of rainfall.

The second set of synthetic events consists of design storms derived from the local point IDF curve for return periods of 100, 200, and 500 years. For such, the following procedure was adopted: (1) the aerial reduction factor (ARF) was estimated empirically as the conventional ratio between the aerial precipitation observed in the catchment and the

respective maximum-point gauge records for a fixed accumulation interval of 24 h; (2) the accumulation-duration-frequency curve (ADF) was derived from the IDF curve; (3) the mean point accumulated precipitation for each return period was converted into mean

mean point accumulated precipitation for each return period was converted into mean aerial accumulated precipitation and; (4) for each of the converted mean aerial accumulated precipitation values, design storms with the 4 shapes of Huff design storms [32] and based on the alternating blocks method [33] were generated to produce a wide diversity of climatic-based rainfall shapes.

Other more advanced methods for generating synthetic rainfall events, such as the stochastic storm transposition [34], could also have been applied; however, the two abovementioned approaches for generating synthetic rainfall data were selected for being intuitive and of simple implementation.

Figure 3a presents the observation values used to define the ARF. Each 24-h precipitation record with accumulation values higher than 46 mm (i.e., the total accumulated precipitation estimated for rainfalls within a return period of 50 years) is represented as a point. The respective regression line is characterized by a slope (the aerial-point rainfall ratio) of 0.6, which was used as the ARF.



Figure 3. Regression lines related to (**a**) the definition of the point-aerial ratio and (**b**) the 25-h ADF curves, both from reference (solid line) and extrapolated (dashed lines).

For each return period (frequency) f (in years) and total rainfall duration T (in hours), ECCC characterizes the respective point IDF curves by:

$$I(T,f) = A_f * T^{B_f},\tag{1}$$

in which I(T, f) is the mean point rainfall intensity (in mm/hour), and A_f and B_f are site-specific constants. Thus, the associated point ADF curve used to calculate $P_p(T, f)$ is given by:

$$P_p(T,f) = I(T,f) * T = A_f * T^{1+B_f}.$$
(2)

ECCC provides estimations of A_f and B_f for return periods of up to 100 years for the Toronto International Airport. To obtain the same coefficient values for return periods of 200 and 500 years, a linear extrapolation was performed using the values of A_f and B_f available for the longest return periods available, i.e., 25, 50, and 100 years (Table 1). Using Equation 2, the obtained accumulated point rainfall for a duration *T* of 24 h and return periods of 100, 200, and 500 years was 139 mm, 153 mm, and 172 mm, respectively. Considering the reduction factor (0.6), the mean aerial accumulated precipitation for the 100-, 200-, and 500-years return periods was estimated as 83 mm, 92 mm, and 103 mm, respectively (Figure 3b). For each of the three mean aerial accumulated precipitation values, five storm designs were created based on the four Huffs shapes and one based on the alternating blocks method. The 15 resulting pluviograms were used as spatially uniform inputs for simulations in the hydrodynamic model of the catchment.

Coefficient _	Return Period (Years)				
	25	50	100	200	500
A_{f}	41.0	46.0	50.9	55.7	61.9
B_{f}	-0.689	-0.686	-0.684	-0.683	-0.680

Table 1. Coefficient values of the IDF curve estimated by ECCC (regular font) and extrapolated (bold).

3.3.2. Construction of the Simulations Database

The hydrodynamical model was used to simulate the response of the catchment to the rainfall events of the hybrid dataset. All model runs started from a stable baseflow condition, generating both discharge hydrographs at the two main inlet points of the flood-prone area (Q_1 and Q_2 , Figure 1c) and instant inundation maps for the 36 h of each rainfall event at 15-min intervals, resulting in a total of 144 inundation maps per rainfall event. An additional scenario of no-rainfall event was also included for the sake of completeness of the dataset.

For each simulated instant *t*, the resulting inundations map (IM_t), the discharge values simulated for points Q_1 and Q_2 , and the 15-min accumulated mean aerial precipitation $P_{15\text{min}, t}$ were stored in the database. It is worth noting that a complete IM_t stored in the database is composed of water depth values in all 101,577 cells of the 2D space, including points both within and outside the river boundaries. Considering that flood conditions in the considered area are predominantly driven by the overflow of the Don River, which is triggered by intense precipitation in the upstream area, other usually relevant hydrological components—such as evapotranspiration, soil moisture, and drainage/sewer pipeline—were not stored or considered in further steps. The same applies to precipitation occurring over the 2D domain, as such an area represents a small fraction of the total area of the catchment; thus the influence of the runoff generated in this component is assumed to be negligible compared to the runoff routed to Q_1 and Q_2 .

Once all of the inundation maps were generated, the 2D cells in the floodplain domain were classified into three groups. The first group, referred to as "wet cells", is composed of the 2D cells that presented non-zero water depths on the maps produced by the simulation without rainfall forcing (i.e., the 2D cells within the river boundaries). The second group, referred to as "dry cells", is composed of the 2D cells that presented zero water depths during all instances of all simulations (i.e., points that are extremely unlikely to be inundated). The remaining 2D cells, referred to as "inundation cells", represent locations on the land surface that can be potentially flooded during intense rainfall events. The inundation maps considered in further steps are comprised solely of the inundation cells to reduce the overall complexity and computational burden of the machine learning models.

3.3.3. Selection of the Training/Validation and Test Dataset

For each IM_t , an average inundation depth (AID_t) was calculated as the simple mean of the instant water depths of all inundation cells of IM_t . The AID_t is used in this work as a univariate value representing the overall instant magnitude of the inundation process.

The historical rainfall event that produced the IM_t with highest AID_t was considered the most extreme real event in the database and was reserved for testing. Such an event represents a real scenario outside the historical events in the learning space of the surrogate models. Additionally, one event of intermediate magnitude and one event that occurred posteriorly to all historical events were included in the test set so that the performance of the model could include one rainfall event that was not expected to trigger responsive actions and one event temporarily outside the training set.

