Towards urban pluvial flood mapping using data-driven models

Cumulative dissertation

submitted for obtaining the degree of

Doctor of Engineering (Dr.-Ing.)

by

Omar Mahmoud Abdelwahab Seleem

Institute of Environmental Science and Geography
Faculty of Science
University of Potsdam

February 2023

First Supervisor:	Prof. Dr. Axel Bronstert
Second Supervisor:	PD Dr. Maik Heistermann
Mentor	Prof Dr. Oliver Korup
First Reviewer:	Prof. Dr. Axel Bronstert
Second Reviewer:	PD Dr. Maik Heistermann
Independent Reviewer:	Prof. Dr. Reinhard Hinkelmann

Examination board members:

1. Prof. Dr. Axel Bronstert
2. PD Dr. Maik Heistermann
3. Prof. Dr. Reinhard Hinkelmann
4. Prof. Dr. Annegret Thieken
5. Prof. Dr. Thorsten Wagener
6. Prof. Dr. Bruno Merz

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-59813
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-598137

Declaration of originality

I, Omar Seleem, hereby declare that, to the best of my knowledge, this work does not bear resemblance to any other work in whole or in part and has been completed by myself. I did not use any other sources and means than specified. Furthermore, this work has not been previously submitted to any university. All sources have been referred to and this work gives adequate credit to others for their work. I, in no way, claim to have created this information myself.

Location and Date	Omar Seleem

Abstract

Casualties and damages from urban pluvial flooding are increasing. Triggered by short, localized, and intensive rainfall events, urban pluvial floods can occur anywhere, even in areas without a history of flooding. Urban pluvial floods have relatively small temporal and spatial scales. Although cumulative losses from urban pluvial floods are comparable, most flood risk management and mitigation strategies focus on fluvial and coastal flooding. Numerical-physical-hydrodynamic models are considered the best tool to represent the complex nature of urban pluvial floods; however, they are computationally expensive and time-consuming. These sophisticated models make large-scale analysis and operational forecasting prohibitive. Therefore, it is crucial to evaluate and benchmark the performance of other alternative methods.

The findings of this cumulative thesis are represented in three research articles. The first study evaluates two topographic-based methods to map urban pluvial flooding, fill–spill–merge (FSM) and topographic wetness index (TWI), by comparing them against a sophisticated hydrodynamic model. The FSM method identifies flood-prone areas within topographic depressions while the TWI method employs maximum likelihood estimation to calibrate a TWI threshold (τ) based on inundation maps from the 2D hydrodynamic model. The results point out that the FSM method outperforms the TWI method. The study highlights then the advantage and limitations of both methods.

Data-driven models provide a promising alternative to computationally expensive hydrodynamic models. However, the literature lacks benchmarking studies to evaluate the different models' performance, advantages and limitations. Model transferability in space is a crucial problem. Most studies focus on river flooding, likely due to the relative availability of flow and rain gauge records for training and validation. Furthermore, they consider these models as black boxes. The second study uses a flood inventory for the city of Berlin and 11 predictive features which potentially indicate an increased pluvial flooding hazard to map urban pluvial flood susceptibility using a convolutional neural network (CNN), an artificial neural network (ANN) and the benchmarking machine learning models random forest (RF) and support vector machine (SVM). I investigate the influence of spatial resolution on the implemented models, the models' transfer-

ability in space and the importance of the predictive features. The results show that all models perform well and the RF models are superior to the other models within and outside the training domain. The models developed using fine spatial resolution (2 and 5 m) could better identify flood-prone areas. Finally, the results point out that aspect is the most important predictive feature for the CNN models, and altitude is for the other models.

While flood susceptibility maps identify flood-prone areas, they do not represent flood variables such as velocity and depth which are necessary for effective flood risk management. To address this, the third study investigates data-driven models' transferability to predict urban pluvial floodwater depth and the models' ability to enhance their predictions using transfer learning techniques. It compares the performance of RF (the best-performing model in the previous study) and CNN models using 12 predictive features and output from a hydrodynamic model. The findings in the third study suggest that while CNN models tend to generalise and smooth the target function on the training dataset, RF models suffer from overfitting. Hence, RF models are superior for predictions inside the training domains but fail outside them while CNN models could control the relative loss in performance outside the training domains. Finally, the CNN models benefit more from transfer learning techniques than RF models, boosting their performance outside training domains.

In conclusion, this thesis has evaluated both topographic-based methods and data-driven models to map urban pluvial flooding. However, further studies are crucial to have methods that completely overcome the limitation of 2D hydrodynamic models.

Zusammenfassung

Die Zahl der Todesopfer und Schäden durch Überschwemmungen in Städten nimmt zu. Ausgelöst durch kurze, lokal begrenzte und intensive Niederschlagsereignisse können urbane pluviale Überschwemmungen überall auftreten - sogar in Gebieten, in denen es in der Vergangenheit keine Überschwemmungen gab. Urbane pluviale Überschwemmungen haben eine relativ geringe zeitliche und räumliche Ausdehnung. Obwohl die kumulativen Verluste durch urbane pluviale Überschwemmungen vergleichbar sind, konzentrieren sich die meisten Hochwasserrisikomanagement- und -minderungsstrategien auf Fluss- und Küstenüberschwemmungen. Numerisch-physikalisch-hydrodynamische Modelle gelten als das beste Instrument zur Darstellung der komplexen Natur städtischer pluvialer Überschwemmungen; sie sind jedoch rechenintensiv und zeitaufwändig. Diese anspruchsvollen Modelle machen groß angelegte Analysen und operationelle Vorhersagen unerschwinglich. Daher ist es von entscheidender Bedeutung, die Leistung anderer Methoden zu bewerten und zu vergleichen, die komplexe hydrodynamische Modelle ersetzen könnten.

Die Ergebnisse dieser kumulativen Arbeit werden in drei Forschungsartikeln dargestellt. In der ersten Studie bewerte ich zwei topografiebasierte Methoden zur Kartierung von Überschwemmungen in Städten, die Fill-Spill-Merge-Methode (FSM) und den topografischen Nässeindex (TWI), indem ich sie mit einem hochentwickelten hydrodynamischen Modell vergleiche. Die FSM-Methode identifiziert überschwemmungsgefährdete Gebiete innerhalb topografischer Senken, während die TWI-Methode eine Maximum-Likelihood-Schätzung verwendet, um einen TWI-Schwellenwert (τ) auf der Grundlage von Überschwemmungskarten aus dem hydrodynamischen 2D-Modell zu kalibrieren. Die Ergebnisse zeigen, dass die FSM-Methode die TWI-Methode übertrifft. Anschließend werden die Vorteile und Grenzen beider Methoden aufgezeigt.

Datengesteuerte Modelle stellen eine vielversprechende Alternative zu rechenintensiven hydrodynamischen Modellen dar. In der Literatur fehlt es jedoch an Benchmarking-Studien zur Bewertung der Leistung, Vorteile und Grenzen der verschiedenen Modelle. Die räumliche Übertragbarkeit von Modellen ist ein entscheidendes Problem. Die meisten Studien konzentrieren sich auf Flussüberschwemmungen, was wahrscheinlich auf

die relative Verfügbarkeit von Abfluss- und Regenmesserdaten für Training und Validierung zurückzuführen ist. Außerdem betrachten sie diese Modelle als Black Boxes. In der zweiten Studie verwende ich ein Hochwasserinventar für die Stadt Berlin und 11 prädiktive Merkmale, die potenziell auf eine erhöhte pluviale Hochwassergefahr hinweisen, um die Anfälligkeit für pluviale Überschwemmungen in Städten zu kartieren. Dazu verwende ich ein Faltungsneuronales Netzwerk (CNN), ein Künstliches Neuronales Netzwerk (ANN) und die Benchmarking-Modelle Random Forest (RF) und Support Vector Machine (SVM). Ich untersuche den Einfluss der räumlichen Auflösung auf die implementierten Modelle, die Übertragbarkeit der Modelle im Raum und die Bedeutung der prädiktiven Merkmale. Die Ergebnisse zeigen, dass alle Modelle gut abschneiden und die RF-Modelle den anderen Modellen innerhalb und außerhalb des Trainingsbereichs überlegen sind. Die Modelle, die mit feiner räumlicher Auflösung (2 und 5 m) entwickelt wurden, konnten hochwassergefährdete Gebiete besser identifizieren. Schließlich zeigen die Ergebnisse, dass der Aspekt das wichtigste Vorhersagemerkmal für die CNN-Modelle ist, und die Höhe für die anderen Modelle.

Während Hochwasseranfälligkeitskarten überschwemmungsgefährdete Gebiete identifizieren, stellen sie keine Hochwasservariablen wie Geschwindigkeit und Wassertiefe dar, die für ein effektives Hochwasserrisikomanagement notwendig sind. Um dieses Problem anzugehen, untersuche ich in der dritten Studie die Übertragbarkeit datengesteuerter Modelle auf die Vorhersage der Überschwemmungstiefe in städtischen Gebieten und die Fähigkeit der Modelle, ihre Vorhersagen durch Transfer-Learning-Techniken zu verbessern. Ich vergleiche die Leistung von RF- (das beste Modell in der vorherigen Studie) und CNN-Modellen anhand von 12 Vorhersagemerkmalen und den Ergebnissen eines hydrodynamischen Modells. Die Ergebnisse der dritten Studie deuten darauf hin, dass CNN-Modelle dazu neigen, die Zielfunktion auf dem Trainingsdatensatz zu verallgemeinern und zu glätten, während RF-Modelle unter Overfitting leiden. Daher sind RF-Modelle für Vorhersagen innerhalb der Trainingsbereiche überlegen, versagen aber außerhalb davon, während CNN-Modelle den relativen Leistungsverlust außerhalb der Trainingsdomänen kontrollieren können. Schließlich profitieren die CNN-Modelle mehr von Transfer-Learning-Techniken als RF-Modelle, was ihre Leistung außerhalb der Trainingsbereiche erhöht.

Zusammenfassend lässt sich sagen, dass in dieser Arbeit sowohl topografiebasierte

Methoden als auch datengesteuerte Modelle zur Kartierung von Überschwemmungen in Städten bewertet wurden. Weitere Studien sind jedoch von entscheidender Bedeutung, um Methoden zu entwickeln, die die Beschränkungen von 2D-hydrodynamischen Modellen vollständig überwinden.

Acknowledgement

I would like to express my deepest gratitude to the people who supported me in different ways throughout this PhD thesis.

To Maik Heistermann for his unwavering support and guidance, especially when I felt overwhelmed or defeated. I thank him for the fruitful discussions and critical reviewing of the manuscripts.

To Axel, for support in all forms, for keeping me always motivated and optimistic during difficult times, for believing in me when I had lost faith in myself.

I thank Prof. Reinhard Hinkelmann for agreeing to act as an external reviewer on this thesis.

I would like to thank all members of NatRiskChange and the Havel working group - Bora, Sophia, Arthur, Paul, and Georgy for discussions about work and non-work subjects.

My warmest thanks go to Seth Bryant for proofreading parts of the thesis.

To my best friends – Dissoky, Salah, Abdel Aziz and Mostafa– for their unconditional support during the ups and downs. I thank my friends in Berlin-Potsdam for providing a sense of home away from home.

To my family for their unwavering support and belief in me, for being close despite the distance.

To Aise, for supporting me, providing strength and love during this entire time.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Background and motivation

Urban pluvial flooding has caused heavy damage and loss of life around the world, e.g. Hull (UK) in 2007 (Coulthard & Frostick 2010), Rome (Italy) in multiple events over the past decades (Di Salvo et al. 2017), Beijing (China) in 2012 (Zhang et al. 2015), and several cities in Turkey in the last decade (Koç et al. 2021). In Germany, heavy rainfall events caused damage in cities such as in Dortmund 2008 (Bung et al. 2011), Muenster 2014 (Spekkers et al. 2017), Berlin in 2017 and 2019 (Berghäuser et al. 2021, Dillenardt et al. 2022, Caldas-Alvarez et al. 2022) and Aachen in 2018 (Hofmann & Schüttrumpf 2021).

Urbanization changes natural surfaces to impervious surfaces such as buildings and pavements, reducing the landscape's capacity for infiltration, and accelerating runoff generation. On the other hand, climate change has been increasing the frequency and severity of heavy rainfall events (Myhre et al. 2019). These factors combine to significantly increase the frequency and severity of urban pluvial flooding (Ashley et al. 2005, Skougaard Kaspersen et al. 2017), and hence the caused economic damage and casualties (Merz et al. 2010, Smith 2006, Nofal & Van De Lindt 2020*b*). The Intergovernmental Panel on Climate Change reports support this presumption (Qin et al. 2007, Bates et al. 2008, Pörtner et al. 2022). Furthermore, regional climate models predict future changes in the frequency and severity of heavy rainfall events in many regions in Europe (Pfeifer et al. 2015, Martel et al. 2020). Seneviratne et al. (2006) predicted an increase in extreme rainfall events in Germany under the changing climate. The ongoing urbanization and climate change, along with the densification of the human population and assets, will exacerbate the risk of urban pluvial flooding, making cities more vulnerable. (Rosenzweig et al. 2018).

Urban pluvial flooding occurs when the rainfall intensity exceeds the capacity of the urban stormwater drainage system and the ground's capacity to infiltrate water (Houston et al. 2011) (see Figure 1.1). These floods are typically caused by short intense local rainfall such as the flood events in Muenster in 2014 (Spekkers et al. 2017) and Aachen in 2018 (Hofmann & Schüttrumpf 2021). They are less often caused by long events; but

such flooding can occur such as the flooding in Berlin in 2017 (Berghäuser et al. 2021, Caldas-Alvarez et al. 2022). Urban pluvial flooding could theoretically occur in any location within an urbanized watershed, even in areas without flooding history (Zhang & Pan 2014). It depends on the local meteorology and the catchment's characteristics such as topography, soil type, and land cover (Zhang & Pan 2014, Skougaard Kaspersen et al. 2017). Falconer et al. (2009) emphasized the importance of not confusing pluvial flooding with surface water flooding, with the latter referring to combined flooding in urban areas during heavy intense rainfall. It is caused by a combination of factors such as blocked drainage systems, sewer flooding, pluvial flooding, and excessive groundwater overflow (Falconer et al. 2009, Prokic et al. 2019). Pluvial flooding also differs from flash flooding, which typically arises from a watercourse and is associated with heavy intense rainfall (Borga et al. 2010, Prokic et al. 2019).

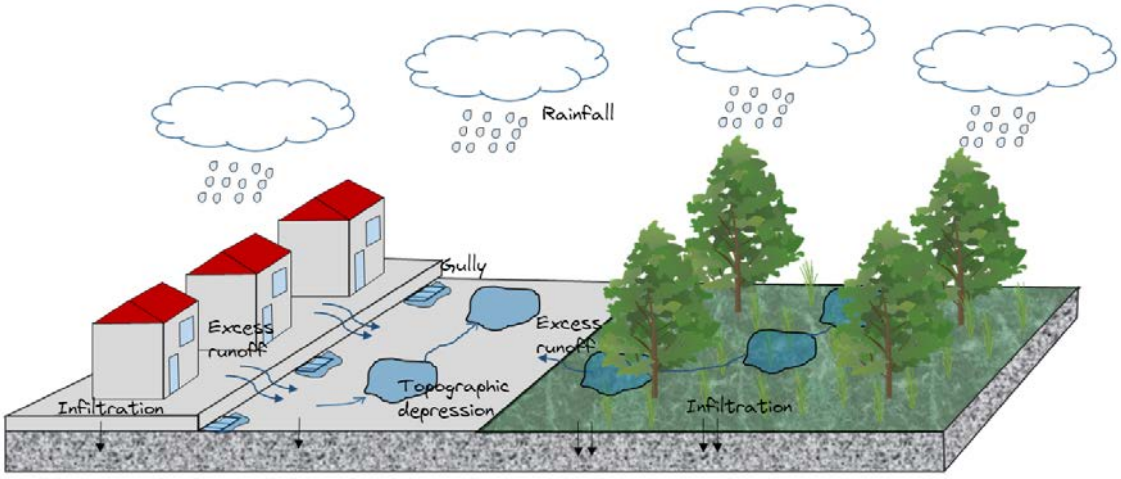


Fig. 1.1 Permeable surfaces such as natural ground, facilitate the infiltration of rainfall, which reduces the volume and velocity of runoff. Conversely, sealed surfaces such as buildings and pavements reduce water infiltration, causing an increase in excess runoff. Some of the excess runoff is drained by the storm drainage system, while the remaining becomes a surface flow that accumulates in topographic depressions.

Urban pluvial floods usually occur at small temporal and spatial scales compared to river and coastal floods. However, they occur frequently at different locations and cause comparable cumulative damage over years to the damage caused by river and coastal floodings (Ten Veldhuis 2011, Tanaka et al. 2020). In recent years, urban pluvial flooding has caused property damage in the hundreds of millions of euros, traffic disruptions, and in some cases fatalities as shown in Table 1.1. Unlike river floods, people affected by urban pluvial floods are less well informed about possible dangers and take fewer precautions themselves (Thieken et al. 2022). Although the literature is rich with risk management and mitigation strategies for river and coastal floods (Samela et al. 2020), there are fewer strategies to face the increasing pluvial flood risk (Penning-Rowsell et al. 2005, Zhou et al. 2012, Hammond et al. 2015). Furthermore, the European Floods Di-

Table 1.1 Examples of flood damage and casualties caused by urban pluvial flooding in Germany in the last decades.

City	Event	Fatalities	Damage (million)	Reference
Dortmund	26th July 2008	3	€17	Rözer et al. (2016)
Muenster	28th July 2014	2	€140	Burrough et al. (2015)
Berlin and Potsdam	29th and 30th June 2017	-	€60	Dillenardt et al. (2022)
Berlin	Summer 2019	-	€50	Berghäuser et al. (2021)

rective (2007/60/EC) obliges all member states to prepare flood hazard maps for major rivers. While pluvial floods are excluded from the directive, municipalities in Germany are responsible for preparing pluvial flood maps (Kaiser 2021). There are several types of flood maps that can be used to identify and understand flood risk. Bentivoglio et al. (2022) divided the flood maps into three categories as shown in Figure 1.2:

1. Inundation extent maps (see Figure 1.2a) often obtained from remote sensing analysis. They show the observed inundation extent for a specific flood event.

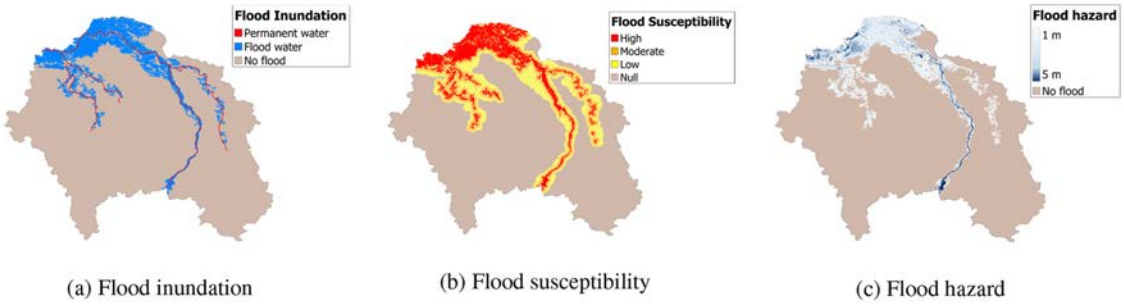
2. Flood susceptibility maps (see Figure 1.2b) which identify areas that are likely to be affected by floods. They are created by analysing the combination of factors that contribute to flood risk such as topography, soil type, land use, and historical flood inventories. They can be created using different methodologies, such as statistical analysis and data-driven models.

3. Flood hazard maps (see Figure 1.2c) which are often created using physical numerical models. They show the spatial distribution of variables that define the flood hazard of a specific event such as floodwater depth and velocity.

These maps are widely used, each with its own limitations. For example, remote sensing data are not always able to capture the flood peak and can require manual refinement (Notti et al. 2018). Historical flood inventories are not always available to map flood susceptibility (Zhao et al. 2020). Finally, numerical hydrodynamic simulations are computationally expensive and time-consuming (Löwe et al. 2021).

1.2 Modelling urban pluvial flooding

Modelling urban pluvial flooding is crucial to flood risk management in urban areas. It involves simulating the runoff generation and the movement of water through the urban environment. It considers factors such as rainfall patterns, topography and infrastructure. The model estimates the flood extent, water depth and the volume of generated



(a) Flood inundation (b) Flood susceptibility (c) Flood hazard

Fig. 1.2 Examples of flood map types. (a) flood inundation map, (b) flood susceptibility map, and (c) flood hazard map. The figure is obtained from Bentivoglio et al. (2022).

runoff by a specific rainfall event. The complex nature of urban pluvial floods is considered the major challenge to developing a fast and accurate flood model at both spatial and temporal resolution (Merz & Thieken 2005, Teng et al. 2017). Conventional urban pluvial flood models could be grouped into hydrodynamic models and simplified topographic-based models (Bulti & Abebe 2020). Finally, data-driven models are rising as a potential that overcomes the limitations of the computationally expensive hydrodynamic models.

1.2.1 Hydrodynamic models

Hydrodynamic models are widely used to simulate flood dynamics (Teng et al. 2017). They solve physical equations with various degrees of complexity. They predict variables such as flood extent, flood water depth, and velocity, which are necessary for flood risk management strategies (Néelz et al. 2010, Néelz & Pender 2013). They can be categorised as one-dimensional (1D), two-dimensional (2D), and three-dimensional (3D) models (Teng et al. 2017, Bulti & Abebe 2020). The 1D hydrodynamic models are used to model the flow in a pipe or a confined channel. They are considered too simple to represent floodplains. The 2D models are considered the best representation of the floodplain flow in two dimensions assuming that the third dimension is shallow compared to the two others (Teng et al. 2017). Finally, the 3D hydrodynamic models are more suitable to simulate floods due to dam breaks, levee breaches and tsunamis. They are regarded as complexly unnecessary for urban pluvial flooding (Teng et al. 2017).

The rainfall-runoff process in urban areas describes how rainwater becomes surface runoff. During this process, some of the rainwater evaporates or infiltrates into the ground, while the rest will flow into the storm drainage system and over the surface as runoff (Bedient et al. 2008). The process is influenced by various factors ranging from rainfall patterns and terrain profiles to building footprints, urban storm drainage capacity, and even gullies and manhole locations (Houston et al. 2011). The best-established approach to simulate the rainfall-runoff process in a complex urban environment is coupling the 1D and 2D hydrodynamic models (see Figure 1.3) where the 1D model

simulates the flow in the storm drainage system and the 2D model simulates the surface flow (Tayefi et al. 2007, Maksimović et al. 2009, Patel et al. 2017, Fan et al. 2017, Bermúdez et al. 2018). However, the storm drainage system data is often unavailable (Rangari et al. 2018) and the 2D hydrodynamic models can be computationally demanding. They are suitable neither for large urban watersheds (Costabile et al. 2017, Bates 2022) nor for real-time predictions (Hsu et al. 1995, Bhola et al. 2019).

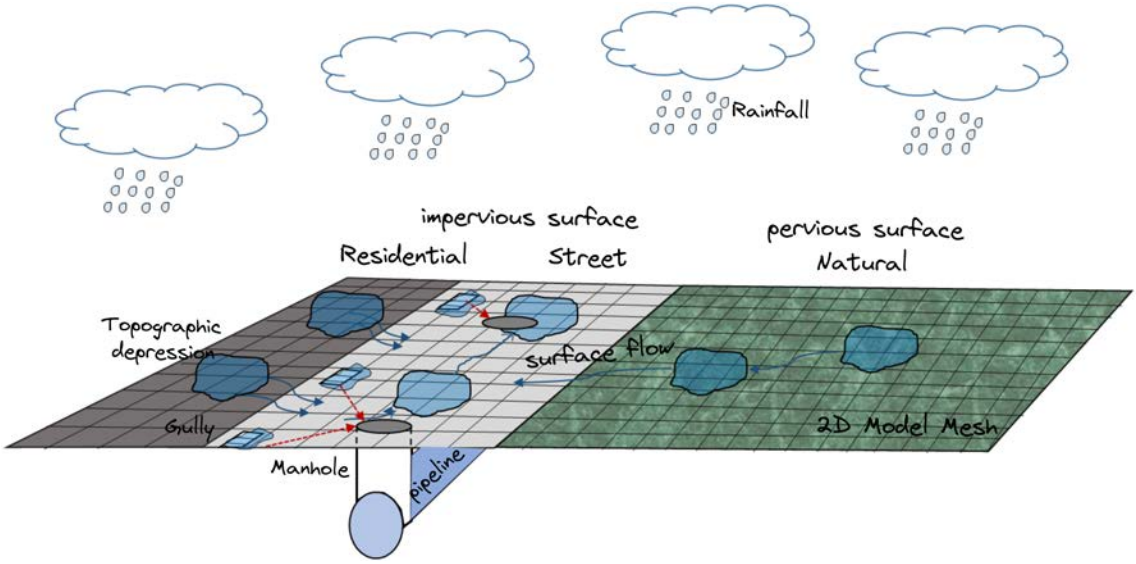


Fig. 1.3 Schematisation of a coupled 1D-2D hydrodynamic model. The 1D model simulates the part of the generated runoff that enters the storm drainage system through gullies and flows to the nearest manhole, while the 2D model simulates the remaining runoff that becomes a surface flow.

1.2.2 Simplified topography-based models

Simplified topographic-based models simulate the behaviour of urban flooding by first filling topographic depressions and then spilling the excess volume to travel downhill and accumulate in the downstream topographic depressions (Zhang & Pan 2014, Samela et al. 2017, Bulti & Abebe 2020). These models are relatively fast and only require terrain and rainfall data (Liu et al. 2015, Yang et al. 2015, Shen et al. 2015); however, model outputs are limited to final flood depth and extent. The output does not include other flood variables, such as velocity, flood duration, and time of the peak, upon which flood risk management decisions are based (Fritsch et al. 2016).

Previous studies proposed methods to identify urban pluvial flood-prone areas based on topographic indices (Samela et al. 2018, Manfreda & Samela 2019, Huang et al. 2019, Kelleher & McPhillips 2020). Kelleher & McPhillips (2020) showed that the topographic wetness index (TWI) could identify urban pluvial flood-prone areas. Huang et al. (2019) proposed a topography-based index to identify urban pluvial flood-prone areas, namely the topographic control index (TCI). De Risi et al. (2015, 2018) used

the maximum likelihood function to estimate a TWI threshold (τ) based on inundation maps from a 2D hydrodynamic model for a given window within the city. Locations with TWI above the estimated threshold (τ) are then considered flood-prone areas. These methods are often used to identify flood-prone areas in large regions. However, there are no benchmark studies that evaluate these methods.

1.2.3 Data-driven models

Recently, using data-driven models to map flooding is gaining growing attention (Guo et al. 2021). The implemented models to map flooding in the literature could be classified into the point and image-based models. Point-based traditional machine learning algorithms such as support vector machine (SVM) (Cortes & Vapnik 1995) and random forest (RF) (Breiman 2001) have shown their ability to map flood susceptibility (Tehrany et al. 2014, 2015, Zhao et al. 2020) and to predict floodwater depth (Zahura et al. 2020, Hou et al. 2021). They establish a relationship between the local flood predictive features and flood occurrence (Zhao et al. 2020). However, they have not been tested outside their training domains and are mostly applied to map fluvial flooding as shown in Figure 1.4. It points out that previous studies tended to use artificial neural networks (ANN) and multilayer perceptron (MLP) rather than classification and regression trees (CART), support vector machine (SVM), support vector regression (SVR), decision trees (DT), and random forest (RF).

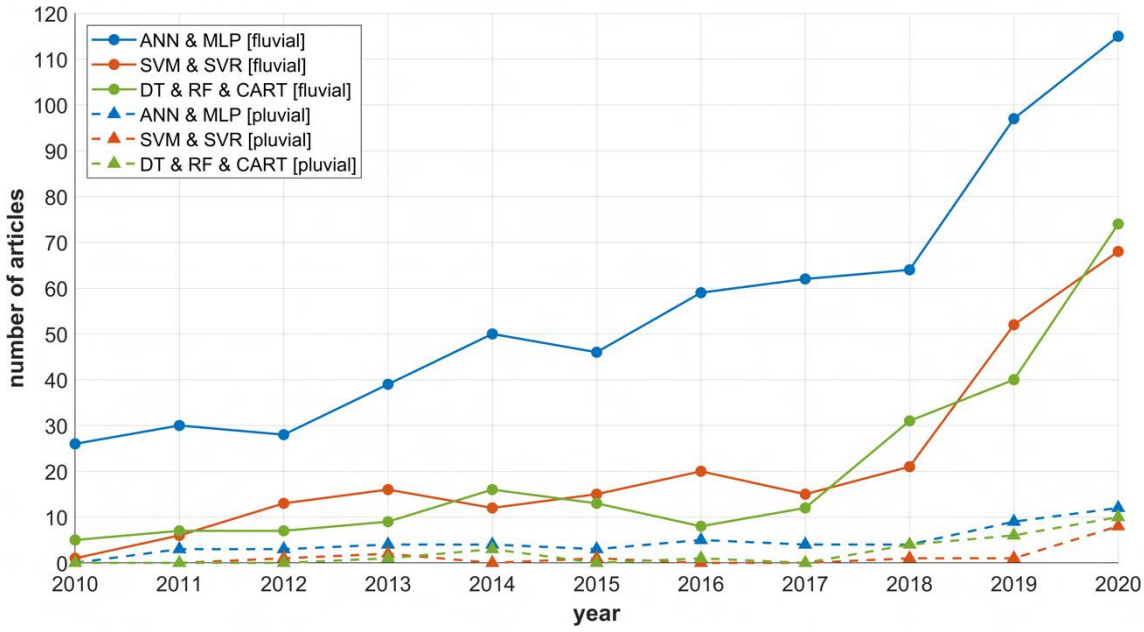


Fig. 1.4 Overview of studies using data-driven models for fluvial and pluvial flood prediction based on Web Scopus (search on 09.06.2021). The figure is obtained from Hofmann & Schüttrumpf (2021).

Recently, deep learning represented in convolutional neural networks (CNN) (image-based model) has been used to map floods (Wang et al. 2020, Zhao et al. 2020, Löwe et al. 2021, Guo et al. 2022). CNN models have the advantage over traditional machine learning algorithms that they could learn patterns in two dimensions and understand relevant patterns around a predicted location. Recent publications claimed that deep learning models outperform traditional machine learning algorithms. However, Grinsztajn et al. (2022) put these claims into question.

Hofmann & Schüttrumpf (2021) showed that data-driven models developed by previous studies aimed mostly to predict fluvial flooding due to the availability of historical hydrologic data from stream gauge stations and their corresponding meteorologic data. Similar datasets are not available for urban pluvial flooding. Therefore, data generated from hydrodynamic models and flood inventories (reported flood locations) are being used to develop data-driven models to predict floodwater depth and map flood susceptibility respectively caused by urban pluvial flooding. The models' performance is limited with the amount of available data and the model transferability to new areas outside the training domains is a major challenge (Bentivoglio et al. 2022).

In summary, data-driven models are raising as a surrogate that might overcome the limitations of the computationally expensive hydrodynamic models (Löwe et al. 2021, Guo et al. 2022, Bentivoglio et al. 2022). However, a data-driven model that generalizes outside the training domain is still a major challenge (Bentivoglio et al. 2022). In addition, these models are not commonly used to map urban pluvial flooding due to the lack of reliable flood inventories (Yang et al. 2016) and in-situ measurements (Hofmann & Schüttrumpf 2019).

1.3 Research questions and structure

The computationally expensive hydrodynamic models are suitable neither for identifying urban pluvial flood-prone areas nor mapping flood hazards for a large urban watershed using fine spatial resolution. Hence the overarching aim of this thesis is to evaluate alternatives to map urban pluvial flooding. In regard to the overarching aim, this thesis comprises three main Chapters in the form of manuscripts to facilitate answering the following research questions:

RQ1: Can simplified topographic-based methods mimic the two-dimensional hydrodynamic models to simulate urban pluvial flooding?

The research question is answered in the second chapter, where I evaluate and compare the fill-spill-merge (FSM) and the TWI methods to identify urban pluvial flood-prone areas. FSM is a simplified topographic-based method while the TWI method

identifies a TWI threshold (τ) which defines flood-prone areas. Chapter 2 highlights the advantage and limitations of each method.

RQ2.1: Does deep learning outperform traditional machine learning algorithms to map urban flood susceptibility?

RQ2.2: Does the spatial resolution of the predictive features impact the data-driven models' prediction?

RQ2.3: How transferable are data-driven models to map urban flood susceptibility?

RQ2.4: Which predictive features are most useful for flood susceptibility mapping?

