USING MACHINE LEARNING TECHNIQUES TO IMPROVE OPERATIONAL FLASH FLOOD FORECASTING

Using Machine Learning Techniques to Improve Operational Flash Flood Forecasting

By Andre Della Libera Zanchetta, M.Sc., B.Sc.

Faculty of Engineering,

Civil Engineering

A Thesis Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements for the Degree

of Doctor of Philosophy.

DECEMBER 2022

DOCTOR OF PHILOSOPHY (2022), McMaster University, Hamilton, Ontario

TITLE: Using Machine Learning Techniques to Improve Operational Flash Flood Forecasting

AUTHOR: Andre Della Libera Zanchetta

M.Sc. (The University of Iowa, United States)

B.Sc. (UNESP and UNICAMP, Brazil)

SUPERVISOR: Dr. Paulin Coulibaly

NUMBER OF PAGES: xxvii, 216

Lay Abstract

Flash floods are among the most hazardous and impactful environmental disasters faced by different societies across the globe. The timely adoption of mitigation actions by decision makers and response teams is particularly challenging due to the rapid development of such events after (or even during) the occurrence of an intense rainfall. The short time interval available for response teams imposes a constraint for the direct use of computationally demanding components in real-time forecasting chains. Examples of such are high-resolution 2D hydraulic models based on physics laws, which are capable to produce valuable flood inundation maps dynamically. This research explores the potential of using machine learning models to reproduce the behavior of hydraulic models designed to simulate the evolution of flood inundation maps in a configuration suitable for operational flash flood forecasting application. Contributions of this thesis include (1) a comprehensive literature review on the recent advances and approaches adopted in operational flash flood forecasting systems with the identification and the highlighting of the main research gaps on this topic, (2) the identification of evidences that machine learning models have the potential to identify patterns in multiple quantitative precipitation estimates from different sources for enhancing the performance of rainfall-runoff estimation in urban catchments prone to flash floods, (3) the assessment that hybrid data driven structures based on self-organizing maps (SOM) and nonlinear autoregressive with exogenous inputs (NARX), originally proposed for large scale and slow-developing flood scenarios, can be successfully applied on flashy catchments, and (4) the proposal of using

k-folding ensemble as a technique to produce probabilistic flood inundation forecasts in which the uncertainty inherent to the surrogating step is represented.

Abstract

Compared with other types of floods, timely and accurately predicting flash floods is particularly challenging due to the small spatiotemporal scales in which the hydrologic and hydraulic processes tend to develop, and to the short lead time between the causative event and the inundation scenario. With continuous increased availability of data and computational power, the interest in applying techniques based on machine learning for hydrologic purposes in the context of operational forecasting has also been increasing. The primary goal of the research activities developed in the context of this thesis is to explore the use of emerging machine learning techniques for enhancing flash flood forecasting.

The studies presented start with a review on the state-of-the-art of documented forecasting systems suitable for flash floods, followed by an assessment of the potential of using multiple concurrent precipitation estimates for early prediction of high-discharge scenarios in a flashy catchment. Then, the problem of rapidly producing realistic high-resolution flood inundation maps is explored through the use of hybrid machine learning models based on Non-linear AutoRegressive with eXogenous inputs (NARX) and Self-Organizing Maps (SOM) structures as surrogates of a 2D hydraulic model. In this context, the use of k-fold ensemble is proposed and evaluated as an approach for estimating uncertainties related to the surrogating of a physics-based model.

The results indicate that, in a small and flashy catchment, the abstract nature of data processing in machine learning models benefits from the presentation of multiple concurrent precipitation products to perform rainfall-runoff simulations when compared to the business-as-usual single-precipitation approach. Also, it was found that the hybrid

NARX-SOM models, previously explored for slowly developing flood scenarios, present acceptable performances for surrogating high-resolution models in rapidly evolving inundation events for the production of both deterministic and probabilistic inundation maps in which uncertainties are adequately estimated.

Acknowledgments

I would like to express my gratitude to Dr. Paulin Coulibaly for accepting me as one of his graduate students, for his guidance and for his continuous support during the course of my Ph.D. It is an honor for me to have him as my supervisor.

I would also like thank Dr. Yipping Guo and Dr. Zoe Li as members of my supervisory committee for their time, attention and constructive feedbacks. Their contributions were very positive and insightful for the development of my studies.

Additional thanks for all the members of the McMasters' Water Resources and Hydroclimatic Modelling Lab with whom I had the pleasure to work with and to have interesting discussions during the last years, namely: Dr. Jongho Keum, Dr. Shasha Han, Dr. James Leach, Dr. Frezer Awol, Dr. Dayal Wijayarathne, Dr. Pedram Darbandsari, M.Sc. Jacob Ursulak, M.Sc. Rasel Sheikh and M.Sc. Ruth Sankofi. Their company greatly enhanced my experience as a graduate student.

This research was made possible by the financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC), and by the administrative and structural support from the Department of Civil Engineering, from the School of Earth, Environment & Society and the McMaster University as a whole to provide an exceptional work environment.

I am eternally indebted to my parents, Roseli and Pedro, for their resolute support to my decisions and availability to provide their warm company and words during both the calmest and the toughest times, even being thousands of kilometers away.

And finally, I would like to deeply thank my wife Lizbeth for always being on my side in this journey since even before our first steps together in Canadian soils, for accepting me with all my faults, and for continuously teaching me the value and beauty of love.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Meaning
ADF	Accumulation-Duration-Frequency
AID	Average Inundation Depth
ARF	Aerial Reduction Factor
CADEX	Computer Aided Design of Experiments
CaPA	Canadian Precipitation Analysis
CAPE	Convective Available Potential Energy
CMMF	Conditional Median/Mean Function
CRPS	Continuous Ranked Probability Score
CSI	Critical Success Index
DEM	Digital Elevation Model
ECCC	Environment and Climate Change Canada
ELM	Extreme Learning Machine
EPIC	European Precipitation Index based on simulated Climatology
FC	Flow Comparison
FEWS	Flood Early Warning System
FF	Flash Flood
FFEWS	Flash Flood Early Warning System
FFG	Flash Flood Guidance
FFSA	Flash Flood Susceptibility Assessment
GFFG	Gridded Flash Flood Guidance
HAND	Height Above Nearest Drainage
IDF	Intensity-Duration-Frequency

IM	Inundation Map
KGE	Kling-Gupta Efficiency
LFFG	Lumped Flash Flood Guidance
PW	Precipitable Water
QPE	Quantitative Precipitation Estimations
QPF	Quantitative Precipitation Forecast
NARX	Nonlinear AutoRegressive Neural network with eXogenous inputs
NWM	Numerical Weather Model
POD	Probability of Detection
POI	Point of Interest
RAP	Rapid Refresh system
RC-SC	Rainfall Comparisons with Surface conditions Considered
RC-SN	Rainfall Comparison with Surface conditions Neglected
RE-O	Rainfall Events – Observations
RE-OH	Rainfall Events – Observed Highest
RE-SE	Rainfall Events – Synthetic by Extrapolation
RE-SS	Rainfall Events – Synthetic by Shuffling
ReLu	Rectified Linear Unit
RMSE	Root Mean Square Error
RH	Relative Humidity
SSM	Surface Soil Moisture
SOM	Self-Organizing Map
SR	Success Ratio
SWMM	Storm Water Management Model
US	United States

List of Symbols

Symbol	Meaning
A_f and B_f	Coefficients that define the shape of a IDF curve for a return period f
T_d	Dew point temperature
CR_b	Containing ratio for a confidence interval b
C_{pred}	Total number of correct predictions
$d_{c,t}$	Water depth in the inundation cell c at a time t
D_w	Multivariate variable of all instant water depths of an inundation map w
EP_w	Estimated probability of an inundation map w to be the best for a prediction
F	Number of "false alarms" (false positive predictions) in a contingency table
H	Number of "hits" (true positive predictions) in a contingency table
M	Number of "misses" (false negative predictions) in a contingency table
N_{pred}	Total number of predictions
P_t	Accumulated precipitation at a time t
Q_t	Instant river discharge at a time t
R^2	Coefficient of determination
$TN_{w,f}$	Topological node w of a SOM f
$\hat{y}_t$	Predicted value for a time t
y_t	Observed value at a time t
Δ_t	Temporal resolution of a timeseries
$\mathbb{P}(d_{c,t,L})$	Probability distribution of the predicted water depth for an inundation cell c at a time t issued at a lead time L
$\tau^i_{c,t,L}$	The i-th quantile value in a probability distribution $\mathbb{P}(d_{c,t,L})$
ε_t	Error of a predicted value for a time t

Declaration of Academic Achievement

This thesis was structured following a sandwich style and according to the guidelines established by the School of Graduate Studies of McMaster University as presented in the "Guide for the Preparation of Master's and Doctoral Theses - 2021".

It includes the following articles published in peer-reviewed scientific journals:

Chapter 2: Recent Advances in Real-Time Pluvial Flash Flood Forecasting, by A. D. L. Zanchetta and P. Coulibaly (2020), Water 12(2):570, DOI: 10.3390/w12020570.

Chapter 3: Forecasting High-Flow Discharges in a Flashy Catchment Using Multiple Precipitation Estimates as Predictors in Machine Learning Models, by A. D. L. Zanchetta, P. Coulibaly and V. Fortin (2022), Hydrology 9(12), DOI: 10.3390/hydrology9120216.

Chapter 4: Hybrid Surrogate Model for Timely Prediction of Flash Flood Inundation Maps Caused by Rapid River Overflow, by A. D. L. Zanchetta and P. Coulibaly (2022), Forecasting 4(1):126-148, DOI:10.3390/forecast4010007.

Chapter 5: Probabilistic Forecasts of Flood Inundation Maps Using Surrogate Models, by A. D. L. Zanchetta and P. Coulibaly (2022), Geosciences 12 (11), DOI: 10.3390/geosciences12110426.

A. D. L. Zanchetta was the writer of all chapters; P. Coulibaly was the proponent of the main research topic and supervised, providing research guidance and reviewing the content of all chapters. V. Fortin provided research guidance and support with data access for the work presented in the Chapter 3.

The research activities reported in this work were developed between September 2018 and November 2022.

Chapter 1. Introduction

1.1. Background

Floods caused by river overflow are events observed in natural environments with frequencies that vary by location. Due to the adaptiveness of different species of animals and vegetation resident in floodplains to the recurrent inundations, most of times such floods are not considered to be disastrous. However, human occupation in the vicinity of streams may expose local residents, commuters, and exogenous fauna and flora anthropogenically introduced to hazard conditions in the event of riverine floods, which can lead to devastating consequences [1]. While the number of fatalities caused by floods is varied across the globe, the rate of occurrence of such events, the number of affected people and total flood-induced mortality worldwide has been increasing [2]. Potential causes include (1) climate change, which has the potential of inducing extreme rainfall events with higher intensity and/or frequency [3], (2) intense urbanization, which is associated with a decrease of the soil perviousness and suppression of vegetation [4], and (3) the continuous occupation of flood-prone areas [5].

Large-scale, slow-onset floods, which are usually driven by sustained rainfall, extensive snowmelt or high tides have the potential to cause massive displacement and material damage [6,7]. However, one type of high impacting inundation event of special concern due to dynamicity and areas occurrence is flash floods. Both the World Meteorological Association (WMA) and the United States of America (USA) National

Weather Service (NWS) define flash floods as the events in which the stream overbank condition occurs up to six hours from the causative event, which can be, for example, intense precipitation or dam failures [8,9].

1.2. Current Mitigation and Responsive Approaches

Several structural interventions in urban areas have been proposed and implemented to mitigate the impact of intensified soil imperviousness on peak river discharges. In this context, the low-impact development practices (LID) consist of implementing local structures capable of improving the overall stormwater management in a specific area, such as green roofs [10], infiltration trenches [11], permeable pavements [12]. Additionally, larger-scale constructions designed to allow the regulation of stream flows, such as dams and reservoirs, are historically implemented to protect sensitive areas downstream [13].

Despite of an increasing adoption of the aforementioned techniques and their associated positive impact in the reduction of peak flows and of instant runoff generation, they are not enough to completely suppress the occurrence of flash floods, as documented in the Emergency Disasters Database (Em-DAT) [14] and represented in Figure 1-1. Thus, it is of paramount importance for many societies to maintain a state of readiness so that appropriate responsive actions can be taken in the imminence of an upcoming hazardous flood scenario. Such actions commonly include the evacuation of buildings, the closure of roads, the interruption of massive transport systems operation, and the mobilization of monitoring and rescue teams.

Flood early warning systems (FEWS) have been implemented worldwide to support decision makers on taking effective responsive actions to flood hazard [15,16]. One of the key elements of FEWSs is a forecasting system capable to accurately predict the occurrence of flood events, which needs to suit the environmental characteristics and the spatiotemporal scales of the monitored area. When such an area is characterized by catchments prone to flash floods, high temporal (hourly or finer) and high spatial (a few meters or finer) data resolutions are required to properly represent the development of inundation events [17]. In addition, forecasts of flash floods are expected to be updated in periods as short as possible, preferentially at hourly or sub-hourly rates, so that the rapid changes in the hydrometeorological conditions of the monitored region are taken into consideration by the prediction system.

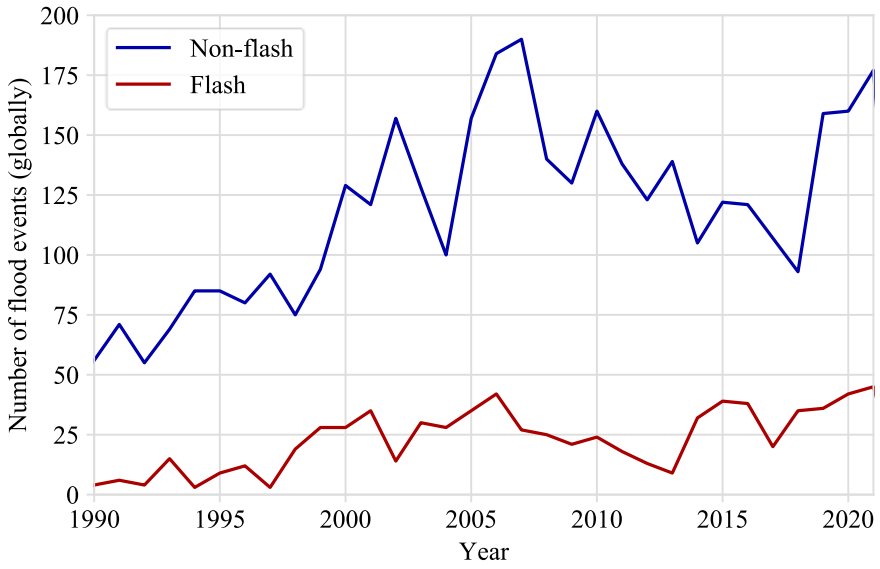

Figure 0-1. Flood frequency worldwide from 1990 to 2021.
Source: EM-DAT [14].

1.3. Problem Definition

The rapid development and the localized area of flash floods are characteristics that impose specific challenges for their appropriate prediction when compared to their non-flashy flood counterparts. Considering that localized and intense rainfall over an area is usually the dominant factor that triggers these flooding events, that the magnitude of the floods is influenced by multiple characteristics of the contributing catchment(s), and that the shape and depth of the resulting inundated area is highly correlated with the characteristics of the underlying land surface, forecasting flash floods can be interpreted as a complex problem that involves members of both meteorological and hydrological communities.

From a meteorological perspective, major challenges consist of anticipating the occurrence, the timing and the location of intense rainfall events capable to trigger flash floods [18]. Thus, improving the ability to forecast flash floods is intimately related to improving the quality and availability of quantitative precipitation estimates (QPE) and forecasts (QPF) products, especially in areas where precipitation is driven by local convection and/or by orography [19].

From a hydrological perspective, atmospheric data including QPEs and QPFs are considered as inputs for water-related processes occurring both in the land surface and, when relevant, in the soil subsurface. Here, major challenges consist of timely anticipating if the runoff produced by a catchment in response to an intense QPE or QPF record will exceed overbank conditions and, if so, which areas can be potentially affected by the resulting flood inundation. For such, conceptual or physics-based hydrologic models are

conventionally used to simulate the local rainfall-runoff process, and hydraulic models solving the computationally expensive shallow water equations are usually applied for simulating the development of the flood inundation resulting from overbank conditions. In this context, improving the ability to forecast flash floods is directly related to improving the quality and availability of hydrographs and flood inundation maps.

The availability of hydrometeorological data is considerably heterogeneous across the globe in terms of quality, variety and historical records length, which results in different groups of meteorologists and hydrologists approaching the task of forecasting flash floods with distinct restrictions and resources. In North America and Europe, extensive weather radar networks have been operational for more than three decades [20,21] in addition to long maintained and relatively dense meteorological monitoring stations. Such data availability has enabled the development and implementation of operational numerical weather models (NWM) capable of producing realistic QPEs and QPFs at spatiotemporal resolution finer than the global products derived from satellites [22]. In face of this variety of sources, forecasting system designers have to take decisions on which of them and how the data will be pre-processed and used as input.

Forecasting systems are usually designed to have a hydrological model being recurrently executed in continuous mode to simulate the rainfall-runoff of the covered catchment(s) and thus producing point hydrographs. While hydrodynamic models have long been assessed as capable to produce realistic flood inundation maps even when high motion is present, as it is the case for flash floods, these models are not usually used in operational time due to their restrictive computational costs [23].

In recent years, due to an increase in the computational resources availability, an ever-growing volume of data and the development of new algorithms, the use of data-driven techniques based in machine learning is gaining a crescent attention both from the research [24] and the operational [25] communities. However, despite being a field of study for more than 30 years [26,27], most of the works exploring the use of machine-learning techniques to predict floods is focused on large-scale inundation events.

The works presented in this thesis have the broad objective of documenting the current state-of-the-art of operational flash flood forecasting systems and exploring how different machine learning techniques can be applied to improving the prediction of flood events with the specific conditions, requirements and circumstances of flash floods.

1.4. Research Scope and Objectives

A total of four studies were conducted with the broad objective of potentially improving the predictive power and capabilities of flash flood forecasting systems through the application of machine learning methods. The specific objectives of each research work are:

1. To provide a broad and comprehensive perspective of the state-of-the-art and of the advances documented in topics related to pluvial flash flood forecasting, with focus on publications issued in the decade of 2010-2020.
2. To explore the potential of using multiple concurrent precipitation products for enhancing the performance of a rainfall-runoff extreme learning machine model as an alternative to the business-as-usual single-precipitation input approach.

3. To address the issue of generating deterministic flood inundation maps in a time frame compatible with operational forecasting chains by evaluating the performance of hybrid network structures composed by nonlinear autoregressive neural networks with exogenous inputs (NARX) and self-organizing maps (SOM) in high spatiotemporal resolution.
4. To propose and evaluate the use of k-fold ensembles of NARX-SOM hybrid surrogate models for timely producing probabilistic flood inundation maps, and to assess the impact of including QPF data as an additional predictor of the data-driven surrogate models.

The first review study explores almost 200 research articles and technical documents to present a broad perspective on the operational and openly documented flash flood forecasting systems worldwide, contrasting their differences in terms of decision criteria, input data used, and particularities of the different spatial scopes covered. The second, third and fourth studies used the Don River Basin located in the Greater Toronto Area as the study case scenario for the research activities on assessing and proposing the abovementioned methodologies explored in this work. The choice for the Don River Basin is due to its propensity to flash flooding and to the availability of observational hydrometeorological data in spatiotemporal resolutions compatible with the requirements for forecasting such events.

1.5. Thesis Outline

The thesis is organized in six chapters, each of which subdivided in specific sections and subsections. In this first chapter, it is presented an introductory discussion on the needs and challenges of forecasting flash floods, as well the motivations on exploring the applicability of machine learning techniques for such a task. The second chapter provides a comprehensive picture of the advances in research and operational forecasting tools designed or compatible with the requirements of a flash flood prediction system. A research activity in which the use of precipitation data captured in rain gauges, estimated from ground-based weather radar signals and simulated by a numerical weather model are used concurrently for rainfall-runoff modeling with focus on the detection of high flows composes the third chapter. In the fourth chapter, the results of a study evaluating the applicability of a hybrid recurrent/classification network for timely producing deterministic high-resolution flood inundation maps are presented and discussed. Such a study is extended so that the generation of probabilistic flood inundation maps and the inclusion of precipitation forecasts as part of the input set are assessed, with results provided in the fifth chapter. The thesis is finalized in the chapter six, in which the main conclusions and contributions of the research work developed are summarized and suggestions for future works are discussed.

1.6. References

1. 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; AMS: Portland, OR, USA, November 2016.

2. Hu, P.; Zhang, Q.; Shi, P.; Chen, B.; Fang, J. Flood-Induced Mortality across the Globe: Spatiotemporal Pattern and Influencing Factors. *Science of the Total Environment* **2018**, *643*, 171–182, doi:10.1016/j.scitotenv.2018.06.197.
3. Sun, Q.; Zhang, X.; Zwiers, F.; Westra, S.; Alexander, L. V. A Global, Continental, and Regional Analysis of Changes in Extreme Precipitation. *J Clim* **2021**, *34*, 243–258, doi:10.1175/JCLI-D-19-0892.1.
4. 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, doi:10.1038/srep40527.
5. Jongman, B. The Fraction of the Global Population at Risk of Floods Is Growing. *Nature* **2021**, *596*, 37–38, doi:10.1038/d41586-021-01974-0.
6. Yu, Q.; Wang, Y.; Li, N. Extreme Flood Disasters: Comprehensive Impact and Assessment. *Water (Switzerland)* **2022**, *14*, 1–14, doi:10.3390/w14081211.
7. Kakinuma, K.; Puma, M.J.; Hirabayashi, Y.; Tanoue, M.; Baptista, E.A.; Kanae, S. Flood-Induced Population Displacements in the World. *Environmental Research Letters* **2020**, *15*, doi:10.1088/1748-9326/abc586.
8. WMO — World Meteorological Organization. *Climate and Water*. WMO Bulletin Vol. 69 (1), 2020. WMO: Geneve, Switzerland.
9. NWS — National Weather Service. NWS Glossary. Available online: https://w1.weather.gov/glossary/ (accessed on 4 November 2022).

10. Ercolani, G.; Chiaradia, E.A.; Gandolfi, C.; Castelli, F.; Masseroni, D. Evaluating Performances of Green Roofs for Stormwater Runoff Mitigation in a High Flood Risk Urban Catchment. *J Hydrol (Amst)* **2018**, *566*, 830–845, doi:10.1016/j.jhydrol.2018.09.050.

11. Bonneau, J.; Fletcher, T.D.; Costelloe, J.F.; Burns, M.J. Stormwater Infiltration and the 'Urban Karst' – A Review. *J Hydrol (Amst)* **2017**, *552*, 141–150, doi:10.1016/j.jhydrol.2017.06.043.

12. Alsubih, M.; Arthur, S.; Wright, G.; Allen, D. Experimental Study on the Hydrological Performance of a Permeable Pavement. *Urban Water J* **2017**, *14*, 427–434, doi:10.1080/1573062X.2016.1176221.

13. Khaddor, I.; Achab, M.; Soumali, M.R.; Benjbara, A.; Alaoui, A.H. The Impact of the Construction of a Dam on Flood Management. *Civil Engineering Journal (Iran)* **2021**, *7*, 343–356, doi:10.28991/cej-2021-03091658.

14. Guha-Sapir, D.; Below, R.; Hoyois, Ph. EM-DAT: The CRED/OFDA International Disaster Database. Available online: www.emdat.be (accessed on 4 November 2022).

15. Perera, D.; Seidou, O.; Agnihotri, J.; Rasmy, M.; Smakhtin, V.; Coulibaly, P.; Mehmood, H. *Flood Early Warning Systems: A Review Of Benefits, Challenges And Prospects 08*; United Nations University: Hamilton, ON, Canada, 2019.

16. Georgakakos, K.P.; Modrick, T.M.; Shamir, E.; Campbell, R.; Cheng, Z.; Jubach, R.; Sperfslage, J.A.; Spencer, C.R.; Banks, R. The Flash Flood Guidance System Implementation Worldwide: A Successful Multidecadal Research-to-Operations

Effort. *Bull Am Meteorol Soc* **2022**, *103*, E665–E679, doi:10.1175/BAMS-D-20-0241.1.

17. Hapuarachchi, H.A.P.; Wang, Q.J.; Pagano, T.C. A Review of Advances in Flash Flood Forecasting. *Hydrol Process* **2011**, *25*, 2771–2784, doi:10.1002/hyp.8040.
18. Barthold, F.E.; Workoff, T.E.; Cosgrove, B.A.; Gourley, J.J.; Novak, D.R.; Mahoney, K.M. Improving Flash Flood Forecasts : The HMT-WPC Flash Flood and Intense Rainfall Experiment. *Bull Am Meteorol Soc* **2015**, *96*, 1859–1866, doi:10.1175/BAMS-D-14-00201.1.
19. Yu, W.; Nakakita, E.; Kim, S.; Yamaguchi, K. Improvement of Rainfall and Flood Forecasts by Blending Ensemble NWP Rainfall with Radar Prediction Considering Orographic Rainfall. *J Hydrol (Amst)* **2015**, *531*, 494–507, doi:10.1016/j.jhydrol.2015.04.055.
20. Saltikoff, E.; Haase, G.; Delobbe, L.; Gaussiat, N.; Martet, M.; Idziorek, D.; Leijnse, H.; Novák, P.; Lukach, M.; Stephan, K. OPERA the Radar Project. *Atmosphere (Basel)* **2019**, *10*, 1–13, doi:10.3390/atmos10060320.
21. Whiton, R.C.; Smith, P.L.; Bigler, S.G.; Wilk, K.E.; Harbuck, A.C. History of Operational Use of Weather Radar by U.S. Weather Services. Part I: The Pre-NEXRAD Era. *Weather Forecast* **1998**, *13*, 219–243, doi:10.1175/1520-0434(1998)013<0219:HOOUOW>2.0.CO;2.
22. Ramsauer, T.; Weiß, T.; Marzahn, P. Comparison of the GPM IMERG Final Precipitation Product to RADOLAN Weather Radar Data over the Topographically

and Climatically Diverse Germany. *Remote Sens (Basel)* **2018**, *10*, doi:10.3390/rs10122029.

23. 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. *Environmental Modelling and Software* **2017**, *90*, 201–216, doi:10.1016/j.envsoft.2017.01.006.
24. Mosavi, A.; Ozturk, P.; Chau, K.W. Flood Prediction Using Machine Learning Models: Literature Review. *Water (Basel)* **2018**, *10*, 1–40, doi:10.3390/w10111536.
25. Nevo, S.; Morin, E.; Rosenthal, A.G.; Metzger, A.; Barshai, C.; Weitzner, D.; Kratzert, F.; Elidan, G.; Dror, G.; Begelman, G.; et al. Flood Forecasting with Machine Learning Models in an Operational Framework. *Hydrol Earth Syst Sci* **2021**, 1–31.
26. Dawson, C.W.; Wilby, R.L. Hydrological Modelling Using Artificial Neural Networks. *Prog Phys Geogr* **2001**, *25*, 80–108, doi:10.1177/030913330102500104.
27. Abrahart, R.J.; Anctil, F.; Coulibaly, P.; Dawson, C.W.; Mount, N.J.; See, L.M.; Shamseldin, A.Y.; Solomatine, D.P.; Toth, E.; Wilby, R.L. Two Decades of Anarchy? Emerging Themes and Outstanding Challenges for Neural Network River Forecasting. *Prog Phys Geogr* **2012**, *36*, 480–513, doi:10.1177/0309133312444943.

Chapter 2. Recent Advances in Real-Time Pluvial Flash Flood Forecasting

Summary of the review article: Zanchetta, A.D.L. and Coulibaly, P. (2020) Recent Advances in Real-Time Pluvial Flash Flood Forecasting. Water 12(2): 570. DOI: 10.3390/w12020570.

This literature review characterizes the state-of-the-art of pluvial flash flood forecasting systems with focus on the developments and studies documented in the decade of 2010 to 2020. The broad questions explored are:

- How has flash flood forecasting evolved in the last decade of 2010 — 2020?
- Which are the current major gaps and trends in this field?

Key findings include:

- Multiple flood forecasting systems were implemented worldwide during the decade using a high variety of inputs, model structure and decision criteria that are driven by local availability of data and by the characteristics of the region covered by the system.
- There is a continuous search for higher resolution data, with X-band weather radar being assessed as providing more adequate for predicting flash floods than the conventional C- and S-band counterparts, and the implementation of a dense constellation of autonomous monitoring sensors designed following the concept of Internet of Things is being considered as alternative to the conventional centralized networks.

- Rainfall-runoff is usually simulated using physics-based or conceptual models. Machine learning models were not documented in operational applications and were considered as potential tools for timely generation of inundation maps as surrogate models.

2.1. Abstract

Recent years have witnessed considerable developments in multiple fields with the potential to enhance our capability of forecasting pluvial flash floods, one of the most costly environmental hazards in terms of both property damage and loss of life. This work provides a summary and description of recent advances related to insights on atmospheric conditions that precede extreme rainfall events, to the development of monitoring systems of relevant hydrometeorological parameters, and to the operational adoption of weather and hydrological models towards the prediction of flash floods. With the exponential increase of available data and computational power, most of the efforts are being directed towards the improvement of multi-source data blending and assimilation techniques, as well as assembling approaches for uncertainty estimation. For urban environments, in which the need for high-resolution simulations demands computationally expensive systems, query-based approaches have been explored for the timely retrieval of pre-simulated flood inundation forecasts. Within the concept of the Internet of Things, the extensive deployment of low-cost sensors opens opportunities from the perspective of denser monitoring capabilities. However, different environmental conditions and uneven

distribution of data and resources usually leads to the adoption of site-specific solutions for flash flood forecasting in the context of early warning systems.

2.2. Introduction

Flash floods (FFs) are among the most damaging types of weather-related disasters faced nowadays. They may be caused either by extreme precipitation, by the failure of human-made structures, such as dams, or by complex water-snow interactions. The fast development of FFs imposes additional challenges for early prediction when compared to riverine floods. Structural measures adopted to reduce the impact of these events include the construction of physical components aimed to enhance the overall resilience of drainage systems, such as levees and detention ponds. Nonstructural solutions include the adoption of regulations for land use/occupation, personal training for responsive actions, and the implementation of operational flash flood early warning systems (FFEWSs).

A core feature of FFEWSs is the capability to perform timely and accurate FF forecasts. Methods for FF forecasting demand continuous improvement, mainly in the current context of progressive changes in urbanization and climate patterns that lead to an increased susceptibility to FFs observed in different locations worldwide [1–3]. The work developed by Hapuarachchi et al. in 2011 [4] brings comprehensive state-of-the-art for its time in the topics of input data, modeling approaches and estimation of uncertainties related to FF forecasting. Since then, several advances were observed in multiple related fields driven by an expansion of monitoring capabilities, consolidation of extensive datasets, and the establishment of higher resolution models due to increasing computational power.

The objective of this review paper is to provide a centralized summary of recent developments associated with FF forecasting with a focus on existing or potential real-time operational applications and to discuss the latest insights on promising opportunities for their enhancement. The main contribution of this paper relies on answering the questions, "How has flash flood forecasting evolved in the last decade?" and "Which are the current major gaps and trends in this field?".

As such, a non-structured literature review was performed over a selected number of papers published in renowned peer-reviewed journals, official technical reports, and conference abstracts considered relevant for topics related to the enhancement of operational FF forecasting.

2.2.1. Scope, Definitions, and Work Structure

Flash floods are defined by the United States' (US) National Weather Service as "A rapid and extreme flow of high water into a normally dry area, or a rapid water level rise in a stream or creek above a predetermined flood level, beginning within six hours of the causative event (e.g., intense rainfall, dam failure, ice jam)" [5]. From an operational perspective, priority is usually given to the capability of predicting their occurrence, while for riverine floods primary importance is given to the prediction of their magnitudes [6].

Various terminologies can be applied for specific activities associated with prediction systems. In this work, "anticipation" is used to qualitatively describe the expected occurrence/non-occurrence of an event in the near future, with no (or very few) details about the upcoming scenario. The term "forecast" is used for the activities that generate quantitative information in time and space. Specifically, the terms "short-term forecast"

and “nowcast” are used for forecast windows of up to 6 h [7,8], and “long-term forecast” is applied when the forecast window is longer.

This work considers only events driven by extreme precipitation due to their significantly higher occurrence when compared to the ones triggered by other circumstances. Pluvial FFs can be caused by deep and local convective precipitation, orographic effects, storm surges, and cyclones. As cyclones are usually associated with synoptic-scale patterns and have their specific and extensive research field, they are not explicitly discussed in this work. From this perspective, FF forecasting is highly related to the challenging meteorological problem of predicting extreme local rainfall events [9].

Multiple approaches were proposed and implemented for FF prediction as environments with different configurations and available datasets are prone to such types of hazards, usually leading to the adoption of customized solutions. FF prone areas include non-urban steep catchments [10], urban or semi-urban catchments [2], urban neighborhoods served [11] (or not [12]) by a central drainage channel, and coastal urban zones [13], as illustrated in Figure 0-1. In this work, we use the expression “catchment” (and “sub-catchment”) for areas in which runoff is directed to a single outlet point (and an inlet point is present). When the boundaries of the study area take into consideration non-geomorphological delineations (as administrative borders), the term “neighborhood” is used. When both boundary definitions are acceptable, the expressions “zone” or “environment” are interchangeably used.

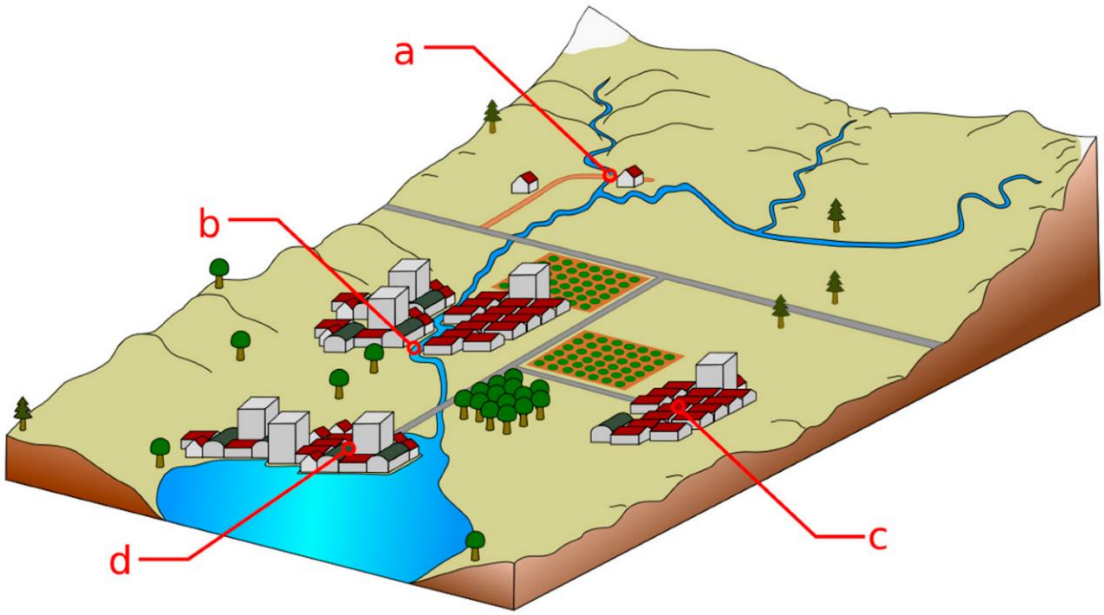

Figure 0-1. Different types of flash flood-prone environments include (a) non-urban steep catchments and (b) urban neighborhoods served or (c) not by a central drainage channel, and (d) coastal urban zones.

As this type of hazard is mainly characterized by occurring with short development time and on small catchments, advances towards: (1) increase in the spatiotemporal resolution and precision of input and output data, (2) increase of overall lead time and awareness, and (3) reduction of the total computational expenses for the generation of relevant products are assumed to be of interest to the problem and are explored in this work.

The discussion is presented as follows. In Section 2.3, the different approaches usually adopted to determine whether a FF is expected or not are described. In Section 2.4, we comment on results from a selected number of recently published exploratory analyses on atmospheric contributing factors for extreme precipitation events. In Section 2.5, recent advances and developments on remote sensing techniques relevant to the subject are commented on. Precipitation prediction and hydrodynamic models involved in forecasting chains are presented in Section 2.6 and Section 2.7, respectively. The work is concluded

with a summarizing discussion of the key findings from the presented literature review and recommendations of future work in the field.

2.3. Criteria for Deciding Flash Flood Occurrence

At the operational time, the resolution of whether or not an FF event is expected to happen in the near future at a given location can be determined through different approaches and it is usually responsible for triggering (or not) the first responsive actions to the upcoming hazard. The choice of which method to adopt depends on multiple factors, including resource availability, previous experience acquired, and even personal preferences of the operational team. The different approaches are presented in four classes (Sections 2.3.1 to 2.3.4), sorted by an increasing level of complexity. Such division is derived from the classification adopted by Hapuarachchi et al. [4], with the difference that two families of rainfall comparison methods are defined, taking into consideration whether surface conditions are considered or not in the representation of the rainfall-runoff process. A simplified diagram of the different families of approaches is presented in Figure 2-2.

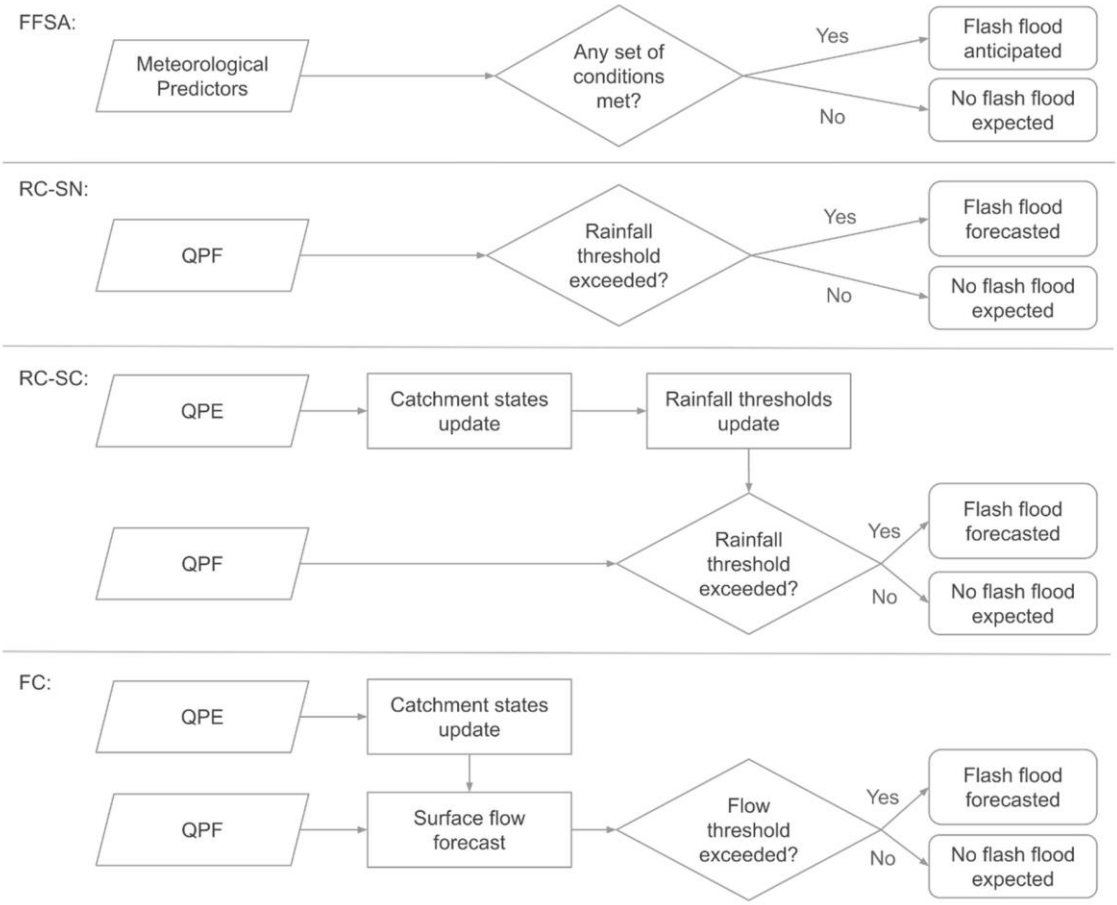

Figure 0-2. Overall workflows adopted by the different decision approaches usually present in flash flood early warning systems (FFEWS).

2.3.1. Flash Flood Susceptibility Assessment (FFSA)

One relatively simple approach that can be used as a first step for anticipating FF events is based on the assessment of multiple hydrometeorological conditions known to precede extreme precipitation scenarios. It can be performed through ingredients-based [14], checklists [15], or scoring techniques [16].

Quantitative precipitation forecast (QPF) products are frequently considered to be part of the data available for FF forecasters. However, when they are missing or considered not

reliable, a core objective becomes the prediction of extreme rainfall events itself. Meteorological parameters traditionally explored for such include precipitable water (PW), relative humidity (RH), dew point temperature (T_d), convective available potential energy (CAPE), and the so-called K-index, which describes the local potential for thunderstorms [17]. When antecedent surface conditions, such as surface soil moisture (SSM), are part of the predictors, they are accounted for through simplified means, such as the integration of recently observed precipitation. Due to its meteorological-driven basis, this family of monitoring activities is usually performed by weather service teams (i.e., meteorologists) instead of by river forecast centers (i.e., hydrologists).

Research areas of benefit for the flash flood susceptibility assessment (FFSA) include the search for optimal meteorological FF predictors, the application of data provided by newly available meteorological monitoring systems by operational teams, and the proposal of strategies to communicate uncertainties associated with the eventual unavailability of part of the data used [18].

2.3.2. Rainfall Comparison with Surface Conditions Neglected (RC-SN)

When QPF products are available, the decision of whether a FF is expected to happen on a catchment can be made based solely on a threshold-exceedance assessment of the predicted peak precipitation accumulation value [19].

Rainfall frequency analysis approaches can be used for establishing the raw static thresholds. Rainfall return periods are usually adopted on FFEWS designed for large areas covering multiple low-urbanized catchments since such locations are usually poorly monitored and the dataset needed for empirical definitions is thus unavailable. QPF

intensity values are translated to their respective estimated return periods based on reforecast analysis (e.g., European Precipitation Index based on simulated Climatology — EPIC [20,21]) or on intensity duration frequency (IDF) curves retrieved either from radar (e.g., Guadalhorce basin Flood Warning System — GFWS [22], European Rainfall-InduCed Hazard Assessment system — ERICHA [23]) or rain gauge [24] observations. When the monitored catchments are assumed to share similar rainfall-runoff response behavior, a common return period value is usually applied as the threshold for all gauges covered by the system, which favors a fast interpretation of the data.

For urban environments, in which rainfall-runoff response can be highly heterogeneous in space, rainfall thresholds can be defined at neighborhood level if past FF events were properly documented. Under such circumstances, recent works have obtained acceptable results by simply performing graphical analysis of historical events [25] or by updating first-guess thresholds established from simplified hydrodynamic simulations in a hit-and-miss fashion [26]. These works illustrate the importance and applicability of high-quality disaster datasets for FF forecasting.

Since the QPF was the sole input considered by this approach, advances in precipitation forecast capabilities are considered of special benefit for rainfall comparison with surface conditions neglected (RC-SN) approaches.

2.3.3. Rainfall Comparison with Surface Conditions Considered (RC-SC)

The family of approaches based on rainfall comparisons with surface conditions considered (RC-SC) evaluates the raw rainfall forecasted taking into consideration the respective effective rainfall to be generated. For each location, a static flood-initiating

runoff threshold (Thresh-R) value is defined. In real-time, catchment states (e.g., SSM and channel storage) are continually updated taking into consideration remotely sensed data. To consider such transient conditions, FF-generating rainfall thresholds are also dynamically updated and then compared against QPF peak values (backward comparison). The additional component, usually a hydrologic model, in the prediction chain used for tracking the estimated surface parameters increases the overall complexity of such systems. At gauged locations in which a reliable rating curve is available bank full water level values can be used as the Thresh-R [27]. For ungauged sites, Thresh-R values can be estimated from the flow frequency analysis of simulation datasets [28].

The flash flood guidance (FFG) is probably the most prominent framework of this approach. It was adopted by river forecast centers in the US in the 1970s and has been recently implemented operationally in different countries [6]. It is based on the recurrent estimation of the total raw precipitation needed to occur during specific time intervals (usually 1, 3, and 6 h) to cause flood scenarios, and the rainfall-runoff transformation is usually performed by a continuous hydrological model, be it spatially lumped flash flood guidance (LFFG) or gridded flash flood guidance (GFFG) [29,30]. Regardless of the discretization used to represent the terrain in the hydrological model, rainfall is finally represented as uniformly distributed so that a single precipitation value can be used as a threshold. The uncertainty estimation for these methods should account for both the rainfall data aggregation step [31] and the uncertainties of the own input data used [32]. Recent efforts applied to enhance the representativeness of the uncertainty associated to FFG values include the consideration of the spatial rainfall features defined through statistical

analysis [33] and the adoption of Bayesian probabilistic approaches to consider the fact that the same amount of accumulated rainfall may or may not trigger floods [34].

To simplify the operational forecasting chains, promising probabilistic approaches based on Bayesian utility [35] and risk entropy [36] paradigms were presented. The probabilistic functions used to dynamically update the critical rainfall values were fitted using extensive offline hydrological model simulations assuming simplified antecedent SSM conditions.

2.3.4. Flow Comparison (FC) Flow

Flow comparison (FC) approaches use QPF products as input to real-time running hydrological models so that the simulated output surface flow (expressed as surface runoff or channel discharge) is directly compared against static Thresh-R values (forward comparison). Such a direct comparison has the advantage of also providing information related to the magnitude of the upcoming event.