To reduce the redundancy of the data used for training/validation, a subset of the remaining simulated events was selected using the conventional Computer Aided Design of Experiments (CADEX) sampling method [35]. Given a set *S* in which each of its records is defined by *F* features f_1, f_2, \ldots, f_N , the objective of CADEX is to select a sample *Z* maximizing the heterogeneity of the selected records in the feature space. For such, a

function $\Delta(r_i, r_j)$ is defined to estimate the distance between two records r_i and r_j in the feature space. Initially, the two records of *S* that are the most distant from each other in terms of Δ are selected to compose *Z*. Additional records are iteratively added to *Z* based on the criteria of maximizing the total mutual distance among all members of *Z* until *Z* reaches a size defined a priori. The reader is referred to Kennard and Stone [35] for further details on the method.

In this work, for the application of the CADEX method, each simulated rainfall event r_i is represented as a record with 144 features, each feature being the *AID* of the *y*-th *IM* of r_i (i.e., $AID_{i,y}$). The distance between two rainfall events r_i and r_j is given by:

$$\Delta(r_i, r_j) = \frac{\sum_{y=1}^{144} |AID_{i,y} - AID_{j,y}|}{144}$$
(3)

which can be interpreted as the mean absolute distance between the *AID* timeseries of the two events.

The CADEX method requires the size of the sample to be defined a priori. To evaluate multiple values for *k* in the k-folding implemented in posterior steps, as further discussed in Section 3.3.5, the size of the sample was set to be 36. Other sampling algorithms, such as SELECT [36] and Poisson Disk Sampling [37], adopt stopping criteria that do not ensure a number of elements in the selected subsample defined a priori, thus they were not considered in this study.

Figure 4 presents the timeseries of the *AID* of all 108 simulated rainfall events in the simulations database and of the 36 rainfall events selected using the CADEX method. It is possible to note that, despite being composed of only one-third of the total number of records, the *AID* timeseries of the rainfall events in the training/validation set (Figure 4b) present a variety of forms comparable with the full dataset (Figure 4a). The majority of events that were not included in the training/validation set are characterized by their lower intensity and high recurrency due to their mutual similitude. Conversely, all simulations using design storms, which are designed to have heterogeneous shapes and less recurrent intensities, were included in the training/validation set. A summary of the composition of each set is given in Table 2.



Figure 4. Timeseries of the *AID* of the simulated rainfall events in (**a**) the entire database and (**b**) in the selection using the CADEX algorithm.

Table 2. Number of events in each set of simulations.

		Type of Precipitation		
Set of Simulations	Observed	Disturbed Observation	Design Storm	Total
Full database	31	62	15	108
Training/Validation	5	16	15	36

3.3.4. Establishing the Hyperparameters of the Surrogate Models

Each surrogate model member of the ensemble forecasting system consists of a hybrid structure composed of a NARX and a SOM [38] in a configuration similar to the one adopted by Zanchetta and Coulibaly [19], which demonstrated the suitability of the NARX-SOM approach for surrogating the same hydrodynamic model of this study. It is worth noting that in the aforementioned study only deterministic predictions were generated and assessed, while this work targets the generation and the analysis of probabilistic predictions. In addition, in the previous work the use of QPFs as one of the predictors is not considered, while in this paper the performance of QPF-aware models is compared with the performance of their QPF-absent counterpart.

The SOM component has the objective of reducing the dimensionality of flood inundation maps, which is usually composed of hundreds or thousands of water depth values, one for each cell in the modeled 2D space. For such, an extensive collection of instant flood inundation maps is used to train the SOM, which is a non-supervised clustering method capable of efficiently handling highly dimensional datasets using a rectangular 2D topological space [39]. Before training, the number of topological nodes (in terms of *W* columns, *H* rows in a rectangular topological organization) must be defined as a hyperparameter. After training, the content of each topological node can be interpreted as an inundation map that represents the shared characteristics of the inundation maps assigned to it.

There is not a consensus on how to determine the number of topological nodes of a SOM. In this work, an empirical approach is adopted taking into consideration that a SOM model is valuable as long as it is able to identify patterns shared by different inundation maps (generalization power) without losing the capability to differentiate records distant between each other in the feature space (discretization power). For such, the following algorithm was applied: (1) a SOM with small topological dimension, W = 3 and H = 3, is trained using all of the training/validation dataset; (2) if all of the topological nodes had 2 or more inundation maps associated to them, the trained SOM is considered "valid", the value of W (or H if H < W) is increased by 1, and the algorithm returns to step 1; and (3) the iterations proceed until at least one topological node in the trained SOM is composed of a single record of the training dataset (SOM considered "invalid"). The values of H and W of the last "valid" SOM are then fixed and adopted in further steps. Figure 5 presents how the number of records of each topological node varied with the change of the topological map size. As the SOM with topological dimensions of 05×05 was considered "invalid", it was not included in the plot and the immediate antecedent topological configuration (of 05×04) was selected as the fixed dimensionality for the SOMs trained and used in the subsequent steps.



Figure 5. Distribution of the records in the training/validation dataset assigned to different topological nodes for multiple configurations of SOM dimensions. Diamond markers indicate the median of each distribution.

The NARX model adopted in this work consists of a regular feed-forward neural network with three neuron layers (input, hidden, and output) that predicts the topological node of its associated SOM given limited antecedent and forecasted data.

Each NARX model was trained to perform a prediction at a specific lead time *L*. The total number of input neurons (one per predictor) varied from 7 to 10 (Table 3). All models have, as part of their predictor set, antecedent values of mean aerial quantitative precipitation estimate (*QPE*), simulated inflow discharge at points Q_1 and Q_2 , and antecedent simulated *AID*. The number of hourly accumulated quantitative precipitation forecasted (*QPF*) values ranged from 1 to 4 depending on the *L*, as represented in Figure 6.

Table 3. Listing of all potential NARX predictors.

Predictor	Meaning	On Lead Time L
QPE_L	Mean estimated precipitation, 2-h accumulation	All
$Q_{1,L}$	Earlier inflow discharge at Q_1 , 30-min mean	All
$Q_{1,L-1}$	Later inflow discharge at Q_1 , 30-min mean	All
$Q_{2,L}$	Earlier inflow discharge at Q_2 , 30-min mean	All
$Q_{2,L-1}$	Later inflow discharge at Q_2 , 30-min mean	All
AID_{-1} (or \overline{D}_{-1})	Average antecedent simulated inundated depth, instant	All
QPF_{L+1}	Mean predicted forecast, 1-h accumulation, 1 h ahead	All
QPF_{L+2}	Mean predicted forecast, 1-h accumulation, 2 h ahead	L > 60 min
QPF_{L+3}	Mean predicted forecast, 1-h accumulation, 3 h ahead	L > 120 min
QPF_{L+4}	Mean predicted forecast, 1-h accumulation, 4 h ahead	L > 180 min

(a) Predictors for 15-minutes lead time

(b) Predictors for 90-minutes lead time



Figure 6. Temporal representation of the predictors of the NARX models trained for lead times of (a) 15 min and (b) 90 min assuming an issue time t_0 .