These questions are addressed in the third chapter, where I compare the performance of various data-driven models to map urban pluvial flood susceptibility. Chapter 3 shows the implemented data-driven models and predictive features used to train, validate and test the models. The model's output is a map which shows the flood susceptibility of a certain location. These maps are useful to identify flood-prone areas but they do not show the important variables for flood hazard mapping, e.g. floodwater depth and velocity which led us to the following research questions:

RQ3.1: How transferable are data-driven models to predict urban pluvial floodwater depth?

RQ3.2: Can transfer learning techniques improve the model performance outside the training domain and thus help to overcome the issue of limited training data?

I approach these questions in the fourth chapter. I compare the performance of random forest (RF) (the best-performing model in the previous study) with a deep learning model to predict urban pluvial floodwater depth among and outside the training domain. Then, I investigate if the data-driven models could boost their performance outside the training domain beneficially from transfer learning techniques.

1.4 Author Contributions

The main core of this thesis consists of three research papers, which have been published, or are submitted and intended to be published in international peer-reviewed journals. The manuscripts are adjusted to the formatting of this thesis. I would like to express my appreciation and gratitude to the co-authors, namely Axel Bronstert, Maik Heistermann, Georgy Ayzel and Arthur Costa Tomaz de Souza who contributed to these manuscripts. The contributions are as follows:

Chapter 2: Seleem, O., Heistermann, M. and Bronstert, A., 2021. Efficient Hazard Assessment for Pluvial Floods in Urban Environments: A Benchmarking Case Study for the City of Berlin, Germany. Water, 13(18), p.2476.

Conceptualization: OS and MH.; Methodology, software, formal analysis: OS; Writing—original draft preparation: OS and MH; Visualization: OS and MH; Supervision: AB and MH; Writing -review & editing: OS, MH, AB

Chapter 3: Seleem, O., Ayzel, G., de Souza, A.C.T., Bronstert, A. and Heistermann, M., 2022. Towards urban flood susceptibility mapping using data-driven models in Berlin, Germany. Geomatics, Natural Hazards and Risk, 13(1), pp.1640-1662.

Conceptualization: OS, GA and MH.; Methodology, software, formal analysis: OS and AC; Writing—original draft preparation: OS and MH; Visualization: OS and MH; Supervision: AB and MH; Writing -review & editing: OS, GA, MH, AB

Chapter 4: Seleem, O., Ayzel, G., de Souza, A.C.T., Bronstert, A. and Heistermann, M., 2022. Towards urban flood susceptibility mapping using data-driven models in Berlin, Germany. Geomatics, Natural Hazards and Risk, 13(1), pp.1640-1662.

Conceptualization: OS, GA and MH.; Methodology, software, formal analysis: OS; Writing—original draft preparation: OS and MH; Visualization: OS and MH; Supervision: AB and MH; Writing -review & editing: OS, GA, MH, AB

In addition to the above-mentioned manuscripts, I also contributed to the following publication, which is not included in the thesis:

Berghäuser, L., Schoppa, L., Ulrich, J., Dillenardt, L., Jurado, O.E., Passow, C., Samprogna Mohor, G., **Seleem, O.,** Petrow, T. and Thieken, A.H., 2021. Starkregen in Berlin: Meteorologische Ereignisrekonstruktion und Betroffenenbefragung.Task force report, Universität Potsdam, https://doi.org/10.25932/publishup-50056, 2021.

CHAPTER 2

Efficient hazard assessment for pluvial floods in urban environments: A benchmarking case study for the city of Berlin, Germany

---

This chapter is published as:

Seleem, O., Heistermann, M. and Bronstert, A., 2021. Efficient Hazard Assessment for Pluvial Floods in Urban Environments: A Benchmarking Case Study for the City of Berlin, Germany. Water, 13(18), p.2476.

Abstract

The presence of impermeable surfaces in urban areas hinders natural drainage and directs the surface runoff to storm drainage systems with finite capacity, which makes these areas prone to pluvial flooding. The occurrence of pluvial flooding depends on the existence of minimal areas for surface runoff generation and concentration. Detailed hydrologic and hydrodynamic simulations are computationally expensive and require intensive resources. This study compared and evaluated the performance of two simplified methods to identify urban pluvial flood-prone areas, namely the Fill–Spill–Merge (FSM) method and the topographic wetness index (TWI) method and used the TELEMAC-2D hydrodynamic numerical model for benchmarking and validation. The FSM method uses common GIS operations to identify flood-prone depressions from a high-resolution digital elevation model (DEM). The TWI method employs the maximum likelihood method (MLE) to probabilistically calibrate a TWI threshold (τ) based on the inundation maps from a 2D hydrodynamic model for a given spatial window (W) within the urban area. We found that the FSM method clearly outperforms the TWI method both conceptually and effectively in terms of model performance.

2.1 Introduction

Pluvial flooding causes massive human and economic losses (Smith 2006, Nofal & Van De Lindt 2020*b*). They are considered as an omnipresent hazard both in urban and ru-

ral areas, e.g., (Tabari et al. 2021, Bronstert et al. 2018). Pluvial floods are caused by intensive and short duration precipitation storms due to the limited capacity of storm drainage systems, typically called "minor systems", which result in the generation of overland flow that travels through the terrain, depressions and the road networks, creating a surface flow network typically called the "major system". The major system could transfer the overland flow for a long distance, which can cause flooding in an area that is located far from the location where the minor system's capacity was exceeded (Maksimović et al. 2009, Guerreiro et al. 2017).

Pluvial flood-prone areas can be defined as the low elevated areas within the urban watershed where excess runoff accumulates (Schanze 2018). It is challenging to predict pluvial floods as they could strike areas with no flood records with little warning time. They have no defined flood plains such as rivers and seashores (Di Salvo et al. 2017). Furthermore, they depend on how well the storm drainage system and the associated infrastructure respond to a sudden precipitation storm. The damage depends not only on the event intensity and duration but also on physical and social urban characteristics (Leal et al. 2019, Kumar et al. 2021, Nofal & van de Lindt 2020*a*). The European Flood Directive widely implements flood hazard maps for rivers and coastal areas, but corresponding efforts are scarce for pluvial flooding (Union 2017, Zhou et al. 2012, Di Salvo et al. 2017).

Urban pluvial floods are affected by several local factors, for example, the presence and maintenance conditions of the storm drainage system, the existence of underground infrastructures, and the extent of impervious surfaces within the urban watershed (Di Salvo et al. 2017). Sustainable urban development requires efficient pluvial flood modeling (Zhang & Pan 2014). Urban pluvial flood hazard mapping is commonly carried out by performing hydrologic modeling to estimate the excess runoff followed by a 1D-2D hydrodynamic simulation to compute the inundation extent and depth and runoff velocity. This method is considered the most accurate representation of the storm drainage system. However, this method is computationally expensive, requires comprehensive knowledge of hydrology and hydraulics, as well as detailed input data such as topography, land cover, soil type, and storm drainage system data, which hamper its widespread application in a large urban watershed for pluvial flood hazard mapping and early warnings (Schumann et al. 2011, Neal et al. 2009, Hunter et al. 2005).

In recent years, the increasing availability of higher resolution DEM has encouraged not only the advanced application of the 2D hydrodynamic models but also the development of quick and simplified methods that are based on DEM as valid alternatives for flood hazard mapping. These simplified methods have been mainly developed and applied for mapping fluvial flood hazards, and their application for pluvial flooding hazard mapping is still sparse in the literature (Samela et al. 2020). They are less accurate than using detailed hydrologic and 1D-2D hydrodynamic models, but they are

faster and require fewer input data. Therefore, they are suitable for flood hazard mapping and early warnings for urban areas. Many of the simplified models are based on the fill–spill–merge (FSM) concept where the excess runoff concentrates in the DEM depressions and when a depression is completely filled with excess runoff, the excess runoff spills, and flows downstream towards lower elevated depressions. (Zhang & Pan 2014, Balstrøm & Crawford 2018, Samela et al. 2020) applied the FSM method to different case studies in China, Italy, and Denmark. FSM can map the pluvial flood hazard in urban watersheds quickly as it only shows the final inundation extent. Samela et al. (2020) evaluated the FSM method performance based on evaluating the estimated inundation area extent with the obtained inundation area extent from a 2D hydrodynamic model. However, they did not evaluate the water depth, which is vital for damage assessment. Therefore, this study evaluates the FSM method performance based on both flood extent and the water depth for pluvial flooding.

Other previous studies proposed methods to delineate flood-prone areas based on geomorphic classifiers (Manfreda & Samela 2019, Samela et al. 2018, 2017). They are suitable for data-scarce regions and to analyze large geographic regions that have limited information on flooding potential. Kelleher & McPhillips (2020) showed that urban flood reported areas tend to have high TWI values and suggested using the TWI as an index for urban pluvial flooding. On the contrary, Huang et al. (2019) showed that TWI failed to map the urban pluvial flooding and proposed a DEM-based index to map the urban pluvial flooding, namely the topographic control index (TCI). References (Jalayer et al. 2014, De Risi et al. 2015, 2018) proposed a framework for identifying urban flood-prone areas based on TWI. The framework (hereafter, the TWI method) uses the maximum likelihood method (MLE) to probabilistically calibrate the TWI threshold (τ) based on the inundation maps from a 2D hydrodynamic model for a given spatial window (W) within the urban area, where areas with TWI values that exceed the TWI threshold (τ) are defined as flood-prone areas.

The objective of this study is to evaluate and compare the FSM method and the TWI method to assess and map pluvial flooding and highlight the limitations of each method. We applied the two methods to two urban areas in Berlin for different precipitation depths ranging from 30 to 150 mm (10 mm increments) and used the TELEMAC-2D hydrodynamic numerical model for benchmarking and validation to show how these methods can mimic the behavior of a 2D-hydrodynamic model with regard to inundation depth and inundation extent. Finally, we discuss the advantages and limitations of each method.

2.2 Data

The high-resolution DEM was downloaded from Berlin's city geographical information systems spatial databases (ATKIS 2020). The DEM has a 1 $\times$ 1 m horizontal resolution and 20 cm vertical accuracy, as shown in Figure 2.1a. The soil type, hydrologic condition (Ross et al. 2018) and land use data from open street map (Haklay & Weber 2008) were used to generate the curve number map. Furthermore, the radar precipitation data for the period between 2005 and 2019 were obtained from the German Weather Service (DWD) (Winterrath et al. 2017, Kreklow et al. 2019).

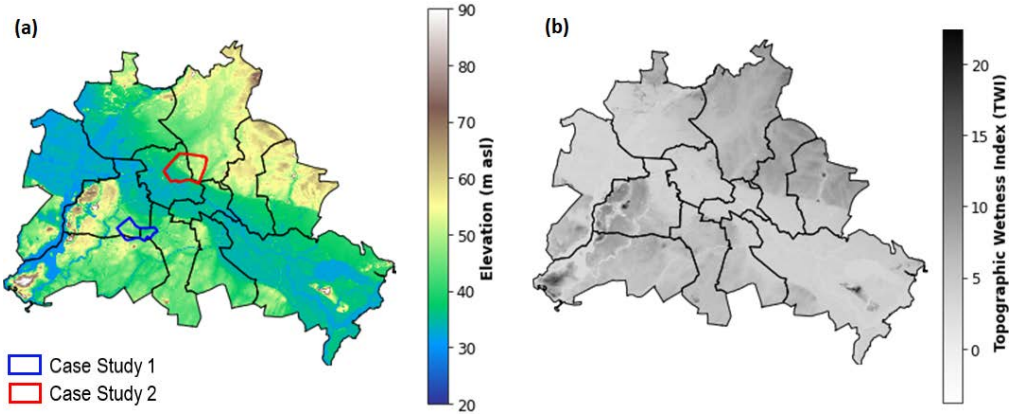


Fig. 2.1 (**a**) Digital elevation model (DEM) of Berlin, the two case studies and the city districts. (**b**) The spatial distribution of TWI for Berlin and the city districts.

2.3 Models

2.3.1 Two-Dimensional Hydrodynamic Simulations

We used the TELEMAC-2D model to perform the 2D hydrodynamic simulations. The TELEMAC-2D model (Hervouet 2007, Galland et al. 1991) is an open-source model and included in the TELEMAC-MASCARET suite. It solves the shallow water equations using both the finite volume method (FVM) and the finite element method (FEM) over non-structured triangular grids. Urban flood inundation mapping is considered a challenge because of the complex characteristics of the urban watersheds. Buildings are known as one of the most significant components of urban flood modeling because they withstand the flow of excess runoff. Li et al. (2019) showed that the building resistance (BR) method is the best representation of buildings in numerical models. The BR method is based on assigning a high roughness coefficient value to the buildings to allow excess runoff to flow through the buildings. Furthermore, the roughness coefficient is also a significant uncertainty factor in urban flood modeling. In the current study, the Manning roughness coefficient was defined based on the surface type as follows: 10^4

$m^{-1/3} \cdot s$ for buildings and 0.013 for other surface types (Testa et al. 2007, Papaioannou et al. 2018, French & French 1985).

The Blue Kenue software was used to define the boundary conditions and to generate the non-structured triangular mesh with a maximum side length of 1 m for the TELEMAC-2D model (Kenue 2021). A one-hour event duration and precipitation depths ranging from 30 to 150 mm (10 mm increments) were used to perform the simulations. The precipitation depth range was selected based on the analysis of the extreme precipitation events that caused urban flooding in Berlin in the period between 2005 and 2019 (Winterrath et al. 2017, Kreklow et al. 2019). Infiltration and losses were estimated by the SCS-CN method per pixel. We tested both the FVM and FEM. Eventually, FVM was used for all the numerical simulations to obtain better simulation accuracy (R. 2019). Finally, a variable computation time step was defined according to the Courant's criterion (Cu $\leq$ 1.0). We used the inundated areas from the TELEMAC-2D model as a "true" reference to evaluate the other implemented methods.

2.3.2 The Fill–Spill–Merge (FSM) Method

The FSM method identifies flood-prone areas based on nested depressions as extracted from a DEM, the accumulated precipitation volume in depressions, and the vertical and horizontal connectivity between the depressions: when a depression is filled, excess runoff water spills and flows downstream to depressions at a lower elevation, as shown in Figure 2.2a. Figure 2.2b, and c show the depression filling process and the vertical hierarchy structure, respectively. The FSM method considers flood-prone areas as areas within the depression with an elevation below a specified water level, as shown in Figure 2.2.

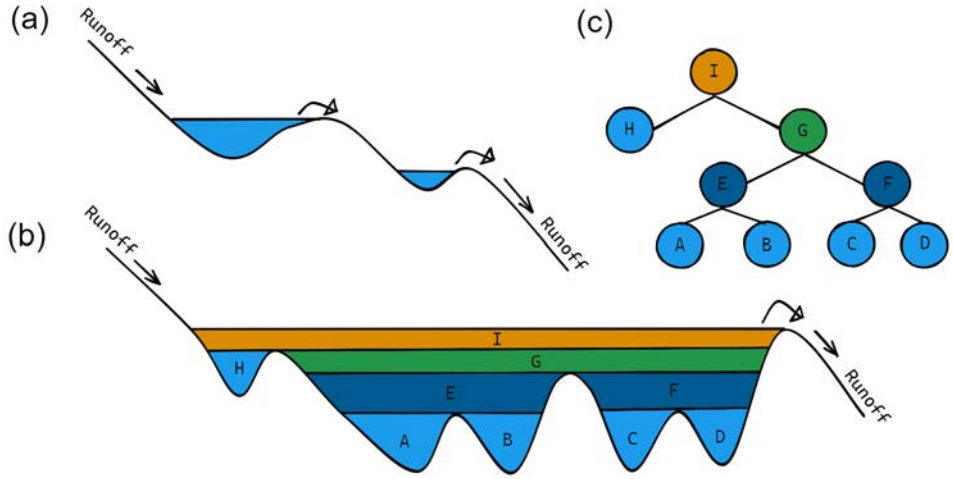


Fig. 2.2 Process of the FSM method. (**a**) Runoff flowing process between depressions. (**b**) Schematic of the nested depressions. (**c**) Vertical hierarchical tree.

We implemented three main steps to estimate the inundation depth, extent, and volume in a study area, as shown in Figure 2.3: (1) DEM prepossessing and derivation of effective rainfall, (2) flood routing and (3) hazard mapping.

1. **DEM prepossessing and derivation of effective rainfall**: The first step in delineating a watershed is to fill sinks, in which existing depressions (called also sinks and pits) in the DEM are filled to guarantee stream connectivity. However, excess runoff is accumulated in these depressions in urban watersheds. Therefore, it is vital to identify them in urban areas as they are more prone to urban flooding. Depressions were identified by computing the elevation difference between the filled DEM and the original DEM in ArcGIS (v. 10.5.1). After that, only depressions with a depth of more than 0.20 m (the vertical accuracy of the original DEM) and surface areas bigger than 1000 m^2 were considered. We calculated the flow direction and the flow accumulation using the D8 algorithm from the filled DEM and used them to estimate the flow paths, spilling points, and the vertical hierarchy of the depressions. Finally, we obtained the partially filled DEM by filling the depressions in the original DEM except for the selected depressions. The partially filled DEM was used to estimate the contributing watershed for each depression. The curve number method was used to estimate the effective rainfall for each pixel. Therefore, the curve number map for Berlin was created based on the soil type, hydrologic condition from Ross et al. (2018) and land use from an open street map (Haklay & Weber 2008), Figure 2.3 shows the implemented workflow.

2. **Flood routing**: In general, rainfall is transformed to either evapotranspiration, infiltration, or surface and sewer system runoff. Runoff at the surface and in the urban storm drainage system could contribute to pluvial flooding in terms of inundation of areas in depressions. This study neglected the latter because of the difficulties of obtaining the detailed urban storm drainage system data if they exist. In addition, the urban storm drainage system tends to be ineffective in the case of intensive rainstorms (Di Salvo et al. 2018, Falconer et al. 2009). Using the Hydrologic Engineering Center—Hydrologic Modeling System (HEC-HMS) model (Scharffenberg & Harris 2008), we estimated the excess runoff based on the SCS method (Cronshey 1986) for precipitation depths ranging from 30 to 150 mm (10 mm increments), performed the runoff routing, and estimated the stored depth and volume at each depression.

3. **Hazard mapping**: We estimated the water depth and inundation extent inside the depressions by subtracting the obtained water level minus the elevation for each pixel from the original DEM in ArcGIS (v. 10.5.1).

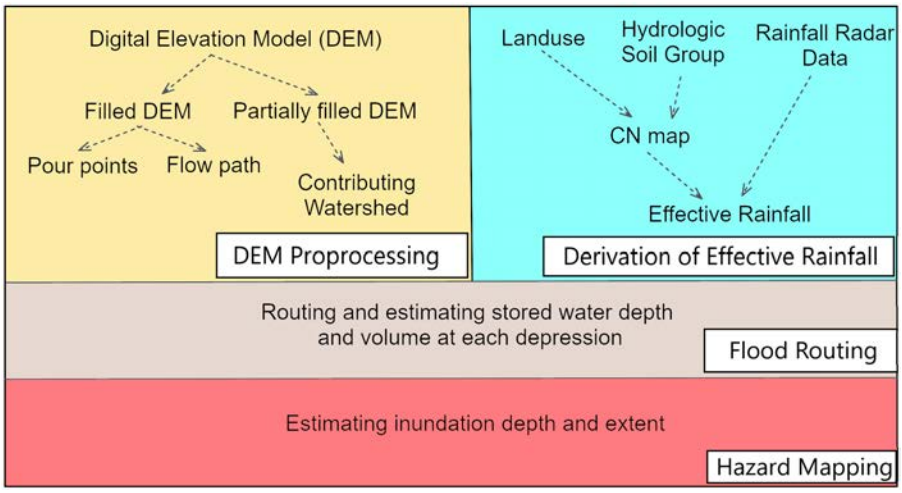


Fig. 2.3 FSM method's workflow.

2.3.3 TWI Method

The TWI was initially introduced by Kirkby (1975) for conceptual hydrological modeling at the catchment scale on mountainous and hilly terrain. It was intended to reflect the tendency of a location along a hillslope to generate surface runoff from saturation excess but was later used to generally represent the tendency of a location to accumulate water subject to the local slope and the upstream contributing area. To that end, it was also suggested as an index for pluvial flooding in urban areas (Kelleher & McPhillips 2020). We obtained the TWI from a DEM using ArcGIS (v. 10.5.1): the eight-direction (D8) flow model was used to calculate the flow direction from the fully filled DEM, upslope area and local slope. Finally, we estimated the TWI on a cell-by-cell basis, as shown in Figure 2.1b:

$$TWI = ln(a/tan\beta), \tag{2.1}$$

The parameter a represents the upslope contributing area per grid length, and β is the local slope angle.

We applied the TWI method in Berlin to identify flood-prone areas (referred to as FP). The underlying assumption is that a location is flood-prone if its TWI exceeds a certain threshold τ. The estimation of τ employs the maximum likelihood method, as described in detail by (Jalayer et al. 2014, De Risi et al. 2018, 2015). We used the inundation maps from the TELEMAC-2D model for two case studies within Berlin, which have frequent flooding, to estimate the TWI threshold (τ) for each case independently.

Hence, FP represents the pluvial flood-prone area that meets the condition TWI $> \tau$, and IN represents the inundated area as obtained from the TELEMAC-2D model for a certain precipitation event depth (simulated water depth h is larger than 0, $h > 0$) as shown in Figure 2.4a. The probability of correctly identifying the flood-prone area,

or the likelihood function for τ (L(τ|W)), can be calculated for various values of τ as follows:

$$L(\tau|W) = P(FP, IN|\tau, W) + P(\overline{FP}, \overline{IN}|\tau, W) \tag{2.2}$$

where $P(FP, IN|\tau, W)$ indicates the probability that a certain pixel within a spatial window W is identified both as inundated IN (based on the calculated inundation maps from TELEMAC-2D model) and FP (based on the TWI method), and conditioned on a given value of τ. Furthermore, $P(\overline{FP}, \overline{IN}|\tau, W)$ indicates the probability that a certain pixel within the spatial window W is neither identified as IN nor as FP conditioned on a given τ.

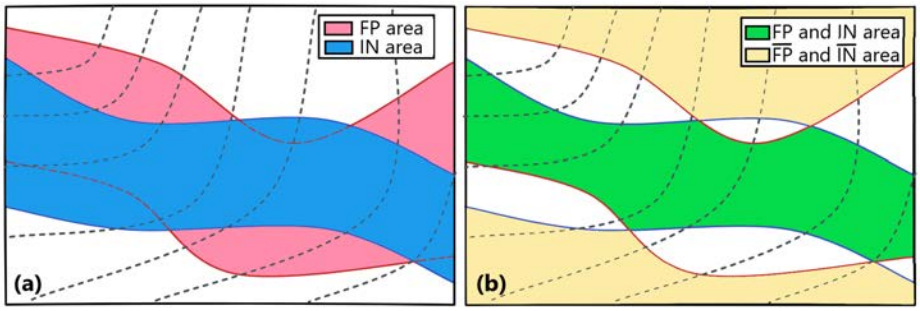


Fig. 2.4 Flood-prone (FP) and inundated (IN) areas.

$P(FP, IN|\tau, W)$ and $P(\overline{FP}, \overline{IN}|\tau, W)$ could be expanded by utilizing the product rule of the probability theory (Jaynes 2003), as follows:

$$P(FP, IN|\tau, W) = P(FP|\tau, W) \cdot P(IN|FP, \tau, W) \tag{2.3}$$

$$P(\overline{FP}, \overline{IN}|\tau, W) = P(\overline{FP}|\tau, W) \cdot P(\overline{IN}|\overline{FP}, \tau, W) \tag{2.4}$$

where $P(FP|\tau, W)$ and $P(\overline{FP}|\tau, W)$ indicate the probability that a certain pixel within the spatial window W is FP and $\overline{FP}$, respectively, conditioned on a given value of τ. The $P(IN|FP, \tau, W)$ indicates the probability that a certain pixel within the spatial window W is identified as IN on the condition that it is FP, and $P(\overline{IN}|\overline{FP}, \tau, W)$ indicates the probability that a certain pixel within the spatial window W is identified as $\overline{IN}$ on the condition that it is $\overline{FP}$, for a given value of τ.

Maximum Likelihood Estimation

1. **Mesoscale estimations**: The terrain in Berlin is relatively flat. Therefore, we calculated the $P(FP|\tau, W)$ using the spatial extent of the case study area as follows:

$$P(FP|\tau, W) = \frac{A_{\text{urban}}(FP)}{A_{\text{urban}}} \tag{2.5}$$

where $A_{\text{urban}}(FP)$ indicates the FP area within the case study area (area with TWI values $> \tau$) and A_{urban} is the case study area.

2. **Microscale estimations**: $P(IN|FP,\tau,W)$ and $P(\overline{IN}|\overline{FP},\tau,W)$ were calculated using the spatial extent of the selected window within the study area, as shown in Figure 2.5 as follows:

$$P(IN|FP,\tau,W) = \frac{A_{\text{W}}(IN,FP)}{A_{\text{w}}(FP)} \tag{2.6}$$

$$P(\overline{IN}|\overline{FP},\tau,W) = \frac{A_{\text{W}}(\overline{IN},\overline{FP})}{A_{\text{w}}(\overline{FP})} \tag{2.7}$$

where $A_{\text{W}}(IN,FP)$ indicates the area that is defined as both IN and FP within the selected window (the green-colored area in Figure 2.4b), and $A_{\text{W}}(FP)$ indicates the area that is defined as FP within the selected window. Furthermore, $A_{\text{W}}(\overline{IN},\overline{FP})$ indicates the area that is defined as neither IN nor FP within the selected window (the yellow-colored area in Figure 2.4b) and $A_{\text{W}}(\overline{FP})$ indicates the area that is defined as $\overline{FP}$ within the selected window.

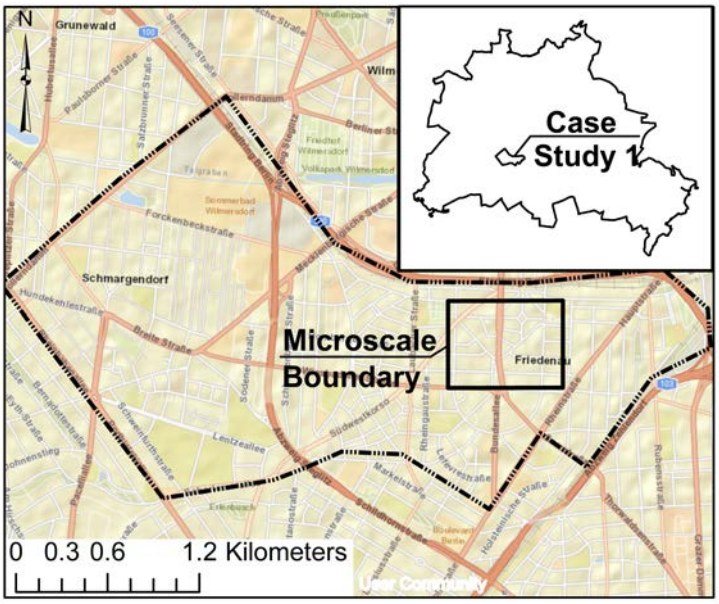


Fig. 2.5 The spatial extent for the mesoscale and microscale.

Finally, we calculated the likelihood function Equation (2.2) and estimated the TWI threshold value (τ) as the value that maximizes the likelihood function.

2.3.4 Benchmarking Experiments

1—Case Studies

We selected two case studies within Berlin, as shown in Figure 2.1a. They have been frequently inundated in the last decades (Berghäuser et al. 2021). Case study 1 and 2 areas are 6.3 and 12.5 km^2, respectively. Depressions represent 14% and 8% of case study 1 and 2 areas, respectively. Depressions with a surface area more than 1000 m^2 represent approximately 89% of the area of the total depression in the two case studies. The estimated historical precipitation depth that caused flooding in the period between 2005 and 2019 for the two case studies ranges from 30 to 150 mm based on the rainfall radar data analysis (Winterrath et al. 2017, Kreklow et al. 2019).

2—Evaluation Metrics

We compared the output of the two methods, FSM and TWI, with the output from the TELEMAC-2D model and considered the latter as a true reference. Considering the uncertainty of the FSM method, TWI method, and TELEMAC-2D model and the difference in the computational domain discretization (i.e., pixel-based for FSM method and TWI method, and unstructured triangular mesh for TELEMAC-2D model), we considered pixels with a simulated water depth (h) higher than 0.10 m as flooded for both the FSM method and the TWI method for comparing and evaluating the performance of the two methods. Then, we categorized every pixel into one of the following four categories:

1. True positive (TP): correctly classified as flooded;
2. True Negative (TN): correctly classified as non-flooded;
3. False positive (FP): incorrectly classified as flooded;
4. False negative (FN): incorrectly classified as non-flooded.

The performance of the applied methods was then evaluated based on the following metrics (Table 2.1):

Table 2.1 Performance indices.

height**Index**	**Equation**	**Range**
Sensitivity(TPR)	$\frac{TP}{TP+FN}$	[0 - 1]
Matthews correlation coefficient (MCC)	$\frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP) \times (TP+FN) \times (TN+FP) \times (TN+FN)}}$	[-1 - 1]
Flood Area Index (FAI)	$\frac{TP}{TP+FP+FN}$	[0 - 1]

The Sensitivity (TPR) refers to the proportion of the pixels that are correctly identified as inundated among all the inundated pixels from the true reference. The Matthews correlation coefficient (MCC) measures the correlation between binary classifications, and the Flood Area Index (FAI) refers to the agreement in identifying the inundated area with the true reference.

We evaluated the output water depth (h) from the FSM method and TELEMAC-2D model by the Nash Sutcliffe Efficiency index (NSE) (Nash & Sutcliffe 1970).

$$NSE = 1 - \frac{\sum_{i=1}^{n}(W_i^{Obs} - W_i^{sim})^2}{\sum_{i=1}^{n}(W_i^{Obs} - W^{mean})^2} \tag{2.8}$$

where W_i^{Obs} is the water depth from the TELEMAC-2D model; W_i^{sim} is the water depth from the FSM method; W^{mean} is the mean water depth from TELEMAC-2D model, and n is the total number of pixels. NSE value is between$-\infty$ and 1.0, values less than or equal to zero indicate unacceptable performance, a value between 0 and 1.0 indicates acceptable performance and 1.0 indicates total agreement.

2.4 Results and Discussion

We applied the FSM and TWI methods to identify pluvial flood prone areas for two case studies in Berlin by using input precipitation depths from 30 to 150 mm (10 mm increments) and compared the results to the inundation extent obtained from the TELEMAC-2D model. We structured the result section as follows: First, we estimated the TWI threshold (τ) for the two case studies separately, considering IN areas as areas with water depths (h) greater than zero (h $>$ 0). We then estimated the TWI threshold (τ) for various definitions of IN areas (water depth (h) greater than zero, 1, 5, and 10 cm) for case study 1. Finally, we evaluated the performance of the FSM method and the TWI method compared to the TELEMAC-2D model performance.

2.4.1 Maximum Likelihood Estimation

We calculated the likelihood function $L(\tau|w)$ for all possible values of τ by using Equation (2.2). This was performed separately for different event precipitation depths between 30 and 150 mm. Figure 2.6 shows the estimated maximum likelihood for a 100 mm precipitation depth for case study region 1. The TWI threshold (τ) value associated with the maximum likelihood is τ = 5.3. This means, hence, that for a 100 mm storm, we would expect all areas in region 1 with a $TWI > 5.3$ to inundate. Furthermore, we defined the maximum likelihood interval as the τ interval that represents the 99 th percentile of the estimated maximum likelihood values. The maximum likelihood interval is [4.8, 6.4] for case study 1 for the 100 mm precipitation depth, as shown in Figure 2.6.

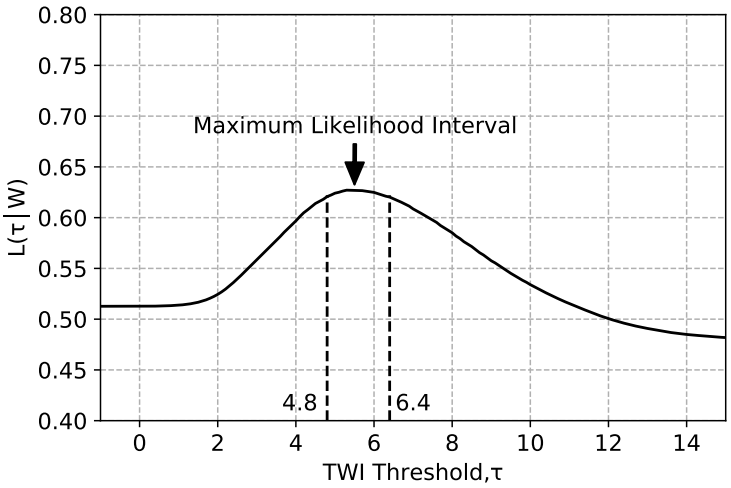


Fig. 2.6 Likelihood function L(τ/W) for a 100 mm precipitation depth (case study 1).