To enhance communication among stakeholders, Thresh-R values are usually presented in terms of the return period. Flow duration curves can be derived from grid-based statistical analysis of multiple historical simulations [37,38] or from flow quantile regionalization of gauged data [39,40].

2.3.5. Performance Comparison and Multi-Approach Tools

The increasing number of proposed and implemented FF forecasting systems motivated recent works based on comparative analysis to identify the most accurate approaches adopted. The critical success index (CSI) is a common metric used for assessing the

efficiency of an operational system on detecting the occurrence of real FF events (hits) in studies developed over areas where a considerable number of FFs are registered. CSI also takes into consideration the observed events that were not detected (misses) and false alarms issued by:

$$CSI = \frac{hits}{hits + misses + false\ alarms} \quad (2\text{-}1)$$

in which a CSI value of 1.0 means a perfect performance, while a value of 0.0 means a total lack of skill.

A set of selected operational FFEWS is presented with their core features in Table 2-1. The results summarized in Table 2-2 from recently published comparison works illustrate how the best criterion for detecting FF scenarios depends on the target domain. Overall, the increase in complexity associated with the transition from a RC to a FC approach is justified by a significant gain in performance for systems covering large domains [38,41]. Such results motivate the inclusion of hydrological models into forecasting chains of FFEWSs under implementation [42]. However, the same pattern was not observed in studies developed over more restricted scopes [22,43]. One explanation can be that systems covering large domains may include several catchments that, despite being small, are not urbanized and present smooth relief. Such catchments are more prone to floods originated from the runoff concentration downstream of the rainfall location and are better represented by hydrologic models due to their capability to identify floods not occurring at the exact same location as the causative rainfall cell. Specific-domain systems, on the other hand, are usually implemented to cover regions known to be extremely "flashy", be it due to the presence of urban areas and/or mountains, and the almost instantaneous process of rainfall-

runoff transformation significantly reduces the role of hydrological models in the forecasting chain. However, most of the comparative works do not take into consideration RC methods based on empirical and probabilistic approaches, despite their potential to perform FF forecasts [36]. Besides the performance, another critical element to be considered is the usual lead time for FF detection associated with each approach. As reported by Lincoln [44], despite the potential for less accurately representing the actual reports of FFs, a RC-SN system using uncorrected radar rainfall estimates as the input was considered more helpful for river forecast centers operators than its FC-based counterpart fed with gauge-corrected quantitative precipitation estimations (QPE) due to the fact that the update time of the former was much shorter than the latter, thus increasing the response time of decision-makers for responsive actions.

Table 0-1. Summary of selected reported operational FFEWS systems sorted by criteria.

Reference	Criteria	Operational System	Coverage	QPE Source	QPF Source	Resolution	Forecast Window
[20,21,45]	RC-SN	EPIC-EFAS[1]	Continental Europe	N/A	NWP[2]	6 h/7 km	5 days
[46]	RC-SN	ERICHA-EFAS	Continental Europe	N/A	Radar extrapolation	15 min/1 km	6 hours
[47]	RC-SC	FFG-BSMEFFG	Multinational Middle East	Satellite, radars	CP-NWP[4]	1 h/50 km^2	6 hours
[48]	RC-SC	FFG-HDRFFGS	Haiti, Dominican Republic	Satellite	CP-NWP[3]	1 h/70 km^2	36 hours
[38]	FC	ERIC-EFAS	Continental Europe	N/A[6]	NWP[2]	6 h/1 km	5 days
[37,49]	FC	DHM-TF-FFMP	Single large basin (US)	Radar mosaic	Radar extrapolation	1 h/4 km	1 hour
[39,40,50]	FC	AIGA[8] -Vigicrues Flash	National (France)	Radar mosaic	N/A	15 min/1 km	6 hours
[51]	FC	Flood-PROOFS	Regional (Liguria, Italy)	Satellite	Stat. down. NWP[10]	30 min/1 km	3 days

[1] The European Precipitation Index based on Climatology (EPIC) was replaced operationally by the European Runoff Index based on Climatology (ERIC) in the context of the European Flood Awareness System (EFAS). [2] Numerical weather prediction (NWP). [3] Black Sea and Middle East FFG system (BSMEFFG). [4] Convection-permitting NWP (CP-NWP). [5] Haiti and Dominican Republic FFG (HDRFFGS). [6] Soil moisture of ERIC is updated daily with NWP estimations. [7] Distributed Hydrologic Model-Threshold Frequency (DHM-TF) in the context of the Flash Flood Monitoring and Prediction (FFMP) program. [8] Geographic information adaptation for flood warning (in French: Adaptation d'Information géographique pourl'Alerte en crue—AIGA). [9] Flood-PRObabilistic Operational Forecasting System (Flood-PROOFS). [10] Statistical downscaling of an NWP product.

Table 0-2. Summary of results from selected comparative works of systems based on different flash flood identification criteria.

Reference	Criteria	Description	Best CSI *	Resolution	Coverage	Conclusion/Highlights
[52]	RC-SN RC-SC RC-SC RC-SC	Empirical rainfall thresholds Online model simulation Bayesian utility function Risk entropy function	0.29-0.45 0.20-0.57 0.14-0.56 0.14-0.68	30 min / lumped	Three non-urban catchments, Italy	Simple empirically-based thresholds presented the best performance for catchments with limited datasets. Others outperformed depending on available data.
[41]	RC-SC RC-SC FC	LFFG GFFG DHM-TF	0.19-0.34 0.20-0.22 0.32-0.47	1 h / 4 km	Large monitored rural basin, US (70 stations)	Clear overperformance of the FC approach when compared to RC-SC (FFG) methods.
[38]	RC-SN FC	EPIC ERIC	0.34 0.49	6 h / 1 km	Continental Europe	The cost-benefit of the FC approach was positive.
[43]	RC-SN FC	ERICHA Flood-PROOFS	N/A	10 min / 1 km	Mountainous periurban region, Italy	No significant differences found in performance between the two systems.
[23]	RC-SN FC	Rainfall IDF curves Online model simulation	N/A	10 min / 1 km	Large poorly gauged periurban basin, Spain	Both approaches were efficient for FF forecasting, but FC is also efficient for non-FF forecasts.

* CSI values are presented in ranges when studies considered multiple scenarios. The best CSI value is underlined.

With that perspective, interactive toolsets were proposed and implemented operationally to communicate multiple metrics concurrently on clear graphic user interfaces to support decision-makers (e.g., Hydrometeorological Risks in Mediterranean and Mountainous Areas (in French: Risques Hydrométéorologiques en Territoires Montagnards et Méditerranéens — RHYTMME) [53], Flooded Locations and Simulated Hydrographs — FLASH [54]).

2.4. Insights into Meteorological Contributors to Flash Floods

The problem of identifying and explaining meteorological contributing conditions to FFs causing extreme precipitation has long been explored. Extensive analysis of overall synoptic and mesoscale atmospheric patterns [9], together with the consolidation of observed and modeled datasets, motivated the development of studies focused on quantitatively identifying effective antecedent FF atmospheric descriptors to support decision-makers and response teams (the FFSA approach).

Despite overall agreement that descriptors associated to air moisture (e.g., RH, PW, Td) and atmospheric stability (e.g., K-index, CAPE) have the high predictive potential for extreme rainfall events, and thus to pluvial FFs, there is still not a “silver bullet” combination of parameters that are widely applicable. Rather, this seems to be a problem to which solutions are either scope-, resolution- or data source-dependent. Recent works exploring observed rawinsonde data [55], atmospheric model forecasts [56,57], and reanalysis [15,58] outputs reached different optimal sets of best descriptors. These results make the use of techniques such as sensitivity analysis for feature selection [56,57], which

is almost a mandatory step for each activity related to FF forecasting due to the increasing volume of data and candidates. However, metrics associated with PW, such as absolute [56], anomaly [55], or spatial inhomogeneity [59] values, are consistently considered powerful predictors. These findings highlight the importance of developing and enhancing methods to take into consideration such meteorological factors into forecasting chains.

2.5. Remote Sensing Techniques

2.5.1. Precipitable Water (PW)

A monitoring approach of special interest for FF forecasting due to its short update time (around 15 min) is through the analysis of the travel time delays of communication signals from dense Global Navigation Satellite System (GNSS) networks, such as the Global Positioning System (GPS).

This concept has long been explored [60] and has led to the implementation of near real-time national monitoring systems by different countries, including the US, Germany, and China [61–63]. Those systems make use of their respective national networks to complement the existing international positioning stations managed as part of the International GNSS Service. A remarkable active research field targets the development of methods to integrate additional constellations of GNSS satellites (e.g., GLONASS, BeiDou and Galileo) towards the improvement of data resolution and error reduction [64,65]. However, it has to be clarified as to how the enhancements obtained from such integration can be reflected as enhancements towards the early identification of local convective systems.

Recent lines of research explore the gains and prediction power of GPS-based PW monitoring approaches for anticipating extreme precipitation through the proposal of threshold-based techniques for issuing FF warnings [59,66] and through the use of trend analysis as a complement to radiosonde observations [67]. Despite promising results that suggest that similar experiments should be developed for further locations, there is a consensus that PW-related values alone are not sufficient to act as a pluvial FF predictor [68].

2.5.2. Quantitative Precipitation Estimation (QPE)

Due to the high spatiotemporal variability of precipitation events usually associated with FFs,

several limitations emerge for the use of rain gauge data alone, mainly due to the low density of sensors deployed at most of the sites [69]. In this context, the use of weather radar and satellite products have been recognized for the purpose of performing FF forecasts due to their capability to describe rainfall fields of large areas.

Recent years have witnessed the consolidation of weather radar networks with national coverage in different countries [70–73]. Those systems are totally or partially based on C- or S-band with Doppler and dual-polarization technologies, usually generating observation products with temporal resolution in the order of 5 to 10 min and spatial discretization of 0.25 to 1 km. Many operational FFEWSs with extensive coverage of ungauged basins rely on mosaic QPE products derived from those radar systems, such as the multi-sensor precipitation estimate (MPE), the high-resolution precipitation estimator (HPE) [74,75], and the precipitation composites from the Operational Program on the Exchange of

Weather Radar Information (OPERA) [73]. However, some studies [76,77] indicate that radar products should present a temporal resolution of 1 min to be considered suitable for urban modeling. From that perspective, the increasing popularization of X-Band weather stations, with resolutions of up to 0.1 km/1 min, appears as the most promising advance for urban FF forecasting in the context of monitoring systems over the upcoming years [78–80]. Recently presented study cases have assessed the accuracy and positive impact of using dual-polarized X-band radar data as complements to the regional large-scale networks, mainly for dense urban areas [81–83], despite the known issues associated with the high noise and susceptibility to signal attenuation that demands careful attention.

Continuous research activities have been developed to estimate the propagation of weather radar-originated uncertainties, which are considered high for both mountainous [84] and urban environments [85], on the issuing of FF warnings [32,86]. The process of merging rain gauge and radar data in real-time is a continuously researched topic, and the choice of the approach used operationally may be influenced by multiple environmental factors, such as rain gauge density, rainfall event features, proximity to the radar station and temporal resolution of the products [87], and by the level of complexity of the techniques [88]. Methods proven to enhance radar data operationally at the hourly or sub-hourly scales required for FF forecasting include mean field bias (MFB) [89], kriging with external drift (KED) [90], and Bayesian combination (BAY) [91]. However, as suggested by Ochoa-Rodriguez et al. [88], due to the ever-growing volume of heterogeneous available data, the research field would benefit from more studies exploring data-driven methods and integration of multi-source, multi-resolution datasets.

Satellite-based observations can also be considered valuable for near real-time QPE due to their usual global coverage. Rainfall rate estimations based on passive microwave (PMW) data tend to be more accurate than their infrared (IR) -based counterparts but are generated with longer latency [92] and thus may provide limited support for FF forecasting. With the continuous increase of data-availability, multi-sourced products have emerged and been improved. The Global Hydro-Estimator (GHE) from the National Oceanic and Atmospheric Administration (NOAA) is an example of a multi-constellation, IR-based operational QPE product with 15–20 min latency time and spatial resolution in the order of 4 km [93]. It is used as the input for the FFG systems installed on poorly monitored countries [6]. The deployment of new satellites equipped with sensors capable to detect a wider range of IR spectral bands, such as the undergoing replacement of the Geostationary Operational Environmental Satellites (GOES) by GOES-R series, have the potential to increase the spatial discretization of the precipitation products by up to 2 km [94]. The precipitation estimation from remotely sensed information using artificial neural networks with a cloud classification system (PERSIANN-CCS) algorithm uses a multilayer feed-forward neural network to generate QPE from IR data [95]. It was recently implemented operationally, producing rainfall products with 1 h/0.04◦ (~4 km) resolution and 1 h latency using PMW calibrations [96]. Despite evidencing the potential for using neural network systems towards the generation of QPE products, recent evaluation works found that such products can contain considerable underestimation of precipitation estimates and suggestion caution on their current applicability for FF forecasting [97,98].

2.5.3. Surface Soil Moisture (SSM)

When SSM is estimated taking into consideration precipitation as the sole water input, the neglection of other potentially relevant processes such as irrigation may result in a decreased performance of FF forecasting systems [99] into forecasting chains. This motivated the search for approaches to assimilate SSM observations [100] and to assess their gains [101] towards runoff predictions. Non-urban headwater catchments, where SSM plays a most significant role in the generation of surface runoff, are usually insufficiently equipped with in situ soil monitoring equipment capable of the needed real-time transmission. An alternative is the use of satellite data, which usually have global coverage of the top layers of soil (up to 10 cm depth) and are made freely available by their respective spatial agencies. Passive microwave-based remote sensing products have been developed and assessed during the last decades with a proper agreement with local observations [102].

The last decade witnessed the launching of a considerable number of new satellite missions capable of performing near-surface soil moisture measurements (e.g., Soil Moisture Ocean Salinity (SMOS), Soil Moisture Active Passive (SMAP), Sentinel-1). Most of the currently active products from independent satellite constellations are provided either with coarse spatial resolutions and short revisiting time (in the order of 10 km/1 day) [103–105] or the opposite (e.g., 500 m/12-days [106]), depending on the swath width of the respective sensor. For FF forecasting systems based on hydrologic models, a shorter update time was found to play a more relevant role than finer spatial discretization [107]; however, daily updates can still be considered a significant operational constraint [101,108]. In this perspective, products based on the data blending of multiple satellites

missions, such as the Soil Moisture Operational Product System (SMOPS) [109], offer interesting opportunities to improve the accuracy of FF forecasts, with their potential to provide sub-daily resolution data, and their applicability for such purpose deserves to be assessed.

2.5.4. Drainage Network Monitoring and Controlling Systems

The data assimilation of the flow discharge observed in monitored drainage networks has the

potential of reducing the uncertainties of FF forecasting systems [110,111]. Due to the potential to transport damaging objects at high velocity during extreme events, the use of non-intrusive ultrasonic or radar sensors is preferred for flash flood-prone streams over their submersible counterparts [112]. Traditionally, only the higher magnitude channels are gauged by official agencies due to the high costs associated with the acquisition and maintenance of precise equipment. Such sparse observations may not provide valuable information for small-sized neighborhoods not served by a central discharge link or for headwaters catchments prone to FFs.

Recent technological advances have led to the development of low-cost electronic systems capable of transmitting significant volumes of information through the internet making use of now widely available Wi-Fi connections, which made the deployment of several flow monitoring sensors more feasible even for channels of lower magnitude. In this perspective, the emerging concept of the Internet of Things (IoT) can be seen as an alternative to traditional supervisory control and data acquisition (SCADA) systems due to the fact that representatives of the latter tend to be isolated platforms characterized by lower

levels of interoperability and scalability, while the former usually makes use of the expanding wireless availability and open communication protocols to achieve higher levels of flexibility. The IoT has been explored for the proposal and development of integrated systems capable of supporting multiple flood-related sensed data sources, with case studies presented mainly in the context of urban floods [113–115], some of them with the potential of performing autonomous decisions towards the optimization of flood-mitigating structures [115,116] in real-time.

The increased popularization of densely monitored drainage networks can thus enhance the efficiency of FF forecasting systems through the early identification of channels in overbank conditions and of sewer systems operating above their capacity, but the migration of currently implemented prototypes to effectively operational systems is yet to be assessed.

2.6. Precipitation Modeling and Prediction

In most cases, QPF products are the main inputs for FF forecasting chains. Methods for obtaining products with the high resolution required for FF forecasting are usually based on the downscaling of coarser numerical weather prediction (NWP) models outputs, on the temporal extrapolation of distributed remote sense observations (Figure 2-3), or on the integration of both approaches.

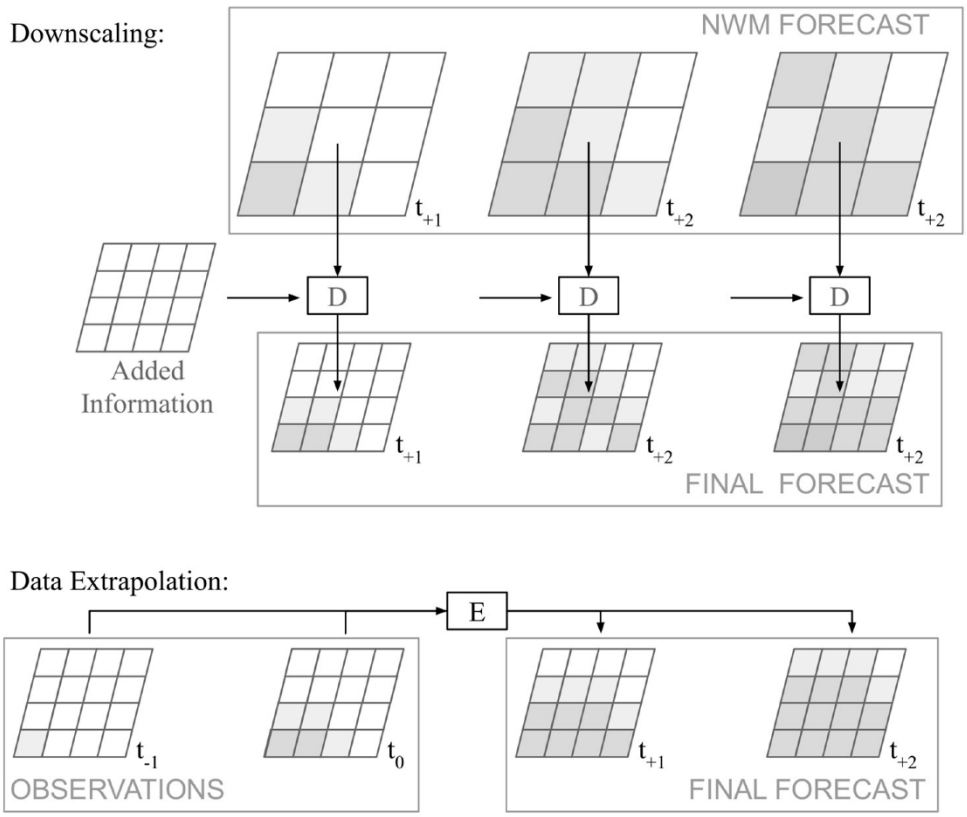

Figure 0-3. Schematic representation of downscaling (D) and data extrapolation (E) processes to obtain high-resolution precipitation products. The "Added Information" can be either sub-grid physics or statistical relationships.

2.6.1. Dynamical Downscaling

The first operational meteorological systems were limited to simulate synoptic-scale flows through hydrostatic primitive equations, in which sub-grid convective phenomena are represented indirectly using specific sets of parameters. Recent increases of overall computational power, data availability, and understanding of physical atmospheric processes allowed the consolidation of the so-called convection-permitting NWP (CP-NWP) models, which are based on hydrodynamic processes with a spatial resolution of approximately 4 km or higher, enabling the explicit representation of mesoscale events and

local convection [117–120]. An illustrative selection of currently CP-NWP models in operation is summarized in Table 2-3.

Table 0-3. Selected CP-NWP operational quantitative precipitation forecast (QPF) products sorted by spatial resolution.

Reference	Model (Product)	Agency	Coverage	Resolution	Update cycle
[117]	WRF (HRRR)	NOAA	US	1 h/3 km	1 hour
[118]	AROME (France)	Météo, France	France	1 h/1.3 km	1 hour
[120]	COSMO (DE)	DWD[1]	Germany	15 min/2.8 km	3 hours
[119]	HRDPS	MSC[2]	Canada	1 h/2.5 km	1 hour
[48]	WRF	CIMH[3]	Hispaniola	1 h/4 km	6 hours
[130]	ALARO (Turkey)	MGM[4]	Turkey	1 h/4.5 km	6 hours

[1] German Weather Service (DWD); [2] Meteorological Services of Canada (MSC); [3] Caribbean Institute for Meteorology and Hydrology (CIMH); [4] Turkish State Meteorological Service (MGM).

With the experience gained with operational CP-NWP systems, multiple works were developed to assess their advantages. It has been observed that, despite increasing the overall performance when compared to their background model products, some considerations need to be taken and addressed: (1) The finer spatial scale of CP-NWP models results in higher uncertainties at the grid-scale due to spatial noise, which demands the use of ensemble systems and assessment procedures that go beyond the pixel-to-pixel comparison [121,122]; (2) CP-NWP products tend to be positively biased when compared to products from synoptic-scale models, overestimating the magnitude of extreme precipitation events [123–125], thus demanding data assimilation procedures; and (3) large-scale convective events may be better represented by synoptic-scale models than by CP-NWP [124].

The potential of overestimating extreme rainfall events was, unsurprisingly, reflected as an increase in false alarms when the products of those models were applied to FF forecasting without intermediate processing [126,127], which reasserts the need of bias correcting these products when hydrologic models are involved in the FFEWS (RC-RC or FC approaches). Due to the complex physical basis of these methods, uncertainties are usually expected to be quantified by ensemble products [128]. However, the high computational cost demanded by the higher resolution models leads to a usually reduced number of realizations being available. The work developed by Corazza et al. [129] illustrates how the Poor Man's Ensemble (PME) approach can be used to address this issue by considering ensembles composed by deterministic products originated from multiple agencies and models. The authors obtained estimated probabilities that were well correlated to observations, but a certain level of underestimation detected was associated with the fact that members of both hydrostatic and CP-NWP models were included in the ensemble. The assessment of applying the PME approach using only CP-NWP model products is promising and yet to be developed.

2.6.2. Statistical Downscaling

One of the main advantages of using a statistical downscaling approach is the extremely low computational cost at the operational time when compared to dynamical downscaling. Statistical downscaling approaches are based on performing regression analysis between coinciding NWP-generated aerial estimations and gauge point observations. Methods successfully explored for obtaining precipitation time series with hourly temporal resolution include filtered autoregression [131], neural networks [132], and adaptable

random forests [133]. Due to their statistical nature, the outputs obtained from such methods are usually directly associated with the estimation of uncertainty of physics-based models [134,135]. Flood-PRObabilistic Operational Forecasting System (Flood-PROOFS) [51] is an example of a flood forecast system that includes the downscaling model RainFARM (acronym for Rainfall Filtered Autoregressive Model) to obtain ensemble QPF of 30 min/1 km resolution from a deterministic NWP model of 7 km/3 h. The estimated uncertainties were shown to acceptably represent the errors associated with the original QPE product and illustrate the potential gains of applying statistical downscaling approaches to perform FF forecasts. However, a remarkable drawback is the recurrent need for performing statistical reanalysis every time a component of the source large-scale NWP system is changed, which limits their adoption on operational forecasting chains.

2.6.3. Distributed Remote Sense Data Extrapolation

Extrapolating weather radar observations of precipitation in time is similar to the computer-vision problem of predicting the next frames of a recorded video. It can be performed by extrapolating the reflectivity values observed, which has output values that later need to be converted into precipitation (radar echo extrapolation, REE) or by directly producing QPF as output values out of the observed reflectivity, thus implicitly embedding the so-called Z–R relationship.

Motion tracking functions based on optical flow techniques have long been explored for REE. Examples of operating systems include the use of variational echo tracking [136,137] and combinations of the traditional Horn and Schunck approach with the Lucas–Kanade method [138].

The consolidation of extensive weather radar datasets allowed the development of data-driven techniques based on analog-based approaches, with promising results being obtained for locations with high orographic influence in the formation of precipitation [7,139]. Methods based on deep learning, mainly exploring the capabilities of convolutional neural networks (CNNs), have started to be explored in the last 5 years [140] and were proposed as benchmarks for precipitation nowcast methods [141]. The integration of long short-term memory (LSTM) neural network approaches with the satellite-based precipitation estimation algorithm PERSIANN-CCS was also shown to outperform optical flow-based and NWP models, mainly for capturing the patterns of convective precipitation systems [8]. However, as this is a still-emerging field, some relatively basic issues associated with the implementation of deep learning algorithms for the task of frame prediction, such as which assessment metric is the best to be used [142], are still under discussion in the community.

2.6.4. Multi-Model

The assessment that data extrapolation methods tend to overperform NWP models for lead times of up to 2 or 3 h (and that NWP models are more reliable for longer horizons) [143,144] motivated the exploration techniques for integrating both types of precipitation prediction products.

In the United Kingdom (UK), a dual-system approach has been adopted. By default, the stratiform-focused Nimrod [145], based on Lagrangian persistence extrapolation of radar data, is continuously executed and evaluated. When convective patterns are identified, the Generating Advanced Nowcasts for Deployment in Operational Land-based

Flood forecasts (GANDOLF) [146] system, which is based on object tracking considering NWM estimated advection, is activated. Such a strategy is justified by the best individual scenario-specific performances of each model [147,148].

Recently proposed approaches with an operational adoption include Integrated Nowcasting through Comprehensive Analysis (INCA) [149–151], in which the weighting between two deterministic nowcast/forecast models is constant and dependent on the lead time. In INCA, only radar extrapolation is considered from 0 to 2 h lead time; from 3 to 6 h, both the radar extrapolation and the NWP-based products are considered, with linearly increasing of importance (weighting) of the latter with respect to the former; for 7 h onwards, only the NWP-based product is considered.

The Short-Term Ensemble Prediction System (STEPS) [152] is a probabilistic blending approach, in which uncertainties from multiple scales and sources are considered in a fractal cascade for the generation of a dynamically weighted ensemble product. The uncertainty level of each component is calculated taking into consideration the climatological analysis of the forecasted value. The method was adopted operationally in the UK and Australia in 2008 [153] and was successfully assessed for urban hydrology [154,155].

Works exploring new blending techniques are undergoing, with the proposition of approaches based on volume-correction [156] and harmony search adaptive weighting [157], for example. However, the applicability of such methods for operational FF forecasting systems requires further assessment.

2.7. Hydrologic-Hydraulic Modeling and Forecasting

2.7.1. Runoff Simulation

Mostly for non-urban, ungauged and/or data-scarce catchments, lumped models were the first to be explored and adopted operationally due to their simplicity, low level of data requirement and computational demand. Sacramento soil moisture accounting (SAC-SMA), for example, was widely used in the early versions of FFG systems in the US and is still used as part of large-domain systems with online simulations [48,54].

In the last decade, distributed rainfall-runoff and routing models were adopted operationally. Multiple river forecasting centers in the US replaced their SAC-SMA-based models with Hydrology Laboratory Research Distributed Hydrologic Model (HL-RDHM)-based counterparts. In Europe, the LISFLOOD model is used in the operational FC-based approach ERIC, which is adopted as part of EFAS [158]. Those systems operate on spatial scales in the order of 1 km to match the QPE and QPF forcings involved in the forecasting chains. However, the high number of parameters demanded by a physics-based approach may become a constraint due to the need for extensive, often unavailable, descriptive datasets or challenging calibration procedures that may result in high levels of uncertainty [159]. HL-RDHM, for example, is reported to require the calibration of 15 parameters per grid cell [160].

New models have been proposed using more hybrid conceptual-physical-based approaches to obtain more parsimonious representations of the hydrological processes. Such models intensely rely on topographic features to reduce the number of parameters

used to describe the runoff and routing processes [161] and have shown to be also suitable for sub-kilometer simulations of flash flood events [162]. Operationally applied examples include the Continuum [163] and the Coupled Routing and Excess Storage (CREST) [164] models, part of the Flood-PROOFS [43] and FLASH [54] systems that require the calibration of 6 and 10 parameters per grid cell, respectively.

Arid and semi-arid regions are characterized by having more dynamic rainfall-runoff responses due to the lack of vegetation coverage and organic matter, which makes the runoff generation more dependent on the varying SSM conditions. In such regions, data assimilation of SSM can be particularly beneficial when compared to humid environments [165]. The performance of the recently presented hybrid conceptual-physical models over these specific conditions of parameter variability is yet to be assessed.

In urban areas, floods are usually initiated when the drainage systems operate above their capacity. Hydrological models can be coupled with hydraulic models so that the runoff estimated as the output by the former is used as the input flow by the latter [166]. The Storm Water Management Model (SWMM) [167], MIKE URBAN [168] and Infoworks CS are examples of established frameworks with one-dimensional (1D) components for representing hydraulic systems such as sewer networks and the presence of low-impact-development (LID) structures. The use of 1D drainage models alone is considered not suitable for representing the overflow phenomenon, but they can be used to estimate the locations of their occurrence through the identification of manholes in the overflow state, for example [169].

Hydraulic sewer models can operate at spatiotemporal resolutions in the order of centimeters/sub-minutes, which results in the need for high data availability and computational resources for online simulations. Few locations have proper and available documentation of the installed sub-surface drainage systems for an accurate model development, and the use of synthetic sewer networks derived from digital elevation models (DEMs) and structural analysis have been proposed [170] and were able to acceptably estimate the location of the underground pipes, but proper dimensioning of the diameters of the conduits is still a challenging question that may limit the applicability of such approaches for modeling extreme precipitation events [171].

Regardless of the purpose, the quantification of uncertainty associated with the model output is exceptionally being valued, which is usually achieved through the use of ensemble forecasting chains. In the context of FF forecasting, attention has been given to assess the uncertainty related to the use of precipitation nowcasts [172,173] and to scenario-specific error patterns associated with seasonality [174].

Due to the assumption that floods happen after water accumulates in the serving drainage channel, channel discharge flow modeling may not properly represent complex scenarios in which local surface ponding is observed, and flooding conditions are not necessarily associated with a channel state. Additionally, the absence of associated inundation mapping data of systems based on flow-communication only may limit the responsive steps that succeed in the forecasting phase of FFEWSs.

2.7.2. Flood Inundation Simulation

Accurate flood inundation maps are extremely valuable products as they have the potential to direct response efforts during emergency scenarios, particularly for dense urban areas.

Physics-based models used for the surface flow may be classified considering the spatial dimensionality supported. Two-dimensional (2D) models represent only the surface flows, thus abstracting the behavior of potentially existing drainage systems and relying on the high-quality representation of the surface through high-resolution elevation and land cover classification maps. In order to include the influence of well-described sewer networks, coupling approaches are continuously explored to integrate 1D hydraulic models. In 1D–2D modeling, the two-dimensional surface water spread is linked to the linear pipe and channel flows through connection points, such as manholes and culverts. Surface structures such as streets can also be represented as channels with linear water flows so that overland components can be simplified as 1D formulas, leading to the so-called 1D–1D coupling approaches [169]. In this context, LISFLOOD-FP [175] is a remarkable example of a widely adopted 2D model that has been assessed in urban environments at high-resolution simulations for scenarios of drainage systems overflow [176].

FF-related inundations are characterized by their high dynamicity and complexity, which demands proper representation of momentum conservation. However, solving the full 2D-forms of the Saint–Venant equation (SVE), mainly for the high spatial resolutions required in urban environments, tend to become so computationally expensive that real-

time online simulations are considered unfeasible [169]. Several methods have been successfully explored to reduce the total simulation time of surface flow models with acceptable accuracy loss based on meta-modeling (or model surrogating). Physically-simplified surrogated models are based on the suppression of acceleration components of the SVE and thus are only suited for floods driven by slow flows [177,178]. The few data-driven surrogate inundation models proposed so far with adequate performance for representing spatial resolution in the order of meters were based on feed-forward neural networks and present promising results [179,180]. So far, those models have only used the simulated output maps of water depth as the training target, thus the dynamic components (such as flow velocity) associated to the flood inundation simulation have only been considered implicitly as part of the black-box training process. Their explicit consideration should be explored as a way to increase the replicability of high inertial flows, a typical feature for FFs.

The cellular automata (CA) concept explores simplified, parallelizable grid-based operations for solving field-propagation problems. In the last decade, it has been applied both for simulating sewer network flow [181] and for rapid flood inundation modeling using the Manning equation [182,183] or simplified topographic-driven water spread [184] as part of the routing process. Recent works have been developed to propose 1D–2D coupling approaches for a more realistic representation of the drainage system [185] and to improve parallelization capabilities on complex networks [186] with remarkable gains in computational time. However, the CA-based methods are still unable to replicate

inundation events characterized by high inertia due to the usual neglection of momentum conservation, which can limit their applicability towards FF forecasting.

Considering that most of the physics-based inundation models assume that flood happens due to the overflows of a drainage system, which is not a necessary condition for urban environments, a new approach based on the surface water spread of instantaneous runoff generation at impervious regions was recently proposed [12]. Despite presenting promising results, the performance of the model has not been explored for scenarios of a spatial resolution higher than 50 m and its computational costs are yet to be addressed for its potential online applicability in FFEWSs.

Advances in the adoption of LiDAR technology for DEM production [187] motivated recent studies evaluating the adoption of high-resolution surface representation for urban flood inundation simulations. The positive impact of using 10 cm resolution topographic data over their respective 1 m resolution counterpart to properly represent relevant features such as curbs was assessed, but the computational cost associated with such a detailed representation may be considered a limitation for real-time operational purposes [188,189]. The use of such a hyper-resolution dataset may require special processing so that relevant "hidden" water passways are properly represented in the DEM. For such, structure-from-motion (SfM) techniques, which are based on photogrammetric and computer vision interpretation of overlapping photos, have showed promising results as supplementary sources of information for LiDAR [190] and for regular ground-based field surveys [191].

There is an interest in accounting for and communicating uncertainties of flood inundation maps [192]. However, the computational cost associated with the forecast runs

of multiple model realizations results in unfeasible real-time operational applicability of the traditional ensemble approach. As suggested by Teng et al. [178], Gaussian processes and polynomial chaos emulation are appealing approaches to be explored in this context; however, to the best of the authors' knowledge, no work has been developed on such a topic.

2.7.3. Query-Based Approaches

A family of approaches focused on making use of the high-quality simulation products obtained from full physical models without facing their expensive computational requirements in real-time is based on pre-simulating (offline) flood events caused by several different "what-if" feasible precipitation forcing scenarios. Pairs of input/output values used/obtained are stored in some sort of database schema so that, during operational time, the predicted flooding conditions (outputs) can be retrieved almost instantly from observed/forecasted conditions by a similarity search with the indexed inputs (querying) in the assumption of constant cause/consequence relationships (Figure 2-4). In this context, multiple combinations of methods used to represent, store, index, and query procedures have been explored.

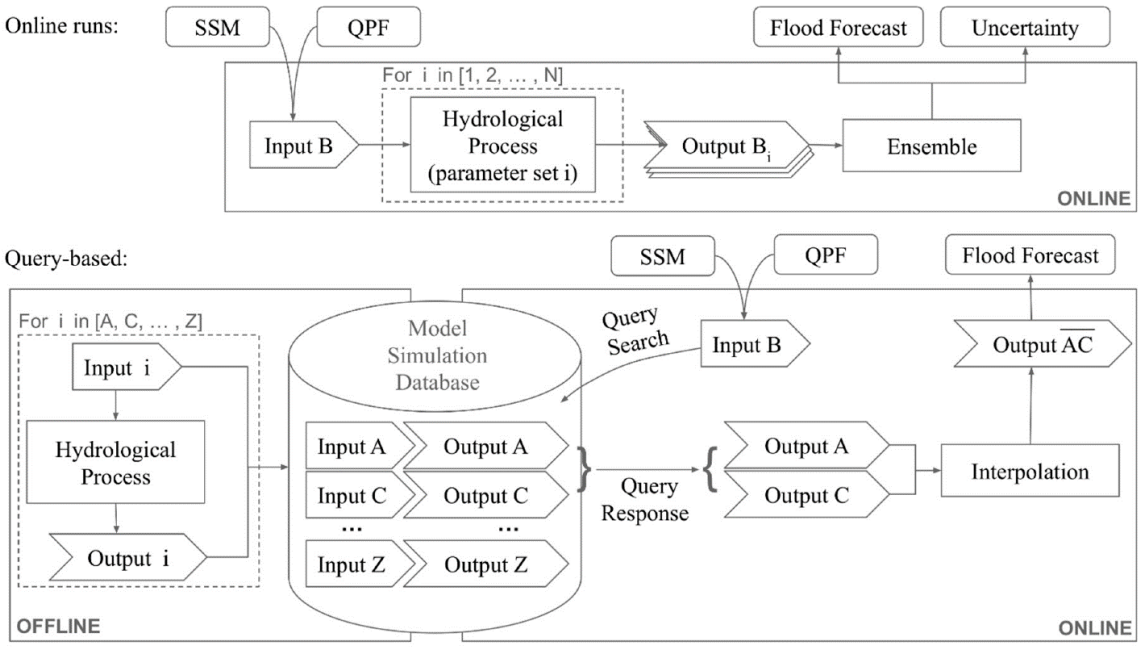

Figure 0-4. Comparison diagram of hypothetical online-run and query-based systems, both driven by SSM and QPF values.

One example of such an approach was presented by Song et al. [11]. The authors used the observed water depth at two locations of a channel crossing an urban area to correlate, using a 2D matrix, the Thresh-R needed to increase the water level in the systems up to a flood-triggering threshold value. The Thresh-R was then mapped backward, based on pre-simulated scenarios, into its minimum total generation precipitation intensity (P_{min}). P_{min} was finally compared with QPF values for warning issuing purposes.

Other works use a combination of self-organizing maps and recurrent neural networks to identify the most recurrent flood inundation map features simulated, thus reducing the total number of maps stored. Optimistic results were obtained for resolutions of 3 h/75 m [193] and 1 h/5 m [194] and included the modeling of a dense sewer network.

In a query-based approach, the limited number of pre-simulated scenarios does not cover all possible observable combinations of input, thus the adoption of approximation or

interpolation procedures are needed. The quantification of the uncertainties derived from such an interpolation is a problem yet to be addressed.

2.8. Discussion and Summary

Forecasting pluvial flash floods with sufficient accuracy and lead time to support effective response actions is a challenging hydrometeorological problem that involves multiple scientific and technological fields. This work presents a non-exhaustive overview of some of the main documented operational flash flood early warning systems, of advances observed in the related topics during the last years, and of opportunities for further development.

Different criteria can be used for the prediction of a flash flood scenario. Existing operational systems may take into account meteorological patterns known to precede extreme precipitation events and apply rainfall or runoff threshold comparisons over forecasted values to identify upcoming flood events. The choice of the "best" criteria to be adopted is mainly driven by the area covered by the forecasting team and by the available resources.

For systems with national coverage in which weather radar precipitation mosaic products are available, runoff-threshold exceedance usually performs the best as the decision criterion. In this context, distributed hydrological models based on mixed conceptual-physical representations of the rainfall-runoff/routing processes are being preferred to their counterparts based in physics due to the reduced number of parameters to be calibrated. When radar coverage is insufficient, systems are being set up using

satellite-derived rainfall data derived from infrared radiation measurements and using rainfall-threshold exceedance with the consideration of antecedent soil moisture conditions as warning criteria.

When a system is designed for a region or a catchment known to be dominated by flash floods, there is little evidence to justify the inclusion of a rainfall-runoff/routing model in a forecasting chain. In this context, decisions made based solely on the exceedance of a regionalized rainfall threshold may present sufficient accuracy and appropriate timely response. Raw rainfall threshold exceedance has also been shown to be a particularly efficient criterion for issuing early warnings for urban environments but, in that case, the critical values are location-specific and require proper documentation at neighborhood level of past events to be defined.

Some enhancements on deployed monitoring capabilities with the potential to enhance flash flood forecasting activities deserve to be highlighted:

- The recent expansion of global navigation satellite systems (GNSSs) allowed the development of promising multi-constellation approaches for retrieving more accurate real-time estimations of precipitable water vapor, a meteorological factor determinant on the occurrence of convective storms. However, it is not clear yet how this improvement can be reflected in our capability to anticipate precipitation events that can result in flash floods, and the research field would benefit from study cases developed in several parts of the globe to access such regional gains.
- The blending of the observations performed by recently launched satellite missions directed to monitor surface soil moisture allows products such as NOAA's SMOPS

to be produced with an update time of 6 h or less, a remarkable gain to be explored towards the estimation of antecedent soil moisture conditions when compared to the usual constraining daily revisit interval characteristic of individual missions.

- Recently presented study cases making use of dual-polarized X-band radar stations over urban areas assessed the use of such higher-resolution, short-range equipment as a positive contribution to the already established C- and S-band wide range networks. Further deployments are seen as an essential step towards accurate flash flood monitoring in urban areas, and the use of radar extrapolation techniques for precipitation nowcasting at such a finer scale is the expected direction of further research.
- The concept of the Internet of Things, based on the reduction of the production cost of autonomous sensors capable of communicating in real-time through widely available wireless networks, has the potential to increase our deployed capability to monitor small urban channels and drainage pipes. Such dense measurement may enhance hydrological forecasts through data assimilation, but few locations have sensor networks of such type already installed and few are under study.

The main input product and the primary source of uncertainties for most of the scenarios is still the forecasted extreme precipitation products. During the last decade, we have observed the consolidation and broad operational adoption of numerical weather models with spatiotemporal resolution high enough to explicitly represent local convection. However, such higher resolution is attained with increased computational cost and spatial noise. The use of multi-model approaches, such as the Poor Man's Ensemble,

applied to a set composed only by convection-permitting numeric weather models should be further considered for generating uncertainty estimations since more products of this type are expected to be made available operationally in the near future.

However, for precipitation nowcast the extrapolation of remotely sensed rainfall fields is still considered the most accurate family of approaches. Here, novel deep learning-based techniques have been proposed as options to the traditional Lagrangian methods with promising results but, as a novel research field, some core questions are still under discussion, including which training and validating metrics should be used by the community and the size of training datasets needed for proper application of such techniques.

Flood inundation forecasts are precious products for decision-makers, but their consideration during flash floods is usually neglected due to the usually high computational time associated with the execution of traditional physics-based hydraulic models. Approaches based on data-driven surrogate modeling using machine learning, on query-based and on cellular automata have been proposed for the timely generation or retrieval of inundation maps. Despite being successfully presented in study cases with spatiotemporal resolution suitable for urban environments, further assessment is needed to validate the capability of such methods to represent the influence of high momentum flows, a characteristic of flash floods.

2.9. Conclusions

In the last decade, the field of flash flood forecasting has benefitted from a significant increase in overall data availability due to continuous improvements in the coverage and in spatio-temporal resolutions of monitoring and modeling systems. However, the data of such recently deployed systems is, in many cases, limited in terms of archived volume and length for proper calibration and for the adoption of data-driven modeling, a limitation that is less restrictive on older, coarser, and still well-maintained systems. The perspective presented in this work suggests that the exploration of techniques for blending multiresolution, multisource data in real-time is a major trend to benefit both the operational and research community.

The continuous deployment of early warning systems resulted in scenarios in which multiple forecasting techniques, sometimes even based on different criteria, were being executed in parallel towards the prediction of the same observed flash flood events. Such approaches facilitated the intercomparison of the performances of different methods under operational environments, and the findings suggest that systems covering large domains tend to deliver better performance when hydrological models are part of the forecast chain, while systems designed for restrict domains considered particularly flashy may present comparable efficiency by only relying on rainfall-exceedance criteria. This observation is based on a very restricted number of published comparative papers, which also illustrates the main limitation of this work, i.e., the limited number of existing FFEWSs properly and publicly documented. From that perspective, the authors highlight the importance of

comparing and documenting multiple different forecasting approaches on the implementation of FFEWSs.

2.10. References

1. Llasat, M.C.; Marcos, R.; Turco, M.; Gilabert, J.; Llasat-Botija, M. Trends in flash flood events versus convective precipitation in the Mediterranean region: The case of Catalonia. *J. Hydrol.* **2016**, *541*, 24–37.
2. 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.
3. Shanableh, A.; Al-Ruzouq, R.; Yilmaz, A.G.; Siddique, M.; Merabtene, T.; Imteaz, M.A. Effects of land cover change on urban floods and rainwater harvesting: A case study in Sharjah, UAE. *Water* **2018**, *10*.
4. Hapuarachchi, H.A.P.; Wang, Q.J.; Pagano, T.C. A review of advances in flash flood forecasting. *Hydrol. Process.* **2011**, *25*, 2771–2784.
5. National Weather Service NWS Glossary Available online: https://w1.weather.gov/glossary/.
6. Modrick, T.M.; Graham, R.; Shamir, E.; Jubach, R.; Spencer, C.R.; Sperfslage, J.A.; Georgakakos, K.P. Operational flash flood warning systems with global applicability. In Proceedings of the 7th International Congress on Environmental Modelling and Software: Bold Visions for Environmental Modeling (iEMSs 2014); San Diego, CA, USA, 2014; pp. 694–701.

7. Panziera, L.; Germann, U.; Gabella, M.; Mandapaka, P. V. NORA-Nowcasting of Orographic Rainfall by means of analogues. *Q. J. R. Meteorol. Soc.* **2011**, *137*, 2106–2123.

8. Akbari Asanjan, A.; Yang, T.; Hsu, K.; Sorooshian, S.; Lin, J.; Peng, Q. Short-term precipitation forecast based on the PERSIANN system and LSTM recurrent neural networks. *J. Geophys. Res. Atmos.* **2018**, *123*, 12,543-12,563.

9. Maddox, R.A.; Chappell, C.F.; Hoxit, L.R. Synoptic and meso-α scale aspects of flash flood events. *Bull. Am. Meteorol. Soc.* **1979**, *60*, 155–176.