It is worth mentioning that other hydrological forcings usually considered relevant for predicting river discharge were not included in this work due to the particularities of the study catchment and the events selected. Examples of such include the influence of (1) snow, usually represented in the form of snow water equivalent (SWE), which was neglected due to the fact that the rainfall events used were restricted to the warm seasons; (2) temperature, which was expected to have neglectable impact on the time span of the events; and (3) soil moisture, which is expected to have limited impact on the rainfall-runoff process due to the low level of soil permeability driven by intense urbanization.

The number of nodes in the hidden layer was defined empirically, with multiple values ranging from 10 to 50 being tested on the training of each network so that the configuration that presented the highest validation performance (in terms of minimum loss) was selected.

Softmax is used as the activation function in the output layer with a total of 20 (5 × 4) neurons, each output neuron representing one topological node of the associated SOM. As the output in each neuron of a softmax layer provides the probability of such a neuron being the correct one in a classification problem, we consider that the two topological nodes that were assigned with highest probabilities by the NARX model are the best candidates to represent the forecasted *IM*. The *IM* effectively produced by the hybrid model is composed of a weighted average of such pair of best candidates *IMs*. A schematic representation of the application of such a system operationally is provided in Figure 7.



Figure 7. Diagram representing the dataflow of the hypothetical operational setup. Eⁱ indicates the i-th member of the ensemble system, comprehended by a hybrid NARX-SOM model (represented within the dashed rectangle). Blue boxes indicate the changes from Zanchetta and Coulibaly [19]. Number of inputs, neurons, and topological nodes are hypothetical and for illustrative purposes only.

3.3.5. Training the Surrogate Models

The conventional k-fold cross-validation method was applied to train multiple surrogate models. In such an approach, the full training/validation dataset is split into k equally sized subsets. For each subset ("fold"), a model is tuned using all of the other subsets ("fold-in") for training and its own subset ("fold-out") for validation, resulting in k models trained at the end of all iterations. The establishment of multiple surrogate models using a limited dataset can also be performed by having the subsets used for training and validation being selected independently, i.e., without the segmentation in equally sized folds. In this case, each subset is composed of elements of the full dataset selected randomly, thus it is not possible to ensure that the subsets will be composed of a significant number of different elements, which ultimately may result in the undesired scenario of different surrogate models being trained with very similar sets of inputs. The splitting of the dataset in k-folds ensures a minimum overlap between the subsets used for training.

There are multiple approaches for selecting the number k. In our work, for the sake of simplicity, the most empirical approach is applied, i.e., multiple values of k are explored and the one that resulted in the best performance in terms of Continuous Ranked Probability Score (CRPS, described in Section 3.5) is selected. The training/validation dataset was set to have a size of 36 simulated rainfall events so that four values of k (4, 6, 9, 18) could be tested under the condition of equal number of records per fold.

As can be observed in Table 4, the CRPS of the configuration composed of 12 folds has the lowest value for the 4 lead times evaluated, indicating that such a data split leads to the best performance among the explored alternatives and justifies the fixing of k = 12 in the following steps.

Lead Time (h)					
Number of Folds	1	2	3	4	Mean
04	0.026	0.030	0.034	0.032	0.030
06	0.033	0.029	0.029	0.029	0.030
09	0.026	0.034	0.042	0.049	0.038
12	0.021	0.023	0.026	0.029	0.024
18	0.021	0.027	0.031	0.032	0.028

Table 4. CRPS of the ensemble surrogate models for different cross-fold ensemble configurations in the training/validation dataset. The lowest (best) value of each column is highlighted in bold.

3.4. Forecasting the Probabilistic Inundation Maps

3.4.1. Generating Ensemble Forecasts

The ensemble of trained surrogate models was used to forecast the three events in the test dataset. RAP precipitation forecasts, bias-corrected through quantile mapping against gauge records, were used as QPF values. Thus, 16 sets of *k* flood inundations maps were produced, one for each lead time distant 15 min apart from 15 min to 4 h in the future. The water depth value predicted by the *i*-th ensemble member for an inundation cell *c* at a time *t* for a lead time *L* is hereafter denoted as $D_{c,t,L}^{i}$.

3.4.2. Converting Ensembles into Probabilistic Forecasts

In probabilistic forecasts, values are provided in the form of a probability distribution rather than a univariate numeric value. In this work, the predicted probability distributions of the water depth for an inundation cell *c* for a time *t* at a lead time *L*, $\mathbb{P}(D_{c,t,L})$, is defined by nine values $\tau_{c,t,L}^{0.1}$, $\tau_{c,t,L}^{0.2}$, ..., $\tau_{c,t,L}^{0.9}$ in which $\tau_{c,t,L}^{i}$ indicates the *i*-th quantile value in the distribution. In this work, for the sake of simplification, the model ensemble members are assumed to be equally likely to issue the correct forecast, and the quantile estimation from an ensemble of predicted values is performed by simple linear interpolation.