Figure 2.7 shows the estimated TWI threshold (τ) for the two case studies together with the 99% confidence interval, depending on precipitation depth. The values of τ are very similar for the two case studies. It also appears that after exceeding a certain precipitation threshold (100 mm in the two case studies), τ does not decrease much more with increasing the precipitation. A possible explanation is that a further increase in precipitation does not increase the inundation *extent* but only the inundation *depth*. This is shown in Figure 2.8 with the inundation-extent for 100 and 150 mm precipitation depths and the inundation depth difference from the two precipitation depth scenarios.

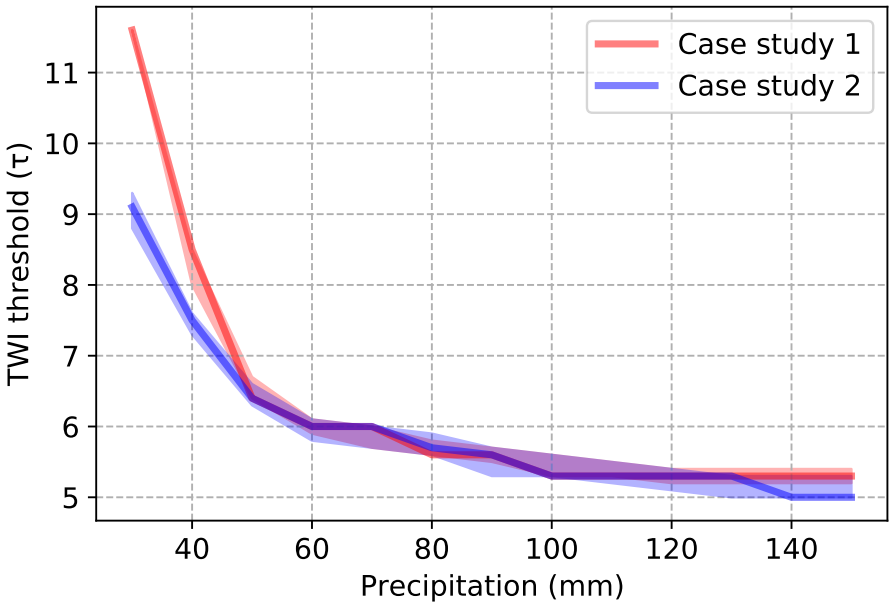


Fig. 2.7 TWI threshold (τ) for the two case studies with the 99% confidence interval.

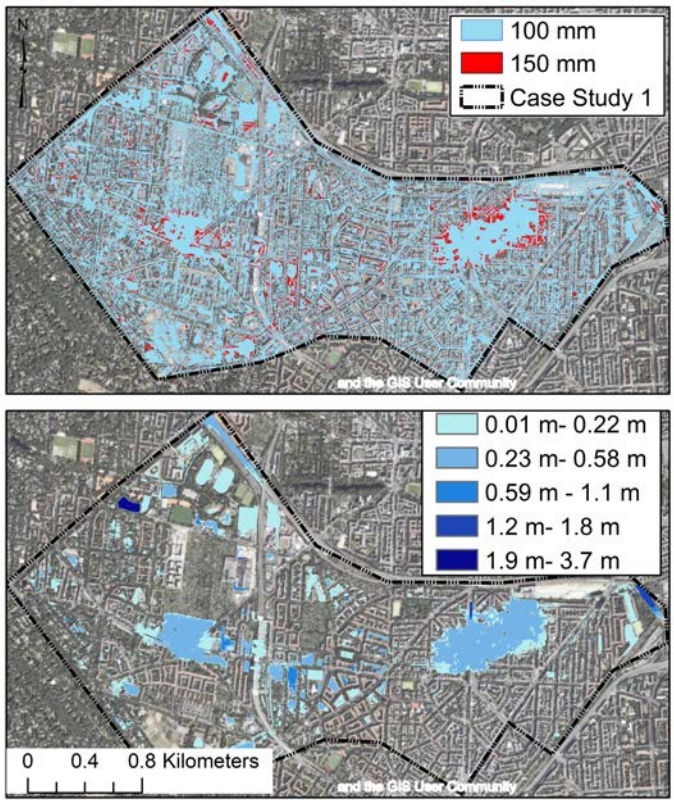


Fig. 2.8 Inundation-extent for 100 and 150 mm precipitation depths for case study 1.

2.4.2 Water Depth (h) Threshold

(De Risi et al. 2015, 2018, Jalayer et al. 2014) defined inundated areas as areas with a water depth greater than zero. However, inundation starts to become a disturbance for pedestrians when the water depth exceeds 10 cm (height of sidewalk) and cars when the water depth is typically higher than 30 cm (depth at which the floodwater reaches the car electronics and causes traffic chaos) (Huang et al. 2019). Therefore, we estimated the TWI threshold (τ) by defining the inundated areas as areas with water depths greater than zero, 1, 5, and 10 cm for case study 1. The likelihood function $L(\tau|W)$ for case study 1 for a precipitation depth 100 mm is shown in Figure 2.9 for the different water depth thresholds as a function of τ. As expected, the τ value corresponding to the maximum likelihood increased with increasing the water depth threshold. As a consequence, flood-prone areas as identified by the TWI method decrease with an increasing water depth threshold, as shown in Figure 2.10. Furthermore, Figure 2.10 shows that estimated flood-prone areas did not change with increasing the water depth threshold from 1 to 10 cm, which could be explained by the low difference between the estimated TWI thresholds (τ) for these water depth thresholds.

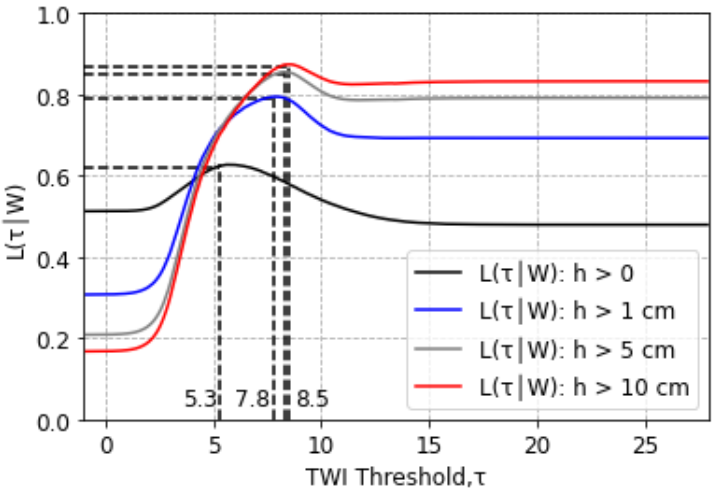


Fig. 2.9 Likelihood function L(τ/W) for a 100 mm precipitation depth (case study 1).

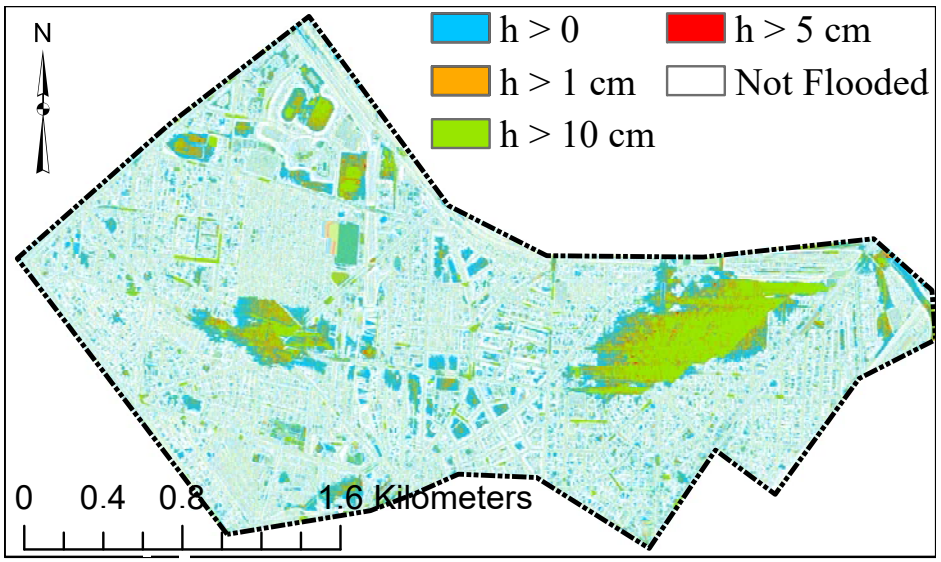


Fig. 2.10 Flood-prone area (FP) for different water depth thresholds for a 100 mm precipitation depth.

Figure 2.11a shows the estimated TWI threshold (τ) for different water depths thresholds and precipitation depths for case study 1. As we can see, the TWI method could not find a local maximum for the likelihood function and hence no τ value for lower precipitation depths together with water depth thresholds greater than zero (h $>$ 0). Obviously, this limits the applicability of the TWI method to identify flood-prone areas for impact-relevant water depth thresholds in combination with rather low (but also impact-relevant) precipitation scenarios. The reason behind the inability to identify a value for τ in such cases (i.e., the inability to find a local maximum in the likelihood function) can be explained by looking at the two components of the likelihood function Equation (2.2): the $P(FP, IN|\tau, W)$ (the probability of areas being identified both as flood-prone by the TWI method and as inundated based on the TELEMAC-2D model), and $P(\overline{FP}, \overline{IN}|\tau, W)$ (the probability of areas being neither identified as flood-prone nor modeled as inundated). As we can see in Figure 2.12a, an increase in the water depth threshold (h) beyond 0 cm reduces the inundated area (IN) so much that the probability term $P(FP, IN|\tau, W)$ is reduced to just a negligible contribution to the maximum likelihood value $L(\tau|W)$. As a consequence, $L(\tau|W)$ is dominated by the probability of identifying an area as neither flood-prone nor inundated ($P(\overline{FP}, \overline{IN}|\tau, W)$), as shown in Figure 2.12b,c. In that case, maximizing $L(\tau|W)$ corresponds to choosing the maximum possible TWI value in the spatial domain W.

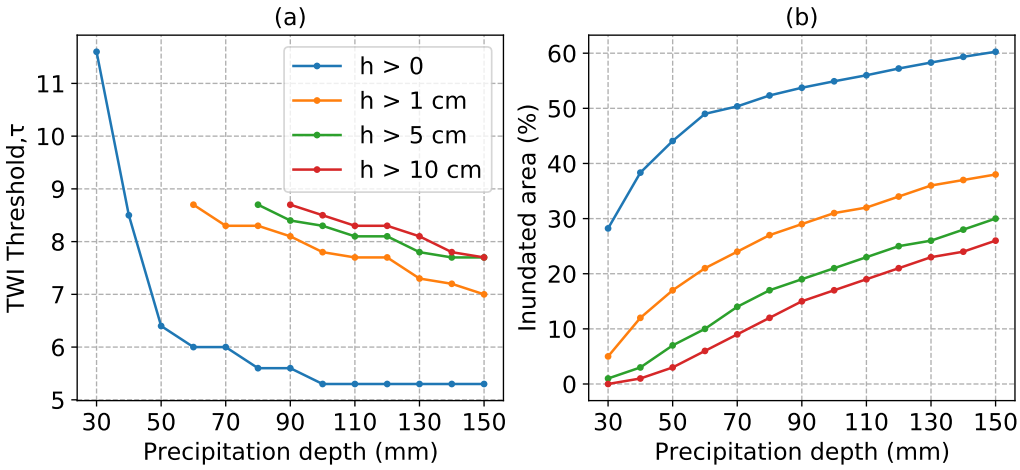


Fig. 2.11 (**a**) TWI threshold (τ) for different water depth thresholds (h). (**b**) Percentage of inundated areas for different water depth thresholds (h).

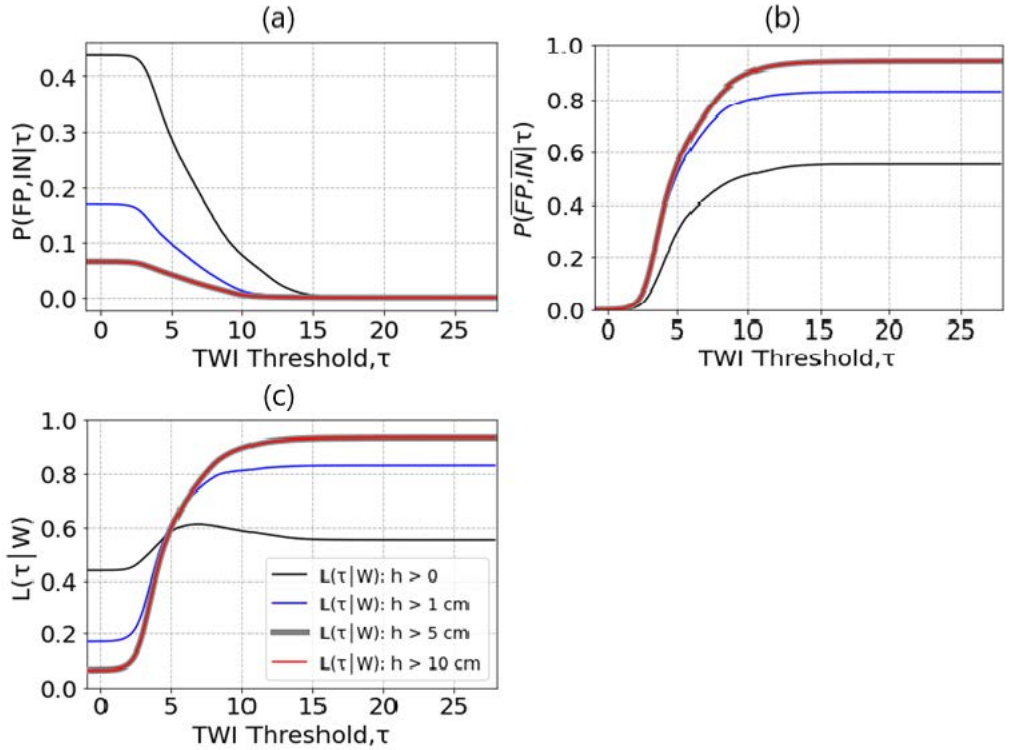


Fig. 2.12 Probabilities of 50 mm precipitation event.

2.4.3 Evaluating FSM and TWI Methods Performance

Figure 2.13 shows the performance of the FSM and TWI methods with regards to their ability to represent the inundation extent as simulated by the TELEMAC-2D model for the different precipitation scenarios (from 30 to 150 mm) and case study areas. A pixel is considered as flooded if the simulated water depth (h) exceeds 0.10 m. For this

threshold of h, the performance of the TWI method could only be measured for precipitation depths of 90 mm and more (see the above discussion about how the estimation of τ is limited).

Figure 2.13 highlights two main results: first, the FSM method clearly outperforms the TWI method for those precipitation depths for which both methods are applicable (i.e., 90 mm and more). This holds for both case studies and all performance metrics (sensitivity TPR, Matthew Correlation Coefficient MCC, Flood Area Index FAI). For a 100 mm precipitation event, e.g., the FSM method has a TPR of 0.72 and 0.68 (case studies 1 and 2) while the TWI method only achieves TPR values of 0.37 and 0.26 for the same event.

Second, the performance of all methods tends to increase with precipitation depth. This is most pronounced for case study 1 with both methods across all performance metrics: TPR increases from 0.31 to 0.79, MCC from 0.08 to 0.76, and FAI from 0.02 to 0.67. For case study 2, the performance metrics increase steeply at first. Then, MCC and FAI values of the FSM method slightly decline at precipitation depths of more than 110 mm, and TPR exhibits a more pronounced decrease beyond 100 mm of precipitation. The reason for this behavior of the FSM method in case study 2 is that the depressions in this area were completely filled with excess runoff at a precipitation depth around 100 m, so that inundation spread to areas outside the depressions. The FSM method, however, only represents inundation within depressions. Therefore, the method fails to represent flooding beyond depressions and hence shows a performance decrease for such situations. For the TWI method, such a decrease in performance does not occur as it is not explicitly limited to represent depressions (but obviously, the performance of the TWI method continues to increase at a much lower performance level). Please note that, for a more comprehensive evaluation, we provided additional evaluation metrics (precision, specificity, accuracy, and balanced accuracy) in Figure S1 in the supplementary material. However, these additional metrics do not provide new insights as to the differences in performance for the different models, case studies, and precipitation thresholds, so we do not discuss them here in further detail.

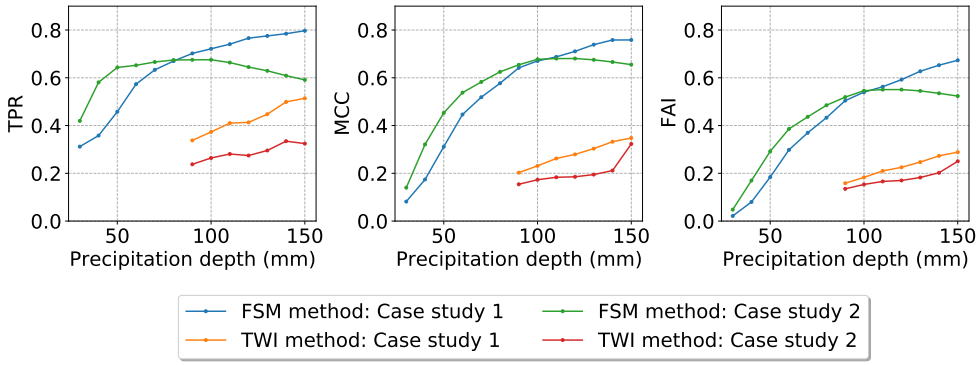


Fig. 2.13 Performance indices for the FSM and TWI methods for different precipitation scenarios.

In Figure 2.14, we compare the spatial distribution of inundated areas from FSM and TWI methods with the inundation obtained from the TELEMAC-2D model for a 100 mm precipitation depth scenario. With this figure, we can better understand why the FSM method outperforms the TWI method so clearly. Obviously, the TWI method fails to identify the inundated areas since the TWI concept does not include the concept of a "depression" as a flood-prone structure and hence fails to correctly detect depressions in terms of flood-prone areas. It is exactly the inherent strength of the FSM method to identify such depressions, so that the inundation extent obtained from the FSM method agrees well with the TELEMAC-2D model.

In order to evaluate the ability of the FSM method not only to identify flood-prone areas but also the water depth in inundated areas, we calculated the Nash Sutcliffe Efficiency (NSE) for the two case studies, as shown in Figure 2.15. The NSE values for case study 1 continuously increase with precipitation depth from -68.66 to 0.78. For case study 2, the NSE increases from -4.42 to 0.47, and starts to saturate (and slightly decrease) at precipitation depths above 70 mm. Given that the NSE indicates skill only for values above 0 (as compared to just assuming an average observation), the FSM method can be considered as skillful to quantify inundation depth only for precipitation depths beyond 90 (case study 1) and 50 mm (case study 2).

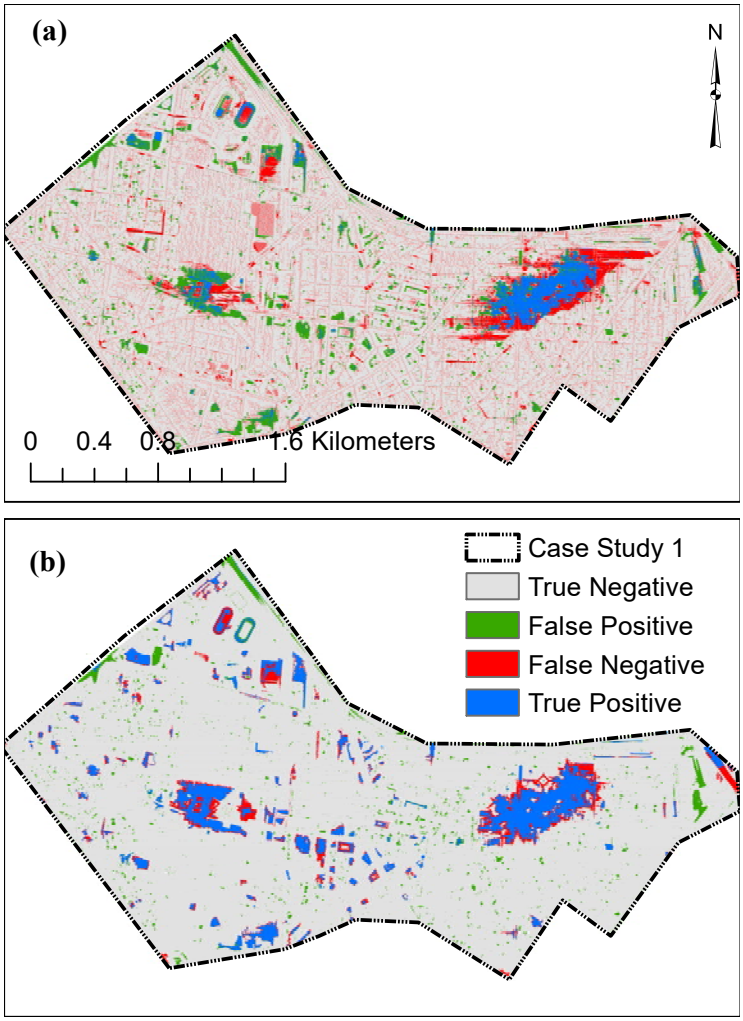


Fig. 2.14 Comparison of the inundation extent for a 100 mm precipitation event: (**a**) Comparison of TELEMAC-2D model and the TWI method inundation extents. (**b**) Comparison of TELEMAC-2D model and the FSM method inundation extents.

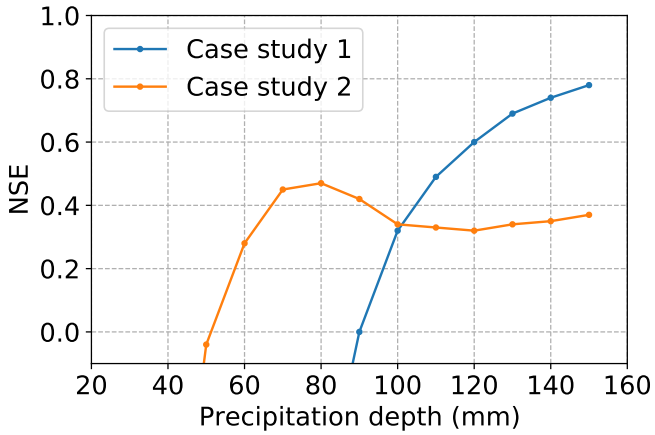


Fig. 2.15 NSE values for the simulation results of the two case studies.

2.5 Conclusions

Loss and damage caused by pluvial floods are increasing globally, yet the establishment of accurate hazard and risk maps is still impeded by high computational costs and the resource-intensive model setup of hydrodynamic models. At the same time, the availability of high-resolution DEMs is increasing, which means that limitations in terms of the availability of high-quality data are not as acute as, e.g., a decade ago. Still, we require effective and efficient techniques to capitalize on the potential of these data. In the present study, we evaluated two methods that have been suggested in the recent literature to identify urban areas that are prone to pluvial flooding, namely the Fill–Spill–Merge (FSM) method and the Topographic Wetness Index (TWI) method. As a reference for this evaluation, we used the hydrodynamic model TELEMAC-2D. As case studies for this evaluation, we selected two areas in the city of Berlin, Germany.

The TWI method has the obvious advantage that a map of TWI values can be created with minimal effort from a DEM. The TWI method, however, requires a "calibration" in the sense that a TWI threshold τ has to be estimated above which a location is considered as flood-prone (by using maximum likelihood estimation). This calibration has to be repeated for each precipitation depth of interest. To this end, a reference is required, here the simulation results of a hydrodynamic model (TELEMAC-2D). While this requirement is certainly undesirable, as it again involves the effort to set up and apply a hydrodynamic model, we found that the identified values of τ were similar for the two case study areas, so that there might be potential to transfer threshold values between regions of application. Unfortunately, though, we found serious limitations in the applicability of the TWI method, at least in a rather flat terrain such as Berlin. Using maximum likelihood estimation, we were unable to identify a local likelihood maximum and hence a value for τ for lower but still impact-relevant precipitation depths. We were only able to estimate τ across the entire range of precipitation event depths from 30 to 150 mm in the case we defined an "inundated area" as an area where the water depth exceeded 0 m. However, if we defined inundation as a water depth exceeding 0.1 m, we were only able to identify τ for precipitation depths of 90 mm and more.

Apart from this serious limitation, the FSM method clearly outperformed the TWI method when it came to detecting flood-prone areas in terms of inundation above 0.1 m. This superiority of the FSM method held for both case study regions and across all considered performance metrics (sensitivity TPR, Matthew Correlation MCC, Flood Area index FAI). The main reason for the failure of the TWI method was that inundation patterns in urban areas are dominated by depressions, while the TWI method has no "notion" of depressions as a structural element that is specifically prone to the accumulation of excess runoff.

While the FSM method was clearly superior to the TWI method, we also found that its performance tends to be higher with increasing event precipitation. For case study 2, however, we also found that the performance of the FSM method can start to decline with very high precipitation depths in case inundation starts to spread *beyond* depression structures. While the FSM method is theoretically able to point out flood-prone areas and quantify the water depth in flooded areas, this ability was only recognizable beyond minimum precipitation depths of 50 and 90 mm (case study 2 and 1, respectively).

In summary, however, our study showed that at least for the investigated case study areas, the simple and easy to apply TWI method cannot at all compete with the FSM method. Another large advantage of the FSM method is that it does not need to be calibrated to the output of a hydrodynamic model, while its application is about 600 times faster than the hydrodynamic model. Therefore, while the transferability of the FSM method appears high, it is still limited by the effort to actually set up the required model chain: this requires the DEM pre-processing and depression analysis as well as the implementation of runoff routing in a hydrological model. Prospective research should hence aim at providing more efficient and reproducible software tools so that future efforts can advance with more informative and extensive benchmarking experiments.

Funding

This research was funded by the Deutscher Akademischer Austauschdienst (DAAD). We acknowledge the support of the Deutsche Forschungsgemeinschaft and Open Access Publishing Fund of the University of Potsdam.

Supplementary Materials

The supplementary materials are available online at `https://www.https://www.mdpi.com/2073-4441/13/18/2476`

Conflicts of Interest

The authors declare no conflict of interest.

CHAPTER 3

Towards urban flood susceptibility mapping using data-driven models in Berlin, Germany

---

This chapter is published as:

Seleem, O., Ayzel, G., de Souza, A.C.T., Bronstert, A. and Heistermann, M., 2022. Towards urban flood susceptibility mapping using data-driven models in Berlin, Germany. Geomatics, Natural Hazards and Risk, 13(1), pp.1640-1662.

Abstract

Identifying urban flood-prone area is necessary but the application of hydrodynamic models is limited to small areas. Data-driven models have shown their ability for flood susceptibility mapping. A flood inventory (4333 flooded locations) and 11 flood influencing factors were used as the predictor variables to implement convolutional neural network (CNN), artificial neural network (ANN), random forest (RF) and support vector machine (SVM) to: (1) Map flood susceptibility in Berlin at 30, 10, 5, and 2 m spatial resolutions. (2) Evaluate the trained models transferability in space. (3) Estimate the most useful factors for flood susceptibility mapping. The models' performance was validated using the accuracy, F−score, Kappa, and the area under the receiver operating characteristic curve (AUC). The results indicated that all models perform very well (minimum AUC = 0.87 for the testing dataset). The RF models outperformed all other models at all spatial resolutions and the RF model at 2 m spatial resolution was superior for the present flood inventory and predictor variables. Aspect and altitude were the most influencing factors on the image-based and point-based models respectively. We conclude that data-driven models can be a reliable tool for urban flood susceptibility mapping wherever a flood inventory is available.

3.1 Introduction

Urbanization increases both life and damage losses caused by floods (Karamouz et al. 2011, Cherqui et al. 2015, Zhou et al. 2019). Floods can be classified based on their generation mechanisms into different types such as river floods, urban pluvial floods,

flash floods, and coastal floods (Kundzewicz et al. 2014). Urban pluvial floods usually occur due to inundation caused by excess runoff before it enters stormwater drainage system (Falconer & De Hemptinne 2009) or due to intense rainfall that leads to overwhelming the stormwater drainage system's capacity (Schmitt et al. 2004). They could occur anywhere subject to the occurrence of high intensity rainstorms and the existence of a critical area for runoff generation (Zhang & Pan 2014). The ubiquity of this hazard highlights the importance of accurate flood susceptibility mapping to support urban pluvial flood risk management.

Commonly, physical hydrodynamic models are used to simulate urban pluvial flooding. They can be divided into one-dimensional (1D) hydrodynamic models such as SWMM model (Barco et al. 2008), two-dimensional (2D) hydrodynamic models such as TELEMAC-2D model (Seleem et al. 2021), and 1D-2D hydrodynamic models such as MIKE URBAN (Bisht et al. 2016). These models solve the shallow water equation numerically, and are considered the best representation of the involved processes; yet, the computational costs are high. Therefore, they can only be applied to small areas using a fine spatial resolution and can't be scaled to produce flood hazard maps for large areas (Petroselli 2012).

To overcome this limitation, (Zhang & Pan 2014, Balstrøm & Crawford 2018, Samela et al. 2020) applied simplified methods based on digital elevation models (DEM), where depressions in the terrain were considered as the inundated areas. (Jalayer et al. 2014, Huang et al. 2019, Kelleher & McPhillips 2020) proposed using topographic indices as indicators of urban pluvial flooding locations. However, these methods have limitations, for example, they perform poorly with low precipitation events, consider inundation only within the topographic depressions, and some parameters need to be calibrated for every precipitation depth of interest (Seleem et al. 2021).

Point-based data-driven models such as logistic regression (Al-Juaidi et al. 2018), the statistical index (Wi) method (Shafapour Tehrany et al. 2019), random forests (RF) (Wang et al. 2015, Lee et al. 2017, Chen et al. 2020), support vector machines (SVM) (Tehrany et al. 2014, 2015), and artificial neural networks (ANN) (Bui et al. 2020) have been used as alternatives to map flood susceptibility for large areas. They can incrementally create high-level features from a raw dataset, and capture complex patterns in the dataset (Bui et al. 2020). They have demonstrated powerful performance in several areas worldwide (Tehrany et al. 2014, 2015, Zhao et al. 2018, Chen et al. 2020, Vafakhah et al. 2020, Costache et al. 2021). However, few studies have utilized such models for flood susceptibility mapping in urban areas because of the lack of inundation data and reliable flood inventories (Yang et al. 2016). Furthermore, the fundamental assumption of point-based models is that a relationship between local flood influencing factors and the local occurrence of flooding could be established (Zhao et al. 2020).

Recently, convolutional neural networks (CNN) have been used for flood suscepti-

bility mapping (Wang et al. 2020, Zhao et al. 2020, Lei et al. 2021). Zhao et al. (2020) demonstrated that convolutional neural networks (CNN) could outperform point-based models. They could learn patterns in two dimensions and understand relevant patterns around a predicted location. However, (Zhao et al. 2020, Lei et al. 2021) used a dataset at a coarse spatial resolution (30 and 25 m respectively). The coarse spatial resolution could hide essential details of the urban surface that affect flow paths and inundation patterns (Komolafe et al. 2018, Arrighi & Campo 2019). So far, the impact of spatial resolution on the results of data-driven flood susceptibility mapping has scarcely been studied (Avand et al. 2021). Moreover, (Zhao et al. 2020) considered the CNN model as a black box and did not investigate the importance of the flood influencing factors.

In previous studies, data-driven models for flood susceptibility mapping were typically evaluated in the same area used to train the models. (Zhao et al. 2021) showed that using transfer learning technique could improve the CNN model performance. The trained CNN performed poorly outside the training area, but its performance improved by adding more training data from the area outside the training area. Such techniques have not been yet investigated for RF and SVM which are considered as a benchmark for other models in the literature (Tehrany et al. 2015).

In summary, the use of data-driven models for urban pluvial flood susceptibility mapping still lacks a sufficient understanding of: the effect of spatial resolution on different model types, the transferability of models in space, and the importance of specific predictor sets. On the basis of these research gaps, this study aims to address the following research questions:

1. How do image-based models (CNN) and point-based models (RF, SVM and ANN) compare with regard to the spatial resolution (i.e., 30, 10, 5, and 2 m) of the input data?

2. How transferable (in space) are the trained models?

3. Which factors are most useful for flood susceptibility mapping?

We will investigate these questions on the basis of a unique flood inventory that is available for the city of Berlin, Germany (Berghäuser et al. 2021).