10. Borga, M.; Stoffel, M.; Marchi, L.; Marra, F.; Jakob, M. Hydrogeomorphic response to extreme rainfall in headwater systems: Flash floods and debris flows. *J. Hydrol.* **2014**, *518*, 194–205.

11. Song, Y.; Park, Y.; Lee, J.; Park, M.; Song, Y. Flood forecasting and warning system structures: Procedure and application to a small urban stream in South Korea. *Water* **2019**, *11*.

12. Wang, X.; Kinsland, G.; Poudel, D.; Fenech, A. Urban flood prediction under heavy precipitation. *J. Hydrol.* **2019**, *577*.

13. 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; AMS: Portland, OR, USA, 2016.

14. Doswell, C.A.; Brooks, H.E.; Maddox, R.A. Flash flood forecasting: An ingredients-based methodology. *Weather Forecast.* **1996**, *11*, 560–581.

15. Jessup, S.M.; DeGaetano, A.T. A statistical comparison of the properties of flash

flooding and nonflooding precipitation events in portions of New York and Pennsylvania. *Weather Forecast.* **2008**, *23*, 114–130.

16. Collier, C.G.; Fox, N.I. Assessing the flooding susceptibility of river catchments to extreme rainfall in the United Kingdom. *Int. J. River Basin Manag.* **2003**, *1*, 225–235.
17. George, J.J. Weather forecasting for aeronautics. *Q. J. R. Meteorol. Soc.* **1961**, *87*.
18. Murray, S.J.; Smith, A.D.; Phillips, J.C. A modified flood severity assessment for enhanced decision support: Application to the Boscastle flash flood of 2004. *Weather Forecast.* **2012**, *27*, 1290–1297.
19. Llort, X.; Sánchez-Diezma, R.; Rodríguez, Á.; Sancho, D.; Berenguer, M.; Sempere-Torres, D. FloodAlert: a simplified radar-based EWS for urban flood warning. In Proceedings of the 11th International Conference on Hydroinformatics; IAHR-IWA, 2014.
20. Alfieri, L.; Velasco, D.; Thielen, J. Flash flood detection through a multi-stage probabilistic warning system for heavy precipitation events. *Adv. Geosci.* **2011**, *29*, 69–75.
21. Alfieri, L.; Thielen, J. A European precipitation index for extreme rain-storm and flash flood early warning. *Meteorol. Appl.* **2015**, *22*, 3–13.
22. Versini, P.-A.; Berenguer, M.; Corral, C.; Sempere-Torres, D. An operational flood warning system for poorly gauged basins: demonstration in the Guadalhorce basin (Spain). *Nat. Hazards* **2014**, *71*, 1355–1378.
23. Park, S.; Berenguer, M.; Sempere-Torres, D. Long-term analysis of gauge-adjusted

radar rainfall accumulations at European scale. *J. Hydrol.* **2019**, *573*, 768–777.

24. Bezak, N.; Šraj, M.; Mikoš, M. Copula-based IDF curves and empirical rainfall thresholds for flash floods and rainfall-induced landslides. *J. Hydrol.* **2016**, *541*, 272–284.
25. Papagiannaki, K.; Lagouvardos, K.; Kotroni, V.; Bezes, A. Flash flood occurrence and relation to the rainfall hazard in a highly urbanized area. *Nat. Hazards Earth Syst. Sci.* **2015**, *15*, 1859–1871.
26. Jang, J.-H. An advanced method to apply multiple rainfall thresholds for urban flood warnings. *Water* **2015**, *7*, 6056–6078.
27. Zhai, X.; Guo, L.; Liu, R.; Zhang, Y. Rainfall threshold determination for flash flood warning in mountainous catchments with consideration of antecedent soil moisture and rainfall pattern. *Nat. Hazards* **2018**, *94*, 605–625.
28. Norbiato, D.; Borga, M.; Dinale, R. Flash flood warning in ungauged basins by use of the Flash Flood Guidance and model-based runoff thresholds. *Meteorol. Appl.* **2009**, *16*, 65–75.
29. Schmidt, J.A.; Anderson, A.J.; Paul, J.H. Spatially-variable, physically-derived, flash flood guidance. In Proceedings of the 21st Conference on Hydrology; AMS: San Antonio, TX, USA, 2007.
30. Gourley, J.J.; Erlingis, J.M.; Hong, Y.; Wells, E.B. Evaluation of tools used for monitoring and forecasting flash floods in the united states. *Weather Forecast.* **2012**, *27*, 158–173.
31. Ntelekos, A.A.; Georgakakos, K.P.; Krajewski, W.F. On the uncertainties of Flash

Flood Guidance: Toward probabilistic forecasting of flash floods. *J. Hydrometeorol.* **2006**, *7*, 896–915.

32. Villarini, G.; Krajewski, W.F.; Ntelekos, A.A.; Georgakakos, K.P.; Smith, J.A. Towards probabilistic forecasting of flash floods: The combined effects of uncertainty in radar-rainfall and flash flood guidance. *J. Hydrol.* **2010**, *394*, 275–284.
33. Lee, B.J.; Kim, S. Gridded flash flood risk index coupling statistical approaches and TOPLATS land surface model for mountainous areas. *Water* **2019**, *11*.
34. Ghadua, Z.; Bhattacharya, B. Improving flash flood forecasting with a Bayesian probabilistic approach: A case study on the Posina Basin in Italy. *Int. J. Environ. Ecol. Eng.* **2019**, *13*, 331–337.
35. Martina, M.L.V.; Todini, E.; Libralon, A. A Bayesian decision approach to rainfall thresholds based flood warning. *Hydrol. Earth Syst. Sci.* **2006**, *10*, 413–426.
36. Montesarchio, V.; Ridolfi, E.; Russo, F.; Napolitano, F. Rainfall threshold definition using an entropy decision approach and radar data. *Nat. Hazards Earth Syst. Sci.* **2011**, *11*, 2061–2074.
37. Reed, S.; Schaake, J.; Zhang, Z. A distributed hydrologic model and threshold frequency-based method for flash flood forecasting at ungauged locations. *J. Hydrol.* **2007**, *337*, 402–420.
38. Raynaud, D.; Thielen, J.; Salamon, P.; Burek, P.; Anquetin, S.; Alfieri, L. A dynamic runoff co-efficient to improve flash flood early warning in Europe: Evaluation on the 2013 central European floods in Germany. *Meteorol. Appl.* **2015**, *22*, 410–418.

39. Javelle, P.; Demargne, J.; Defrance, D.; Pansu, J.; Arnaud, P. Evaluating flash-flood warnings at ungauged locations using post-event surveys: a case study with the AIGA warning system. *Hydrol. Sci. J.* **2014**, *59*, 1390–1402.

40. Javelle, P.; Organde, D.; Demargne, J.; Saint-Martin, C.; de Saint-Aubin, C.; Garandeau, L.; Janet, B. Setting up a French national flash flood warning system for ungauged catchments based on the AIGA method. In Proceedings of the 3rd European Conference on Flood Risk Management; Lyon, France, 2016.

41. Gourley, J.J.; Flamig, Z.L.; Hong, Y.; Howard, K.W. Evaluation of past, present and future tools for radar-based flash-flood prediction in the USA. *Hydrol. Sci. J.* **2014**, *59*, 1377–1389.

42. Liu, C.; Guo, L.; Ye, L.; Zhang, S.; Zhao, Y.; Song, T. A review of advances in China's flash flood early-warning system. *Nat. Hazards* **2018**, *92*, 619–634.

43. Corral, C.; Berenguer, M.; Sempere-Torres, D.; Poletti, L.; Silvestro, F.; Rebora, N. Comparison of two early warning systems for regional flash flood hazard forecasting. *J. Hydrol.* **2019**, *572*, 603–619.

44. Lincoln, S. Analysis of the 15 June 2013 isolated extreme rainfall event in Springfield, Missouri. *J. Oper. Meteorol.* **2014**, *2*, 233–245.

45. Smith, P.J.; Pappenberger, F.; Wetterhall, F.; Thielen Del Pozo, J.; Krzeminski, B.; Salamon, P.; Muraro, D.; Kalas, M.; Baugh, C. On the operational implementation of the European Flood Awareness System (EFAS). In *Flood Forecasting - A Global Perspective*; Academic Press: Reading, UK, 2016; pp. 313–348 ISBN 9780128018842.

46. Park, S.; Berenguer, M.; Sempere-torres, D.; Baugh, C.; Smith, P. Toward seamless high-resolution flash flood forecasting over Europe based on radar nowcasting and NWP: An evaluation with case studies. In Proceedings of the EGU General Assembly; EGU: Vienna, Austria, 2017.

47. Ulupinar, Y.; Akbas, A.I.; Gulsoy, E.; Celik, S.; Kose, S.; Aksoy, M. *Black Sea and Middle East Flash Flood Guidance System - User guide*; Turkish Meteorological Service: Ankara, Turkey, 2015;

48. Shamir, E.; Georgakakos, K.P.; Spencer, C.; Modrick, T.M.; Murphy, M.J.; Jubach, R. Evaluation of real-time flash flood forecasts for Haiti during the passage of Hurricane Tomas, November 4-6, 2010. *Nat. Hazards* **2013**, *67*, 459–482.

49. Cosgrove, B.A.; Clark, E.; Reed, S.; Koren, V.; Zhang, Z.; Cui, Z.; Smith, M. *Overview and initial evaluation of the Distributed Hydrologic Model Threshold Frequency (DHM-TF) flash flood forecasting system*; US Department of Commerce: Silver Spring, MD, USA, 2012;

50. Demargne, J.; Javelle, P.; Organde, D.; Fouchier, C.; Janet, B. Enhancements of the French operational flash flood warning system, Vigicrues Flash. In Proceedings of the EGU General Assembly; EGU: Vienna, Austria, 2019.

51. Laiolo, P.; Gabellani, S.; Rebora, N.; Rudari, R.; Ferraris, L.; Ratto, S.; Stevenin, H.; Cauduro, M. Validation of the Flood-PROOFS probabilistic forecasting system. *Hydrol. Process.* **2014**, *28*, 3466–3481.

52. Montesarchio, V.; Napolitano, F.; Rianna, M.; Ridolfi, E.; Russo, F.; Sebastianelli, S. Comparison of methodologies for flood rainfall thresholds estimation. *Nat.*

Hazards **2015**, *75*, 909–934.

53. Fouchier, C.; Mériaux, P.; Atger, F.; Ecrepont, S.; Liébault, F.; Bertrand, M.; Batista, D.; Azemard, P. Implementation of a real-time warning and mapping system for natural hazards triggered by rainfall in mountainous and Mediterranean areas of Southeastern France. In Proceedings of the 10th International Workshop on Precipitation in Urban Areas (UrbanRain15); ETHZ: Pontresina, Switzerland, 2015.
54. Gourley, J.J.; Flamig, Z.L.; Vergara, H.; Kirstetter, P.E.; Clark, R.A.; Argyle, E.; Arthur, A.; Martinaitis, S.; Terti, G.; Erlingis, J.M.; et al. The FLASH project - Improving the tools for flash flood monitoring and prediction across the United States. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 361–372.
55. Schroeder, A.; Basara, J.; Marshall Shepherd, J.; Nelson, S. Insights into atmospheric contributors to urban flash flooding across the United States using an analysis of rawinsonde data and associated calculated parameters. *J. Appl. Meteorol. Climatol.* **2016**, *55*, 313–323.
56. Clark III, R.A. Machine Learning Predictions of Flash Floods. Ph.D., School of Meteorology, University of Oklahoma: Norman, OK, USA, 2016.
57. Herman, G.R.; Schumacher, R.S. Money doesn't grow on trees, but forecasts do: Forecasting extreme precipitation with random forests. *Mon. Weather Rev.* **2018**, *146*, 1571–1600.
58. Turkington, T.; Ettema, J.; Van Westen, C.J.; Breinl, K. Empirical atmospheric thresholds for debris flows and flash floods in the southern French Alps. *Nat. Hazards Earth Syst. Sci.* **2014**, *14*, 1517–1530.

59. Shoji, Y. Retrieval of water vapor inhomogeneity using the japanese nationwide GPS array and its potential for prediction of convective precipitation. *J. Meteorol. Soc. Japan* **2013**, *91*, 43–62.

60. Bevis, M.; Businger, S.; Herring, T.A.; Rocken, C.; Anthes, R.A.; Ware, R.H. GPS meteorology: remote sensing of atmospheric water vapor using the Global Positioning System. *J. Geophys. Res.* **1992**, *97*, 787–801.

61. Ware, R.H.; Fulker, D.W.; Stein, S.A.; Anderson, D.N.; Avery, S.K.; Clark, R.D.; Droegemeier, K.K.; Kuettner, J.P.; Minster, J.B.; Sorooshian, S. Suominet: A real-time national GPS network for atmospheric research and education. *Bull. Am. Meteorol. Soc.* **2000**, *81*, 677–694.

62. Gendt, G.; Dick, G.; Reigber, C.; Tomassini, M.; Liu, Y.; Ramatschi, M. Near real time GPS water vapor monitoring for numerical weather prediction in Germany. *J. Meteorol. Soc. Japan* **2004**, *82*, 361–370.

63. Zhang, H.; Yuan, Y.; Li, W.; Zhang, B. A real-time precipitable water vapor monitoring system using the national GNSS network of China: Method and preliminary results. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 1587–1598.

64. Li, X.; Zus, F.; Lu, C.; Dick, G.; Ning, T.; Ge, M.; Wickert, J.; Schuh, H. Retrieving of atmospheric parameters from multi-GNSS in real time: Validation with water vapor radiometer and numerical weather model. *J. Geophys. Res. Atmos.* **2015**, *120*, 7189–7204.

65. Wang, J.; Wu, Z.; Semmling, M.; Zus, F.; Gerland, S.; Ramatschi, M.; Ge, M.;

Wickert, J.; Schuh, H. Retrieving precipitable water vapor from shipborne multi-GNSS observations. *Geophys. Res. Lett.* **2019**, *46*, 5000–5008.

66. Shi, J.; Xu, C.; Guo, J.; Gao, Y. Real-Time GPS precise point positioning-based precipitable water vapor estimation for rainfall monitoring and forecasting. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 3452–3459.
67. Moore, A.W.; Small, I.J.; Gutman, S.I.; Bock, Y.; Dumas, J.L.; Fang, P.; Haase, J.S.; Jackson, M.E.; Laber, J.L. National Weather Service forecasters use GPS precipitable water vapor for enhanced situational awareness during the Southern California Summer monsoon. *Bull. Am. Meteorol. Soc.* **2015**, *96*, 1867–1877.
68. Benevides, P.; Catalao, J.; Miranda, P.M.A. On the inclusion of GPS precipitable water vapour in the nowcasting of rainfall. *Nat. Hazards Earth Syst. Sci.* **2015**, *15*, 2605–2616.
69. Mishra, A.K.; Coulibaly, P. Developments in hydrometric network design: A review. *Rev. Geophys.* **2009**, *47*.
70. Saltikoff, E.; Cho, J.Y.N.; Tristant, P.; Huuskonen, A.; Allmon, L.; Cook, R.; Becker, E.; Joe, P. The threat to weather radars by wireless technology. *Bull. Am. Meteorol. Soc.* **2016**, *97*, 1159–1167.
71. ECCC - Environment Climate Change Canada Modernizing Canada's weather-radar network. Available online: https://www.canada.ca/en/environment-climate-change/services/weather-general-tools-resources/radar-overview/modernizing-network.html.
72. Min, C.; Chen, S.; Gourley, J.J.; Chen, H.; Zhang, A.; Huang, Y.; Huang, C.

Coverage of China new generation weather radar network. *Adv. Meteorol.* **2019**, *2019*.

73. Saltikoff, E.; Haase, G.; Delobbe, L.; Gaussiat, N.; Martet, M.; Idziorek, D.; Leijnse, H.; Novák, P.; Lukach, M.; Stephan, K. OPERA the radar project. *Atmosphere (Basel).* **2019**, *10*, 1–13.

74. Kitzmiller, D.; Miller, D.; Fulton, R.; Ding, F. Radar and multisensor precipitation estimation techniques in national weather service hydrologic operations. *J. Hydrol. Eng.* **2013**, *18*, 133–142.

75. Willie, D.; Chen, H.; Chandrasekar, V.; Cifelli, R.; Campbell, C.; Reynolds, D.; Matrosov, S.; Zhang, Y. Evaluation of multisensor quantitative precipitation estimation in Russian river basin. *J. Hydrol. Eng.* **2017**, *22*, 1–11.

76. Einfalt, T.; Arnbjergnielsen, K.; Golz, C.; Jensen, N.; Quirmbach, M.; Vaes, G.; Vieux, B. Towards a roadmap for use of radar rainfall data in urban drainage. *J. Hydrol.* **2004**, *299*, 186–202.

77. Thorndahl, S.; Einfalt, T.; Willems, P.; Nielsen, J.E.; ten Veldhuis, M.-C.; Arnbjerg-Nielsen, K.; Rasmussen, M.R.; Molnar, P. Weather radar rainfall data in urban hydrology. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 1359–1380.

78. Allegretti, M. X-Band Mini Radar for Observing and Monitoring Rainfall Events. *Atmos. Clim. Sci.* **2012**, *02*, 290–297.

79. Chandrasekar, V.; Chen, H.; Philips, B. Principles of High-Resolution Radar Network for Hazard Mitigation and Disaster Management in an Urban Environment. *J. Meteorol. Soc. Japan. Ser. II* **2018**.

80. Hirano, K.; Maki, M. Imminent nowcasting for severe rainfall using vertically integrated liquid water content derived from X-band polarimetric radar. *J. Meteorol. Soc. Japan* **2018**, *96A*, 201–220.

81. Chen, H.; Chandrasekar, V. The quantitative precipitation estimation system for Dallas-Fort Worth (DFW) urban remote sensing network. *J. Hydrol.* **2015**, *531*, 259–271.

82. Chen, H.; Lim, S.; Chandrasekar, V.; Jang, B.J. Urban hydrological applications of dual-polarization X-band radar: Case study in Korea. *J. Hydrol. Eng.* **2017**, *22*, 1–13.

83. Cifelli, R.; Chandrasekar, V.; Chen, H.; Johnson, L.E. High resolution radar quantitative precipitation estimation in the San Francisco bay area: Rainfall monitoring for the urban environment. *J. Meteorol. Soc. Japan* **2018**, *96A*, 141–155.

84. Volkmann, T.H.M.; Lyon, S.W.; Gupta, H. V.; Troch, P.A. Multicriteria design of rain gauge networks for flash flood prediction in semiarid catchments with complex terrain. *Water Resour. Res.* **2010**, *46*, 1–16.

85. Thorndahl, S.; Beven, K.J.; Jensen, J.B.; Schaarup-Jensen, K. Event based uncertainty assessment in urban drainage modelling, applying the GLUE methodology. *J. Hydrol.* **2008**, *357*, 421–437.

86. Rafieeinasab, A.; Norouzi, A.; Kim, S.; Habibi, H.; Nazari, B.; Seo, D.J.; Lee, H.; Cosgrove, B.; Cui, Z. Toward high-resolution flash flood prediction in large urban areas - Analysis of sensitivity to spatiotemporal resolution of rainfall input and hydrologic modeling. *J. Hydrol.* **2015**, *531*, 370–388.

87. McKee, J.L.; Binns, A.D. A review of gauge–radar merging methods for quantitative precipitation estimation in hydrology. *Can. Water Resour. J.* **2016**, *41*, 186–203.

88. Ochoa-Rodriguez, S.; Wang, L. -P.; Willems, P.; Onof, C. A review of radar-rain gauge data merging methods and their potential for urban hydrological applications. *Water Resour. Res.* **2019**, *55*, 6356–6391.

89. Mckee, J.L. Evaluation of gauge-radar merging methods for quantitative precipitation estimation in hydrology : a case study in the Upper Thames River basin. Masters, Department of Civil and Environmental Engineering, The University of Western Ontario: London, ON, Canada, 2015.

90. Jewell, S.A.; Gaussiat, N. An assessment of kriging-based rain-gauge-radar merging techniques. *Q. J. R. Meteorol. Soc.* **2015**, *141*, 2300–2313.

91. Wang, L.P.; Ochoa-Rodríguez, S.; Van Assel, J.; Pina, R.D.; Pessemier, M.; Kroll, S.; Willems, P.; Onof, C. Enhancement of radar rainfall estimates for urban hydrology through optical flow temporal interpolation and Bayesian gauge-based adjustment. *J. Hydrol.* **2015**, *531*, 408–426.

92. Hossain, F.; Anagnostou, E.N. Assessment of current passive-microwave- and infrared-based satellite rainfall remote sensing for flood prediction. *J. Geophys. Res. D Atmos.* **2004**, *109*, 1–14.

93. Scofield, R.A.; Kuligowski, R.J. Status and outlook of operational satellite precipitation algorithms for extreme-precipitation events. *Weather Forecast.* **2003**, *18*, 1037–1051.

94. Kuligowski, R.J.; Li, Y.; Hao, Y.; Zhang, Y. Improvements to the GOES-R rainfall rate algorithm. *J. Hydrometeorol.* **2016**, *17*, 1693–1704.

95. Hong, Y.; Hsu, K.L.; Sorooshian, S.; Gao, X. Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. *J. Appl. Meteorol.* **2004**, *43*, 1834–1852.

96. Nguyen, P.; Ombadi, M.; Sorooshian, S.; Hsu, K.; AghaKouchak, A.; Braithwaite, D.; Ashouri, H.; Rose Thorstensen, A. The PERSIANN family of global satellite precipitation data: A review and evaluation of products. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 5801–5816.

97. Chen, S.; Liu, H.; You, Y.; Mullens, E.; Hu, J.; Yuan, Y.; Huang, M.; He, L.; Luo, Y.; Zeng, X.; et al. Evaluation of high-resolution precipitation estimates from satellites during July 2012 Beijing flood event using dense rain gauge observations. *PLoS One* **2014**, *9*, 1–7.

98. Cánovas-García, F.; García-Galiano, S.; Karbalaee, N. Validation of a global satellite rainfall product for real time monitoring of meteorological extremes. In Proceedings of the Remote Sensing for Agriculture, Ecosystems, and Hydrology XIX, SPIE; Warsaw, Poland, 2017; Vol. 10421, p. 8.

99. Seo, D.; Lakhankar, T.; Mejia, J.; Cosgrove, B.; Khanbilvardi, R. Evaluation of operational national weather service gridded flash flood guidance over the Arkansas Red River basin. *J. Am. Water Resour. Assoc.* **2013**, *49*, 1296–1307.

100. Seo, D.; Lakhankar, T.; Cosgrove, B.; Khanbilvardi, R.; Zhan, X. Applying SMOS soil moisture data into the National Weather Service (NWS)'s Research Distributed

Hydrologic Model (HL-RDHM) for flash flood guidance application. *Remote Sens. Appl. Soc. Environ.* **2017**, *8*, 182–192.

101. Crow, W.T.; Chen, F.; Reichle, R.H.; Liu, Q. L band microwave remote sensing and land data assimilation improve the representation of prestorm soil moisture conditions for hydrologic forecasting. *Geophys. Res. Lett.* **2017**, *44*, 5495–5503.
102. Owe, M.; van de Griend, A.A.; Chang, A.T.C. Surface moisture and satellite microwave observations in semiarid southern Africa. *Water Resour. Res.* **1992**, *28*, 829–839.
103. Bartalis, Z.; Wagner, W.; Naeimi, V.; Hasenauer, S.; Scipal, K.; Bonekamp, H.; Figa, J.; Anderson, C. Initial soil moisture retrievals from the METOP-A Advanced Scatterometer (ASCAT). *Geophys. Res. Lett.* **2007**, *34*, 5–9.
104. Entekhabi, D.; Njoku, E.G.; O'Neill, P.E.; Kellogg, K.H.; Crow, W.T.; Edelstein, W.N.; Entin, J.K.; Goodman, S.D.; Jackson, T.J.; Johnson, J.; et al. The soil moisture active passive (SMAP) mission. *Proc. IEEE* **2010**, *98*, 704–716.
105. Kerr, Y.H.; Waldteufel, P.; Wigneron, J.P.; Delwart, S.; Cabot, F.; Boutin, J.; Escorihuela, M.J.; Font, J.; Reul, N.; Gruhier, C.; et al. The SMOS L: New tool for monitoring key elements of the global water cycle. *Proc. IEEE* **2010**, *98*, 666–687.
106. Torres, R.; Snoeij, P.; Geudtner, D.; Bibby, D.; Davidson, M.; Attema, E.; Potin, P.; Rommen, B.Ö.; Floury, N.; Brown, M.; et al. GMES Sentinel-1 mission. *Remote Sens. Environ.* **2012**, *120*, 9–24.
107. Cenci, L.; Pulvirenti, L.; Boni, G.; Chini, M.; Matgen, P.; Gabellani, S.; Squicciarino, G.; Pierdicca, N. An evaluation of the potential of Sentinel 1 for

improving flash flood predictions via soil moisture-data assimilation. *Adv. Geosci.* **2017**, *44*, 89–100.

108. Srivastava, P.K. Satellite soil moisture: Review of theory and applications in water resources. *Water Resour. Manag.* **2017**, *31*, 3161–3176.

109. Liu, J.; Zhan, X.; Hain, C.; Yin, J.; Fang, L.; Li, Z.; Zhao, L. NOaa Soil Moisture Operational Product System (SMOPS) and its validations. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS); IEEE: Beijing, China, 2016; pp. 3477–3480.

110. Hansen, L.S.; Borup, M.; Møller, A.; Mikkelsen, P.S. Flow forecasting using deterministic updating of water levels in distributed hydrodynamic urban drainage models. *Water* **2014**, *6*, 2195–2211.

111. Fava, M.C.; Mazzoleni, M.; Abe, N.; Mendiono, E.M.; Solomatine, D. An approach for urban catchment model updating. **2018**, *3*, 692–685.

112. Acosta-Coll, M.; Ballester-Merelo, F.; Martinez-Peiró, M.; De la Hoz-Franco, E. Real-time early warning system design for pluvial flash floods—a review. *Sensors* **2018**, *18*.

113. Yang, T.-H.; Yang, S.-C.; Kao, H.-M.; Wu, M.-C.; Hsu, H.-M. Cyber-physical-system-based smart water system to prevent flood hazards. *Smart Water* **2018**, *3*, 1–13.

114. Chang, C.H.; Chung, M.K.; Yang, S.Y.; Hsu, C.T.; Wu, S.J. A case study for the application of an operational two-dimensional real-time flooding forecasting system and smart water level gauges on roads in Tainan City, Taiwan. *Water* **2018**, *10*.

115. Bartos, M.; Wong, B.; Kerkez, B. Open storm: A complete framework for sensing and control of urban watersheds. *Environ. Sci. Water Res. Technol.* **2018**, *4*, 346–358.

116. Kerkez, B.; Gruden, C.; Lewis, M.; Montestruque, L.; Quigley, M.; Wong, B.; Bedig, A.; Kertesz, R.; Braun, T.; Cadwalader, O.; et al. Smarter stormwater systems. *Environ. Sci. Technol.* **2016**, *50*, 7267–7273.

117. Powers, J.G.; Klemp, J.B.; Skamarock, W.C.; Davis, C.A.; Dudhia, J.; Gill, D.O.; Coen, J.L.; Gochis, D.J.; Ahmadov, R.; Peckham, S.E.; et al. The weather research and forecasting model: Overview, system efforts, and future directions. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 1717–1737.

118. Seity, Y.; Malardel, S.; Hello, G.; Bénard, P.; Bouttier, F.; Lac, C.; Masson, V. The AROME-France convective-scale operational model. *Mon. Weather Rev.* **2011**, *139*, 976–991.

119. Milbrandt, J.A.; Bélair, S.; Faucher, M.; Vallée, M.; Carrera, M.L.; Glazer, A. The pan-canadian high resolution (2.5 km) deterministic prediction system. *Weather Forecast.* **2016**, *31*, 1791–1816.

120. Baldauf, M.; Seifert, A.; Förstner, J.; Majewski, D.; Raschendorfer, M.; Reinhardt, T. Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities. *Mon. Weather Rev.* **2011**, *139*, 3887–3905.

121. Mittermaier, M.; Roberts, N.; Thompson, S.A. A long-term assessment of precipitation forecast skill using the Fractions Skill Score. *Meteorol. Appl.* **2013**, *20*, 176–186.

122. Clark, P.; Roberts, N.; Lean, H.; Ballard, S.P.; Charlton-Perez, C. Convection-permitting models: A step-change in rainfall forecasting. *Meteorol. Appl.* **2016**, *23*, 165–181.

123. Chamberlain, J.M.; Bain, C.L.; Boyd, D.F.A.; Mccourt, K.; Butcher, T.; Palmer, S. Forecasting storms over Lake Victoria using a high resolution model. *Meteorol. Appl.* **2014**, *21*, 419–430.

124. Herman, G.R.; Schumacher, R.S. Extreme precipitation in models: An evaluation. *Weather Forecast.* **2016**, *31*, 1853–1879.

125. Woodhams, B.J.; Birch, C.E.; Marsham, J.H.; Bain, C.L.; Roberts, N.M.; Boyd, D.F.A. What is the added value of a convection-permitting model for forecasting extreme rainfall over tropical East Africa? *Mon. Weather Rev.* **2018**, *146*, 2757–2780.

126. Rogelis, M.C.; Werner, M. Streamflow forecasts from WRF precipitation for flood early warning in mountain tropical areas. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 853–870.

127. Snook, N.; Kong, F.; Brewster, K.A.; Xue, M.; Thomas, K.W.; Supinie, T.A.; Perfater, S.; Albright, B. Evaluation of convection-permitting precipitation forecast products using WRF, NMMB, and FV3 for the 2016-17 NOAA hydrometeorology testbed flash flood and intense rainfall experiments. *Weather Forecast.* **2019**, *34*, 781–804.

128. Yussouf, N.; Knopfmeier, K.H. Application of the Warn-on-Forecast system for flash-flood-producing heavy convective rainfall events. *Q. J. R. Meteorol. Soc.*

2019, 2385–2403.

129. Corazza, M.; Sacchetti, D.; Antonelli, M.; Drofa, O. The ARPAL operational high resolution Poor Man's Ensemble, description and validation. *Atmos. Res.* **2018**, *203*, 1–15.

130. Termonia, P.; Fischer, C.; Bazile, E.; Bouyssel, F.; Brožková, R.; Bénard, P.; Bochenek, B.; Degrauwe, D.; Derková, M.; El Khatib, R.; et al. The ALADIN System and its canonical model configurations AROME CY41T1 and ALARO CY40T1. *Geosci. Model Dev.* **2018**, *11*, 257–281.

131. Rebora, N.; Ferraris, L.; von Hardenberg, J.; Provenzale, A. RainFARM: Rainfall downscaling by a Filtered Autoregressive Model. *J. Hydrometeorol.* **2006**, *7*, 724–738.

132. Tomassetti, B.; Verdecchia, M.; Giorgi, F. NN5: A neural network based approach for the downscaling of precipitation fields - Model description and preliminary results. *J. Hydrol.* **2009**, *367*, 14–26.

133. He, X.; Chaney, N.W.; Schleiss, M.; Sheffield, J. Spatial downscaling of precipitation using adaptable random forests. *Water Resour. Res.* **2016**, *52*, 8217–8237.

134. Silvestro, F.; Rebora, N. Impact of precipitation forecast uncertainties and initial soil moisture conditions on a probabilistic flood forecasting chain. *J. Hydrol.* **2014**, *519*, 1052–1067.

135. Davolio, S.; Silvestro, F.; Malguzzi, P. Effects of increasing horizontal resolution in a convection-permitting model on flood forecasting: The 2011 dramatic events in

Liguria, Italy. *J. Hydrometeorol.* **2015**, *16*, 1843–1856.

136. Germann, U.; Zawadzki, I. Scale-dependence of the predictability of precipitation from continental radar images. Part I: Description of the methodology. *Mon. Weather Rev.* **2002**, *130*, 2859–2873.

137. Yoo, C.; Lee, J.; Chang, K.; Yang, D. Sensitivity evaluation of the flash flood warning system introduced to ungauged small mountainous basins in Korea. *J. Mt. Sci.* **2019**, *16*, 971–990.

138. Woo, W.C.; Wong, W.K. Operational application of optical flow techniques to radar-based rainfall nowcasting. *Atmosphere (Basel).* **2017**, *8*.

139. Foresti, L.; Panziera, L.; Mandapaka, P. V.; Germann, U.; Seed, A. Retrieval of analogue radar images for ensemble nowcasting of orographic rainfall. *Meteorol. Appl.* **2015**, *22*, 141–155.

140. Shi, X.; Chen, Z.; Wang, H. Convolutional LSTM Network: A machine learning approach for precipitation nowcasting. In Proceedings of the 28th International Conference on Neural Information Processing Systems - NIPS; Montreal, QC, Canada, 2015.

141. Shi, X.; Gao, Z.; Lausen, L.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.C. Deep learning for precipitation nowcasting: A benchmark and a new model. In Proceedings of the 31st International Conference on Neural Information Processing Systems - NIPS; Long Beach, CA, USA, 2017.

142. Tran, Q.-K.; Song, S. Computer vision in precipitation nowcasting: Applying image quality assessment metrics for training deep neural networks. *Atmosphere (Basel).*

2019, *10*, 244.

143. Mandapaka, P. V.; Germann, U.; Panziera, L.; Hering, A. Can Lagrangian extrapolation of radar fields be used for precipitation nowcasting over complex alpine orography? *Weather Forecast.* **2012**, *27*, 28–49.

144. Smith, P.J.; Panziera, L.; Beven, K.J. Forecasting flash floods using data-based mechanistic models and NORA radar rainfall forecasts. *Hydrol. Sci. J.* **2014**, *59*, 1403–1417.

145. Golding, B.W. Nimrod: A system for generating automated very short range forecasts. *Meteorol. Appl.* **1998**, *5*, 1–16.

146. Pierce, C.E.; Hardaker, P.J.; Collier, C.G.; Haggett, C.M. GANDOLF: A system for generating automated nowcasts of convective precipitation. *Meteorol. Appl.* **2000**, *7*, 341–360.

147. Pierce, C.E.; Ebert, E.; Seed, A.W.; Sleigh, M.; Collier, C.G.; Fox, N.I.; Donaldson, N.; Wilson, J.W.; Roberts, R.; Mueller, C.K. The nowcasting of precipitation during Sydney 2000: An appraisal of the QPF algorithms. *Weather Forecast.* **2004**, *19*, 7–21.

148. Jensen, D.G. Combining weather radar nowcasts and numerical weather prediction models to estimate short-term quantitative precipitation and uncertainty. Ph.D., Faculty of Engineering and Science, Aalborg University: Aalborg, Denmark, 2015.

149. Haiden, T.; Kann, A.; Wittmann, C.; Pistotnik, G.; Bica, B.; Gruber, C. The integrated nowcasting through comprehensive analysis (INCA) system and its validation over the Eastern Alpine region. *Weather Forecast.* **2011**, *26*, 166–183.

150. Reyniers, M.; Delobbe, L. The nowcasting system INCA-BE in Belgium and its performance in different synoptic situations. In Proceedings of the 7th European Conference on radar in Meteorology and Hydrology; Météo France: Toulouse, France, 2012.

151. Kann, A.; Pistotnik, G.; Bica, B. INCA-CE: a Central European initiative in nowcasting severe weather and its applications. *Adv. Sci. Res.* **2012**, *8*, 67–75.

152. Bowler, N.E.; Pierce, C.E.; Seed, A.W. STEPS: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP. *Q. J. R. Meteorol. Soc.* **2006**, *132*, 2127–2155.

153. Seed, A.W.; Pierce, C.E.; Norman, K. Formulation and evaluation of a scale decomposition-based stochastic precipitation nowcast scheme. *Water Resour. Res.* **2013**, *49*, 6624–6641.

154. Liguori, S.; Rico-Ramirez, M.A. A practical approach to the assessment of probabilistic flow predictions. *Hydrol. Process.* **2013**, *27*, 18–32.

155. Foresti, L.; Reyniers, M.; Seed, A.; Delobbe, L. Development and verification of a real-time stochastic precipitation nowcasting system for urban hydrology in Belgium. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 505–527.

156. Poletti, M.L.; Silvestro, F.; Davolio, S.; Pignone, F.; Rebora, N. Using nowcasting technique and data assimilation in a meteorological model to improve very short range hydrological forecasts. *Hydrol. Earth Syst. Sci.* **2019**, *23*, 3823–3841.

157. Yoon, S.-S. Adaptive blending method of radar-based and numerical weather prediction QPFs for urban flood forecasting. *Remote Sens.* **2019**, *11*, 642.

158. European Commission - Joint Research Centre. *EFAS upgrade for the extended model domain – technical documentation*; Publications Office of the European Union: Luxembourg, 2019;

159. Quintero, F.; Versini, P.A.; Baltas, E.; Berenguer, M.; Sempere-Torres, D. A radar-based flash flood forecasting for the Llobregat river basin in the Catalonia region (Spain). In Proceedings of the 35th Conference on Radar Meteorology; AMS: Pittsburgh, PA, USA, 2011.

160. Thorstensen, A.; Nguyen, P.; Hsu, K.; Sorooshian, S. Using densely distributed soil moisture observations for calibration of a hydrologic model. *J. Hydrometeorol.* **2016**, *17*, 571–590.

161. Mantilla, R.; Gupta, V.K. A GIS numerical framework to study the process basis of scaling statistics in river networks. *IEEE Geosci. Remote Sens. Lett.* **2005**, *2*, 404–408.

162. Bartsotas, N.S.; Nikolopoulos, E.I.; Anagnostou, E.N.; Solomos, S.; Kallos, G. Moving toward subkilometer modeling grid spacings: Impacts on atmospheric and hydrological simulations of extreme flash flood-inducing storms. *J. Hydrometeorol.* **2017**, *18*, 209–226.

163. Silvestro, F.; Gabellani, S.; Delogu, F.; Rudari, R.; Boni, G. Exploiting remote sensing land surface temperature in distributed hydrological modelling: The example of the Continuum model. *Hydrol. Earth Syst. Sci.* **2013**, *17*, 39–62.

164. Wang, J.; Yang, H.; Li, L.; Gourley, J.J.; Sadiq, K.I.; Yilmaz, K.K.; Adler, R.F.; Policelli, F.S.; Habib, S.; Irwn, D.; et al. The coupled routing and excess storage

(CREST) distributed hydrological model. *Hydrol. Sci. J.* **2011**, *56*, 84–98.

165. Jiang, D.; Wang, K. The role of satellite-based remote sensing in improving simulated streamflow: A review. *Water* **2019**, *11*.

166. Nanía, L.S.; León, A.S.; García, M.H. Hydrologic-hydraulic model for simulating dual drainage and flooding in urban areas: Application to a catchment in the metropolitan area of Chicago. *J. Hydrol. Eng.* **2015**, *20*, 1–13.

167. Rossman, L.A. *Storm Water Management Model - User's manual version 5.1*; US EPA: Cincinnati, OH, 2015;

168. Danish Hydraulic Institute - DHI *Mike Urban Model Manager - User Guide*; DHI: Hørsholm, Denmark, 2019;

169. Henonin, J.; Russo, B.; Mark, O.; Gourbesville, P. Real-time urban flood forecasting and modelling - A state of the art. *J. Hydroinformatics* **2013**, *15*, 717–736.

170. Blumensaat, F.; Wolfram, M.; Krebs, P. Sewer model development under minimum data requirements. *Environ. Earth Sci.* **2012**, *65*, 1427–1437.

171. Mair, M.; Zischg, J.; Rauch, W.; Sitzenfrei, R. Where to find water pipes and sewers?-On the correlation of infrastructure networks in the urban environment. *Water* **2017**, *9*.

172. Silvestro, F.; Rebora, N. Operational verification of a framework for the probabilistic nowcasting of river discharge in small and medium size basins. *Nat. Hazards Earth Syst. Sci.* **2012**, *12*, 763–776.

173. Silvestro, F.; Rebora, N.; Cummings, G.; Ferraris, L. Experiences of dealing with flash floods using an ensemble hydrological nowcasting chain: implications of

communication, accessibility and distribution of the results. *J. Flood Risk Manag.* **2017**, *10*, 446–462.

174. Quintero, F.; Sempere-Torres, D.; Berenguer, M.; Baltas, E. A scenario-incorporating analysis of the propagation of uncertainty to flash flood simulations. *J. Hydrol.* **2012**, *460–461*, 90–102.

175. Bates, P.D.; De Roo, A.P.J. A simple raster-based model for flood inundation simulation. *J. Hydrol.* **2000**, *236*, 54–77.

176. Fewtrell, T.J.; Bates, P.D.; Horritt, M.; Hunter, N.M. Evaluating the effect of scale in flood inundation modelling in urban environments. *Hydrol. Process.* **2008**, *22*, 5107–5118.

177. Néelz, S.; Pender, G. *Benchmarking the latest generation of 2D hydraulic modelling packages*; Bristol, UK, 2013;

178. 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.

179. Berkhahn, S.; Fuchs, L.; Neuweiler, I. An ensemble neural network model for real-time prediction of urban floods. *J. Hydrol.* **2019**, *575*, 743–754.

180. Chu, H.; Wu, W.; Wang, Q.J.; Nathan, R.; Wei, J. An ANN-based emulation modelling framework for flood inundation modelling: Application, challenges and future directions. *Environ. Model. Softw.* **2019**.

181. Austin, R.J.; Chen, A.S.; Savić, D.A.; Djordjević, S. Quick and accurate Cellular Automata sewer simulator. *J. Hydroinformatics* **2014**, *16*, 1359–1374.

182. Dottori, F.; Todini, E. A 2D flood inundation model based on Cellular Automata approach. In Proceedings of the XVIII International Conference on Water Resources; Barcelona, Spain, 2010.

183. Guidolin, M.; Chen, A.S.; Ghimire, B.; Keedwell, E.C.; Djordjević, S.; Savić, D.A. A weighted cellular automata 2D inundation model for rapid flood analysis. *Environ. Model. Softw.* **2016**, *84*, 378–394.

184. Jamali, B.; Bach, P.M.; Cunningham, L.; Deletic, A. A Cellular Automata Fast Flood Evaluation (CA-ffé) Model. *Water Resour. Res.* **2019**, 4936–4953.

185. Liu, L.; Liu, Y.; Wang, X.; Yu, D.; Liu, K.; Huang, H.; Hu, G. Developing an effective 2-D urban flood inundation model for city emergency management based on cellular automata. *Nat. Hazards Earth Syst. Sci.* **2015**, *15*, 381–391.

186. Noh, S.J.; Lee, J.H.; Lee, S.; Kawaike, K.; Seo, D.J. Hyper-resolution 1D-2D urban flood modelling using LiDAR data and hybrid parallelization. *Environ. Model. Softw.* **2018**, *103*, 131–145.

187. Chen, Z.; Gao, B.; Devereux, B. State-of-the-art: DTM generation using airborne LIDAR data. *Sensors* **2017**, *17*.

188. Sampson, C.C.; Fewtrell, T.J.; Duncan, A.; Shaad, K.; Horritt, M.S.; Bates, P.D. Use of terrestrial laser scanning data to drive decimetric resolution urban inundation models. *Adv. Water Resour.* **2012**, *41*, 1–17.

189. de Almeida, G.A.M.; Bates, P.; Ozdemir, H. Modelling urban floods at submetre resolution: challenges or opportunities for flood risk management? *J. Flood Risk Manag.* **2018**, *11*, S855–S865.

190. Meesuk, V.; Vojinovic, Z.; Mynett, A.E.; Abdullah, A.F. Urban flood modelling combining top-view LiDAR data with ground-view SfM observations. *Adv. Water Resour.* **2015**, *75*, 105–117.

191. Diakakis, M.; Andreadakis, E.; Nikolopoulos, E.I.; Spyrou, N.I.; Gogou, M.E.; Deligiannakis, G.; Katsetsiadou, N.K.; Antoniadis, Z.; Melaki, M.; Georgakopoulos, A.; et al. An integrated approach of ground and aerial observations in flash flood disaster investigations. The case of the 2017 Mandra flash flood in Greece. *Int. J. Disaster Risk Reduct.* **2019**, *33*, 290–309.

192. Schumann, G.J. -P. The need for scientific rigour and accountability in flood mapping to better support disaster response. *Hydrol. Process.* **2019**, 1–5.

193. 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.

194. Kim, H. Il; Keum, H.J.; Han, K.Y. Real-time urban inundation prediction combining hydraulic and probabilistic methods. *Water* **2019**, *11*.

Chapter 3. Using Multiple Precipitation Estimates to Predict High-Flow Discharges in a Flashy Catchment Using Machine Learning

Summary of the research article: Zanchetta, A.D.L.; Coulibaly, P.; Fortin, V. (2022) Forecasting High-Flow Discharges in a Flashy Catchment Using Multiple Precipitation Estimates as Predictors in Machine Learning Models. Hydrology, 9(12). DOI: 10.3390/hydrology9120216.

This technical note describes the potential of using concurrent precipitation estimates from different sources (namely: rain gauges, ground-based weather radar and numerical weather models) to improve the prediction of the discharge in the outlet of an urban catchment prone to flash floods. The broad questions explored are:

- Can a machine learning model benefit from the additional information potentially present in more than one precipitation estimate to reduce the errors in the forecasting of river discharge?
- How the inclusion of additional precipitation estimates in the predictor set of a machine learning affects its performance for detecting high flows?

Key findings include:

- Using multiple concurrent precipitation estimates has the potential to improve the goodness of fit of the models (in terms of reduction of error and bias) results.

- The higher is the number of predictors used, the higher is the overall skill (estimated based in Critical Success Index metric) and precision of the predictor for detecting high flows. However, such gains may be accompanied by a reduction in the sensitivity of the models.