3.5. Evaluation

In the absence of observed flood inundation maps, the probabilistic flood inundation maps produced by the hybrid surrogate models for the three events in the test set were compared against the maps produced by the hydrodynamic model. Thus, what is evaluated in this work is the ability of the surrogate model to properly reproduce the behavior of the hydrodynamic model in a significantly reduced time interval and capture the additional uncertainties resulting from the surrogating process. The mean CRPS is used to evaluate the overall goodness of fit of the surrogate models. Assume a simulated inundation map representing the instant *t* and composed of *C* deterministic water depth values $D_{1,t}$, $D_{2,t}$, ..., $D_{C,t}$, with *C* being the total number of inundation cells. For a probabilistic forecast map issued for *t* at a lead time *L* and consisting of *C* random variables $D'_{1,t,L}$, $D'_{2,t,L'}$, ..., $D'_{C,t,L}$, the mean CRPS is calculated as:

$$\overline{CRPS}_{t,L} = \frac{1}{C} \sum_{c=1}^{C} \int_{x=-\infty}^{x=\infty} \left(Prob. \left(D'_{C,t,L} \le x \right) - H(D_{c,t},x) \right)^2 dx \tag{4}$$

in which *H* is the Heaviside step function, i.e.,:

$$H(D_{c,t} \le x) = \begin{cases} 0 & if \ D_{c,t} > x, \\ 1 & otherwise. \end{cases}$$
(5)

 $\overline{CRPS}_{t,L}$ values range from 0 (perfect fit) to ∞ , unitless. In this work, CRPS is applied in two sets of data. The first set consists of every inundation cell *c* at every instant time *t*, regardless of the values of $D_{c,t}$ and $D'_{1,t,L}$. The second set consists only of the pairs of *c* and *t* in which $D_{c,t} > D_{threshold}$, i.e., only when a local inundation was effectively present in the

simulation. The constant $D_{threshold}$ represents the minimum water depth for an inundation cell to be considered "inundated" (or "wet"), fixed as 0.01 m in this work.

The accuracy of the probabilistic model to predict the condition of an inundation cell in terms of dry/wet is measured using the mean Brier Score (BS). The BS is similar to the CRPS, with the difference that only one value of x ($D_{threshold}$ in this work) is evaluated, i.e.,:

$$\overline{BS}_{t,L} = \frac{1}{C} \sum_{c=1}^{C} \left(Prob. \left(D'_{c,t,L} \le D_{threshold} \right) - H(D_{c,t}, D_{threshold}) \right)^2$$
(6)

with values ranging from 0 (perfect accuracy) to 1 (null accuracy), unitless.

The reliability of the forecasts is estimated based on the containing ratio (CR_{α}) [40], which is defined as the percentage of times that observed values fall within specific predicted bounds α . If α represents a confidence interval, the closer the value of CR_{α} is to α , the more reliable the predictor is considered. In this work, as the lower and higher predicted quantiles are 0.1 ($\tau^{0.1}$) and 0.9 ($\tau^{0.9}$), respectively, the bandwidth of the 80% confidence interval (CR_{80}) is used. Thus, given N water depth records calculated by the hydraulic model, N_h of which have values between the $\tau^{0.1}$ and $\tau^{0.9}$ quantiles predicted by a hybrid surrogate model, $CR_{80,t,L}$ is given by a percent value as:

$$\overline{CR}_{80,t,L} = \frac{N_h}{N} \times 100\%, \tag{7}$$

in which the closer $CR_{80,t,L}$ is to 80%, the more reliable the forecast is considered.

Average Bandwidth [40] is used to estimate the sharpness of a prediction. Similar to the CR_{80} , in this work the bandwidth of the 80% confidence interval (B_{80}) is used. Given the quantiles $\tau_{c,t,L}^{0.9} - \tau_{c,t,L}^{0.1}$ forecasted for an inundation cell *c* at instant *t* issued at a lead time *L*, $B_{80,c,t,L}$ is given by:

$$B_{80,c,t,L} = \tau_{c,t,L}^{0.9} - \tau_{c,t,L}^{0.1}.$$
(8)

The value of $B_{80,c,t,L}$ is always non-negative, and the higher the value of the bandwidth, the lower is the sharpness of the prediction. As B_{80} values are in the same unit as the analyzed variable and the water depths values associated with each inundation cell have different magnitudes, this metric is assessed pointwise.

Two metrics are considered for bias. For the general case, in which all records are considered, the Mean Fractional Bias (MFB) is used [41]. It is given by:

$$\overline{MFB}_{t,L} = \frac{1}{C} \sum_{c=1}^{C} \frac{\tau_{c,t,L}^{0.5} - D_{c,t}}{\tau_{c,t,L}^{0.5} + D_{c,t}}$$
(9)

and has unitless values bounded by +2 (biased high) and -2 (biased low), with a value of zero meaning a perfectly unbiased model. MFB is used in this work for the general case due to the fact that the evaluated variable (surface water depth in individual inundation cells) recurrently has value zero, which would result in divisions by zero if other more conventional metric biases were used. Additionally, it is applied in a specific case metric, named event Peak Bias (PB), which is calculated by:

$$PB_{c,E,L} = \frac{\max\left(\tau_{c,E,L}^{0.5}\right) - \max(D_{c,E})}{\max(D_{c,E})} * 100\%$$
(10)

in which, for a cell *c* during an event *E*, $\max(D_{c,E})$ is the maximum water depth value simulated by the hydrodynamic model and $\max(\tau_{c,E,L}^{0.5})$ is the respective maximum median value forecasted at a lead time *L*.

4. Results and Discussion

To facilitate visual interpretation, only timeseries and inundation maps of forecasts issued for lead times of 1, 2, 3, and 4 h are presented, despite results being available at a 15-min time step. For the sake of simplicity, the surrogate models that do not include QPF values in their set of predictors are referred to as "no-QPF" or "No QPF", while the models that have RAP QPF values as part of their predictors are referred to as "QPF-aware" or "RAP QPF" hereafter.

4.1. Overall Performance

When all data is considered, the forecasts produced by the no-QPF models tend to have a better goodness of fit, lower bias, higher reliability, and higher accuracy than their QPF-aware counterparts for most of the lead times (Figure 8a,c,e,f, respectively). This result contradicts the initial expectations that including QPFs would improve the performance of the surrogate models at longer lead times. Such a decay in performance is driven by the presence of additional inputs of precipitation from the QPF products that are not present in the QPE, which leads to the prediction of false inundation points (further illustrated in Section 4.2). Interestingly, for the instants when an inundation is present in the simulation, outputs from the QPF-aware models presented an overall better fit to the simulations, especially for the lead time of four (Figure 8b). Such a gain in performance for the longer lead time is probably due to the increase in confidence (lower $\overline{B_{80}}$) that is derived from the additional information present in the QPF products (Figure 9). Another relevant difference is that the peak water depths predicted for each rainfall event are significantly less biased in the outputs of the QPF-aware model then in its no-QPF counterpart (Figure 8d).