3.2 Study area

Berlin is the capital and the largest city in Germany. It has 12 administrative districts as shown in Figure 3.1. The city's population was around 3.6 million in 2020. 55 % of the city consists of built-up areas (Kottmeier et al. 2007). It is located in the northeast of Germany and has a relatively flat topography: 95 % of the city has an altitude between 30 and 60 m above sea level, and 55 % of the city has a slope angle less

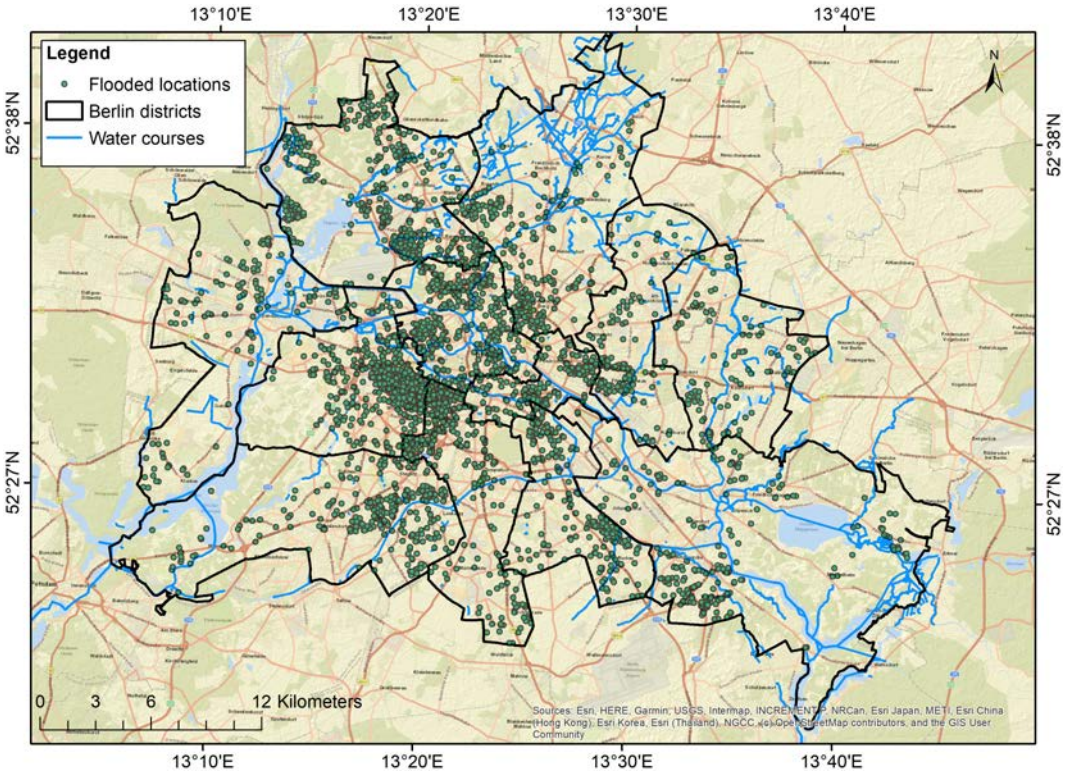


Fig. 3.1 Spatial distribution of 4333 reported flooded locations between 2005 –2017 and administrative districts in Berlin, Germany.

than 2°(as shown in Figure S1 in the supporting information). It has an oceanic climate (Köppen: Cfb) (Peel et al. 2007), with an average annual precipitation around 570 millimetres (Berghäuser et al. 2021), see also Figure S2 in the supporting information. Several pluvial urban flood events occurred in the last decades as a consequence of heavy summer precipitation. For example, the 24 hr precipitation depth for the 29th and 30th of the June 2017 event was around 170 mm and the 3 hr precipitation depth for the 2nd August 2019 event was around 45 mm (Berghäuser et al. 2021).

3.3 Data and methods

The overall approach implemented in this study is as follow: Firstly, the flood inventory and eleven factors that potentially influence flood occurrence were used to prepare the training, validation, and testing data sets for models development. Then, the models performance was compared based on selected performance indices. After that, the flood susceptibility maps were compared for a selected area within the training area and the ability of the trained models to map flood susceptibility for the whole city was evaluated. Finally, the importance of the flood influencing factors was estimated for all the implemented models. Figure S3 in the supporting information provides a graphical overview of the implemented methodology.

3.3.1 Flood inventory for Berlin

The flood inventory from Berlin Wasserbetriebe includes 4333 reported flood locations distributed all over the city as shown in Figure 3.1. These were compiled based on reports of the fire brigade, from social media and from customer reports between the years 2005 to 2017.

We selected an area within the city (170 km^2) that has a high density of flooded locations to develop the models as shown in Figure 3.2. There were 1967 reported flood locations within this area. Additionally, 1967 non-flooded locations were selected in area free of flooding. In comparison to previous studies, this is a very large and unique dataset (e.g., Termeh et al. (2018) with 53 flooded locations in an area of 5737 km^2; Choubin et al. (2019): 51 locations in 126 km^2; Zhao et al. (2020): 216 locations in 131 km^2). Flooded and non-flooded locations within the model training area were randomly split into a training set (60 %), a validation set (20 %), and a testing set (20 %) (Yacoub et al. 2003, Trost et al. 2014, Raschka 2015). The training dataset was used to fit the different models, the models' hyper-parameters were estimated based on the model performance on the validation dataset, then the testing dataset was used to evaluate the models' performance. Evaluating the models based on a testing dataset that the models have not seen before allowed us to get a less biased estimate of their ability to generalize to new data. Flooded locations outside the models' training area and an equivalent number of non-flooded locations were used to evaluate the models' transferability.

3.3.2 Flood influencing factors

According to the available data for Berlin and a literature review (Arabameri et al. 2019, Khosravi et al. 2019, Zhao et al. 2020), 11 factors were identified which potentially indicate an increased hazard for pluvial flooding. These factors represent the topographical, infrastructural, and hydrometeorological conditions: altitude, slope, topographic wetness index (TWI), curvature (Curve), distance to the river (DTRiver), distance to the road (DTRoad), distance to stormwater drainage system (DTDrainage), curve number (CN), and the frequency (FP) and magnitude (AP) of extreme precipitation events.

- Altitude is one of the most important flooding triggering factors (Tehrany et al. 2014). In general, runoff tends to accumulate at lower elevation (Zhang & Pan 2014, Seleem et al. 2021). A digital elevation model (DEM) with 1×1 m pixel size is available for the entire city of Berlin (ATKIS 2020).

- Slope affects runoff velocity and thus time available for infiltration, and also the speed for runoff concentration and hence accumulation (Rahmati et al. 2016).

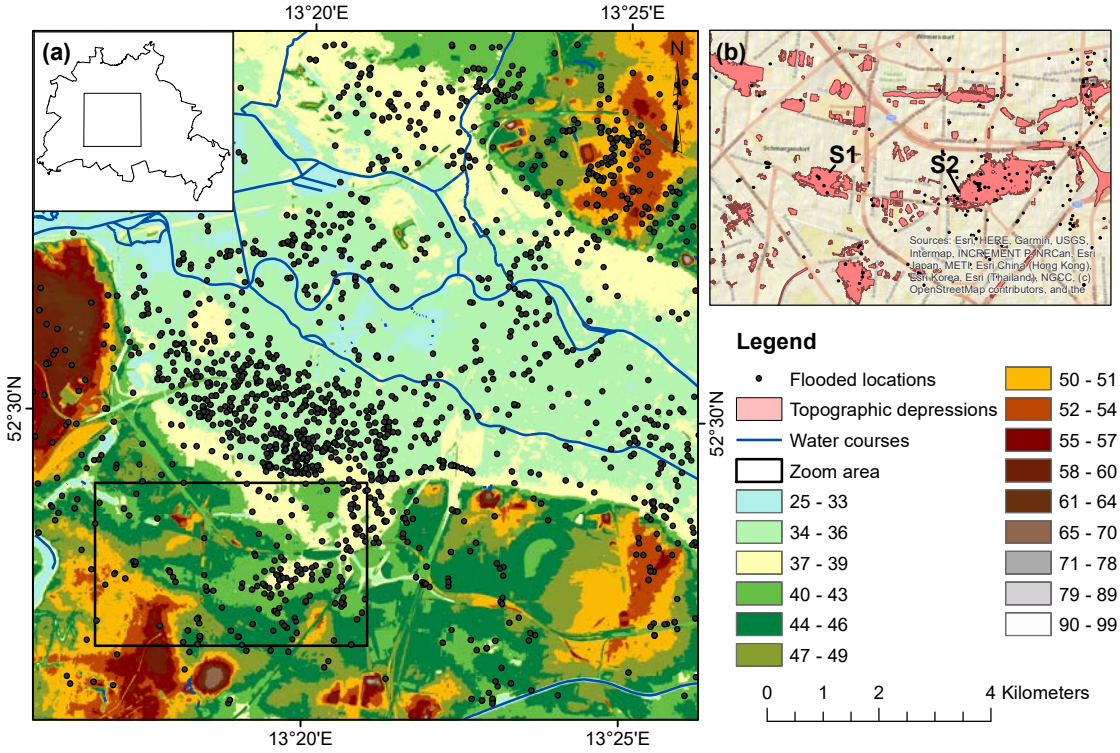


Fig. 3.2 (a) Location of the models' training area and spatial distribution of the reported flooded locations within it. (b) Zoom area (the area used to compare flood susceptibility maps at different spatial resolutions) showing the topographic depressions and the flooded locations within it.

- The TWI was originally proposed by Kirkby (1975) for hydrological modelling in mountainous and hilly terrain. It is a physical property that indicates the level of geotechnical wetness (Chapi et al. 2017) and can be used to identify flood-prone areas (Jalayer et al. 2014, Seleem et al. 2021). It is calculated as:

$$TWI = ln(a/tan\beta) \tag{3.1}$$

  Where a and β are the upslope contributing area per grid length and the slope gradient respectively. TWI map was calculated in ArcGIS.

- The DTRiver indicates the Euclidean distance between a point and the nearest river. The DTRiver is considered as one of the most important factors to map flood susceptibility (O'Neill et al. 2016). The river network was obtained from open street maps (Haklay & Weber 2008).

- Aspect indicates the direction of the maximum slope. It is considered relevant because it is directly related to the water flow direction and indicates the flat areas (Regmi et al. 2014, Jaafari et al. 2015, Shafapour Tehrany et al. 2019, Choubin et al. 2019).

- Curvature (Curve) represents the changes in slope inclination (Wilson & Gallant 2000). The curvature value indicates whether the surface is convex, concave or flat. Flood water tends to retain in concave surfaces, potentially increasing flooding susceptibility (Rejith et al. 2019).

- DTRoad indicates the Euclidean distance between a point and the nearest road. In pluvial flooding, the limited capacity of the stormwater drainage system can generate runoff that travels through the road network converting it to a preferential path for runoff (Yin et al. 2016, Singh et al. 2018). The road network was obtained from open street maps (Haklay & Weber 2008).

- DTDrainage indicates the Euclidean distance between a point and the nearest inlet to the stormwater drainage system. Pluvial flooding can occur when the capacity of the stormwater drainage system is exceeded. We downloaded the gullies' locations from (ATKIS 2020).

- CN is an empirical parameter that is used to calculate the direct runoff (Cronshey 1986). It represents the ability of the land surface to retain water, and is calculated using the land-cover and the soil type and texture. We used the CN map for Berlin generated by (Seleem et al. 2021).

- By definition, urban pluvial flooding is caused by heavy precipitation. Therefore, flood susceptibility mapping should consider both the spatial and temporal precipitation patterns (Wang et al. 2015, Zhao et al. 2018). To that end, we selected the annual maximum daily precipitation (AP) and the frequency of extreme precipitation storm (FP) (Zhao et al. 2020). AP indicates the maximum daily precipitation depth recorded between 2005 and 2017 and FP indicates the frequency of occurrence of daily precipitation depth greater than 50 mm between 2005 and 2017. Both AP and FP were calculated from radar precipitation estimates from the German Weather Service (DWD) in the period between 2005 and 2017 (Winterrath et al. 2017, Kreklow et al. 2019).

Figure 3.3 shows the spatial distribution of flood influencing factors (predictors) in Berlin.

3.3.3 Models

The following sub-sections provide a brief summary of the different model types, designs and set-ups, including hyper-parameters which were applied and which can influence both model performance and feature importance (Probst et al. 2019), and need to be set by the user.

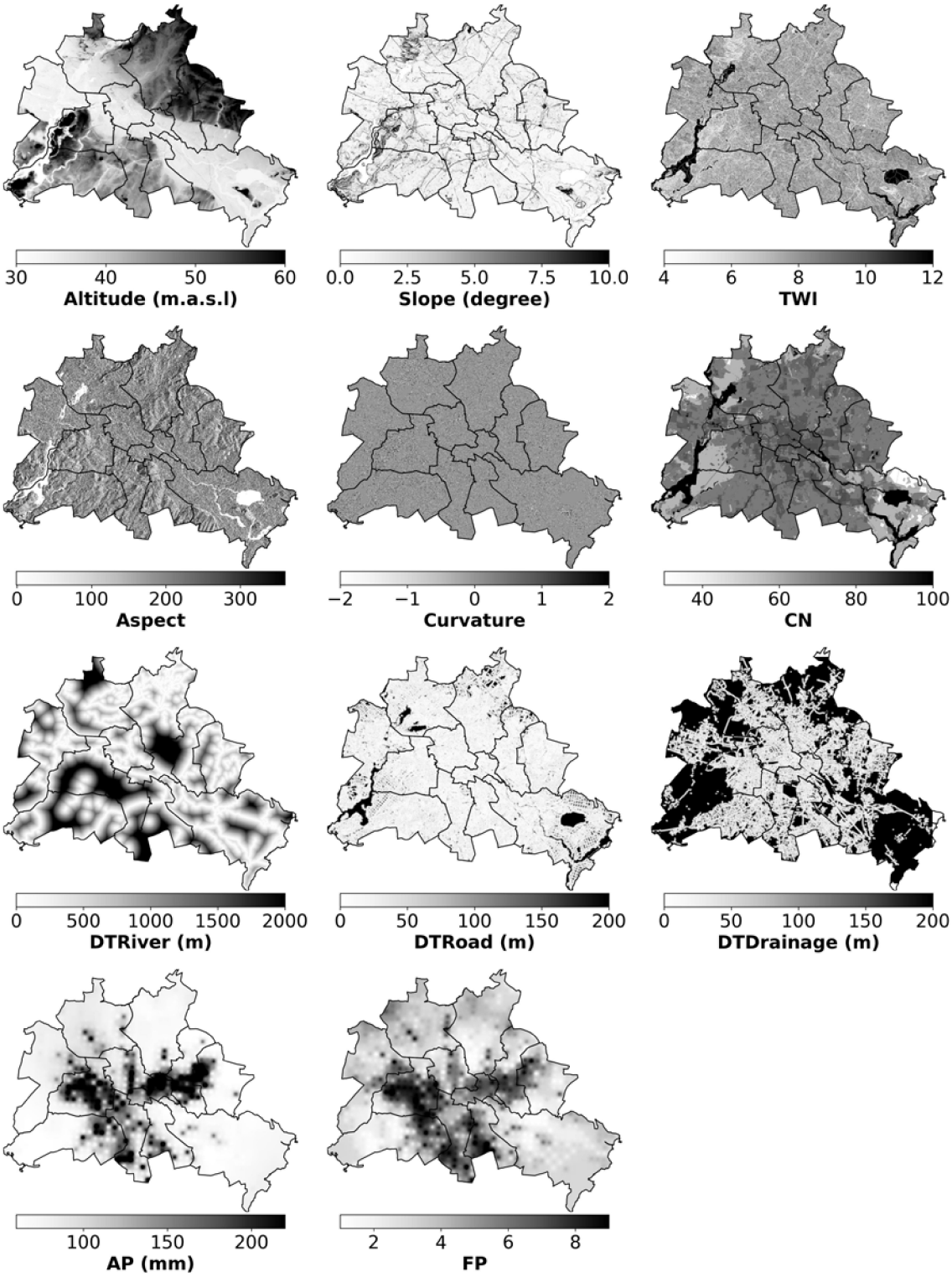


Fig. 3.3 Spatial distribution of flood influencing factors used to develop the models.

3.3.3.1 Convolutional neural networks (CNN)

The application of CNNs for flood susceptibility mapping is still rare in the literature (Zhao et al. 2020). In our study, we adopted the LeNet-5 architecture with an input image size of 23×23 pixels (Zhao et al. 2020). LeNet-5 has one input layer which is followed by two convolutional layers, each convolutional layer is followed by one pooling layer; then two fully connected neural network layers, and, a final output layer. The design is shown in Figure 3.4a. This study used the Rectified Linear Unit (ReLU) and the softmax functions as the activation and transfer functions respectively. The Adaptive moment estimation (Adam) (Kingma & Ba 2014) was used to update and optimize the weights of the CNN. A drop-out strategy with a drop rate of 0.4 was implemented to the convolutional layers and the fully connected layer to avoid overfitting. We considered the batch size and the learning rate as the hyper-parameters for the CNN models setup (Table S1 in the supporting information shows the best hyper-parameters combinations for the implemented CNN models).

3.3.3.2 Artificial neural network

This study adopted the Artificial Neural Network (ANN) architecture from Bui et al. (2020) to generate urban pluvial flood susceptibility maps. ANN includes 3 hidden layers and 192 neurons as shown in Figure 3.4b. The ReLU and the sigmoid functions were used as the activation and transfer functions respectively, while the weights of the ANN were updated and optimized using Adam (Kingma & Ba 2014). Similar to the CNN, a drop-out strategy with a drop rate of 0.4 was implemented to the hidden layers to avoid overfitting. We considered the batch size and the learning rate as the hyper-parameters for the ANN models setup (Table S2 in the supporting information shows the best hyper-parameters combinations for the implemented ANN models).

3.3.3.3 Random Forest (RF)

RF was proposed by Breiman (2001). It has been widely used for flood susceptibility mapping (Zhao et al. 2020, Lee et al. 2017, Chen et al. 2020). It implements the bootstrap technique which divides the input data to several sub-samples and develops a tree model for each sub-sample. The final result is determined based on the majority result of all the tree models. This allows RF models to avoid problems such as outliers, noise, and overfitting. We considered the number of trees in the forest, the minimum number of samples necessary to split an internal node, the minimum number of samples required to be at a leaf node, and the maximum depth of the tree as the hyper-parameters for the RF models setup (Table S3 in the supporting information shows the best hyper-parameters combinations for the implemented RF models).

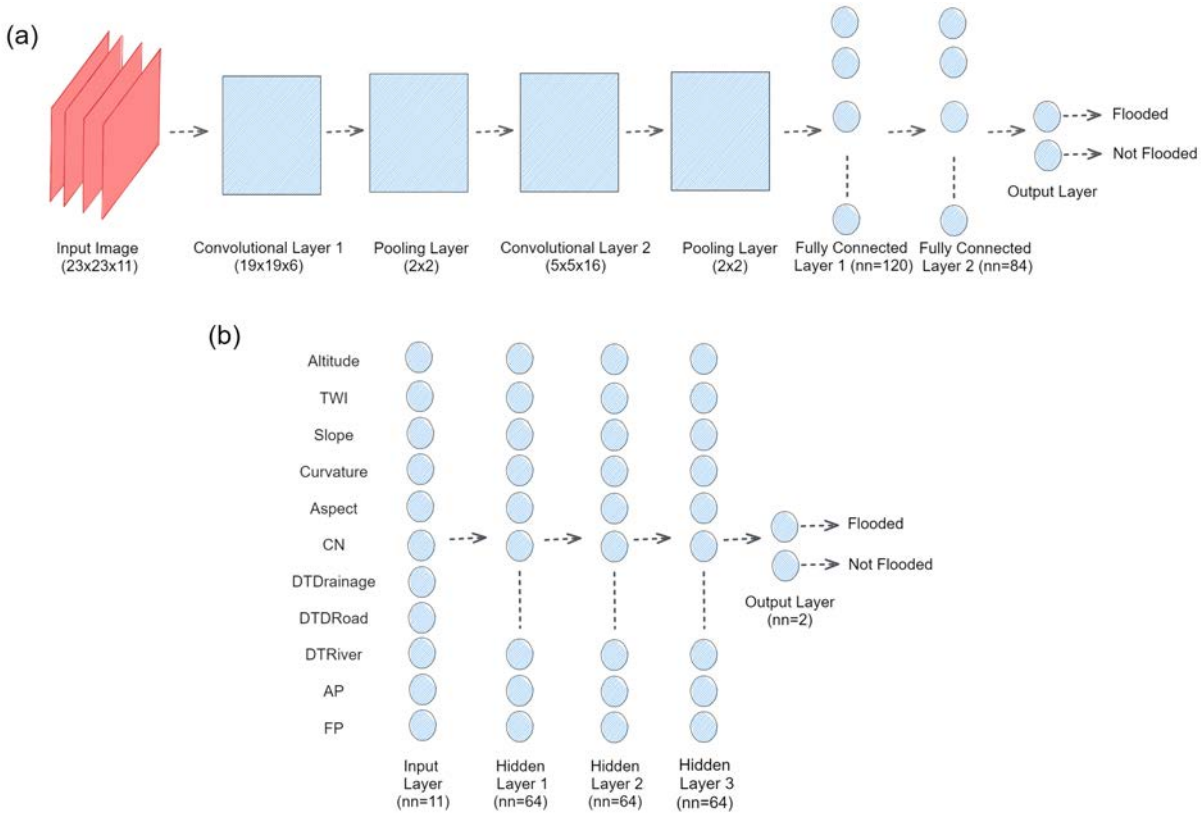


Fig. 3.4 CNN and ANN network architectures.

3.3.3.4 Support vector machine (SVM)

SVM is a machine learning technique proposed by Cortes & Vapnik (1995). It is based on the risk minimization and statistical learning theory (Tien Bui et al. 2012). It has been widely implemented in flood susceptibility mapping (Tehrany et al. 2014, 2015, Wang et al. 2020, Zhao et al. 2020). It works on finding the optimal hyperplane that separates the non-flooded and flooded classes $[0, 1]$ (Choubin et al. 2019).

This study used the RBF kernel of the SVM as it outperformed other kernels in flood susceptibility mapping (linear, sigmoid, polynomial, see Tien Bui et al. 2012, Tehrany et al. 2015, Hong et al. 2018, Wang et al. 2020). We considered the penalty coefficient and the radial basis function bandwidth as the hyper-parameters for the SVM models setup (Table S4 in the supporting information shows the best hyper-parameters combinations for the implemented SVM models).

3.3.4 Feature importance

Previous studies have not investigated the importance of individual factors to predict flood susceptibility. In this study, the SHAP (SHapley Additive exPlanations) python package (Lundberg & Lee 2017) was used to determine the feature importance for all models. SHAP assigns an importance value for each feature for each prediction (Lundberg & Lee 2017). It can be used with a wide range of models, including tree-based

models, linear models and neural networks. Compared to other techniques, SHAP does not only show feature importance, but also determines whether a feature has a positive or negative effect on the predicted values.

3.3.5 Evaluation of model performance

We used Kappa and the area under the receiver operating characteristic (ROC) curve (AUC) to evaluate model accuracy (additional indices are shown in Table S5 supporting information). Both indices have been widely used to evaluate the flood susceptibility maps in literature (Tehrany et al. 2015, Zhao et al. 2020, Bui et al. 2020, Zhao et al. 2021). Kappa is calculated as follow:

$$Kappa = \frac{p_{\mathrm{o}} - p_{\mathrm{e}}}{1 - p_{\mathrm{e}}} \tag{3.2}$$

Where observed agreement (p_0) and hypothetical probability of chance agreement (p_e) (Viera et al. 2005) can be calculated by comparing observations to model predictions. Kappa can range between -1 (less than chance agreement) and 1 (almost perfect agreement) while the AUC can range between 0 and 1, where 1 indicates a perfect model while a value of 0.5 markes the performance of a random prediction.

3.3.6 Computational details

The maps of the flood influencing factors were created using ArcGIS. The Keras python package (Chollet et al. 2015) was used to implement both the CNN and ANN models while the RF and SVM models were developed using the sklearn.ensemble.RandomForestClassifier (Pedregosa et al. 2011) and the sklearn.svm.SVC python modules (Chang & Lin 2011) respectively. K-fold cross validation was applied to quantify the model performance.

3.4 Results and discussion

3.4.1 Comparison of predicted flood susceptibility

To produce flood susceptibility maps for Berlin utilizing different data-driven models at different horizontal resolutions. We used 11 factors which we assumed to affect flood susceptibility (altitude, slope, curvature, TWI, CN, DTRoad, DTRiver, DTDrainage, FP and AP), and used 60 % of the flooded locations as training data. Flood susceptibility values range from 0 (lowest) to 1 (highest). For visualisation, the flood susceptibility was categorized into five classes using the natural breaks (Jenks) method (Jenks 1967) which is widely utilized in flood susceptibility mapping (Chapi et al. 2017, Zhao et al. 2020, Wang et al. 2020): very low, low, moderate, high and very high. Figure 3.5 shows the flood susceptibility maps using the different models and the different horizontal

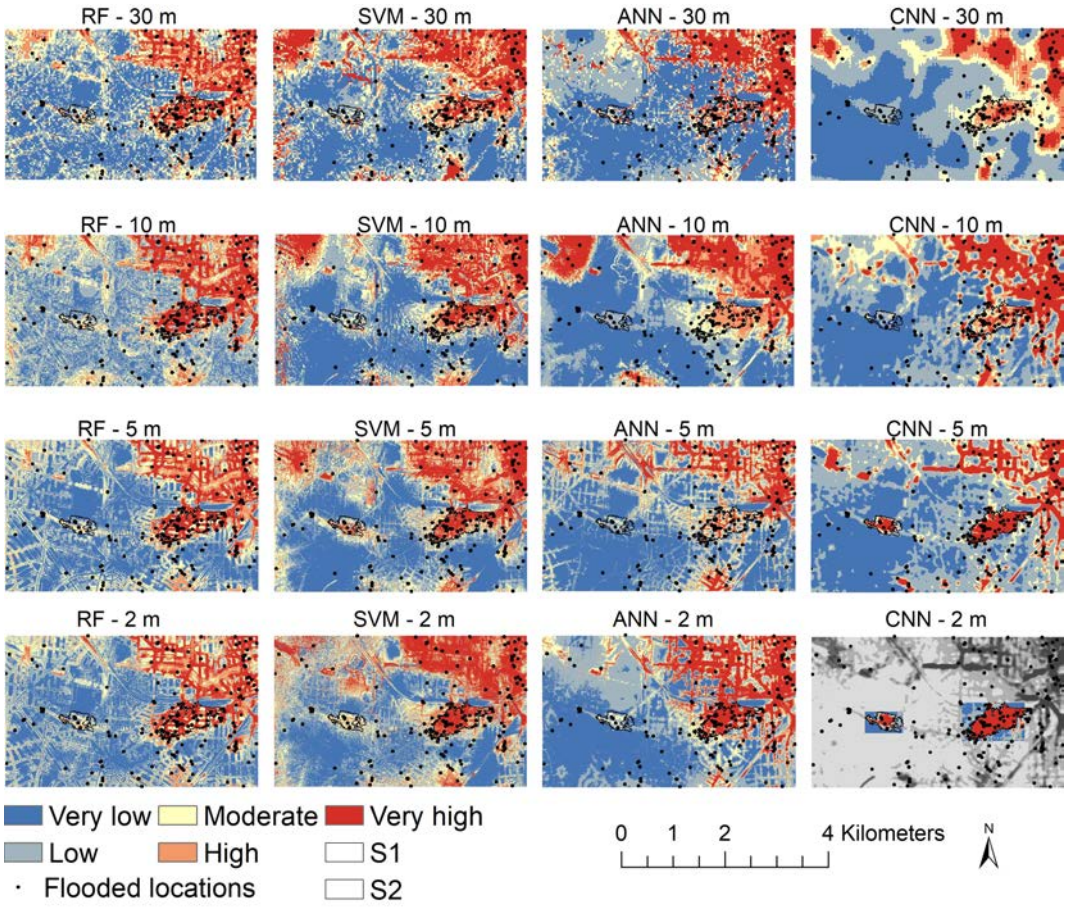


Fig. 3.5 Flood susceptibility maps from all models at different spatial resolution for the zoom area in the training area.

resolutions for the zoom area inside the training area (Figure 3.2). Visually, the majority of the flooded locations coincide with high flood susceptibility values. Due to excessive computation costs, the CNN model at 2 m resolution was not applied for the entire area. Instead, we produced a flood susceptibility map only for topographic depressions S1 and S2 (Figure 3.2b).

We selected topographic depressions S1 and S2 to evaluate the models' ability to detect topographic depressions (where normally excess runoff would accumulate). Figure 3.6 shows the predicted flood susceptibility maps for S1. The figure highlights the impact of the spatial resolution, for example, for the RF models, the impact of the road network on the identification of locations with high flood susceptibility becomes obvious only at resolutions of 5 and 2 m. Moreover, it is very interesting to see that the CNN models (image-based) could only recognize S1 as a flood-susceptible area at the 5 and 2 m spatial resolutions. Figure 3.7 shows the produced flood susceptibility maps for S2. Again, it becomes obvious that a higher resolution allows the CNN model to identify the flood-prone area more clearly.

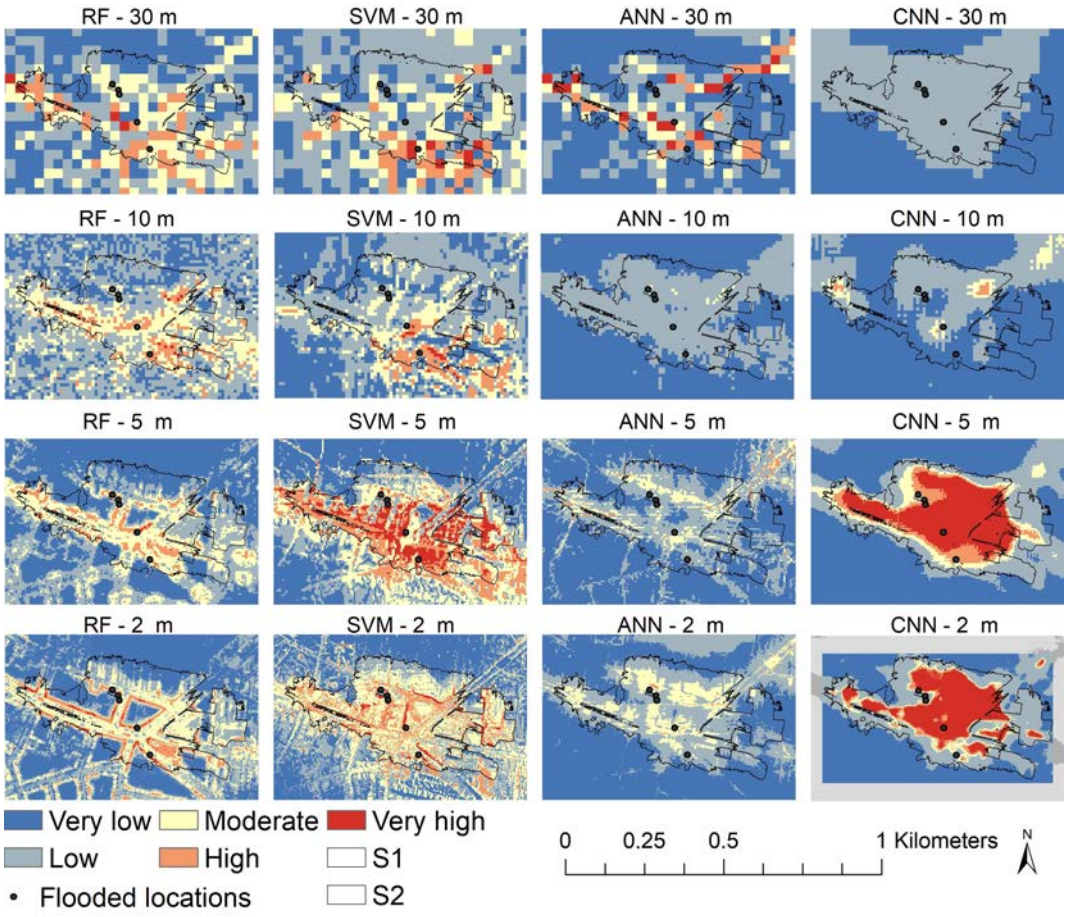


Fig. 3.6 Flood susceptibility maps from all models at different spatial resolution for topographic depression S1.

3.4.2 Model validation

The performance of the CNN, ANN, RF and SVM models with 30, 10, 5, and 2 m spatial resolutions was quantified based on the metrics AUC and Kappa (other metrics are shown in Table S5 in the supporting information). Figure 3.8 shows the results for the training dataset, testing dataset (which was reserved as 20 % of the flooded locations), and the points located outside the training area. Generally, all models achieved AUC values higher than 0.88, 0.87, and 0.8 for the three datasets respectively.

For the training data set, the RF models outperformed the other models at all spatial resolutions. All the performance indices calculated for the RF models were equal to 1 which demonstrate the models' ability to perfectly distinguish between the flooded and non-flooded locations. The CNN - 2 m model had the lowest performance indices: AUC=0.88 and kappa=0.59 which represent a moderate performance ($0.41 <$ kappa $<$ 0.60) based on the kappa evaluation criteria (Viera et al. 2005).