3.1. Abstract

The use of machine learning for predicting high river flow events is gaining prominence among research and operational communities. Among the non-trivial design decisions of such models is the definition of the input dataset, including the rainfall estimates to be used. This study proposes and evaluates the use of multiple concurrent precipitation products to improve the performance of a machine learning model towards the forecasting of the river discharge in a flashy urban catchment. Multiple extreme learning machines were trained with distinct combinations of quantitative precipitation estimates from radar, reanalysis, and gauge datasets. Their performance was then assessed in terms of goodness of fit and contingency analysis for the prediction of high flows. It was found that models that use multiple precipitation estimates simultaneously over-performed the best of its single-precipitation counterparts for the lead times in which forecasts were considered "useful", with larger improvement achieved at time-lags close to the catchment's response time. The novelty of these results suggest that the implementation of data-driven models could achieve better performance if the predictive features related to rainfall data were more diverse in terms of data sources when compared with the currently predominant use of a single rainfall product.

3.2. Introduction

Floods are among the most hazardous events for human life and infrastructures, and according to estimations from the United Nations their occurrence increased by more than 50% in the current decade following an upward trend apparently driven by climate change [1]. To support the decisions of emergency response teams, governmental agencies implement flood early warning systems (FEWS), which are characterized as integrated components designed for hazard monitoring, forecasting, disaster risk assessment and communication [2,3].

The river discharge forecasting component of FEWSs historically rely on conventional hydrological models (CHMs) to predict river discharges through simulations of physical or conceptual representations of the hydrological cycle [4,5]. As data-driven alternatives for CHMs, the potential of using supervised machine learning (ML) models for tasks related to rainfall-runoff modelling has been studied for more than 25 years [6,7]. Driven by a continuously increasing of data and computational power availability, an expanding number of ML methods and algorithms has been explored with success, with performances comparable or superior to their CHM counterparts [8]. Examples of ML structures already applied for such a purpose include multi-layer perceptrons (MLP) [9,10], long-short term memory neural networks (LSTM) [11,12], random and decision tree forests [13,14], support vector machines (SVM) [15,16] and extreme learning machines (ELM) [17,18]. This scenario motivates the exploration of techniques and approaches to improve their performance for an implementation in operational forecasting chains [19].

Replicating the design approach adopted for CHMs, in which each input data represents univocally a material or energetic flux relevant for the local hydrological processes, ML models have been trained and tested with quantitative precipitation estimates (QPE) derived from a single source, be it from rain gauges [12,15], ground-based weather radar [20] or satellite data [21,22]. However, each of such products has specific advantages and limitations: rain gauges are considered the most accurate estimator of rainfall water depth while being restricted to provide measurements for point locations; ground-based weather radars are able to provide relevant information about the spatial distribution and motion of rainfall by recording microwaves reflected from the upper atmosphere, which results in elevated uncertainties concerning the amount of precipitated water that reaches the ground; and numerical weather models (NWM) include concepts of physics to fill gaps and bridge different sources of observational data, potentially adding epistemic uncertainties to the estimated rainfall value [23,24].

These differences usually raise questions on which of these estimations is the best to be used to represent precipitation as input in discharge forecasting systems. They also motivate the search for techniques for combining different QPEs in a way to efficiently maintain their positive characteristics and mitigate their errors. The combination of multiple QPEs is usually performed separately from the rainfall-runoff simulation model, in an additional component that produces a single and supposedly improved QPE. The post-processed QPE is then used as the input for the rainfall-runoff model [25,26]. The use of a post processor external to the hydrological model, however, has the drawbacks of (1) increasing the complexity of an operational systems, as two components need to be set up

and maintained, and of (2) assuming that one of the rainfall products (usually from rain gauges) is available and has a quality high enough to be considered a reference for the other QPEs, which may become a constraint for data scarce regions. If a rainfall-runoff model could encapsulate both the QPE combination and the discharge estimation components through data-driven methods (thus without prior assumptions on the data quality of individual products), both the drawbacks would be overcome.

Different from CHMs, conventional ML models are designed to abstractly identify patterns in the training dataset, can be constructed to have an arbitrary number of inputs, and are not constraint by physical laws or conceptualizations of the hydrological behavior of a catchment. This flexibility allows the construction of models that estimates future discharge having multiple concurrent precipitation estimates as part of its input. Such a configuration fulfils the suggested encapsulation, however, to the best of the author's knowledge, no previous work has explored the potential effectiveness of including more than one source of QPE products in the predictors set for improving discharge forecasting in ML models.

3.2.1. Research Questions and Objectives

The purpose of this work is to answer the question: can the use of concurrent QPE products improve the performance of rainfall-runoff ML models? And more specifically: can it improve the prediction of high flows? The hypothesis evaluated is that ML tools can be efficient on identifying how underlying patterns in concurrent QPE products can be reflected in the discharge of a small catchment. This raises the expectations that the more QPE products used as predictive features, the higher is the expected performance of the

model. However, the inclusion of an additional predictor highly correlated with other input data may not result in performance gains and thus leading to an unjustified increase in the model complexity. Even worse: it can result in the overfitting of the data-driven model and consequently in the deterioration of the model performance. Such a duality is usually referred to as "the curse of dimensionality" [27] and demands case-specific analysis on whether a feature should be included or not, which justifies the need for the study presented in this paper.

Answering the abovementioned research questions can serve as an important factor for designers of river forecasting systems based on artificial intelligence, particularly on their decision to consider (or not) using predictor sets composed by more than one rainfall product instead of relying on the use of single QPEs as usually observed in the literature. While the experiment is performed in a densely monitored catchment, the insights presented can also benefit the development of models of poorly gauged or ungauged basins given the current availability of multiple precipitation products with global coverage that could be used as a replacement for the gauged data in this study. Examples of such precipitation products available in near-real time include the estimates derived from satellites, such as the Integrated Multi-satellitE Retrievals for Global precipitation measurement (IMERG) [28], the Global Satellite Mapping of Precipitation (GSMaP) [29], the and the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks - Cloud Classification System (PERSIAN- CCS) [22,30,31].

For evaluating our hypothesis, a computational experiment was conducted to compare the performance of different ELMs models, each taking distinct sets of precipitation

estimates as input. The conventional single-QPE design is included in the comparison to represent the business-as-usual approach, and the Don River Basin, located in Toronto, Canada, is taken as a study case.

3.3. Materials

3.3.1. Study Area

The densely occupied and predominantly urbanized Don River basin (Toronto, Canada) houses approximately 1.4 million residents in total area of nearly 350 km^2 upstream the river gauge HY019 (Figure 0-1), where the average baseflow is estimated to be in the order of 5 m^3/s. The longest flow path has 38 km length and an average slope of 6.6 * 10^{-3}, which reflects the overall smoothness of the relief in the surroundings of the Great Lakes.

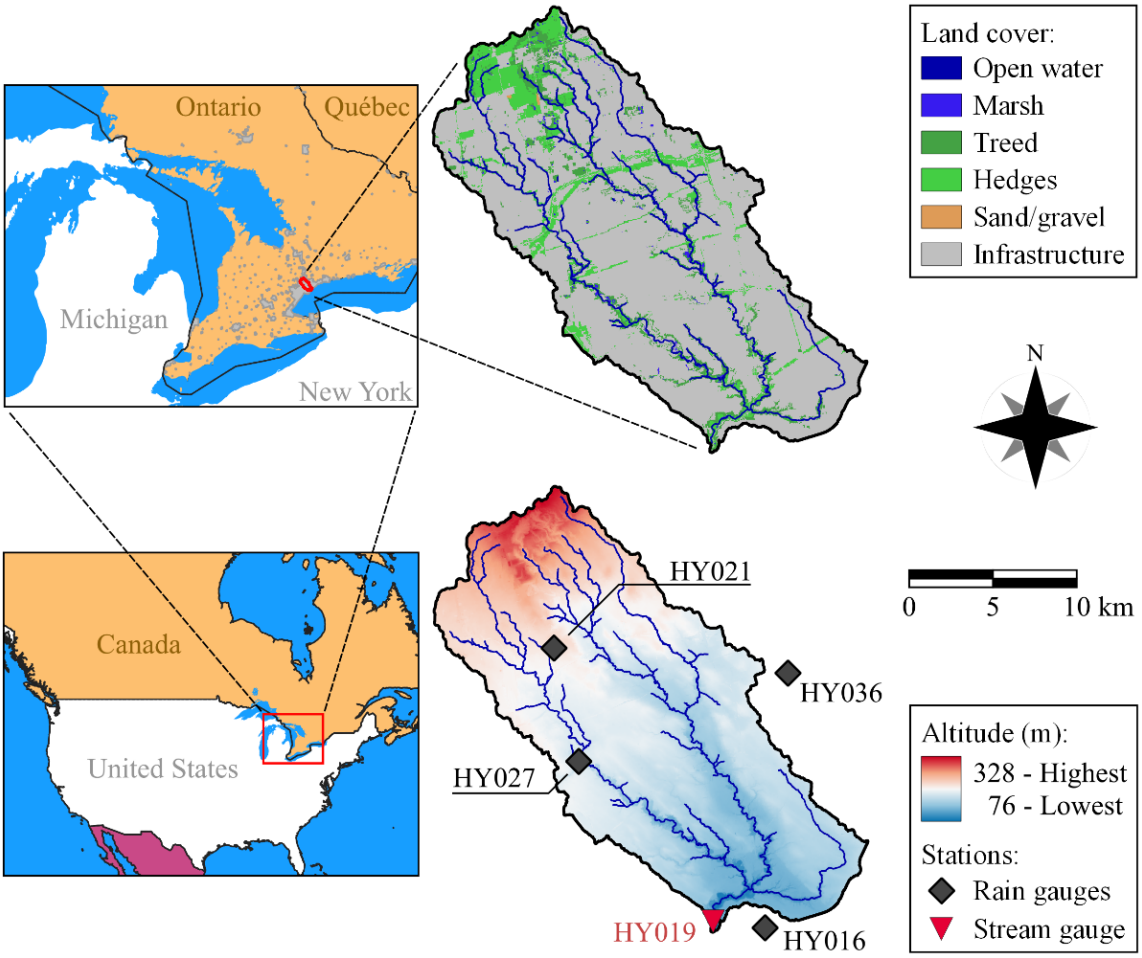

Figure 0-1. Location, landcover and elevations of the Don River basin with the rain and stream gauges used in this work.

The area just downstream the gauge HY019, known as "lower Don", frequently becomes flooded due to river overbank conditions triggered by intense precipitation events during the warm season of the years (May to October, both months included). Such flooding scenarios recurrently result in disturbances in the operation of high-traffic roads and urban railways in the riverside of the lower Don, causing significant economical and material losses, besides overall life-threatening conditions for the local population. The response time of the catchment at the gauge HY019 is the order of 3 hours, therefore the

Don River basin can be considered “flashy” according to the criterion adopted by the United States’ National Weather Service for defining flash floods, i.e., floods “beginning within six hours of the causative event (e.g., intense rainfall, dam failure, ice jam)” [32].

3.3.2. Dataset

3.3.2.1. Data Description

The discharge data (Q) used was collected by the stream gauge HY019. Four meteorological stations equipped with tipping buckets and located within (or close to) the catchment (HY016, HY021, HY027, HY036 as shown in Figure 3-1) recorded the data used as a representative of rain gauge estimates (referred as G). The Toronto and Region Conservation Authority (TRCA) is the agency responsible for managing the Don River Basin, for maintaining all the above-mentioned gauges, and for providing the collected data to the public at 15-minutes temporal resolution.

QPEs from data collected by the National Oceanic and Atmospheric Administration (NOAA) NEXRAD S-Band weather radar located in Buffalo, USA, were used to represent products from ground-based radars (referred as R). The distance between the station and the most distant point of the catchment is 128 km, which is within the usable range of 180 km from this ground-based weather radar [33]. The specific product used, referred as N1P, is calculated by NOAA’s Precipitation Processing System algorithm [34] and provided at temporal and spatial resolutions of 1 hour and approximately 2 km over the Don River basin, respectively.

The Canadian Precipitation Analysis (CaPA) products produced by the Environment and Climate Change Canada (ECCC) in its 10 km, hourly resolution version [35] is used as the NWM representative (referred as C). The version of CaPA products used in this work integrates observation data from North American Surface Synoptic Observations (SYNOP), Surface Weather Observations (SWOB) and METeorological Aerodrome Reports (METAR) gauge networks with numerical weather simulations based on the ECCC's Global Environmental Multiscale model (GEM) [36]. It is worth noting that the aforementioned gauge networks do not include stations maintained by TRCA. A summary of the dataset is given in Table 3-1.

Table 3-1. Summary of the dataset used in this work.

				Original resolution	
Variable	**Data type**	**Data source**	**Acronym**	**Temporal (hour)**	**Spatial (km)**
Precipitation	Observation	Rain gauges	G	0.25	Point
Precipitation	Observation	Weather radar	R	1	~ 2 km
Precipitation	Modelling	NWM (CaPA)	C	1	~ 10 km
Discharge	Observation	River gauge	Q	0.25	Point

3.3.2.2. Data Pre-Processing

The radar and the NWM products were used "as-is". The decision of not applying common data treatments (such as bias correction) is based on the assumption that the ELM is capable of "learning" by its own the characteristics and patterns of the raw QPE products that are relevant to the task of predicting stream discharge. All precipitation records were spatially aggregated into a single uniform value following a lumped representation of the catchment given its small area. Particularly, the rain gauge data was aggregated using the conventional Thiessen polygons method [37] and the weights of each gauge in the weighted

average is given in Table 3-2. The spatial granularity of the different precipitation products over the Don River basin is presented in Figure 3-2.

Table 3-2. Aerial representation and relative weight of each rain gauge.

Rain gauge	Aerial representativity (km^2)	Weight
HY016	59.2	0.17
HY021	114.9	0.33
HY027	107.9	0.31
HY036	66.2	0.19

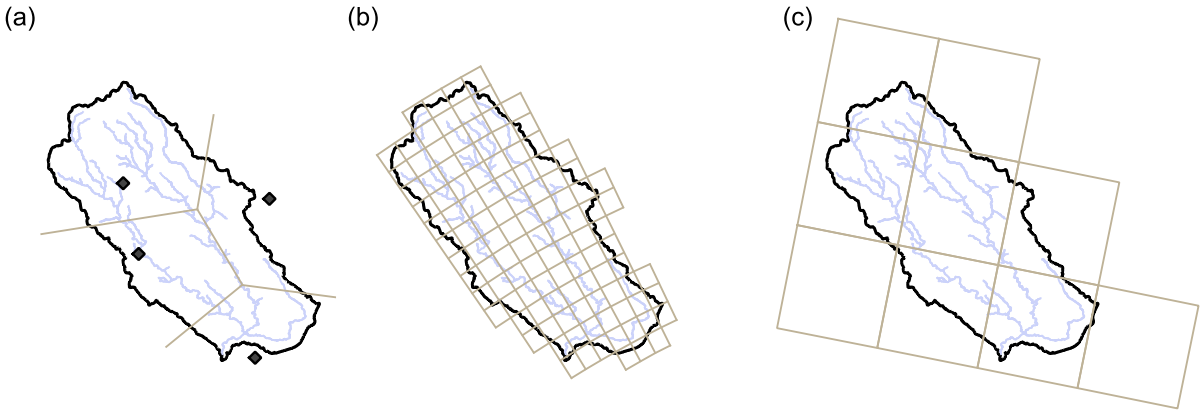

Figure 0-2. Spatial granularity (before aggregation) of the data precipitation products from (a) rain gauges' Thiessen polygons, (b) ground-based weather radar's derived N1P product grid cells, and (c) NWM product grid cells.

All timeseries were temporarily aggregated into hourly resolution to match the resolution of the CaPA data as it is the coarsest among the products considered. A total of 63 intense rainfall-runoff events observed in the warm seasons between the years of 2011 to 2017 were identified and used to compose the train/validation/test dataset. A rainfall-runoff event is considered intense when the discharge posterior to a rainfall record exceeds 10 times the baseflow value, and each event is defined in time as the 24 hours interval centered in the peak discharge.

3.4. Methods

3.4.1. Extreme Learning Machine (ELM) Models

ELM models [38] are single-layer feed-forward neural networks that are calibrated analytically instead of through the conventional stochastic gradient descent approach. In a summary, the output $\vec{y}'$ for an input $\vec{x}$ produced by an already trained ELM is given by:

$$\vec{y}'(\vec{x}) = \sum_{i=1}^{N_h} \omega_i \, F_i(\vec{x}) = \sum_{i=1}^{N_h} \omega_i \, f(s(\vec{\alpha}_i, \vec{x})) = \sum_{i=1}^{N_h} \omega_i \, f(\sum_{j=1}^{N_x} \alpha_{i,j} x_j) \qquad (0\text{-}1)$$

in which N_h is the number of hidden neurons; the output value and output weight of the i-th hidden node are given by F_i and ω_i, respectively; F_i is calculated as the output of the activation function f for the weighted sum s of $\vec{x}$; $\vec{\alpha}_i$ is the input weights for the i-th hidden node; both $\vec{x}$ and $\vec{\alpha}_i$ have N_x components; and the j-th components $\vec{x}$ and $\vec{\alpha}_i$ is represented by x_j and $\alpha_{j,i}$, respectively. A visual representation of the structure is given in Figure 0-3.

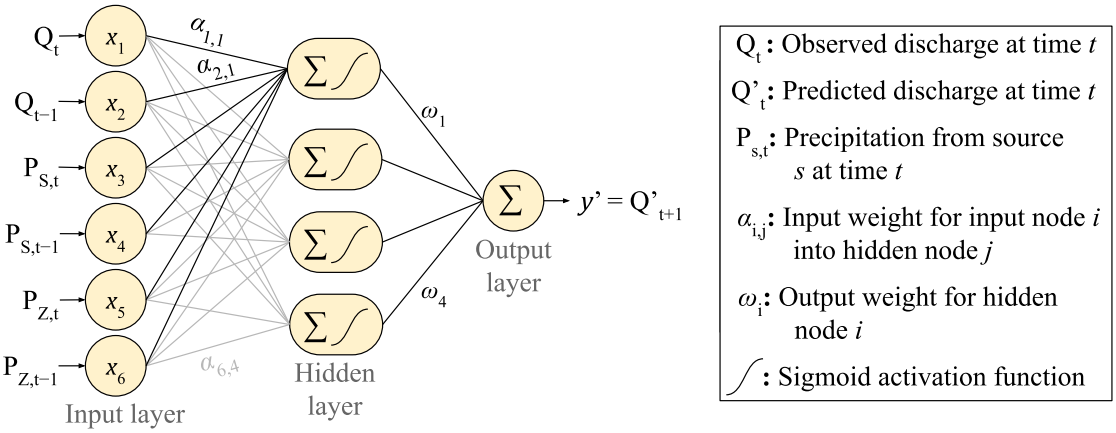

Figure 0-3. Visual representation of a hypothetical ELM structure as adopted in this work. It takes two QPE products (S and Z) and antecedent discharge as input to predict the instant discharge one unit of time (hour) in the future.

As a data-driven model, the tuning of an ELM is performed by fitting its intrinsic mathematical formula (in the form of Equation 3-1) using a set X of N_t training samples ($\vec{x}_1$, $\vec{x}_2$, …, $\vec{x}_{N_t}$) as arguments and a set Y of their respective target values ($y_1, y_2, \dots, y_{N_t}$). The set $H(X)$ of output values of the hidden layers for all samples in X, the set W of output weights of the hidden nodes, and Y can be represented in matrices as:

$$H(X) = \begin{bmatrix} F_1(\vec{x}_1) & \cdots & F_{N_h}(\vec{x}_1) \\ \vdots & \ddots & \vdots \\ F_1(\vec{x}_{N_t}) & \cdots & F_{N_h}(\vec{x}_{N_t}) \end{bmatrix}, \quad W = \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_{N_h} \end{bmatrix} \text{ and } \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_{N_t} \end{bmatrix}$$

so that the vectorial representation of the solutions for the training sample is given by:

$$Y = H(X) \times W. \tag{0-2}$$

As previously discussed, the function F is parameterized by the input-to-hidden node weights ($\alpha_{j,i}$ in Equation 3-1). In ELM, a random value is set for each $\alpha_{j,i}$ and are kept constant, which makes W the only unknown to be defined during the training stage. The inverse of $H(X)$, $H(X)^+$, is obtained through the Moon-Penrose generalized inversion function so that an approximation W' is calculated as:

$$W' = H(X)^+ \times Y. \tag{0-3}$$

The values of W' (ω'_1, ω'_2, …, ω'_{N_h}), which were therefore defined using the training sets X and Y, are used as the calibrated values of ω_1,ω_2,…,ω_{N_h} in Equation 1-1. As discussed by Huang et al. [38], such an approach provides the minimum training error, calculated as least-square error, for a network characterized by its number of hidden nodes and activation functions. In this work, we used ELM with the sigmoid activation function based on previous successful application on rainfall-runoff modelling [18,39], while the number of nodes in the hidden layer is considered a hyperparameter.

3.4.2. Experiment Set Up

A total of 8 sets of inputs were composed by all possible combinations of the 3 QPE products considered ([G], [R], [C], [G, R], [G, C], [R, C], [G, R, C], [None]) and two records of antecedent discharge (Q_{t-L} and Q_{t-L-1}). For each input set and for each lead time L ranging from 1 to 5 hours, multiple ELM models were trained to map the predictors P_{t-L}, P_{t-L-1}, Q_{t-L}, Q_{t-L-1} to the predictand instant discharge Q_t (t being an instant time and P the univariate, multivariate or absent QPE in the input set).

As ELM models are fit analytically to the data used for training, they are known to be highly prone to overfitting and conventional generalization methods designed for in iterative trainings, like early stopping [40], are not applicable. The approach adopted for overcoming this issue is to use the mean of the outputs of an ensemble of ELM predictors and with each of the ensemble members being fit to a different subset of the training data, as presented by Liu and Wang [41]. In this work, for creating the ensemble of models and for testing them, a nested k-fold cross validation approach was adopted.

The entire dataset was randomly segmented in 6 folds (subsets) so that each possible combination of 4 folds (66.6% of the data) can be used to train one individual ML model using 1 of the other folds for testing (16.7% of the data) and the 1 remaining fold for validation (16.7% of the data). Such a data splitting schema was adopted to resemble the data distribution of 70% for training and 30% for testing/validating as commonly adopted in literature [42–44]. Based on such a data division, 6 ensemble sets of ELM models were set up, each of the ensembles having a folding configuration in which 5 of the subsets are used for training its ensemble members (*ensemble fold-in*), and the remaining subset is

reserved for testing the ensemble forecasts (*ensemble fold-out*, or *member out-of-the-sample*). The 5 *ensemble folds-in* of each folding configuration are used to fit 5 ELM models, each ELM model using 4 of the folds for producing multiple ELM model candidates (*member fold-in*) and the remaining fold (*member fold-out*) to select the trained candidate with best generalization power. A visual representation of the aforementioned data splitting is given in Figure 0-4.

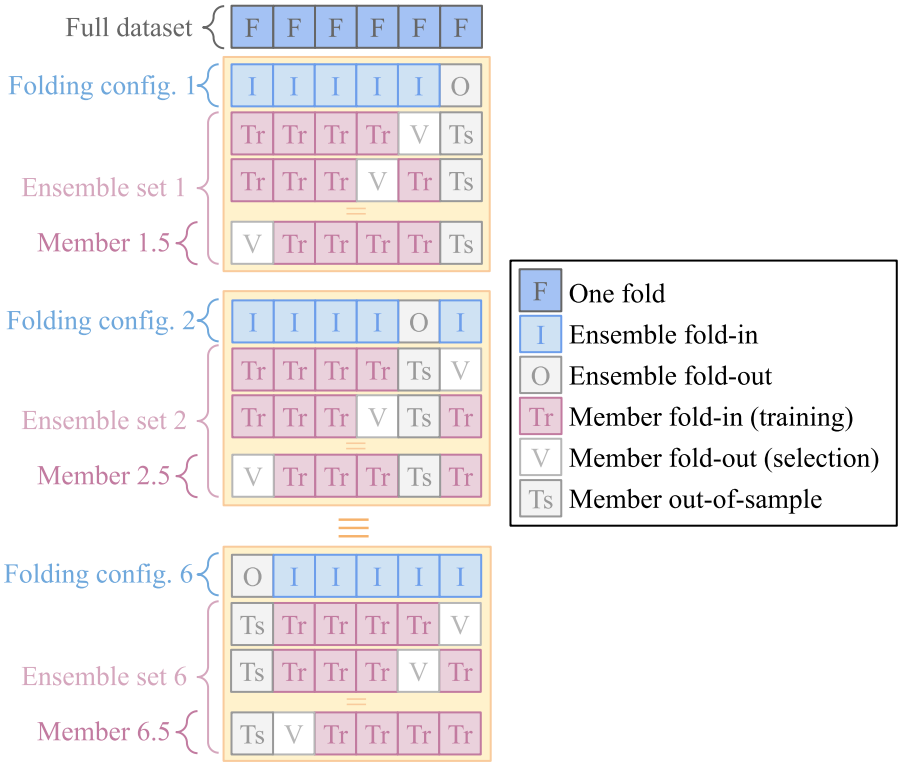

Figure 0-4. Data organization scheme of the cross-folding configurations and ensemble members training setup.

To tune the only hyperparameter of the ELM models (i.e., the number of nodes in the hidden layer), multiple ELM candidate models for each *member fold-in* subset were trained, each of which having a different number of hidden nodes ranging from 5 to 300.

The candidate model that presented the best performance in terms of lower root mean squared error (RMSE) using the respective *member fold-out* is selected. The aforementioned range of explored number of nodes in the hidden layer was determined from preliminary empirical experiments based on try and error (not shown). As it can be observed in Figure 3-5, most of the mean number of hidden nodes selected as optimal ranges around 100 and 150 depending on the lead time to which the ML models are trained.

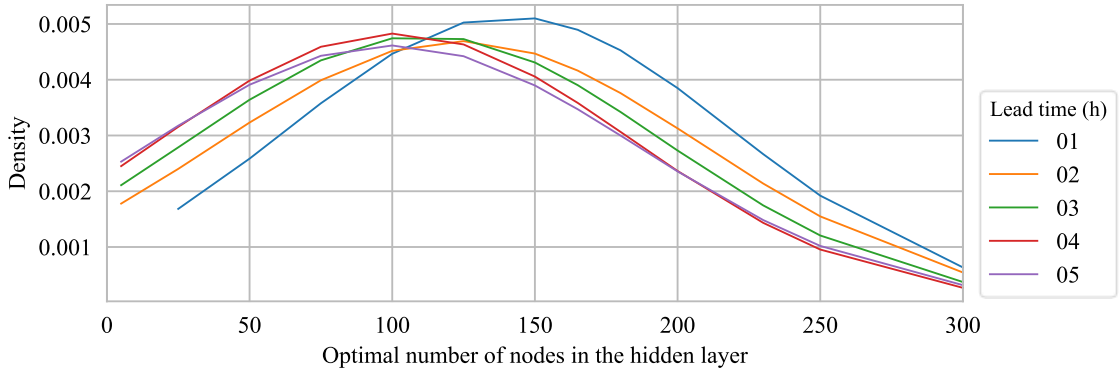

Figure 0-5. Distribution of the number of nodes in the hidden layer of the ELM model selected as optimal for the different lead times.

After the fitting and the hyperparameter tunning, for each of the 6 folds of the entire dataset, an ensemble of 5 ELM models is set up using only data from the other folds. The comparison of the model ensembles is performed using the data in the fold that was not included in the training/validation/selection stage (i.e., *member out-of-the-sample*), as represented in Figure 0-6.

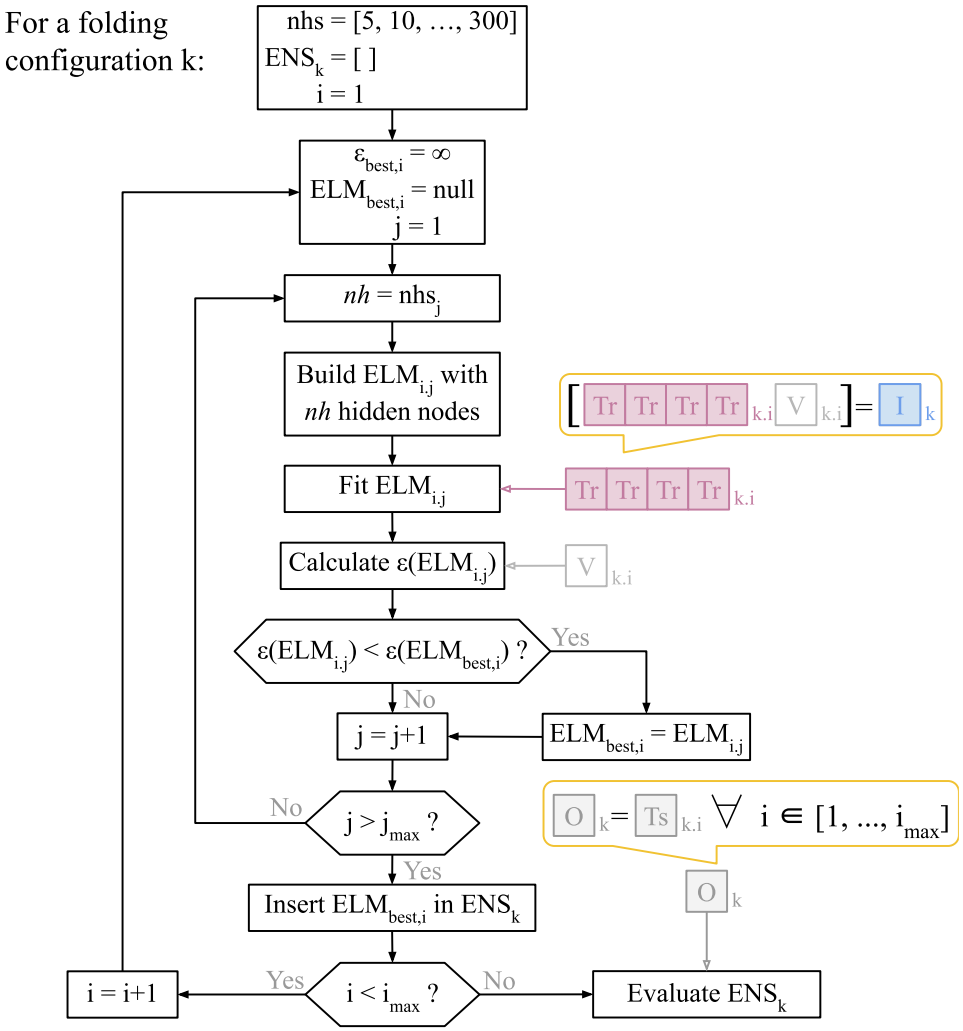

Figure 0-6. Diagram representing the methodology for training, selecting and evaluating the ELM members of ensemble k. ε(x) indicates the validation error of the ELM model x, nhsi indicates the i-th item in the sequence of assessed number of hidden nodes nhs., ENSk is the ensemble of models k.

3.4.3. Models Comparison

The overall performance of the models trained with different predictors were compared in terms of goodness of fit using Pearson's correlation coefficient (r, Equation 3-4) [45], root mean squared error (RMSE, Equation 3-5) [46] and the Kling-Gupta Efficiency (KGE, Equation 3-6) [47]. In a summary: r measures the linear correlation between the predicted and observed discharge values, assuming values between 0 (no correlation) to 1 (perfect

correlation); RMSE represents the square root of the average of squared error of a prediction dataset, has values in the same units as the evaluated variable (m^3/s in this work) and range from 0 to infinity (the higher, the worse); and KGE is an efficiency metric that balances the agreement between the predicted and observed values in terms linear correlation, standard deviation and mean, with unitless values bounded between 0 (worse agreement possible) and 1 (perfect agreement). Additionally, fractional bias (Equation 3-7) [46] is used to estimate the magnitude of the systematic overestimation (if value is positive) or underestimation (if value is negative) of the ELM models.

The above-mentioned metrics are calculated as:

$$r = \frac{\sum_{i=1}^{N}(Q_i - \overline{Q_i}) * (Q_1' - \overline{Q_i'})}{\sqrt{\sum_{i=1}^{N}(Q_i - \overline{Q_i})^2 * \sum_{i=1}^{N}(Q_i' - \overline{Q_i'})^2}}, \quad (0\text{-}4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Q_i' - Q_i)^2}{N}}, \quad (0\text{-}5)$$

$$KGE = 1 - \sqrt{(r-1)^2 - \left(\frac{\mu'}{\mu} - 1\right)^2 - \left(\frac{\sigma'}{\sigma} - 1\right)^2}, \quad (0\text{-}6)$$

$$Fractional\ bias = \frac{2}{N}\sum_{i=1}^{N}\frac{Q_i' - Q_i}{Q_i' + Q_i}. \quad (0\text{-}7)$$

in which Q_i and Q_i' are respectively the i-th observed and i-th predicted discharges; N is the total number of records in Q_i and Q_i'; μ and μ' are the mean of the Q and Q' values, respectively; and σ and σ' are the standard deviation of Q and Q', respectively.

Contingency metrics of critical success index (CSI, Equation 3-8), sensitivity (also known as *recall*, Equation 3-9) and precision (Equation 3-10) were used to assess the applicability of the model to predict high-flow events. In this work, a high-flow event is an individual moment in time during an intense rainfall-runoff event in which Q exceeds 50 m^3/s, equivalent to 10 times the baseflow discharge. The contingency metrics used in this work take into account the total of observed high-flow events that were correctly predicted ($hits$), the total of observed events that were not predicted ($misses$), and the total of events that were predicted but not observed ($false\ alerts$) as:

$$CSI = \frac{hits}{hits + misses + false\ alerts}, \tag{3-8}$$

$$sensitivity = \frac{hits}{hits + misses}, \tag{3-9}$$

$$precision = \frac{hits}{hits + false\ alerts}. \tag{3-10}$$

3.5. Results and Discussion

3.5.1. Overall Performance Statistics

The average performance of the different ELM ensemble models in terms of KGE is presented in Figure 3-7. As expected, the ELM models that uses none of the QPEs presented the worst performance among all input configurations considered, with absolute KGE differences of up to 0.14 when compared with the worst-performing single-QPE models (Figure 3-7d). This observation illustrates the magnitude of gains in performance that the ELM models can attain by "learning", up to a certain level, the influence that precipitation has on the discharge of the catchment for the different lead times.

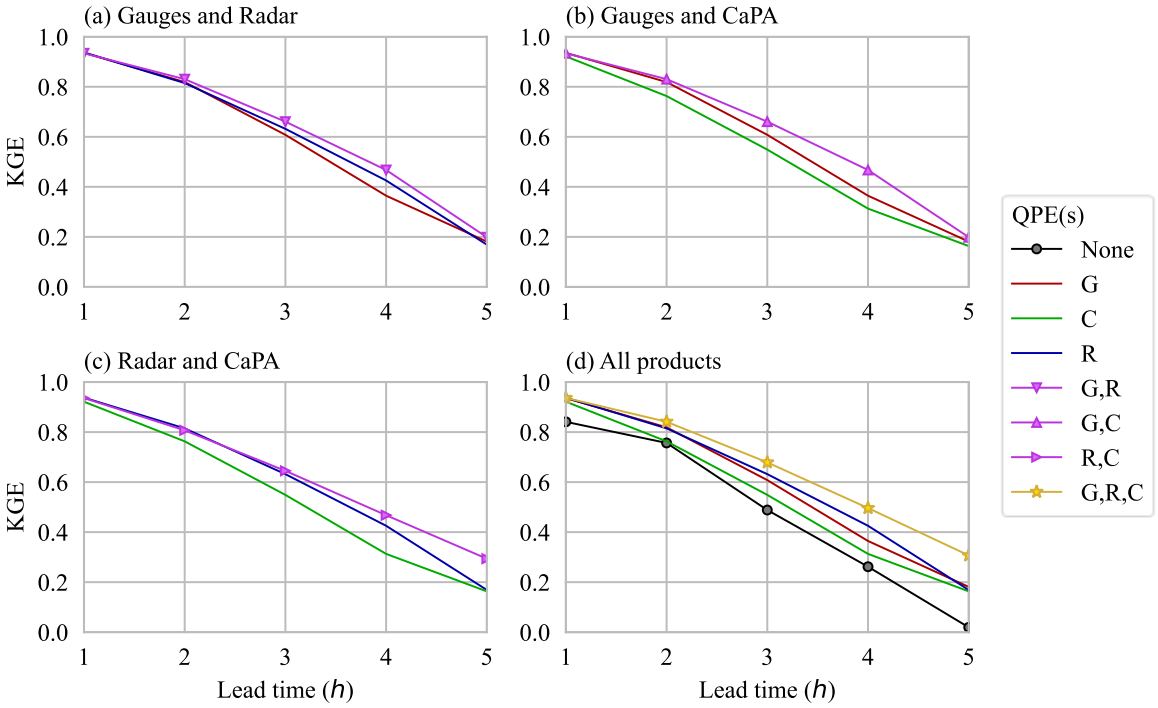

Figure 0-7. Mean KGE of the models that use, besides Q_{t-L} and Q_{t-L-1}, QPE from (a) gauges and radar, (b) gauge and CaPA, (c) radar and CaPA, and (d) all of them, including the no-QPE (None) used as benchmark.

Among the single-QPE ELM models, the ones using radar data presented the best performance in terms of KGE for lead times of 3 and 4 hours, and a performance comparable with the ELM that uses only gauges data for the other lead times. ELM models using CaPA data presented lower performance for lead times of 2, 3 and 4 hours when compared with its radar-based counterpart. A remarkable characteristic of the timeseries of all single-QPE ELMs is their comparable performances for both the shortest (1 hour) and longest (5 hours) lead time. A probable explanation for such a pattern may be related to the response time of the catchment of approximately 3 hours. The water that enters the catchment as precipitation in a moment t tends to have limited influence on the discharge at the outlet of the catchment in a moment t+1 hour as the majority of its volume is still in

transit through the basin. Most of the runoff volume reaches the outlet between t+2 hours and t+4 hours, thus differences in precipitation patterns are more likely to be reflected in the discharge at these instants. At t+5 hours, most of the water entering at time t is expected to have already left the catchment, leading to a condition similar to t+1. The major decrease in performance of the models with lead time L is likely driven by the Q_{t-L} and Q_{t-L-1} predictors (present in all models) following their decaying correlation with the predicted variable Q_t as L increases (Figure 3-8).

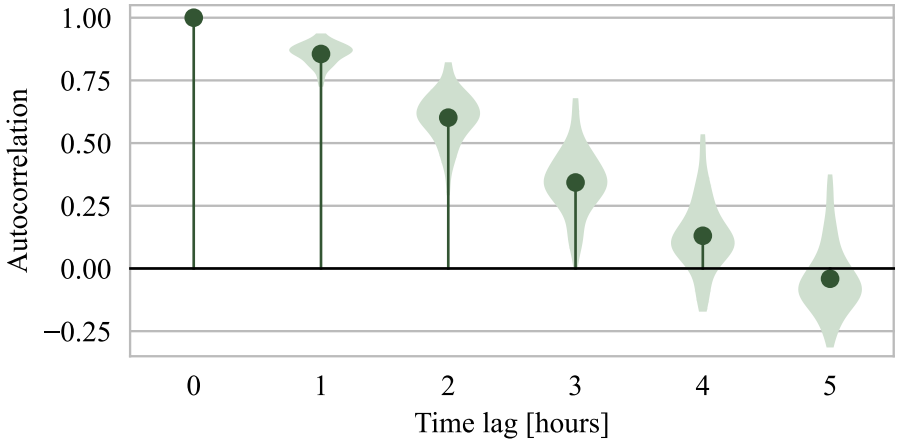

Figure 0-8. Correlogram of the observed discharge (Q) variable. Violin plots represent the distribution of values across the individual rainfall-runoff events.

All ELM models that use two of the QPE products concurrently presented equal or higher KGE values than the models that use only one of the QPEs individually (Figure 3-7a,b,c). The differences in performance between the single-QPE models and the two-QPE models is almost imperceptible for the shortest lead time. For longer lead times, the gain in performance with the use of more precipitation input varied for each pair of QPE products. Gains in absolute KGE are up to 0.04 for the gauge and radar pair (Figure 3-7a),

up to 0.11 for the gauge and CaPA pair (Figure 3-7b) and up to 0.12 for the radar and CaPA pair (Figure 3-7c).

As observed for the two-products scenarios, the models that use all QPEs presented a better performance than the best single-QPE models for all lead times, except for t+1 hour, when the differences are imperceptible (Figure 3-7d).

Taylor diagrams created using the Python library Skill Metrics [48] summarize the statistical parameters of standard deviation, coefficient of correlation with observations, and RMSE of the ELM models that use no QPE, the ones that use only one QPE product, and the ones that use all three QPEs (Figure 3-9). As observed in the KGE analysis, the models using the three QPE products performed in average better or as good as their best single-QPE counterparts for all lead times except for 1 hour, in which all QPE-aware ELMs practically coincide in all the three metrics. This consistency in the results from different metrics is a good indicator that the gains in performance with additional QPEs are not concentrated in specific scenarios nor produce undesired significant drawbacks.

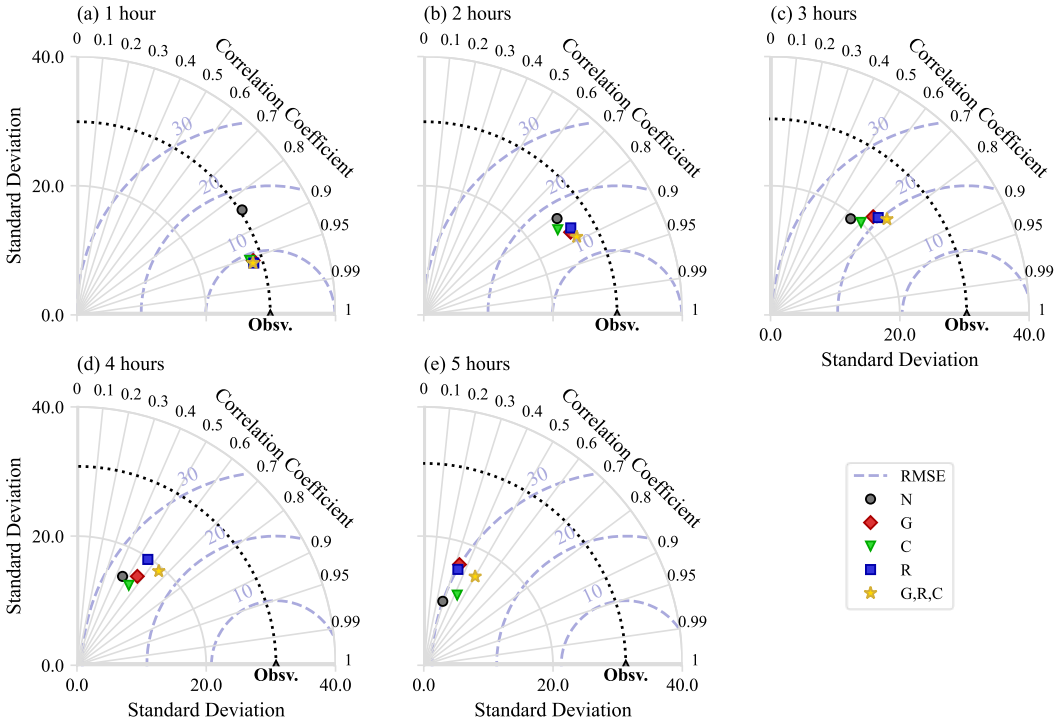

Figure 0-9. Taylor diagrams for the ELM models with different QPE products in their feature set at forecasting lead times ranging from 1 to 5 hours (a – e).

Table 0-3 summarizes the mean of the main metrics for evaluating the goodness-of-fit of the models, including the percentual gain of using the three QPEs when compared to the best performing single-QPE models. Out of the 15 scenarios (KGE, RMSE and *r* metrics at 5 lead times each), the multi-QPE didn't present the best performance in only 1 (RMSE at 1 hours lead time), had a performance comparable to the best of the single-QPE model in other 2 scenarios (KGE and *r* at 1 hour lead time), and was the best in the remaining 12 scenarios, making it the "clear winner" in terms of overall performance.

Table 3-3. Summary of mean performance gains for metrics related to goodness of fit when comparing single-QPE ML models with three-QPEs ML models.

		Best single QPE		3 QPEs	
Metric	**Lead time (*h*)**	**QPE**	**Value**	**Value**	**Δ (%)**
KGE	1	R	0.94	0.94	0.00

	2	G	0.82	0.84	2.44
	3	R	0.63	0.68	4.76
	4	R	0.43	0.50	16.28
	5	G	0.18	0.31	72.22
RMSE [m^3/s]	1	R	8.44	8.47	-0.35
	2	G	14.69	13.61	7.35
	3	R	20.37	19.30	5.25
	4	G	25.49	23.28	8.67
	5	C	28.26	27.07	4.21
r	1	R	0.96	0.96	0.00
	2	G	0.87	0.89	2.30
	3	R	0.74	0.77	4.05
	4	G	0.56	0.66	17.86
	5	C	0.43	0.50	16.28

Note: Δ(%) = 100% * (value(3 QPEs) - value(Best single QPE)) ÷ value(Best single QPE).

As presented in Table 3-4, the set of models that consider only radar data presented the lowest bias among the single-QPE group of models. Nevertheless, when all the three QPE products are used as input, a reduction in the overall bias is observed for all lead times. It worth noting, though, that all model setups are characterized as being biased low (negative bias value), a pattern that can be due to the fact that each predicted value used is the mean of an ensemble of model outputs, which tends to lead to a "smoothening" of the predictions, especially for the peak values. As improvements were also observed in other metrics, it is possible to deduce that the use of multiple QPE products resulted in an appropriate increase of the overall values predicted (higher bias value).

Table 3-4. Bias values of the ML models that use only one QPE and of the models that use all three QPEs for different lead times. Best values (closest to zero) for a lead time are highlighted in bold.