Figure 8. General performance metrics of the no-QPF and the QPF-aware surrogate models.





For all metrics, there is an overall trend of loss in performance with the increase of the forecast lead time. Such a decay is expected, given that the longer the time interval between the last observed data and forecasted map, the lower the amount of information potentially relevant to the predictor.

4.2. Study Cases

The performance of the surrogate models was assessed at the POIs and their surrounding areas in two events of the test set. The event of 8 July 2013, is the most extreme of the observed dataset, and the resulting flood caused the stranding of an urban train carrying passengers in the POI 1 and the stranding of several cars in the POI 2 [42]. The second event, of 2 August 2020, raised local flood warning alerts and lead to the closure of the road at the POI 2; however, no interruption of the urban train services was noticed [43,44].

4.2.1. 8 July 2013

The hydrographs in Figures 10 and 11 show the predicted water depths at POIs 1 and 2, respectively, for the event of 8 July 2013, issued at different lead times. In both figures, it is possible to note a first peak in the predictions issued by the QPF-aware models for longer lead times. Such first peaks are not reproduced by the hydrodynamic model and are absent in the no-QPF forecasts, likely being driven by the higher bias and higher overall number of errors with respect to longer lead times already reported in RAP products [45]. As lead time decreases, the over-forecasted first peak also decreases and the similarity between the effective peak and the predictions produced by the QPF-aware products also increases.

An additional difference between the QPF-aware and no-QPF scenarios is that the inclusion of QPF reduced the spread of the ensemble, which indicates that the additional information increases the confidence of the forecasts. Such a decrease of the ensemble spread is more pronounced for longer lead times and illustrates the general metrics obtained for the bandwidth of the forecasts (Section 4.1, Figure 9).

For both no-QPF and QPF-aware scenarios, the overall shape of the main water depth curve simulated by the hydrodynamic model is within, or very close to, the boundaries of the 80% confidence interval of the ensemble forecasts, which indicates an appropriate representation of the uncertainties originated from the surrogating of the hydrodynamic model. As observed in the overall performance analysis (Section 4.1), major disagreements are observed in longer lead times, which, as indicated by the reliability measurements, can be related to an overconfidence of the models (Figure 8e).

The overall shape of the probability of exceedance maps produced by both the no-QPF and the QPF-aware scenarios is similar to the simulated water depth exceedance map, considering the maximum depth at each location as threshold (Figures 12 and 13). While some overestimation is observed at lead times up to 2 h in both cases, such overestimation is also present in the forecasts for 3 and 4 h in the future when the surrogate model is based solely on QPEs.



Figure 10. Water depth simulated and ensemble forecasts for POI 1 at lead times of (**a**,**b**) 1 h, (**c**,**d**) 2 h, (**e**,**f**) 3 h and (**g**,**h**) 4 h for the event of 8 July 2013.



Figure 11. (**a**–**h**) Same as Figure 10, but for POI 2.



Figure 12. Inundation maps forecasted by the surrogate models without QPF (**a**–**d**), with QPF (**e**–**h**), and simulated by the hydrodynamic model (**i**) at the peak water level of the event of 8 July 2013, at POI 1 (red circle in the maps).



Figure 13. (a–i) Same as Figure 12, but for the POI 2 (red circle in the maps).

4.2.2. 2 August 2020

The forecasting of this event shares many similarities with the forecasts issued for the event of 8 July 2013. The no-QPF surrogate model produced higher peaks than its QPF-aware counterpart for longer lead times; however, the inclusion of QPF products resulted in preliminary forecasted peaks that are not produced by the hydrodynamic model in both the POI 1 (Figure 14) and POI 2 (Figure 15). Conversely, for earlier lead times, the shape of



the QPF-aware ensemble timeseries resembles more the output from the simulation than the no-QPF counterpart, indicating a gain in performance for less intense events.

Figure 14. Water depth simulated and ensemble forecasts for POI 1 at lead times of (**a**,**b**) 1 h, (**c**,**d**) 2 h, (**e**,**f**) 3 h and (**g**,**h**) 4 h for the event of 2 August 2020.



Figure 15. (**a**–**h**) Same as Figure 14, but for POI 2.

The inundation maps forecasted by both surrogate models are also characterized by overestimating the flooded area at shorter lead times (Figures 16 and 17). The overall

shape of the simulated and forecasted maps is comparable for the POI 1. However, it is possible to note that, regardless of the inclusion of the QPF in the feature set, the maps incorrectly forecasted a relatively large area as flooded in the south-west of the POI 2 (lower-left corner of the maps in Figure 17). Such an area has a significantly lower elevation compared to its surroundings in the DEM, which results in recurrent retention of inundated water in the form of a "pound" for long periods of time. Thus, a significant number of the inundation maps that compose the simulations database represent this region as inundated, which probably leads the data-driven model to overestimate the water depths for this area. Such an overestimation is not observed for the inundation cells related to traffic surfaces, however.



Figure 16. Inundation maps forecasted by the surrogate models without QPF (**a**–**d**), with QPF (**e**–**h**), and simulated by the hydrodynamic model (**i**) at the peak water level of the 2 August 2020, event at POI 1 (red circle in the maps).



Figure 17. (a-i) Same as Figure 16, but for POI 2 (red circle in the maps).

4.3. Discussion Summary

The overall spread of ensemble forecasts is significantly low (overconfidence) due to the fact that all ensemble members are trained to mimic the behavior of the same hydrodynamic model and from the same set of predictors; moreover, they all share similar network structures. This can be interpreted as a limitation of the single-model *k*-fold ensemble approach, in which the difference between the ensemble members lies solely in the configuration of the subsets used for their training and validation.

From an operational perspective, the inclusion of QPF products does not significantly impact the performance of the surrogate models for predictions for lead times up to two hours. For longer lead times, however, outputs from the QPF-aware setup tend to produce early false inundations, which may lead to the undesirable issue of false alerts and to the adoption of unnecessary preventive actions. Yet, the maximum event water depths predicted by the QPF-aware models tend to be closer to the peak simulated. The peak inundation may be considered the most important variable for decision makers as it represents the total extent of an inundation at locations that deserve specific actions in the upcoming hours. Thus, forecasting centers may consider that the benefit of improving the prediction of such a variable overcomes the drawback of potential false early inundations associated with the adoption of QPF-aware surrogate models.