For the testing dataset, the RF models outperformed the other models at all spatial resolutions, too. The RF-model at 2 m resolution had the highest AUC (0.96) while

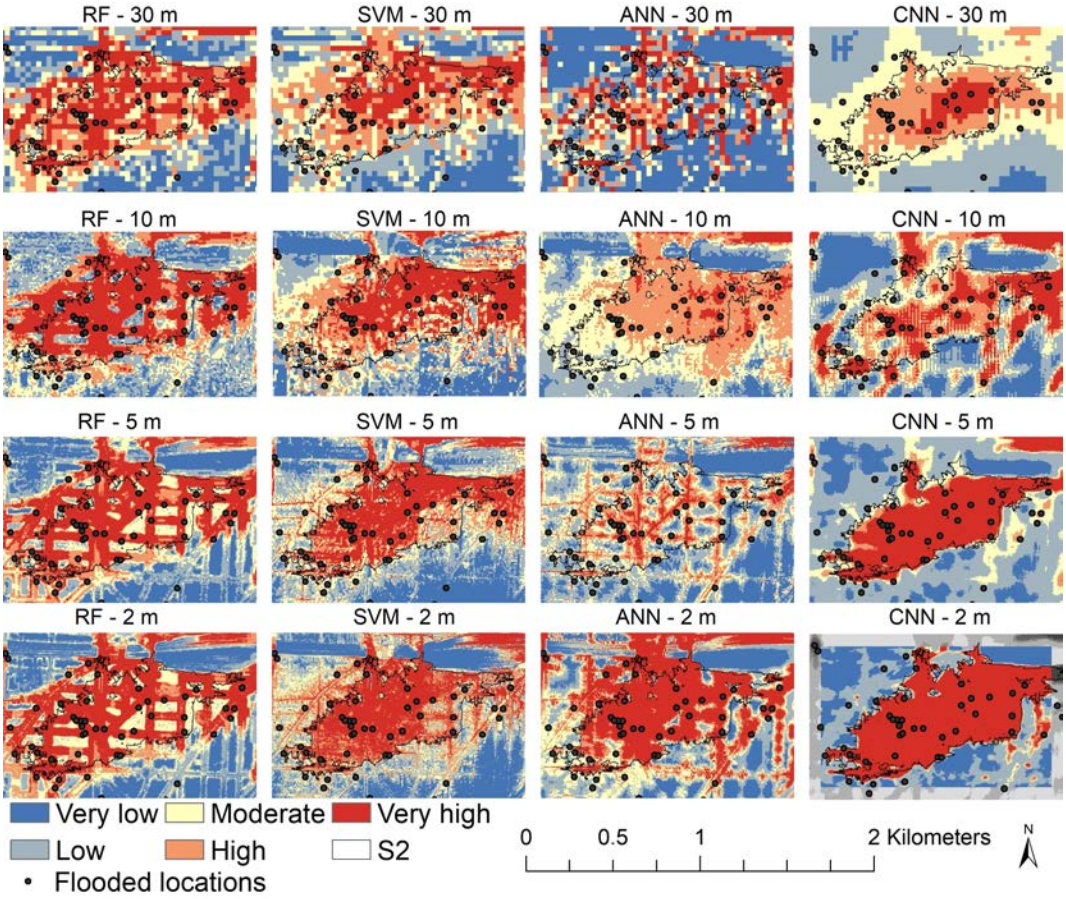


Fig. 3.7 Flood susceptibility maps from all models at different spatial resolution for topographic depression S2.

the CNN-model at 5 m resolution had the lowest AUC (0.87). According to the kappa evaluation criteria, the predictions from the RF - 2 m model (kappa = 0.79) demonstrated substantial agreement with the observations (0.61 $<$ kappa $<$ 0.80) while the predictions from CNN - 5 m model (kappa = 0.58) showed moderate agreement.

3.4.3 Flood susceptibility for Berlin

The trained models were then used to predict flood susceptibility for all of Berlin. The flooded and sampled non-flooded locations outside the training area were then used to assess the model performance outside the training area, and hence the model transferability in space. For illustration, Figure 3.9 shows flood susceptibility from the RF model at 2 m resolution (best model on training and testing data sets). Visually, areas with high flood susceptibility coincide with the flooded locations.

Figure 3.8 shows the performance indices for the points located outside the training area. The RF models were superior to the other models. While the RF - 2 m had the best performance (AUC =0.92 Kappa =0.61), ANN - 30 m had the least performance (AUC = 0.8 Kappa =0.02). Figure 3.8 and Figure S4 and S5 in the supporting file show that

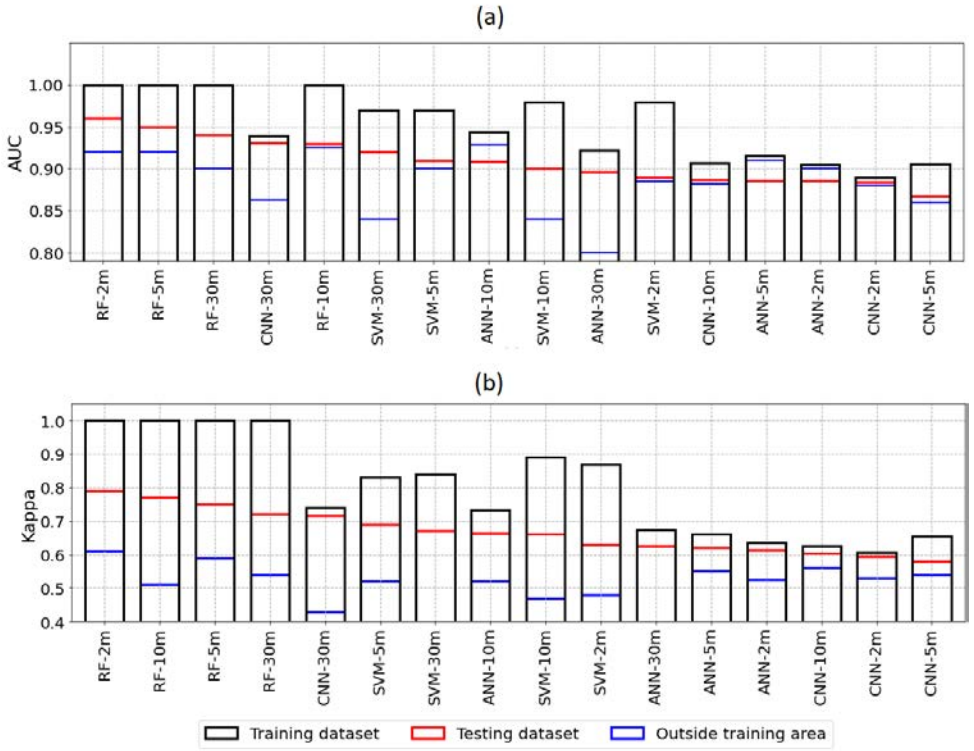


Fig. 3.8 Calculated performance indices for the training dataset, testing dataset and the locations located outside training area, models are arranged horizontally in descending order according to their score for the testing dataset.

despite the superiority of the RF models to other models, the CNN and ANN models had smaller relative loss for the testing dataset and locations outside the training area.

3.4.4 Feature importance

We used the SHAP algorithm to evaluate feature importance for all competing models. As mentioned before, the SHAP values do not only show the feature's importance but also whether a feature affects the predicted values positively or negatively. Flood susceptibility values range from 0 to 1. A balanced training dataset (the same number of flooded and non-flooded locations) was used to develop the models. Therefore, the default model prediction would be 0.5 if we did not know the input feature values at the predicted location. Then, the prediction would change from 0.5 to the final prediction based on the values of the input feature as shown in Figure 3.10. The feature effects are represented by the SHAP values in Figure 3.11 for the RF model at 2 m resolution. Figure 3.11 shows that the predictor variables affect the model prediction (flooded or non-flooded) and floods tend to occur in locations at low altitudes close to drainage system inlets and roads (RF - 2 m model).

For the image-based CNN model, the SHAP algorithm can detect which pixels increase the probability of a certain prediction (flooded or non-flooded) as shown in Fig-

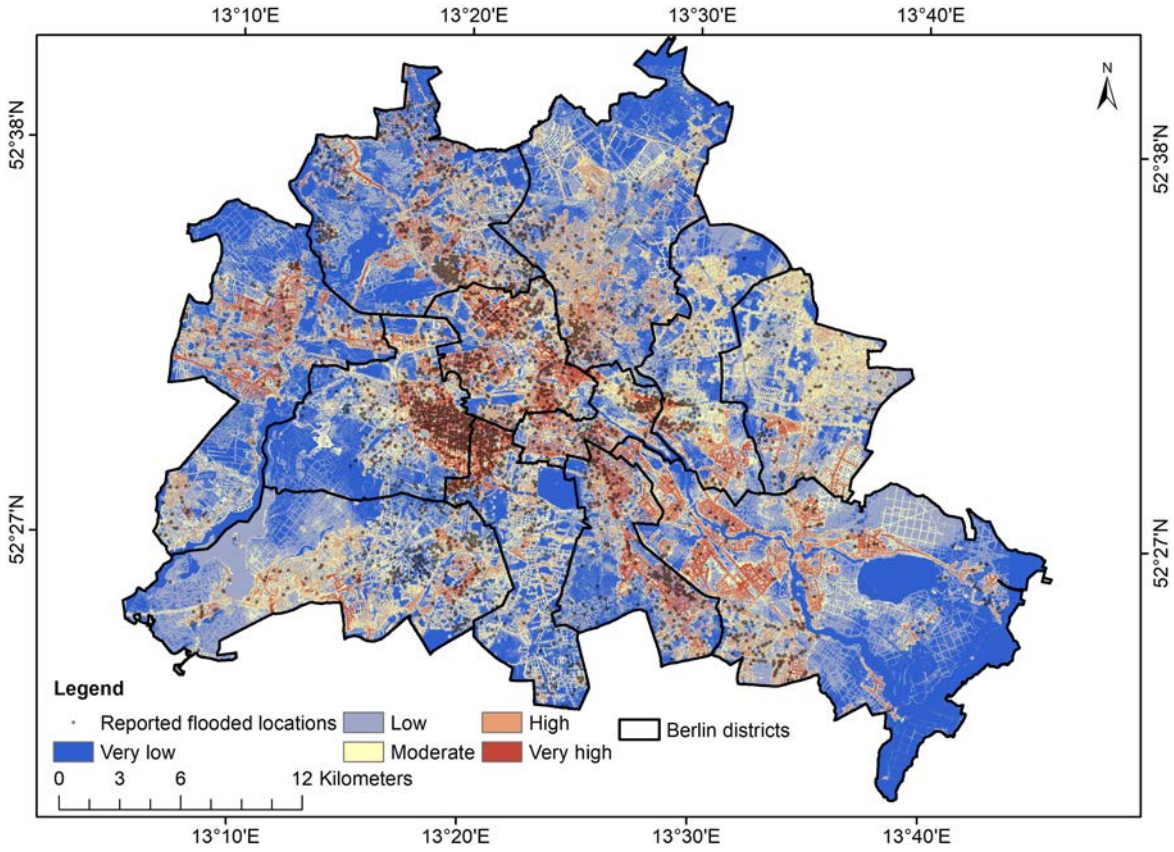


Fig. 3.9 Flood susceptibility map for Berlin from RF - 2 m model (best performance model).

ure 3.12. Figure 3.12 shows images for two locations, the top images represent a flooded location while the bottom images represent a non-flooded location. Both locations were correctly predicted by the model. Figure 3.12 shows an advantage of using the image-based CNN model as it shows the importance of considering the area surrounding a location and how they influenced the prediction.

SHAP value can be positive or negative, depending on whether the feature increase or decrease the probability of a certain class (flooded or non-flooded). The higher the absolute SHAP value, the more impact the feature has on the model prediction. Therefore, the mean absolute SHAP values could indicate the importance of each feature on the models' prediction. Figure 3.13 and Table S6 in the supporting information show the calculated mean absolute SHAP values for each feature for all the developed models. For the point-based models (RF, SVM, and ANN), the feature importance depended on the model and the horizontal resolution of the used dataset. The mean absolute SHAP value for the altitude was always significantly higher than for other features. Altitude, DTDrainage, DTRiver, AP and FP had high values while TWI, slope, curve, aspect and CN had low values which demonstrated that they had a low impact on the models' prediction. Despite the CN had no impact on the models' prediction at 30 m spatial resolution, it had impact at finer spatial resolutions. Similarly, the importance of DTRoad

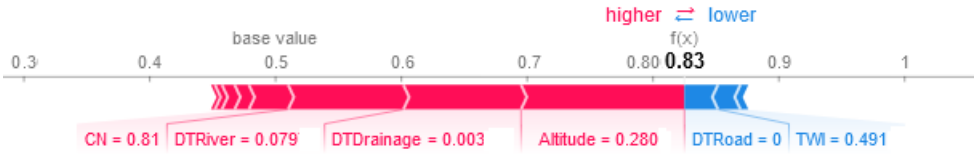


Fig. 3.10 SHAP values of each feature and their impact on the model prediction for a certain location by the RF model at 2 m resolution. The values of the input features at this location (shown here as normalized values between 0 and 1) moved the model prediction from 0.5 to 0.83 (final prediction). Features that decreased the probability of classifying the location as flooded are colored in blue while features increased the probability of classifying the location as flooded are colored in red. The features visual size show the magnitude of their impact on the prediction. For example, altitude had the largest impact on the prediction at this location.

on the RF models prediction increases with finer resolutions which demonstrates that the importance of the features could change with different spatial resolutions.

The common most important features for the CNN models were aspect, DTRiver, FP, and AP while curvature, slope and DTRoad were the common least important features as shown in Figure 3.13d. Although altitude was the most important feature for the point-based models, it is of moderate importance for the CNN models. The SHAP values are not explaining how the CNN models are working but only show which features and pixels influenced the model predictions. Aspect value equals -1 represents a flat area and 55 % of Berlin has a slope less than 2°as shown in Figure S1 in the supporting information. Therefore, we claimed that the CNN models had found a relationship between flood occurrence and the aspect as a flood predicting factor for the used flood inventory and predictor factors. The magnitude of SHAP values from the CNN models was different from other models because SHAP uses different algorithms to explain the prediction of different models (Lundberg & Lee 2017).

3.5 Conclusions

CNN, ANN, RF and SVM models were used to map flood susceptibility for the city of Berlin using 11 predictor variables at 30, 10, 5, and 2 m spatial resolution. A detailed urban flood inventory served as reference data for training, testing, and validation. The key findings are summarized in the following sections:

3.5.1 Model performance

Based on the calculated performance indices, we found that all models performed well on the training and testing data sets, while the RF model outperformed the others at all spatial resolutions, and the RF model at 2 m spatial resolution performed best. We evaluated CNN and ANN architectures that had been used in the literature (Zhao et al.

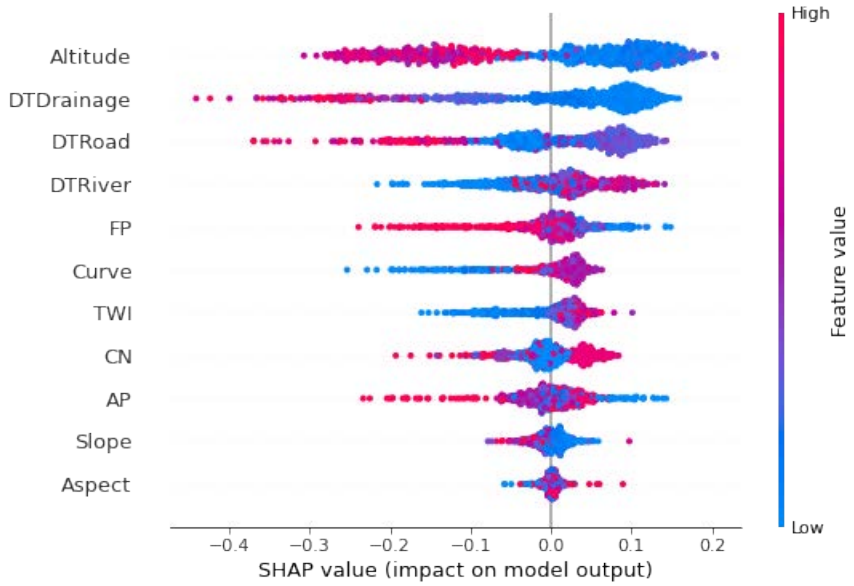


Fig. 3.11 SHAP values for the testing dataset using RF - 2 m model. The features are arranged vertically based on their importance in descending order. The horizontal axis shows the SHAP values, a positive value means that it increased the probability of classifying the location as flooded and a negative value means that it increased the probability of classifying the location as non-flooded, the colour shows whether the feature value is low or high, the SHAP values for each feature at every location are represented by dots. The dots tend to pile up along each feature row to represent the dots density.

2020, 2021, Bui et al. 2020). Both architectures performed well. However, we expect a better model performance by examining other architectures that fit with the used flood inventory and the predictor data for Berlin. In contrast to Zhao et al. (2020) findings, we found that the RF and SVM models outperformed the CNN models.

The models' ability to identify topographic depressions was evaluated using two topographic depressions (S1 and S2). Although all the models could predict S2 as a flood-susceptible area at all spatial resolutions, the performance varied for the S1. RF models at fine resolution (2 and 5 m) could recognize the streets in S1 as a flood-susceptible area, the CNN models recognized the topographic depression as a flood-susceptible area only at fine resolution (5 and 2 m). The maps showed that the models could better understand the complex urban environment using finer horizontal resolution and the 30 m coarse resolution is not recommended to be used for urban areas. The literature still lacks models which use fine resolution data sets (Zhao et al. 2020, Lei et al. 2021).

3.5.2 Model transferability in space

Model transferability in space could enable the prediction of flood susceptibility for areas outside the model's training area. It is still a new rising topic for flood mapping (Zhao et al. 2021). Our findings show that the majority of the models had a moderate

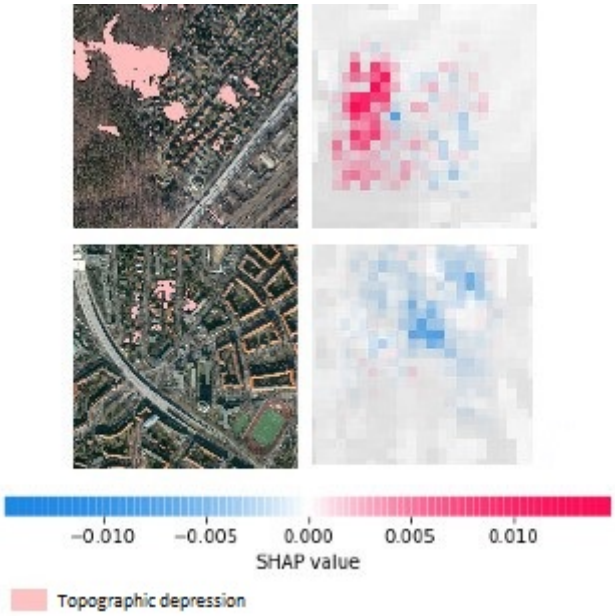


Fig. 3.12 The right figures show images of two locations for the CNN model including the topographic depression in the images and the left figures show the calculated SHAP values for the two images based on the CNN - 30 m model. The intense red coloured pixels increase the probability that the image is predicted as flooded, whereas intense blue coloured pixels increase the probability that the image is predicted as non-flooded.

performance for predictions outside the training area based on Kappa evaluation. Although the predicted maps had a relative error up to 20 % (minimum AUC = 0.8), it is a quick method to generate flood susceptibility maps for urban areas. The RF at 2 m spatial resolution outperformed all the other models. It had substantial performance-based on kappa evaluation (kappa = 0.61) and AUC = 0.92. The model performance outside the training area could be improved by adding more training data to the trained model or using transfer learning techniques (Zhao et al. 2021). Future research requires testing transferability further in environments with different characteristics (particularly with cities in more mountainous environments).

3.5.3 Feature importance

So far, data-driven models in pluvial flood susceptibility mapping were considered as black boxes. In this study, we investigated the importance of individual factors to predict pluvial flood susceptibility and explained the models' predictions based on the input features values using SHAP.

The spatial resolution affected the importance of the feature for the model prediction. SHAP values showed how each feature affected the model prediction and for the CNN models, it showed which pixels affected the model prediction. The point-based models agreed with the findings in the literature that low-lying areas located closer to the stormwater drainage system, to river banks and roads are more flood-prone. In con-

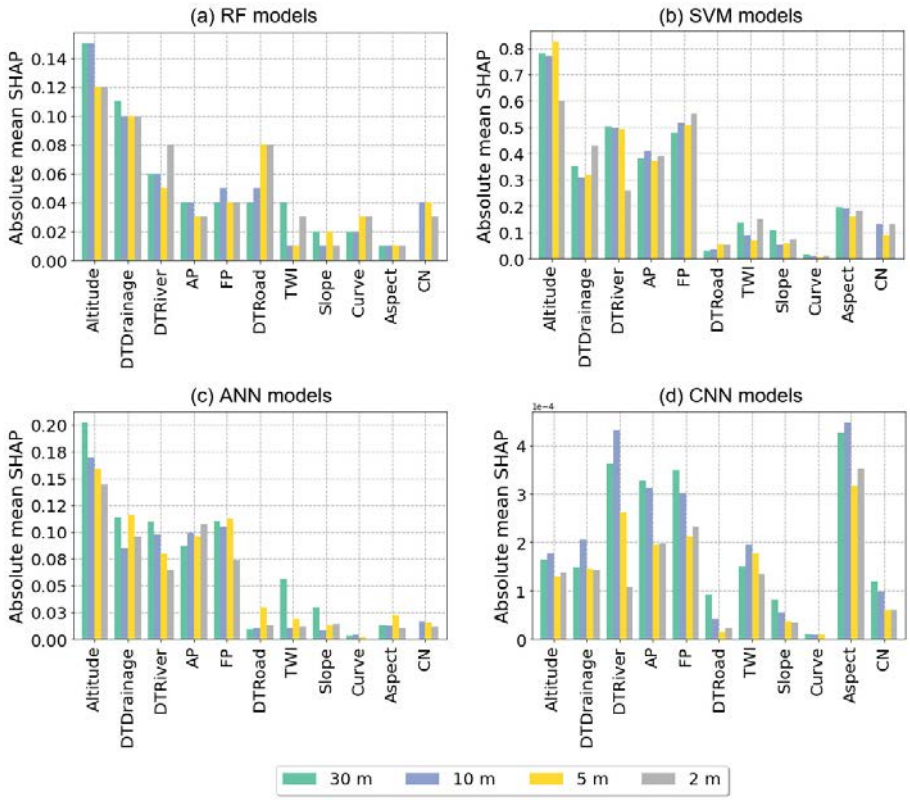


Fig. 3.13 Feature importance for all models based on the calculated absolute mean SHAP values.

trast, the CNN models identified aspect as the most important flood predicting factors in Berlin which confirm (Löwe et al. 2021) findings. We used 11 flood influencing factors that are widely used in the literature for flood susceptibility mapping. However, we found that many features are not important for the point-based model while more features affected on the CNN model predictions. Löwe et al. (2021) showed that using fewer predicting features would improve the CNN model performance. Therefore, we recommend further research to carry out a detailed feature selection analysis to consider only features that would strongly impact the model prediction.

Flood susceptibility mapping using a data-driven model is considered as an alternative for the complicated hydrodynamic simulations. However, it should be emphasized that the reported flood locations were compiled by citizens, socioeconomic factors (such as education level, age and past pluvial flooding experience) may have an impact on the data. These models also consider the topographic factors as the most influencing features on model prediction reducing the role of precipitation in causing flooding. Therefore, we recommend future research to use output of hydrodynamic simulations to develop data-driven models for urban flood management.

Overall, all the used models could map the urban flood susceptibility efficiently. The point-based model would be recommended for flood susceptibility mapping for large areas because the CNN models were both computationally and time-consuming

in terms of input data preparation especially for fine resolution which is necessary to show the urban watershed characteristics.

Data availability statement

The DEM is openly available to download for the entire city of Berlin (ATKIS 2020). The used flood inventory that supports the findings of this study is available from Berlin Wasserbetriebe. Restrictions apply to the availability of the flood inventory data, which was used under license for this study.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Supplemental online material

Supplementary data to this article can be found online at `https://doi.org/10.6084/m9.figshare.19668756.v1`

Funding

This research was funded by the Deutscher Akademischer Austauschdienst (DAAD).We acknowledge the support of the Deutsche Forschungsgemeinschaft and Open Access Publishing Fund of the University of Potsdam.

CHAPTER 4

Transferability of data-driven models to predict urban pluvial flood water depth in Berlin, Germany

---

This chapter is under review as:

Seleem, O., Ayzel, G., Bronstert, A. and Heistermann, M., 2022. Transferability of data-driven models to predict urban pluvial flood water depth in Berlin, Germany. Natural Hazards and Earth System Sciences Discussions, pp.1-23.

Abstract

Data-driven models have been recently suggested to surrogate computationally expensive hydrodynamic models to map flood hazards. However, most studies focused on developing models for the same area or the same precipitation event. It is hence not obvious how transferable the models are in space. This study evaluates the performance of a convolutional neural network (CNN) based on the U-Net architecture and the random forest (RF) algorithm to predict flood water depth, the models' transferability in space and performance improvement using transfer learning techniques. We used three study areas in Berlin to train, validate and test the models. The results showed that (1) the RF models outperformed the CNN models for predictions within the training domain, presumable at the cost of overfitting; (2) the CNN models had significantly higher potential than the RF models to generalize beyond the training domain; and (3) the CNN models could better benefit from transfer learning technique to boost their performance outside training domains than RF models.

4.1 Introduction

Urbanization increases the frequency and severity of extreme urban pluvial flood events (Skougaard Kaspersen et al. 2017). Therefore, it is crucial to quantify the flood water depth and extent due to pluvial flooding in urban environments. While 2-dimensional hydrodynamic models are effective and robust in estimating urban floodwater depth, they are difficult to scale due to prohibitive computational costs (Costabile et al. 2017). Data-driven models are raising as a surrogate might overcome the limitations of the

computationally expensive numerical models (Hou et al. 2021, Guo et al. 2021, Löwe et al. 2021, Guo et al. 2022, Bentivoglio et al. 2022). They do not simulate the physical process of runoff generation and concentration, but find patterns between the input and output data. The model's accuracy depends on the amount, quality and diversity of the available data. They could predict water depth with a sufficient level of accuracy within seconds. Consequently, they are a helpful tool that can support decision-makers with a real-time forecast.

Data-driven models used to address urban pluvial flood hazards in the literature can be grouped into models that use only rainfall input to map flood hazards (Hou et al. 2021, Hofmann & Schüttrumpf 2021), and models that account for the topographic characteristics of the urban landscape (Löwe et al. 2021, Guo et al. 2022). The former group interpolates the flood response between rainfall events that were used to train the model and hence can only predict flood hazards within the training domain while the latter has the potential to generalize and make accurate predictions outside the training domain (Bentivoglio et al. 2022).

Point-based data-driven models such as the random forest (RF) algorithm have been widely used in the literature to map susceptibility for pluvial flooding (Lee et al. 2017, Chen et al. 2020, Zhao et al. 2020, Seleem et al. 2022). RF models outperformed convolutional neural networks (CNN) to map flood susceptibility in Berlin at various spatial resolutions, and showed promising results outside the training domain (Seleem et al. 2022). (Hou et al. 2021) trained RF and K-nearest neighbour (KNN) algorithms to predicate urban pluvial flood water depth using only the rainfall characteristics as inputs, and Zahura et al. (2020) trained a RF model to predict flood water depth in an urban coastal area using three topographic predictive features. However, both studies evaluated the model performance inside the training domain only. The algorithm performance to map urban pluvial flood hazards using different topographic characteristics of the urban area and its ability to generalize to other areas than the training domain have not been systematically investigated in the literature, yet.

CNNs have recently demonstrated the potential to map urban pluvial flood susceptibility (Zhao et al. 2020, 2021, Seleem et al. 2022) and flood hazard (Löwe et al. 2021, Guo et al. 2022).They are designed to extract spatial information from the input data and to handle image (raster) data without an unwarranted growth in the model complexity. Löwe et al. (2021) trained a CNN model based on the U-Net architecture (Ronneberger et al. 2015) to predict urban pluvial flood water depth. They divided the city into a grid, used part of it for training and the rest for testing. The testing areas were close to or surrounded by training areas which guaranteed that the testing dataset had minimal diversity from the training dataset. Guo et al. (2022) used four topographic predictive features and one precipitation event to train a CNN model. The model performed well outside the training domain for the same precipitation event used to train the model.

Deep learning uses transfer learning techniques to mitigate the problem of insufficient training data (Tan et al. 2018). Zhao et al. (2021) applied transfer learning techniques to map urban pluvial flood susceptibility using the LeNet-5 network architecture. A model that was trained on a certain part of the city (pre-trained model) performed poorly outside the training domain. A transferred model trained by freezing the pre-trained model weights and allowing only a few weights to be re-trained using a few new training data from the new area improved the model performance. The transferred model used the knowledge learned from the pre-trained model and outperformed a model that was only trained for the new area. These techniques have not yet been investigated for predicting flood water depth or for shallow machine learning algorithms such as RF.

In summary, deep learning was consistently superior to shallow machine learning in literature but recent studies showed the contrary (Seleem et al. 2022, Grinsztajn et al. 2022). However, shallow machine learning algorithms have not been systematically challenged in terms of transferability for urban flood modelling. A data-driven model that generalizes outside the training domain is still a major challenge in literature (Bentivoglio et al. 2022). While previous studies tried to examine the transferability of CNN in space to predict flood water depth under certain limitations (Löwe et al. 2021, Guo et al. 2022) and use transfer learning techniques to improve the CNN performance outside the training domain to map flood susceptibility (Zhao et al. 2021), such efforts have been examined neither for RF models nor for surrogates of physical numerical 2D hydrodynamic models. It is not obvious how transfer learning techniques could improve the data-driven model performance and be a useful tool to overcome the limitations of applying computationally expensive 2D hydrodynamic models to a big region. In this study, we investigate the transferability of data-driven models to surrogate the physical numerical 2D hydrodynamic models by addressing the following research questions:

(1) How does the performance of RF and CNN models in predicting urban pluvial flood water depth compare inside and outside the training domain?

(2) Can transfer learning techniques improve the model performance outside the training domain and thus help to overcome the issue of limited training data?

4.2 Methodology

4.2.1 Study design

The overall design of this study was as follows: firstly, we selected three areas (Figure. 4.1) that have frequently been flooded in the last decades based on a flood inventory (Seleem et al. 2022) gathered between 2005 and 2017. 2D hydrodynamic simulations were carried out in these areas. Then, the precipitation depth, topographic predictive features and water depth from the 2D hydrodynamic simulations were used to prepare the train-

ing, validation and testing datasets. We randomly selected 10000 images (raster with spatial extent 256 $\times$ 256) and 10 $\%$ of the available data (number of pixels within the training domain $\times$ number of training precipitation events) to develop both the U-Net and RF models respectively. We split the data into training (60 $\%$), validation (20 $\%$) and testing (20 $\%$) datasets. The validation dataset was used to estimate the optimal hyperparameter combinations. The testing dataset included data from three precipitation events (50, 100, and 140 mm) which were not included in the training and validation datasets. Next, we defined six combinations of training and testing datasets as shown in Table 4.1, and evaluated the model performance inside each training domain and the models' spatial transferability to other testing domains, hence we evaluated the transferability between precipitation events (at the same training domains) and the transferability in space between study areas. Afterwards, we selected the best hyperparameter combinations for the data-driven model that best fit the validation dataset. Finally, we investigated whether the learned knowledge from the pre-trained models can improve urban flood hazard mapping outside the training domain using transfer learning techniques and which predictive features are mostly influencing the model predictions.

Table 4.1 Examined training data combinations to train the data-driven models.

Training domain	Testing domain	Training domain	Testing domain
SA0	SA0*, SA1, & SA2	SA0 & 1	SA0*, SA1*, & SA2
SA1	SA0, SA1*, & SA2	SA0 & 2	SA0*, SA1, & SA2*
SA2	SA0, SA1, & SA2*	SA1 & 2	SA0, SA1*, & SA2*

* refers to testing the model with precipitation events that were not included in the training dataset.

4.2.2 Study area and hydrodynamic model

Berlin is the capital of Germany and has around 3.6 million inhabitants. The city has a relatively flat topography (Seleem et al. 2022) and has an oceanic climate (Köppen: Cfb) (Peel et al. 2007). The average annual precipitation is around 570 mm (Berghäuser et al. 2021). Heavy summer precipitation caused several urban pluvial floods in the last decades, for example, the 170 mm precipitation depth event on the 29$^{\text{th}}$ and 30$^{\text{th}}$ of June 2017 (Berghäuser et al. 2021). The selected study areas are between 6, 11, and 12 km^2. Seleem et al. (2021) showed that SA0 has large deep topographic depressions where flood water tends to accumulate, while flood water spills outside the topographic depressions after a certain precipitation depth threshold in SA2.