	Lead time (hours)				
Feature set	**1**	**2**	**3**	**4**	**5**
G	-0.06	-0.10	-0.17	-0.22	-0.22
R	-0.05	-0.08	-0.15	-0.19	-0.20

C	-0.07	-0.13	-0.18	-0.23	-0.22
G,R,C	**-0.03**	**-0.06**	**-0.12**	**-0.18**	**-0.19**

These results answer our first research question by evidencing that ML models indeed have the potential to "learning" patterns in the rainfall data estimated by different sources concurrently and improve their performance in reproducing rainfall-runoff processes. Additionally, it is possible to note that if a ML designer is constrained to use a single QPE product, discussions may rise on which data source should be chosen depending on the metrics considered. For example: models using only radar data and models using only gauge data were each considered as "the best model" for 2 lead times in terms of RMSE and r, while for KGE, radar-based models overperformed gauge-based models with a tight difference (3 times the first overperforming the later against 2 the other way around).

3.5.2. Contingency Analysis

As described in Section 3.3, the threshold adopted in this study to identify high-flows events was the value representing 10 times the baseflow discharge (i.e., 50 m^3/s), and the calculated contingency metrics for the identification of high-flows events is presented in Figure 3-10. For all lead times, the EML models that use the three QPE products as predictors presented a median CSI value that is higher than its best-performing single-QPE counterparts, which indicates an overall best performance of the first over the later on predicting the upcoming occurrence of a high-flow record. Reflecting the results obtained in the analysis of the goodness of fit, the single-QPE model that uses CaPA products consistently presented the worst results. Models using only gauge and only radar data

presented competitive performances between each other, with the first performing better in lead times of 2 hours while the later overperformed in the remaining lead times.

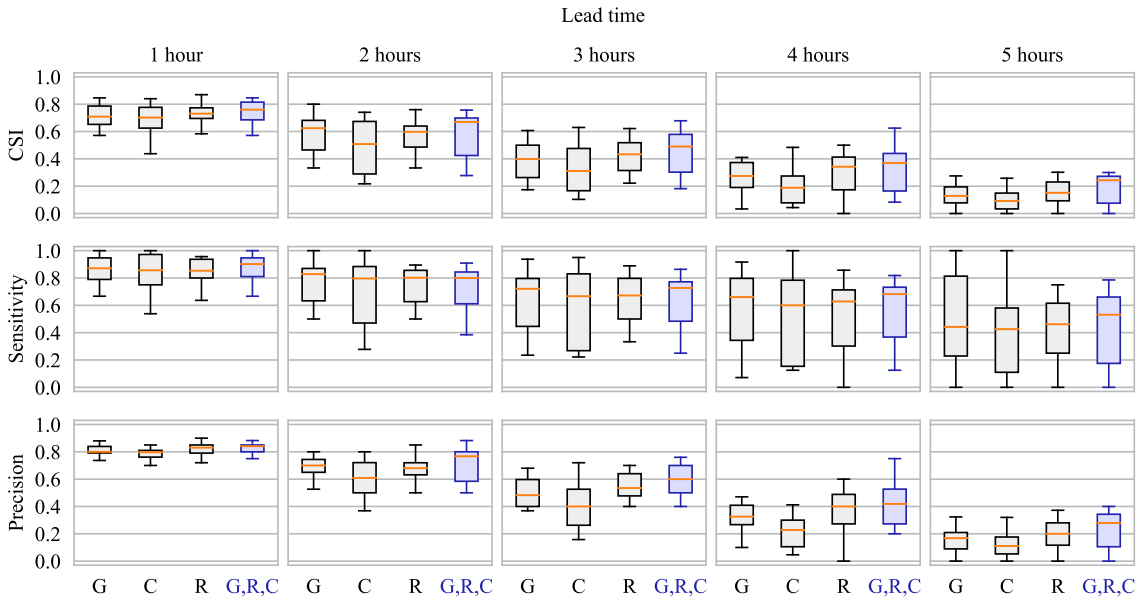

Figure 0-10. Contingency analysis on predicting high-flow events for the ML models using only QPEs from gauge (G), only from ground-based weather radar (R), only from CaPA as representative of NWM (C), and using all of them (G,R,C).

The boxplots representing the sensitivity values of the models that use multiple QPEs have little differences when compared to the boxplots of its best-performing single-QPE counterpart. The forecast precision of all models remains high for lead time 1 and 2 hour at about 75% and then deteriorate significantly for subsequent lead times. If the criterion for deciding if an ensemble of models is that both precision and sensitivity should be higher than 0.5, the maximum lead time to which they could be deemed “useful” is 3 hours. The resulting CSI values of the models that use multiple precipitation inputs is the highest (or equivalent to other highest model with only one precipitation product) for all lead times – indicating the superior potential benefit in using all precipitation products.

The performance of single-QPE and three-QPE models for detecting high flows is summarized in Table 3-5. If only the ability to anticipate the occurrence of an upcoming high-flow event is to be considered, regardless of the number of false alarms issued, modelers could be inclined to select the models that use only QPEs from gauges as their sensitivity was the highest for 2 of the lead times. On the other hand, if avoiding the emission of false alarms is the main interest for the forecasters, the single-QPE model that uses only radar data overcomes its gauge data-based counterpart for the majority of the lead times due to its best performance in terms of precision. However, usually both the sensitivity and precision of forecasting models are important, and a balanced metric such as CSI is used as tiebreaker. In this work, however, CSI values indicate, as observed with KGE, that the radar-based ELMs overperform the gauges-based ELMs with just 3 out of the 5 lead times, which could still rise questions in the selection of the single-QPE product to be used.

The three-QPEs models overperformed the other three single-QPE ones in terms of precision for all led times, scored equally to the best ELMs (the ones using gauge data only) in two of the lead times and underperformed it in the remaining three lead times. If CSI is used as a tiebreaker, the three-QPEs configuration would emerge as a “clear winner” as it presents the best metrics for 4 out of 5 lead times though. Taking the single-QPE model that uses radar data as a reference, it is possible to interpret from these results that the addition of the other precipitation products provided useful information for skillfully reduce the number of false alarms issued, thus increasing the precision of the model in a way that outweighs the reduction of the number of missed events. These results answer our

second question: using concurrent QPE as input of the ML models also improved the prediction of high-flow events mainly by increasing its precision.

Table 3-5. Summary of mean performance gains – metrics related to contingency analysis.

		Best single QPE		3 QPEs	
Metric	**Lead time (*h*)**	**QPE**	**Value**	**Value**	**Δ (%)**
CSI	1	G	0.72	0.73	1.38
	2	G	0.59	0.58	-1.69
	3	R	0.43	0.45	4.65
	4	R	0.30	0.32	6.67
	5	R	0.16	0.19	18.75
Sensitivity	1	G	0.86	0.86	0.00
	2	G	0.78	0.74	-5.12
	3	R	0.65	0.64	-1.54
	4	G	0.56	0.56	0.00
	5	G	0.50	0.45	-10.0
Precision	1	R	0.82	0.83	1.22
	2	G	0.70	0.72	2.86
	3	R	0.55	0.59	7.27
	4	R	0.38	0.41	7.89
	5	R	0.20	0.24	20.00

Note: Δ(%) = 100% * (value(3 QPEs) - value(Best single QPE)) ÷ value(Best single QPE).

3.6. Conclusions

The performance of machine learning models to predict high-flow river discharge has the potential to benefit from the concurrent use of different QPE products in the input set of features when compared to the business-as-usual configuration of a single source for QPEs. For the "flashy" and small urban catchment taken in this work as study case, the model using three QPE products presented an overall goodness of fit to the observed discharge that outperformed the best single-QPE models for most of the forecasted lead times (ranging from 1 to 5 hours in the future) and metrics (KGE, coefficient of correlation,

RMSE and bias) evaluated. For anticipating high river flows, the multi-QPE models tended to present a better trade-off between number of errors of false alarms and missed events when compared to each of their best-performing single-QPE counterparts. The eventual performance loss usually expected when highly correlated variables are used as predictors of data-driven models (due to overfitting or excessive increase in model complexity) were overcome by the gains in considering the complementary information provided by each individual QPE source. Thus, our study suggests that if a region is served by more than one precipitation estimate, which is becoming more feasible due to the consolidation of the different satellite-based programs, then it worth considering the use of multiple QPEs as predictors of ML-based rainfall-runoff models. Therefore, the answer to the question "which of the available QPEs should be used to reach the best overall performance?" may be just "all of them" when using ML models to predict stream discharges.

The gains in performance observed, however, are limited by the information available in the input data as only three QPE products were considered. Also, results were presented only using one machine learning technique and for only one gauged catchment in which summer floods are mainly driven by intense rainfall followed by nearly instant surface runoff, and in which the influence of other hydrological factors such as soil moisture and evapotranspiration can be neglected.

Considering the aforementioned limitations of this work, future studies are recommended for exploring:

1. How the use of multiple precipitation products may benefit poorly gauged or ungauged basins using QPEs based on satellite data (e.g., PERSIANN-CCS, IMERG, GSMaP);
2. How does the performance of ML-based rainfall-runoff models that use multiple concurrent QPEs as input compares with the performance of conventional hydrological models that use a post-processed QPE based on the combination of different QPE products (e.g., radar data corrected with gauged data);
3. What are the potential gains in performance that rainfall-runoff models based on other popular ML structures (e.g., MLP, LSTM, SVM, random forests) may achieve with the inclusion of concurrent QPEs in the set of predictors;
4. How the ML-based rainfall-runoff modeling of basins with characteristics different from the flashy urban catchment used in this study may benefit from the use of multiple concurrent QPEs as input. Examples of such include mountainous catchments and non-flashy basins with larger areas.

3.7. References

1. UN-Water *United Nations World Water Development Report 2020: Water and Climate Change*; UNESCO: Paris, 2020; ISBN 9789231003714.
2. WMO - World Meteorological Organization *Multi-Hazard Early Warning Systems: A Checklist*; Geneva, 2017;
3. UN - United Nations *Report of the Open-Ended Intergovernmental Expert Working Group on Indicators and Terminology Relating to Disaster Risk Reduction*; 2016;

4. Zahmatkesh, Z.; Kumar Jha, S.; Coulibaly, P.; Stadnyk, T. An Overview of River Flood Forecasting Procedures in Canadian Watersheds. *Canadian Water Resources Journal* **2019**, *44*, 213–229, doi:10.1080/07011784.2019.1601598.
5. Zanchetta, A.D.L.; Coulibaly, P. Recent Advances in Real-Time Pluvial Flash Flood Forecasting. *Water (Basel)* **2020**, *12*, 570, doi:10.3390/w12020570.
6. Minns, A.W.; Hall, M.J. Modélisation Pluie-Débit Par Des Réseaux Neuroneaux Artificiels. *Hydrological Sciences Journal* **1996**, *41*, 399–417, doi:10.1080/02626669609491511.
7. Mosavi, A.; Ozturk, P.; Chau, K.W. Flood Prediction Using Machine Learning Models: Literature Review. *Water (Basel)* **2018**, *10*, 1–40, doi:10.3390/w10111536.
8. Frame, J.M.; Kratzert, F.; Klotz, D.; Gauch, M.; Shalev, G.; Gilon, O.; Qualls, L.M.; Gupta, H. v.; Nearing, G.S. Deep Learning Rainfall–Runoff Predictions of Extreme Events. *Hydrol Earth Syst Sci* **2022**, *26*, 3377–3392, doi:10.5194/hess-26-3377-2022.
9. Choi, H.S.; Kim, J.H.; Lee, E.H.; Yoon, S.K. Development of a Revised Multi-Layer Perceptron Model for Dam Inflow Prediction. *Water (Switzerland)* **2022**, *14*, doi:10.3390/w14121878.
10. Dawson, C.W.; Wilby, R.L. Hydrological Modelling Using Artificial Neural Networks. *Prog Phys Geogr* **2001**, *25*, 80–108, doi:10.1177/030913330102500104.
11. 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*, doi:10.3390/hydrology9060105.

12. Song, T.; Ding, W.; Wu, J.; Liu, H.; Zhou, H.; Chu, J. Flash Flood Forecasting Based on Long Short-Term Memory Networks. *Water (Basel)* **2019**, *12*, 109, doi:10.3390/w12010109.

13. Khan, M.T.; Shoaib, M.; Hammad, M.; Salahudin, H.; Ahmad, F.; Ahmad, S. Application of Machine Learning Techniques in Rainfall–Runoff Modelling of the Soan River Basin, Pakistan. *Water (Switzerland)* **2021**, *13*, doi:10.3390/w13243528.

14. Singh, A.K.; Kumar, P.; Ali, R.; Al-Ansari, N.; Vishwakarma, D.K.; Kushwaha, K.S.; Panda, K.C.; Sagar, A.; Mirzania, E.; Elbeltagi, A.; et al. An Integrated Statistical-Machine Learning Approach for Runoff Prediction. *Sustainability (Switzerland)* **2022**, *14*, doi:10.3390/su14138209.

15. Wu, J.; Liu, H.; Wei, G.; Song, T.; Zhang, C.; Zhou, H. Flash Flood Forecasting Using Support Vector Regression Model in a Small Mountainous Catchment. *Water (Switzerland)* **2019**, *11*, doi:10.3390/w11071327.

16. Alquraish, M.M.; Khadr, M. Remote-Sensing-Based Streamflow Forecasting Using Artificial Neural Network and Support Vector Machine Models. *Remote Sens (Basel)* **2021**, *13*, 1–20, doi:10.3390/rs13204147.

17. Atiquzzaman, M.; Kandasamy, J. Prediction of Hydrological Time-Series Using Extreme Learning Machine. *Journal of Hydroinformatics* **2016**, *18*, 345–353, doi:10.2166/hydro.2015.020.

18. Yeditha, P.K.; Rathinasamy, M.; Neelamsetty, S.S.; Bhattacharya, B.; Agarwal, A. Investigation of Satellite Precipitation Product Driven Rainfall-Runoff Model Using

Deep Learning Approaches in Two Different Catchments of India. *Journal of Hydroinformatics* **2021**, *00*, 1–22, doi:10.2166/hydro.2021.067.

19. Muñoz, P.; Orellana-Alvear, J.; Bendix, J.; Feyen, J.; Célleri, R. Flood Early Warning Systems Using Machine Learning Techniques: The Case of the Tomebamba Catchment at the Southern Andes of Ecuador. *Hydrology* **2021**, *8*, doi:10.3390/hydrology8040183.
20. Ke, Q.; Tian, X.; Bricker, J.; Tian, Z.; Guan, G.; Cai, H.; Huang, X.; Yang, H.; Liu, J. Urban Pluvial Flooding Prediction by Machine Learning Approaches – a Case Study of Shenzhen City, China. *Adv Water Resour* **2020**, *145*, 103719, doi:10.1016/j.advwatres.2020.103719.
21. Kumar, A.; Ramsankaran, R.A.A.J.; Brocca, L.; Muñoz-Arriola, F. A Simple Machine Learning Approach to Model Real-Time Streamflow Using Satellite Inputs: Demonstration in a Data Scarce Catchment. *J Hydrol (Amst)* **2021**, *595*, doi:10.1016/j.jhydrol.2021.126046.
22. Bhusal, A.; Parajuli, U.; Regmi, S.; Kalra, A. Application of Machine Learning and Process-Based Models for Rainfall-Runoff Simulation in DuPage River Basin, Illinois. *Hydrology* **2022**, *9*, doi:10.3390/hydrology9070117.
23. Sun, Q.; Miao, C.; Duan, Q.; Ashouri, H.; Sorooshian, S.; Hsu, K.L. A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons. *Reviews of Geophysics* **2018**, *56*, 79–107, doi:10.1002/2017RG000574.
24. Li, Z.; Chen, M.; Gao, S.; Hong, Z.; Tang, G.; Wen, Y.; Gourley, J.J.; Hong, Y. Cross-Examination of Similarity, Difference and Deficiency of Gauge, Radar and

Satellite Precipitation Measuring Uncertainties for Extreme Events Using Conventional Metrics and Multiplicative Triple Collocation. *Remote Sens (Basel)* **2020**, *12*, doi:10.3390/RS12081258.

25. Gabriele, S.; Chiaravalloti, F.; Procopio, A. Radar-Rain-Gauge Rainfall Estimation for Hydrological Applications in Small Catchments. *Advances in Geosciences* **2017**, *44*, 61–66, doi:10.5194/adgeo-44-61-2017.
26. McKee, J.L.; Binns, A.D. A Review of Gauge–Radar Merging Methods for Quantitative Precipitation Estimation in Hydrology. *Canadian Water Resources Journal* **2016**, *41*, 186–203, doi:10.1080/07011784.2015.1064786.
27. Friedman, J.H. On Bias, Variance, 0/1-Loss, and the Curse-of-Dimensionality. *Data Min Knowl Discov* **1997**, *1*, 55–77, doi:10.1023/A:1009778005914.
28. Huffman, G.J.; Stocker, E.F.; Bolvin, D.T.; Nelkin, E.J.; Tan, J. GPM IMERG Final Precipitation L3 Half Hourly 0.1 Degree x 0.1 Degree V06.
29. Kubota, T.; Aonashi, K.; Ushio, T.; Shige, S.; Takayabu, Y.N.; Kachi, M.; Arai, Y.; Tashima, T.; Masaki, T.; Kawamoto, N.; et al. Global Satellite Mapping of Precipitation (GSMaP) Products in the GPM Era. In *Satellite Precipitation Measurement*; Levizzani, V., Kidd, C., Kirschbaum, D.B., Kummerow, C.D., Nakamura, K., Turk, F.J., Eds.; Springer, Cham, 2020; pp. 355–373.
30. Hong, Y.; Hsu, K.L.; Sorooshian, S.; Gao, X. Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network Cloud Classification System. *Journal of Applied Meteorology* **2004**, *43*, 1834–1852, doi:10.1175/jam2173.1.

31. Hsu, K.L.; Gao, X.; Sorooshian, S.; Gupta, H. v. Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks. *Journal of Applied Meteorology* **1997**, *36*, 1176–1190, doi:10.1175/1520-0450(1997)036<1176:PEFRSI>2.0.CO;2.

32. National Weather Service NWS Glossary Available online: https://w1.weather.gov/glossary/.

33. Wijayarathne, D.; Coulibaly, P.; Boodoo, S.; Sills, D. Evaluation of Radar-Gauge Merging Techniques to Be Used in Operational Flood Forecasting in Urban Watersheds. *Water (Switzerland)* **2020**, *12*, doi:10.3390/w12051494.

34. Fulton, R.A.; Breidenbach, J.P.; Seo, D.J.; Miller, D.A.; O'Bannon, T. The WSR-88D Rainfall Algorithm. *Weather Forecast* **1998**, *13*, 377–395, doi:10.1175/1520-0434(1998)013<0377:TWRA>2.0.CO;2.

35. Gasset, N.; Fortin, V.; Dimitrijevic, M.; Carrera, M.; Bilodeau, B.; Muncaster, R.; Gaborit, É.; Roy, G.; Pentcheva, N.; Bulat, M.; et al. A 10 Km North American Precipitation and Land-Surface Reanalysis Based on the GEM Atmospheric Model. *Hydrol Earth Syst Sci* **2021**, *25*, 4917–4945, doi:10.5194/hess-25-4917-2021.

36. Côté, J.; Gravel, S.; Méthot, A.; Patoine, A.; Roch, M.; Staniforth, A. The Operational CMC-MRB Global Environmental Multiscale (GEM) Model. Part I: Design Considerations and Formulation. *Mon Weather Rev* **1998**, *126*, 1373–1395, doi:10.1175/1520-0493(1998)126<1373:TOCMGE>2.0.CO;2.

37. Thiessen, A.H. Precipitation Averages for Large Areas. *Mon Weather Rev* **1911**, *39*, 1082–1089, doi:10.1175/1520-0493(1911)39<1082b:PAFLA>2.0.CO;2.

38. Huang, G. bin; Zhu, Q.Y.; Siew, C.K. Extreme Learning Machine: Theory and Applications. *Neurocomputing* **2006**, *70*, 489–501, doi:10.1016/j.neucom.2005.12.126.

39. Deo, R.C.; Şahin, M. Application of the Extreme Learning Machine Algorithm for the Prediction of Monthly Effective Drought Index in Eastern Australia. *Atmos Res* **2015**, *153*, 512–525, doi:10.1016/j.atmosres.2014.10.016.

40. Coulibaly, P.; Anctil, F.; Bobée, B. Daily Reservoir Inflow Forecasting Using Artificial Neural Networks with Stopped Training Approach. *J Hydrol (Amst)* **2000**, *230*, 244–257, doi:10.1016/S0022-1694(00)00214-6.

41. Liu, N.; Wang, H. Ensemble Based Extreme Learning Machine. *IEEE Signal Process Lett* **2010**, *17*, 754–757, doi:10.1109/LSP.2010.2053356.

42. Leach, J.M.; Coulibaly, P. An Extension of Data Assimilation into the Short-Term Hydrologic Forecast for Improved Prediction Reliability. *Adv Water Resour* **2019**, *134*, doi:10.1016/j.advwatres.2019.103443.

43. Dahigamuwa, T.; Yu, Q.; Gunaratne, M. Feasibility Study of Land Cover Classification Based on Normalized Difference Vegetation Index for Landslide Risk Assessment. *Geosciences (Switzerland)* **2016**, *6*, 1–14, doi:10.3390/geosciences6040045.

44. Conforti, M.; Ietto, F. Modeling Shallow Landslide Susceptibility and Assessment of the Relative Importance of Predisposing Factors, through a Gis-based Statistical Analysis. *Geosciences (Switzerland)* **2021**, *11*, 1–28, doi:10.3390/geosciences11080333.

45. Pearson's Correlation Coefficient. In *Encyclopedia of Public Health*; Kirch, W., Ed.; Springer Netherlands: Dordrecht, The Netherlands, 2008; pp. 1090–1091 ISBN 978-1-4020-5614-7.

46. Meddage, D.P.P.; Ekanayake, I.U.; Herath, S.; Gobirahavan, R.; Muttil, N.; Rathnayake, U. Predicting Bulk Average Velocity with Rigid Vegetation in Open Channels Using Tree-Based Machine Learning: A Novel Approach Using Explainable Artificial Intelligence. *Sensors* **2022**, *22*, doi:10.3390/s22124398.

47. Gupta, H. v.; Kling, H.; Yilmaz, K.K.; Martinez, G.F. Decomposition of the Mean Squared Error and NSE Performance Criteria: Implications for Improving Hydrological Modelling. *J Hydrol (Amst)* **2009**, *377*, 80–91, doi:10.1016/j.jhydrol.2009.08.003.

48. Rochford, P.A. SkillMetrics: A Python Package for Calculating the Skill of Model Predictions against Observations Available online: http://github.com/PeterRochford/SkillMetrics.

Chapter 4. Hybrid Surrogate Model for Timely Prediction of Flash Flood Inundation Maps Caused by Rapid River Overflow

Summary of the research article: Zanchetta, A. D. L. and Coulibaly, P. (2022) Hybrid Surrogate Model for Timely Prediction of Flash Flood Inundation Maps Caused by Rapid River Overflow. Forecasting 4(1):126-148, DOI:10.3390/forecast4010007.

In this research work, a hybrid machine learning type of structure composed by a Nonlinear AutoRegressive Neural network with eXogenous inputs (NARX) and a Self-Organizing Maps (SOM) network is implemented as a surrogate of a computationally expensive 2D hydrodynamics model for producing deterministic flood inundation maps. The main questions explored are:

- Are networks based in the hybrid NARX-SOM structure able to acceptably reproduce the outputs of a 2D hydrodynamic model for the same set of input hydrologic data in the scope of a catchment prone to flash floods?
- Is the surrogate model capable of producing high-resolution forecasts of flood inundation maps in a time interval compatible with real-time forecasting?

The key findings are:

- The evaluated surrogate models were able to mimic the behavior of the hydrodynamic model for lead times ranging from 30 minutes to 2.5 hours, which can be considered acceptable for the purposes of supporting decisions (e.g., related to the closure of roads).

- The surrogate models produce flood inundation maps for a period of 4 hours in approximately 13 minutes, which can be considered compatible with forecasting chains and is around 18 times faster than the 2D hydrodynamic model used as reference.

4.1. Abstract

Timely generation of accurate and reliable forecasts of flash flood events is of paramount importance for flood early warning systems in urban areas. Although physically based models are able to provide realistic reproductions of fast-developing inundation maps in high resolutions, the high computational demand of such hydraulic models makes them difficult to be implemented as part of real-time forecasting systems. This paper evaluates the use of a hybrid machine learning approach as a surrogate of a quasi-2D urban flood inundation model developed in PCSWMM for an urban catchment located in Toronto (Ontario, Canada). The capability to replicate the behavior of the hydraulic model was evaluated through multiple performance metrics considering error, bias, correlation, and contingency table analysis. Results indicate that the surrogate system can provide useful forecasts for decision makers by rapidly generating future flood inundation maps comparable to the simulations of physically based models. The experimental tool developed can issue reliable alerts of upcoming inundation depths on traffic locations within one to two hours of lead time, which is sufficient for the adoption of important preventive actions. These promising outcomes were achieved in a deterministic setup and use only past records of precipitation and discharge as input during runtime.

4.2. Introduction

The World Meteorological Organization estimates that flash floods are responsible for more than 5,000 human deaths every year, making them the most frequent and with highest mortality rate type of flooding [1]. The continuous densification of urban areas associated with the increasing recurrence of high intensity precipitation driven by climate change has the potential to intensify the frequency of such events [2–4]. To mitigate impacts, flash flood early warning systems have been implemented around the world since the 1970s mainly by governmental agencies (e.g., provincial or national flood/flow forecasting centers) [5], and the development of techniques and technologies to improve their predictive performance is an active research topic [6,7]. A pivotal component of such systems is the real-time forecasting of hazardous scenarios, which is decisive for decision makers on the task of choosing whether or not to take actions. Preventive actions such as closing roads, evacuating buildings, and changing the current operation of a dam have the potential to cause significant social disturbance and economical loss, thus a system that issues an excessive number of false alarms is undesired. To be operationally applicable, a forecasting system is expected to produce trustable and timely predictions, especially in the context of flash floods due to their rapid development, which happens "within six hours of the causative event (e.g., intense rainfall, dam failure, ice jam)", as defined by the United States' National Weather Service [5].

The capability of forecasting flood inundation maps (IMs) is usually considered a valuable asset to complement the usual river discharge hydrographs due to the direct representation of the locations to be (or not) inundated. As flash floods are characterized

by flows with high motion, realistic IMs of such events are attained through the simulation of physically based hydraulic models that solve the two-dimensional (2D) or quasi-2D Saint-Venant equation considering both mass and momentum conservation [8]. However, the computational cost of such models usually becomes a constraint for their direct use as part of toolchains of forecasting systems. One of the strategies developed for obtaining IMs in a time interval compatible to the update cycles of real-time flood forecasting systems is based on the development of simplified 2D models such as the Height Above Nearest Drainage (HAND) [9], AutoRAPID [10] and Safer_RAIN [11], which adopt the approach of "spilling and filling" digital elevation models at specific water heights. Despite providing realistic results for flood events of large scale with applicability demonstrated in real-time scenarios [12], the lack of physics for representing momentum results in limited applicability for rapidly evolving flash flood events [13,14]. Other modelling approaches developed for reducing the computational burden of high-resolution models that do not neglect the momentum component include sub-grid techniques [15,16], the use of cellular automata frameworks [17], and the development of new systems and models specifically designed to explore the processing power of graphic processing units [18,19]. The implementation of such promising techniques, however, may require expensive tasks of reimplementing already existing hydrological models using such new tools and the training of the technical team for the use of such frameworks.

In this context, the development of response surface surrogate models (hereafter simply referred to as "surrogate models") have been explored as workarounds to rapidly estimate realistic IMs. Surrogate models are mathematical systems calibrated to map values in the

input space to values in the output space of more complex, physically based and computationally expensive simulation systems, thus reproducing their behavior through the use of a simplified set of equations derived from statistical or data-driven approaches [20].

The use of machine learning techniques for the development of surrogate inundation models has presented significant evolution in recent years. Here, a hydraulic model in which both mass and momentum components of the surface water flow are present is implemented for the region of interest. Such a hydraulic model is assumed to be a realistic representation of the environment and is used to simulate the behavior of the system when forced with a wide range of realistic forcing values. The resulting IMs are stored, and a data-driven system is then trained to map each forcing data used to the respective IM produced. The high number of 2D cells in regular IMs to which a water depth value must be estimated, associated with the costly task of generating a large number of simulation products, usually leads to the challenging issue of properly tuning very complex models with limited input/output datasets. A common approach to reduce the number of variables to be predicted by the data-driven method is to train multiple independent simpler networks covering smaller parts of the grid, be it one model trained for each individual grid cell [21] or one model per groups of neighboring grid cells [22]. Such a strategy has as drawbacks: (1) the need to train a potentially extensive number of subnetworks, which can and up being a constraint in operational environments when the hydraulic/hydrological model is continuously updated (thus demanding continuous retraining of the surrogate model); and

(2) the fact that the spatial correlation of nearby grids trained by independent predictors is lost, potentially resulting in unrealistic discontinuities in the estimated IM.

In this context, the use of a Self-Organizing Maps (SOM) model was proposed as a dimensionality reduction method by Chang et al. [23] for regional flood inundation forecasting triggered by river overflow. In their work, SOMs are trained to cluster the inundation maps of the training dataset so that the topologic nodes of the network hold what can be interpreted as representative inundation instant conditions of the covered area. The average inundation depth (AID) of each topological node is calculated and stored as an associated variable after the tuning step so that a univariate predicted AID value can be used for querying complete IMs during forecasting time. A recurrent model is then trained to predict the univariate AID variable out of descriptive hydrologic observed features to provide the temporal development of the event. Chang et al. [23] reported promising results for the prediction of inundation maps with spatiotemporal resolution of 75 km/3 h for a lead time of up to 12 h, a configuration suitable for the regional scope evaluated in their work (total area of 100 km^2 in the outlet of the nearly 2,000 km^2 Kemaman River Basin, Malaysia). Kim et al. [24] applied a similar approach on a finer grid resolution and assessed its effectiveness on timely predicting flash flood inundation scenarios caused by direct runoff generation and consequent sewer overflow in a densely occupied urban area in Seoul, South Korea, under heavy rain. In their work, the overflow of 103 manholes were used as the querying variables and 122 rainfall events were applied for training the networks. Consistent results were obtained when the analysis was later extended to 24 districts [25]. Nonlinear Autoregressive neural network with exogenous inputs (NARX) is

the recurrent structure commonly chosen by the abovementioned works due to its simplicity and efficiency.

To the best of the author's knowledge, results for the application of such a hybrid approach are yet to be reported for urban catchments prone to flash flood events caused by fast-response river overflow. Our study aims to fill this research gap by evaluating the applicability of a hybrid NARX-SOM structure to surrogate a coupled hydrologic/hydraulic model in the prediction of IMs in a timely and accurate manner compatible with operational purposes for the Don River Basin in Toronto, Canada.

4.3. Study Area

The Don River basin (DRB), located in the Greater Toronto Area, Ontario, Canada, has a total drainage area of nearly 350 km^2, approximately 300 km^2 of which is located upstream from the river gauge HY019. The baseflow of the Don River recorded by HY019 is approximately 4.5 m^3/s. As represented in Figure 4-1, its landcover is characterized predominantly by urban infrastructure (80%), with disperse portions of sparse tree clusters, wetlands and crops [26]. The region just downstream the HY019 river gauge is often inundated during flash flood events with reports of stranding cars and urban trains in streets and railroads along the riverbanks of the Don River [27,28]. Two locations were selected as points of interest (POI) for further analysis: POI 1 is a segment of the local urban train railway, while the POI 2 is a section of the Don Valley Parkway.

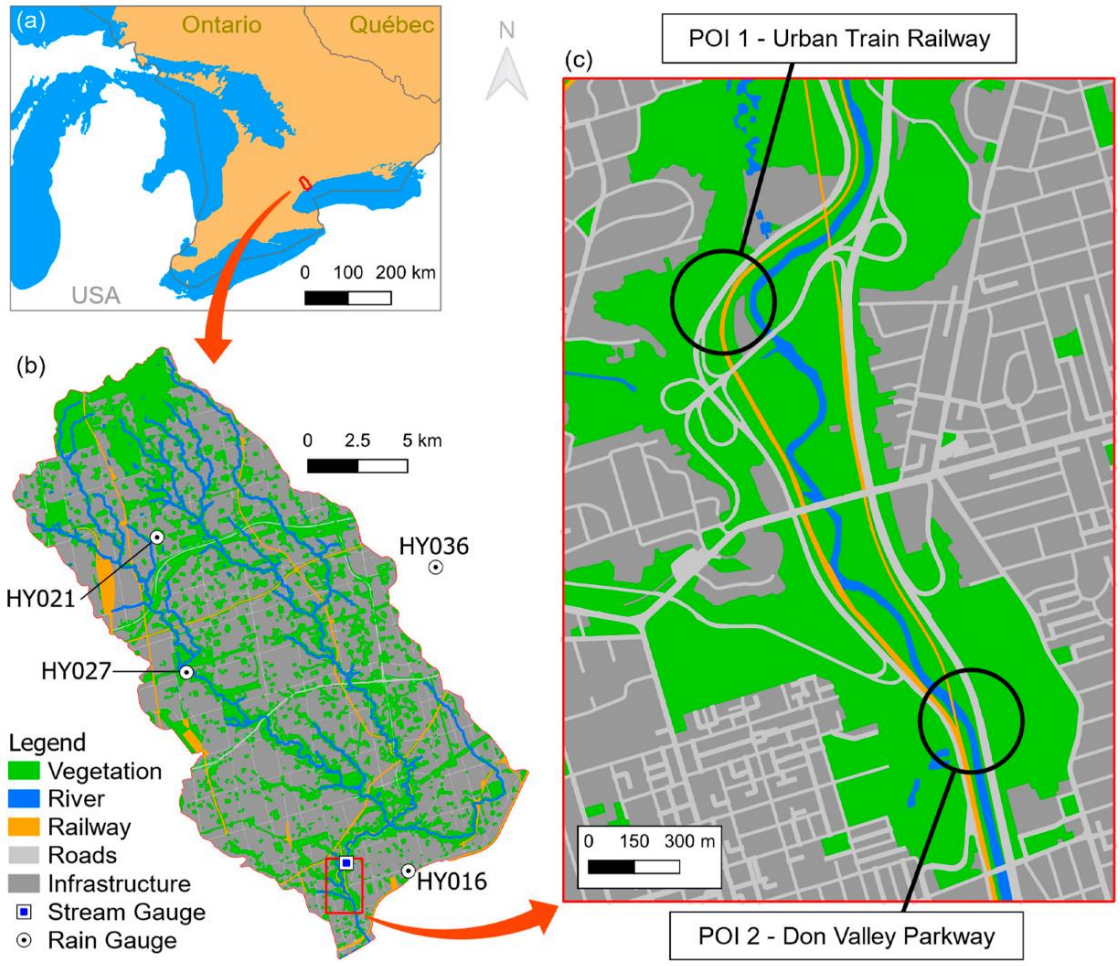

Figure 0-1. Representation of the (a) location of the Don River Basin, (b) its landcover with stream and rain gauges, and (c) the domain of the 2D hydraulic model with the two points of interest (POI) considered in this study.

Precipitation data from four rain gauges (HY016, HY021, HY027, HY036) and river discharge data from the river gauge HY019, both with 15 min resolution and ranging from 2011 to 2019, were used to select high-flow rainfall-runoff events. As this study is focused on pluvial flash floods, only the data of the warm seasons (from May to October) of each year were considered to avoid potential influence of snowfall and snowmelt.

4.4. Materials and Methods

4.4.1. Physically Based Model Used as Reference

A calibrated semi-distributed hydrologic model of the entire DRB (Figure 4-1b) was initially developed and provided by the Toronto and Region Conservation Authority (TRCA). The model is based on the Storm Water Management Model (SWMM) [29] and was implemented with the PCSWMM software [30]. The basin is discretized into 462 sub-catchments, each of them set to receive the same precipitation timeseries observed by its nearest precipitation gauge as if it was spatially uniform within the sub-catchment. Water flow is routed through 2,703 SWMM conduits (river segments) that reproduce the major rivers and pathways by solving the complete one-dimensional Saint-Venant equation, thus considering the momentum component but without the explicit modeling of surface inundation spread.

The original DRB model was modified to include a hydraulic surface flow component (hereafter referred to as "hydraulic component") of the aforementioned flood-prone area (Figure 4-1c). The Digital Elevation Model (DEM) used was obtained from the aggregation of LiDAR products originally provided publicly by the governmental agency Natural Resources Canada at 1 m resolution. The spatial aggregation into 2 m was performed to reduce the negative impact that relatively small surface artefacts, such as sudden spikes or depressions generated by signal noise, may cause to the final simulation results as they can be wrongly interpreted as barriers to the surface water flow [31,32].

In PCSWMM, surface flow is simulated through a dense network of regular SWMM components. A mesh of junctions is established from the underlying DEM, each junction being connected to up to 6 neighbors by 1-segment conduits. Each junction reports the water depth of a location, thus acting (and hereafter referred) as “surface flow cells”. The conduits are modeled as open rectangular cross-sections and perform bi-directional 1D flow routing between two neighboring surface flow cells. Such a configuration allows the simulation of both the lateral and longitudinal water flows needed to represent a 2D surface water movement by solving multiple one-dimensional equations — an approach usually referred to as “quasi-2D”. The flow under crossing bridges is represented by concrete conduits with closed rectangular cross-sections. Manning’s roughness coefficient of the river channel and surrounding vegetated areas were set as 0.01 and 0.05 following the values suggested by Ricketts et al. [33], for open-channel flow over concrete and over light brush on banks, respectively. To represent the flow resistance offered by piers, Manning’s roughness in the channel segments under passing bridges was set to 0.035. Buildings were represented as barriers for the water in the form of regions in the 2D space without surface flow cells.

The hydraulic component of the coupled model is composed by a total of 101,577 cells, 6394 of which are within the river boundaries, 40,482 are floodable cells (i.e., land cells that were inundated in at least one simulation), and the remaining cells are land cells that were not inundated in any simulation.

4.4.2. Self-Organizing Map (SOM)

SOMs, also known as Kohonen maps after its proposer [34], are among one of the most traditional unsupervised clustering methods in machine learning. An SOM is a specific structure of feed forward artificial neural networks in which the output nodes, representing the clusters, are organized in a 2D topological map. As a clustering technique, each output node of the network refers to a cluster identified for the dataset and has the same number of features as the number of input features. The output nodes in the topological map (hereafter referred simply as "topological nodes", or TNs) are tuned to be similar to their topological neighbors and, conversely, dissimilar to topologically distant nodes. The process of training starts with a network with randomly defined weights. The records in the training set are successively clustered by the network, i.e., the TN most similar to the clustered record, also called winner node, is identified using a distance function. The winner node then has its weights updated to become more similar to the clustered record through the reduction of the distance value between them by a learning rate α. The other TNs are updated with learning rates that gradually reduce as their topological distance to the winner node increases. Such a gradual reduction is governed by a neighborhood function with parameters defined by the user.

One of its most highlighted features is the capability to successfully cluster [35,36] and provide visual insights [37,38] on high dimensional datasets, which are assets for problems related to flood IMs due to the extensive number of water depth values predicted, one for each inundation cell in the domain. Other applications in water resources include but are not limited to the regionalization of groundwater types [38] and of hydrologic

homogeneous watersheds [39], the simulation of daily precipitation [40], and the support of decisions for the operation of reservoirs [41]. The reader is referred to the work of Clark et al. [42] for a further discussion on the use of SOMs in water resources.

The number of TNs of an SOM is defined by the designer during the model implementation step. As a clustering technique, the higher the number of TNs (clusters), the higher is the expected sharpness of the results and the lower is the degree of generalization of the patterns detected. In this work, SOMs with 12 TNs organized in a 3 × 4 conventional rectangular shape were trained using the water depth values of the 40,482 floodable cells as input features so that each output node can be interpreted as a complete representative IM itself (Figure 4-2). The total number of 12 TNs was determined heuristically based on the rule-of-thumb suggested by Kohonen [43] of an average of 50 training records per node. Considering the high number of features involved, the total number of training records per node was increased to 100 for a more consistent detection of patterns in the data. The networks were constructed and trained with the MiniSOM library for Python [44] using Gaussian neighborhood and Euclidian distance functions.

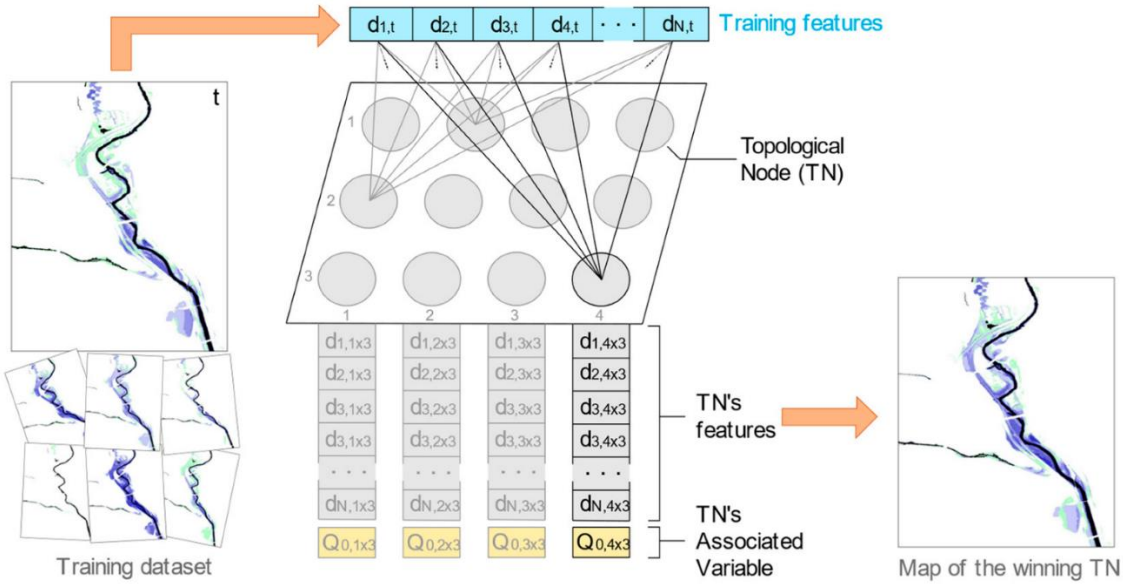

Figure 0-2. Illustration of an SOM utilized in this work during the training step, when the IM of time t, composed by N (40,482) instant water depths $d_{1,t}$, $d_{2,t}$, …, $d_{N,t}$, is associated to the TN 4 × 3 by similitude. The values of such winning TN ($d_{1,4\times3}$, $d_{2,4\times3}$, …, $d_{N,4\times3}$) are updated following the new learned values.

To ensure that all records in the train/validation dataset are used for training the data-driven models, 9 SOMs were tuned following the k-fold cross-validation approach, i.e., the records in the train/validation dataset were split into 9 subgroups, then each SOM was trained using 8 of such groups (fold in) and validated against the remaining subgroup (fold out). Hereafter, SOM_f indicates the SOM validated against the fold-out group f. The choice of splitting the records into 9 subgroups was done so that the conventional proportion of 90% training and 10% validation records was achieved (the effective dataset has 17 rainfall events, as discussed in Section 4.1). The training and validation steps are performed concurrently so that, after every fold-in record is presented to the SOM in learning mode, all records in the fold-out group are clustered in prediction mode and the sum of their distances to their respective winner nodes is calculated as Δ_t. The fold-in records have their

order shuffled and the process is repeated so that Δ_{t+1} is calculated. The iteration continues indefinitely while gains in similarity with the validation dataset are observed, i.e., $\Delta_t > \Delta_{t+1}$. Such approach is usually referred as "early stopping criteria" and tends to improve the generalization power of data-driven models [45].

4.4.3. Recurrent Network

4.4.3.1. Phase Space Composition

The objective of the recurrent models in this work is to predict the best topological node (and, consequently, the best IM cluster representative) in a specific time in the future given the limited number of hydrological features available in the present. For such, the available dataset needs to be organized in a tabular format so that all the predictive features (inputs) and the respective topological node expected to be predicted (output) are in the same tuple.

Three types of hydrological data were chosen to compose the input set: precipitation (P), discharge (Q) and average inundation depth (AID). P is given as the accumulated average areal precipitation in the sub-catchments of the hydrological model. Initially, at 15 min resolution, P is reconstructed so that P_t is the accumulated precipitation in the interval $[t - 2\text{ h}, t)$. Q is the discharge value estimated by the hydrological model in the Don River at the input of the hydraulic domain. Originally at 15 min resolution, Q is reconstructed so that Q_t is the mean of the discharge values within the interval $[t - 30\text{ min}, t)$. The instantaneous AID_t of each IM_t calculated by the hydraulic model for time t is given simply by:

$$AID_t = \sum_{n=1}^{N} \frac{d_{n,t}}{N} \ , \tag{4-1}$$

in which $d_{n,t}$ is the water depth in the nth floodable cell at time t and N is the total number of floodable cells in the model.

Each IM_t of the train/validation dataset assigned to a record in the fold f was clustered by SOM_f (Section 4.4.2), thus being associated to a winning topological node w of SOM_f ($TN_{w,f}$). Then, 17 phase spaces were constructed, one for each lead time $L \in [0, 15, 30, \ldots, 240]$ (values in minutes), so that a record for the time t in the phase space for L is composed by the tuple: P_t, Q_{t-1}, Q_t, AID_{t+L-1}, IM_{t+L}, $TN_{w,f}$. The first four elements compose the predictor feature set while the last is the predictand. The data organization is illustrated in Figure 4-3.