The *k*-fold ensemble approach is intuitive, easily implemented, and model-agnostic, and it showed acceptable performance on estimating the uncertainties associated with the process of surrogating 2D inundation models; thus, it can be taken as a benchmark for future research in this field.

4.4. Runtime

Both deterministic simulations using the hydrodynamic model and ensemble forecasts from surrogate models were generated using a desktop computer with 64 GB randomaccess memory (RAM) and a CPU Intel I9 with 3.6 GHz, eight codes, and 16 logical processors. While the runtime of the hydrodynamic model demanded approximately 4 h and 30 min to produce 4 h of inundation maps, the ensemble of forecasts for the same time interval required between 13 to 17 min to be generated, which may be considered applicable for real-time setups.

5. Conclusions, Limitations, and Future Works

The present work evaluates the applicability of the NARX-SOM hybrid surrogate models for forecasting probabilistic flood inundation maps at a flashy catchment in the region of the Great Lakes, as well as analyzing the performance impact of including RAP QPF as a predictor. A *k*-fold approach is used to produce ensemble models that are trained to surrogate an SWMM-based hydrodynamic model in a forecasting setup. The forecasted maps are compared with the simulated maps to assess the efficiency of the surrogate models on rapidly reproducing the hydrodynamic model outputs.

For the most part of the simulated timeseries, the outputs produced by the hydrodynamic model were within, or close to, the 80% confidence interval of the forecasts produced by the surrogate models, indicating that the use of the *k*-fold ensemble was successful in capturing the additional uncertainties of the surrogating step. The inclusion of QPF products did not significantly impact the maps forecasted for lead times up to 2 h. For longer lead times, the no-QPF models tend to produce forecasted peaks biased high and with high spread. Conversely, the inclusion of QPF results in less biased peaks with the tread-off of producing more peaks that were not present in the hydrodynamic simulations, which could trigger false alarms during operational time. Such findings suggest that a forecasting system composed of a combination of no-QPF and QPF-aware surrogate models has the potential to produce more accurate and less biased forecasts for longer lead times; however, exploring strategies for such combination is beyond the scope of this study.

A limitation identified for the *k*-fold ensemble approach is that, by using a single hydrodynamic model as reference and a single approach for model surrogating (SOM-NARX

hybrid structures), the forecasts were characterized by overconfidence (low spread), which limits the potential gains in performance of a post-processing step based on dynamic model weighting, for example. Such an observation motivates the use of multi-model ensemble forecasts; however, the availability of multiple hydrodynamic models for the same floodprone area may be uncommon in forecasting centers given the highly demanding tasks of producing and maintaining each individual model and keeping them updated to reflect changes in the land cover. Alternatively, surrogate models with different structures can be explored to compose the multi-model ensemble. Other recurrent algorithms commonly applied for river flow forecasting, such as the Gated Recurrent Unit (GRU) [46] and the Long Short-Term Memory (LSTM) [47–50], are suggested alternatives for the NARX structure adopted in this work, while convolutional neural networks can be used to compose the flood inundation maps [51]. Additionally, Mosavi et al. [52] (Section 4.2 of their work) listed a series of hybrid models already explored for short-term forecasting that potentially can be adapted for the prediction of flood inundation maps. How the use of alternative structures in the hybrid model impacts the model outputs is beyond the scope of this study and suggested for future work. The results presented are specific for the Don River Basin and for the data products utilized, and the evaluation of this approach at a broader scope is suggested as future research.

Author Contributions: Conceptualization, implementation, and writing—original draft preparation, A.D.L.Z.; writing—review and editing, supervision, project administration, funding acquisition, P.C.; all authors contributed equally to the discussion over the results and contributed to the final manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Natural Science and Engineering Research Council (NSERC) of Canada, grant NSERC Canadian FloodNet (NETGP-451456).

Data Availability Statement: The observed data from the rain gauges and the stream gauge used in this work can be found at TRCA's Open Data Portal (https://data.trca.ca/dataset, accessed on 18 October 2022). Rainfall forecast data from RAP can be found at: NOAA's Archive Information Request System (https://www.ncei.noaa.gov/has/HAS.DsSelect, accessed on 18 October 2022). The hydrological-hydraulic model used in this work is not publicly available as it is a property of the Toronto and Region Conservation Authority (TRCA).

Acknowledgments: The authors would like to thank the Toronto and Region Conservation Authority (TRCA) for providing the original hydrological model and the data used in this work, and Computational Hydraulics International (CHI) for providing an academic license for PCSWMM software for this study.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

Acronym	Meaning
2D	Two-dimensional
ADF	Accumulation-Duration-Frequency
AID	Average Inundation Depth
ARF	Aerial Reduction Factor
B_{α}	Bandwidth of α confidence interval
BS	Brier Score
CADEX	Computer Aided Design of Experiments
CR_{α}	Containing Ratio of α confidence interval
CRPS	Continuous Ranked Probability Score
DEM	Digital Elevation Model
ECCC	Environment and Climate Change Canada
GRU	Gated Recurrent Unit
IDF	Intensity-Duration-Frequency
IM	Inundations Map

LSTM	Long Short-Term Memory
MFB	Mean Fractional Bias
NARX	Nonlinear Autoregressive Recurrent Networks with eXogenous inputs
POI	Points of Interest
QPE	Quantitative Precipitation Estimate
QPF	Quantitative Precipitation Forecasts
PB	Peak Bias
RAM	Random-Access Memory
RAP	Rapid Refresh system
SOM	Self-Organizing Maps
SWE	Snow Water Equivalent
SWMM	Storm Water Management Model
TRCA	Toronto and Region Conservation Authority