The maximum water depths were obtained from TELEMAC-2D (Galland et al. 1991) hydrodynamic simulations (for SA0 and SA2) performed by (Seleem et al. 2021). We performed additional simulations for SA1 using the same model setup. We used the

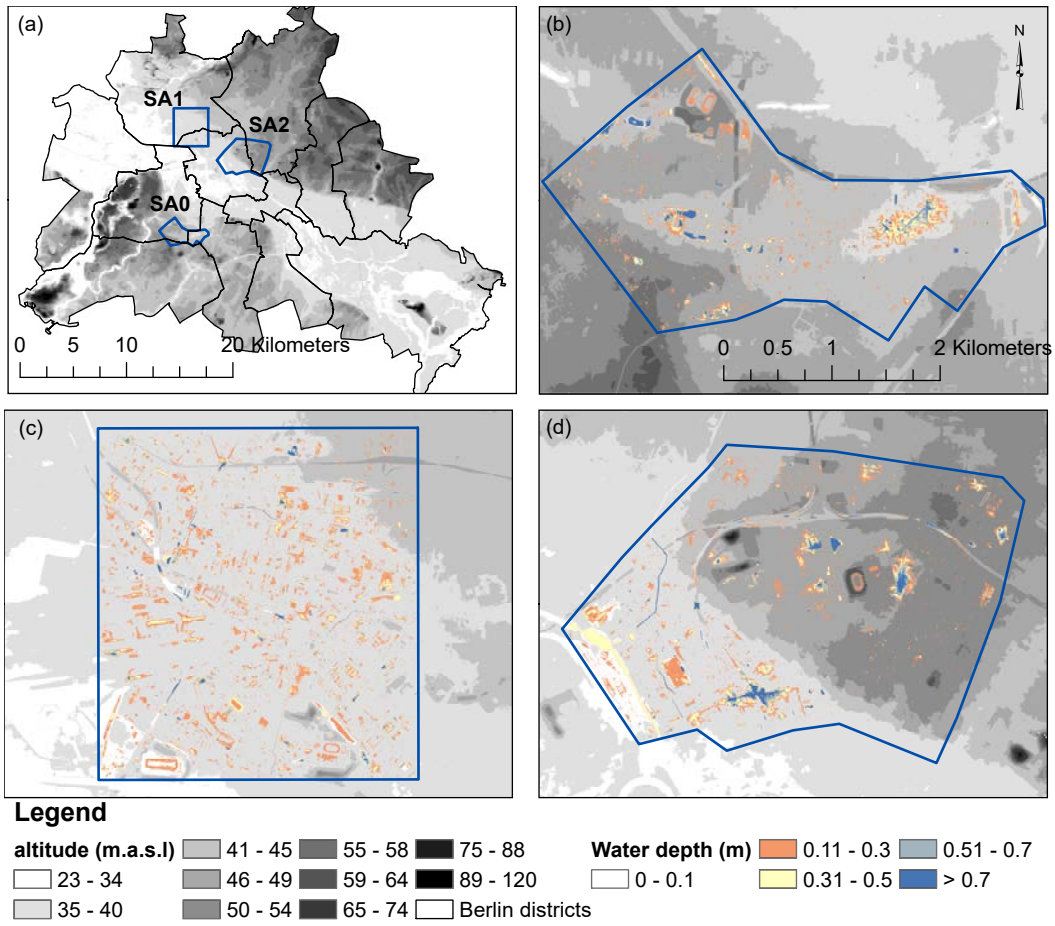


Fig. 4.1 (a) The three study areas in Berlin and the altitude map in the background. (b), (c) and (d) show the water depth map from TELEMAC-2D simulation for a one-hour block rainfall 100 mm precipitation event for SA0, SA1, and SA2 respectively and the altitude map in the background.

finite volume scheme to solve the shallow water equations over non-structured triangular grids (1 m maximum side length). The simulations were carried out using one-hour duration precipitation events (block rainfall) with precipitation depths ranging from 20 to 150 mm (10 mm increments), the one-hour intensive precipitation event in 2019 caused pluvial flooding (Berghäuser et al. 2021). We used the SCS-CN method (Cronshey 1986) to estimate excess runoff. The storm drainage system was not included in the TELEMAC-2D simulations due to the unavailability of detailed data of the storm drainage system. Additionally, the city of Berlin has a relatively flat topography and van Dijk et al. (2014) showed that there was no significant difference between the results of 2D and coupled 1D-2D hydrodynamic models in urban areas with flat terrain. For more information about the model setup, please see (Seleem et al. 2021).

4.2.3 Predictive features

While data-driven models do not ”understand” the physical processes of runoff generation and concentration, they are designed to detect relationships between input and target variables (Grant & Wischik 2020), in this case simulated inundation depth. Therefore, predictive features should represent the surface characteristics of the study area which could inform the model of governing hydrological and hydrodynamic patterns. Table 4.2 shows the selected 12 predictive features that we considered potentially relevant for mapping urban floods and their description. The topographic predictive features were generated from a digital elevation model (DEM) with a 1 x 1 m pixel size which is openly available to download for the entire city of Berlin (ATKIS 2020).

Table 4.2 Spatial predictive features used to train the data-driven models.

Predictive feature	Data adjustment	Description
Altitude	Normalized to [0,1]	Surface elevation is important for flood hazard mapping because runoff tends to accumulate at low elevation (Zhao et al. 2020, Seleem et al. 2021, Löwe et al. 2021, Seleem et al. 2022).
Slope	Normalized to [0,1]	Slope impacts the runoff velocity and the available time for infiltration (Rahmati et al. 2016).
Aspect	Scaled to [-1,1]	Aspect indicates the flow direction. We used the cosine and sine of aspect as two separate predictive features to deal with the cyclic behaviour of flow direction (Löwe et al. 2021). (Löwe et al. 2021, Seleem et al. 2022) found that aspect was the most important predictive feature for mapping urban floods using CNNs.
TWI	Normalized to [0,1]	Topographic wetness index was proposed by (Kirkby 1975). It indicates the geotechnical wetness level and is being used to identify urban flood-prone areas (Jalayer et al. 2014, Seleem et al. 2021).
Curvature	Normalized to [-1,1]	Depending on the curvature value, the surface is flat, concave or convex. (Guo et al. 2021, Löwe et al. 2021) used it to predict urban flooding using data-driven models.
SDepth	Normalized to [0,1]	Depth of topographic depression impacts the volume of excess runoff that can be accumulated in it (Zhang & Pan 2014, Seleem et al. 2021, 2022, Löwe et al. 2021).
FLACC	Normalized to [0,1]	Flow accumulation indicates the number of pixels draining into a certain pixel. We used the upper cutoff at 250 ha because very large values represent natural streams (Löwe et al. 2021).

TPI	Normalized to [-1,1]	Topographic position index is defined as the difference between the pixel elevation and the mean elevation of the surrounding pixels (Lei et al. 2021). A positive value denotes that the pixel is higher than the neighbouring pixels while a negative value indicates that the pixel is lower that the neighbouring pixels and a zero value represents flat areas (Weiss 2001).
CN	Normalized to [0,1]	Curve number is an empirical parameter that is computed using land-cover and soil hydrologic group (Cronshey 1986). It is used to estimate the direct runoff. We used the CN map produced by (Seleem et al. 2021).
Roughness	Normalized to [0,1]	Roughness impacts the excess runoff flow over the surface. We used the Manning roughness coefficient map produced by (Seleem et al. 2021). Buildings were defined by a high roughness coefficient similar to the TELEMAC - 2D model setup (Seleem et al. 2021).
DEML	Normalized to [0,1]	It is computed as the difference between the elevation of a pixel and the focal mean of elevation within 100 m radius. Urban pluvial floods occur on a small spatial scale ($<$ 1 km) and are connected to the local variation in elevation (Löwe et al. 2021).
Precipitation depth	Normalized to [0,1]	We used one-hour duration precipitation events with precipitation depths ranging from 20 to 150 mm (10 mm increments) (Seleem et al. 2021).

4.2.4 Models

4.2.4.1 U-Net

The application of CNNs for mapping urban flood hazards is still rare in the literature (Löwe et al. 2021). This study adopted the U-Net architecture (Ronneberger et al. 2015) as shown in Figure 4.2. The U-Net architecture showed a good performance to predict water depth in the literature (Löwe et al. 2021, Guo et al. 2022). The model input is a

terrain raster with 13 image channels (13 channels represent the predictive features) and the output is the resulting water depth at the surface. The U-Net architecture belongs to encoder/decoder architectures. The encoder follows the typical architecture of a convolutional neural network and uses pooling to downscale the spatial resolution, while the decoder uses upsampling to upscale the learned patterns. Skip connections concatenate the output of each encoder layer to its corresponding decoding layer to provide the spatial information (Srivastava et al. 2015).

We applied LeakyReLU with an activation threshold of 0.2 to all layers except the output layer (Maas et al. 2013, Löwe et al. 2021, Guo et al. 2022) and adaptive moment estimation (Adam; Kingma & Ba 2014) to update and optimize the network weights. We used average pooling because it showed better performance than maximum pooling (Löwe et al. 2021), and added a batch normalization layer after each convolutional layer to stabilize and speed up the training process (Ioffe & Szegedy 2015, Santurkar et al. 2018). A drop-out strategy was implemented with a rate of 0.5 to the convolutional layers (Löwe et al. 2021, Seleem et al. 2022), and early stopping to prevent overfitting (Prechelt 1998). We used a batch size of 10 and the mean squared loss as a loss function to train the models (Löwe et al. 2021).

The success of CNN relies on finding a suitable architecture that fits a given task (Miikkulainen et al. 2019). Therefore, we varied three parameters similar to (Löwe et al. 2021) to obtain the most suitable network architecture, namely the network depth (i.e. number of encoding and decoding blocks) (varied between 3 and 4), number of filters in the first convolutional layer (varied between 16, 32 and 64) and the size of the kernels in the convolutional layers (varied between 3, 5, and 7). Using a deeper network and more filters increases the number of parameters and the computational expense. Moreover, using a larger kernel size allows the network to perform spatial aggregation on a larger region, again, however, at increasing computational cost. All the implemented models were validated based on the holdout validation method. Löwe et al. (2021) showed that a model trained using the holdout validations method was superior to models trained using the k-fold cross-validation method to predict urban floodwater depth.

We implemented an input image size of 256 $\times$ 256 pixels (1 $\times$ 1 m spatial resolution). Löwe et al. (2021) used the same image size but with a 5 m spatial resolution. We understand that this image size may be not sufficient to fully capture urban watersheds or topographic depressions. Then again, the selected study areas are small (area ranges from 6 to 12 km^2). We also used 12 predictive features to guarantee that the input data are well representing both the terrain and hydrological characteristics. The predictive features were calculated for the whole city and hence the calculated rasters consider the characteristics of the upstream urban catchment. Finally, training models with larger images is also limited by the memory of the graphic card.

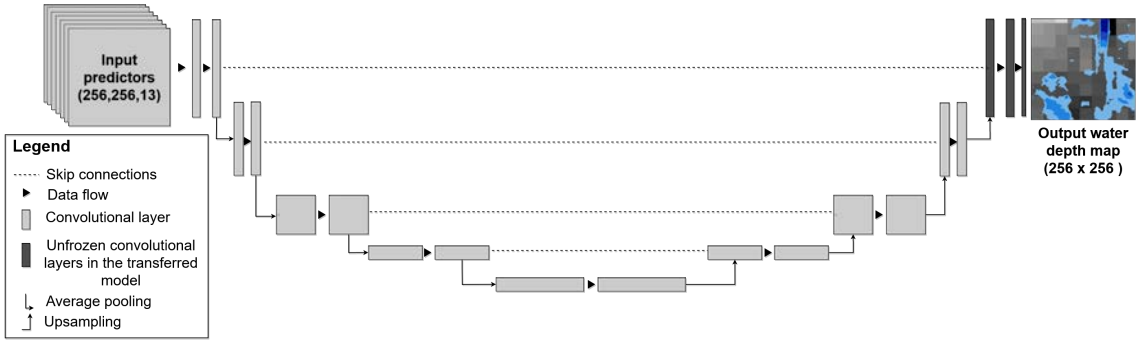


Fig. 4.2 Schematic diagram of the applied U-Net architecture for a network of depth = 4 (4 blocks of encoder and decoder). The transferred model obtained the weights from the pre-trained model except for the weights in the last decoder block (black colour). Then, the new training data was used to train the remaining untrained weights.

4.2.4.2 Random forest

The random forest (RF) is a decision tree algorithm that was proposed by (Breiman 2001). It solves both classification and regression problems by combining several randomized decision trees and averaging their predictions. RF divides the training data into several sub-datasets. Then, a tree model is developed for each dataset. Finally, a prediction is determined based on the majority result of the decision trees as shown in Figure 4.3. This approach intends to prevent overfitting (Biau & Scornet 2016).

It is well known that RF performs relatively well with default hyper-parameter values. Still, hyperparameter tuning may improve model performance (Probst et al. 2019). This study used the default values for the hyperparameters such as the minimum number of samples in a node and the maximum depth of each tree in the sklearn.ensemble.RandomForestRegres (Pedregosa et al. 2011), and varied the number of trees in the forest (between 10, 100, 200 and 300) (Zahura et al. 2020). Finally, an increasing number of training data points increases the training time and the model size dramatically. We used 10 $\%$ of the available training data (number of pixels within the training domain $\times$ number of training precipitation events) to train the RF model for all the simulations carried out in the study. We also tried to use larger portions of the training data, but without a significant improvement in model performance. In addition, we performed hyperparameter tuning using the k-fold cross-validation method using a smaller training dataset (number of samples = 100,000) to investigate the models' performance and their transferability.

4.2.5 Transfer learning

The transfer learning technique is a vital tool in deep learning to overcome the problem of insufficient training data (Tan et al. 2018). It is based on the idea that a model is firstly trained for a certain task (called the pre-trained model). Then, a new model is implemented (the transferred model) where some of its layers are frozen (they use

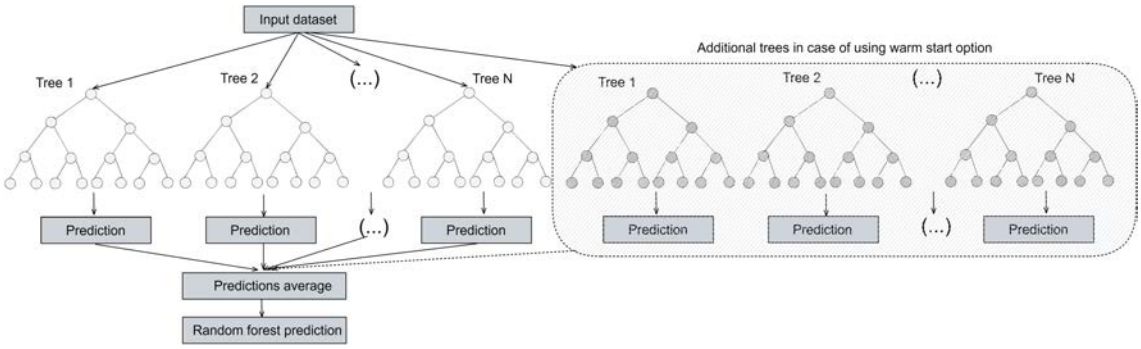


Fig. 4.3 Schematic diagram of the random forest algorithm and the additional trees that are added to the model in case of a warm start. The additional trees are trained using the new training data while the old trees (from the pre-trained model) remain unchanged.

the same weights from the pre-trained model) and the remaining layers (weights) are trained using new training data and/or a new task. This technique hence extends the application of data-driven models outside the training domain of the pre-trained model. It also reduces the training time because of the reuse of the weights from the pre-trained model. In this study, we froze all the layers in the U-Net model except the layers in the last decoding block which were then re-trained using new training data (see Figure 4.2) (Adiba et al. 2019).

The majority of shallow machine learning algorithms do not support transfer learning techniques because training the model is always fast and not complicated. However, RF offers the warm start option which allows adding more trees to the forest to be fitted using a new training dataset which means a model can be trained (pre-trained model) then new trees can be added to the forest and trained using the new training data (transferred model) without changing the trees in the pre-trained model as shown in Figure 4.3.

4.2.6 Performance evaluation

The models' performance was assessed based on predicting water depth and inundation extent. For computing the performance indices, we compared the water depth and extent obtained from the TELEMAC-2D model to the results of the competing data-driven models. Table 4.3 gives an overview of performance metrics. We computed other indices like balanced accuracy, mean absolute error and the total flooded area ratio. However, we found that root mean square error (RMSE), Nash Sutcliffe efficiency (NSE) and critical success index (CSI) are well representing the model performance. A 10 cm threshold was applied for the CSI calculation.

Table 4.3 Performance indices used to evaluate the models' predictions. The y_i and $\hat{y}_i$ denote the water depth from the TELEMAC-2D model and the data-driven model respectively. $\overline{\hat{y}}_i$ is the average of water depths from the data-driven model. Hits, misses and false alarms are estimated by the contingency table.

Index	Equation	Range	Description
RMSE	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}$	[0, ∞]	Root mean square error measures the difference between the predicted and observed values. The optimal RMSE is zero.
NSE	$1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}\left(y_i-\overline{\hat{y}}_i\right)^2}$	[-∞, 1]	Nash Sutcliffe efficiency shows how well the observed values are predicted by the model (Nash & Sutcliffe 1970). The optimal NSE value is one.
CSI	$\frac{hits}{hits+misses+falsealarms}$	[0, 1]	Critical success index is a binary index calculated based on pixel basis. The optimal value is one.

4.2.7 Predictive feature importance

We adopted the forward selection process from Löwe et al. (2021) to estimate the most important topographical predictive features for the U-Net model. Firstly, we trained 11 models, each of which considered one of the 11 topographical predictive features (precipitation depth was included in all models) from Table 4.2. Then, we evaluated the model performance based on the performance indices in Table 4.3 and selected the best model. After that, we trained 10 new models based on the best model from the previous step by adding one of the remaining 10 predictive features to the inputs. We repeated this procedure three times to get the three most important predictive features for the U-Net model.

One of the advantages of the RF algorithm is the ability to compute the importance of predictive features, hence no forward selection process was required to estimate the importance of specific features for the RF models. We used the built-in feature importance in the RF model, which is implemented in scikit-learn Python package (Pedregosa et al. 2011). The importance of the predictive features is calculated as the mean and standard deviation of accumulation of the impurity decrease within each tree (Pedregosa et al. 2011).

4.2.8 Computational details

The U-Net models were implemented using the Keras Python package (Chollet et al. 2015) while the RF models were implemented using the method ensemble.RandomForestRegressor from the Python package scikit-learn (Pedregosa et al. 2011). The U-Net models were trained using a high-performance machine with NVIDIA Quadro P4000 GPU while RF

models were trained using a machine with Intel(R) Xeon(R) CPU E5-2667 v3@ 3.20 GHz. The training time ranged from 20 minutes to 48 hours and from 10 minutes to 2 hours for the U-Net and RF models respectively. The U-net models needed around 20 seconds to predict and map the water depth while the RF models took around 3 minutes.

4.3 Results and discussion

4.3.1 Evaluating different combinations of training data

In order to evaluate model transferability between spatial domains, we used a U-Net model with the following combination of hyperparameters: depth = 4, kernel size = 3, number of filters in the first encoding block = 32. This combination showed reasonable performance with the training datasets and had performed well in previous studies (Guo et al. 2021, Löwe et al. 2021). For the RF model, we used the holdout validation method with a number of trees in the forest = 10 as it shows also reasonable results and training time (around 10 minutes).

Figure 4.4 compares the performance indices for each study area (SA) and for all combinations of training/testing datasets, for both the U-Net and RF models. The NSE values show that the RF models outperformed the U-Net models for predicting water depth within the training domains; however, they failed to predict water depth outside the training domains. It is obvious from Figure 4.4 that the RF models were overfitted to the training data while the U-Net models tended to generalise better. The CSI and RMSE values are in line with that finding: they show that the RF models could predict the inundation extent better than the U-Net models in some training combinations despite failing to predict the water depth outside the training domain accurately. Allowing the decision tree to have unlimited maximum depth may cause overfitting, so we performed multiple simulations varying it (as shown in the supporting information). The simulations showed that reducing the maximum depth of the decision tree improved the model performance outside the training domain at the cost of lower performance inside the training domain. We also trained RF models using the K-fold validation method. The results indicated that the models were not able to generalise outside the training domain as demonstrated in the supporting information. Finally, it is clear from Figure 4.4 that the models U-Net - SA1 and RF - SA1 performed best outside the training domain, compared to models trained using training data from the SA0 and SA2 separately. The U-Net-SA1 & 2 model had the best performance within and outside the training domain.

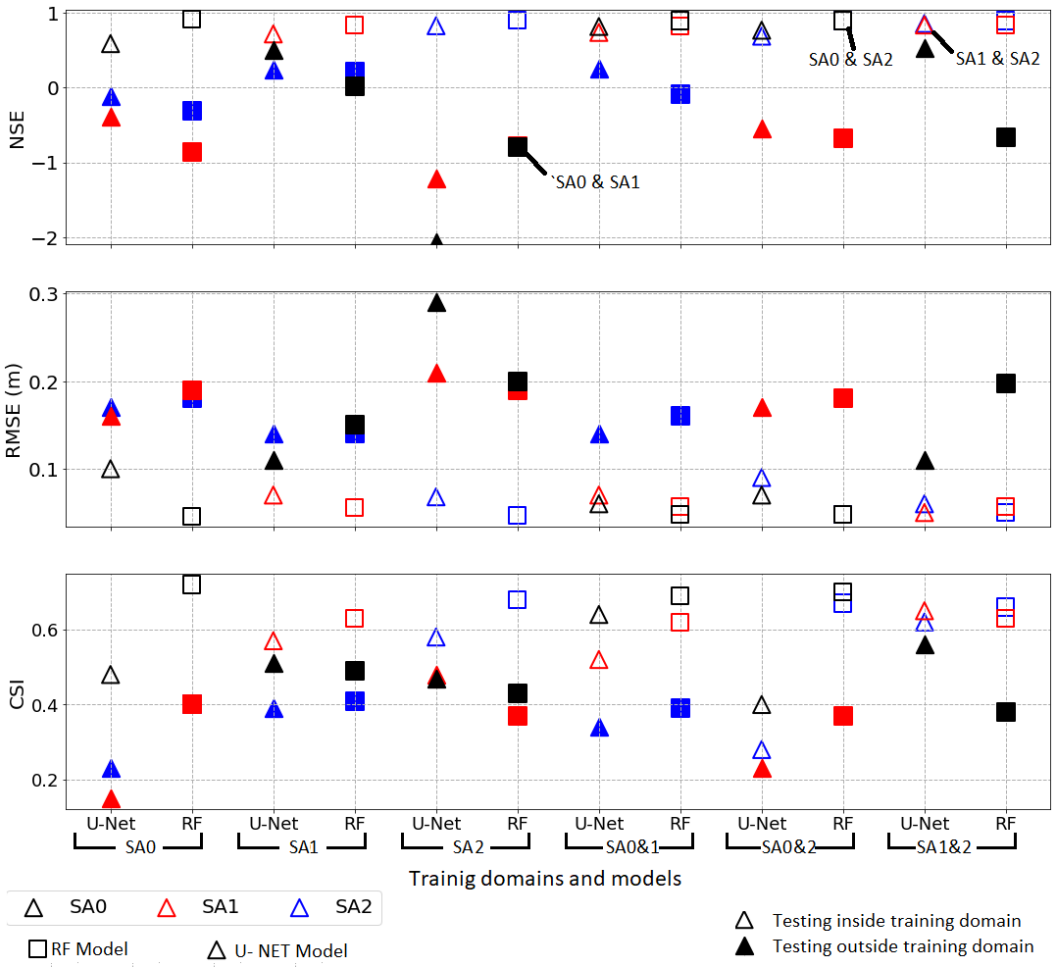


Fig. 4.4 Computed performance indices (based on the testing dataset) for different combinations of training datasets for both the U-Net and RF models. The X-axis shows the used model and the training domain while the Y-axis shows the perfromacne indices. The U-Net-SA1 & 2 model had the best performance within and outside the training domain.

4.3.2 Transfer learning

We evaluated how transfer learning could improve model performance outside the original training domain. Specifically, we investigated how the improvement from transfer learning depends on the percentage of data that was used from the target domain of the transfer. Figure 4.5 compares the performance of the transferred U-Net and RF models to the models trained exclusively on the target domain of the transfer. The figure shows that the transfer learning technique boosted the U-Net and RF model performance outside the training domain of the pre-trained models. Another advantage for transfer learning for U-Net models is that training of the transferred models (20 minutes to two hours) was faster than training the whole network from scratch.

All U-Net models transferred to the SA0 domain outperformed the U-Net-SA0 model for all performance indices. This applies even if we used only 10 % of the

available training data (from SA0) for transfer learning (in contrast to using 100 % of the SA0 training data for training the U-Net-SA0 model). We could conclude from Figure 4.5 that the transferred model could use the previously learned knowledge from the U-Net-SA1&2 model to predict water depth in SA0. Contrary to U-Net, the trained RF models for each SA separately outperformed all the transferred RF models. All RF models transferred to the SA0 domain performed better than the RF-SA1&2 model, but worse than the RF-SA0 model. Figure 4.5 confirms the previous findings that RF models are prone to overfitting.

4.3.3 Flood maps

In order to convey a visual idea of the resulting flood maps, Figure 4.6 compares the water depth as predicted by the different models to the water depths as simulated by the TELEMAC-2D model for region SA0 and for a precipitation depth of 100 mm (Figures S2 and S3 in the supporting information show the flood maps for 50 and 140 mm precipitation depths). Apparently, all models could identify topographic depressions and predict flood water within them. The U-Net - SA0 model underestimates the water depth as shown in Figure 4.6b. Figure 4.6c and d show the predicted water depth from the best performance U-Net - SA1&2 model and the transferred model (U-Net - SA1&2 $\rightarrow$ SA0) using 10 % of the training data of SA0 (including only 40 and 120 mm precipitation depths) to train the weights in the transferred model. The transferred model outperformed both U-Net-SA0 and U-Net-SA1&2. It predicted the most identical inundation extent as the TELEMAC-2D model. Finally, Figure 4.6e, f and g show the predicted water depth from the RF - SA0, RF - SA1&2 and RF - SA1&2 $\rightarrow$ SA0 models respectively. The RF - SA0 model memorised the training data as shown in Figure 4.4 and thus predicted the water depth accurately while the RF - SA1&2 model could not predict the flood water inside the topographical depressions correctly and the RF - SA1&2 $\rightarrow$ SA0 model overestimated the water depth.

4.3.4 Feature importance

Figure 4.7 shows the NSE for SA1 and SA2 for the first three rounds in the predictive feature forward selection process for the best performance model U-Net-SA1&2 (other indices were computed but not shown here since the results regarding feature importance did not change). We stopped after three rounds because the process was computationally expensive and we aimed to obtain just the most important topographical predictive features. These were TWI, SDepth, roughness and altitude. TWI showed the best performance in the first round for both SA1 and SA2, while a model trained with TWI and SDepth was superior to other models in round two. Finally, training a model with TWI, SDepth and altitude outperformed the other models in round three.

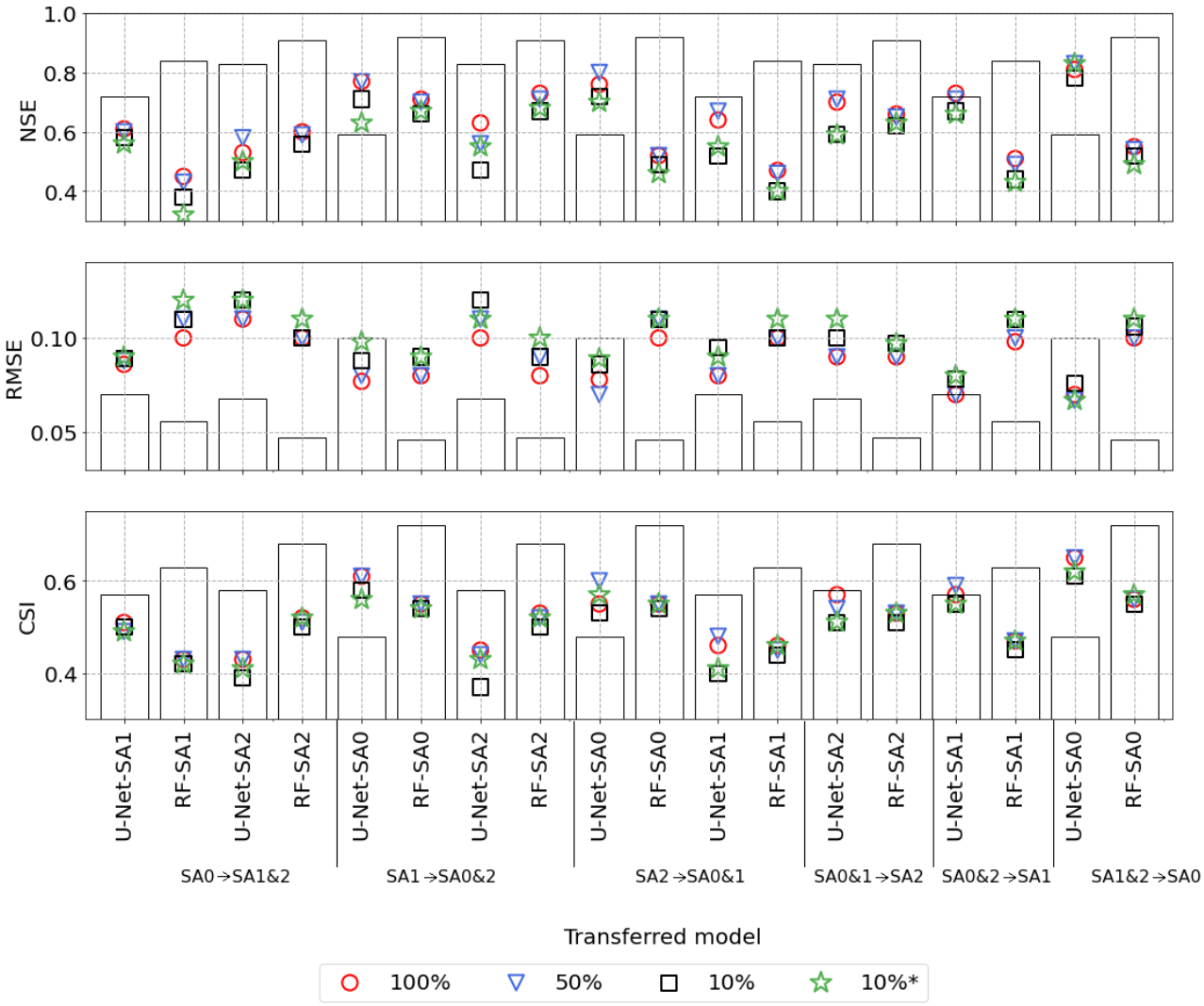


Fig. 4.5 Evaluation of transfer learning: The colored markers represent the performance indices for transferred models with different percentages of data from the domain where the model has been transferred to. For example, SA0→SA1&2 refers to a model pre-trained on SA0, and then transferred (re-trained) on SA1 and SA2. The bars show the performance indices for the models trained exclusively on the transfer target domains. 10%* denotes that the training data from the transferred domain was generated using only two precipitation events (40 and 120 mm). The transferred U-Net-SA1&2→SA0 (pre-trained model SA1&2 and transfer target SA0) models outperformed the U-Net-SA0 model but the RF-SA0 model was superior to the transferred RF-SA1&2→SA0 models for all used percentages of new training data from SA0.