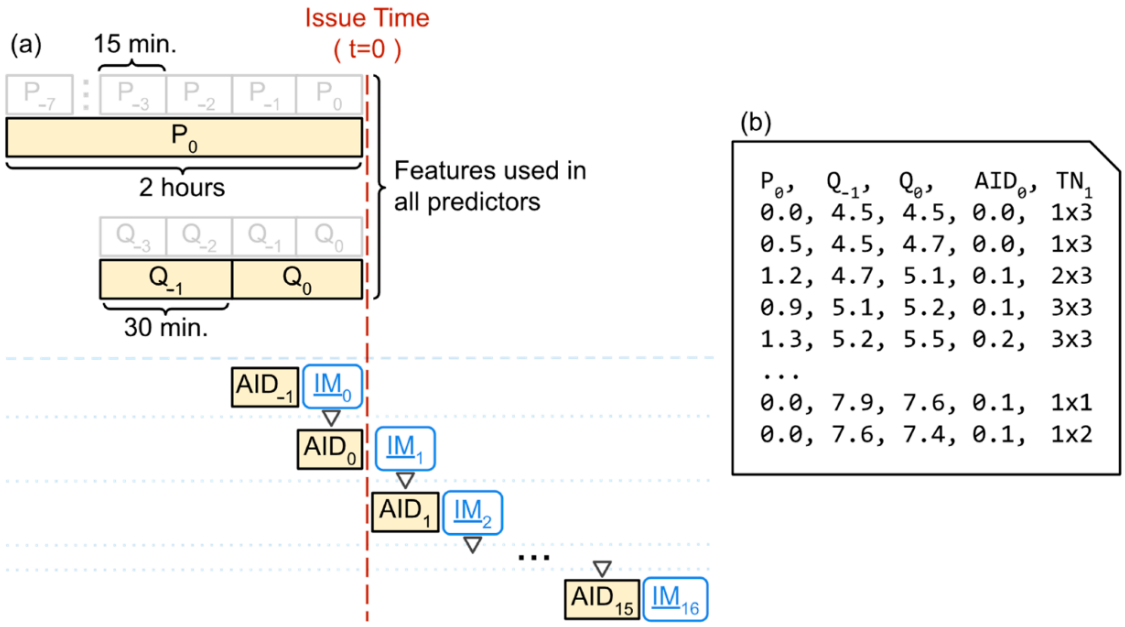

Figure 0-3. Organization of the phase space used as input for the NARX. In the scheme (a), data in the state space (light grey boxes) are aggregated and used as predictive features (yellow boxes) in different combinations to predict IMs (blue rounded-corner boxes). In

the output file (b), the first four columns (P_0, Q_{-1}, Q_1, AID_0) are the predictors and the last is the target (TN_1).

4.4.3.2. Nonlinear Autoregressive Neural Network with Exogenous Inputs (NARX)

A NARX is a simple recurrent model structured as a conventional feed forward neural network. Its autoregressive component is due the fact that part of the input features depends on the output values predicted by the same model for an antecedent time step. Additional exogenous explanatory variables are included in the input set to provide more information to the model. NARX models can be generically described as a mathematical function $f()$ given by:

$$\hat{y}_t = f(\hat{y}_{t-1}, \hat{y}_{t-2}, \hat{y}_{t-3}, \dots, \hat{y}_{t-M_y}, x_{t-L-1}, \dots, x_{t-L-M_x}) + \varepsilon_t \quad , \qquad (4\text{-}2)$$

in which $\hat{y}_t$ is the value predicted for time t, L is the lead time of the forecast (always positive), x_t is an exogenous variable at time t, M_y and M_x are the number of previous time steps considered (memory) of the variables y and x, respectively, and ε_t is the estimation error at time t that is minimized during the training step. It is worth noting that: (1) multiple exogenous variables can be part of the input feature set, and (2) that the recurrent values may be derived from the previous output values, i.e., an input feature can be, for example, $g(\hat{y}_{t-1})$ instead of the direct output $\hat{y}_{t-1}$.

In the original work of Chang et al. [23], a NARX model is trained and used to estimate future AID values using the most recent three-hour accumulated precipitation records observed in three sub-catchments in addition to the last AID value estimated. Such configuration is set up four times, one for each of the four lead times forecasted.

As described in Section 4.4.3.1, the configuration adopted in our work uses P_t, Q_{t-1}, Q_t, $AID_{n,\mathrm{t+L-1}}$ to estimate the probability for each TN to be the winner ($TN_{w,\mathrm{t+L}}$) for such inputs. From such a set of probabilities and the associated IMs of each TN, the respective $IM_{t+\mathrm{L}}$ is retrieved through interpolation or extrapolation (Section 4.4.5), from which the $AID_{t+\mathrm{L}}$ required for the prediction of the following step is estimated. Similar to the SOMs configuration, 9 sets of NARX models were tuned according to the training/validation dataset groups. Each set f is composed by 17 $NARX_{f,\mathrm{L}}$ models, one for each phase space constructed towards a lead time L.

All models trained were composed by a three-layer feed forward neural network structure (one input, one hidden and one output layers). Rectified Linear Unit (ReLu) [46] was set as the activation function of the input and hidden layers, while Softmax [47] was used as the activation function of the output layer. The size of the output layer was 12, each output node of the $NARX_{f,\mathrm{L}}$ matching a topological node of the respective SOM_f model. In such a configuration, the 12 values predicted by the Softmax function represent the probability of each SOM topological map to be the correct one for a given input. The networks were trained to optimize the Categorical Cross-Entropy loss function, which is usually considered as a standard loss function for multi-class classification tasks. The criterion adopted for early stopping is the increase of loss in the validation dataset between each epoch.

The number of nodes in the hidden layer was not defined a priori. For each combination of f and L, 8 neural networks were trained, each with a different number of hidden nodes

ranging from 5 to 12. The configuration that presented the best (lowest) validation loss was taken as the $NARX_{f,\mathrm{L}}$ and the others were discarded.

4.4.4. Updating SOMs with Associated Variables

After training, the SOMs were updated to have values of Q_0 linked to their topological nodes as additional variables. $Q_{0,\mathrm{t}}$ is the instantaneous discharge value at time t, thus each IM_t has a respective $Q_{0,\mathrm{t}}$. The Q_0 of a topological node TN is calculated as the mean of the $Q_{0,\mathrm{t}}$ values associated with all IM_t clustered as part of TN.

4.4.5. Hybrid Model Structure

Each pair of $NARX_{f,\mathrm{L}}$ and SOM_f is considered a hybrid model that receives the 4-variables input for the NARX models and outputs an IM derived from the SOM topological nodes for the lead time L. The 9 hybrid models set up for a lead time L are used simultaneously to produce ensemble forecasts. Such a strategy is often named multiple predicting cross-validation and the merged product of the ensemble members tends to be more consistent than the prediction performed by a single model through conventional approaches [48]. A schematic representation of the reproduced forecasting system applying the dual NARX-SOM is presented in Figure 4-4. Due to the lack of descriptive data of observed inundation flood maps, the discharge produced by the hydrologic model was assumed to be a realistic response from the catchment to a forcing precipitation data. In an operational environment, however, the simulated discharge data can be replaced by observations if real-time stream gauges are timely available, or discharge forecasts produced by the hydrologic model.

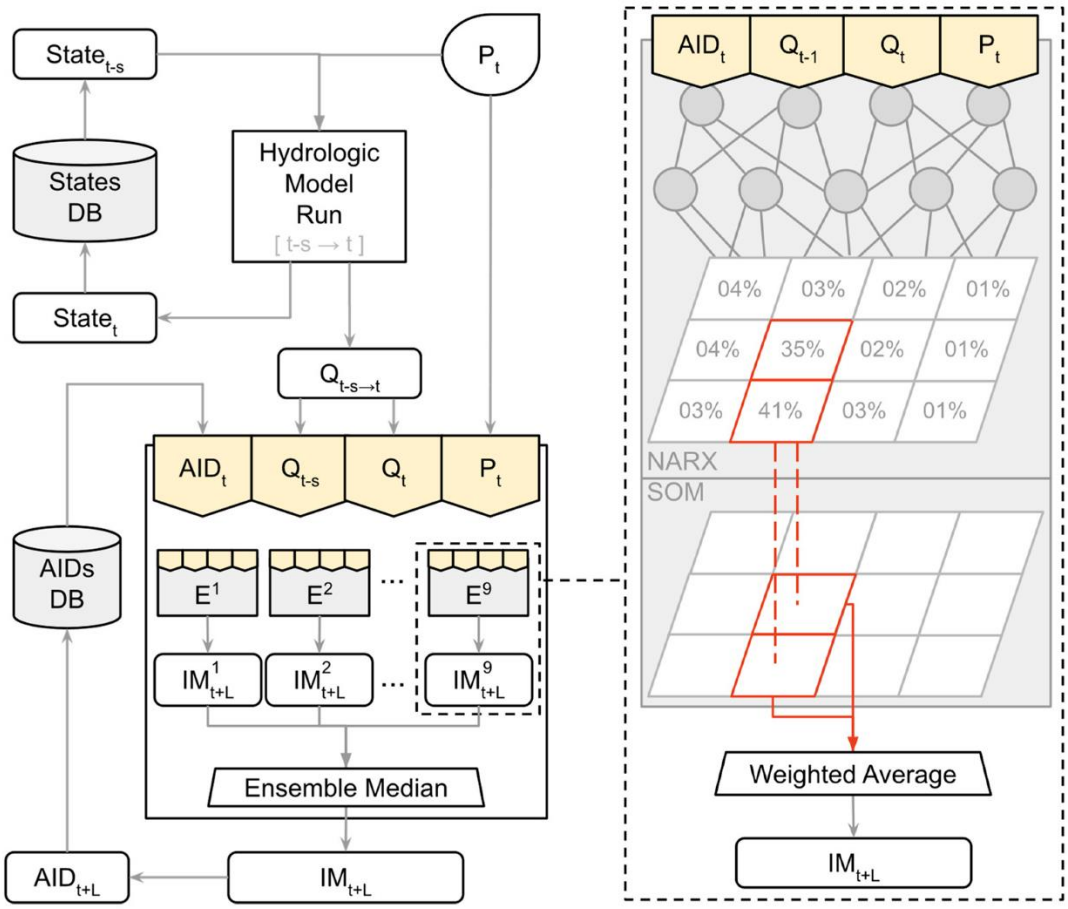

Figure 0-4. Overview of the system in the runtime setup demonstrating how the inundation map (IM) forecasted for a lead time L is obtained from the current data available at time t and from previous steps s. It represents an ensemble system composed by 9 predictors (E^1, …, E^9), NARX components have 4 input and 5 hidden nodes (interconnected circles), and SOMs have 3 × 4 topological nodes (square grids).

The 9 individual inundation maps forecasted for a certain lead time are aggregated into a single, deterministic map through the Conditional Median/Mean Function (CMMF). CMMF simply defines the water depth value of the floodable cell n as either the mean or the median value of the water depths predicted by all ensemble members to the same node n, represented as the multivariate D_n, by:

$$CMMF_n(D_n) = \begin{cases} 0 & if\ median(D_n) = 0, \\ mean(D_n) & otherwise. \end{cases} \quad (4\text{-}3)$$

CMMF is used as an alternative to the direct application of the elementwise mean value to ensure that nodes more likely to be dry (i.e., that had its water depth predicted to be zero by most of the ensemble members) are not set as wet due to the results of a minority of the members. The aggregation of multiple ensemble forecasts has the potential to smooth out oscillations detected in independent realization that are potentially caused by overfitting [22].

The topologic node with higher value of AID of each SOM model trained is labeled as the most extreme inundation scenario learned. Such identification is used to determine whether the predicted inundation map should be estimated by the interpolation or extrapolation of the learned patterns. If the winner topologic node is the most extreme one and its estimated probability is above 50%, then it is assumed that the input will produce an inundation condition potentially not covered in the train/validation dataset and thus an extrapolation from the most extreme scenario shall be performed. The extrapolation is performed by the multiplication of the water depths of the most extreme inundation scenario learned by a factor α, given by:

$$\alpha = \frac{\mathrm{Q}_0}{TN(Q_0)_{1,L}} * (1 - \frac{\beta * L}{\Delta_t}) \quad , \tag{4-4}$$

in which Q_0 is the most recent input discharge in the as predictive feature set, $TN(Q_0)_{1,\mathrm{L}}$ is the most recent input discharge associated with the winner topologic node, β is a decay factor, L is the lead time of the forecast, and Δ_t is the temporal resolution. The inclusion of a decay factor aims to simulate an expected reduction of future input discharge after the observation of extreme conditions (the higher the value of β, the more intense is the decay).

The β value used influences the performance of the maps generated outside of the training data space and reflects the expectations of the implementers concerning the intensity of possible future extreme rainfall events. In this work, a decay factor of 0.03 was utilized after experiments performed in the train/validation dataset.

For all other scenarios, the IMs of the two topologic nodes with highest estimated probabilities are interpolated by the weighted averaging, so that the interpolated IM estimated for a lead time L, INT_L, is calculated as:

$$INT_L = \frac{(TN(IM)_{1,L} * EP_{1,L}) + (TN(IM)_{2,L} * EP_{2,L})}{EP_{1,L} + EP_{2,L}}, \quad (4\text{-}5)$$

in which $TN(IM)_{1,\mathrm{L}}$ and $TN(IM)_{2,\mathrm{L}}$ are the water depth values of the floodable cells of the topologic node with the first and second highest estimated probabilities $EP_{1,\mathrm{L}}$ and $EP_{2,\mathrm{L}}$, respectively.

4.4.6. Train/Validation and Test Datasets

A total of 87 significant rainfall-runoff events were initially extracted from the available gauge data (described in Section 4.3). In this work, a rainfall-runoff event is characterized as "significant" if a precipitation record is followed by an increase in the discharge recorded at the HY019 gauge of 2 times the magnitude of the regular baseflow or more. All intense rainfall events were equally set to have total duration of 24 h, starting 12 h before and ending 12 h after its peak precipitation record. As the observed precipitation data were kept unmodified, this set of rainfall events are named RE-O (Rainfall Events—Observations) hereafter.

A set of 87 synthetic rainfall events were generated out of the RE-O dataset through the shuffling of the precipitation timeseries observations within the same event (e.g., as if the gauge HY036 had recorded what was recorded by the HY027 gauge, and as if HY027 had recorded what was recorded by HY021, and so on until all timeseries were set to a gauge different from its original gauge). Such a spatial disturbance of past observations through shuffling the location of observations is assumed to be a realistic "what-if" scenario. This set of records is named RE-SS (Rainfall Events—Synthetic by Shuffling) hereafter.

The 10 RE-O events with highest observed peak flows were selected to compose the RE-OH (Rainfall Event—Observed Highest) set. An additional augmentation was performed on the RE-OH events by a simple increase of 10% of all values of the timeseries. The resulting synthetic events represent scenarios of more intense, less recurrent events that could be observed but were not captured in the considered data period. This set is named RE-SE (Rainfall Events—Synthetic by Extrapolation) hereafter. A summary of the intermediate datasets is presented in Table 4-1.

Table 0-1. Summary of sub datasets.

Acronym	Description	Number of events
RE-O	Rainfall Events — Observations	87
RE-OH	Rainfall Events — Observations (Highest)	10
RE-SS	Rainfall Events — Synthetic by Shuffling	87
RE-SE	Rainfall Events — Synthetic by Extrapolation	10

All records of RE-O, RE-OH, RE-SS and RE-SE were used as the precipitation forcing for the hydrologic-hydraulic model starting from a baseflow steady condition. The simulated IMs and river discharge at HY019 were recorded with a temporal resolution of 15 min to match the temporal resolution of the precipitation dataset.

4.4.7. Evaluation Metrics

The performance of the NARX models on predicting the correct SOM topological node was performed through a binary approach using accuracy as the metric, i.e., the topological node deterministically predicted by the NARX model was compared to the topological node obtained by direct SOM classification of the respective inundation map. The total number of correct predictions (C_{pred}) and the total number of predictions evaluated (N_{pred}) were used to estimate the overall accuracy as:

$$Accuracy = \frac{C_{pred}}{N_{pred}} \quad . \tag{4-6}$$

Accuracy values range from 0 (no correct predictions detected) to 1 (perfect predictive power). Three metrics were used to evaluate the performance of the hybrid model so that average error, variance replicability and bias of the studied models can be assessed. The coefficient of determination (R^2) calculates the goodness of the fit of an estimation to a reference dataset in terms of variance replicability. It has values ranging from 0 (worse possible) to 1 (perfect) and is calculated as the square of the Pearson correlation coefficient ρ, i.e.,

$$R^2 = \left(\frac{\sum_{i=1}^{N}(y_i - \bar{y})\left(\hat{y}_i - \bar{\hat{y}}\right)}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2 \ \ \sum_{i=1}^{N}\left(\hat{y}_i - \bar{\hat{y}}\right)^2}} \right)^2 , \tag{4-7}$$

in which N is the total number of records, $\bar{y}$ is the mean value of the reference dataset, and y_i and $\hat{y}_i$ are the i-th reference and estimated values, respectively.

The root mean square error (RMSE) estimates the average error magnitude in the estimations from a model given in the same unit as the evaluated variable (meters in this work), with 0 meaning perfect match and no upstream value constraint. RMSE is given by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N}} \quad . \tag{4-8}$$

Bias is calculated by the relative percent difference of the predictions to the observations, being calculated by:

$$Bias = \sum_{i=1}^{N}\left(2\frac{\hat{y}_i - y_i}{|\hat{y}_i| + |y_i|}\right) \div N \quad . \tag{4-9}$$

A positive bias value represents overestimation while a negative value indicates underestimation, both bounded by 2 and −2, respectively. The closer the bias value is to 0, the less biased is considered the model. This approach to calculate bias was selected due to its tolerance to zero values, be it in the reference or in estimation.

Contingency tables synthetize the performance of a model in predicting the occurrence of a certain type of event through the number of true positives (or "hits", H), false negatives (or "misses", M), false positives (or "false alarms", F) and true negatives (or "correct no-alarm", N). In the context of contingency analysis of this work, the occurrence of an "event" is determined by the exceedance of the water depth in a floodable cell over a given threshold during any moment of a time interval. In this work, water depths of 10, 25, 50, 100 cm were considered as thresholds representing the respective conditions of minor, moderate, high and major threat for the traffic of cars and urban trains.

Contingency metrics considered in this work are the probability of detection (POD), the Success Ratio (SR) and the Critical Success Index (CSI) at the 2 POIS for the prediction of multiple maximum water depths. POD is a metric of sensibility and has values ranging from 0 (completely insensitive, i.e., unable to predict any tested event) to 1 (completely sensitive, i.e., able to predict all tested events) and is given by:

$$POD = \frac{H}{H + M}, \tag{4-10}$$

while SR is a metric of reliability also with values between 0 (completely unreliable, i.e., no forecasted events were observed) and 1 (completely reliable, i.e., all forecasted events were observed), and is calculated by:

$$SR = \frac{H}{H + F}. \tag{4-11}$$

There is a usual trade-off between the POD and the SR metrics: while models characterized by overestimated predictions tend to have high POD and low SR scores, models with underestimated prediction forecasts tend to present the opposite POD/SR relationship. As both sensibility and reliability are important features for predicting floods, such duality may turn the task of evaluating the overall performance of models to be subjective. The CSI metric can be seen taken as an objective metric that balances both the abovementioned characteristics and is calculated as:

$$CSI = \frac{H}{H + F + M}, \tag{4-12}$$

with values ranging from 0 (worse possible with no correct prediction) to 1 (best possible with no miss predictions). By not taking into consideration the component N, CSI values tend to be numerically lower than accuracy values in circumstances in which non-event

records significantly outnumber event records, as for the prediction of low recurrence flood events. Given such consideration, we assume in this work that a model must have a minimum CSI of 0.8 to be considered useful for flood forecasting from the perspective of decision makers. This can be considered a conservative threshold after the minimum accuracy of the same value as reported by [49] for a catchment also managed by TRCA.

4.5. Results and Discussion

4.5.1. Selected Rainfall-Runoff Events

Table 4-2 presents the overall characteristics of the 24 h events selected for the tuning and evaluation of the data-driven models. To reduce overbias of the model towards highly recurrent scenarios, several rainfall events with low maximum AID (lower than 0.15 m) were discarded. A total of 30 events were effectively used out of the 194 initial simulations performed, 17 of which composed the train/validation dataset while the remaining 13 composed the test dataset. The most extreme event, identified as *e41-vDr_x11*, was included in the test dataset so that the performance of the model for extrapolating results could be assessed. Figure 4-5 shows the AID timeseries of all selected events and illustrates the data distribution of each dataset, with the timeseries of the two events used as study case scenarios (Section 4.5.4.2) highlighted.

Table 0-2. Events characteristics.

		Total P (mm)		**Peak Q (m3/s)**		**Peak AID (m)**	
Dataset	**Events**	**Min**	**Max**	**Min**	**Max**	**Min**	**Max**
Train/validation	17	84.1	238.6	83.6	294.9	0.13	1.39
Test	13	95.2	246.9	64.1	247.9	0.06	1.52

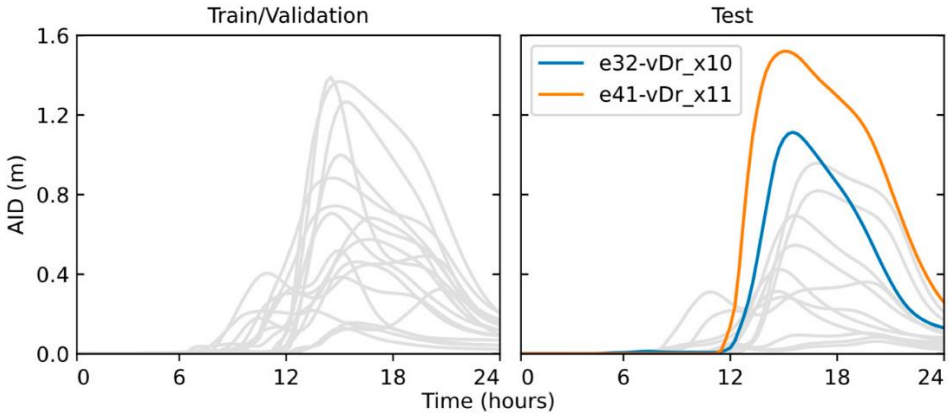

Figure 0-5. AID timeseries of the rainfall-runoff events selected to compose the train/validation (left) and test (right) datasets. Highlighted timeseries are from events selected as study cases.

4.5.2. Assessment of the SOM Models Trained

As discussed in Section 4.4.2, nine SOM networks were trained using the k-fold cross validation approach, each of which have 3 × 4 topological dimension and output node weights representing the water depth at individual cells of the 2D predicted inundation map. Figure 4-6 presents the topological nodes of one of the nine SOMs trained to illustrate their overall organization. As expected, it can be observed a gradual increase of the flood extension and overall water depth from the map with virtually no flooding (node 1 × 3) to the map related to the most extreme conditions (node 4 × 1). The consistency of the inundation maps obtained demonstrate that SOM is able to properly cluster high-resolution grids (i.e., high number of features) with a relatively limited number of inundation maps used as samples for training (1,320 IMs) and validating (176 IMs). Linear discontinuities through the inundated area and river path are due to the existence of crossing bridges, which are modeled in SWMM as closed conduits to represent their underneath water pathways instead of nodes that are interpreted as floodable cells.

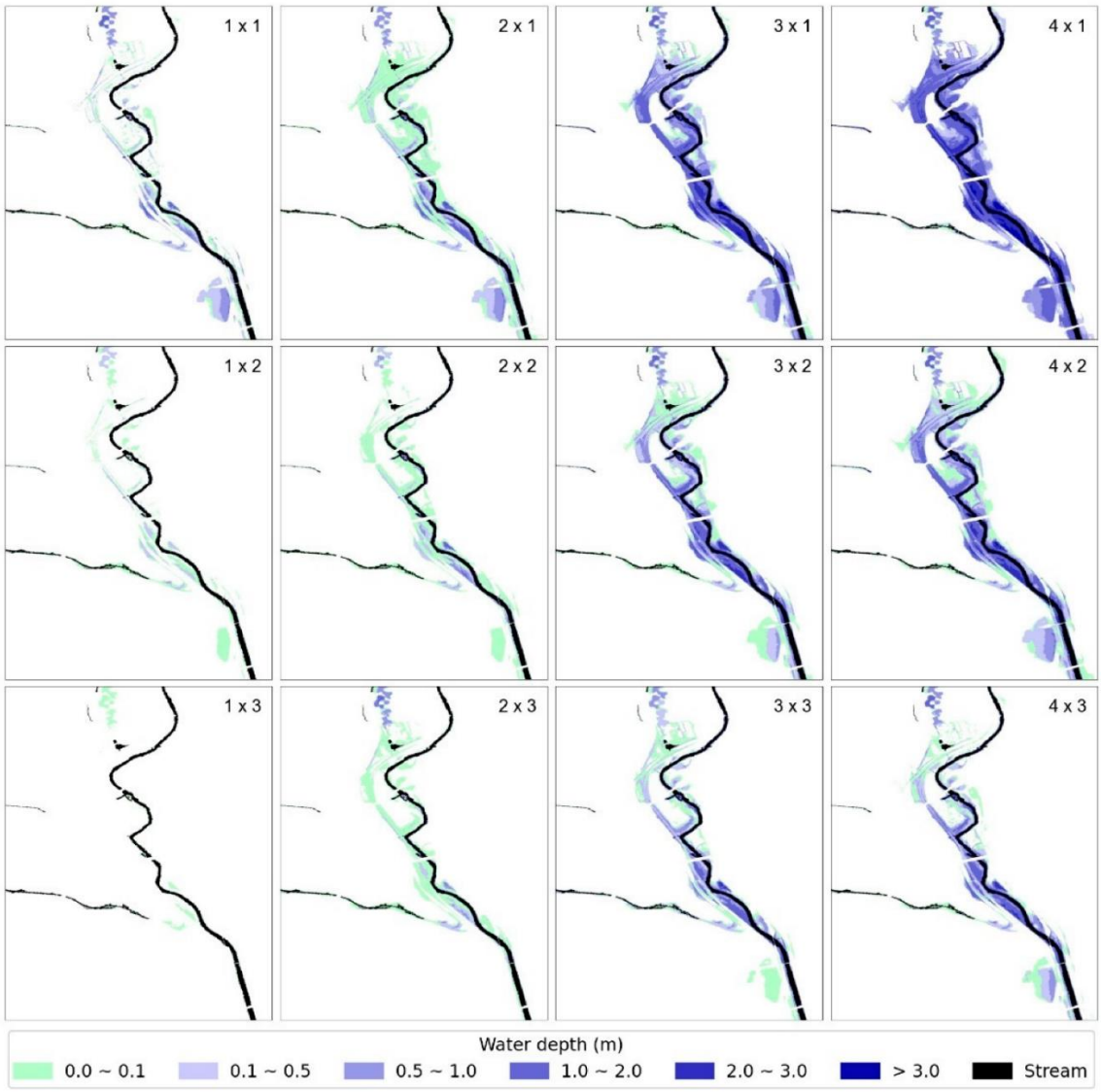

Figure 0-6. Topological maps of one of the SOMs trained.

It is also interesting to note that the four maps in the top left corner of Figure 4-5 may be taken as scenarios of concern for the POI 2 while all scenarios in columns 3 and 4 may result in the need of actions for POI 1. In an operational setting, such a configuration could be used as additional information to improve the perception of imminent risk through the interpretation that the topologically closer a predicted winner node is to these “clusters of clusters”, the higher should be the awareness level of the forecasters.

4.5.3. Assessment of the NARX Models Trained

As discussed in Section 4.4.3, eight NARX models were trained for each of the 17 lead times predicted for each of the nine folding groups, resulting in a total of 1,224 models trained, 153 of which were selected as the ones with the best performance. There was not a predominant set of hyperparameter values in the selected models, which corroborates with the general concept that each independent problem may have different optimal network structures.

The accuracy of the selected NARX models on predicting the correct topological node of its related SOM out of the hydrological features is presented in Figure 4-7. As expected, the predictive performance on the training dataset is superior to the performance on the validation dataset, but the relatively small difference between their accuracy values indicates an acceptable degree of generalization. As also expected, the overall performance of the classifier decreases as the lead time of the prediction increases from the lowest lead time up to a lead time of 2 h, after which the performance only oscillates around a constant value. As the response time of the catchment is also in the order of 2 h, the discontinuity can be explained by the influence of precipitation and discharge in the earlier lead times, while for longer lead times the performance is influenced almost completely by the recurrent variable of AID. The performance obtained was superior to 80% for all lead times; however, such a result must be observed with discretion as the AID values used as part of the predictive feature sets were calculated from the IMs simulated by the physically based model instead of being derived from estimations of the outputs from the surrogate

model for previous time steps. Such a configuration represents an ideal scenario in which the surrogate model reproduces with perfection the physically based model.

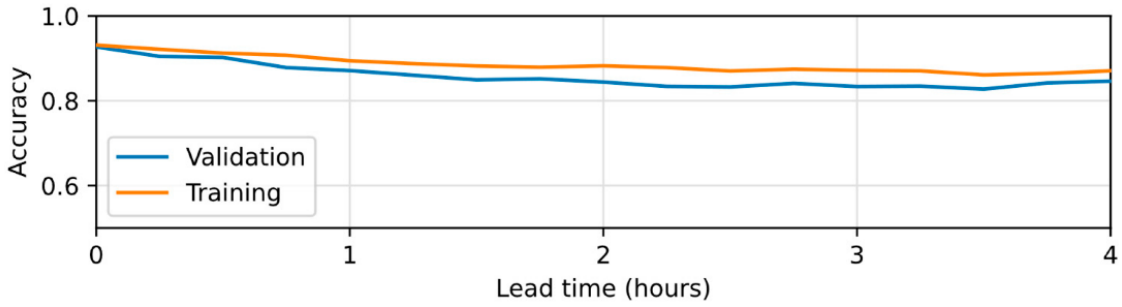

Figure 0-7. Accuracy of the NARX model in predicting the topological node of the SOM network in the cross-validation datasets.

It is worth noting that a misclassification may still be considered a useful classification if the predicted TN neighbored the correct TN, which would result in a low overall error due to their similarities. The global error of the combined model is explored in Section 4.5.4.

4.5.4. Assessment of the Hybrid Model

4.5.4.1. Global Performance

After both SOM and NARX models were trained using the train/validation dataset, the effective performance of the hybrid model was assessed using the testing dataset in an operational setup, i.e., with the value of the recurrent feature AID being obtained from the inundation map latest predicted in previous steps.

Figure 4-8 shows the performance of the forecasts in terms of R^2, RMSE and bias for the different lead times considered. The higher degradation in performance with respect to the lead time when compared to the NARX analysis is due the cumulative error carried

with the sequential estimation of AID values. Results were considered acceptable for the lead time of up to 2 h (120 min) in terms of both R^2 (above 0.8) and RMSE (bellow 0.25 m). All lead times presented low overall bias; however, the significant spread for lead times longer than 1.5 h (90 min) indicates a high level of uncertainty that can be considered a constraint for decision makers. The slight trend towards overestimation can be induced by the loss function used during the training of the NARX models, which prioritizes the correct classification of less recurrent (in our case, more intense) events over more common (and thus less intense) events.

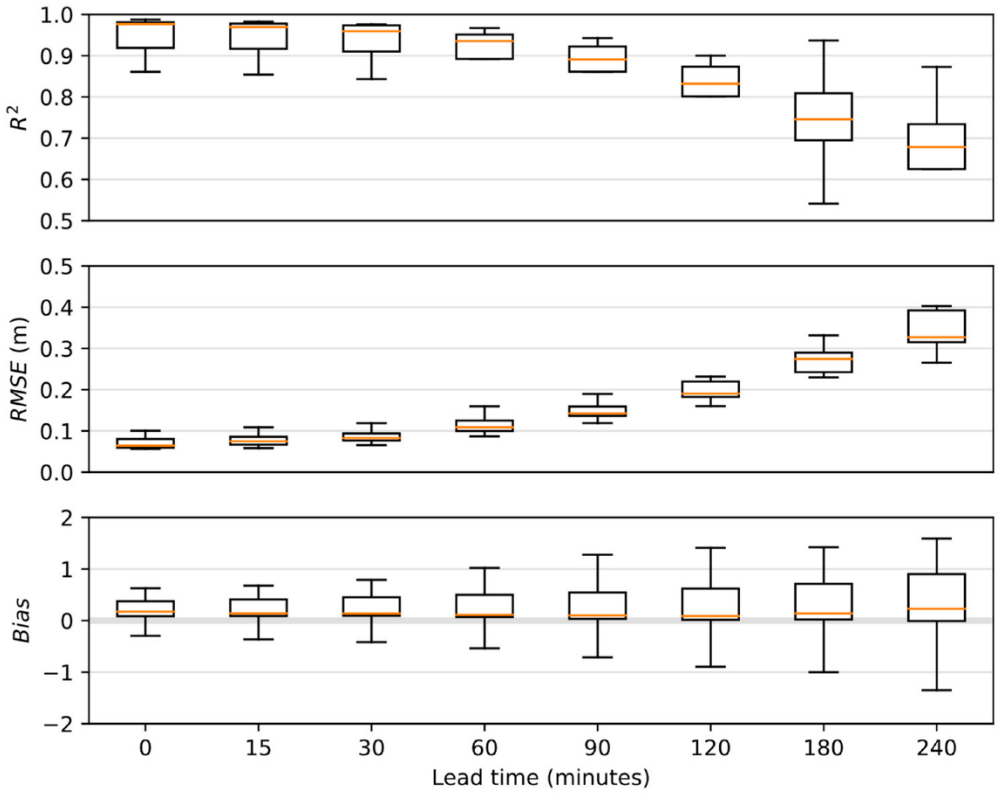

Figure 0-8. Overall performance of the hybrid model for the test dataset. Values were calculated for each event independently.

The forecasts were able to satisfactorily predict the occurrence of flood conditions at different depths on both POIs (Tables 4-3 and 4-4). The aforementioned tendency to

overestimate the forecast is reflected in highs level of sensitivity (POD), allowing the early detection of more than 90% of the upcoming events in most of the cases and increasing with longer lead times. However, as the sensitivity increases with longer lead times, the number of false alarms issued also increases, which disproportionately decreases the reliability (SR) of the predictions and consequently reduces the overall skill of the model (CSI). CSI is higher than 0.8 for maximum lead times ranging from 30 min (POI 1 for a threshold of 50 cm) up to 150 min (POI 1 for threshold of 25 cm), which indicates that the proposed system can be useful to forecasters. The overall better performance to detect events in the POIS 2 for early lead times when compared to the POIS 1 (Figure 4-9) can be explained by the different location and surrounding areas of each point. POIS 2 is located close to the margin of stream and is not subject to pounded water, which leads to its water depth to be highly correlated to the instantaneous river water depth. Conversely, the POIS 1 has its behavior more correlated to longstanding flow conditions due to its longer distance from the stream edge and its surrounding micro-relief to be subject to pounded water, which results in better prediction performances at longer lead times.

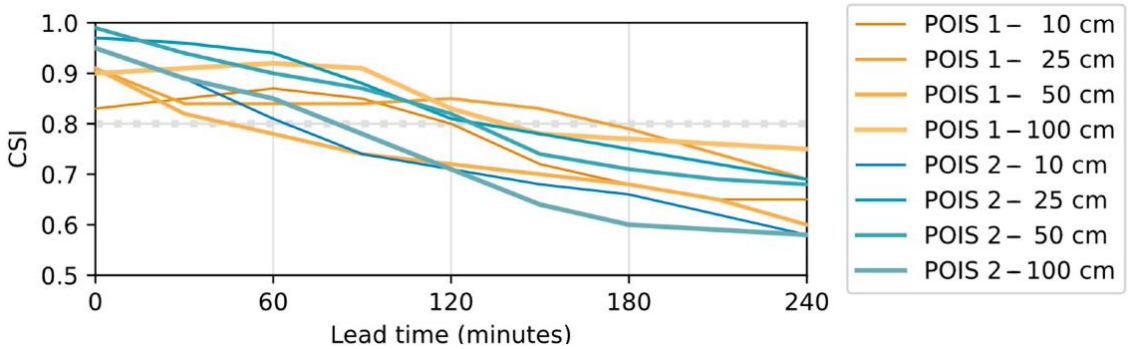

Figure 0-9. CSI values on predicting the exceedance of the water depth over different thresholds at the two POISs. A threshold of 0.8 CSI is considered as minimal for a forecasting system to be considered useful.

4.5.4.2. Performance on Selected Events

The two most extreme events of the testing dataset (highlighted in Figure 4-5) were selected as study cases. The first event (*e32-r-x10*) refers to the storm event observed during the night of 29 May 2013, which caused a temporary interruption of the urban train operations and the partial submersion of cars due to inundations in the POIs [50]. Such a rainfall event has a magnitude below the most extreme event in the train/validation dataset, thus it is expected to be reproduced by the hybrid model from the interpolation of its "known" events. The second event (*e41-r-x11*) simulates the storm event that was observed less than two months after the first one, on 9 July 2013, when several passengers had to be rescued from an urban train strained in the POI 1, and the road traffic was interrupted in the POI 2 due to river overbank conditions [51]. It has a magnitude above the most extreme scenario of the train/validation dataset, thus serves to assess the capabilities of the hybrid model to perform predictions through extrapolation of learned patterns.

Table 0-3. Contingency metrics at different lead times for the POI 1.

Metric	**Water Depth (cm)**	**Lead Time (minutes)**							
		0	**30**	**60**	**90**	**120**	**150**	**180**	**240**
POD	10	0.97	0.97	0.98	0.98	0.99	0.99	0.99	1.00
	25	0.98	0.98	0.99	0.99	0.99	0.99	0.99	0.99
	50	0.97	0.98	0.98	0.99	0.99	0.99	0.99	0.99
	100	0.91	0.93	0.95	0.96	0.97	0.97	0.97	0.97
SR	10	0.86	0.87	0.89	0.86	0.80	0.73	0.68	0.65
	25	0.93	0.85	0.85	0.85	0.85	0.83	0.79	0.70
	50	0.94	0.84	0.79	0.75	0.73	0.71	0.68	0.61
	100	0.99	0.98	0.97	0.94	0.85	0.81	0.79	0.76
CSI	10	0.83	0.85	0.87	0.85	0.80	0.72	0.68	0.65
	25	0.91	0.84	0.84	0.84	0.85	0.83	0.79	0.69
	50	0.91	0.82	0.78	0.74	0.72	0.70	0.68	0.60
	100	0.90	0.91	0.92	0.91	0.83	0.78	0.77	0.75

Table 0-4. Contingency metrics at different lead times for the POI 2.

Metric	Water Depth (cm)	Lead Time (minutes)							
		0	**30**	**60**	**90**	**120**	**150**	**180**	**240**
POD	10	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00
	25	0.97	0.99	0.99	0.99	0.99	0.99	0.99	1.00
	50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	100	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SR	10	0.95	0.90	0.81	0.74	0.71	0.68	0.66	0.58
	25	1.00	0.97	0.95	0.88	0.82	0.79	0.75	0.70
	50	0.99	0.94	0.90	0.87	0.82	0.74	0.71	0.68
	100	0.97	0.89	0.85	0.78	0.71	0.64	0.60	0.58
CSI	10	0.95	0.89	0.81	0.74	0.71	0.68	0.66	0.58
	25	0.97	0.96	0.94	0.88	0.81	0.78	0.75	0.69
	50	0.99	0.94	0.90	0.87	0.82	0.74	0.71	0.68
	100	0.95	0.89	0.85	0.78	0.71	0.64	0.60	0.58

A comparison between Ims produced by the physically based model for the event e32-r-x10 and the respective predictions generated by the surrogate model at a fixed issue time is shown in Figure 4-10. It is possible to observe that there is little deviation from the simulation in the estimation for the present-moment condition, and as that the error due to overestimation tends to increase as lead time increases. Despite such errors, the overall development of the inundation is properly reproduced, and decision makers responsible for closing (or not) the road at the POI 2 would be properly informed that a minor to moderate inundation scenario would occur in the upcoming 30 min, while a moderate to major condition would be attained within 60 min. A similar result can be observed at the POI 1 in spite of its early underestimation of the water depth.

The overall maps issued at different lead times for the peak inundation condition of the event *e41-r-x11* and their comparison with the maps produced by the physically based model are shown in Figure 4-11. As it can be observed, Ims with high overall similarity

are forecasted up to 2 h ahead of the most extreme condition despite some significant underestimation. Figure 4-12 shows a local view of the POI 2 with the maps produced directly by PCSWMM and by the surrogate model, illustrating how the overall inundation pattern is properly reproduced, and how the forecasters would be notified that the roads along the riverbank would have points in which water height ranges between 1 to 2 m. As the lead time decreases (i.e., as the issue time approximates the peak inundation time) and more precipitation and discharge data become available, the overall underestimation of the water depth also gradually significantly decreases, thus illustrating the effects of recent data on "correcting" previously issued predictions (Figure 4-11b,c,e,f).

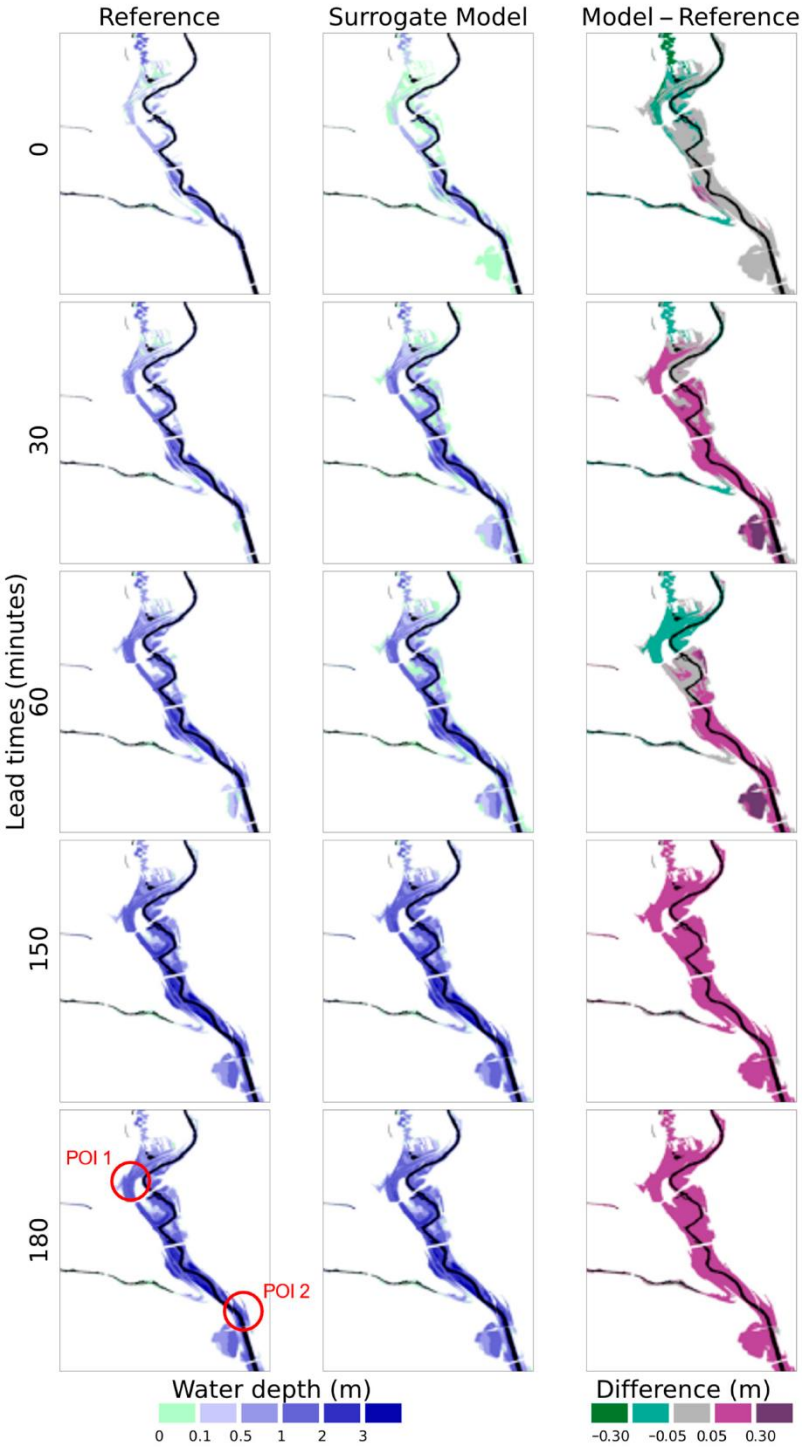

Figure 0-10. Inundation maps produced by PCSWWM used as reference compared with the maps forecasted by the NARX-SOM surrogate model issued on 150 before the peak inundation time of event *e32-r-x10*.

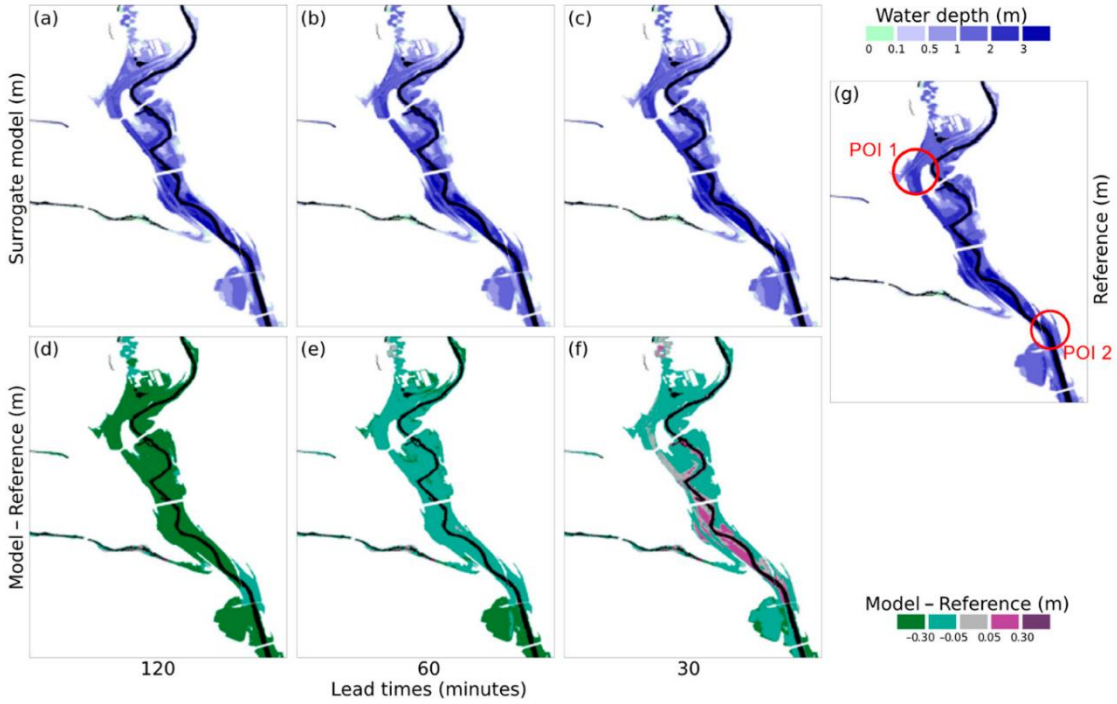

Figure 0-11. Flood inundation maps predicted by the NARX-SOM surrogate model to the same instant issued at different lead times (a–c) with their respective differences (d–f) to the map produced by the PCSWMM model used as reference (g) for the peak of the *e41-vDr_x11* event.