References

- 1. Fofana, M.; Adounkpe, J.; Larbi, I.; Hounkpe, J.; Djan'na Koubodana, H.; Toure, A.; Bokar, H.; Dotse, S.Q.; Limantol, A.M. Urban Flash Flood and Extreme Rainfall Events Trend Analysis in Bamako, Mali. *Environ. Chall.* **2022**, *6*, 100449. [CrossRef]
- Sofia, G.; Roder, G.; Dalla Fontana, G.; Tarolli, P. Flood Dynamics in Urbanised Landscapes: 100 Years of Climate and Humans' Interaction. Sci. Rep. 2017, 7, 40527. [CrossRef] [PubMed]
- Yang, L.; Smith, J.A.; Wright, D.B.; Baeck, M.L.; Villarini, G.; Tian, F.; Hu, H. Urbanization and Climate Change: An Examination of Nonstationarities in Urban Flooding. J. Hydrometeorol. 2013, 14, 1791–1809. [CrossRef]
- Zanchetta, A.D.L.; Coulibaly, P. Recent Advances in Real-Time Pluvial Flash Flood Forecasting. *Water* 2020, *12*, 570. [CrossRef]
 Teng, J.; Jakeman, A.J.; Vaze, J.; Croke, B.F.W.; Dutta, D.; Kim, S. Flood Inundation Modelling: A Review of Methods, Recent
- Advances and Uncertainty Analysis. Environ. Model. Softw. 2017, 90, 201–216. [CrossRef]
- 6. Kaya, C.M.; Tayfur, G.; Gungor, O. Predicting Flood Plain Inundation for Natural Channels Having No Upstream Gauged Stations. *J. Water Clim. Change* **2019**, *10*, 360–372. [CrossRef]
- Zarzar, C.M.; Hosseiny, H.; Siddique, R.; Gomez, M.; Smith, V.; Mejia, A.; Dyer, J. A Hydraulic MultiModel Ensemble Framework for Visualizing Flood Inundation Uncertainty. J. Am. Water Resour. Assoc. 2018, 54, 807–819. [CrossRef]
- Aureli, F.; Prost, F.; Vacondio, R.; Dazzi, S.; Ferrari, A. A GPU-Accelerated Shallow-Water Scheme for Surface Runoff Simulations. Water 2020, 12, 637. [CrossRef]
- 9. Ming, X.; Liang, Q.; Xia, X.; Li, D.; Fowler, H.J. Real-Time Flood Forecasting Based on a High-Performance 2-D Hydrodynamic Model and Numerical Weather Predictions. *Water Resour. Res.* **2020**, *56*, e2019WR025583. [CrossRef]
- Morsy, M.M.; Goodall, J.L.; O'Neil, G.L.; Sadler, J.M.; Voce, D.; Hassan, G.; Huxley, C. A Cloud-Based Flood Warning System for Forecasting Impacts to Transportation Infrastructure Systems. *Environ. Model. Softw.* 2018, 107, 231–244. [CrossRef]
- 11. Bhola, P.K.; Leandro, J.; Disse, M. Framework for Offline Flood Inundation Forecasts for Two-Dimensional Hydrodynamic Models. *Geosciences* 2018, *8*, 346. [CrossRef]
- 12. Crotti, G.; Leandro, J.; Bhola, P.K. A 2D Real-Time Flood Forecast Framework Based on a Hybrid Historical and Synthetic Runoff Database. *Water* **2020**, *12*, 114. [CrossRef]
- 13. Ying, X. An Overview of Overfitting and Its Solutions. J. Phys. Conf. Ser. 2019, 1168, 022022. [CrossRef]
- Bermúdez, M.; Cea, L.; Puertas, J. A Rapid Flood Inundation Model for Hazard Mapping Based on Least Squares Support Vector Machine Regression. J. Flood Risk Manag. 2019, 12, 1–14. [CrossRef]
- Berkhahn, S.; Fuchs, L.; Neuweiler, I. An Ensemble Neural Network Model for Real-Time Prediction of Urban Floods. *J. Hydrol.* 2019, 575, 743–754. [CrossRef]
- Chang, L.-C.; Amin, M.; Yang, S.-N.; Chang, F.-J. Building ANN-Based Regional Multi-Step-Ahead Flood Inundation Forecast Models. *Water* 2018, 10, 1283. [CrossRef]
- Kim, H.I.; Keum, H.J.; Han, K.Y. Real-Time Urban Inundation Prediction Combining Hydraulic and Probabilistic Methods. *Water* 2019, 11, 293. [CrossRef]
- 18. Kim, H.I.; Han, K.Y. Data-Driven Approach for the Rapid Simulation of Urban Flood Prediction. *KSCE J. Civ. Eng.* **2020**, *24*, 1932–1943. [CrossRef]
- 19. Zanchetta, A.D.L.; Coulibaly, P. Hybrid Surrogate Model for Timely Prediction of Flash Flood Inundation Maps Caused by Rapid River Overflow. *Forecasting* **2022**, *4*, 126–148. [CrossRef]
- 20. Bales, J.D.; Wagner, C.R. Sources of Uncertainty in Flood Inundation Maps. J. Flood Risk Manag. 2009, 2, 139–147. [CrossRef]
- 21. Zahmatkesh, Z.; Han, S.; Coulibaly, P. Understanding Uncertainty in Probabilistic Floodplain Mapping in the Time of Climate Change. *Water* **2021**, *13*, 1248. [CrossRef]
- 22. Brandt, S.A. Modeling and Visualizing Uncertainties of Flood Boundary Delineation: Algorithm for Slope and DEM Resolution Dependencies of 1D Hydraulic Models. *Stoch. Environ. Res. Risk Assess* **2016**, *30*, 1677–1690. [CrossRef]
- Rincón, D.; Khan, U.T.; Armenakis, C. Flood Risk Mapping Using GIS and Multi-Criteria Analysis: A Greater Toronto Area Case Study. *Geosciences* 2018, 8, 275. [CrossRef]