While the gained knowledge in round three by adding altitude and roughness was the same for SA1, adding roughness reduced the model performance in SA2. It is explainable that roughness influenced the models' prediction because buildings were defined in the input dataset by having high roughness values. The precipitation depth was added as a predictive feature to all the trained models but not included in Figure 4.7 because the main objective was to estimate the most important topographical predictive features. In contrast to (Löwe et al. 2021, Seleem et al. 2022), aspect was not among the most

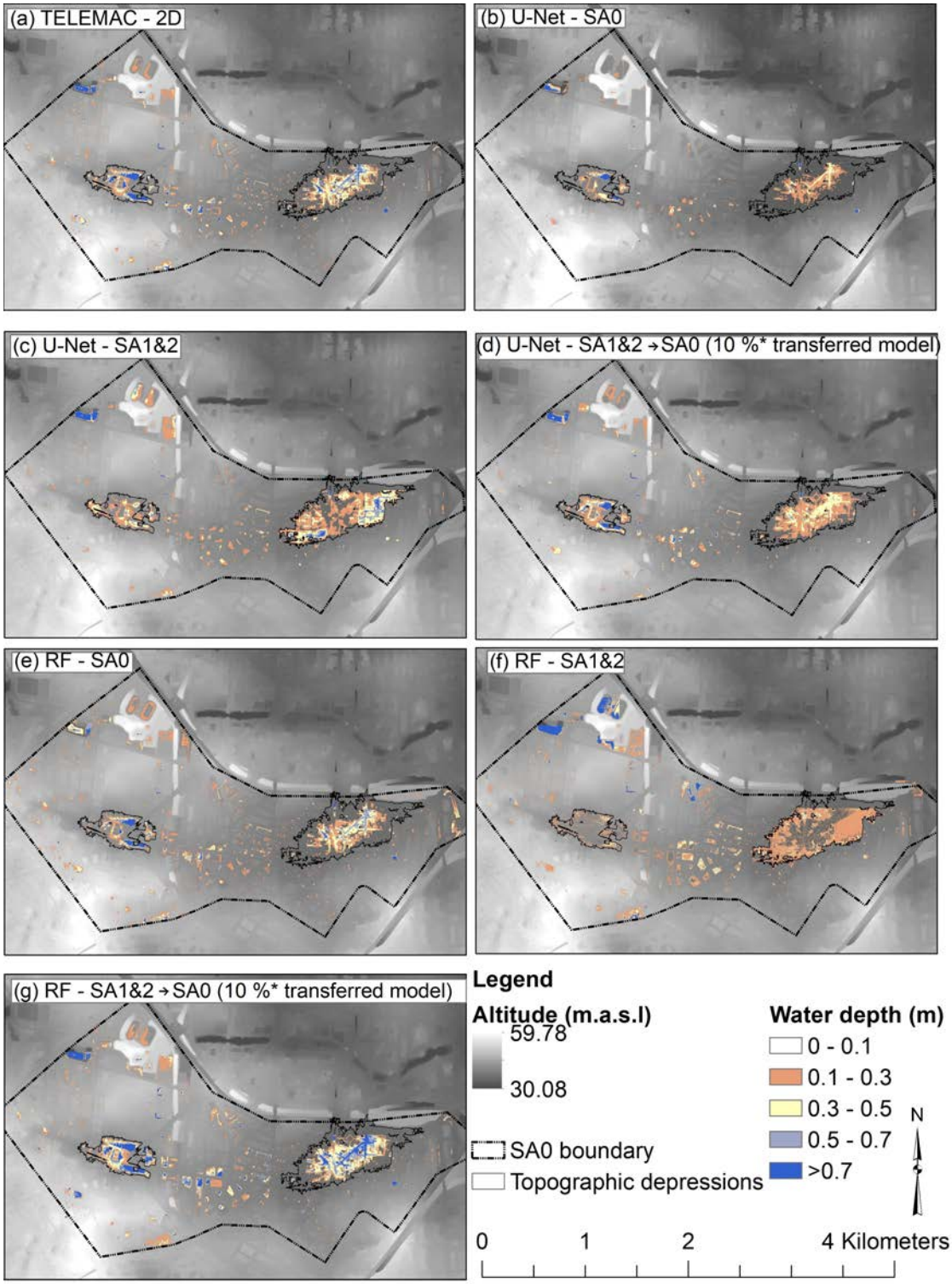


Fig. 4.6 Comparison of water depths from different models and TELEMAC-2D model for a 100 mm precipitation event for SA0. The figure highlights the boundary of two topographic depressions within SA0 where runoff accumulates. The altitude is shown in the background.

important features.

Figure 4.8 shows the feature importance for the RF-SA1&2 model. SDepth, altitude and CN were the most important predictive features. In contrast to U-Net models, TWI was not among the most important predictive features for the RF models. The estimated best predictive features from the U-Net and RF models were not the same but the results agree with the findings in the literature that TWI (Jalayer et al. 2014, Seleem et al. 2021, Bentivoglio et al. 2022), SDepth (Zhang & Pan 2014, Seleem et al. 2021) and altitude (Zhang & Pan 2014, Seleem et al. 2021, 2022) are indicators for urban flood-prone areas.

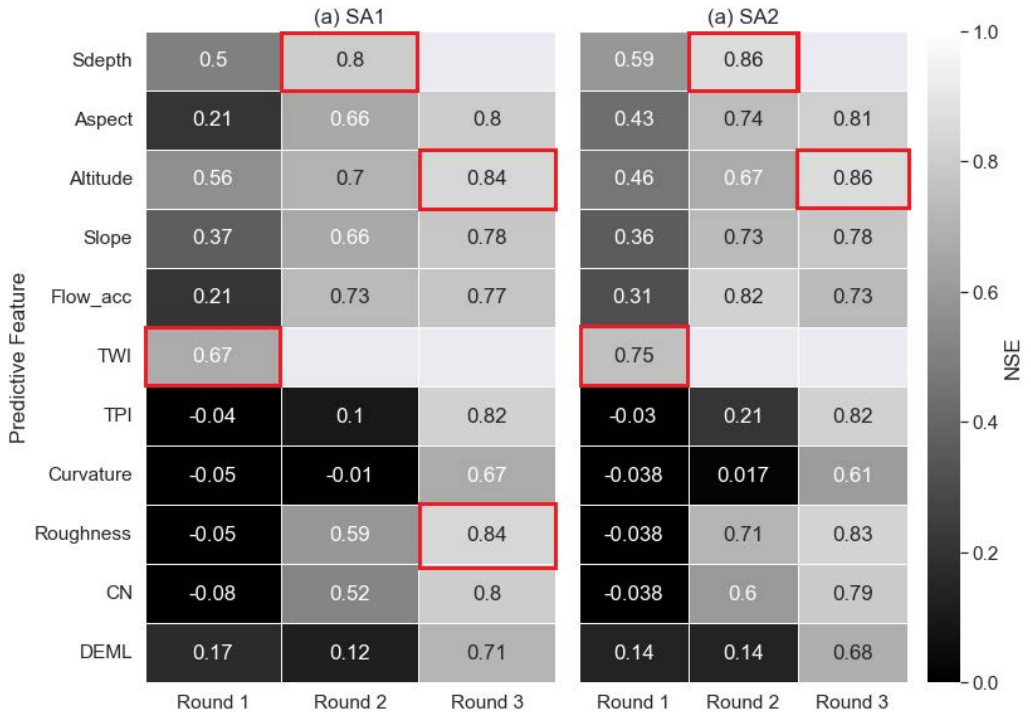


Fig. 4.7 NSE values for SA1 (a) and SA2 (b) for the models trained in the forward selection process for the best performance training data combination (U-Net - SA1&2). The best-performing model in every round is marked in red.

4.4 Conclusions

This study developed and tested CNN models (based on the U-Net architecture) and RF models to emulate the output of a 2-D hydrodynamic model (TELEMAC-2D) for three selected areas within the city of Berlin. We trained the data-driven surrogate models to map topographic, land cover and precipitation variables to flood water depths as obtained from 2D hydrodynamic model simulations. The evaluation of model performance was designed to assess the transferability of trained models to areas outside the training domain. It is worth mentioning that the accuracy of the predicted flood maps by a data-driven model highly depends on the accuracy of the used hydrodynamic model

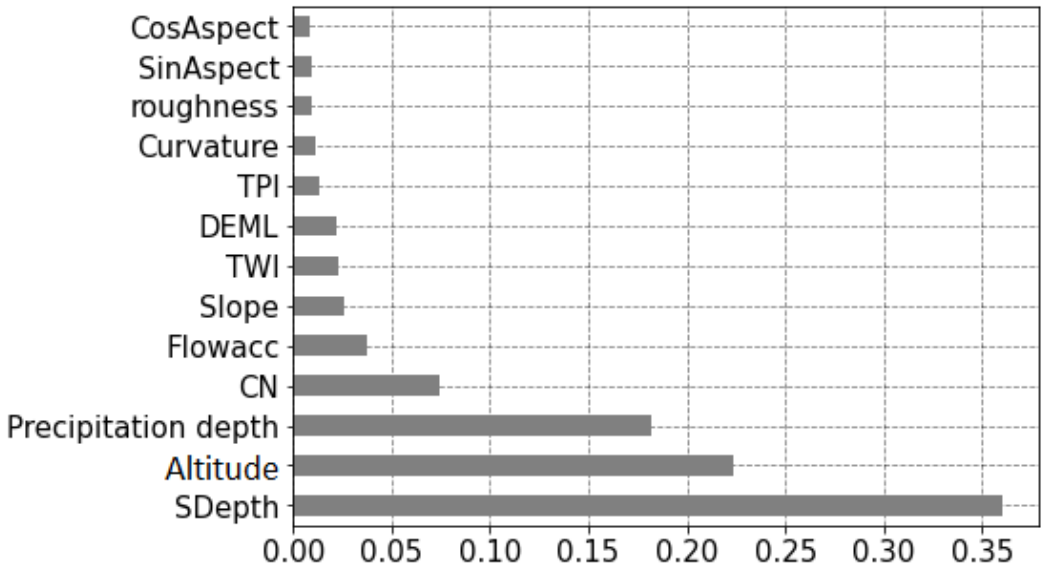


Fig. 4.8 Predictive feature importance for RF-SA1&2 model.

simulations to train the model. While urban area lacks monitoring devices, crowd-sourced data and fine-resolution satellite images could be helpful tools to validate the hydrodynamic models.

Both U-Net and RF models were skilful in predicting water depth within the training domain (minimum NSE=0.6). Contrary to the hypothesis that deep learning algorithms were superior to shallow machine learning algorithms (Bentivoglio et al. 2022), the results suggested that the RF models outperformed the U-Net models for predictions within the training domain. However, we found that the high performance of RF models was largely owed to overfitting: outside of the training domains, RF models exhibited a substantial performance loss for all considered metrics (NSE, RMSE, and CSI). Although some RF models showed better performance outside the training domain (as shown in the supporting information), this study aimed to evaluate the model transferability, not to optimize the model to generalize. For the CNN models, the loss of performance was also considerable, but clearly less pronounced than for the RF models. We hence conclude that the potential of CNN models to generalize beyond the training domain is significantly higher than for RF models. Further research requires testing the data-driven model's transferability further in environments with different characteristics (particularly with cities in more mountainous environments).

Furthermore, we found that the CNN models' ability to generalize and hence to be transferred beyond the training domain could be boosted by transfer learning: by providing only a small fraction of training data from a target domain, transfer learning improved the performance of some pre-trained CNN models in a way it outperformed a CNN that was trained from scratch with the full amount of training data from that domain. This outcome clearly distinguishes deep learning models such as CNN from

shallow models such as RF which could not benefit from transfer learning in a similar fashion. Transfer learning hence provides a promising perspective to efficiently use additional training data to adjust deep learning models to specific target areas or to provide an additional level of generalization, at a minimum computational expense.

Analyzing the results showed that the depth of a depression (SDepth) is a strong predictive feature for both the U-Net and RF models. SDepth, altitude and CN were the most influencing topographical predictors for the RF model while TWI, SDepth,roughness and altitude were the most influencing topographical predictive features for the U-Net model. This is in contrast to Löwe et al. (2021) and Seleem et al. (2022) who found the aspect to be the most important predictive feature for flood hazard and susceptibility mapping using CNN. We hence suggest a detailed future study to systematically explore the suitability of different topographical predictive features for data-driven models of urban flood hazard.

Altogether, this study confirms that deep learning could be a skilful tool for upscaling flood hazard maps in urban environments. Given the excessive costs of providing complete high-resolution 2-D hydrodynamic model coverage, deep learning, namely CNN, has shown the ability to learn transferable knowledge of simulated inundation patterns. This puts into perspective previous study results by Seleem et al. (2022) that highlighted the performance of random forest models – which we now found less able to generalize. Given the apparent potential of CNN for generalization, however, it is all the more important to collect training and testing data from many and diverse regions in order to capitalize on this learning capability. This could be a community effort, and the basis for future benchmarking experiments that move beyond the boundaries of isolated cities. In order to start this process, we provided the output of the 2-D hydrodynamic simulations along with this paper.

Code and Data Availability

The predictive features and water depth from the TELEMAC 2D model simulations are available at `https://doi.org/10.5281/zenodo.7516408` (last access: 10 January 2023); the source code for the models are provided through a GitHub repository `https://github.com/omarseleem92/Urban_flooding.git`.

Competing interests The authors declare that no competing interests are present

Acknowledgements

This research was funded by the Deutscher Akademischer Austauschdienst (DAAD). We acknowledge the support of the Deutsche Forschungsgemeinschaft and Open Access Publishing Fund of the University of Potsdam.

CHAPTER 5

Conclusions

5.1 Main findings and conclusions

Urban pluvial flooding poses a growing threat to cities worldwide. Hence, accurate urban pluvial flood hazard maps are required. Hydrodynamic models can effectively predict flood variables such as floodwater depth and velocity. However, they are computationally expensive and not practical for fast predictions in urban watersheds. This thesis aims to investigate and evaluate alternative methods for mapping urban pluvial flooding. The following summarizes and discusses the main findings and conclusions of this doctoral thesis.

Simplified topography-based methods

This section addresses the following research question:

- RQ1: Can simplified topographic-based methods mimic the two-dimensional hydrodynamic models to simulate urban pluvial flooding?

The first study (Chapter 2) evaluated two simplified methods that have been suggested by recent studies to identify areas prone to urban pluvial flooding: the fill-spill-merge (FSM) and the topographic wetness index (TWI). The simulation results of the TELEMAC-2D model for two case studies in Berlin were used as a reference for the evaluation. The TWI method has the advantage of requiring minimal effort to create a TWI map from a digital elevation model (DEM). However, the TWI method requires to be calibrated to the output of a hydrodynamic model for each precipitation depth of interest. This requirement necessitates the task of setting up and applying the hydrodynamic model. The results showed that the estimated values of the TWI threshold (τ) were similar for the two case study areas, indicating the potential for transferring threshold values between different regions of application. However, the analysis also revealed several limitations of the TWI method. The TWI threshold τ could only be estimated across the examined rainfall depths (from 30 to 150 mm with 10 mm incremental steps) by defining an ”inundated area” as an area with a water depth higher than 0 m. By

defining the inundation area as an area with a water depth exceeding 0.1 m, τ could only be estimated for rainfall depths greater than 90 mm. This can be attributed to the fact that runoff tends to accumulate in topographic depressions within urban watersheds and the TWI method was not able to identify depressions as urban pluvial flood-prone areas.

The results pointed out that the FSM method is clearly superior to the TWI method in identifying urban pluvial flood-prone areas. The FSM method has an advantage over the TWI method of being able to predict the floodwater depth, while the TWI method prediction is limited to identifying flood-prone areas. While the FSM method is much faster than the 2D hydrodynamic models (about 600 times), it requires effort to set up the required model chain. Additionally, its performance may decline with extremely heavy rainfall when flooding begins to spread beyond topographic depressions. Finally, the results showed that the performance of both methods tended to improve with increasing the rainfall depth.

Data-driven models performance and transferability

This section addresses the following research questions:

- RQ2.1: Does deep learning outperform traditional machine learning algorithms to map urban flood susceptibility?
- RQ2.3: How transferable are data-driven models to map urban flood susceptibility?
- RQ3.1: How transferable are data-driven models to predict urban pluvial floodwater depth?

Chapters 3 and 4 compared the performance of deep learning and traditional machine learning algorithms such as random forest (RF) and support vector machine (SVM) to map flood susceptibility and predict floodwater depth. I evaluated the models' performance in the training domain and their transferability in space. In Chapter 4, I investigated the models' ability to boost their performance outside the training domain using transfer learning techniques.

Previous studies claimed that deep learning models are superior to traditional machine learning algorithms (Zhao et al. 2020). However, this study disputes this claim. The results from Chapters 3 and 4 contradict each other. Specifically, the results in Chapter 3 pointed out that the RF models outperformed their comparative models within and outside the training domain for mapping urban flood susceptibility. In Chapter 4, the RF models outperformed deep learning models again in predicting floodwater depth within the training domain. However, the high performance of RF models was mainly due to overfitting when evaluating their performance outside the training domains.

The fact that the RF models were superior to their comparative models in the second study (Chapter 3) aligns with the findings of recent studies such as Grinsztajn et al. (2022). These studies found that tree-based models are still state-of-the-art on medium-sized datasets (around 10,000 samples). Deep learning models tend to smooth out the target function on the training data (Grinsztajn et al. 2022) and also tend to bias towards less complex functions (Rahaman et al. 2019). In contrast, RF models learn piece-wise constant functions and do not have this bias. It is worth noting that while traditional machine learning models may be the better choice in the specific case of flood susceptibility mapping due to limited data availability, deep learning models have shown success in other areas where large datasets are available, such as precipitation nowcasting (Ehsani et al. 2021). In conclusion, the superiority of either model is problem-specific and dependent on the amount of data available.

Breiman (2001) claimed that RF models do not overfit. This claim is correct when there is no significant difference in the distribution of training and testing datasets (Gyori et al. 2022). Altitude, distance to the road (DTRoad) and distance to the storm drainage system (DTDrainage) were the most influencing predictive features for the best-performance RF model (at 2 m spatial resolution) in Chapter 3. Figure 5.1 shows that these predictive features have the same distribution within the training domain. Contrary to altitude and DTRoad, there is a difference in the distribution of DTDrainage within and outside the training domain. This can explain the decline in the model's performance for predictions outside the training domain (the area under a receiver operating characteristic curve (AUC) dropped from 1 on the training dataset to 0.92 on the dataset outside the training domain). Finally, the poor performance of the RF models outside the training domains in Chapter 4 can be attributed to the difference in the study areas (SA) characteristics as shown in Figure 5.2. CNN models which included data from SA1 in the training dataset could better generalize to other study areas. SA1 has deeper topographic depressions compared to the other study areas, as illustrated in Figure 5.2, this might have aided the CNN models in recognizing the relationship between the depth of topographic depressions and floodwater depth and hence could better generalize.

Transfer learning techniques

This section addresses the following research question:

- RQ3.2: Can transfer learning techniques improve the model performance outside the training domain and thus help to overcome the issue of limited training data?

A transferred model is a model that has been trained on one task or a dataset (called a pre-trained model) and then adapted for a different task or dataset. This approach is particularly useful when limited data are available for training a model from scratch.

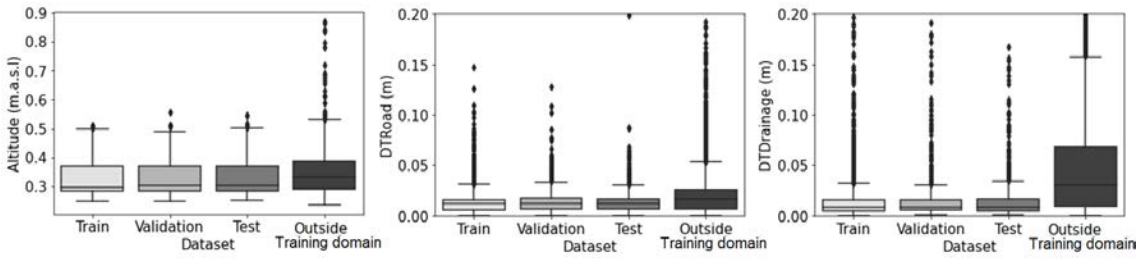


Fig. 5.1 Characteristics of the most influencing predictive features on the best performing RF model to map urban flood susceptibility (Chapter 3), i.e., the Y-axis shows the normalized values between 0 and 1.

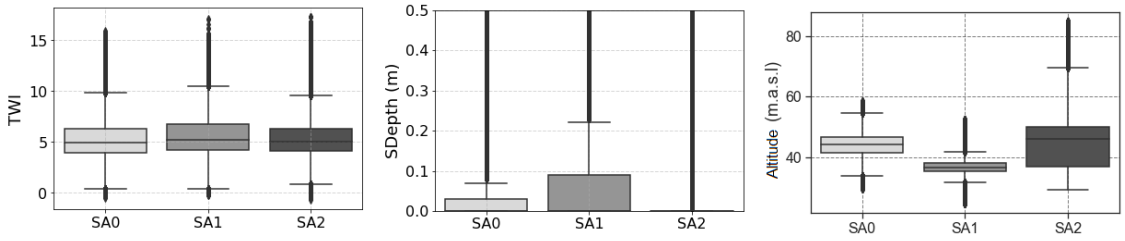


Fig. 5.2 Characteristics of the three study areas (SA) in Chapter 4. Topographic wetness index (TWI) and topographic depression depth (SDepth) were the most influencing features for the convolutional neural network (CNN) models while SDepth and altitude were the most important for RF models.

In Chapter 4, transfer learning techniques were applied to both convolutional neural networks (CNN) and random forest (RF) models to enhance their performance for predicting floodwater depth outside the training domain. The results indicated that deep learning models benefit more from transfer learning techniques than traditional machine learning models. Moreover, a transferred deep learning model that was trained using a small fraction of the training dataset beat a deep learning model that had been trained from scratch with the complete training dataset from the target domain. This shows that a pre-trained deep-learning model can be easily adapted to a specific target area with minimal computational cost.

Spatial resolution

This section addresses the research question:

- RQ2.2: Does the spatial resolution of the predictive features impact the data-driven models' prediction?

In the second paper (Chapter 3), the impact of spatial resolution of the predictive features on the data-driven models' performance was evaluated. The CNN models could only recognize topographic depressions as a flood-susceptible area at the 5 and 2 m spatial resolution. However, this fine spatial resolution was computationally expensive and

time-consuming to map flood susceptibility using CNN. Furthermore, the best performance models (RF at 5 and 2 m spatial resolution) could recognize the streets in the topographic depression as the flood-susceptible area. The relative loss in performance from inside to outside the training domain was particularly high for models at 30 m spatial resolution. The results highlighted the importance of using fine spatial resolution for predictions in a complex urban environment.

Predictive features

This section addresses the research question:

- RQ2.4: Which predictive features are most useful for urban pluvial flood mapping?

In Chapters 3 and 4, I used three methods - Shapley values (Lundberg & Lee 2017), forward feature selection and random forest feature importance (Pedregosa et al. 2011) - to assess the importance of the predictive features for all implemented models. Shapley values are particularly useful because they can explain the complex data-driven models that are often considered a black box. The analysis in Chapter 3 revealed that the importance of the predictive features could vary depending on the used spatial resolution. Point-based models identified flood-susceptible areas based on altitude, while image-based models were strongly influenced by the aspect values. Aspect values indicate the direction of the slope and hence the flow direction. They pinpoint where the flow direction changes and hence the locations where the terrain has a concave shape and where floodwater thus accumulates. In Chapter 4, the CNN model found a strong relationship between topographic wetness index (TWI), topographic depression depth (SDepth) and floodwater depth, while RF models were most influenced by SDepth, altitude and curve number.

Predictive features denoting changes in elevation, such as altitude in Chapter 3 and topographic depression depth and altitude in Chapter 4, were the most influencing predictive features for the RF models. On the other hand, the CNN models were mostly influenced by the aspect to map urban pluvial flood susceptibility (Chapter 3), but the aspect was not among the most three influencing predictive features to predict urban floodwater depth (Chapter 4). The importance of the predictive features can vary depending on the used dataset and model. It is worth mentioning that using different network architectures (LeNet 5 in Chapter 3 and U-Net in Chapter 4) may have also influenced the importance of the predictive features.

5.2 Outlook

This thesis demonstrated that data-driven models can be a useful tool for mapping urban pluvial flooding and highlighted the advantages and limitations of topographic-based

methods. The availability of data such as high-resolution digital elevation models, land use, soil type and precipitation is increasing, which makes data-driven models a valuable tool for urban pluvial flood hazards mapping. This thesis has made some progress, but further research is needed for a comprehensive understanding of data-driven models' capabilities and limitations in flood hazard mapping. Further studies should focus on understanding how these models can be used to improve urban pluvial flood hazard mapping.

Topographical based methods to predict urban flooding

Topographical-based methods mainly require a digital elevation model (DEM) and rainfall data to identify urban pluvial flood-prone areas. The availability of these data has been increasing, reducing the limitations that were more acute in the past due to data unavailability. While urban pluvial floods could occur anywhere within an urban watershed, even in areas without flooding history, hydrodynamic models can only be applied to small areas (a few square kilometres) using fine spatial resolution (less than 5 m). The Fill-Spill-Merge (FSM) method, for example, can predict urban pluvial floodwater depth 600 times faster than hydrodynamic models. However, there is currently no open-source software tool to automate this method. Further research should also investigate these methods in other urban environments with different characteristics, particularly in cities located in more mountainous regions. To support these efforts, future research should focus on developing more efficient and reproducible software tools, allowing for more informative and extensive benchmarking experiments.

Data-driven models for flood modelling

Previous studies focused on using shallow machine learning algorithms (point-based) to map urban flooding while recent studies claimed that deep learning models represented in convolutional neural networks (CNN) (image-based) outperform the point-based models. This thesis focused on investigating the transferability of data-driven models because it is the major challenge that they face. Future research needs to be conducted to address the following:

- While random forest (RF) models were superior to other models to map urban flood susceptibility in Chapter 3, they overfitted the training dataset in Chapter 4 and could not generalize outside the training domain. Chapters 3 and 4 emphasize the importance of conducting benchmark studies to determine which models are most effective for different types of flood maps.
- A simple CNN architecture (LeNet-5) was used to map urban flood susceptibility. Future studies are recommended to investigate deeper network architectures.

- Data-driven model's transferability in space is still a new rising topic for flood mapping. This thesis evaluated the models' transferability within the same city to map urban flood susceptibility and predict urban floodwater depth in Chapters 3 and 4 respectively. Future research requires testing transferability further in environments with different characteristics (cities in more mountainous environments).

- Data-driven models' ability to make accurate predictions strongly depends on the used predictive features to train the model. This thesis showed that the importance of the predictive features varies with the implemented models. While point-based models were influenced by predictive features representing the change in elevation, the importance of the predictive features for the CNN varied between Chapters 3 and 4. Therefore, further studies are needed to investigate which predictive features are relevant for urban pluvial flood mapping.

- This thesis used transfer learning techniques to improve the data-driven models' performance outside the training domains in Chapter 4. While flood inventories and flood hazard maps showing floodwater depth are not often available, transfer learning techniques can be a useful tool to overcome this problem. Therefore, future research is necessary to assess their applicability across different regions.

Finally, this thesis shows that the data-driven model has the potential to generalize to map urban pluvial floods. It is crucial to collect training and testing data from a variety of regions to fully take advantage of this learning capability. This could be a collaborative effort and could serve as the foundation for future benchmarking experiments that extend beyond individual cities.

Towards open science

I would like to express my deep gratitude to the open science community. This research would not have been possible without the support, open data and tools from the community. In support of reproducibility, transparency and open science in flood modelling, the scripts with hands-on tutorials and data used in this thesis are made available in publicly available GitHub repositories (https://github.com/omarseleem92?tab=repositories) and the work done in this thesis is explained in a series of publicly available articles (https://medium.com/@omar.abdwahab92).

REFERENCES

Adiba, A., Hajji, H. & Maatouk, M. (2019), Transfer learning and u-net for buildings segmentation, *in* 'Proceedings of the New Challenges in Data Sciences: Acts of the Second Conference of the Moroccan Classification Society', pp. 1–6.

Al-Juaidi, A. E., Nassar, A. M. & Al-Juaidi, O. E. (2018), 'Evaluation of flood susceptibility mapping using logistic regression and gis conditioning factors', *Arabian Journal of Geosciences* **11**(24), 1–10.

Arabameri, A., Rezaei, K., Cerdà, A., Conoscenti, C. & Kalantari, Z. (2019), 'A comparison of statistical methods and multi-criteria decision making to map flood hazard susceptibility in northern iran', *Science of the Total Environment* **660**, 443–458.

Arrighi, C. & Campo, L. (2019), 'Effects of digital terrain model uncertainties on high-resolution urban flood damage assessment', *Journal of Flood Risk Management* **12**(S2), e12530.

Ashley, R. M., Balmforth, D. J., Saul, A. J. & Blanskby, J. (2005), 'Flooding in the future–predicting climate change, risks and responses in urban areas', *Water Science and Technology* **52**(5), 265–273.

ATKIS (2020), 'Digitale geländemodelle – atkis dgm', `http://www.stadtentwicklung.berlin.de/geoinformation/landesvermessung/atkis/de/dgm.shtml`.

Avand, M., Kuriqi, A., Khazaei, M. & Ghorbanzadeh, O. (2021), 'Dem resolution effects on machine learning performance for flood probability mapping', *Journal of Hydro-environment Research* .

Balstrøm, T. & Crawford, D. (2018), 'Arc-malstrøm: A 1d hydrologic screening method for stormwater assessments based on geometric networks', *Computers & Geosciences* **116**, 64–73.

Barco, J., Wong, K. M. & Stenstrom, M. K. (2008), 'Automatic calibration of the us epa swmm model for a large urban catchment', *Journal of Hydraulic Engineering* **134**(4), 466–474.

Bates, B., Kundzewicz, Z. & Wu, S. (2008), *Climate change and water*, Intergovernmental Panel on Climate Change Secretariat.

Bates, P. (2022), 'Uneven burden of urban flooding', *Nature Sustainability* pp. 1–2.

Bedient, P. B., Huber, W. C., Vieux, B. E. et al. (2008), *Hydrology and floodplain analysis*, Vol. 816, Prentice Hall Upper Saddle River, NJ.

Bentivoglio, R., Isufi, E., Jonkman, S. N. & Taormina, R. (2022), 'Deep learning methods for flood mapping: A review of existing applications and future research directions', *Hydrology and Earth System Sciences Discussions* pp. 1–50.

Berghäuser, L., Schoppa, L., Ulrich, J., Dillenardt, L., Jurado, O. E., Passow, C., Mohor, G. S., Seleem, O., Petrow, T. & Thieken, A. H. (2021), 'Starkregen in berlin: Meteorologische ereignisrekonstruktion und betroffenenbefragung'.

Bermúdez, M., Ntegeka, V., Wolfs, V. & Willems, P. (2018), 'Development and comparison of two fast surrogate models for urban pluvial flood simulations', *Water Resources Management* **32**(8), 2801–2815.

Bhola, P. K., Nair, B. B., Leandro, J., Rao, S. N. & Disse, M. (2019), 'Flood inundation forecasts using validation data generated with the assistance of computer vision', *Journal of Hydroinformatics* **21**(2), 240–256.

Biau, G. & Scornet, E. (2016), 'A random forest guided tour', *Test* **25**(2), 197–227.

Bisht, D. S., Chatterjee, C., Kalakoti, S., Upadhyay, P., Sahoo, M. & Panda, A. (2016), 'Modeling urban floods and drainage using swmm and mike urban: a case study', *Natural Hazards* **84**(2), 749–776.

Borga, M., ANAGNOSTOU, E. N., BLOSCHL, G. & CREUTIN, J. D. (2010), 'Flash floods: observations and analysis of hydrometeorological controls', *Journal of hydrology (Amsterdam)* **394**(1-2).

Breiman, L. (2001), 'Random forests', *Machine learning* **45**(1), 5–32.

Bronstert, A., Agarwal, A., Boessenkool, B., Crisologo, I., Fischer, M., Heistermann, M., Köhn-Reich, L., López-Tarazón, J. A., Moran, T., Ozturk, U. et al. (2018), 'Forensic hydro-meteorological analysis of an extreme flash flood: The 2016-05-29 event in braunsbach, sw germany', *Science of the total environment* **630**, 977–991.

Bui, D. T., Hoang, N.-D., Martínez-Álvarez, F., Ngo, P.-T. T., Hoa, P. V., Pham, T. D., Samui, P. & Costache, R. (2020), 'A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area', *Science of The Total Environment* **701**, 134413.

Bulti, D. T. & Abebe, B. G. (2020), 'A review of flood modeling methods for urban pluvial flood application', *Modeling earth systems and environment* **6**(3), 1293–1302.

Bung, D., Oertel, M., Schlenkhoff, A. & Schlurmann, T. (2011), 'Flash flood awareness and prevention in germany', *Early Warning for Flash Floods* pp. 34–40.

Burrough, P. A., McDonnell, R. A. & Lloyd, C. D. (2015), *Principles of geographical information systems*, Oxford university press.

Caldas-Alvarez, A., Augenstein, M., Ayzel, G., Barfus, K., Cherian, R., Dillenardt, L., Fauer, F., Feldmann, H., Heistermann, M., Karwat, A. et al. (2022), 'Meteorological, impact and climate perspectives of the 29 june 2017 heavy precipitation event in the berlin metropolitan area', *Natural Hazards and Earth System Sciences Discussions* pp. 1–39.

Chang, C.-C. & Lin, C.-J. (2011), 'Libsvm: a library for support vector machines', *ACM transactions on intelligent systems and technology (TIST)* **2**(3), 1–27.