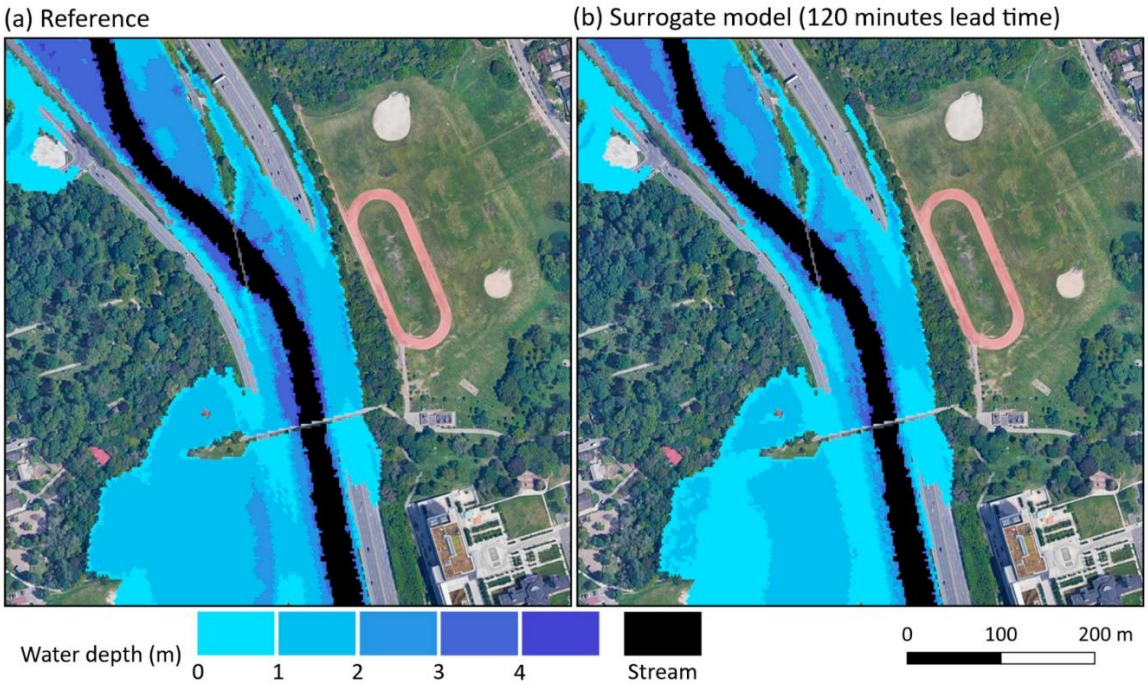

Figure 0-12. Flood inundation maps centered in the POI 2 of the Figure 4-11g (a) and of the Figure 4-11a (b) applied to a Google Earth satellite image (© 2021 Google).

4.5.5. Runtime Comparison

A computer with the following specifications was used to compare the runtime of both the hydrologic-hydraulic model and its surrogate counterpart: CPU Intel I9 with 3.6 GHz, eight cores and 16 logical processors; 64 GB memory RAM. The execution of the physically based hydraulic model for a simulation time of 4 h took 4.5 h to complete, confirming its unsuitability for real-time operational purposes. On the other hand, the generation of the inundation maps for all lead times up to 4 h ahead of the same event took the order of 11 to 15 min, which can be considered compatible to operational forecasting setups. It is worth mentioning that the hybrid models were executed without parallel processing and thus have the potential to have their execution time reduced if such techniques are applied.

4.5.6. Constraints and Limitations of the Methodology

As a surrogate model, the operational adoption of the proposed methodology depends on a pre-existing hydrological/hydraulic model to be available and properly configured. The implementation of such models, however, depends on the availability of detailed information of the study area, such as high-resolution DEMs and drainage systems, which may not be guaranteed for specific locations.

To maintain a proper replication of the physically based model used as reference, the surrogate model needs to be retrained every time the physically based model is modified. Such updates may be performed with certain recurrence to include, for example, significant structural changes in the covered domain (such as changes in the land use) or recalibrations to consider changes in the rainfall regimes induced by climate change.

4.6. Conclusions and Future Works

This work presents evidence that hybrid NARX-SOM models can be effective in predicting inundation maps for flash floods caused by river overflow of catchments with short response time. The promising results were obtained using a restrict number of training events, which can be considered an appealing feature as eventual retraining of the hybrid models would require limited re-simulations of the hydraulic model. Forecast products were considered valuable for decision makers for lead times ranging from 30 min to 2.5 h, which can be considered acceptable for the purposes of closing streets to traffic and stopping the operation of trains. To increase the lead time, the use of precipitation forecasts should be considered and are the potential subject 158uturee studies. In the current study,

multiple predictors were trained independently using the cross-fold approach and a simple aggregation of their predictions was used as deterministic forecasts. The production of probabilistic flood inundation maps [13] is of continued interest, and future works will explore the potentials of the dual NARX-SOM approach to produce inundation maps accompanied with consistent uncertainties. In addition, the presented results are limited to a single study scope and further experiments are expected to be developed to further assess its applicability to other scenarios.

4.7. References

1. World Meteorological Organization *Climate and water*; WMO: Geneve, Switzerland, 2020; Vol. 69;.
2. Lehmann, J.; Coumou, D.; Frieler, K. Increased record-breaking precipitation events under global warming. *Clim. Change* **2015**, *132*, 501–515.
3. Zhao, G.; Gao, H.; Cuo, L. Effects of urbanization and climate change on peak flows over the San Antonio River basin, Texas. *J. Hydrometeorol.* **2016**, *17*, 2371–2389.
4. 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.
5. Modrick, T.M.; Graham, R.; Shamir, E.; Jubach, R.; Spencer, C.R.; Sperfslage, J.A.; Georgakakos, K.P. Operational flash flood warning systems with global applicability. In Proceedings of the 7th International Congress on Environmental Modelling and Software: Bold Visions for Environmental Modeling (iEMSs 2014); San Diego, CA, USA, 2014; pp. 694–701.

6. Hapuarachchi, H.A.P.; Wang, Q.J.; Pagano, T.C. A review of advances in flash flood forecasting. *Hydrol. Process.* **2011**, *25*, 2771–2784.

7. Zanchetta, A.D.L.; Coulibaly, P. Recent Advances in Real-Time Pluvial Flash Flood Forecasting. *Water* **2020**, *12*, 570.

8. Costabile, P.; Costanzo, C.; De Lorenzo, G.; Macchione, F. Is local flood hazard assessment in urban areas significantly influenced by the physical complexity of the hydrodynamic inundation model? *J. Hydrol.* **2020**, *580*.

9. Nobre, A.D.; Cuartas, L.A.; Hodnett, M.; Rennó, C.D.; Rodrigues, G.; Silveira, A.; Waterloo, M.; Saleska, S. Height Above the Nearest Drainage - a hydrologically relevant new terrain model. *J. Hydrol.* **2011**, *404*, 13–29.

10. Follum, M.L.; Tavakoly, A.A.; Niemann, J.D.; Snow, A.D. AutoRAPID: A Model for Prompt Streamflow Estimation and Flood Inundation Mapping over Regional to Continental Extents. *J. Am. Water Resour. Assoc.* **2017**, *53*, 280–299.

11. Samela, C.; Persiano, S.; Bagli, S.; Luzzi, V.; Mazzoli, P.; Humer, G.; Reithofer, A.; Essenfelder, A.; Amadio, M.; Mysiak, J.; et al. Safer_RAIN: A DEM-based hierarchical filling-&-spilling algorithm for pluvial flood hazard assessment and mapping across large urban areas. *Water (Switzerland)* **2020**, *12*.

12. Hu, A.; Demir, I. Real-time flood mapping on client-side web systems using hand model. *Hydrology* **2021**, *8*.

13. 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.

14. Hocini, N.; Payrastre, O.; Bourgin, F.; Gaume, E.; Davy, P.; Lague, D.; Poinsignon, L.; Pons, F.; Bouguenais, F.-; Rennes, G.; et al. Performance of automated methods for flash flood inundation mapping : a comparison of a digital terrain model (DTM) filling and two hydrodynamic methods. **2021**, 2979–2995.

15. Jan, A.; Coon, E.T.; Graham, J.D.; Painter, S.L. A Subgrid Approach for Modeling Microtopography Effects on Overland Flow. *Water Resour. Res.* **2018**, *54*, 6153–6167.

16. Cao, X.; Ni, G.; Qi, Y.; Liu, B. Does subgrid routing information matter for urban flood forecasting? A multiscenario analysis at the land parcel scale. *J. Hydrometeorol.* **2020**, *21*, 2083–2099.

17. Nkwunonwo, U.C.; Whitworth, M.; Baily, B. Urban flood modelling combining cellular automata framework with semi-implicit finite difference numerical formulation. *J. African Earth Sci.* **2019**, *150*, 272–281.

18. Dazzi, S.; Vacondio, R.; Dal Palù, A.; Mignosa, P. A local time stepping algorithm for GPU-accelerated 2D shallow water models. *Adv. Water Resour.* **2018**, *111*, 274–288.

19. 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*.

20. Razavi, S.; Tolson, B.A.; Burn, D.H. Review of surrogate modeling in water resources. *Water Resour. Res.* 2012, *48*.

21. 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.

22. Berkhahn, S.; Fuchs, L.; Neuweiler, I. An ensemble neural network model for real-time prediction of urban floods. *J. Hydrol.* **2019**, *575*, 743–754.
23. 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.
24. Kim, H. Il; Keum, H.J.; Han, K.Y. Real-time urban inundation prediction combining hydraulic and probabilistic methods. *Water* **2019**, *11*.
25. Kim, H. Il; Han, K.Y. Data-Driven Approach for the Rapid Simulation of Urban Flood Prediction. *KSCE J. Civ. Eng.* **2020**, *24*, 1932–1943.
26. Ontario Ministry of Natural Resources and Forestry Ontario Land Cover Compilation Data Specifications Available online: https://geohub.lio.gov.on.ca/datasets/7aa998fdf100434da27a41f1c637382c.
27. Nirupama, N.; Armenakis, C.; Montpetit, M. Is flooding in Toronto a concern? *Nat. Hazards* **2014**, *72*, 1259–1264.
28. DailyHive News Rain causes flooding on low-lying Toronto highway ramps. Available online: https://dailyhive.com/toronto/highway-ramp-flooding-rain (accessed on Nov 10, 2011).
29. Rossman, L.A. *Storm Water Management Model - User's manual version 5.1*; US EPA: Cincinnati, OH, 2015;
30. Computational Hydraulics International - CHI PCSWMM. Available online: https://www.pcswmm.com/.

31. Meesuk, V.; Vojinovic, Z.; Mynett, A.E.; Abdullah, A.F. Urban flood modelling combining top-view LiDAR data with ground-view SfM observations. *Adv. Water Resour.* **2015**, *75*, 105–117.

32. Diakakis, M.; Andreadakis, E.; Nikolopoulos, E.I.; Spyrou, N.I.; Gogou, M.E.; Deligiannakis, G.; Katsetsiadou, N.K.; Antoniadis, Z.; Melaki, M.; Georgakopoulos, A.; et al. An integrated approach of ground and aerial observations in flash flood disaster investigations. The case of the 2017 Mandra flash flood in Greece. *Int. J. Disaster Risk Reduct.* **2019**, *33*, 290–309.

33. Ricketts, J.; Loftin, M.K.; Merritt, F. *Standard Handbook for Civil Engineers*; 5th ed.; McGraw-Hill Education, 2004; ISBN 0071364730.

34. Kohonen, T. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **1982**, *43*, 59–69.

35. Ultsch, A.; Lötsch, J. Machine-learned cluster identification in high-dimensional data. *J. Biomed. Inform.* **2017**, *66*, 95–104.

36. Araújo, A.F.R.; Antonino, V.O.; Ponce-Guevara, K.L. Self-organizing subspace clustering for high-dimensional and multi-view data. *Neural Networks* **2020**, *130*, 253–268.

37. Koua, E.L.; Maceachren, A.; Kraak, M.J. Evaluating the usability of visualization methods in an exploratory geovisualization environment. *Int. J. Geogr. Inf. Sci.* **2006**, *20*, 425–448.

38. Belkhiri, L.; Mouni, L.; Tiri, A.; Narany, T.S.; Nouibet, R. Spatial analysis of groundwater quality using self-organizing maps. *Groundw. Sustain. Dev.* **2018**, *7*,

121–132.

39. Farsadnia, F.; Rostami Kamrood, M.; Moghaddam Nia, A.; Modarres, R.; Bray, M.T.; Han, D.; Sadatinejad, J. Identification of homogeneous regions for regionalization of watersheds by two-level self-organizing feature maps. *J. Hydrol.* **2014**, *509*, 387–397.
40. Li, M.; Jiang, Z.; Zhou, P.; Le Treut, H.; Li, L. Projection and possible causes of summer precipitation in eastern China using self-organizing map. *Clim. Dyn.* **2020**, *54*, 2815–2830.
41. Rodríguez-Alarcón, R.; Lozano, S. SOM-Based Decision Support System for Reservoir Operation Management. *J. Hydrol. Eng.* **2017**, *22*, 04017012.
42. Clark, S.; Sisson, S.A.; Sharma, A. Tools for enhancing the application of self-organizing maps in water resources research and engineering. *Adv. Water Resour.* **2020**, *143*.
43. Kohonen, T. *MATLAB Implementations and Applications of the Self-Organizing Map*; Unigrafia Bookstore Helsinki: Unigrafia Oy, Helsinki, Finland, 2014; ISBN 9789526036786.
44. Vettigli, G. MiniSom: minimalistic and NumPy-based implementation of the Self Organizing Map. Available online: https://github.com/JustGlowing/minisom/.
45. Coulibaly, P.; Anctil, F.; Bobée, B. Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. *J. Hydrol.* **2000**, *230*, 244–257.
46. Fukushima, K. Visual Feature Extraction by a Multilayered Network of Analog Threshold Elements. *IEEE Trans. Syst. Sci. Cybern.* **1969**, *5*, 322–333.

47. Bridle, J.S. Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition. In *Neurocomputing*; Springer Berlin Heidelberg: Berlin, Heidelberg, 1990; pp. 227–236.

48. Jung, Y. Multiple predicting K-fold cross-validation for model selection. *J. Nonparametr. Stat.* **2018**, *30*, 197–215.

49. Erechtchoukova, M.; Khaiter, P.; Saffarpour, S. Short-Term Predictions of Hydrological Events on an Urbanized Watershed Using Supervised Classification. *Water Resour. Manag.* **2016**, *30*, 4329–4343.

50. CBC News Toronto's Don Valley Parkway reopens after severe flooding. *https://www.cbc.ca/news/canada/toronto/toronto-s-don-valley-parkway-reopens-after-severe-flooding-1.1361421* 2013.

51. CBC News Toronto's All Wet: Some Images From The Flash Floods That Hit T.O. Last Nigh. *https://www.cbc.ca/strombo/news/torontos-all-wet-some-images-from-the-flash-floods-that-hit-to-last-night.h* 2013.

Chapter 5. Probabilistic Forecasts of Flood Inundation Maps Using Surrogate Models

Summary of the research article: Zanchetta, A. D. L. and Coulibaly, P. (2022) Probabilistic Forecasts of Flood Inundation Maps Using Surrogate Models. Geosciences, 12(11). DOI: 10.3390/geosciences12110426.

A cross-validation ensemble of hybrid surrogate models for flood inundation forecasting is implemented for the generation of probabilistic forecasts with the inclusion of precipitation forecasts as part of the inputs of a hypothetical prediction system. An ensemble model weighting method based in the clustering of inundation cells by their hydraulic response is proposed for a rapid generation of probabilistic flood maps. Main questions explored are:

- Is the proposed ensemble model weighting method able to represent the uncertainties of surrogating a physics-based model?
- How the inclusion of precipitation forecasts affects the flood inundation maps produced by the surrogate model.

And the keys findings are:

- The proposed ensemble model weighting is able to represent acceptably uncertainties of surrogating a model, however the spread of the resulting probabilistic forecasts is limited by the overconfidence of the ensemble members outputs.

- The inclusion of precipitation forecasts may support the early prediction of flood conditions, but overpredictions characteristics of the product used can be reflected in issues of additional false alarms in longer lead times.

5.1. Abstract

The use of data-driven surrogate models to timely produce flood inundation maps 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 towards the generation of probabilistic outputs and assesses the impact of including quantitative precipitation forecasts (QPFs) in the set of predictors. A database of simulations produced by a quasi-2D hydraulic model is used to train an ensemble of multiple surrogate models following a k-fold cross-validation approach. 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 the quasi-2D 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 of increasing the sharpness with the tread-off of also increasing the bias of the forecasts issued for lead times longer than 2 hours.

5.2. Introduction

Changes in land cover related to urbanization and an expected higher frequency of intense 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 material 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 condition [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 spatiotemporal data for first responders. Using conventional 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 the water movement [5]. Such simulations require solving large sets of intercorrelated Saint-Venant equations, which leads to extensive computational demands that limits the real-time execution of the hydraulic

models as part of operational flood forecasting chains. While new models are being proposed to explore growing sources of processing power such as graphic processor units and cloud computing [6–8], 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 walkarounds 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) inexpensively application of such relationships in real-time (online) as hydrological observations or forecasts become available. In this context, Bhola et al. [9] and Crotti et al. [10] 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 of 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. One approach to overcome such high dimensionality issue is to set-up multiple lower dimensional machine learning models for individual [11] or spatially close 2D cells [12], which has the drawback of requiring the training and

maintenance of a potentially high number of independent models, one model 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. [13], 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 [14,15] and by the river overflow in a flashy catchment [16].

While multiple approaches have been proposed for rapidly producing flood inundation maps, for the best of the authors knowledge, only results for deterministic forecasts were reported despite the recognized importance of representing prediction uncertainties [17–19]. In this context, the surrogating of a 2D hydraulic model, 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. Each surrogate model is trained with a different subset of records of a hydraulic model simulation database, and all individual 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 such events being further analyzed as study cases.

5.3. Study Area

The Don River Basin is located in the Greater Toronto Area, Ontario, Canada (Figure 5-1a), has a total area of approximately 350 km^2, baseflow of approximately 4.5 m^3/s, and its land cover is predominantly characterized by urban infrastructure (Figure 5-1b). 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 results in a scenario of high propensity to flash flood [20], as also observed in other urban catchments surrounding the Great Lakes [21]. The response time of the catchment is in the order of 2.5 to 3 hours 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 [22]. 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 at the Bayview Avenue usually closed due to floods (Figure 5-1c).

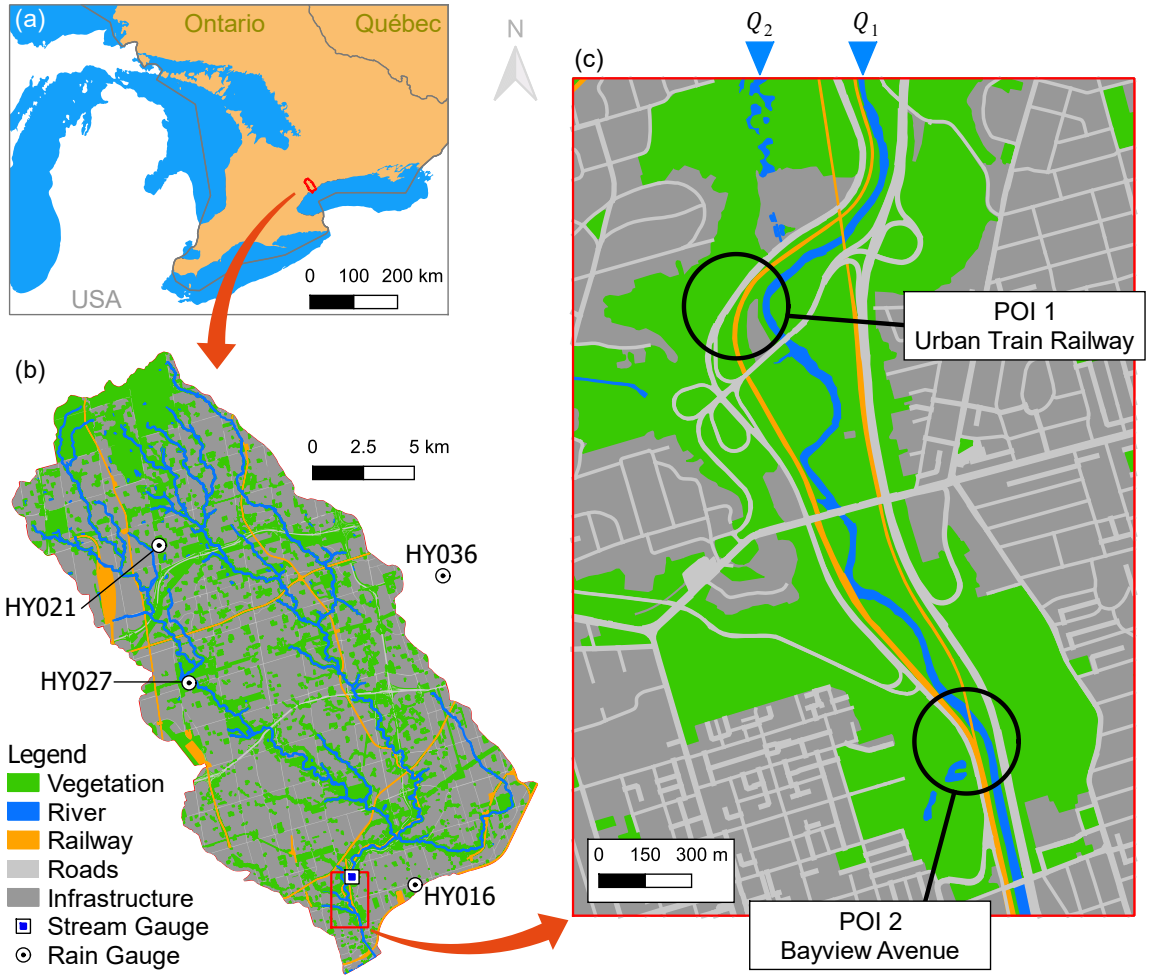

Figure 1-1. Representation of the Don River Basin in terms of (a) its location, (b) its land coverage and (c) its region prone to flash floods. Adapted from Zanchetta and Coulibaly [16].

5.4. Materials and Methods

5.4.1. Materials

5.4.1.1. 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 5-1b). The observed timeseries of such gages are made publicly available at a temporal resolution of 15 minutes. 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) [23] 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 long 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 to the needs of a flash flood forecasting system [24].

Official point intensity-duration-frequency (IDF) curves are provided to the public by Environment and Climate Change Canada [25] for the Toronto Pearson International Airport, located near to the study catchment. Such IDF is used in the design of synthetic storms.

5.4.1.2. Hydrodynamic Model

The catchment management agency developed a calibrated hydrological model of the Don River Basin in Storm Water Management Model (SWMM) [26] using the software PCSWMM [27]. The hydrological model is originally composed of 462 sub catchments, 2,703 conduits (river or channel segments) that represent the water flow unidirectionally and does not count with a 2D hydraulic component to simulate flood inundation maps. In this work it is used the modified version described by Zanchetta and Coulibaly [16] in which a hydraulic flow surface component for the flood-prone area (region of Figure 5-1c) is included. The spatial resolution of the hydraulic surface flow component is in the order of 2 meters, following the granularity of the digital elevation model (DEM) in which it is based and meeting the high degree of spatial discretization required for urban environments [28]. Such a model is hereafter referred simply as “hydrodynamic”.

5.4.2. Methodology

This work is organized in three major stages: the set-up (offline), the emulation of the operational use (online) and the performance assessment of the surrogate models, as represented in Figure 5-2 and as further described in the following sections.

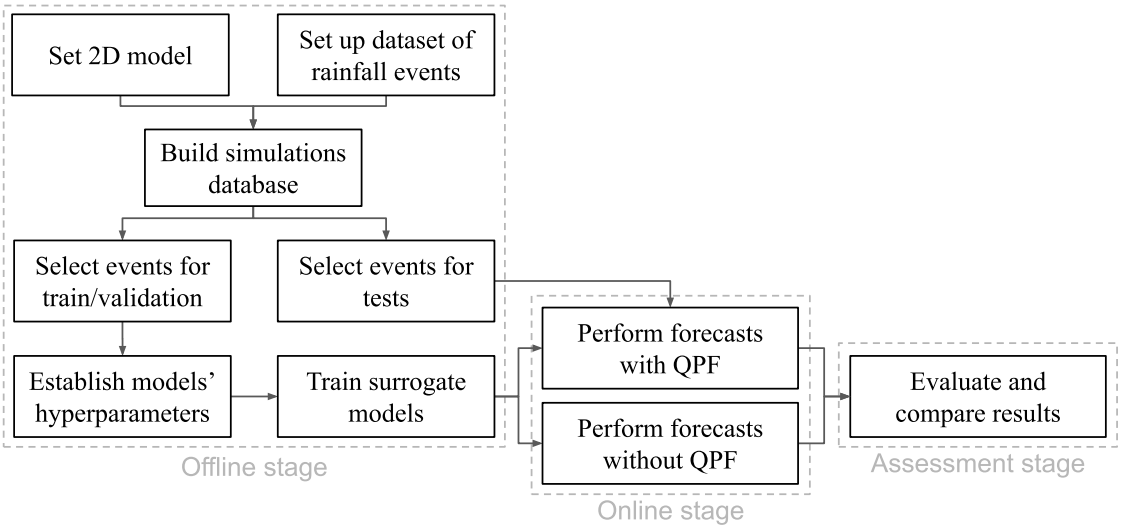

Figure 1-2. Flowchart of the methodology used in this study.

5.4.3. Setting-up the Ensemble Surrogate Model System (offline stage)

This work follows a modified sequence of steps adopted by Zanchetta and Coulibaly [16] for setting up the surrogate models used. Initially, an extensive set of significant observed and synthetic rainfall events is established. For each of such events, 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 train/validation dataset, which is split in subsets of equal size using k-fold approach. For each of such subsets, one hybrid surrogate model system, composed by pairs of one recurrent and one classifier network, is trained.

5.4.3.1. 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 m^3/s) within 3 hours from a rainfall input recorded by at least one of the rain gauges. In order to capture eventual long standing precipitation records and most part of the recession curve, all rainfall and discharge data from 36 hours 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 occurred 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. [10], which identified that using datasets composed by 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 set simulates scenarios with the same rainfall intensity that was observed could have resulted in different outcomes if they have had different spatial configurations.

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 hours, (2)

the accumulation-duration-frequency (ADF) curve was derived from the IDF curve, (3) the 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 [29] and based on alternating blocks method [30] were generated to produce a wide diversity of climatic-based rainfall shapes.

Figure 5-3a presents the observation values used to define the ARF. Each 24-hours precipitation record with accumulation values higher than 46 mm (i.e., the total accumulated precipitation estimated for rainfalls with 50 years return period) 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.

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

$$I(T,f) = A_f * T^{B_f}, \tag{5-1}$$

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:

$$P_p(T,f) = I(T,f) * T = A_f * T^{1+B_f}. \tag{5-2}$$

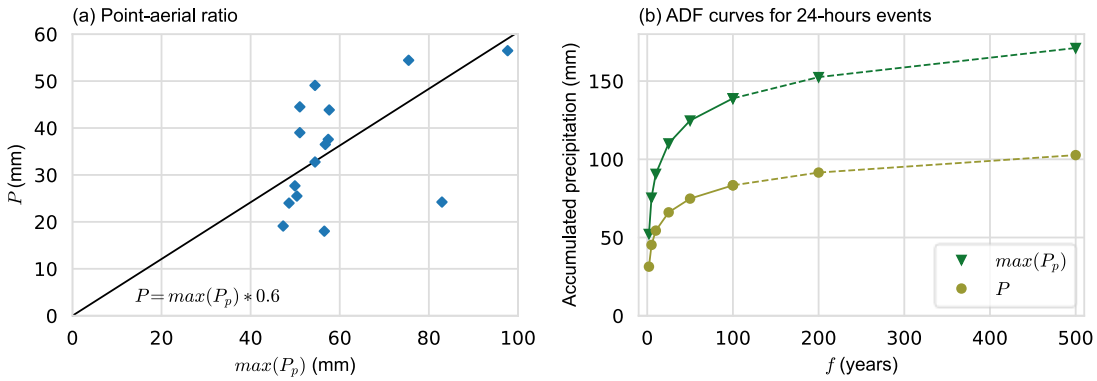

Figure 1-3. Regression lines related to (a) the definition of the point-aerial ratio and (b) the 24-hours ADF curves, both from reference (solid line) and extrapolated (dashed lines).

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 5-1). Using Equation 5-2, the obtained accumulated point rainfall for a duration T of 24 hours 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 5-3b). For each of the three mean aerial accumulated precipitation values, five storm designs were created designed based on the four Huffs shapes and one based on the alternating blocks method. The 15 resulting pluviograms were used as spatially uniform inputs on simulations in the hydrodynamic model of the catchment.

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

	Return period (years)				
Coefficient	**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**

5.4.3.2. 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 5-1c) and instant inundation maps for the 36 hours of each rainfall event at 15-minutes 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-minutes accumulated mean aerial precipitation $P_{15min,t}$ were stored in the database. It worth noting that a complete IM_t stored in the database is composed by 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 neglectable compared to the runoff routed to Q_1 and Q_2.

Once all the inundation maps were generated, the 2D cells in the floodplain domain were classified in three groups. The first group, referred as "wet cells", is composed by 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 as "dry cells", is composed by the 2D cells that presented zero water depths during all instants of all simulations (i.e., points that are extremely unlikely to be inundated). The remaining 2D cells, referred as "inundation cells", represent locations in 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.

5.4.3.3. Selection of the Train/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 an 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 occurred posteriorly to all historical events are 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, respectively.

To reduce the redundancy of the data used for train/validation, a subset of the remaining simulated events was selected using the conventional Computer Aided Design of Experiments (CADEX) sampling method [31]. Given a set S in which each of its records is defined by F features f_1, f_2, …, 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 the 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 [31] for further details on the method.

In this work, for the application of 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:

$$\Delta(r_i, r_j) = \frac{\sum_{y=1}^{144} |AID_{i,y} - AID_{j,y}|}{144} \tag{5-3}$$

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 defines $a\ priori$. To evaluate multiple values for *k* in the k-folding implemented in posterior steps, as further discussed in Section 5.4.3.5, the size of the sample was set to be 36.

Figure 5-4 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 of being composed by only one third of the total number of records, the AID timeseries of the rainfall events in the train/validation set (Figure 5-4b) present a variety of forms comparable with the full dataset (Figure 5-4a). The majority of events that were not included in the train/validation set are characterized by their lower intensity and highly 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 train/validation set. A summary of the composition of each set is given in Table 5-2.

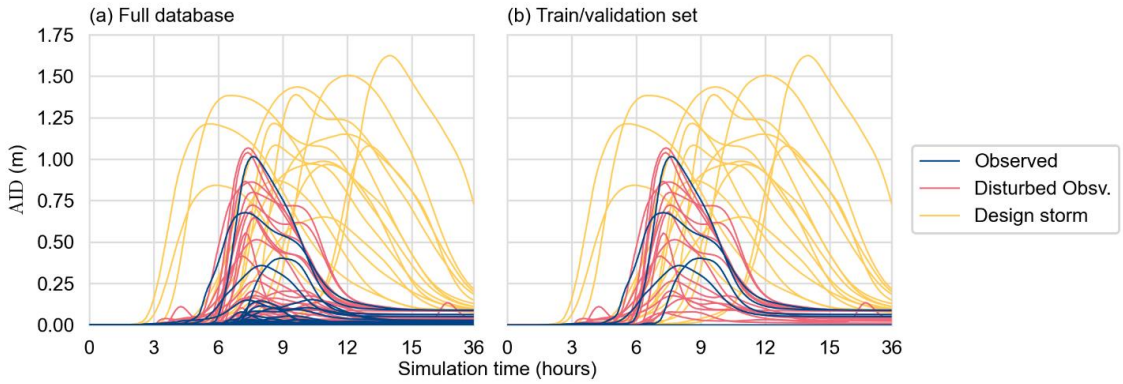

Figure 1-4. Timeseries of the AID of the simulated rainfall events in (a) the entire database and (b) in the selection using the CADEX algorithm.

Table 1-2. 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
Train/Validation	5	16	15	36

5.4.3.4. Establishing the Hyperparameters of the Surrogate Models

Each surrogate model member of the ensemble forecasting system consists of and hybrid structure composed by a nonlinear autoregressive neural network with exogenous inputs (NARX) and a self-organizing map (SOM) [32] in a configuration similar to the one adopted by Zanchetta and Coulibaly [16], which demonstrated the suitability of the NARX-SOM approach for surrogating the same hydrodynamic model of this study in the context of deterministic forecasts.

The SOM component has the objective of reducing the dimensionality of flood inundation maps, which is usually composed by 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 handle highly dimensional datasets using a rectangular 2D topological space [33]. 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 inundations 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 the train/validation dataset; (2) if all the topological nodes had 2 or more inundation maps associated to it, 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; the iterations proceed until (3) at least one topological node in the trained SOM is composed by 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 5-5 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 being 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.

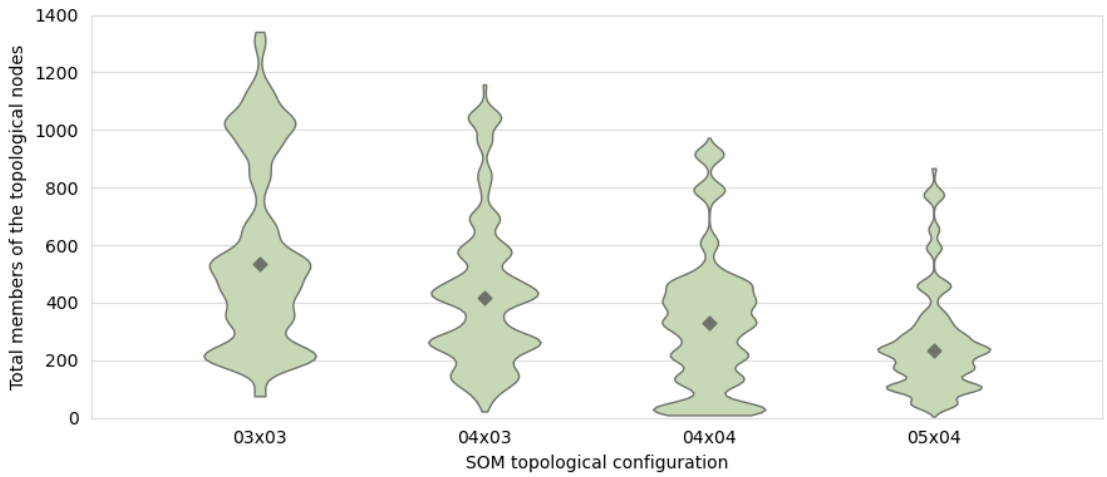

Figure 1-5. Distribution of the records in the train/validation dataset assigned to different topological nodes for multiple configurations of SOM dimensions. Diamond markers indicate the median of each distribution.

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 5-3). All models have, as part of its predictor set, antecedent values of mean aerial quantitative estimated precipitation (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 5-6.

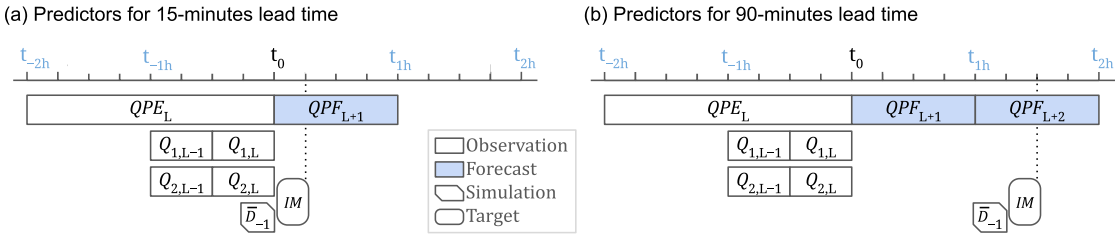

Figure 1-6. Temporal representation of the predictors of the NARX models trained for lead times of (a) 15 minutes and (b) 90 minutes assuming an issue time t_0.

Table 1-3. Listing of all potential NARX predictors.

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

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 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 to be 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 by 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 5-7.

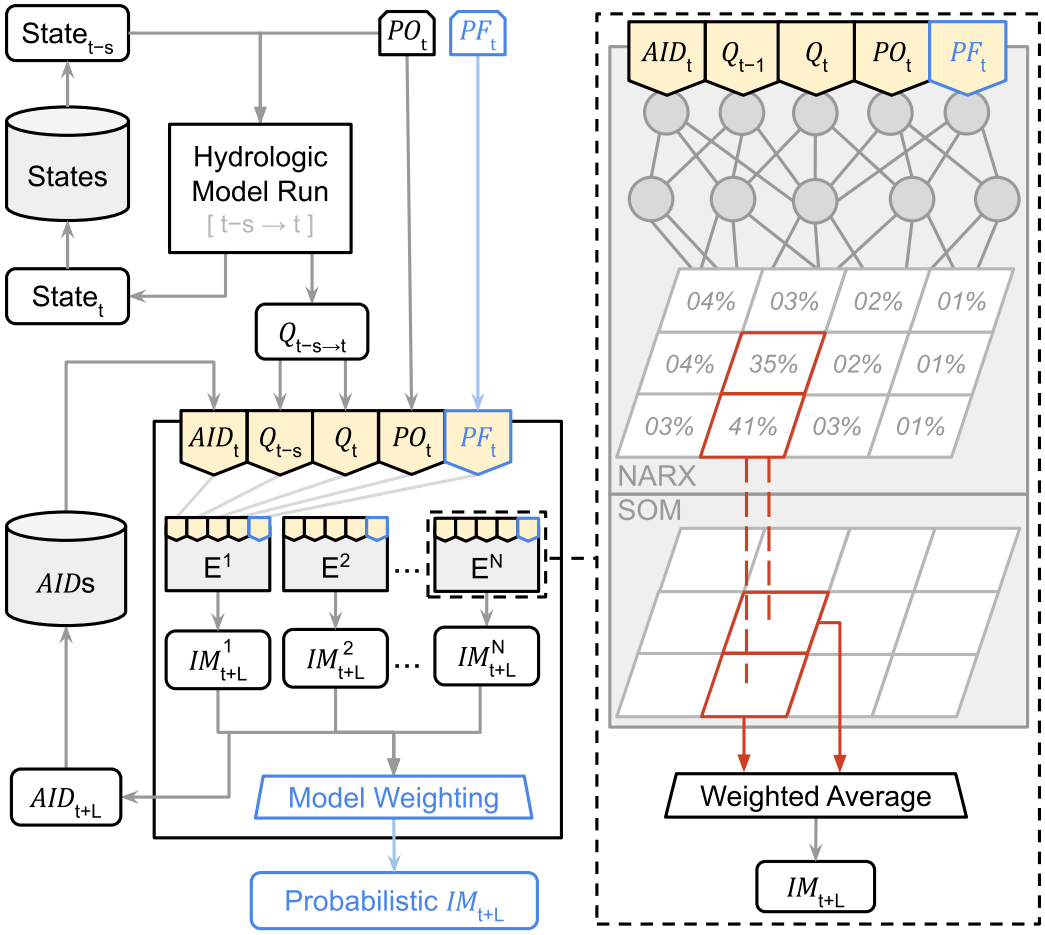

Figure 1-7. Diagram representing the dataflow of the hypothetical operational setup. E^i indicates the *i*-th member of the ensemble system, comprehended by a hybrid NARX-SOM model (represented within the dashed rectangle). Blue boxes indicate the changes from Zanchetta and Coulibaly [16]. Number of inputs, neurons and topological nodes are hypothetical and for illustrative purposes only.

5.4.3.5. Training the Surrogate Models

The conventional *k*-fold cross-validation method was applied to train multiple surrogate models. In such an approach, the full train/validation dataset is split into *k* equally

sized subsets. For each subset ("fold"), a model is tuned using all the other subsets ("fold-in") for training and the own subset ("fold-out") for validation, resulting in k models trained at the end of all iterations.

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 best performance is terms of Continuous Ranked Probability Score (CRPS, described in Section 5.4.5) is selected. The train/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 it can be observed in Table 5-4, the CRPS of the configuration composed by 12 folds has the lowest value for the four 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.

Table 1-4. CRPS of the ensemble surrogate models for different cross-fold ensemble configurations in the train/validation dataset. The lowest (best) value of each column is highlighted in bold.

	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

5.4.4. Forecasting the Probabilistic Inundation Maps

5.4.4.1. 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 minutes apart from 15 minutes to 4 hours 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 $D^{i}_{c,t,L}$.

5.4.4.2. Converting Ensemble 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 $\tau^{0.1}_{c,t,L}$, $\tau^{0.2}_{c,t,L}$, …, $\tau^{0.9}_{c,t,L}$, in which $\tau^{i}_{c,t,L}$ 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.

5.4.5. 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 by 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:

$$\overline{CRPS}_{t,L} = \frac{1}{C}\sum_{c=1}^{C}\int_{x=-\infty}^{x=\infty}\left(Prob.\left(D'_{C,t,L} \le x\right) - H(D_{c,t}, x)\right)^2 dx \qquad (5\text{-}4)$$

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

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

$\overline{CRPS}_{t,L}$ values range from 0 (perfect fit) to ∞, unitless. In this work, CRPS is applied into two sets of data. The first set consists of every inundation cell c at every instant time t, regardless 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 0.01m 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.:

$$\overline{BS}_{t,L} = \frac{1}{C}\sum_{c=1}^{C}\left(Prob.(D'_{c,t,L} \leq D_{threshold}) - H(D_{c,t}, D_{threshold})\right)^2 \quad (5\text{-}6)$$

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, [34], 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 is considered the predictor. 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 having 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:

$$\overline{CR}_{80,t,L} = \frac{N_h}{N} * 100\%, \quad (5\text{-}7)$$

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

Average Bandwidth [34] 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 $\tau^{0.9}_{c,t,L} - \tau^{0.1}_{c,t,L}$ forecasted for an inundation cell c at instant t issued at a lead time L, $B_{80,c,t,L}$ is given by:

$$B_{80,c,t,L} = \tau^{0.9}_{c,t,L} - \tau^{0.1}_{c,t,L}. \quad (5\text{-}8)$$

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 to 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, it is used the Mean Fractional Bias (MFB) [35]. It is given by:

$$\overline{MFB}_{t,L} = \frac{1}{C}\sum_{c=1}^{C}\frac{\tau_{c,t,L}^{0.5} - D_{c,t}}{\tau_{c,t,L}^{0.5} + D_{c,t}} \tag{5-9}$$

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 bias were used. Additionally, it is applied a specific case metric, named event Peak Bias (PB), which is calculated by:

$$PB_{c,E,L} = \frac{\max\left(\tau_{c,E,L}^{0.5}\right) - \max\left(D_{c,E}\right)}{\max\left(D_{c,E}\right)} * 100\% \tag{5-10}$$

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.

5.5. Results and Discussion

To facilitates visual interpretation, only timeseries and inundation maps of forecasts issued for lead times of 1, 2, 3 and 4 hours are presented, despite of results being available

at a 15-minutes time step. For the sake of simplicity, the surrogate models that don't 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 as "QPF-aware" or "RAP QPF" hereafter.

5.5.1. 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 its QPF-aware counterparts for most of lead times (Figure 5-8a,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 5.5.2). Interestingly, for the instants when an inundation is present in the simulation, outputs from the QPF-aware models presented and overall better fit to the simulations, especially for the lead time of 4 (Figure 5-8b). 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 5-9). 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 5-8d).