- 24. Krajewski, M.; Brown, D.; Gibbons, E. Flash Flooding, Stormwater, and Decision Making Flash Flooding, Stormwater, and Decision Making for Cities in the Great Lakes; University of Michigan—Climate Center: Ann Arbor, MI, USA, 2015.
- Sills, D.; Ashton, A.; Knott, S.; Boodoo, S.; Klaassen, J. A Billion Dollar Flash Flood in Toronto-Challenges for Forecasting and Nowcasting. In Proceedings of the 28th Conference on Severe Local Storms, Portland, OR, USA, 7–11 November 2016.
- Benjamin, S.G.; Weygandt, S.S.; Brown, J.M.; Hu, M.; Alexander, C.R.; Smirnova, T.G.; Olson, J.B.; James, E.P.; Dowell, D.C.; Grell, G.A.; et al. A North American Hourly Assimilation and Model Forecast Cycle: The Rapid Refresh. *Mon. Weather Rev.* 2016, 144, 1669–1694. [CrossRef]
- 27. Hapuarachchi, H.A.P.; Wang, Q.J.; Pagano, T.C. A Review of Advances in Flash Flood Forecasting. *Hydrol Process.* 2011, 25, 2771–2784. [CrossRef]
- 28. ECCC-Environment Climate Change Canada Engineering Climate Datasets. Available online: https://climate.weather.gc.ca/prods_servs/engineering_e.html (accessed on 1 May 2022).
- 29. Rossman, L.A. Storm Water Management Model.-User's Manual Version 5.1; US EPA: Cincinnati, OH, USA, 2015.
- Computational Hydraulics International-CHI PCSWMM. Available online: https://www.pcswmm.com/ (accessed on 10 June 2022).
- 31. Dinu, C.; Sîrbu, N.; Drobot, R. Delineation of the Flooded Areas in Urban Environments Based on a Simplified Approach. *Appl. Sci.* **2022**, *12*, 3174. [CrossRef]
- 32. Huff, F.A. Time Distribution of Rainfall in Heavy Storms. *Water Resour Res.* **1967**, *3*, 1007–1019. [CrossRef]
- 33. Chow, V.T. Applied Hydrology; Clark, B.J., Morriss, J., Eds.; McGraw-Hill: New York, NY, USA, 1988; ISBN 0 07-010810-2.
- Wright, D.B.; Mantilla, R.; Peters-Lidard, C.D. A Remote Sensing-Based Tool for Assessing Rainfall-Driven Hazards. *Environ. Model. Softw.* 2017, 90, 34–54. [CrossRef]
- 35. Kennard, R.W.; Stone, L.A. Computer Aided Design of Experiments. Technometrics 1969, 11, 137. [CrossRef]
- Shenk, J.S.; Westerhaus, M.O. Population Definition, Sample Selection, and Calibration Procedures for Near Infrared Reflectance Spectroscopy. Crop. Sci. 1991, 31, 469–474. [CrossRef]
- 37. Cook, R.L. Stochastic Sampling in Computer Graphics. ACM Trans. Graph. 1986, 5, 51–72. [CrossRef]
- 38. Kohonen, T. Self-Organized Formation of Topologically Correct Feature Maps. Biol. Cybern. 1982, 43, 59–69. [CrossRef]
- 39. Kohonen, T. MATLAB Implementations and Applications of the Self-Organizing Map; Unigrafia Oy: Helsinki, Finland, 2014; ISBN 9789526036786.
- Xiong, L.; Wan, M.; Wei, X.; O'Conno, K.M. Indices for Assessing the Prediction Bounds of Hydrological Models and Application by Generalised Likelihood Uncertainty Estimation. *Hydrol. Sci. J.* 2009, *54*, 852–871. [CrossRef]
- Boylan, J.W.; Russell, A.G. PM and Light Extinction Model Performance Metrics, Goals, and Criteria for Three-Dimensional Air Quality Models. *Atmos. Environ.* 2006, 40, 4946–4959. [CrossRef]
- CBC News. Toronto's All Wet: Some Images From The Flash Floods That Hit T.O. Last Night. 2013. Available online: https://www.cbc.ca/strombo/news/torontos-all-wet-some-images-from-the-flash-floods-that-hit-to-last-night.h (accessed on 12 September 2022).
- CTV News. What's Open and Closed This Holiday Monday in Toronto? 2020. Available online: https://toronto.ctvnews.ca/ what-s-open-and-closed-this-civic-holiday-monday-in-toronto-1.6008664 (accessed on 12 September 2022).
- DailyHive News. Rain Causes Flooding on Low-Lying Toronto Highway Ramps. 2020. Available online: https://dailyhive.com/ toronto/highway-ramp-flooding-rain (accessed on 12 September 2022).
- Burg, T.; Elmore, K.L.; Grams, H.M. Assessing the Skill of Updated Precipitation-Type Diagnostics for the Rapid Refresh with MPING. Weather 2017, 32, 725–732. [CrossRef]
- 46. Chen, C.; Jiang, J.; Zhou, Y.; Lv, N.; Liang, X.; Wan, S. An Edge Intelligence Empowered Flooding Process Prediction Using Internet of Things in Smart City. J. Parallel Distrib. Comput. 2022, 165, 66–78. [CrossRef]
- 47. Fu, M.; Fan, T.; Ding, Z.; Salih, S.Q.; Al-Ansari, N.; Yaseen, Z.M. Deep Learning Data-Intelligence Model Based on Adjusted Forecasting Window Scale: Application in Daily Streamflow Simulation. *IEEE Access* **2020**, *8*, 32632–32651. [CrossRef]
- Song, T.; Ding, W.; Wu, J.; Liu, H.; Zhou, H.; Chu, J. Flash Flood Forecasting Based on Long Short-Term Memory Networks. Water 2019, 12, 109. [CrossRef]
- 49. Arsenault, R.; Martel, J.; Mai, J. Continuous Streamflow Prediction in Ungauged Basins: Long Short- Term Memory Neural Networks Clearly Outperform Hydrological Models. *Hydrol. Earth Syst. Sci.* 2022; *in review.* [CrossRef]
- 50. Kilsdonk, R.A.H.; Bomers, A.; Wijnberg, K.M. Predicting Urban Flooding Due to Extreme Precipitation Using a Long Short-Term Memory Neural Network. *Hydrology* **2022**, *9*, 105. [CrossRef]
- Ghaith, M.; Yosri, A.; El-Dakhakhni, W. Synchronization-Enhanced Deep Learning Early Flood Risk Predictions: The Core of Data-Driven City Digital Twins for Climate Resilience Planning. *Water* 2022, 14, 3619. [CrossRef]
- 52. Mosavi, A.; Ozturk, P.; Chau, K.W. Flood Prediction Using Machine Learning Models: Literature Review. *Water* **2018**, *10*, 1536. [CrossRef]