Chapi, K., Singh, V. P., Shirzadi, A., Shahabi, H., Bui, D. T., Pham, B. T. & Khosravi, K. (2017), 'A novel hybrid artificial intelligence approach for flood susceptibility assessment', *Environmental modelling & software* **95**, 229–245.

Chen, W., Li, Y., Xue, W., Shahabi, H., Li, S., Hong, H., Wang, X., Bian, H., Zhang, S., Pradhan, B. et al. (2020), 'Modeling flood susceptibility using data-driven approaches of naïve bayes tree, alternating decision tree, and random forest methods', *Science of The Total Environment* **701**, 134979.

Cherqui, F., Belmeziti, A., Granger, D., Sourdril, A. & Le Gauffre, P. (2015), 'Assessing urban potential flooding risk and identifying effective risk-reduction measures', *Science of the Total Environment* **514**, 418–425.

Chollet, F. et al. (2015), 'Keras', `https://keras.io`.

Choubin, B., Moradi, E., Golshan, M., Adamowski, J., Sajedi-Hosseini, F. & Mosavi, A. (2019), 'An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines', *Science of the Total Environment* **651**, 2087–2096.

Cortes, C. & Vapnik, V. (1995), 'Support vector machine', *Machine learning* **20**(3), 273–297.

Costabile, P., Costanzo, C. & Macchione, F. (2017), 'Performances and limitations of the diffusive approximation of the 2-d shallow water equations for flood simulation in urban and rural areas', *Applied Numerical Mathematics* **116**, 141–156.

Costache, R., Arabameri, A., Elkhrachy, I., Ghorbanzadeh, O. & Pham, Q. B. (2021), 'Detection of areas prone to flood risk using state-of-the-art machine learning models', *Geomatics, Natural Hazards and Risk* **12**(1), 1488–1507.

Coulthard, T. & Frostick, L. (2010), 'The hull floods of 2007: implications for the governance and management of urban drainage systems', *Journal of Flood Risk Management* **3**(3), 223–231.

Cronshey, R. (1986), Urban hydrology for small watersheds, Technical report, US Dept. of Agriculture, Soil Conservation Service, Engineering Division.

De Risi, R., Jalayer, F. & De Paola, F. (2015), 'Meso-scale hazard zoning of potentially flood prone areas', *Journal of Hydrology* **527**, 316–325.

De Risi, R., Jalayer, F., De Paola, F. & Lindley, S. (2018), 'Delineation of flooding risk hotspots based on digital elevation model, calculated and historical flooding extents: the case of ouagadougou', *Stochastic Environmental Research and Risk Assessment* **32**(6), 1545–1559.

Di Salvo, C., Ciotoli, G., Pennica, F. & Cavinato, G. P. (2017), 'Pluvial flood hazard in the city of rome (italy)', *Journal of Maps* **13**(2), 545–553.

Di Salvo, C., Pennica, F., Ciotoli, G. & Cavinato, G. P. (2018), 'A gis-based procedure for preliminary mapping of pluvial flood risk at metropolitan scale', *Environmental modelling & software* **107**, 64–84.

Dillenardt, L., Hudson, P. & Thieken, A. H. (2022), 'Urban pluvial flood adaptation: Results of a household survey across four german municipalities', *Journal of Flood Risk Management* **15**(3), e12748.

Ehsani, M. R., Zarei, A., Gupta, H. V., Barnard, K. & Behrangi, A. (2021), 'Nowcasting-nets: Deep neural network structures for precipitation nowcasting using imerg', *arXiv preprint arXiv:2108.06868* .

Falconer, R. & De Hemptinne, F. (2009), 'Ewa expert meeting on pluvial flooding in europe-final report', *European Water Association* .

Falconer, R. H., Cobby, D., Smyth, P., Astle, G., Dent, J. & Golding, B. (2009), 'Pluvial flooding: new approaches in flood warning, mapping and risk management', *Journal of Flood Risk Management* **2**(3), 198–208.

Fan, Y., Ao, T., Yu, H., Huang, G. & Li, X. (2017), 'A coupled 1d-2d hydrodynamic model for urban flood inundation', *Advances in Meteorology* **2017**.

French, R. H. & French, R. H. (1985), *Open-channel hydraulics*, McGraw-Hill New York.

Fritsch, K., Assmann, A. & Tyrna, B. (2016), Long-term experiences with pluvial flood risk management, *in* 'E3S Web of Conferences', Vol. 7, EDP Sciences, p. 04017.

Galland, J.-C., Goutal, N. & Hervouet, J.-M. (1991), 'Telemac: A new numerical model for solving shallow water equations', *Advances in water resources* **14**(3), 138–148.

Grant, T. D. & Wischik, D. J. (2020), Finding patterns as the path from input to output, *in* 'On the path to AI', Springer, pp. 41–48.

Grinsztajn, L., Oyallon, E. & Varoquaux, G. (2022), 'Why do tree-based models still outperform deep learning on tabular data?', *arXiv preprint arXiv:2207.08815* .

Guerreiro, S. B., Glenis, V., Dawson, R. J. & Kilsby, C. (2017), 'Pluvial flooding in european cities—a continental approach to urban flood modelling', *Water* **9**(4), 296.

Guo, Z., Leitão, J. P., Simões, N. E. & Moosavi, V. (2021), 'Data-driven flood emulation: Speeding up urban flood predictions by deep convolutional neural networks', *Journal of Flood Risk Management* **14**(1), e12684.

Guo, Z., Moosavi, V. & Leitão, J. P. (2022), 'Data-driven rapid flood prediction mapping with catchment generalizability', *Journal of Hydrology* **609**, 127726.

Gyori, N. G., Palombo, M., Clark, C. A., Zhang, H. & Alexander, D. C. (2022), 'Training data distribution significantly impacts the estimation of tissue microstructure with machine learning', *Magnetic resonance in medicine* **87**(2), 932–947.

Haklay, M. & Weber, P. (2008), 'Openstreetmap: User-generated street maps', *IEEE Pervasive computing* **7**(4), 12–18.

Hammond, M. J., Chen, A. S., Djordjević, S., Butler, D. & Mark, O. (2015), 'Urban flood impact assessment: A state-of-the-art review', *Urban Water Journal* **12**(1), 14–29.

Hervouet, J.-M. (2007), *Hydrodynamics of free surface flows: modelling with the finite element method*, John Wiley & Sons.

Hofmann, J. & Schüttrumpf, H. (2019), 'Risk-based early warning system for pluvial flash floods: Approaches and foundations', *Geosciences* **9**(3), 127.

Hofmann, J. & Schüttrumpf, H. (2021), 'floodgan: Using deep adversarial learning to predict pluvial flooding in real time', *Water* **13**(16), 2255.

Hong, H., Tsangaratos, P., Ilia, I., Liu, J., Zhu, A.-X. & Chen, W. (2018), 'Application of fuzzy weight of evidence and data mining techniques in construction of flood susceptibility map of poyang county, china', *Science of the total environment* **625**, 575–588.

Hou, J., Zhou, N., Chen, G., Huang, M. & Bai, G. (2021), 'Rapid forecasting of urban flood inundation using multiple machine learning models', *Natural Hazards* **108**(2), 2335–2356.

Houston, D., Werrity, A., Bassett, D., Geddes, A., Hoolachan, A. & McMillan, M. (2011), 'Pluvial (rain-related) flooding in urban areas: the invisible hazard'.

Hsu, K.-l., Gupta, H. V. & Sorooshian, S. (1995), 'Artificial neural network modeling of the rainfall-runoff process', *Water resources research* **31**(10), 2517–2530.

Huang, H., Chen, X., Wang, X., Wang, X. & Liu, L. (2019), 'A depression-based index to represent topographic control in urban pluvial flooding', *Water* **11**(10), 2115.

Hunter, N. M., Bates, P. D., Horritt, M. S., De Roo, A. & Werner, M. G. (2005), 'Utility of different data types for calibrating flood inundation models within a glue framework', *Hydrology and Earth System Sciences* **9**(4), 412–430.

Ioffe, S. & Szegedy, C. (2015), Batch normalization: Accelerating deep network training by reducing internal covariate shift, *in* 'International conference on machine learning', PMLR, pp. 448–456.

Jaafari, A., Najafi, A., Rezaeian, J. & Sattarian, A. (2015), 'Modeling erosion and sediment delivery from unpaved roads in the north mountainous forest of iran', *GEM-International Journal on Geomathematics* **6**(2), 343–356.

Jalayer, F., De Risi, R., De Paola, F., Giugni, M., Manfredi, G., Gasparini, P., Topa, M. E., Yonas, N., Yeshitela, K., Nebebe, A. et al. (2014), 'Probabilistic gis-based method for delineation of urban flooding risk hotspots', *Natural hazards* **73**(2), 975–1001.

Jaynes, E. T. (2003), *Probability theory: The logic of science*, Cambridge university press.

Jenks, G. F. (1967), 'The data model concept in statistical mapping', *International yearbook of cartography* **7**, 186–190.

Kaiser, M. H. E. (2021), Data-driven modeling of areas prone to heavy rain-induced floods, PhD thesis, Technische Universität München.

Karamouz, M., Hosseinpour, A. & Nazif, S. (2011), 'Improvement of urban drainage system performance under climate change impact: Case study', *Journal of Hydrologic Engineering* **16**(5), 395–412.

Kelleher, C. & McPhillips, L. (2020), 'Exploring the application of topographic indices in urban areas as indicators of pluvial flooding locations', *Hydrological Processes* **34**(3), 780–794.

Kenue, B. (2021), 'Blue kenue: software tool for hydraulic modellers', `http://nrc.canada.ca/en/research-development/products-services/software-applications/blue-kenuetm-software-tool-hydraulic-modellers`.

Khosravi, K., Shahabi, H., Pham, B. T., Adamowski, J., Shirzadi, A., Pradhan, B., Dou, J., Ly, H.-B., Gróf, G., Ho, H. L. et al. (2019), 'A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods', *Journal of Hydrology* **573**, 311–323.

Kingma, D. P. & Ba, J. (2014), 'Adam: A method for stochastic optimization', *arXiv preprint arXiv:1412.6980* .

Kirkby, M. (1975), 'Hydrograph modelling strategies, 69–90', *Peel R, Chisholm M, Haggert P, Processes in Physical and Human Geography, Heineman, London* .

Koç, G., Natho, S. & Thieken, A. H. (2021), 'Estimating direct economic impacts of severe flood events in turkey (2015–2020)', *International Journal of Disaster Risk Reduction* **58**, 102222.

Komolafe, A., Herath, S. & Avtar, R. (2018), 'Sensitivity of flood damage estimation to spatial resolution', *Journal of flood risk management* **11**, S370–S381.

Kottmeier, C., Biegert, C. & Corsmeier, U. (2007), 'Effects of urban land use on surface temperature in berlin: Case study', *Journal of urban planning and development* **133**(2), 128–137.

Kreklow, J., Tetzlaff, B., Kuhnt, G. & Burkhard, B. (2019), 'A rainfall data intercomparison dataset of radklim, radolan, and rain gauge data for germany', *Data* **4**(3), 118.

Kumar, S., Agarwal, A., Villuri, V. G. K., Pasupuleti, S., Kumar, D., Kaushal, D. R., Gosain, A. K., Bronstert, A. & Sivakumar, B. (2021), 'Constructed wetland management in urban catchments for mitigating floods', *Stochastic Environmental Research and Risk Assessment* pp. 1–20.

Kundzewicz, Z. W., Kanae, S., Seneviratne, S. I., Handmer, J., Nicholls, N., Peduzzi, P., Mechler, R., Bouwer, L. M., Arnell, N., Mach, K. et al. (2014), 'Flood risk and climate change: global and regional perspectives', *Hydrological Sciences Journal* **59**(1), 1–28.

Leal, M., Boavida-Portugal, I., Fragoso, M. & Ramos, C. (2019), 'How much does an extreme rainfall event cost? material damage and relationships between insurance, rainfall, land cover and urban flooding', *Hydrological Sciences Journal* **64**(6), 673–689.

Lee, S., Kim, J.-C., Jung, H.-S., Lee, M. J. & Lee, S. (2017), 'Spatial prediction of flood susceptibility using random-forest and boosted-tree models in seoul metropolitan city, korea', *Geomatics, Natural Hazards and Risk* **8**(2), 1185–1203.

Lei, X., Chen, W., Panahi, M., Falah, F., Rahmati, O., Uuemaa, E., Kalantari, Z., Ferreira, C. S. S., Rezaie, F., Tiefenbacher, J. P. et al. (2021), 'Urban flood modeling using deep-learning approaches in seoul, south korea', *Journal of Hydrology* **601**, 126684.

Li, Z., Liu, J., Mei, C., Shao, W., Wang, H. & Yan, D. (2019), 'Comparative analysis of building representations in telemac-2d for flood inundation in idealized urban districts', *Water* **11**(9), 1840.

Liu, L., Liu, Y., Wang, X., Yu, D., Liu, K., Huang, H. & Hu, G. (2015), 'Developing an effective 2-d urban flood inundation model for city emergency management based on cellular automata', *Natural hazards and earth system sciences* **15**(3), 381–391.

Löwe, R., Böhm, J., Jensen, D. G., Leandro, J. & Rasmussen, S. H. (2021), 'U-flood–topographic deep learning for predicting urban pluvial flood water depth', *Journal of Hydrology* **603**, 126898.

Lundberg, S. M. & Lee, S.-I. (2017), A unified approach to interpreting model predictions, *in* 'Proceedings of the 31st international conference on neural information processing systems', pp. 4768–4777.

Maas, A. L., Hannun, A. Y., Ng, A. Y. et al. (2013), Rectifier nonlinearities improve neural network acoustic models, *in* 'Proc. icml', Vol. 30, Citeseer, p. 3.

Maksimović, Č., Prodanović, D., Boonya-Aroonnet, S., Leitao, J. P., Djordjević, S. & Allitt, R. (2009), 'Overland flow and pathway analysis for modelling of urban pluvial flooding', *Journal of Hydraulic Research* **47**(4), 512–523.

Manfreda, S. & Samela, C. (2019), 'A digital elevation model based method for a rapid estimation of flood inundation depth', *Journal of Flood Risk Management* **12**, e12541.

Martel, J.-L., Mailhot, A. & Brissette, F. (2020), 'Global and regional projected changes in 100-yr subdaily, daily, and multiday precipitation extremes estimated from three large ensembles of climate simulations', *Journal of Climate* **33**(3), 1089–1103.

Merz, B., Kreibich, H., Schwarze, R. & Thieken, A. (2010), 'Review article" assessment of economic flood damage"', *Natural Hazards and Earth System Sciences* **10**(8), 1697–1724.

Merz, B. & Thieken, A. H. (2005), 'Separating natural and epistemic uncertainty in flood frequency analysis', *Journal of Hydrology* **309**(1-4), 114–132.

Miikkulainen, R., Liang, J., Meyerson, E., Rawal, A., Fink, D., Francon, O., Raju, B., Shahrzad, H., Navruzyan, A., Duffy, N. et al. (2019), Evolving deep neural networks, *in* 'Artificial intelligence in the age of neural networks and brain computing', Elsevier, pp. 293–312.

Myhre, G., Alterskjær, K., Stjern, C. W., Hodnebrog, Ø., Marelle, L., Samset, B. H., Sillmann, J., Schaller, N., Fischer, E., Schulz, M. et al. (2019), 'Frequency of extreme precipitation increases extensively with event rareness under global warming', *Scientific reports* **9**(1), 16063.

Nash, J. E. & Sutcliffe, J. V. (1970), 'River flow forecasting through conceptual models part i—a discussion of principles', *Journal of hydrology* **10**(3), 282–290.

Neal, J. C., Bates, P. D., Fewtrell, T. J., Hunter, N. M., Wilson, M. D. & Horritt, M. S. (2009), 'Distributed whole city water level measurements from the carlisle 2005 urban flood event and comparison with hydraulic model simulations', *Journal of Hydrology* **368**(1-4), 42–55.

Néelz, S. & Pender, G. (2013), 'Benchmarking the latest generation of 2d hydraulic modelling packages', *Environment Agency: Bristol, UK* .

Néelz, S., Pender, G. et al. (2010), 'Benchmarking of 2d hydraulic modelling packages'.

Nofal, O. M. & van de Lindt, J. W. (2020*a*), 'Probabilistic flood loss assessment at the community scale: Case study of 2016 flooding in lumberton, north carolina', *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering* **6**(2), 05020001.

Nofal, O. M. & Van De Lindt, J. W. (2020*b*), 'Understanding flood risk in the context of community resilience modeling for the built environment: Research needs and trends', *Sustainable and Resilient Infrastructure* pp. 1–17.

Notti, D., Giordan, D., Caló, F., Pepe, A., Zucca, F. & Galve, J. P. (2018), 'Potential and limitations of open satellite data for flood mapping', *Remote sensing* **10**(11), 1673.

O'Neill, E., Brereton, F., Shahumyan, H. & Clinch, J. P. (2016), 'The impact of perceived flood exposure on flood-risk perception: The role of distance', *Risk Analysis* **36**(11), 2158–2186.

Papaioannou, G., Efstratiadis, A., Vasiliades, L., Loukas, A., Papalexiou, S. M., Koukouvinos, A., Tsoukalas, I. & Kossieris, P. (2018), 'An operational method for flood directive implementation in ungauged urban areas', *Hydrology* **5**(2), 24.

Patel, D. P., Ramirez, J. A., Srivastava, P. K., Bray, M. & Han, D. (2017), 'Assessment of flood inundation mapping of surat city by coupled 1d/2d hydrodynamic modeling: a case application of the new hec-ras 5', *Natural Hazards* **89**(1), 93–130.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. et al. (2011), 'Scikit-learn: Machine learning in python', *the Journal of machine Learning research* **12**, 2825–2830.

Peel, M. C., Finlayson, B. L. & McMahon, T. A. (2007), 'Updated world map of the köppen-geiger climate classification', *Hydrology and earth system sciences* **11**(5), 1633–1644.

Penning-Rowsell, E., Johnson, C., Tunstall, S., Tapsell, S., Morris, J., Chatterton, J. & Green, C. (2005), 'The benefits of flood and coastal risk management: a handbook of assessment techniques', *ISBN 1904750516* .

Petroselli, A. (2012), 'Lidar data and hydrological applications at the basin scale', *GIScience & Remote Sensing* **49**(1), 139–162.

Pfeifer, S., Bülow, K., Gobiet, A., Hänsler, A., Mudelsee, M., Otto, J., Rechid, D., Teichmann, C. & Jacob, D. (2015), 'Robustness of ensemble climate projections analyzed with climate signal maps: seasonal and extreme precipitation for germany', *Atmosphere* **6**(5), 677–698.

Pörtner, H., Roberts, D., Poloczanska, E., Mintenbeck, K., Tignor, M., Alegría, A., Craig, M., Langsdorf, S., Löschke, S., Möller, V. et al. (2022), 'Ipcc, 2022: Summary for policymakers', *Climate Change* .

Prechelt, L. (1998), Early stopping-but when?, *in* 'Neural Networks: Tricks of the trade', Springer, pp. 55–69.

Probst, P., Wright, M. N. & Boulesteix, A.-L. (2019), 'Hyperparameters and tuning strategies for random forest', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **9**(3), e1301.

Prokic, M. N., Savić, S. & Pavić, D. (2019), 'Pluvial flooding in urban areas across the european continent', *Geographica Pannonica* **23**(4).

Qin, D., Chen, Z., Averyt, K., Miller, H., Solomon, S., Manning, M., Marquis, M. & Tignor, M. (2007), 'Ipcc, 2007: Summary for policymakers'.

R., H. J. A. (2019), User manual of opensource software telemac-2d, Report, EDF-R&D, www.opentelemac.org. V8P1.

Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., Bengio, Y. & Courville, A. (2019), On the spectral bias of neural networks, *in* 'International Conference on Machine Learning', PMLR, pp. 5301–5310.

Rahmati, O., Pourghasemi, H. R. & Zeinivand, H. (2016), 'Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the golastan province, iran', *Geocarto International* **31**(1), 42–70.

Rangari, V. A., Veerendra Gopi, K., Umamahesh, N. & Patel, A. K. (2018), Simulation of urban drainage system using disaggregated rainfall data, *in* 'Hydrologic modeling', Springer, pp. 123–133.

Raschka, S. (2015), *Python machine learning*, Packt publishing ltd.

Regmi, A. D., Devkota, K. C., Yoshida, K., Pradhan, B., Pourghasemi, H. R., Kumamoto, T. & Akgun, A. (2014), 'Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in central nepal himalaya', *Arabian Journal of Geosciences* **7**(2), 725–742.

Rejith, R., Anirudhan, S. & Sundararajan, M. (2019), 'Delineation of groundwater potential zones in hard rock terrain using integrated remote sensing, gis and mcdm techniques: a case study from vamanapuram river basin, kerala, india', *S. Venkatramanan, VM Prasanna, & YS Chung, GIS and Geostatistical Techniques for Groundwater Science* pp. 349–364.

Ronneberger, O., Fischer, P. & Brox, T. (2015), U-net: Convolutional networks for biomedical image segmentation, *in* 'International Conference on Medical image computing and computer-assisted intervention', Springer, pp. 234–241.

Rosenzweig, B. R., McPhillips, L., Chang, H., Cheng, C., Welty, C., Matsler, M., Iwaniec, D. & Davidson, C. I. (2018), 'Pluvial flood risk and opportunities for resilience', *Wiley Interdisciplinary Reviews: Water* **5**(6), e1302.

Ross, C., Prihodko, L., Anchang, J., Kumar, S., Ji, W. & Hanan, N. (2018), 'Hysogs250m, global gridded hydrologic soil groups for curve-number-based runoff modeling, scientific data, 5, 180091'.

Rözer, V., Müller, M., Bubeck, P., Kienzler, S., Thieken, A., Pech, I., Schröter, K., Buchholz, O. & Kreibich, H. (2016), 'Coping with pluvial floods by private households', *Water* **8**(7), 304.

Samela, C., Albano, R., Sole, A. & Manfreda, S. (2018), 'A gis tool for cost-effective delineation of flood-prone areas', *Computers, Environment and Urban Systems* **70**, 43–52.

Samela, C., Persiano, S., Bagli, S., Luzzi, V., Mazzoli, P., Humer, G., Reithofer, A., Essenfelder, A., Amadio, M., Mysiak, J. et al. (2020), 'Safer_rain: a dem-based hierarchical filling-&-spilling algorithm for pluvial flood hazard assessment and mapping across large urban areas', *Water* **12**(6), 1514.

Samela, C., Troy, T. J. & Manfreda, S. (2017), 'Geomorphic classifiers for flood-prone areas delineation for data-scarce environments', *Advances in water resources* **102**, 13–28.

Santurkar, S., Tsipras, D., Ilyas, A. & Madry, A. (2018), 'How does batch normalization help optimization?', *Advances in neural information processing systems* **31**.

Schanze, J. (2018), 'Pluvial flood risk management: an evolving and specific field'.

Scharffenberg, W. & Harris, J. (2008), Hydrologic engineering center hydrologic modeling system, hec-hms: interior flood modeling, *in* 'World Environmental and Water Resources Congress 2008: Ahupua'A', pp. 1–3.

Schmitt, T. G., Thomas, M. & Ettrich, N. (2004), 'Analysis and modeling of flooding in urban drainage systems', *Journal of hydrology* **299**(3-4), 300–311.

Schumann, G. J.-P., Neal, J. C., Mason, D. C. & Bates, P. D. (2011), 'The accuracy of sequential aerial photography and sar data for observing urban flood dynamics, a case study of the uk summer 2007 floods', *Remote Sensing of Environment* **115**(10), 2536–2546.

Seleem, O., Ayzel, G., de Souza, A. C. T., Bronstert, A. & Heistermann, M. (2022), 'Towards urban flood susceptibility mapping using data-driven models in berlin, germany', *Geomatics, Natural Hazards and Risk* **13**(1), 1640–1662.

Seleem, O., Heistermann, M. & Bronstert, A. (2021), 'Efficient hazard assessment for pluvial floods in urban environments: A benchmarking case study for the city of berlin, germany', *Water* **13**(18), 2476.

Seneviratne, S. I., Lüthi, D., Litschi, M. & Schär, C. (2006), 'Land–atmosphere coupling and climate change in europe', *Nature* **443**(7108), 205–209.

Shafapour Tehrany, M., Kumar, L., Neamah Jebur, M. & Shabani, F. (2019), 'Evaluating the application of the statistical index method in flood susceptibility mapping and its comparison with frequency ratio and logistic regression methods', *Geomatics, Natural Hazards and Risk* **10**(1), 79–101.

Shen, D., Wang, J., Cheng, X., Rui, Y. & Ye, S. (2015), 'Integration of 2-d hydraulic model and high-resolution lidar-derived dem for floodplain flow modeling', *Hydrology and Earth System Sciences* **19**(8), 3605–3616.

Singh, P., Sinha, V. S. P., Vijhani, A. & Pahuja, N. (2018), 'Vulnerability assessment of urban road network from urban flood', *International journal of disaster risk reduction* **28**, 237–250.

Skougaard Kaspersen, P., Høegh Ravn, N., Arnbjerg-Nielsen, K., Madsen, H. & Drews, M. (2017), 'Comparison of the impacts of urban development and climate change on exposing european cities to pluvial flooding', *Hydrology and Earth System Sciences* **21**(8), 4131–4147.

Smith, M. B. (2006), 'Comment on'analysis and modeling of flooding in urban drainage systems", *Journal of Hydrology* .

Spekkers, M., Rözer, V., Thieken, A., ten Veldhuis, M.-C. & Kreibich, H. (2017), 'A comparative survey of the impacts of extreme rainfall in two international case studies', *Natural Hazards and Earth System Sciences* **17**(8), 1337–1355.

Srivastava, R. K., Greff, K. & Schmidhuber, J. (2015), 'Training very deep networks', *Advances in neural information processing systems* **28**.

Tabari, H., Asr, N. M. & Willems, P. (2021), 'Developing a framework for attribution analysis of urban pluvial flooding to human-induced climate impacts', *Journal of Hydrology* **598**, 126352.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C. & Liu, C. (2018), A survey on deep transfer learning, *in* 'International conference on artificial neural networks', Springer, pp. 270–279.

Tanaka, T., Kiyohara, K. & Tachikawa, Y. (2020), 'Comparison of fluvial and pluvial flood risk curves in urban cities derived from a large ensemble climate simulation dataset: A case study in nagoya, japan', *Journal of Hydrology* **584**, 124706.

Tayefi, V., Lane, S., Hardy, R. & Yu, D. (2007), 'A comparison of one-and two-dimensional approaches to modelling flood inundation over complex upland floodplains', *Hydrological Processes: An International Journal* **21**(23), 3190–3202.

Tehrany, M. S., Pradhan, B. & Jebur, M. N. (2014), 'Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in gis', *Journal of hydrology* **512**, 332–343.

Tehrany, M. S., Pradhan, B., Mansor, S. & Ahmad, N. (2015), 'Flood susceptibility assessment using gis-based support vector machine model with different kernel types', *Catena* **125**, 91–101.

Ten Veldhuis, J. (2011), 'How the choice of flood damage metrics influences urban flood risk assessment', *Journal of Flood Risk Management* **4**(4), 281–287.

Teng, J., Jakeman, A. J., Vaze, J., Croke, B. F., Dutta, D. & Kim, S. (2017), 'Flood inundation modelling: A review of methods, recent advances and uncertainty analysis', *Environmental modelling & software* **90**, 201–216.

Termeh, S. V. R., Kornejady, A., Pourghasemi, H. R. & Keesstra, S. (2018), 'Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms', *Science of the Total Environment* **615**, 438–451.

Testa, G., Zuccalà, D., Alcrudo, F., Mulet, J. & Soares-Frazão, S. (2007), 'Flash flood flow experiment in a simplified urban district', *Journal of Hydraulic Research* **45**(sup1), 37–44.

Thieken, A. H., Samprogna Mohor, G., Kreibich, H. & Müller, M. (2022), 'Compound inland flood events: different pathways, different impacts and different coping options', *Natural Hazards and Earth System Sciences* **22**(1), 165–185.

Tien Bui, D., Pradhan, B., Lofman, O. & Revhaug, I. (2012), 'Landslide susceptibility assessment in vietnam using support vector machines, decision tree, and naive bayes models', *Mathematical problems in Engineering* **2012**.

Trost, S. G., Zheng, Y. & Wong, W.-K. (2014), 'Machine learning for activity recognition: hip versus wrist data', *Physiological measurement* **35**(11), 2183.

Union, E. (2017), 'Directive 2007/60/ec on the assessment and management of flood risks.', *Journal of the European Union Official, Brussels* .

Vafakhah, M., Mohammad Hasani Loor, S., Pourghasemi, H. & Katebikord, A. (2020), 'Comparing performance of random forest and adaptive neuro-fuzzy inference system data mining models for flood susceptibility mapping', *Arabian Journal of Geosciences* **13**, 1–16.

van Dijk, E., van der Meulen, J., Kluck, J. & Straatman, J. (2014), 'Comparing modelling techniques for analysing urban pluvial flooding', *Water science and technology* **69**(2), 305–311.

Viera, A. J., Garrett, J. M. et al. (2005), 'Understanding interobserver agreement: the kappa statistic', *Fam med* **37**(5), 360–363.

Wang, Y., Fang, Z., Hong, H. & Peng, L. (2020), 'Flood susceptibility mapping using convolutional neural network frameworks', *Journal of Hydrology* **582**, 124482.

Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S. & Bai, X. (2015), 'Flood hazard risk assessment model based on random forest', *Journal of Hydrology* **527**, 1130–1141.

Weiss, A. (2001), Topographic position and landforms analysis, *in* 'Poster presentation, ESRI user conference, San Diego, CA', Vol. 200.

Wilson, J. P. & Gallant, J. C. (2000), 'Secondary topographic attributes', *Terrain analysis: Principles and applications* pp. 87–131.

Winterrath, T., Brendel, C., Hafer, M., Junghänel, T., Klameth, A., Walawender, E., Weigl, E. & Becker, A. (2017), *Erstellung einer radargestützten Niederschlagsklimatologie*.

Yacoub, S. M., Simske, S. J., Lin, X. & Burns, J. (2003), Recognition of emotions in interactive voice response systems., *in* 'Interspeech'.

Yang, T.-H., Chen, Y.-C., Chang, Y.-C., Yang, S.-C. & Ho, J.-Y. (2015), 'Comparison of different grid cell ordering approaches in a simplified inundation model', *Water* **7**(2), 438–454.

Yang, T.-H., Hwang, G.-D., Tsai, C.-C. & Ho, J.-Y. (2016), 'Using rainfall thresholds and ensemble precipitation forecasts to issue and improve urban inundation alerts', *Hydrology and Earth System Sciences* **20**(12), 4731–4745.

Yin, J., Yu, D., Yin, Z., Liu, M. & He, Q. (2016), 'Evaluating the impact and risk of pluvial flash flood on intra-urban road network: A case study in the city center of shanghai, china', *Journal of hydrology* **537**, 138–145.

Zahura, F. T., Goodall, J. L., Sadler, J. M., Shen, Y., Morsy, M. M. & Behl, M. (2020), 'Training machine learning surrogate models from a high-fidelity physics-based model: Application for real-time street-scale flood prediction in an urban coastal community', *Water Resources Research* **56**(10), e2019WR027038.

Zhang, S. & Pan, B. (2014), 'An urban storm-inundation simulation method based on gis', *Journal of hydrology* **517**, 260–268.

Zhang, Y., Hong, Y., Wang, X., Gourley, J. J., Xue, X., Saharia, M., Ni, G., Wang, G., Huang, Y., Chen, S. et al. (2015), 'Hydrometeorological analysis and remote sensing of extremes: Was the july 2012 beijing flood event detectable and predictable by global satellite observing and global weather modeling systems?', *Journal of hydrometeorology* **16**(1), 381–395.

Zhao, G., Pang, B., Xu, Z., Cui, L., Wang, J., Zuo, D. & Peng, D. (2021), 'Improving urban flood susceptibility mapping using transfer learning', *Journal of Hydrology* **602**, 126777.

Zhao, G., Pang, B., Xu, Z., Peng, D. & Zuo, D. (2020), 'Urban flood susceptibility assessment based on convolutional neural networks', *Journal of Hydrology* **590**, 125235.

Zhao, G., Pang, B., Xu, Z., Yue, J. & Tu, T. (2018), 'Mapping flood susceptibility in mountainous areas on a national scale in china', *Science of the Total Environment* **615**, 1133–1142.

Zhou, Q., Leng, G., Su, J. & Ren, Y. (2019), 'Comparison of urbanization and climate change impacts on urban flood volumes: Importance of urban planning and drainage adaptation', *Science of the Total Environment* **658**, 24–33.

Zhou, Q., Mikkelsen, P. S., Halsnæs, K. & Arnbjerg-Nielsen, K. (2012), 'Framework for economic pluvial flood risk assessment considering climate change effects and adaptation benefits', *Journal of Hydrology* **414**, 539–549.