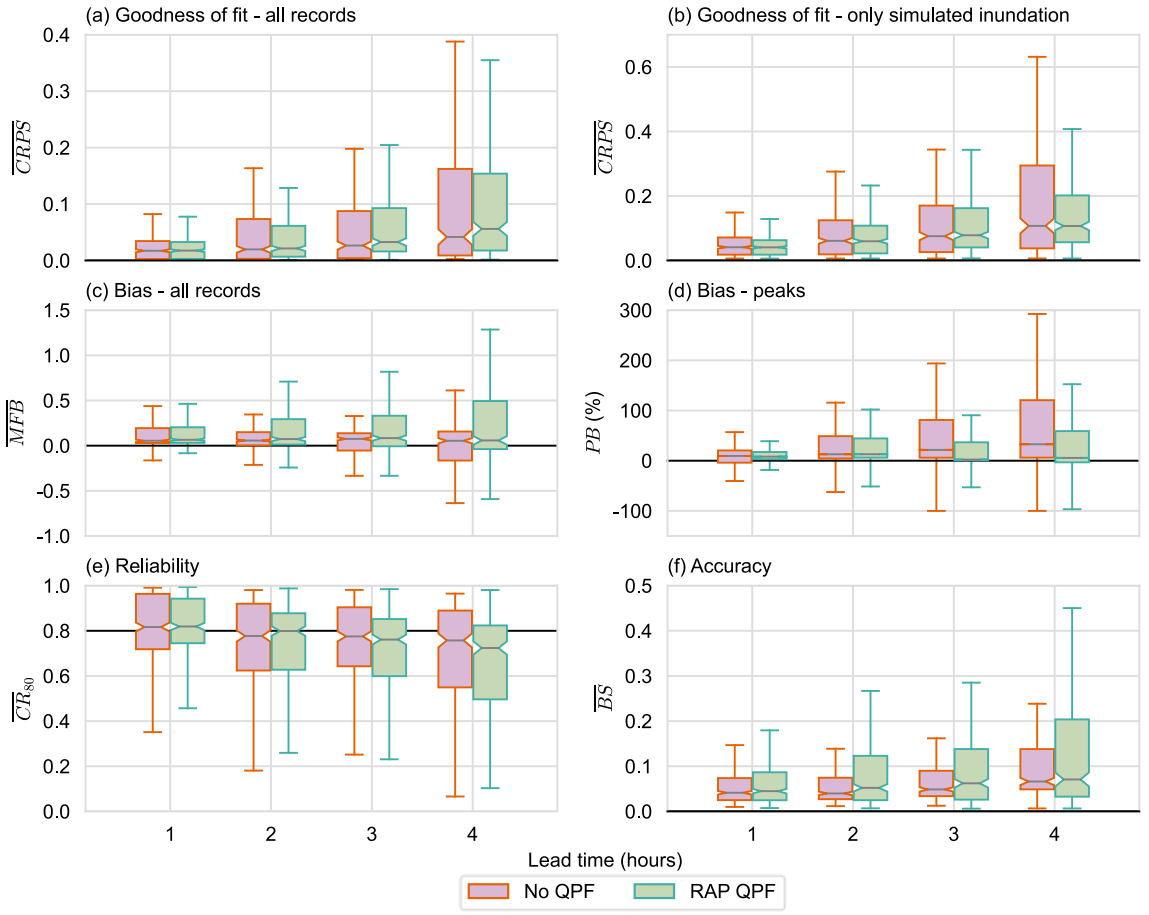

Figure 1-8. General performance metrics of the no-QPF and the QPF-aware surrogate models. The darker horizontal line indicates the "perfect forecast" value.

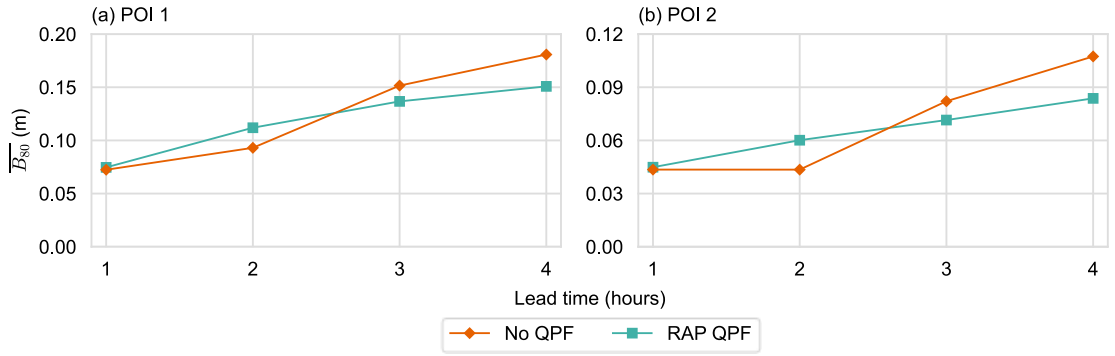

Figure 1-9. Mean B_{80} for the three test events at (a) POI 1 and (b) POI 2.

5.5.2. 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 July 8th, 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 [36]. The second event, of August 2nd, 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 [37,38].

5.5.2.1. July 8th, 2013

The hydrographs in Figures 5-10 and 5-11 show the predicted water depths at POIs 1 and 2, respectively, for the event of July 08th, 2013, issued at different lead times. On both figures, it is possible to note a first peak at 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, being likely driven by the higher bias and higher overall errors with respect to longer lead times already reported in RAP products [39]. As lead time decreases, also decreases the over-forecasted first peak and also increases the similarity between the effective peak and the predictions produced by the QPF-aware products.

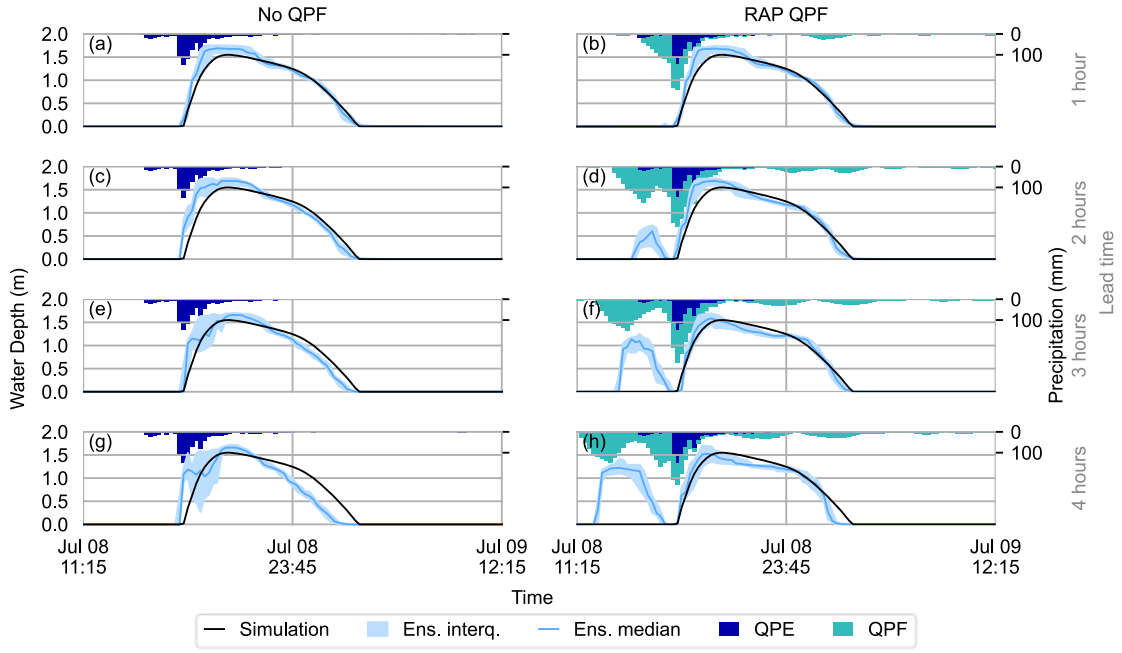

Figure 1-10. Water depth simulated and ensemble forecasts for POI 1 at lead times of (a,b) 1 hour, (c,d) 2 hours, (e,f) 3 hours and (g,h) 4 hours for the event of July 8th, 2013.

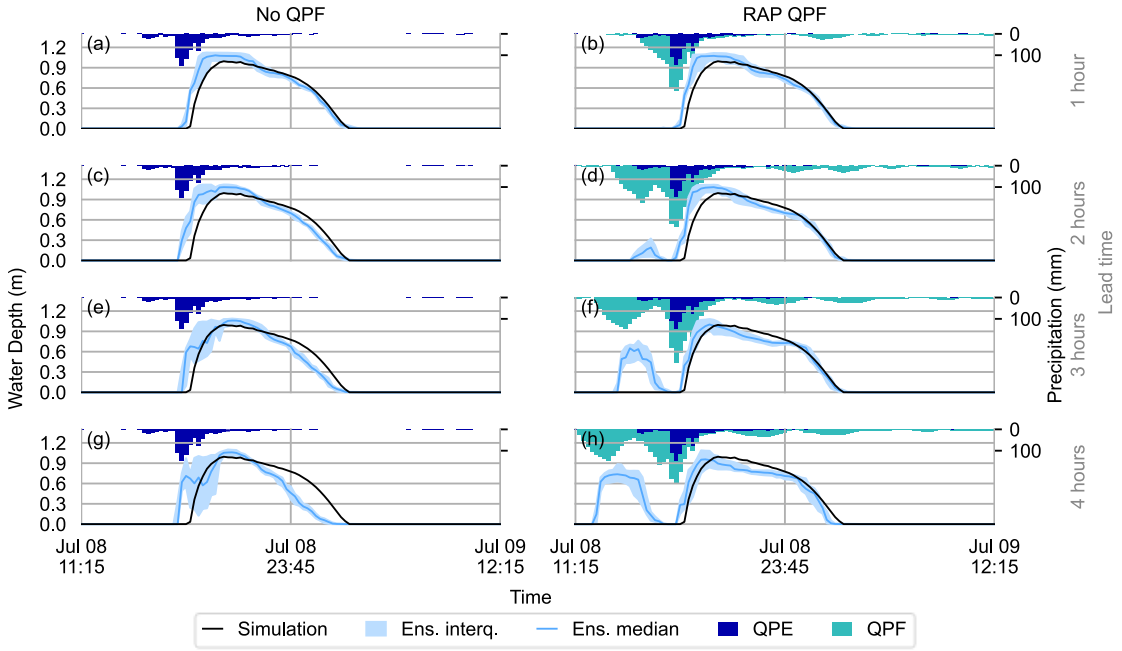

Figure 1-11. Same as Figure 5-10, but for POI 2.

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 5.5.1, Figure 5-9).

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. 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 5-8e).

The overall shape of the probability of exceedance maps produced by both the no-QPF and the QPF-aware is similar to the simulated water depth exceedance map considering the maximum depth at each location as threshold (Figures 5-12 and 5-13). While some overestimation is observed at lead times up to 2 hours in both cases, such overestimation is also present in the forecasts for 3 and 4 hours in the future when the surrogate model is based solely on QPEs.

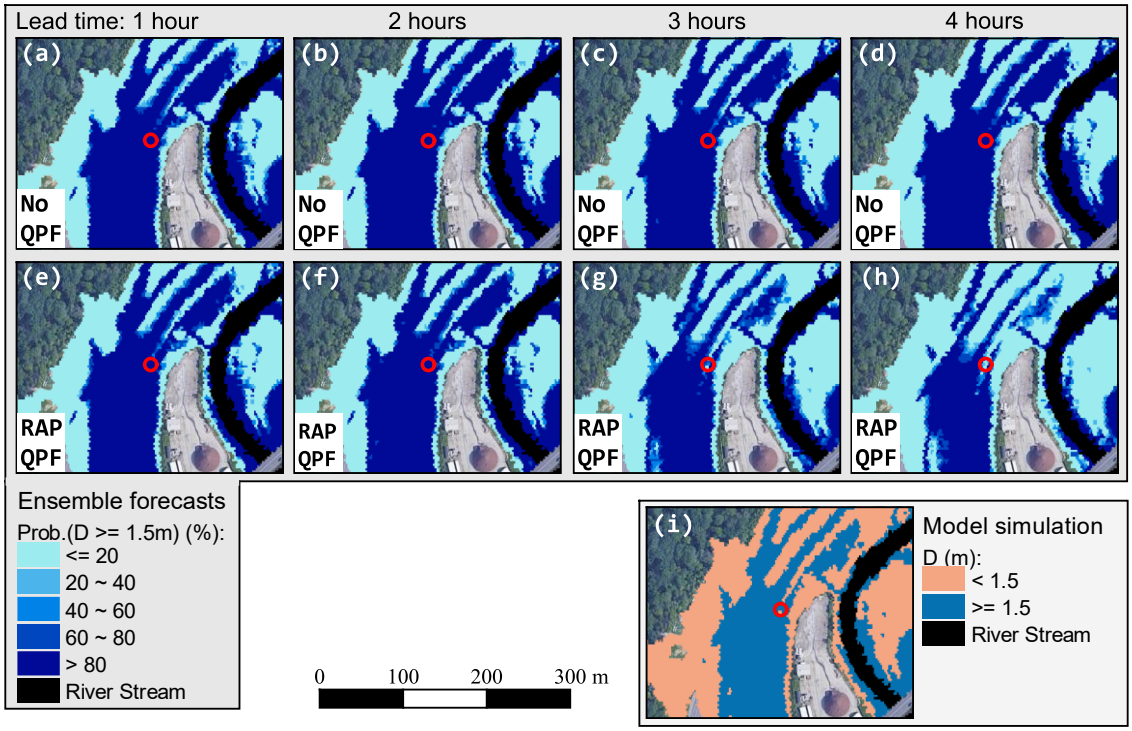

Figure 1-12. Inundation maps forecasted by the surrogate models without QPF (a-d), with QPF (e-h), and simulated by the hydrodynamic model (i) at the peak water level of the event of July 08th, 2013, at POI 1.

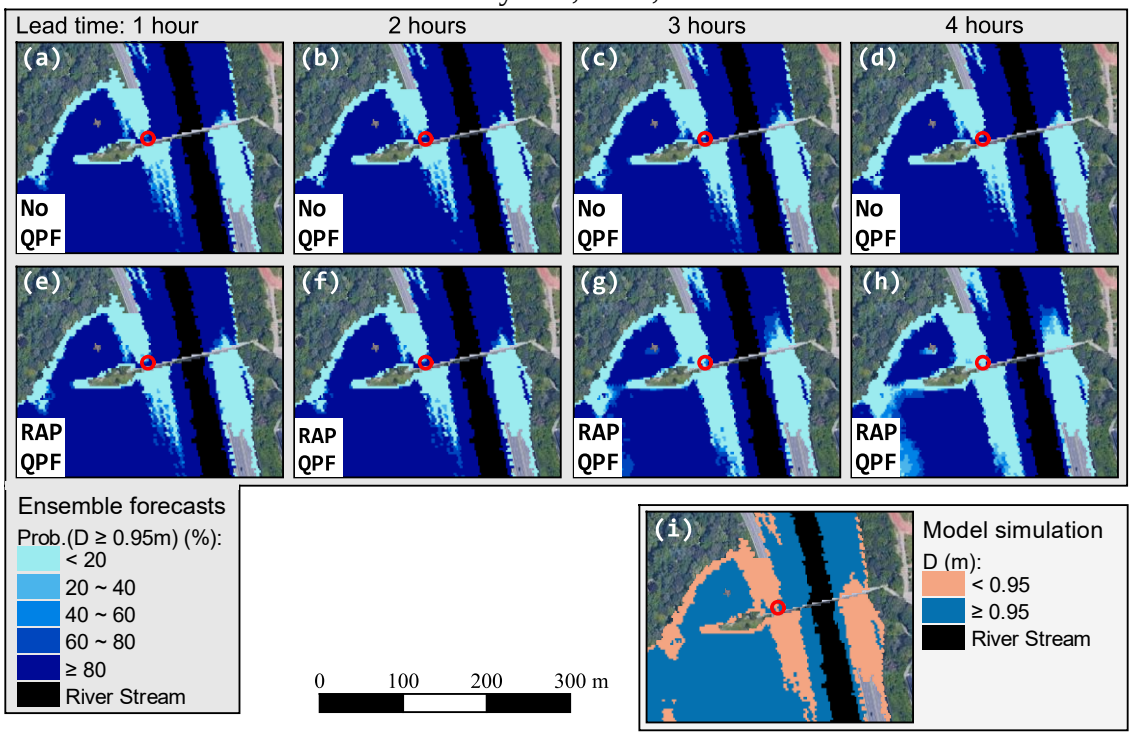

Figure 1-13. Same as Figure 12, but for the POI 2.

5.5.2.2. August 2nd, 2020

The forecasting of this event shares many similarities with the forecasts issued for the event of July 8th, 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 5-14) and POI 2 (Figure 5-15). 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.

The inundation maps forecasted by both surrogate models are also characterized by overestimating the flooded area at shorter lead times (Figures 5-16 and 5-17). The overall shape of the simulated and the forecasted maps are comparable for the POI 1. However, it is possible to note that, regardless the inclusion of the QPF in the feature set, the maps forecasted wrongly a relatively large area as flooded in the south-west of the POI 2 (lower-left corner of the maps in Figure 5-17). Such an area has a significant 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 the area. Such an overestimation is not observed for the inundation cells related to traffic surfaces, though.

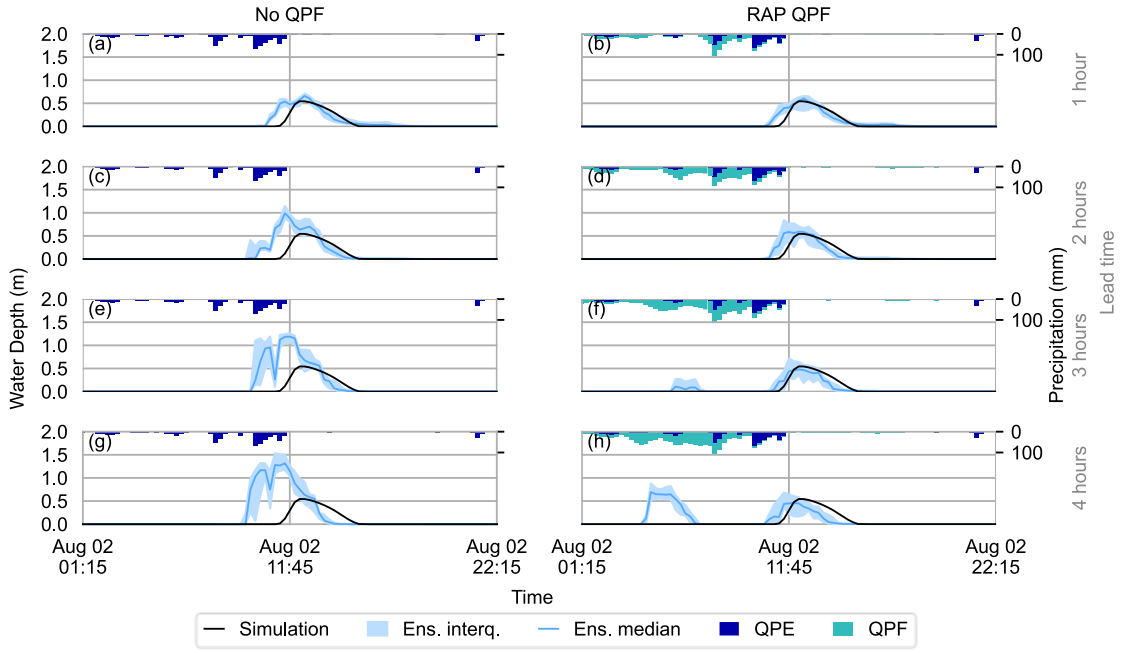

Figure 1-14. Water depth simulated and ensemble forecasts for POI 1 at lead times of (a,b) 1 hour, (c,d) 2 hours, (e,f) 3 hours and (g,h) 4 hours for the event of August 2nd, 2020.

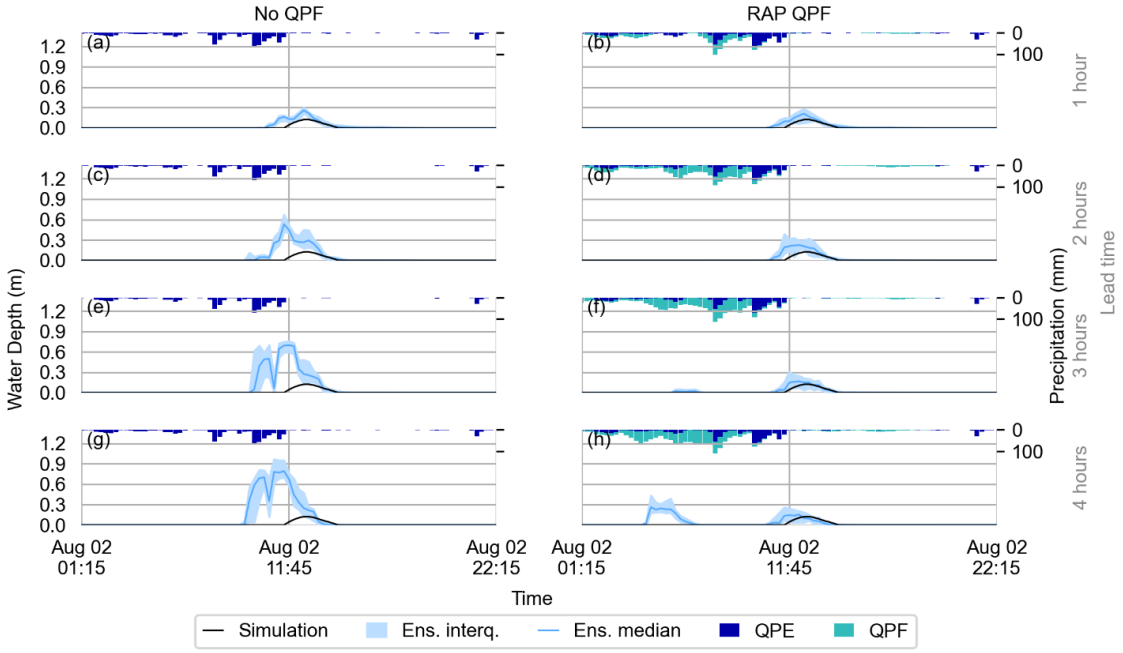

Figure 1-15. Same as Figure 5-14, but for POI 2.

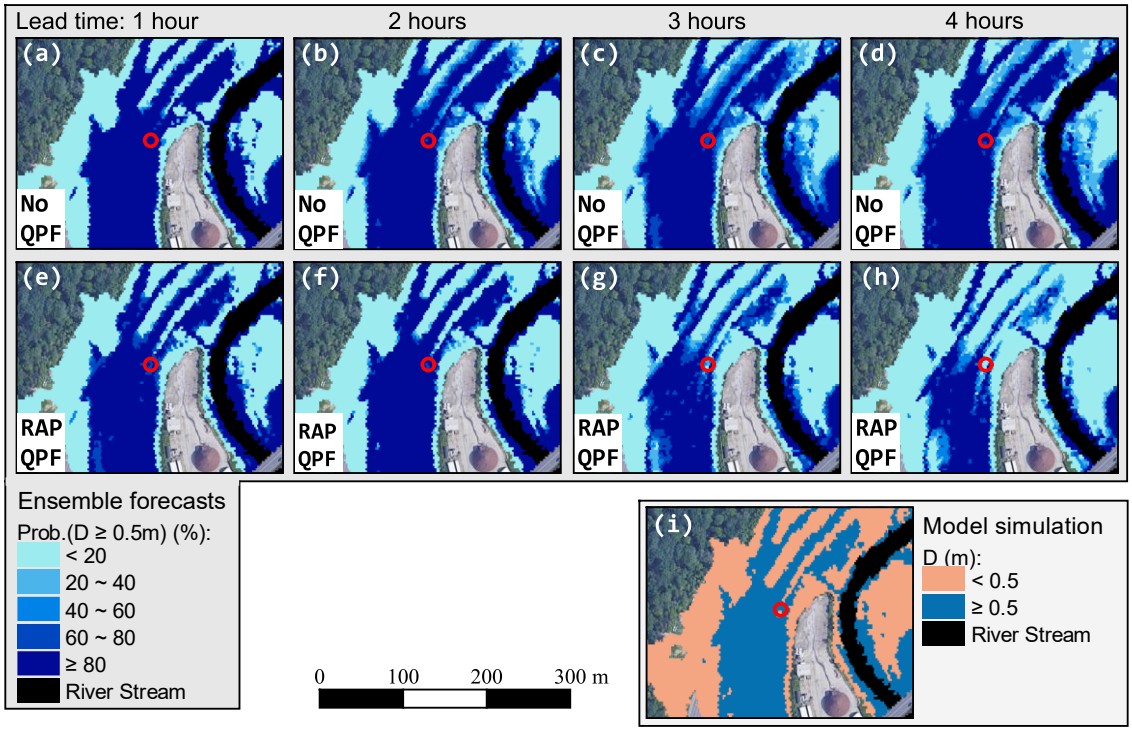

Figure 1-16. Inundation maps forecasted by the surrogate models without QPF (a-d), with QPF (e-h), and simulated by the hydrodynamic model (i) at the peak water level of the August 02nd, 2020, event at POI 1.

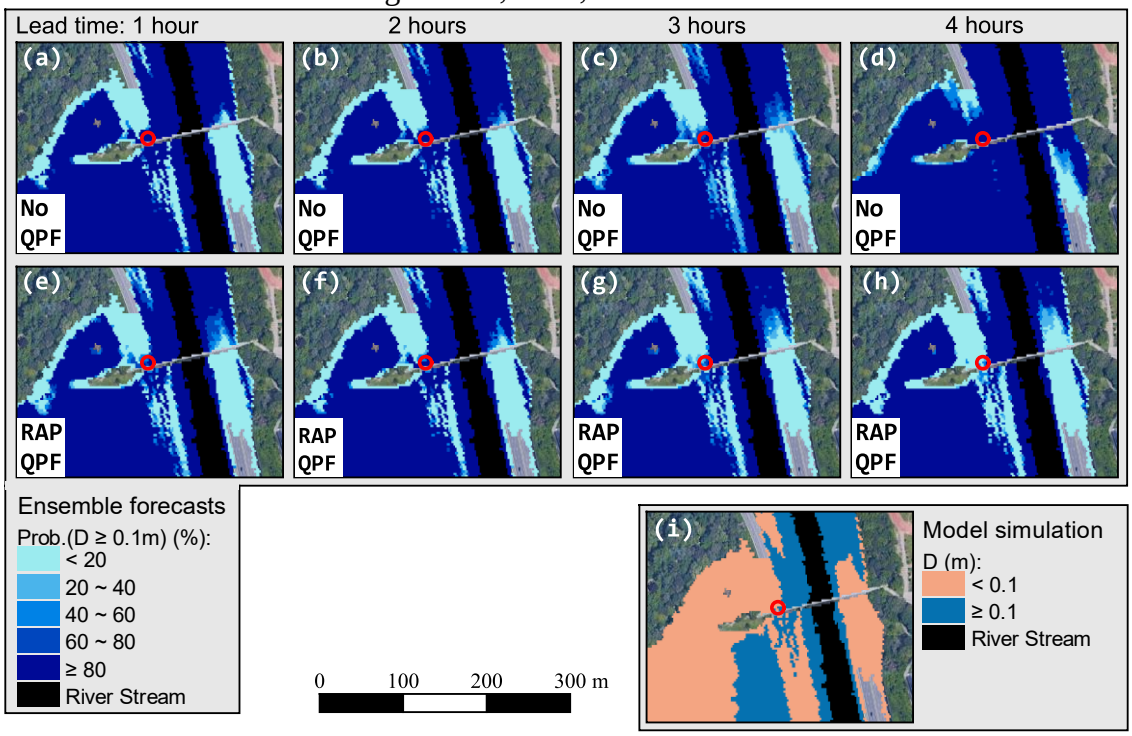

Figure 1-17. Same as Figure 5-16, but for POI 2.

5.5.3. Discussion

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, besides sharing similar network structures. This can be interpreted as a limitation of the 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, though, 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. However, 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.

5.5.4. Runtime

Both deterministic simulations using the hydrodynamic model and ensemble forecasts from surrogate models were generated using a desktop computer with 64 GB rapid access

memory (RAM) and CPU Intel I9 with 3.6 GHz, eight codes and 16 logical processors. While the runtime of the hydrodynamic model demanded approximately 4 hours and 30 minutes to produce 4 hours of inundation maps, the ensemble of forecasts for the same time interval required between 13 to 17 minutes to be generated, which may be considered applicable for real-time setups.

5.6. Conclusion, 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, besides 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 a SWMM-based hydrodynamic model in a forecasting setup. The forecasted maps were compared with the simulated maps to assess the efficiency of the surrogate models on rapidly reproducing the hydrodynamic model outputs.

For 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 k-fold ensemble was successful in capturing the additional uncertainties of the surrogating step. The inclusion of QPF products didn't impact significantly the maps forecasted for lead times up to 2 hours. 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 suggests that a forecasting system composed by 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, the forecasts were characterized by over confidence (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 flood-prone area may be uncommon in forecasting centers giving the highly demanding tasks of producing and maintaining each individual model updated to reflect changes in the land cover. The results presented are specific for the Don River Basin and data products utilized, and the evaluation of this approach at a broader scope is suggested as future research.

5.7. 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. *Environmental Challenges* **2022**, *6*, 0–8, doi:10.1016/j.envc.2022.100449.

2. 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, doi:10.1038/srep40527.

3. 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, doi:10.1175/JHM-D-12-095.1.

4. Zanchetta, A.D.L.; Coulibaly, P. Recent Advances in Real-Time Pluvial Flash Flood Forecasting. *Water (Basel)* **2020**, *12*, 570, doi:10.3390/w12020570.

5. 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. *Environmental Modelling and Software* **2017**, *90*, 201–216, doi:10.1016/j.envsoft.2017.01.006.

6. Aureli, F.; Prost, F.; Vacondio, R.; Dazzi, S.; Ferrari, A. A GPU-Accelerated Shallow-Water Scheme for Surface Runoff Simulations. *Water (Switzerland)* **2020**, *12*, doi:10.3390/w12030637.

7. 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*, doi:10.1029/2019WR025583.

8. 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. *Environmental Modelling and Software* **2018**, *107*, 231–244, doi:10.1016/j.envsoft.2018.05.007.

9. Bhola, P.K.; Leandro, J.; Disse, M. Framework for Offline Flood Inundation Forecasts for Two-Dimensional Hydrodynamic Models. *Geosciences (Switzerland)* **2018**, *8*, doi:10.3390/geosciences8090346.

10. Crotti, G.; Leandro, J.; Bhola, P.K. A 2D Real-Time Flood Forecast Framework Based on a Hybrid Historical and Synthetic Runoff Database. *Water (Switzerland)* **2020**, *12*, doi:10.3390/w12010114.

11. 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, doi:10.1111/jfr3.12522.

12. Berkhahn, S.; Fuchs, L.; Neuweiler, I. An Ensemble Neural Network Model for Real-Time Prediction of Urban Floods. *J Hydrol (Amst)* **2019**, *575*, 743–754, doi:10.1016/j.jhydrol.2019.05.066.

13. Chang, L.-C.; Amin, M.; Yang, S.-N.; Chang, F.-J. Building ANN-Based Regional Multi-Step-Ahead Flood Inundation Forecast Models. *Water (Basel)* **2018**, *10*, 1283, doi:10.3390/w10091283.

14. Kim, H. il; Keum, H.J.; Han, K.Y. Real-Time Urban Inundation Prediction Combining Hydraulic and Probabilistic Methods. *Water (Basel)* **2019**, *11*, doi:10.3390/w11020293.

15. Kim, H. il; Han, K.Y. Data-Driven Approach for the Rapid Simulation of Urban Flood Prediction. *KSCE Journal of Civil Engineering* **2020**, *24*, 1932–1943, doi:10.1007/s12205-020-1304-7.

16. 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, doi:10.3390/forecast4010007.

17. Bales, J.D.; Wagner, C.R. Sources of Uncertainty in Flood Inundation Maps. *J Flood Risk Manag* **2009**, *2*, 139–147, doi:10.1111/j.1753-318X.2009.01029.x.

18. Zahmatkesh, Z.; Han, S.; Coulibaly, P. Understanding Uncertainty in Probabilistic Floodplain Mapping in the Time of Climate Change. *Water (Switzerland)* **2021**, *13*, doi:10.3390/w13091248.

19. Brandt, S.A. Modeling and Visualizing Uncertainties of Flood Boundary Delineation: Algorithm for Slope and DEM Resolution Dependencies of 1D Hydraulic Models. *Stochastic Environmental Research and Risk Assessment* **2016**, *30*, 1677–1690, doi:10.1007/s00477-016-1212-z.

20. Rincón, D.; Khan, U.T.; Armenakis, C. Flood Risk Mapping Using GIS and Multi-Criteria Analysis: A Greater Toronto Area Case Study. *Geosciences (Switzerland)* **2018**, *8*, doi:10.3390/geosciences8080275.

21. Krajewski, M.; Brown, D.; Gibbons, E. *Flash Flooding , Stormwater , and Decision Making Flash Flooding , Stormwater , and Decision Making for Cities in the Great Lakes*; Ann Arbor, MI, 2015;

22. 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; AMS: Portland, OR, USA, November 2016.

23. 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, doi:10.1175/MWR-D-15-0242.1.

24. Hapuarachchi, H.A.P.; Wang, Q.J.; Pagano, T.C. A Review of Advances in Flash Flood Forecasting. *Hydrol Process* **2011**, *25*, 2771–2784, doi:10.1002/hyp.8040.

25. 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).

26. Rossman, L.A. *Storm Water Management Model - User's Manual Version 5.1*; US EPA: Cincinnati, OH, 2015;

27. Computational Hydraulics International - CHI PCSWMM. Available online: https://www.pcswmm.com/.

28. Dinu, C.; Sîrbu, N.; Drobot, R. Delineation of the Flooded Areas in Urban Environments Based on a Simplified Approach. *Applied Sciences (Switzerland)* **2022**, *12*, doi:10.3390/app12063174.

29. Huff, F.A. Time Distribution of Rainfall in Heavy Storms. *Water Resour Res* **1967**, *3*, 1007–1019, doi:10.1029/WR003i004p01007.

30. Chow, V. te *Applied Hydrology*; Clark, B.J., Morriss, J., Eds.; McGraw-Hill, 1988; ISBN 0 07-010810-2.

31. Kennard, R.W.; Stone, L.A. Computer Aided Design of Experiments. *Technometrics* **1969**, *11*, 137, doi:10.2307/1266770.

32. Kohonen, T. Self-Organized Formation of Topologically Correct Feature Maps. *Biol Cybern* **1982**, *43*, 59–69, doi:10.1007/BF00337288.

33. Kohonen, T. *MATLAB Implementations and Applications of the Self-Organizing Map*; Unigrafia Bookstore Helsinki: Unigrafia Oy, Helsinki, Finland, 2014; ISBN 9789526036786.

34. 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. *Hydrological Sciences Journal* **2009**, *54*, 852–871, doi:10.1623/hysj.54.5.852.

35. 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, doi:10.1016/j.atmosenv.2005.09.087.

36. CBC News Toronto's All Wet: Some Images From The Flash Floods That Hit T.O. Last Nigh. *https://www.cbc.ca/strombo/news/torontos-all-wet-some-images-from-the-flash-floods-that-hit-to-last-night.h* 2013.

37. CTV News What's Open and Closed This Holiday Monday in Toronto? 2020.

38. DailyHive News Rain Causes Flooding on Low-Lying Toronto Highway Ramps 2020.

39. Burg, T.; Elmore, K.L.; Grams, H.M. Assessing the Skill of Updated Precipitation-Type Diagnostics for the Rapid Refresh with MPING. *Weather Forecast* **2017**, *32*, 725–732, doi:10.1175/waf-d-16-0132.1.

Chapter 6. Conclusions and Recommendations

6.1. Conclusions

The research activities that compose this thesis focus on exploring the potential of using machine learning methods for enhancing the forecasting of flash flood events driven by intense rainfall. The first chapter presents the state-of-art of the currently implemented and openly documented flash flood early warning systems. Each of the following chapters present a study related to one part of a hypothetical forecasting chain in which different data-driven techniques are applied for variate purposes. Figure 6-1 illustrates how the tools presented or explored in the research articles could be used in a single operational system and thus represents the components of prediction chains that are expected to benefit from the findings of this research.

The main conclusions of the thesis are summarized as follows:

- The state-of-the-art of flood forecasting systems suitable for the prediction of flash flood events indicates a high heterogeneity in terms of models applied, input data used, and decision criteria adopted among the documented forecasting chains.
- Instrumental and computational resources capable of producing high resolution precipitation estimates and forecasts are not implemented evenly across the globe, which justify the different design decisions and research directions adopted for enhancing of flash flood forecasting system.
- Machine learning models implemented for reproducing the rainfall-runoff behavior of catchments prone to flash floods are usually designed to use precipitation input

data in the same configuration as it is used to force conventional hydrological models, however the increasing availability of concurrent precipitation estimates can be directly explored by machine learning models as additional and valuable input features, with benefits in performance related to an increase in the overall goodness of fit and in the precision of detecting high flows.

- Hybrid flood inundation surrogate models based on the integration of NARX and SOM structures, previously documented and assessed for large scale and slowly developing flood events, present an acceptable performance for timely producing deterministic flood inundation maps at the high spatiotemporal resolution required for catchments prone to flash floods causes by rapid river overflow.
- The use of cross-validation ensemble forecasts for producing probabilistic flood inundation maps adequately captures the forecasting uncertainties related to the process of surrogating hydrodynamic models based on physics.
- Including quantitative precipitation forecast products as predictors of data-driven surrogate models has the potential of improving their performance in terms of estimating the maximum water depth of upcoming flood events. However, if the precipitation forecast product applied is characterized as biased high, it can be propagated through the forecasting chain and result in the wrongly over prediction of flooding conditions.

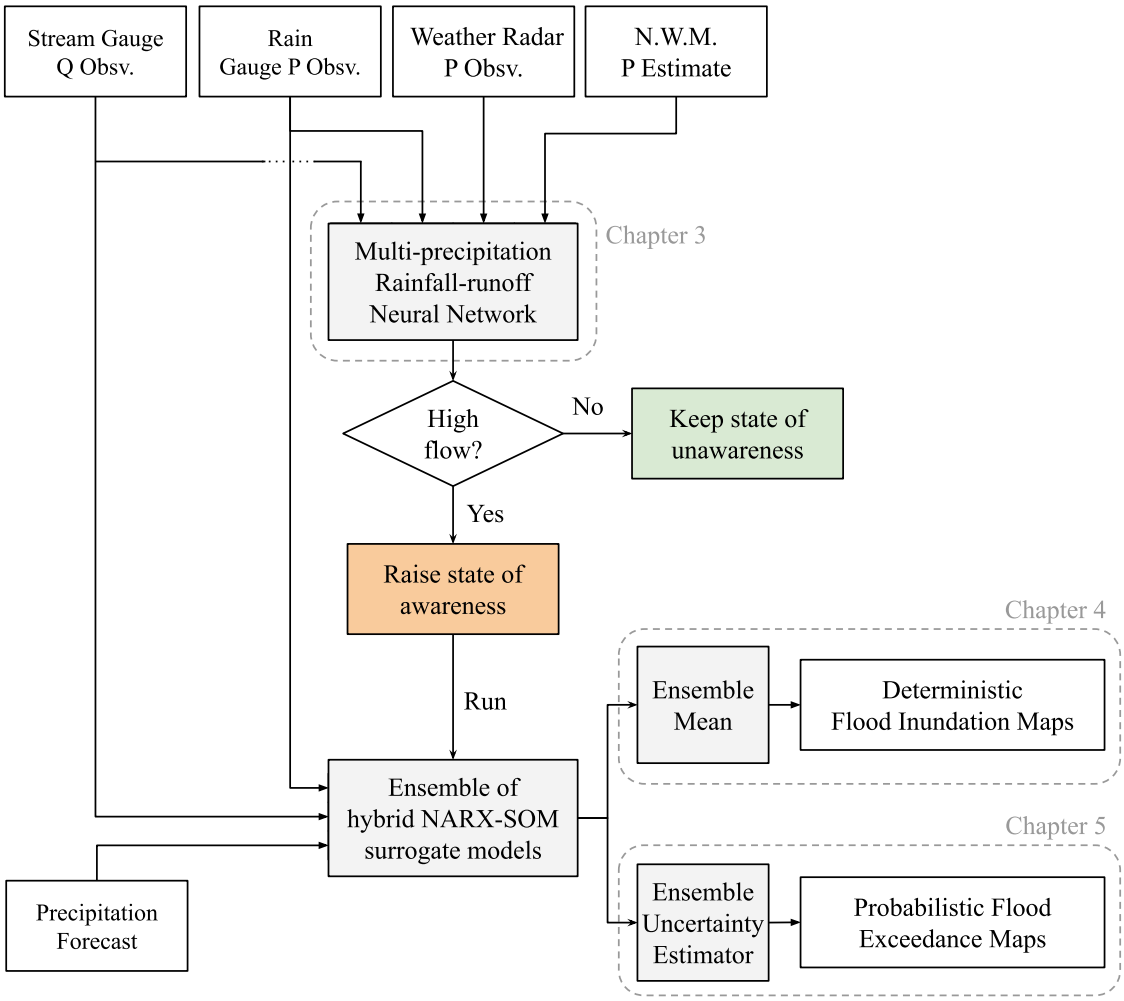

Figure 0-1. Hypothetical system representing all components of a flash flood forecasting system in which the techniques evaluated or proposed in this work are applied simultaneously.

6.2. Limitations and Recommendations for Future Research

The studies developed in Chapters 3, 4 and 5 used only a single catchment as study case. The experiments performed using the Don River Basin should be replicated in other catchments prone to flash floods in order to assess the generalizability of the findings.

Specifically concerning the study evaluating the use of multiple precipitation products as input for rainfall-runoff forecasting, only one type of neural network structure (extreme learning machine) was used among several alternatives possible. Other machine learning structures also popular among forecasters, such as the long-short term memory (LSTM) [1,2] and support vector machines [3], are also expected to benefit from the use of multiple precipitation estimates, and the assessment is expected to be developed in a future research work.

Regarding the generation of probabilistic flood inundation maps, in the presented work only the simple approach of quantile estimation from equally weighted members was applied. Several other post-processing methods for translating ensemble of values into probabilistic distributions can be explored. Examples of such include techniques based on minimizing the variance of the combination errors [4], on Bayesian model averaging [5] and on quantile model averaging [6], and exploring their applicability for the task of flood inundation forecasting is subject to future research works. Additionally, the water level distributions in the probabilistic forecast express only the uncertainties related to the process of surrogating the model, while other major uncertainties in hydrological forecasts are usually associated with quantitative precipitation estimates and forecasts [7]. In the availability of a study environment in which the development of flood inundation events is adequately documented, further studies comparing the different sources of uncertainties are encouraged to be developed with the inclusion of the surrogating uncertainties presented.

Similar to the conventional approaches in which calibrated hydrologic and hydraulic models are used in real time forecasting systems, models based on machine learning are expected to present a poor performance when subject to inputs out of the space of values used in the tunning step [8]. In a context of climate change, the data used for calibrating a model usually reflects the climatologic characteristics in the historical or present period, which may not represent adequately the future climate patterns [9,10]. Aware of such a nonstationary, the operational implementation of the methodologies discussed in this thesis would likely demand a recurrent reassessment and retraining of the calibrated models as new observational data becomes available. Considering that significant changes in the frequency and/or intensity of extreme precipitation events (i.e., the main trigger of the flash flood events explored in this thesis) may be detected in time spans of the order of decades, an annual retraining of the models is expected to be suitable for maintaining the predictions up to date with potential climate changes and variability.

6.3. References

1. Song, T.; Ding, W.; Wu, J.; Liu, H.; Zhou, H.; Chu, J. Flash Flood Forecasting Based on Long Short-Term Memory Networks. *Water (Basel)* **2019**, *12*, 109, doi:10.3390/w12010109.
2. Nevo, S.; Morin, E.; Rosenthal, A.G.; Metzger, A.; Barshai, C.; Weitzner, D.; Kratzert, F.; Elidan, G.; Dror, G.; Begelman, G.; et al. Flood Forecasting with Machine Learning Models in an Operational Framework. *Hydrol Earth Syst Sci* **2021**, 1–31.

3. Yan, J.; Jin, J.; Chen, F.; Yu, G.; Yin, H.; Wang, W. Urban Flash Flood Forecast Using Support Vector Machine and Numerical Simulation. *Journal of Hydroinformatics* **2018**, *20*, 232–245, doi:10.2166/hydro.2017.175.

4. Coulibaly, P.; Haché, M.; Fortin, V.; Bobée, B. Improving Daily Reservoir Inflow Forecasts with Model Combination. *Journal of Hydrologic Engineering* **2005**, *10*, 91–99, doi:10.1016/(ASCE)1084-0699(2005)10:2(91).

5. Muhammad, A.; Stadnyk, T.A.; Unduche, F.; Coulibaly, P. Multi-Model Approaches for Improving Seasonal Ensemble Streamflow Prediction Scheme with Various Statistical Post-Processing Techniques in the Canadian Prairie Region. *Water (Switzerland)* **2018**, *10*, doi:10.3390/w10111604.

6. Schepen, A.; Wang, Q.J. Model Averaging Methods to Merge Operational Statistical and Dynamic Seasonal Streamflow Forecasts in Australia. *Water Resources Research* **2015**, *51*, 1797–1812, doi:10.1002/2014WR016163.

7. Zappa, M.; Jaun, S.; Germann, U.; Walser, A.; Fundel, F. Superposition of Three Sources of Uncertainties in Operational Flood Forecasting Chains. *Atmospheric Research* **2011**, *100*, 246–262, doi:10.1016/j.atmosres.2010.12.005.

8. Tayfur, G.; Moramarco, T.; Singh, V.P. Predicting and Forecasting Flow Discharge at Sites Receiving Significant Lateral Inflow. *Hydrological Processes* **2007**, *21*, 1848–1859, doi:10.1002/hyp.6320.

9. Switzman, H.; Razavi, T.; Traore, S.; Coulibaly, P.; Burn, D.H.; Henderson, J.; Fausto, E.; Ness, R. Variability of Future Extreme Rainfall Statistics: Comparison

of Multiple IDF Projections. *Journal of Hydrologic Engineering* **2017**, *22*, 04017046, doi:10.1061/(asce)he.1943-5584.0001561.

10. Silva, D.F.; Simonovic, S.P.; Schardong, A.; Goldenfum, J.A. Assessment of Non-Stationary IDF Curves under a Changing Climate: Case Study of Different Climatic Zones in Canada. *Journal of Hydrology: Regional Studies* **2021**, *36*, 100870, doi:10.1016/j.ejrh.2021.100870.